Estimating the Effects of Government Programs with Machine Learning

a new approach with an application to labor supply and child benefits

Filip Premik*
Department of Economics, Monash University
FAME|GRAPE

April 15, 2024

Abstract

This paper proposes a method of evaluating universal government programs in which the lack of appropriate control groups limits the usability of traditional evaluation methods. The method predicts individuals' behavior based on a general form discrete choice model, estimated non-parametrically using the Generalized Random Forest. I apply my framework to estimate the effects of a large-scale child benefit program on the female labor supply. Focusing on labor market flows and aggregating individual counterfactual decisions, I show that the program led to a 2–3 percentage points decrease in labor supply among eligible women, driven by discouragement among job seekers.

Keywords: generalized random forest, large-scale government programs, program evaluation, child benefit programs, female labor supply

Statements

Data availability. This paper uses confidential data from the Labor Force Survey, which can obtained by applying to the Central Statistical Office in Poland.

Funding. The author gratefully acknowledges the support of the Polish National Science Centre (grant #2016/23/N/HS4/03637).

Conflicts of interest. The author reports there are no competing interests to declare.

Supplementary materials. The paper is accompanied by an **Appendix** providing more detailed results and intuitions, and a **code-pack** with programs used in the analysis.

*filip.premik@monash.edu. The author thanks Joanna Tyrowicz without whose support and effort this paper would have never been written, as well as multiple discussants whose comments improved its quality. The author gratefully acknowledges the support of the Polish National Science Centre (grant #2016/23/N/HS4/03637).

1 Introduction

Large-scale government programs have become increasingly prevalent around the world, particularly since the COVID-19 pandemic and subsequent energy price crisis. Broad eligibility implies that an appropriate control group does not exist and thus standard program evaluation methods cannot be utilized. Specifically, the behavior of the ineligible group is rarely a good counterfactual to the behavior of the eligible group in the absence of the program due to significant differences in the decision processes among individuals related to various stages of life-cycle, systematic heterogeneity in unobservable characteristics or incentives motivating their decisions. Given that these untargeted or broadly targeted transfers and subsidies often constitute a significant cost to the taxpayers, understanding the effects of such policies is of paramount interest.

In this paper, I employ a data-driven approach exploiting large amounts of information regarding individuals' socio-economic backgrounds to approximate individuals' decisions as closely as possible. Datasets including such information are increasingly available for researchers evaluating the impacts of large-scale government policies. The counterfactual outcomes necessary to estimate the effects of interests rely on a general form discrete choice model. My framework shares the flexibility of reduced form approaches (by conditioning on potentially a large set of state variables and not requiring functional form assumptions) and benefits from the appealing interpretation of the underlying choice model (which is a typical aspect of structural models) without the necessity to impose restrictive assumptions on expectations or future evolution of state variables, hence it provides a convenient framework to evaluate the effects of universal government programs. The flexibility of my approach allows researchers to study various nuances related to program evaluation including effect heterogeneity, or measuring changes in the importance of particular characteristics in explaining the individual decisions. It also provides a convenient tool to simulate various counterfactual scenarios on the aggregate level.

I use my framework to evaluate the effects of a universal large-scale child benefit program in Poland on the female labor supply. The program – *Family 500 Plus* (henceforth: *P500*, or the *intervention*) – costs approximately 2% GDP yearly and provides transfers to families raising kids. From 2016 on, families raising two or more children receive a monthly nonequivalent transfer of approximately 20% of the median wage per second and any further child. Even though I focus primarily on changes in labor force participation among Polish women in response to the program *P500*, the methodology proposed in this paper generalizes easily to study many other universal or near-universal government programs. The key required feature is an underlying individual decision problem.

I proceed in three steps. First, I derive counterfactual decisions using a flexible choice model, leveraging a rich set of control variables. The model is identified through a set of local moment conditions and estimated using Generalized Random Forest methodology by Athey, Tibshirani, and Wager (2019). Second, the counterfactuals focus on labor market flows, which is appropriate to study decisions as compared to stocks. Individual decisions behind the labor market flow reflect individual optimizations, capturing better the variation driving changes in the labor force participation upon the introduction of the benefits. Third, and final, once individual counterfactuals are obtained, I aggregate labor force participation permitting the comparisons of the actual and counterfactual aggregate paths over time.

The results of my analysis indicate that the program *P500* led to a decrease in female labor force participation, mainly by discouraging the activation of women outside of the labor force. This discouragement was driven by changes in trade-offs and future beliefs. There is also evidence of 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 approximately a 2 percentage point drop in the labor force participation rate among eligible women in two years and over a 3 percentage point drop in four years after the introduction of the program.

The decrease in labor force participation may have affected employers who improved the working conditions facing increased difficulty in maintaining the staff. If this is true, one would expect an implied decrease in outflow rates driven by changes in the economic environment. Such an outflow is confirmed by the decomposition. Hence, the program *P500* impacted the labor supply also in an indirect manner. Removing this effect in a counterfactual labor force participation path counteracts the further propagation of initial shocks resulting in the estimate of a 1.2 percentage point drop in the labor force participation among the eligible women after four years since the introduction of the program. However, due to a large amount of uncertainty, this estimate lacks statistical precision.

My paper speaks to a few strands of the literature. The first two focus on evaluating large-scale government programs using reduced-form and structural approaches respectively. I also relate to papers using machine learning methods in studying labor market outcomes. Lastly, I add to the prior research on *P500* in Poland.

In related studies concerning reduced-form large-scale program evaluations, Schirle (2015); Koebel and Schirle (2016) show that the Canadian Universal Child Care Benefit decreases the labor supply of married women. Baker, Messacar, and Stabile (2021) provide an overview of a few reforms of the Canadian child benefit system showing a reduction in child poverty and no evidence of labor supply response on both extensive and intensive margins. González (2013) investigates a universal child benefit program in Spain and finds a decrease in the 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; Immervoll, Kleven, Kreiner, and Saez, 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 the data-generating process associated with the potential outcomes model. In addition, my results have an appealing interpretation of a micro-founded model.

Another strand of the literature uses structural modeling as a tool to evaluate the impacts of large child support programs. Blundell, Duncan, McCrae, and Meghir (2000) study Working Families' Tax Credit program in the UK. Using 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 the Japanese child benefit system on household wealth accumulation. A fully specified structural model requires several assumptions regarding agent expectations and the 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 the decision problem. Moreover, it uses machine learning techniques that allow to tractably condition women's decisions on a large number of observed state variables.

My paper applies machine learning methods to study labor force participation. In a related setting, Cengiz, Dube, Lindner, and Zentler-Munro (2021) uses similar tools to predict which individuals are likely to be affected by the minimum wage reforms. Sigurdsson (2019) applies 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. A common denominator of these studies is that they focus on recovering causal parameters in the potential outcomes framework. My study differs from these papers by using the machine learning algorithm to estimate a flexible structural choice model and then simulate counterfactual decisions.

