Optimal Dynamic Treatment Regimes and Partial Welfare Ordering*

Sukjin (Vincent) Han
Department of Economics
University of Bristol

First Draft: August 8, 2019
This Draft: October 9, 2020

Abstract

Dynamic treatment regimes are treatment allocations tailored to heterogeneous individuals. The optimal dynamic treatment regime is a regime that maximizes counterfactual welfare. We introduce a framework in which we can partially learn the optimal dynamic regime from observational data, relaxing the sequential randomization assumption commonly employed in the literature but instead using (binary) instrumental variables. We propose the notion of sharp partial ordering of counterfactual welfares with respect to dynamic regimes and establish mapping from data to partial ordering via a set of linear programs. We then characterize the identified set of the optimal regime as the set of maximal elements associated with the partial ordering. One main contribution of this paper is that we develop simple analytical conditions to establish the ordering, which bypass solving a large number of large-scale linear programs, and thus facilitate estimation and inference. This paper's analytical framework has broader applicability beyond the current context, e.g., in establishing signs of various treatment effects and rankings of policies across different counterfactual scenarios.

JEL Numbers: C14, C32, C33, C36

Keywords: Optimal dynamic treatment regimes, endogenous treatments, dynamic treatment effect, partial identification, 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, Shenshen Yang, participants in the 2020 World Congress of the Econometric Society, 2019 CEMMAP & WISE conference, the 2019 Asian Meeting of the Econometric Society, the Bristol Econometrics Study Group Conference, the 2019 Midwest Econometrics Group Conference, the 2019 Southern Economic Association Conference, and in seminars at UCL, Rice, UIUC, Bristol, Queen Mary London, NUS, and SMU.
vincent.han@bristol.ac.uk

1 Introduction

Dynamic (i.e., adaptive) treatment regimes are personalized treatment allocations tailored to heterogenous individuals in order to improve welfare. Define a dynamic treatment regime $\boldsymbol{\delta}(\cdot)$ as a sequence of binary rules $\delta_t(\cdot)$ that map previous outcomes and treatments (and possibly other covariates) onto current allocation decisions: $\delta_t(y_1, ..., y_{t-1}, d_1, ..., d_{t-1}) = d_t \in \{0, 1\}$ for $t = 1, ..., T$. Then the optimal dynamic treatment regime, which is this paper's main parameter of interest, is defined as a regime that maximizes certain counterfactual welfare:

$$\boldsymbol{\delta}^*(\cdot) = \arg\max_{\boldsymbol{\delta}(\cdot)} W_{\boldsymbol{\delta}}. \tag{1.1}$$

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

Optimal treatment regimes have been extensively studied in the biostatistics literature (Murphy et al. (2001), Murphy (2003), and Robins (2004), among others). These studies typically rely on an ideal multi-stage experimental environment that satisfies sequential randomization. Based on such experimental data, they identify optimal regimes that maximize welfare, defined as the average counterfactual outcome. However, non-compliance is prevalent in experiments, and more generally, treatment endogeneity is a marked feature in observational studies. This may be one reason the vast biostatistics literature has not yet gained traction in economic analyses, despite the potentially fruitful applications of optimal dynamic regimes in policy evaluation.

To illustrate the policy relevance of the optimal dynamic regime, consider the labor market returns to types of high school and college. A policymaker may be interested in learning a schedule of school allocation rules $\boldsymbol{\delta}(\cdot) = (\delta_1, \delta_2(\cdot))$ that maximizes the employment rate $W_{\boldsymbol{\delta}} = E[Y_2(\boldsymbol{\delta})]$, where $\delta_1 \in \{0, 1\}$ assigns the high school type (vocational or academic), $\delta_2(y_1, \delta_1) \in \{0, 1\}$ assigns the college type (two-year or four-year) based on δ_1 and the high school GPA $y_1 \in \{0, 1\}$ (low or high), and $Y_2(\boldsymbol{\delta})$ indicates the counterfactual employment status under regime $\boldsymbol{\delta}(\cdot)$. Suppose the optimal regime $\boldsymbol{\delta}^*(\cdot)$ is such that $\delta_2^*(1, 1) = 1$ and $\delta_2^*(0, 0) = 0$; i.e., it turns out optimal to assign a four-year college to a student who was assigned to an academic high school and achieved a high GPA, and to assign a two-year college to a student with the opposite history. Even if assigning schools is mostly infeasible in reality, one of the policy implications of such $\boldsymbol{\delta}^*(\cdot)$ is that the average job market performance can be improved by a merit-based tuition subsidy for four-year college. This type of policy questions is hard to answer from learning an optimal static regime where $\delta_t(\cdot)$ is a constant

