Dynamic Treatment Effects and Dynamic Selection with Time-Varying Instruments*

Joshua Shea[1] and Abigail Stocker[2]

[1]Department of Economics, University of Illinois at Urbana-Champaign
[2]Department of Economics, College of William & Mary

March 13, 2026

[Preliminary draft, please do not circulate]

Abstract

We develop a framework to identify dynamic treatment effects in a panel setting with endogenous treatment timing and repeated binary instrumental variable (IV) shocks. Agents respond dynamically to the instrument when selecting timing of treatment, generating multiple latent complier and non-complier types. Using the sequential IV shocks, we nonparametrically identify the distribution of types and dynamic treatment effects. Heterogeneous effects can be separately identified across complier types and treatment cohorts. Neither conventional monotonicity nor parallel trend assumptions are required. We propose nonparametric estimators for the distribution of latent types and treatment effects and derive their asymptotic properties. We additionally propose a model misspecification test to aid researchers in modeling selection into treatment. Simulations indicate good finite-sample performance, even in the presence of many latent types.

Keywords: Dynamic treatment effects; dynamic selection; sequential instrumental variables; treatment effect heterogeneity; nonparametric; panel data; noncompliance

JEL Classification: C22, C23, C26

---

*We thank participants at the Southern Economics Association 2025 Annual Meetings for their feedback. We are also grateful to Ashrita Mohanty for her research assistance. Contact: Shea: jkc-shea@illinois.edu; Stocker: ajstocker@wm.edu

1 Introduction

Instrumental variables (IVs) are often used to identify dynamic effects for an endogenous treatment. This empirical strategy has been applied to evaluate job training programs (Schochet et al., 2008; Alzúa et al., 2016; Hirshleifer et al., 2016; Das, 2021), study the effects of fertility on labor market outcomes for women (Bronars and Grogger, 1994; Silles, 2016; Lundborg et al., 2017; Bensnes et al., 2023; Angrist et al., 2024; Gallen et al., 2025; Ilciukas, 2025), and examine the impact of law enforcement on deforestation (Assunção et al., 2023; Picchetti, 2024). Recently, an econometric framework for the case of dynamic selection into treatment under time-invariant IVs has emerged (Angrist et al., 2024; Ferman and Tecchio, 2025; Gallen et al., 2025). However, despite the use of time-varying IVs in observational settings (Miguel et al., 2004; Autor et al., 2013; Acemoglu et al., 2013, 2016; Assunção et al., 2023), as well as the use of repeated randomization and nudges in experimental settings (Fehr and Goette, 2007; Thaler and Sunstein, 2009; Klasnja et al., 2015; Karlan et al., 2016; Antinyan et al., 2021; Milkman et al., 2025), no formal framework exists for the case of time-varying IVs. In this paper, we propose a methodology to identify and estimate dynamic effects under dynamic selection into treatment when a time-varying instrument is available.

We consider a setting where an absorbing treatment is available over multiple periods and agents endogenously select when to enroll in treatment. Each period, each agent experiences a binary IV shock affecting their enrollment decision. Agents respond dynamically to these shocks, with their responses depending on the history of past shocks. This generates many latent agent types defined by their choice of treatment timing under each potential sequence of shocks. We allow untreated outcomes and treatment effects to arbitrarily depend on latent types. This complicates the identification of treatment effects as the composition of treated agents changes non-randomly over time, increasing the potential for selection bias. Additionally, a parallel trend assumption allowing us to infer counterfactual outcomes of one group of agents from the outcomes of another loses its credibility. We propose a general framework that uses sequential IV shocks to identify the distribution of latent types and dynamic treatment effects. Moreover, we show that treatment effects can be separately identified across different complier types (Imbens and Angrist, 1994) whose timing of treatment is impacted by the IV, as well as across different treatment cohorts. We show that the latent type distribution and dynamic treatment effects can be nonparametrically estimated using a straightforward GMM-style estimator. Our framework also permits a simple model misspecification test to assist researchers in modeling selection into treatment.

Our paper relates to several branches of the econometrics and applied econometrics literature. We relate most closely to the recent literature on dynamic non-compliance (Angrist et al., 2024; Ferman and Tecchio, 2025; Gallen et al., 2025; Ilciukas, 2025). These papers consider panels where agents are randomized into treatment in an initial period (i.e., a time-invariant IV shock), but potentially deviate from their initial assignment over time. Selection into treatment and non-compliance are dynamic in the sense that agents' responses to the initial assignment vary across time. This also creates a multitude of latent agent types with potentially heterogeneous treatment effects. These papers invoke two key assumptions. The first is the monotonicity assumption of Imbens and Angrist (1994), which restricts the IV to weakly induce all agents to enroll in treatment at an earlier period. The second is the *wave ignorability* assumption of Angrist et al. (2024), which states that treatment effects depend on time since treatment rather than the timing of take-up.[1] Under these assumptions, the papers above identify a weighted average of the dynamic treatment effects across complier groups. We contribute to this strand of literature in four ways: (i) we extend the selection model to allow agents to respond dynamically to repeated shocks; (ii) we relax the monotonicity assumption so that the IV can induce some agents to enroll in treatment later and others to enroll earlier; (iii) when treatment effects are heterogeneous across agent types, we separately identify the dynamic effects for different complier types; (iv) we consider an alternative framework for treatment effect heterogeneity where effects vary across cohorts rather than latent types, and we separately identify the effects for different cohorts.

Closely related is the literature combining IV methods and difference-in-differences (DID) to recover the dynamic treatment effect for compliers/switchers when treatment is endogenous, but the response to the IV—which may be time-invariant or time-varying—is static (De Chaisemartin and d'Haultfoeuille, 2018; Ye et al., 2023; Bensnes et al., 2023; Picchetti, 2024; Miyaji, 2024). These papers employ a combination of parallel trends, monotonicity, and a static selection model to obtain identification. Under our setting of sequential IV shocks, we show that dynamic treatment effects can be separately identified for complier types with neither parallel trends nor monotonicity under a dynamic selection model.

Finally, we relate to the conventional event studies literature (Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Borusyak et al., 2024) and its extensions that question or relax the parallel trends assumption (Abadie, 2005; Callaway and Tsyawo, 2023; Rambachan and Roth, 2023; Ye et al., 2024; Ghanem et al., 2025). Intuitively, our identification strategy can be thought of as applying an event study design to one agent type at a time, despite agent type being latent. Different treatment cohorts within agent types are generated by the IV. Treatment effects are thus recovered via within-agent type comparisons, which enables us to relax the parallel trends assumption and allow for un-

[1]The wave ignorability assumption is implicit in event study settings where dynamic treatment effects are indexed by time since treatment rather than calendar time or period of enrollment.

restricted heterogeneity in untreated outcomes across agent types. Whereas some earlier papers relaxing parallel trends have used a partial identification approach (Rambachan and Roth, 2023; Ye et al., 2024) or imposed a factor model (Callaway and Tsyawo, 2023), our sequential IV shocks permit us to nonparametrically point identify treatment effect trajectories for multiple agent types.

The remainder of the paper is organized as follows. Section 2 presents the dynamic selection model. Section 3 establishes general identification results for the latent type distribution, as well as for dynamic treatment effects under two frameworks for treatment effect heterogeneity. We additionally provide examples of selection models and the parameters identified under each model. Section 4 proposes estimators for the latent type distribution and dynamic treatment effects, and derives the asymptotic distributions of the estimators. Section 5 presents simulations showing the finite-sample properties of the proposed estimators. Section 6 concludes.

2 Selection model

We first propose a simple, flexible model for when and whether agents select into treatment. Once treated, agents remain treated. We then provide examples of restrictions generating certain selection patterns. In Section 3.2.2, we discuss whether these selection patterns generate testable implications for model misspecification.

2.1 Model

Let $i \in \mathcal{I}$ index economic agents and $t \in \mathcal{T} = \{1, 2, \ldots, \bar{t}\}$ index time period. Each period, untreated agents decide whether to enroll in an absorbing treatment. Affecting this decision is the instrumental variable $Z_{i,t} \in \{0, 1\}$, which represents a binary shock realized each period. Let $Z_i \equiv (Z_{i,1}, \ldots, Z_{i,\bar{t}}) \in \mathcal{Z}$ denote the vector of IV shocks across all time periods for agent i. In each period t, agents select into treatment according to the decision rule

$$D_{i,t} = \max\left\{D_{i,t-1},\ \mathbb{1}\left[0 \leq h_t(Z_{i,1}, \ldots, Z_{i,t}; A_i)\right]\right\}, \tag{1}$$

where $D_{i,t} = 0$ for $t < 1$, $A_i \in \mathbb{R}^{d_A}$ for $d_A \in \mathbb{N}$, and $h_t : \mathbb{R}^{t+d_A} \to \mathbb{R}$. The max operator accounts for the absorbing nature of treatment, ensuring the agent stays treated once taking up treatment. The function $h_t(\cdot)$ generates a latent index based on the IV shocks up to period t and a vector of unobserved agent characteristics A_i. We allow A_i to arbitrarily depend on the full trajectory of outcomes and treatment effects.[2]

[2]The selection model is similar to a Roy model with a non-separable index. See Roy (1951); Heckman (1979); Heckman and Vytlacil (2005).

Define $E_i(z) \in \mathcal{E} \equiv \{1, \dots, \bar{t}, \infty\}$ to be the first period agent i takes up treatment when Z_i is fixed to be z.[3] We use the convention that $E_i(z) = \infty$ if the agent does not take up treatment before $\bar{t}$. Each value of A_i implies a specific $E_i(z)$ for each $z \in \mathcal{Z}$, and different combinations of A_i can generate the same combination of potential treatment take-up periods, $\{E_i(z)\}$. Agents can thus be typified by their combination $\{E_i(z)\}$. We show below how average outcomes and average treatment effects specific to types can be identified. However, this becomes difficult when the number of types is large. This can be addressed by imposing restrictions on A_i to exclude certain combinations of $\{E_i(z)\}$ and reduce the number of types.[4] Alternatively, we can implicitly restrict A_i by instead restricting $\{E_i(z)\}$ directly, i.e., explicitly state the combinations of $\{E_i(z)\}$ that are permissible.

Remark 2.1. The model can be generalized so that treatment is not available in every period prior to $\bar{t}$. IV shocks also do not need to be as varied as described above. For instance, not all $2^{\bar{t}}$ sequences of IV shocks must be realized. Similarly, there does not need to be a new shock every period. ■

2.2 An interpretable model

The generality of $h_t(\cdot)$ in (1) permits a wide variety of selection patterns but limits the interpretability of the model. However, for cases where interpretability is desired, we propose the following parameterization of $h_t(\cdot)$ that models selection into treatment along two dimensions: sensitivity to time and sensitivity to the IV. Sensitivity to time pertains to an agent's desire to take up treatment by a certain period. Sensitivity to the IV pertains to how an agent adjusts their timing of treatment in response to each IV shock.

Let

$$h_t(Z_{i,1}, \dots, Z_{i,t}; A_i) = -A_{1i} + \sum_{s=1}^{t} A_{2i,s} + \sum_{s'=1}^{t} A_{3i,s'} Z_{i,s'},$$

so that

$$D_{i,t} = \max \left\{ D_{i,t-1},\ \mathbb{1}\left[A_{i1} \leq \sum_{s=1}^{t} A_{2i,s} + \sum_{s'=1}^{t} A_{3i,s'} Z_{i,s'} \right] \right\}. \tag{2}$$

[3] More explicitly, define $D_{i,t}(z)$ to be the potential treatment decision in period t fixing $Z_i = z$. Then $E_i(z) \equiv \min\{t : D_{i,t}(z) = 1\}$.

[4] Since $Z_{i,t}$ is binary and there are $\bar{t}$ periods, there are up to $2^{\bar{t}}$ possible values of Z_i. For each Z_i, there are $\bar{t}+1$ possible periods the agent takes up treatment, with $E_i(Z_i) \in \{1, \dots, \bar{t}, \infty\}$. Absent any restrictions on A_i, there can be up to $(\bar{t}+1)^{2^{\bar{t}}}$ distinct types.

A_{1i} plays the role of a threshold and can be thought of as the agent's cost of taking treatment. The sum $\sum_{s=1}^{t} A_{2is} + \sum_{s'=1}^{t} A_{3i,s'} Z_{i,s'}$ is a latent index and can be thought of as the agent's benefit of taking up treatment in period t. $A_{2i,t}$ is the agent's sensitivity to time, with $A_{2i,t} > 0$ indicating greater benefit of taking up treatment after period t and $A_{2i,t} < 0$ indicating lesser benefit. $A_{3i,t}$ is the agent's sensitivity to the IV, with $A_{3i,t} > 0$ indicating greater benefit of taking up treatment after an IV shock in period t and $A_{3i,t} < 0$ indicating lesser benefit. Selection into treatment is thus characterized by $A_i \equiv (A_{1i}, \{A_{2i,t}\}, \{A_{3i,t}\})$. The summations of $A_{2i,t}$ and $A_{3i,t}$ across t allow for dynamic responses to time and the IV, whereby an agent's treatment decision depends on the time periods elapsed and their history of IV shocks.

Equation (2) permits a rich set of models for when and whether agents select into treatment, some of which we demonstrate below.

Remark 2.2. The monotonicity assumption of Imbens and Angrist (1994) restricts $A_{3i,t}$ to have the same sign for all i. Such an assumption can aid identification of the distribution of agent types and treatment effects by reducing the number of agent types. We show below that the monotonicity assumption is not required in our setting. Moreover, we show that we can identify the distribution of agent types and treatment effects even when $A_{3i,t}$ varies in sign across t within i. That is, our framework allows IV shocks from one period to encourage a subset of agents to take up treatment earlier, and shocks from another period to encourage the same subset of agents to take up treatment later. We provide an example of this in Section 3.3.3. ■

2.2.1 Threshold model

The first set of restrictions generates a threshold model whereby agents either take up treatment immediately or never take up treatment after experiencing a sufficient number of shocks.

Suppose agent i intends to enroll in treatment in period $E_i^\star \in \mathcal{E}$ absent any IV shocks. This event is realized when $Z_i = (0, \dots, 0)$ and implies $E_i(0, \dots, 0) = E_i^\star$. Suppose agent i is willing to deviate from $E_i^\star$ and take up treatment immediately if they have experienced $K_i \in \mathbb{N}$ shocks. Agents with small K_i are sensitive to the IV and agents with large K_i are insensitive to the IV. We adopt the convention that $K_i = \infty$ for agents who are unresponsive to the IV. For agents with $E_i^\star \leq \bar{t}$, the behavior described above can be characterized by[5]

$$\begin{aligned}
A_{1i} &> 0, \\
A_{2i,t} &= 0 \quad \text{for } t < E_i^\star, \\
A_{2i,t} &> A_{1i} \quad \text{for } t \geq E_i^\star, \\
A_{3i,t} &= \frac{A_{1i}}{K_i} \quad \text{for all } t.
\end{aligned}$$

For agents with $E_i^\star = \infty$, we update the restrictions above to have $A_{2i,t} = 0$ for all t. We refer to agents who are induced into treatment by the IV as *in-compliers*, and agents who are unresponsive to the IV as *non-compliers*.

The effect of the IV on treatment timing can also go in the opposite direction, where agents deviate from $E_i^\star$ by never taking up treatment once experiencing K_i shocks. For agents with $E_i^\star < \infty$, we update $A_{2i,t}$ and $A_{3i,t}$ so that[6]

$$\begin{aligned}
A_{1i} < &A_{2i,t} < 2A_{1i} \quad \text{for } t = E_i^\star, \\
&A_{2i,t} = 0 \quad \text{for } t > E_i^\star \\
&A_{3i,t} = -\frac{A_{1i}}{K_i} \quad \text{for all } t.
\end{aligned}$$

For agents with $E_i^\star = \infty$, we further update $A_{2i,t}$ so that $A_{2i,t} = 0$ for all t. We refer to agents who are induced out of treatment by the IV as *out-compliers*.

Table 1 presents the complete set of agent types under this threshold model when $\bar{t} = 2$. In the top panel, each row corresponds to a sequence of IV shocks, each column corresponds to an agent type, and each entry of the table corresponds to the period of treatment take-up. The bottom panel shows the number of shocks required to induce each agent type into or out of treatment.

When the number of types is large, it may be impossible to identify average treatment effects for any type of agent. A natural solution is to further restrict the number of types by imposing assumptions such as monotonicity. For instance, if the IV is assumed to weakly induce all agents into treatment, then agent types are restricted to non-compliers and in-

[5]The restrictions $A_{1i} > 0$ and $A_{2i,t} = 0$ for $t < E_i^\star$ ensure agents do not take up treatment prior to $E_i^\star$ absent any IV shocks. The restriction $A_{2i,t} > A_{1i}$ for $t \geq E_i^\star$ ensures the agent takes up treatment no later than $E_i^\star$, even in the absence of IV shocks. The restriction $A_{3i,t} = \frac{A_{1i}}{K_i}$ ensures agents must experience K_i shocks by period $t < E_i^\star$ if they are to take up treatment prior to $E_i^\star$.

[6]The restriction $A_{1i} < A_{2i,t}$ for $t = E_i^\star$ ensures agents take up treatment by period $E_i^\star$ absent any IV shocks. The restrictions $A_{2i,t} < 2A_i$ for $t = E_i^\star$ and $A_{3i,t} = -\frac{A_{1i}}{K_i}$ for all t ensure that the agent does not take treatment in period $E_i^\star$ if she has received K_i shocks by period $E_i^\star$. The restriction $A_{2i,t} = 0$ for $t > E_i^\star$ ensures the agent remains untreated in periods $t > E_i^\star$ if she has received K_i shocks by period $E_i^\star$.

Table 1: Example of agent types under a threshold model

	Type								
	Non-compliers			In-compliers			Out-compliers		
Z_i	1	2	3	4	5	6	7	8	9
$(0,0)$	1	2	∞	2	∞	∞	1	2	2
$(0,1)$	1	2	∞	2	2	∞	1	∞	2
$(1,0)$	1	2	∞	1	1	∞	∞	∞	2
$(1,1)$	1	2	∞	1	1	2	∞	∞	∞
K_i	∞	∞	∞	1	1	2	1	1	2

Notes: Each row corresponds to a sequence of IV shocks. Each column corresponds to a type of agents. Each entry in the table indicates the period a type of agent takes up treatment fixing their sequence of IV shocks.

compliers. If instead the IV is assumed to weakly induce all agents out of treatment, then agent types are restricted to non-compliers and out-compliers. However, monotonicity is not a necessary condition for identification. In Section 3.3.3, we provide examples where identification is obtained even when the IV induces some agents into treatment and other agents out of treatment.

2.2.2 Incremental model

An alternative response to the IV is an incremental deviation from $E_i^\star$, whereby K_i shocks are required to shift the timing of treatment one period earlier or later. Unlike the threshold model where $K_i \in \mathbb{N}$, the incremental model permits $K_i \in \mathbb{R}_{>0}$, with $K_i \in [0,1)$ implying that each IV shock shifts the timing of treatment $\lfloor K_i^{-1} \rfloor \geq 1$ periods earlier or later. For agents for whom the IV shifts the timing of treatment earlier and $E_i^\star < \bar{t}$, this behavior can be characterized by[7]

$$\begin{aligned} A_{1i} &> 0, \\ A_{2i,t} &= \frac{A_{1i}}{E_i^\star} \quad \text{for all } t, \\ A_{3i,t} &= \frac{A_{2i,t}}{K_i} \quad \text{for all } t. \end{aligned}$$

If instead the IV shifts the timing of the treatment later, then we simply update the restrictions above so that $A_{3i,t} = -A_{2i,t}$ for all t. For agents with $E_i^\star = \infty$, we update the restrictions above so that $A_{2i,t} = A_{1i}/(\bar{t}+1)$.

[7] The restriction $A_{2i,t} = \frac{A_{1i}}{E_i^\star}$ implies the agents require $E_i^\star$ periods to pass absent any IV shocks in order to take up treatment. The restriction $A_{3i,t} = \frac{A_{2i,t}}{K_i}$ implies that K_i shocks has the same effect as one period passing.

Table 2: Example of agent types under an incremental model

	Type								
	Non-compliers			In-compliers			Out-compliers		
Z_i	1	2	3	4	5	6	7	8	9
$(0,0)$	1	2	∞	2	∞	∞	1	2	2
$(0,1)$	1	2	∞	2	2	∞	1	∞	2
$(1,0)$	1	2	∞	1	2	∞	2	∞	2
$(1,1)$	1	2	∞	1	2	2	∞	∞	∞
K_i	∞	∞	∞	1	1	2	1	1	2

Notes: Each row corresponds to a sequence of IV shocks. Each column corresponds to a type of agents. Each entry in the table indicates the period a type of agent takes up treatment fixing their sequence of IV shocks.

Table 2 presents the set of agent types under the incremental model when $\bar{t} = 2$, $K_i \in \{1, 2, \infty\}$. Note that other than types 5 and 7, all other agent types are also found in Table 1. This is because there are only two periods of treatment available, which restricts how agents can respond to the IV. For larger values of $\bar{t}$, differences between the threshold and incremental models become more evident. This is demonstrated in the examples in Section 3.3.3.

2.2.3 More general models

An alternative approach to modeling selection is to directly construct $\{E_i(z)\}$ for each type of agent, similar to the principal stratification framework of Angrist et al. (1996). So rather than derive the agent types from (2) and restrictions on A_i, we instead construct a table of agent types similar to Tables 1–2 and check whether the agent types can be generated under (2). Since (2) is linear in A_i, checking whether the agent types can be generated by (2) entails checking the feasibility of a set of linear constraints, which is easily done using modern solvers.[8] If the linear constraints are feasible, then the construction of $\{E_i(z)\}$ is consistent with (2). If the linear constraints are not feasible, then selection

[8]Under (2), if there exists an agent i taking up treatment in period $e \in \mathcal{E}$, then there must exist A_i satisfying

$$A_{i1} > \sum_{s=1}^{t} A_{2i,s} + \sum_{s'=1}^{t} A_{3i,s'} Z_{i,s'} \text{ for } t < e,$$
$$A_{i1} \leq \sum_{s=1}^{e} A_{2i,s} + \sum_{s'=1}^{e} A_{3i,s'} Z_{i,s'}.$$

Checking the feasibility of such linear constraints is easily done by solvers such as Gurobi, Mosek, and CPLEX.

into treatment should instead be interpreted using a more general model.

3 Identification

We begin by stating the econometric assumptions imposed throughout the paper. As part of this, we propose two frameworks for treatment effect heterogeneity. Let *treatment cohort* refer to the set of agents taking up treatment in a particular period $e \in \mathcal{E}$. The first framework is akin to the local average treatment effect (LATE) framework of Imbens and Angrist (1994), but we extend the framework to accommodate multiple complier and non-complier types, and we allow average treatment effects to vary by agent type but not by treatment cohort. The second framework is akin to the event study framework, and instead allows average treatment effects to vary by treatment cohort but not by agent type.[9]

Our first identification result shows that the distribution of agent types can be identified under either framework. We also discuss testable implications of the selection model from Section 2. Our second identification result pertains to the treatment effects. We show that, under the first framework, dynamic treatment effects specific to certain complier types can be identified. Similarly, under the second framework, dynamic treatment effects specific to certain cohorts can be identified.

3.1 Notation and assumptions

Let $G_i \in \mathcal{G} = \{1, 2, \ldots, \overline{g}\}$ denote the agent type based on $\{E_i(z)\}$. The total number of types $\overline{g}$ is equal to the number of unique combinations of $\{E_i(z)\}$, which is either indirectly determined by the restrictions imposed on A_i or directly determined by the set of values of $\{E_i(z)\}$ permitted. Since $\{E_i(z)\}$ is unobserved in the data, agent type G_i is also unobserved.

Assumption 1. $\mathbb{P}[G_i = g] > 0$ *for all* $g \in \mathcal{G}$.

Assumption 1 states that there must be a strictly positive share of each agent type. This is a mild assumption as we focus our analysis on types that exist in the population and do not seek to extrapolate to types outside of the population.

Let $Y_{i,t}(\infty)$ denote the potential untreated outcome of agent i in period t. We impose no restrictions on the evolution of $Y_{i,t}(\infty)$ over time nor across agent types. This contrasts with the DID framework, where average untreated outcomes are assumed to evolve in parallel across treated and untreated groups. This parallel trends assumption allows

[9] Allowing for treatment effect heterogeneity across both agent type and treatment cohort is ongoing work.

researchers to infer the untreated counterfactual for the treated group from the untreated units. However, because treatment is endogenous in our setting, agents who opt into treatment in one period can differ from agents who opt into treatment in another period or do not opt into treatment at all. For this reason, we neither invoke the parallel trend assumption nor impose any restrictions on the untreated trends across agent types.