My study adds to the discussion concerning the effects of the program *p500*. 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. Results indicate a drop in the labor force supply of roughly 150 thousand women, or approx. 2% of economically active women, which is similar to my findings concerning early-stage program evaluation. The simulations are obtained in the short-run and partial equilibrium, that is they ignore potential changes in the wage structure and working conditions. My approach allows for implicit consideration of these effects. Magda, Kiełczewska, and Brandt (2018) use a difference-in-difference approach to provide an early evaluation of the effects on the female labor supply. Their identifying assumption relies on the short term between the introduction of the program and the measurement of their effects. They find treatment effects implying a 2-3 percentage points drop in the female labor force supply as a result of introducing the program, which again is consistent with my estimates of the initial effects of the program.

The remainder of the paper is structured as follows. Section 2 describes the policy design. Section 3 reviews the data. Section 4 presents the model and explains its use in the program evaluation exercise. Section 5 introduces details of the estimation routine. Empirical results are discussed in section 6. Section 7 measures the effects of the program on aggregate labor force participation. Section 8 concludes.

2 The Program Family 500Plus

The program Family 500Plus 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.

To obtain the benefit, an eligible household is supposed to apply for it at the local administration. Approximately 95% of households with two or more children below 18 are obtaining the benefit.

The program was 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 were 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 was announced and gradually introduced. 2017 and subsequent years belong to the post-intervention period.

3 Data

Data comes from the Labor Force Survey conducted by the 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.

[Figure 1 near here]

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 with two or more children below 18 (shortly: $\geq$2 children) who are surely eligible to receive the benefit at least for one child. I also investigate the behavior of females who do not have children below 18 (shortly: childless) and therefore are not eligible for the benefit. This allows me to keep track of indirect changes that the program may induce on the non-participating individuals through changes in the labor market environment. Moreover, I need a model of untreated behavior to simulate the paths of the aggregated labor force in the last part of my paper.

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 participate given the short panel dimension in my data.

3.1 Labor Market flows

In this paper, I identify the effects of child benefits on labor force participation through changes in the labor market 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 and stock of labor force participation. Differences in their dynamics suggest that the inflows are driven by other economic processes than the outflows. In addition, flows exhibit more variability compared to the levels of labor force participation. All of these factors motivate focusing on the flows in analyzing the program's impacts.

[Figure 2 near here]

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 1. 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. Therefore, even relatively smaller changes in the outflows may translate into significant shocks to the aggregate labor supply.

[Table 1 near here]

Figure 2 presents also 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. In the following analysis, we capture how much of this variation can be attributed to the program *p500*.

3.2 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 the 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 the mother's and father's sources of income and subjective evaluation of their labor market status.

[Table 2 near here]

Table 2 provides a summary of available predictors of women's labor market decisions. In total, I take a set of 379 observed state variables to the estimation. Appendix B.1 provides a detailed description of these variables.

4 The Model

In this section, I present a general discrete choice model of a woman's decision of whether to be a part of the labor force, highlighting potential differences in behavior depending on whether the woman is in or out of the labor force at the moment of making the decision.

4.1 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. I model inflows into and outflows from the labor force separately. In analyzing inflows, $y = 1$ describes a 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.

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

$$v_t(y, s, \varepsilon; G_t) \equiv v_t(y, s, \varepsilon) \tag{1}$$

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:

$$V_t(s, \varepsilon) = \max_{y \in \{0,1\}} \Big\{ v_t(y, s, \varepsilon) \Big\} \tag{2}$$

The optimal policy function is:

$$y_t(s, \varepsilon) = \mathbf{1} \Big[v_t(1, s, \varepsilon) \geq v_t(0, s, \varepsilon) \Big] \tag{3}$$

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:

$$\mathrm{P}[y = 1 | s, t] = \int\limits_{\varepsilon : v_t(1,s,\varepsilon) - v_t(0,s,\varepsilon) \geq 0} \mathrm{d}F_t(\varepsilon | s) \quad \equiv \quad \varrho_t(s) \tag{4}$$

$\varrho_t(s)$ is a conditional choice probability and describes woman's decision rule given s. The estimated decision rules serve to generate woman's counterfactual choices at various t and s. I use these counterfactuals to evaluate the impacts of the program *P500*.

4.2 Decomposing Differences in Choice Probability

Evaluating the effects of the child benefit program is essentially asking how did a woman change her labor force participation in 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:

$$\begin{aligned}\mathbb{E}_\varepsilon\big[y_1(s_1,\varepsilon_1) - y_0(s_0,\varepsilon_0)\big] &= \varrho_1(s_1) - \varrho_0(s_0) = \\ &= \underbrace{\varrho_1(s_0) - \varrho_0(s_0)}_{\beta(s_0)} + \underbrace{\varrho_1(s_1) - \varrho_1(s_0)}_{\gamma_1(s_1,s_0)}\end{aligned} \tag{5}$$

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. I refer to the parameter $\beta(s_0)$ as to *decision rule* effect to accentuate its dependence on the evolution of conditional choice probabilities.

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. By adjusting elements of s_1, a woman may increase her probability of receiving a benefit, which in turn affects her labor force participation. For example, consider an ineligible woman with one child. Suppose she derives high utility from staying out of the 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. I label $\gamma_1(s_1, s_0)$ the *composition* effect to emphasize its dependence on the shift in observed characteristics.

4.3 Sample Decomposition and Specification Test

The decomposition given by equation (5) 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:

$$\hat{\xi}(s_1, s_0) \equiv \big(\overline{y}_1(s_1) - \overline{y}_0(s_0)\big) - \big(\hat{\varrho}_1(s_1) - \hat{\varrho}_0(s_0)\big) \tag{6}$$

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. This error refers to the variation in the unobservables. I call it the *residual* parameter.

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

$$\begin{aligned} \overline{y}_1(s_1) - \overline{y}_0(s_0) &= \hat{\varrho}_1(s_1) - \hat{\varrho}_0(s_0) + \hat{\xi}(s_1, s_0) = \\ &= \underbrace{\hat{\beta}(s_0)}_{\text{decision rule}} + \underbrace{\hat{\gamma}_1(s_1, s_0)}_{\text{composition}} + \underbrace{\hat{\xi}(s_1, s_0)}_{\text{residual}} \end{aligned} \tag{7}$$

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 $H_0 : \hat{\xi}(s_1, s_0) = 0$.

4.4 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 a higher level of aggregation. Both equations (5) and (7) can easily be 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 *decision rule*, *composition* and *residual* effects. This approach allows for uncovering more detailed patterns of dynamics in the evolution of conditional choice probabilities and *ipso facto* also the components of its decomposition. This can be helpful for a researcher to understand various channels of program's impacts on woman's decisions. In the empirical part, I estimate quarterly time series of the decomposition to inform the identification of changes in the women's labor supply that are likely attributed to the intervention.

