



Optimal Dynamic Treatment Regimes and Partial Welfare Ordering

Sukjin Han

To cite this article: Sukjin Han (2023): Optimal Dynamic Treatment Regimes and Partial Welfare Ordering, Journal of the American Statistical Association, DOI: 10.1080/01621459.2023.2238941

To link to this article: <https://doi.org/10.1080/01621459.2023.2238941>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 06 Sep 2023.



[Submit your article to this journal](#)



[View related articles](#)



[View Crossmark data](#)

Optimal Dynamic Treatment Regimes and Partial Welfare Ordering

Sukjin Han

School of Economics, University of Bristol, Bristol, UK

ABSTRACT

Dynamic treatment regimes are treatment allocations tailored to heterogeneous individuals (e.g., via previous outcomes and covariates). The optimal dynamic treatment regime is a regime that maximizes counterfactual welfare. We introduce a framework in which we can partially learn the optimal dynamic regime from observational data, relaxing the sequential randomization assumption commonly employed in the literature but instead using (binary) instrumental variables. We propose the notion of sharp partial ordering of counterfactual welfares with respect to dynamic regimes and establish mapping from data to partial ordering via a set of linear programs. We then characterize the identified set of the optimal regime as the set of maximal elements associated with the partial ordering. We relate the notion of partial ordering with a more conventional notion of partial identification using topological sorts. Practically, topological sorts can be served as a policy benchmark for a policymaker. We apply our method to understand returns to schooling and post-school training as a sequence of treatments by combining data from multiple sources. The framework of this article can be used beyond the current context, for example, in establishing rankings of multiple treatments or policies across different counterfactual scenarios. Supplementary materials for this article are available online.

ARTICLE HISTORY

Received July 2021
Accepted July 2023

KEYWORDS

Dynamic treatment effect;
Endogenous treatments;
Instrumental variable; Linear
programming; Optimal
dynamic treatment regimes;
Partial identification

1. Introduction

Dynamic treatment regimes are dynamically personalized treatment allocations. Given that individuals are heterogeneous, allocations tailored to heterogeneity can improve overall welfare. Define a dynamic treatment regime $\delta(\cdot)$ as a sequence of binary rules $\delta_t(\cdot)$ that map the previous outcome and treatment (and possibly other covariates) onto current allocation decisions: $\delta_t(y_{t-1}, d_{t-1}) = d_t \in \{0, 1\}$ for $t = 1, \dots, T$. The motivation for being adaptive to the previous outcome is that it may contain information on unobserved heterogeneity that is not captured in covariates. Then the optimal dynamic treatment regime, which is this article's main parameter of interest, is defined as a regime that maximizes certain counterfactual welfare:

$$\delta^*(\cdot) = \arg \max_{\delta(\cdot)} W_{\delta}.$$

This article investigates the possibility of identifiability of the optimal dynamic regime $\delta^*(\cdot)$ from data that are generated from randomized experiments in the presence of noncompliance or more generally from observational studies in multi-period settings.

Optimal treatment regimes have been extensively studied in the biostatistics literature (Murphy et al. 2001; Murphy 2003; Robins 2004, among others). These studies typically rely on an ideal multi-stage experimental environment that satisfies sequential randomization. Based on such experimental data, they identify optimal regimes that maximize welfare, defined as the average counterfactual outcome. However, noncompliance is prevalent in experiments, and more generally, treatment

endogeneity is a marked feature in observational studies. This may be one reason the vast biostatistics literature has not yet gained traction in other fields of social science, despite the potentially fruitful applications of optimal dynamic regimes in various policy evaluations.

This article proposes a nonparametric framework, in which we can at least partially learn the ranking of counterfactual welfares W_{δ} 's and hence the optimal dynamic regime $\delta^*(\cdot)$. We view that it is important to avoid making stringent modeling assumptions in the analysis of personalized treatments, because the core motivation of the analysis is individual heterogeneity, which we want to keep intact as much as possible. Instead, we embrace the partial identification approach. Given the observed distribution of sequences of outcomes and endogenous treatments and using the instrumental variable (IV) method, we establish sharp partial ordering of welfares, and characterize the identified set of optimal regimes as a discrete subset of all possible regimes. We define welfare as a linear functional of the joint distribution of counterfactual outcomes across periods. Examples of welfare include the average counterfactual terminal (i.e., distal) outcome commonly considered in the literature and as shown above. We assume we are equipped with some IVs that are possibly binary. We show that it is helpful to have a sequence of IVs generated from sequential experiments or quasi-experiments. Examples of the former are increasingly common as forms of random assignments or encouragements in medical trials, public health and educational interventions, and A/B testing on digital platforms. Examples of the latter can be some combinations of traditional IVs and regression discontinuity designs.

CONTACT Sukjin Han  sukjin.han@gmail.com  School of Economics, University of Bristol, Bristol, UK

 Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Our framework also accommodates a single binary IV in the context of dynamic treatments and outcomes (e.g., Cellini et al. 2010). The identifying power in such a case is investigated in simulation. The partial ordering and identified set proposed in this article enable “sensitivity analyses.” That is, by comparing a chosen regime (e.g., from a parametric approach) with these benchmark objects, one can determine how much the former is led by assumptions and how much is informed by data. Such a practice also allows us to gain insight into data requirements to achieve a certain level of informativeness.

The identification analysis is 2-fold. In the first part, we establish mapping from data to sharp partial ordering of counterfactual welfares with respect to possible regimes. The point identification of $\delta^*(\cdot)$ will be achieved by establishing the total ordering of welfares, which is not generally possible in this flexible nonparametric framework with limited exogenous variation. To establish the partial ordering, we first characterize bounds on the difference between each pair of welfares as the set of optima of linear programs, and we do so for all possible welfare pairs. The bounds on welfare gaps are informative about whether welfares are comparable or not, and when they are, how to rank them. Then we show that although the bounds are calculated from separate optimizations, the partial ordering is consistent with *common* data-generating processes. The partial ordering obtained in this way is shown to be sharp in the sense that will become clear later. Note that each welfare gap measures the *dynamic treatment effect*. The partial ordering concisely (and tightly) summarizes the identified signs of these treatment effects, and thus can be a parameter of independent interest.

In the second part of the analysis, given the sharp partial ordering, we show that the identified set can be characterized as the set of maximal elements associated with the partial ordering, that is, the set of regimes that are *not inferior*. Given the partial ordering, we also calculate topological sorts, which are total orderings that do not violate the underlying partial ordering. Theoretically, topological sorts can be viewed as observationally equivalent total orderings, which insight relates the partial ordering we consider with a more conventional notion of partial identification. Practically, topological sorts can be served as a policy benchmark that a policymaker can be equipped with. If desired, linear programming can be solved to calculate bounds on a small number of sorted welfares (e.g., top-tier welfares).

Given the minimal structure we impose in the data-generating process, the size of the identified set may be large in some cases. Such an identified set may still be useful in eliminating suboptimal regimes or warning about the lack of informativeness of the data. Often, however, researchers are willing to impose additional assumptions to gain identifying power. We propose identifying assumptions, such as uniformity assumptions that generalize the monotonicity assumption in Imbens and Angrist (1994) and Angrist et al. (1996), Markovian structure, and stationarity. These assumptions tighten the identified set by reducing the dimension of the simplex in the linear programming, thus, producing a denser partial ordering. We show that these assumptions are easy to impose in our framework.

This article makes several contributions. To our best knowledge, this article is first in the literature that considers the

identifiability of optimal *dynamically adaptive* regimes under treatment endogeneity. Murphy (2003) and subsequent works consider point identification of optimal dynamic regimes, but under the sequential randomization assumption. This article brings that literature to observational contexts. Recently, Han (2021b), Han (2021a), Cui and Tchetgen Tchetgen (2021), and Qiu et al. (2021) relax sequential randomization and establish identification of dynamic average treatment effects and/or optimal regimes using IVs. They consider a regime that is a mapping only from covariates, but not previous outcomes and treatments, to an allocation. They focus on point identification by imposing assumptions such as the existence of additional exogenous variables in a multi-period setup (Han (2021b)), or the zero correlation between unmeasured confounders and compliance types (Cui and Tchetgen Tchetgen 2021; Qiu et al. 2021) or uniformity (Han 2021a) in a single-period setup. The dynamic effects of treatment timing (i.e., irreversible treatments) have been considered in Heckman and Navarro (2007) and Heckman et al. (2016) who use exclusion restrictions and infinite support assumptions. A related staggered adoption design was recently studied in multi-period difference-in-differences settings under treatment heterogeneity by Athey and Imbens (2022), Callaway and Sant’Anna (2021), and Sun and Abraham (2021). de Chaisemartin and d’Haultfoeuille (2020) consider a similar problem but without necessarily assuming staggered adoption. This article complements these papers by considering treatment scenarios of multiple dimensions with adaptivity as the key ingredient.