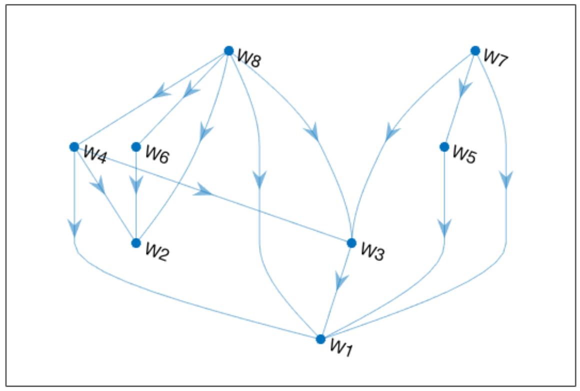


Figure 1: An Example of Sharp Partial Ordering of Welfares

function.[1] In learning $\boldsymbol{\delta}^*(\cdot)$ in this example, observational data may only be available where the observed treatments (schooling decisions) are endogenous.

This paper proposes a nonparametric framework of partial identification, in which we can partially learn the ranking of counterfactual welfares $W_{\boldsymbol{\delta}}$'s and hence the optimal dynamic regime $\boldsymbol{\delta}^*(\cdot)$. Given the observed distribution of sequences of outcomes and endogenous treatments and using the instrumental variable (IV) method, we establish sharp partial ordering of welfares, and characterize the identified set of optimal regimes as a discrete subset of all possible regimes. We define welfare as a linear functional of the joint distribution of counterfactual outcomes across periods. Examples of welfare include the average counterfactual terminal outcome commonly considered in the literature and as shown above. We assume we are equipped with some IVs that are possibly binary. We show that it is helpful to have a sequence of IVs generated from sequential experiments or quasi-experiments. Examples of the former are increasingly common as forms of random assignments or encouragements in medical trials, public health and educational interventions, and A/B testing on digital platforms. Examples of the latter can be some combinations of traditional IVs and regression discontinuity designs. Our framework also accommodates a single binary IV in the context of dynamic treatments and outcomes (e.g., Cellini et al. (2010)). The identifying power in such a case is investigated in simulation. The framework of this paper allows us to gain insight into data requirements to achieve a certain level of informativeness.

The identification analysis is twofold. In the first part, we establish mapping from data to sharp partial ordering of counterfactual welfares with respect to possible regimes, representing the partial ordering as a directed acyclic graph (DAG).[2] The point identification of $\boldsymbol{\delta}^*(\cdot)$ will

[1] In this case, the optimal regime will only reveal information about, e.g., dynamic complementarity (Cunha and Heckman (2007), Almond and Mazumder (2013), Johnson and Jackson (2019)).

[2] The way directed graphs are used in this paper is completely unrelated to causal graphical models in the

be achieved by establishing the total ordering of welfares, which is not generally possible in this flexible nonparametric framework with limited exogenous variation. Figure 1 is an example of partial ordering (interchangeably, a DAG) that we calculated by applying this paper's theory and using simulated data. Here, we consider a two-period case as in the school-type example, which yields eight possible $\boldsymbol{\delta}(\cdot)$'s and corresponding welfares, and "$\rightarrow$" corresponds to the relation "$>$". To establish the partial ordering, we first characterize bounds on the difference between two welfares as the set of optima of linear programs, and we do so for all possible welfare pairs. The bounds on welfare gaps are informative about whether welfares are comparable or not, and when they are, how to rank them. Then we show that although the bounds are calculated from separate optimizations, the partial ordering is consistent with *common* data-generating processes. The DAG obtained in this way is shown to be sharp (in the sense that will become clear later). 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 as the number of possible welfare pairs, which is also large due to adaptivity. Instead, we provide a simple analytical condition that identifies the signs of the optima of each linear program, which are enough to establish the sharp DAG. Note that each welfare gap measures the *dynamic treatment effect*. The DAG concisely (and tightly) summarizes the identified signs of these treatment effects, and thus can be a parameter of independent interest.