My approach combines the advantages of both structural and reduced-form approaches in estimating the effects of government policy. 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 typically has to be independent from s. In my framework, this independence assumption can be relaxed.

5 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:

$$\mathbb{E}\Big[y - \varrho | t, s\Big] = 0 \tag{8}$$

Estimation based on conditional moment restriction is often subject to the curse of dimensionality, which effectively limits the analysis to very few state variables. 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 A contains an intuitive description of the mechanics behind the GRF estimator.

5.1 Estimating Conditional Choice Probabilities

Motivated by the differences in flows' dynamics presented in the descriptive analysis, I estimate separate decision models of inflows and outflows. In the model of 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 the model of 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 provide a good source of variation relevant to women's choices available in my data.

The empirical strategy relies on uncovering the underlying women'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. 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, but my approach does not preclude joint decisions in the data-generating process. In the empirical part, I condition women's decisions also on the characteristics of other household members in a previous period.

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 (in terms of the size of a single tree or number of trees holding the tree size constant) and resulting computational complexity. With a forest large enough, it is sufficient to estimate one model of female 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 an estimate of $\varrho_t(s)$ 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 200 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. I use the pre-post model estimates to quantify the overall changes in parameters of the decomposition (7) between pre- and post-intervention periods and to study heterogeneous impacts among women with a different socio-economic background. Pre-post models provide also a de-

[1] I use `R` package `grf` developed by Tibshirani, Athey, and Wager (2020).

scription of the most important predictors of the labor force flows and changes in their significance after the intervention, which provides additional insights into how the impacts of the program propagated in women's choices.

Next, I turn to investigate quarterly dynamics in parameters of decomposition (7). 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. 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.

6 Results

In this section, I present the results of estimation. I start with interpreting the parameters of *pre-post* effects and then complement the analysis with the *quarterly* estimates. Next, I come back to the *pre-post* frequency and investigate effect heterogeneity by women's socio-economic status. Lastly, I track changes in the list of important predictors of women's decisions and draw conclusions relating these changes to the impact of the program.

6.1 Pre-Post Effects

I summarize changes in the female labor market flows using the pre-post framework. Components of decomposition (7) 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 (*decision rule* parameter), their observed characteristics (*composition* parameter), and residual factors (*residual* parameter).

Table 3 presents estimated parameters in the pre-post framework. A key driving force affecting changes in the inflows among eligible women is the *decision rule* channel. The pre-intervention population of females with two or more children below 18 decreased their inflow rate in the post-intervention period by 2.22 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.

[Table 3 near here]

Changes in women's observed state variables described by the *composition* channel affected both inflows and outflows. The parameter estimates indicate approximately 1 percentage point increase in the post-intervention period inflows – resulting from changes in their decision rules, and approximately 1 percentage point decrease in post-intervention period outflows among both eligible and ineligible females – resulting from changes in their observed characteristics. Notably, the effects are fairly similar for both eligible and ineligible groups.

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 3) imply that the model explains the data satisfactorily well.

6.2 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 3 presents the time series of estimates.

[Figure 3 near here]

The *pre-post* estimation reveals over 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 3). 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 promptly internalized. The arrival of benefits discouraged labor activation 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. Estimation results reflect this presumption. The *decision rule* 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 *composition* parameters in the decomposition of the inflows show that neither eligible nor ineligible women changed their labor market activation rates due to changes in their observed characteristics. This result suggests that women did not self-select to the program on the inflows margin.

The *decision rule* 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 *decision rule* 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 *decision rule* 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 *decision rule* parameters in the outflow rate changes decomposition at the end of the sample is likely to be a result of the *P500*.

The estimated *composition* 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 *composition* channel, quantitatively stronger. 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 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 *composition* channel.

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.

6.3 Heterogeneous Effects

The elasticity of labor supply with respect to the benefit is likely to vary with observed state variables. This is reflected in the estimated parameters of the model analyzed in subpopulations defined by women's demographics.

The *decision rule* channel consistently leads to stronger 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 who are divorced. 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 inflow 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 relatively 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 who belong to these subpopulations are more likely to opt out of labor force participation regardless of the benefits program.

The *composition* mechanisms in shaping the dynamics in outflows are also heterogeneous. The strongest effect is observed among the youngest women and those who were divorced. This result is in line with economic intuition. These two subpopulations consist of women who are likely to depend on their own income. Additional benefits may enable more eligible women to organize child care and hence decrease their labor market withdrawal rates. In turn, possibly improved conditions in the labor market probably created higher incentives for keeping the job by the ineligible and women in their 20's. In turn, there are no discouragement effects due to changes in their observed characteristics among eligible females who are in the labor force and whose youngest child is below three. Intuitively, mothers of toddlers are in general unlikely to be in the labor force, and the benefits are not expected to change this.

A full set of heterogenous effects is provided in section B.2 of the Appendix.

6.4 Important Predictors

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 section A.1 of the appendix. A detailed summary of the most important predictors is presented in Appendix B.3.

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.

Changes in the split significance measure revealed that the benefits play an important role in shaping eligible females' decisions in the post-intervention period. One of the most important predictors of inflows among the eligible after the introduction of the program *p500*, but not the ineligible, was *family duties as subjective evalution of past labor market status*. *family duties as subjective evaluation of current labor status* or *benefits as the main source of income* observed one of the highest increases in power in predicting outflows. Notably, these variables do not show up as important predictors among ineligible women either in the pre-intervention or in the post-intervention period.

Moreover, two spousal characteristics observe a significant increase in split significance measure in predicting inflows among women with two or more children: *monthly wage* and *declared willingness to work more to earn more*. Also, a household level variable *# individuals in household* also reports a high increase in significance. All of these variables refer to the actual and potential level of income that the spouse (and other household members) contribute 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 are features related to job safety, including *permanent horizon of employment* (equivalent to tenure), and *salaried worker*. This result is consistent with the hypothesis that postulates that employment safety contributes to the decrease in outflows through the *decision rule* channel in 2018-2019.

7 Measuring the Effect on Aggregate Labor Force Participation

In this section, I investigate the impact of the *P500* on the aggregate labor supply by shutting down the variation in parameters of decomposition (7) attributable to the program and simulating counterfactual paths of the labor force using the implied flows for both the eligible and ineligible.

I distinguish three potential channels of how the program *P500* may have impacted the aggregate labor market flows. First, the benefits may have changed the relative profitability of costly job search, discouraging labor market activation 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 *decision rule* 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.

