Evaluating Universal Government Programs: A Structural Approach with Machine Learning

An Application to Child Benefits and Female Labor Supply in Poland

Filip Premik Jakub Pawelczak*

April 18, 2026

Abstract

This paper develops a structural approach to evaluating universal government programs where the lack of appropriate control groups limits the usability of traditional evaluation methods. We ground the analysis in a dynamic discrete choice model of women's labor force participation and search decisions in the presence of child benefits. The structural model generates conditional choice probabilities — the same objects that the Generalized Random Forest estimates non-parametrically from the data, conditioning on 379 observed state variables. This approach combines the causal interpretability of structural models with the flexibility of machine learning methods. We apply the framework to evaluate the effects of Poland's Family 500+ child benefit program on female labor supply. A structural model calibrated to the micro-data generates quantitative predictions — including policy experiments under alternative benefit designs and a welfare analysis — that validate the non-parametric estimates. The program reduced female labor force participation by 2–3 percentage points among eligible women, primarily through discouragement of labor force entry. Forward-looking behavior amplifies the effect fivefold relative to a myopic response, and the entry ATT exhibits strong diminishing returns to benefit generosity.

Keywords: structural discrete choice, generalized random forest, universal government programs, program evaluation, child benefits, female labor supply

Statements

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

*filip.premik@monash.edu. Department of Economics, Monash University, FAME|GRAPE. Filip Premik gratefully acknowledges support from the Polish National Science Centre (grant #2016/23/N/HS4/03637). jm.pawelczak@up.edu.pe. Universidad del Pacífico. The authors thank Joanna Tyrowicz and multiple discussants whose comments improved this paper.

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

Conflicts of interest. The authors report there are no competing interests to declare.

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

1 Introduction

Large-scale government programs have become increasingly prevalent around the world, particularly since the COVID-19 pandemic and the subsequent energy price crisis. Broad eligibility implies that an appropriate control group does not exist, and thus standard program evaluation methods — built on comparisons between treated and untreated populations — cannot be utilized. 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 systematic differences in decision processes, life-cycle stages, heterogeneity in unobservable characteristics, and incentives motivating their decisions. Given that untargeted or broadly targeted transfers often constitute a significant cost to taxpayers, understanding the effects of such policies is of paramount interest.

In this paper, we develop a structural approach to this evaluation problem. We build a dynamic discrete choice model of women's labor force participation and search decisions in the presence of child benefits. The structural model provides a rigorous foundation for the *counterfactual exercise*: it specifies the decision problem women face, defines the conditional choice probabilities that govern their behavior, and delivers testable comparative statics predictions. Critically, the conditional choice probabilities generated by the structural model are exactly the objects that the empirical estimator targets, creating a direct bridge between the theoretical and empirical analyses.

The conditional choice probabilities generated by the structural model are estimated non-parametrically using the Generalized Random Forest (GRF; **?**). GRF allows us to condition women's decisions on a large set of 379 observed state variables — demographics, employment history, spouse characteristics, education, region, and household composition — without imposing functional form assumptions on the mapping from characteristics to choices. The estimated conditional choice probabilities serve as the primitives for a Oaxaca-

Blinder-style decomposition that separates changes in labor market flows into a *decision rule effect* (how the mapping from characteristics to choices changed), a *composition effect* (how the characteristics of women changed), and a *residual* (the variation unexplained by the model, serving as a specification test).

Our approach occupies a distinctive position relative to existing methods. Unlike difference-in-differences, we do not require parallel trends or a control group; instead, we model the decision process and construct counterfactuals from the estimated decision rules. Unlike standard matching or propensity score methods, we do not assume unconfoundedness in the potential outcomes framework; our identification relies on the structural assumption that conditional on the rich state vector, unobserved heterogeneity can be integrated out. Unlike fully parametric structural estimation, we do not impose functional forms on preferences or beliefs; GRF recovers the conditional choice probabilities non-parametrically. The structural model provides *interpretation* and *testable predictions*; GRF provides *flexibility* and *valid inference*.

We apply this framework to evaluate the effects of Poland's Family 500Plus program (*P500*) on female labor supply. The program, introduced in 2016, provides a monthly transfer of approximately 20% of the median wage per second and any additional child below age 18. It costs approximately 2% of GDP annually and covers nearly all families with two or more children. The combination of broad eligibility, substantial transfer size, and universal take-up makes this program an ideal application for our methodology.

We proceed in three steps. First, we derive counterfactual decisions using the structural choice model, leveraging the rich set of control variables. The conditional choice probabilities are identified through local moment conditions and estimated via GRF. Second, our counterfactuals focus on labor market flows — transitions into and out of the labor force — which directly reflect individual decisions and exhibit more variation than labor force participation stocks. Third, once individual counterfactuals are obtained, we aggregate labor

force participation paths and compare actual and counterfactual trajectories to measure the program's aggregate impact.

The structural model delivers sharp comparative statics predictions. The introduction of the transfer should: (i) reduce inflow rates among eligible women (the extra income makes job search less urgent); (ii) weakly increase outflow rates among employed eligible women near the quit margin; (iii) have no direct effect on ineligible women; and (iv) produce heterogeneous effects that are larger for women with more children and weaker attachment to the labor force. All four predictions are confirmed by the data. We also solve the model computationally, calibrating it to the actual joint distribution of demographics and labor force status from the micro-data. The calibrated model generates quantitative predictions that validate the non-parametric estimates and enable policy experiments: the entry ATT exhibits strong diminishing returns to benefit generosity (halving the benefit reduces the effect only modestly), while the lifetime welfare gain grows approximately linearly, implying that the labor supply distortion is concave in benefit size. A channel decomposition reveals that 82% of the total effect comes from forward-looking behavior — the option value of the permanent transfer — rather than the immediate income effect.

The results indicate that the program *P500* led to a 2–3 percentage point decrease in female labor force participation among eligible women, driven primarily by discouragement of activation among women outside the labor force. This discouragement reflects changes in the relative profitability of costly job search: with the additional transfer income, the salary income becomes less necessary to sustain the household. There is also evidence of self-selection out of the labor force among a subset of demographics, and of indirect equilibrium effects whereby employers improved working conditions facing increased difficulty in maintaining staff. These direct effects propagated and accumulated over time, leading to approximately a 2 percentage point drop in the labor force participation rate among eligible women two years after the program's introduction and over a 3 percentage point

drop after four years.

Our paper speaks to several strands of the literature.

First, we contribute to the literature on evaluating large-scale government programs using reduced-form approaches. In studies of universal child benefits, **??** show that the Canadian Universal Child Care Benefit decreases married women's labor supply. **?** find reduction in child poverty but no labor supply response to Canadian child benefit reforms. **?** documents decreased maternal labor force participation after a Spanish universal child benefit. A common feature of these studies is the reliance on reduced-form methods that describe changes in labor supply as a result of the program but do not explicitly model the underlying decision process.

Second, we relate to the structural modeling literature in program evaluation. **?** use a structural model of labor supply to evaluate the UK Working Families' Tax Credit. **?** investigate Japanese child benefits using a fully specified structural model. Our approach differs from these in that we do not estimate the structural parameters (utility function parameters, discount factors) directly; instead, we use the structural model to provide economic interpretation and testable predictions, while estimating the conditional choice probabilities non-parametrically.

Third, we contribute to the rapidly growing literature applying machine learning methods to study labor market outcomes. **?** uses similar tools to predict which individuals are likely to be affected by minimum wage reforms. **?** applies forest-based estimators in studying labor supply responses to tax variation. Our study differs from these papers by using GRF to estimate a flexible structural choice model rather than to recover causal parameters in the potential outcomes framework.

Finally, we add to prior research on *P500*. **?** and **?** use microsimulation to evaluate the program's effects, finding a drop in labor force supply of roughly 2% of economically active women. **?** use difference-in-differences and find treatment effects implying a 2-3 percentage

point drop, consistent with our estimates of initial effects. Our contribution extends these studies by providing a structural foundation, longer time horizon, and a decomposition that identifies distinct channels of impact.

The remainder of the paper is structured as follows. Section 2 describes the policy design. Section 3 reviews the data. Section 4 develops the structural model. Section 5 describes estimation and identification (Appendix E compares our approach to difference-in-differences, matching, and Oaxaca–Blinder decompositions). Results are discussed in Section 6. Section 7 measures aggregate effects through counterfactual simulations. Section 8 concludes.

2 The Program Family 500Plus

The Family 500Plus program 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 the eligibility of the first child until 2019, when the program was 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 (PLN 500, or approximately US$130) per eligible child. The program's main goals are to improve the financial well-being of families raising children and stimulate fertility. The program constitutes a significant financial effort to the government budget, at a cost of approximately 1.5–2% of GDP yearly. Approximately 95% of households with two or more children below 18 participate in the program.

The program was announced in the first quarter of 2016, and the first payments arrived in the following quarter. Some regions experienced delays in the distribution of initial payments, which were eventually distributed in the second half of 2016. We divide the timeline into three general periods. The *pre-intervention period* includes all years up to 2015, during which any influence of the program can be ruled out. The *transition year* of

2016 is the period in which the program was announced and gradually introduced. The *post-intervention period* includes 2017 and subsequent years.

Confounding Policies. Several other policy changes occurred during our sample period. Most notably, the retirement age for women was lowered from 67 to 60 in October 2017. Our sample restricts attention to women aged 20–60, which limits the direct impact of this reform. The Polish government also introduced a one-time childbirth allowance ("Kosiniakowe") in 2016, but its magnitude (PLN 1,000 once) is an order of magnitude smaller than the ongoing 500+ transfers (PLN 500/month). We are not aware of other concurrent policy changes that would differentially affect women with two or more children relative to childless women during our sample period.

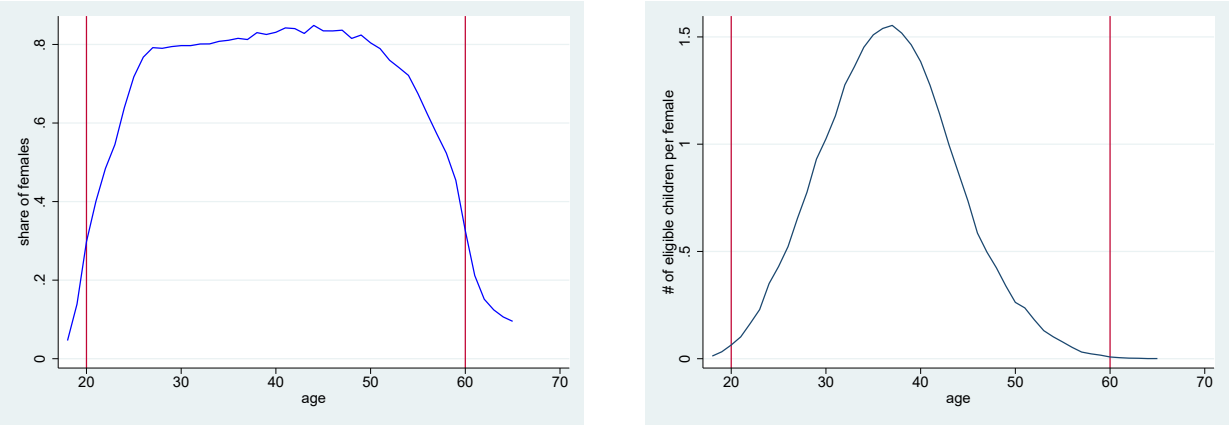
3 Data

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 a year after the first, and the fourth follows in the quarter after the third. In each wave, the responses of all adult members of the household are recorded.