Second, this article contributes to the literature on partial identification of treatment effects that uses linear programming approach, which has early examples as Balke and Pearl (1997) and Manski (2007), and appears recently in Mogstad et al. (2018), Torgovitsky (2019), Machado et al. (2019), Kamat (2019), and Han and Yang (2023) to name a few. The advantages of this approach is that (i) bounds can be automatically obtained even when analytical derivation is not possible, (ii) the proof of sharpness is straightforward and not case-by-case, and (iii) it can streamline the analysis of different identifying assumptions. The dynamic framework of this article complicates the identification analysis and thus fully benefits from these advantages. However, a distinct feature of the present article is that the linear programming approach is used in establishing a sharp partial ordering across counterfactual objects—a novel concept in the literature—and in such a way that separate optimizations yield a common object, namely the partial ordering. The framework of this article can also be useful in other settings where the goal is to compare welfares across multiple treatments and regimes—for example, personalized treatment rules—or more generally, to establish rankings of policies across different counterfactual scenarios and find the best ones.

Third, we apply our method to conduct a policy analysis with schooling and post-school training as a sequence of treatments, which is to our knowledge a novel attempt in the literature. We consider dynamic treatment regimes of allocating a high

school diploma and, given pre-program earnings, a job training program for economically disadvantaged population. By combining data from the Job Training Partnership Act (JTPA), the US Census, and the National Center for Education Statistics (NCES), we construct a dataset with a sequence of IVs that is used to estimate the partial ordering of expected earnings and the identified set of the optimal regime. Even though only partial orderings are recovered, we can conclude with certainty that allocating the job training program only to the low earning type is welfare optimal. We also find that more costly regimes are not necessarily welfare-improving.

The dynamic treatment regime considered in this article is broadly related to the literature on statistical treatment rules, for example, Manski (2004), Hirano and Porter (2009), Bhattacharya and Dupas (2012), Stoye (2012), Kitagawa and Tetenov (2018), Kasy (2016), and Athey and Wager (2021). However, our setting, assumptions, and goals are different from those in these papers. In a single-period setting, they consider allocation rules that map covariates to decisions. They impose assumptions that ensure point identification, such as (conditional) unconfoundedness or homogeneity, and focus on establishing the asymptotic optimality of the treatment rules, with Kasy (2016) the exception. Kasy (2016) focuses on establishing partial ranking by comparing a pair of treatment-allocating probabilities as policies. The notion of partial identification of ranking relates to ours, but we introduce the notion of sharpness of a partially ordered set with discrete policies and a linear programming approach to achieve that. Another distinction is that we consider a dynamic setup. In the sense of constructing a set of optimal dynamic treatment regimes, the current article also relates to the approach in biostatistics, most notably in Ertefaie et al. (2016) and Chao et al. (2022). However, the fundamental difference is that, in the latter approach, the set consists of regimes that cannot be differentiated from the best regime due to sampling uncertainty (i.e., the set is a confidence set) while, in our approach, it results from model uncertainty (i.e., the set is an identified set). Finally, in order to focus on the challenge with endogeneity, we consider a simple setup where the exploration and exploitation stages are separated, unlike in the literature on bandit problems (Athey and Imbens 2019; Kasy and Sautmann 2021; Kock et al. 2021). We believe the current setup is a good starting point.

In the next section, we introduce the dynamic regimes and related counterfactual outcomes, which define the welfare and the optimal regime. Section 3 conducts the main identification analysis by constructing the partial ordering and characterizing the identified set. Section 4 presents the empirical application on returns to schooling and job training. In the online supplemental appendix, the analysis with binary outcomes and discrete covariates is extended to continuous outcomes and covariates, and stochastic regimes are discussed. The supplemental appendix also presents numerical studies and discusses topological sorts, cardinality reduction for the set of regimes, and inference. Most proofs are collected in the supplemental appendix.

2. Dynamic Regimes and Counterfactual Welfares

2.1. Dynamic Regimes

Let t be the index for a period or stage. For each $t = 2, \dots, T$ with fixed T , define an *adaptive treatment rule* $\delta_t : \{0, 1\}^{t-1} \times$

Table 1. Dynamic regimes $\delta(\cdot)$ when $T = 2$ and $\delta_1(x) = \delta_1$.

Regime #	δ_1	$\delta_2(1, \delta_1)$	$\delta_2(0, \delta_1)$
1	0	0	0
2	1	0	0
3	0	1	0
4	1	1	0
5	0	0	1
6	1	0	1
7	0	1	1
8	1	1	1

$\{0, 1\}^{t-1} \rightarrow \{0, 1\}$ that maps the lags of the realized binary outcomes and treatments $\mathbf{y}^{t-1} \equiv (y_1, \dots, y_{t-1})$ and $\mathbf{d}^{t-1} \equiv (d_1, \dots, d_{t-1})$ onto a deterministic treatment allocation $d_t \in \{0, 1\}$:

$$\delta_t(\mathbf{y}^{t-1}, \mathbf{d}^{t-1}) = d_t. \tag{2.1}$$

This adaptive rule also appears in, for example, Murphy (2003). When $t = 1$, the adaptive rule $\delta_1 : \mathcal{X} \rightarrow \{0, 1\}$ maps discrete pre-treatment covariate vector x onto an allocation $d_1 \in \{0, 1\}$:

$$\delta_1(x) = d_1. \tag{2.2}$$

The treatment rules above are dynamic in the sense that it is a function of previous outcomes, treatments and covariates. Special cases of (2.1)–(2.2) are a dynamic rule that is only a function of covariates but not $(\mathbf{y}^{t-1}, \mathbf{d}^{t-1})$ (Han 2021b; Cui and Tchetgen Tchetgen 2021) and a static rule where $\delta_t(\cdot)$ is a constant function. Binary outcomes and treatments are prevalent, and they are helpful in analyzing, interpreting, and implementing dynamic regimes (Zhang et al. 2015). Later, we extend the framework to allow for continuous outcome variables and covariates and time-varying covariates; see Appendices A.1 and A.2. We only consider deterministic rules $\delta_t(\cdot) \in \{0, 1\}$. In Appendix A.3, we extend this to stochastic rules and show why it is enough to consider deterministic rules in some cases. Then, a *dynamic regime* up to period t is defined as a vector of all treatment rules:

$$\boldsymbol{\delta}^t(\cdot) \equiv (\delta_1(\cdot), \delta_2(\cdot), \dots, \delta_t(\cdot)).$$

Let $\boldsymbol{\delta}(\cdot) \equiv \boldsymbol{\delta}^T(\cdot) \in \mathcal{D}$ where \mathcal{D} is the set of all possible regimes. We can allow \mathcal{D} to be a strict subset of the set of all possible regimes due to institutional or practical purposes; see Appendix F.4 for relevant discussions. Throughout the article, we will mostly focus on the leading case with $T = 2$ for simplicity. Also, this case already captures the essence of the dynamic features, such as adaptivity and complementarity. Table 1 lists all possible dynamic regimes $\boldsymbol{\delta}(\cdot) \equiv (\delta_1, \delta_2(\cdot))$ (with constant function $\delta_1(x) = \delta_1$) as contingency plans, and there are eight of them. When $\delta_1(x)$ is a function of binary $x \in \{0, 1\}$, it is easy to see that there will be 16 regimes in total.

2.2. Counterfactual Welfares and Optimal Regimes

To define welfare with respect to (w.r.t.) this dynamic regime, we first introduce a counterfactual outcome as a function of a dynamic regime. Because of the adaptivity intrinsic in dynamic regimes, expressing counterfactual outcomes is more involved than that with static regimes d_t , that is, $Y_t(\mathbf{d}^t)$ with $\mathbf{d}^t \equiv (d_1, \dots, d_t)$. Let $\mathbf{Y}^t(\mathbf{d}^t) \equiv (Y_1(d_1), Y_2(d_2), \dots, Y_t(\mathbf{d}^t))$. In terms

of notation throughout the article, for an arbitrary r.v. R_t , we let $\mathbf{R}^t \equiv (R_1, \dots, R_t)$ denote a vector that collects R_t across time up to t , and let \mathbf{r}^t be its realization. Most of the time, we write $\mathbf{R} \equiv \mathbf{R}^T$ for convenience. We express a counterfactual outcome with adaptive regime $\delta^t(\cdot)$ and covariate values x as follows:

$$Y_t(\delta^t(\cdot)) \equiv Y_t(\mathbf{d}^t), \quad (2.3)$$

where the “bridge variables” $\mathbf{d}^t \equiv (d_1, \dots, d_t)$ satisfy

$$\begin{aligned} d_1 &= \delta_1(x), \\ d_2 &= \delta_2(Y_1(d_1), d_1), \\ d_3 &= \delta_3(Y^2(\mathbf{d}^2), \mathbf{d}^2), \\ &\vdots \\ d_t &= \delta_t(Y^{t-1}(\mathbf{d}^{t-1}), \mathbf{d}^{t-1}). \end{aligned} \quad (2.4)$$

Suppose $T = 2$. Then, the two counterfactual outcomes are defined as $Y_1(\delta_1(\cdot)) = Y_1(\delta_1(x))$ and $Y_2(\delta^2(\cdot)) = Y_2(\delta_1, \delta_2(Y_1(\delta_1), \delta_1))$. As the notation suggests, we implicitly assume the “no anticipation” condition.