In the second part of the analysis, given the sharp partial ordering, we show that the identified set can be characterized as the set of maximal elements associated with the partial ordering, i.e., the set of regimes that are *not inferior*. For example, according to Figure 1, the identified set consists of regimes 7 and 8. Given the DAG, we also calculate topological sorts, which are total orderings that do not violate the underlying partial ordering. Theoretically, topological sorts can be viewed as observationally equivalent total orderings, which insight relates the partial ordering we consider with a more conventional notion of partial identification. Practically, topological sorts can be served as a policy menu that a policymaker can be equipped with. If desired, linear programming can be solved to calculate bounds on a small number of sorted welfares (e.g., top-tier welfares).

Given the minimal structure we impose in the data-generating process, the size of the identified set may be large in some cases. Such an identified set may still be useful in eliminating suboptimal regimes or warning about the lack of informativeness of the data. Often, however, researchers are willing to impose additional assumptions to gain identifying power. We propose identifying assumptions, such as uniformity assumptions that generalize

literature.

the monotonicity assumption in Imbens and Angrist (1994), an assumption about an agent's learning, Markovian structure, and stationarity. These assumptions tighten the identified set by reducing the dimension of the simplex in the linear programming, thus producing a denser DAG. We show that these assumptions are easy to impose in our framework.

This paper makes several contributions. To our best knowledge, this paper is first in the literature that considers the identifiability of optimal dynamic regimes under treatment endogeneity. Murphy (2003) and subsequent works consider point identification of optimal dynamic regimes, but under the sequential randomization assumption. This paper brings that literature to observational contexts. Recently, Han (forthcoming), Cui and Tchetgen Tchetgen (2020) and Qiu et al. (2020) relax sequential randomization and establish identification of dynamic average treatment effects and/or optimal regimes using instrumental variables. They consider a regime that is a mapping only from covariates, but not previous outcomes and treatments, to an allocation. They focus on point identification by imposing assumptions such as the existence of additional exogenous variables in a multi-period setup (Han (forthcoming)) or the zero correlation between unmeasured confounders and compliance types in a single-period setup (Cui and Tchetgen Tchetgen (2020); Qiu et al. (2020)). Relatedly, the dynamic effects of treatment timing (i.e., irreversible treatments) have been considered in Heckman and Navarro (2007) and Heckman et al. (2016) who utilize exclusion restrictions and infinite support assumptions, and in Athey and Imbens (2018), Callaway and Sant'Anna (2019), and Abraham and Sun (2020), who extend the difference-in-differences approach to dynamic settings. This paper complements these papers by considering treatment scenarios of multiple dimensions with adaptivity as the key ingredient.

Second, the simple analytical procedure of establishing sharp ordering of counterfactual welfares has broader applicability beyond the context of this paper. The linear programming approach to partially identifying counterfactuals has early examples as Balke and Pearl (1997) and Manski (2007), and appears recently in Torgovitsky (2019), Deb et al. (2017), Mogstad et al. (2018), Kitamura and Stoye (2019), Machado et al. (2019), Tebaldi et al. (2019), Kamat (2019), and Gunsilius (2019), to name a few. In the contexts of these studies, if one is interested in the signs of the parameters (thus the signs of the optima in linear programs) rather than the magnitude (thus the values of the optima), our analytical method can be useful. The conditions to be checked depend only on the distribution of data and known components of each linear program, and not on the solution of the program. As a result, we show that the estimation and inference (where resampling methods are typically used) are relatively straightforward and computationally light. The method can also be useful in other settings where the goal is to identify signs of various treatment effects, to compare welfares across multiple treatments and regimes—e.g., personalized treatment rules—or more generally, to establish rankings of policies across different counterfactual scenarios and find the best ones.

The dynamic treatment regime considered in this paper is broadly related to the literature on statistical treatment rules, e.g., Manski (2004), Hirano and Porter (2009), Bhattacharya and Dupas (2012), Stoye (2012), Kitagawa and Tetenov (2018), Kasy (2016), and Athey and Wager (2017). However, our setting, assumptions, and goals are different from those in these papers. In a single-period setting, they consider allocation rules that map covariates to decisions. They impose assumptions that ensure point identification, such as (conditional) unconfoundedness, and focus on establishing the asymptotic optimality of the treatment rules, with Kasy (2016) the exception.[3] Kasy (2016) focuses on establishing partial ranking by comparing a pair of treatment-allocating probabilities as policies. The notion of partial identification of ranking is related to ours, but we introduce the notion of sharpness of a partially ordered set with discrete policies and a linear programming approach to achieve that. Finally, in order to focus on the challenge with endogeneity, we consider a simple setup where the exploration and exploitation stages are separated, unlike in the literature on bandit problems (Kock and Thyrsgaard (2017), Kasy and Sautmann (forthcoming), Athey and Imbens (2019)). We believe the current setup is a good starting point.

