Optimal Dynamic Treatment Regimes from Studies with Imperfect Compliance*

Sukjin Han
Department of Economics
University of Texas at Austin
sukjin.han@austin.utexas.edu

This Draft: July 29, 2019

Abstract

Dynamic treatment regimes are treatment assignments tailored to heterogeneous individuals. The optimal dynamic treatment regime is a sequence of adaptive assignment rules that maximizes average welfares. This paper investigates the possibility of identification of optimal dynamic regimes when the data (i.e., observed outcomes and treatments) are generated from multi-period experiments with possible non-compliance, or more generally from observational studies. The biostatistics literature that studies optimal treatment regimes relies on the sequential randomization assumption, although non-compliance is prevalent especially in multi-period settings. This paper relaxes sequential randomization proposes a simple nonparametric framework with monotonicity, under which we can learn optimal dynamic regimes via exclusion restrictions. We establish the partial ordering of the average potential outcomes using binary excluded instruments, which are then used to construct the identified set of optimal dynamic regimes. By viewing the partial ordering as a directed acyclic graph, we show

*The author is grateful to Donald Andrews, Toru Kitagawa, Ed Vytlacil, participants in the 2019 CEMMAP & WISE conference, the 2019 Asian Meeting of the Econometric Society, the Bristol Econometrics Study Group Conference, and in seminars at National U of Singapore, Singapore Management U, and U of Nebraska for helpful comments and discussions. This is a preliminary draft. Comments are welcome.

that the identified set can be computed using a topological sorting algorithm. In this paper, we also calculate bounds on the optimal welfares and regrets, and show how additional variation in exogenous variables helps shrink the identified set of the objects of interest.

JEL Numbers: C14, C32, C33, C36
Keywords: Optimal dynamic treatment regimes, endogenous treatments, non-compliance, average treatment effect, instrumental variable.

1 Introduction

Dynamic treatment regimes are treatment assignments tailored to individual need. When individuals are heterogeneous, their welfare may be improved by assigning treatments that are adaptive to individual heterogeneity, compared to that from assigning pre-determined treatments. Typically, a dynamic (i.e., adaptive) treatment regime is defined as a sequence of assignment rules that map previous outcomes and treatments onto current allocation decisions. The optimal dynamic treatment regime is then defined as a sequence of assignment rules that maximizes a social planner's objective function, such as average welfare. This paper investigate the possibility of identification of optimal dynamic regimes when a panel of outcomes and treatments is generated from multi-period experiments in the presence of non-compliance, or more generally from observational studies.

To clarify this paper's framework, consider the notion of exploration and exploitation. Exploration refers to an analyst's problem who wants to learn the effects of treatments (e.g., via randomized trials). On the other hand, exploitation is a planner's problem who wants to allocate treatments that are beneficial to individuals (e.g., via optimal allocation rules). As there is a clear trade-off between the two (i.e., when we explore, we cannot exploit, and vice versa), we consider a simple framework that separates the data generation stage and the stage of applying optimal regimes. That is, we want to learn about optimal regimes from the data generated by exploration, so that the planner can use them for future exploitation. In particular, in order to learn dynamic regimes, we need to conduct multi-period experiments. Examples of multi-period experiments can be easily found in medical interventions, educational interventions, or online advertisements.

Optimal treatment regimes have been extensively studied in the biostatistics literature (Robins (1997), Murphy et al. (2001), and Murphy (2003), among others). These studies often critically rely on an ideal multi-period experimental environment that satisfies sequential randomization, namely, that the treatment is randomized every period within those individuals who have the same history of outcomes and treatments, and such an assignment is fully complied. Based on this assumption, they identify optimal regimes that maximize welfare, written as the average of potential outcomes. Non-compliance, however, is prevalent in experiments, especially in multi-period settings, e.g., due to the cost of enforcement or subjects' learning, and therefore it is important to be allowed for. For example, in Although we focus on randomized control trials in the presence of non-compliance as a leading example, the framework covers observational studies.

This paper relaxes sequential randomization and proposes a simple nonparametric framework with monotonicity, under which we can partially learn optimal dynamic treatment regimes via exclusion restrictions. Following the literature, we define the welfare as the average potential outcome in, e.g., the terminal period, which is a function of a dynamic regime. We consider binary outcomes and treatments to introduce feasible dynamic regimes by reducing the cardinality of possible regimes. The structure we impose on the data generating process is a monotonicity/uniformity assumption that generalizes the local average treatment effect (LATE) monotonicity assumption in Imbens and Angrist (1994). By extending Vytlacil (2002), we show that the monotonicity assumption is equivalent to imposing a nonparametric threshold-crossing structure. Using a range of monotonicity assumptions and a sequence of binary instruments, we characterize the identified set of optimal regimes as a discrete subset of all possible regimes, and subsequently calculate bounds on the optimal welfares and regrets. We then show how additional variation in exogenous variables can shrink the identified sets of the objects of interest. The existence of the exogenous variables is motivated by an assumption that agent's information set is limited in anticipating factors that affect outcomes in next periods.

The analysis is conducted in two steps. As a first step, we establish the partial orderings of the joint distributions of all potential outcomes across periods with respect to the static (i.e., non-adaptive) regimes. Establishing the ordering the joint distributions is closely related to a single-period problem of identifying the sign of the ATE (or more generally establishing bounds on the ATE). The approach in the first step of our analysis can therefore be viewed as a dynamic generalization of the analysis pioneered by Balke and Pearl (1997) and extended by Machado et al. (Forthcoming). The second step analysis is based on the fact that potential outcomes with an adaptive regime are defined by viewing each adaptive rule as a contingency plan that takes values of static regimes as inputs and outputs. Exploiting this relationship between adaptive and non-adaptive regimes, we show how the orderings established in the first step can be utilized to establish the partial ordering of the average potential outcome. By viewing the partial ordering as a directed acyclic graph, we characterize the identified set of the optimal adaptive regimes as a set of maxima of all subgraphs that are directed paths (i.e., that are totally ordered). In practice, the set of possible adaptive regimes may be too large, even if we have made the problem discrete. This is particularly true in the case of more than two periods, in which analytical derivation of the identified set is cumbersome. Using the directed acyclic graph representation, we show that the identified set can be easily computed using a topological sorting algorithm (e.g., Kahn (1962)).

In the paper, we also discuss how to reduce the cardinality of the set of possible regimes for computational, institutional, and practical reasons. We consider a dynamic regime, which is only a function of the lagged outcome and treatment, as opposed to a function of the full history considered above. We also consider a dynamic regime where the rule is adaptive only in later periods. This is a reasonable setting to consider, since an adaptive regime can costly and it is cost-effective to apply once the knowledge of individual heterogeneity is sufficiently accumulated over time.

To our best knowledge, this paper is first in the literature that considers identification of optimal dynamic regimes under treatment endogeneity. Robins (1997), Murphy et al. (2001), and Murphy (2003) identify optimal dynamic regimes but under the sequential randomization assumption. Recently, Han (2019) and Wang and Tchetgen Tchetgen (2018) relax sequential randomization (and thus allow non-compliance) and considers identification of average treatment effects, but only as functions of non-adaptive regimes which greatly simplify the analysis. Also, Han (2019) considers point identification which requires conditions on the support of exogenous variables, which we avoid in the current paper. Besides Balke and Pearl (1997) and Machado et al. (Forthcoming), this paper's strategy for partial identification is also related to Vytlacil and Yildiz (2007), Shaikh and Vytlacil (2011), Jun et al. (2016), and Balat and Han (2018), which all consider single-period settings. Similar to our approach, Heckman and Navarro (2007) and Heckman et al. (2016) utilize exclusion restrictions to recover dynamic treatment effects, but they rely on infinite support assumptions and consider irreversible treatments.

In terms of notation, let $\boldsymbol{W}^t \equiv (W_1, .., W_t)$ denote a row vector that collects r.v.'s W_t across time up to t, and let $\boldsymbol{w}^t$ be its realization. Most of the time, we write $\boldsymbol{W} \equiv \boldsymbol{W}^T$ for convenience. For a vector $\boldsymbol{W}$ without the t-th element, we write $\boldsymbol{W}_{-t} \equiv (W_1, ..., W_{t-1}, W_{t+1}, ..., W_T)$ with realization $\boldsymbol{w}_{-t}$. For r.v.'s Y and W, we sometimes abbreviate $\Pr[Y = y|W = w]$ to $\Pr[Y = y|w]$. We abbreviate "with probability one" as "w.p.1" and "with respect to" as "w.r.t." The symbol "$\perp$" denotes statistical independence.

2 Dynamic Regimes and Counterfactual Outcomes

For a finite horizon $t = 1, ..., T$ with fixed T, let Y_t be a binary outcome at t and A_t be a binary treatment at t, with realizations y_t and a_t, respectively. For example, Y_t is a symptom indicator of a patient and A_t is a medical treatment received. We consider binary outcomes and treatments since they are helpful in defining, analyzing and implementing dynamic regimes by reducing the cardinality of possible regimes. We consider a small T large N panel, and suppress the individual unit i throughout the paper. For each t, define an *adaptive treatment rule* $d_t : \{0,1\}^{t-1} \times \{0,1\}^{t-1} \to \{0,1\}$ that maps the lags of realized outcomes and treatments onto a non-stochastic treatment allocation $a_t \in \{0,1\}$:

$$d_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}) = a_t. \tag{2.1}$$

This adaptive rule also appears in, e.g., Murphy (2003). A special case of (2.1) is a non-adaptive rule where $d_t(\cdot)$ is just a constant function (Han (2019), Wang and Tchetgen Tchetgen (2018)). Whether the rule is adaptive or non-adaptive, we only consider non-stochastic rules.[1] Then, a *dynamic regime* up to period t is defined as a vector of all treatment rules (given $\boldsymbol{y}^{t-1}$)

$$\boldsymbol{d}^t(\cdot) \equiv \left(d_1, d_2(y_1, d_1), d_3(\boldsymbol{y}^2, \boldsymbol{d}^2(\cdot)), ..., d_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}(\cdot))\right)$$

[1] A stochastic rule allocates the probability of treatment and is considered in, e.g., Murphy et al. (2001), Murphy (2003), and Manski (2004). Our analysis can be extended to this case, although we do not pursue in this paper.

in the class $\mathcal{D}$ of all possible regimes.