Let $q_\delta(\mathbf{y}) \equiv \Pr[\mathbf{Y}(\delta(\cdot)) = \mathbf{y}]$ be the joint distribution of counterfactual outcome vector $\mathbf{Y}(\delta(\cdot)) \equiv (Y_1(\delta_1(\cdot)), Y_2(\delta^2(\cdot)), \dots, Y_T(\delta^T(\cdot)))$. We define counterfactual welfare as a linear functional of $q_\delta(\mathbf{y})$:

$$W_\delta \equiv f(q_\delta). \quad (2.5)$$

Examples of the functional include the average counterfactual terminal outcome $E[Y_T(\delta(\cdot))] = \Pr[Y_T(\delta(\cdot)) = 1]$, our leading case and which is common in the literature, and the weighted average of counterfactuals $\sum_{t=1}^T \omega_t E[Y_t(\delta^t(\cdot))]$. Then, the *optimal dynamic regime* is a regime that maximizes the welfare:

$$\delta^*(\cdot) = \arg \max_{\delta(\cdot) \in \mathcal{D}} W_\delta. \quad (2.6)$$

We assume that the optimal dynamic regime is unique by simply ruling out a knife-edge case in which two regimes deliver the same welfare. In the case of $W_\delta = E[Y_T(\delta(\cdot))]$, the solution $\delta^*(\cdot)$ can be justified by backward induction in finite-horizon dynamic programming. Moreover in this case, the regime with deterministic rules $\delta_t(\cdot) \in \{0, 1\}$ achieves the same optimal regime and optimized welfare as the regime with stochastic rules $\delta_t(\cdot) \in [0, 1]$; see Theorem A.1 in Appendix A.3.

The identification analysis of the optimal regime is closely related to the identification of welfare for each regime and welfare gaps, which also contain information for policy. Some interesting special cases are the following: (i) the *optimal welfare*, W_{δ^*} , which in turn yields (ii) the *regret* from following individual decisions, $W_{\delta^*} - W_{\mathbf{D}}$, where $W_{\mathbf{D}}$ is simply $f(\Pr[\mathbf{Y}(\mathbf{D}) = \cdot]) = f(\Pr[\mathbf{Y} = \cdot])$, and (iii) the *gain from adaptivity*, $W_{\delta^*} - W_{\mathbf{d}^*}$, where $W_{\mathbf{d}^*} = \max_{\mathbf{d}} W_{\mathbf{d}}$ is the optimum of the welfare with a static rule, $W_{\mathbf{d}} = f(\Pr[\mathbf{Y}(\mathbf{d}) = \cdot])$. If the cost of treatments is not considered, the gain in (iii) is nonnegative as the set of all \mathbf{d} is a subset of \mathcal{D} .

To illustrate the policy relevance of the optimal dynamic regime, consider an example of the labor market returns to high school education and post-school training for disadvantaged individuals. First, consider a static regime, which is a schedule $\mathbf{d} = (d_1, d_2) \in \{0, 1\}^2$ of first assigning a high school

diploma ($d_1 \in \{0, 1\}$) and then a job training ($d_2 \in \{0, 1\}$). Define associated welfare, which is the employment rate $W_{\mathbf{d}} = E[Y_2(\mathbf{d})]$ where Y_2 is an indicator of employment status with value 1 if being employed. This setup is already useful in learning, for example, $E[Y_2(1, 0)] - E[Y_2(0, 1)]$ or complementarity (i.e., $E[Y_2(0, 1)] - E[Y_2(0, 0)]$ versus $E[Y_2(1, 1)] - E[Y_2(1, 0)]$), which cannot be learned from period-specific treatment effects. However, because d_1 and d_2 are not simultaneously given but d_1 precedes d_2 , the allocation d_2 can be more informed by incorporating the knowledge about the individual’s response to d_1 . An example of such a response to d_1 would be employment status y_1 after high school and before the training program. This motivates the dynamic regime, which is the schedule $\delta(\cdot) = (\delta_1(\cdot), \delta_2(\cdot)) \in \mathcal{D}$ of allocation *rules* that first assigns a high school diploma ($\delta_1(x) \in \{0, 1\}$) depending on individual characteristics x and then assigns a job training ($\delta_2(y_1, \delta_1) \in \{0, 1\}$) depending on δ_1 and the employment status y_1 . Then, the optimal regime with adaptivity $\delta^*(\cdot)$ is the one that maximizes $W_\delta = E[Y_2(\delta)]$. Suppose the optimal regime $\delta^*(\cdot)$ is such that $\delta_1^* = 1$, $\delta_2^*(0, \delta_1^*) = 1$, and $\delta_2^*(1, \delta_1^*) = 0$; that is, it turns out optimal to assign a high school diploma to all individuals and a training program to unemployed individuals. One of the policy implications of such $\delta^*(\cdot)$ is that the average job market performance can be improved by job trainings focusing on low performance individuals complementing with high school education. As a static regime—where $\delta_t(\cdot)$ is a constant function—is a special case of a dynamic regime, the optimal dynamic regime provides richer policy candidates than what can be learned from the optimal static regime \mathbf{d}^* . In this sense, the optimal dynamic regime provides richer policy candidates than what can be learned from dynamic complementarity (Cunha and Heckman 2007; Cellini et al. 2010; Almond and Mazumder 2013; Johnson and Jackson 2019).

3. Partial Ordering and Partial Identification

3.1. Observables

We introduce observables based on which we want to identify the optimal regime and counterfactual welfares. Assume that the time length of the observables is equal to T , the length of the optimal regime to be identified; in general, we may allow $\tilde{T} \geq T$ where \tilde{T} is the length of the observables. For each period or stage $t = 1, \dots, T$, assume that we observe the binary instrument Z_t , the binary endogenous treatment decision D_t , and the binary outcome $Y_t = \sum_{\mathbf{d}^t \in \{0, 1\}^t} 1\{\mathbf{D}^t = \mathbf{d}^t\} Y_t(\mathbf{d}^t)$. Also, we observe discrete pre-treatment covariates X that are potentially endogenous. As an example, Y_t is a symptom indicator for a patient, D_t is the medical treatment received, and Z_t is generated by a multi-period medical trial. Importantly, the framework does not preclude the case in which Z_t exists only for some t but not all; see Appendix E for related discussions. In this case, Z_t for the other periods is understood to be degenerate. Let $D_t(\mathbf{z}^t)$ be the counterfactual treatment given $\mathbf{z}^t \equiv (z_1, \dots, z_t) \in \{0, 1\}^t$. Then, $D_t = \sum_{\mathbf{z}^t \in \mathcal{Z}^t} D_t(\mathbf{z}^t)$. Let $\mathbf{Z} \equiv (Z_1, \dots, Z_T)$, $\mathbf{Y}(\mathbf{d}) \equiv (Y_1(d_1), Y_2(\mathbf{d}^2), \dots, Y_T(\mathbf{d}))$, and $\mathbf{D}(\mathbf{z}) \equiv (D_1(z_1), D_2(\mathbf{z}^2), \dots, D_T(\mathbf{z}))$ and let “ \perp ” denote statistical independence.

Assumption SX. $\mathbf{Z} \perp (\mathbf{Y}(\mathbf{d}), \mathbf{D}(\mathbf{z})) | X$.