Let $\Delta_i(e, s)$ denote the treatment effect for the agent s periods after taking up treatment in period e. $\Delta_i(e, 0)$ corresponds to the treatment effect at the period of take-up.

Assumption 2. $\Delta_i(e, s) = 0$ *for* $s < 0$.

Assumption 2 imposes that the treatment effects are 0 prior to treatment take-up. This differs from the no-anticipation assumption standard in the DID framework, which restricts the average untreated potential outcomes to be invariant across timing of treatment. The no-anticipation assumption supports the parallel trends assumption and allows researchers to infer the untreated counterfactual for the treated units from the observed untreated outcomes of the not-yet-treated units. We do not impose the no-anticipation assumption since treatment is endogenous in our setting, suggesting agents likely anticipate treatment and potentially adjust their behavior shortly before treatment. The treatment effects we identify are thus relative to the untreated outcomes containing any anticipatory effects of treatment.[10] This does not invalidate our method as we do not rely on parallel trends for identification, but only requires that we interpret the treatment effects as previously described.

Let $E_i \equiv \sum_{z \in \mathcal{Z}} \mathbb{1}[Z_i = z]\ E_i(z)$ denote the observed period agent i takes up treatment. The observed outcome in the data is

$$Y_{i,t} \equiv \underbrace{Y_{i,t}(\infty)}_{\text{Untreated outcome}} + \overbrace{\mathbb{1}[t \geq E_i]\ \Delta_i(E_i, t - E_i)}^{\text{Treatment effect, if treated}}.$$

The first framework we propose allows the average treatment effects to vary across types G_i but not cohort E_i. This requires us to restrict $\Delta_i(e, s)$ to depend only on s.[11]

Assumption 3. $\Delta_i(e, s) = \Delta_i(s)$.

[10] Anticipation effects can be incorporated into the model as additional parameters, but treatment effects may only be partially identified in such cases.

[11] Assumption 3 is similar to the *wave ignorability* assumption of Angrist et al. (2024) and Ferman and Tecchio (2025).

The parameter of interest in this setting is

$$\Delta^1(g, s) \equiv \mathbb{E}[\Delta_i(s) \mid G_i = g]$$

for $g \in \mathcal{G}$ and $s \in \{0, \ldots, \bar{t} - 1\}$. $\Delta^1(g, s)$ is the type-specific average treatment effect s periods after treatment and is analogous to the LATE. Identification of $\{\Delta^1(g, s)\}$ is obtained by invoking the following assumption on Z_i.

Assumption 4a.

(i) (Exogeneity/exclusion) $Z_i \perp\!\!\!\perp (G_i, \{Y_{i,t}(\infty)\}, \{\Delta_i(s)\})$.

(ii) (Relevance) $Z_i \not\perp\!\!\!\perp E_i$.

Assumption 4a(i) states that the instrument Z_i is independent of agent types, untreated potential outcomes, and treatment effects. Assumption 4a(ii) states that the instrument can induce variation in the timing of treatment take-up.

The second framework we propose allows the average treatment effects to vary across cohort E_i but not types G_i. By restricting the average effects to be homogeneous across agent types, we do not require Assumption 3. The parameter of interest is

$$\Delta^2(e, s) \equiv \mathbb{E}[\Delta_i(e, s)]$$

for $e \in \mathcal{T}$ and $s \in \{0, \ldots, \bar{t} - 1\}$. $\Delta^2(e, s)$ is the cohort-specific treatment effect s periods after treatment. Identification of $\{\Delta^2(g, s)\}$ is obtained by invoking the following assumption on Z_i.

Assumption 4b.

(i) (Exogeneity/exclusion 1) $Z_i \perp\!\!\!\perp (G_i, \{Y_{i,t}(\infty)\})$.

(ii) (Exogeneity/exclusion 2) $(Z_i, G_i) \perp\!\!\!\perp \{\Delta_i(e, s)\}$.

(iii) (Relevance) $Z_i \not\perp\!\!\!\perp E_i$.

Assumption 4b(i) states that the instrument is independent of agent types and untreated potential outcomes. Assumption 4b(ii) states that both the instrument and agent types are independent of treatment effects. Assumption 4b(iii) states that the instrument can induce variation in the timing of treatment take-up.

3.2 First stage: Identifying distribution of types

3.2.1 Identification of types

The distribution of agent types is identified by exogenously varying E_i via Z_i. The response of E_i to changes in Z_i reveals information on the distribution of G_i since the response depends on type G_i. For example, if the distribution of E_i changes greatly with Z_i, that not only suggests the instrument is relevant, but also that there may be a large share of agent types who are highly responsive to the IV.

To formalize this argument, define $\check{e}(z,g) \equiv \{E_i(z) \mid G_i = g\}$ to be the (deterministic) function returning the enrollment period for agents of type g when fixing the IV to be z. Assumptions 4a(i) and 4b(i) imply $Z_i \perp\!\!\!\perp G_i$, which in turn implies $\mathbb{P}[G_i = g \mid Z_i] = \mathbb{P}[G_i = g]$ for all $g \in \mathcal{G}$. This allows us to express the share of agents taking up treatment in a specific period conditional on Z_i as the total mass of some combination of agent types,

$$\begin{aligned}\mathbb{P}[E_i = e \mid Z_i = z] &= \sum_{g \in \mathcal{G}} \mathbb{1}[\check{e}(z,g) = e] \underbrace{\mathbb{P}[G_i = g \mid Z_i = z]}_{\text{As. 4a, 4b} \Rightarrow Z_i \perp\!\!\!\perp G_i} \\ &= \sum_{g \in \mathcal{G}} \mathbb{1}[\check{e}(z,g) = e]\ \mathbb{P}[G_i = g]. \end{aligned} \tag{3}$$

Moreover, the combination of agent types is known since selection into treatment is modeled in Section 2. Differences in the observed probabilities $\mathbb{P}[E_i = e \mid Z_i = z]$ across $e \in \mathcal{E}$ and $z \in \mathcal{Z}$ can therefore be mapped to differences in the combinations of agents. The share of a particular agent type is identified when differences in $\mathbb{P}[E_i = e \mid Z_i = z]$ can be mapped to a single agent type.

To state our identification result on the distribution of types, define the matrix $\mathbf{M}^e \in \{0,1\}^{|\mathcal{Z}| \times \bar{g}}$ to be a matrix of indicators for whether each type of agent takes up treatment in period e under each value of the IV. Specifically, define the entry in the j^{th} row and k^{th} column of $\mathbf{M}^e$ to be

$$\mathbf{M}^e_{jk} \equiv \mathbb{1}[\check{e}(z_j, g_k) = e],$$

where $z_j \in \mathcal{Z}$ and $g_k \in \mathcal{G}$. Define

$$\mathbf{A}^1 \equiv \begin{bmatrix} \mathbf{M}^1 \\ \vdots \\ \mathbf{M}^{\bar{t}} \\ \mathbf{M}^{\infty} \end{bmatrix}$$

to be the binary matrix stacking $\mathbf{M}^e$ for all $e \in \mathcal{E}$. By construction, each row of $\mathbf{A}^1$ lists the combination of agent types selected when conditioning on E_i and Z_i. If a particular type can be isolated using elementary row operations on $\mathbf{A}^1$, then the mass of that type is identified. This leads to our first identification result.

Proposition 1. *Suppose Assumptions 1, and either 4a or 4b, hold. If* $\mathbf{A}^1$ *is full rank, then the distribution of* G_i *is identified.*

All proofs can be found in Appendix A.

To provide an example of the identification argument, consider the case where $\bar{t} = 2$ and selection is modeled using the threshold model from Section 2.2.1. Further impose the monotonicity restriction that the IV weakly induces agents into treatment earlier so that only the non-compliers and in-compliers of Table 1 exist. The implied forms of $\mathbf{M}^1$, $\mathbf{M}^2$, $\mathbf{M}^\infty$, and (3) are shown in Figure 1. The order of the columns in Figure 1 have been adjusted and differ from that of Table 1, revealing triangular systems in $\mathbf{M}^1$, $\mathbf{M}^2$, and $\mathbf{M}^\infty$ (see the boxed entries). These triangular systems make it straightforward to solve for the distribution of agent types.

3.2.2 Testable implications of model misspecfication

The identification argument in Proposition 1 involves solving a linear system of $|\mathcal{Z}| \times |\mathcal{E}|$ equations and $\overline{g}$ unknowns. If the selection model results in $\overline{g} < |\mathcal{Z}| \times |\mathcal{E}|$, then the overidentifying restrictions may be used to test for model misspecification.

A simple example of an overidentifying restriction is the set of rows in $\mathbf{M}^e$ for $e \in \mathcal{E}$ that are identical to each other. For example, in Figure 1, the rows corresponding to z_1 and z_2 in $\mathbf{M}^1$ are identical, which implies $\mathbb{P}[E_i = 1 \mid Z_i = z_1] = \mathbb{P}[E_i = 1 \mid Z_i = z_2]$. Similarly, the rows corresponding to z_2 and z_3 in $\mathbf{M}^\infty$ are identical, which implies $\mathbb{P}[E_i = \infty \mid Z_i = z_2] = \mathbb{P}[E_i = \infty \mid Z_i = z_3]$. If these equalities do not hold in the population, that implies the selection model is misspecified.

More generally, any linearly dependent combination of rows from $\mathbf{M}^e$ implying

$$\mathbb{P}[E_i = e \mid Z_i = z] = \mathbb{P}[E_i = e' \mid Z_i = z'] \tag{4}$$

for $e, e' \in \mathcal{E}$ and $z, z' \in \mathcal{Z}$ may be used to test for model misspecification. If the equalities do not hold in the population, that implies the selection model is misspecified. We propose a statistical test for model misspecification in Section 4.4.

Figure 1: Example of $\mathbf{M}^1$, $\mathbf{M}^2$, $\mathbf{M}^\infty$

(a) Equation (3) under $\mathbf{M}^1$

$$\underbrace{\begin{array}{c} \\ z_1 \\ z_2 \\ z_3 \\ z_4 \end{array}\begin{array}{c} \begin{array}{cccccc} g_1 & g_2 & g_3 & g_4 & g_5 & g_6 \end{array} \\ \begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array}}_{\mathbf{M}^1} \begin{bmatrix} \mathbb{P}[G_i = g_1] \\ \vdots \\ \mathbb{P}[G_i = g_6] \end{bmatrix} = \begin{bmatrix} \mathbb{P}[E_i = 1 \mid Z_i = z_1] \\ \vdots \\ \mathbb{P}[E_i = 1 \mid Z_i = z_4] \end{bmatrix}$$

(b) Equation (3) under $\mathbf{M}^2$

$$\underbrace{\begin{array}{c} \\ z_1 \\ z_2 \\ z_3 \\ z_4 \end{array}\begin{array}{c} \begin{array}{cccccc} g_1 & g_2 & g_3 & g_4 & g_5 & g_6 \end{array} \\ \begin{bmatrix} 0 & \boxed{1} & \boxed{1} & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \end{array}}_{\mathbf{M}^2} \begin{bmatrix} \mathbb{P}[G_i = g_1] \\ \vdots \\ \mathbb{P}[G_i = g_6] \end{bmatrix} = \begin{bmatrix} \mathbb{P}[E_i = 2 \mid Z_i = z_1] \\ \vdots \\ \mathbb{P}[E_i = 2 \mid Z_i = z_4] \end{bmatrix}$$

(c) Equation (3) under $\mathbf{M}^\infty$

$$\underbrace{\begin{array}{c} \\ z_1 \\ z_2 \\ z_3 \\ z_4 \end{array}\begin{array}{c} \begin{array}{cccccc} g_1 & g_2 & g_3 & g_4 & g_5 & g_6 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & \boxed{1} & \boxed{1} & \boxed{1} \\ 0 & 0 & 0 & 0 & \boxed{1} & \boxed{1} \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & \boxed{1} \end{bmatrix} \end{array}}_{\mathbf{M}^\infty} \begin{bmatrix} \mathbb{P}[G_i = g_1] \\ \vdots \\ \mathbb{P}[G_i = g_6] \end{bmatrix} = \begin{bmatrix} \mathbb{P}[E_i = \infty \mid Z_i = z_1] \\ \vdots \\ \mathbb{P}[E_i = \infty \mid Z_i = z_4] \end{bmatrix}$$

Note: The rows of each matrix correspond to different values of the instrument, with $z_1 = (0, 0)$, $z_2 = (0, 1)$, $z_3 = (1, 0)$, $z_4 = (1, 1)$. The columns of each matrix correspond to a different type of agent. The boxed numbers reveal triangular systems in $\mathbf{M}^1$, $\mathbf{M}^2$, $\mathbf{M}^\infty$ that can be solved to recover the share of each type of agent.

3.2.3 Testable implications of IV relevance

The relevance condition requires the IV affect the timing of treatment for some share of agents. This implies there must exist some $e \in \mathcal{E}$ and $z, z' \in \mathcal{Z}$ such that

$$\{g \in \mathcal{G} : \check{e}(z, g) = e\} \neq \{g \in \mathcal{G} : \check{e}(z', g) = e\}. \tag{5}$$

If (5) does not hold, then the composition of agents taking up treatment each period is the same across all values of the IV, indicating the IV does not induce any agent types to change their treatment timing. To test whether (5) is satisfied, suppose the selection model is correctly specified. If there exist $e \in \mathcal{E}$ and $z, z' \in \mathcal{Z}$ such that

$$\mathbb{P}[E_i = e \mid Z_i = z] \neq \mathbb{P}[E_i = e \mid Z_i = z'], \tag{6}$$

then it must be the case that the composition of agents taking up treatment in period e has changed across the IV values and (5) is satisfied.[12] Following this logic, we propose a statistical test for instrumental relevance in Section 4.4.

3.3 Second stage: Identifying average treatment effects

In this section we discuss identification of dynamic treatment effects under both frameworks for treatment effect heterogeneity. Depending on the selection model, we are able to point identify either the full trajectory of treatment effects for certain agent types and cohorts, or a subset of treatment effects.[13]

3.3.1 Heterogeneous treatment effects across agent types

We first consider the case where the effects depend on agent type but not timing of treatment. Throughout, we assume that Assumptions 1–3 and 4a hold.

From the data, we observe

$$\mathbb{E}[Y_{i,t} \mid E_i, Z_i] = \mathbb{E}[Y_{i,t}(\infty) \mid E_i, Z_i] + \underbrace{\mathbb{E}[\Delta_i(t - E_i) \mid E_i, Z_i]}_{\text{As. 3 applies}}.$$

[12] Note that if there does not exist $e \in \mathcal{E}$ and $z, z' \in \mathcal{Z}$ satisfying (6), that need not imply (5) is not satisfied. If (5) holds, (6) can still fail to be satisfied if the mass of agents switching out of taking treatment in one period is replaced by an equal mass of agents switching into taking treatment that period when changing values of the IV.

[13] Since identification is obtained by solving a linear system, it is straightforward to obtain bounds on treatment effects that are not point identified.

We can write

$$\begin{aligned}
&\mathbb{E}[Y_{it}(\infty) \mid E_i = e, Z_i = z] \\
&= \sum_{g \in \mathcal{G}} \mathbb{E}[Y_{it}(\infty) \mid \underbrace{E_i = e, Z_i = z, G_i = g}_{E_i \text{ is determined given } Z_i,\, G_i}] \; \underbrace{\mathbb{P}[G_i = g \mid E_i = e, Z_i = z]}_{0 \text{ for } g \text{ s.t. } \check{e}(z,g) \neq e} \\
&= \sum_{g \in \mathcal{G}} \underbrace{\mathbb{E}[Y_{it}(\infty) \mid Z_i = z, G_i = g]}_{\text{As. 4a} \Rightarrow Y_{it}(\infty) \perp\!\!\!\perp Z_i \mid G_i} \; \frac{\mathbb{1}[\check{e}(z,g) = e] \; \mathbb{P}[G_i = g]}{\sum_{g' \in \mathcal{G}} \mathbb{1}[\check{e}(z,g') = e] \; \mathbb{P}[G_i = g']} \\
&= \sum_{g \in \mathcal{G}} \mathbb{E}[Y_{it}(\infty) \mid G_i = g] \; \frac{\mathbb{1}[\check{e}(z,g) = e] \; \mathbb{P}[G_i = g]}{\sum_{g' \in \mathcal{G}} \mathbb{1}[\check{e}(z,g') = e] \; \mathbb{P}[G_i = g']}. \qquad (7)
\end{aligned}$$

The first equality follows by the law of iterated expectations (LIE). The second equality follows from E_i being fully determined by Z_i and G_i, and rewriting $\mathbb{P}[G_i = g \mid E_i = e, Z_i = z]$. The final equality follows from Assumption 4a(i), which implies $Y_{it}(\infty) \perp\!\!\!\perp Z_i \mid G_i$.

By the same logic, we can write

$$\begin{aligned}
&\mathbb{E}[\Delta_i(t - E_i) \mid E_i = e, Z_i = z] \\
&= \sum_{g \in \mathcal{G}} \mathbb{E}[\Delta_i(t-e) \mid \underbrace{E_i = e, Z_i = z, G_i = g}_{E_i \text{ is determined given } Z_i,\, G_i}] \; \underbrace{\mathbb{P}[G_i = g \mid E_i = e, Z_i = z]}_{0 \text{ for } g \text{ s.t. } \check{e}(z,g) \neq e} \\
&= \sum_{g \in \mathcal{G}} \underbrace{\mathbb{E}[\Delta_i(t-e) \mid Z_i = z, G_i = g]}_{\text{As. 4a} \Rightarrow \Delta_i(t-e) \perp\!\!\!\perp Z_i \mid G_i} \; \frac{\mathbb{1}[\check{e}(z,g) = e] \; \mathbb{P}[G_i = g]}{\sum_{g' \in \mathcal{G}} \mathbb{1}[\check{e}(z,g') = e] \; \mathbb{P}[G_i = g']} \\
&= \sum_{g \in \mathcal{G}} \underbrace{\mathbb{E}[\Delta_i(t-e) \mid G_i = g]}_{\Delta^1(g,t-e)} \; \frac{\mathbb{1}[\check{e}(z,g) = e] \; \mathbb{P}[G_i = g]}{\sum_{g' \in \mathcal{G}} \mathbb{1}[\check{e}(z,g') = e] \; \mathbb{P}[G_i = g']} \\
&= \sum_{g \in \mathcal{G}} \Delta^1(g, t-e) \; \frac{\mathbb{1}[\check{e}(z,g) = e] \; \mathbb{P}[G_i = g]}{\sum_{g' \in \mathcal{G}} \mathbb{1}[\check{e}(z,g') = e] \; \mathbb{P}[G_i = g']}. \qquad (8)
\end{aligned}$$

Identification of treatment effects is established by expressing $\mathbb{E}[Y_{i,t} \mid E_i, Z_i]$ as a linear combination of $\{\mathbb{E}[Y_{i,t}(\infty) \mid G_i = g]\}$ and $\{\Delta^1(g,s)\}$, with the weight for each agent type being $\frac{\mathbb{1}[\check{e}(z,g)=e] \; \mathbb{P}[G_i=g]}{\sum_{g' \in \mathcal{G}} \mathbb{1}[\check{e}(z,g')=e] \; \mathbb{P}[G_i=g']}$. Define

$$\mathbf{b}^2 \equiv \begin{bmatrix} \mathbb{E}[Y_{i,1} \mid E_i = 1, Z_i = z_1] \\ \vdots \\ \mathbb{E}[Y_{i,\bar{t}} \mid E_i = \bar{t}, Z_i = z_{|\mathcal{Z}|}] \end{bmatrix} \qquad (9)$$

to be the vector of moments $\{\mathbb{E}[Y_{i,t} \mid E_i = e, Z_i = z]\}$ observed in the data.[14] Define

$$\mathbf{x}_t^{Y(\infty)} \equiv \begin{bmatrix} \mathbb{E}[Y_{i,t}(\infty) \mid G = g_1] \\ \vdots \\ \mathbb{E}[Y_{i,t}(\infty) \mid G = \overline{g}] \end{bmatrix}, \quad \mathbf{x}_s^{\Delta^1} \equiv \begin{bmatrix} \Delta^1(g_1, s) \\ \vdots \\ \Delta^1(\overline{g}, s) \end{bmatrix}$$

to be the model parameters averaged over in (7)–(8), and define

$$\mathbf{x}^{Y(\infty)} \equiv \begin{bmatrix} \mathbf{x}_1^{Y(\infty)} \\ \vdots \\ \mathbf{x}_{\overline{t}}^{Y(\infty)} \end{bmatrix}, \quad \mathbf{x}^{\Delta^1} \equiv \begin{bmatrix} \mathbf{x}_0^{\Delta^1} \\ \vdots \\ \mathbf{x}_{\overline{t}-1}^{\Delta^1} \end{bmatrix}, \quad \mathbf{x}^{\text{All},1} \equiv \begin{bmatrix} \mathbf{x}^{Y(\infty)} \\ \mathbf{x}^{\Delta^1} \end{bmatrix}$$

to be the vectors stacking the model parameters across all time periods.

Proposition 2. *Suppose Assumptions 1–3 and 4a hold, and $\mathbf{A}^1$ is full rank. There exists a known matrix $\mathbf{A}^{\text{All},1}$ that is a function of $\mathbf{x}^1$ such that*

$$\mathbf{A}^{\text{All},1}\mathbf{x}^{\text{All},1} = \mathbf{b}^2. \tag{10}$$

In general, $\mathbf{A}^{\text{All},1}$ is not full rank, implying not all entries of $\mathbf{x}^{\text{All},1}$ are point identified. This is expected as $\mathbf{x}^{\text{All},1}$ potentially includes treatment effects for agent types that are never treated. Since (14) is a linear system of equations, it is straightforward to determine which entries of $\mathbf{x}^{\text{All},1}$ are uniquely determined from the system and solve for them.

Remark 3.1. Intuitively, our identification strategy can be thought of as a within-agent type DID. Conditional on Z_i and E_i, we observe a weighted average of time series across agent types. Since the weights are identified from the first stage, we are able to disaggregate the average time series and recover the time series specific to certain agent types. By varying Z_i, we can potentially recover time series for complier types over multiple treatment take-up periods. We can then compare the time series with later take-up against the time series with earlier take-up to recover type-specific treatment effects. ■

Remark 3.2. For clarity, we present our results in the setting where agents are eligible for treatment in all periods. Our results generalize to settings where treatment eligibility is limited to a window of time. Consider a panel where all agents are initially ineligible for treatment for some periods (pre-treatment periods), then become eligible to enroll in treatment (treatment/post-treatment periods), before becoming ineligible again to enroll (post-eligibility period). Agents who enroll in treatment during the treatment periods remain treated in the post-eligibility periods. Agents who do not enroll during

[14] The vector $\mathbf{b}^2$ stacks $\mathbb{E}[Y_{i,t} \mid E_i = e, Z_i = z]$ across all time periods, $e \in \mathcal{E}$, and $z \in \mathcal{Z}$.

the treatment periods remain untreated during the post-eligibility periods. Our method can be applied to panels containing the treatment periods, and any combination of pre-treatment and post-eligibility periods. All proofs are written in this generalized context. ■

Remark 3.3. For entries in $\mathbf{x}^{\text{All},1}$ that are not uniquely determined, we can obtain bounds on those entries by solving the optimization problem

$$\begin{aligned} \min_{\mathbf{x}^{\text{All},1}}/\max \quad & \mathbf{x}_j^{\text{All},1} \\ \text{s.t.} \quad & \mathbf{A}^{\text{All},1}\mathbf{x}^{\text{All},1} = \mathbf{b}^2 \\ & \mathbf{C}^{\text{All},1}\mathbf{x}^{\text{All},1} \leq \mathbf{d}^{\text{All},1} \end{aligned} \tag{11}$$

where $\mathbf{x}_j^{\text{All},1}$ is the j^{th} entry of $\mathbf{x}^{\text{All},1}$; and $\mathbf{C}^{\text{All},1}$ is a known matrix and $\mathbf{d}^{\text{All},1}$ is a known vector characterizing additional restrictions on the model. The additional restrictions are necessary in order for the bounds to be nontrivial. ■

3.3.2 Heterogeneous treatment effects across cohorts

In the framework where effects are heterogeneous across cohorts, the identification argument is the same as above. Throughout, we assume that Assumptions 1–3 and 4b hold.