The law of motion of labor force participation y_t is given by:

$$\mathrm{P}[y_t = 1] = \underbrace{\mathrm{P}[y_t = 1 | y_{t-1} = 0]}_{\text{inflow rate}} \cdot \mathrm{P}[y_{t-1} = 0] + \underbrace{\mathrm{P}[y_t = 1 | y_{t-1} = 1]}_{\text{(negative) outflow rate}} \cdot \mathrm{P}[y_{t-1} = 1] \tag{9}$$

The flow rates can be further decomposed into:

$$\mathrm{P}[y_t = 1 | y_{t-1} = y] = \underbrace{\mathrm{P}[y_{t-1} = 1 | y_{t-2} = y]}_{\text{lagged flow rate}} + \underbrace{\mathrm{P}[y_t = 1 | y_{t-1} = y] - \mathrm{P}[y_{t-1} = 1 | y_{t-2} = y]}_{\text{change in flow rate}} \tag{10}$$

for $y \in \{0,1\}$. Given the decomposition (7), the change in flows can be written as:

$$\text{change in flow rate}_{t,t-1} = \beta^{flow}(s_{t-1}) + \gamma_1^{flow}(s_t, s_{t-1}) + \xi^{flow}(s_t, s_{t-1}) \tag{11}$$

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 *decision rule* 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 *composition* 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 *decision rule* 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)).

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 4 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. The latter summarizes the overall effects of the program. The complete time series of projected labor participation rate paths and the simulated flows are shown in Appendix B.4

[Table 4 near here]

The results indicate that in the absence of the program, the labor force participation rate among eligible women would be 1.74-2.044 percentage points higher at the end of 2017. 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 3.25-3.37 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 small t-statistics in table 4. 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 0.472 drop in the labor force participation among the ineligible women to the intervention. This estimate is also statistically insignificant.

8 Conclusion

In this paper, I propose a simple but comprehensive approach to analyze the effects of large-scale government programs and apply the framework to study how a child benefit program impacts women's 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 a 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 the impacts of the program on labor force participation.

My estimates reveal a moderate but statistically significant drop in labor force participation among eligible women as a result of the program, driven mainly by changes in women's decision rules accommodating perceived trade-offs and beliefs that discouraged labor force inflows immediately after the introduction of the program. Changes in flow rates induced by the program are not large quarterly but accumulate over time leading to significant changes in the labor force participation.

The flexibility of my approach may enable researchers to study many other large-scale programs in which it is difficult to define appropriate control groups and hence the identification strategies based on traditional frameworks may seem less appealing. The only trade-off for the researchers is that due to the relative simplicity of my model, an analysis of more elaborative counterfactual scenarios at the aggregate level requires more ad-hoc assumptions about the market and may lead to an increase in estimation uncertainty.

References

Angrist, J. and B. Frandsen (2019). Machine labor. Technical report, National Bureau of Economic Research.

Athey, S., J. Tibshirani, and S. Wager (2019). Generalized random forests. *The Annals of Statistics 47*(2), 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. Technical report, 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*(1), 75–104.

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. Technical report, 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*(3), 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*(516), 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*(1), 49–64.

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*(03), 36–44.

Schirle, T. (2015). The effect of universal child benefits on labour supply. *Canadian Journal of Economics/Revue canadienne d'économique 48*(2), 437–463.

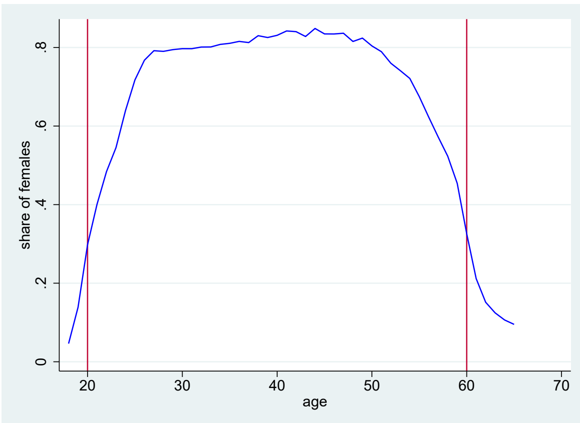
Sigurdsson, J. (2019). Labor supply responses and adjustment frictions: A tax-free year in iceland. *Available at SSRN 3278308*.

Stephens Jr, M. and T. Unayama (2015). Child benefit payments and household wealth accumulation. *The Japanese Economic Review 66*(4), 447–465.

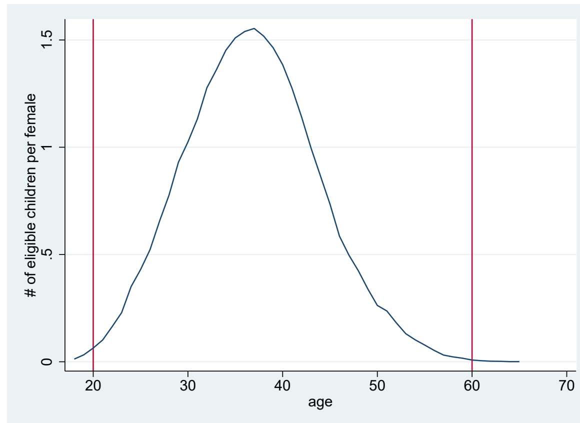
Tibshirani, J., S. Athey, and S. Wager (2020). *grf: Generalized Random Forests*. R package version 1.2.0.

Figures and Tables

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



(a) female labor force participation



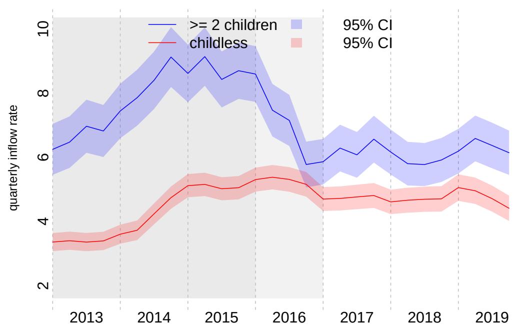
(b) number of children below 18 per female

Table 1: Labor force participation - inflows and outflows.

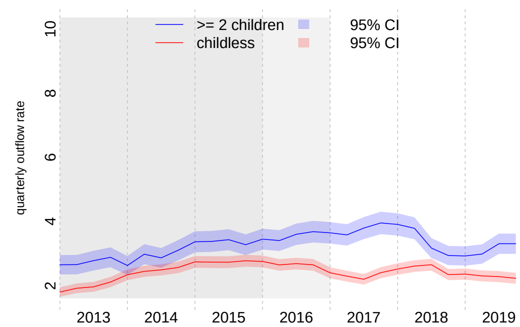
	>=2 children		childless	
	inflows	outflows	inflows	outflows
post-int. (2017-2019)	6.202	3.393	4.648	2.325
pre-int. (2014-2015)	8.901	3.188	4.893	2.669
difference	-2.699	.205	-.245	-.343

Note: Inflows and outflows expressed in percentage points.