Assumption SX assumes the strict exogeneity and exclusion restriction. A single IV with conditional independence trivially satisfies this assumption. For a sequence of IVs, this assumption is satisfied in typical sequential randomized experiments, as well as quasi-experiments. Returning to our illustrative example, let $D_{i1} = 1$ if student i has a high school diploma and $D_{i1} = 0$ otherwise; let $D_{i2} = 1$ if i participates in a job training program and $D_{i2} = 0$ if not. Also, let $Y_{i1} = 1$ if i is employed before the training program and $Y_{i1} = 0$ if not; let $Y_{i2} = 1$ if i is employed after the program and $Y_{i2} = 0$ if not. Finally, let X_i be i 's observable characteristics. Given the observational data, suppose we are interested in recovering regimes that maximize the employment rate as welfare. As D_1 and D_2 are endogenous, $\{D_{i1}, Y_{i1}, D_{i2}, Y_{i2}\}$ are not useful by themselves to identify W_δ 's and $\delta^*(\cdot)$. Therefore, we employ the approach of using IVs, either a single IV (e.g., in the initial period) or a sequence of IVs. In this particular example, we can use the distance to high schools or the number of high schools per square mile as an instrument Z_1 for D_1 conditional on X . Then, a random assignment of the job training in a field experiment can be used as an instrument Z_2 for the compliance decision D_2 . Assumption SX requires that conditional on individual characteristics, these instruments are jointly independent of the unobserved confounders (e.g., ability, personality) that are present in the outcome formation and treatment selection processes. In Section 4, we empirically study schooling and job training as a sequence of treatments and combine IVs from experimental and observational data. In observational settings as this example, one can use IVs from quasi-experiments, those from RD design, or a combination of them. In experimental settings, examples of a sequence of IVs can be found in multi-stage experiments, such as the Fast Track Prevention Program (Conduct Problems Prevention Research Group 1992), the Elderly Program randomized trial for the Systolic Hypertension (The Systolic Hypertension in the Elderly Program (SHEP) Cooperative Research Group 1988), and Promotion of Breastfeeding Intervention Trial (Kramer et al. 2001). It is also possible to combine multiple experiments as in Johnson and Jackson (2019).

Let (Y, D, Z, X) be the vector of observables (Y_t, D_t, Z_t, X_t) for the entire T periods and X , and let p be its distribution. We assume that (Y_i, D_i, Z_i, X_i) is independent and identically distributed and $\{(Y_i, D_i, Z_i) : i = 1, \dots, N\}$ is a small T large N panel. We mostly suppress the individual unit i throughout the article. For empirical applications, the data structure can be more general than a panel and the kinds of $Y_t, D_t,$ and Z_t are allowed to be different across time; recall the above illustrative example. For the population from which the data are drawn, we are interested in learning the optimal regime and related welfares.

3.2. Partial Ordering of Welfares

Given the distribution p of the data (Y, D, Z, X) and under Assumption SX, we show how the optimal dynamic regime and welfares can be partially recovered. The identified set of $\delta^*(\cdot)$ will be characterized as a subset of the discrete set \mathcal{D} . As the first step, we establish *partial ordering* of W_δ w.r.t. $\delta(\cdot) \in \mathcal{D}$ as a function of p . The partial ordering can be represented

by a *directed acyclic graph* (DAG).¹ The DAG summarizes the identified signs of the dynamic treatment effects, as will become clear later. Moreover, the DAG representation is fruitful for introducing the notion of the sharpness of partial ordering and later to translate it into the identified set of $\delta^*(\cdot)$.

To facilitate this analysis, we enumerate all $|\mathcal{D}| = 2^{2^T - 2} \times 2^{|\mathcal{X}|}$ possible regimes. For index $k \in \mathcal{K} \equiv \{k : 1 \leq k \leq |\mathcal{D}|\}$ (and thus $|\mathcal{K}| = |\mathcal{D}|$), let $\delta_k(\cdot)$ denote the k th regime in \mathcal{D} . For $T = 2$ and $\delta_1(x) = \delta_1$, Table 1 indexes all possible dynamic regimes $\delta(\cdot) \equiv (\delta_1, \delta_2(\cdot))$. Let $W_k \equiv W_{\delta_k}$ be the corresponding welfare. Figure 1 illustrates examples of the partially ordered set of welfares as DAGs where each edge " $W_k \rightarrow W_{k'}$ " indicates the relation " $W_k > W_{k'}$ ".

In general, the point identification of $\delta^*(\cdot)$ is achieved by establishing the total ordering of W_k . Without strong additional assumptions, this is only possible if instruments has infinite support. Since we allow for instruments with minimal variation (i.e., binary instruments), we may only recover a partial ordering. We want the partial ordering to be sharp in the sense that it cannot be improved given the data and maintained assumptions. To formally state this, let $G(\mathcal{K}, \mathcal{E})$ be a DAG where \mathcal{K} is the set of welfare (or regime) indices and \mathcal{E} is the set of edges. For example, in Figure 1(a), we have $\mathcal{E} = \{(W_1, W_2), (W_2, W_3), (W_4, W_2)\}$.

Definition 3.1. Given the data distribution p , a partial ordering $G(\mathcal{K}, \mathcal{E}_p)$ is sharp under the maintained assumptions if there exists no partial ordering $G(\mathcal{K}, \mathcal{E}'_p)$ such that $\mathcal{E}'_p \supsetneq \mathcal{E}_p$ without imposing additional assumptions.

Establishing sharp partial ordering amounts to determining whether we can tightly identify the sign of a counterfactual welfare gap $W_k - W_{k'}$ (i.e., the dynamic treatment effects) for $k, k' \in \mathcal{K}$, and if we can, what the sign is. The sharp identification of the sign is possible when we can construct sharp bounds on the counterfactual welfare gap. This motivates the following analysis.

3.3. Data-Generating Framework

We introduce a simple data-generating framework and formally define the identified set. First, we introduce latent state variables that generate (Y, D) . A latent state of the world will determine specific maps $\mathbf{d}^t \mapsto y_t$ and $\mathbf{z}^t \mapsto d_t$ for $t = 1, \dots, T$ under the exclusion restriction in Assumption SX. A more primitive state of the world would determine maps $(\mathbf{y}^{t-1}, \mathbf{d}^t) \mapsto y_t$ and $(\mathbf{y}^{t-1}, \mathbf{d}^{t-1}, \mathbf{z}^t) \mapsto d_t$ for $t = 1, \dots, T$, but we do not consider them as they not relevant to our objective as shown below. We introduce the latent state variable \tilde{S}_t whose realization represents a latent state. We define \tilde{S}_t as follows. For given $(\mathbf{d}^t, \mathbf{z}^t)$, recall $Y_t(\mathbf{d}^t)$ and $D_t(\mathbf{z}^t)$ denote the counterfactual outcomes and treatments, respectively. Let $\{Y_t(\mathbf{d}^t)\}_{\mathbf{d}^t}$ and $\{D_t(\mathbf{z}^t)\}_{\mathbf{z}^t}$ denote their sequences w.r.t. \mathbf{d}^t and \mathbf{z}^t . Then, by concatenating the two sequences, define $\tilde{S}_t \equiv (\{Y_t(\mathbf{d}^t)\}, \{D_t(\mathbf{z}^t)\}) \in \{0, 1\}^{2^t} \times \{0, 1\}^{2^t}$. For example, $\tilde{S}_1 = (Y_1(0), Y_1(1), D_1(0), D_1(1)) \in \{0, 1\}^2 \times \{0, 1\}^2$, whose realization specifies particular maps $d_1 \mapsto y_1$ and $z_1 \mapsto d_1$. It is convenient to transform $\tilde{\mathbf{S}} \equiv (\tilde{S}_1, \dots, \tilde{S}_T)$ into

¹The way directed graphs are used in this article is completely unrelated to causal graphical models in the literature.

(a) $\delta^*(\cdot)$ is partially identified(b) $\delta^*(\cdot)$ is point identified**Figure 1.** Partially ordered sets of welfares.

a scalar (discrete) latent variable in \mathbb{N} as $S \equiv \beta(\tilde{S}) \in \mathcal{S} \subset \mathbb{N}$, where $\beta(\cdot)$ is a one-to-one map that transforms a binary sequence into a decimal value. Define

$$q_s(x) \equiv \Pr[S = s | X = x],$$

and define the vector $q(x) \equiv \{q_s(x)\}_{s \in \mathcal{S}}$, which represents the distribution of S conditional on $X = x$, namely the true data-generating process. Then the vector $q \equiv \{q(x)\}_{x \in \mathcal{X}}$ resides in $\mathcal{Q} \equiv \{q : \sum_s q_s(x) = 1 \forall x \text{ and } q_s(x) \geq 0 \forall s, x\}$ of dimension $d_q - |\mathcal{X}|$ where $d_q \equiv \dim(q)$. A useful fact is that the joint distributions of counterfactuals (conditional on $X = x$) can be written as linear functionals of $q(x)$:

$$\begin{aligned} \Pr[\mathbf{Y}(\mathbf{d}) = \mathbf{y}, \mathbf{D}(\mathbf{z}) = \mathbf{d} | X = x] &= \Pr[S \in \mathcal{S} : \mathbf{Y}(\mathbf{d}) = \mathbf{y}, \mathbf{D}(\mathbf{z}) \\ &= \mathbf{d} | X = x] \\ &= \Pr[S \in \mathcal{S} : Y_t(\mathbf{d}^t) = y_t, D_t(\mathbf{z}^t) \\ &= d_t \forall t | X = x] \\ &= \sum_{s \in \mathcal{S}_{y,d|\mathbf{z}}} q_s(x), \end{aligned} \quad (3.1)$$

where $\mathcal{S}_{y,d}$ and $\mathcal{S}_{y,d|\mathbf{z}}$ are constructed by using the definition of \mathcal{S} ; their expressions can be found in Appendix B.

