Optimal Dynamic Treatment Regimes and Partial Welfare Ordering*

Sukjin Han
Department of Economics
University of Texas at Austin

This Draft: January 28, 2020
Click Here for the Most Recent Version

Abstract

Dynamic treatment regimes are treatment allocations tailored to heterogeneous individuals. The optimal dynamic treatment regime is a regime that maximizes counterfactual welfare. This paper investigates the possibility of identification of optimal dynamic regimes when data are generated from sequential (quasi-) experiments. We propose a framework in which we can partially learn the optimal dynamic regime and ordering of welfares, relaxing sequential randomization assumptions commonly employed in the literature. We establish the sharp partial ordering of counterfactual welfares with respect to dynamic regimes by using a series of linear programs. A distinct feature of our approach is that, instead of solving a large number of large-scale linear programs, we provide simple analytical conditions for the ordering. The identified set of the optimal regime is then characterized as the set of maximal elements of the partial order. We also propose topological sorts of the partial order as a policy menu. We show how policymaking can be further guided by imposing assumptions such as monotonicity/uniformity of different stringency, agent's learning, Markovian structure, and stationarity.

JEL Numbers: C14, C32, C33, C36

Keywords: Optimal dynamic treatment regimes, endogenous treatments, dynamic treatment effect, instrumental variable, linear programming.

*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, Ed Vytlacil, participants in the 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 Rice, U of Illinois Urbana-Champaign, National U of Singapore and Singapore Management U. sukjin.han@austin.utexas.edu

1 Introduction

Dynamic (i.e., adaptive) treatment regimes are treatment assignments tailored to individual heterogeneity to improve welfare. Typically, a dynamic treatment regime is defined as a sequence of assignment rules that maps previous outcomes and treatments onto current allocation decisions. Then the optimal dynamic treatment regime is defined as a regime that maximizes a social planner's objective function, such as counterfactual welfare. This paper investigate the possibility of identification of optimal dynamic regimes, when panel data are generated from sequential (quasi-) experiments, i.e., multi-stage experiments in the presence of non-compliance, or more generally from observational studies. Examples of the former are medical trials, public health and educational interventions in field experiments, or A/B testing in digital platforms, and examples of the latter are squences of policy shocks or what we call *sequential regression discontinuity (RD) designs*.

Optimal treatment regimes have been extensively studied in the biostatistics literature (Murphy et al. (2001), Murphy (2003), and Robins (2004) among others). These studies critically rely on an ideal multi-stage experimental environment that satisfies sequential randomization. It assumes that the treatment is randomly assigned in each period conditional on the history and that such an assignment is fully complied. Based on the data that satisfy this assumption, they identify optimal regimes that maximize welfare, written as the average counterfactual outcomes. Non-compliance, however, is prevalent in experiments, especially in multi-stage settings, e.g., due to the cost of enforcement or subjects' learning, and therefore should be allowed for. More generally, treatment endogeneity is a marked feature in observational studies, and this may be one reason why the vast biostatistics literature has not yet gain traction in economic analyses.

This paper proposes a framework in which we can partially learn the optimal dynamic regime and ordering of welfares. We define welfare as a linear functional of the joint distribution of counterfactual outcomes of all periods, and the optimal dynamic regime as its maximizer. An example of the welfare is the average counterfactual terminal outcome considered in the literature. Using a panel of outcomes, endogenous treatments, and instruments, we establish the sharp partial ordering of welfares, and characterize the identified set of optimal regimes as a discrete subset of all possible regimes as well as the bounds on welfares. We focus on binary outcomes and treatments for policy sets to be feasible by reducing the cardinality of possible regimes. A sequence of binary instruments is assumed to be generated by sequential randomized trials or sequential RD.

The analysis is conducted in two steps. In the first step, we establish the partial ordering of counterfactual welfares with respect to possible regimes, representing it as a directed acyclic graph (DAG). For this purpose, we characterize bounds on the difference of welfares for possible regime pairs via a sequence of linear programs. These bounds on welfare gaps are informative about whether welfares are comparable or not, and when they are, how to order them. The DAG obtained in this way is shown to be sharp (in the sense that will become clear). A novel feature of this analysis is that we do *not* numerically solve the linear programming problems. Solving them is computationally costly because each linear program is large-scale and there are as many linear programs to solve as the number of possible welfare pairs, which is also large due to adaptivity. Instead, as one of the main contributions of this paper, we provide simple analytical conditions for incomparability of welfare pairs and conditions that determine the signs of welfare gaps. Note that each welfare gap measures the *dynamic treatment effect*. The DAG concisely summarizes the identified signs of the treatment effects, and thus it is a parameter of independent interest in this paper.

In the second step, given the DAG representation, we show that the identified set can be characterized as the set of maximal elements of the partial order, i.e., regimes that are *not inferior*. We show that the set can be easily computed from the adjacency matrix of the DAG. Given the DAG, we also calculate topological sorts, which are linear orderings that do not violate the underlying partial order and thus can be viewed as a policy menu. We then solve linear programming only to calculate bounds on a small number of sorted welfares as well as regrets.

Often, the researcher is willing to impose additional assumptions to gain identification power. We propose identifying assumptions, such as monotonicity/uniformity assumptions that generalize the monotonicity assumption in Imbens and Angrist (1994), an assumption on agent's learning, Markovian structure, and stationarity. These assumptions tighten the identified set by reducing the dimension of the simplex in the linear programming, and thus producing a denser DAG.

To our best knowledge, this paper is first in the literature that considers the identifiability of optimal dynamic regimes under treatment endogeneity. As mentioned, Robins (1997), Murphy et al. (2001), and Murphy (2003) identify optimal dynamic regimes but under the sequential randomization assumption. Recently, Han (Forthcoming) and Wang and Tchetgen Tchetgen (2018) relax sequential randomization (and thus allow non-compliance) and consider identification of average treatment effects, but only as functions of non-adaptive regimes which greatly simplify the analysis. Relatedly, 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. Athey and Imbens (2018), Abraham and Sun (2019), Callaway and Sant'Anna (2019) extend the difference-in-differences approach to dynamic settings and consider the effects of treatment timing (i.e., irreversible treatments) on the treated.

The linear programming approach to partial identification of counterfactuals has early examples as Balke and Pearl (1997) and Manski (2007), and more recently appears in Torgovitsky (2019), Kamat (2017), Deb et al. (2017), Mogstad et al. (2018), Kitamura and Stoye (2019), Machado et al. (Forthcoming), and Gunsilius (2019). A distinct feature of this paper is that, even though our problem produces the large number of large-scale programs, we do not numerically solve them but derive analytical conditions that determine the signs of the optima, which are sufficient to construct the partial order and the identified set. These conditions have simple sample counterparts, and thus estimation and inference of the DAG and the identified set are shown to be relatively straightforward, without relying on objects that are solutions to linear programs. The notion of sharp partial ordering introduced in this paper and its analytical derivation have broader applicability beyond the context of this paper. They can be used in settings where linear programming is involved and the goal is to compare welfares across multiple treatments or, more generally, to establish a counterfactual ordering across different scenarios and find the best one.

In the next section, we introduce the dynamic regimes and related counterfactual outcomes, which define the welfare and the optimal regime. Section 3 provides a motivating example. Section 4 conducts the main identification analysis by constructing the DAG and characterizing the identified set. Sections 5–7 introduce topological sorts, additional identifying assumptions, and discuss cardinality reduction for the set of regimes. Section 8 illustrates our analysis with numerical exercises and Section 9 discusses estimation and inference. Most proofs are collected in the Appendix.

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. 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 Welfares

2.1 Dynamic Regimes

Let t be the index for a period or stage. For each $t = 1, ..., T$ with fixed T, define an *adaptive treatment rule* $\delta_t : \{0,1\}^{t-1} \times \{0,1\}^{t-1} \to \{0,1\}$ that maps the lags of realized binary outcomes and binary treatments $\boldsymbol{y}^{t-1} \equiv (y_1, ..., y_{t-1})$ and $\boldsymbol{d}^{t-1} \equiv (d_1, ..., d_{t-1})$ onto a

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

Table 1: Dynamic Regimes $\boldsymbol{\delta}(\cdot)$ with $T = 2$

non-stochastic treatment allocation $d_t \in \{0, 1\}$:

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

This adaptive rule also appears in, e.g., Murphy (2003). A special case of (1) is a non-adaptive rule where $\delta_t(\cdot)$ is just a constant function (Han (Forthcoming), Wang and Tchetgen Tchetgen (2018)). Whether the rule is adaptive or non-adaptive, we only consider non-stochastic rules.[1] The rule can also be a function of other observables, which we do not consider here for succinctness. Then, a *dynamic regime* up to period t is defined as a vector of all treatment rules:

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

where $\mathcal{D}^t$ is the class of all possible regimes. Let $\boldsymbol{\delta}(\cdot) \equiv \boldsymbol{\delta}^T(\cdot) \in \mathcal{D} \equiv \mathcal{D}^T$. With $T = 2$, Table 1 lists all possible dynamic regimes $\boldsymbol{\delta}(\cdot) \equiv (\delta_1, \delta_2(\cdot))$ as contingency plans.

2.2 Counterfactual Welfares and Optimal Regimes

To define welfare w.r.t. this dynamic regime, we first introduce a counterfactual outcome as a function of the dynamic regime. Because of the adaptivity intrinsic in dynamic regimes, expressing counterfactual outcomes is more involved than that with non-adaptive regimes d_t, i.e., $Y_t(\boldsymbol{d}^t)$ with $\boldsymbol{d}^t \equiv (d_1, ..., d_t)$. Let $\boldsymbol{Y}^t(\boldsymbol{d}^t) \equiv (Y_1(d_1), Y_2(\boldsymbol{d}^2), ..., Y_t(\boldsymbol{d}^t))$. We express a

[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.

counterfactual outcome with adaptive regime $\boldsymbol{\delta}^t(\cdot)$ as follows:

$$Y_t(\boldsymbol{\delta}^t(\cdot)) \equiv Y_t(\boldsymbol{d}^t) \tag{2}$$

where the bridge variables $\boldsymbol{d}^t \equiv (d_1, ..., d_t)$ satisfies

$$\begin{aligned} d_1 &= \delta_1, \\ d_2 &= \delta_2(Y_1(d_1), d_1), \\ d_3 &= \delta_3(\boldsymbol{Y}^2(\boldsymbol{d}^2), \boldsymbol{d}^2), \\ &\vdots \\ d_t &= \delta_t(\boldsymbol{Y}^{t-1}(\boldsymbol{d}^{t-1}), \boldsymbol{d}^{t-1}). \end{aligned} \tag{3}$$

In this recursive expression, for each t, the adaptive regime $\boldsymbol{\delta}^t(\cdot)$ take a value $\boldsymbol{d}^t$ which is fed into the next period's rule as an argument itself and as an argument of counterfactual outcome vector. Suppose $T = 2$. Then, the two counterfactual outcomes are defined as $Y_1(\delta_1) = Y_1(d_1)$ and $Y_2(\boldsymbol{\delta}^2(\cdot)) = Y_2(\delta_1, \delta_2(Y_1(\delta_1), \delta_1))$.

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

$$W_{\boldsymbol{\delta}} \equiv f(q_{\boldsymbol{\delta}}).$$

Examples of the functional include the average counterfactual terminal outcome $E[Y_T(\boldsymbol{\delta}(\cdot))] = \Pr[Y_T(\boldsymbol{\delta}(\cdot)) = 1]$, our leading case, the weighted average of counterfactuals $\sum_{t=1}^{T} \omega_t E[Y_t(\boldsymbol{\delta}^t(\cdot))]$, and these benefits less the cost of treatments $\sum_{t=1}^{T} p_t \boldsymbol{\delta}_t(\cdot)$ where p_t is a known cost at t. Then, the *optimal dynamic regime* is a regime that maximizes the welfare:[2]

$$\boldsymbol{\delta}^*(\cdot) = \arg\max_{\boldsymbol{\delta}(\cdot)\in\mathcal{D}} W_{\boldsymbol{\delta}} \tag{4}$$

In some cases, the solution $\boldsymbol{\delta}^*(\cdot)$ can be justified by backward induction in a finite-horizon dynamic programming; see Appendix A.1.

The identification analysis of the optimal regime is closely related to identification of welfare at each regime and welfare gaps, which also contain information for policy. Some interesting special cases are the followings: (i) the *optimal welfare*, $W_{\boldsymbol{\delta}^*}$, which in turn yields (ii)

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

the *regret* from following individual decisions, $W_{\boldsymbol{\delta}^*} - W_{\boldsymbol{D}}$, where $W_{\boldsymbol{D}}$ is simply $f(\Pr[\boldsymbol{Y}(\boldsymbol{D}) = \cdot]) = f(\Pr[\boldsymbol{Y} = \cdot])$, and (iii) the *gain from adaptivity*, $W_{\boldsymbol{\delta}^*} - W_{\boldsymbol{d}^*}$, where $W_{\boldsymbol{d}^*} = \max_{\boldsymbol{d}} W_{\boldsymbol{d}}$ is the optimum of the welfare with a non-adaptive rule, $W_{\boldsymbol{d}} = f(\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \cdot])$. If the cost of treatments is not considered, the gain in (iii) is non-negative since the set of all $\boldsymbol{d}$ is a subset of $\mathcal{D}$.

