Attention vs Choice in Welfare Take-up: What Works for WIC?

Lei Bill Wang
Ohio State University
wang.13945@osu.edu

Sooa Ahn
Ohio State University
ahn.352@osu.edu

Job Market Paper

Abstract

Incomplete take-up of welfare benefits remains a major policy puzzle. This paper decomposes the causes of incomplete welfare take-up into two mechanisms: inattention, where households *do not consider* program participation, and active choice, where households consider participation but find it *not worthwhile*. To capture these two mechanisms, we model households' take-up decision as a two-stage process: attention followed by choice. Applied to NLSY97 data on the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), our model reveals substantial household-level heterogeneity in both attention and choice probabilities. Furthermore, counterfactual simulations predict that choice-nudging policies outperform attention-boosting policies. We test this prediction using data from the WIC2Five pilot program that sent choice-nudging and attention-boosting text messages to different households. Consistent with the counterfactual prediction, choice-nudging messages increased retention much more effectively than attention-boosting messages. (**JEL**: I38, H75, D12, D91, C33, C54)

Disclaimer: This research was conducted with restricted access to Bureau of Labor Statistics (BLS) confidential data. The views expressed here do not necessarily reflect the views of the BLS.

1 Introduction

Welfare programs significantly improve participants' well-being (Banerjee et al., 2024). Yet it remains a policy puzzle why participation rates of many programs remain stubbornly low (Ko and Moffitt, 2024). One prominent example is the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC). Kreider, Pepper and Roy (2016); Sonchak (2016) show that WIC successfully reduces childhood food insecurity and increases important health measures such as birth weight. Despite these positive impacts, the take-up rate of WIC has hovered around 50% over the past two decades, leaving millions of eligible households out of the program and more than a billion dollars worth of benefits unclaimed per year.

To address the gap between high program effectiveness and low take-up rates, policymakers have long pursued two broad strategies: attention-raising policies and choice-inducing policies. Attention-raising policies increase the probability that eligible households *consider* participation. They typically involve running informational campaigns that promote program awareness or informing selected households about likely eligibility. In contrast, choice-inducing policies increase the probability that households *choose* participation conditional on consideration. They typically involve improving the perceived value of the program benefits or reducing the hassle involved in the sign-up and eligibility recertification processes. For example, in 2016, the Vermont WIC office sent out attention-boosting and choice-nudging text messages to WIC participants at different sites. Attention-boosting messages remind recently dropped-out households that they might still be eligible for WIC. Choice-nudging messages inform participating households that continued enrollment in WIC is beneficial for the future of the children, improving participants' perceived value of the program benefits.[1] Both interventions aim at increasing participation by improving program retention. i.e., the intensive margin of program participation.[2]

The dichotomy of attention-raising versus choice-inducing policies motivates two empirical questions: What is the primary driver for incomplete take-up, inattention or a low choice tendency? What kind of policies, attention-raising or choice-inducing, promote take-up more effectively? We study these two questions in the context of WIC.

To answer the first question, we develop a two-stage inattention model that explicitly quantifies the contributions of inattention and low choice tendency to WIC nonparticipation. In our two-stage inattention model, households must first pay attention before making a choice about program participation. For a household to participate in WIC, it must pay attention at the attention stage *and* think that it is worthwhile to participate in WIC at the choice stage. We allow almost all covariates, including the lagged WIC take-up decision and household-level unobserved heterogeneity, to enter simultaneously into the attention and choice stages. This setup creates an identification challenge: when a covariate correlates with take-up, it is not immediately clear whether it operates through attention, choice, or both.

[1] It will become clear why these two interventions are attention-boosting and choice-nudging, respectively, once the two-stage inattention model is introduced.

[2] See examples of other attention-raising and choice-inducing policy interventions for various welfare programs in Bhargava and Manoli (2015); Ganong and Liebman (2018); Finkelstein and Notowidigdo (2019). These interventions increase participation by increasing new program recruitment. i.e., the extensive margin of program participation. Our two-stage inattention model can accommodate policy counterfactuals for both policies that increase program retention and policies that increase new recruitment.

We exploit two institutional structures of welfare programs to devise a novel strategy that separately identifies attention and choice parameters using standard panel data: full attention among participants and the presence of a recertification process. These two institutional structures are common in many welfare programs, making our model generalizable.

Full attention among participants: Many welfare programs require ongoing engagement with the program. To use the WIC benefit package, enrollees actively redeem food vouchers or use the benefit card provided by the program when buying groceries. Therefore, current program participants are fully attentive in the next period due to their ongoing engagement, while non-participants are only stochastically attentive. This institutional structure allows us to decompose the information in welfare participation panel data along two distinct dimensions. The first dimension is the probability of *staying* in the program. This probability is a function of choice parameters only, since current participants are fully attentive and proceed directly to the choice stage. We show that this first dimension identifies all choice parameters except for the hassle parameters.[3] The second dimension is the probability of *starting* participation, which depends on both attention and choice. By comparing these two dimensions, we identify the attention parameters.

Recertification: Welfare programs are means-tested, involving a periodic eligibility recertification process. Households required to recertify their eligibility need to pay a hassle cost to continue participation. All else equal, comparing the continuation behavior of households in recertification periods to those not in recertification periods provides clean variation that identifies the hassle parameters.

We apply our model to study WIC participation using data from the National Longitudinal Survey of Youth 1997 (NLSY97). Our model finds substantial household-level heterogeneity in both attention and choice behaviors. Notably, we find substantial household-level *unobserved* heterogeneity. This finding suggests that the incorporation of unobserved heterogeneity into both stages of the inattention model, as practiced by Heiss et al. (2021); Agarwal and Somaini (2025), can be crucial for counterfactual analysis based on such models.

To answer the second question, we simulate various attention-raising and choice-inducing policies using our model estimates. We use our preferred specification to simulate two types of counterfactuals: a population-wide perfect intervention and a realistic intervention that reflects feasible policy options.

Population-wide perfect intervention result indicates that at the *population level*, inattention is the largest barrier to WIC participation. Under the population-wide perfect intervention counterfactual, we simulate two extreme policies: one raises the attention probability for all households to 100% during all periods, and the other induces a 100% choice probability for all households during all periods. Forcing attention increases the take-up rate by 42pp, whereas forcing choice increases it by 8pp.

Given the infeasibility of population-wide perfect interventions, we simulate more realistic policies that reflect real-world constraints. WIC policymakers often focus retention efforts on currently-enrolled households due to budget limitations. We therefore design two policy coun-

[3] The choice stage design partitions the choice parameters into three types: benefit parameters, hassle parameters, and usage cost parameters.

terfactuals that resemble the likely impact of the attention-boosting and choice-nudging text message interventions in the 2016 Vermont pilot program. One, an attention boosting intervention increases the attention probability of the household that recently exited the program to 100%, giving the household one more chance to choose participation. Two, a choice-nudging intervention increases the current participants' next period choice probability by a small amount.

Strikingly, we find that a choice-nudging intervention that increases participating households' choice probability every month by a mere 0.45pp is as effective at raising overall take-up as an attention-boosting intervention that ensures full attention—but only for households in the month after they exit the program. The comparison suggests that under the constraints that policymakers can only intervene *current participants*, choice nudge is likely a more effective policy.

We reconcile the different implications of the two counterfactuals. The population-wide perfect intervention counterfactual suggests that inattention is a bigger barrier at the population level, whereas the more realistic intervention counterfactual suggests that choice nudge is likely a more effective policy under the constraint that policymakers can intervene with current participants only. This difference is primarily driven by an indirect mechanism: a choice nudge increases current-period retention, which mechanically induces full attention in the next period. Hence, a choice-nudging policy not only increases choice probability directly but also leads to higher future attention probability indirectly.

We test the hypothesis that choice-nudging is more effective than attention-boosting using the Vermont pilot program data. In 2016 and 2017, the Vermont WIC policymakers experimented with two types of text messages in different counties, both of which aimed at increasing the WIC retention rate. The first type of message was attention-boosting. Such messages reminded households who recently dropped out of WIC that they are still likely eligible for the benefits, essentially asking the households to reconsider their choice. The second type of message was choice-nudging. Each month, the WIC policymakers sent out educational messages to inform households of the benefits of continuing WIC participation, improving these households' perceived utility of the program benefit.

The WIC2Five report concludes that the choice-nudging intervention "did not meet our goal of increasing child retention" by comparing the *levels* from July 2015 to July 2017. We reanalyze the retention rates data provided by the WIC2Five report by comparing the *differences* between treated and control sites. Using difference-in-differences (DiD), event study, and permutation test designs, our analysis yields opposing results from the WIC2Five report. Sites that received attention-boosting text messages saw virtually no change in their retention rate; sites that received choice-nudging messages witnessed a significant 8-10pp increase in retention. The results corroborate our counterfactual analysis: a choice-nudging policy is likely to be more effective than an attention-boosting policy when it comes to promoting WIC take-up.

We contribute to three sets of literature. First, the existing works in the welfare take-up literature focus on estimating the causal relationship between observed covariates and the take-up rate (Rossin-Slater, 2013; Chareyron, Domingues and Lieno-Gaillardon, 2021) or the treatment effect of various policy interventions (Friedrichsen, König and Schmacker, 2018; Shepard and Wagner, 2025). Our work complements this literature by providing a structural framework that disentangles the attention and choice mechanisms behind these causal findings. Second, we propose new identifying conditions that have not been considered by existing inattention model literature (Hortaçsu, Madanizadeh and Puller, 2017; Heiss et al., 2021; Barseghyan, Molinari and Thirkettle, 2021; Abaluck and Adams-Prassl, 2021; Einav, Klopack and Mahoney, 2025; Agarwal and Somaini, 2025). Our identifying conditions are tied to the two institutional structures of welfare programs, making our model applicable to many other welfare take-up contexts. We further compare our methodological contribution to the existing inattention model literature in Appendix A. Third, we contribute to the long-standing empirical research on WIC (Bitler, Currie and Scholz, 2003; Rossin-Slater, 2013; Tiehen, 2014; Carlson and Neuberger, 2021; Davis, Leavitt and Chau, 2022). Using different methodologies and datasets, we present two pieces of new evidence that a choice-nudging intervention is likely to be more effective than an attention-boosting intervention in terms of promoting WIC take-up. Further experimental studies to test this hypothesis can be an interesting future research direction.

The rest of the paper is organized as follows: Section 2 and Section 3 explain the setup, identification, and estimation of the preliminary model and the full model, respectively. In Section 2 and Section 3, we provide intuitions for all theoretical results and present their proofs in Appendix B. Section 4 introduces the dataset and the institutional details of WIC, the welfare program to which we apply our model. Section 5 details the empirical findings pointed out by our empirical analysis, which is validated in Section 6. Section 7 concludes the paper.

2 Preliminary model

We first introduce a preliminary model where we do not include unobserved household-level heterogeneity. The purpose of setting up the preliminary model is twofold. First, the model is sufficient to illustrate the key intuition behind the identification strategy for the full model. Second, it clarifies a crucial mathematical assumption imposed when unobserved household-level heterogeneity is not taken into account.

2.1 Model setup

We introduce the notations and model premises that we use throughout the paper. We observe the household i's last period welfare take-up decision D_{it-1} and the current decision D_{it}, and $X_{it}, Z_{it} \in \mathcal{X} \times \{0,1\}$. X_{it} and Z_{it} are assumed to be strictly exogenous. X_{it} is the household characteristics, such as program benefit amount, demographic information, income, and education. It also contains a column of ones as the intercept term. X_{it} enters both stages simultaneously. Z_{it} is the indicator whether the household needs to recertify its eligibility if the household is already enrolled in the program in the last period. Z_{it} is a policy construct as the policymakers specify what type of households need recertification, i.e., $Z_{it} = 1$. Z_{it} is a choice stage exclusive shifter.

Households have to pay attention, denoted as $A_{it} = 1$, before choosing whether to participate in the welfare program, denoted as C_{it}. An econometrician observes that the household i participates in the program if and only if household i is attentive and thinks that it is worthwhile to participate in the program, i.e., $D_{it} = A_{it}C_{it}$. All observed variables and random shocks introduced later are assumed to be independent across $i \in \mathcal{I}$, where $\mathcal{I}$ represents all households. We assume that the attention probability and choice probability are strictly positive.

Attention stage

If household i has participated last period, it pays attention for sure; otherwise, the household has a probability to be attentive, and the probability depends on the household's characteristics X_{it}. We specify the attention probability as

$$P_a(D_{it-1}, X_{it}) := P(A_{it} = 1 \mid D_{it-1}, X_{it}) = P(X_{it}\gamma > \mathcal{E}_{it})^{1-D_{it-1}}$$

Our specification assumes that A_{it} is conditionally independent from Z_{it}, hence, we do not include Z_{it} in the specification of $P_a(D_{it-1}, X_{it})$.

Choice stage

We divide household characteristics, X_{it}, into three components: those related to the utility from using the program benefits, X_{it}^{υ}, e.g., benefit dollar amount; those related to disutility from the sign-up/recertification hassle, X_{it}^{κ}, e.g., local accessibility measures; and those related to disutility from the usage cost X_{it}^{χ}, e.g., stigma-related measures. Correspondingly, we specify utility from using the program benefits as $X_{it}^{\upsilon}\beta^{\upsilon}$, disutility from sign-up/recertification hassle as $X_{it}^{\kappa}\beta^{\kappa}$, disutility from usage cost as $X_{it}^{\chi}\beta^{\chi}$. Assuming that the sign-up and recertification processes are identical, we do not differentiate between sign-up hassle and recertification hassle; we refer to both as hassle cost and do not differentiate them.[4]

Households make a *static* utility-maximizing decision when they are attentive.A household compares the utility of using the program benefit and the (disutility of hassle + disutility from usage) when they need to sign-up or recertify their eligibility. It compares utility of using the program benefit and disutility from usage when it does not need to sign up or recertify to participate. The comparison is subject to a choice utility shock, Ξ_{it}.

$$P_c(D_{it-1}, X_{it}, Z_{it}) := P(C_{it} = 1 \mid D_{it-1}, X_{it}, Z_{it}) = P(X_{it}^{\upsilon}\beta^{\upsilon} + S_{it}X_{it}^{\kappa}\beta^{\kappa} + X_{it}^{\chi}\beta^{\chi} > \Xi_{it})$$

where $S_{it} = \max\{1 - D_{it-1}, D_{it-1}Z_{it}\}$ characterizes whether the households need to pay a hassle cost to participate/stay in the program. $S_{it} = 1$ in two cases. One, when households are not enrolled in the program last period, then it has to sign up for the program and pay a hassle cost. Two, when the households are enrolled in the program last period but are required to recertify their eligibility, i.e., $Z_{it} = 1$.

We include a column of ones into X_{it}^{υ} and another column of ones into X_{it}^{κ} to represent the baseline utility when a household participates in the welfare program and the baseline hassle cost when sign-up/recertification happens, respectively. As shown by the inclusion of two columns of ones, we do not rule out overlap between X_{it}^{υ} and X_{it}^{κ} (or between X_{it}^{χ} and X_{it}^{κ}). Hence, $X_{it} = X_{it}^{\upsilon} \cup X_{it}^{\kappa} \cup X_{it}^{\chi}$.

[4]We discuss how to accommodate different sign-up and recertification hassles in the extension section.

2.2 Identification

Following the consideration set literature (Abaluck and Adams-Prassl, 2021; Barseghyan, Molinari and Thirkettle, 2021; Agarwal and Somaini, 2025), we assume that we observe the take-up probability, $P(D_{it} = 1 \mid D_{it-1}, X_{it}, Z_{it})$, of a household for any given $\{D_{it-1}, X_{it}, Z_{it}\} \in \{0, 1\} \times \mathcal{X} \times \{0, 1\}$. Given the sequential nature of attention and choice, we decompose the observed single-period take-up probability into attention and choice probabilities under Assumption 2.1 imposed on $\{\mathcal{E}_{it}, \Xi_{it}\}$. We will explicitly relax this assumption in the full model.

Assumption 2.1.

(a) $\mathcal{E}_{it} \overset{iid}{\sim} G$ over time. G is strictly increasing and has full support.

(b) $\Xi_{it} \overset{iid}{\sim} H$ over time. H is strictly increasing and has full support.

(c) $\mathcal{E}_{it} \perp \Xi_{it}$ for all t.

Lemma 2.1. *Under Assumption 2.1, $P(D_{it} = 1 \mid D_{it-1}, X_{it}, Z_{it})$ is the product of attention and choice probabilities:*

$$P(D_{it} = 1 \mid D_{it-1}, X_{it}, Z_{it}) = \underbrace{G(X_{it}\gamma)^{1-D_{it-1}}}_{P_a(D_{it-1}, X_{it})} \underbrace{H(X^{\upsilon}_{it}\beta^{\upsilon} + S_{it}X^{\kappa}_{it}\beta^{\kappa} + X^{\chi}_{it}\beta^{\chi})}_{P_c(D_{it-1}, X_{it}, Z_{it})}.$$

With Lemma 2.1 in hand, by setting D_{it-1} and Z_{it} to different values, we can identify all γ and β. When $D_{it-1} = 1$, a household is fully attentive. Hence, the observed take-up probability is a function of choice parameters β only.

$$P(D_{it} = 1 \mid D_{it-1} = 1, X_{it}, Z_{it}) = H(X^{\upsilon}_{it}\beta^{\upsilon} + Z_{it}X^{\kappa}_{it}\beta^{\kappa} + X^{\chi}_{it}\beta^{\chi}).$$

We discuss identification of $\{\beta^{\upsilon}, \beta^{\chi}\}$ and β^{κ} separately by varying Z_{it}.[5] We first set $Z_{it} = 0$, the observed take-up probability

$$P(D_{it} = 1 \mid D_{it-1} = 1, X_{it}, Z_{it} = 0) = H(X^{\upsilon}_{it}\beta^{\upsilon} + X^{\chi}_{it}\beta^{\chi}).$$

Since H is strictly increasing, we can invert it to obtain a linear equation of β^{υ} and β^{χ}. We obtain a system of linear equations of β^{υ} and β^{χ} by varying X^{υ}_{it} and X^{χ}_{it}. Assuming that we observe sufficient variations in these two variables such that the system of linear equations is solvable, we can identify β^{υ} and β^{χ}. Then, we set $Z_{it} = 1$. The observed take-up probability is

$$P(D_{it} = 1 \mid D_{it-1} = 1, X_{it}, Z_{it} = 1) = H(X^{\upsilon}_{it}\beta^{\upsilon} + X^{\kappa}_{it}\beta^{\kappa} + X^{\chi}_{it}\beta^{\chi}).$$

Again, we can invert H and obtain a system of linear equations of β. Treating β^{υ} and β^{χ} as known (since they are already identified), we can identify β^{κ} with sufficient variations in X^{κ}_{it}. This identification result makes intuitive sense: we need some households to experience recertification hassle, i.e., $Z_{it} = 1$, to identify the hassle cost.

[5] We recognize that it is possible to discuss the identification of three types of β together. We discuss the identification of $\{\beta^{\upsilon}, \beta^{\chi}\}$ and β^{κ} separately to highlight the necessity of the recertification process for our identification results.