In the next section, we introduce the dynamic regimes and related counterfactual outcomes, which define the welfare and the optimal regime. Section 3 provides a motivating example. Section 4 conducts the main identification analysis by constructing the DAG and characterizing the identified set. Section 5 provides the analytical conditions for linear programming. Sections 6–8 introduce topological sorts and additional identifying assumptions and discuss cardinality reduction for the set of regimes. Section 9 illustrates the analysis with numerical exercises, and Section 10 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 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 the realized binary

[3] Athey and Wager (2017)'s framework allows observational data with endogenous treatments as a special case, but the conditional homogeneity of treatment effects is assumed.

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 $\boldsymbol{\delta}(\cdot)$ When $T = 2$

outcomes and treatments $\boldsymbol{y}^{t-1} \equiv (y_1, ..., y_{t-1})$ and $\boldsymbol{d}^{t-1} \equiv (d_1, ..., d_{t-1})$ onto a deterministic treatment allocation $d_t \in \{0, 1\}$:

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

This adaptive rule also appears in, e.g., Murphy (2003). The rule can also be a function of other discrete covariates, which we do not consider here for brevity. A special case of (2.1) is a static rule where $\delta_t(\cdot)$ is only a function of covariates but not $(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1})$ (Han (forthcoming), Cui and Tchetgen Tchetgen (2020)) or a constant function.[4] Binary outcomes and treatments are prevalent, and they are helpful in analyzing, interpreting, and implementing dynamic regimes (Zhang et al. (2015)). Still, extending the framework to allow for multi-valued discrete variables is possible. Whether the rule is dynamic or static, we only consider deterministic rules $\delta_t(\cdot) \in \{0, 1\}$. In Appendix A.1, we extend this to stochastic rules $\tilde{\delta}_t(\cdot) \in [0, 1]$ and show why it is enough to consider deterministic rules in some cases. Then, a *dynamic regime* up to period t is defined as a vector of all treatment rules:

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

Let $\boldsymbol{\delta}(\cdot) \equiv \boldsymbol{\delta}^T(\cdot) \in \mathcal{D}$ where $\mathcal{D}$ is the set of all possible regimes.[5] For $T = 2$, 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 a dynamic regime. Because of the adaptivity intrinsic in dynamic regimes, ex-

[4] This means that our term of "static regime" is narrowly defined than in the literature. In the literature, a regime is sometimes called dynamic even if it is only a function of covariates.

[5] We can allow $\mathcal{D}$ to be a strict subset of the set of all possible regimes; see Section 8 for this relaxation.

pressing counterfactual outcomes is more involved than that with static 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 counterfactual outcome with adaptive regime $\boldsymbol{\delta}^t(\cdot)$ as follows[6]:

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

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

$$\begin{aligned} d_1 &= \delta_1, \\ d_2 &= \delta_2(Y_1(d_1), d_1), \\ d_3 &= \delta_3(\boldsymbol{Y}^2(\boldsymbol{d}^2), \boldsymbol{d}^2), \\ &\vdots \\ d_t &= \delta_t(\boldsymbol{Y}^{t-1}(\boldsymbol{d}^{t-1}), \boldsymbol{d}^{t-1}). \end{aligned} \tag{2.3}$$

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 counterfactual welfare as a linear 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 and which is common in the literature, and the weighted average of counterfactuals $\sum_{t=1}^T \omega_t E[Y_t(\boldsymbol{\delta}^t(\cdot))]$. Then, the *optimal dynamic regime* is a regime that maximizes the welfare as defined in (1.1):[7]

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

In the case of $W_{\boldsymbol{\delta}} = E[Y_T(\boldsymbol{\delta}(\cdot))]$, the solution $\boldsymbol{\delta}^*(\cdot)$ can be justified by backward induction in finite-horizon dynamic programming. Moreover in this case, the regime with deterministic rules $\delta_t(\cdot) \in \{0, 1\}$ achieves the same optimal regime and optimized welfare as the regime with stochastic rules $\delta_t(\cdot) \in [0, 1]$; see Theorem A.1 in Appendix A.1.