3 Motivating Example: Returns to School Types

We provide a stylized example in an observational setting to motivate the policy relevance of the optimal dynamic regime and the type of data that is useful to recover it.[3] Consider labor market returns to the types of high schools and colleges. Let $D_{i1} = 1$ if student i enrolls in an academic high school and $D_{i1} = 0$ if a vocational high school; let $Y_{i1} = 1$ if i achieves above-median GPA in high school and $Y_{i1} = 0$ if below-median. Also, let $D_{i2} = 1$ if i enrolls in a four-year college and $D_{i2} = 0$ if a two-year college. Finally, let $Y_{i2} = 1$ if i is employed at age 25 and $Y_{i2} = 0$ if not. Given the data, suppose we are interested in recovering regimes that maximize the employment rate as welfare.

Compared to the optimal non-adaptive regime, the optimal regime with adaptivity provides rich policy implications. Consider the optimal non-adaptive regime first. This will be the schedule $\boldsymbol{d} = (d_1, d_2) \in \{0, 1\}^2$ of school allocations that maximizes the employment rate $W_{\boldsymbol{d}} = E[Y_2(\boldsymbol{d})]$. In contrast, the optimal dynamic regime is the schedule $\boldsymbol{\delta}(\cdot) = (\delta_1, \delta_2(\cdot)) \in \mathcal{D}$ of school allocation *rules* that maximizes the employment rate $W_{\boldsymbol{\delta}} = E[Y_2(\boldsymbol{\delta})]$. The schedule of allocation rules would first assign either an academic or vocational high school ($\delta_1 \in \{0, 1\}$) and then assign either a four-year or two-year college ($\delta_2(y_1, \delta_1) \in \{0, 1\}$) depending on the high school type δ_1 and performance y_1. Suppose $\boldsymbol{\delta}^*(\cdot)$ that is identified is such that $\delta_2(1, 1) = 1$ and $\delta_2(0, 0) = 0$. That is, it turns out to be optimal to assign a four-year college to a student who enrolled in an academic high school and achieved high GPA, and to assign a two-year college to a student who enrolled in a vocational high school and achieved low GPA. In reality, a policy maker rarely literally assigns schools to students. Such $\boldsymbol{\delta}^*(\cdot)$, however, may have a policy implication that the average job market performance will be improved by a merit-based tuition subsidy for four-year college. Note that this type of policy questions cannot be answered from the optimal non-adaptive regime.

Since D_1 and D_2 are endogenous, the data $\{D_{i1}, Y_{i1}, D_{i2}, Y_{i2}\}$ above is not useful by themselves to identify $W_{\boldsymbol{\delta}}$'s and $\boldsymbol{\delta}^*(\cdot)$. We propose a sequential version of the fuzzy RD

[3] Examples in multi-stage experimental studies are the Fast Track Prevention Program (Conduct Problems Prevention Research Group (1992)) and the Elderly Program randomized trial for the Systolic Hypertension (The Systolic Hypertension in the Elderly Program (SHEP) Cooperative Research Group (1988)).

design as one possible source of exogenous variation. The sequence of high school and college entrance exams would generate running variables, i.e., test scores, that define eligibility for admission. Let $Z_{i1} = 1$ if student i landed slightly above the cutoff of the academic high school entrance exam and $Z_{i1} = 0$ if slightly below; let $Z_{i2} = 1$ if i landed slightly above cutoff for the four-year college entrance exam, $Z_{i2} = 0$ if slightly below. Then (Z_{i1}, Z_{i2}) can serve as the sequence of binary instruments that satisfy Assumption SX.[4]

4 Partial Ordering and Partial Identification

4.1 Observables

We introduce observables based on which we wish to identify the optimal regime and counterfactual welfares. For each period/stage $t = 1, ..., T$ with fixed T, assume that we observe the binary randomized treatment assignment Z_t, the binary endogenous treatment decision D_t, and the binary outcome $Y_t = \sum_{\boldsymbol{d}^t \in \{0,1\}^t} Y_t(\boldsymbol{d}^t)$. For example, Y_t is a symptom indicator of a patient, D_t is a medical treatment received, and Z_t is generated by a multi-period medical trial. Binary variables are helpful to reduce the cardinality of possible regimes and to define linear programs.[5] Let $D_t(\boldsymbol{z}^t)$ be the counterfactual treatment given $\boldsymbol{z}^t \equiv (z_1, ..., z_t) \in \{0, 1\}^t$. Then, $D_t = \sum_{\boldsymbol{z}^t \in \mathcal{Z}^t} D_t(\boldsymbol{z}^t)$. Let $\boldsymbol{Y}(\boldsymbol{d}) \equiv (Y_1(d_1), Y_2(\boldsymbol{d}^2), ..., Y_T(\boldsymbol{d}))$ and $\boldsymbol{D}(\boldsymbol{z}) \equiv (D_1(z_1), D_2(\boldsymbol{z}^2), ..., D_T(\boldsymbol{z}))$.

Assumption SX. $Z_t \perp (\boldsymbol{Y}(\boldsymbol{d}), \boldsymbol{D}(\boldsymbol{z}))|\boldsymbol{Z}^{t-1}$.

Assumption SX assumes the strict exogeneity and exclusion restriction. This assumption is satisfied in typical sequential randomized experiment designs as well as in sequential fuzzy RD designs;[6] see Section 3 for the example of the latter. Let $(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z})$ be the vector of observables (Y_t, D_t, Z_t) for the entire T periods and let p be its distribution. We assume that $\{(\boldsymbol{Y}_i, \boldsymbol{D}_i, \boldsymbol{Z}_i) : i = 1, ..., N\}$ is a small T large N panel. We mostly suppress the individual unit i throughout the paper. For empirical applications, it is important to note that 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; Section 3 contains such an example. For the population where the data is drawn, we are interested in learning the optimal adaptive allocation sequence.

[4] Alternatively, the discretized version of the distance to (or the tuition cost of) these schools can serve as Z_1 and Z_2.

[5] Extending the analysis to multi-valued discrete variables is possible, but we keep the setting simple.

[6] There may be other covariates available for the researcher but we suppress them for succinctness. All the stated assumptions and the analyses of this paper can be followed conditional on covariates.

(a) $\boldsymbol{\delta}^*(\cdot)$ is partially identified (b) $\boldsymbol{\delta}^*(\cdot)$ is point identified

Figure 1: Partially Ordered Sets as Directed Acyclic Graphs

4.2 Partial Ordering as Directed Acyclic Graphs

Given the distribution p of the data $(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z})$ and under Assumption SX, we show how the optimal dynamic regime and welfares can be partially recovered. The identified set $\mathcal{D}^*$ of $\boldsymbol{\delta}^*(\cdot)$ will be formally defined as a subset of the discrete set $\mathcal{D}$. As the first step in characterizing this set, we establish a *partial ordering* of $W_{\boldsymbol{\delta}}$ w.r.t. $\boldsymbol{\delta}(\cdot) \in \mathcal{D}$ using p. A partial ordering can be represented by a *directed acyclic graph* (DAG), $G(V, E)$, where V is the set of all welfares and E is the set of directed links governed by "$\geq$." The DAG summarizes the identified signs of the dynamic treatment effects, as will become clear later. Moreover, the DAG representation is fruitful to introduce the notion of sharpness of a partial ordering and later to translate it into the identified set of $\boldsymbol{\delta}^*(\cdot)$.

To facilitate this analysis, we enumerate all $|\mathcal{D}| = 2^{2^T-1}$ possible regimes. For index k ($1 \leq k \leq |\mathcal{D}|$), let $\boldsymbol{\delta}_k(\cdot)$ denote the k-th regime in $\mathcal{D}$. With $T = 2$, Table 1 indexes all possible dynamic regimes $\boldsymbol{\delta}(\cdot) \equiv (\delta_1, \delta_2(\cdot))$. Let $W_k \equiv W_{\boldsymbol{\delta}_k}$ be the corresponding welfare. Then, most of the time, we conveniently define the set of vertices V of a DAG as the set of welfare (or regime) indices $\{k : 1 \leq k \leq |\mathcal{D}|\}$ instead of welfares themselves $\{W_k : 1 \leq k \leq |\mathcal{D}|\}$. Figure 1 illustrates examples of the partially ordered set of welfares as DAGs where each edge "$W_k \to W_{k'}$" indicates the relation "$W_k \geq W_{k'}$."

In general, the point identification of $\boldsymbol{\delta}^*(\cdot)$ will be achieved by establishing a total ordering of W_k, which is not possible with instruments of limited support. Instead, we only recover a partial ordering. We want the partial ordering to be sharp in the sense that it cannot be improved upon given the data and maintained assumptions.

Definition 4.1. *Given the data distribution p, a partial ordering $G(V, E_p)$ is sharp under maintained assumptions if there exists no partial ordering $G(V, E_p')$ such that $E_p' \supsetneq E_p$ without imposing additional assumptions.*

Establishing the sharp partial ordering amounts to determining whether we can identify the sign of a counterfactual welfare gap $W_k - W_{k'}$ (i.e., the dynamic treatment effects) for

$k, k' \in V$, and if we can, what the sign is.

4.3 Data-Generating Framework

We introduce a simple data-generating framework and formally define the identified set. First, we introduce latent state variables that generate $(\boldsymbol{Y}, \boldsymbol{D})$. A latent state of the world will determine specific maps $(\boldsymbol{y}^{t-1}, \boldsymbol{d}^t) \mapsto y_t$ and $(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^t) \mapsto d_t$ for $t = 1, ..., T$ under the exclusion restriction in Assumption SX. We introduce the latent state variable $\tilde{S}_t$ whose realization represents such a state. We define $\tilde{S}_t$ as follows. For given $(\boldsymbol{y}^{t-1}, \boldsymbol{d}^t, \boldsymbol{z}^t)$, let $Y_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^t)$ and $D_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^t)$ respectively be the extended counterfactual outcomes and treatments, and let $\{Y_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^t)\}$ and $\{D_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^t)\}$ and their sequences w.r.t. $(\boldsymbol{y}^{t-1}, \boldsymbol{d}^t, \boldsymbol{z}^t)$. Then $\tilde{S}_t \equiv (\{Y_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^t)\}, \{D_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^t)\}) \in \{0,1\}^{2^{2t-1}} \times \{0,1\}^{2^{3t-2}}$ concatenates the two sequences. 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 $\tilde{\boldsymbol{S}} \equiv (\tilde{S}_1, ..., \tilde{S}_T)$ into a scalar (discrete) latent variable in $\mathbb{N}$ as $S \equiv \beta(\tilde{\boldsymbol{S}}) \in \mathcal{S}$, where $\beta(\cdot)$ is a one-to-one map that transforms a binary sequence into a decimal value. Define

$$q_s \equiv \Pr[S = s],$$

and define the vector q of q_s which represents the distribution of S or the true data-generating process. The vector q resides in a standard simplex $\mathcal{Q} \equiv \{q : \sum_s q_s = 1 \text{ and } q_s \geq 0 \ \forall s\}$ of dimension $d_q - 1$ where $d_q \equiv \dim(q)$. A useful fact is that the joint distribution of counterfactuals can be written as a linear functional of q:

$$\begin{aligned}\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}, \boldsymbol{D}(\boldsymbol{z}) = \boldsymbol{d}] &= \Pr[S \in \mathcal{S} : \boldsymbol{Y}(\boldsymbol{y}^{T-1}, \boldsymbol{d}) = \boldsymbol{y}, \boldsymbol{D}(\boldsymbol{y}^{T-1}, \boldsymbol{d}^{T-1}, \boldsymbol{z}) = \boldsymbol{d}] \\ &= \Pr[S \in \mathcal{S} : Y_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^t) = y_t, D_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^t) = d_t \quad \forall t] \\ &= \sum_{s \in \mathcal{S}_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}}} q_s, \end{aligned} \tag{5}$$

where $\mathcal{S}_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}}$ is constructed by using the definition of S; its expression can be found in Appendix A.2.