After identifying β, we set $D_{it-1} = 0$ to identify γ. The observed take-up probability is

$$P(D_{it} = 1 \mid D_{it-1} = 0, X_{it}, Z_{it}) = G(X_{it}\gamma)H(X_{it}^{\upsilon}\beta^{\upsilon} + X_{it}^{\kappa}\beta^{\kappa} + X_{it}^{\chi}\beta^{\chi})$$

We already identified all β, so we can pin down $G(X_{it}\gamma)$ as the ratio between the other two terms in the equation. By the strict monotonicity of G, we can invert G and obtain a system of linear equations of γ and identify all γ with sufficient variations in X_{it}. The three conditions in Proposition 2.1 spell out what is "sufficient variations" in the previous discussion.

Proposition 2.1. *Under Assumption 2.1,*

(1) Identification of β:

(1a) β^{υ} and β^{χ} are identifiable when the take-up probabilities are observed for at least $dim(\beta^{\upsilon}) + dim(\beta^{\chi})$ linear independent values of $\{X_{it}^{\upsilon}, X_{it}^{\chi}\}$ when $D_{it-1} = 1, Z_{it} = 0$,

(1b) β^{κ} is identifiable when the take-up probabilities are observed for at least $dim(\beta^{\upsilon}) + dim(\beta^{\chi})$ linear independent values of $\{X_{it}^{\upsilon}, X_{it}^{\chi}\}$ when $D_{it-1} = 1, Z_{it} = 0$, and $dim(\beta^{\kappa})$ linear independent values of X_{it}^{κ} when $D_{it-1} = 1, Z_{it} = 1$,

(2) Identification of γ:

γ is identifiable when the take-up probabilities are observed for at least $dim(\beta^{\upsilon}) + dim(\beta^{\chi})$ linear independent values of $\{X_{it}^{\upsilon}, X_{it}^{\chi}\}$ when $D_{it-1} = 1, Z_{it} = 0$, and $dim(\beta^{\kappa})$ linear independent values of X_{it}^{κ} when $D_{it-1} = 1, Z_{it} = 1$, and $dim(\gamma)$ linear independent values of X_{it} when $D_{it-1} = 0$.

Though Proposition 2.1 is heavy in notation, its intuition is straightforward. We partition the data into two dimensions: probability of staying in the program, corresponding to $D_{it-1} = 1$ and probability of starting program participation, corresponding to $D_{it-1} = 0$. The probability of staying is a function of choice parameters only since these households skip over the attention stage and proceed into the choice stage. This dimension of the data identifies choice parameters, corresponding to (1). One caveat is that the hassle parameter can only be identified if some households experience recertification hassle. Therefore, we further require variations in Z_{it} to separately identify $\{\beta_{\upsilon}, \beta^{\chi}\}$ and β^{κ}. This divides part (1) into (1a) and (1b). The probability of starting is a function of both attention and choice parameters because the household needs to overcome both the attention and choice hurdles. By comparing the first and second dimensions of the data, we identify the attention parameters. This corresponds to part (2).

3 Full model

3.1 Model setup

Assumption 2.1 required that the attention utility shock $\mathcal{E}_{it}$ and the choice utility shock Ξ_{it} are i.i.d. over time and mutually independent. The full model relaxes each of these restrictions by allowing for persistent, unobserved household heterogeneity captured by a random effect Q_i. Intuitively, Q_i represents latent time-invariant household traits such as thriftiness (or some part of thriftiness). One component of thriftiness may be explained by observed covariates entering the utility functions, while the orthogonal residual component is unobserved and summarized by Q_i. In two of the model extensions, we discuss how to relax the orthogonality assumption.

Assumption 3.1.

(a) $\mathcal{E}_{it} = \epsilon_{it} - \sigma_1 Q_i$, $\Xi_{it} = \xi_{it} - \sigma_2 Q_i$.

(b) $Q_i \overset{iid}{\sim} F$ over time, is independent from $\{X_{is}, Z_{is}\}_{s=1}^{T_i}$. F has full support.

(c) $\epsilon_{it} \overset{iid}{\sim} G$ over time. G is strictly increasing, differentiable, and has full support.

(d) $\xi_{it} \overset{iid}{\sim} H$ over time. H is strictly increasing, differentiable, and has full support.

(e) $\epsilon_{it} \perp \xi_{it}$ for all t.

Assumption 3.1 generalizes Assumption 2.1. When $\sigma_1 = \sigma_2 = 0$, the two sets of assumptions coincide, except that Assumption 3.1 additionally requires differentiability of G and H. This difference is largely innocuous in practice, since most applications assume logit, normal, or Type I extreme value shocks, all of which admit differentiable CDFs. We use g and h to denote the derivative of G and H, respectively. After this modification of the assumption on the unobserved random terms, the full model is specified as

$$P_a(D_{it-1}, X_{it}, Q_i) = P(\underbrace{X_{it}\gamma + \sigma_1 Q_i}_{U^a_{it}(X_{it}, Q_i)} > \epsilon_{it})^{1-D_{it-1}} \tag{3.1}$$

$$P_c(D_{it-1}, X_{it}, Z_{it}, Q_i) = P(\underbrace{X^{\upsilon}_{it}\beta^{\upsilon} + S_{it}X^{\kappa}_{it}\beta^{\kappa} + X^{\chi}_{it}\beta^{\chi} + \sigma_2 Q_i}_{U^c_{it}(D_{it-1}, X_{it}, Z_{it}, Q_i)} > \xi_{it}) \tag{3.2}$$

where $U^a_{it}(X_{it}, Q_i)$ and $U^c_{it}(D_{it-1}, X_{it}, Z_{it}, Q_i)$ denote the latent utilities from the attention and choice stages, respectively. The current period decision is characterized by $D_{it} = A_{it}C_{it}$, where A_{it} and C_{it} are binary random variables characterized by Equation (3.1) and Equation (3.2), respectively.

So far, we have described the transition from period $t-1$ to t. To complete the model we must specify the initial condition, which is central in panel models with random effects. We assume the econometrician observes the first period in which each household becomes eligible for the program.[6] We relabel this first eligible period as $t = 1$ and augment each household history with $t = 0$, when the household was not yet eligible. This implies the fixed initial condition $D_{i0} = 0$.

3.2 Full Model vs Preliminary Model

The preliminary model treats the lagged decision D_{it-1} as exogenous, whereas the full model allows D_{it-1} to be endogenous. Assumption 2.1 requires both unobserved terms, $\mathcal{E}_{it}$ and Ξ_{it}, to be uncorrelated with D_{it-1}. In contrast, the full model allows persistent household-level heterogeneity Q_i to influence both the attention and choice utilities in every period, thereby

[6] This assumption is satisfied in our empirical application. NLSY97 follows individuals from adolescence into adulthood and therefore we observe the first period when respondents become eligible for WIC as adults.

affecting take-up decisions across *all* periods. In other words, this design permits correlation between D_{it-1} and the unobserved term, Q_i, for all $t > 1$.

A second difference concerns the relation between attention and choice. Under Assumption 2.1, A_{it} and C_{it} are conditionally independent given X_{it} and D_{it-1}, since their shocks are independent. This independence can be unrealistic: thriftier households (higher Q_i) may both pay more attention to the program and be more inclined to enroll once attentive. The full model captures this by letting Q_i enter both U_{it}^a and U_{it}^c, generating correlation between A_{it} and C_{it} even after conditioning on observables.

Together, these modifications relax the strong exogeneity and independence restrictions of the preliminary model, yielding a more realistic framework in which persistent unobserved heterogeneity drives both dynamics across time and dependence across stages.

3.3 Identification

In the preliminary model, we identify β first using the probability of a household staying in the program and identify γ by comparing the probability of starting program participation and the probability of staying in the program. We show that under the full model, β and γ are identifiable with the same intuition. Last period participation induces full attention in the current period, making the probability of staying a function of choice parameters only. Then, comparing the probability of staying and the probability of starting participation identifies attention parameters. The key difference in the identification strategies of the two models is that instead of analyzing the probability of a single-period transition, we analyze the probability of a take-up decision *sequence*.

Assuming that a population of households which are eligible for a welfare program for T periods. We observe the joint distribution of any take-up sequence $D_{i,1:T} \in \{0,1\}^T$ for and any characteristics sequence $\{X_{i,1:T}, Z_{i,1:T}\} = \{\mathcal{X} \times \{0,1\}\}^T$. Under Assumption 3.1, the random effect can be integrated out of this probability over its marginal distribution.

$$P(D_{i,1:T} \mid X_{i,1:T}, Z_{i,1:T}) = \int P(D_{i,1:T} \mid X_{i,1:T}, Z_{i,1:T}, Q_i = q) dF(q)$$

Once conditional on Q, the probability of the take-up sequence can be decomposed as a product of T single-period transition probabilities.

Lemma 3.1. *Under Assumption 3.1,*

$$P\left(D_{i,1:T} \mid X_{i,1:T}, Z_{i,1:T}, Q_i\right) = \prod_{t=1}^{T} P(D_{it} \mid D_{it-1}, X_{it}, Z_{it}, Q_i)$$

The proof of Lemma 3.1 can be found in Appendix B. One thing to note is that such decomposition does not work without conditioning on Q_i because D_{it-1} is endogenous in general for $t > 1$. $\{D_{is}\}_{s=1}^{T-1}$ correlates with Q_i when σ_1 and/or σ_2 is nonzero.

Consider an infinitesimally small increase in one of the covariates at period τ, X_τ^ω, where ω represents which covariate we vary. τ is a fixed integer, different from t, which is a running index. Given the joint distribution of $D_{i,1:T}$ and $\{X_{i,1:T}, Z_{i,1:T}\}$, we observe $\frac{\partial}{\partial X_{i\tau}^\omega} P(D_{i,1:T} \mid$

$X_{i,1:T}, Z_{i,1:T})$. Leveraging the product structure in Lemma 3.1, we decompose this observed first-order derivative (FOD).

Lemma 3.2. *Under Assumption 3.1,* $\frac{\partial}{\partial X^{\omega}_{i\tau}} P(D_{i,1:T} \mid X_{i,1:T}, Z_{i,1:T})$ *is*

$$\int \underbrace{\prod_{t=1}^{T} P(D_{it} \mid D_{it-1}, X_{it}, Z_{it}, Q_i = q)}_{CPL(D_{i,1:T}, X_{i,1:T}, Z_{i,1:T}, Q_i=q)} \underbrace{\frac{\frac{\partial}{\partial X^{\omega}_{i\tau}} P(D_{i\tau} \mid D_{i\tau-1}, X_{i\tau}, Z_{i\tau}, Q_i = q)}{P(D_{i\tau} \mid D_{i\tau-1}, X_{i\tau}, Z_{i\tau}, Q_i = q)}}_{SE^{\omega}(D_{i\tau}, D_{i\tau-1}, X_{i\tau}, Z_{i\tau}, Q_i=q)} dF(q). \tag{3.3}$$

where CPL stands for conditional probability level and SE stands for semi-elasticity.[7]

Assumption 3.1 does not impose location or scale normalization, so we can freely normalize one β and one γ. With little loss of generalization on the sign, we normalize that $\beta^B = 1$ and $\gamma^B = 1$ where B is the benefit dollar amount. The loss of generality is that we restrict the sign of B to be positive for both attention and choice. Such a restriction is mild as the benefit amount is likely to have a positive influence on attention and choice utilities.

Identification of β

In the preliminary model, we set $D_{it-1} = 1, Z_{it} = 0$ to skip over the attention stage and hence, to identify β^v and β^X. We will leverage the same intuition for the full model. Consider $\tilde{d} = \{\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_{\tau-2}, 1, 1, \tilde{d}_{\tau+1}, \ldots, \tilde{d}_T\}$ where $\tilde{d}_{\tau-1} = \tilde{d}_\tau = 1$ and the rest of $\tilde{d}_t \in \{0,1\}$; $\tilde{z} = \{\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_{\tau-2}, \tilde{z}_{\tau-1}, 0, \tilde{z}_{\tau+1}, \ldots, \tilde{z}_T\}$ where $\tilde{z}_\tau = 0$ and the rest of $\tilde{z}_t \in \{0,1\}$. Here, instead of considering a single period transition of staying in the program as in the preliminary model, we consider a transition of staying in the program *within a sequence* of take-up decisions. At period τ, a household skips over the attention stage and hence, the period-τ semi-elasticity for $\tilde{d}_{t-1} = \tilde{d}_t = 1$ should be a function of β only.

$$SE^{\omega}(1, 1, X_{i\tau}, 0, Q_i = q) = \frac{h(U^c_{i\tau}(1, X_{it}, 0, q))\beta^{\omega}}{H(U^c_{i\tau}(1, X_{it}, 0, q))}$$

where $h(U^c_{it}(1, X_{it}, 0, q)) = \frac{\partial H(U^c_{it}(1, X_{it}, 0, q))}{\partial U^c_{it}(1, X_{it}, 0, q)}$. Substituting this semi-elasticity expression into Equation (3.3), we obtain an identifying equation

$$\frac{\partial}{\partial X^{\omega}_{i\tau}} P(D_{i,1:T} = \tilde{d} \mid X_{i,1:T}, Z_{i,1:T} = \tilde{z}) = \int CPL(\tilde{d}, X_{i,1:T}, \tilde{z}, q) \frac{h(U^c_{i\tau}(1, X_{it}, 0, q))}{H(U^c_{i\tau}(1, X_{it}, 0, q))} dF(q)\beta^{\omega}$$

The integral term on the right hand side does not depend on the choice of ω, i.e., which variable we vary. Hence, by normalizing $\beta^B = 1$ and varying benefit amount at period τ, we can pin down the integral as $\frac{\partial}{\partial X^B_{i\tau}} P(D_{i,1:T} = \tilde{d} \mid X_{i,1:T}, Z_{i,1:T} = \tilde{z})$. With the knowledge of the integral, we can identify β^{ω} in general by varying other choices of ω.

In the following theorems, we simplify the expression $X_{i,1:T} = x$ to x to make our results more compact. The simplification should not cause confusion since the alphabet used for the

[7] The term semi-elasticity is borrowed from the consumer demand estimation literature. It means the FOD of the market share divided by the market share. In our context, take-up probability is equivalent to the market share concept in consumer demand estimation literature.

random variable $X_{i,1:T}$ matches the alphabet used for the running variable x. We make the same simplification to $D_{i,1:T} = d$ and $Z_{i,1:T} = z$.

Theorem 3.1. *Under Assumption 3.1, $\beta^\omega \in \beta^\upsilon \cup \beta^\chi$ are constructively identified as*

$$\beta^\omega = \frac{\mathbb{E}_{d\in\tilde{\mathcal{D}},x\in\mathcal{X},z\in\tilde{\mathcal{Z}}}\left[\frac{\partial}{\partial X^\omega_{i\tau}}P(d \mid x,z)\right]}{\mathbb{E}_{d\in\tilde{\mathcal{D}},x\in\mathcal{X},z\in\tilde{\mathcal{Z}}}\left[\frac{\partial}{\partial X^B_{i\tau}}P(d \mid x,z)\right]}$$

where $\tilde{\mathcal{Z}}$ is the set of all $Z_{i,1:T}$ with $z_\tau = 0$ and $z_t \in \{0,1\}$ for all $t \neq \tau$, $\tilde{\mathcal{D}}$ is the set of all $D_{i,1:T}$ with $d_{\tau-1} = d_\tau = 1$ and $d_t \in \{0,1\}$ for all $t \notin \{\tau-1,\tau\}$, and β^B is normalized to 1.

Theorem 3.1 identifies all β^υ and β^χ associated with some $X^\omega_{i\tau}$. Theorem 3.1 corresponds to part (1a) in Proposition 2.1. Similar to part (1b) in Proposition 2.1, we need some observed probabilities with $Z_{i\tau} = 1$ to identify β^κ for the full model.

Theorem 3.2. *Under Assumption 3.1, $\beta^\omega \in \beta^\kappa$ are constructively identified as*

$$\beta^\omega = \frac{\mathbb{E}_{d\in\tilde{\mathcal{D}},x\in\mathcal{X},z\in\tilde{\mathcal{Z}}}\left[\frac{\partial}{\partial X^\omega_{i\tau}}P(d \mid x,z)\right]}{\mathbb{E}_{d\in\tilde{\mathcal{D}},x\in\mathcal{X},z\in\tilde{\mathcal{Z}}}\left[\frac{\partial}{\partial X^B_{i\tau}}P(d \mid x,z)\right]}$$

where $\tilde{\mathcal{Z}}$ is the set of all $Z_{i,1:T}$ with $z_\tau = 1$ and $z_t \in \{0,1\}$ for all $t \neq \tau$, $\tilde{\mathcal{D}}$ is the set of all $D_{i,1:T}$ with $d_{\tau-1} = d_\tau = 1$ and $d_t \in \{0,1\}$ for all $t \notin \{\tau-1,\tau\}$, and β^B is normalized to 1.

Identification of γ

Following the intuition of part (2) of Proposition 2.1, we compare the probability of staying and the probability of starting. Consider two sequences, $\mathring{d} = \{\mathring{d}_1, \mathring{d}_2, \ldots, \mathring{d}_{\tau-2}, 1, 1, \ldots, \mathring{d}_T\}$ with $\mathring{d}_{\tau-1} = \mathring{d}_\tau = 1$ and $\mathring{d}_t \in \{0,1\}$ for all $t \notin \{\tau-1,\tau\}$ and $\ddot{d} = \{\ddot{d}_1, \ddot{d}_2, \ldots, \ddot{d}_{\tau-2}, 0, 1, \ldots, \ddot{d}_T\}$ with $\ddot{d}_{\tau-1} = 0, \ddot{d}_\tau = 1$ and $\ddot{d}_t \in \{0,1\}$ for all $t \notin \{\tau-1,\tau\}$. $\mathring{d}$ is associated with $\{\mathring{x}, \mathring{z}\}$ and $\ddot{d}$ is associated with $\{\ddot{x}, \ddot{z}\}$. Then, we compare two semi-elasticities for period τ

$$\begin{aligned} SE^\omega(1,1,\mathring{x}_\tau,\mathring{z}_\tau,q) &= \overbrace{\frac{h(U^c_{i\tau}(1,\mathring{x}_\tau,\mathring{z}_\tau,q))\beta^\omega}{H(U^c_{i\tau}(1,\mathring{x}_\tau,\mathring{z}_\tau,q))}}^{\text{Choice}} \\ SE^\omega(1,0,\ddot{x}_\tau,\ddot{z}_\tau,q) &= \underbrace{\frac{g(U^a_{i\tau}(\ddot{x}_\tau,q))\gamma^\omega}{G(U^a_{i\tau}(\ddot{x}_\tau,q))}}_{\text{Attention}} + \underbrace{\frac{h(U^c_{i\tau}(0,\ddot{x}_\tau,\ddot{z}_\tau,q))\beta^\omega}{H(U^c_{i\tau}(0,\ddot{x}_\tau,\ddot{z}_\tau,q))}}_{\text{Choice}} \end{aligned}$$

Under high-level assumptions H1 and H2, we can difference out the choice components in the two semi-elasticities.