The identification analysis of the optimal regime is closely related to the identification of

[6] As the notation suggests, we implicitly assume the "no anticipation" condition.

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

welfare for each regime and welfare gaps, which also contain information for policy. Some interesting special cases are the following: (i) the *optimal welfare*, $W_{\boldsymbol{\delta}^*}$, which in turn yields (ii) 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 static 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 as the set of all $\boldsymbol{d}$ is a subset of $\mathcal{D}$.

3 Motivating Examples

For illustration, we continue discussing the example in the Introduction. This stylized example in an observational setting is meant to motivate the policy relevance of the optimal dynamic regime and the type of data that are useful for recovering it. Again, 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 an above-median GPA in high school and $Y_{i1} = 0$ if below-median. In addition, 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.

First, consider the optimal static regime $\boldsymbol{d}^*$. 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 regime with adaptivity $\boldsymbol{\delta}^*(\cdot)$ 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. As argued in the Introduction, $\boldsymbol{\delta}^*(\cdot)$ provides policy implications that $\boldsymbol{d}^*$ cannot.

As D_1 and D_2 are endogenous, $\{D_{i1}, Y_{i1}, D_{i2}, Y_{i2}\}$ above are not useful by themselves to identify $W_{\boldsymbol{\delta}}$'s and $\boldsymbol{\delta}^*(\cdot)$. We employ the approach of using IVs, either a single IV (e.g., in the initial period) or a sequence of IVs. For the latter, we propose a sequential version of the fuzzy RD design. 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 lands slightly above the cutoff for the academic high school entrance exam and $Z_{i1} = 0$ if slightly below; let $Z_{i2} = 1$ if i lands slightly above cutoff for the four-year college entrance exam and $Z_{i2} = 0$ if slightly below. Then (Z_{i1}, Z_{i2}) can serve as the sequence of binary instruments that satisfy the assumption for IVs below. Alternatively, the distance to (or the tuition cost of) these schools or some combination with the fuzzy RD above can serve as Z_1 and Z_2.

In experimental settings, examples of a sequence of IVs can be found in multi-stage experiments, such as the Fast Track Prevention Program (Conduct Problems Prevention Research Group (1992)), the Elderly Program randomized trial for the Systolic Hypertension (The Systolic Hypertension in the Elderly Program (SHEP) Cooperative Research Group (1988)), and Promotion of Breastfeeding Intervention Trial (Kramer et al. (2001)). It is also possible to combine multiple experiments as in Johnson and Jackson (2019).

4 Partial Ordering and Partial Identification

4.1 Observables

We introduce observables based on which we want to identify the optimal regime and counterfactual welfares. Assume that the time length of the observables is equal to T, the length of the optimal regime to be identified.[8] For each period or stage $t = 1, ..., T$, assume that we observe the binary instrument Z_t, the binary endogenous treatment decision D_t, and the binary outcome $Y_t = \sum_{\boldsymbol{d}^t \in \{0,1\}^t} 1\{\boldsymbol{D}^t = \boldsymbol{d}^t\} Y_t(\boldsymbol{d}^t)$. These variables are motivated in the previous section. As another example, Y_t is a symptom indicator for a patient, D_t is the medical treatment received, and Z_t is generated by a multi-period medical trial. Importantly, the framework does not preclude the case in which Z_t exists only for some t but not all; see Section 9 for related discussions. In this case, Z_t for the other periods is understood to be degenerate. 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.[9] A single IV with full independence trivially satisfies this assumption. For a sequence of IVs, this assumption is satisfied in typical sequential randomized experiments, as well as quasi-experiments as discussed in Section 3. 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 $(\boldsymbol{Y}_i, \boldsymbol{D}_i, \boldsymbol{Z}_i)$ is independent and identically distributed and $\{(\boldsymbol{Y}_i, \boldsymbol{D}_i, \boldsymbol{Z}_i) : i = 1, ..., N\}$ is a small T large N panel. We mostly suppress the individual unit i throughout the paper. For empirical applications, the data structure can be more general than a panel and the kinds of Y_t, D_t and Z_t are allowed

[8] In general, we may allow $\tilde{T} \geq T$ where $\tilde{T}$ is the length of the observables.