From the data, we observe

$$\mathbb{E}[Y_{i,t} \mid E_i, Z_i] = \mathbb{E}[Y_{i,t}(\infty) \mid E_i, Z_i] + \mathbb{E}[\Delta_i(E_i, t - E_i) \mid E_i, Z_i].$$

The term $\mathbb{E}[Y_{i,t}(\infty) \mid E_i, Z_i]$ can be written as in (7). The term $\mathbb{E}[\Delta_i(E_i, t - E_i) \mid E_i, Z_i]$ can be written as

$$\begin{aligned} &\mathbb{E}[\Delta_i(E_i, t - E_i) \mid E_i = e, Z_i = z] \\ &= \sum_{g \in \mathcal{G}} \mathbb{E}[\Delta_i(e, t - e) \mid \underbrace{E_i = e, Z_i = z, G_i = g}_{E_i \text{ is determined given } Z_i,\, G_i}]\ \mathbb{P}[G_i = g \mid E_i = e, Z_i = z] \\ &= \sum_{g \in \mathcal{G}} \underbrace{\mathbb{E}[\Delta_i(e, t - e) \mid Z_i = z, G_i = g]}_{\text{As. 4b} \Rightarrow \Delta_i(e, t-e) \perp\!\!\!\perp (Z_i, G_i)}\ \mathbb{P}[G_i = g \mid E_i = e, Z_i = z] \\ &= \sum_{g \in \mathcal{G}} \overbrace{\mathbb{E}[\Delta_i(e, t - e)]}^{\Delta^2(e, t-e),\text{ does not depend on } g}\ \mathbb{P}[G_i = g \mid E_i = e, Z_i = z] \\ &= \Delta^2(e, t - e). \end{aligned} \tag{12}$$

Define

$$\mathbf{x}_e^{\Delta^2} \equiv \begin{bmatrix} \Delta^2(e,0) \\ \vdots \\ \Delta^2(e,\bar{t}-1) \end{bmatrix}, \quad \mathbf{x}^{\Delta^2} \equiv \begin{bmatrix} \mathbf{x}_1^{\Delta^2} \\ \vdots \\ \mathbf{x}_{\bar{t}}^{\Delta^2} \end{bmatrix}$$

to be the vector of treatment effect parameters for all cohorts, and

$$\mathbf{x}^{\text{All},2} \equiv \begin{bmatrix} \mathbf{x}^{Y(\infty)} \\ \mathbf{x}^{\Delta^2} \end{bmatrix}$$

to be the vector stacking all the model parameters across all time periods.

Proposition 3. *Suppose Assumptions 1–3 and 4b hold, and $\mathbf{A}^1$ is full rank. There exists a known matrix $\mathbf{A}^{\text{All},2}$ that is a function of $\mathbf{x}^1$ such that*

$$\mathbf{A}^{\text{All},2}\mathbf{x}^{\text{All},2} = \mathbf{b}^2. \tag{13}$$

As before, $\mathbf{A}^{\text{All},2}$ is not full rank in general, so not all entries of $\mathbf{x}^{\text{All},2}$ are point identified. Entries of $\mathbf{x}^{\text{All},2}$ that are uniquely determined can be solved for.

3.3.3 Examples of selection models and identified parameters

We now present four examples of selection models and their identified treatment effects. As part of these examples, we show that longer treatment effect trajectories can be identified if the panel extends beyond the $\bar{t}$ periods when enrolling into treatment is possible. Let t^{Extend} denote the number of periods beyond $\bar{t}$ that the panel extends to. For example, if treatment is available in the first $\bar{t} = 2$ periods and the panel extends to period 5, then $t^{\text{Extend}} = 3$. In the examples below, we also show that not all possible sequences of Z_i are required to define the selection model and identify treatment effects.

Example 3.1. Suppose $\bar{t} = 2$ (i.e., treatment is available in periods 1 and 2). Consider a threshold selection model (Section 2.2.1) where agents intend to take up treatment at a given period absent IV shocks, but may immediately switch into treatment once experiencing a sufficient number of shocks. Different agent types have different intended take-up periods and require different numbers of shocks before deviating from their intended take-up period, but the IV satistifies monotonicity so that all agents are weakly induced to take up treatment earlier.

Table 3 shows all possible agent types under the model described.[15] When treatment

[15] Table 3 shows all combinations of $\{E_i(z)\}$ under the threshold model when $\bar{t} = 2$ and IV incentivizes

Table 3: Example of a threshold selection model

	Agent type, G_i					
IV shocks, Z_i	1	2	3	4	5	6
$(0,0)$	1	2	2	∞	∞	∞
$(0,1)$	1	2	2	2	∞	∞
$(1,0)$	1	1	2	1	∞	∞
$(1,1)$	1	1	2	1	2	∞
Intended	1	2	2	∞	∞	∞
Threshold	∞	1	∞	1	2	∞

Notes: Each column corresponds to an agent type. Each row corresponds to a sequence of IV shocks. Each entry in the table shows the period an agent type would take up treatment given a sequence of IV shocks. Black columns indicate compliers, gray columns indicate non-compliers. The row labeled "Intended" shows the period an agent type would take treatment subject to no IV shocks. The row labeled "Threshold" shows the number of IV shocks the agent must receive prior to the intended take-up period in order to immediately take up treatment.

effects are heterogeneous across types, then $\Delta^1(g,s)$ is identified for $s \in \{0,\ldots,t^{\text{Extend}}\}$ for all complier types $g \in \{2,4,5\}$. When treatment effects are heterogeneous across cohorts, then $\Delta^2(e,s)$ is identified for $s \in \{0,\ldots,\bar{t}+t^{\text{Extend}}-e\}$ for all $e \in \mathcal{E}$. Treatment effects are thus identified over the longest possible horizon for all cohorts.[16] ■

Example 3.2. Suppose $\bar{t} = 3$. Consider an incremental selection model (Section 2.2.2) where agents intend to take up treatment at a given period absent IV shocks, but may delay treatment with each shock. Different agent types have different intended periods of take-up and require different numbers of shocks before delaying treatment by one period, but the IV is monotonic in that all agents are weakly induced to take up treatment later. We restrict our model to agents who require one, two, or infinite (i.e., non-compliers) shocks before delaying treatment.

Table 4 shows the agent types under the model described. When treatment effects are heterogeneous across types, then $\Delta^1(g,s)$ is identified for $s \in \{0,\ldots,\bar{t}+t^{\text{Extend}}\}$ for $g=1$; $s \in \{0,\ldots,\bar{t}+t^{\text{Extend}}-1\}$ for $g \in \{3,4\}$; and $s \in \{0,\ldots,\bar{t}+t^{\text{Extend}}-2\}$ for $g \in \{6,7\}$. Treatment effects are thus identified for all complier types but for different horizons.

all agents to take up treatment earlier.

[16] A longer horizon is not possible since $\Delta_i(e,s)$ for $s > \bar{t}+t^{\text{Extend}}-e$ is never realized in the panel for any $e \in \mathcal{E}$.

Table 4: Example of an incremental selection model

	Agent type, G_i								
IV shocks, Z_i	1	2	3	4	5	6	7	8	9
$(0,0,0)$	1	1	2	2	2	3	3	3	∞
$(0,0,1)$	1	1	2	2	2	∞	3	3	∞
$(0,1,0)$	1	1	3	2	2	∞	3	3	∞
$(0,1,1)$	1	1	∞	2	2	∞	∞	3	∞
$(1,0,0)$	2	1	3	2	2	∞	3	3	∞
$(1,0,1)$	2	1	∞	2	2	∞	∞	3	∞
$(1,1,0)$	3	1	∞	3	2	∞	∞	3	∞
$(1,1,1)$	∞	1	∞	3	2	∞	∞	3	∞
Intended	1	1	2	2	2	3	3	3	∞
Increment	1	0	1	0.5	0	1	0.5	0	0

Notes: Each column corresponds to an agent type. Each row corresponds to a sequence of IV shocks. Each entry in the table shows the period an agent type would take up treatment given a sequence of IV shocks. Black columns indicate compliers, gray columns indicate non-compliers. The row labeled "Intended" shows the period an agent type would take treatment subject to no IV shocks. The row labeled "Increment" shows the number of periods each IV shock delays treatment by. Types with an increment of 0.5 must experience two shocks before delaying treatment by one period.

When treatment effects are heterogeneous across cohorts, then $\Delta^2(e, s)$ is identified over the longest possible horizon for all cohorts. ■

Example 3.3. Suppose $\bar{t} = 3$. Consider an incremental model where some agents are non-responsive to the shock, some delay treatment by one period after one shock, and some advance treatment by one period after one shock. The IV thus violates monotonicity.

Table 5 shows the agent types under the model described. When treatment effects are heterogeneous across types, then $\Delta^1(g, s)$ is identified for $s \in \{0, \dots, \bar{t} + t^{\text{Extend}} - 3\}$ for $g \in \{2, 3, 5, 6, 8, 9\}$. Treatment effects are thus identified for all complier types. When treatment effects are heterogeneous across cohorts, then $\Delta^2(e, s)$ is identified over the longest possible horizon for all cohorts. ■

Example 3.4. Suppose $\bar{t} = 2$. Table 6 is an example of a selection model where we design each agent type (Section 2.2.3). Unlike the previous examples, agents do not abide to the same threshold or incremental response rule. Type 1 agents respond to the first shock by delaying treatment by one period, but do not respond to the second shock, which can be viewed as both a threshold response and incremental response.[17] Type 3 agents

[17] Agents may be permanently switching to enrolling in period 2 once experiencing a single shock

Table 5: Example of a non-monotonic incremental selection model

	Agent type, G_i									
IV shocks, Z_i	1	2	3	4	5	6	7	8	9	10
$(0,0,0)$	1	1	2	2	2	3	3	3	∞	∞
$(0,0,1)$	1	1	2	2	2	3	3	∞	3	∞
$(0,1,0)$	1	1	2	2	3	2	3	∞	3	∞
$(0,1,1)$	1	1	2	2	∞	2	3	∞	3	∞
$(1,0,0)$	1	2	1	2	3	2	3	∞	3	∞
$(1,0,1)$	1	2	1	2	∞	2	3	∞	3	∞
$(1,1,0)$	1	3	1	2	∞	2	3	∞	2	∞
$(1,1,1)$	1	∞	1	2	∞	2	3	∞	2	∞
Intended	1	1	2	2	2	3	3	3	∞	∞
Increment	0	1	-1	0	1	-1	0	1	-1	0

Notes: Each column corresponds to an agent type. Each row corresponds to a sequence of IV shocks. Each entry in the table shows the period an agent type would take up treatment given a sequence of IV shocks. Black columns indicate compliers, gray columns indicate non-compliers. The row labeled "Intended" shows the period an agent type would take treatment subject to no IV shocks. The row labeled "Increment" shows the number of periods each IV shock delays treatment by. Types with an increment of 0.5 must experience two shocks before delaying treatment by one period.

respond to shocks according to a threshold rule, switching into treatment immediately after experiencing one shock. Type 4 agents respond to the first shock by advancing treatment by one period, but respond to the second shock by delaying treatment by one period, which follows neither the threshold nor incremental rule. Types 2 and 5 are non-compliers. Among these five agent types, multiple response rules are present, monotonicity across agents is violated (types 1 and 3), and monotonicity within types is violated (type 4).

Furthermore, Table 6 does not include all four possible sequences of Z_i. This is to demonstrate that our proposed method is compatible with selection models defined using a subset of the $2^{\bar{t}}$ possible sequences of Z_i. Such selection models maybe by design, such as when the shocks received in one period may depend on the history of shocks.[18] Such selection models may also be required due to insufficient variation of Z_i in the population, such as when $\bar{t}$ is large so not all sequences of Z_i are realized. A natural consequence of restricting the support of Z_i is that fewer agent types can be defined, which limits the extent to which we can model and study treatment effect heterogeneity.[19]

(threshold), or are only willing to delay treatment for one period (incremental).

[18] This is similar to the two-stage designs of Karlan and Zinman (2009), Ashraf et al. (2010), Cohen and Dupas (2010).

[19] Let $\mathcal{Z}^{\text{Full}}$ denote the set of all $2^{\bar{t}}$ possible sequences of Z_i, and $\mathcal{Z}$ denote the set of sequences realized

Table 6: Example of a mixed selection model with restricted support for Z_i

	Agent type, G_i				
IV shocks, Z_i	1	2	3	4	5
$(0,0)$	1	2	∞	∞	∞
$(1,0)$	2	2	1	2	∞
$(1,1)$	2	2	1	∞	∞

Notes: Each column corresponds to an agent type. Each row corresponds to a sequence of IV shocks. Each entry in the table shows the period an agent type would take up treatment given a sequence of IV shocks. Black columns indicate compliers, gray columns indicate non-compliers.

Under the selection model of Table 6, when treatment effects are heterogeneous across types, then $\Delta^1(g,s)$ is identified for $s \in \{0,\dots,\bar{t}+t^{\text{Extend}}-1\}$ for $g \in \{1,3\}$; and $s \in \{0,\dots,\bar{t}+t^{\text{Extend}}-2\}$ for $g=4$. Treatment effects are thus identified for all complier types but for different horizons. When treatment effects are heterogeneous across cohorts, then $\Delta^2(e,s)$ is identified over the longest possible horizon for all cohorts. ■

3.4 Incorporating covariates

Let $W_{i,t} \in \mathcal{W}$ denote a vector of potentially time-varying covariates. We propose two approaches to incorporating covariates in the analysis, although this is ongoing work.

The first approach is to be fully nonparametric and condition the previous results on $W_{i,t}$. This requires the covariates to be time-invariant, i.e., $W_{i,t}=W_i$, but allows G, $Y_{i,t}(\infty)$, and Δ_i to depend on W_i. The identification arguments above can then be applied conditional on W_i. If covariates were instead time-varying, then the identification arguments above cannot be applied since they do not account for the dependence between the treatment effects and covariates.

The second approach to incorporating covariates is to model $Y_{i,t}(\infty)$ and limit the dependence between $W_{i,t}$, Δ_i, and G_i. We impose the following assumptions.

Assumption 5. *$\check{e}(z,g)$ does not depend on $W_{i,t}$.*

in the data. If $\mathcal{Z} \subset \mathcal{Z}^{\text{Full}}$, then certain agent types defined under $\mathcal{Z}$ will be an aggregation of types defined under $\mathcal{Z}^{\text{Full}}$. The type-specific treatment effects recovered under $\mathcal{Z}$ thus become a weighted average of the type-specific treatment effects defined under $\mathcal{Z}^{\text{Full}}$.

Assumption 6.

(i) $Y_{i,t}(\infty) = h(W_{i,t}) + \alpha_{G_i,t} + U_{i,t}$, *where* $h : \mathcal{W} \to \mathbb{R}$, $\{\alpha_{gt}\}$ *are constants, and* $U_{i,t} \in \mathbb{R}$.

(ii) $\mathbb{E}[U_{i,t} \mid Z_i, G_i, W_{i,t}] = 0$.

Assumption 5 states that the timing of treatment E_i is only determined by the agent's type G_i and the instrument Z_i. It follows that $\mathbb{P}[E_i \mid Z_i, G_i, W_{i,t}] = \mathbb{P}[E_i \mid Z_i, G_i]$ and that the first stage identification argument outlined above can be applied without covariates. Assumption 6 places structure on $Y_{i,t}(\infty)$. In particular, the unobserved sources of variation, $\alpha_{g,t}$ and $U_{i,t}$, are additively separable from the covariates. The error term $U_{i,t}$ is also assumed to be mean zero conditional on Z_i, G_i, and $W_{i,t}$.

When treatment effects vary across agent types, we additionally impose the following assumption.

Assumption 7a.

(i) $W_{i,t} \perp\!\!\!\perp G_i$.

(ii) $(W_{i,t}, Z_i) \perp\!\!\!\perp \{\Delta_i(s)\} \mid G_i$.

Assumption 7a(i) states that $W_{i,t}$ is independent of agent types. Assumption 7a(ii) states that treatment effect dynamics are determined by agent type and not covariates.

Proposition 4. *Under Assumptions 1–3, 4a, 5–6, and 7a, there exists a known vector* $\mathbf{b}^{\mathrm{Cov},2}$ *such that*

$$\mathbf{A}^{\mathrm{All},1}\mathbf{x}^{\mathrm{All},1} = \mathbf{b}^{\mathrm{Cov},1}. \tag{14}$$

When treatment effects vary across treatment cohorts, we instead impose the following assumption.

Assumption 7b. $(Z_i, G_i, W_{i,t}) \perp\!\!\!\perp \{\Delta_i(e, s)\}$.

Assumption 7b strengthens Assumption 4b(ii) so that treatment effects are independent of $W_{i,t}$. However, unlike Assumption 7a, covariates are allowed to be correlated across types.

Proposition 5. *Under Assumptions 1–3, 4b, 5–6, and 7b, there exists a known vector*

$\mathbf{b}^{\text{Cov},2}$ *such that*

$$\mathbf{A}^{\text{All},2}\mathbf{x}^{\text{All},2} = \mathbf{b}^{\text{Cov},2}. \tag{15}$$

In the proofs of Propositions 4–5, we show that $\mathbf{b}^{\text{Cov},j}$ for $j = 1, 2$ is equal to the vector of conditional means of $Y_{i,t}$ after projecting off $W_{i,t}$, where the conditioning is on E_i and Z_i.

Remark 3.4. Assumptions 5 and 7a rule out any relationship between the covariates and agent types and do not allow covariates to affect selection into treatment. Relaxing these assumptions is ongoing work. ■

4 Estimation and inference

In this section, we propose estimators for the distribution of agent types and the treatment effects, and derive their asymptotic properties. Throughout, we impose the following assumptions.

Assumption 8. $(\{Y_{i,t}(\infty)\}, \{\Delta_i(e, s)\}, A_i, Z_i, E_i, G_i)$ *are i.i.d. across* i.

Assumption 9. $(\{Y_{i,t}\}, E_i, Z_i)$ *are observed for* $i = 1, \ldots, n$ *and* $t = 1, \ldots, \bar{t}$.

Assumptions 8–9 simply state that we have a panel of n i.i.d. agents over $\bar{t}$ periods and we observe the variables $Y_{i,t}$, E_i, and Z_i.

4.1 First stage

Define

$$\mathbf{b}^{1,e} \equiv \begin{bmatrix} \mathbb{P}[E_i = e \mid Z_i = z_1] \\ \vdots \\ \mathbb{P}[E_i = e \mid Z_i = z_{|\mathcal{Z}|}] \end{bmatrix}$$

to be the probability that agents take up treatment in period e conditional on Z_i, and

$$\mathbf{b}^1 \equiv \begin{bmatrix} \mathbf{b}^{1,1} \\ \vdots \\ \mathbf{b}^{1,\bar{t}} \end{bmatrix}.$$

Let $\mathbf{b}^1_n$ denote the sample analog of $\mathbf{b}^1$.

Proposition 6. *Suppose Assumptions 1 and 8–9 hold, either Assumption 4a or 4b holds, and* $\mathbf{A}^1$ *is full rank. Then the estimator* $\mathbf{x}_n^1 \equiv (\mathbf{A}^{1\top}\mathbf{A}^1)^{-1}\mathbf{A}^{1\top}\mathbf{b}_n^1$ *satisfies*

$$\sqrt{n}\,(\mathbf{x}_n^1 - \mathbf{x}^1) \xrightarrow{\text{d}} N(0, \Psi^1)$$

as $n \to \infty$*, where* Ψ^1 *can be consistently estimated.*

4.2 Second stage

Recall that for $j = 1, 2$, $\mathbf{A}^{\text{All},j}$ is a function of $\mathbf{x}^1$. Define $\mathbf{A}_n^{\text{All},j}$ to be the analog of $\mathbf{A}^{\text{All},j}$ when constructed using $\mathbf{x}_n^1$ in place of $\mathbf{x}^1$, and $\mathbf{b}_n^2$ to be the sample analog of $\mathbf{b}^2$. We cannot immediately use the sample analogs of (14) and (15) for estimation, since $\mathbf{A}_n^{\text{All},j}$ is not full rank. So we propose an estimator that normalizes certain unidentified parameters in $\mathbf{x}^{\text{All},j}$ to 0. Appendix A.3 outlines the normalization. Let $\mathbf{x}^{\text{Sub},j}$ denote the subvector of $\mathbf{x}^{\text{All},j}$ containing all the identified entries in $\mathbf{x}^{\text{All},j}$, as well as the unidentified parameters not normalized to 0. Let $\mathbf{A}_n^{\text{Sub},j}$ denote the columns in $\mathbf{A}_n^{\text{All},j}$ corresponding to $\mathbf{x}^{\text{Sub},j}$.

Proposition 7. *Let* $j = 1, 2$*. Suppose Assumptions 1–3, 8–9 hold. Suppose Assumption 4a holds if* $j = 1$*, and suppose Assumption 4b holds if* $j = 2$*. Suppose* $\mathbf{A}^1$ *is full rank. Then the estimator* $\mathbf{x}_n^{\text{Sub},j} \equiv (\mathbf{A}_n^{\text{Sub},j\top}\mathbf{A}_n^{\text{Sub},j})^{-1}\mathbf{A}_n^{\text{Sub},j\top}\mathbf{b}_n^2$ *satisfies*

$$\sqrt{n}\,(\mathbf{x}_n^{\text{Sub},j} - \mathbf{x}^{\text{Sub},j}) \xrightarrow{\text{d}} N(0, \Psi^{\text{Sub},j})$$

as $n \to \infty$*, where* $\Psi^{\text{Sub},j}$ *can be consistently estimated.*

Remark 4.1. If only bounds on $\mathbf{x}^{\text{All},j}$ for $j = 1, 2$ may be obtained, then the estimation procedure involves solving a linear programming problem, similar to Mogstad et al. (2018). In such cases, Fang and Santos (2019) and Cho and Russell (2024) may be used for statistical inference. ■

4.3 Incorporating covariates

To accommodate covariates, we extend Assumptions 8–9 to include $\{W_{i,t}\}$.

Assumption 10. $(\{Y_{i,t}(\infty)\}, \{\Delta_i(e,s)\}, \{W_{i,t}\}, A_i, Z_i, E_i, G_i)$ *are i.i.d. across* i*.*

Assumption 11. $(\{Y_{i,t}\}, \{W_{i,t}\}, E_i, Z_i)$ *are observed for* $i = 1, \ldots, n$ *and* $t = 1, \ldots, \bar{t}$*.*

The estimators in Proposition 7 may be applied after projecting off $W_{i,t}$ from $Y_{i,t}$. That is, regress $Y_{i,t}$ on E_i, Z_i, and $W_{i,t}$, and subtract from $Y_{i,t}$ the fitted values corresponding to $W_{i,t}$. Then construct $\mathbf{b}_n^2$ from the residualized $Y_{i,t}$. The matrices $\mathbf{A}_n^{\text{Sub},j}$ for $j = 1, 2$

can be constructed as in Proposition 7.

4.4 Testing for instrumental relevance and model misspecification

In Section 3.2.2, we show that the selection model may imply $\mathbb{P}[E_i = e \mid Z_i = z]$ is the same across multiple values of $z \in \mathcal{Z}$ for some $e \in \mathcal{E}$. In Section 3.2.3, we show that instrumental relevance implies $\mathbb{P}[E_i = e \mid Z_i = z]$ differs across values of $z \in \mathcal{Z}$ for some $e \in \mathcal{E}$. Both sets of implications can be characterized by linear restrictions on $\mathbf{b}^1$, which we estimate with $\mathbf{b}^1_n$. Since

$$\sqrt{n}\,(\mathbf{b}^1_n - \mathbf{b}^1) \xrightarrow{\text{d}} N(0, \Psi)$$

as $n \to \infty$, where Ψ is a consistently estimable matrix, both sets of implications can be tested using a Wald test.[20]

5 Simulations

In this section we present simulation results for four data generating processes (DGP) to demonstrate the statistical properties of our estimators. For each DGP, we draw 1000 samples of sizes $n = 1000, 3000, 5000$ and estimate the parameters that are identified.