Based on (5), the counterfactual welfare can be written as a linear combination of q_s's. That is, there exists $1 \times d_q$ vector A_k of 1's and 0's such that

$$W_k = A_k q. \tag{6}$$

Recall $W_k \equiv f(q_{\boldsymbol{\delta}_k})$ where $q_{\boldsymbol{\delta}}(\boldsymbol{y}) \equiv \Pr[\boldsymbol{Y}(\boldsymbol{\delta}(\cdot)) = \boldsymbol{y}]$. The key observation to the result (6) is that $\Pr[\boldsymbol{Y}(\boldsymbol{\delta}(\cdot)) = \boldsymbol{y}]$ can be written as a linear functional of the joint distributions of counterfactual outcomes with *non-adaptive* regime, i.e., $\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}]$'s, which is in turn a linear functional of q. To illustrate with $T = 2$ and welfare $W = E[Y_2(\boldsymbol{\delta}(\cdot))]$, we have

$$\Pr[Y_2(\boldsymbol{\delta}(\cdot)) = 1] = \sum_{y_1 \in \{0,1\}} \Pr[Y_2(\delta_1, \delta_2(y_1, \delta_1)) = 1 | Y_1(\delta_1) = y_1] \Pr[Y_1(\delta_1) = y_1]$$

by the law of iterated expectation. Then, for instance, Regime 8 in Table 1 yields

$$\Pr[Y_2(\boldsymbol{\delta}_8(\cdot)) = 1] = P[\boldsymbol{Y}(1,1) = (1,1)] + P[\boldsymbol{Y}(1,1) = (0,1)], \tag{7}$$

where each $\Pr[\boldsymbol{Y}(d_1, d_2) = (y_1, y_2)]$ is the counterfactual distribution with non-adaptive regime, which in turn is a linear functional of (5).

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

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

and p as the vector of $p_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}}$'s except redundant elements. Let $d_p \equiv \dim(p)$. Since $\Pr[\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{D} = \boldsymbol{d} | \boldsymbol{Z} = \boldsymbol{z}] = \Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}, \boldsymbol{D}(\boldsymbol{z}) = \boldsymbol{d}]$ by Assumption SX, we can readily show by (5) that there exists $d_p \times d_q$ matrix B such that

$$Bq = p, \tag{8}$$

where each row of B is a vector of 1's and 0's. We assume $rank(B) = d_p$ without loss of generality, since redundant constraints do not play a role in restricting $\mathcal{Q}$. We focus on the non-trivial case of $d_p < d_q$. If $d_p \geq d_q$, which rarely is the case, we can solve for $q = (B^\top B)^{-1} B^\top p$, and can trivially point identify $W_k = A_k q$ and thus $\boldsymbol{\delta}^*(\cdot)$. The formal derivation of A_k as well as B can be found in Appendix A.2. It is important to note that the linearity in (6) and (8) is not a restriction but given by the discrete nature of our setting.

The expression (6) and (8) are useful to define the identified set of $\boldsymbol{\delta}^*(\cdot)$. Let $\boldsymbol{\delta}^*(\cdot; q) \equiv \arg\max_{\boldsymbol{\delta}_k(\cdot) \in \mathcal{D}} W_k = A_k q$ be the optimal regime, explicitly written as a function of the data-generating process q.

Definition 4.2. *Under Assumption SX, the identified set of $\boldsymbol{\delta}^*(\cdot)$ given the data distribution p is defined as*

$$\mathcal{D}_p^* \equiv \{\boldsymbol{\delta}^*(\cdot; q) : Bq = p \text{ and } q \in \mathcal{Q}\}, \tag{9}$$

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

4.4 Characterizing Partial Ordering and Identified Set

We establish the partial ordering of W_k's, i.e., generate the DAG as a $|\mathcal{D}| \times |\mathcal{D}|$ adjacency matrix, by determining whether, given p, $W_k \geq W_{k'}$, $W_k < W_{k'}$, or W_k and $W_{k'}$ are not comparable, denoted as $W_k \sim W_{k'}$, for $k, k' \in V$. 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 $k, k' \in V$ and $k > k'$.[7] Then the identified set can be characterized based on the resulting partial order.

The nature of our data generation induces the linear system (6) and (8). This enables us to characterize the bounds on $W_k - W_{k'} = (A_k - A_{k'})q$ as optima of linear programming. Let $U_{k,k'}$ and $L_{k,k'}$ be the upper and lower bounds. Also $\Delta_{k,k'} \equiv A_k - A_{k'}$ for simplicity. Then, we have, for $k, k' \in V$ and $k > k'$,

$$\begin{aligned} U_{k,k'} &= \max\nolimits_{q \in \mathcal{Q}} \Delta_{k,k'} q, \\ L_{k,k'} &= \min\nolimits_{q \in \mathcal{Q}} \Delta_{k,k'} q, \end{aligned} \qquad s.t. \quad Bq = p. \tag{10}$$

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

Assumption B imposes that the model is correctly specified. Under misspecification, the identified set is empty by definition. The next theorem constructs the sharp DAG and the identified set using $U_{k,k'}$ and $L_{k,k'}$ for $k, k' \in V$ and $k > k'$, or equivalently, $L_{k,k'}$ for $k, k' \in V$ and $k \neq k'$.

Theorem 4.1. *Suppose Assumptions SX and B hold. Then, (i) $G(V, E_p)$ with $E_p \equiv \{(k, k') : L_{k,k'} \geq 0$ for $k, k' \in V$ and $k \neq k'\}$ is sharp;*[8] *(ii) $\mathcal{D}_p^*$ defined in (9) satisfies*

$$\mathcal{D}_p^* = \{\boldsymbol{\delta}_{k'}(\cdot) : \nexists k \text{ such that } L_{k,k'} > 0 \text{ for } k, k' \in V \text{ and } k \neq k'\}. \tag{11}$$

Theorem 4.1(i) immediately holds by Definition 4.1, since $L_{k,k'}$ (and $U_{k,k'}$) is sharp in (10). The latter is because $\{q : Bq = p \text{ and } q \in \mathcal{Q}\}$ is convex and thus $\{\Delta_{k,k'} q : Bq = p \text{ and } q \in \mathcal{Q}\}$ is convex, which implies that any point between $[L_{k,k'}, U_{k,k'}]$ is attainable. According to (i), the sharp DAG is constructed as follows: when $L_{k,k'} \geq 0$, then $W_k \geq W_{k'}$ and a directed edge is formed between (k, k'); when $L_{k,k'} < 0 < U_{k,k'}$, then $W_k \sim W_{k'}$ and no edge is formed between (k, k'). The DAG can be represented as a $|\mathcal{D}| \times |\mathcal{D}|$ adjacency matrix Σ such that its element $\Sigma_{k,k'} = 1$ if $W_k \geq W_{k'}$ and $\Sigma_{k,k'} = 0$ otherwise.

[7] Note that directly comparing sharps bounds on welfares themselves will *not* deliver sharp partial ordering.

[8] Notice that $(L_{k,k'}, U_{k,k'})$ for all $k, k' \in V$ and $k > k'$ contain the same amount of information as $L_{k,k'}$ for all $k, k' \in V$ and $k \neq k'$, since $U_{k,k'} = -L_{k',k}$.

In Theorem 4.1(ii), $\mathcal{D}_p^*$ is characterized as the collection of $\boldsymbol{\delta}_k(\cdot)$ where k is in the set of *maximal elements* of the partially ordered set $G(V, E_p)$, i.e., the set of regimes that are *not inferior*. In Figure 1(a), it is easy to see that the set of maximals is $\mathcal{D}_p^* = \{\boldsymbol{\delta}_1(\cdot), \boldsymbol{\delta}_4(\cdot)\}$. Using the adjacency matrix Σ, the set of maximal elements (11) can be obtained by

$$\mathcal{D}_p^* = \{\boldsymbol{\delta}_{k'}(\cdot) : \Sigma_{k,k'} = 0 \text{ for all } k \in V \text{ and } k \neq k' \in V\}. \tag{12}$$

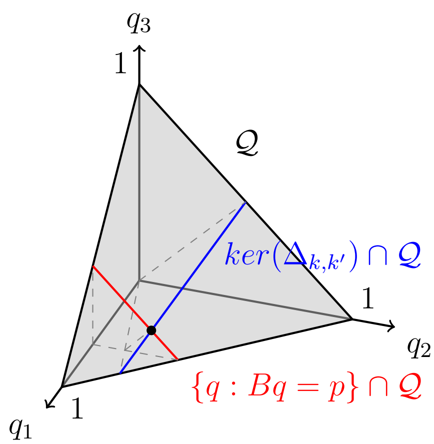
The identified set $\mathcal{D}_p^*$ characterizes the information content of the model. Given the minimal structure we impose in the model, the size of $\mathcal{D}_p^*$ may be large in some cases. We, however, argue that an uninformative $\mathcal{D}_p^*$ still has implications for policy: (i) it recommends the planner to eliminate sub-optimal regimes from her options; (ii) in turn, it warns the planner of 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 non-trivial policy suggestion. As shown in our 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. (Forthcoming) for the average treatment effect in a static model. Section 6 lists further identifying assumptions that help shrink $\mathcal{D}_p^*$.

4.5 Analytical Conditions

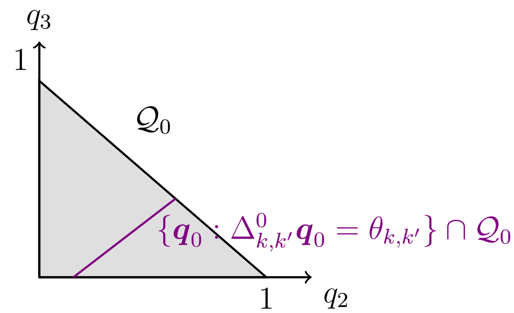
In practice, a naive approach to obtain the sharp DAG and the identified set based on Theorem 4.1 is to directly compute $U_{k,k'}$ and $L_{k,k'}$ by solving linear program (10) for $k, k' \in V$ and $k > k'$. This can be computationally very costly. Note that, to generate the DAG, we need to make at most "$|V| = |\mathcal{D}| = 2^{2^T-1}$ choose 2" pair-wise comparisons of the welfares.[9] With the naive approach, this amounts to solving "2^{2^T-1} choose 2" times two linear programs (10), where (10) is a large-scale linear program. In this program, the dimension of q is $d_q = |\mathcal{Q}| + 1 = |\mathcal{S}| = \prod_{t=1}^T |\mathcal{S}_t|$, which can be immense; e.g., when $T = 2$, $d_q = 2^2 \times 2^2 \times 2^{16} \times 2^8 = 268{,}435{,}456$. Also, the number of constraints is $d_p + d_q + 1$ where $d_p = 2^{3T} - 2^T$. This computational complexity can possibly be mitigated by imposing further assumptions on the data-generating process as shown later. Even then, the naive approach poses nontrivial challenges in developing inference methods for $\boldsymbol{\delta}^*(\cdot)$ and other parameters, because they involve objects produced by solving linear programming.

Instead, we propose a simple analytical procedure to obtain the sharp DAG and the

[9] This procedure is closely related to what is called the *bubble sort*. There are more efficient algorithms, such as the *quick sort*, although they need to be modified to incorporate the distinct feature of our problem: the possible incomparability that stems from partial identification. Note that, for comparable pairs, transitivity can be applied and thus the total number of comparisons can be smaller.



(a) Original Problem with a Kernel, a Hyperplane, and a Simplex



(b) Simpler Problem with a Hyperplane and a Cone (with $\boldsymbol{q}_0 = (q_2, q_3)^\top$)

Figure 2: Illustration of Conditions for $L_{k,k'} < 0 < U_{k,k'}$ (with $q = (q_1, q_2, q_3)^\top$)

identified set. Recall that $\mathcal{Q} \equiv \{q : \sum_s q_s = 1 \text{ and } q_s \geq 0 \ \forall s\} \subset \mathbb{R}^{d_q}$ is a standard simplex of dimension $d_q - 1$, and B is $d_p \times d_q$ matrix with $d_p < d_q$. We assume $B = (B_1 \vdots O)$ for some $d_p \times d_p$ full rank matrix B_1 and $d_p \times (d_q - d_p)$ zero matrix O, which is also without loss of generality.[10]

Fix k and k'. We first investigate the possibility of detecting $W_k \sim W_{k'}$, the incomparability of W_k and $W_{k'}$. Note that $L_{k,k'} < 0 < U_{k,k'}$ if and only if there exists $q \in \mathcal{Q}$ such that $\Delta_{k,k'} q = 0$ and $Bq = p$, or simply,

$$ker(\Delta_{k,k'}) \cap \{q : Bq = p\} \cap \mathcal{Q} \neq \emptyset, \tag{13}$$

where $ker(\Delta)$ denotes the kernel (i.e., the null space) of Δ. That is, we want to find conditions under which the simplex $\mathcal{Q}$, the hyperplanes $\{q : Bq = p\}$, and $ker(\Delta_{k,k'})$ all intersect. Figure 2(a) depicts this intersection for the case of $d_q = 3$.