We define a counterfactual outcome as a function of this dynamic regime. With dynamic regimes, expressing potential outcomes is more involved than with static regimes. Let $Y_t(\boldsymbol{a}^t)$ be a potential outcome with non-adaptive allocation sequence $\boldsymbol{a}^t$, and let $\boldsymbol{Y}^t(\boldsymbol{a}^t) \equiv (Y_1(a_1), Y_2(\boldsymbol{a}^2), ..., Y_t(\boldsymbol{a}^t))$. We express a potential outcome with adaptive regime $\boldsymbol{d}^t(\cdot)$ as follows:

$$Y_t(\boldsymbol{d}^t(\cdot)) \equiv Y_t(\boldsymbol{a}^t) \tag{2.2}$$

where

$$\begin{aligned} a_1 &= d_1, \\ a_2 &= d_2(Y_1(a_1), a_1), \\ a_3 &= d_3(\boldsymbol{Y}^2(\boldsymbol{a}^2), \boldsymbol{a}^2), \\ &\vdots \\ a_t &= d_t(\boldsymbol{Y}^{t-1}(\boldsymbol{a}^{t-1}), \boldsymbol{a}^{t-1}). \end{aligned} \tag{2.3}$$

In this recursive expression, for each t, the adaptive regime $\boldsymbol{d}^t(\cdot)$ take a value $\boldsymbol{a}^t$ which is fed into the next period's rule as an argument itself and as an argument of potential outcome vector.

Let $\boldsymbol{d}(\cdot) \equiv \boldsymbol{d}^T(\cdot)$. Given the definitions above, we can define the *optimal dynamic regime* as the regime that maximizes the average terminal potential outcome:[2]

$$\boldsymbol{d}^*(\cdot) = \arg \max_{\boldsymbol{d}(\cdot) \in \mathcal{D}} E[Y_T(\boldsymbol{d}(\cdot))].$$

It is fruitful for our analysis to rewrite the average potential outcome as

$$E[Y_T(\boldsymbol{d}(\cdot))] = E\left[E\left[\cdots E\left[E[Y_T(\boldsymbol{a})|\boldsymbol{Y}^{T-1}(\boldsymbol{a}^{T-1})]\middle| \boldsymbol{Y}^{T-2}(\boldsymbol{a}^{T-2})\right]\cdots \middle| Y_1(a_1)\right]\right], \tag{2.4}$$

[2] We assume that the optimal dynamic regime is unique by simply ruling out knife-edge cases where two regimes deliver the same welfare.

where $\boldsymbol{a} = (a_1, ..., a_T)$ satisfies

$$\begin{aligned}
a_1 &= d_1, \\
a_2 &= d_2(Y_1(a_1), a_1), \\
a_3 &= d_3(\boldsymbol{Y}^2(\boldsymbol{a}^2), \boldsymbol{a}^2), \\
&\vdots \\
a_T &= d_T(\boldsymbol{Y}^{T-1}(\boldsymbol{a}^{T-1}), \boldsymbol{a}^{T-1}).
\end{aligned}$$

Given this expression, the solution $\boldsymbol{d}^*(\cdot)$ can be justified by backward induction in a finite-horizon dynamic programming. First, the T-th element in $\boldsymbol{d}^*(\cdot)$ corresponds to the optimal rule at the final period:

$$d_T^*(\boldsymbol{y}^{T-1}, \boldsymbol{a}^{T-1}) = \arg\max_{a_T} E[Y_T(\boldsymbol{a})|\boldsymbol{Y}^{T-1}(\boldsymbol{a}^{T-1}) = \boldsymbol{y}^{T-1}].$$

Define a value function at period T as $V_T(\boldsymbol{y}^{T-1}, \boldsymbol{a}^{T-1}) \equiv \max_{a_T} E[Y_T(\boldsymbol{a})|\boldsymbol{Y}^{T-1}(\boldsymbol{a}^{T-1}) = \boldsymbol{y}^{T-1}]$. Similarly, for each $t = 1, ..., T-1$, let

$$d_t^*(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}) = \arg\max_{a_t} E[V_{t+1}(\boldsymbol{Y}^t(\boldsymbol{a}^t), \boldsymbol{a}^t)|\boldsymbol{Y}^{t-1}(\boldsymbol{a}^{t-1}) = \boldsymbol{y}^{t-1}]$$

and $V_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}) \equiv \max_{a_t} E[V_{t+1}(\boldsymbol{Y}^t(\boldsymbol{a}^t), \boldsymbol{a}^t)|\boldsymbol{Y}^{t-1}(\boldsymbol{a}^{t-1}) = \boldsymbol{y}^{t-1}]$, which then iteratively defines all the elements in $\boldsymbol{d}^*(\cdot)$.[3] By definition, $\boldsymbol{d}^*(\cdot)$ is adaptive to past outcomes and treatments.[4]

To illustrate how $\boldsymbol{d}^*(\cdot)$ is defined, suppose $T = 2$. Then, the two potential outcomes are defined as $Y_1(d_1) = Y_1(a_1)$ and $Y_2(\boldsymbol{d}^2(\cdot)) = Y_2(a_1, a_2)$ with $a_1 = d_1$ and $a_2 = d_2(Y_1(a_1), a_1)$, or more succinctly,

$$Y_2(\boldsymbol{d}^2(\cdot)) = Y_2(d_1, d_2(Y_1(d_1), d_1)),$$

and $\boldsymbol{d}^*(\cdot) = \arg\max_{\boldsymbol{d}(\cdot)} E[Y_2(\boldsymbol{d}(\cdot))] = E\left[E[Y_2(d_1, d_2(Y_1(d_1), d_1))|Y_1(d_1)]\right]$. Also, using

[3] Although we consider a stylized objective function here for simplicity, we may be able to have more realistic objective functions (e.g., the welfare function in Kitagawa and Tetenov (2018); Manski (2004) or the net welfare in Han (2019)).

[4] To reduce the dimension of the regime, we may want to consider an unconditional objective function $E[Y_T(\boldsymbol{d}(\cdot))]$ instead, but integrating $E[Y_T(\boldsymbol{d}(\cdot))|\boldsymbol{X}]$ may require an additional support condition.

backward induction, we have

$$d_2^*(y_1, a_1) = \arg\max_{a_2} E[Y_2(\boldsymbol{a})|Y_1(a_1) = y_1], \tag{2.5}$$

and, by defining $V_2(y_1, a_1) \equiv \max_{a_2} E[Y_2(\boldsymbol{a})|Y_1(a_1) = y_1]$,

$$d_1^* = \arg\max_{a_1} E[V_2(Y_1(a_1), a_1)]. \tag{2.6}$$

Therefore, $\boldsymbol{d}^*(\cdot)$ is equal to the collection of these solutions: $\boldsymbol{d}^*(\cdot) = (d_1^*, d_2^*(y_1, d_1^*))$.

Based on $\boldsymbol{d}^*(\cdot)$, we are also interested in calculating the *optimal welfare*:

$$E[Y_T(\boldsymbol{d}^*(\cdot))],$$

as well as the *regret* from following individual decisions:

$$E[Y_T(\boldsymbol{d}^*(\cdot))] - E[Y_T(\boldsymbol{A})] = E[Y_T(\boldsymbol{d}^*(\cdot))] - E[Y_T],$$

and the *gain from the adaptive regime* (compared to a non-adaptive regime):

$$E[Y_T(\boldsymbol{d}^*(\cdot))] - E[Y_T(\boldsymbol{a}^*)],$$

where $E[Y_T(\boldsymbol{a}^*)] = \max_{\boldsymbol{a}} E[Y_T(\boldsymbol{a})]$ is the optimal welfare with a non-adaptive rule.

3 Assumptions

To facilitate the identification analysis of $\boldsymbol{d}^*(\cdot)$ without invoking sequential randomization, we introduce a sequence of instruments and a range of monotonicity assumptions. Let Z_t be a (potentially binary) instrument at t. Let $(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z})$ be the vector of observables for the entire T periods. We posit that $(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z})$ is the panel data from which we want to learn the optimal dynamic regimes and is generated by a data generating progress that satisfies the following set of assumptions. Let $A_t(\boldsymbol{z}^t)$ be the counterfactual treatment had the sequence $\boldsymbol{z}^t$ be assigned. Let $\boldsymbol{Y}(\boldsymbol{a}) \equiv (Y_1(a_1), Y_2(\boldsymbol{a}^2), ..., Y_T(\boldsymbol{a}^T))$ and $\boldsymbol{A}(\boldsymbol{z}) \equiv (A_1(z_1), A_2(\boldsymbol{z}^2), ..., A_T(\boldsymbol{z}^T))$.

Assumption SX. $(\boldsymbol{Y}(\boldsymbol{a}), \boldsymbol{A}(\boldsymbol{z})) \perp \boldsymbol{Z}$.

Assumption R. *(i)* $\Pr[A_t = 1|\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^t]$ *is a nontrivial function of* Z_t*; (ii)*

$\Pr[Y_t = 1 | \boldsymbol{Y}^{t-1}, \boldsymbol{A}^t]$ *is a nontrivial function of* A_t.

Assumption M1. *For each* t*, either* $A_t(\boldsymbol{Z}^{t-1}, 1) \geq A_t(\boldsymbol{Z}^{t-1}, 0)$ *w.p.1 or* $A_t(\boldsymbol{Z}^{t-1}, 1) \leq A_t(\boldsymbol{Z}^{t-1}, 0)$ *w.p.1. conditional on* $(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^{t-1})$.

Assumption SX imposes strict exogeneity for $\boldsymbol{Z}$. Assumption R is a regularity condition that rules out non-relevance of contemporary variables. Assumption M1 imposes monotonicity of A_t in Z_t conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^{t-1})$. This assumption can be viewed as a sequential version of the LATE monotonicity assumption. Without conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^{t-1})$, however, there can be a general non-monotonic pattern in the way that $\boldsymbol{Z}^t$ influences $\boldsymbol{A}^t$. General non-monotonicity related to this is considered in Lee and Salanié (2018). By extending the idea of Vytlacil (2002), we can show that M1 is equivalent of imposing a threshold-crossing model for A_t under SX:

$$A_t = 1\{\pi_t(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^t) \geq V_t\}, \tag{3.1}$$

where $\pi_t(\cdot)$ is an unknown, measurable, and non-trivial function of Z_t.

Lemma 3.1. *Suppose Assumptions SX and R(i) hold. Assumption M1 is equivalent to* (3.1) *being satisfied conditional on* $(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^{t-1})$ *for each* t.

The proofs of this lemma and others below are presented in the Appendix.

The treatment selection model (3.1) should not be confused with the dynamic regime (2.1). Compared to the dynamic regime $A_t = d_t(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1})$, the equation $A_t = 1\{\pi_t(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^t) \geq V_t\}$ denotes each individual's treatment decision, in that it is not only a function of $(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1})$ but also V_t, the individual's unobserved characteristics. We assume that the social planner has no access to $\boldsymbol{V}$. The functional dependence of A_t on the past outcomes and treatments $(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1})$ and a sequence of random assignments $(\boldsymbol{Z}^{t-1})$ reflects the agent's learning.