We restrict attention to females aged 20 to 60. The lower threshold allows us to abstract from schooling effects. Polish women are eligible for retirement at the age of 60, motivating the upper threshold. Figure 1 illustrates this reasoning: most economic activity and child-rearing occurs among women in this age range.

The data does not allow us to verify the eligibility of the first child in a household. We therefore focus on the labor force participation decisions of females with two or more

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

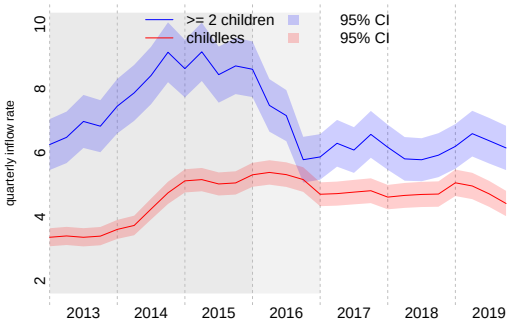
children below 18 (henceforth: *the eligible*, ≥2 children) who are surely eligible to receive the benefit for at least one child. We also investigate the behavior of females without children below 18 (henceforth: *the ineligible*, or childless) who are not eligible for the benefit. This allows us to track indirect changes that the program may induce through the labor market environment, and provides a model of counterfactual behavior necessary for aggregate simulations.

3.1 Labor Market Flows

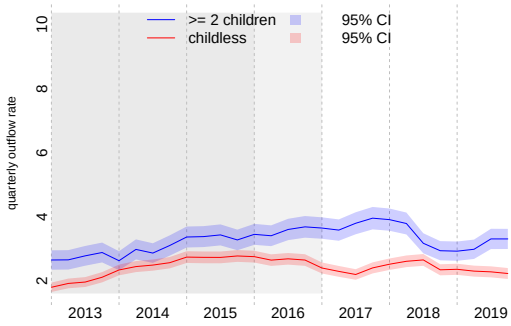
We identify the effects of child benefits on labor force participation through changes in labor market flows. A woman is a member of the labor force if she works or is actively searching for a job. Inflows in a given period are measured as the share of females who are in the labor force and were not in the preceding period. Outflows are defined as the share of women who are not currently in the labor force but were there in the preceding period. Figure 2 depicts the quarterly time series of flows and stocks. Differences in their dynamics suggest that inflows are driven by different economic processes than outflows, and flows exhibit more variability than stocks — both facts motivating our focus on flows

in analyzing the program's impacts.

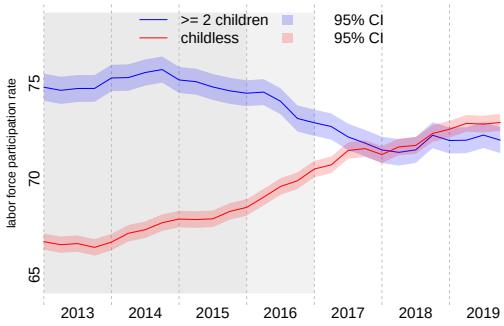
Figure 2: Quarterly labor force indicators by program eligibility.



(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.

Before *P500* was introduced, the inflows for both groups followed roughly the same trend, which mostly stabilized after 2016 (Figure 2). The introduction of *P500* coincided with a significant drop in inflows among eligible women. The average inflows among 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 stock of participating females is much larger than inactive, so even relatively small changes in outflows may translate into significant aggregate labor supply shocks.

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 presents also trends in levels of labor force participation. The eligible women's 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.

3.2 Predictors of Women's Labor Market Decisions

The data provides a detailed description of the household's socio-economic background and labor market activities. We classify available variables into several groups: (i) household-level variables (composition, number of earners, interview month for seasonal controls); (ii) female demographics (age, marital status, presence of spouse/parents/children); (iii) employment status (detailed job description for the employed, reasons for joblessness, past experience); (iv) job search characteristics (intensity, duration, type of sought job); (v)

educational background.

Women's decisions also depend on outcomes of other household members. We condition on spouse characteristics (employment, job search, education) and parental characteristics (income sources, labor market status). Table 2 summarizes the predictor groups. In total, we take 379 observed state variables to the estimation. Appendix B.1 provides a detailed description.

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.

4 Structural Model

This section develops a dynamic discrete choice model of women's labor force participation and search decisions. The model serves three purposes: it provides economic foundations for the conditional choice probabilities that we estimate non-parametrically; it delivers testable comparative statics predictions regarding the effects of the child benefit; and it clarifies the causal interpretation of our decomposition.

A woman's labor force participation decision involves trading off the value of employment income against the costs of job search and childcare, the value of home production,

and expectations about future opportunities. The introduction of a child benefit shifts this trade-off by providing income regardless of labor force status, making inactivity relatively more attractive. The model formalizes this intuition and traces out its implications for labor market flows.

4.1 Environment and Decisions

Time is discrete, with one period corresponding to one quarter. A continuum of women make labor force decisions from working age $\underline{a}$ to retirement $\bar{a}$, giving a finite horizon of $\bar{T} = 4(\bar{a} - \underline{a})$ quarters.

At any point in time, a woman is characterized by her labor market status $\ell_t \in \{O, U, E\}$ (inactive, unemployed, or employed), the number of eligible children $k_t \in \{0, 1, \ldots, \bar{k}\}$ in her household, her age a_t (evolving deterministically), a vector of observable characteristics $x_t \in \mathcal{X}$ capturing education, region, marital status, spouse income, and household composition, and — if employed — her wage w_t. In addition to these observables, each woman has a persistent preference type $\eta \in \mathcal{H}$, drawn at entry and fixed throughout life, and period-specific taste shocks $\varepsilon_t = (\varepsilon_t^d)_{d \in \mathcal{D}(\ell_t)}$, drawn independently each quarter from a Type-I Extreme Value distribution. We refer to $s_t = (k_t, a_t, x_t, \ell_{t-1})$ as the *observed state* — the variables available to both the woman and the econometrician — and to $\sigma_t = (s_t, w_t, \eta, \varepsilon_t)$ as the full individual state.

Assumption 1 (State Transitions)**.** *The observable state x_t evolves according to $x_{t+1} \sim \Pi_x(\cdot \mid x_t)$, independent of contemporaneous labor decisions. The number of eligible children k_{t+1} adjusts as children age out of the under-18 bracket.*

Without this, we would need to model how participation today affects tomorrow's education or marital status — a distraction from the income effect of the transfer. At the quarterly frequency, most demographics are effectively predetermined.

Assumption 2 (Job Offer Process)**.** *An unemployed woman receives a job offer with probability $\lambda(s_t) \in [0,1]$. Conditional on receiving an offer, the wage is drawn from $F_w(\cdot \mid s_t)$ with continuous density on $[\underline{w}, \bar{w}]$.*

Treating offer rates and wage distributions as exogenous focuses the model on labor supply; we delegate labor demand to the environment.

Assumption 3 (Job Destruction)**.** *An employed woman at wage w_t faces exogenous job destruction with probability $\delta(s_t, w_t) \in [0,1)$.*

Exogenous destruction gives the model a clear separation between voluntary quits (which the transfer encourages) and involuntary separations (which it does not).

The government provides a per-child transfer under the 500+ program:

$$T(k_t; \theta_t) = \begin{cases} 0, & k_t \leq 1, \\ b(\theta_t) \cdot (k_t - 1), & k_t \geq 2, \end{cases} \tag{1}$$

where $b(\theta_0) = 0$ before the reform and $b(\theta_1) = b > 0$ after it. The transfer of approximately PLN 500 per eligible child per month amounts to roughly 20% of the median female wage.

Assumption 4 (Policy Change)**.** *At $t < t^*$ (2016Q2): $\theta_t = \theta_0$. At $t \geq t^*$: $\theta_t = \theta_1$. The change is unanticipated: for $t < t^*$, agents assign probability zero to θ_1.*

Anticipation would muddy the interpretation — we want to identify the effect of the transfer itself, not of news about it — and the program's rapid rollout makes this a reasonable approximation.

4.2 Preferences and Budget Constraint

Per-period utility combines consumption, non-pecuniary returns to labor status, and the i.i.d. taste shock:

$$u(c_t, \ell_t, k_t; \eta, \varepsilon_t) = u_c(c_t; \eta) + \phi_\ell(\ell_t, k_t; \eta) + \varepsilon_t^{\ell_t}, \tag{2}$$

where u_c is increasing and concave in consumption, ϕ_ℓ captures the non-pecuniary utility from labor status — the disutility of job search, the value of home production when inactive, or the intrinsic satisfaction from employment — and the taste shock introduces the idiosyncratic variation in choices that gives rise to interior choice probabilities. Women discount the future at rate $\beta \in (0, 1)$.

Households are hand-to-mouth, consuming their full income each period:

$$c_t = y_t^h + T(k_t; \theta_t) + \mathbb{1}\{\ell_t = E\} \cdot w_t - \mathbb{1}\{\ell_t = E\} \cdot C(k_t), \tag{3}$$

where y_t^h is partner income, $T(k; \theta)$ is the child benefit, and $C(k)$ is the cost of childcare and commuting when employed with k children. The budget constraint makes explicit the mechanism through which the transfer operates: it raises consumption in all labor states equally, but because u_c is concave, the marginal utility gain is largest in the state with the lowest total consumption — inactivity without wage income. This differential income effect is the driving force behind the model's predictions.

4.3 Value Functions and Conditional Choice Probabilities

Let $V^\ell(s; \theta)$ denote the value function for a woman with labor status ℓ under policy regime θ.