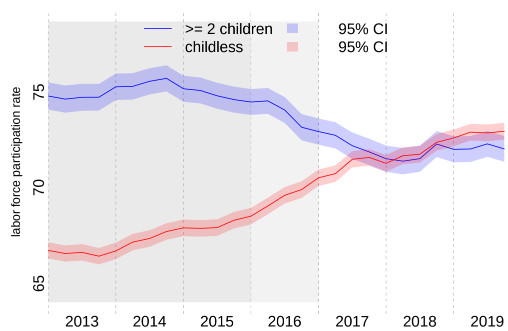
Figure 2: Quarterly labor force indicators by program eligibility.



(a) y_q^0: rate of inflows to labor force.



(b) y_q^1: rate of outflows from labor force.



(c) y_q^1: labor force participation.

Note: All values expressed as percentage points. Dark gray background indicates pre-intervention periods, light gray indicates the transition period.

Table 2: Choice of the Observed State Variables – Summary

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

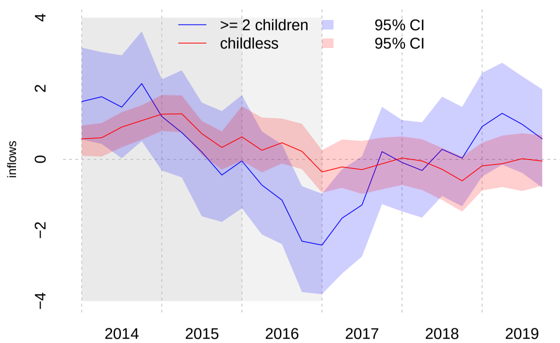
Note: * only a selection of variables in the group is chosen.

Table 3: Flows to the labor force – estimates of pre- and post-intervention differences.

	decision rule	*composition*	*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	-2.22 $(-4.365)^{***}$	-0.16 (-1.832)	-0.319 (-0.653)
childless	0.118 (0.518)	-0.301 $(-3.745)^{***}$	-0.063 (-0.31)
		Outflows from the Labor Force	
>= 2 children	1.165 $(5.391)^{***}$	-1.144 $(-14.974)^{***}$	0.184 (1.042)
childless	0.969 $(7.809)^{***}$	-1.348 $(-25.686)^{***}$	0.035 (0.376)

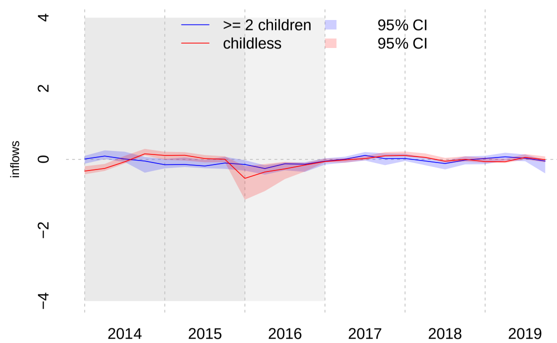
Note: The table presents estimated parameters of the decomposition (7) in pre-post setting. In the inflows part, ϱ denotes the conditional probability of being in the labor force conditionally on being out a quarter before. In the outflows part, ϱ denotes the conditional probability of being out of the 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 200 bootstrap replications presented in the parentheses. Stars denote *** p-val.<0.001, **, p-val.<.005, * p-val.<0.01.

Figure 3: Estimated Parameters of Decomposition (7), Time Series Approach.



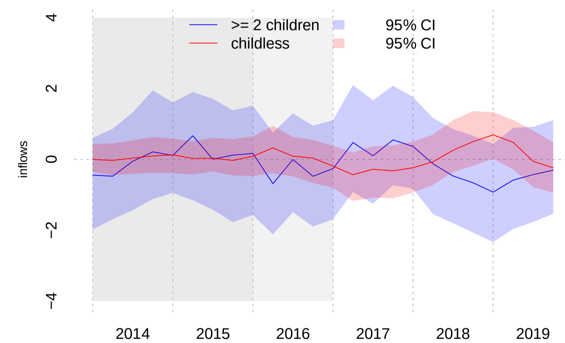
(a) inflows: *decision rule* parameters:

$\hat{\beta}(s_{t-1}) \equiv \hat{\varrho}_t(s_{t-1}) - \hat{\varrho}_{t-1}(s_{t-1})$



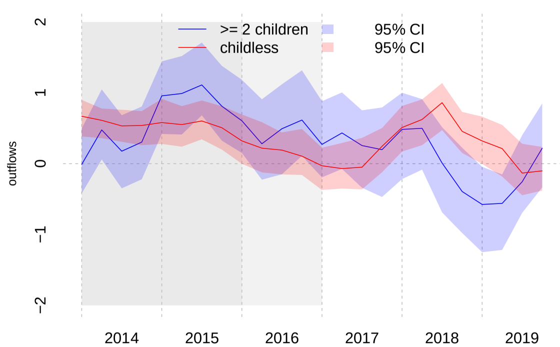
(b) inflows: *composition* parameters:

$\hat{\gamma}_t(s_t, s_{t-1}) \equiv \hat{\varrho}_t(s_t) - \hat{\varrho}_t(s_{t-1})$



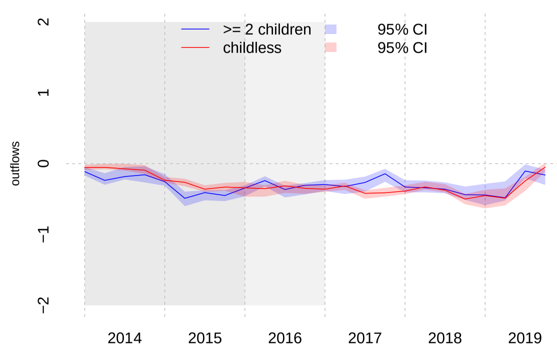
(c) inflows: *idiosyncratic* parameters:

$\hat{\xi}_t(s_t, s_{t-1}) \equiv (\overline{y}_1 - \overline{y}_0) - (\hat{\varrho}_1(s_1) - \hat{\varrho}_0(s_0))$



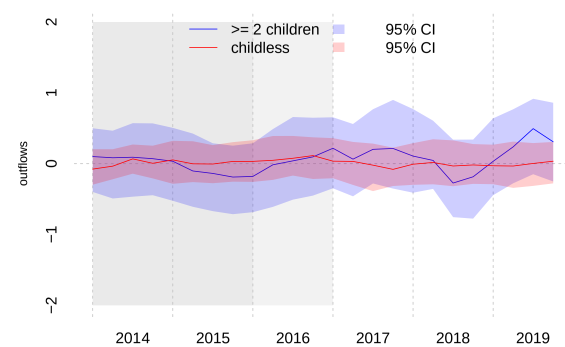
(d) outflows: *decision rule* parameters:

$\hat{\beta}(s_{t-1}) \equiv \hat{\varrho}_t(s_{t-1}) - \hat{\varrho}_{t-1}(s_{t-1})$



(e) outflows: *composition* 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 4: Effects on the labor force participation – simulation results.