Define partitions $\Delta_{k,k'} = (\Delta^1_{k,k'} \vdots \Delta^0_{k,k'})$ and $q = (\boldsymbol{q}_1^\top, \boldsymbol{q}_0^\top)^\top$ according to partition $B = (B_1 \vdots O)$. Then $p = Bq = B_1\boldsymbol{q}_1$ or

$$\boldsymbol{q}_1 = B_1^{-1} p \tag{14}$$

[10]If (iii) does not hold, then we can find an elementary column operating matrix M of order $d_q \times d_q$ such that $\tilde{B} \equiv BM = (B_1 \vdots O)$. Then, using M we can redefine all the relevant quantities and proceed analogously. Let $\tilde{A}_k \equiv A_k M$, $\tilde{A}_{k'} \equiv A_{k'} M$, and $\tilde{q} \equiv M^{-1} q$ as M is invertible. Then, it satisfies that $Bq = BMM^{-1}q = \tilde{B}\tilde{q}$ and $(A_k - A_{k'})q = (A_k - A_{k'})MM^{-1}q = (\tilde{A}_k - \tilde{A}_{k'})\tilde{q}$. Note that $\tilde{\mathcal{Q}} \equiv \{M^{-1}q : q \in \mathcal{Q}\} \subset \mathbb{R}^{d_q}$ is also a standard simplex and $\tilde{\mathcal{Q}}_p \equiv \{\tilde{q} \in \tilde{\mathcal{Q}} : \tilde{B}\tilde{q} = p\}$.

because B_1 has full rank. That is, we can solve for the subvector of the data-generating process as a function of the data distribution. Plugging (14) in, $\Delta_{k,k'}q = \Delta^1_{k,k'}\boldsymbol{q}_1 + \Delta^0_{k,k'}\boldsymbol{q}_0 = 0$ can be rewritten as

$$\Delta^1_{k,k'}B_1^{-1}p + \Delta^0_{k,k'}\boldsymbol{q}_0 = 0.$$

For simplicity, let $\theta_{k,k'} \equiv -\Delta^1_{k,k'}B_1^{-1}p$, which is a scalar. Define $\mathcal{Q}_0 \equiv \{\boldsymbol{q}_0 : q \in \mathcal{Q}\} = \{\boldsymbol{q}_0 : \sum_{s\in\mathcal{S}_0} q_s \leq 1 \text{ and } q_s \geq 0\ \forall s \in \mathcal{S}_0\}$ where $\mathcal{S}_0$ is the set of indices that correspond to the subvector $\boldsymbol{q}_0$. Then, $L_{k,k'} < 0 < U_{k,k'}$ if and only if there exists *nonzero vector* $\boldsymbol{q}_0 \in \mathcal{Q}_0$ such that $\Delta^0_{k,k'}\boldsymbol{q}_0 = \theta_{k,k'}$, or

$$\{\boldsymbol{q}_0 : \Delta^0_{k,k'}\boldsymbol{q}_0 = \theta_{k,k'}\} \cap \mathcal{Q}_0 \backslash \{\boldsymbol{0}\} \neq \emptyset. \tag{15}$$

We want to find nonzero $\boldsymbol{q}_0$, since when $\boldsymbol{q}_0 = \boldsymbol{0}$, then $\Delta_{k,k'}q = -\theta_{k,k'}$ for all $\boldsymbol{q}_1$ and thus we trivially point identify $W_k - W_{k'} = -\theta_{k,k'}$. Since $\mathcal{Q}_0$ is a *finitely generated cone* and independent of the constraints, finding conditions under which (15) holds is mathematically more tractable than directly analyzing (13); see Figure 2(b). It essentially reduces down to checking whether the hyperplane $\Delta^0_{k,k'}\boldsymbol{q}_0 = \theta_{k,k'}$ lies between the vertices of the cone. The next theorem states these conditions of incomparability (i.e., $W_k \sim W_{k'}$). For elements γ_s ($s \in \mathcal{S}_0$) of vector $\Delta^0_{k,k'}$, we define $\underline{\gamma}_{k,k'} \equiv \min\{\gamma_s\}_{s\in\mathcal{S}_0} \in \{-1,0,1\}$ and $\overline{\gamma}_{k,k'} \equiv \max\{\gamma_s\}_{s\in\mathcal{S}_0} \in \{-1,0,1\}$.

Theorem 4.2. *Suppose Assumptions SX and B hold. For $k, k' \in V$ and $k \neq k'$, let $\theta_{k,k'} \equiv -\Delta^1_{k,k'}B_1^{-1}p$ and $\underline{\gamma}_{k,k'} \in \{-1,0,1\}$ and $\overline{\gamma}_{k,k'} \in \{-1,0,1\}$ be the minimum and maximum elements of vector $\Delta^0_{k,k'}$. Then, $L_{k,k'} < 0 < U_{k,k'}$ if and only if either one of the following holds: (i) $\underline{\gamma}_{k,k'} < \theta_{k,k'} < \overline{\gamma}_{k,k'}$, (ii) $\underline{\gamma}_{k,k'} \geq \theta_{k,k'} \geq 0$, or (iii) $\overline{\gamma}_{k,k'} \leq \theta_{k,k'} \leq 0$.*

Since $\theta_{k,k'}$ (up to p) and $\Delta^0_{k,k'}$ are known to the researcher, we can directly detect the incomparability from the data p without solving linear programmings. Furthermore, we can show the following result, which can be used to conclude $W_k \geq W_{k'}$:[11]

Corollary 4.1. *Suppose Assumptions SX and B hold. For $k, k' \in V$ and $k \neq k'$, let $\theta_{k,k'} \equiv -\Delta^1_{k,k'}B_1^{-1}p$ and $\underline{\gamma}_{k,k'} \in \{-1,0,1\}$ be the minimum element of vector $\Delta^0_{k,k'}$. Then, $L_{k,k'} \geq 0$ if and only if*

$$\theta_{k,k'} < \min\{0, \underline{\gamma}_{k,k'}\}. \tag{16}$$

[11]Note Conditions (i)–(iii) are exclusive. These conditions, the condition (16) below, and $\theta_{k,k'} > \max\{0, \overline{\gamma}_{k,k'}\}$ are exhaustive. The last condition guarantees $U_{k,k'} \leq 0$, which is redundant information for the DAG as k, k' are exchangeable.

Theorem 4.1(i) and Corollary 4.1 provide the basis for the systematic computation of the DAG. They suggest an algorithm that generates the DAG as the adjacency matrix Σ by automating the task of checking the condition in the corollary. Compared to directly solving the set of large-scale linear programs, finding $\underline{\gamma}_{k,k'}$ from a large-dimensional vector $\Delta^0_{k,k'}$ is an extremely simple computational task, especially since its value is known to be one of $\{-1, 0, 1\}$. Note that Corollary 4.1 cannot be directly used to construct $\mathcal{D}_p^*$ since we should be able to determine the strictly inequality ($L_{k,k'} > 0$), according to Theorem 4.1(i). Instead, the generated Σ can be used to construct $\mathcal{D}_p^*$ as shown in (12).

5 Topological Sorts and Bounds on Sorted Welfare

5.1 Topological Sorting

The DAG is a useful policy benchmark. For a complicated DAG, it may be easier to examine a linear ordering based on it. A *topological sort* of a DAG is a linear ordering of its vertices such that for every directed edge $k \to k'$, k comes before k' in that ordering. In other words, it is a linear extension of the partial ordering where W_k *cannot* be larger than $W_{k'}$ as long as $k < k'$. Let L_G be the number of topological sorts of $G(V, E_p)$ and, for $1 \leq l \leq L_G$, let $k_{l,1}$ is the initial vertex of the l-th topological sort. For example, given the DAG in Figure 1(a), $(\boldsymbol{\delta}_1, \boldsymbol{\delta}_4, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3)$ is an example of a topological sort (with $k_{l,1} = 1$), but $(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\delta}_4, \boldsymbol{\delta}_3)$ is not. Topological sorts are routinely reported for a given DAG, and there are well-known algorithms that efficiently find topological sorts, such as Kahn (1962)'s algorithm.

The following theorem alternatively characterizes $\mathcal{D}_p^*$ using topological sorts.[12]

Theorem 5.1. *Suppose Assumptions SX and B hold. The identified set $\mathcal{D}_p^*$ defined in (9) satisfies*

$$\mathcal{D}_p^* = \{\boldsymbol{\delta}_{k_{l,1}}(\cdot) : 1 \leq l \leq L_G\},$$

where $k_{l,1}$ is the initial vertex of the l-th topological sort of $G(V, E_p)$.

Suppose the DAG we recover from the data is not too sparse. By definition, a topological sort provides a ranking of regimes that is *not inconsistent* to the welfare ordering. Therefore, for given topological sort l, not just $\boldsymbol{\delta}_{k_{l,1}}(\cdot) \in \mathcal{D}_p^*$ but the full sequence

$$\left(\boldsymbol{\delta}_{k_{l,1}}(\cdot), \boldsymbol{\delta}_{k_{l,2}}(\cdot), ..., \boldsymbol{d}_{k_{l,|\mathcal{D}|}}(\cdot)\right) \tag{17}$$

[12] Theorem A.1 in the Appendix provides an alternative way of obtaining $\mathcal{D}_p^*$ based on directed paths of $G(V, E_p)$.

can be informative. A planner can be equipped with any of such sequences for $1 \leq l \leq L_G$ as a policy menu.

5.2 Bounds on Sorted Welfares

A topological sort provides ordinal information about counterfactual welfares. To gain more comprehensive knowledge about these welfare, a topological sort can be accompanied by cardinal information: bounds on the sorted welfares. One might especially be interested in the bounds on "top-tier" welfares that are associated with the identified set or the first few elements in the topological sort. Bounds on gains from adaptivity and regrets can also be computed. These bounds can be calculated by solving linear programming. For instance, for $k \in V$, the sharp lower and upper bounds on the welfare W_k can be calculated via

$$\begin{aligned} U_k &= \max\nolimits_{q\in\mathcal{Q}} A_k q, \\ L_k &= \min\nolimits_{q\in\mathcal{Q}} A_k q, \end{aligned} \qquad s.t. \quad Bq = p. \tag{18}$$

This computational approach to calculating bounds is inevitable in our context. Unlike in the static case of calculating bound on, e.g., the average treatment effect, calculating bounds on W_k and proving their sharpness are analytically infeasible, especially when $T \geq 3$. Fortunately, since the partial order and thus the topological sort are obtained analytically, we can focus on just a few welfares for which using linear programming is less of a burden than using it for all possible welfare gaps.

6 Additional Assumptions

Often, researchers are willing to impose more assumptions based on priors about agent's behaviors or the data-generating process. Examples are monotonicity/uniformity, agent's learning, Markovian structure, and stationarity. These assumptions are easy to incorporate within the linear programmings (10) and (18). These assumptions tighten the identified set $\mathcal{D}_p^*$ or the bounds on welfares by reducing the dimension of the simplex $\mathcal{Q}$, and thus producing a denser DAG.[13]

To incorporate these assumptions, we slightly revise the framework introduced in Section 4.3. Suppose h is $d_q \times 1$ vector of ones and zeros, where zeros are imposed by given identifying assumptions. Introduce $d_q \times d_q$ diagonal matrix $H = diag(h)$. Then, we can define a standard

[13] Similarly, when these assumptions are incorporated in (10), we obtain tighter bounds on welfares.

simplex for $\bar{q} \equiv Hq$ as

$$\bar{\mathcal{Q}} \equiv \{\bar{q} : \sum_s \bar{q}_s = 1 \text{ and } \bar{q}_s \geq 0 \ \forall s\}. \tag{19}$$

Note that the dimension of this simplex is smaller than the dimension d_q of $\mathcal{Q}$ if h contains zeros. Then we can modify (6) and (8) as

$$B\bar{q} = p,$$
$$W_k = A_k\bar{q},$$

respectively. Let $\boldsymbol{\delta}^*(\cdot;\bar{q}) \equiv \arg\max_{\boldsymbol{\delta}_k(\cdot)\in\mathcal{D}} W_k = A_k\bar{q}$. Then, the identified set with identifying assumptions coded in h is defined as

$$\bar{\mathcal{D}}_p^* \equiv \{\boldsymbol{\delta}^*(\cdot;\bar{q}) : B\bar{q} = p \text{ and } \bar{q} \in \mathcal{Q}\}, \tag{20}$$

which is assumed to be empty when $B\bar{q} \neq p$. Importantly, the latter occurs when any of the identifying assumptions are misspecified. Note that H is idempotent. Define $\bar{\Delta} \equiv \Delta H$ and $\bar{B} \equiv BH$. Then $\Delta\bar{q} = \bar{\Delta}\bar{q}$ and $B\bar{q} = \bar{B}\bar{q}$. Therefore, to generate the DAG and characterize the identified set, Theorem 4.1, Corollary 4.1, and Theorem 5.1 or A.1 can be modified by replacing q, B and Δ with $\bar{q}$, $\bar{B}$ and $\bar{\Delta}$, respectively.

We now list possible identifying assumptions. The first assumption is a sequential version of the monotonicity/uniformity assumption in Imbens and Angrist (1994).