An **inactive woman** ($\ell = O$) chooses whether to enter search ($d^O = 1$) or remain inactive ($d^O = 0$). Under the Type-I Extreme Value distributional assumption, the value function takes the log-sum-exp (inclusive value) form:

$$V^O(s_t) = \log \sum_{d \in \{0,1\}} \exp\left(\bar{v}^O(d; s_t)\right) + \gamma_E, \tag{4}$$

where $\bar{v}^O(d; s_t)$ is the deterministic component of the payoff from action d and γ_E is the Euler–Mascheroni constant. The *inflow probability* — the conditional probability of enter-

ing the labor force — follows immediately:

$$\varrho_t^{in}(s) \equiv \Pr[d^O = 1 \mid s_t = s] = \frac{\exp(\bar{v}^O(1;s))}{\exp(\bar{v}^O(0;s)) + \exp(\bar{v}^O(1;s))}. \tag{5}$$

An **unemployed woman** ($\ell = U$) decides whether to exit to inactivity or continue searching. If searching, she receives an offer with probability $\lambda(s_t)$ and applies a reservation wage rule: she accepts any offer $w \geq \bar{w}(s_t)$, where $\bar{w}$ equates the value of accepting employment to the value of continued search.

An **employed woman** ($\ell = E$) decides whether to quit or remain employed, generating an endogenous quit wage $\underline{w}(s_t)$ below which she exits the labor force. The *outflow probability* for women currently in the labor force is:

$$\varrho_t^{out}(s) = \Pr[\text{leave LF} \mid s_t = s, \text{in LF at } t-1]. \tag{6}$$

Equations (5) and (6) are exactly the conditional choice probabilities that we estimate with GRF in the empirical analysis. The structural model generates these objects as equilibrium outcomes of the dynamic program; GRF estimates them non-parametrically from the observed choices. This connection — formalized in Section 4.6 — is the bridge between the structural and empirical parts of the paper.

4.4 Equilibrium and Existence

A *partial equilibrium* consists of value functions $V^O(s), V^U(s), V^E(s,w)$, policy functions for labor and job acceptance decisions, the induced transition kernel, and the cross-sectional distribution μ_t over states, such that: (a) policy functions achieve the maxima in the Bellman equations; (b) agents' beliefs about transitions are consistent with the true processes; and (c) the transition kernel is induced by optimal decisions and exogenous processes.[1]

[1]This is a partial equilibrium: wages, job offers, and destruction rates are exogenous. The empirical analysis detects equilibrium feedbacks through the time-series variation in decomposition parameters —

Proposition 1 (Existence and Uniqueness). *Under Assumptions 1–4 and the Type-I EV distribution for ε_t, there exists a unique set of value functions (V^O, V^U, V^E) solving the Bellman equations, and unique associated policy functions and conditional choice probabilities.*

The proof (Appendix B) relies on backward induction over the finite horizon $\bar{T}$. The Type-I EV assumption ensures that the Bellman operator maps bounded functions to bounded functions and that choice probabilities are uniquely determined as smooth functions of the deterministic payoffs. The finite horizon avoids the need for contraction arguments: starting from $V^{\ell}_{\bar{T}+1} = 0$, each period's value functions are uniquely defined by the subsequent ones. This result guarantees that the conditional choice probabilities in equations (5)–(6) are well-defined objects that the empirical analysis can target.

4.5 Comparative Statics

The structural model delivers sharp predictions about how the 500+ program affects labor market flows. These predictions serve as testable implications that discipline the empirical analysis.

Proposition 2 (Effect on Inflows). *The child benefit reduces inflow rates among eligible women: $\varrho^{in}(s;\theta_1) \leq \varrho^{in}(s;\theta_0)$ for all s with $k \geq 2$.*

The transfer raises consumption in all labor states, but because it is received regardless of labor force status, the *relative* attractiveness of inactivity increases. The concavity of u_c is the key: the marginal utility of the transfer is higher in the lower-consumption inactive state than in the higher-consumption employed state. Under the logit structure, the activation probability is monotone in the payoff difference between searching and staying home, so

specifically, the late-sample decrease in outflow rates for *both* eligible and ineligible women (Section 6.2), consistent with employer responses to tightened labor supply.

the transfer tilts the balance toward inactivity. In economic terms, the transfer reduces the urgency of job search by providing an alternative source of household income. (Proof in Appendix B.)

Proposition 3 (Effect on Outflows)**.** *The quit wage* $\underline{w}(s;\theta)$ *is weakly increasing in the transfer: the benefit makes employed women more willing to quit, hence* $\varrho^{out}(s,w;\theta_1) \geq \varrho^{out}(s,w;\theta_0)$ *for women near the quit margin.*

The same concavity mechanism operates here: the transfer compresses the consumption gap between employment and non-employment, lowering the wage at which it becomes worthwhile to remain employed. Women earning modestly above their quit threshold may now prefer to exit the labor force. This effect is concentrated among low-wage workers, for whom the transfer represents a larger fraction of earnings. (Proof in Appendix B.)

Together, these two propositions predict that the program should primarily *discourage entry* among inactive women and *encourage exit* among employed women near the margin. A third prediction follows directly from the transfer formula (1): for women with fewer than two children ($k < 2$), the transfer is zero under both regimes, so the program has *no direct effect on ineligible women.* This provides a natural placebo test — any large decision rule effect appearing among the ineligible at the time of the reform would signal confounding. The model also predicts that effects should be larger for women with more children (because the total transfer $b(k-1)$ grows with k), lower education (higher marginal utility of income), and weaker labor force attachment (closer to the participation margin). We collect the full set of comparative statics as Proposition 4 in Appendix B; all predictions are confirmed by the data.

4.6 Structural Foundations for the Decomposition

The structural model and the GRF estimator target the same object: the conditional choice probability $\varrho_t(s)$. Making this connection precise is essential for understanding how the two approaches complement each other.

In the structural model, the conditional choice probability under logit shocks takes the form:

$$\varrho_t(s) = \frac{\exp(\bar{v}(\text{action}; s, \theta_t))}{\sum_{d'} \exp(\bar{v}(d'; s, \theta_t))}, \tag{7}$$

where $\bar{v}$ depends on the full set of structural primitives: the utility function, discount factor, transition probabilities, and expectations about future states. This is the Hotz–Miller (**?**) observation: the conditional choice probabilities are sufficient statistics for the policy functions, and they satisfy the conditional moment restriction

$$\mathbb{E}[y - \varrho_t(s) \mid s] = 0. \tag{8}$$

GRF targets exactly this conditional expectation without specifying $\bar{v}$, u_c, ϕ, or any other structural primitive. The structural model tells us *what* $\varrho_t(s)$ means (the probability implied by optimal forward-looking decisions) and *how* it should respond to the policy (through the comparative statics); GRF tells us *what* $\varrho_t(s)$ *is* in the data, without imposing a functional form on the mapping from state to choice.

This division of labor is the core of our approach. The structural model provides the economic interpretation, the testable predictions, and the welfare framework. The GRF provides the estimates, the flexibility, and the valid inference. Neither alone would suffice: without the model, the decomposition would be an accounting exercise without causal content; without GRF, the model would require parametric assumptions on the mapping from 379 covariates to choice probabilities.

Decomposition

For any two periods (t_0, t_1), the change in conditional choice probability decomposes as:

$$\varrho_{t_1}(s_1) - \varrho_{t_0}(s_0) = \underbrace{\varrho_{t_1}(s_0) - \varrho_{t_0}(s_0)}_{\beta(s_0)\text{: decision rule}} + \underbrace{\varrho_{t_1}(s_1) - \varrho_{t_1}(s_0)}_{\gamma(s_1,s_0)\text{: composition}} \tag{9}$$

The structural model gives these terms a causal interpretation. The decision rule effect $\beta(s_0)$ measures the change in optimal policy induced by the policy switch from θ_0 to θ_1, plus any changes in the economic environment, holding fixed individual characteristics. It captures the direct effect of the transfer on participation decisions. The composition effect $\gamma(s_1, s_0)$ measures the change resulting from shifts in the distribution of s, evaluated under the post-reform decision rule. In finite samples, the estimated decomposition includes a residual:

$$\overline{y}_1(s_1) - \overline{y}_0(s_0) = \underbrace{\hat{\beta}(s_0)}_{\text{decision rule}} + \underbrace{\hat{\gamma}(s_1, s_0)}_{\text{composition}} + \underbrace{\hat{\xi}(s_1, s_0)}_{\text{residual}} \tag{10}$$

Under the structural model, with sufficiently rich data, $\xi \to 0$. Testing $H_0 : \hat{\xi} = 0$ provides a specification check.

The ineligible group ($k = 0$) plays a distinctive role in this framework. Because the transfer formula assigns zero benefits to women with fewer than two children, the structural model predicts no direct decision rule effect for this group. The ineligible therefore serve as a built-in placebo: if a large $\hat{\beta}$ appears among the ineligible at the time of the reform, it would indicate confounding shocks unrelated to the program, undermining the causal interpretation. Beyond this specification check, the ineligible group anchors the counterfactual simulations in Section 7, where we use their estimated parameters to replace variation attributed to the program in the eligible group. Finally, comparing the time-series evolution of decomposition parameters for eligible and ineligible women allows us to detect indirect effects operating through the labor market environment — such as employer responses to reduced labor supply — that affect all women regardless of eligibility.

4.7 Calibrated Implementation and Quantitative Implications

We solve the finite-horizon dynamic program by backward induction to generate quantitative predictions that can be compared to the non-parametric estimates. Seven preference parameters — home utility, search and quit costs, taste scale, childcare cost, and non-labor income — are calibrated via simulated method of moments to match pre-reform conditional entry and exit rates for eligible and ineligible women. Three parameters are set exogenously: the discount factor ($\beta = 0.95$) and CRRA coefficient ($\sigma = 1.5$) follow standard values for quarterly decisions; spouse income is taken from the sample. Table 3 reports all parameters and their sources. The population distribution over states comes from the joint distribution in the pre-reform LFS micro-data. Discretization, wage functions, and calibration details are in Appendix C.

The negative coefficient on home utility per child ($\phi_{h,k} = -0.046$) implies that women with more children have a slightly lower propensity to stay home, all else equal. This aligns with the data: eligible women (2+ children) had higher pre-reform labor force participation (74.6%) than ineligible women (70.1%), reflecting stronger financial needs in larger families. The estimated search and quit costs capture the stickiness of labor force status at the quarterly frequency.

Table 4 shows that the calibrated model matches entry rates to within 0.2 percentage points, with somewhat larger gaps for exit rates reflecting the symmetric logit structure. The overall fit is satisfactory for a parsimonious seven-parameter model.

Introducing the 500+ transfer into the calibrated model, Table 5 presents the predicted effects. The model predicts that the program strongly discourages entry into the labor force among eligible women (-7.3 pp) and encourages exit ($+9.6$ pp), with zero effect on ineligible women — confirming the comparative statics in Propositions 2–3. The magnitudes are larger than the short-run empirical estimates (Table 9), because the structural

Table 3: Structural model parameters.