Assumption (H1). *$CPL(\mathring{d},\mathring{x},\mathring{z},q) = CPL(\ddot{d},\ddot{x},\ddot{z},q)\ \forall\ q \in \mathbb{R}$.*

Assumption (H2). *$U^c_{i\tau}(1,\mathring{x}_\tau,\mathring{z}_\tau,q) = U^c_{i\tau}(0,\ddot{x}_\tau,\ddot{z}_\tau,q)\ \forall\ q \in \mathbb{R}$.*

Under Assumptions H1 and H2, $\frac{\partial}{\partial X^\omega_{i\tau}}P(D_{i,1:T} = \ddot{d} \mid X_{i,1:T} = \ddot{x}, Z_{i,1:T} = \ddot{z}) - \frac{\partial}{\partial X^\omega_{i\tau}}P(D_{i,1:T} = \mathring{d} \mid X_{i,1:T} = \mathring{x}, Z_{i,1:T} = \mathring{z}) = \int CPL(\ddot{d},\ddot{x},\ddot{z},q)\frac{g(U^a_{i\tau}(\ddot{x}_\tau,q))}{G(U^a_{i\tau}(\ddot{x}_\tau,q))}dF(q)\gamma^\omega$. Using the same normalization argument that we use for β identification, we obtain the following results for γ identification.

Theorem 3.3. *Under Assumption 3.1, γ is constructively identified as*

$$\gamma^{\omega} = \frac{\mathbb{E}_{\{\ddot{x},\ddot{z},\mathring{x},\mathring{z}\}\in\mathcal{M}}\left[\mathbb{E}_{\ddot{d}\in\ddot{\mathcal{D}}(\ddot{x},\ddot{z},\mathring{x},\mathring{z})}\left[\frac{\partial}{\partial X^{\omega}_{i\tau}}P(\ddot{d}\mid\ddot{x},\ddot{z})\right]-\mathbb{E}_{\mathring{d}\in\mathring{\mathcal{D}}(\ddot{x},\ddot{z},\mathring{x},\mathring{z})}\left[\frac{\partial}{\partial X^{\omega}_{i\tau}}P(\mathring{d}\mid\mathring{x},\mathring{z})\right]\right]}{\mathbb{E}_{\{\ddot{x},\ddot{z},\mathring{x},\mathring{z}\}\in\mathcal{M}}\left[\mathbb{E}_{\ddot{d}\in\ddot{\mathcal{D}}(\ddot{x},\ddot{z},\mathring{x},\mathring{z})}\left[\frac{\partial}{\partial X^{B}_{i\tau}}P(\ddot{d}\mid\ddot{x},\ddot{z})\right]-\mathbb{E}_{\mathring{d}\in\mathring{\mathcal{D}}(\ddot{x},\ddot{z},\mathring{x},\mathring{z})}\left[\frac{\partial}{\partial X^{B}_{i\tau}}P(\mathring{d}\mid\mathring{x},\mathring{z})\right]\right]}$$

where $\mathcal{M} := \{\ddot{x},\ddot{z},\mathring{x},\mathring{z} \mid \textit{H2 is true}\}$*,* $\ddot{\mathcal{D}}(\ddot{x},\ddot{z},\mathring{x},\mathring{z}) := \{\ddot{d} \mid \ddot{d}_{\tau-1} = 0, \ddot{d}_{\tau} = 1, \exists\ \mathring{d} \textit{ with } \mathring{d}_{\tau-1} = 1, \mathring{d}_{\tau} = 1, \textit{such that H1 is true}\}$*, and* $\mathring{\mathcal{D}}(\ddot{x},\ddot{z},\mathring{x},\mathring{z}) := \{\mathring{d} \mid \mathring{d}_{\tau-1} = 1, \mathring{d}_{\tau} = 1, \exists\ \ddot{d} \textit{ with } \ddot{d}_{\tau-1} = 0, \ddot{d}_{\tau} = 1, \textit{such that H1 is true}\}$*,* γ^{ω} *is normalized to 1.*

Theorem 3.3 requires high-level assumptions and does not directly prove the identifiability of the model. Next, we show that there exists some sequence of $\{\mathring{d},\mathring{x},\mathring{z}\}$ and $\{\ddot{d},\ddot{x},\ddot{z}\}$ that satisfy the high-level assumptions H1 and H2. Consider a population of households that are eligible for a welfare program for four periods. $\{\mathring{d},\mathring{x},\mathring{z}\}$ and $\{\ddot{d},\ddot{x},\ddot{z}\}$ are specified as follows:

t	0	1	2	3	4
$\mathring{d}$	0	1	1	0	0
$\mathring{x}$		x^*	x^*	x^*	x^*
$\mathring{z}$		0	1	0	0

t	0	1	2	3	4
$\ddot{d}$	0	0	1	1	0
$\ddot{x}$		x^*	x^*	x^*	x^*
$\ddot{z}$		0	0	1	0

We deliberately set $X_{i,1:4}$ as one time-invariant, x^*, to simplify our argument; however, such a setup is not necessary. We first show $\{\mathring{d},\mathring{x},\mathring{z}\}$ and $\{\ddot{d},\ddot{x},\ddot{z}\}$ satisfy H1. Under $\{\mathring{d},\mathring{x},\mathring{z}\}$, a household experiences sequentially one sign-up ($D_{it-1} = 0, D_{it} = 1$), one stay in the program with a recertification ($D_{it-1} = 1, D_{it} = 1, Z_{it} = 1$), one exit ($D_{it-1} = 1, D_{it} = 0$), one stay out of the program ($D_{it-1} = 0, D_{it} = 0$). Under $\{\ddot{d},\ddot{x},\ddot{z}\}$, a household experiences sequentially one stay out of the program ($D_{it-1} = 0, D_{it} = 0$), one sign-up ($D_{it-1} = 0, D_{it} = 1$), one stay in the program with a recertification ($D_{it-1} = 1, D_{it} = 1, Z_{it} = 1$), one exit ($D_{it-1} = 1, D_{it} = 0$). Given that $x_t = x^*$ for all t for $\{\mathring{d},\mathring{x},\mathring{z}\}$ and $\{\ddot{d},\ddot{x},\ddot{z}\}$. These two sequences have the same product of probability conditional on Q_i

$$\prod_{t=1}^{4} P(D_{it} = \mathring{d}_t | \mathring{x}_t, \mathring{z}_t, Q_i = q) = \prod_{t=1}^{4} P(D_{it} = \ddot{d}_t | \ddot{x}_t, \ddot{z}_t, Q_i = q)$$

By Lemma 3.1, $\{\mathring{d},\mathring{x},\mathring{z}\}$ and $\{\ddot{d},\ddot{x},\ddot{z}\}$ share the same $CPL(D_{i,1:4}, X_{i,1:4}, Z_{i,1:4}, Q_i)$, satisfying H1. Now, we verify H2. Under $\{\mathring{d},\mathring{x},\mathring{z}\}$, $\tau = 2$, a household needs to sign up and hence needs to pay a hassle cost to participate in the welfare program. Under $\{\ddot{d},\ddot{x},\ddot{z}\}$, a household needs to recertify eligibility and hence needs to pay a hassle cost to continue participating in the welfare program. Given the same x^*, $U^c_{i2}(X_{i2} = \mathring{x}_2, Z_{i2} = \mathring{z}_2, Q_i = q) = U^c_{i2}(X_{i2} = \ddot{x}_2, Z_{i2} = \ddot{z}_2, , Q_i = q) = x^*\beta$, satisfying H2.

Note that all identification results are semiparametric as $\{F, G, H\}$ are not parametrically restricted and are with respect to X variables that have variations both within and across households. For variables that do not vary within or across households, model identifiability relies on the specific parametric assumptions imposed on $\{F, G, H\}$. Nevertheless, the three theorems provide insights into why key structural parameters of the model, such as the influence of household demographics, income, local accessibility on attention and choice, can be learned through a standard panel dataset despite covariates simultaneously entering the attention and choice stages.

Identification of $\{\sigma_1, \sigma_2, \sigma_G, \sigma_H\}$

We intuitively explain the identification of $\{\sigma_1, \sigma_2, G, H\}$ under parametric assumptions. Assuming that all $\{F, G, H\}$ are normally distributed, we only need to identify the means and variances of the three distributions. Given that we include an intercept term in X_{it}, the means of $\{F, G, H\}$ are zero. We only need to identify their variances. Additionally, since Q_i is scaled by σ_1, σ_2, we aim to identify $\{\sigma_1, \sigma_2, \sigma_G, \sigma_H\}$.

σ_1 and σ_2 are identified from cross-household variation. Assuming that $\sigma_1, \sigma_2 > 0$. Note that this assumption aligns with reality. A thriftier household is not only more attentive but also has a higher choice tendency. Very heterogeneous and persistent outcomes among households that share similar observed characteristics indicate that household-level heterogeneity has a large variance. If there is a large variance in the waiting time between being eligible and starting participation, then σ_1 is large. If there is a large variance in the participation duration, then σ_2 is large. On the other hand, if there is little difference between take-up behaviors among households with similar observed characteristics, then household-level heterogeneity has a small variance. The identification of σ_1 and σ_2 follows as discussed previously.

σ_G and σ_H are identified through within-household comparisons. Our data exhibits high persistence within households along two dimensions: first, households that are already in the program are extremely likely to continue participation; second, households that are not currently enrolled in the program are extremely unlikely to start participation. The first dimension shows that there is little variation in the exogenous part of the choice stage utility shock, ϵ_{it}. In other words, σ_H is small. The second dimension shows that there is little variation in the exogenous part of the attention stage utility shock, ξ_{it}. In other words, σ_G is small.

3.4 Model extension

Q_i^a and Q_i^c: In our full model, we assume an identical random effect for the attention and choice stages. This assumption can be relaxed. Consider bivariate iid $\{Q_i^a, Q_i^c\}$ and that they are independent from $X_{1,1:T}$ and $Z_{i,1:T}$, though Q_i^a and Q_i^c do not be mutually independent from each other. This relaxation of identical Q_i for the two stages changes Equation (3.3) from a single integral to a double integral. This change does not alter our identification proof.

Different sign-up and recertification: Our model assumes that recertification and sign-up share the same hassle; therefore, we only need a recertification indicator Z_{it} to guarantee identifiability. This assumption is more likely to hold for programs with administratively identical recertification and sign-up. In some programs, recertification is administratively easier than sign-up. In this case, having a sign-up/recertification complexity measure as proposed in Kleven and Kopczuk (2011) is sufficient for our identification argument to work. For example, we can use the average number of trips households need to take for recertification and sign-up in a given region and a given year. As long as there is overlap between the distributions of the recertification complexity and that of sign-up complexity, then we can condition on the same value for the complexity measures to identify the γ.

Household type fixed effect: Economists may find the orthogonality assumption between the household-level unobserved heterogeneity and observed characteristics too strong. One way to relax this assumption is to allow household-type fixed effects. Many survey datasets, e.g., NLSY97, include households' participation status for *multiple* programs. Suppose the economists are interested in one particular program. In that case, they can use other programs' participation status information to cluster households into types and augment our model with households' latent types fixed effects. This augmentation decomposes the household-level unobserved heterogeneity into a household-type fixed effect that can arbitrarily correlate with observed characteristics and a household-level random effect that is orthogonal to the observed characteristics.

Latent group structure: Households have heterogeneous preferences. Our full tries to capture such heterogeneity through Q_i. Such a random-effect design allows for household-level heterogeneity that is not systematically different across household characteristics. To further allow heterogeneity that can vary with group characteristics, we borrow Bonhomme, Lamadon and Manresa (2022)'s algorithm to uncover the latent group structures in our sample. Taking this latent group structure as the ground truth, the identification argument made in this paper carries through trivially. It remains an interesting future research direction whether the statistical inference results from Bonhomme, Lamadon and Manresa (2022) carry over to our model setup if we augment our full model with a group-fixed-effect design.

3.5 Estimation

We propose an MLE for the model, which maximizes the probability of observing the *sequence* of welfare take-up decisions for all households $\mathcal{I}$ in the sample, $P(D_{i,1:T_i}, X_{i,1:T_i}, Z_{i,1:T_i})$, where T_i is the number of periods that households i is eligible for the welfare program. Since $\{X_{i,1:T_i}, Z_{i,1:T_i}\}$ are strictly exogenous, then equivalently, we maximize the probability of

$$\prod_{i\in\mathcal{I}} P(D_{i,1:T_i} \mid X_{i,1:T_i}, Z_{i,1:T_i}) \tag{3.4}$$

where a sequence of take-up history for household i is denoted as $D_{i,1:T_i} := \{D_{i1}, \dots, D_{iT_i}\}$. $X_{i,1:T_i}$ and $Z_{i,1:T_i}$ are defined similarly to $D_{i,1:T_i}$: it is the history of the strictly exogenous control variables. The derivation of the MLE involves three steps of transforming the likelihood function Equation (3.4). We illustrate a rough sketch of how to decompose the likelihood into an expression that is possible to numerically optimize.

$$
\begin{aligned}
&\log\left(\prod_{i\in\mathcal{I}} P(D_{i,1:T_i} \mid X_{i,1:T_i}, Z_{i,1:T_i})\right) \\
&= \sum_{i\in\mathcal{I}} \log \int P(D_{i,1:T_i} \mid X_{i,1:T_i}, Z_{i,1:T_i}, Q_i = q) dF(q) && \text{Step 1} \\
&= \sum_{i\in\mathcal{I}} \log \int \prod_{t=1}^{T_i} P(D_{it} \mid D_{it-1}, X_{it}, Z_{it}, Q_i = q) dF(q) && \text{Step 2} \\
&= \sum_{i\in\mathcal{I}} \log \int \prod_{t=1}^{T_i} (P_a(D_{it-1}, X_{it}, Q_i) P_c(D_{it-1}, X_{it}, Z_{it}, Q_i))^{D_{it}} \\
&\qquad (1 - P_a(D_{it-1}, X_{it}, Q_i) P_c(D_{it-1}, X_{it}, Z_{it}, Q_i))^{1-D_{it}}\, dF(q) && \text{Step 3}
\end{aligned}
$$

We explain the validity of these three steps in Appendix B.4. In the main text, we formalize the following claim. Once we specify the parametric distribution of the attention and choice utility shocks $\{\epsilon_{it}, \xi_{it}\}$, the likelihood can be optimized numerically. Under Assumption 3.1, the MLE of the likelihood of all households' take-up history $\prod_{i\in\mathcal{I}} P(D_{i,1:T_i} \mid X_{i,1:T_i}, Z_{i,1:T_i})$ is

$$
\begin{aligned}
\{\hat{\gamma}, \hat{\beta}, \hat{\sigma}\} = \underset{\gamma,\beta,\sigma}{\operatorname{argmax}} \sum_{i\in\mathcal{I}} \log \int \prod_{t=1}^{T_i} &\left[G(X_{it}\gamma + \sigma_1 q)^{1-D_{it-1}} H(X_{it}^{\upsilon}\beta^{\upsilon} + S_{it} X_{it}^{\kappa}\beta^{\kappa} + X_{it}^{\chi}\beta^{\chi} + \sigma_2 q) \right]^{D_{it}} \\
&\left[1 - G(X_{it}\gamma + \sigma_1 q)^{1-D_{it-1}} H(X_{it}^{\upsilon}\beta^{\upsilon} + S_{it} X_{it}^{\kappa}\beta^{\kappa} + X_{it}^{\chi}\beta^{\chi} + \sigma_2 q) \right]^{1-D_{it}} dF(q)
\end{aligned}
$$

where $\beta := \{\beta^{\upsilon}, \beta^{\kappa}, \beta^{\chi}\}$, $\sigma := \{\sigma_1, \sigma_2\}$.

For estimation, we set $\{F, G, H\}$ as independent standard normal. The computational challenge of the proposed MLE is that it involves an integration without a closed form. Following Heiss et al. (2021), we use Gaussian-Hermite quadrature to approximate the integral and solve the maximization problem numerically.

4 Data and institutional details

We use the National Longitudinal Survey of Youth 1997 (NLSY97) and its confidential geocode information. Using NLSY97 data merged with county-level WIC accessibility information, this section presents empirical evidence to explain our approach to managing various variables and to motivate our model setup in Section 2 and Section 3.

We select WIC-eligible individuals from the NLSY97 using two legislative requirements: category and income. *Categorical requirement:* To be eligible for the program, a household should have a woman who is either pregnant or 6 weeks after the birth of an infant, or in postpartum up to six months after the birth of the infant or the end of the pregnancy, or breastfeeding, up to the infant's first birthday. In addition, infants up to their first birthday and children up to their fifth birthday are eligible. *Income requirement:* To qualify for the WIC benefit, household income should be below 185 percent of the federal poverty rate. Additionally, households participating in other income-eligible government programs, including SNAP (formerly Food Stamps), Medicaid, and Temporary Assistance for Needy Families (formerly Aid to Families with Dependent Children, or AFDC in short) are automatically eligible for WIC. We select WIC-eligible households from 2000 through 2009 based on the categorical and income requirements since we observe the family structure and annual household income.[8]

We end up with a sample of 3486 unique households and 159199 household-month-level observations. We use this sample for the preliminary model and present the descriptive data of this sample in the rest of this section. For the full model, we drop those household-eligibility-periods which we do not observe the full history. The full model sample has 3229 households and 5004 unique household-eligibility-period combinations, amounting to 131382 household-month-level observations. The descriptive data of the two samples are almost the same.

4.1 Summary statistics

Figure 1 shows that the overall WIC take-up rate is around 50% over our sample period. The trend is generally downward, except after the 2007 food package revision. We give more details about the food package revision in Appendix C. Moreover, the take-up rate among black households is below the national average. The take-up rate among Hispanic households is higher than the national average. All these features of our sample largely align with the national average data from various sources. For example, using statistics from Table 6 in Bitler, Currie and Scholz (2003), the take-up is around 48.5% among eligible households, Swann (2010) recorded a take-up around 55%, USDA also reports a close to 55% during the study period. Table 1 shows that the households with infant has a take-up rate of 59% whereas households without infant has a take-up rate around 41%. This algins with Kreider, Pepper and Roy (2016) whose the corresponding take-up rates in a largely overlapping study period from 1999 to 2008 is 62.9% and 39.2%, see Table 1 of Kreider, Pepper and Roy (2016).

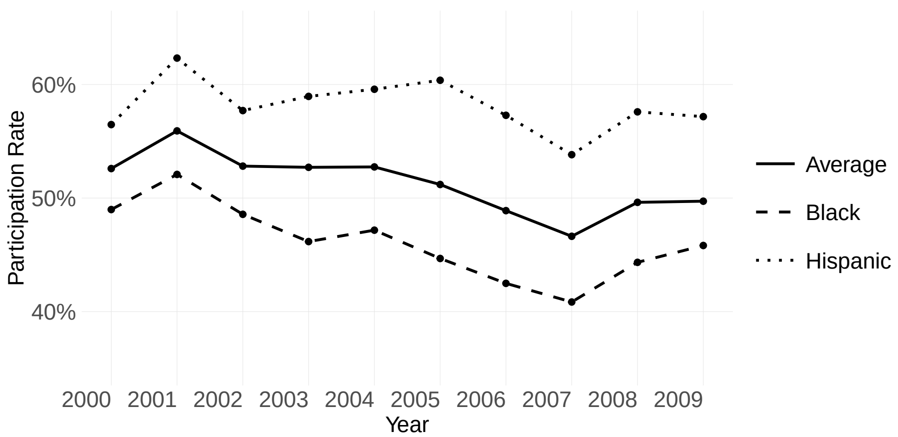


Figure 1: Overall participation rate by year shows a general decreasing trend except for after the 2007 food package revision.