	the eligible		the ineligible	
	2017Q4	2019Q4	2017Q4	2019Q4
		Scenario (1)		
labor force participation	−1.74 (−2.762)**	−3.373 (−2.702)**	—	—
inflow rates	−1.797 (−2.408)*	−1.797 (−2.408)*	—	—
		Scenario (2)		
labor force participation	−2.044 (−2.611)**	−3.253 (−2.092)*	—	—
inflow rates	−1.797 (−2.408)*	−1.797 (−2.408)*	—	—
outflow rates	0.065 (0.481)	−0.068 (−0.377)	—	—
		Scenario (3)		
labor force participation	−2.044 (−2.611)**	−1.157 (−0.686)	—	−0.472 (−1.108)
inflow rates	−1.797 (−2.408)*	−1.797 (−2.408)*	—	—
outflow rates	0.065 (0.481)	−1 (−2.754)**	—	0.11 (0.633)

Note: The table presents the results of the 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 200 bootstrap replications are presented in parentheses. Stars denote *** p-val.<0.001, **, p-val.<.005, * p-val.<0.01.

Estimating the Effects of Government Programs with Machine Learning

a new approach with an application to labor supply and child benefits

Appendix

Filip Premik*
Department of Economics, Monash University
FAME|GRAPE

April 15, 2024

*filip.premik@monash.edu.

Appendix A 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 (equation 7 in the paper) 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 the remaining 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 recursively partitioning the dataset. 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[1] 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 the average of the assigned numbers over

[1] See Hastie, Tibshirani, Friedman, and Friedman (2009) for a detailed treatment of tree methods in machine learning.

all trees. Therefore, the forest weights are obtained by averaging neighborhoods produced by different 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:

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

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 makes GRF particularly useful for econometric applications. Based on the random forest algorithm, GRF is designed to deal with high-dimensional datasets and provides an 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.

Appendix A.1 Split Significance Measure

The random forest algorithm provides a simple framework to evaluate the 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 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 the observed state variable.

Appendix B Graphs and Tables

Appendix B.1 List of Observed State Variables.

Table 1: 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

Continued on next page

Table 1 – continued from previous page

code	description		type
	main description	detailed description	
w6_2		temporary because cannot find permanent job	bin
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

Continued on next page

Table 1 – continued from previous page

code	description		type
	main description	detailed description	
w26_2		farmer	bin
w26_3		entrepreneurship	bin
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

Continued on next page

Table 1 – continued from previous page

code	description: main description	description: detailed description	type
s11_3		wants to work less	bin
s11_4		wants better employment conditions	bin
s11_5		want a permanent job	bin
	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

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

Appendix B.2 Heterogenous Impacts

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. I decompose changes in inflows and outflows on subsamples generated by the value of chosen observed state variables. I consider the woman's education level, the size of the city she lives in, her age and marital status, the number of eligible children, and the age of the youngest child.

The estimated parameters of decomposition for eligible and ineligible women are presented in the tables 2 and 3 respectively in the supplementary material. Consistently with the aggregate effects, I observe significant variation in *decision rule* parameters in decompositions of both inflows and outflows among the eligible women and decomposition of outflows among the ineligible women, as well as *composition* effects while considering outflows of both the eligible and the ineligible.

Table 2: Heterogenous Impacts - eligible women.

	inflows		outflows	
	decision rule $\hat{\beta}(s_0) = \hat{\varrho}_1(s_0) - \hat{\varrho}_0(s_0)$	*composition* $\hat{\gamma}_1(s_1, s_0) = \hat{\varrho}_1(s_1) - \hat{\varrho}_1(s_0)$	*decision rule* $\hat{\beta}(s_0) = \hat{\varrho}_1(s_0) - \hat{\varrho}_0(s_0)$	*composition* $\hat{\gamma}_1(s_1, s_0) = \hat{\varrho}_1(s_1) - \hat{\varrho}_1(s_0)$
education				
tertiary	−2.265 (−2.333)*	−0.306 (−2.332)*	0.719 (4.182)***	−0.621 (−13.939)***
secondary	−2.448 (−4.581)***	−0.173 (−2.138)*	1.292 (4.898)***	−0.868 (−9.992)***
vocational or lower	−1.911 (−3.614)***	−0.316 (−3.309)***	1.792 (4.886)***	−1.507 (−11.994)***
size of the city				
>= 100k	−2.832 (−3.723)***	0.024 (0.189)	1.109 (5.484)***	−1.163 (−15.443)***
20-100k	−2.293 (−3.979)***	−0.197 (−1.303)	1.348 (5.79)***	−1.199 (−13.095)***
< 20k	−2.086 (−3.643)***	−0.177 (−1.416)	1.195 (4.508)***	−0.798 (−8.482)***
< 20k	−2.039 (−4.094)***	−0.211 (−2.234)*	1.115 (4.831)***	−1.352 (−14.316)***
womanś age				
20-29	−2.056 (−3.25)**	−0.306 (−3.151)**	2.117 (3.617)***	−2.27 (−10.883)***
30-39	−2.443 (−4.632)***	−0.054 (−0.562)	1.181 (5.402)***	−0.874 (−11.343)***
40-49	−1.749 (−2.989)**	−0.209 (−1.438)	0.774 (4.631)***	−0.835 (−13.481)***
womanś marital status				
married	−2.161 (−4.269)***	−0.179 (−2.014)*	1.105 (5.244)***	−1.081 (−14.736)***
never married	−2.462 (−3.404)***	−0.095 (−0.545)	2.083 (3.831)***	−1.75 (−7.725)***
divorced	−3.205 (−3.138)**	−0.013 (−0.072)	1.601 (3.205)**	−1.985 (−10.937)***
# of eligible children				
2	−2.421 (−4.368)***	−0.207 (−2.139)*	1.15 (5.434)***	−1.119 (−15.246)***
3	−1.856 (−3.773)***	−0.009 (−0.092)	1.129 (4.297)***	−1.055 (−10.685)***
>= 4	−1.475 (−2.672)**	−0.029 (−0.264)	1.691 (3.672)***	−1.888 (−7.818)***
age of the youngest child				
0-3	−1.795 (−2.307)*	0.251 (1.342)	1.191 (2.627)**	0.132 (1.045)
4-6	−2.502 (−3.661)***	−0.263 (−2.41)*	1.37 (3.749)***	−1.073 (−9.769)***
7-12	−2.401 (−4.512)***	−0.133 (−1.573)	1.272 (5.176)***	−1.381 (−13.102)***
13-17	−1.953 (−3.336)***	−0.226 (−1.797)	0.995 (4.923)***	−1.158 (−16.16)***

Note: The table presents estimated parameters of the decomposition in pre-post setting for the eligible, 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 200 bootstrap replications presented in the parentheses. Stars denote *** p-val.<0.001, **, p-val.<.005, * p-val.<0.01.