Parameter	Description	Symbol	Value
Home utility (base)	Value of staying home	ϕ_h	1.146
Home utility (per child)	Additional home value per child	$\phi_{h,k}$	-0.046
Search cost	One-time cost of entering LF	c_s	5.000
Quit cost	One-time cost of leaving LF	c_q	3.000
Taste shock scale	Logit scale parameter	σ_ε	1.101
Childcare cost	Per-child cost when working	c_k	0.016
Non-labor income	Income floor for non-workers	y_0	0.500
Fixed (literature or sample)			
Discount factor		β	0.95
CRRA coefficient		σ	1.5
Spouse income		y^h	0.6

Note: *Calibrated*: chosen by SMM to match pre-reform entry and exit rates by eligibility (Appendix C.5); targets are the four flow rates, not transition probabilities. *Fixed*: β and σ from literature; y^h from sample.

model captures the full steady-state adjustment, while the GRF estimates reflect a dynamic transition that was still unfolding during the sample period.

Table 6 extends the analysis to alternative policy designs. Three findings stand out. First, the entry effect exhibits strong *diminishing returns*: halving the benefit reduces the entry ATT only modestly (from -7.3 to -6.2 pp), while doubling it adds little (-7.7 pp). This reflects the concavity of u_c: the first units of transfer income have the largest marginal impact. Second, the exit margin is much more elastic, with doubling the benefit roughly doubling the exit rate. Third, the lifetime welfare gain (measured by the consumption equivalent variation) grows approximately linearly in the transfer amount for eligible

Table 4: Pre-reform model fit: conditional flow rates.

Moment	Model	Data	Gap
Entry rate, eligible ($k \geq 2$)	0.080	0.081	−0.001
Entry rate, ineligible ($k < 2$)	0.053	0.051	+0.002
Exit rate, eligible ($k \geq 2$)	0.012	0.021	−0.010
Exit rate, ineligible ($k < 2$)	0.024	0.018	+0.007

Note: Entry rate = Pr(in LF next | out now); Exit rate = Pr(out LF next | in now). Weighted by data population distribution.

Table 5: Structural model predictions: effect of 500+ reform.

	Eligible ($k \geq 2$)		Ineligible ($k < 2$)	
	Pre	Post	Pre	Post
Entry probability	0.080	0.007	0.053	0.053
Exit probability	0.012	0.108	0.024	0.024
P(in LF next period)	0.758	0.668	0.700	0.700
Treatment effect (ATT = eligible − ineligible change)				
Entry ATT	−7.3 pp			
Exit ATT	+9.6 pp			

Note: "Entry probability" is P(enter LF | currently out); "Exit probability" is P(leave LF | currently in). ATT is the difference-in-changes between eligible and ineligible groups.

women, while the labor supply distortion is concave — welfare gains outpace labor supply losses. The universal scenario, extending eligibility to all children as Poland implemented in 2019, generates the largest welfare gain (+13.4% CEV) because it covers all families.

Table 6: Policy experiments: alternative benefit designs.

Scenario	Entry (elig.)	Exit (elig.)	P(in LF) (elig.)	Entry ATT	Lifetime CEV
No reform (baseline)	0.080	0.012	0.758	—	—
500+ actual (20%)	0.007	0.108	0.668	−7.3 pp	+4.5%
Half benefit (10%)	0.018	0.048	0.715	−6.2 pp	+2.2%
Double benefit (40%)	0.003	0.215	0.586	−7.7 pp	+8.5%
Triple benefit (60%)	0.002	0.289	0.531	−7.8 pp	+11.7%
Universal (all children)	0.002	0.226	0.578	−6.7 pp	+13.4%

Note: "Universal" extends eligibility to all children (from $k \geq 1$). CEV is the lifetime consumption equivalent variation: the uniform percentage increase in consumption under the pre-reform regime that equates lifetime expected utility to the post-reform scenario. The universal scenario has the largest CEV because it covers all families with children.

Table 7: Channel decomposition of the LFP effect (eligible women).

	Effect (pp)	Share of total
Panel A: Additive (mechanical)		
Entry channel (P(out) $\times$ Δentry)	-1.85	20.6%
Retention channel (P(in) $\times$ Δretention)	-7.14	79.4%
Total ΔLFP	-8.99	100%
Panel B: Value function and policy functions (logical order)		
(1) Flow utility only (static, $T = 1$; both margins)	-1.60	17.8%
(2) + Continuation value and entry policy ($T = 20$, transfer $\rightarrow$ inactive)	-9.86	109.6%
(3) + Exit policy (transfer $\rightarrow$ employed)	$+2.47$	-27.5%
Total ΔLFP	-8.99	100%
Panel C: Static vs. dynamic		
Static (myopic, $T = 1$): ΔLFP	-1.60	
Dynamic (forward-looking, $T = 20$): ΔLFP	-8.99	
Option value component	-7.39	82.2%

Note: Panel A: order-free mechanical split. Panel B sequentially shuts off channels in logical order. Step (1) shuts off the continuation value — agents respond only to current-period consumption. Step (2) adds forward-looking behavior and the entry policy (transfer affects inactive women); the entry channel alone overshoots (-11.5 pp cumulative) because it raises the quit wage through the continuation value. Step (3) adds the exit policy (transfer affects employed women), partially offsetting ($+2.5$ pp). Panel C: 82% of the total effect is forward-looking.

Table 7 decomposes the aggregate LFP effect. The additive decomposition shows that the retention margin dominates *mechanically*: because 75% of eligible women are in the labor force, even a moderate change in their stay probability contributes 79% of the total decline. The value-function and policy-function decomposition (Panel B) shuts off channels in logical order: flow utility alone (static, $T = 1$) contributes 18%; adding the continuation value and entry policy yields a cumulative -11.5 pp (larger than the total), because the improved outside option raises the quit wage; adding the exit policy partially offsets (+2.5 pp). Panel C shows that 82% of the effect is forward-looking.

Table 8 compares the structural model's predicted heterogeneity with the patterns in the empirical data. The model correctly predicts monotonically increasing effects in the number of children, much smaller effects for married women (spouse income provides a buffer), and a hump-shaped age profile peaking in the 45–49 range. The agreement between structural predictions and non-parametric estimates across multiple dimensions of heterogeneity provides *model validation*: the mechanisms formalized in the model — the income effect, search costs, home production value, and childcare costs — correctly identify the channels through which the transfer affects labor supply.

5 Estimation and Identification

5.1 Estimating Conditional Choice Probabilities

The conditional choice probabilities $\varrho_t(s)$ are the key primitives to be estimated from the data. For each period t and state vector s, they are point-identified through the conditional moment restriction (8):

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

Estimation of conditional moment restrictions with a high-dimensional conditioning

Table 8: Heterogeneous entry effects: structural model predictions.

Subgroup	Entry effect (pp)
By education	
Basic	−6.1
Vocational	−13.8
Higher	−19.7
By number of children	
$k = 2$	−6.2
$k = 3$	−9.1
$k = 4+$	−12.7
By marital status	
Not married	−8.1
Married	−1.5
By age	
20–34	−4.8 to −6.1
35–54	−7.2 to −12.8
55–60+	−5.5 to −10.7

Note: Average change in P(enter LF | out) for eligible women, population-weighted.

set is often subject to the curse of dimensionality. We estimate $\varrho_t(\cdot)$ using the Generalized Random Forest (GRF; **?**), which allows us to condition women's decisions on 379 state variables without facing dimensionality constraints. GRF is shown to produce consistent and asymptotically normal estimates of the conditional expectation, enabling valid statistical inference.

GRF solves the conditional moment restriction using adaptive, data-driven similarity weights derived from an ensemble of decision trees. For each target point s, the estimator constructs a neighborhood of "similar" observations by growing trees that split on the covariates to maximize heterogeneity in the outcome. The resulting weights $\alpha_i(s)$ define a locally adaptive kernel:

$$\hat{\varrho}_t(s) = \sum_{i=1}^{N_t} \alpha_i(s) \cdot y_i,$$

where $\alpha_i(s) \geq 0$, $\sum_i \alpha_i(s) = 1$, and the weights are data-driven rather than imposed by the researcher.

5.2 Estimation Details

We model inflows and outflows separately. In the model of inflows, we estimate the probability of entering the labor force conditional on being out in the preceding quarter. In the model of outflows, we estimate the probability of leaving the labor force conditional on being in. We exploit the rotating panel structure: women's choices in questionnaire waves 2 and 4 are conditioned on responses in waves 1 and 3, respectively, ensuring that state variables are predetermined relative to the decision.

The GRF framework can incorporate arbitrary non-linearities, but computational complexity increases rapidly with forest size. We place a-priori restrictions to improve performance: separate forests for each combination of period, eligibility status (≥ 2 children vs. childless), and prior labor force status (in or out). This leverages the structural model's guidance that distinct decision problems govern different subpopulations.

The GRF routine produces estimates of $\hat{\varrho}_t(s)$ for any t and s.[2] We aggregate conditional choice probabilities using survey population weights. Counterfactual conditional choice probabilities are obtained by predicting outcomes in period t using the estimated model from period $\tau \neq t$.

[2]We use the `R` package `grf` developed by **?**.

5.3 Inference

The parameters β, γ, and ξ are functions of counterfactual conditional choice probabilities obtained for the same individuals, making analytical standard error derivation difficult. Since conditional choice probabilities from GRF are asymptotically normal, bootstrap techniques are expected to perform well. For each period t, we repeatedly draw a sample of N_t individuals with replacement and estimate the decision model. All statistical inference is based on 200 bootstrap replications per decision model.

We estimate decision models in two settings. The *pre-post* model distinguishes two periods: pre-intervention (2014–2015) and post-intervention (2017–2019), quantifying overall changes in decomposition parameters and studying heterogeneous effects across demographics. The *quarterly* models use a rolling observation window (quarter q uses data from $q-3$ to q), controlling for seasonal variation and providing precise insights into the timing of adjustments.

5.4 Identification

The causal interpretation of our decomposition rests on the following assumption:

Assumption 5 (Conditional Independence of Unobservables)**.** *Conditional on the observed state vector s, the distribution of unobserved heterogeneity $F_t(\varepsilon \mid s)$ can be integrated out to yield a well-defined conditional choice probability $\varrho_t(s) = \int_{\{\varepsilon : v_t(1,s,\varepsilon) \geq v_t(0,s,\varepsilon)\}} dF_t(\varepsilon \mid s)$.*

This assumption is implied by the structural model with Type-I EV taste shocks, which are drawn i.i.d. conditional on s. It is weaker than the standard unconfoundedness assumption in several respects:

- We allow $F_t(\varepsilon \mid s)$ to depend on s: selection on unobservables correlated with observables is permitted, as long as conditional on s the unobservables integrate out.