[8]Our sample period ends in 2009 because NLSY stops surveying about welfare program participation in 2010. We do not include Medicaid status in our sample selection procedure. The NLSY97 survey asked respondents about welfare medical insurance coverage since 2005, which does not cover a significant portion of our data. Additionally, the survey question is not specific about Medicaid, but any welfare medical insurance program in general. In addition, we disregard the nutritional risk requirement following Bitler, Currie and Scholz (2003).

Panel A of Table 1 shows that among the eligible households, our sample contained more female family members than male family members. This aligns with the categorical requirement. A single pregnant mother satisfies the categorical requirement, whereas a single father whose child is not yet born is not eligible for WIC. Participating households tend to enjoy higher benefits, have more children (below the age of 5), and more infants (below the age of 1). The average education and LA_{it} (an accessibility measure defined in Section 4.5) are very close for participants and non-participants. We will show, later in this section, that these two variables do differ substantially once we condition on other observed characteristics of the households.

Panel A: Participants vs non-participants

	Overall	$D_{it}=1$	$D_{it}=0$
% Female	0.65	0.75	0.55
% Black	0.41	0.36	0.45
% Hispanic	0.26	0.30	0.22
Benefit ($)	105.81	112.48	98.89
# children	1.38	1.44	1.33
# Infant	0.59	0.69	0.49
Education	1.72	1.74	1.71
LA	12.80	12.78	12.83
Income ($)	7688.82	8062.72	7300.69
Sample	159199	81087	78112

Panel B: Participation rates by groups

	$D_{it}=1$	Sample
$D_{it-1}=1$	0.97	78815
$D_{it-1}=1, Z_{it}=0$	0.97	75954
$D_{it-1}=1, Z_{it}=1$	0.93	2861
$D_{it-1}=0$	0.04	74789
With infant	0.59	71886
No infant	0.41	87313
No infant × <HS	0.36	16838
No infant × HS	0.42	36823
No infant × >HS	0.47	3041

Table 1: Descriptive Statistics of the sample: Panel A compares the average characteristics of participants and non-participants. Panel B presents the average take-up rate for different groups. Education is divided into three groups: less than high school (<HS), high school (HS), more than high school (>HS). We encode them as 1,2,3, respectively. Income is the annual income per household member.

4.2 High persistence of participation status

As shown by Panel B of Table 1, households exhibit a tremendous amount of behavioral inertia: more than 95% of the time, households choose to stick with their previous period decision, D_{it-1}. $P(D_{it}=1 \mid D_{it-1}=1) > 0.95$ and $P(D_{it}=0 \mid D_{it-1}=0) > 0.95$.

In our setup, paying attention is a necessary condition for welfare program participation. Hence, the extreme persistence from $1 \to 1$ transition shows that households that were in the program last period pay attention at least 97% of the time. This strongly supports our setup that participating households are fully attentive next period.

A household with $D_{it-1}=0$ participates in period t with a probability less than 5%. We compare this figure with another fully attentive group, but also needs to pay a hassle cost, $D_{it-1}=1, Z_{it}=1$, which has a take-up rate of more than 90%. This strongly suggests that a large percentage of households that did not participate in the last period do not pay attention to WIC, supporting our stochastic attention setup for those households with $D_{it-1}=0$.

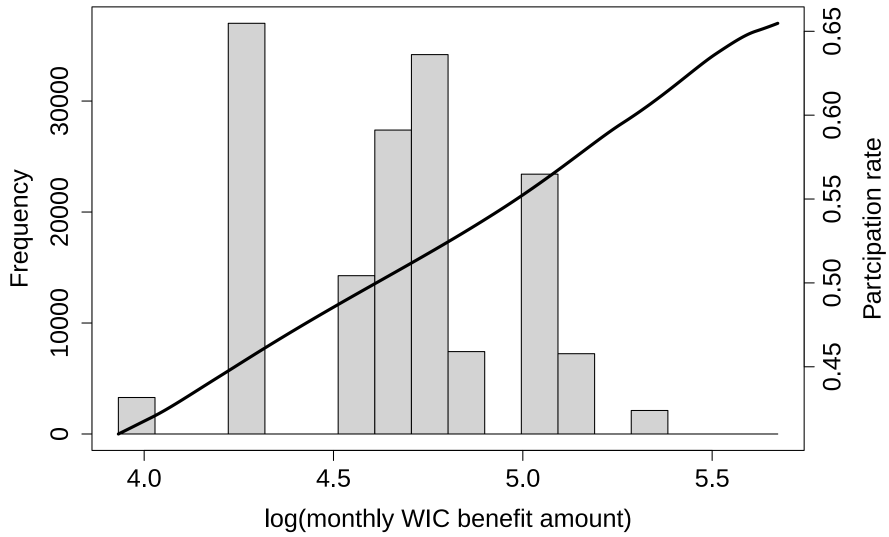


Figure 2: In the histogram of log imputed benefit $\log(B_{it})$, all bins with sample size fewer than 2000 are redacted for data confidentiality. Participation increases steadily with $\log(B_{it})$.

4.3 Benefit imputation

Studies of welfare take-up inherently suffer from a missing data problem. Eligible households who choose not to participate in the program do not report how much they *potentially* can receive from the welfare program. Additionally, some households participating in the welfare program misreport benefit amounts. We provide an imputation procedure for computing the benefit amount, assuming that the economists know which variables affect the (potential) benefit a household receives. We explain benefit imputation details in Appendix C.

Figure 2 depicts the histogram of the log imputed benefit $\log(B_{it})$. There are two points to note: first, there are substantial variations in the benefit amount households are eligible for; second, the rate of participation increases steadily with $\log(B_{it})$, which aligns with intuition.

4.4 Recertification process

Like many means-tested programs, WIC requires households to recertify their eligibility regularly. Administratively, the recertification is identical to the signup process. A household should visit the doctor's office for a health checkup and verify income eligibility. Then, the participants have an appointment with the WIC office to recertify their eligibility. There are two conditions under which households have to recertify their eligibility: (i) households are required to recertify their eligibility after 12 consecutive months of participation, and (ii) households with an infant below the age of 12 months are required to recertify their eligibility when the baby reaches 13 months old.

$Z_{it} = 1$ denotes that the household i is required to recertify their eligibility at period t when they are already enrolled $D_{it-1} = 1$. Denote the youngest child's age in household i at period t as YA_{it}. We define $Z_{it} = \mathbb{1}\{YA_{it} \in \{1, 13, 25, 37\}\}$. In words, it means that the household was in WIC last month $t-1$, and their youngest kid reached the age in the specified set. Next, we justify this definition of Z_{it}.

Figure 3(a) shows that the vast majority of the participants join the program when their youngest children are below the age of 12 months. Moreover, many of them join at the beginning of pregnancy, and there is a spike in the number of participants joining WIC when the infants are just born (one month old shown as the pink column in Figure 3(a)). Therefore, given recertification conditions (i) and (ii), most of the households would have their first recertification period when $YA_{it} = 1$ or 13. If the household does not have a newborn anymore, then they would need to recertify their eligibility when $YA_{it} = 25$ or 37; if the household has a newborn, then it would reset its recertification period counting based on YA_{it}. The reset is reasonable because it is in the interest of both the administrators and the participants to adopt the practice of signing up/recertifying multiple kids simultaneously (fewer signup/recertification appointments).

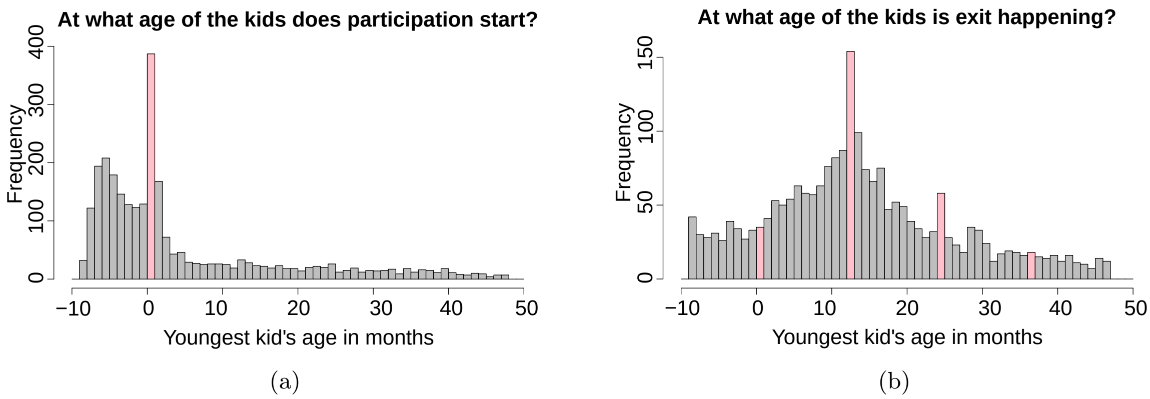


Figure 3: In the first panel, the pink bin is month 1 when the baby is just born. In the second panel, the pink bins are when the youngest kid is $\{1, 13, 25, 37\}$ months old. The pink bins are the periods for which we define $R_{it} = 1$.

The definition of Z_{it} is further supported by Figure 3(b). There are two spikes in the number of exits in months 13 and 25. In these two months, households face additional recertification hassle cost and hence, are more likely to exit. We argue that Z_{it} exclusively shifts whether the households need to pay for a sign-up hassle, and does not affect anything else after controlling for other variables related to YA_{it}.

4.5 County-level accessibility and (re)-sign-up hassle κ_{it}

Current participants go straight into the choice stage and evaluate whether they should participate in the welfare program or not. Moreover, at the choice stage, participating households periodically face recertification costs, which informs us about hassle. To model hassle cost, we construct a county-level local accessibility variable as follows.

By exploiting the geographic variation in the NLSY geocode data, we study how much county-level accessibility to WIC reduces hassle. To consider the impact of social resources on participation in government welfare programs, we merge the NLSY97 to the number of outpatient care centers (NAICS CODE 6214), social assistance establishments (NAICS CODE 624), urban transit facilities (NAICS CODE 4851), and grocery and convenience retailers stores (NAICS CODE 4451) from county-level Quarterly Census of Employment and Wages data by the U.S.

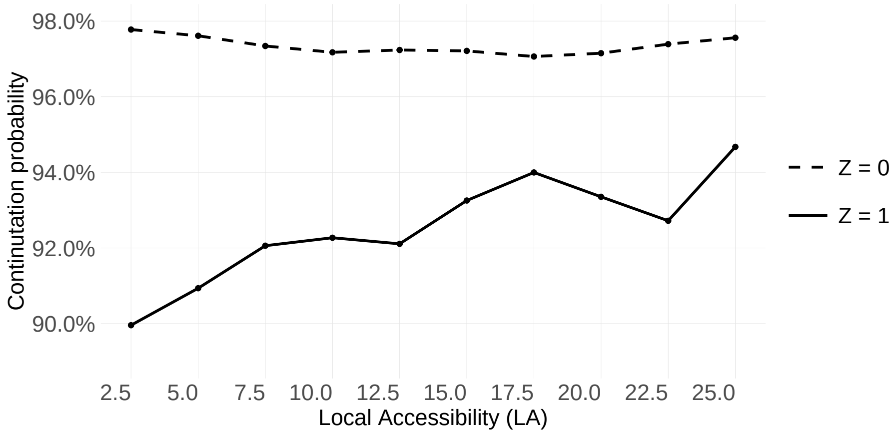


Figure 4: The continuation probability for those with $R_{it} = 1$ and those with $R_{it} = 0$.

Bureau of Labor Statistics. The county-level local accessibility variable is defined as

$$LA_{it} := \log(NC_{it,6214} + 1) + \log(NC_{it,624} + 1) + \log(NC_{it,4851} + 1) + \log(NC_{it,4451} + 1).$$

Figure 4 shows that the LA_{it} is associated with a higher probability of recertification rate, $P(D_{it} = 1 \mid D_{it-1} = 1, Z_{it} = 1)$, but not with a the probability of continuing participation when recertification is not required, $P(D_{it} = 1 \mid D_{it-1} = 1, Z_{it} = 0)$. As LA_{it} increases from 0 to 25, the continuation rate for those with $Z_{it} = 1$ increases steadily from 90% to 94%, whereas for those with $Z_{it} = 0$, the continuation rate remains virtually constant at slightly below 98%.

4.6 Education

Panel A of Table 1 shows that the average education levels among participating households and non-participating households are very close. Panel B shows that once we zoom into households without infants, higher education is associated with a higher participation rate. There can be two reasons. One, the benefit amount of WIC sharply decreases once the infant ages from 12 months to 13 months. Higher education households may be more health-conscious and hence value the benefit more than less-educated households after the benefit drop, i.e., higher utility of using the benefit even though the benefit amount is equal. Two, the benefit structure becomes more complex once the infant ages from 12 months to 13 months. Higher education households may be better at navigating the more complex benefit structure, i.e., lower usage cost. Since we model benefit and usage cost every period, these two interpretations result in isomorphic choice stage specifications.

5 Estimation results and policy counterfactual

We implement both the preliminary model and the full model with the NLSY data. In this section, we present the results of only the full model. Readers can find the preliminary model results in Appendix D.

5.1 Estimation results

Table 2 summarizes the model fitting results of the full model. All results are stable across different specifications. We will highlight the qualitative findings in all specifications and numerically interpret specification (6), which uses the benefit amount in dollars B_{it}. This makes the interpretation of the choice stage much more straightforward. On the other hand, using $\log(B_{it})$ is likely more appropriate since raw B_{it} is skewed.

Specification (1) presents our baseline model, which includes $\log(B_{it})$, an indicator for households without infants $\mathbb{1}\{YA_{it} > 12\}$, local accessibility LA_{it}, and education. We find that most covariates influence attention and choice probabilities in the same direction. Higher benefit levels are associated with increased probabilities of both attention and choice. Conversely, households without infants exhibit sharply lower attention and choice utilities. Two variables, however, exert opposing effects across stages. Local accessibility LA_{it} has little impact on attention probability but significantly influences choice: while the direct effect of hassle (S_{it}) is negative—indicating substantial hassle costs—this effect is mitigated by higher local accessibility, as shown by the positive coefficient on $S_{it} \times LA_{it}$. Education also displays differential effects: it has little impact on attention but significantly increases choice probability. These findings are robust throughout all specifications.

Specification (2) adds SNAP into the model. SNAP has a significantly positive impact on attention probability. Social workers of SNAP likely inform SNAP participants whom they deem to satisfy the categorical eligibility about WIC. On the other hand, SNAP participation does not influence choice utility significantly. We provide two opposing mechanisms that potentially explain it. First, SNAP participation reduces the hassle cost for WIC sign-up as SNAP participants are adjunctively eligible for WIC and they do not need further income verification process. This mechanism increases the choice utility of SNAP participants. Second, SNAP participants are already collecting substantial amount of benefit from SNAP, the additional benefit from WIC has less marginal utility to them as compare to another households that has the same observable characteristics but do not participate in SNAP. This decreases the choice utility of the SNAP-participating households. These two mechanisms likely cancel each other in the choice stage.

Specification (3) adds the racial group indicator for Black and Hispanic households. Black households are less attentive and have a lower choice tendency; Hispanic households are more attentive and have a higher choice tendency. This aligns with the raw data pattern. Black households have a lower than sample average take-up rate, and Hispanic households have a higher than sample average take-up rate. Such racial disparity is persistent even to date, suggesting that outreach to the Black community might be strategic for promoting WIC take-up.

Specification (4) introduces household income in the form of log income per household member as an additional covariate. Counterintuitively, we find that higher income is associated with both greater attention and a higher choice tendency. This aligns with the raw data in Table 1, which shows that participating households have a higher average income than non-participating households. We hypothesize that this result is driven by car ownership, which likely correlates with income. Vehicle access significantly reduces the hassle costs of traveling to appointments and WIC vendors, potentially outweighing the lower marginal utility of benefits

Parameter	Model Specification (1)	(2)	(3)	(4)	(5)	(6)
Attention Stage						
γ_0	-2.716	-2.336	-2.707	-3.581	-3.192	-2.426
	(0.250)	(0.253)	(0.251)	(0.303)	(0.306)	(0.147)
$\log(B_{it})$	0.321	0.226	0.325	0.323	0.212	–
	(0.052)	(0.053)	(0.052)	(0.058)	(0.058)	–
B_{it}	–	–	–	–	–	0.002
	–	–	–	–	–	(0.0005)
No infant	-0.732	-0.808	-0.728	-0.717	-0.801	-0.801
	(0.031)	(0.033)	(0.031)	(0.034)	(0.036)	(0.034)
LA_{it}	0.001	0.001	-0.000	-0.002	-0.003	-0.003
	(0.008)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Education	0.034	0.043	0.056	0.007	0.050	0.049
	(0.032)	(0.031)	(0.031)	(0.034)	(0.033)	(0.033)
Black	–	–	-0.314	–	-0.269	-0.269
	–	–	(0.047)	–	(0.047)	(0.047)
Hispanic	–	–	0.205	–	0.197	0.198
	–	–	(0.052)	–	(0.052)	(0.052)
SNAP	–	0.291	–	–	0.371	0.367
	–	(0.030)	–	–	(0.034)	(0.034)
Income	–	–	–	0.099	0.106	0.106
	–	–	–	(0.014)	(0.014)	(0.014)
Choice Stage						
β_0	1.134	1.124	1.101	0.399	0.430	1.151
	(0.259)	(0.260)	(0.260)	(0.302)	(0.305)	(0.137)
$\log(B_{it})$	0.141	0.154	0.143	0.195	0.200	–
	(0.053)	(0.053)	(0.052)	(0.057)	(0.058)	–
B_{it}	–	–	–	–	–	0.002
	–	–	–	–	–	(0.0005)
No infant	-0.381	-0.373	-0.380	-0.343	-0.343	-0.342
	(0.034)	(0.034)	(0.039)	(0.037)	(0.037)	(0.036)
S_{it}	-0.360	-0.355	-0.332	-0.443	-0.421	-0.420
	(0.101)	(0.101)	(0.102)	(0.109)	(0.110)	(0.110)
$S_{it} \times LA_{it}$	0.010	0.010	0.008	0.017	0.016	0.016
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
Education	0.082	0.070	0.095	0.065	0.073	0.073
	(0.026)	(0.026)	(0.026)	(0.028)	(0.029)	(0.029)
Black	–	–	-0.125	–	-0.073	-0.073
	–	–	(0.036)	–	(0.038)	(0.038)
Hispanic	–	–	0.176	–	0.136	0.137
	–	–	(0.040)	–	(0.040)	(0.040)
SNAP	–	-0.049	–	–	-0.009	-0.011
	–	(0.025)	–	–	(0.029)	(0.029)
Income	–	–	–	0.058	0.056	0.051
	–	–	–	(0.013)	(0.013)	(0.013)
Random effect						
$\log(\sigma_1)$	0.452	0.450	0.442	0.318	0.290	0.293
	(0.027)	(0.028)	(0.029)	(0.035)	(0.036)	(0.036)
$\log(\sigma_2)$	-0.114	-0.119	-0.123	-0.292	-0.305	-0.301
	(0.046)	(0.046)	(0.048)	(0.060)	(0.059)	(0.058)
Sample Size	131,382	131,382	131,382	106,055	106,055	106,055

Table 2: Comparison of full model results across specifications

for higher-income households. At the same time, not having a vehicle may make the sign-up process prohibitively troublesome, making households without a car less attentive. While we lack direct data on vehicle ownership to test this mechanism, we find supportive indirect evidence: controlling for income doubles the coefficient on the $S_{it} \times LA_{it}$ interaction. If income proxies for car ownership, this suggests that local accessibility—such as the density of medical services, social assistance offices, and public transit—matters most for households without cars, precisely the group for whom these local amenities are critical for reducing hassle costs.