Table 3: Heterogenous Impacts - ineligible women.

	inflows		outflows	
	decision rule $\hat{\beta}(s_0) = \hat{\varrho}_1(s_0) - \hat{\varrho}_0(s_0)$	*composition* $\hat{\gamma}_1(s_1, s_0) = \hat{\varrho}_1(s_1) - \hat{\varrho}_1(s_0)$	*decision rule* $\hat{\beta}(s_0) = \hat{\varrho}_1(s_0) - \hat{\varrho}_0(s_0)$	*composition* $\hat{\gamma}_1(s_1, s_0) = \hat{\varrho}_1(s_1) - \hat{\varrho}_1(s_0)$
		education		
tertiary	0.28 (0.427)	−0.284 (−2.002)*	0.42 (4.031)***	−0.633 (−19.055)***
secondary	−0.093 (−0.329)	−0.151 (−1.952)	1.09 (6.841)***	−1.369 (−22.229)***
vocational or lower	0.156 (0.766)	−0.252 (−2.79)**	1.584 (7.742)***	−2.101 (−21.276)***
		size of the city		
>= 100k	0.162 (0.599)	−0.301 (−3.384)***	0.74 (6.055)***	−1.279 (−25.271)***
20-100k	0.09 (0.416)	−0.051 (−0.534)	1.183 (8.155)***	−1.562 (−23.053)***
< 20k	0.079 (0.342)	−0.289 (−2.88)**	1.085 (7.636)***	−1.469 (−21.241)***
< 20k	0.114 (0.45)	−0.359 (−4.226)***	1.026 (6.996)***	−1.348 (−22.069)***
		womanś age		
20-29	0.258 (0.482)	−0.198 (−1.808)	1.262 (4.568)***	−1.689 (−17.262)***
30-39	−0.477 (−0.83)	−1.015 (−5.321)***	0.814 (6.513)***	−0.974 (−18.94)***
40-49	−0.356 (−0.686)	−0.884 (−6.646)***	0.962 (6.783)***	−1.393 (−19.962)***
		womanś marital status		
married	−0.072 (−0.29)	−0.335 (−3.065)**	0.741 (5.533)***	−1.169 (−19.127)***
never married	0.29 (0.639)	−0.441 (−4.625)***	1.225 (5.305)***	−1.566 (−18.421)***
divorced	0.27 (0.725)	0.368 (3)**	1.313 (5.642)***	−1.453 (−15.885)***

Note: The table presents estimated parameters of the decomposition in pre-post setting for the 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 200 bootstrap replications are presented in the parentheses. Stars denote *** p-val.<0.001, **, p-val.<.005, * p-val.<0.01.

Appendix B.3 Most Important Predictors

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 detailed results are presented in tables 4 - 7.

Necessity to provide care as a reason to quit the job experiences the second 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.

Table 4: 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	total working experience (years)	
2	woman	subjective evaluation of labor status in the previous year	employed
3	woman	# years worked in the previous job	
4	woman	can start a job within 2 weeks	yes
5	woman	# of months since lost the job	
6	woman	occupation	missing
7	woman	learned profession	specialists
8	woman	education achieved	master degree
9	woman	used to work in the past	but not anymore
10	husband	monthly wage	
		Top 10 pre-intervention predictors of the inflows	
1	woman	total working experience (years)	
2	woman	occupation	missing
3	woman	can start a job within 2 weeks	yes
4	woman	# years worked in the previous job	
5	woman	# of months since lost the job	
6	woman	subjective evaluation of labor status in the previous year	family duties
7	woman	reasons for losing a job	necessity to provide care
8	woman	sum of years until reaching 18 – female's children	
9	woman	# of child-years of other children until reaching 18	
10	woman	subjective evaluation of current labor status	unemployed
		Top 10 predictors of the inflows to increase their importance	
1	woman	subjective evaluation of labor status in the previous year	employed
2	woman	learned profession	specialists
3	woman	education achieved	master degree
4	husband	monthly wage	
5	woman	# years worked in the previous job	
6	household	# of individuals in the household:	all
7	woman	occupation - current or most recent	specialists
8	husband	wants to work more in order to earn more	yes
9	woman	years since education completed	
10	household	household lives in a rural zone	
		Top 10 predictors of the inflows to dencrease their importance	
1	woman	occupation	missing
2	woman	reasons for losing a job	necessity to provide care
3	woman	registered as unemployed	no
4	woman	subjective evaluation of labor status in the previous year	unemployed
5	woman	subjective evaluation of current labor status	unemployed
6	household	voivodship	lubelskie
7	woman	# of child-years of other children until reaching 18	
8	woman	sum of years until reaching 18 – female's children	
9	woman	has never worked	
10	woman	reasons for not looking for a job	failed to find before

Table 5: 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	subjective evaluation of current labor status	unemployed
2	woman	can start a job within 2 weeks	yes
3	woman	years since education completed	
4	woman	age	
5	woman	registered as unemployed	yes
6	woman	# of months since lost the job	
7	husband	years since education completed	
8	woman	subjective evaluation of labor status in the previous year	unemployed
9	woman	reasons for losing a job	necessity to provide care
10	woman	main source of income	dependent
		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	subjective evaluation of labor status in the previous year	unemployed
4	woman	age	
5	woman	# of months since lost the job	
6	woman	can start a job within 2 weeks	yes
7	woman	years since education completed	
8	woman	main source of income	pension
9	woman	total working experience (years)	
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	husband	years since education completed	
3	woman	can start a job within 2 weeks	yes
4	woman	marital status	married
5	woman	years since education completed	
6	woman	subjective evaluation of labor status in the previous year	student
7	woman	occupation	missing
8	husband	total working experience (years)	
9	woman	lives with:	husband/wife
10	father	subjective evaluation of labor status in the previous year	employed
		Top 10 predictors of the inflows to dencrease their importance	
1	woman	subjective evaluation of labor status in the previous year	employed
2	woman	registered as unemployed	yes
3	woman	main source of income	pension
4	woman	subjective evaluation of current labor status	pensioner
5	mother	subjective evaluation of current labor status	employed
6	woman	total working experience (years)	
7	husband	# of months since lost the job	
8	woman	lives with:	father
9	woman	marital status	unmarried
10	husband	main source of income	pension

Table 6: 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.

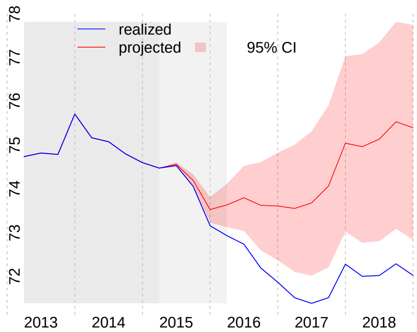
	who	description	details
		Top post-intervention predictors of the outflows	
1	woman	# of employees in the workplace	
2	woman	# of hours usually worked	
3	woman	# of months since lost the job	
4	woman	subjective evaluation of current labor status	employed
5	woman	has an additional job	no
6	woman	# years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

Table 7: 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.