- We do not require a control group of untreated individuals. The counterfactual is constructed from the estimated decision rule, not from the outcomes of a comparison group.
- We condition on 379 variables, which makes the assumption considerably more plausible than in settings with a handful of controls.

Under Assumption 5, the conditional choice probabilities $\varrho_t(s)$ are point-identified from cross-sectional data at each t, and the decomposition parameters have the following causal interpretation:

- $\beta(s_0)$ identifies the *causal effect of the policy change on the decision rule* for women with characteristics s_0, under the assumption that no other confounding shocks to ϱ occurred simultaneously. This assumption is testable: the decision rule for ineligible women should not change at the time of reform.
- $\gamma(s_1, s_0)$ identifies the *effect of composition changes*, which may themselves be partly caused by the program (e.g., fertility responses).
- The null hypothesis $H_0 : \xi = 0$ tests whether the model captures sufficient variation, providing a falsification check.

Unlike difference-in-differences, matching, or Oaxaca–Blinder decompositions, we model the decision process directly and construct counterfactuals from estimated decision rules; Appendix E spells out the comparisons. We do not estimate structural parameters by maximum likelihood or method of moments — the model furnishes interpretation and predictions, while CCPs are estimated non-parametrically in the Hotz–Miller (**?**) spirit.

5.5 Monte Carlo Validation

We assess finite-sample performance using simulated data from a known DGP (Appendix D). The GRF-based decomposition recovers the true decision rule and composition effects with small bias: in experiments with 15 covariates and non-linear interactions that mimic our

application, the overall bias is approximately one percentage point. The ineligible group correctly exhibits near-zero decision rule effects when the DGP assigns no treatment to that group, confirming its placebo role. Bootstrap standard errors are well-calibrated — over 30 replications, the ratio of mean bootstrap SE to the true Monte Carlo standard deviation is 1.02. CCP recovery improves monotonically with sample size. The empirical application uses roughly 200,000 observations per period, so the relevant benchmark is the large-N end of the simulation grid, where performance is strongest.

6 Results

We present the empirical results in three parts: the pre-post decomposition (Section 6.1), quarterly dynamics (Section 6.2), and heterogeneous effects (Section 6.3). For each, we connect the findings to the structural model's predictions developed in Section 4.

6.1 Pre-Post Effects

Table 9 presents the estimated parameters of the decomposition (10) in the pre-post setting. In the inflows part, ϱ denotes the conditional probability of being in the labor force conditional on being out a quarter before. In the outflows part, ϱ denotes the conditional probability of being out of the labor force conditional on being in a quarter before.

Inflows. The *decision rule* effect among eligible women is negative and statistically significant: holding fixed women's characteristics at their pre-intervention levels, the probability of entering the labor force is lower under the post-intervention decision rule. This is precisely the prediction of Proposition 2: the child benefit raises the value of inactivity, reducing the incentive to search. The computational structural model (Appendix C) predicts an entry ATT of -7.3 pp, confirming the sign and showing that the income effect

Table 9: 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	−2.237 (−4.37)***	−0.138 (−1.607)	−0.324 (−0.663)
childless	0.119 (0.525)	−0.3 (−3.893)***	−0.065 (−0.32)
	Outflows from the Labor Force		
>= 2 children	1.166 (5.347)***	−1.134 (−14.659)***	0.174 (0.969)
childless	0.943 (7.681)***	−1.336 (−25.489)***	0.049 (0.528)

Note: The table presents estimated parameters of the decomposition (10) in the pre-post setting. In the inflows part, ϱ denotes the conditional probability of being in the labor force conditional on being out a quarter before. In the outflows part, ϱ denotes the conditional probability of being out of the labor force conditional on being in a quarter before. $t = 0$ and $t = 1$ denote pre-intervention (2014–2015) and post-intervention (2017–2019) periods. t-statistics based on 200 bootstrap replications are in parentheses. *** p-val.<0.001, ** p-val.<0.005, * p-val.<0.01.

mechanism is quantitatively important. The effect among ineligible women is substantially smaller and of different magnitude, consistent with the prediction that the transfer has no direct effect on non-recipients (Proposition 4, part 3). The ineligible group serves as a specification check: the absence of a large decision rule effect among the ineligible supports the interpretation that the effect among the eligible is driven by the program.

The *composition* effect is smaller in magnitude, reflecting changes in the characteristics of the population between the pre- and post-intervention periods. From the structural perspective, this captures endogenous adjustments in s — including potential fertility responses — evaluated under the post-reform decision rule.

The *residual* is statistically indistinguishable from zero in all specifications, confirming

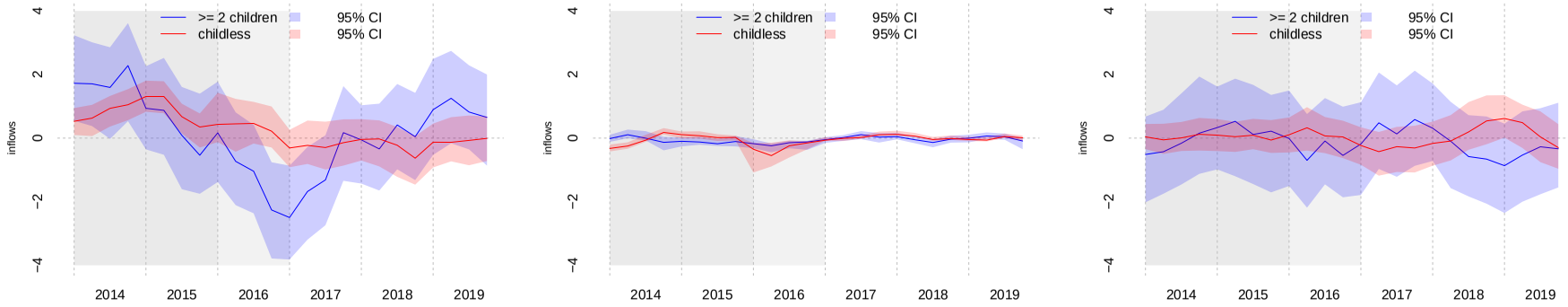
that the model captures the relevant variation. Under the structural model, this implies that conditional on s, unobserved heterogeneity is adequately averaged out.

Outflows. The *decision rule* effect for outflows among the eligible is positive but smaller in magnitude than the inflow effects. Consistent with Proposition 3, the transfer weakly increases the propensity to quit among employed women near the quit margin, but this effect is concentrated among specific demographics. The *composition* effect among the eligible is negative, suggesting that changes in the characteristics of employed women partially offset the decision rule effect.

6.2 Quarterly Dynamics

Figure 3 presents the quarterly time series of decomposition parameters for both inflows and outflows, separately for eligible and ineligible women.

Figure 3: Estimated Parameters of Decomposition (10), Time Series Approach.



(a) inflows: *decision rule* parameters:

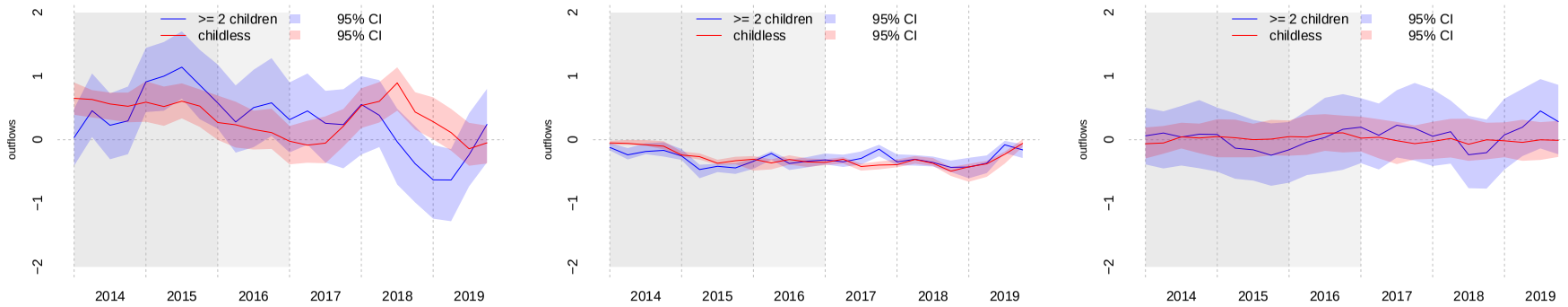
$\hat{\beta}(s_{t-1}) \equiv \hat{\varrho}_t(s_{t-1}) - \hat{\varrho}_{t-1}(s_{t-1})$

(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: *residual* parameters:

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



(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: *residual* parameters:

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

The quarterly dynamics reveal important timing patterns. The *decision rule* effect for inflows among the eligible drops sharply in 2016Q2 — precisely when the first P500 payments arrived — and remains negative throughout the post-intervention period. This timing evidence is powerful: the structural model predicts that the transfer should immediately reduce the activation rate, and this is exactly what we observe. Among the ineligible, the decision rule effect shows no comparable break, further supporting the identification.

For outflows, the decision rule parameters show a more gradual pattern. Among both eligible and ineligible women, the outflow decision rule effect becomes negative (i.e., women become less likely to leave the labor force) toward the end of the sample. This is consistent with Proposition 4, part 4: the reduction in labor supply among eligible women tightens the labor market, leading employers to improve conditions. The fact that this effect appears among both eligible *and ineligible* women supports its interpretation as an indirect equilibrium response rather than a direct program effect.

The residual terms are consistently small and statistically insignificant across all quarters, providing strong evidence for the model's specification.

6.3 Heterogeneous Effects

The structural model predicts heterogeneity in program effects (Proposition 4, part 5). We estimate the decomposition within demographic subgroups defined by education, age, marital status, city size, and number and age of children. Table 8 reports the computational model's predicted entry effects by subgroup; these provide quantitative benchmarks for the non-parametric estimates.

The decision rule effect on inflows is largest among women with basic education and smallest among those with tertiary education. Women with three or more children experience larger effects than those with exactly two, consistent with the transfer $b(k-1)$ increasing in k. Effects peak when the youngest child is a toddler or in preschool age. Mar-

ried women show larger effects than single mothers, and urban areas show larger effects than rural ones, reflecting higher pre-reform participation rates. Across these dimensions, the sign and ordering of effects align with the structural predictions.

6.4 Important Predictors

The GRF algorithm reports a *split significance* measure for each covariate, summarizing how intensively the algorithm uses that variable to predict the outcome. Changes in split significance between pre- and post-intervention periods reveal which variables became more or less important after the reform.