Based on (3.1), the counterfactual welfare can be written as a linear combination of $q_s(x)$'s. That is, there exists $1 \times d_q$ vector A_k of 1's and 0's such that

$$W_k = A_k q. \quad (3.2)$$

The formal derivation of A_k can be found in Appendix B, but the intuition is as follows. Recall $W_k \equiv f(q_{\delta_k})$ where $q_{\delta_k}(\mathbf{y}) \equiv \Pr[\mathbf{Y}(\delta(\cdot)) = \mathbf{y}]$. The key observation in deriving the result (3.2) is that $\Pr[\mathbf{Y}(\delta(\cdot)) = \mathbf{y}]$ can be written as a linear functional of the joint distributions of counterfactual outcomes with a *static* regime, that is, $\Pr[\mathbf{Y}(\mathbf{d}) = \mathbf{y} | X = x]$'s, which in turn is a linear functional of $q(x)$. To illustrate this when $T = 2$ and welfare $W_{\delta} = E[Y_2(\delta(\cdot))]$ with $\delta_1(x) = \delta_1$, we have

$$\begin{aligned} \Pr[Y_2(\delta(\cdot)) = 1 | X = x] &= \sum_{y_1 \in \{0,1\}} \Pr[Y_2(\delta_1, \delta_2(Y_1(\delta_1), \delta_1)) \\ &= 1 | Y_1(\delta_1) = y_1, X = x] \Pr[Y_1(\delta_1) = y_1 | X = x] \end{aligned}$$

by the law of iterated expectation. Then, for instance, Regime 4 in Table 1 yields

$$\begin{aligned} \Pr[Y_2(\delta_4(\cdot)) = 1 | X = x] &= P[\mathbf{Y}(1, 1)] \\ &= (1, 1) | X = x + P[\mathbf{Y}(1, 0) = (0, 1) | X = x], \end{aligned} \quad (3.3)$$

where each $\Pr[\mathbf{Y}(d_1, d_2) = (y_1, y_2) | X = x]$ is the counterfactual distribution with a *static* regime, which in turn is a linear combination of $q_s(x)$'s as in (3.1). Finally, $\Pr[Y_2(\delta(\cdot)) = 1] = \sum_{x \in \mathcal{X}} p(x) \Pr[Y_2(\delta(\cdot)) = 1 | X = x]$ where $p(x) \equiv \Pr[X = x]$, and therefore the welfare is a linear function of q .

The data impose restrictions on $q \in \mathcal{Q}$. Define

$$p_{y,d|\mathbf{z},x} \equiv p(\mathbf{y}, \mathbf{d} | \mathbf{z}, x) \equiv \Pr[\mathbf{Y} = \mathbf{y}, \mathbf{D} = \mathbf{d} | \mathbf{Z} = \mathbf{z}, X = x],$$

and p as the vector of $p_{y,d|\mathbf{z},x}$'s except redundant elements. Let $d_p \equiv \dim(p)$. Since $\Pr[\mathbf{Y} = \mathbf{y}, \mathbf{D} = \mathbf{d} | \mathbf{Z} = \mathbf{z}, X = x] = \Pr[\mathbf{Y}(\mathbf{d}) = \mathbf{y}, \mathbf{D}(\mathbf{z}) = \mathbf{d} | X = x]$ by Assumption SX, we can readily show by (3.1) that there exists $d_p \times d_q$ matrix B such that

$$Bq = p, \quad (3.4)$$

where B is a matrix of 1's and 0's; the formal derivation of B can be found in Appendix B. It is worth noting that the linearity in (3.2) and (3.4) is *not* a restriction but given by the discrete nature of the setting. We assume $\text{rank}(B) = d_p$ without loss of generality, because redundant constraints do not play a role in restricting \mathcal{Q} . We focus on the nontrivial case of $d_p < d_q$. If $d_p \geq d_q$, which rarely holds, we can solve for $q = (B^\top B)^{-1} B^\top p$, and can trivially point identify $W_k = A_k q$ and thus $\delta^*(\cdot)$. Otherwise, we have a set of observationally equivalent q 's, which is the source of partial identification and motivates the following definition of the identified set. For simplicity, we use the same notation for the true q and its observational equivalence.

For a given q , let $\delta^*(\cdot; q) \equiv \arg \max_{\delta_k(\cdot) \in \mathcal{D}} W_k = A_k q$ be the optimal regime, explicitly written as a function of the data-generating process.

Definition 3.2. Under Assumption SX, the identified set of $\delta^*(\cdot)$ given the data distribution p is

$$\mathcal{D}_p^* \equiv \{\delta^*(\cdot; q) : Bq = p \text{ and } q \in \mathcal{Q}\} \subset \mathcal{D}, \quad (3.5)$$

which is assumed to be empty when $Bq \neq p$.

3.4. Characterizing Partial Ordering and the Identified Set

Given p , we establish the partial ordering of W_k 's, that is, generate the DAG, by determining whether $W_k > W_{k'}$, $W_k < W_{k'}$, or W_k and $W_{k'}$ are not comparable, denoted as $W_k \sim W_{k'}$, for $k, k' \in \mathcal{K}$. As described in the next theorem, this procedure can be accomplished by determining the signs of the bounds on the welfare gap $W_k - W_{k'}$ for $k, k' \in \mathcal{K}$ and $k > k'$. Note that directly comparing sharp bounds on welfares themselves will

not deliver sharp partial ordering. Then the identified set can be characterized based on the resulting partial ordering.

The nature of the data generation induces the linear system (3.2) and (3.4). This enables us to characterize the bounds on $W_k - W_{k'} = (A_k - A_{k'})q$ as the optima in linear programming. Let $U_{k,k'}$ and $L_{k,k'}$ be the upper and lower bounds. Also let $\Delta_{k,k'} \equiv A_k - A_{k'}$ for simplicity, and thus the welfare gap is expressed as $W_k - W_{k'} = \Delta_{k,k'}q$. Then, for $k, k' \in \mathcal{K}$, we have the main linear programs:

$$\begin{aligned} U_{k,k'} &= \max_{q \in \mathcal{Q}} \Delta_{k,k'}q, \\ L_{k,k'} &= \min_{q \in \mathcal{Q}} \Delta_{k,k'}q, \end{aligned} \quad \text{s.t. } Bq = p. \quad (3.6)$$

Assumption B. $\{q : Bq = p\} \cap \mathcal{Q} \neq \emptyset$.

Assumption B imposes that the model is correctly specified. In particular, this means Assumption SX is correctly specified because the relationship $Bq = p$ is derived under this assumption. Under misspecification, the identified set is empty by definition. The next theorem constructs the sharp DAG and characterize the identified set using $U_{k,k'}$ and $L_{k,k'}$ for $k, k' \in \mathcal{K}$ and $k > k'$, or equivalently, $L_{k,k'}$ for $k, k' \in \mathcal{K}$ and $k \neq k'$ since $U_{k,k'} = -L_{k',k}$.

Theorem 3.1. Suppose Assumptions SX and B hold. Then, (i) $G(\mathcal{K}, \mathcal{E}_p)$ with $\mathcal{E}_p \equiv \{(k, k') \in \mathcal{K} : L_{k,k'} > 0 \text{ and } k \neq k'\}$ is sharp; (ii) \mathcal{D}_p^* defined in (3.5) satisfies

$$\mathcal{D}_p^* = \{\delta_{k'}(\cdot) : \nexists k \in \mathcal{K} \text{ such that } L_{k,k'} > 0 \text{ and } k \neq k'\} \quad (3.7)$$

$$= \{\delta_{k'}(\cdot) : L_{k,k'} \leq 0 \text{ for all } k \in \mathcal{K} \text{ and } k \neq k'\}, \quad (3.8)$$

and therefore the sets on the right-hand side are sharp.

The proof of **Theorem 3.1** is shown in Appendix C. The key insight of the proof is that even though the bounds on the welfare gaps are calculated from separate optimizations, the partial ordering is governed by *common q's* (each of which generates all the welfares) that are observationally equivalent; see Appendix F.2 for related discussions.

Theorem 3.1(i) prescribes how to calculate the sharp DAG as a function of data. The DAG can be conveniently represented in terms of a $|\mathcal{K}| \times |\mathcal{K}|$ adjacency matrix Ω such that its element $\Omega_{k,k'} = 1$ if $W_k \geq W_{k'}$ and $\Omega_{k,k'} = 0$ otherwise. According to (3.7) in (ii), \mathcal{D}_p^* is characterized as the collection of $\delta_k(\cdot)$ where k is in the set of *maximal elements* of the partially ordered set $G(\mathcal{K}, \mathcal{E}_p)$, that is, the set of regimes that are *not inferior*. In **Figure 1**, it is easy to see that the set of maximals is $\mathcal{D}_p^* = \{\delta_1(\cdot), \delta_4(\cdot)\}$ in panel (a) and $\mathcal{D}_p^* = \{\delta_1(\cdot)\}$ in panel (b).