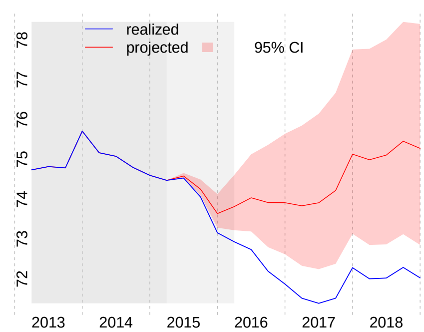
	who	description	details
		Top post-intervention predictors of the outflows	
1	woman	# of hours usually worked	
2	woman	subjective evaluation of current labor status	employed
3	woman	searched for a job within past month	no and has not found a job
4	woman	duration of search (months)	
5	woman	worked as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

Appendix B.4 Simulation Graphs

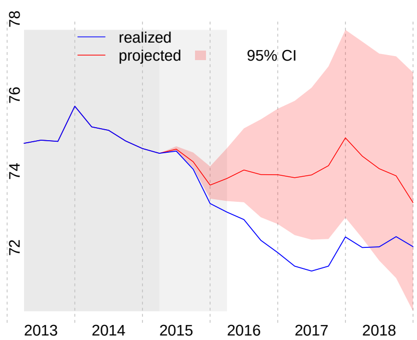
Figure 1: Effects on the labor force participation – simulation results.



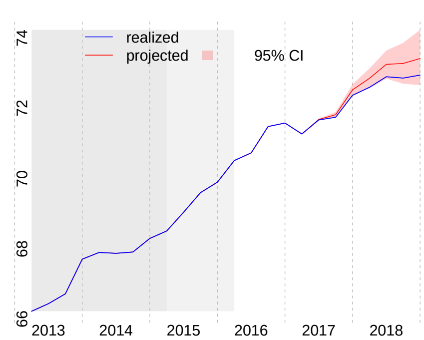
(a) Effects on the eligible - Scenario (1)



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

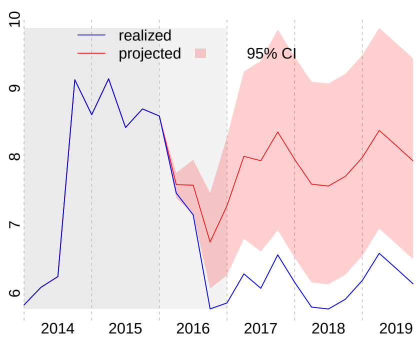


(c) Effects on the eligible - Scenario (3)

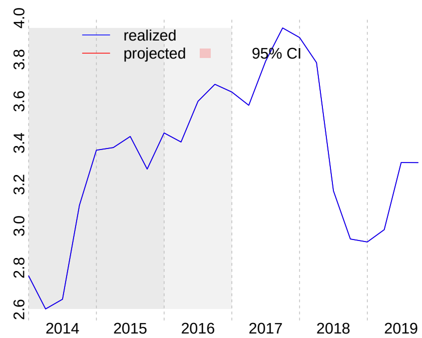


(d) Effects on the ineligible - Scenario (3)

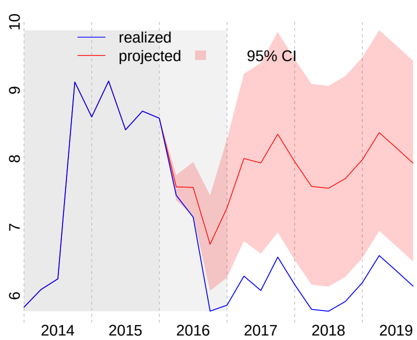
Figure 2: Effects on the flows – simulation results.



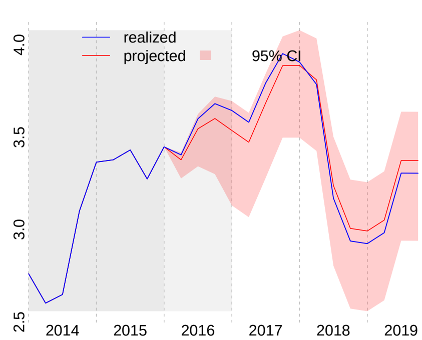
(a) Inflow rate dynamics among the eligible - Scenario (1)



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

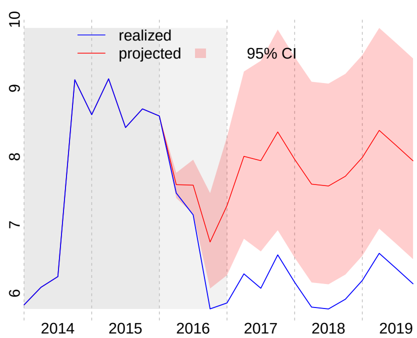


(c) Inflow rate dynamics among the eligible - Scenario (2)

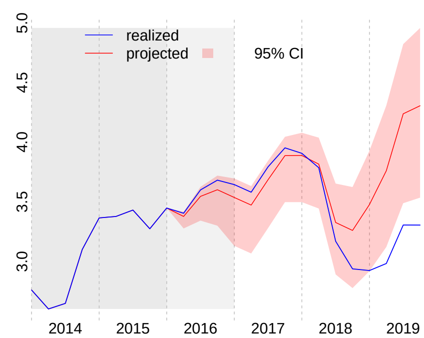


(d) Outflow rate dynamics among the eligible - Scenario (2)

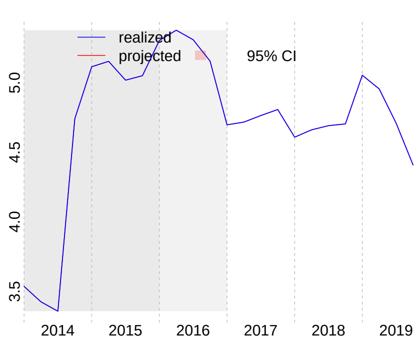
Figure 3: Effects on the flows – simulation results.



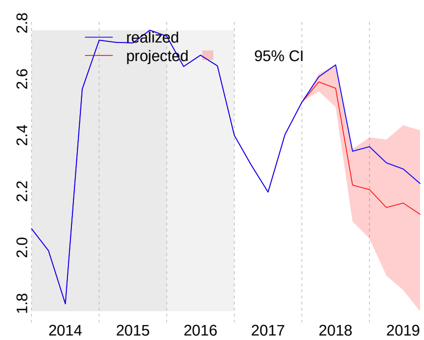
(a) Inflow rate dynamics among the eligible - Scenario (3)



(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)

References

Breiman, L. (2001). Random forests. *Machine learning 45*(1), 5–32.

Hastie, T., R. Tibshirani, J. H. Friedman, and J. H. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*, Volume 2. Springer.