[9] There may be other covariates available for the researcher, but we suppress them for brevity. All the stated assumptions and the analyses of this paper can be followed conditional on the covariates. A sufficient condition for Assumption SX is that $\boldsymbol{Z} \perp (\boldsymbol{Y}(\boldsymbol{d}), \boldsymbol{D}(\boldsymbol{z}))$.

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

Figure 2: Partially Ordered Sets of Welfares

to be different across time; Section 3 contains such an example. For the population from which the data are drawn, we are interested in learning the optimal regime.

4.2 Partial Ordering of Welfares

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 of $\boldsymbol{\delta}^*(\cdot)$ will be characterized as a subset of the discrete set $\mathcal{D}$. As the first step, we establish *partial ordering* of $W_{\boldsymbol{\delta}}$ w.r.t. $\boldsymbol{\delta}(\cdot) \in \mathcal{D}$ as a function of p. The partial ordering can be represented by a *directed acyclic graph* (DAG). The DAG summarizes the identified signs of the dynamic treatment effects, as will become clear later. Moreover, the DAG representation is fruitful for introducing the notion of the sharpness of partial ordering and later to translate it into the identified set of $\boldsymbol{\delta}^*(\cdot)$.

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

In general, the point identification of $\boldsymbol{\delta}^*(\cdot)$ is achieved by establishing the 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 given the data and maintained assumptions. To formally state this, let $G(\mathcal{K}, \mathcal{E})$ be a DAG where $\mathcal{K}$ is the set of welfare (or regime) indices and $\mathcal{E}$ is the set of edges.

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

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

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 $D_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^t)$ denote the extended counterfactual outcomes and treatments, respectively, 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, by concatenating the two sequences, define $\tilde{S}_t \equiv (\{Y_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^t)\}, \{D_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^t)\}) \in \{0,1\}^{2^{2t-1}} \times \{0,1\}^{2^{3t-2}}$. For example, $\tilde{S}_1 = (Y_1(0), Y_1(1), D_1(0), D_1(1)) \in \{0,1\}^2 \times \{0,1\}^2$, 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 $S \equiv \beta(\tilde{\boldsymbol{S}}) \in \mathcal{S} \subset \mathbb{N}$, 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, namely the true data-generating process. The vector q resides in $\mathcal{Q} \equiv \{q : \sum_s q_s = 1 \text{ and } q_s \geq 0 \ \forall s\}$ 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:

$$\begin{aligned}
\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}, \boldsymbol{D}(\boldsymbol{z}) = \boldsymbol{d}] &= \Pr[S \in \mathcal{S} : \boldsymbol{Y}(\boldsymbol{y}^{T-1}, \boldsymbol{d}) = \boldsymbol{y}, \boldsymbol{D}(\boldsymbol{y}^{T-1}, \boldsymbol{d}^{T-1}, \boldsymbol{z}) = \boldsymbol{d}] \\
&= \Pr[S \in \mathcal{S} : Y_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^t) = y_t, D_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^t) = d_t \quad \forall t] \\
&= \sum_{s \in \mathcal{S}_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}}} q_s,
\end{aligned} \tag{4.1}$$

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 (4.1), 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

$$W_k = A_k q. \tag{4.2}$$

The formal derivation of A_k can be found in Appendix A.2, but the intuition is as follows. 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 in deriving the result (4.2) 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 a *static* regime, i.e., $\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}]$'s, which in turn is a linear functional of q. To illustrate with $T = 2$ and welfare $W_{\boldsymbol{\delta}} = E[Y_2(\boldsymbol{\delta}(\cdot))]$, we have

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

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

$$\Pr[Y_2(\boldsymbol{\delta}_8(\cdot)) = 1] = P[\boldsymbol{Y}(1,1) = (1,1)] + P[\boldsymbol{Y}(1,1) = (0,1)], \tag{4.3}$$

where each $\Pr[\boldsymbol{Y}(d_1, d_2) = (y_1, y_2)]$ is the counterfactual distribution with a static regime, which in turn is a linear functional of (4.1).