Specifications (5) and (6) include the full set of covariates: all covariates in the baseline model, SNAP participation status, race, and income. All results remain robust. We interpret the numerical results for specification (6) choice stage in dollar amounts. A household with an infant values the WIC package \$182 more than another without an infant, ceteris paribus. The hassle cost for a household living in a county with no local accessibility is \$224, but each unit increase in LA_{it} reduces hassle cost by more than \$8. Consider a typical household with $LA_{it} = 15$, its hassle cost is about \$97. Note that the benefit data is skewed; hence, these dollar amount results are not necessarily interpretable. Next, we consider the attention and choice probabilities under specification (5).

5.2 Heterogeneity in attention and choice probabilities

We first define a baseline household to serve as a benchmark for all our heterogeneity analysis. A baseline household is defined as one infant, no preschooler, in a county where the local accessibility is 15, high school education, non-Black, non-Hispanic, does not participate in SNAP, with an annual income of \$6600 per person in the year 2009. Our model involves a random effect. Hence, we consider a range of random effects Q_i ranging from -2 to 2, which covers more than 95% of the standard normal distribution. We will show heterogeneity in both attention and choice probabilities along two dimensions: first, between baseline households and others, second, across different values of Q.

Baseline vs no infant

The baseline household has a monthly benefit of \$150, and a household without an infant, all else equal, has a monthly benefit of \$100. We plot the attention probabilities and choice probabilities of the baseline household and a household without an infant but has a preschooler. The rest of the characteristics of these two households are identical. Having an infant is a large attention trigger and increases choice probability substantially, as shown by Figure 5. Without the infant, households are much less attentive among the lower random effect subpopulation and have a lower choice tendency among the higher random effect subpopulation.

Another dimension of the heterogeneity is across different levels of Q. When $Q < -1$, the households are hardly attentive, leading to close to huge nonparticipation among this subpopulation. In contrast, as Q increases above -1, households become much more attentive. A similar pattern is observed for the choice probability. When $Q > 1$, then the choice probability is close to 100%; this subpopulation always chooses participation over nonparticipation. When Q decreases below 1, then choice probability quickly drops, making these households likely to exit WIC prematurely.

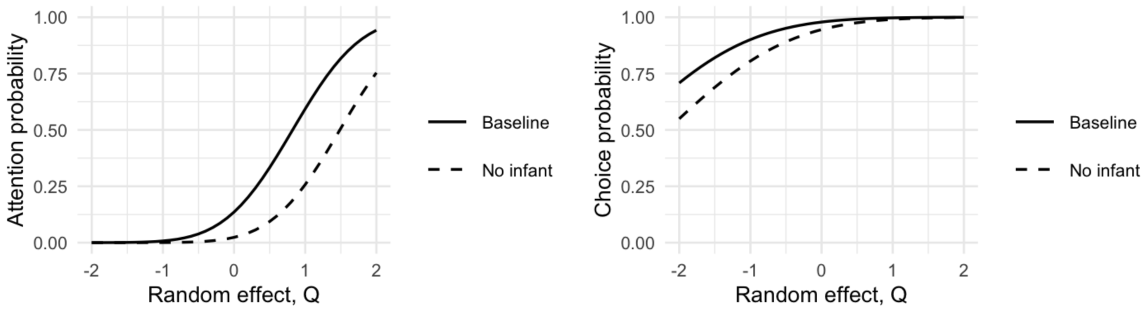


Figure 5: Left panel shows the attention probability of the baseline household and a household without an infant across different values of Q; right panel shows the choice probability of the baseline household and a household without an infant across different values of Q.

Baseline vs SNAP participants

As shown by the left panel of Figure 6, the attention probability of a SNAP-participating household is substantially higher than that of the baseline household when $Q > -1$. When $0 < Q < 1$, the difference between the two types of households is as large as 10%. In contrast, the SNAP-participating household's choice probability is almost identical to the baseline household's, as shown by the overlapping solid and dashed lines in the right panel of Figure 6. These two results show that the difference between the take-up rates between SNAP-participating households and other households is mainly driven by attention, not choice.

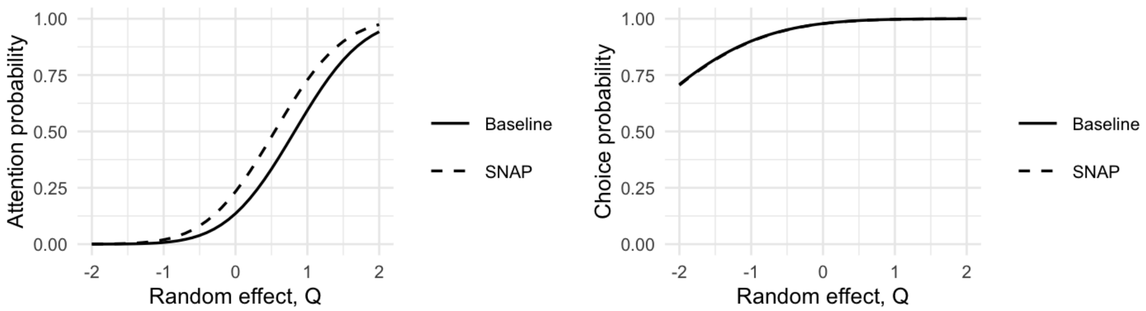


Figure 6: Left panel shows the attention probability of the baseline household and a household in SNAP across different values of Q; right panel shows the choice probability of the baseline household and a household in SNAP across different values of Q.

Racial disparity

Figure 7 displays racial disparities in both attention and choice probabilities. Hispanic households are more attentive and have a higher choice tendency than average. Black households are less attentive and have a lower choice tendency than average. There is more racial disparity in attention probability among households with infants than among households with no infants. The disparity kicks in around $Q = -1$ for the top left panel, whereas it kicks in around $Q = 0$ for the top right panel. Conversely, there is more racial disparity in choice probability among households with no infants than among households with infants.

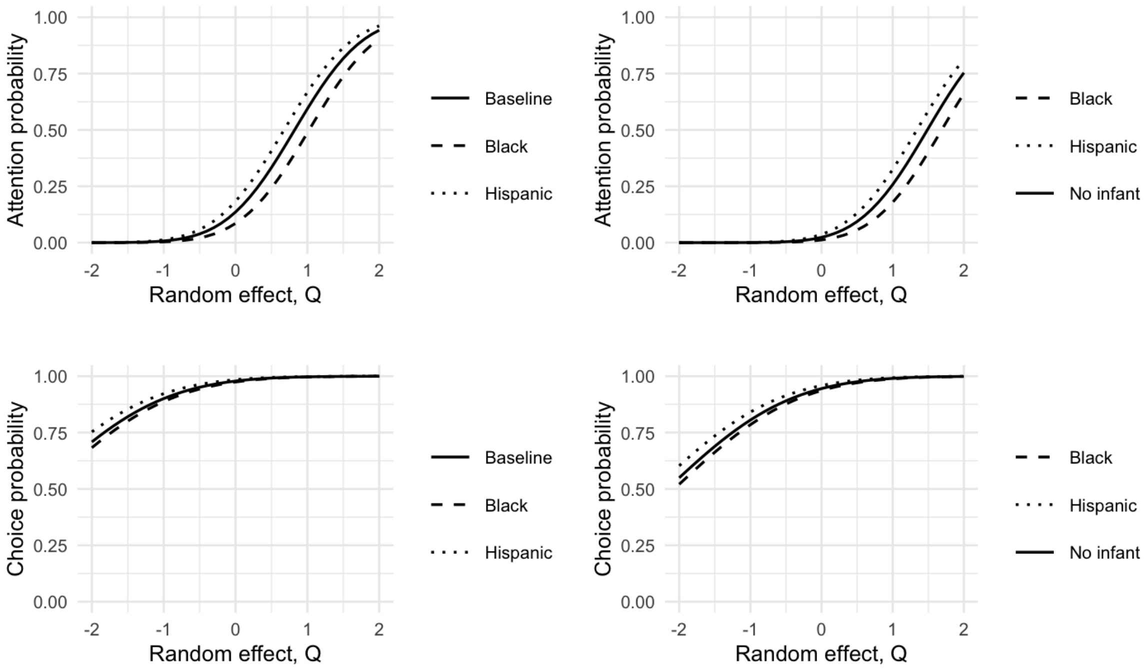


Figure 7: Top left panel shows the attention probability of the household by races across different values of Q; top right panel shows the attention probability of the household without an infant by races across different values of Q. Bottom left panel shows the choice probability of the household by races across different values of Q; bottom right panel shows the choice probability of the household without an infant by races across different values of Q.

Hassle cost and LA_{it}

We compare the choice probability of a household who do not need to pay a hassle, those who need to pay a hassle and stay in a county with $LA_{it} = 5$ (LA05), and those who need to pay a hassle and stay in a county with $LA_{it} = 15$ (LA15). The comparison shows that hassle cost significantly lowers choice probability when Q is negative. On the other hand, LA_{it} significantly reduces the influence of hassle cost.

5.3 Counterfactual simulations

Since our model involves a random effect, we have to simulate households. We divide education into 3 levels: less than high school, high school, and more than high school, denoted as $\{1, 2, 3\}$. We set LA_{it} as one of the three values: 10, 15, and 20. All households start with one child at the age of $\{3, 6, 9, 12, 15\}$ months. Half of the households will have a newborn at the 12th month. In total, there are 90 types of households. For each type, we simulate 5 of them. In total, we simulate 450 households. Each household has a time-invariant random effect drawn from a standard normal. We simulate each household's decisions for 30 months. We simulate the counterfactuals using specification (1). The general conclusions are robust using different specifications.

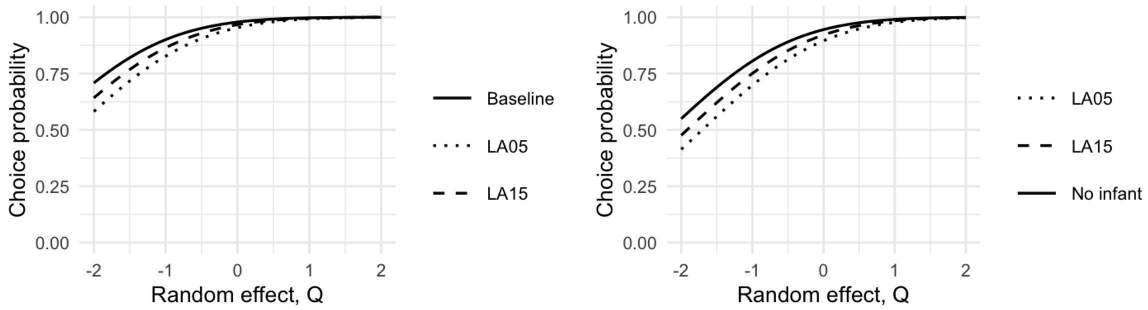


Figure 8: Left panel shows the choice probability of the baseline household and a household without an infant across different values of Q when they need to pay a hassle cost to participate in WIC; right panel shows the choice probability of the baseline household and a household without an infant across different values of Q when they need to pay a hassle cost to participate in WIC.

Forcing attention vs Forcing choice

We first compare two policy counterfactuals following the attention-choice model literature convention. In the first policy simulation setup, households are assumed to be always fully attentive. We call this intervention forcing attention. In the second policy simulation setup, the households are assumed always to choose participation over non-participation if they are attentive. We call this intervention forcing choice.

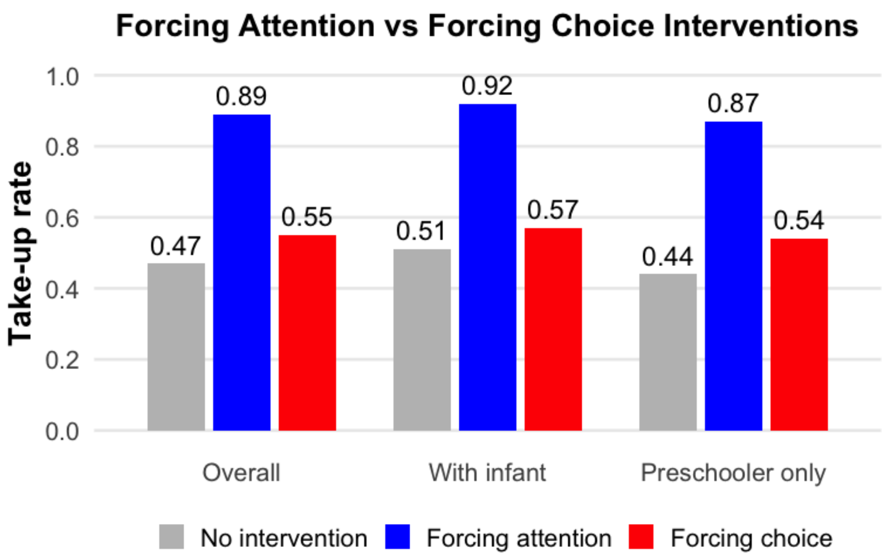


Figure 9: Forcing attention increases take-up by 40%; forcing choice increases take-up by 9%.

Forcing attention increases the participation rate from 47% to 89%. This 42% increase dwarfs the increase from the forcing choice, which is merely 8% (from 47% to 55%). The comparison suggests that attention is the major bottleneck for promoting welfare take-up.

Why forcing attention is more effective than forcing choice?

In Figure 10, we plot the histograms of attention probabilities and choice probabilities of all households under no intervention. The attention probabilities cluster close to 0. Around 48% of the households have their attention probabilities lower than 5%, close to 70% of the households have their attention probabilities lower than 25%. Raising attention probability to 100% induces a huge shift in attention probability's distribution, so it is extremely effective. In contrast, many households' choice probabilities cluster close to 1. Around 48% of the households have their attention probabilities higher than 95%, and close to 80% of the households have their attention probabilities higher than 75%. Hence, raising choice probability to 100% is much less effective since many households' initial choice probabilities are already close to 100%.

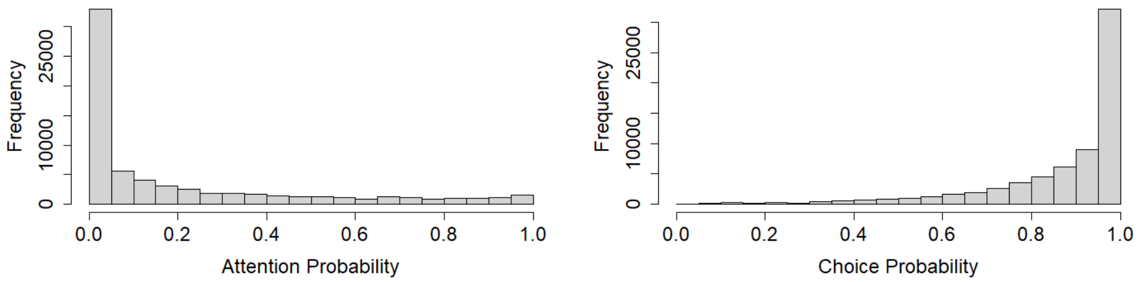


Figure 10: Histograms of attention probabilities and choice probabilities under no intervention.

Attention boost vs Choice nudge

This section simulates two realistic policies under the constraint that WIC policymakers can only intervene with currently participating households, as contact information for eligible non-participants is unavailable. Consequently, the only viable strategy is to extend participation spells among enrolled families. The first policy is an attention boost, which raises the attention probability to 100% in the period after a household exits the program. This mimics reminder messages sent to households that exit the program prematurely. The second policy is a choice nudge, which increases the choice probability by a small margin for all participating households with a child under 30 months ($YA_{it} \leq 30$). We test a range of increases, mirroring messages that inform households that continued WIC enrollment is beneficial for the future of the children. Results show that the attention boost raises the participation rate from 46.9% to only 47.4%—a gain of just half a percentage point. In contrast, a choice nudge that increases choice probability by a mere 0.45% achieves the same effect as the attention boost, as shown in Figure 11.

The effectiveness of choice nudge comes from two mechanisms. One, it directly increases the household's current period choice probability. Hence, current period retention increases among households who are already enrolled last period. Two, once the households are retained in WIC, they are fully attentive next period due to the ongoing engagement that they have with WIC. This indirectly increases their future attention probability. The combination of the two mechanisms drives the effectiveness of a choice-nudging policy.

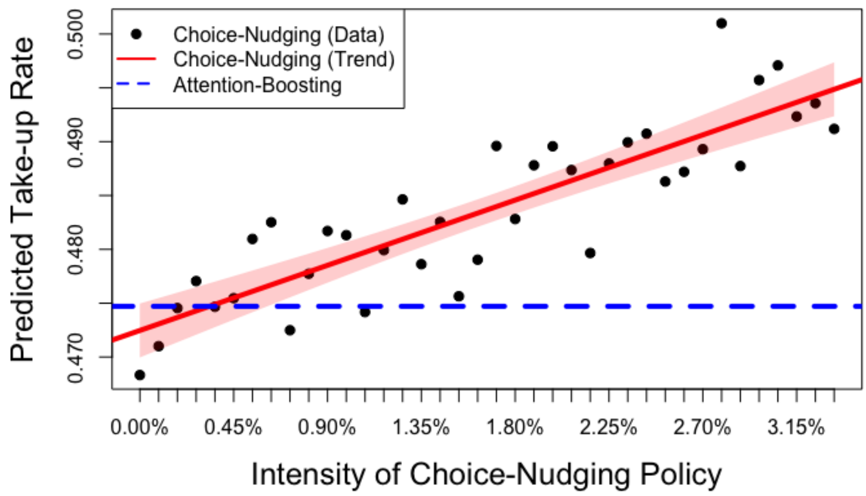


Figure 11: Comparing attention boost against different intensities of choice nudge

Limitations of the Counterfactual Analysis

Our counterfactual analysis indicates that choice-nudging policies may be more effective than attention-boosting interventions, as even a small (0.45%) increase in choice probability can match the effect of a large attention boost. However, this finding rests on a key empirical question: do real-world policies exist that can reliably increase choice probability by such a margin? This limitation—the gap between theoretical policy predictions and their practical implementation—is a common challenge in models of inattention (Hortaçsu, Madanizadeh and Puller, 2017; Heiss et al., 2021; Einav, Klopack and Mahoney, 2025). To validate our theoretical prediction, it is necessary to compare the actual treatment effects of attention-boosting and choice-nudging policies in a real-world setting. In Section 6, we use experimental data from the WIC2Five pilot program to provide exactly this empirical test.