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

Assumption M1 imposes that there is no defying (complying) behavior in the decision D_t conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^{t-1})$. Without conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^{t-1})$, however, there can be a general non-monotonic pattern in the way that $\boldsymbol{Z}^t$ influences $\boldsymbol{D}^t$. Recall $\tilde{S}_t \equiv (\{Y_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^t)\}, \{D_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^t)\}) \in \{0,1\}^{2^{2t-1}} \times \{0,1\}^{2^{3t-2}}$. For example, the no-defier assumption can be incorporated in (19) with h whose elements satisfy $h_s = 0$ for $s \in \{S = \beta(\tilde{\boldsymbol{S}}) : D_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^{t-1}, 1) = 0 \text{ and } D_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^{t-1}, 0) = 1 \ \forall t\}$ and $h_s = 1$ otherwise. By extending the idea of Vytlacil (2002), we can show that M1 is equivalent of imposing a threshold-crossing model for D_t under Assumption SX:

$$D_t = 1\{\pi_t(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^t) \geq \nu_t\}, \tag{21}$$

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

Lemma 6.1. *Suppose Assumption SX holds and* $\Pr[D_t = 1|\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^t]$ *is a nontrivial function of* Z_t*. Assumption M1 is equivalent to* (21) *being satisfied conditional on* $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^{t-1})$ *for each* t*.*

The dynamic selection model (21) should not be confused with the dynamic regime (1). Compared to the dynamic regime $d_t = \delta_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1})$, which is a hypothetical quantity, the equation (21) models each individual's observed treatment decision, in that it is not only a function of $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1})$ but also ν_t, the individual's unobserved characteristics. We assume that the social planner has no access to $\boldsymbol{\nu}$. The functional dependence of D_t on the past outcomes and treatments $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1})$ and a sequence of random assignments $(\boldsymbol{Z}^{t-1})$ reflects the agent's learning. Indeed, a specific version of such learning can be imposed as an additional identifying assumption:

Assumption L. *For each* t *and given* $\boldsymbol{z}^t$*,* $D_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^t) \geq D_t(\tilde{\boldsymbol{y}}^{t-1}, \tilde{\boldsymbol{d}}^{t-1}, \boldsymbol{z}^t)$ *w.p.1 for* $(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1})$ *and* $(\tilde{\boldsymbol{y}}^{t-1}, \tilde{\boldsymbol{d}}^{t-1})$ *such that* $\left\|\boldsymbol{y}^{t-1} - \boldsymbol{d}^{t-1}\right\| < \left\|\tilde{\boldsymbol{y}}^{t-1} - \tilde{\boldsymbol{d}}^{t-1}\right\|$ *(long memory) or* $y_{t-1} - d_{t-1} < \tilde{y}_{t-1} - \tilde{d}_{t-1}$ *(short memory).*

According to Assumption L, agents have the ability to revise his next period's decision based on his memory. To illustrate, consider the second period's decision, $D_2(y_1, d_1)$. Under Assumption L, an agent who would switch his treatment decision at $t = 2$ had he experienced bad health ($y_1 = 0$) after receiving the treatment ($d_1 = 1$), i.e., $D_2(0, 1) = 0$, would remain to take the treatment had he experienced good health, i.e., $D_2(1, 1) = 1$. More importantly, if an agent has not switched even after bad health, i.e., $D_2(0, 1) = 1$, it should only because of his unobserved preference, not because he cannot learn from the past, i.e., $D_2(1, 1) = 0$ cannot happen.[14]

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

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

As before, without conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^{t-1})$, there can be a general non-monotonic pattern in the way that $\boldsymbol{D}^t$ influences $\boldsymbol{Y}^t$. It is important to note that Assumption M2 (and M1) 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

[14] As suggested in this example, it is implicit in Assumption L that Y_t and D_t are of the same (or at least similar) types over time, which is not generally required for the analysis of this paper.

the direction. By a similar argument as before, Assumption M2 is equivalent of a dynamic version of a nonparametric triangular model under Assumption SX:

$$Y_t = 1\{\mu_t(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^t) \geq \varepsilon_t\}, \tag{22}$$

$$D_t = 1\{\pi_t(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^t) \geq \nu_t\}, \tag{23}$$

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

Lemma 6.2. *Suppose Assumption SX holds,* $\Pr[D_t = 1|\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^t]$ *is a nontrivial function of* Z_t*, and* $\Pr[Y_t = 1|\boldsymbol{Y}^{t-1}, \boldsymbol{D}^t]$ *is a nontrivial function of* D_t*. Assumption M2 is equivalent to (22)–(23) being satisfied conditional on* $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^{t-1})$ *for each* t*.*

As clearly seen in (22), Assumption M2 imposes non-trivial restrictions on treatment heterogeneity. To illustrate this point, consider an alternative specification for Y_t:

$$Y_t = 1\{\mu_t(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^t) \geq \varepsilon_t(D_t)\}, \tag{24}$$

where $\varepsilon_t(D_t) = D_t\varepsilon_t(1) + (1 - D_t)\varepsilon_t(0)$, which allows different unobservables for different treatment state d_t. This specification is more general than (22) as it effectively incorporates vector unobservables. We can slightly relax Assumption M2 by imposing (24) and assuming a sequential version of rank similarity (Chernozhukov and Hansen (2005)) that $\boldsymbol{\varepsilon}(1, \boldsymbol{d}_{-t}) \overset{d}{=} \boldsymbol{\varepsilon}(0, \boldsymbol{d}_{-t})$, conditional on $(\boldsymbol{\nu}^t, \boldsymbol{Z})$ for each t, where $\boldsymbol{\varepsilon}(\boldsymbol{d}) \equiv (\varepsilon_1(d_1), ..., \varepsilon_T(d_T))$. This assumption can be found in Han (Forthcoming).[15] Note that (22) postulates that $\varepsilon_t(d_t) = \varepsilon_t$ for all $d_t \in \{0, 1\}$ and t.

The last assumption imposes a Markov-type structure in the Y_t and D_t processes.

Assumption K. $Y_t|(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^t) \overset{d}{=} Y_t|(Y_{t-1}, D_t)$ *and* $D_t|(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^t) \overset{d}{=} D_t|(Y_{t-1}, D_{t-1}, Z_t)$ *for each* t*.*

In terms of the triangular model (22)–(23), Assumption K implies

$$Y_t = 1\{\mu_t(Y_{t-1}, D_t) \geq \varepsilon_t\},$$
$$D_t = 1\{\pi_t(Y_{t-1}, D_{t-1}, Z_t) \geq \nu_t\},$$

which yields a familiar structure of dynamic discrete choice models in the literature. When there are more than two periods, an assumption that imposes stationarity can be helpful for

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

identification. Such an assumption can be found in Torgovitsky (2019), which is not stated here for succinctness.

7 Cardinality Reduction

The typical time horizons we consider in this paper are short, say, $T \leq 5$. For example, a multi-stage experiment called the Fast Track Prevention Program (Conduct Problems Prevention Research Group (1992)) considers $T = 4$. When T is not small, the cardinality of $\mathcal{D}$ ($|\mathcal{D}| = 2^{2^T-1}$) may be too large and we may want to reduce it for computational, institutional, and practical purposes.

One way to reduce the cardinality is to reduce the dimension of the adaptivity. 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 $d_t \in \{0,1\}$:

$$\tilde{d}_t(y_{t-1}, d_{t-1}) = d_t$$

in the class $\tilde{\mathcal{D}}$. In this case, we have $\left|\tilde{\mathcal{D}}\right| = 2^{2T-1}$. An even simpler rule, $\tilde{d}_t(y_{t-1}) = d_t$, appears in Murphy et al. (2001).

Another possibility is to consider a strict subset $\tilde{\mathcal{D}}$ 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 the cardinality to $\left|\tilde{\mathcal{D}}\right| = 2 \times 2 \times \cdots \times 2 \times (2^{T-1} \cdot 2) = 2^{2T-1}$.

8 Numerical Studies

We conduct numerical exercises to illustrate (i) the main theoretical result developed in Section 4, (ii) the role of assumptions introduced in Section 6, and (iii) the overall computational scale of the problem. For $T = 2$, we consider the following data-generating process:

$$D_{i1} = 1\{\pi_1 Z_{i1} + \alpha_i + v_{i1} \geq 0\}, \tag{25}$$

$$Y_{i1} = 1\{\mu_1 D_{i1} + \alpha_i + e_{i1} \geq 0\}, \tag{26}$$

$$D_{i2} = 1\{\pi_{21} Y_{i1} + \pi_{22} D_{i1} + \pi_{23} Z_{i2} + \alpha_i + v_{i2} \geq 0\}, \tag{27}$$

$$Y_{i2} = 1\{\mu_{21} Y_{i1} + \mu_{22} D_{i2} + \alpha_i + e_{i2} \geq 0\}, \tag{28}$$

where $(v_1, e_1, v_2, e_2, \alpha)$ are mutually independent and jointly normally distributed, the endogeneity of D_{i1} and D_{i2} as well as the serial correlation is captured by the individual effect α_i, and (Z_1, Z_2) are Bernoulli, independent of $(v_1, e_1, v_2, e_2, \alpha)$. Notice that the process is intended to satisfy Assumptions SX, M2 and K. We consider a data-generating process where all the coefficients in (25)–(28) take positive values. In this exercise, we consider the welfare $W_k = E[Y_2(\boldsymbol{\delta}_k(\cdot))]$.

As shown in Table 1, there are eight possible regimes, i.e., $|\mathcal{D}| = 8$. Since the current exercise is of a small scale, instead of using the analytical algorithm proposed in Corollary 4.1 to generate the DAG, we directly calculate the lower and upper bounds $(L_{k,k'}, U_{k,k'})$ on welfare gap $W_k - W_{k'}$ for all pair $k, k' \in \{1, ..., 8\}$ $(k < k')$. This is also to illustrate the role of assumptions in improving the bounds. We conduct the bubble sort, which makes $\binom{8}{2} = 28$ pair-wise comparisons. That is, there are 28×2 linear programs to run. As a researcher, we impose Assumption K. Then, for each linear program, the dimension of q is $|\mathcal{Q}| + 1 = |\mathcal{S}| = |\mathcal{S}_1| \times |\mathcal{S}_2| = 2^2 \times 2^2 \times 2^8 \times 2^4 = 65,536$. The number of main constraints is $\dim(p) = 2^{3\times 2} = 16$. There are $1 + 65,536$ additional constraints that define the simplex, i.e., $\sum_s q_s = 1$ and $q_s \geq 0$ for all $s \in \mathcal{S}$. Each linear program takes less than a second to calculate $L_{k,k'}$ or $U_{k,k'}$ in a computer with 2.2 GHz single-core processor and 16 GB memory and with a modern solver such as CPLEX, MOSEK and GUROBI.

Figure 3 reports the bounds $(L_{k,k'}, U_{k,k'})$ on $W_k - W_{k'}$ for all $(k, k') \in \{1, ..., 8\}$ under Assumption M1 (in black) and Assumption M2 (in red). In the figure, we can determine the sign of the welfare gap for those bounds that do not include zero. The difference in black and red bounds illustrates the role of Assumption M2 relative to M1. That is, there are more bounds that avoid the zero vertical line, which is consistent with the theory. The bounds generate associated DAGs (produced as 8×8 adjacency matrices). We proceed with M2 for succinctness.

Figure 4 depicts the sharp DAG generated from $(L_{k,k'}, U_{k,k'})$'s under M2, based on Theorem 4.1(a). By (12), the identified set of $\boldsymbol{\delta}^*(\cdot)$ is

$$\mathcal{D}_p^* = \{\boldsymbol{\delta}_7(\cdot), \boldsymbol{\delta}_8(\cdot)\}.$$

Finally, the following is one of the topological sorts produced from the DAG:

$$(\boldsymbol{\delta}_8(\cdot), \boldsymbol{\delta}_4(\cdot), \boldsymbol{\delta}_7(\cdot), \boldsymbol{\delta}_3(\cdot), \boldsymbol{\delta}_5(\cdot), \boldsymbol{\delta}_1(\cdot), \boldsymbol{\delta}_6(\cdot), \boldsymbol{\delta}_2(\cdot)).$$

The bounds on the welfares in the order of this topological sort are shown in Figure 5.