5.1 Data generating processes

We obtain our four DGPs by varying the level of treatment effect heterogeneity across two selection models. DGPs 1 and 2 impose the threshold selection model from Example 3.1, where agents intend to take up treatment at a given period absent IV shocks, but may immediately take up treatment once experiencing a sufficient number of shocks. Treatment is available in periods 1 and 2 ($\bar{t} = 2$), but we extend the panel to include three additional periods ($t^{\text{Extend}} = 3$) so we can estimate longer treatment effect trajectories. The effects are heterogeneous across agent types in DGP 1 and heterogeneous across cohorts in DGP 2. Table 3 shows the agent types in these simulations, and Example 3.1 discusses the treatment effect parameters identified.

[20] To test for model misspecification, define $\mathbf{Q}$ to be such that each row of $\mathbf{Q}\mathbf{b}^1$ corresponds to a restriction as in (4). To test for IV relevance, instead define $\mathbf{Q}$ to be such that each row of $\mathbf{Q}\mathbf{b}^1$ corresponds to a restriction as in (6). For both cases, test the hypotheses

$$H_0 : \mathbf{Q}\mathbf{b}^1 = 0,$$
$$H_1 : \mathbf{Q}\mathbf{b}^1 \neq 0.$$

The model is misspecified if H_0 is rejected. The instrument is irrelevant if H_0 fails to be rejected.

DGPs 3 and 4 impose the incremental selection model of Example 3.2, where agents intend to take up treatment at a given period absent IV shocks, but may delay treatment with each shock. Treatment is available in periods 1, 2, and 3 ($\bar{t} = 3$), but we again extend the panel to include three additional periods ($t^{\text{Extend}} = 3$). The effects are heterogeneous across agent types in DGP 3 and heterogeneous across cohorts in DGP 4. Table 4 shows the agent types in these simulations, and Example 3.2 discusses the treatment effect parameters identified.

In DGPs 1 and 3, outcomes are generated as

$$Y_{i,t} = \underbrace{\mathbb{E}[Y_{i,t}(\infty) \mid G_i]}_{\text{Untreated trend}} + \underbrace{\mathbb{1}[t \geq \check{e}(Z_i, G_i)]\ \Delta^1(G_i, t - \check{e}(Z_i, G_i))}_{\text{Treatment effect, if treated}} + \underbrace{U_{i,t}}_{\text{Idiosyncratic shock}},$$

where $\mathbb{E}[Y_{i,t}(\infty) \mid G_i = g]$ is randomly drawn from a $\mathcal{N}(1, 1)$ distribution for all $g \in \mathcal{G}$; $\Delta^1(g, s)$ is randomly drawn from a $\mathcal{N}(0, 1)$ distribution for all $g \in \mathcal{G}$ and $s > 0$; and $U_{i,t}$ follows the MA(1) process

$$U_{i,t} = \varepsilon_{i,t} + 0.2\varepsilon_{i,t-1},$$

where $\varepsilon_{i,t} \sim N(0, 0.2)$. The draws of $\{\mathbb{E}[Y_{i,t}(\infty) \mid G_i = g]\}$, $\{\Delta^1(g, s)\}$ are fixed across all 1000 simulations. To verify size control, we fix $\Delta^1(g, 0) = 0$ for all $g \in \mathcal{G}$ so that there is no treatment effect in the period of take-up.

Similarly, in DGPs 2 and 4, outcomes are generated as

$$Y_{i,t} = \underbrace{\mathbb{E}[Y_{i,t}(\infty) \mid G_i]}_{\text{Untreated trend}} + \underbrace{\mathbb{1}[t \geq \check{e}(Z_i, G_i)]\ \Delta^2(\check{e}(Z_i, G_i), t - \check{e}(Z_i, G_i))}_{\text{Treatment effect, if treated}} + \underbrace{U_{i,t}}_{\text{Idiosyncratic shock}},$$

where $\mathbb{E}[Y_{i,t}(\infty) \mid G_i = g]$ and $U_{i,t}$ are generated as above, and $\Delta^2(e, s)$ are drawn from a $\mathcal{N}(0, 1)$ distribution for all $e \in \mathcal{E}$ and $s > 0$. Again, the draws of $\{\mathbb{E}[Y_{i,t}(\infty) \mid G_i = g]\}$, $\{\Delta^2(e, s)\}$ are fixed across all 1000 simulations, and we fix $\Delta^2(e, 0) = 0$ for all $e \in \mathcal{E}$ to verify size control.

For each DGP, we randomly generate the population distribution of agent types subject to Assumption 1, and fix the distribution across all 1000 simulations. Similarly, we randomly generate the distribution of IV shocks such that all possible sequences of shocks over the $\bar{t}$ periods that treatment is available have non-zero probability, and fix the distribution across all 1000 simulations.

5.2 Results

Table 7 shows the performance of the first stage estimator under DGPs 1 and 2 (threshold model), and Table 8 shows the performance under DGPs 3 and 4 (incremental model).[21] Column (1) indexes the agent types; column (2) presents the true distribution of agent types; column (3) presents the relative bias, which is equal to the bias as a fraction of the true parameter value;[22] column (4) presents the average length of the 95% confidence intervals relative to the true parameter value;[23] and column (5) presents the coverage probabilities of the 95% confidence intervals. A good estimator exhibits small relative biases and coverage probabilities of 95%. For all four DGPs and sample sizes, we find the relative bias to be approximately 1% or less, suggesting our first stage estimator is unbiased. We also find the coverage probabilities of the 95% confidence intervals match their nominal size, even when the number of agent types is large.

Tables 9–12 show the performance of the second stage estimator across the four DGPs. In each table, column (1) indexes the treatment effect parameter; column (2) presents the true effect; column (3) presents the relative bias; column (4) presents the relative length of the 95% confidence intervals; column (5) presents the coverage probabilities of the 95% confidence intervals; and column (6) presents the rejection probability when testing the null hypothesis of no effect at the 5% significance level. In addition to small relative biases and correct coverage probabilities, a good estimator exhibits rejection probabilties of 0.05 for $\Delta^1(g, 0)$ and $\Delta^2(e, 0)$, which indicate correct size control, and large rejection probabilities for all other treatment effect parameters, which indicate high power. In Tables 11–12, where there are many treatment effects to estimate, we focus on samples of size $n = 3000, 5000$ for brevity.

In DGPs 1 and 3 (Tables 9 and 11), where effects vary across agent types, the estimator performs well. For many parameters, when $n \geq 3000$, the estimator exhibits relative biases of approximately 10% or less, and the confidence intervals have the correct coverage probability. The rejection rates for $\Delta^1(g, 0)$ and $\Delta^2(e, 0)$ are also consistent with the nominal size of 5%. We find the estimator struggles to accurately estimate effects at the tail of the treatment effect trajectory. For instance, in DGP 1 (Table 9), the estimator performs well for $\Delta^1(4, s)$ for $s = 0, 1, 2$, but exhibits large bias when $s = 3$. Similarly, in DPG 3, the estimator performs well for $\Delta^1(4, s)$ for $s = 0, 1, 2$ but exhibits large bias when $s = 3$.

[21] The distribution of agent types is the same across simulations 1 and 2. Likewise, the distributions are the same across simulations 3 and 4. The first stage estimates do not depend on whether treatment effects are heterogeneous across agent types or cohorts.

[22] For example, a relative bias of -0.01 means the estimator underestimates the parameter by 1% on average.

[23] For example, a relative length of 1 means the average difference between the lower and upper bound of the confidence interval is equal to the true parameter value.

Table 7: Simulation results for DGPs 1 and 2, first stage

			95% confidence interval	
Type (1)	$\mathbb{P}[G_i = g]$ (2)	Relative bias (3)	Relative length (4)	Coverage (5)
$n = 1000$				
1	0.173	−0.001	0.366	0.957
2	0.129	−0.002	1.253	0.956
3	0.17	−0.008	0.572	0.939
4	0.213	0.002	0.69	0.958
5	0.159	0.013	0.778	0.942
6	0.156	−0.005	0.695	0.949
$n = 3000$				
1	0.173	0	0.212	0.951
2	0.129	0.005	0.725	0.966
3	0.17	−0.003	0.332	0.948
4	0.213	−0.001	0.399	0.96
5	0.159	0.002	0.449	0.956
6	0.156	−0.002	0.402	0.938
$n = 5000$				
1	0.173	−0.001	0.164	0.942
2	0.129	0.001	0.561	0.952
3	0.17	0.001	0.257	0.954
4	0.213	−0.001	0.309	0.943
5	0.159	−0.001	0.348	0.957
6	0.156	0.001	0.312	0.955

Notes: Column (1) indexes the agent types. Column (2) presents the true distribution of agent types. Column (3) presents the bias as a proportion of column (2). Column (4) presents the average length of the 95% confidence intervals relative to column (2). Column (5) presents the coverage probabilities of the 95% confidence intervals. All simulations consist of 1000 iterations of the same first stage data generating process.

Table 8: Simulation results for DGPs 3 and 4, first stage

			95% confidence interval	
Type (1)	$\mathbb{P}[G_i = g]$ (2)	Relative bias (3)	Relative length (4)	Coverage (5)
$n = 1000$				
1	0.052	−0.016	1.747	0.947
2	0.078	0.004	0.794	0.952
3	0.123	0.001	1.093	0.949
4	0.077	0.008	1.428	0.939
5	0.059	0	1.309	0.942
6	0.32	−0.003	0.396	0.951
7	0.1	−0.01	1.77	0.948
8	0.125	0.012	0.721	0.956
9	0.066	0.005	1.195	0.939
$n = 3000$				
1	0.052	0.008	1.008	0.943
2	0.078	−0.003	0.458	0.939
3	0.123	0.001	0.63	0.95
4	0.077	−0.001	0.822	0.953
5	0.059	0.001	0.756	0.953
6	0.32	0.003	0.228	0.947
7	0.1	−0.007	1.02	0.948
8	0.125	−0.003	0.414	0.955
9	0.066	−0.005	0.69	0.939
$n = 5000$				
1	0.052	0.003	0.781	0.955
2	0.078	0	0.356	0.953
3	0.123	0.005	0.488	0.948
4	0.077	−0.002	0.637	0.949
5	0.059	−0.001	0.586	0.94
6	0.32	0	0.177	0.948
7	0.1	−0.01	0.79	0.936
8	0.125	0.002	0.322	0.956
9	0.066	0	0.535	0.948

Notes: Column (1) indexes the agent types. Column (2) presents the true distribution of agent types. Column (3) presents the bias as a proportion of column (2). Column (4) presents the average length of the 95% confidence intervals relative to column (2). Column (5) presents the coverage probabilities of the 95% confidence intervals. All simulations consist of 1000 iterations of the same first stage data generating process.

Table 9: Simulation results for DGP 1, second stage

Parameter (1)	True effect (2)	Relative bias (3)	95% confidence interval: Relative length (4)	95% confidence interval: Coverage (5)	Rejection rate (6)
$n = 1000$					
$\Delta^1(2,0)$	0			0.983	0.017
$\Delta^1(2,1)$	0.487	0.009	5.639	0.989	0.103
$\Delta^1(2,2)$	−0.189	−1.563	27.314	0.918	0.111
$\Delta^1(4,0)$	0			0.969	0.031
$\Delta^1(4,1)$	−0.098	0.498	13.301	0.948	0.02
$\Delta^1(4,2)$	−0.543	−0.106	2.769	0.954	0.38
$\Delta^1(4,3)$	0.635	−0.68	1.241	0.396	0.182
$\Delta^1(5,0)$	0			0.936	0.064
$\Delta^1(5,1)$	0.37	−0.088	3.479	0.938	0.337
$\Delta^1(5,2)$	−0.19	−0.333	13.08	0.945	0.097
$n = 3000$					
$\Delta^1(2,0)$	0			0.962	0.038
$\Delta^1(2,1)$	0.487	−0.012	2.676	0.967	0.34
$\Delta^1(2,2)$	−0.189	−0.324	11.135	0.944	0.119
$\Delta^1(4,0)$	0			0.958	0.042
$\Delta^1(4,1)$	−0.098	0.184	7.234	0.948	0.042
$\Delta^1(4,2)$	−0.543	−0.041	1.509	0.946	0.684
$\Delta^1(4,3)$	0.635	−0.681	0.698	0.033	0.459
$\Delta^1(5,0)$	0			0.953	0.047
$\Delta^1(5,1)$	0.37	−0.032	1.891	0.95	0.562
$\Delta^1(5,2)$	−0.19	−0.117	7.169	0.951	0.116
$n = 5000$					
$\Delta^1(2,0)$	0			0.965	0.035
$\Delta^1(2,1)$	0.487	0.009	2.04	0.953	0.513
$\Delta^1(2,2)$	−0.189	−0.211	8.365	0.94	0.129
$\Delta^1(4,0)$	0			0.956	0.044
$\Delta^1(4,1)$	−0.098	0.112	5.542	0.957	0.064
$\Delta^1(4,2)$	−0.543	−0.03	1.158	0.954	0.851
$\Delta^1(4,3)$	0.635	−0.677	0.538	0.002	0.665
$\Delta^1(5,0)$	0			0.945	0.055
$\Delta^1(5,1)$	0.37	−0.025	1.453	0.953	0.718
$\Delta^1(5,2)$	−0.19	−0.101	5.518	0.962	0.134

Notes: Column (1) presents the treatment effect parameter. Column (2) presents the true effect. Column (3) presents the bias as a proportion of column (2). Column (4) presents the average length of the 95% confidence intervals relative to column (2). Column (5) presents the coverage probabilities of the 95% confidence intervals. Column (6) presents the rejection rate when testing the null hypothesis of no effect at the 5% significance level. All simulations consist of 1000 iterations of the same data generating process.

Table 10: Simulation results for DGP 2, second stage

Parameter (1)	True effect (2)	Relative bias (3)	95% confidence interval: Relative length (4)	95% confidence interval: Coverage (5)	Rejection rate (6)
$n = 1000$					
$\Delta^2(1,0)$	0			0.96	0.04
$\Delta^2(1,1)$	0.64	−0.007	1.098	0.961	0.918
$\Delta^2(1,2)$	1.327	−0.001	0.477	0.969	0.999
$\Delta^2(1,3)$	0.152	0.055	7.776	0.958	0.042
$\Delta^2(2,0)$	0			0.956	0.044
$\Delta^2(2,1)$	0.487	−0.016	1.382	0.947	0.778
$\Delta^2(2,2)$	−0.189	−0.089	7.546	0.96	0.108
$n = 3000$					
$\Delta^2(1,0)$	0			0.945	0.055
$\Delta^2(1,1)$	0.64	0.01	0.622	0.945	0.999
$\Delta^2(1,2)$	1.327	0	0.271	0.938	1
$\Delta^2(1,3)$	0.152	−0.039	4.41	0.956	0.113
$\Delta^2(2,0)$	0			0.946	0.054
$\Delta^2(2,1)$	0.487	−0.004	0.795	0.934	0.99
$\Delta^2(2,2)$	−0.189	0.013	4.335	0.94	0.177
$n = 5000$					
$\Delta^2(1,0)$	0			0.953	0.047
$\Delta^2(1,1)$	0.64	0.001	0.479	0.95	1
$\Delta^2(1,2)$	1.327	0.001	0.209	0.935	1
$\Delta^2(1,3)$	0.152	−0.008	3.409	0.954	0.185
$\Delta^2(2,0)$	0			0.949	0.051
$\Delta^2(2,1)$	0.487	−0.003	0.615	0.933	0.999
$\Delta^2(2,2)$	−0.189	−0.01	3.359	0.936	0.243

Notes: Column (1) presents the treatment effect parameter. Column (2) presents the true effect. Column (3) presents the bias as a proportion of column (2). Column (4) presents the average length of the 95% confidence intervals relative to column (2). Column (5) presents the coverage probabilities of the 95% confidence intervals. Column (6) presents the rejection rate when testing the null hypothesis of no effect at the 5% significance level. All simulations consist of 1000 iterations of the same data generating process.

Table 11: Simulation results for DGP 3, second stage

			95% confidence interval		
Parameter (1)	True effect (2)	Relative bias (3)	Relative length (4)	Coverage (5)	Rejection rate (6)
$n = 3000$					
$\Delta^1(1,0)$	0			0.966	0.034
$\Delta^1(1,1)$	0.546	−0.005	2.681	0.975	0.368
$\Delta^1(1,2)$	1.178	−0.073	3.044	0.957	0.383
$\Delta^1(1,3)$	−0.743	0.138	5.912	0.951	0.03
$\Delta^1(1,4)$	−1.76	0.19	4.102	0.941	0.014
$\Delta^1(3,0)$	0			0.961	0.039
$\Delta^1(3,1)$	−1.446	−0.048	1.668	0.951	0.636
$\Delta^1(3,2)$	−1.381	−0.062	2.446	0.946	0.434
$\Delta^1(3,3)$	2.002	0.107	2.843	0.947	0.171
$\Delta^1(4,0)$	0			0.972	0.028
$\Delta^1(4,1)$	−2.485	0.021	1.309	0.949	0.839
$\Delta^1(4,2)$	−1.512	0.034	3.484	0.949	0.277
$\Delta^1(4,3)$	−0.202	0.043	30.278	0.946	0.066
$\Delta^1(6,0)$	0			0.947	0.053
$\Delta^1(6,1)$	0.253	0.011	1.066	0.934	0.974
$\Delta^1(6,2)$	0.443	−0.03	1.993	0.95	0.491
$\Delta^1(7,0)$	0			0.955	0.045
$\Delta^1(7,1)$	−0.881	−0.119	4.557	0.974	0.215
$\Delta^1(7,2)$	0.393	0.746	17.019	0.957	0.02
$n = 5000$					
$\Delta^1(1,0)$	0			0.966	0.034
$\Delta^1(1,1)$	0.546	0.001	1.968	0.969	0.551
$\Delta^1(1,2)$	1.178	−0.007	2.184	0.973	0.515
$\Delta^1(1,3)$	−0.743	0.016	4.242	0.973	0.065
$\Delta^1(1,4)$	−1.76	0.083	2.862	0.948	0.121
$\Delta^1(3,0)$	0			0.946	0.054
$\Delta^1(3,1)$	−1.446	−0.032	1.258	0.956	0.792
$\Delta^1(3,2)$	−1.381	−0.042	1.845	0.952	0.572
$\Delta^1(3,3)$	2.002	0.048	2.118	0.94	0.471
$\Delta^1(4,0)$	0			0.949	0.051
$\Delta^1(4,1)$	−2.485	−0.004	0.993	0.959	0.952
$\Delta^1(4,2)$	−1.512	−0.011	2.646	0.951	0.358
$\Delta^1(4,3)$	−0.202	−0.275	23.005	0.947	0.061
$\Delta^1(6,0)$	0			0.951	0.049
$\Delta^1(6,1)$	0.253	−0.008	0.823	0.953	0.998
$\Delta^1(6,2)$	0.443	0.017	1.539	0.951	0.728
$\Delta^1(7,0)$	0			0.958	0.042
$\Delta^1(7,1)$	−0.881	−0.064	3.295	0.959	0.326
$\Delta^1(7,2)$	0.393	0.508	11.76	0.945	0.025

Notes: Column (1) presents the treatment effect parameter. Column (2) presents the true effect. Column (3) presents the bias as a proportion of column (2). Column (4) presents the average length of the 95% confidence intervals relative to column (2). Column (5) presents the coverage probabilities of the 95% confidence intervals. Column (6) presents the rejection rate when testing the null hypothesis of no effect at the 5% significance level. All simulations consist of 1000 iterations of the same data generating process.

Table 12: Simulation results for DGP 4, second stage

Parameter (1)	True effect (2)	Relative bias (3)	95% confidence interval: Relative length (4)	95% confidence interval: Coverage (5)	Rejection rate (6)
$n = 3000$					
$\Delta^2(1,0)$	0			0.973	0.027
$\Delta^2(1,1)$	0.546	0.072	1.517	0.973	0.893
$\Delta^2(1,2)$	1.178	−0.116	2.372	0.952	0.52
$\Delta^2(1,3)$	−0.743	−0.159	3.127	0.933	0.429
$\Delta^2(1,4)$	−1.76	0.082	1.434	0.96	0.95
$\Delta^2(2,0)$	0			0.954	0.046
$\Delta^2(2,1)$	1.846	−0.007	0.729	0.948	1
$\Delta^2(2,2)$	0.456	−0.022	1.152	0.938	0.96
$\Delta^2(2,3)$	−0.144	0.036	6.87	0.938	0.07
$\Delta^2(3,0)$	0			0.946	0.054
$\Delta^2(3,1)$	−1.446	−0.003	0.128	0.95	1
$\Delta^2(3,2)$	−1.381	−0.005	0.419	0.945	1
$n = 3000$					
$\Delta^2(1,0)$	0			0.965	0.035
$\Delta^2(1,1)$	0.546	0.018	1.043	0.957	0.993
$\Delta^2(1,2)$	1.178	−0.051	1.585	0.96	0.687
$\Delta^2(1,3)$	−0.743	−0.076	2.071	0.928	0.544
$\Delta^2(1,4)$	−1.76	0.028	0.969	0.946	1
$\Delta^2(2,0)$	0			0.948	0.052
$\Delta^2(2,1)$	1.846	−0.009	0.566	0.937	1
$\Delta^2(2,2)$	0.456	−0.005	0.899	0.945	1
$\Delta^2(2,3)$	−0.144	−0.022	5.329	0.952	0.08
$\Delta^2(3,0)$	0			0.953	0.047
$\Delta^2(3,1)$	−1.446	−0.002	0.099	0.959	1
$\Delta^2(3,2)$	−1.381	−0.001	0.324	0.955	1

Notes: Column (1) presents the treatment effect parameter. Column (2) presents the true effect. Column (3) presents the bias as a proportion of column (2). Column (4) presents the average length of the 95% confidence intervals relative to column (2). Column (5) presents the coverage probabilities of the 95% confidence intervals. Column (6) presents the rejection rate when testing the null hypothesis of no effect at the 5% significance level. All simulations consist of 1000 iterations of the same data generating process.

In the case of DGPs 2 and 4 (Tables 10 and 12), where effects vary across treatment cohorts, the estimator also performs well. For almost all parameters, regardless of sample size, the relative bias is approximately 10% or less, the confidence intervals have the correct coverage probability, and the test exhibits either high power or correct size. Similar to DGPs 1 and 3, the estimator struggles to precisely estimate some effects at the tail of the treatment effect trajectory. For instance, in DGP 2 when $n \geq 3000$, $\Delta^2(2,2)$ is estimated with approximately 1% bias, but the power is less than 25%. Similarly, in DGP 4 when $n \geq 3000$, $\Delta^2(2,3)$ is estimated with less than 4% bias but the power is less than 10%.

Our simulations suggest the estimator performs better when effects are heterogeneous across cohorts rather than agent types. In the former case, we find smaller relative biases, shorter relative lengths of the confidence intervals, and higher power compared to the latter case. The coverage probabilities and size of the tests are comparable and equal to their nominal levels. This result is not surprising as each treatment effect parameter is effectively estimated using a larger subset of the data when effects are heterogeneous across cohorts rather than types. This suggests that the performance of the estimator depends less on the selection model and more on the level of treatment effect heterogeneity.

Overall, the estimator performs well across all four DGPs. For many treatment effect parameters, the estimator exhibits small bias, the confidence intervals have the correct coverage probability, and hypothesis tests show high power and correct size.

6 Conclusion

We consider a setting where treatment is available in multiple periods and agents select when to enroll in treatment, if at all. Each period, agents experience an IV shock affecting their enrollment decision and respond dynamically to these shocks. Heterogeneity in their responses to the shocks induces multiple latent agent types. We propose a general framework to identify and estimate dynamic treatment effects, and our framework is compatible with a broad range of selection models that researchers can easily specify.

Our method is compatible with two frameworks for treatment effect heterogeneity. In the first framework, treatment effects are heterogeneous across agent types. In the second framework, treatment effects are heterogeneous across treatment cohorts. In both cases, we show that the distribution of agent types can be identified, and the dynamic treatment effects can be separately identified across either agent types or cohorts.

We propose a misspecification test to aid researchers in choosing a selection model. We also propose nonparametric estimators for all identified parameters of the model and derive their asymptotic properties. Simulation results suggest the estimators perform well under both frameworks for treatment effect heterogeneity, even when there are many latent types and many distinct treatment effect trajectories to estimate.

We hope our methodology will prove helpful in panel settings where timing of treatment is endogenous—thereby ruling out the common trend assumption standard in the difference-in-differences literature—and agents respond dynamically to repeated IV shocks affecting their treatment decisions.

References

Abadie, A. (2005): “Semiparametric difference-in-differences estimators,” *Review of Economic Studies*, 72, 1–19.