Sometimes, we want to further impose monotonicity of Y_t in A_t on top of Assumption M1:

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

As before, without conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^{t-1})$, there can be a general non-monotonic pattern in the way that $\boldsymbol{A}^t$ influences $\boldsymbol{Y}^t$. It is important to note that Assumption M2 does not assume the direction of monotonicity. It rather assumes the uniformity in the way that individuals' outcomes at t are affected by the contemporary treatment. This is in contrast to the monotone treatment response condition in e.g., Manski (1997), which assumes the direction. By a similar argument as before, Assumption M2 is equivalent of a dynamic version of a nonparametric triangular model under SX:

$$Y_t = 1\{\mu_t(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^t) \geq U_t\}, \tag{3.2}$$

$$A_t = 1\{\pi_t(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^t) \geq V_t\}, \tag{3.3}$$

where $\mu_t(\cdot)$ and $\pi_t(\cdot)$ are unknown, measurable and non-trivial functions of Y_t and A_t, respectively.

Lemma 3.2. *Suppose Assumptions SX and R hold. Assumption M2 is equivalent to* (3.2)–(3.3) *being satisfied conditional on* $(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^{t-1})$ *for each* t*.*

As clearly seen in (3.2), Assumption M2 imposes non-trivial restrictions on treatment heterogeneity. To illustrate this point, consider an alternative specification for Y_t:

$$Y_t = 1\{\mu_t(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^t) \geq U_t(A_t)\}, \tag{3.4}$$

where $U_t(A_t) = A_t U_t(1) + (1 - A_t) U_t(0)$, which allows different unobservables for different treatment state a_t. This specification is more general than (3.2) as it effectively incorporates vector unobservables. On the other hand, (3.2) delivers weak monotonicity by relying on scalar unobservable U_t. We can relax Assumption M2 by imposing (3.4) and assuming a sequential version of rank similarity (Chernozhukov and Hansen (2005)) that $\boldsymbol{U}(1, \boldsymbol{a}_{-t}) \overset{d}{=} \boldsymbol{U}(0, \boldsymbol{a}_{-t})$, conditional on $(\boldsymbol{V}^t, \boldsymbol{Z})$ for each t, where $\boldsymbol{U}(\boldsymbol{a}) \equiv (U_1(a_1), ..., U_T(a_T))$. This assumption can be found in Han (2019).[5] Note that (3.2) postulates that $U_t(a_t) = U_t$ for all $a_t \in \{0, 1\}$ and t.

4 Example: Medical Interventions

Suppose we are interested in two types of treatments for cancer: A_t indicates whether radiation therapy (R) or chemotherapy (C) is received. We want to know which treatment regime is best to improve the symptoms of cancer patients, e.g., $\boldsymbol{a} =$

[5] See Remark 5.3 of Han (2019) for more discussions on sequential rank similarity.

(R, R, R, C, C, C) or (R, C, R, C, R, C). Moreover, we want the regime to be adaptive to the history of symptoms and treatments ($\boldsymbol{d}(\boldsymbol{y}^{T-1}, \boldsymbol{a}^{T-1})$). Therefore, a medical trial is conducted by sequentially randomly assigning R or C for multiple periods. Patients, however, may deviate from the assignments (because of their habit, learning, fear, etc.), which creates noncompliance. In this example, Z_t's are assigned treatments, and A_t's are received treatments, which is endogenous due to possible noncompliance. As for the data, we have a sequence of observed symptoms, treatments, and instruments, $(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z})$. Using the data, we the planner want to find the optimal $a_t \in \{R, C\}$ in each period, given the patient's history and knowing the continuation value, i.e.,

$$d_t^*(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}) = \arg\max_{a_t} E[V_{t+1}(\boldsymbol{Y}^t(\boldsymbol{a}^t), \boldsymbol{a}^t)|\boldsymbol{Y}^{t-1}(\boldsymbol{a}^{t-1}) = \boldsymbol{y}^{t-1}].$$

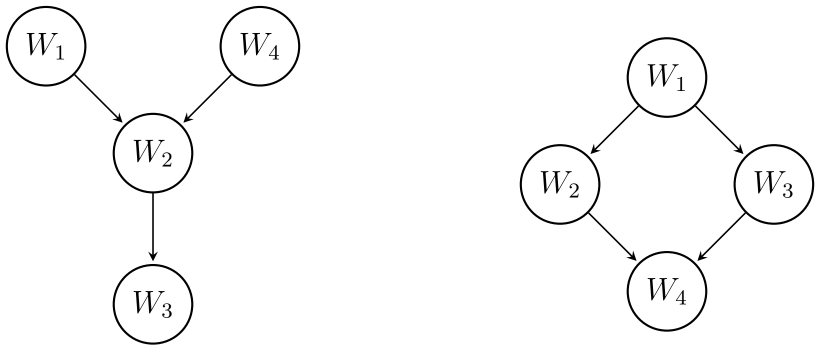
In this example, Assumption M1 states that, among patients with the same history $(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^{t-1})$, there are either no defiers or no compliers in terms of the response A_t to the assignment Z_t. Without conditional on the history, there can be both compliers and defiers in this patient population in terms of the responses $\boldsymbol{A}^t$ to the assignments $\boldsymbol{Z}^t$.

5 Partial Orderings and Partial Identification

We show how optimal dynamic regimes and related welfares can be partially recovered. In this analysis, the identified set of $\boldsymbol{d}^*(\cdot)$ will be characterized as a subset of the discrete set $\mathcal{D}$:

$$\mathcal{D}^* \subset \mathcal{D}.$$

We construct this set by establishing a *partial ordering* of $E[Y_T(\boldsymbol{d}(\cdot))]$ w.r.t. $\boldsymbol{d}(\cdot) \in \mathcal{D}$. Figure 5 illustrates examples of the *partially ordered set* of welfares, $\{E[Y_T(\boldsymbol{d}(\cdot))] : \boldsymbol{d}(\cdot) \in \mathcal{D}\}$. In this figure, we use directed acyclic graphs to depict the ordering, where each directed link "$A \to B$" indicates the relation "$A \geq B$" between two nodes (i.e., welfares) A and B. Note that in order to point identify $\boldsymbol{d}^*(\cdot)$, we need to establish a total ordering of $E[Y_T(\boldsymbol{d}(\cdot))]$. With binary instruments, this is not available in this challenging situation of partial compliance, and only a partial ordering is possible. Given $\mathcal{D}^*$, we can calculate the bounds on the optimal welfare $E[Y_T(\boldsymbol{d}^*(\cdot))]$ and the regrets.



(a) $\boldsymbol{d}^*(\cdot)$ is partially identified

(b) $\boldsymbol{d}^*(\cdot)$ is point identified

Figure 1: Partially Ordered Set, $\{W_1, W_2, W_3, W_4\}$, as Directed Acyclic Graphs

Recall $Y_t(\boldsymbol{a}^t)$ be a potential outcome with non-adaptive allocation sequence $\boldsymbol{a}^t$. Note that this is *not* the potential outcome under dynamic regime $\boldsymbol{d}^t(\cdot)$, but is useful in defining it via (2.2). We write $\boldsymbol{Y}(\boldsymbol{a}) \equiv (Y_1(a_1), Y_2(\boldsymbol{a}^2), ..., Y_T(\boldsymbol{a}^T))$ for the full vector of potential outcomes with fixed allocations. By repetitively applying the law of iterated expectation, we can show that the r.h.s. of (2.4) can be expressed as

$$\sum_{y_1} \cdots \sum_{\boldsymbol{y}^{T-2}} \sum_{\boldsymbol{y}^{T-1}} \Pr[Y_T(\boldsymbol{a}) = 1 | \boldsymbol{Y}^{T-1}(\boldsymbol{a}^{T-1}) = \boldsymbol{y}^{T-1}] \\ \times \Pr[\boldsymbol{Y}^{T-1}(\boldsymbol{a}^{T-1}) = \boldsymbol{y}^{T-1} | \boldsymbol{Y}^{T-2}(\boldsymbol{a}^{T-2}) = \boldsymbol{y}^{T-2}] \times \cdots \times \Pr[Y_1(a_1) = y_1], \tag{5.1}$$

Therefore, as a first step of our goal, we establish the ordering of the joint distribution of the potential outcomes with fixed regimes, i.e., the ordering of $\Pr[\boldsymbol{Y}(\boldsymbol{a}) = \boldsymbol{y}]$ w.r.t. $\boldsymbol{a} \in \{0,1\}^2$ for each $\boldsymbol{y} \in \{0,1\}^2$. Here we illustrate our analysis under Assumption M2, or equivalently, (3.2)–(3.3). The analysis under Assumption M1 can be analogously followed. Let $O_t \equiv (Y_t, A_t, Z_t)$ with realization o_t. Define

$$h_t(\boldsymbol{o}^{t-1}) \equiv \Pr[Y_t = 1 | Z_t = 1, \boldsymbol{o}^{t-1}] - \Pr[Y_t = 1 | Z_t = 0, \boldsymbol{o}^{t-1}], \tag{5.2}$$

which is a reduced-from quantity directly identified from the data.

Lemma 5.1. *Suppose Assumptions SX, R and M2 hold. Then, for given t and $\boldsymbol{o}^{t-1}$, the sign of $h_t(\boldsymbol{o}^{t-1})$ is equal to the sign of $Y_t(\boldsymbol{a}^{t-1}, 1) - Y_t(\boldsymbol{a}^{t-1}, 0)$ w.p.1. conditional on $\boldsymbol{O}^{t-1} = \boldsymbol{o}^{t-1}$.*

This lemma can be proved by slightly modifying the proof of Lemma 5.1 in Han (2019); see the Appendix. Lemma 5.1 is useful to establish the ordering of $\Pr[\boldsymbol{Y}(\boldsymbol{a}) = \boldsymbol{y}]$. Suppose $T = 2$ for illustration; a parallel analysis can be conducted with general T. Suppose that the data satisfy the following:

$$\begin{aligned} h_1 &> 0, \\ h_2(o_1) &> 0, \end{aligned}$$

for all o_1.[6] Then, by Lemma 5.1, $Y_1(1) > Y_1(0)$ and $Y_2(a_1, 1) > Y_2(a_1, 0)$ w.p.1 conditional on $O_1 = o_1$ for all o_1, or equivalently by Lemma (3.2), $\mu_1(1) > \mu_1(0)$ and $\mu_2(y_1, a_1, 1) > \mu_2(y_1, a_1, 0)$ for all (y_1, a_1). Using this information, we establish the following partial orderings for $\Pr[\boldsymbol{Y}(\boldsymbol{a}) = \boldsymbol{y}]$ w.r.t. $\boldsymbol{a} \in \{0, 1\}^2$ for each $\boldsymbol{y} \in \{0, 1\}^2$: For $\boldsymbol{y} = (1, 1)$,

$$\begin{aligned} &P[\boldsymbol{Y}(1, 1) = (1, 1)] \\ &\geq \max\{\boldsymbol{Y}(0, 1) = (1, 1)], P[\boldsymbol{Y}(1, 0) = (1, 1)]\} \\ &\geq \min\{\boldsymbol{Y}(0, 1) = (1, 1)], P[\boldsymbol{Y}(1, 0) = (1, 1)]\} \\ &\geq P[\boldsymbol{Y}(0, 0) = (1, 1)]. \end{aligned}$$