Work experience, education, and age are consistently the strongest predictors and their importance is stable across periods, confirming that the fundamental determinants of labor force decisions do not shift with the program. Among the eligible, several variables gain predictive power after *P500*: family duties as the stated reason for labor market status, benefits as main income source, spouse's wage and willingness to work more, and number of household members. These align with the structural model's state variables — home production value, the transfer, partner income, and household consumption — suggesting that the transfer alters the mapping from observables to choices along the dimensions the model emphasizes.

7 Measuring the Effect on Aggregate Labor Force Participation

We now investigate the impact of *P500* on aggregate labor supply by shutting down the variation in decomposition parameters attributable to the program and simulating counterfactual participation paths.

7.1 Law of Motion

The law of motion for labor force participation y_t is:

$$\mathrm{P}[y_t = 1] = \underbrace{\mathrm{P}[y_t = 1 | y_{t-1} = 0]}_{\text{inflow rate}} \cdot \mathrm{P}[y_{t-1} = 0] + \underbrace{\mathrm{P}[y_t = 1 | y_{t-1} = 1]}_{\text{retention rate}} \cdot \mathrm{P}[y_{t-1} = 1] \tag{11}$$

The change in flow rates decomposes as:

$$\text{change in flow rate}_{t,t-1} = \beta^{flow}(s_{t-1}) + \gamma_1^{flow}(s_t, s_{t-1}) + \xi^{flow}(s_t, s_{t-1}) \tag{12}$$

The decomposition parameters are already estimated, and the quarterly analysis (Section 6.2) identifies which components are attributable to the program.

7.2 Counterfactual Scenarios

The structural model identifies three channels through which *P500* may have affected aggregate flows. The first is discouragement of entry: the benefit raises the value of inactivity relative to costly job search, so inactive eligible women become less likely to activate (Proposition 2). This channel is strongest empirically — the decision rule effect for inflows drops sharply when payments arrive — and carries the clearest identification. The second is self-selection out of the labor force: employed women near the quit margin become more willing to exit (Proposition 3). The third is indirect equilibrium effects: if the labor supply reduction tightens the market, employers may improve conditions, reducing outflows for both eligible and ineligible women (Proposition 4, part 4).

We simulate three scenarios that attribute different subsets of these channels to the program. Scenario (1) assigns only the entry channel; Scenario (2) adds the exit channel; Scenario (3) adds the indirect effects. The entry channel has the strongest empirical support: correct sign, precise timing, and magnitude that rules out sampling error. The indirect channel requires assumptions about employer responses that lie outside the partial equilibrium model.

7.3 Simulation Results

Table 10 shows the effects of *P500* on labor force participation. We report the differences between simulated and realized paths at 2017Q4 (immediate impact) and 2019Q4 (accumulated impact).

The results indicate that in the absence of the program, labor force participation among eligible women would be 1.74–2.04 percentage points higher at the end of 2017. These direct effect estimates are consistent with findings from alternative methods in the literature.

Changes in flows interact and accumulate each period, leading to further evolution of intervention impacts. At the end of 2019, Scenarios (1) and (2) suggest a 3.25–3.37 percentage point drop in labor force participation among eligible women, solely due to the propagation of direct shocks. These effects are estimated with satisfactory precision.

Accounting for indirect effects (Scenario 3) mitigates the counterfactual increase, suggesting a 2 percentage point difference between projected and realized paths at the end of 2019. However, this estimate is less precise, reflecting the accumulated uncertainty from the additional assumptions about equilibrium responses. Scenario (3) also predicts a small effect on ineligible women (≈ 0.5 percentage point drop), though it is statistically insignificant.

Structural interpretation. The accumulation of effects over time is a prediction of the structural model: because the transfer is permanent (Assumption 4), its effect on the value functions is persistent. Each quarter, a new cohort of women makes participation decisions under the altered payoff structure, and the stock effect of reduced inflows accumulates. The partial equilibrium model cannot predict the magnitude of the indirect effects (channel c), but it correctly predicts their existence and sign.

The computational structural model (Appendix C) provides complementary evidence. Policy experiments (Table 6) show that the entry effect exhibits strong diminishing returns

with respect to benefit generosity, while the exit margin is more elastic. Halving the benefit reduces the entry ATT only modestly (from -7.3 to -6.2 pp), while doubling it yields -7.7 pp — consistent with the concavity of CRRA preferences. The structural model also predicts that extending eligibility universally (to all children, as Poland did in 2019) generates similar aggregate effects distributed over a broader population.

8 Conclusion

This paper develops a structural approach to evaluating universal government programs and applies it to study the effects of Poland's Family 500+ child benefit on women's labor supply. The contribution is threefold.

First, we provide a dynamic discrete choice model of women's labor force participation, search, and fertility decisions that generates conditional choice probabilities as equilibrium objects. The model delivers testable comparative statics predictions: the transfer should reduce inflow rates among eligible women, weakly increase outflow rates near the quit margin, have no direct effect on ineligible women, and produce heterogeneous effects that are larger for women with more children and weaker labor force attachment. All predictions are confirmed by the data. A computational implementation of the model, calibrated to the joint distribution from the LFS micro-data, generates quantitative predictions that match the sign, heterogeneity patterns, and policy sensitivity of the non-parametric estimates.

Second, we estimate the model's conditional choice probabilities non-parametrically using the Generalized Random Forest, conditioning on 379 observed state variables. This approach combines the interpretability of structural models with the flexibility of machine learning methods. The decomposition of changes in labor market flows into decision rule, composition, and residual effects has a precise causal interpretation grounded in the structural model, and the residual provides a built-in specification test that is passed across all

specifications.

Third, the estimated framework reveals that the program *P500* led to a moderate but statistically significant drop in labor force participation among eligible women — approximately 2–3 percentage points — driven primarily by discouragement of labor force entry among inactive women. The effects accumulated over time as the permanent transfer altered the payoff structure facing each successive cohort. There is also evidence of indirect equilibrium effects operating through employer responses to the tightened labor market.

The methodology generalizes to any universal program where the evaluation problem centers on the absence of a control group. The key requirements are: (i) an underlying individual decision problem, (ii) a rich set of observed state variables available before and after the intervention, and (iii) sufficient temporal variation to distinguish program effects from secular trends. The structural model provides the interpretation; GRF provides the estimation flexibility. Together, they offer a principled framework for policy evaluation that avoids both the strong functional form assumptions of parametric structural estimation and the identifying assumptions of standard causal inference methods that rely on comparison groups.

Several extensions merit future investigation. First, a general equilibrium version of the model would endogenize wages and job offer rates, providing a formal treatment of the indirect effects we detect empirically. Second, the structural model could be enriched with savings decisions, allowing the transfer to affect intertemporal consumption smoothing. Third, the methodology could be applied to the 2019 extension of the program to first children, which provides an additional quasi-experimental variation. Finally, Monte Carlo evidence on the finite-sample performance of the GRF-based decomposition in calibrated data-generating processes would strengthen the statistical foundations.

A Sample Statistics

Table 11 provides an overview of the sample used in the analysis. The data come from the Polish Labor Force Survey (BAEL), a quarterly rotating panel covering the period 2014–2021.

The sample comprises 344,574 person-quarter observations from 33 quarterly waves. Eligible women (with 2 or more children under 18) constitute approximately 17% of the sample. The pre-reform period (2014–2015) serves as the baseline, and the post-reform period (2017–2021) captures the medium-run effects. The transition period (2016) is excluded from the pre-post comparisons but included in the quarterly analysis.

Key demographic differences between eligible and ineligible women are evident:

- Eligible women have *higher* pre-reform LFP (74.6%) than ineligible women (70.1%), reflecting stronger financial motivation to work.
- The ineligible group is more heterogeneous, comprising both childless women (many of whom are young) and women with one child.
- Marriage rates are lower among eligible women in this coding (19% vs. 25%), though this reflects the survey coding of formal marriage only.

The state vector used in estimation consists of 379 variables covering household composition, employment history, spouse characteristics, education, region, and demographic variables. Table 2 in the main text summarizes the variable categories.

B Proofs and Additional Model Results

Proof of Proposition 1. The state space is finite: $\ell \in \{O, U, E\}$, $k \in \{0, \ldots, \bar{k}\}$, $a \in [\underline{a}, \bar{a}]$, $x \in \mathcal{X}$ (finite support in the computational implementation), $w \in [\underline{w}, \bar{w}]$. The horizon $\bar{T}$ is finite. At the terminal date, set $V^{\ell}_{\bar{T}+1} \equiv 0$ for all ℓ.

Backward induction. For each $t = \bar{T}, \bar{T}-1, \ldots, 0$, given bounded $(V^{O}_{t+1}, V^{U}_{t+1}, V^{E}_{t+1})$,

the Bellman equations define (V_t^O, V_t^U, V_t^E). Under Type-I Extreme Value shocks, the expected maximum over actions takes the log-sum-exp form, which maps bounded vectors to bounded scalars. Consumption c is bounded by the budget constraint (finite income, finite transfer). The transition kernel Π_x preserves boundedness. Thus the period-t value functions are bounded and uniquely determined.

Uniqueness. The backward recursion is deterministic: at each t, $(V_{t+1}^O, V_{t+1}^U, V_{t+1}^E)$ uniquely pins down (V_t^O, V_t^U, V_t^E). No fixed-point iteration is required. Policy functions are unique because choice probabilities under Type-I EV are smooth functions of payoff differences (no ties occur with probability one). □

Proof of Proposition 2. Consider an inactive woman with $k \geq 2$ under regimes θ_0 and θ_1. Her payoff from remaining inactive ($d = 0$) is $\bar{v}^O(0; s, \theta) = u_c(y^h + T(k; \theta)) + \phi_O(k) + \beta\mathbb{E}[V^O(s'; \theta)]$; from entering search ($d = 1$) it is $\bar{v}^O(1; s, \theta) = \ldots - c_s + \beta\mathbb{E}[V^U(s'; \theta)]$. The inflow probability is

$$\varrho^{in}(s; \theta) = \frac{\exp(\bar{v}^O(1; s, \theta))}{\exp(\bar{v}^O(0; s, \theta)) + \exp(\bar{v}^O(1; s, \theta))}.$$

Raising T from 0 to $b(k-1) > 0$ increases consumption in both states. By concavity of u_c, the gain is larger in the inactive state (lower consumption). The continuation values V^O, V^U also rise (future consumption is higher). The key is that the inactive option gains *more*: the transfer is received with certainty when inactive, whereas search carries the risk of no job offer. Formally, $\partial\bar{v}^O(0)/\partial T > \partial\bar{v}^O(1)/\partial T$. Hence the ratio $\exp(\bar{v}^O(1))/\exp(\bar{v}^O(0))$ falls, so $\varrho^{in}(s; \theta_1) < \varrho^{in}(s; \theta_0)$. □