Acemoglu, D., D. Autor, D. Dorn, G. H. Hanson, and B. Price (2016): “Import competition and the great US employment sag of the 2000s,” *Journal of Labor Economics*, 34, S141–S198.

Acemoglu, D., A. Finkelstein, and M. J. Notowidigdo (2013): “Income and health spending: Evidence from oil price shocks,” *Review of Economics and Statistics*, 95, 1079–1095.

Alzúa, M. L., G. Cruces, and C. Lopez (2016): “Long-run effects of youth training programs: Experimental evidence from Argentina,” *Economic inquiry*, 54, 1839–1859.

Angrist, J., B. Ferman, C. Gao, P. Hull, O. L. Tecchio, and R. W. Yeh (2024): “Instrumental variables with time-varying exposure: New estimates of revascularization effects on quality of life,” Tech. rep., National Bureau of Economic Research.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996): “Identification of causal effects using instrumental variables,” *Journal of the American statistical Association*, 91, 444–455.

Antinyan, A., Z. Asatryan, Z. Dai, and K. Wang (2021): “Does the frequency of reminders matter for their effectiveness? A randomized controlled trial,” *Journal of Economic Behavior & Organization*, 191, 752–764.

Ashraf, N., J. Berry, and J. M. Shapiro (2010): “Can higher prices stimulate product use? Evidence from a field experiment in Zambia,” *American Economic Review*, 100, 2383–2413.

Assunção, J., C. Gandour, and R. Rocha (2023): “DETER-ing Deforestation in the Amazon: Environmental Monitoring and Law Enforcement,” *American Economic Journal: Applied Economics*, 15, 125–156.

Autor, D. H., D. Dorn, and G. H. Hanson (2013): “The China syndrome: Local labor market effects of import competition in the United States,” *American economic review*, 103, 2121–2168.

Bensnes, S., I. Huitfeldt, and E. Leuven (2023): “Reconciling Estimates of the Long-Term Earnings Effect of Fertility,” Tech. Rep. 16174, Institute of Labor Economics (IZA).

Borusyak, K., X. Jaravel, and J. Spiess (2024): "Revisiting event-study designs: Robust and efficient estimation," *Review of Economic Studies*, 91, 3253–3285.

Bronars, S. G. and J. Grogger (1994): "The economic consequences of unwed motherhood: Using twin births as a natural experiment," *The American Economic Review*, 1141–1156.

Callaway, B. and P. H. Sant'Anna (2021): "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 225, 200–230.

Callaway, B. and E. S. Tsyawo (2023): "Treatment effects in staggered adoption designs with non-parallel trends," *arXiv preprint arXiv:2308.02899.*

Cho, J. and T. M. Russell (2024): "Simple inference on functionals of set-identified parameters defined by linear moments," *Journal of Business & Economic Statistics*, 42, 563–578.

Cohen, J. and P. Dupas (2010): "Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment," *The Quarterly Journal of Economics*, 125, 1–45.

Das, N. (2021): "Training the disadvantaged youth and labor market outcomes: Evidence from Bangladesh," *Journal of Development Economics*, 149, 102585.

De Chaisemartin, C. and X. d'Haultfoeuille (2018): "Fuzzy differences-in-differences," *The Review of Economic Studies*, 85, 999–1028.

Fang, Z. and A. Santos (2019): "Inference on directionally differentiable functions," *The Review of Economic Studies*, 86, 377–412.

Fehr, E. and L. Goette (2007): "Do workers work more if wages are high? Evidence from a randomized field experiment," *American Economic Review*, 97, 298–317.

Ferman, B. and O. Tecchio (2025): "Dynamic LATEs with a static instrument," *arXiv preprint arXiv:2305.18114.*

Gallen, Y., J. S. Joensen, E. R. Johansen, and G. F. Veramendi (2025): "The Labor Market Returns to Delaying Pregnancy," *SSRN Electronic Journal*, posted August 26, 2023; Last revised September 10, 2025.

Ghanem, D., P. H. Sant'Anna, and K. Wüthrich (2025): "Selection and parallel trends," *arXiv preprint arXiv:2203.09001.*

Heckman, J. J. (1979): "Sample selection bias as a specification error," *Econometrica: Journal of the econometric society*, 153–161.

Heckman, J. J. and E. Vytlacil (2005): "Structural equations, treatment effects, and econometric policy evaluation 1," *Econometrica*, 73, 669–738.

Hirshleifer, S., D. McKenzie, R. Almeida, and C. Ridao-Cano (2016): "The impact of vocational training for the unemployed: Experimental evidence from Turkey," *The Economic Journal*, 126, 2115–2146.

Ilciukas, J. (2025): "Parenthood Timing and Gender Inequality," Working paper, first version January 27, 2024; accessed December 3, 2025.

Imbens, G. W. and J. D. Angrist (1994): "Identification and estimation of local average treatment effects," *Econometrica*, 62, 467–475.

Karlan, D., M. McConnell, S. Mullainathan, and J. Zinman (2016): "Getting to the top of mind: How reminders increase saving," *Management science*, 62, 3393–3411.

Karlan, D. and J. Zinman (2009): "Observing unobservables: Identifying information asymmetries with a consumer credit field experiment," *Econometrica*, 77, 1993–2008.

Klasnja, P., E. B. Hekler, S. Shiffman, A. Boruvka, D. Almirall, A. Tewari, and S. A. Murphy (2015): "Microrandomized trials: An experimental design for developing just-in-time adaptive interventions." *Health Psychology*, 34, 1220.

Lundborg, P., E. Plug, and A. W. Rasmussen (2017): "Can women have children and a career? IV evidence from IVF treatments," *American Economic Review*, 107, 1611–1637.

Miguel, E., S. Satyanath, and E. Sergenti (2004): "Economic shocks and civil conflict: An instrumental variables approach," *Journal of political Economy*, 112, 725–753.

Milkman, K. L., S. F. Ellis, D. M. Gromet, I. M. DeMay, H. N. Graci, Y. Jung, R. S. Mobarak, R. A. Silvera Zumaran, M. N. Simmons, C. Van den Bulte, et al. (2025): "Can reminder emails compel Americans to save? A two-million-person megastudy," *PNAS nexus*, 4, pgaf280.

Miyaji, S. (2024): "Instrumented Difference-in-Differences with Heterogeneous Treatment Effects," *arXiv preprint*, arXiv:2405.12083, last revised January 7, 2025 (v5).

Mogstad, M., A. Santos, and A. Torgovitsky (2018): "Using instrumental variables for inference about policy relevant treatment parameters," *Econometrica*, 86, 1589–1619.

Picchetti, P. (2024): "Instrumental Variables with Multiple Time Periods," SSRN working paper, last revised October 10, 2024. Available at SSRN: https://ssrn.com/abstract=4832234.

Rambachan, A. and J. Roth (2023): "A more credible approach to parallel trends," *Review of Economic Studies*, 90, 2555–2591.

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

Schochet, P. Z., J. Burghardt, and S. McConnell (2008): "Does job corps work? Impact findings from the national job corps study," *American economic review*, 98, 1864–1886.

Silles, M. A. (2016): "The impact of children on women's labour supply and earnings in the UK: Evidence using twin births," *Oxford economic papers*, 68, 197–216.

Sun, L. and S. Abraham (2021): "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 225, 175–199.

Thaler, R. H. and C. R. Sunstein (2009): *Nudge: Improving decisions about health, wealth, and happiness*, Penguin.

Ye, T., A. Ertefaie, J. Flory, S. Hennessy, and D. S. Small (2023): "Instrumented difference-in-differences," *Biometrics*, 79, 569–581.

Ye, T., L. Keele, R. Hasegawa, and D. S. Small (2024): "A negative correlation strategy for bracketing in difference-in-differences," *Journal of the American Statistical Association*, 119, 2256–2268.

A Proofs

A.1 Proofs of identification

Proof of Proposition 1. Define

$$\mathbf{b}^{1,e} \equiv \begin{bmatrix} \mathbb{P}[E_i = e \mid Z_i = z_1] \\ \vdots \\ \mathbb{P}[E_i = e \mid Z_i = z_{|\mathcal{Z}|}] \end{bmatrix}$$

to be the observed vector of probabilities that agents take up treatment in period e conditional on Z_i, and

$$\mathbf{b}^1 \equiv \begin{bmatrix} \mathbf{b}^{1,1} \\ \vdots \\ \mathbf{b}^{1,\infty} \end{bmatrix}.$$

Define

$$\mathbf{x}^1 \equiv \begin{bmatrix} \mathbb{P}[G_i = 1] \\ \vdots \\ \mathbb{P}[G_i = \overline{g}] \end{bmatrix} \tag{A.1}$$

to be the distribution of G_i. Under either Assumptions 4a or 4b, $Z_i \perp\!\!\!\perp G_i$ and (3) holds for all $e \in \mathcal{E}$ and $z \in \mathcal{Z}$. We can jointly express (3) for all $e \in \mathcal{E}$ and $z \in \mathcal{Z}$ as

$$\mathbf{A}^1\, \mathbf{x}^1 = \mathbf{b}^1.$$

Since $\mathbf{A}^1$ is known and $\mathbf{b}^1$ is observed from the data, we can can uniquely solve for $\mathbf{x}^1$ if $\mathbf{A}^1$ is full rank. ■

We provide a more general proofs of Propositions 2 and 3 for the setting where treatment is only available during a window of time. Agents are observed during the periods when treatment is available, and may also be observed before and/or after treatment is available.[24]

Let $\mathcal{T} \equiv \{\underline{t}, \underline{t}+1, \dots, 0, \dots, \overline{t}-1, \overline{t}\}$ be the set of all time indices, with $\underline{t} \leq 1 < \overline{t}$. Let $\mathcal{E} \equiv \{1, \dots, \overline{e}-1, \overline{e}, \infty\}$ be the set of all time periods agents can enroll into an absorbing treatment, where $\overline{e} \leq \overline{t}$. We refer to periods $t \leq 0$ as pre-treatment periods, periods $t \in \mathcal{E}$ as post-treatment periods, and periods $t > \overline{e}$ as post-eligibility periods.

[24]For example, consider a panel following agents from 2021–2026 where agents can enroll in an absorbing treatment between 2023–2024.

Proof of Proposition 2. From the data, we observe $\mathbb{E}[Y_{i,t} \mid E_i = e, Z_i = z]$ for $t \in \mathcal{T}$, $e \in \mathcal{E}$, and $z \in \mathcal{Z}$. We can write $\mathbb{E}[Y_{i,t} \mid E_i = e, Z_i = z]$ as

$$\mathbb{E}[Y_{i,t} \mid E_i, Z_i] = \mathbb{E}[Y_{i,t}(\infty) \mid E_i, Z_i] + \mathbb{E}[\Delta_i(t - E_i) \mid E_i, Z_i]. \tag{A.2}$$

In equations (7)–(8) of the main paper, we show that

$$\begin{aligned} &\mathbb{E}[Y_{i,t}(\infty) \mid E_i = e, Z_i = z] \\ &= \sum_{g \in \mathcal{G}} \mathbb{E}[Y_{i,t}(\infty) \mid G_i = g] \, \frac{\mathbb{1}[\check{e}(z,g) = e] \; \mathbb{P}[G_i = g]}{\sum_{g' \in \mathcal{G}} \mathbb{1}[\check{e}(z,g') = e] \; \mathbb{P}[G_i = g']} \end{aligned} \tag{A.3}$$

and

$$\begin{aligned} &\mathbb{E}[\Delta_i(t - E_i) \mid E_i = e, Z_i = z] \\ &= \sum_{g \in \mathcal{G}} \Delta^1(g, t - e) \, \frac{\mathbb{1}[\check{e}(z,g) = e] \; \mathbb{P}[G_i = g]}{\sum_{g' \in \mathcal{G}} \mathbb{1}[\check{e}(z,g') = e] \; \mathbb{P}[G_i = g']}. \end{aligned} \tag{A.4}$$

It follows that we can express the observed moments $\mathbb{E}[Y_{i,t} \mid E_i = e, Z_i = z]$ as weighted averages of $\{\mathbb{E}[Y_{i,t}(\infty) \mid G_i = g]\}$ and $\{\Delta^1(g, t-e)\}$. We show how this expression can be jointly written for all t, e, and z as $\mathbf{A}^{\text{All},1}\mathbf{x}^{\text{All},1} = \mathbf{b}^2$.

In (9) of the main paper, we define $\mathbf{b}^2$ to be the vector stacking all observed moments $\{\mathbb{E}[Y_{i,t} \mid E_i = e, Z_i = z]\}$.[25] In the remainder of this proof, we construct $\mathbf{A}^{\text{All},1}$ and $\mathbf{x}^{\text{All},1}$.

Notice that the weights for $\{\mathbb{E}[Y_{i,t}(\infty) \mid G_i = g]\}$ and $\Delta^1(g, t-e)$ in (A.3)–(A.4) are the same and equal to

$$\left[\sum_{g \in \mathcal{G}} \mathbb{1}[\check{e}(z,g) = e]\right]^{-1} \mathbb{1}[\check{e}(z,g) = e] \; \mathbb{P}[G_i = g]. \tag{A.5}$$

The term $\mathbb{1}[\check{e}(z,g) = e]$ in the weights ensures that non-zero weights are only assigned to types with $\check{e}(z,g) = e$ for some z and e. This information is contained in $\mathbf{M}^e$, which indicates which agent types take up treatment in e for each z. The term $\mathbb{P}[G_i = g]$ in the weights is the distribution of types and is identified in Proposition 1. The vector $\mathbf{x}^1$, defined in (A.1), denotes the distribution of types. The weights (A.5) for a given $e \in \mathcal{E}$ and all $z \in \mathcal{Z}$ can then be jointly expressed as

$$\mathbf{W}^e = \underbrace{\operatorname{diag}\left(\mathbf{M}^e \, \mathbf{x}^1\right)^{-1}}_{\left[\sum_{g \in \mathcal{G}} \mathbb{1}[\check{e}(z,g') = e] \; \mathbb{P}[G_i = g]\right]^{-1} \text{ for all } z \in \mathcal{Z}} \overbrace{\mathbf{M}^e \operatorname{diag}\left(\mathbf{x}^1\right)}^{\mathbb{1}[\check{e}(z,g) = e] \; \mathbb{P}[G_i = g] \text{ for all } z \in \mathcal{Z} \text{ for all } g \in \mathcal{G},\, z \in \mathcal{G}},$$

[25]The vector $\mathbf{b}^2$ is ordered by (e, t, z).

where each column of $\mathbf{W}^e$ corresponds to an agent type, each row corresponds to a particular value of Z_i, and each entry is the weight assigned to an agent type.

Let $\mathcal{S} \equiv \{0, \ldots, \bar{t}-1\}$ contain all indexes for time relative to treatment take up.[26] Define

$$\mathbf{x}_t^{Y(\infty)} \equiv \begin{bmatrix} \mathbb{E}[Y_{i,t}(\infty) \mid G = 1] \\ \vdots \\ \mathbb{E}[Y_{i,t}(\infty) \mid G = \bar{g}] \end{bmatrix},$$

$$\mathbf{x}_s^{\Delta^1} \equiv \begin{bmatrix} \Delta^1(1, s) \\ \vdots \\ \Delta^1(\bar{g}, s) \end{bmatrix}$$

to be the parameters that (7)–(8) are averaging over, where $s \in \mathcal{S}$. Then for a given $t \in \mathcal{T}$ and $e \in \mathcal{E}$, we can jointly express (7) across all z as

$$\begin{bmatrix} \mathbb{E}[Y_{i,t}(\infty) \mid E_i = e, Z_i = z_1] \\ \vdots \\ \mathbb{E}[Y_{i,t}(\infty) \mid E_i = e, Z_i = z_{|\mathcal{Z}|}] \end{bmatrix} = \mathbf{W}^e \mathbf{x}_t^{Y(\infty)}.$$

Likewise, we can jointly express (8) across all z as

$$\begin{bmatrix} \mathbb{E}[\Delta_i(t - E_i) \mid E_i = e, Z_i = z_1] \\ \vdots \\ \mathbb{E}[\Delta_i(t - E_i) \mid E_i = e, Z_i = z_{|\mathcal{Z}|}] \end{bmatrix} = \mathbf{W}^e \mathbf{x}_{t-e}^{\Delta^1}.$$

Combining the previous two expressions, we can write (A.2) as

$$\begin{aligned} &\mathbb{E}[Y_{i,t} \mid E_i = e, Z_i = z] \\ &= \mathbf{W}^e \mathbf{x}_t^{Y(\infty)} + \mathbf{W}^e \mathbf{x}_{t-e}^{\Delta^1} \\ &= \mathbf{W}^e (\mathbf{x}_t^{Y(\infty)} + \underbrace{\mathbf{x}_{t-e}^{\Delta^1}}_{\text{As. 2} \Rightarrow \mathbf{x}_{t-e}^{\Delta^1} \neq 0 \text{ only if } t-e \geq 0}) \\ &= \mathbf{W}^e \left(\mathbf{x}_t^{Y(\infty)} + \mathbb{1}[t - e \geq 0]\ \mathbf{x}_{t-e}^{\Delta^1}\right), \end{aligned}$$

where the final equality follows from Assumption 2, which assumes $\Delta^1(g, s) = 0$ for $s < 0$ and implies $\mathbf{x}_{t-e}^{\Delta^1} = 0$ for $t - e < 0$.

[26] For example, $s = 0$ indexes the period where agent takes up treatment, $s = 1$ indexes one period after the agent takes up treatment.

We now stack the previous system of equations to get a single system. Define

$$\mathbf{x}^{Y(\infty)} \equiv \begin{bmatrix} \mathbf{x}_{\underline{t}}^{Y(\infty)} \\ \vdots \\ \mathbf{x}_{\bar{t}}^{Y(\infty)} \end{bmatrix}, \tag{A.6}$$

$$\mathbf{x}^{\Delta^1} \equiv \begin{bmatrix} \mathbf{x}_{0}^{\Delta^1} \\ \vdots \\ \mathbf{x}_{\bar{t}-1}^{\Delta^1} \end{bmatrix},$$

$$\mathbf{x}^{\text{All},1} \equiv \begin{bmatrix} \mathbf{x}^{Y(\infty)} \\ \mathbf{x}^{\Delta^1} \end{bmatrix}, \tag{A.7}$$

where $\mathbf{x}^{Y(\infty)}$ contains the average untreated outcomes, $\mathbf{x}^{\Delta^1}$ contains the treatment effect dynamics, and $\mathbf{x}^{\text{All},1}$ is the full vector of unknown parameters.[27]

We stack $\mathbf{M}^e$ to select the parameters in $\mathbf{x}^{Y(\infty)}$. Define

$$\mathbf{R}_e^{Y(\infty)} \equiv \mathbf{I}_{|\mathcal{T}|} \otimes \mathbf{M}^e,$$

$$\mathbf{R}^{Y(\infty)} \equiv \begin{bmatrix} \mathbf{R}_1^{Y(\infty)} \\ \vdots \\ \mathbf{R}_{\bar{e}}^{Y(\infty)} \end{bmatrix}, \tag{A.8}$$

where $\mathbf{I}_m$ is an $m \times m$ identity matrix and $\otimes$ is the Kronecker product. Each diagonal entry in $\mathbf{I}_{|\mathcal{T}|}$ corresponds to a single time period. It follows that

$$\mathbf{R}_e^{Y(\infty)}\mathbf{x}^{Y(\infty)} = \begin{bmatrix} \sum_{g\in\mathcal{G}} \mathbb{1}[\check{e}(z_1, g) = e]\ \mathbb{E}[Y_{i,\underline{t}}(\infty) \mid G_i = g] \\ \vdots \\ \sum_{g\in\mathcal{G}} \mathbb{1}[\check{e}(z_{|\mathcal{G}|}, g) = e]\ \mathbb{E}[Y_{i,\bar{t}}(\infty) \mid G_i = g] \end{bmatrix}.$$

$\mathbf{R}^{Y(\infty)}\mathbf{x}^{Y(\infty)}$ simply stacks the expression above across all $e \in \mathcal{E}$.

Similarly, we stack $\mathbf{M}^e$ to select the parameters in $\mathbf{x}^{\Delta^1}$. Define

$$\mathbf{R}_e^{\Delta^1} \equiv \left[\begin{array}{cc} \multicolumn{2}{c}{\mathbf{0}_{(-\underline{t}+1)\times|\mathcal{S}|}} \\ \multicolumn{2}{c}{\mathbf{0}_{(e-1)\times|\mathcal{S}|}} \\ \hdashline \mathbf{I}_{(|\mathcal{S}|-e+1)} & \mathbf{0}_{(s-e+1)\times(e-1)} \end{array}\right] \otimes \mathbf{M}^e,$$

$$\mathbf{R}^{\Delta^1} \equiv \begin{bmatrix} \mathbf{R}_1^{\Delta_1} \\ \vdots \\ \mathbf{R}_{\bar{e}}^{\Delta_1} \end{bmatrix},$$

[27] $\mathbf{x}_s^{\Delta^1}$ for $s < 0$ are omitted since they are assumed to be 0 under Assumption 2.

where $\mathbf{0}_{(-\underline{t}+1)\times|\mathcal{S}|}$ is a $(-\underline{t}+1)\times|\mathcal{S}|$ matrix of 0s that accounts for how there are no treatment effects in the pre-treatment periods; $\mathbf{0}_{(e-1)\times|\mathcal{S}|}$ accounts for the periods when treatment was available but the types had yet to take up treatment; $\mathbf{I}_{|\mathcal{S}|-e+1}$ selects the treatment effects $\Delta^1(g,s)$ for $s=0,\ldots,\bar{t}-e$; and $\mathbf{0}_{(s-e+1)\times(e-1)}$ accounts for how the final $e-1$ treatment effects in the sequence of treatment effect dynamics are not realized before $\bar{t}$. It follows that

$$\mathbf{R}_e^{\Delta^1}\mathbf{x}^{\Delta^1}=\begin{bmatrix}\mathbf{0}_{(-\underline{t}+e)\times 1}\\ \sum_{g\in\mathcal{G}}\mathbb{1}[\check{e}(z_1,g)=e]\ \Delta^1(g,0)\\ \vdots\\ \sum_{g\in\mathcal{G}}\mathbb{1}[\check{e}(z_{|\mathcal{Z}|},g)=e]\ \Delta^1(g,\bar{t}-e)\end{bmatrix}$$

$\mathbf{R}^{\Delta^1}\mathbf{x}^{\Delta^1}$ simply stacks the expression above across all e.

Combining the results above, the matrix

$$\mathbf{R}^{\mathrm{M}}\equiv\begin{bmatrix}\mathbf{R}^{Y(\infty)} & \mathbf{R}^{\Delta^1}\end{bmatrix}$$

selects the parameters in $\mathbf{x}^{Y(\infty)}$ and $\mathbf{x}^{\Delta^1}$ out of $\mathbf{x}^{\mathrm{All},1}$. That is, each entry of $\mathbf{R}^{\mathrm{M}}\mathbf{x}^{\mathrm{All},1}$ is equal to

$$\sum_{g\in\mathcal{G}}\mathbb{1}[\check{e}(z,g)=e]\ (\mathbb{E}[Y_{it}(\infty)\mid G=g]+\mathbb{1}[t-e>0]\ \Delta^1(g,t-e)) \tag{A.9}$$

for some e, z, and t.

We now incorporate the weights (A.5) into (A.9). We first construct the denominator of the weights. Note that $\mathbf{R}^{Y(\infty)}$ selects the relevant agent types for all e, z, and t.[28] It follows that

$$\mathbf{R}^{Y(\infty)}\ (\mathbf{1}_{|\mathcal{T}\times 1|}\otimes\mathbf{x}^1)=\begin{bmatrix}\sum_{g\in\mathcal{G}}\mathbb{1}[\check{e}(z_1,g)=1]\ \mathbb{P}[G_i=g]\\ \vdots\\ \sum_{g\in\mathcal{G}}\mathbb{1}[\check{e}(z_{|\mathcal{Z}|},g)=\bar{e}]\ \mathbb{P}[G_i=g]\end{bmatrix},$$

where $\mathbf{1}_{|\mathcal{T}|}$ is an $|\mathcal{T}|\times 1$ vector of 1s. The expression above thus contains the values in the denominator of the weights in (A.5). For convenience, define

$$\mathbf{R}^{\mathrm{L,Arg}}\equiv\mathbf{R}^{Y(\infty)}(\mathbf{1}_{|\mathcal{T}|}\otimes\mathbf{I}_{|\mathcal{G}|}) \tag{A.10}$$

[28]In contrast, $\mathbf{R}^{\Delta^1}$ does not select the relevant types for each e and z across all periods. This is because $\mathbf{R}^{\Delta^1}$ only selects types in the periods where treatment effects are realized.

so that

$$\mathbf{R}^{\text{L,Arg}}\ \mathbf{x}^1 = \mathbf{R}^{Y(\infty)}\ (\mathbf{1}_{|\mathcal{T}|} \otimes \mathbf{x}^1)$$

contains the denominators of all the weights.