5.4 Additional findings

On top of showing that choice nudge is likely more effective than attention boost, the full model generates two other interesting empirical findings. First, we find that the salient targeting pattern of choice nudge is on the unobserved random effect, but not on the observables. This is different from the findings from Finkelstein and Notowidigdo (2019), which find a targeting pattern on the observed household characteristics. We provide a unifying explanation for these two targeting patterns. Second, we note that higher education is associated with lower attention but higher choice tendency. This finding opens a scope for heterogeneous policy targeting. We defer the details of these additional findings to Appendix E.

6 Policy comparison validation

In this section, we validate our policy recommendation that the key to increasing WIC take-up is inducing choice, not raising attention among participating households. We show strong evidence from the publicly available data from Vermont WIC Program (2017), which aligns with our policy recommendations. Moreover, the pilot program was conducted on a limited budget from a WIC mini-grant, suggesting that a choice-nudging solution can be **both effective and inexpensive**. Both CNM and ABM interventions cost $4000/year for the entire state of Vermont, about $0.30 per household per year.

In 2016 and 2017, the Vermont Department of Health WIC Program State Office implemented a pilot program aiming to increase the WIC participation duration among households who are already in the program. The pilot program sent out two types of text messages to WIC participants: one was a choice-inducing message from Aug 2015 to Aug 2016, and the other was an attention-raising message from Mar 2017 to May 2017. Here, we quote from the message track file and final report of the pilot program one choice-nudging message (CNM) and one attention-boosting message (ABM).[9]

CNM Example

Welcome to WIC2Five! Get texts on family nutrition, health, activities and more. Kids who stick with WIC for the first 5 years grow healthy, happy and smart!

ABM Example

Hi Lynne! We missed seeing you at WIC today. Keep your benefits active. Call us at 802-388-4644 to reschedule. Thanks, the Middlebury Health Department.

CNM increases households' perceived value of the program. It informs them that WIC participation is good for the future development of the kids. On the other hand, ABM is a straightforward attention trigger that reminds participants of the necessary steps that need to be taken in order to continue participation. Under our two-stage model, ABM essentially ask the household who choose to exit WIC to reconsider their choices. Our estimation and counterfactual analyses in Section 5 suggest that CNM is more effective than ABM.

We anticipate the treatment effect of CNM to kick in *later* than the end of the intervention period, Aug 2016. This is due to two reasons. One, the vast majority of households start participation at a very young age of the infant, many even during pregnancy; two, households tend to participate until their youngest baby ages until 13 months or 25 months old. In contrast, if there is any treatment effect of ABM, we expect the treatment effect to kick in *immediately* during the year of intervention, 2017. This is because the ABM is sent to households that are about to recertify or have already missed their recertification appointment.

The outcome variable that is used by the Vermont WIC office to measure the policy effectiveness of ABM and CNM is the retention rate. We contacted the Vermont WIC office to check which month the retention rate is measured. The retention rates were measured in July 2015, July 2016, and July 2017.

6.1 Site selection

In this section, we discuss the treatment assignment of CNM and ABM. Table 3 shows which sites are selected for which treatment. In total, five sites are selected for CNM treatment and four sites are selected for ABM treatment.

[9]More examples of CNM and ABM can be found in Appendix F.

Location	CNM Treatment	ABM Treatment	EBT Timing	Retention Rates		
				2015	2016	2017
Rutland			Jun 2015	69.9%	65.9%	59.2%
Springfield	✓	✓	Oct 2015	66.9%	62.5%	64.7%
Bennington			Oct 2015	74.8%	73.7%	63.3%
White River	✓		Nov 2015	70.8%	63.3%	66.3%
Brattleboro	✓		Nov 2015	72.6%	70.3%	67.6%
St. Johnsbury		✓	Dec 2015	79.9%	77.3%	70.0%
Newport			Jan 2016	81.5%	73.7%	59.4%
Morrisville		✓	Jan 2016	83.8%	84.3%	77.9%
St. Albans			Feb 2016	73.2%	70.4%	59.0%
Burlington	✓		Feb 2016	68.1%	62.6%	62.3%
Middlebury		✓	Mar 2016	84.0%	79.7%	72.4%
Barre	✓		Mar 2016	66.8%	61.4%	61.9%

Table 3: WIC Child Retention Rates and Intervention Timeline (2015-2017), Ordered by EBT Rollout Date

We argue that the CNM-treated sites are negatively selected, hence, our estimates are conservative. On the other hand, ABM-treated sites are self-selected and hence, we are less confident about our claims regarding this policy. In addition, we show that the staggered EBT rollout during the intervention period (June 2015 to March 2016) is unlikely to have an impact on our treatment effect estimates.

Negative selection on CNM treatment

The pilot program selected Barre, Brattleboro, Springfield, White River districts for CNM intervention because these four sites "reported the largest decreases in caseload in 2014, and were also the offices with the largest difference between the number of WIC-eligible children receiving Medicaid and the number of children who were enrolled in WIC" (Vermont WIC Program, 2017).[10] The selection process was based on the outcome variable retention rate, implying that selected sites for the CNM treatment tend to have more downward-sloping trends in the retention rate of participating households. This coincides with our permutation test and three-period event study plot in Section 6.2. Hence, our estimate for CNM's treatment effect is potentially conservative and *understates* the effectiveness of CNM. Nevertheless, we show that the treatment effect is extremely salient and statistically positive in Section 6.2.

Self-enrollment on ABM treatment

All 12 districts were asked to take part in the ABM treatment. "Four offices, Middlebury, Morrisville, Springfield, and St. Johnsbury, quickly stepped up and began sending reminder texts one to three days before a scheduled appointment." Unlike the CNM treatment, it is difficult to determine whether the selection is positive or negative for the ABM treatment. The counties may volunteer because they have more responsible WIC staff, implying positive selection; the

[10]Burlington was asked to test the CNM intervention because it was the "largest and most culturally diverse" site.

counties may also have experienced or were expecting to experience a large reduction in retention rate and self-selected into the ABM treatment, implying negative selection.

Staggered rollout of EBT during the intervention period

Vermont WIC Program (2017) states that it is hard to estimate the true effectiveness of the pilot program because there was a staggered rollout of EBT during the intervention period from Jun 2015 to Mar 2016. Using the EBT rollout timing from Vermont WIC Program (2015) meeting minutes, we show that the two treated groups are not different from the control groups in terms of their EBT implementation timing.

We conduct the Wilcoxon rank-sum test for the EBT rollout timings for both treatments. The test results are summarized in Table 4. The ranks of the rollout timings in Table 4 can be computed based on Table Table 3. For example, Rutland is the first district to implement EBT and has a rank of 1. The next districts are Springfield and Bennington, both of which implemented EBT in the same month, so their ranks are (2+3)/2 = 2.5. The tests fail to reject the hypothesis that the treated and control groups for each treatment differ by their treatment timing. Hence, it is unlikely that the staggered rollout will contaminate the evaluation of the policy effectiveness for CNM and ABM.

	Ranks of the treated group	Ranks of the control group	Two-sided test p-value
CNM	$\{2.5, 4.5, 4.5, 9.5, 11.5\}$	$\{1, 2.5, 6, 7, 9.5, 11.5\}$	1
ABM	$\{2.5, 6, 7, 11.5\}$	$\{1, 2.5, 4.5, 4.5, 7, 9.5, 9.5, 11.5\}$	0.5861

Table 4: The Wilcoxon rank-sum test shows that we do **not** have enough evidence to reject that the treated sites and control sites have the same EBT rollout timing distribution.

6.2 CNM treatment evaluation

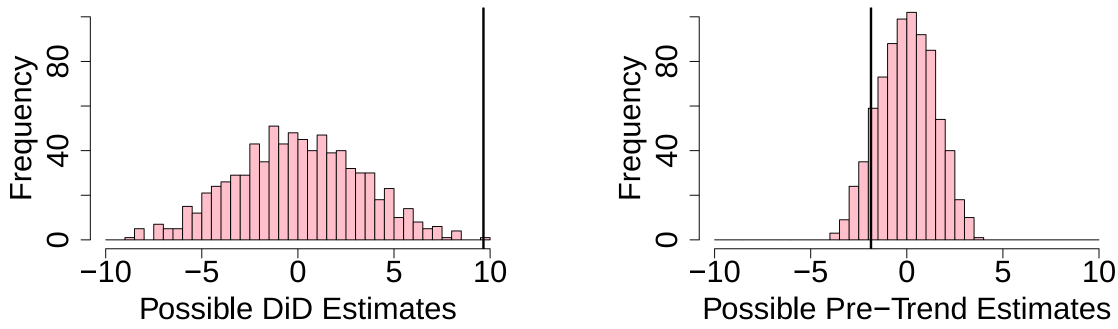
We combine the permutation test and DiD estimand to test whether there is a *positive* treatment effect of CNM and whether there is a *negative* pretrend, both of which are *one-sided* hypothesis tests. We supplement our causal analysis with the canonical DiD estimate and an event-study plot to support our findings further.

For the permutation test, we exhaust all possible $\binom{12}{5} = 792$ treatment assignments. For each treatment assignment, we compute the average of the treated sites' differences between 2017 and 2015 retention rates. We perform the same computation for the control sites. Then, we take the difference of the two differences. To be precise, denote the set of all 792 possible treatment assignments as $\mathcal{D}$, each possible treatment as $\mathbb{d}$, and the retention rate of site i in year t as RR_{it}. Since there are 5 sites treated out of 12, each $\mathbb{d}$ is a vector with length 12 with 5 entries equal to 1 and 7 entries equal to 0. The DiD estimand that we compute is

$$DiD_{\mathbb{d}} = \frac{1}{5}\sum_{i=1}^{12} \mathbb{d}_i(RR_{i,2017} - RR_{i,2015}) - \frac{1}{7}\sum_{i=1}^{12}(1 - \mathbb{d}_i)(RR_{i,2017} - RR_{i,2015}).$$

We compute $DiD_{\mathbb{d}}$ for all $\mathbb{d} \in \mathcal{D}$ and plot the the empirical distribution of $DiD_{\mathcal{D}}$ in Figure

12(a). Out of all the 792 possible assignments, the $DiD_{\mathbb{d}}$ for the actual treatment assignment is ranked **first**, indicating strongly that the treatment effect is positive with a p-value of $\frac{1}{792}$. As a robustness check, We replace $RR_{i,2017} - RR_{i,2015}$ with $RR_{i,2017} - RR_{i,2016}$ and perform the same permutation test, the actual DiD estimate is still ranked first and the p-value is still $\frac{1}{792}$.



(a) The black line indicates the actual $DiD_{\mathbb{d}}$.

(b) The black line indicates the actual $PreTrend_{\mathbb{d}}$.

Figure 12: Permutation test results for the CNM treatment effect evaluation

Just like the canonical DiD setup, one might be concerned about the parallel trend assumption and test the pretrend (as a proxy test for the parallel trend assumption). We perform the same test with permutation. We replace $DiD_{\mathbb{d}}$ with

$$PreTrend_{\mathbb{d}} = \frac{1}{5}\sum_{i=1}^{12} \mathbb{d}_i(RR_{2016} - RR_{2015}) - \frac{1}{7}\sum_{i=1}^{12}(1 - \mathbb{d}_i)(RR_{2016} - RR_{2015}).$$

We plot the empirical distribution of $PreTrend_{\mathcal{D}}$ in Figure 14(b). As suggested by Section 6.1, there is a negative selection on the treated sites. There is a downward trend among the sites selected for the CNM treatment. The negative pretrend is marginally significant with a p-value of 10.1%. Hence, our treatment effect estimate will likely understate its true value due to the downward-sloping trend.

We conduct canonical DiD and event-study (ES) to support our permutation test results: first, CNM's treatment effect is positive and statistically significant; second, its pretrend is negative and marginally significant. The estimation equations are as follows:

$$\begin{aligned} &\text{DiD} \qquad RR_{it} = \theta_i + \phi_t + \beta_{DiD}\mathbb{1}\{t = 2017\} \times \mathbb{d}_i + \epsilon_{it} \\ &\text{ES} \qquad RR_{it} = \theta_i + \phi_t + \beta_{pre}\mathbb{1}\{t = 2015\} \times \mathbb{d}_i + \beta_{post}\mathbb{1}\{t = 2017\} \times \mathbb{d}_i + \epsilon_{it} \end{aligned}$$

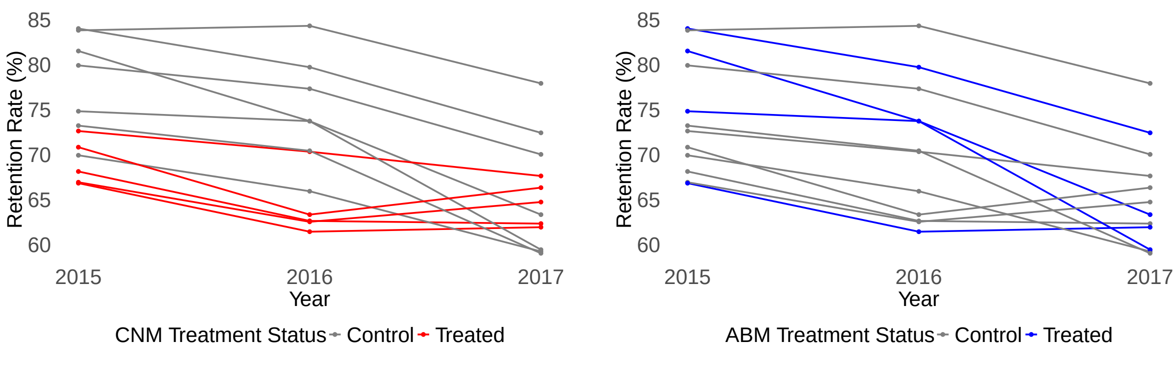
Table 5 shows that the estimation results for DiD and ES align with the permutation test results. DiD estimates that the CNM intervention increases the child retention rate by 8.723%, whereas ES estimates the treatment effect to be even higher at 9.654%. Both methods indicate strong statistical significance for the positive treatment effect of the CNM intervention.

	CNM		ABM	
	DiD	ES	DiD	ES
β_{DiD}	8.723***		−4.713	
	(1.646)		(3.485)	
β_{pre}		1.863		1.075
		(1.323)		(1.597)
β_{post}		9.654***		−4.175
		(1.527)		(3.546)
Observations	36	36	36	36
Clusters (site)	12	12	12	12

Table 5: Difference-in-Differences and Event Study Estimation Results

Addressing concerns on mean reversion

In Section 6.1, we claim that the sites are negatively selected, hence, our estimate is likely an understatement of the true treatment effect. One potential threat to the correctness of this statement is the mean reversion phenomenon. If sites that previously experienced a large decrease in retention rate are more likely to bounce back up to their mean retention rate, then the negatively selected sites will experience an increment in retention rate in the absence of the treatment. In this case, the estimated positive treatment effect does not have a causal interpretation. We address this concern by plotting Figure 13(a) and taking a closer look at the retention rate changes over 2015 to 2017 for both the treated and control sites. For completeness, we make the same plot for the ABM treatment (See Figure 13(b)).



(a)

(b)

Figure 13: 12 sites' retention rates over the three years. Lines are colored by treatment statuses.

Assuming a common mean for all 12 sites, if mean reversion is true, then the slower decline (from 2016 to 2017) in retention rate for orange lines would not necessarily indicate a positive treatment effect, but simply that those which previous decline more (from 2015-2016) bounce back to the mean of the 12 sites' retention rates. However, if this were to be true, the lower gray lines should also exhibit similar "bounce back" behavior which we do not observe in Figure 13(a). We observe the opposite, lower gray lines decline more than upper gray lines. Hence, it is unlikely that mean reversion plays a big role in our treatment effect evaluation.

6.3 ABM treatment evaluation

We test whether there is *any* treatment and whether there is *any* pretrend, both of which are *two-sided* hypothesis tests. We use the same methodologies to evaluate the effectiveness of the ABM treatment: permutation with a DiD estimand, canonical DiD, and event study. The results are reported in Figure 14 and Table 5,. All three sets of results agree that there is neither a significant pretrend nor a significant treatment effect.

Since we expect that the treatment effect of ABM kicks in immediately, we use $RR_{i,2017} - RR_{i,2016}$ instead of $RR_{i,2017} - RR_{i,2015}$. Figure 14(a) shows that the actual DiD estimate is negative with a p-value $119/495 \approx 24.02\%$; Figure 14(b) shows that the actual pretrend estimate has a p-value $245/495 \approx 49.50\%$. Our permutation tests show that both the treatment effect and the pretrend are not statistically significant.

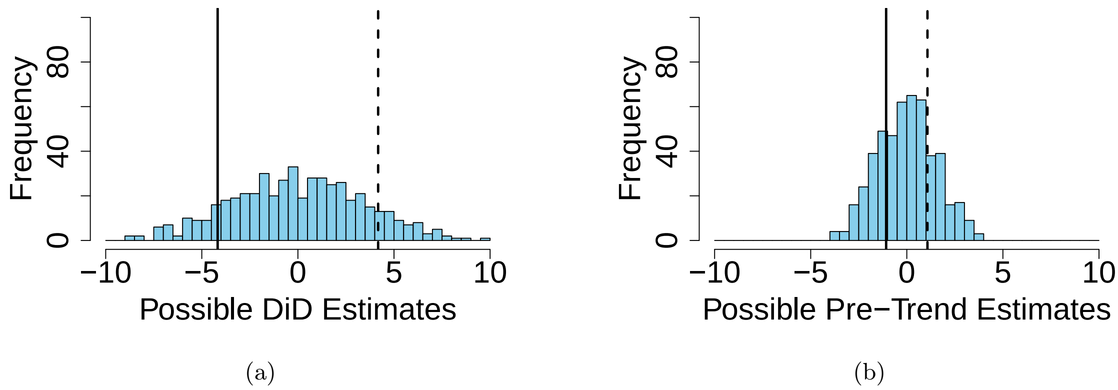


Figure 14: Permutation test results for the ABM treatment effect evaluation

Table 5 also show that both the pretrend and treatment effect are statistically insignificant. We plot the event study estimates and their 90% confidence intervals on Figure 14(b). Both confidence intervals cover 0.

Though the treatment effect is estimated to be insignificant, we do not claim that the ABM intervention is ineffective. Vermont WIC Program (2017) shows that ABM is useful in decreasing the rate of rescheduling and canceling appointments, decreasing workload from WIC staff. This allows the staff to focus on better serving other aspects of the program, potentially having positive long-term impacts, which are out of the scope of our study.

7 Conclusion

This paper proposes a two-stage inattention model to disentangle the causes of welfare program nonparticipation among eligible households into two latent mechanisms: inattention and active choice. Our identification strategy leverages the institutional structure of welfare programs. First, because current participants maintain ongoing program engagement, we assume they are fully attentive. This allows us to decompose observed data along two dimensions: the probability of staying in the program and the probability of starting participation. The former identifies choice parameters except hassle costs, while comparing the two dimensions identifies attention parameters. Second, the recertification process provides a clean source of variation that identifies hassle costs.