The identified set \mathcal{D}_p^* characterizes the information content of the model. Given the minimal structure we impose in the model, \mathcal{D}_p^* may be large in some cases. However, we argue that an uninformative \mathcal{D}_p^* still has implications for policy: (i) such set may recommend the policymaker eliminate sub-optimal regimes from her options; (ii) in turn, it warns the policymaker about her lack of information (e.g., even if she has access to the experimental data); when $\mathcal{D}_p^* = \mathcal{D}$ as one extreme, “no recommendation” can be given as a nontrivial policy suggestion of the need for better data. As shown in the numerical exercise, the size of \mathcal{D}_p^* is related to the strength of Z_t (i.e., the size of the complier group at t) and the strength of the dynamic treatment

effects. This is reminiscent of the findings in Machado et al. (2019) for the average treatment effect in a static model.

3.5. Additional Assumptions

Often, researchers are willing to impose more assumptions based on priors about the data-generating process, for example, agent's behaviors. Examples are uniformity, Markovian structure, and stationarity. These assumptions are easy to incorporate within the linear programming (3.6); see Appendix D for details. These assumptions tighten the identified set \mathcal{D}_p^* by reducing the dimension of simplex \mathcal{Q} , and thus producing a denser DAG. The list of identifying assumptions here is far from complete, and there may be other assumptions on how (Y, D, Z, X) are generated.

The first assumption is a sequential version of the uniformity assumption (i.e., the monotonicity assumption) in Imbens and Angrist (1994) and Angrist et al. (1996). Let “w.p.1” stand for “with probability one.”

Assumption M1. For each t , either $D_t(Z^{t-1}, 1) \geq D_t(Z^{t-1}, 0)$ w.p.1 or $D_t(Z^{t-1}, 1) \leq D_t(Z^{t-1}, 0)$ w.p.1 conditional on $(Y^{t-1}, D^{t-1}, Z^{t-1}, X)$.

Assumption M1 postulates that there is no defying (or complying) behavior in decision D_t conditional on $(Y^{t-1}, D^{t-1}, Z^{t-1}, X)$. In our illustrative example, M1 assumes that (conditional on the history) there are no individuals with perverse behavior who would participate in the job training when not eligible but would not participate when eligible. We exclude the same perverse behavior in attending high school. Without being conditional on $(Y^{t-1}, D^{t-1}, Z^{t-1}, X)$, however, there can be a general non-monotonic pattern in the way that Z^t influences D^t . For example, we can have $D_t(Z^{t-1}, 1) \geq D_t(Z^{t-1}, 0)$ for $D_{t-1} = 1$ while $D_t(Z^{t-1}, 1) < D_t(Z^{t-1}, 0)$ for $D_{t-1} = 0$. By extending the idea of Vytlacil (2002), we can show that M1 is the equivalent of imposing a threshold-crossing model for D_t :

$$D_t = 1\{\pi_t(Y^{t-1}, D^{t-1}, Z^t, X) \geq \nu_t\}, \quad (3.9)$$

where $\pi_t(\cdot)$ is an unknown, measurable, and nontrivial function of Z_t . The equivalence is formally established in Appendix D. The dynamic selection model (3.9) should not be confused with the dynamic regime (2.1). Compared to the dynamic regime $d_t = \delta_t(y^{t-1}, d^{t-1})$, which is a hypothetical quantity, (3.9) models each individual's *observed* treatment decision, in that it is not only a function of (Y^{t-1}, D^{t-1}) but also ν_t , the individual's unobserved characteristics. We assume that the policymaker has no access to $\nu \equiv (\nu_1, \dots, \nu_T)$. The functional dependence of D_t on $(Y^{t-1}, D^{t-1}, Z^{t-1})$ reflects the agent's learning. Sometimes, we want to further impose uniformity in the formation of Y_t on top of Assumption M1:

Assumption M2. Assumption M1 holds, and for each t , either $Y_t(D^{t-1}, 1) \geq Y_t(D^{t-1}, 0)$ w.p.1 or $Y_t(D^{t-1}, 1) \leq Y_t(D^{t-1}, 0)$ w.p.1 conditional on (Y^{t-1}, D^{t-1}, X) .

This assumption postulates uniformity in a way that restricts heterogeneity of the contemporaneous treatment effect. However, similarly as before, without being conditional

on $(\mathbf{Y}^{t-1}, \mathbf{D}^{t-1}, X)$, there can be a general non-monotonic pattern in the way that \mathbf{D}^t influences \mathbf{Y}^t . For example, we can have $Y_t(\mathbf{D}^{t-1}, 1) \geq Y_t(\mathbf{D}^{t-1}, 0)$ for $Y_{t-1} = 1$ while $Y_t(\mathbf{D}^{t-1}, 1) \leq Y_t(\mathbf{D}^{t-1}, 0)$ for $Y_{t-1} = 0$. In our illustrative example, this implies that the job training program should have a homogeneous influence over the labor market performance across individuals conditional on the history, but it may have heterogeneous influences unconditionally. It is also worth noting that Assumption M2 (and M1) does not assume the direction of monotonicity, but the direction may be recovered from the data. Using a similar argument as before, Assumption M2 is the equivalent of a dynamic version of a nonparametric triangular model:

$$Y_t = 1\{\mu_t(\mathbf{Y}^{t-1}, \mathbf{D}^t, X) \geq \varepsilon_t\}, \quad (3.10)$$

$$D_t = 1\{\pi_t(\mathbf{Y}^{t-1}, \mathbf{D}^{t-1}, \mathbf{Z}^t, X) \geq \nu_t\}, \quad (3.11)$$

where $\mu_t(\cdot)$ and $\pi_t(\cdot)$ are unknown, measurable, and nontrivial functions of D_t and Z_t , respectively. Again, the equivalence is formally established in Appendix D. The next assumption imposes a Markov-type structure in the Y_t and D_t processes.

Assumption K. Conditional on X , $Y_t | (\mathbf{Y}^{t-1}, \mathbf{D}^t) \stackrel{d}{=} Y_t | (Y_{t-1}, D_t)$ and $D_t | (\mathbf{Y}^{t-1}, \mathbf{D}^{t-1}, \mathbf{Z}^t) \stackrel{d}{=} D_t | (Y_{t-1}, D_{t-1}, Z_t)$ for each t .

In terms of the triangular model (3.10)–(3.11), Assumption K implies $Y_t = 1\{\mu_t(Y_{t-1}, D_t, X) \geq \varepsilon_t\}$ and $D_t = 1\{\pi_t(Y_{t-1}, D_{t-1}, Z_t, X) \geq \nu_t\}$, which yields the familiar structure of dynamic discrete choice models found in the literature. Lastly, when there are more than two periods, an assumption that imposes stationarity can be helpful for identification. Such an assumption can be found in Torgovitsky (2019).

4. Application

We apply the framework of this article to understand returns to schooling and post-school training as a sequence of treatments and to conduct a policy analysis. Schooling and post-school training are two major interventions that affect various labor market outcomes, such as earnings and employment status (Ashenfelter and Card 2010). These treatments also have influences on health outcomes, either directly or through the labor market outcomes, and thus of interest for public health policies (Backlund et al. 1996; McDonough et al. 1997; Case et al. 2002). We find that the Job Training Partnership Act (JTPA) is an appropriate setting for our analysis. The JTPA program is one of the largest publicly funded training programs in the United States for economically disadvantaged individuals. Unfortunately, the JTPA only concerns post-school trainings, which have been the main focus in the literature (Bloom et al. 1997; Abadie et al. 2002; Kitagawa and Tetenov 2018). In this article, we combine the JTPA Title II data with those from other sources regarding high school education to create a dataset that allows us to study the effects of a high school (HS) diploma (or its equivalents) and the subsidized job trainings as a sequence of treatments. We consider high school diplomas rather than college degrees because the former is more relevant for the disadvantaged population of Title II of the JTPA program.

We are interested in the dynamic treatment regime $\delta(\cdot) = (\delta_1, \delta_2(\cdot))$, where δ_1 is a HS diploma and $\delta_2(y_1)$ is the job training program given pre-program earning type y_1 . The motivation of having δ_2 as a function of y_1 comes from acknowledging the dynamic nature of how earnings are formed under education and training. The first-stage allocation δ_1 will affect the pre-program earning. This response may contain information about unobserved characteristics of the individuals. Therefore, the allocation of δ_2 can be informed by being adaptive to y_1 . Then, the counterfactual earning type in the terminal stage given $\delta(\cdot)$ can be expressed as $Y_2(\delta(\cdot)) = Y_2(\delta_1, \delta_2(Y_1(\delta_1)))$ where $Y_1(\delta_1)$ is the counterfactual earning type in the first stage given δ_1 . We are interested in the optimal regime δ^* that maximizes each of the following welfares: the average terminal earning $E[Y_2(\delta(\cdot))]$ and the average lifetime earning $E[Y_1(\delta_1)] + E[Y_2(\delta(\cdot))]$.