To incorporate the numerator of the weights, define

$$\mathbf{R}^{\text{R,Arg}} \equiv \mathbf{1}_{|\mathcal{T}|+|\mathcal{S}|} \otimes \mathbf{I}_{|\mathcal{G}|},$$

where each entry in $\mathbf{1}_{|\mathcal{T}|+|\mathcal{S}|}$ corresponds to $\mathbf{x}_t^{Y(\infty)}$ for a particular $t \in \mathcal{T}$ or $\mathbf{x}_s^{\Delta^1}$ for a particular $s \in \mathcal{S}$. Each row of $\mathbf{R}^{\text{R,Arg}}$ thus corresponds to a particular parameter in $\mathbf{x}^{\text{All},1}$. It follows that

$$\mathbf{R}^{\text{R,Arg}}\ \mathbf{x}^1 = \begin{bmatrix} \mathbb{P}[G_i = g_1] \\ \vdots \\ \mathbb{P}[G_i = g_{|\mathcal{G}|}] \\ \vdots \\ \mathbb{P}[G_i = g_1] \\ \vdots \\ \mathbb{P}[G_i = g_{|\mathcal{G}|}] \end{bmatrix}.$$

and

$$\begin{aligned} &\operatorname{diag}\left(\mathbf{R}^{\text{R,Arg}}\ \mathbf{x}^1\right)\ \mathbf{x}^{\text{All},1} \\ &= \begin{bmatrix} \mathbb{P}[G_i = 1]\ \mathbb{E}[Y_{i,\underline{t}}(\infty) \mid G = 1] \\ \vdots \\ \mathbb{P}[G_i = \overline{g}]\ \mathbb{E}[Y_{i,\underline{t}}(\infty) \mid G = \overline{g}] \\ \vdots \\ \mathbb{P}[G_i = 1]\ \mathbb{E}[Y_{i,\overline{t}}(\infty) \mid G = 1] \\ \vdots \\ \mathbb{P}[G_i = \overline{g}]\ \mathbb{E}[Y_{i,\overline{t}}(\infty) \mid G = \overline{g}] \\ \mathbb{P}[G_i = 1]\ \Delta^1(1, 0) \\ \vdots \\ \mathbb{P}[G_i = \overline{g}]\ \Delta^1(\overline{g}, 0) \\ \vdots \\ \mathbb{P}[G_i = 1]\ \Delta^1(1, \overline{t} - 1) \\ \vdots \\ \mathbb{P}[G_i = \overline{g}]\ \Delta^1(\overline{g}, \overline{t} - 1) \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \text{Parameters for } t = \underline{t} \\ \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \text{Parameters for } t = \overline{t} \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \text{Parameters for } s = 0 \\ \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \text{Parameters for } s = \overline{t} - 1 \end{matrix} \end{aligned}$$

Finally, define

$$\mathbf{R}^{\mathrm{L}} \equiv \operatorname{diag}\left(\mathbf{R}^{\mathrm{L,Arg}}\ \mathbf{x}^{1}\right)^{-1}, \tag{A.11}$$

$$\mathbf{R}^{\mathrm{R}} \equiv \operatorname{diag}\left(\mathbf{R}^{\mathrm{R,Arg}}\ \mathbf{x}^{1}\right), \tag{A.12}$$

$$\mathbf{A}^{\mathrm{All},1} \equiv \mathbf{R}^{\mathrm{L}}\ \mathbf{R}^{\mathrm{M}}\ \mathbf{R}^{\mathrm{R}}, \tag{A.13}$$

where $\mathbf{R}^{\mathrm{M}}$ generates the weightless expression in (A.9), and $\mathbf{R}^{\mathrm{L}}$ and $\mathbf{R}^{\mathrm{R}}$ incorporate the weights back in. The full system of equations for (A.2) may then be written as

$$\mathbf{A}^{\mathrm{All},1}\ \mathbf{x}^{\mathrm{All},1} = \underbrace{\mathbf{R}^{\mathrm{L}}}_{\text{Contains } \left[\sum_{g\in\mathcal{G}} \mathbb{1}[\check{e}(z,g)=e]\ \mathbb{P}[G_i=g]\right]^{-1}} \overbrace{\mathbf{R}^{\mathrm{M}}\ \mathbf{R}^{\mathrm{R}}}^{\text{Performs } \sum_{g\in\mathcal{G}} \mathbb{1}[\check{e}(z,g)=e]\ \mathbb{P}[G_i=g]}\ \mathbf{x}^{\mathrm{All},1} = \mathbf{b}^{2}.$$

■

Proof of Proposition 3. The proof is similar to that of Proposition 2 and reuses some of its derivations. In the main paper, we show that we can write $\mathbb{E}[Y_{i,t} \mid E_i, Z_i]$ as

$$\mathbb{E}[Y_{i,t} \mid E_i, Z_i] = \mathbb{E}[Y_{i,t}(\infty) \mid E_i, Z_i] + \mathbb{E}[\Delta_i(E_i, t-E_i) \mid E_i, Z_i].$$

We also show that term $\mathbb{E}[Y_{i,t}(\infty) \mid E_i, Z_i]$ can be written as in (A.3), and $\mathbb{E}[\Delta_i(E_i, t-E_i) \mid E_i, Z_i] = \Delta^2(e, t-e)$.

We reuse the derivations from the proof of Proposition 2 to express $\{\mathbb{E}[Y_{i,t}(\infty) \mid E_i, Z_i]\}$ as a linear system of equations. Define $\mathbf{x}^{Y(\infty)}$ as in (A.6), $\mathbf{R}^{Y(\infty)}$ as in (A.8), $\mathbf{R}^{\mathrm{L,Arg}}$ as in (A.10), and

$$\mathbf{R}^{\mathrm{R,Arg}} \equiv \mathbf{1}_{|\mathcal{T}|\times 1} \otimes \mathbf{I}_{|\mathcal{G}|},$$

$$\mathbf{R}^{\mathrm{M}} \equiv \mathbf{R}^{Y(\infty)}. \tag{A.14}$$

Then by defining $\mathbf{R}^{\mathrm{L}}$ and $\mathbf{R}^{\mathrm{R}}$ as in (A.11)–(A.12),[29]

$$\mathbf{R}^{\mathrm{L}}\ \mathbf{R}^{\mathrm{M}}\ \mathbf{R}^{\mathrm{R}}\ \mathbf{x}^{Y(\infty)}$$
$$= \begin{bmatrix} \mathbb{E}[Y_{i,\underline{t}}(\infty) \mid E_i = 1, Z_i = z_1] \\ \vdots \\ \mathbb{E}[Y_{i,\underline{t}}(\infty) \mid E_i = 1, Z_i = z_{|\mathcal{Z}|}] \\ \vdots \\ \mathbb{E}[Y_{i,\bar{t}}(\infty) \mid E_i = \bar{e}, Z_i = z_1] \\ \vdots \\ \mathbb{E}[Y_{i,\bar{t}}(\infty) \mid E_i = \bar{e}, Z_i = z_{|\mathcal{Z}|}] \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} \text{Moments for } t = \underline{t},\ E_i = 1 \\ \\ \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} \text{Moments for } t = \bar{t},\ E_i = \bar{e} \end{matrix}$$

To select the appropriate treatment effects, define

$$\mathbf{M}^{\Delta^2}(e,t) \equiv \begin{bmatrix} \mathbf{0}_{|\mathcal{Z}|\times(t-e)} & \mathbf{1}_{|\mathcal{Z}|\times 1} & \mathbf{0}_{|\mathcal{Z}|\times(|\mathcal{S}|-(t-e+1))} \end{bmatrix},$$
$$\mathbf{R}^{\Delta^2}_e \equiv \begin{bmatrix} \mathbb{1}[e=1]\ \mathbf{M}^{\Delta^2}(1,1) & \dots & \mathbb{1}[e=\bar{e}]\ \mathbf{M}^{\Delta^2}(\bar{e},1) \\ \vdots & \ddots & \vdots \\ \mathbb{1}[e=1]\ \mathbf{M}^{\Delta^2}(1,\bar{t}) & \dots & \mathbb{1}[e=\bar{e}]\ \mathbf{M}^{\Delta^2}(\bar{e},\bar{t}) \end{bmatrix},$$
$$\mathbf{R}^{\Delta^2} \equiv \begin{bmatrix} \mathbf{R}^{\Delta^2}_1 \\ \vdots \\ \mathbf{R}^{\Delta^2}_{\bar{e}} \end{bmatrix},$$

where each column of $\mathbf{M}^{\Delta^2}$ corresponds to $\Delta^2(e,s)$ for a particular s; and each block $\mathbf{R}^{\Delta^2}_e$ selects $\Delta^2(e,t-e)$ for an $e \in \mathcal{E}$ and $t \in \{1,\dots,\bar{t}\}$. Also define

$$\mathbf{x}^{\Delta^2}_e \equiv \begin{bmatrix} \Delta^2(e,0) \\ \vdots \\ \Delta^2(e,|\mathcal{S}|-1) \end{bmatrix},$$
$$\mathbf{x}^{\Delta^2} \equiv \begin{bmatrix} \mathbf{x}^{\Delta^2}_1 \\ \vdots \\ \mathbf{x}^{\Delta^2}_{\bar{e}} \end{bmatrix}$$

to be the vector of treatment effect parameters when treatment effects are heterogeneous across cohorts. Then $\mathbf{R}^{\Delta^2}\ \mathbf{x}^{\Delta^2}$ selects the appropriate treatment effects conditional on

[29]The vector $\mathbf{R}^{\mathrm{L}}\ \mathbf{R}^{\mathrm{M}}\ \mathbf{R}^{\mathrm{R}}\ \mathbf{x}^{Y(\infty)}$ is sorted by (e,t,z), as in $\mathbf{b}^2$ in (9).

E_i and Z_i. To see this, note that

$$\mathbf{R}_e^{\Delta^2}\ \mathbf{x}^{\Delta^2} = \begin{bmatrix} \mathbb{1}[e=1]\ \mathbf{M}^{\Delta^2}(1,1)\ \mathbf{x}_1^{\Delta^2} & \dots & \mathbb{1}[e=\bar{e}]\ \mathbf{M}^{\Delta^2}(\bar{e},1)\ \mathbf{x}_{\bar{e}}^{\Delta^2} \\ \vdots & \ddots & \vdots \\ \mathbb{1}[e=1]\ \mathbf{M}^{\Delta^2}(1,\bar{t})\ \mathbf{x}_1^{\Delta^2} & \dots & \mathbb{1}[e=\bar{e}]\ \mathbf{M}^{\Delta^2}(\bar{e},\bar{t})\ \mathbf{x}_{\bar{e}}^{\Delta^2} \end{bmatrix},$$

where the only non-zero column in the block matrix above corresponds to the period of take up e. Moreover, by construction of $\mathbf{M}^{\Delta^2}(e,t)$, the product $\mathbf{M}^{\Delta^2}(e,t)\ \mathbf{x}_e^{\Delta^2}$ only selects the $(t-e)^{\text{th}}$ entry in $\mathbf{x}_e^{\Delta^2}$, which is $\Delta^2(e,t-e)$.

Then if we define $\mathbf{b}^2$ as in (9), and

$$\mathbf{x}^{\text{All},2} \equiv \begin{bmatrix} \mathbf{x}^{Y(\infty)} \\ \mathbf{x}^{\Delta^2} \end{bmatrix},$$
$$\mathbf{A}^{\text{All},2} \equiv \begin{bmatrix} \mathbf{R}^{\text{L}}\ \mathbf{R}^{\text{M}}\ \mathbf{R}^{\text{R}} & \mathbf{R}^{\Delta^2} \end{bmatrix}$$

we have

$$\begin{aligned} \mathbf{A}^{\text{All},2}\ \mathbf{x}^{\text{All},2} &= \mathbf{R}^{\text{L}}\ \mathbf{R}^{\text{M}}\ \mathbf{R}^{\text{R}}\ \mathbf{x}^{Y(\infty)} + \mathbf{R}^{\Delta^2}\ \mathbf{x}^{\Delta^2} \\ &= \mathbf{b}^2. \end{aligned}$$

■

Proof of Proposition 4. When there are covariates $W_{i,t}$, the terms in (A.5) potentially depend on $W_{i,t}$, implying the weights for the agent types potentially depend on $W_{i,t}$. Let

$$\omega(e,z,g,w) \equiv \frac{\mathbb{1}[\check{e}(z,g,w)=e]\ \mathbb{P}[G_i=g \mid W_{i,t}=w]}{\sum_{g'\in\mathcal{G}} \mathbb{1}[\check{e}(z,g',w)=e]\ \mathbb{P}[G_i=g' \mid W_{i,t}=w]}$$

denote the weights when there are covariates. Under Assumption 5, $\check{e}(z,g,w)=\check{e}(z,g)$. Under Assumption 7a(i), $\mathbb{P}[G_i=g \mid W_{i,t}=w]=\mathbb{P}[G_i=g]$. It follows that

$$\omega(e,z,g,w) \equiv \frac{\mathbb{1}[\check{e}(z,g)=e]\ \mathbb{P}[G_i=g]}{\sum_{g'\in\mathcal{G}} \mathbb{1}[\check{e}(z,g')=e]\ \mathbb{P}[G_i=g']} = \omega(e,z,g).$$

Then under Assumptions 1–3, 4a, 5–7a, we can decompose $\mathbb{E}[Y_{i,t} \mid E_i=e, Z_i=z, W_{i,t}=$

$w]$ as

$$
\begin{aligned}
&\mathbb{E}[Y_{i,t} \mid E_i = e, Z_i = z, W_{i,t} = w] \\
&= \mathbb{E}[Y_{i,t}(\infty) \mid E_i = e, Z_i, W_{i,t} = w] + \mathbb{E}[\Delta_i(t-e) \mid E_i = e, Z_i = z, W_{i,t} = w] \\
&= \sum_{g \in \mathcal{G}} \overbrace{\omega(e,z,g,w)}^{\text{As. 5, 7a} \Rightarrow \omega(e,z,g)} \; \mathbb{E}[\overbrace{Y_{i,t}(\infty)}^{\text{Apply As. 6}} \mid \underbrace{E_i = e}_{E_i \text{ is determined conditional on } Z_i,\, G_i}, Z_i = z, G_i = g, W_{i,t} = w] \\
&\qquad + \sum_{g \in \mathcal{G}} \overbrace{\omega(e,z,g,w)}^{\text{As. 5, 7a} \Rightarrow \omega(e,z,g)} \; \underbrace{\mathbb{E}[\Delta_i(t-e) \mid \underbrace{E_i = e}_{E_i \text{ is determined conditional on } Z_i,\, G_i}, Z_i = z, G_i = g, W_{i,t} = w]}_{\text{As. 7a} \Rightarrow \Delta_i(t-e) \perp\!\!\!\perp (Z_i, W_{i,t}) \mid G_i} \\
&= \sum_{g \in \mathcal{G}} \omega(e,z,g) \; \mathbb{E}[h(W_{i,t}) + \alpha_{G_i t} + U_{i,t} \mid Z_i = z, G_i = g, W_{i,t} = w] \\
&\qquad + \sum_{g \in \mathcal{G}} \omega(e,z,g) \; \underbrace{\mathbb{E}[\Delta_i(t-e) \mid G_i = g]}_{\Delta^1(g,t-e)} \\
&= \sum_{g \in \mathcal{G}} \omega(e,z,g) \; (h(w) + \alpha_{gt} + \underbrace{\mathbb{E}[U_{i,t} \mid Z_i = z, G_i = g, W_{i,t} = w]}_{0 \text{ by As. 6}}) \\
&\qquad + \sum_{g \in \mathcal{G}} \omega(e,z,g) \; \Delta^1(g,t-e) \\
&= h(w) + \underbrace{\sum_{g \in \mathcal{G}} \omega(e,z,g) \; \alpha_{gt} + \sum_{g \in \mathcal{G}} \omega(e,z,g) \; \Delta^1(g,t-e)}_{\text{Depends on } E_i,\, Z_i, \text{ but not } W_{i,t}},
\end{aligned}
$$

where the first equality follows by definition of $Y_{i,t}$; the second equality follows by LIE; the third equality follows from applying Assumptions 5 and 7a(i) to $\omega(e,z,g,w)$, replacing $Y_{i,t}(\infty)$ with the model specified in Assumption 6(i), and applying Assumption 7a(ii) to the expectation containing $\Delta_i(t-e)$; the fourth equality follows from expanding the conditional expectation in the first sum; and the final equality follows by applying Assumption 6(ii).

It follows that

$$
\begin{aligned}
&\mathbb{E}[Y_{i,t} - h(W_{i,t}) \mid E_i = e, Z_i = z, W_i] \\
&= \mathbb{E}[Y_{i,t} - h(W_{i,t}) \mid E_i = e, Z_i = z] \\
&= \sum_{g \in \mathcal{G}} \omega(e,z,g) \; \alpha_{gt} + \sum_{g \in \mathcal{G}} \omega(e,z,g) \; \Delta^1(g,t-e). \qquad \text{(A.15)}
\end{aligned}
$$

We can then apply Proposition 2 to (A.15) to recover the treatment effects. So define

$$\mathbf{b}^{\text{Cov},1} = \begin{bmatrix} \mathbb{E}[Y_{i,1} - h(W_{i,t}) \mid E_i = 1, Z_i = z_1] \\ \vdots \\ \mathbb{E}[Y_{i,\bar{t}} - h(W_{i,t}) \mid E_i = \bar{t}, Z_i = z_{|\mathcal{Z}|}] \end{bmatrix},$$

where $h(\cdot)$ is identified by nonparametrically regressing $Y_{i,t}$ on $W_{i,t}$, controlling for E_i and Z_i. Then $\mathbf{A}^{\text{All},1}\mathbf{x}^{\text{All},1} = \mathbf{b}^{\text{Cov},1}$. ■

Proof of Proposition 5. The proof is similar to that of Proposition 4. Under Assumptions 1–3, 4b, 5–6, and 7b, agents are weighted by $\omega(e, z, g)$ (see proof of Proposition 4), and we can decompose $\mathbb{E}[Y_{i,t} \mid E_i = e, Z_i = z, W_{i,t} = w]$ as

$$\begin{aligned}
&\mathbb{E}[Y_{i,t} \mid E_i = e, Z_i = z, W_{i,t} = w] \\
&= \underbrace{\mathbb{E}[Y_{i,t}(\infty) \mid E_i = e, Z_i = z, W_{i,t} = w]}_{\text{Decompose as in (A.15)}} + \mathbb{E}[\Delta_i(t-e) \mid E_i = e, Z_i = z, W_{i,t} = w] \\
&= h(w) + \sum_{g \in \mathcal{G}} \omega(e, z, g)\alpha_{gt} \\
&\qquad + \sum_{g \in \mathcal{G}} \omega(e, z, g) \underbrace{\mathbb{E}[\Delta_i(t-e) \mid E_i = e, Z_i = z, G_i = g, W_{i,t} = w]}_{\text{As. 7b} \Rightarrow \Delta_i(t-e) \perp\!\!\!\perp (Z_i, G_i, W_{i,t}) \mid E_i} \\
&= h(w) + \sum_{g \in \mathcal{G}} \omega(e, z, g)\alpha_{gt} + \underbrace{\sum_{g \in \mathcal{G}} \omega(e, z, g) \underbrace{\mathbb{E}[\Delta_i(t-e) \mid E_i = e]}_{\Delta^2(e, t-e), \text{ does not depend on } g}}_{\text{Sums to } \Delta^2(e, t-e)} \\
&= \underbrace{h(w) + \sum_{g \in \mathcal{G}} \omega(e, z, g)\alpha_{gt} + \Delta^2(e, t-e)}_{\text{Depends on } E_i,\ Z_i, \text{ but not } W_{i,t}}, \qquad \text{(A.16)}
\end{aligned}$$

where the first equality follows from the definition of $Y_{i,t}$; the second equality follows from the decomposition of $\mathbb{E}[Y_{i,t}(\infty) \mid E_i = e, Z_i = z, W_{i,t} = w]$ from (A.15), and LIE; the third equality follows from Assumption 7b; and the final equality follows from collecting terms. As in Proposition 4, equation (A.16) suggests that we can identify the treatment effects by first residualizing $Y_{i,t}$ with respect to $W_{i,t}$ before applying the identification strategy described above. ■

A.2 Proofs for estimation and inference

Let $\mathbb{E}_n$, $\mathbb{P}_n$, and $\mathbb{V}_n$ denote the sample analogs to $\mathbb{E}$, $\mathbb{P}$, and $\mathbb{V}$ for a sample of size n. Let $\mathbf{x}_n^1$ and $\mathbf{x}_n^{\text{Sub},j}$ for $j = 1, 2$ denote the first and second stage estimates. Their asymptotic distributions may be derived using the delta method. The following conventions and derivations will be useful.

Let $\mathbf{v} \in \mathbb{R}^p$, $\mathbf{b}(\mathbf{v}) \in \mathbb{R}^m$, and $\mathbf{A}(\mathbf{v}) \in \mathbb{R}^{m\times n}$. Then

$$\frac{\partial}{\partial \mathbf{v}}\mathbf{b}(\mathbf{v}) \in \mathbb{R}^{m\times p}, \quad \frac{\partial}{\partial \mathbf{v}}\mathbf{A}(\mathbf{v}) \in \mathbb{R}^{m\times n\times p}.$$

Let $\mathsf{D} \in \mathbb{R}^{m\times n\times p}$ be an array, $\mathbf{A}^1 \in \mathbb{R}^{q\times m}$ and $\mathbf{A}^2 \in \mathbb{R}^{n\times r}$ be matrices. Then

$$[\mathbf{A}^1\ \mathsf{D}\ \mathbf{A}^2]_{\bullet\bullet i} = \mathbf{A}^1\ \mathsf{D}_{\bullet\bullet i}\mathbf{A}^2$$

and is a $q\times r\times p$ matrix. We use the notation $\mathsf{D}^{(d_1,d_2,d_3)}$ to rearrange the dimensions of the array in the order (d_1, d_2, d_3). For example, $\mathsf{D}^{(3,2,1)}$ is a $p\times n\times m$ matrix. $\left[\mathsf{D}^{(1,3,2)}\right]^{(1,3,2)}$ is a $m\times n\times p$ matrix.

Finally, let $\mathbf{A}(\mathbf{v}) = \text{diag}\left(\mathbf{R}\mathbf{v}\right)$ for some arbitrary matrix $\mathbf{R} \in \mathbb{R}^{m\times p}$ and vector $\mathbf{v} \in \mathbb{R}^p$. Then

$$\left[\frac{d}{d\mathbf{v}}\mathbf{A}(\mathbf{v})\right]_{ij\bullet} = \begin{cases} \mathbf{R}_{i\bullet} & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases} \tag{A.17}$$

$$\frac{d}{d\mathbf{v}}\mathbf{A}(\mathbf{v})^{-1} = -\mathbf{A}(\mathbf{v})^{-1}\ \left[\frac{d}{d\mathbf{v}}\mathbf{A}(\mathbf{v})\right]\ \mathbf{A}(\mathbf{v})^{-1}, \tag{A.18}$$

where the second line follows rules on differentiating inverses of matrices. Note that $\frac{d}{d\mathbf{v}}\mathbf{A}$ is an $m\times m\times p$ array.