For $\boldsymbol{y} = (1, 0)$,

$$\begin{aligned} &P[\boldsymbol{Y}(1, 0) = (1, 0)] \\ &\geq \max\{P[\boldsymbol{Y}(1, 1) = (1, 0)], P[\boldsymbol{Y}(0, 0) = (1, 0)]\} \\ &\geq \min\{P[\boldsymbol{Y}(1, 1) = (1, 0)], P[\boldsymbol{Y}(0, 0) = (1, 0)]\} \\ &\geq P[\boldsymbol{Y}(0, 1) = (1, 0)]. \end{aligned}$$

For $\boldsymbol{y} = (0, 1)$,

$$\begin{aligned} &P[\boldsymbol{Y}(0, 1) = (0, 1)] \\ &\geq \max\{P[\boldsymbol{Y}(1, 1) = (0, 1)], P[\boldsymbol{Y}(0, 0) = (0, 1)]\} \\ &\geq \min\{P[\boldsymbol{Y}(1, 1) = (0, 1)], P[\boldsymbol{Y}(0, 0) = (0, 1)]\} \\ &\geq P[\boldsymbol{Y}(1, 0) = (0, 1)]. \end{aligned}$$

[6] Note that the strict inequality holds because of Assumption R.

Finally, for $\boldsymbol{y} = (0,0)$,

$$\begin{aligned}
&P[\boldsymbol{Y}(0,0) = (0,0)] \\
&\geq \max\{\boldsymbol{Y}(0,1) = (0,0)], P[\boldsymbol{Y}(1,0) = (0,0)]\} \\
&\geq \min\{\boldsymbol{Y}(0,1) = (0,0)], P[\boldsymbol{Y}(1,0) = (0,0)]\} \\
&\geq P[\boldsymbol{Y}(1,1) = (0,0)].
\end{aligned}$$

That is, for each $\boldsymbol{y}$, $\{\Pr[\boldsymbol{Y}(\boldsymbol{a}) = \boldsymbol{y}] : \boldsymbol{a} \in \{0,1\}^2\}$ is a partially ordered set. Note that establishing these orderings are related to identification of the signs of the ATE's (Abrevaya et al. (2010), Machado et al. (Forthcoming)). In a single-period triangular model, Bhattacharya et al. (2008) and Machado et al. (Forthcoming) discuss how threshold-crossing structure in the selection process (or equivalently, the LATE monotonicity) can help identify the sign of the ATE, without assuming monotone treatment response (MTR) of Manski and Pepper (2000). Machado et al. (Forthcoming) also shows that when threshold-crossing structure in the outcome formation process is additionally assumed, then the sign of the ATE is always identified. We show that the partial orderings above are "sharp." That is, the orderings cannot be improved based on the data-generating-process and the maintained assumptions, i.e., the orderings are the best that we can achieve. This can be shown by extending the linear programming argument in Balke and Pearl (1997), which consider treatment effects in single-period experiments with partial compliance. The sharp partial orderings of the joint distributions will later imply that the set $\mathcal{D}^*$ is sharp, i.e., it is the identified set of $\boldsymbol{d}^*(\cdot)$. We show this in Theorem 5.1 below.

Now, based on the partially ordered sets above, we can establish the partial ordering of $E[Y_2(\boldsymbol{d}(\cdot))]$ w.r.t. $\boldsymbol{d}(\cdot) = (d_1, d_2(y_1, d_1)) \in \mathcal{D}$ where

$$\begin{aligned}\mathcal{D} = \Big\{ &[d_1 = 1, d_2(1,1) = 1, d_2(0,1) = 1], \\ &[d_1 = 1, d_2(1,1) = 1, d_2(0,1) = 0], \\ &[d_1 = 1, d_2(1,1) = 0, d_2(0,1) = 1], \\ &[d_1 = 1, d_2(1,1) = 0, d_2(0,1) = 0], \\ &[d_1 = 0, d_2(1,0) = 1, d_2(0,0) = 1], \\ &[d_1 = 0, d_2(1,0) = 1, d_2(0,0) = 0], \\ &[d_1 = 0, d_2(1,0) = 0, d_2(0,0) = 1], \\ &[d_1 = 0, d_2(1,0) = 0, d_2(0,0) = 0] \Big\}, \end{aligned} \tag{5.3}$$

with the function values of each possible $\boldsymbol{d}(\cdot)$ being collected in $[\cdot]$. By (5.1), we can express

$$\begin{aligned} E[Y_2(\boldsymbol{d}(\cdot))] &= \sum_{y_1} \Pr[Y_2(d_1, d_2(y_1, d_1)) = 1 | Y_1(d_1) = y_1] \Pr[Y_1(d_1) = y_1] \\ &= \sum_{y_1} \Pr[Y_2(d_1, d_2(y_1, d_1)) = 1, Y_1(d_1) = y_1]. \end{aligned} \tag{5.4}$$

Therefore, using the partial orderings of $P[\boldsymbol{Y}(\boldsymbol{a}) = \boldsymbol{y}]$, we can establish partial ordering of the following eight objects, corresponding to the eight elements in $\mathcal{D}$, respectively:

$$\begin{aligned} &P[\boldsymbol{Y}(1,1) = (1,1)] + P[\boldsymbol{Y}(1,1) = (0,1)], \\ &P[\boldsymbol{Y}(1,1) = (1,1)] + P[\boldsymbol{Y}(1,0) = (0,1)], \\ &P[\boldsymbol{Y}(1,0) = (1,1)] + P[\boldsymbol{Y}(1,1) = (0,1)], \\ &P[\boldsymbol{Y}(1,0) = (1,1)] + P[\boldsymbol{Y}(1,0) = (0,1)], \\ &P[\boldsymbol{Y}(0,1) = (1,1)] + P[\boldsymbol{Y}(0,1) = (0,1)], \\ &P[\boldsymbol{Y}(0,1) = (1,1)] + P[\boldsymbol{Y}(0,0) = (0,1)], \\ &P[\boldsymbol{Y}(0,0) = (1,1)] + P[\boldsymbol{Y}(0,1) = (0,1)], \\ &P[\boldsymbol{Y}(0,0) = (1,1)] + P[\boldsymbol{Y}(0,0) = (0,1)] \end{aligned}$$

The orderings for the joint distributions we established above (the first and third orderings in particular) imply that $P[\boldsymbol{Y}(1,1) = (1,1)] + P[\boldsymbol{Y}(1,1) = (0,1)]$ is the largest among the first four in this list and $P[\boldsymbol{Y}(0,1)=(1,1)]+P[\boldsymbol{Y}(0,1)=(0,1)]$ is the largest among the last four. Consequently, we obtain the identified set of $\boldsymbol{d}^*(\cdot)$ as

$$\mathcal{D}^*=\{[d_1=1,d_2(1,1)=1,d_2(0,1)=1],[d_1=0,d_2(1,0)=1,d_2(0,0)=1]\}. \tag{5.5}$$

This result deserves some discussion. When $h_1>0$ and under M2, we know from Lemma 5.1 (or the existing result in a cross-sectional setting, e.g., Machado et al. (Forthcoming)) that the sign of the ATE, $\Pr[Y_1(1)=1]-\Pr[Y_1(0)=1]$, is positive. That is, $1=\arg\max_{a_1}\Pr[Y_1(a_1)=1]$. However, (5.5) shows that d_1^* is not necessarily equal to 1. This is because our goal is to maximize the average potential terminal outcome $\Pr[Y_2(\boldsymbol{d}(\cdot))=1]$ w.r.t. adaptive regimes. There are two important aspects related to this goal: (i) First, due to the adaptivity in the regime, the influence of d_1^* on the next period should be taken into account. This is clearly seen from the backward induction argument, (2.6). Here, we need to gain knowledge on $V_2(y_1,a_1)$, which can be obtained from the sign of $h_2(o_1)$. Generally, it is impossible to determine $d_t^*(\cdot)$ based solely on the sign of $h_t(\boldsymbol{o}^{t-1})$, and knowledge of the signs of $h_s(\boldsymbol{o}^{s-1})$'s for all $s\geq t$ needs to be reflected.[7] (ii) Second, even though our objective function is the *average* outcome $\Pr[Y_2(\boldsymbol{d}(\cdot))=1]$, due to the point (i), not just the mean but the entire distribution of $Y_1(a_1)$ needs to be taken into account. This can be seen from (2.6) or from (5.4). In sum, by (i) and (ii), the two elements in $\mathcal{D}^*$ are equally plausible in the current example.

Now we present a main result that generalizes this illustrative exercise. Note that $\{E[Y_T(\boldsymbol{d}(\cdot))]:\boldsymbol{d}(\cdot)\in\mathcal{D}\}$ is in general a partially ordered set. For each j that satisfies $1\leq j\leq J\leq 2^{|\mathcal{D}|}$, suppose $\mathcal{D}_j$ is a subset of $\mathcal{D}$ such that $\{E[Y_T(\boldsymbol{d}(\cdot))]:\boldsymbol{d}(\cdot)\in\mathcal{D}_j\}$ is a totally ordered set, which can be derived from an algorithm that exploits M1 or M2 (as illustrated above using Lemma 5.1). Let G be a *directed acyclic graph* (DAG) that represents the partial ordered set $\{E[Y_T(\boldsymbol{d}(\cdot))]:\boldsymbol{d}(\cdot)\in\mathcal{D}\}$, and for $1\leq j\leq J$, let G_j be a subgraph of G that is a *directed path*, representing a total ordered set of $\{E[Y_T(\boldsymbol{d}(\cdot))]:\boldsymbol{d}(\cdot)\in\mathcal{D}_j\}$. In the above example, there are two directed paths G_1 and

[7] In fact, if we are interested in the optimal non-adaptive regime $\boldsymbol{a}^*$ instead, we can recover each a_t^* by inspecting the sign of $h_t(\boldsymbol{o}^{t-1})$ only.