For the purpose of our analysis, we combine the JTPA data with data from the U.S. Census and the National Center for Education Statistics (NCES), from which we construct the following set of variables: Y_2 above or below median of 30-month earnings, D_2 the job training program, Z_2 a random assignment of the program, Y_1 above or below 80th percentile of pre-program earnings, D_1 the HS diploma or GED, and Z_1 the number of high schools per square mile above or below 35.² The instrument Z_1 for the HS treatment appears in the literature (e.g., Neal 1997). The number of individuals in the sample is 9223. We impose Assumptions SX and M2 throughout the analysis.

The estimation of the DAG and the identified set \mathcal{D}_p^* is straightforward given the conditions in Theorem 3.1 and the linear programs (3.6). The only unknown object is p , the joint distribution of $(\mathbf{Y}, \mathbf{D}, \mathbf{Z})$, which can be estimated as \hat{p} , a vector of $\hat{p}_{y,d|z} = \sum_{i=1}^N 1\{Y_i = y, D_i = d, Z_i = z\} / \sum_{i=1}^N 1\{Z_i = z\}$.

Figure 2 reports the estimated partial ordering of welfare $W_\delta = E[Y_2(\delta(\cdot))]$ (left) and the resulting estimated set $\hat{\mathcal{D}}$ (right, highlighted in red and starred) that we estimate using $\{(Y_i, D_i, Z_i)\}_{i=1}^{9,223}$. Although there exist welfares that cannot be ordered, we can conclude *with certainty* that allocating the program only to the low earning type ($Y_2 = 0$) is welfare optimal, as it is the common implication of Regimes 5 and 6 in $\hat{\mathcal{D}}$. Also, the second best policy is to either allocate the program to the entire population or none, while allocating it only to the high earning type ($Y_2 = 1$) produces the lowest welfare. This result is consistent with the eligibility of Title II of the JTPA, which concerns individuals with “barriers to employment” where the most common barriers are unemployment spells and high-school dropout status (Abadie et al. 2002). Possibly due to the fact that the first-stage instrument Z_1 is not strong enough, we have the two disconnected sub-DAGs and thus the two elements in $\hat{\mathcal{D}}$, which are agnostic about the optimal allocation in the first stage or the complementarity between the first- and second-stage allocations.

Figure 3 reports the estimated partial ordering and the estimated set with $W_\delta = E[Y_1(\delta_1)] + E[Y_2(\delta(\cdot))]$. Despite the partial ordering, $\hat{\mathcal{D}}$ is a singleton for this welfare and δ^* is estimated

²For Y_1 , the 80th percentile cutoff is chosen as it is found to be relevant in defining subpopulations that have contrasting effects of the program. There are other covariates in the constructed dataset, but we omit them for the simplicity of our analysis. These variables can be incorporated as pre-treatment covariates so that the first-stage treatment is adaptive to them.

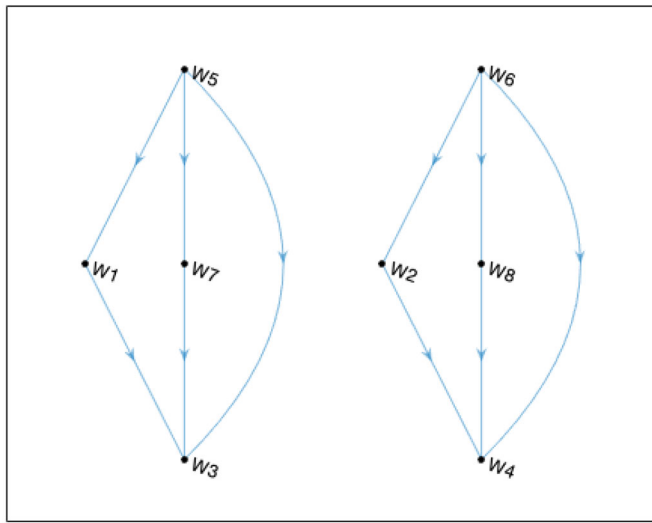


Figure 2. Estimated DAG of $W_\delta = E[Y_2(\delta(\cdot))]$ and estimated set for δ^* (red and starred).

Regime #	δ_1	$\delta_2(1, \delta_1)$	$\delta_2(0, \delta_1)$
1	0	0	0
2	1	0	0
3	0	1	0
4	1	1	0
5*	0	0	1
6*	1	0	1
7	0	1	1
8	1	1	1

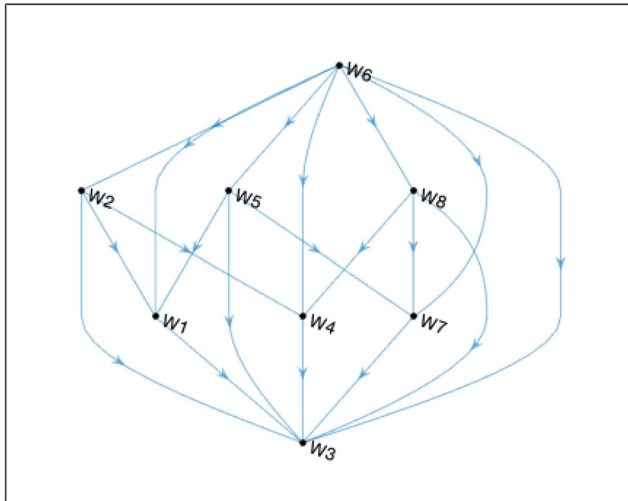


Figure 3. Estimated DAG of $W_\delta = E[Y_1(\delta_1)] + E[Y_2(\delta(\cdot))]$ and Estimated Set for δ^* (red and starred).

Regime #	δ_1	$\delta_2(1, \delta_1)$	$\delta_2(0, \delta_1)$
1	0	0	0
2	1	0	0
3	0	1	0
4	1	1	0
5	0	0	1
6*	1	0	1
7	0	1	1
8	1	1	1

to be Regime 6. According to this regime, the average lifetime earning is maximized by allocating HS education to all individuals and the training program to individuals with low pre-program earnings. As discussed earlier, additional policy implications can be obtained by inspecting suboptimal regimes. Interestingly, Regime 8, which allocates the treatments regardless, is inferior to Regime 6. This can be useful knowledge for policy makers especially because Regime 8 is the most “expensive” regime. Similarly, Regime 1, which does not allocate any treatments regardless and thus is the least expensive regime, is superior to Regime 3, which allocates the program to high-earning individuals. The estimated DAG shows how more expensive policies do not necessarily achieve greater welfare. Moreover, these conclusions can be compelling as they are drawn without making arbitrary parametric restrictions nor strong identifying assumptions.

Finally, as an alternative approach, we use $\{(Y_i, D_i, Z_{2i})\}_{i=1}^{9,223}$ for estimation, that is, we drop Z_1 and only use the exogenous variation from Z_2 . This reflects a possible concern that Z_1 may not be as valid as Z_2 . Then, the estimated DAG looks identical to the left panel of Figure 2 whether the targeted welfare is

$E[Y_2(\delta(\cdot))]$ or $E[Y_1(\delta_1)] + E[Y_2(\delta(\cdot))]$. Clearly, without Z_1 , the procedure lacks the ability to determine the first stage’s best treatment. Note that, even though the DAG for $E[Y_2(\delta(\cdot))]$ is identical for the case of one versus two instruments, the inference results will reflect such difference by producing a larger confidence set for the former case.

Supplementary Materials

In the online supplemental appendix, the analysis with binary outcomes and discrete covariates is extended to continuous outcomes and covariates, and stochastic regimes are discussed. The supplemental appendix also presents numerical studies and discusses topological sorts, cardinality reduction for the set of regimes, and inference. Most proofs are collected in the supplemental appendix.

Acknowledgments

For helpful comments and discussions, the author is grateful to Donald Andrews, Isaiah Andrews, Junehyuk Jung, Yuichi Kitamura, Hiro Kaido, Shakeeb Khan, Adam McCloskey, Susan Murphy, Takuya Ura, Ed Vytlačil, Shenshen Yang, participants in the 2021 Cowles Conference, the 2021 North American Winter Meeting, the 2020 European Winter Meeting, and the

2020 World Congress of the Econometric Society, 2019 CEMMAP & WISE conference, the 2019 Asian Meeting of the Econometric Society, the Bristol Econometrics Study Group Conference, the 2019 Midwest Econometrics Group Conference, the 2019 Southern Economic Association Conference, and in seminars at Harvard, MIT, LSE, UCL, U of Toronto, Simon Fraser U, Rice U, UIUC, U of Bristol, Queen Mary London, U of Colorado Boulder, NUS, and SMU.