Proof of Proposition 6. We estimate $\mathbf{x}^1$ using the sample analog of the identification result. Define[30]

$$v^1_{i,e,z} \equiv \mathbb{1}[E_i = e, Z_i = z],$$

$$\mathbf{v}^1_i \equiv \begin{bmatrix} v^1_{i,1,z_1} \\ \vdots \\ v^1_{i,1,z_{|\mathcal{Z}|}} \\ \vdots \\ v^1_{i,\bar{e},z_1} \\ \vdots \\ v^1_{i,\bar{e},z_{|\mathcal{Z}|}} \end{bmatrix} \quad \begin{matrix} \left.\vphantom{\begin{matrix} a \\ b \\ c \end{matrix}}\right\} v^1_{i,e,z} \text{ for } e = 1,\ z \in \mathcal{Z}, \\ \\ \left.\vphantom{\begin{matrix} a \\ b \\ c \end{matrix}}\right\} v^1_{i,e,z} \text{ for } e = \bar{e},\ z \in \mathcal{Z}, \end{matrix}$$

$$\overline{\mathbf{v}}^1_n \equiv \mathbb{E}_n[\mathbf{v}^1_i],$$

$$\Sigma^1 \equiv \mathbb{V}[\mathbf{v}^1_i],$$

$$\Sigma^1_n \equiv \mathbb{V}_n[\mathbf{v}^1_i],$$

[30] The vectors $\mathbf{v}^1_i$ and $\overline{\mathbf{v}}^1_n$ are ordered by (e, z).

where each entry of $\overline{\mathbf{v}}_n^1$ is equal to $\mathbb{P}_n[E_i = e, Z_i = z]$ for some $e \in \mathcal{E}$, $z \in \mathcal{Z}$. Define

$$\mathbf{R}^1 \equiv \mathbf{1}_{|\mathcal{E}|\times|\mathcal{E}|} \otimes \mathbf{I}_{|\mathcal{Z}|}.$$

Then

$$\mathbf{R}^1\, \overline{\mathbf{v}}_n^1 = \begin{bmatrix} \sum_{e\in\mathcal{E}} \mathbb{P}_n[E_i = e, Z_i = z_1] \\ \vdots \\ \sum_{e\in\mathcal{E}} \mathbb{P}_n[E_i = e, Z_i = z_{|\mathcal{Z}|}] \\ \vdots \\ \sum_{e\in\mathcal{E}} \mathbb{P}_n[E_i = e, Z_i = z_1] \\ \vdots \\ \sum_{e\in\mathcal{E}} \mathbb{P}_n[E_i = e, Z_i = z_{|\mathcal{Z}|}] \end{bmatrix} \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\}\text{Repeated } |\mathcal{E}| \text{ times}$$

$$= \begin{bmatrix} \mathbb{P}_n[Z_i = z_1] \\ \vdots \\ \mathbb{P}_n[Z_i = z_{|\mathcal{Z}|}] \\ \vdots \\ \mathbb{P}_n[Z_i = z_1] \\ \vdots \\ \mathbb{P}_n[Z_i = z_{|\mathcal{Z}|}] \end{bmatrix} \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\}\text{Repeated } |\mathcal{E}| \text{ times.}$$

Then $\mathbf{b}^1$ may be estimated by

$$\begin{aligned} \mathbf{b}_n^1 &\equiv \operatorname{diag}\left(\mathbf{R}^1\overline{\mathbf{v}}_n^1\right)^{-1}\ \overline{\mathbf{v}}_n^1 \\ &= \begin{bmatrix} \frac{\mathbb{P}_n[E_i=1,Z_i=z_1]}{\mathbb{P}_n[Z_i=z_1]} \\ \vdots \\ \frac{\mathbb{P}_n[E_i=1,Z_i=z_{|\mathcal{Z}|}]}{\mathbb{P}_n[Z_i=z_{|\mathcal{Z}|}]} \\ \vdots \\ \frac{\mathbb{P}_n[E_i=\overline{e},Z_i=z_1]}{\mathbb{P}_n[Z_i=z_1]} \\ \vdots \\ \frac{\mathbb{P}_n[E_i=\overline{e},Z_i=z_{|\mathcal{Z}|}]}{\mathbb{P}_n[Z_i=z_{|\mathcal{Z}|}]} \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\}\mathbb{P}_n[E_i = 1 \mid Z_i = z] \text{ for } z \in \mathcal{Z}, \\ \\ \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\}\mathbb{P}_n[E_i = \overline{e} \mid Z_i = z] \text{ for } z \in \mathcal{Z}. \end{matrix} \end{aligned}$$

Define the estimator $h : \mathbb{R}^{\dim(\mathbf{v}^1)} \to \mathbb{R}^{\overline{g}}$ as

$$h^1(\mathbf{v}) \equiv (\mathbf{A}^{1\top}\mathbf{A}^1)^{-1}\mathbf{A}^{1\top}\operatorname{diag}\left(\mathbf{R}^1\mathbf{v}\right)^{-1}\ \mathbf{v}. \tag{A.19}$$

Then

$$\mathbf{x}^1 = h^1(\mathbb{E}[\mathbf{v}_i^1]),$$

which we estimate using

$$\mathbf{x}_n^1 = h^1(\overline{\mathbf{v}}_n^1). \tag{A.20}$$

If $\mathbf{A}^1$ is full rank and $\mathbb{P}[Z_i = z] > 0$ for all $z \in \mathcal{Z}$, then $h^1(\cdot)$ is continuous. Then by the weak law of large numbers and the continuous mapping theorem, we have $\overline{\mathbf{v}}_n^1 \xrightarrow{\text{p}} \mathbb{E}[\mathbf{v}_i^1]$ and $\mathbf{x}_n^1 \xrightarrow{\text{p}} \mathbf{x}^1$ as $n \to \infty$.

To conduct inference, we use the delta method. The matrices $\mathbf{A}^1$ and $\mathbf{R}^1$ in (A.19) are nonrandom, so $\mathbf{x}_n^1$ is a function of $\overline{\mathbf{v}}_n^1$, for which

$$\sqrt{n}(\overline{\mathbf{v}}_n^1 - \mathbb{E}[\mathbf{v}_i^1]) \xrightarrow{\text{d}} N(0, \Sigma^1).$$

Taking the derivative of (A.19) with respect to $\mathbf{v}$, we have

$$\begin{aligned}\frac{d}{d\mathbf{v}} h^1(\mathbf{v}) &= \frac{d}{d\mathbf{v}} \underbrace{(\mathbf{A}^{1\top}\mathbf{A}^1)^{-1}\mathbf{A}^{1\top}}_{\text{Nonrandom}} \operatorname{diag}\left(\mathbf{R}^1\mathbf{v}\right)^{-1} \mathbf{v} \\ &= (\mathbf{A}^{1\top}\mathbf{A}^1)^{-1}\mathbf{A}^{1\top} \frac{d}{d\mathbf{v}} \left[\operatorname{diag}\left(\mathbf{R}^1\mathbf{v}\right)^{-1} \mathbf{v}\right].\end{aligned}$$

The derivative of $\operatorname{diag}\left(\mathbf{R}^1\mathbf{v}\right)^{-1}\mathbf{v}$ is

$$\begin{aligned}&\frac{d}{d\mathbf{v}}\operatorname{diag}\left(\mathbf{R}^1\mathbf{v}\right)^{-1} \mathbf{v} \\ &= \underbrace{\left[\frac{d}{d\mathbf{v}}\operatorname{diag}\left(\mathbf{R}^1\mathbf{v}\right)^{-1}\right]}_{\text{Apply (A.18)}} \mathbf{v} + \operatorname{diag}\left(\mathbf{R}^1\mathbf{v}\right)^{-1} \underbrace{\left[\frac{d}{d\mathbf{v}}\mathbf{v}\right]}_{\mathbf{I}_{|\mathcal{E}||\mathcal{Z}|}} \\ &= -\operatorname{diag}\left(\mathbf{R}^{-1}\mathbf{v}\right) \underbrace{\left[\frac{d}{d\mathbf{v}}\operatorname{diag}\left(\mathbf{R}^1\mathbf{v}\right)\right]}_{\text{Apply (A.17)}} \operatorname{diag}\left(\mathbf{R}^{-1}\mathbf{v}\right) + \operatorname{diag}\left(\mathbf{R}^1\mathbf{v}\right)^{-1},\end{aligned}$$

where the first equality is by the product rule; the second equality follows from (A.18). Putting everything together, define the derivative of (A.19) as

$$\begin{aligned}\mathbf{D}^1(\mathbf{v}) &\equiv \frac{d}{d\mathbf{v}} h^1(\mathbf{v}) \\ &= (\mathbf{A}^{1\top}\mathbf{A}^1)^{-1}\mathbf{A}^{1\top} \left[-\operatorname{diag}\left(\mathbf{R}^{-1}\mathbf{v}\right)\left[\frac{d}{d\mathbf{v}}\operatorname{diag}\left(\mathbf{R}^1\mathbf{v}\right)\right]\operatorname{diag}\left(\mathbf{R}^{-1}\mathbf{v}\right) \right. \\ &\qquad \left. + \operatorname{diag}\left(\mathbf{R}^1\mathbf{v}\right)^{-1}\right]\end{aligned}$$

and

$$\Psi^1 \equiv \mathbf{D}^1(\mathbb{E}[\mathbf{v}_i^1])\ \Sigma^1\ \mathbf{D}^1(\mathbb{E}[\mathbf{v}_i^1])^\top.$$

Then

$$\sqrt{n}(\mathbf{x}_n^1 - \mathbf{x}^1) \xrightarrow{\text{d}} N(0, \Psi^1).$$

Since $\overline{\mathbf{v}}_n^1 \xrightarrow{\text{p}} \mathbb{E}[\mathbf{v}_i^1]$, $\mathbf{D}^1(\cdot)$ is continuous, and $\Sigma_n^1 \xrightarrow{\text{p}} \Sigma^1$, we can consistently estimate Ψ^1 by

$$\Psi_n^1 \equiv \mathbf{D}^1(\overline{\mathbf{v}}_n^1)\ \Sigma_n^1\ \mathbf{D}^1(\overline{\mathbf{v}}_n^1)^\top.$$

■

Proof of Proposition 7. We estimate $\mathbf{x}^{\text{Sub},j}$ for $j = 1, 2$ using the sample analog of the identification results. We begin by showing how to estimate $\mathbf{b}^2$. Define[31]

$$\begin{aligned}
v_{i,t,e,z}^2 &\equiv Y_{it}\ \mathbb{1}[E_i = e, Z_i = z], \\
\mathbf{v}_i^2 &\equiv \begin{bmatrix} v_{i,\underline{t},1,z_1}^2 \\ \vdots \\ v_{i,\underline{t},1,z_{|\mathcal{Z}|}}^2 \\ \vdots \\ v_{i,\bar{t},\bar{e},z_1}^2 \\ \vdots \\ v_{i,\bar{t},\bar{e},z_{|\mathcal{Z}|}}^2 \end{bmatrix}
\begin{matrix} \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} v_{i,t,e,z}^2 \text{ for } t = \underline{t},\ e = 1,\ z \in \mathcal{Z}, \\ \\ \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} v_{i,t,e,z}^2 \text{ for } t = \bar{t},\ e = \bar{e},\ z \in \mathcal{Z}, \end{matrix} \\
\overline{\mathbf{v}}_n^2 &\equiv \mathbb{E}_n[\mathbf{v}_i^2], \\
\Sigma^2 &\equiv \mathbb{V}[(\mathbf{v}_i^{1\top}, \mathbf{v}_i^{2\top})^\top], \\
\Sigma_n^2 &\equiv \mathbb{V}_n[(\mathbf{v}_i^{1\top}, \mathbf{v}_i^{2\top})^\top].
\end{aligned}$$

Also define

$$\mathbf{R}^2 \equiv \mathbf{I}_{|\mathcal{E}|} \otimes \mathbf{1}_{|\mathcal{T}|\times 1} \otimes \mathbf{I}_{|\mathcal{Z}|}.$$

[31] The vector $\mathbf{v}_i^2$ is ordered by (e, t, z).

Then[32]

$$\mathbf{R}^2\overline{\mathbf{v}}^1 = \begin{bmatrix} \mathbb{P}_n[E_i = 1, Z_i = z_1] \\ \vdots \\ \mathbb{P}_n[E_i = 1, Z_i = z_{|\mathcal{Z}|}] \\ \vdots \\ \mathbb{P}_n[E_i = \overline{e}, Z_i = z_1] \\ \vdots \\ \mathbb{P}_n[E_i = \overline{e}, Z_i = z_{|\mathcal{Z}|}] \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \mathbb{P}_n[E_i = e, Z_i = z] \text{ for } e = 1,\ z \in \mathcal{Z},\ t = \underline{t}, \\ \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \mathbb{P}_n[E_i = e, Z_i = z] \text{ for } e = \overline{e},\ z \in \mathcal{Z},\ t = \overline{t}, \end{matrix}$$

and $\mathbf{b}^2$ may be estimated by

$$\begin{aligned} \mathbf{b}^2_n &\equiv \text{diag}\left(\mathbf{R}^2\overline{\mathbf{v}}^1_n\right)^{-1}\overline{\mathbf{v}}^2_n \\ &= \begin{bmatrix} \frac{\mathbb{E}_n[Y_{i,\underline{t}}\,\mathbb{1}[E_i=1,Z_i=z_1]]}{\mathbb{P}_n[E_i=1,Z_i=z_1]} \\ \vdots \\ \frac{\mathbb{E}_n[Y_{i,\underline{t}}\,\mathbb{1}[E_i=1,Z_i=z_{|\mathcal{Z}|}]]}{\mathbb{P}_n[E_i=1,Z_i=z_{|\mathcal{Z}|}]} \\ \vdots \\ \frac{\mathbb{E}_n[Y_{i,\overline{t}}\,\mathbb{1}[E_i=\overline{e},Z_i=z_1]]}{\mathbb{P}_n[E_i=\overline{e},Z_i=z_1]} \\ \vdots \\ \frac{\mathbb{E}_n[Y_{i,\overline{t}}\,\mathbb{1}[E_i=\overline{e},Z_i=z_{|\mathcal{Z}|}]]}{\mathbb{P}_n[E_i=\overline{e},Z_i=z_{|\mathcal{Z}|}]} \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \mathbb{E}_n[Y_{i,\underline{t}} \mid E_i = e, Z_i = z] \text{ for } e = 1,\ z \in \mathcal{Z},\ t = \underline{t}, \\ \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \mathbb{E}_n[Y_{i,\overline{t}} \mid E_i = e, Z_i = z] \text{ for } e = \overline{e},\ z \in \mathcal{Z},\ t = \underline{t}. \end{matrix} \end{aligned}$$

The matrix $\mathbf{A}^{\text{Sub},j}$ is a "normalized" version of $\mathbf{A}^{\text{All},j}$. The normalization simply drops certain rows from $\mathbf{A}^{\text{All},j}$, which is equivalent to dropping certain rows and columns in the matrices used to construct $\mathbf{A}^{\text{All},j}$. We leave the full discussion of the normalization to Appendix A.3, but show how to construct $\mathbf{A}^{\text{Sub},j}$ with the normalization.

In the proof to Propositions 2 and 3, we show that $\mathbf{A}^{\text{All},j}$ for $j = 1, 2$ consists of the matrices $\mathbf{R}^{\text{L}}$, $\mathbf{R}^{\text{M}}$, $\mathbf{R}^{\text{R}}$, and $\mathbf{R}^{\Delta^2}$. The matrices $\mathbf{R}^{\text{M}}$, $\mathbf{R}^{\Delta^2}$ are deterministic and can be constructed as originally defined. The matrices $\mathbf{R}^{\text{L}}$, $\mathbf{R}^{\text{R}}$ involve $\mathbf{x}^1$, which we estimate using $\mathbf{x}^1_n$. The normalization entails dropping a set of columns from $\mathbf{R}^{\text{M}}$, a set of rows from $\mathbf{R}^{\text{R,Arg}}$ (equivalent to dropping a set of columns from $\mathbf{R}^{\text{R}}$), and a set of columns from $\mathbf{R}^{\Delta^2}$. Let $\mathbf{R}^{\text{M,Norm}}$, $\mathbf{R}^{\text{R,Arg,Norm}}$, and $\mathbf{R}^{\Delta^2\text{,Norm}}$ denote the corresponding "normalized" matrices.[33]

[32] The vector $\mathbf{R}^2\overline{\mathbf{v}}^1$ is sorted by (e, t, z).

[33] Note that $\mathbf{R}^{\text{M}}$ is constructed differently depending on whether treatment effects are heterogeneous across agent types or cohorts. The normalized matrices $\mathbf{R}^{\text{M,Norm}}$ will also differ across the two frameworks.

Define $g^L : \mathbb{R}^{\dim(\mathbf{v}^1)} \to \mathbb{R}^{\dim(\mathbf{R}^{\mathrm{L}})}$ and $g^R : \mathbb{R}^{\dim(\mathbf{v}^1)} \to \mathbb{R}^{\dim(\mathbf{R}^{\mathrm{R}})}$ as

$$g^{\mathrm{L}}(\mathbf{v}) \equiv \mathrm{diag}\left(\mathbf{R}^{\mathrm{L,Arg}}\ h^1(\mathbf{v})\right)^{-1},$$
$$g^{\mathrm{R}}(\mathbf{v}) \equiv \mathrm{diag}\left(\mathbf{R}^{\mathrm{R,Arg,Norm}}\ h^1(\mathbf{v})\right),$$

where $h^1(\cdot)$ is defined in (A.19). Then we can estimate $\mathbf{R}^{\mathrm{L}}$ and the normalized version of $\mathbf{R}^{\mathrm{R}}$ as

$$\mathbf{R}^{\mathrm{L}}_n = g^L(\overline{v}^1_n),$$
$$\mathbf{R}^{\mathrm{R,Norm}}_n = g^R(\overline{v}^1_n).$$

Define $g^{\mathrm{Sub},1} : \mathbb{R}^{\dim(\mathbf{v}^1)} \to \mathbb{R}^{\dim(\mathbf{A}^{\mathrm{Sub},1})}$ as

$$g^{\mathrm{Sub},1}(\mathbf{v}) \equiv g^{\mathrm{L}}(\mathbf{v})\ \mathbf{R}^{\mathrm{M,Norm}}\ g^{\mathrm{R}}(\mathbf{v}), \tag{A.21}$$

where $\mathbf{R}^{\mathrm{M,Norm}}$ is the normalized counterpart to $\mathbf{R}^{\mathrm{M}}$ defined in (**??**). Then we can estimate $\mathbf{A}^{\mathrm{Sub},1}$ by

$$\mathbf{A}^{\mathrm{Sub},1}_n \equiv g^{\mathrm{Sub},1}(\overline{\mathbf{v}}^1_n). \tag{A.22}$$

Similarly, define $g^{\mathrm{Sub},2} : \mathbb{R}^{\dim(\mathbf{v}^1)} \to \mathbb{R}^{\dim(\mathbf{A}^{\mathrm{All},2})}$ as

$$g^{\mathrm{Sub},2}(\mathbf{v}) \equiv \left[\ g^{\mathrm{L}}(\mathbf{v})\ \mathbf{R}^{\mathrm{M,Norm}}\ g^{\mathrm{R}}(\mathbf{v}) \quad \mathbf{R}^{\Delta^2,\mathrm{Norm}}\ \right], \tag{A.23}$$

where $\mathbf{R}^{\mathrm{M,Norm}}$ is the normalized counterpart to $\mathbf{R}^{\mathrm{M}}$ defined in (A.14). Then we can estimate $\mathbf{A}^{\mathrm{Sub},2}$ by

$$\mathbf{A}^{\mathrm{Sub},2}_n \equiv g^{\mathrm{Sub},2}(\overline{\mathbf{v}}^1_n). \tag{A.24}$$

Finally, for $j = 1, 2$, define $h^{\mathrm{Sub},j} : \mathbb{R}^{\dim(\mathbf{v}^1_i)} \times \mathbb{R}^{\dim(\mathbf{v}^2_i)} \to \mathbb{R}^{\dim(\mathbf{x}^{\mathrm{Sub},j})}$ as

$$h^{\mathrm{Sub},j}(\mathbf{v}^1, \mathbf{v}^2) \equiv \left[g^{\mathrm{Sub},j}(\mathbf{v}^1)^\top g^{\mathrm{Sub},j}(\mathbf{v}^1)\right]^{-1} g^{\mathrm{Sub},j}(\mathbf{v}^1)^\top \underbrace{\mathrm{diag}\left(\mathbf{R}^2\mathbf{v}^1\right)^{-1}\ \mathbf{v}^2}_{\text{Estimator for } \mathbf{b}^2}, \tag{A.25}$$

By construction of $h^{\mathrm{Sub},j}(\cdot)$, it follows that

$$h^{\mathrm{Sub},j}(\mathbb{E}[\mathbf{v}^1_i], \mathbb{E}[\mathbf{v}^2_i]) = (\mathbf{A}^{\mathrm{Sub},j\top}\mathbf{A}^{\mathrm{Sub},j})^{-1}\mathbf{A}^{\mathrm{Sub},j\top}\mathbf{b}^2$$
$$= \mathbf{x}^{\mathrm{Sub},j}.$$

So we estimate $\mathbf{x}^{\text{Sub},j}$ with

$$\mathbf{x}_n^{\text{Sub},j} = h^{\text{Sub},j}(\overline{\mathbf{v}}_n^1, \overline{\mathbf{v}}_n^2). \tag{A.26}$$

If $\mathbf{A}^{\text{Sub},j}$ is full rank, then by weak law of large numbers and continuous mapping theorem, we have $(\overline{\mathbf{v}}_n^{1\top}, \overline{\mathbf{v}}_n^{2\top})^\top \xrightarrow{\text{p}} (\mathbb{E}[\mathbf{v}_i^1]^\top, \mathbb{E}[\mathbf{v}_i^2]^\top)^\top$ and $\mathbf{x}_n^{\text{Sub},j} \xrightarrow{\text{p}} \mathbf{x}^{\text{Sub},j}$ as $n \to \infty$.

To conduct inference, we use the delta method to obtain the distribution of the estimator. The estimator (A.26) is a function of $(\overline{\mathbf{v}}_n^{1\top}, \overline{\mathbf{v}}_n^{2\top})^\top$, for which

$$\sqrt{n}\left(\begin{bmatrix} \overline{\mathbf{v}}_n^1 \\ \overline{\mathbf{v}}_n^2 \end{bmatrix} - \mathbb{E}\begin{bmatrix} \mathbf{v}_i^1 \\ \mathbf{v}_i^2 \end{bmatrix}\right) \xrightarrow{\text{d}} N(0, \Sigma^2).$$

Define $\mathbf{D}^{\text{Sub},j}$ to be the derivative of $h^{\text{Sub},j}(\mathbf{v}^1, \mathbf{v}^2)$ with respect to $(\mathbf{v}^{1\top}, \mathbf{v}^{2\top})^\top$,

$$\mathbf{D}^{\text{Sub},j}(\mathbf{v}^1, \mathbf{v}^2) \equiv \begin{bmatrix} \frac{d}{d\mathbf{v}^1} h^{\text{Sub},j}(\mathbf{v}^1, \mathbf{v}^2) & \frac{d}{d\mathbf{v}^2} h^{\text{Sub},j}(\mathbf{v}^1, \mathbf{v}^2) \end{bmatrix}. \tag{A.27}$$

We show below how to derive $\mathbf{D}^{\text{Sub},j}(\mathbf{v}^1, \mathbf{v}^2)$, which is continuous in its arguments. Define

$$\Psi^{\text{Sub},j} \equiv \mathbf{D}^{\text{Sub},j}(\mathbb{E}[\mathbf{v}_i^1], \mathbb{E}[\mathbf{v}_i^2])\ \Sigma^2\ \mathbf{D}^{\text{Sub},j}(\mathbb{E}[\mathbf{v}_i^1], \mathbb{E}[\mathbf{v}_i^2])^\top,$$

which can be consistently estimated by

$$\Psi_n^{\text{Sub},j} \equiv \mathbf{D}^{\text{Sub},j}(\overline{\mathbf{v}}^1, \overline{\mathbf{v}}^2)\ \Sigma_n^2\ \mathbf{D}^{\text{Sub},j}(\overline{\mathbf{v}}^1, \overline{\mathbf{v}}^2)^\top.$$

Then

$$\sqrt{n}(\mathbf{x}_n^{\text{Sub},j} - \mathbf{x}^{\text{Sub},j}) \xrightarrow{\text{d}} N(0, \Psi^{\text{Sub},j}). \tag{A.28}$$

We now derive the $\mathbf{D}^{\text{Sub},j}(\mathbf{v}^1, \mathbf{v}^2)$ for $j = 1, 2$. We first consider the case of $j = 1$. The case of $j = 2$ is almost identical and we consider it at the end of the proof. For brevity, define

$$g^{\text{Inv},j}(\mathbf{v}^1) \equiv \left[g^{\text{Sub},j}(\mathbf{v}^1)^\top g^{\text{Sub},j}(\mathbf{v}^1)\right]^{-1}$$

for $j = 1, 2$. Differentiating (A.25) with respect to $\mathbf{v}^1$ when $j = 1$, we have

$$\begin{aligned}
&\frac{d}{d\mathbf{v}^1} h^{\text{Sub},1}(\mathbf{v}^1, \mathbf{v}^2) \\
&= \frac{d}{d\mathbf{v}^1} \underbrace{\left[g^{\text{Sub},1}(\mathbf{v}^1)^\top g^{\text{Sub},1}(\mathbf{v}^1)\right]^{-1}}_{g^{\text{Inv},1}(\mathbf{v}^1)} g^{\text{Sub},1}(\mathbf{v}^1)^\top \text{diag}\left(\mathbf{R}^2\mathbf{v}^1\right)^{-1} \ \mathbf{v}^2 \qquad \text{(A.29)} \\
&= \underbrace{\left[\frac{d}{d\mathbf{v}^1} g^{\text{Inv},1}(\mathbf{v}^1)\right]}_{\text{(A)}} \left[g^{\text{Sub},1}(\mathbf{v}^1)^\top \text{diag}\left(\mathbf{R}^2\mathbf{v}^1\right)^{-1} \mathbf{v}^2\right] \\
&\quad + g^{\text{Inv},1}(\mathbf{v}^1) \underbrace{\left[\frac{d}{d\mathbf{v}^1} g^{\text{Sub},1}(\mathbf{v}^1)^\top \text{diag}\left(\mathbf{R}^2\mathbf{v}^1\right)^{-1} \mathbf{v}^2\right]}_{\text{(B)}}, \qquad \text{(A.30)}
\end{aligned}$$

where the second equality follows from the product rule. Next, we expand the terms (A) and (B).