G_2 (with $J = 2$) and corresponding $\mathcal{D}_1$ and $\mathcal{D}_2$ are

$$\begin{aligned}\mathcal{D}_1 = \Big\{ & [d_1 = 1, d_2(1,1) = 1, d_2(0,1) = 1], \\ & [d_1 = 1, d_2(1,1) = 1, d_2(0,1) = 0], \\ & [d_1 = 1, d_2(1,1) = 0, d_2(0,1) = 1], \\ & [d_1 = 1, d_2(1,1) = 0, d_2(0,1) = 0] \Big\}, \end{aligned} \tag{5.6}$$

and

$$\begin{aligned}\mathcal{D}_2 = \Big\{ & [d_1 = 0, d_2(1,0) = 1, d_2(0,0) = 1], \\ & [d_1 = 0, d_2(1,0) = 1, d_2(0,0) = 0], \\ & [d_1 = 0, d_2(1,0) = 0, d_2(0,0) = 1], \\ & [d_1 = 0, d_2(1,0) = 0, d_2(0,0) = 0] \Big\}. \end{aligned} \tag{5.7}$$

Finally, we define $\mathcal{D}^*$ as a collection of all the maxima in directed paths:

$$\mathcal{D}^* \equiv \{\boldsymbol{d}(\cdot) : \boldsymbol{d}(\cdot) = \arg \max_{\boldsymbol{d}(\cdot) \in \mathcal{D}_j} E[Y_T(\boldsymbol{d}(\cdot))] \text{ for } 1 \leq j \leq J\}. \tag{5.8}$$

The construction of $\mathcal{D}^*$ in this definition mimics the illustrative exercise above.

Theorem 5.1. *Under Assumptions SX, R and M1 or M2, it satisfies that $\mathcal{D}^*$ is sharp.*

Now, we can calculate the bounds on the optimal welfare. Without a general result, we only continue with the illustrative example for $T = 2$ above. Let $\boldsymbol{d}_1^*(\cdot)$ and $\boldsymbol{d}_2^*(\cdot)$ be the two elements in (5.5) (in the same order), the optimal welfare is obtained as

$$\begin{aligned} E[Y_T(\boldsymbol{d}_1^*(\cdot))] &= P[\boldsymbol{Y}(1,1) = (1,1)] + P[\boldsymbol{Y}(1,1) = (0,1)] \\ &= P[Y_2(1,1) = 1] \end{aligned}$$

and

$$\begin{aligned} E[Y_T(\boldsymbol{d}_2^*(\cdot))] &= P[\boldsymbol{Y}(0,1) = (1,1)] + P[\boldsymbol{Y}(0,1) = (0,1)] \\ &= P[Y_2(0,1) = 1]. \end{aligned}$$

For the first candidate, the lower bound is

$$\begin{aligned}L(\boldsymbol{d}_1^*(\cdot)) =& P[Y_2=1, D=(1,1)] + P[Y_2=1, D=(1,0)]\\ &+ P[Y_2=1, Y_1=1, D=(0,1)] + P[Y_2=1, Y_1=1, D=(0,0)]\end{aligned}$$

and upper bound is $U(\boldsymbol{d}_1^*(\cdot)) = 1$. For the second candidate, the lower bound is

$$\begin{aligned}L(\boldsymbol{d}_2^*(\cdot)) =& P[Y_2=1, D=(0,1)] + P[Y_2=1, D=(0,0)]\\ &+ P[Y_2=1, Y_1=0, D=(1,1)] + P[Y_2=1, Y_1=0, D=(1,0)]\end{aligned}$$

and upper bound is $U(\boldsymbol{d}_2^*(\cdot)) = 1$. Therefore the lower and upper bounds on the optimal welfare are:

$$\begin{aligned}L(\boldsymbol{d}^*(\cdot)) &= \min\{L(\boldsymbol{d}_1^*(\cdot)), L(\boldsymbol{d}_2^*(\cdot))\},\\ U(\boldsymbol{d}^*(\cdot)) &= \max\{U(\boldsymbol{d}_1^*(\cdot)), U(\boldsymbol{d}_2^*(\cdot))\} = 1.\end{aligned} \tag{5.9}$$

Since the optimal welfare is defined using the dominant elements of the partially ordered sets, we may not generally be able to derive a nontrivial upper bound on the optimal welfare. Note that we can refine $\mathcal{D}^*$ in (5.5) above by comparing $L(\boldsymbol{d}_1^*(\cdot))$ and $L(\boldsymbol{d}_2^*(\cdot))$ as long as we are willing to assume that the social planner is conservative in a minimax sense. That is, by minimizing the maximum possible mistake of inferring the optimal welfare, we can uniquely choose $\boldsymbol{d}^*(\cdot)$ by

$$\boldsymbol{d}^*(\cdot) = \arg\min_{\boldsymbol{d}_2^*(\cdot), \boldsymbol{d}_1^*(\cdot)} \{-L(\boldsymbol{d}_1^*(\cdot)), -L(\boldsymbol{d}_2^*(\cdot))\}.$$

Obviously, the lower bound on the optimal welfare under this regime is identical to $L(\boldsymbol{d}^*(\cdot))$ in (5.9).

Before closing this section, note that in this partial identification analysis, we only used the variation of the instruments $\boldsymbol{Z} \in \{0,1\}^T$. Nonetheless, the identified set is sometimes a singleton when $T = 2$ (even without the minimax refinement discussed above); see Appendix A for this case. When $T > 2$, this situation may be less likely to occur. Even if the identified set may not be small in those cases, we can still suggest the social planner remove suboptimal regimes $\boldsymbol{d}^\circ(\cdot)$ s.t.

$$E[Y_T(\boldsymbol{d}^\circ(\cdot))] \leq E[Y_T(\boldsymbol{d}(\cdot))]$$

for some $\boldsymbol{d}(\cdot)$. This may still be a useful policy recommendation. When we have extra exogenous variables, we may be able to shrink $\mathcal{D}^*$ further, which is considered later.

6 Directed Acyclic Graphs and Topological Sorting

When $T \geq 3$, it may be too cumbersome to fine $\boldsymbol{d}^*(\cdot)$ analytically. Using the DAG representation introduced in the previous section, we propose to use the *topological sorting* in the computer science literature (e.g., Kahn (1962)). The topological sorting of a DAG is a linear ordering of its vertices such that for every directed edge $W_l \to W_m$, W_l comes before W_m in the ordering. Let K be the number of all possible topological sorts of G that corresponds to the partially ordered set $\{E[Y_T(\boldsymbol{d}(\cdot))] : \boldsymbol{d}(\cdot) \in \mathcal{D}\}$. For $1 \leq k \leq K$, let $\mathcal{S}_k$ be a topological sort of G:

$$\mathcal{S}_k \equiv \{E[Y_T(\boldsymbol{d}_{k,1}(\cdot))], E[Y_T(\boldsymbol{d}_{k,2}(\cdot))], ..., E[Y_T(\boldsymbol{d}_{k,|\mathcal{D}|}(\cdot))]\}.$$

The following theorem is useful to justify the use of the topological sorting to find $\boldsymbol{d}^*(\cdot)$.

Theorem 6.1. *$\mathcal{D}^*$ defined in (5.8) is equivalent to*

$$\{\boldsymbol{d}_{k,1}(\cdot) : 1 \leq k \leq K\}.$$

Based on this theorem, $\mathcal{D}^*$ can be obtained by finding all topological sorts of G and collect $\boldsymbol{d}(\cdot)$ that produces the first element in each $\mathcal{S}_k$. There are well-known algorithms that efficiently find topological sorts, such as Kahn (1962)'s algorithm.

By definition, $\mathcal{S}_k$ is a linear extension of the partial ordering where $E[Y_T(\boldsymbol{d}_{k,l}(\cdot))]$ *cannot* be larger than $E[Y_T(\boldsymbol{d}_{k,m}(\cdot))]$ as long as $l < m$. Therefore, not just $\boldsymbol{d}_{k,1}(\cdot)$ but the full sequence

$$\{\boldsymbol{d}_{k,1}(\cdot), \boldsymbol{d}_{k,2}(\cdot), ..., \boldsymbol{d}_{k,|\mathcal{D}|}(\cdot)\}$$

can be a useful policy benchmark as it provides a ranking of regimes that is not inconsistent to welfare maximization. All possible such sequences for $1 \leq k \leq K$ are the menu of benchmarks a policy maker can be equipped with, and they are reported as outputs in those algorithms.

7 Cardinality Reduction

The typical time horizons we consider in this paper are short, say, $T \leq 5$. For example, a multi-period experiment called the Fast Track Prevention Program (Conduct Problems Prevention Research Group (1992)) considers $T = 4$. A secondary school interventions for college admission has $T = 3$. When T is not small, the size of $\mathcal{D}$ may be too large and we may want to consider reducing its dimension for computational, institutional, and practical purposes.

One way to reduce the cardinality is to reduce the dimension of the adaptivity. That is, we define a simpler adaptive treatment rule $\tilde{d}_t : \{0,1\} \times \{0,1\} \to \{0,1\}$ that maps only the lagged outcome and treatment onto a treatment allocation $a_t \in \{0,1\}$:

$$\tilde{d}_t(y_{t-1}, a_{t-1}) = a_t$$

in the class $\tilde{\mathcal{D}}$, or an even simpler rule, $\tilde{d}_t(y_{t-1}) = a_t$. The latter rule appears in Murphy et al. (2001).

Another possibility is to consider a strict subset of $\mathcal{D}$, motivated by institutional constraints. For example, it may be the case that adaptive allocation is available every second period or only later in the horizon due to cost consideration. For example, suppose that the social planner decides to introduce the adaptive rule at $t = T$ while maintaining non-adaptive rules for $t \leq T - 1$. Then, we reduce $|\mathcal{D}| = 2^{2^T - 1}$ to $|\mathcal{D}| = 2 \times 2 \times \cdots \times 2 \times (2^{T-1} \cdot 2) = 2^{2T-1}$.

8 Additional Exogenous Variables

When there exists exogenous variation in the model in addition to the excluded instruments, we show that the identified set for $\boldsymbol{d}^*(\cdot)$ can be shrunk. Introduce a model that extends (3.2)–(3.3):

$$Y_t = 1\{\mu_t(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^t, X_t) \geq U_t\}, \tag{8.1}$$

$$A_t = 1\{\pi_t(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^t) \geq V_t\}. \tag{8.2}$$

Note that we can consider an extension of Assumption M2 that is equivalent to (8.1)–(8.2), but we use the latter in the subsequent analysis for convenience. In this extended model, Z_t is a vector that contains binary instrument and X_t is set of exogenous variables that may not included in Z_t. A slightly restrictive model similar to (8.1)–(8.2) is considered in Han (2019). The existence of X_t can be guaranteed by making a behavioral/information assumption. That is, we assume that there exist outcome-determining factors that agent cannot fully anticipate when making treatment decision or earlier. We argue that this may be plausible in dynamic settings.