Applied to NLSY97 data on WIC take-up, the model finds substantial observed and unobserved household-level heterogeneity in both attention and choice probabilities. Counterfactual simulations show that inattention is the primary barrier to WIC participation at the population level. Under a more realistic policy scenario—where administrators can only contact currently enrolled households—we find that a small choice nudge that increases monthly choice probability by 0.45% is as effective as a large attention boost one month after exit. This suggests that choice-nudging policies may be more effective than attention-boosting ones. We test this hypothesis using data from the WIC2Five pilot program. The permutation tests provide strong support in favor of choice nudges, highlighting the critical role of choice nudges in improving program take-up.

References

Abaluck, Jason, and Abi Adams-Prassl. 2021. "What do consumers consider before they choose? Identification from asymmetric demand responses." *The Quarterly Journal of Economics*, 136(3): 1611–1663.

Agarwal, Nikhil, and Paulo J Somaini. 2025. "Demand analysis under latent choice constraints." *Review of Economic Studies.*

Banerjee, Abhijit, Rema Hanna, Benjamin A Olken, and Diana Sverdlin Lisker. 2024. "Social protection in the developing world." *Journal of Economic Literature*, 62(4): 1349–1421.

Barseghyan, Levon, Francesca Molinari, and Matthew Thirkettle. 2021. "Discrete choice under risk with limited consideration." *American Economic Review*, 111(6): 1972–2006.

Bere, Elling, Frank Van Lenthe, Knut-Inge Klepp, and Johannes Brug. 2008. "Why do parents' education level and income affect the amount of fruits and vegetables adolescents eat?" *The European Journal of Public Health*, 18(6): 611–615.

Bhargava, Saurabh, and Dayanand Manoli. 2015. "Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment." *American Economic Review*, 105(11): 3489–3529.

Bitler, Marianne P, Janet Currie, and John Karl Scholz. 2003. "WIC eligibility and participation." *Journal of Human resources*, 1139–1179.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2022. "Discretizing unobserved heterogeneity." *Econometrica*, 90(2): 625–643.

Carlson, Steven, and Zoë Neuberger. 2021. "WIC works: addressing the nutrition and health needs of low-income families for more than four decades." *Center on Budget and Policy Priorities.*

Chareyron, Sylvain, Patrick Domingues, and Lucie Fotsa Lieno-Gaillardon. 2021. "Does social interaction matter for welfare participation?" *Annals of Economics and Statistics*, 141: 49–70.

Cho, Sun-Joo, Sarah Brown-Schmidt, and Woo-yeol Lee. 2018. "Autoregressive generalized linear mixed effect models with crossed random effects: An application to intensive binary time series eye-tracking data." *Psychometrika*, 83(3): 751–771.

Cragg, John G. 1971. "Some statistical models for limited dependent variables with application to the demand for durable goods." *Econometrica: journal of the Econometric Society*, 829–844.

Crawford, Gregory S, Rachel Griffith, and Alessandro Iaria. 2021. "A survey of preference estimation with unobserved choice set heterogeneity." *Journal of Econometrics*, 222(1): 4–43.

Davis, David E. 2012. "Bidding for WIC Infant Formula Contracts: Do Non-WIC Customers Subsidize WIC Customers?" *American Journal of Agricultural Economics*, 94(1): 80–96.

Davis, Rebekah A, Hannah B Leavitt, and Melissa Chau. 2022. "A review of interventions to increase WIC enrollment and participation." *Journal of Community Health*, 47(6): 990–1000.

Einav, Liran, Ben Klopack, and Neale Mahoney. 2025. "Selling subscriptions." *American Economic Review*, 115(5): 1650–1671.

Escaramís, Geòrgia, Josep L Carrasco, and Carlos Ascaso. 2008. "Detection of significant disease risks using a spatial conditional autoregressive model." *Biometrics*, 64(4): 1043–1053.

Finkelstein, Amy, and Matthew J Notowidigdo. 2019. "Take-up and targeting: Experimental evidence from SNAP." *The Quarterly Journal of Economics*, 134(3): 1505–1556.

Friedrichsen, Jana, Tobias König, and Renke Schmacker. 2018. "Social image concerns and welfare take-up." *Journal of Public Economics*, 168: 174–192.

Ganong, Peter, and Jeffrey B Liebman. 2018. "The decline, rebound, and further rise in SNAP enrollment: Disentangling business cycle fluctuations and policy changes." *American Economic Journal: Economic Policy*, 10(4): 153–176.

Heiss, Florian, Daniel McFadden, Joachim Winter, Amelie Wuppermann, and Bo Zhou. 2021. "Inattention and switching costs as sources of inertia in Medicare Part D." *American Economic Review*, 111(9): 2737–2781.

Honka, Elisabeth, Ali Hortaçsu, and Maria Ana Vitorino. 2017. "Advertising, consumer awareness, and choice: Evidence from the US banking industry." *The RAND Journal of Economics*, 48(3): 611–646.

Hortaçsu, Ali, Seyed Ali Madanizadeh, and Steven L Puller. 2017. "Power to choose? An analysis of consumer inertia in the residential electricity market." *American Economic Journal: Economic Policy*, 9(4): 192–226.

Kleven, Henrik Jacobsen, and Wojciech Kopczuk. 2011. "Transfer program complexity and the take-up of social benefits." *American Economic Journal: Economic Policy*, 3(1): 54–90.

Ko, Wonsik, and Robert A Moffitt. 2024. "Take-up of social benefits." *Handbook of Labor, Human Resources and Population Economics*, 1–42.

Kreider, Brent, John V Pepper, and Manan Roy. 2016. "Identifying the effects of WIC on food insecurity among infants and children." *Southern Economic Journal*, 82(4): 1106–1122.

Lora, Karina R, Leslie Hodges, Cayley Ryan, Michele Ver Ploeg, and Joanne Guthrie. 2023. "Factors that influence children's exits from the Special Supplemental Nutrition Program for Women, Infants, and Children: A systematic review." *Nutrients*, 15(3): 766.

Rossin-Slater, Maya. 2013. "WIC in your neighborhood: New evidence on the impacts of geographic access to clinics." *Journal of Public Economics*, 102: 51–69.

Shepard, Mark, and Myles Wagner. 2025. "Do ordeals work for selection markets? Evidence from health insurance auto-enrollment." *American Economic Review*, 115(3): 772–822.

Sonchak, Lyudmyla. 2016. "The impact of WIC on birth outcomes: new evidence from South Carolina." *Maternal and child health journal*, 20: 1518–1525.

Swann, Christopher A. 2010. "WIC eligibility and participation: The roles of changing policies, economic conditions, and demographics." *The BE Journal of Economic Analysis & Policy*, 10(1): 1935–1682.

Tiehen, Laura. 2014. *WIC participation patterns: An investigation of delayed entry and early exit.* DIANE Publishing.

Vermont WIC Program. 2015. "Vermont WIC Retail Vendor Council January 2015 Quarterly Update." *Meeting minutes.*

Vermont WIC Program. 2017. "WIC2FIVE: Using Mobile Health Education Messaging to Support Program Retention." Vermont Department of Health Final Report. Vermont 2014 WIC Special Project Mini-Grant.

Wooldridge, Jeffrey M. 2005. "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity." *Journal of applied econometrics*, 20(1): 39–54.

A Methodological contribution to the attention-choice model

A.1 Combining marketing and welfare take-up

We believe that this is the first work that applies the attention-choice model from the marketing literature (Hortaçsu, Madanizadeh and Puller, 2017; Heiss et al., 2021; Einav, Klopack and Mahoney, 2025) to the incomplete welfare take-up problem (see Ko and Moffitt (2024) for a recent survey on the welfare take-up literature). Ko and Moffitt (2024) lists four major mechanisms investigated by the welfare take-up literature: unawareness, limited usefulness, sign-up hassle, and stigma. We incorporate all four mechanisms into our setup: unawareness is subsumed in the attention stage, whereas the other three mechanisms are explicitly modeled in the choice stage.

A.2 Inattention models

Our model introduces two innovations relative to existing two-stage inattention models in the marketing literature. First, our model does not require an exclusive shifter for the attention stage and requires only one binary exclusive shifter for the choice stage. The exclusive shifter is a binary indicator for whether a household needs to recertify its eligibility. Given that welfare programs are means-tested, such a data requirement is unlikely to be restrictive. This contrasts with the typical approach in the inattention model literature, which relies on two sets of exclusive shifters for both attention and choice stages (Heiss et al., 2021; Agarwal and Somaini, 2025). Second, we allow unobserved household-level heterogeneity to influence both attention and choice probabilities simultaneously. In contrast, most inattention models rule out unobserved heterogeneity (Honka, Hortaçsu and Vitorino, 2017; Crawford, Griffith and Iaria, 2021). We compare in detail our contribution and two leadings papers in this field.

A.2.1 Abaluck and Adams-Prassl (2021)

Abaluck and Adams-Prassl (2021) exploits the Slutsky asymmetry when *prices* of goods vary. Importantly, the price variation of the default good (the good that inattentive consumers choose by default) has a different impact from the price variation of non-default goods. In the welfare take-up context, the default option (nonparticipation for SA) *does not have a price.* Hence, our context calls for a different identification strategy.

We find the full model and Abaluck and Adams-Prassl (2021) similar in the constructing-with-derivative strategy for identifying functions of attention probability. Abaluck and Adams-Prassl (2021) leverages the Slutsky asymmetry in first-order derivatives (FODs) when prices vary for the default good versus for non-default good. The former affects both attention probability and choice probability; the latter affects choice probability only. Hence, their difference identifies a function of attention probability. We adopt a similar strategy to identify the ratio of attention parameters. However, looking for relevant FODs is trickier in the presence of unobserved household heterogeneity, which is ruled out in Abaluck and Adams-Prassl (2021). To accommodate household-level unobserved heterogeneity, we look for two *sequences* of take-up decisions that are *coherent shuffle* of each other. The ratio of the two relevant FODs identifies the ratio between attention parameters. The price we pay for allowing unobserved heterogeneity is the linear specification of the utility functions.

A.2.2 Agarwal and Somaini (2025)

Agarwal and Somaini (2025) assumes two sets of exclusive shifters, one for the attention stage and the other for the choice stage. In contrast, our key identifying assumption is the existence of an observed fully attentive subpopulation. In our propositions, the attention shifter is not required to be exclusive but has to generate an observed fully attentive subgroup.[11] Moreover, the exclusive attention shifter in Agarwal and Somaini (2025) is assumed to have large support. In our model, the non-exclusive attention shifter is the last period *binary* participation status.

Admittedly, though our theorems have a lower data requirement, our model setup, which requires the presence of a full attention group and a recertification process, is much more restrictive. Hence, we anticipate that readers find our work a better fit for specific empirical setups, such as welfare take-up or some other program participation contexts, whereas Agarwal and Somaini (2025)'s theoretical results apply to a wider range of topics but may impose unrealistic data assumptions in the empirical contexts that this paper specializes in.

A.3 Double hurdle model

Some readers may find our model similar to the classical double hurdle model (Cragg, 1971). We clarify that our setup is substantially different from the classical double hurdle model as our outcome variable is binary, whereas Cragg (1971) explicitly states that part of the observed outcome is a “continuous and positive random variable”.

A.4 Dynamic panel discrete choice with random effect

The full model introduces a dynamic panel (AR1) covariate into a generalized linear mixed-effect model (GLMM). Though such a design has been used by other works from econometrics, biometrics, and psychometrics (Wooldridge, 2005; Escaramís, Carrasco and Ascaso, 2008; Cho, Brown-Schmidt and Lee, 2018), the full model allows the AR1 term (i.e., last period take-up decision) and the random-effect component to *simultaneously* enter the attention and choice stages. This is a special feature of the full model that, to the best of our knowledge, has only been used by one paper: Heiss et al. (2021). However, Heiss et al. (2021) argues that the identifiability of their model comes from the presence of two sets of large-support exclusive shifters; we show that certain aspects of the model identifiability can be established semiparametrically/nonparametrically when an FA subpopulation is created by the *non-exclusive* attention shifter. Furthermore, the attention shifter in our paper is the binary-support lagged outcome, imposing virtually no data requirements if the data is longitudinal.

[11] We further modify the main identification results for the preliminary model in Appendix B to show that if the attention shifter is exclusive, exogenous and generates a fully attentive subgroup, then an exclusive choice shifter is not required.

B Mathematical Appendix

B.1 Proof of Lemma 2.1

Proof.

$$
\begin{aligned}
& P(D_{it} = 1 \mid D_{it-1}, X_{it}, Z_{it}) \\
= & P(A_{it} = 1, C_{it} = 1 \mid D_{it-1}, X_{it}, Z_{it}) \\
= & \begin{cases} P(X_{it}^{\upsilon}\beta^{\upsilon} + S_{it}X_{it}^{\kappa}\beta^{\kappa} + X_{it}^{\chi}\beta^{\chi} > \Xi_{it} \mid D_{it-1}, X_{it}, Z_{it}) & \text{when } D_{it-1} = 1 \\ P(X_{it}\gamma > \mathcal{E}_{it}, X_{it}^{\upsilon}\beta^{\upsilon} + S_{it}X_{it}^{\kappa}\beta^{\kappa} + X_{it}^{\chi}\beta^{\chi} > \Xi_{it} \mid D_{it-1}, X_{it}, Z_{it}) & \text{when } D_{it-1} = 0 \end{cases} \\
= & \begin{cases} H(X_{it}^{\upsilon}\beta^{\upsilon} + S_{it}X_{it}^{\kappa}\beta^{\kappa} + X_{it}^{\chi}\beta^{\chi}) & \text{when } D_{it-1} = 1 \\ G(X_{it}\gamma)H(X_{it}^{\upsilon}\beta^{\upsilon} + S_{it}X_{it}^{\kappa}\beta^{\kappa} + X_{it}^{\chi}\beta^{\chi}) & \text{when } D_{it-1} = 0 \end{cases} \\
= & G(X_{it}\gamma)^{1-D_{it-1}} H(X_{it}^{\upsilon}\beta^{\upsilon} + S_{it}X_{it}^{\kappa}\beta^{\kappa} + X_{it}^{\chi}\beta^{\chi})
\end{aligned}
$$

The third equality is from Assumption 2.1. The independence between $\mathcal{E}_{it}$ and Ξ_{it} allows us to write the observed take-up probability when $D_{it-1} = 0$ as a product of attention and choice probabilities. Then, we substitute the cdfs of $\mathcal{E}_{it}$ and Ξ_{it}, denoted as G and H into the equality. $\square$

B.2 Proof of Lemma 3.1

Proof.

$$
\begin{aligned}
& P(D_{i,1:T} \mid X_{i,1:T}, Z_{i,1:T}, Q) \\
= & P(D_{iT}, D_{i,1:(T-1)} \mid X_{i,1:T}, Z_{i,1:T}, Q_i) \\
= & P(D_{iT} \mid D_{i,1:(T-1)}, X_{i,1:T}, Z_{i,1:T}, Q_i)P(D_{i,1:(T-1)} \mid X_{i,1:T}, Z_{i,1:T}, Q_i) && \text{(B.1)} \\
= & P(D_{iT} \mid D_{iT-1}, X_{iT}, Z_{iT}, Q_i)P(D_{i,1:(T-1)} \mid X_{i,1:(T-1)}, Z_{i,1:(T-1)}, Q_i) && \text{(B.2)} \\
= & \prod_{t=1}^{T} P(D_{it} \mid D_{it-1}, X_{it}, Z_{it}, Q_i) && \text{(B.3)}
\end{aligned}
$$

Equation (B.1) is by the definition of conditional probability. Equation (B.2) is by the Markov property of the model setup. This step requires conditioning on Q_i. Recall that Q_i is correlated with all D_{it-1} with $t > 1$, so we need to condition on Q_i to drop $D_{i,1:(T-2)}$ in the conditioning set. Equation (B.3) is by recursively applying the previous reasoning to the last term in Equation (B.2). $\square$

B.3 Proof of Lemma 3.2

Proof.

$$
\begin{aligned}
&\frac{\partial}{\partial X^{\omega}_{i\tau}} P(D_{i,1:T} \mid X_{i,1:T}, Z_{i,1:T}) \\
&= \frac{\partial}{\partial X^{\omega}_{i\tau}} \int P(D_{i,1:T} \mid X_{i,1:T}, Z_{i,1:T}, Q_i = q) dF(q) \\
&= \frac{\partial}{\partial X^{\omega}_{i\tau}} \int \prod_{t=1}^{T} P(D_{it} \mid D_{it-1}, X_{it}, Z_{it}, Q_i) dF(q) \text{ by Lemma 3.2} \\
&= \int \frac{\prod_{t=1}^{T} P(D_{it} \mid D_{it-1}, X_{it}, Z_{it}, Q_i = q)}{P(D_{i\tau} \mid D_{i\tau-1}, X_{i\tau}, Z_{i\tau}, Q_i = q)} \frac{\partial}{\partial X^{\omega}_{i\tau}} P(D_{i\tau} \mid D_{i\tau-1}, X_{i\tau}, Z_{i\tau}, Q_i = q) dF(q) \\
&= \int \prod_{t=1}^{T} P(D_{it} \mid D_{it-1}, X_{it}, Z_{it}, Q_i = q) \frac{\frac{\partial}{\partial X^{\omega}_{i\tau}} P(D_{i\tau} \mid D_{i\tau-1}, X_{i\tau}, Z_{i\tau}, Q_i = q)}{P(D_{i\tau} \mid D_{i\tau-1}, X_{i\tau}, Z_{i\tau}, Q_i = q)} dF(q).
\end{aligned}
$$

□

B.4 Explanations of the three steps

Step 1: Integrating over Q_i

All the conditioning variables except for Q_i are assumed to be independent from Q_i, allowing us to integrate out Q_i over its marginal distribution. Though we will assume that Q_i follows standard normal for estimation, we do not need to impose any parametric assumption for identification. Therefore, we will use F to denote the distribution of Q_i throughout the paper, instead of a specific parametric family.

Step 2: Multiple period probability of take-up history conditional on Q_i

Lemma 3.1 decomposes the probability of a sequence of decisions as a product of many transition probabilities.

Step 3: Plug in single period transition probability

We express a single transition probability as the product of attention and choice probabilities.

$$
\begin{aligned}
&P(D_{it} = 1 \mid X_{it}, D_{it-1}, Z_{it}, Q_i) \\
&= P(A_{it} = 1 \mid X_{it}, D_{it-1}, Z_{it}, Q_i) P(C_{it} = 1 \mid A_{it} = 1, X_{it}, D_{it-1}, Z_{it}, Q_i) \qquad \text{(B.4)} \\
&= P_a(D_{it-1}, X_{it}, Q_i) P_c(D_{it-1}, X_{it}, Z_{it}, Q_i) \qquad \text{(B.5)}
\end{aligned}
$$

Equation (B.4) is established similarly to Lemma 2.1.

C Data details

C.1 Benefit imputation variable details

Using NLSY as an example, we illustrate the missing data and misreporting problems. **Missing data:** 56.57% of the eligible households do not report their WIC benefit amount, and most do not participate in WIC. **Misreporting:** Out of those who do report, the maximum monthly benefit amount is $13000, which is blatant misreporting.