We also conducted a parallel analysis but with a slightly different data-generating process,

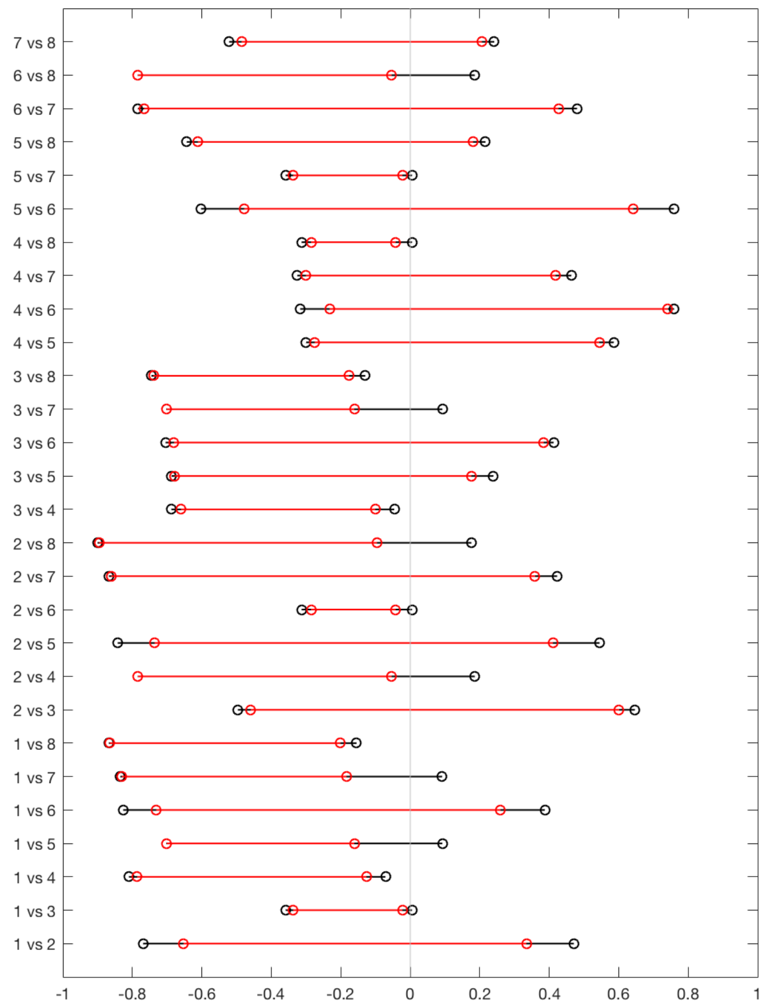


Figure 3: Sharp Bounds on Welfare Gaps under M1 (black) and M2 (red)

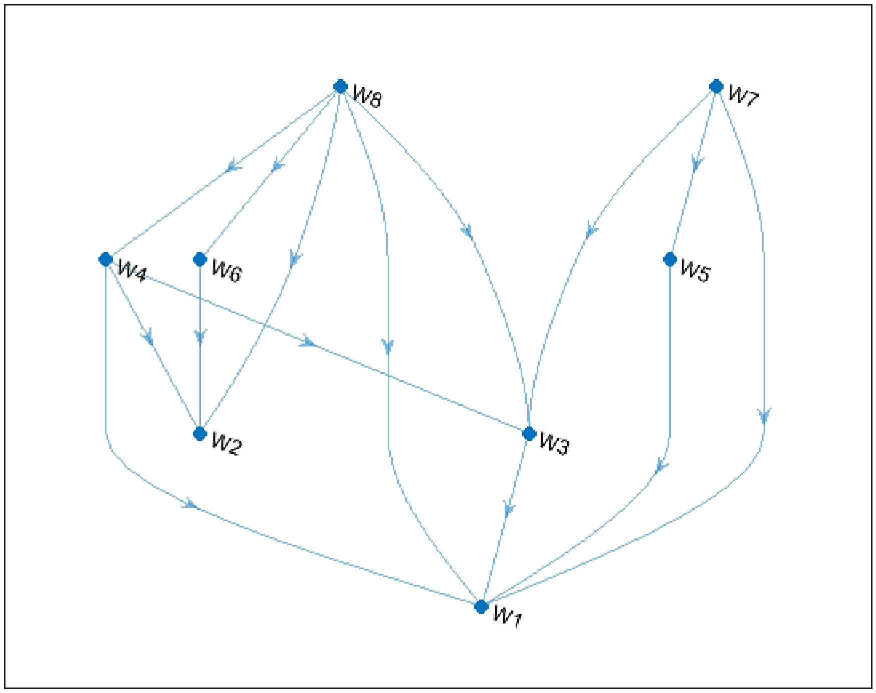


Figure 4: Sharp Directed Acyclic Graph under M2

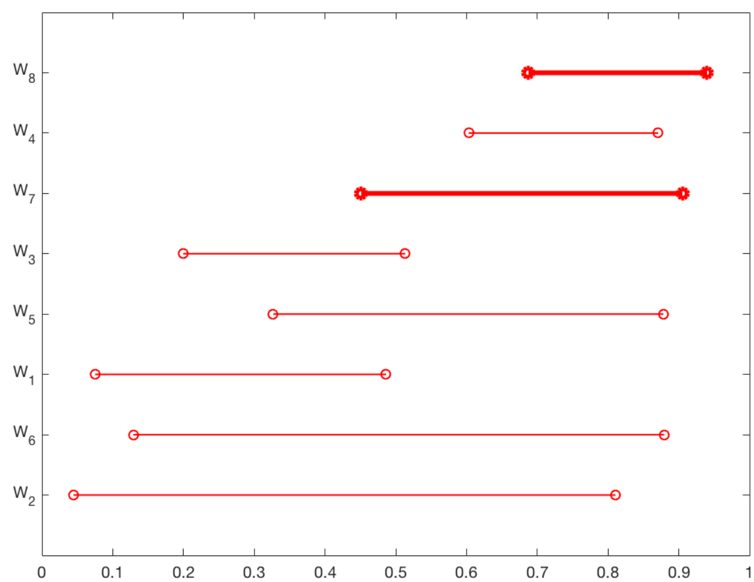


Figure 5: Sharp Bounds on Sorted Welfares under M2 (bold: for welfares with $\boldsymbol{\delta}(\cdot) \in \mathcal{D}_p^*$)

where all the coefficients in (25)–(28) are positive except $\mu_{22} < 0$. In this case, we obtain $\mathcal{D}_p^* = \{\boldsymbol{\delta}_2(\cdot)\}$ as a singleton, i.e., we point identify $\boldsymbol{\delta}^*(\cdot) = \boldsymbol{\delta}_2(\cdot)$.

9 Estimation and Inference

The estimation of the identified set $\mathcal{D}_p^*$ straightforward given the condition (16) of Corollary 4.1: $\theta_{k,k'} < \min\{0, \underline{\gamma}_{k,k'}\}$. The only unknown object in the condition is p, the joint distribution of $(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z})$, which can be estimated by $\hat{p}$, a vector of $\hat{p}_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}} = \sum_{i=1}^{N} 1\{\boldsymbol{Y}_i = \boldsymbol{y}, \boldsymbol{D}_i = \boldsymbol{d}, \boldsymbol{Z}_i = \boldsymbol{z}\} / \sum_{i=1}^{N} 1\{\boldsymbol{Z}_i = \boldsymbol{z}\}$. Then, with $\hat{\theta}_{k,k'} \equiv -\Delta_{k,k'}^{1} B_1^{-1} \hat{p}$, the estimated DAG is $G(V, \hat{E}_p)$, where

$$\hat{E}_p = \{(k, k') : \hat{\theta}_{k,k'} < \min\{0, \underline{\gamma}_{k,k'}\} \text{ for } k, k' \in V \text{ and } k \neq k'\}.$$

Let $\hat{\Sigma}$ be the resulting estimated adjacency matrix. Then, based on (12), $\mathcal{D}_p^*$ can be estimated as

$$\widehat{\mathcal{D}}_p^* = \{\boldsymbol{\delta}_{k'}(\cdot) : \hat{\Sigma}_{k,k'} = 0 \text{ for all } k \in V \text{ and } k \neq k' \in V\}.$$

Although we do not fully investigate in the current paper, we briefly discuss inference. To conduct inference on the optimal regime $\boldsymbol{\delta}^*(\cdot)$, we can construct a confidence set (CS) for $\mathcal{D}_p^*$ by the following procedure. We consider a sequence of hypothesis tests, where we eliminate regimes that are (statistically) significantly inferior to others. This is a statistical analog of the elimination procedure encoded in (12). This inference procedure extends the approach of Hansen et al. (2011) on the model confidence set, but in this novel context. For each test given $\tilde{V} \subset V$, we construct a null hypothesis that $W_k \sim W_{k'}$ for all $k, k' \in \tilde{V}$. According to (15), this hypothesis restricts the range of $\theta_{k,k'}$ so that the hyperplane $\theta_{k,k'} = \Delta_{k,k'}^{0} \boldsymbol{q}_0$ lies within the cone $\mathcal{Q}_2$. Based on the conditions (i)–(iii) in Theorem 4.2, this results in a one-sided test for

$$H_{0,\tilde{V}} : \left|\theta_{k,k'} - l_1(\Delta_{k,k'}^{0})\right| - l_2(\Delta_{k,k'}^{0}) \leq 0 \text{ for all } k, k' \in \tilde{V},$$

where l_1 and l_2 satisfy (i) $l_1(\Delta_{k,k'}^{0}) = (\overline{\gamma}_{k,k'} + \underline{\gamma}_{k,k'})/2$ and $l_2(\Delta_{k,k'}^{0}) = (\overline{\gamma}_{k,k'} - \underline{\gamma}_{k,k'})/2$ if $\underline{\gamma}_{k,k'} < 0 < \overline{\gamma}_{k,k'}$; (ii) $l_1(\Delta_{k,k'}^{0}) = \underline{\gamma}_{k,k'}/2$ and $l_2(\Delta_{k,k'}^{0}) = \underline{\gamma}_{k,k'}/2$ if $\underline{\gamma}_{k,k'} \geq 0$; (iii) $l_1(\Delta_{k,k'}^{0}) = \overline{\gamma}_{k,k'}/2$ and $l_2(\Delta_{k,k'}^{0}) = -\overline{\gamma}_{k,k'}/2$ if $\overline{\gamma}_{k,k'} \leq 0$, corresponding to the conditions (i)–(iii) in Theorem 4.2.

Then, the procedure of constructing the CS, denoted as $\widehat{\mathcal{D}}_{CS}$, is as follows: *Step 0. Initially set $\tilde{V} = V$. Step 1. Test $H_{0,\tilde{V}}$ at level α with a test function $\phi_{\tilde{V}} \in \{0, 1\}$. Step 2. If $H_{0,\tilde{V}}$ is not rejected, define $\widehat{\mathcal{D}}_{CS} = \{\boldsymbol{\delta}_k(\cdot) : k \in \tilde{V}\}$; otherwise eliminate a vertex $k_{\tilde{V}}$ from $\tilde{V}$ and repeat from Step 1.* In Step 1, $T_{\tilde{V}} \equiv \max_{k,k' \in \tilde{V}} t_{k,k'}$ can be used as the test statistic for $H_{0,\tilde{V}}$

where $t_{k,k'}$ is a standard t-statistic, i.e., the ratio between $\left|\hat{\theta}_{k,k'} - l_1(\Delta^0_{k,k'})\right| - l_2(\Delta^0_{k,k'})$ and its standard error. The distribution of $T_{\tilde{V}}$ can be estimated using bootstrap. In Step 2, a candidate for $k_{\tilde{V}}$ is $k_{\tilde{V}} \equiv \arg\max_{k \in \tilde{V}} \max_{k' \in \tilde{V}} t_{k,k'}$. Following Hansen et al. (2011), we can show that the resulting CS has desirable properties. Let $H_{A,\tilde{V}}$ be the alternative hypothesis.

Assumption CS. *For any $\tilde{V}$, (i) $\limsup_{n\to\infty} \Pr[\phi_{\tilde{V}} = 1 | H_{0,\tilde{V}}] \le \alpha$, (ii) $\lim_{n\to\infty} \Pr[\phi_{\tilde{V}} = 1 | H_{A,\tilde{V}}] = 1$, and (iii) $\lim_{n\to\infty} \Pr[\boldsymbol{\delta}_{k_{\tilde{V}}}(\cdot) \in \mathcal{D}^*_p | H_{A,\tilde{V}}] = 0$.*

Proposition 9.1. *Under Assumption CS, it satisfies that $\liminf_{n\to\infty} \Pr[\mathcal{D}^*_p \subset \widehat{\mathcal{D}}_{CS}] \ge 1 - \alpha$ and $\lim_{n\to\infty} \Pr[\boldsymbol{\delta}(\cdot) \in \widehat{\mathcal{D}}_{CS}] = 0$ for all $\boldsymbol{\delta}(\cdot) \notin \mathcal{D}^*_p$.*

The procedure of the CS construction does not suffer from the problem of multiple testings. This is because the procedure stops as soon as the first hypothesis is not rejected, and asymptotically, maximal elements will not be questioned before all sub-optimal regimes are eliminated; see Hansen et al. (2011) for related discussions. The resulting CS can also be used to conduct a specification test for a less palatable assumption such as Assumption M2. We can reject the assumption, when the CS under that assumption is empty.

Inference on the welfare bounds can be conducted by using recent results as in Deb et al. (2017), who develop uniformly valid inference for bounds obtained via linear programming. Inference on optimized welfare $W_{\boldsymbol{\delta}^*}$ or $\max_{\boldsymbol{\delta}(\cdot) \in \widehat{\mathcal{D}}_{CS}} W_{\boldsymbol{\delta}}$ can also be an interesting problem. Andrews et al. (2019) considers inference on optimized welfare (evaluated at the estimated policy) in the context of Kitagawa and Tetenov (2018), but with point identified welfare under the unconfoundedness assumption for the treatment. Extending the framework to a setting with partially identified welfare and dynamic regimes will be another interesting future work.