Redefine the adaptive rule as

$$d_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, \boldsymbol{x}^{t-1}) = a_t.$$

Then, the optimal dynamic regime is defined as

$$\boldsymbol{d}^*(\cdot) = \arg\max_{\boldsymbol{d}(\cdot)\in\mathcal{D}} E[Y_T(\boldsymbol{d}(\cdot))]$$

with

$$\begin{aligned} &E[Y_T(\boldsymbol{d}(\cdot))] \\ =&E\left[E\left[\cdots E\left[E[Y_T(\boldsymbol{a})|\boldsymbol{Y}^{T-1}(\boldsymbol{a}^{T-1}), \boldsymbol{X}^{T-1}]\right| \boldsymbol{Y}^{T-2}(\boldsymbol{a}^{T-2}), \boldsymbol{X}^{T-2}\right]\cdots \right| Y_1(a_1), X_1\right]\right], \end{aligned}$$

where $\boldsymbol{a} = (a_1, ..., a_T)$ satisfies

$$\begin{aligned} a_1 &= d_1, \\ a_2 &= d_2(Y_1(a_1), a_1, X_1), \\ a_3 &= d_3(\boldsymbol{Y}^2(\boldsymbol{a}^2), \boldsymbol{a}^2, \boldsymbol{X}^2), \\ &\vdots \\ a_T &= d_T(\boldsymbol{Y}^{T-1}(\boldsymbol{a}^{T-1}), \boldsymbol{a}^{T-1}, \boldsymbol{X}^{T-1}). \end{aligned}$$

Suppose Z_t and X_t have no common element. We modify Assumption SX as follows:

Assumption SX′. $(\boldsymbol{Y}(\boldsymbol{a}), \boldsymbol{A}(\boldsymbol{z})) \perp (\boldsymbol{Z}, \boldsymbol{X})$.

Redefine $O_t \equiv (Y_t, A_t, X_t, Z_t)$ to be the vector of the observables with realization

o_t, and define

$$\begin{aligned} h_t(x_t, x_t'; \boldsymbol{o}^{t-1}) \equiv & \Pr[Y_t = 1, A_t = 1 | Z_t = 1, x_t, \boldsymbol{o}^{t-1}] \\ & - \Pr[Y_t = 1, A_t = 1 | Z_t = 0, x_t, \boldsymbol{o}^{t-1}] \\ & + \Pr[Y_t = 1, A_t = 0 | Z_t = 1, x_t', \boldsymbol{o}^{t-1}] \\ & - \Pr[Y_t = 1, A_t = 0 | Z_t = 0, x_t', \boldsymbol{o}^{t-1}]. \end{aligned}$$

For given t and $\boldsymbol{o}^{t-1}$, the sign of $h_t(x_t, x_t'; \boldsymbol{o}^{t-1})$ is equal to the sign of

$$\mu_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, 1, x_t) - \mu_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, 0, x_t')$$

by a slight modification of Lemma 5.1 in Han (2019). Recall

$$\begin{aligned} & E[Y_T(\boldsymbol{d}(\cdot))] \\ =& E\left[E\left[\cdots E\left[E[Y_T(\boldsymbol{a}) | \boldsymbol{Y}^{T-1}(\boldsymbol{a}^{T-1}), \boldsymbol{X}^{T-1}] \middle| \boldsymbol{Y}^{T-2}(\boldsymbol{a}^{T-2}), \boldsymbol{X}^{T-2} \right] \cdots \middle| Y_1(a_1), X_1 \right] \right]. \end{aligned}$$

Therefore, the knowledge on the sign of $\mu_t(\boldsymbol{y}^{t-1}, 1, x_t) - \mu_t(\boldsymbol{y}^{t-1}, 0, x_t')$ will establish the partial orderings for $\Pr[\boldsymbol{Y}^t(\boldsymbol{a}) = \boldsymbol{y}^t | \boldsymbol{x}^{t-1}]$ w.r.t. $\boldsymbol{a} \in \{0,1\}^2$ for each $\boldsymbol{y} \in \{0,1\}^2$ and $\boldsymbol{x}^{t-1}$. This will improve the identified set $\mathcal{D}^*$ of optimal regimes.

9 Discussions: Estimation and Inference

Estimation can be done by the sample analog principle. As for estimation of $\mathcal{D}^*$, we simply need to estimate

$$h_t(\boldsymbol{o}^{t-1}) \equiv \Pr[Y_t = 1 | Z_t = 1, \boldsymbol{o}^{t-1}] - \Pr[Y_t = 1 | Z_t = 0, \boldsymbol{o}^{t-1}],$$

which amounts to calculating the difference of sample means. Estimating $h_t(x_t, x_t'; \boldsymbol{o}^{t-1})$ involves estimating nonparametric mean functions when X_t is continuously distributed. Similarly, estimation of the bounds on $E[Y_T(\boldsymbol{d}^*(\cdot))]$ involves estimation of (conditional) probabilities.

Inference on $\boldsymbol{d}^*(\cdot)$ is a more challenging and interesting problem, and is related to inference on $E[Y_T(\boldsymbol{d}(\cdot))]$'s. To construct a (discrete) confidence set (CS) for $\boldsymbol{d}^*(\cdot)$, we can invert the following test:

$$H_0 : E[Y_T(\boldsymbol{d}(\cdot))] \geq \max_{\tilde{\boldsymbol{d}}(\cdot)\in\mathcal{D}} E[Y_T(\tilde{\boldsymbol{d}}(\cdot))],$$

which is related to testing with moment inequalities, since the bounds on $E[Y_T(\boldsymbol{d}(\cdot))]$ are written in terms of moment inequalities. A resulting CS can also be used for a specification test for a less palatable assumption such as Assumption M2. That is, when the CS under M2 is empty, we can reject the assumption. Inference on optimized welfare $E[Y_T(\boldsymbol{d}^*(\cdot))]$ can also be an interesting problem. Andrews et al. (2019) considers inference on optimized welfare in the context of Kitagawa and Tetenov (2018), but with point identified welfare under the unconfoundedness assumption for treatment. Extending the framework to a multi-period setting with partially identified welfare with dynamic regimes will be interesting future work.

A Appendix

A.1 Proof of Lemma 3.1

Conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^{t-1}) = (\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, \boldsymbol{z}^{t-1})$, it is easy to show that (3.1) implies Assumption M1. Suppose $\pi_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, \boldsymbol{z}^{t-1}, 1) > \pi_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, \boldsymbol{z}^{t-1}, 1)$ as $\pi_t(\cdot)$ is a nontrivial function of Z_t. Then, we have

$$1\{\pi_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, \boldsymbol{z}^{t-1}, 1) \geq V_t\} \geq 1\{\pi_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, \boldsymbol{z}^{t-1}, 0) \geq V_t\}$$

w.p.1, or equivalently, $A_t(\boldsymbol{z}^{t-1}, 1) \geq A_t(\boldsymbol{z}^{t-1}, 0)$ w.p.1. Suppose $\pi_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, \boldsymbol{z}^{t-1}, 1) < \pi_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, \boldsymbol{z}^{t-1}, 1)$. Then, by a parallel argument, $A_t(\boldsymbol{z}^{t-1}, 1) \leq A_t(\boldsymbol{z}^{t-1}, 0)$ w.p.1.

Now, we show that Assumption M1 implies (3.1) conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^{t-1})$. For each t, Assumption SX implies $Y_t(\boldsymbol{a}^t), A_t(\boldsymbol{z}^t) \perp \boldsymbol{Z}^t | (\boldsymbol{Y}^{t-1}(\boldsymbol{a}^{t-1}), \boldsymbol{A}^{t-1}(\boldsymbol{z}^{t-1}), \boldsymbol{Z}^{t-1})$, which in turn implies the following conditional independence:

$$Y_t(\boldsymbol{a}^t), A_t(\boldsymbol{z}^t) \perp \boldsymbol{Z}^t | (\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^{t-1}). \tag{A.1}$$

Conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^{t-1})$, (3.1) and (A.1) correspond to Assumption S-1 in Vytlacil (2002). Assumption R(i) and (A.1) correspond to Assumption L-1, and Assumption M1 corresponds to Assumption L-2 in Vytlacil (2002). Therefore, the desired result follows by Theorem 1 of Vytlacil (2002). □

A.2 Proof of Lemma 3.2

We are remained to prove that, conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^{t-1})$, (3.2) is equivalent to the second part of Assumption M2. But this proof is analogous to the proof of Lemma 3.1 by replacing the roles of A_t and Z_t with those of Y_t and A_t, respectively. Therefore, we have the desired result. □

A.3 Proof of Lemma 5.1

Since Assumption M2 is equivalent to (3.2)–(3.3) conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^{t-1})$ by Lemma 3.2, we prove Lemma 5.1 using (3.2)–(3.3). We first define the sets of unobservables in the model. Define

$$\mathcal{U}^t(\boldsymbol{a}^t, \boldsymbol{y}^t) \equiv \{\boldsymbol{U}^t : y_s = Y_s(\boldsymbol{a}^s) \text{ for all } s \le t\}.$$

for $t \ge 1$. Then, $\boldsymbol{Y}^t = \boldsymbol{y}^t$ if and only if $\boldsymbol{U}^t \in \mathcal{U}^t(\boldsymbol{a}^t, \boldsymbol{y}^t)$, conditional on $\boldsymbol{A}^t = \boldsymbol{a}^t$. Realizing the dependence of $Y_{s-1}(\boldsymbol{a}^{s-1})$ on $(\boldsymbol{U}^{s-1}, \boldsymbol{a}^{s-1})$, let

$$\pi_s^*(\boldsymbol{U}^{s-1}, \boldsymbol{a}^{s-1}, \boldsymbol{z}^s) \equiv \pi_s(\boldsymbol{Y}^{s-1}(\boldsymbol{a}^{s-1}), \boldsymbol{a}^{s-1}, \boldsymbol{z}^s),$$

and define the set of $\boldsymbol{V}^t$ as

$$\mathcal{V}^t(\boldsymbol{a}^t, \boldsymbol{u}^{t-1}) \equiv \mathcal{V}^t(\boldsymbol{a}^t, \boldsymbol{u}^{t-1}; \boldsymbol{z}^t) \equiv \{\boldsymbol{V}^t : a_s = 1\{V_s \le \pi_s^*(\boldsymbol{u}^{s-1}, \boldsymbol{a}^{s-1}, \boldsymbol{z}^s)\} \text{ for all } s \le t\}$$

for $t \ge 2$. Fix $t \ge 3$. Recall $o_t = (y_t, a_t, z_t)$. Since $\pi_t(\cdot)$ is a non-trivial function of A_t, consider the case $\Pr[A_t = 1 | Z_t = 1, \boldsymbol{o}^{t-1}] > \Pr[A_t = 1 | Z_t = 0, \boldsymbol{o}^{t-1}]$; the opposite case is symmetric. Using the definitions of the sets above, we have