Imputation procedure

To address both the missing data and the misreporting problems, we propose an imputation procedure for the monthly benefit amount. We first clip all data to between $(0, 500)$ and then use the clipped data as the response variable of the imputation procedure. The variables we use to impute the data are state-fixed effect, time trend, number of children, and age of children. Next, we explain our choice of variables.

We impute the monthly benefit amount, B_{it}, for all households including those who have reported their monthly benefit by estimating Equation (C.1) with the least absolute deviation regression (LAD). Imputing for households who do not report their potential benefit amount (mostly those who do not participate in WIC) helps resolve the missing data problem. Imputing for those who have reported their monthly benefit using LAD alleviates the misreporting problem.

$$B_{it} = \phi_s + \tau \mathbb{1}\{\text{year} > 2007\} + \gamma c_{it} + \delta d_{it} + \epsilon_{it} \tag{C.1}$$

where B_{it} is the monthly WIC benefit amount in dollar value reported by household i in month t. ϕ_s is the state-group fixed effect for state group $s \in \{\text{low, medium, high}\}$. c_{it} is the number of kids under the age of 1, and d_{it} is the number of kids between the ages of 1 and 5.

- ϕ_s: WIC is administered by state government.
- year: There is a food package revision that takes effect from 2008.
- c_{it} and d_{it}: each child is entitled to her own benefit and children under the age of 1 are entitled a much greater amount of benefit from the baby formula package.

State groupings WIC is a federal assistance program administered by the United States Department of Agriculture (USDA). However, the actual implementation is managed at the state level. While the federal government provides a framework for the WIC program, including maximum monthly allowances for different food packages, states have some flexibility in implementing the program. For example, currently, USDA specifies that "partially breastfed infants may receive up to 104 fl. oz. of infant formula". The state government chooses which infant formula brand can be purchased with a WIC voucher/EBT card (Davis, 2012). Different infant formula brands have different prices, consequently, the dollar value of WIC benefits can vary across states within the guidelines set by the federal government.

We categorize the state into three groups in terms of monthly benefit amount: low, medium, and high. On top of data confidentiality reasons, the grouping serves two related purposes: (i) to reduce variability for states that have very few reported benefit amounts and (ii) to avoid high dimensional state fixed effect which may lead to overfitting.

We further justify the grouping by observing the political affiliation of these three groups. The state groupings in our analysis reflect a clear alignment with the established party lines in the United States. Most of the states categorized as low-benefit consistently voted Republican in the four presidential elections from 1996 to 2008, which coincides with the period covered by the NLSY97 survey. Among the three groups, the high-benefit group exhibits the highest proportion of states that voted Democrat.

2007 WIC package revision In 2007, USDA introduced a new set of food packages via an Interim Rule based on recommendations from the Institute of Medicine. This event explains the increase in the median value of B_{it} in years 2008 and 2009 as in Figure 15(a). As such, we include the indicator for the event year > 2007 in Equation (C.1).

Number of kids and age of kids WIC benefits are allocated based on the nutritional needs of each child. Therefore, in general, a household with more kids will get more benefits as evidenced by Figure 15(b). Infants from birth to one year of age receive WIC food packages designed to support their growth and development during this critical period. The benefits include infant formula, infant cereal, and baby foods appropriate for their age and stage of development. Children between the ages of one and five years old receive WIC packages that focus on promoting healthy eating habits and meeting their nutritional needs as they transition to solid foods. The benefits include a variety of nutritious foods such as milk, cheese, eggs, fruits, vegetables, and whole grains. In general, the (perceived) value of food packages for infants is larger than that for children between one and five (Lora et al., 2023).

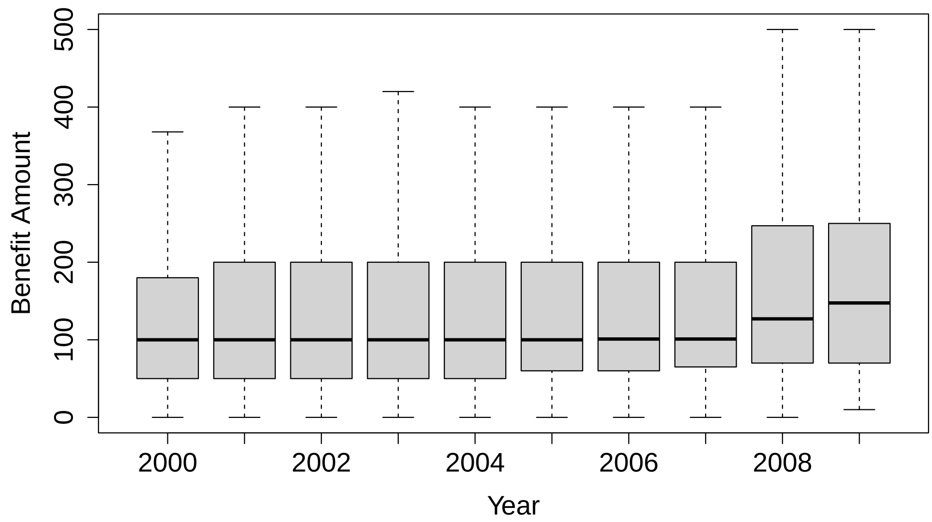
D Preliminary model estimation results

D.1 Choice stage parameters

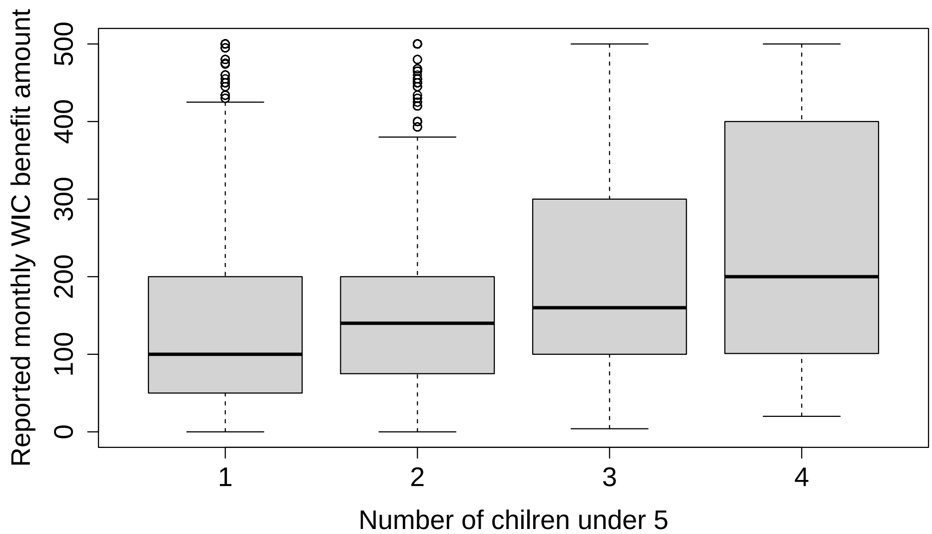
In the simplest logistic model (1) for the preliminary model, we include $\log(B_{it})$, R_{it}, the interaction term between R_{it} and LA_{it}, the surveyed individual's education level. We augment regression specification (1) with the number of infants in the household (denoted as "Infants") and the indicator of whether the year is larger than 2007 (denoted as "After 2007"), both of the which are added to address the special benefit structure of WIC.[12] All regression results are summarized in Table 6. All estimated coefficients' signs align with our intuition and the raw data pattern.

The sign of $\log(B_{it})$'s coefficient is positive: a higher benefit induces a higher participation rate. The positivity stays true throughout all specifications, but is not statistically significant when we condition on both Infant and After 2007. This is likely due to the fact that we use Infant and After 2007 to impute B_{it}; hence, most of the variation in $\log(B_{it})$ comes from Infant and After 2007. The signs of Infant and After 2007 are both positive, indicating that having an infant increases the choice utility of the program benefit and that the 2007 revision of the

[12] As mentioned in Section 4, the benefit amount for the infant package is much greater than that for the preschooler package; and in 2007, the package structure of WIC was significantly revised and the package change was implemented in 2008.



(a) B_{it} variation across years



(b) B_{it} variation across number of kids

Figure 15: The first panel shows B_{it} variation across years. The second panel shows that households with more children collect higher B_{it} from the WIC program.

	Dependent Variable: D_{it}				
	(1)	(2)	(3)	(4)	(5)
$\log(\mathbf{B_{it}})$	0.701^{***}	0.273^{***}	0.616^{***}	0.074	0.137
	(0.083)	(0.096)	(0.084)	(0.101)	(0.102)
Infant		0.548^{***}		0.616^{***}	
		(0.045)		(0.047)	
LeftBF					0.057^{***}
					(0.008)
After 2007			0.257^{***}	0.375^{***}	
			(0.051)	(0.052)	
$\mathbf{S_{it}}$	-1.303^{***}	-1.052^{***}	-1.300^{***}	-1.019^{***}	-1.334^{***}
	(0.160)	(0.161)	(0.160)	(0.161)	(0.163)
$\mathbf{S_{it}} \times LA_{it}$	0.027^{**}	0.028^{**}	0.025^{**}	0.026^{**}	0.032^{**}
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Education	0.144^{***}	0.161^{***}	0.110^{***}	0.112^{***}	0.167^{***}
	(0.039)	(0.040)	(0.040)	(0.040)	(0.041)
Constant	0.042	1.697^{***}	0.426	2.570^{***}	2.776^{***}
	(0.394)	(0.445)	(0.402)	(0.469)	(0.484)
Year fixed effect					✓
Observations	78,815	78,815	78,815	78,815	78,815
Log Likelihood	−10,087.020	−10,008.260	−10,073.820	−9,980.911	−9,976.421

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors are in parentheses.

Table 6: Logistic regression results for the preliminary model

benefit structure succeeds in promoting take-up. In the fullest specification (5), we include a term called "LeftBF" which is the number of remaining months that the household is still eligible for the baby formula benefit. Since less than 5% of the FA households have more than 1 infant, LeftBF is largely a finer partition of Infant; that's why we drop Infant. We also replace the indicator for years 2008 and 2009 with a finer partition of year fixed effects. Specification (5) shows that LeftBF increases choice utility, which indicates that households do consider the future value of the program. The recertification process is costly. Throughout all specifications, the coefficients of R_{it} are significantly negative, and the magnitude is large relative to the coefficient of $R_{it} \times LA_{it}$. On the other hand, county-level accessibility LA_{it} can curb κ_{it}, moving from a low accessibility county to a high accessibility county (LA_{it} increases from 0 to 25) can roughly half the hassle cost of (re)-sign-up. In addition, education reduces χ_{it}, though the interpretation might

D.2 Attention stage estimation results

Since the estimation of the attention probability is nonparametric, it is hard to summarize the estimation results. Using specification (5) for the choice stage, we report two findings concerning attention probability that have direct policy implications. The two findings are best represented by Figure 16(a). (1) The attention probability stays slightly below 10% during pregnancy, and surges to 20% when the baby is delivered. Then, it declines sharply within 3 months to below 5%. (2) Higher education is associated with lower attention probability. This finding is intuitive because lower-educated households are more likely to interact with other low-income households who are also eligible or are participating in WIC, hence, there is a greater network effect for the lower-educated households.

There are two corresponding findings on the choice probability. (1) The choice probability stays between 85% to 90% during. The constant choice probability of households from each education level is determined by the model design. The choice probability surges to 95% when the baby is delivered, and then it steadily declines as LeftBF decreases to 0. (2) Higher-educated households have higher choice probability. This finding aligns with some empirical research that higher-educated parents value nutrition for their children more than lower-educated parents Bere et al. (2008). Next, we show that the full model outputs the same qualitative results.

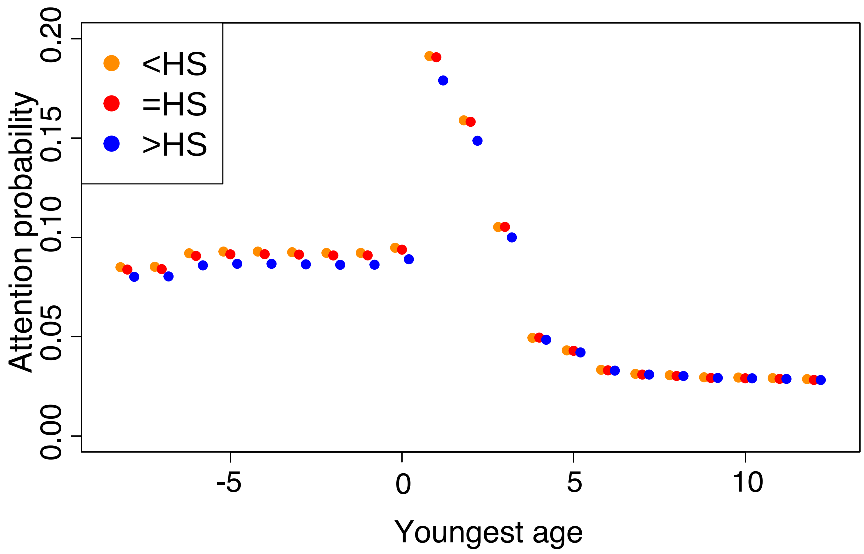
E Additional findings from the full model

E.1 Targeting pattern in comparison to Finkelstein and Notowidigdo (2019)

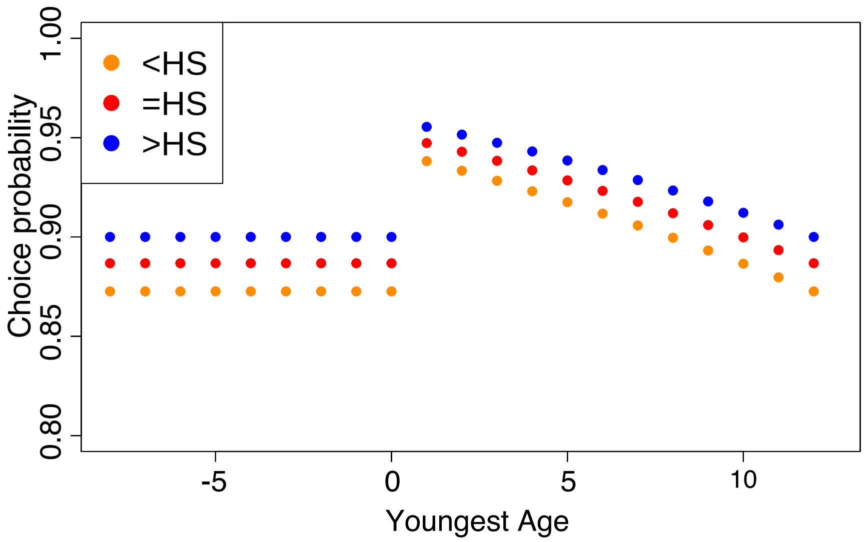
Finkelstein and Notowidigdo (2019) finds that intervention that aims at promoting welfare take-up tend to help less needy households who receive lower benefits from the program and are likely to be less socioeconomically disadvantaged. We simulate the policy intervention in Finkelstein and Notowidigdo (2019) as a one-shot attention boost by 30% for *all* households at period 1.

We summarize in Table 7 the average characteristics of all participating households under different interventions. The average observed characteristics, such as benefit, number of kids below one, education, and local accessibility, virtually do not vary across these four cases. In other words, we do not find the same targeting pattern in the households' observed characteristics. Instead, we find a targeting pattern on the random effect when the policy increases the take-up rate substantially. We provide a unifying explanation for the targeting on observable patterns in Finkelstein and Notowidigdo (2019) and the targeting on unobservable patterns in our paper.

Both types of targeting tend to promote take-up more among households with lower choice probability. In Finkelstein and Notowidigdo (2019), households with lower benefits and lower needs for welfare benefits have a lower choice tendency. On the other hand, a lower random effect also leads to a lower choice tendency. Hence, we can unify the two targeting patterns as helping out households with lower choice tendency. This unifying explanation is intuitive. Households with higher choice tendency are more likely to join the program in the absence of any intervention. Any intervention is likely to induce more participation among households with lower choice tendency.



(a)



(b)

Figure 16: The first panel shows the attention probability of a household with one infant, living in a medium benefit state and median county level accessibility. The second panel shows the choice probability of that household. The attention and choice probabilities pattern is the same for the other types of households.

	No intervention	Attention boost	Choice nudge	FN QJE 2019
Participation rate	0.48	0.49	0.54	0.53
Benefit (log)	4.86	4.86	4.85	4.86
Number of kids below one	0.39	0.39	0.36	0.40
Education	1.99	1.99	1.99	2.00
Local accessibility	14.90	14.92	14.93	14.88
Random effect	0.74	0.73	0.65	0.66

Table 7: Policy targeting pattern comparison

E.2 Heterogeneity across education

γ^{edu} is insignificant, but β^{edu} is significantly positive. Hence, higher-educated households are more likely to enroll if they are attentive. In contrast, less-educated households exhibit a lower choice tendency. This finding suggests heterogeneous policy targeting. When the WIC policymakers have resources to conduct informational campaigns, they may want to collaborate with parenting student groups at community colleges/universities. This outreach strategy can effectively target a group of eligible households that would participate with high probability and likely stay in the program for a long duration once they become attentive. On the other hand, when the policymakers have resources for sign-up assistance or health education, policymakers may want to prioritize lower-educated households or neighborhoods with a lower average education level. These households are more likely to experience difficulty during sign-up or undervalue the program benefits, which leads to a lower choice tendency.

Our estimation results show that under our conditions, the exclusive shifter conditions required for most of the inattention model identification literature are not met. Table 2 shows that SNAP is an exclusive shifter for the attention stage. Agarwal and Somaini (2025) shows that a two-stage inattention model is identified with two sets of exclusive shifters and requires that the exclusive shifter for the attention stage needs to have large support. SNAP has a binary support; hence, the identifying conditions in Agarwal and Somaini (2025) are not met in our empirical context.

F Example of text messages from Vermont WIC Program (2017)

In Table 8, we present additional examples for the CNM sent out by the WIC office during the WIC2Five pilot program.

One other example of ABM is

- Hi Lynne! Reminding you to complete your Nutrition Education before March 31, 2017 to keep your benefits current. It's easy! Complete a lesson online at wichealth.org. Your WIC household ID is 123456. Or call the Middlebury Health Department, 802-388-4644 for more options

1 Year Old	**2 Year Old**	**3 Year Old**	**4 Year Old**
Active play helps your toddler build more than muscle. Build her brain with activities like stomping, waddling and marching. Run and jump every day!	Active play helps your child build more than muscle. Build her brain with activities like scurrying, chasing and trudging. Run and jump every day!	Active play helps your preschooler build more than muscle. Build her brain with activities like hopping, leaping and dashing. Run and jump every day!	Active play helps your preschooler build more than muscle. Build her brain with activities like skipping, prancing, and galloping. Run and jump every day!
Get your free copy of *Playing with Your Toddler*. Text Fitwic now and we will send you one.	Get your free copy of the *Fit WIC Activity Pyramid*. Text Fitwic now and we will send you one.	Get your free copy of the *Fit WIC Activities Book*. Text Fitwic now and we will send you one.	Get your free copy of the *Fit WIC Activities Book*. Text Fitwic now and we will send you one.

Table 8: Example of CNM sent to different age groups