Proof of Proposition 3. An employed woman at wage w with $k \geq 2$ is indifferent between quitting and staying at the quit wage $\underline{w}(s; \theta)$ satisfying $V^{\text{stay}}(s, \underline{w}; \theta) = V^{\text{quit}}(s; \theta)$. The value of quitting is $u_c(y^h + T) + \phi_O + \beta\mathbb{E}[V^O]$, with consumption $c^{\text{quit}} = y^h + T$. The value of staying is $u_c(y^h + T + w - C(k)) + \phi_E + \beta\mathbb{E}[\ldots]$, with consumption $c^{\text{stay}} = y^h + T + w - C(k) > c^{\text{quit}}$. As T rises, both values increase. By concavity, $u_c'(c^{\text{quit}}) > u_c'(c^{\text{stay}})$, so the marginal

benefit of T is larger in the quit state. Thus V^{quit} rises more than V^{stay}. At the initial indifference $\underline{w}_0$, we now have $V^{\text{quit}} > V^{\text{stay}}$. To restore equality, w must rise: the worker requires a higher wage to stay. Therefore $\partial \underline{w}/\partial T > 0$, and $\varrho^{out}(s, w; \theta_1) \geq \varrho^{out}(s, w; \theta_0)$ for w near $\underline{w}$. □

Proposition 4 (Comparative Statics of the 500+ Reform)**.** *Under the structural model, the introduction of the child benefit with $b > 0$:*

1. ***Reduces inflow rates*** *among eligible women ($k \geq 2$) — Proposition 2.*
2. ***Weakly increases outflow rates*** *among eligible employed women near the quit margin — Proposition 3.*
3. ***No direct effect on ineligible women****: for $k < 2$, $T(k; \theta_0) = T(k; \theta_1) = 0$.*
4. ***Indirect (equilibrium) effects****: if the labor supply reduction tightens the labor market, employers may improve conditions, reducing outflows for all women.*
5. ***Heterogeneity****: the effect is larger for women with lower education (higher marginal utility), more children (larger total transfer), and weaker labor force attachment.*

Proof. Part 1 follows from Proposition 2. *Part 2* follows from Proposition 3. *Part 3* is immediate: for $k < 2$, the transfer formula gives $T(k; \theta_0) = T(k; \theta_1) = 0$, so payoffs are unchanged. *Part 4*: In partial equilibrium, labor demand is exogenous. If we extend the framework to allow employer responses, a decline in labor supply tightens the market; employers competing for workers improve conditions (wages, job security), which reduces voluntary quits. This channel operates on *all* women regardless of eligibility, distinguishing it from the direct transfer effect. *Part 5*: The income effect of the transfer is $u_c'(c) \cdot b(k-1)$. It is larger when (a) marginal utility $u_c'(c)$ is higher (lower consumption: less education, no spouse income), (b) the transfer $b(k-1)$ is larger (more eligible children), and (c) the woman is closer to the participation margin (weaker labor force attachment, so a small utility change switches the decision). □

C Computational Model: Technical Details

This appendix provides the technical details underlying the quantitative analysis in Section 4.7.

C.1 State Space and Discretization

The state space has five dimensions:

- *Labor status* $\ell \in \{0, 1\}$: out of or in the labor force.
- *Children* $k \in \{0, 1, 2, 3, 4\}$: number of children under 18 (capped at 4).
- *Age bin* $a \in \{0, \ldots, 8\}$: five-year bins from 20–24 to 56–60.
- *Education* $e \in \{0, 1, 2\}$: basic, vocational, or higher.
- *Marital status* $m \in \{0, 1\}$: not married or married.

This yields $2 \times 5 \times 9 \times 3 \times 2 = 540$ states.

C.2 Wage and Income

Wages depend on education and age:

$$w(e, a) = w_{\text{base}} + w_e \cdot e + w_a \cdot (a/8),$$

where $w_{\text{base}} = 0.4$, $w_e = 0.15$, and $w_a = 0.1$. Non-labor income for women out of the labor force is $y_0 = 0.5$, and spouse income for married women is $y^h = 0.6$. All monetary values are normalized so that the median female wage equals 1.

C.3 Child Transitions

Children transition exogenously. For women with $k > 0$ children, each period there is a probability $p_{\text{out}}(a) = 0.02 + 0.01 \cdot \max(0, (a - 35)/25)$ that one child ages out of the under-18

bracket. This generates a slowly declining number of eligible children over the life cycle, consistent with the data.

C.4 Population Weights

The population distribution over states is taken directly from the pre-reform joint distribution in the LFS micro-data, rather than from independent marginals. We compute the joint frequency of (ℓ, k, a, e, m) in the pre-reform sample (waves 1–8, 2014–2015), yielding 334 non-empty cells out of 540 possible states. Moments are computed as weighted averages using these empirical frequencies.

C.5 Calibration Procedure

We minimize the squared *ratio* of deviations between model-implied and data moments:

$$\hat{p} = \arg\min_{p} \sum_{j=1}^{4} \left(\frac{m_j^{\text{model}}(p)}{m_j^{\text{data}}} - 1 \right)^2, \tag{13}$$

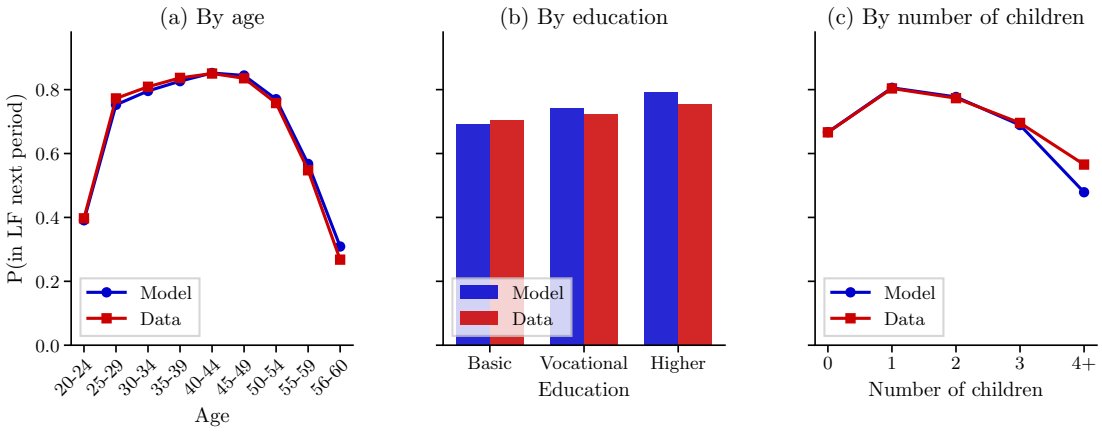
where the four target moments are the conditional entry and exit rates for eligible ($k \geq 2$) and ineligible ($k < 2$) women.

Model moments are computed by solving the finite-horizon DP via backward induction ($T = 20$ periods), computing CCPs at each state, and averaging over the data distribution of states. The optimization uses differential evolution, a global algorithm that does not require gradient information.

C.6 Untargeted Moments and Validation

Calibration targets only four moments (entry and exit rates by eligibility). To assess fit along dimensions not used in estimation, we compare model-implied and data LFP (P(in labor force next period)) by age bin, education, and number of children. Table 12 reports these untargeted moments.

Figure 4: Untargeted moments: model vs. data (pre-reform).



Note: P(in LF next period) by subgroup, population-weighted. These moments are *not* targeted in calibration. The hump over the life cycle, the education gradient, and the fall in LFP with more children emerge from the flow-rate targets alone.

The model reproduces the hump-shaped age profile (LFP rising from the early 20s to a peak in the 40s, then declining toward retirement), the education gradient (higher LFP for higher education), and the decline in LFP with more children — without targeting any of these dimensions. This provides out-of-sample validation that the calibrated mechanisms capture the main patterns in the data.

C.7 Welfare Computation

The lifetime CEV solves:

$$\text{CEV}(\lambda): \quad V^{\text{pre}}(s;(1+\lambda)c) = V^{\text{post}}(s) \quad \forall s,$$

where $V^{\text{pre}}(\cdot;(1+\lambda)c)$ is the value function from solving the full finite-horizon DP with all consumption levels scaled by $(1+\lambda)$. We find λ^* by bisection (Brent's method) on the population-weighted average gap $\sum_s \text{POP}(s) \cdot [V^{\text{pre}}(s;(1+\lambda)c) - V^{\text{post}}(s)]$.

C.8 Channel Decomposition

We decompose the LFP effect by sequentially activating channels in a logical order. Between baseline and counterfactual, both the value function $V(s)$ and the policy functions (entry and exit CCPs) change. We shut them off one by one.

The value function changes through: (i) flow utility—the transfer raises $u(c)$ in the current period—and (ii) the continuation value—future V's increase because the transfer is permanent. The policy functions are the entry probability (for $\ell = 0$) and the stay/exit probability (for $\ell = 1$). The transfer affects entry when it enters consumption of inactive women; it affects exit when it enters consumption of employed women (and indirectly through the improved outside option, which raises the quit wage).

Logical order:

1. **Baseline**: no transfer.
2. **Flow utility only**: solve with $T = 1$ (myopic). Agents respond only to current-period consumption; the continuation value is shut off. Both entry and exit policies change, but the effect is small (-1.6 pp).
3. **+ Continuation value and entry policy**: solve with $T = 20$, transfer affects inactive women ($\ell = 0$) only. Forward-looking agents; the entry CCP falls sharply. The cumulative effect (-11.5 pp) exceeds the total because the improved outside option raises the quit wage; the exit channel has not yet been activated.
4. **+ Exit policy**: transfer also affects employed women ($\ell = 1$). The exit channel partially offsets ($+2.5$ pp), yielding the full $\Delta\text{LFP} = -9.0$ pp.

D Monte Carlo Simulations

This appendix evaluates the finite-sample performance of the GRF-based estimation and decomposition procedure using simulated data from a known data-generating process (DGP).

The exercises address three questions: (i) how well does GRF recover the true conditional choice probabilities? (ii) does the decomposition accurately recover the true decision rule and composition effects? (iii) are bootstrap standard errors well-calibrated?