$$\begin{aligned}
&\Pr[A_t = 1 | z_t, \boldsymbol{o}^{t-1}] \\
=&\Pr[V_t \le \pi_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, \boldsymbol{z}^t) | \boldsymbol{z}^t, \mathcal{V}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{U}^{t-2}), \mathcal{U}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{y}^{t-1})] \\
=&\Pr[V_t \le \pi_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, \boldsymbol{z}^t) | \boldsymbol{z}^{t-1}, \mathcal{V}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{U}^{t-2}), \mathcal{U}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{y}^{t-1})],
\end{aligned}$$

where the first equality is by Lemma 3.2 and the last equality is by Assumption SX and because of the following: conditional on $\boldsymbol{Z}^t = \boldsymbol{z}^t$, (i) $\boldsymbol{Y}^{t-1} = \boldsymbol{y}^{t-1}$ is equivalent to $\boldsymbol{U}^{t-1} \in \mathcal{U}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{y}^{t-1})$ conditional on $\boldsymbol{A}^{t-1} = \boldsymbol{a}^{t-1}$; (ii) $\boldsymbol{A}^{t-1} = \boldsymbol{a}^{t-1}$ is equivalent of $\boldsymbol{V}^{t-1} \in \mathcal{V}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{U}^{t-2})$ conditional on $\boldsymbol{Y}^{t-1} = \boldsymbol{y}^{t-1}$. Note that the sets $\mathcal{V}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{U}^{t-2})$ and $\mathcal{U}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{y}^{t-1})$ do not change with the change in z_t. Therefore, with $\pi_t^1 \equiv \pi_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, \boldsymbol{z}^{t-1}, 1)$ and $\pi_t^0 \equiv \pi_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, \boldsymbol{z}^{t-1}, 0)$,

$$\begin{aligned}
0 <{} & \Pr[A_t = 1 | Z_t = 1, \boldsymbol{o}^{t-1}] - \Pr[A_t = 1 | Z_t = 0, \boldsymbol{o}^{t-1}] \\
={} & \Pr[V_t \leq \pi_t^1 | \boldsymbol{z}^{t-1}, \mathcal{V}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{U}^{t-2}), \mathcal{U}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{y}^{t-1})] \\
& - \Pr[V_t \leq \pi_t^0 | \boldsymbol{z}^{t-1}, \mathcal{V}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{U}^{t-2}), \mathcal{U}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{y}^{t-1})],
\end{aligned}$$

which implies $\pi_t^1 > \pi_t^0$. Next, we have

$$\begin{aligned}
& \Pr[Y_t = 1, A_t = 1 | z_t, \boldsymbol{o}^{t-1}] \\
={} & \Pr[U_t \leq \mu_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, 1), V_t \leq \pi_t | \boldsymbol{z}^{t-1}, \mathcal{V}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{U}^{t-2}), \mathcal{U}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{y}^{t-1})]
\end{aligned}$$

by Assumption SX and Lemma 3.2. Again, note that $\mathcal{V}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{U}^{t-2})$ and $\mathcal{U}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{y}^{t-1})$ do not change with the change in z_t. Therefore, similar expressions can be derived for the other terms involved in h_t, and we have

$$\begin{aligned}
& h_t(\boldsymbol{o}^{t-1}) \\
={} & \Pr[U_t \leq \mu_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, 1), \pi_t^0 \leq V_t \leq \pi_t^1 | \mathcal{V}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{U}^{t-2}), \mathcal{U}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{y}^{t-1})] \\
& - \Pr[U_t \leq \mu_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, 0), \pi_t^0 \leq V_t \leq \pi_t^1 | \mathcal{V}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{U}^{t-2}), \mathcal{U}^{t-1}(\boldsymbol{a}^{t-1}, \boldsymbol{y}^{t-1})],
\end{aligned}$$

the sign of which identifies the sign of $\mu_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, 1) - \mu_t(\boldsymbol{y}^{t-1}, \boldsymbol{a}^{t-1}, 0)$. Therefore, the sign of $Y_t(\boldsymbol{a}^{t-1}, 1) - Y_t(\boldsymbol{a}^{t-1}, 0)$ is identified w.p.1 conditional on $\boldsymbol{O}^{t-1} = \boldsymbol{o}^{t-1}$ by Lemma 3.2. The case $t \leq 2$ can be shown analogously with $\mathcal{V}^1(a_1) \equiv \mathcal{V}^1(a_1; z_1) \equiv \{V_1 : a_1 = 1\{V_1 \leq \pi_1(z_1)\}\}$. $\square$

A.4 Proof of Theorem 5.1

As the first step of this proof, we show that the ordering of $\Pr[\boldsymbol{Y}(\boldsymbol{a}) = \boldsymbol{y}]$ w.r.t. $\boldsymbol{a} \in \{0,1\}^2$ is sharp for each $\boldsymbol{y} \in \{0,1\}^2$. Under Assumptions SX, R and M1 or M2, for any given $\boldsymbol{y} \in \{0,1\}^2$, the ATE, $\Pr[Y(a_t, \boldsymbol{a}_{-t}) = \boldsymbol{y}] - \Pr[Y(\tilde{a}_t, \boldsymbol{a}_{-t}) = \boldsymbol{y}]$, has sharp bounds for any $a_t \neq \tilde{a}_t$ and t, by applying the linear programming method as in Theorems 3.1 and 3.2 in Machado et al. (Forthcoming) or, generally, in Balke and Pearl (1997). That is, either $\Pr[Y(a_t, \boldsymbol{a}_{-t}) = \boldsymbol{y}] \geq \Pr[Y(\tilde{a}_t, \boldsymbol{a}_{-t}) = \boldsymbol{y}]$, $\Pr[Y(a_t, \boldsymbol{a}_{-t}) =$

$\boldsymbol{y}] \le \Pr[Y(\tilde{a}_t, \boldsymbol{a}_{-t}) = \boldsymbol{y}]$ or the incomparableness between $\Pr[Y(a_t, \boldsymbol{a}_{-t}) = \boldsymbol{y}]$ and $\Pr[Y(\tilde{a}_t, \boldsymbol{a}_{-t}) = \boldsymbol{y}]$ is sharp, depending on whether the sharp ATE bounds contains zero or not and what the sign of the ATE is. Therefore, the partial ordering of $\Pr[\boldsymbol{Y}(\boldsymbol{a}) = \boldsymbol{y}]$ w.r.t. $\boldsymbol{a} \in \{0,1\}^2$ is sharp for each $\boldsymbol{y} \in \{0,1\}^2$. This in turn implies that the ordering of any conditional or marginal distribution derived from $\Pr[\boldsymbol{Y}(\boldsymbol{a}) = \boldsymbol{y}]$ is also sharp. Therefore, the ordering of $E[Y_T(\boldsymbol{d}(\cdot))]$ w.r.t. $\boldsymbol{d}(\cdot) \in \mathcal{D}$ is sharp due to (2.4) and (5.1), which in turn implies that, for each $1 \le j \le J$, the total ordering of $E[Y_T(\boldsymbol{d}(\cdot))]$ w.r.t. $\boldsymbol{d}(\cdot) \in \mathcal{D}_j$ is sharp. Consequently, $\mathcal{D}^* \equiv \{\boldsymbol{d}(\cdot) : \boldsymbol{d}(\cdot) = \arg\max_{\boldsymbol{d}(\cdot) \in \mathcal{D}_j} E[Y_T(\boldsymbol{d}(\cdot))] \text{ for } 1 \le j \le J\}$ is sharp. □

A.5 Proof of Theorem 6.1

When all topological sorts are singletons, the proof is trivial so we rule out this possibility. Suppose $\mathcal{D}^* \supset \{\boldsymbol{d}_{k,1}(\cdot) : 1 \le k \le K\}$. Then, for some k, there should exist $\boldsymbol{d}_{k,l}(\cdot)$ for some $l \ne 1$ that is contained in $\mathcal{D}^*$ but not in $\{\boldsymbol{d}_{k,1}(\cdot) : 1 \le k \le K\}$, i.e., that satisfies either (i) $E[Y_T(\boldsymbol{d}_{k,1}(\cdot))] > E[Y_T(\boldsymbol{d}_{k,l}(\cdot))]$ for some l or (ii) $E[Y_T(\boldsymbol{d}_{k,1}(\cdot))]$ and $E[Y_T(\boldsymbol{d}_{k,l}(\cdot))]$ are incomparable and thus either $E[Y_T(\boldsymbol{d}_{k',1}(\cdot))] > E[Y_T(\boldsymbol{d}_{k,l}(\cdot))]$ for some $k' \ne k$ or $E[Y_T(\boldsymbol{d}_{k,l}(\cdot))]$ is a singleton in another topological sort. Consider case (i). If $\boldsymbol{d}_{k,1}(\cdot) \in \mathcal{D}_j$ for some j, then it should be that $\boldsymbol{d}_{k,l}(\cdot) \in \mathcal{D}_j$ as $\boldsymbol{d}_{k,1}(\cdot)$ and $\boldsymbol{d}_{k,l}(\cdot)$ are comparable in terms of welfare, but then $\boldsymbol{d}_{k,l}(\cdot) \in \mathcal{D}^*$ contradicts the fact that $\boldsymbol{d}_{k,1}(\cdot)$ delivers the largest welfare. Consider case (ii). The singleton case is trivially rejected since if the topological sort a singleton, then $\boldsymbol{d}_{k,l}(\cdot)$ should have been already in $\{\boldsymbol{d}_{k,1}(\cdot) : 1 \le k \le K\}$. In the other case, since the two welfares are not comparable, it should be that $\boldsymbol{d}_{k,l}(\cdot) \in \mathcal{D}_{j'}$ for $j' \ne j$. But $\boldsymbol{d}_{k,l}(\cdot)$ cannot be the one that delivers the largest welfare since $E[Y_T(\boldsymbol{d}_{k',1}(\cdot))] > E[Y_T(\boldsymbol{d}_{k,l}(\cdot))]$ where $\boldsymbol{d}_{k',1}(\cdot) \in \mathcal{D}_{j'}$. Therefore $\boldsymbol{d}_{k,l}(\cdot) \in \mathcal{D}^*$ is contradiction. Therefore there is no element in $\mathcal{D}^*$ that is not in $\{\boldsymbol{d}_{k,1}(\cdot) : 1 \le k \le K\}$.