Beginning with (A), the derivative of $g^{\text{Inv},1}(\mathbf{v}^1)$ is

$$\begin{aligned}
&\frac{d}{d\mathbf{v}^1} g^{\text{Inv},1}(\mathbf{v}^1) \\
&= -g^{\text{Inv},1}(\mathbf{v}^1) \underbrace{\left[\frac{d}{d\mathbf{v}^1} g^{\text{Sub},1}(\mathbf{v}^1)^\top g^{\text{Sub},1}(\mathbf{v}^1)\right]}_{\text{Apply product rule}} g^{\text{Inv},1}(\mathbf{v}^1) \\
&= -g^{\text{Inv},1}(\mathbf{v}^1) \\
&\quad \times \left(\frac{d}{d\mathbf{v}^1} g^{\text{Sub},1}(\mathbf{v}^1)^\top\right) g^{\text{Sub},1}(\mathbf{v}^1) + g^{\text{Sub},1}(\mathbf{v}^1)^\top \left(\frac{d}{d\mathbf{v}^1} g^{\text{Sub},1}(\mathbf{v}^1)\right) \qquad \text{(A.31)} \\
&\quad \times g^{\text{Inv},1}(\mathbf{v}^1),
\end{aligned}$$

where the first line follows from the rules on differentiating inverted matrices; the second line follows from the chain rule. The derivative of $g^{\text{Sub},1}(\mathbf{v}^1)$ is

$$\begin{aligned}
&\frac{d}{d\mathbf{v}^1} g^{\text{Sub},1}(\mathbf{v}^1) \\
&= \frac{d}{d\mathbf{v}^1} \left[g^{\text{L}}(\mathbf{v}^1)\mathbf{R}^{\text{M,Norm}} g^{\text{R}}(\mathbf{v}^1)\right] \\
&= \left[\frac{d}{d\mathbf{v}^1} g^{\text{L}}(\mathbf{v}^1)\right] \mathbf{R}^{\text{M,Norm}} g^{\text{R}}(\mathbf{v}^1) + g^{\text{L}}(\mathbf{v}^1)\mathbf{R}^{\text{M,Norm}} \left[\frac{d}{d\mathbf{v}^1} g^{\text{R}}(\mathbf{v}^1)\right],
\end{aligned}$$

where $\mathbf{R}^{\text{M,Norm}}$ is the normalized counterpart to $\mathbf{R}^{\text{M}}$ defined in (**??**). The expressions for

$\frac{d}{d\mathbf{v}^1}g^L(\mathbf{v}^1)$ and $\frac{d}{d\mathbf{v}^1}g^R(\mathbf{v}^1)$ are straightforward to derive using (A.17)–(A.18). Define

$$\begin{aligned}
\mathbf{D}^{\mathrm{L}}(\mathbf{v}^1) &\equiv \frac{d}{d\mathbf{v}^1} g^{\mathrm{L}}(\mathbf{v}^1) \\
&= \frac{d}{d\mathbf{v}^1}\mathrm{diag}\left(\mathbf{R}^{\mathrm{L,Arg}}\ h^1(\mathbf{v}^1)\right)^{-1} \\
&= \Bigg[\underbrace{\Bigg[\frac{d}{dh^1(\mathbf{v}^1)}\mathrm{diag}\left(\mathbf{R}^{\mathrm{L,Arg}}\ h^1(\mathbf{v}^1)\right)^{-1}\Bigg]^{(1,3,2)}}_{\text{Obtained by (A.18)}} \underbrace{\frac{d}{d\mathbf{v}^1}h^1(\mathbf{v}^1)}_{\mathbf{D}^1(\mathbf{v}^1)}\Bigg]^{(1,3,2)},
\end{aligned}$$

where the final line follows from the chain rule. The transposes in the inner bracket align the dimensions corresponding to $h^1(\mathbf{v}^1)$, and the outer bracket reverts the dimensions of the array so the derivatives make up the depth of the array. Similarly, define

$$\begin{aligned}
\mathbf{D}^{\mathrm{R}}(\mathbf{v}^1) &\equiv \frac{d}{d\mathbf{v}^1} g^{\mathrm{R}}(\mathbf{v}^1) \\
&= \frac{d}{d\mathbf{v}^1}\mathrm{diag}\left(\mathbf{R}^{\mathrm{R,Arg,Norm}}\ h^1(\mathbf{v}^1)\right) \\
&= \Bigg[\underbrace{\Bigg[\frac{d}{dh^1(\mathbf{v}^1)}\mathrm{diag}\left(\mathbf{R}^{\mathrm{R,Arg,Norm}}\ h^1(\mathbf{v}^1)\right)\Bigg]^{(1,3,2)}}_{\text{Obtained by (A.17)}} \underbrace{\frac{d}{d\mathbf{v}^1}h^1(\mathbf{v}^1)}_{\mathbf{D}^1(\mathbf{v}^1)}\Bigg]^{(1,3,2)},
\end{aligned}$$

Then the derivative of $g^{\mathrm{Sub},j}(\mathbf{v}^1)$ with respect to $\mathbf{v}^1$ when $j=1$ is

$$\mathbf{D}^{\mathrm{A},1}(\mathbf{v}^1) \equiv \mathbf{D}^{\mathrm{L}}(\mathbf{v}^1)\mathbf{R}^{\mathrm{M,Norm}}g^{\mathrm{R}}(\mathbf{v}^1) + g^{\mathrm{L}}(\mathbf{v}^1)\mathbf{R}^{\mathrm{M,Norm}}\mathbf{D}^{\mathrm{R}}(\mathbf{v}^1). \tag{A.32}$$

Now we turn to (B). Define

$$\mathbf{D}^{\mathrm{B}} \equiv \frac{d}{d\mathbf{v}^1}\mathrm{diag}\left(\mathbf{R}^2\mathbf{v}^1\right)^{-1},$$

which can be obtained from (A.18). Then (B) is equal to

$$\begin{aligned}
&\frac{d}{d\mathbf{v}^1} g^{\mathrm{Sub},1}(\mathbf{v}^1)^\top \mathrm{diag}\left(\mathbf{R}^2\mathbf{v}^1\right)^{-1}\mathbf{v}^2 \\
&= \Bigg[\underbrace{\Bigg[\frac{d}{d\mathbf{v}^1} g^{\mathrm{Sub},1}(\mathbf{v}^1)^\top\Bigg]}_{\mathbf{D}^{\mathrm{A},1\,(2,1,3)}}\mathrm{diag}\left(\mathbf{R}^2\mathbf{v}^1\right)^{-1} + g^{\mathrm{Sub},1}(\mathbf{v}^1)\underbrace{\Bigg[\frac{d}{d\mathbf{v}^1}\mathrm{diag}\left(\mathbf{R}^2\mathbf{v}^1\right)^{-1}\Bigg]}_{\mathbf{D}^{\mathrm{B}}}\Bigg]\mathbf{v}^2 \\
&= \Big[\mathbf{D}^{\mathrm{A},1\,(2,1,3)}\mathrm{diag}\left(\mathbf{R}^2\mathbf{v}^1\right)^{-1} + g^{\mathrm{Sub},1}(\mathbf{v}^1)\mathbf{D}^{\mathrm{B}}\Big]\mathbf{v}^2
\end{aligned}$$

Combining everything, we have

$$
\begin{aligned}
&\frac{d}{d\mathbf{v}^1} h^{\text{Sub},1}(\mathbf{v}^1,\mathbf{v}^2) \\
&= -g^{\text{Inv},1}(\mathbf{v}^1)\left(\mathbf{D}^{\text{A},1}(\mathbf{v}^1)^{(2,1,3)} g^{\text{Sub},j}(\mathbf{v}^1) + g^{\text{Sub},j}(\mathbf{v}^1)^\top \mathbf{D}^{\text{A},1}(\mathbf{v}^1)\right) \\
&\quad\; \times \underbrace{g^{\text{Inv},1}(\mathbf{v}^1) g^{\text{Sub},1}(\mathbf{v}^1)^\top \text{diag}\left(\mathbf{R}^2\mathbf{v}^1\right)^{-1} \mathbf{v}^2}_{h^{\text{Sub},1}(\mathbf{v}^1,\mathbf{v}^1)} \\
&\quad\; + g^{\text{Inv},1}(\mathbf{v}^1)\left[\mathbf{D}^{\text{A},1(2,1,3)} \text{diag}\left(\mathbf{R}^2\mathbf{v}^1\right)^{-1} + g^{\text{Sub},1}(\mathbf{v}^1)\mathbf{D}^{\text{B}}\right]\mathbf{v}^2 \\
&= g^{\text{Inv},1}(\mathbf{v}^1)\Bigg[-\left(\mathbf{D}^{\text{A},1}(\mathbf{v}^1)^{(2,1,3)} g^{\text{Sub},j}(\mathbf{v}^1) + g^{\text{Sub},j}(\mathbf{v}^1)^\top \mathbf{D}^{\text{A},1}(\mathbf{v}^1)\right) h^{\text{Sub},1}(\mathbf{v}^1,\mathbf{v}^2) \\
&\quad\; + \left(\mathbf{D}^{\text{A},1(2,1,3)} \text{diag}\left(\mathbf{R}^2\mathbf{v}^1\right)^{-1} + g^{\text{Sub},1}(\mathbf{v}^1)\mathbf{D}^{\text{B}}\right)\mathbf{v}^2\Bigg].
\end{aligned}
\tag{A.33}
$$

The derivative of $h^{\text{Sub},j}(\mathbf{v}^1,\mathbf{v}^2)$ with respect to $\mathbf{v}^2$ is much simpler, has the same expression for $j = 1, 2$, and is equal to

$$
\begin{aligned}
&\frac{d}{d\mathbf{v}^2} h^{\text{Sub},j}(\mathbf{v}^1,\mathbf{v}^2) \\
&= \frac{d}{d\mathbf{v}^1} \underbrace{g^{\text{Inv},j}(\mathbf{v}^1) g^{\text{Sub},j}(\mathbf{v}^1)^\top \text{diag}\left(\mathbf{R}^2\mathbf{v}^1\right)^{-1}}_{\text{Does not depend on } \mathbf{v}^2} \mathbf{v}^2 \\
&= g^{\text{Inv},j}(\mathbf{v}^1) g^{\text{Sub},j}(\mathbf{v}^1)^\top \text{diag}\left(\mathbf{R}^2\mathbf{v}^1\right)^{-1} \underbrace{\frac{d}{d\mathbf{v}^2}\mathbf{v}^2}_{\mathbf{I}_{|\mathcal{E}||\mathcal{T}||\mathcal{Z}|}} \\
&= g^{\text{Inv},j}(\mathbf{v}^1) g^{\text{Sub},j}(\mathbf{v}^1)^\top \text{diag}\left(\mathbf{R}^2\mathbf{v}^1\right)^{-1}.
\end{aligned}
$$

The expressions derived above for $\frac{d}{d\mathbf{v}^1} h^{\text{Sub},1}(\mathbf{v}^1,\mathbf{v}^2)$, $\frac{d}{d\mathbf{v}^2} h^{\text{Sub},1}(\mathbf{v}^1,\mathbf{v}^2)$ may then be used to construct $\mathbf{D}^{\text{Sub},1}(\mathbf{v}^1,\mathbf{v}^2)$.

To construct $\mathbf{D}^{\text{Sub},2}(\mathbf{v}^1,\mathbf{v}^2)$, we are only missing $\frac{d}{d\mathbf{v}^1} h^{\text{Sub},2}(\mathbf{v}^1,\mathbf{v}^2)$. Much of the derivation can be reused to derive $\frac{d}{d\mathbf{v}^1} h^{\text{Sub},2}(\mathbf{v}^1,\mathbf{v}^2)$. The only difference between $h^{\text{Sub},j}(\mathbf{v}^1,\mathbf{v}^2)$ for $j = 1, 2$ can be seen in (A.21) and (A.23), which shows $g^{\text{Sub},1}(\mathbf{v}^1,\mathbf{v}^2)$ comprises only the left block of $g^{\text{Sub},2}(\mathbf{v}^1,\mathbf{v}^2)$. The right block of $g^{\text{Sub},2}(\mathbf{v}^1,\mathbf{v}^2)$ is a nonrandom matrix and does not depend on $\mathbf{v}^1$, $\mathbf{v}^2$. Thus, to construct $\mathbf{D}^{\text{Sub},2}(\mathbf{v}^1,\mathbf{v}^2)$, we simply need to expand $\mathbf{D}^{\text{A},1}(\mathbf{v}^1,\mathbf{v}^2)$ in (A.32). Define

$$
\mathbf{D}^{\text{A},2}(\mathbf{v}^1,\mathbf{v}^2) \equiv \left[\; \mathbf{D}^{\text{L}}(\mathbf{v}^1)\mathbf{R}^{\text{M,Norm}} g^{\text{R}}(\mathbf{v}^1) + g^{\text{L}}(\mathbf{v}^1)\mathbf{R}^{\text{M,Norm}}\mathbf{D}^{\text{R}}(\mathbf{v}^1) \quad \mathbf{0}_{(|\mathcal{Z}||\mathcal{E}|\bar{t})\times(|\mathcal{S}|\bar{e})\times\bar{g}} \;\right]
$$

to be the derivative of $g^{\text{Sub},2}(\mathbf{v}^1,\mathbf{v}^2)$ with respect to $\mathbf{v}^1$, where $\mathbf{R}^{\text{M,Norm}}$ is the normalized counterpart to $\mathbf{R}^{\text{M}}$ defined in (A.14), and $\mathbf{0}_{(|\mathcal{Z}||\mathcal{E}|\bar{t})\times(|\mathcal{S}|\bar{e})\times\bar{g}}$ reflects that $\mathbf{R}^{\Delta^2}$ does not depend on $\mathbf{v}^1$, $\mathbf{v}^2$. The derivative $\frac{d}{d\mathbf{v}^1}h^{\text{Sub},2}(\mathbf{v}^1,\mathbf{v}^2)$ can then be obtained by replacing $\mathbf{D}^{\text{A},1}(\mathbf{v}^1,\mathbf{v}^2)$ in (A.33) with $\mathbf{D}^{\text{A},2}(\mathbf{v}^1,\mathbf{v}^2)$. ■

A.3 Constructing an invertible system for estimation

In this section, we discuss how to construct an invertible system of equations,

$$\mathbf{A}_n^{\text{Sub},j}\mathbf{x}_n^{\text{Sub},j} = \mathbf{b}^2 \quad \text{for } j = 1,2, \tag{A.34}$$

where $\mathbf{A}_n^{\text{Sub,j}}$ is full column rank and $\mathbf{x}_n^{\text{Sub},j}$ contains all the identified parameters in $\mathbf{x}_n^{\text{All},j}$. It follows that $\mathbf{x}_n^{\text{Sub},j}$ can be solved as

$$\mathbf{x}_n^{\text{Sub},j} = (\mathbf{A}_n^{\text{Sub},j\top}\mathbf{A}_n^{\text{Sub},j})^{-1}\mathbf{A}^{\text{Sub},j\top}\mathbf{b}^2. \tag{A.35}$$

Let $\mathcal{X}^{\text{NID}}$ be the set of parameters in $\mathbf{x}^{\text{All},j}$ for $j = 1,2$ that are unidentified. For brevity, $\mathcal{X}^{\text{NID}}$ contains the set of unidentified parameters for either framework for treatment effect heterogeneity. $\mathcal{X}^{\text{NID}}$ can be divided into four categories,

$$\mathcal{X}^{\text{NID}} \equiv \mathcal{X}^{\text{NID},1} \cup \mathcal{X}^{\text{NID},2} \cup \mathcal{X}^{\text{NID},3} \cup \mathcal{X}^{\text{NID},4}.$$

We now define each category.

The first category $\mathcal{X}^{\text{NID},1}$ arises when treatment effects are heterogeneous across agent types and there exists $g \in \mathcal{G}$ such that $\check{e}(g,z) = \infty$ for all $z \in \mathcal{Z}$. These agent types never takes up treatment, meaning $\Delta^1(g,s)$ for $s \in \mathcal{S}$ are never realized in the panel and thus unidentified.

The second category $\mathcal{X}^{\text{NID},2}$ is treatment effects sufficiently long after taking up treatment. These parameters arises under either framework for heterogeneity. When treatment effects are heterogeneous across agent types, the number of these types of parameters varies across agent types. For instance, suppose agent type g_1 takes up treatment in period 1 (the first period treatment is available) for some value of z. Then, conditional on that value of z, the panel will contain information $\Delta^1(g_1,s)$ for $s \in \mathcal{S}$. In contrast, suppose the earliest period type g_2 takes up treatment is the final period treatment is available, $\bar{e}$. Then the panel contains information on $\Delta^1(g_2,0)$, but contains no information on $\Delta^1(g,s)$ for $s > 0$ since the panel does not extend beyond the first period g_2 takes up treatment. More generally, when treatment effects are heterogeneous across types, for any $g \in \mathcal{G}$,

$$\Delta^1(g,s) \text{ for } s > \bar{t} - \min_{z\in\mathcal{Z}} \check{e}(z,g)$$

are not realized in the panel and thus unidentified. Similarly, when treatment effects are heterogeneous across cohorts,

$$\Delta^2(e, s) \text{ for } s > \bar{t} - e$$

are unidentified.

The third category $\mathcal{X}^{\text{NID},3}$ is treatment effects for noncompliers and only arises when treatment effects are heterogeneous across agent types. Under this setting, identification of treatment effects requires separating $\Delta^1(g, s)$ from $\mathbb{E}[Y_{i,e+s} \mid G_i = g]$ for $g \in \mathcal{G}$. However, this is not possible for noncompliers since, for any period in the panel, there are no two values of the instrument such that the agent type is treated under one value of the instrument and untreated under the other.[34] Thus, $\Delta(g, s)$ for $s \geq 0$ is unidentified for any noncomplier type $g \in \mathcal{G}$. Let $\mathcal{X}^{\text{NID},3}$ denote these parameters.[35]

The final category $\mathcal{X}^{\text{NID},4}$ is simply the remaining unidentified parameters in $\mathbf{x}^{\text{All},j}$ that are not in $\mathcal{X}^{\text{NID},1}$, $\mathcal{X}^{\text{NID},2}$, $\mathcal{X}^{\text{NID},3}$.

The invertible system in (A.34) is constructed simply by removing a subset of parameters from $\mathcal{X}^{\text{NID}}$ from the systems in Propositions 2 and 3. We now discuss which parameters to remove.

The parameters in $\mathcal{X}^{\text{NID},1}$ and $\mathcal{X}^{\text{NID},2}$ are never realized in the panel, so their corresponding columns in $\mathbf{A}^{\text{Al},j}$ are vectors of 0s. We thus remove their columns from $\mathbf{A}^{\text{All},j}$ and their entries from $\mathbf{x}^{\text{All},j}$.

The parameters in $\mathcal{X}^{\text{NID},3}$ may also be removed from the system as they cannot be separately identified from the concomitant untreated outcomes. However, the remaining entries in $\mathbf{x}^{\text{All},j}$ that previously corresponded to the concomitant untreated outcomes are now equal to $\mathbb{E}[Y_{i,e+s} \mid G_i = g]$. That is, $\mathbb{E}[Y_{i,e+s}(\infty) \mid G_i = g]$ and $\Delta^1(g, s)$ are jointly identified at best.

Only a subset of parameters in $\mathcal{X}^{\text{NID},4}$ should be removed from the system. We determine which parameters to remove by sequentially checking whether each parameter in $\mathcal{X}^{\text{NID},4}$ is uniquely determined by the system. In practice, we do this using the reduced row echelon form (RREF) of $\mathbf{A}^{\text{Sub},j}$). If the parameter is uniquely determined, then the parameter

[34] For any noncomplier type g, $\check{e}(z, g) = e$ for all $z \in \mathcal{Z}$ and $e \in \mathcal{E}$. Thus, for $s \geq 0$, any row in (binary matrix) $\mathbf{R}^{\text{M}}$ containing an entry for $\Delta^1(g, s)$ also contains an entry for $\mathbb{E}[Y_{i,e+s}(\infty) \mid G_i = g]$. Likewise, any row in $\mathbf{R}^{\text{M}}$ containing an entry for $\mathbb{E}[Y_{i,e+s}(\infty) \mid G_i = g]$ also contains an entry for $\Delta^1(g, s)$. It is thus impossible to separate the treatment effect $\Delta(g, s)$ from its unique concomitant untreated outcome $\mathbb{E}[Y_{i,e+s}(\infty) \mid G_i = g]$.

[35] While it is not possible to recover the treatment effects for noncomplier types, it may be possible to recover $\mathbb{E}[Y_{i,t} \mid G_i = g]$ for the noncomplier types in the periods before they take up treatment.

stays in the system. Otherwise, it is dropped. The order in which the parameters are checked is not important.[36] It follows that all remaining parameters in the system are uniquely determined and the system is invertible.

Algorithm 1 summarizes how to construct $\mathbf{A}^{\text{Sub},j}$ and $\mathbf{x}^{\text{Sub},j}$ for $j = 1, 2$.

Algorithm 1 Constructing $\mathbf{A}^{\text{Sub},j}$, $\mathbf{x}^{\text{Sub},j}$ in (A.34) for $j = 1, 2$

$\mathbf{A}^{\text{Sub},j} \leftarrow \mathbf{A}^{\text{All},j}$ without columns corresponding to $\mathcal{X}^{\text{NID},1} \cup \mathcal{X}^{\text{NID},2} \cup \mathcal{X}^{\text{NID},3}$
$\mathbf{x}^{\text{Sub},j} \leftarrow \mathbf{x}^{\text{All},j}$ without entries corresponding to $\mathcal{X}^{\text{NID},1} \cup \mathcal{X}^{\text{NID},2} \cup \mathcal{X}^{\text{NID},3}$
for $x \in \mathcal{X}^{\text{NID},4}$ **do**
    Check if x is uniquely determined using RREF of $\mathbf{A}^{\text{Sub},j}$
    if x is not identified **then**
        $\mathbf{A}^{\text{Sub},j} \leftarrow \mathbf{A}^{\text{Sub},j}$ after removing column corresponding to x
        $\mathbf{x}^{\text{Sub},j} \leftarrow \mathbf{x}^{\text{Sub},j}$ after removing entries corresponding to x
    end if
end for

Remark A.1. The procedure in Algorithm 1 does not allow us to identify more parameters. It simply allows us to construct the invertible system (A.34) and obtain a closed form solution for the identified parameters. ■

[36] The parameters in $\mathcal{X}^{\text{ID}}$ are uniquely determined regardless of the order this final step is performed. However, which parameters in $\mathcal{X}^{\text{NID},3}$ and $\mathcal{X}^{\text{NID},4}$ that remain in the system—and thus the values of $\mathbf{x}$ for parameters $\mathcal{X}^{\text{NID},3}$ and $\mathcal{X}^{\text{NID},4}$—may change with the order in which this final step is performed.