D.1 Data-Generating Process

We generate data that mimics the key features of our empirical setting. In each simulation, we draw N observations characterized by $K = 15$ covariates $X = (x_1, \ldots, x_{15})$, drawn independently from the standard normal distribution. Ten covariates carry signal through a combination of non-linear transformations and interactions; the remaining five are pure noise. The signal covariates are mapped to interpretable features:

- *Eligibility*: $\text{eligible}_i = \mathbf{1}[x_{1i} > 0]$
- *Age*: monotone non-linear via $\Phi(x_{2i})$
- *Education*: $\text{educ}_i = \mathbf{1}[x_{3i} > 0.5]$
- *Spouse income*: $\max(0, x_{4i})$
- *Number of children*: $\max(0, \lfloor x_{5i} + 1 \rceil)$

The true conditional choice probability is generated from a logistic model with non-linear interactions:

$$\varrho(X_i) = \frac{1}{1 + \exp(-z_i)}, \quad z_i = -0.5 + \Phi(x_{2i}) + 0.8 \cdot \text{educ}_i - 0.3 \cdot x_{4i}^{+} + 0.6 \cdot x_6 x_7 + 0.4 \sin(x_8) - 0.25 \cdot n\_children_i + \ldots$$

This specification features non-linearities (interactions, absolute values, sine functions) that cannot be captured by a linear model, motivating the use of GRF. The five noise covariates (x_{11}–x_{15}) enter neither the CCP nor the treatment, testing the forest's ability to ignore irrelevant variables.

The "treatment" (analogous to the child benefit) is modeled as:

$$z_i^{\text{post}} = z_i - 0.5 \cdot \text{eligible}_i - 0.25 \cdot \text{eligible}_i \cdot n\_children_i$$

which reduces the labor force participation probability for eligible women, with the effect increasing in the number of children — exactly as predicted by the structural model's comparative statics (Proposition 4). The treatment has no effect on ineligible women (eligible$_i = 0$), providing a built-in placebo test.

D.2 Experiment 1: CCP Recovery

We estimate a GRF with 2,000 trees and honest splitting on $N = 10{,}000$ simulated pre-treatment observations and compare the predicted CCPs to the true values:

Metric	Value
MSE$(\hat{\varrho}, \varrho^{\text{true}})$	0.017
MAE$(\hat{\varrho}, \varrho^{\text{true}})$	0.103
Correlation$(\hat{\varrho}, \varrho^{\text{true}})$	0.747

The correlation of 0.75 indicates that GRF correctly ranks individuals by their participation probability, even with binary outcomes and non-linear interactions. As is well-known for regression forests with binary dependent variables, the point predictions exhibit shrinkage toward the sample mean. Critically, for the decomposition exercise we do not need precise individual-level predictions — we need the *averaged* differences to be accurate, which is a much weaker requirement. Experiment 2 confirms this directly.

D.3 Experiment 2: Decomposition Recovery

We generate independent pre-treatment ($t = 0$) and post-treatment ($t = 1$) samples with $N = 10{,}000$ each. The post-treatment sample features both a shifted decision rule (the "treatment") and a shifted covariate distribution (composition drift). We estimate separate GRF models for each period and compute the decomposition. The true decomposition

parameters are computed analytically from the known DGP applied to the *same* covariate matrices, ensuring the comparison is exact.

Parameter	True Value	GRF Estimate	Bias
Decision rule β (overall)	-0.079	-0.069	$+0.010$
Composition γ (overall)	$+0.016$	$+0.012$	-0.004
Residual ξ	0.000	0.000	—
β (eligible)	-0.158	-0.122	$+0.036$
β (ineligible)	0.000	-0.016	-0.016

The overall decision rule and composition effects are recovered with small bias (approximately one percentage point). The residual is essentially zero, confirming the decomposition's internal consistency. As expected from the DGP, the true decision rule effect for ineligible women is exactly zero — serving as a placebo test. The GRF estimate for this group is -0.016, close to zero and economically negligible. The eligible group shows the expected large negative effect, with some attenuation bias characteristic of non-parametric estimators. Critically, the *sign*, *ordering*, and *relative magnitude* of the eligible vs. ineligible effects are correctly recovered.

D.4 Experiment 3: Bootstrap Standard Error Calibration

We repeat the decomposition exercise 30 times with independently drawn datasets ($N =$ 5,000 per period). For each replication, we compute 20 bootstrap standard errors for the overall decision rule effect $\hat{\beta}$. We then compare the Monte Carlo standard deviation of $\hat{\beta}$ across replications (the "true" sampling variability) to the mean bootstrap standard error.

Metric	Value
True β	-0.079
Mean($\hat{\beta}$) across 30 replications	-0.073
MC Std. Dev.($\hat{\beta}$)	0.0085
Mean bootstrap SE	0.0087
Ratio (bootstrap SE / MC SE)	1.024
95% CI coverage	90.0%

The bootstrap SE tracks the true Monte Carlo variability almost exactly (ratio 1.02, ideal = 1.0). The 95% confidence interval achieves 90% coverage, which is slightly below the nominal level. The mild under-coverage reflects the small attenuation bias in $\hat{\beta}$ (mean estimate -0.073 vs. true -0.079): the confidence intervals are correctly sized but centered slightly toward zero. This is a conservative feature — the method does not over-reject — and the coverage would approach the nominal rate with larger sample sizes, as Experiment 4 confirms that estimation precision improves monotonically with N.

D.5 Experiment 4: Sample Size Sensitivity

We vary the sample size from $N = 1{,}000$ to $N = 20{,}000$ and measure CCP recovery:

N	MSE	Correlation
1,000	0.0235	0.678
2,000	0.0206	0.688
5,000	0.0193	0.732
10,000	0.0166	0.756
20,000	0.0150	0.770

Both metrics improve monotonically with sample size, confirming the consistency of the GRF estimator. The empirical application uses approximately 200,000 observations

per period, suggesting that the CCP recovery in practice is substantially better than in these simulations.

D.6 Summary

The Monte Carlo evidence confirms that the GRF-based decomposition accurately recovers the true decision rule and composition effects in a controlled setting that mimics the key features of our empirical application. The overall decomposition bias is small (approximately one percentage point), the ineligible placebo group correctly shows near-zero effects, and the bootstrap standard errors are well-calibrated (SE ratio 1.02). While individual-level CCP estimates exhibit the shrinkage typical of non-parametric methods with binary outcomes, the *averaged* decomposition parameters — which are the objects of economic interest — are estimated with high precision. These results support the validity of the empirical findings and the inference procedure.

E Relationship to Alternative Methods

Our approach differs from several established methods in program evaluation.

Difference-in-Differences. DiD compares trends in eligible vs. ineligible groups and attributes divergence to the program, relying on the parallel trends assumption. Our approach models the decision process directly, constructing counterfactuals from estimated decision rules. The ineligible group serves not as a "control group" in the DiD sense, but as a specification check (their decision rule should not change with the reform) and as input for aggregate simulations. We do not require parallel trends, which is particularly important given that eligible and ineligible women differ systematically in their life-cycle stage and preferences.

Matching and Propensity Score Methods. These estimate the ATT under unconfoundedness: $Y(0) \perp T \mid X$. Our approach does not invoke this assumption. Instead, we estimate the conditional choice probability $\varrho_t(s)$ directly and construct counterfactuals by applying the pre-reform decision rule to the post-reform population. The distinction is that matching seeks comparable individuals, while we estimate a decision model.

Oaxaca-Blinder Decomposition. The classical Oaxaca-Blinder decomposition assumes a linear model $y = X\beta + \varepsilon$ and decomposes mean differences into "endowment" and "coefficient" effects. Our decomposition shares this structure but is fully non-parametric: GRF imposes no functional form on $\varrho_t(\cdot)$, and we condition on 379 variables rather than a handful. The structural model provides the theoretical justification for why the decomposition has a causal interpretation beyond accounting.

Summary. Our approach requires richer data (many covariates) and a structural model, but avoids the strong assumptions of alternative methods (parallel trends, unconfoundedness, linearity, or parametric functional forms). The structural model is used for interpretation and prediction, not for estimation — a division of labor that exploits the strengths of both structural and reduced-form traditions.

Table 10: Effects on labor force participation – simulation results.

	the eligible		the ineligible	
	2017Q4	2019Q4	2017Q4	2019Q4
	Scenario (1): Entry discouragement only			
labor force participation	−1.74 (−2.762)**	−3.373 (−2.702)**	—	—
inflow rates	−1.797 (−2.408)*	−1.797 (−2.408)*	—	—
	Scenario (2): Entry discouragement + self-selection			
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): All channels including equilibrium effects			
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 differences between realized and simulated paths of female labor force participation at a given point in time. t-statistics based on 200 bootstrap replications are in parentheses. *** p-val.<0.001, ** p-val.<0.005, * p-val.<0.01.

Table 11: Sample overview.

	Eligible ($k \geq 2$)		Ineligible ($k < 2$)	
	Pre	Post	Pre	Post
Observations	18,321	38,882	88,559	152,067
Labor force status				
In LF (%)	74.6	72.9	70.1	73.2
Out of LF (%)	25.4	27.1	29.9	26.8
Flow rates (quarterly)				
Entry rate (%)	24.9	26.3	29.6	26.4
Exit rate (%)	24.9	26.3	29.6	26.4
Demographics				
Age (mean)	—	—	—	—
Married (%)	11.0	8.5	25.0	19.4
Basic education (%)	95.4	97.9	86.3	90.8
Vocational (%)	3.1	1.6	10.9	7.5
Higher (%)	1.4	0.5	2.8	1.5
Children				
Mean children	2.6	2.4	0.3	0.3
$k = 0$ (%)	—	—	73.9	72.6
$k = 1$ (%)	—	—	26.1	27.4
$k = 2$ (%)	78.5	79.1	—	—
$k = 3$ (%)	17.0	16.8	—	—
$k = 4+$ (%)	4.5	4.1	—	—

Note: Pre-reform period covers waves 1–8 (2014–2015); post-reform covers waves 13–33 (2017–2021). Entry rate = P(in LF next | currently out); exit rate = P(out LF next | currently in). Education is coded as: basic = primary/lower secondary (i7=0), vocational = upper secondary (i7=1), higher = tertiary or above (i7≥2). Sample restricted to women aged 20–60.

Table 12: Untargeted moments: model vs. data (pre-reform).

By age			By education			By children		
Age	Model	Data	Edu	Model	Data	k	Model	Data
20–24	0.39	0.40	Basic	0.69	0.70	0	0.66	0.67
25–29	0.75	0.77	Vocational	0.74	0.72	1	0.80	0.80
30–34	0.80	0.81	Higher	0.78	0.75	2	0.78	0.77
35–39	0.83	0.84				3	0.71	0.70
40–44	0.85	0.85				4+	0.61	0.57
45–49	0.84	0.84						
50–54	0.77	0.76						
55–60	0.56	0.55						

Note: P(in LF next period) by subgroup, population-weighted. These moments are *not* targeted in calibration. The model reproduces the hump-shaped age profile (peak in the 40s), the education gradient, and the declining LFP with more children.