Now suppose $\mathcal{D}^* \subset \{\boldsymbol{d}_{k,1}(\cdot) : 1 \le k \le K\}$. Then for k such that $\boldsymbol{d}_{k,1}(\cdot) \notin \mathcal{D}^*$, either $E[Y_T(\boldsymbol{d}_{k,1}(\cdot))]$ is a singleton or $E[Y_T(\boldsymbol{d}_{k,1}(\cdot))]$ is an element in a non-singleton topological sort. But if it is a singleton, then it is trivially totally ordered and is the maximum welfare, and thus $\boldsymbol{d}_{k,1}(\cdot) \notin \mathcal{D}^*$ is contradiction. In the other case, if $E[Y_T(\boldsymbol{d}_{k,1}(\cdot))]$ is a maximum welfare, then $\boldsymbol{d}_{k,1}(\cdot) \notin \mathcal{D}^*$ is contradiction. If it is not a maximum welfare, then it should be a maximum in another topological sort, which is contradiction in either case of being contained in $\{\boldsymbol{d}_{k,1}(\cdot) : 1 \le k \le K\}$ or not. This concludes the proof that $\mathcal{D}^* = \{\boldsymbol{d}_{k,1}(\cdot) : 1 \le k \le K\}$. □

A.6 Point Identified Case

We consider another case with $T = 2$, where $\boldsymbol{d}^*(\cdot)$ is point identified. Suppose that the data satisfy the following:

$$h_1 > 0, \qquad h_2(1, a_1, z_1) > 0, \qquad h_2(0, a_1, z_1) < 0$$

for all (a_1, z_1). Then, we have

$$\mu_1(1) > \mu_1(0), \qquad \mu_2(1, a_1, 1) > \mu_2(1, a_1, 0), \qquad \mu_2(0, a_1, 1) < \mu_2(0, a_1, 0)$$

for all a_1. Based on this knowledge, we establish the partial orderings for $\Pr[\boldsymbol{Y}(\boldsymbol{a}) = \boldsymbol{y}]$ w.r.t. $\boldsymbol{a} \in \{0, 1\}^2$ for each $\boldsymbol{y} \in \{0, 1\}^2$: For $\boldsymbol{y} = (1, 1)$, since $P[\boldsymbol{Y}(\boldsymbol{a}) = (1, 1)] = P[U_1 \leq \mu_1(y_0, a_1), U_2 \leq \mu_2(1, a_2)]$,

$$\begin{aligned}
&P[\boldsymbol{Y}(1, 1) = (1, 1)] \\
&\geq \max\left\{\boldsymbol{Y}(0, 1) = (1, 1)], P[\boldsymbol{Y}(1, 0) = (1, 1)]\right\} \\
&\geq \min\left\{\boldsymbol{Y}(0, 1) = (1, 1)], P[\boldsymbol{Y}(1, 0) = (1, 1)]\right\} \\
&\geq P[\boldsymbol{Y}(0, 0) = (1, 1)].
\end{aligned}$$

For $\boldsymbol{y} = (1, 0)$, since $P[\boldsymbol{Y}(\boldsymbol{a}) = (1, 0)] = P[U_1 \leq \mu_1(y_0, a_1), U_2 > \mu_2(1, a_2)]$,

$$\begin{aligned}
&P[\boldsymbol{Y}(1, 0) = (1, 0)] \\
&\geq \max\left\{P[\boldsymbol{Y}(1, 1) = (1, 0)], P[\boldsymbol{Y}(0, 0) = (1, 0)]\right\} \\
&\geq \min\left\{P[\boldsymbol{Y}(1, 1) = (1, 0)], P[\boldsymbol{Y}(0, 0) = (1, 0)]\right\} \\
&\geq P[\boldsymbol{Y}(0, 1) = (1, 0)].
\end{aligned}$$

For $\boldsymbol{y} = (0, 1)$, since $P[\boldsymbol{Y}(\boldsymbol{a}) = (0, 1)] = P[U_1 > \mu_1(y_0, a_1), U_2 \leq \mu_2(0, a_2)]$,

$$\begin{aligned}
&P[\boldsymbol{Y}(1, 0) = (0, 1)] \\
&\geq \max\left\{P[\boldsymbol{Y}(0, 0) = (0, 1)], P[\boldsymbol{Y}(1, 1) = (0, 1)]\right\} \\
&\geq \min\left\{P[\boldsymbol{Y}(0, 0) = (0, 1)], P[\boldsymbol{Y}(1, 1) = (0, 1)]\right\} \\
&\geq P[\boldsymbol{Y}(0, 1) = (0, 1)].
\end{aligned}$$

Finally, for $\boldsymbol{y} = (0,0)$, since $P[\boldsymbol{Y}(\boldsymbol{a}) = (0,0)] = P[U_1 > \mu_1(y_0, a_1), U_2 > \mu_2(0, a_2)]$,

$$\begin{aligned}
&P[\boldsymbol{Y}(0,1) = (0,0)] \\
&\geq \max\{\boldsymbol{Y}(0,0) = (0,0)], P[\boldsymbol{Y}(1,1) = (0,0)]\} \\
&\geq \min\{\boldsymbol{Y}(0,0) = (0,0)], P[\boldsymbol{Y}(1,1) = (0,0)]\} \\
&\geq P[\boldsymbol{Y}(1,0) = (0,0)].
\end{aligned}$$

Note that the first two orderings are the same as before. Based on the partially ordered sets above, we can establish the partial ordering of

$$E[Y_2(\boldsymbol{d}(\cdot))] = \sum_{y_1} \Pr[Y_2(d_1, d_2(y_1, d_1)) = 1, Y_1(d_1) = y_1]$$

w.r.t. $\boldsymbol{d}(\cdot) = (d_1, d_2(y_1, d_1)) \in \mathcal{D}$, that is, we can establish partial ordering of

$$\begin{aligned}
&P[\boldsymbol{Y}(1,1) = (1,1)] + P[\boldsymbol{Y}(1,1) = (0,1)], \\
&P[\boldsymbol{Y}(1,1) = (1,1)] + P[\boldsymbol{Y}(1,0) = (0,1)], \\
&P[\boldsymbol{Y}(1,0) = (1,1)] + P[\boldsymbol{Y}(1,1) = (0,1)], \\
&P[\boldsymbol{Y}(1,0) = (1,1)] + P[\boldsymbol{Y}(1,0) = (0,1)], \\
&P[\boldsymbol{Y}(0,1) = (1,1)] + P[\boldsymbol{Y}(0,1) = (0,1)], \\
&P[\boldsymbol{Y}(0,1) = (1,1)] + P[\boldsymbol{Y}(0,0) = (0,1)], \\
&P[\boldsymbol{Y}(0,0) = (1,1)] + P[\boldsymbol{Y}(0,1) = (0,1)], \\
&P[\boldsymbol{Y}(0,0) = (1,1)] + P[\boldsymbol{Y}(0,0) = (0,1)].
\end{aligned}$$

The orderings for the joint distributions we established above imply that $P[\boldsymbol{Y}(1,1) = (1,1)|\boldsymbol{x}] + P[\boldsymbol{Y}(1,0) = (0,1)]$ is the largest among them. Consequently, the identified set of $\boldsymbol{d}^*(\cdot)$ is obtained as

$$\mathcal{D}^* = \{[d_1 = 1, d_2(1,1) = 1, d_2(0,1) = 0]\},$$

which is in fact a singleton in this example, i.e., $\boldsymbol{d}^*(\cdot)$ is point identified. Now, we calculate the bounds on the optimal welfare. Given $\mathcal{D}^*$, the optimal welfare is obtained as

$$E[Y_T(\boldsymbol{d}^*(\cdot))] = P[\boldsymbol{Y}(1,1) = (1,1)] + P[\boldsymbol{Y}(1,0) = (0,1)],$$

whose lower and upper bounds can be calculated as before. □

References

Abrevaya, J., J. A. Hausman, and S. Khan (2010): "Testing for causal effects in a generalized regression model with endogenous regressors," *Econometrica*, 78, 2043–2061. 5

Andrews, I., T. Kitagawa, and A. McCloskey (2019): "Inference on winners," Tech. rep., National Bureau of Economic Research. 9

Balat, J. and S. Han (2018): "Multiple treatments with strategic interaction," *UT Austin*. 1

Balke, A. and J. Pearl (1997): "Bounds on treatment effects from studies with imperfect compliance," *Journal of the American Statistical Association*, 92, 1171–1176. 1, 5, A.4

Bhattacharya, J., A. M. Shaikh, and E. Vytlacil (2008): "Treatment effect bounds under monotonicity assumptions: An application to swan-ganz catheterization," *The American Economic Review*, 98, 351–356. 5

Chernozhukov, V. and C. Hansen (2005): "An IV model of quantile treatment effects," *Econometrica*, 73, 245–261. 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. 7

Han, S. (2019): "Nonparametric Identification in Models for Dynamic Treatment Effects," *UT Austin*. 1, 2, 3, 3, 5, 5, 8

Heckman, J. J., J. E. Humphries, and G. Veramendi (2016): "Dynamic treatment effects," *Journal of Econometrics*, 191, 276–292. 1

Heckman, J. J. and S. Navarro (2007): "Dynamic discrete choice and dynamic treatment effects," *Journal of Econometrics*, 136, 341–396. 1

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475. 1

Jun, S. J., J. Pinkse, H. Xu, and N. Yildiz (2016): "Multiple discrete endogenous variables in weakly-separable triangular models," *Econometrics*, 4, 7. 1

Kahn, A. B. (1962): "Topological sorting of large networks," *Communications of the ACM*, 5, 558–562. 1, 6, 6

Kitagawa, T. and A. Tetenov (2018): "Who should be treated? empirical welfare maximization methods for treatment choice," *Econometrica*, 86, 591–616. 3, 9

Lee, S. and B. Salanié (2018): "Identifying effects of multivalued treatments," *Econometrica*, 86, 1939–1963. 3

Machado, C., A. Shaikh, and E. Vytlacil (Forthcoming): "Instrumental variables and the sign of the average treatment effect," *Journal of Econometrics*. 1, 5, 5, A.4

Manski, C. F. (1997): "Monotone treatment response," *Econometrica: Journal of the Econometric Society*, 1311–1334. 3

——— (2004): "Statistical treatment rules for heterogeneous populations," *Econometrica*, 72, 1221–1246. 1, 3

Manski, C. F. and J. V. Pepper (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010. 5

Murphy, S. A. (2003): "Optimal dynamic treatment regimes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 331–355. 1, 2, 1

Murphy, S. A., M. J. van der Laan, J. M. Robins, and C. P. P. R. Group (2001): "Marginal mean models for dynamic regimes," *Journal of the American Statistical Association*, 96, 1410–1423. 1, 1, 7

Robins, J. M. (1997): "Causal inference from complex longitudinal data," in *Latent Variable Modeling and Applications to Causality*, Springer, 69–117. 1

Shaikh, A. M. and E. J. Vytlacil (2011): "Partial identification in triangular systems of equations with binary dependent variables," *Econometrica*, 79, 949–955. 1

VYTLACIL, E. (2002): “Independence, monotonicity, and latent index models: An equivalence result,” *Econometrica*, 70, 331–341. 1, 3, A.1

VYTLACIL, E. AND N. YILDIZ (2007): “Dummy endogenous variables in weakly separable models,” *Econometrica*, 75, 757–779. 1

WANG, L. AND E. TCHETGEN TCHETGEN (2018): “Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80, 531–550. 1, 2