Disclosure Statement

The author reports there are no competing interests to declare.

References

- Abadie, A., Angrist, J., and Imbens, G. (2002), "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91–117. [8]
- Almond, D., and Mazumder, B. (2013), "Fetal Origins and Parental Responses," *Annual Review of Economics*, 5, 37–56. [4]
- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996), "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–455. [2,7]
- Ashenfelter, O., and Card, D. (2010), *Handbook of Labor Economics*, Amsterdam: Elsevier. [8]
- Athey, S., and Imbens, G. W. (2019), "Machine Learning Methods that Economists Should Know About," *Annual Review of Economics*, 11, 685–725. [3]
- (2022), "Design-based Analysis in Difference-in-Differences Settings with Staggered Adoption," *Journal of Econometrics*, 226, 62–79. [2]
- Athey, S., and Wager, S. (2021), "Policy Learning with Observational Data," *Econometrica*, 89, 133–161. [3]
- Backlund, E., Sorlie, P. D., and Johnson, N. J. (1996), "The Shape of the Relationship between Income and Mortality in the United States: Evidence from the National Longitudinal Mortality Study," *Annals of Epidemiology*, 6, 12–20. [8]
- Balke, A., and Pearl, J. (1997), "Bounds on Treatment Effects from Studies with Imperfect Compliance," *Journal of the American Statistical Association*, 92, 1171–1176. [2]
- Bhattacharya, D., and Dupas, P. (2012), "Inferring Welfare Maximizing Treatment Assignment Under Budget Constraints," *Journal of Econometrics*, 167, 168–196. [3]
- Bloom, H. S., Orr, L. L., Bell, S. H., Cave, G., Doolittle, F., Lin, W., and Bos, J. M. (1997), "The Benefits and Costs of JTPA Title II-A Programs: Key Findings from the National Job Training Partnership Act study," *Journal of Human Resources*, 32, 549–576. [8]
- Callaway, B., and Sant'Anna, P. H. (2021), "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 225, 200–230. [2]
- Case, A., Lubotsky, D., and Paxson, C. (2002), "Economic Status and Health in Childhood: The Origins of the Gradient," *American Economic Review*, 92, 1308–1334. [8]
- Cellini, S. R., Ferreira, F., and Rothstein, J. (2010), "The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design," *The Quarterly Journal of Economics*, 125, 215–261. [2,4]
- Chao, Y.-C., Tran, Q., Tsodikov, A., and Kidwell, K. M. (2022), "Joint Modeling and Multiple Comparisons with the Best of Data from a SMART with Survival Outcomes," *Biostatistics*, 23, 294–313. [3]
- Conduct Problems Prevention Research Group. (1992), "A Developmental and Clinical Model for the Prevention of Conduct Disorder: The FAST Track Program," *Development and Psychopathology*, 4, 509–527. [5]
- Cui, Y., and Tchetgen Tchetgen, E. (2021), "A Semiparametric Instrumental Variable Approach to Optimal Treatment Regimes under Endogeneity," *Journal of the American Statistical Association*, 116, 162–173. [2,3]
- Cunha, F., and Heckman, J. (2007), "The Technology of Skill Formation," *American Economic Review*, 97, 31–47. [4]
- de Chaisemartin, C., and d'Haultfoeuille, X. (2020), "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 110, 2964–2996. [2]
- Ertefaie, A., Wu, T., Lynch, K. G., and Nahum-Shani, I. (2016), "Identifying a Set that Contains the Best Dynamic Treatment Regimes," *Biostatistics*, 17, 135–148. [3]
- Han, S. (2021a), "Comment: Individualized Treatment Rules under Endogeneity," *Journal of the American Statistical Association*, 116, 192–195. [2]
- (2021b), "Identification in Nonparametric Models for Dynamic Treatment Effects," *Journal of Econometrics*, 225, 132–147. [2,3]
- Han, S., and Yang, S. (2023), "A Computational Approach to Identification of Treatment Effects for Policy Evaluation," arXiv preprint arXiv:2009.13861. [2]
- Heckman, J. J., Humphries, J. E., and Veramendi, G. (2016), "Dynamic Treatment Effects," *Journal of Econometrics*, 191, 276–292. [2]
- Heckman, J. J., and Navarro, S. (2007), "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics*, 136, 341–396. [2]
- Hirano, K., and Porter, J. R. (2009), "Asymptotics for Statistical Treatment Rules," *Econometrica*, 77, 1683–1701. [3]
- Imbens, G. W., and Angrist, J. D. (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475. [2,7]
- Johnson, R. C., and Jackson, C. K. (2019), "Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending," *American Economic Journal: Economic Policy*, 11, 310–349. [4,5]
- Kamat, V. (2019), "Identification with Latent Choice Sets: The Case of the Head Start Impact Study," arXiv preprint arXiv:1711.02048. [2]
- Kasy, M. (2016), "Partial Identification, Distributional Preferences, and the Welfare Ranking of Policies," *Review of Economics and Statistics*, 98, 111–131. [3]
- Kasy, M., and Sautmann, A. (2021), "Adaptive Treatment Assignment in Experiments for Policy Choice," *Econometrica*, 89, 113–132. [3]
- Kitagawa, T., and Tetenov, A. (2018), "Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice," *Econometrica*, 86, 591–616. [3,8]
- Kock, A. B., Preinerstorfer, D., and Veliyev, B. (2021), "Functional Sequential Treatment Allocation," *Journal of the American Statistical Association*, 117, 1311–1323. [3]
- Kramer, M. S., Chalmers, B., Hodnett, E. D., Sevkovskaya, Z., Dzikovich, I., Shapiro, S., Collet, J.-P., Vanilovich, I., Mezen, I., Ducruet, T., et al. (2001), "Promotion of Breastfeeding Intervention Trial (PROBIT): A Randomized Trial in the Republic of Belarus," *Journal of the American Medical Association*, 285, 413–420. [5]
- Machado, C., Shaikh, A., and Vytlacil, E. (2019), "Instrumental Variables and the Sign of the Average Treatment Effect," *Journal of Econometrics*, 212, 522–555. [2,7]
- Manski, C. F. (2004), "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, 72, 1221–1246. [3]
- (2007), "Partial Identification of Counterfactual Choice Probabilities," *International Economic Review*, 48, 1393–1410. [2]
- McDonough, P., Duncan, G. J., Williams, D., and House, J. (1997), "Income Dynamics and Adult Mortality in the United States, 1972 through 1989," *American Journal of Public Health*, 87, 1476–1483. [8]
- Mogstad, M., Santos, A., and Torgovitsky, A. (2018), "Using Instrumental Variables for Inference about Policy Relevant Treatment Parameters," *Econometrica*, 86, 1589–1619. [2]
- Murphy, S. A. (2003), "Optimal Dynamic Treatment Regimes," *Journal of the Royal Statistical Society, Series B*, 65, 331–355. [1,2,3]
- Murphy, S. A., van der Laan, M. J., Robins, J. M., and C. P. P. R. Group. (2001), "Marginal Mean Models for Dynamic Regimes," *Journal of the American Statistical Association*, 96, 1410–1423. [1]
- Neal, D. (1997), "The Effects of Catholic Secondary Schooling on Educational Achievement," *Journal of Labor Economics*, 15, 98–123. [8]
- Qiu, H., Carone, M., Sadikova, E., Petukhova, M., Kessler, R. C., and Luedtke, A. (2021), "Optimal Individualized Decision Rules Using Instrumental Variable Methods," *Journal of the American Statistical Association*, 116, 174–191. [2]
- Robins, J. M. (2004), "Optimal Structural Nested Models for Optimal Sequential Decisions," in *Proceedings of the Second Seattle Symposium in Biostatistics*, Springer, pp. 189–326. [1]
- Stoye, J. (2012), "Minimax Regret Treatment Choice with Covariates or with Limited Validity of Experiments," *Journal of Econometrics*, 166, 138–156. [3]

- Sun, L., and Abraham, S. (2021), "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 225, 175–199. [2]
- The Systolic Hypertension in the Elderly Program (SHEP) Cooperative Research Group. (1988), "Rationale and Design of a Randomized Clinical Trial on Prevention of Stroke in Isolated Systolic Hypertension," *Journal of Clinical Epidemiology*, 41, 1197–1208. [5]
- Torgovitsky, A. (2019), "Nonparametric Inference on State Dependence in Unemployment," *Econometrica*, 87, 1475–1505. [2,8]
- Vytlačil, E. (2002), "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341. [7]
- Zhang, Y., Laber, E. B., Tsiatis, A., and Davidian, M. (2015), "Using Decision Lists to Construct Interpretable and Parsimonious Treatment Regimes," *Biometrics*, 71, 895–904. [3]