A Appendix

A.1 Finite-Horizon Dynamic Programming

Suppose $W_{\boldsymbol{\delta}} = E[Y_T(\boldsymbol{\delta}(\cdot))]$. Then, it satisfies that

$$E[Y_T(\boldsymbol{\delta}(\cdot))] = E\left[E\left[\cdots E\left[E[Y_T(\boldsymbol{d}) | \boldsymbol{Y}^{T-1}(\boldsymbol{d}^{T-1})] \middle| \boldsymbol{Y}^{T-2}(\boldsymbol{d}^{T-2})\right] \cdots \middle| Y_1(d_1)\right]\right], \tag{29}$$

where the bridge variables $\boldsymbol{d} = (d_1, ..., d_T)$ satisfies

$$\begin{aligned} d_1 &= \delta_1, \\ d_2 &= \delta_2(Y_1(d_1), d_1), \\ d_3 &= \delta_3(\boldsymbol{Y}^2(\boldsymbol{d}^2), \boldsymbol{d}^2), \\ &\vdots \\ d_T &= \delta_T(\boldsymbol{Y}^{T-1}(\boldsymbol{d}^{T-1}), \boldsymbol{d}^{T-1}). \end{aligned}$$

Given (29), the solution $\boldsymbol{\delta}^*(\cdot)$ can be justified by backward induction in a finite-horizon dynamic programming for $W_{\boldsymbol{\delta}} \equiv E[Y_T(\boldsymbol{\delta}(\cdot))]$. First, the T-th element in $\boldsymbol{\delta}^*(\cdot)$ corresponds to the optimal rule in the final period:

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

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

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

and $V_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}) \equiv \max_{d_t} E[V_{t+1}(\boldsymbol{Y}^t(\boldsymbol{d}^t), \boldsymbol{d}^t)|\boldsymbol{Y}^{t-1}(\boldsymbol{d}^{t-1}) = \boldsymbol{y}^{t-1}]$, which then iteratively defines all the elements in $\boldsymbol{\delta}^*(\cdot)$.[16] By definition, $\boldsymbol{\delta}^*(\cdot)$ is adaptive to past outcomes and treatments. To illustrate, when $T = 2$, we have

$$\delta_2^*(y_1, d_1) = \arg\max_{d_2} E[Y_2(\boldsymbol{d})|Y_1(d_1) = y_1], \tag{30}$$

and, by defining $V_2(y_1, d_1) \equiv \max_{d_2} E[Y_2(\boldsymbol{d})|Y_1(d_1) = y_1]$,

$$\delta_1^* = \arg\max_{d_1} E[V_2(Y_1(d_1), d_1)]. \tag{31}$$

Therefore, $\boldsymbol{\delta}^*(\cdot)$ is equal to the collection of these solutions: $\boldsymbol{\delta}^*(\cdot) = (\delta_1^*, \delta_2^*(\cdot))$. A similar argument can be made with a general $W_{\boldsymbol{\delta}}$.

[16] 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 (Forthcoming)).

A.2 Matrices in Section 4.3

We show how to construct matrices A_k and B in (6) and (8) for the linear programming (10). The construction of A_k and B uses the fact that any linear functional of $\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}, \boldsymbol{D}(\boldsymbol{z}) = \boldsymbol{d}]$ can be characterized as a linear combination of q_s. Although the notation of this section can be somewhat heavy, if one is committed to the use of linear programming instead of an analytic solution, most of the derivation can be systematically reproduced in a standard software such as MATLAB and Python.

Consider B first. By Assumption SX and the definition of S_t and R_t, we have

$$\begin{aligned}
p_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}} &= \Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}, \boldsymbol{D}(\boldsymbol{z}) = \boldsymbol{d}] \\
&= \Pr[\boldsymbol{Y}(\boldsymbol{y}^{T-1}, \boldsymbol{d}) = \boldsymbol{y}, \boldsymbol{D}(\boldsymbol{y}^{T-1}, \boldsymbol{d}^{T-1}, \boldsymbol{z}) = \boldsymbol{d}] \\
&= \Pr[S : Y_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^t) = y_t, D_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^t) = d_t \quad \forall t] \\
&= \sum_{s \in \mathcal{S}_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}}} q_s \qquad (32)
\end{aligned}$$

where $\mathcal{S}_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}} \equiv \{S = \beta(\tilde{\boldsymbol{S}}) : Y_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^t) = y_t, D_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^t) = d_t \ \forall t\}$ and $\tilde{S}_t \equiv (\{Y_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^t)\}, \{D_t(\boldsymbol{y}^{t-1},$ Then, for a $d_q \times 1$ vector $B_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}}$,

$$p_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}} = \sum_{s \in I_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}}} q_s = B_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}} q$$

and the $d_q \times d_p$ matrix B stacks $B_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}}$ so that $p = Bq$.

For A_k, recall $W_{\boldsymbol{\delta}_k}$ is a linear functional of $q_{\boldsymbol{\delta}_k}(\boldsymbol{y}) \equiv \Pr[\boldsymbol{Y}(\boldsymbol{\delta}_k(\cdot)) = \boldsymbol{y}]$. We first find a relationship between $\Pr[\boldsymbol{Y}(\boldsymbol{\delta}(\cdot)) = \boldsymbol{y}]$ and $\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}]$. For fixed $\tilde{\boldsymbol{y}}$, by definition (2), we can rewrite

$$\Pr[\boldsymbol{Y}(\boldsymbol{\delta}(\cdot)) = \tilde{\boldsymbol{y}}] = E\left[E\left[\cdots E\left[\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y} | \boldsymbol{Y}^{T-1}(\boldsymbol{d}^{T-1})] \middle| \boldsymbol{Y}^{T-2}(\boldsymbol{d}^{T-2})\right] \cdots \middle| Y_1(d_1)\right]\right], \qquad (33)$$

where the bridge variables $\boldsymbol{d} = (d_1, ..., d_T)$ satisfies

$$\begin{aligned}
d_1 &= \delta_1, \\
d_2 &= \delta_2(Y_1(d_1), d_1), \\
d_3 &= \delta_3(\boldsymbol{Y}^2(\boldsymbol{d}^2), \boldsymbol{d}^2), \\
&\vdots \\
d_T &= \delta_T(\boldsymbol{Y}^{T-1}(\boldsymbol{d}^{T-1}), \boldsymbol{d}^{T-1}).
\end{aligned}$$

By repetitively applying the law of iterated expectation, we can show that the r.h.s. of (33) can be expressed as

$$\begin{aligned}
&\sum_{y_1} \cdots \sum_{\boldsymbol{y}^{T-2}} \sum_{\boldsymbol{y}^{T-1}} \Pr[\boldsymbol{Y}(\boldsymbol{d}) = \tilde{\boldsymbol{y}} | \boldsymbol{Y}^{T-1}(\boldsymbol{d}^{T-1}) = \boldsymbol{y}^{T-1}] \\
&\qquad\qquad \times \Pr[\boldsymbol{Y}^{T-1}(\boldsymbol{d}^{T-1}) = \boldsymbol{y}^{T-1} | \boldsymbol{Y}^{T-2}(\boldsymbol{d}^{T-2}) = \boldsymbol{y}^{T-2}] \times \cdots \times \Pr[Y_1(d_1) = y_1] \\
=&\sum_{y_1} \cdots \sum_{y_{T-2}} \sum_{y_{T-1}} \Pr[\boldsymbol{Y}(\boldsymbol{d}) = \tilde{\boldsymbol{y}} | \boldsymbol{Y}^{T-1}(\boldsymbol{d}^{T-1}) = \boldsymbol{y}^{T-1}] \\
&\qquad\qquad \times \Pr[Y_{T-1}(\boldsymbol{d}^{T-1}) = y_{T-1} | \boldsymbol{Y}^{T-2}(\boldsymbol{d}^{T-2}) = \boldsymbol{y}^{T-2}] \times \cdots \times \Pr[Y_1(d_1) = y_1],
\end{aligned} \tag{34}$$

where the last summation is simply a linear functional of $\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}]$, since $\boldsymbol{d} = (d_1, ..., d_T)$ satisfies

$$\begin{aligned}
d_1 &= \delta_1, \\
d_2 &= \delta_2(y_1, d_1), \\
d_3 &= \delta_3(\boldsymbol{y}^2, \boldsymbol{d}^2), \\
&\vdots \\
d_T &= \delta_T(\boldsymbol{y}^{T-1}, \boldsymbol{d}^{T-1}).
\end{aligned}$$

Notice that the bridge variables are no longer random due to appropriate conditioning in (34). To illustrate, when $T = 2$, the welfare defined as the average counterfactual terminal outcome satisfies

$$\begin{aligned}
E[Y_T(\boldsymbol{\delta}(\cdot))] &= \sum_{y_1} \Pr[Y_2(\delta_1, \delta_2(y_1, \delta_1)) = 1 | Y_1(\delta_1) = y_1] \Pr[Y_1(\delta_1) = y_1] \\
&= \sum_{y_1} \Pr[Y_2(\delta_1, \delta_2(y_1, \delta_1)) = 1, Y_1(\delta_1) = y_1].
\end{aligned} \tag{35}$$

For a chosen $\boldsymbol{\delta}(\cdot)$, the value d at which $Y_2(d)$ and $Y_1(d_1)$ are defined is given in Table 1 as shown in the main text.

Now, define a linear functional $h_k(\cdot)$ that (i) marginalizes $\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}, \boldsymbol{D}(\boldsymbol{z}) = \boldsymbol{d}]$ into $\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}]$ and then (ii) maps $\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}]$ into $\Pr[\boldsymbol{Y}(\boldsymbol{\delta}_k(\cdot)) = \boldsymbol{y}]$ according to (33) and

(34). But recall that $\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}, \boldsymbol{D}(\boldsymbol{z}) = \boldsymbol{d}] = \sum_{s \in \mathcal{S}_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}}} q_s$ by (32). Consequently, we have

$$
\begin{aligned}
W_k = f(q_{\boldsymbol{\delta}_k}) &= f(\Pr[\boldsymbol{Y}(\boldsymbol{\delta}_k(\cdot)) = \cdot]) \\
&= f \circ h_k(\Pr[\boldsymbol{Y}(\cdot) = \cdot, \boldsymbol{D}(\boldsymbol{z}) = \cdot]), \\
&= f \circ h_k \left(\sum_{s \in \mathcal{S}_{\cdot,\cdot|\boldsymbol{z}}} q_s \right) \equiv A_k q.
\end{aligned}
$$

To continue the illustration (7) in the main text, note that

$$
\Pr[\boldsymbol{Y}(1,1) = (1,1)] = \Pr[S : Y_1(1) = 1, Y_2(1,1) = 1] = \sum_{s \in \mathcal{S}_{11}} q_s,
$$

where $\mathcal{S}_{11} \equiv \{S = \beta(\tilde{S}_1, \tilde{S}_2) : Y_1(1) = 1, Y_2(1,1) = 1\}$. Similarly, we have

$$
\Pr[\boldsymbol{Y}(1,1) = (0,1)] = \Pr[S : Y_1(1) = 0, Y_2(1,1) = 1] = \sum_{s \in \mathcal{S}_{01}} q_s,
$$

where $\mathcal{S}_{01} \equiv \{S = \beta(\tilde{S}_1, \tilde{S}_2) : Y_1(1) = 0, Y_2(1,1) = 1\}$.

A.3 Proof of Theorem 4.2

Since $\mathcal{Q}_0$ is a finitely generated cone, finding conditions under which (15) holds is equivalent to finding conditions under which $\Delta^0_{k,k'} \boldsymbol{q}_0 = \theta_{k,k'}$ intersects one of the edges of the cone: $\{\boldsymbol{q}_0 : q_s + q_{s'} = 1 \text{ for } s, s' \in \mathcal{S}_0 \text{ and other elements are zero}\}$ or $\{\boldsymbol{q}_0 : q_s \in [0,1] \text{ for } s \in \mathcal{S}_0 \text{ and other elements are zero}\}$. First, consider condition (i) in the theorem. Choose $\boldsymbol{q}_0$ such that $q_s = t$, $q_{s'} = 1 - t$, and other elements are equal to zero. Then,

$$
\theta_{k,k'} = \Delta^0_{k,k'} \boldsymbol{q}_0 = \gamma_s t + \gamma_{s'}(1-t) = (\gamma_s - \gamma_{s'})t + \gamma_{s'}
$$

if and only if

$$
t = \frac{\theta_{k,k'} - \gamma_{s'}}{\gamma_s - \gamma_{s'}}.
$$

But then $t \in [0,1]$ by (i), and thus such $\boldsymbol{q}_0 \in \mathcal{Q}_0$. Therefore, (15) holds.

Next, consider condition (ii) in the theorem. Choose $\boldsymbol{q}_0$ such that q_s is possibly nonzero for given $s \in \mathcal{S}_0$, while all other elements are zero. Then,

$$
\theta_{k,k'} = \Delta^0_{k,k'} \boldsymbol{q}_0 = \gamma_s q_s
$$

if and only if $q_s = \theta_{k,k'}/\gamma_s$ (assuming $\gamma_s \neq 0$), which is in $[0,1]$ by (ii), and thus such $\boldsymbol{q}_0 \in \mathcal{Q}_0$. In this case, when $\gamma_s = 0$, then we trivially have $\boldsymbol{q}_0 \in \mathcal{Q}_0$. Therefore, (15) holds. The proof with condition (iii) is symmetric, so omitted. □

A.4 Proof of Corollary 4.1

Note that $\gamma_s > \theta_{k,k'}$ for all $s \in \mathcal{S}_0$, then $\sum_{s\in\mathcal{S}_0} \gamma_s q_s \geq \theta_{k,k'} \sum_{s\in\mathcal{S}_0} q_s$ since $q_s \geq 0$ for all $s \in \mathcal{S}_0$. But $\theta_{k,k'} \sum_{s\in\mathcal{S}_0} q_s \geq \theta_{k,k'}$ since $\sum_{s\in\mathcal{S}_0} q_s \leq 1$ and $\theta_{k,k'} < 0$. Combining these results, we have $\Delta_{k,k'} q = \Delta^0_{k,k'} \boldsymbol{q}_0 - \theta_{k,k'} \geq 0$ for any $q \in \mathcal{Q}$, or equivalently, $L_{k,k'} \geq 0$. Conversely, when (i) is violated, the case falls into either one of the three conditions in Theorem 4.2 or a condition that $\gamma_s < \theta_{k,k'}$ for all $s \in \mathcal{S}_0$ and $\theta_{k,k'} > 0$. The former case implies incomparability which contradicts $L_{k,k'} > 0$. The latter case implies either $L_{k,k'} < 0$ (by a symmetric argument but with $q_s > 0$ for all $s \in \mathcal{S}_0$) which is contradiction, or $L_{k,k'} = -\theta_{k,k'}$ with $q_s = 0$ for all $s \in \mathcal{S}_0$ and thus $L_{k,k'} < 0$, which is again contradiction. This proves necessity and sufficiency of the condition. □

A.5 Alternative Characterization of the Identified Set

Given the DAG, the identified set of $\boldsymbol{\delta}^*(\cdot)$ can also be obtained as the collection of initial vertices of all the directed paths of the DAG. For a DAG $G(V,E)$, a *directed path* is a subgraph $G(V_j, E_j)$ $(1 \leq j \leq J \leq 2^{|\mathcal{D}|})$ where $V_j \subset V$ is a totally ordered set with initial vertex $\tilde{k}_{j,1}$.[17] In stating our main theorem, we make it explicit that the DAG calculated by the linear programming is a function of the data distribution p.

Theorem A.1. *Suppose Assumptions SX and B hold. Then, $\mathcal{D}^*_p$ defined in (9) satisfies*

$$\mathcal{D}^*_p = \{\boldsymbol{\delta}_{\tilde{k}_{j,1}}(\cdot) \in \mathcal{D} : 1 \leq j \leq J\}, \tag{36}$$

where $\tilde{k}_{j,1}$ is the initial vertex of the directed path $G(V_{p,j}, E_{p,j})$ of $G(V, E_p)$.

A.6 Proof of Theorem A.1

Let $\tilde{\mathcal{D}}^* \equiv \{\boldsymbol{\delta}_{\tilde{k}_{j,1}}(\cdot) \in \mathcal{D} : 1 \leq j \leq J\}$. First, note that since $\tilde{k}_{j,1}$ is the initial vertex of directed path j, it should be that $W_{\tilde{k}_{j,1}} \geq W_{\tilde{k}_{j,m}}$ for any $\tilde{k}_{j,m}$ in that path by definition. We begin by supposing $\mathcal{D}^*_p \supset \tilde{\mathcal{D}}^*$. Then, there exist $\boldsymbol{\delta}^*(\cdot; q) = \arg\max_{\boldsymbol{\delta}_k(\cdot)\in\mathcal{D}} A_k q$ for some q that satisfies $Bq = p$ and $q \in \mathcal{Q}$, but which is not the initial vertex of any directed path. Such $\boldsymbol{\delta}^*(\cdot; q)$

[17] For example, in Figure 1(a), there are two directed paths ($J = 2$) with $V_1 = \{1, 2, 3\}$ ($\tilde{k}_{1,1} = 1$) and $V_2 = \{2, 3, 4\}$ ($\tilde{k}_{2,1} = 4$).

cannot be other (non-initial) vertices of any paths as it is contradiction by the definition of $\boldsymbol{\delta}^*(\cdot;q)$. But the union of all directed paths is equal to the original DAG, therefore there cannot exist such $\boldsymbol{\delta}^*(\cdot;q)$.

Now suppose $\mathcal{D}_p^* \subset \tilde{\mathcal{D}}^*$. Then, there exists $\boldsymbol{\delta}_{\tilde{k}_{j,1}}(\cdot) \neq \boldsymbol{\delta}^*(\cdot;q) = \arg\max_{\boldsymbol{\delta}_k(\cdot)\in\mathcal{D}} A_k q$ for some q that satisfies $Bq = p$ and $q \in \mathcal{Q}$. This implies that $W_{\tilde{k}_{j,1}} < W_{\tilde{k}}$ for some $\tilde{k}$. But $\tilde{k}$ should be a vertex of the same directed path (because $W_{\tilde{k}_{j,1}}$ and $W_{\tilde{k}}$ are ordered), but then it is contradiction as $\tilde{k}_{j,1}$ is the initial vertex. Therefore, $\mathcal{D}_p^* = \tilde{\mathcal{D}}^*$. $\square$

A.7 Proof of Theorem 5.1

Given Theorem A.1, proving $\tilde{\mathcal{D}}^* = \{\boldsymbol{\delta}_{k_{l,1}}(\cdot) : 1 \leq l \leq L_G\}$ will suffice. Recall $\tilde{\mathcal{D}}^* \equiv \{\boldsymbol{\delta}_{\tilde{k}_{j,1}}(\cdot) \in \mathcal{D} : 1 \leq j \leq J\}$ where $\tilde{k}_{j,1}$ is the initial vertex of the directed path $G(V_{p,j}, E_{p,j})$. When all topological sorts are singletons, the proof is trivial so we rule out this possibility. Suppose $\tilde{\mathcal{D}}^* \supset \{\boldsymbol{\delta}_{k_{l,1}}(\cdot) : 1 \leq l \leq L_G\}$. Then, for some l, there should exist $\boldsymbol{\delta}_{k_{l,m}}(\cdot)$ for some $m \neq 1$ that is contained in $\tilde{\mathcal{D}}^*$ but not in $\{\boldsymbol{\delta}_{k_{l,1}}(\cdot) : 1 \leq l \leq L_G\}$, i.e., that satisfies either (i) $W_{k_{l,1}} > W_{k_{l,m}}$ or (ii) $W_{k_{l,1}}$ and $W_{k_{l,m}}$ are incomparable and thus either $W_{k_{l',1}} > W_{k_{l,m}}$ for some $l' \neq l$ or $W_{k_{l,m}}$ is a singleton in another topological sort. Consider case (i). If $\boldsymbol{\delta}_{k_{l,1}}(\cdot) \in \mathcal{D}_j$ for some j, then it should be that $\boldsymbol{\delta}_{k_{l,m}}(\cdot) \in \mathcal{D}_j$ as $\boldsymbol{\delta}_{k_{l,1}}(\cdot)$ and $\boldsymbol{\delta}_{k_{l,m}}(\cdot)$ are comparable in terms of welfare, but then $\boldsymbol{\delta}_{k_{l,m}}(\cdot) \in \tilde{\mathcal{D}}^*$ contradicts the fact that $\boldsymbol{\delta}_{k_{l,1}}(\cdot)$ the initial vertex of the topological sort. Consider case (ii). The singleton case is trivially rejected since if the topological sort a singleton, then $\boldsymbol{\delta}_{k_{l,m}}(\cdot)$ should have been already in $\{\boldsymbol{\delta}_{k_{l,1}}(\cdot) : 1 \leq l \leq L_G\}$. In the other case, since the two welfares are not comparable, it should be that $\boldsymbol{\delta}_{k_{l,m}}(\cdot) \in \mathcal{D}_{j'}$ for $j' \neq j$. But $\boldsymbol{\delta}_{k_{l,m}}(\cdot)$ cannot be the one that delivers the largest welfare since $W_{k_{l',1}} > W_{k_{l,m}}$ where $\boldsymbol{\delta}_{k_{l',1}}(\cdot)$. Therefore $\boldsymbol{\delta}_{k_{l,m}}(\cdot) \in \tilde{\mathcal{D}}^*$ is contradiction. Therefore there is no element in $\tilde{\mathcal{D}}^*$ that is not in $\{\boldsymbol{\delta}_{k_{l,1}}(\cdot) : 1 \leq l \leq L_G\}$.

Now suppose $\tilde{\mathcal{D}}^* \subset \{\boldsymbol{\delta}_{k_{l,1}}(\cdot) : 1 \leq l \leq L_G\}$. Then for l such that $\boldsymbol{\delta}_{k_{l,1}}(\cdot) \notin \tilde{\mathcal{D}}^*$, either $W_{k_{l,1}}$ is a singleton or $W_{k_{l,1}}$ 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 $\boldsymbol{\delta}_{k_{l,1}}(\cdot) \notin \tilde{\mathcal{D}}^*$ is contradiction. In the other case, if $W_{k_{l,1}}$ is a maximum welfare, then $\boldsymbol{\delta}_{k_{l,1}}(\cdot) \notin \tilde{\mathcal{D}}^*$ 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 $\{\boldsymbol{\delta}_{k_{l,1}}(\cdot) : 1 \leq l \leq L_G\}$ or not. This concludes the proof that $\tilde{\mathcal{D}}^* = \{\boldsymbol{\delta}_{k_{l,1}}(\cdot) : 1 \leq l \leq L_G\}$. $\square$

A.8 Proof of Lemma 6.1

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 (21) 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 (21) 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:

$$Y_t(\boldsymbol{a}^t), A_t(\boldsymbol{z}^t) \perp \boldsymbol{Z}^t | (\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^{t-1}). \tag{37}$$

Conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^{t-1})$, (21) and (37) correspond to Assumption S-1 in Vytlacil (2002). Assumption R(i) and (37) 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.9 Proof of Lemma 6.2

We are remained to prove that, conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{A}^{t-1}, \boldsymbol{Z}^{t-1})$, (22) is equivalent to the second part of Assumption M2. But this proof is analogous to the proof of Lemma 6.1 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. □

References

Abraham, S. and L. Sun (2019): "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *MIT*. 1

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

Athey, S. and G. W. Imbens (2018): "Design-based Analysis in Difference-In-Differences Settings with Staggered Adoption," *National Bureau of Economic Research*. 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

Callaway, B. and P. H. Sant'Anna (2019): "Difference-in-Differences with Multiple Time Periods and an Application on the Minimum Wage and Employment," *Temple University and Vanderbilt University*. 1

Chernozhukov, V. and C. Hansen (2005): "An IV model of quantile treatment effects," *Econometrica*, 73, 245–261. 6

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. 3, 7

Deb, R., Y. Kitamura, J. K.-H. Quah, and J. Stoye (2017): "Revealed price preference: Theory and stochastic testing," . 1, 9

Gunsilius, F. (2019): "Bounds in continuous instrumental variable models," *arXiv preprint arXiv:1910.09502*. 1

Han, S. (Forthcoming): "Nonparametric Identification in Models for Dynamic Treatment Effects," *Journal of Econometrics*. 1, 2.1, 6, 15, 16

Hansen, P. R., A. Lunde, and J. M. Nason (2011): "The model confidence set," *Econometrica*, 79, 453–497. 9, 9

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, 6

Kahn, A. B. (1962): "Topological sorting of large networks," *Communications of the ACM*, 5, 558–562. 5.1

Kamat, V. (2017): "Identification with latent choice sets: The case of the head start impact study," *arXiv preprint arXiv:1711.02048*. 1

Kitagawa, T. and A. Tetenov (2018): "Who should be treated? empirical welfare maximization methods for treatment choice," *Econometrica*, 86, 591–616. 9, 16

Kitamura, Y. and J. Stoye (2019): "Nonparametric Counterfactuals in Random Utility Models," *arXiv preprint arXiv:1902.08350*. 1

Machado, C., A. Shaikh, and E. Vytlacil (Forthcoming): "Instrumental variables and the sign of the average treatment effect," *Journal of Econometrics*. 1, 4.4

Manski, C. F. (1997): "Monotone treatment response," *Econometrica: Journal of the Econometric Society*, 1311–1334. 6

——— (2004): "Statistical treatment rules for heterogeneous populations," *Econometrica*, 72, 1221–1246. 1, 16

——— (2007): "Partial identification of counterfactual choice probabilities," *International Economic Review*, 48, 1393–1410. 1

Mogstad, M., A. Santos, and A. Torgovitsky (2018): "Using instrumental variables for inference about policy relevant treatment parameters," *Econometrica*, 86, 1589–1619. 1

Murphy, S. A. (2003): "Optimal dynamic treatment regimes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 331–355. 1, 2.1, 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

——— (2004): "Optimal structural nested models for optimal sequential decisions," in *Proceedings of the second seattle Symposium in Biostatistics*, Springer, 189–326. 1

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. 3

Torgovitsky, A. (2019): "Nonparametric Inference on State Dependence in Unemployment," *Econometrica*, 87, 1475–1505. 1, 6

Vytlacil, E. (2002): "Independence, monotonicity, and latent index models: An equivalence result," *Econometrica*, 70, 331–341. 6, A.8

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, 2.1