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 (4.1) that there exists $d_p \times d_q$ matrix B such that

$$Bq = p, \tag{4.4}$$

where each row of B is a vector of 1's and 0's; the formal derivation of B can be found in Appendix A.2. It is worth noting that the linearity in (4.2) and (4.4) is *not* a restriction but given by the discrete nature of the setting. We assume $rank(B) = d_p$ without loss of generality, because 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 is rare, we can solve for $q = (B^\top B)^{-1} B^\top p$, and can trivially point identify $W_k = A_k q$ and thus $\boldsymbol{\delta}^*(\cdot)$. Otherwise, we have a set of observationally equivalent q's, which is the source of partial identification and motivates the following definition of the identified set.[10]

For a given q, 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.

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

[10] For simplicity, we use the same notation for the true q and its observational equivalence.

p is

$$\mathcal{D}_p^* \equiv \{\boldsymbol{\delta}^*(\cdot; q) : Bq = p \text{ and } q \in \mathcal{Q}\} \subset \mathcal{D}, \tag{4.5}$$

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

4.4 Characterizing Partial Ordering and the Identified Set

Given p, we establish the partial ordering of W_k's, i.e., generate the DAG, by determining whether $W_k > W_{k'}$, $W_k < W_{k'}$, or W_k and $W_{k'}$ are not comparable (including $W_k = W_{k'}$), denoted as $W_k \sim W_{k'}$, for $k, k' \in \mathcal{K}$. As described in the next theorem, this procedure can be accomplished by determining the signs of the bounds on the welfare gap $W_k - W_{k'}$ for $k, k' \in \mathcal{K}$ and $k > k'$.[11] Then the identified set can be characterized based on the resulting partial ordering.

The nature of the data generation induces the linear system (4.2) and (4.4). This enables us to characterize the bounds on $W_k - W_{k'} = (A_k - A_{k'})q$ as the optima in linear programming. Let $U_{k,k'}$ and $L_{k,k'}$ be the upper and lower bounds. Also let $\Delta_{k,k'} \equiv A_k - A_{k'}$ for simplicity, and thus the welfare gap is expressed as $W_k - W_{k'} = \Delta_{k,k'} q$. Then, for $k, k' \in \mathcal{K}$, we have the main linear programs:

$$\begin{aligned} U_{k,k'} &= \max\nolimits_{q \in \mathcal{Q}} \Delta_{k,k'} q, \\ L_{k,k'} &= \min\nolimits_{q \in \mathcal{Q}} \Delta_{k,k'} q, \end{aligned} \qquad s.t. \quad Bq = p. \tag{4.6}$$

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

Assumption B imposes that the model, Assumption SX in this case, 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 $k, k' \in \mathcal{K}$ and $k > k'$, or equivalently, $L_{k,k'}$ for $k, k' \in \mathcal{K}$ and $k \neq k'$.[12]

Theorem 4.1. *Suppose Assumptions SX and B hold. Then, (i) $G(\mathcal{K}, \mathcal{E}_p)$ with $\mathcal{E}_p \equiv \{(k, k') \in \mathcal{K} : L_{k,k'} > 0 \text{ and } k \neq k'\}$ is sharp; (ii) $\mathcal{D}_p^*$ defined in (4.5) satisfies*

$$\mathcal{D}_p^* = \{\boldsymbol{\delta}_{k'}(\cdot) : \nexists k \in \mathcal{K} \text{ such that } L_{k,k'} > 0 \text{ and } k \neq k'\}. \tag{4.7}$$

$$= \{\boldsymbol{\delta}_{k'}(\cdot) : L_{k,k'} \leq 0 \text{ for all } k \in \mathcal{K} \text{ and } k \neq k'\} \tag{4.8}$$

The proof of Theorem 4.1 is shown in the Appendix. The key insight of the proof is that even though the bounds on the welfare gaps are calculated from separate optimizations, the

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

[12] Notice that $(L_{k,k'}, U_{k,k'})$ for $k > k'$ contain the same information as $L_{k,k'}$ for $k \neq k'$, since $U_{k,k'} = -L_{k',k}$.

partial ordering is governed by common q's (each of which generates all the welfares) that are observationally equivalent; see Section 6.1 for related discussions.

Theorem 4.1(i) prescribes how to calculate the sharp DAG as a function of data.[13] According to (4.7) in (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 $G(\mathcal{K}, \mathcal{E}_p)$, i.e., the set of regimes that are *not inferior*. In Figure 2, it is easy to see that the set of maximals is $\mathcal{D}_p^* = \{\boldsymbol{\delta}_1(\cdot), \boldsymbol{\delta}_4(\cdot)\}$ in panel (a) and $\mathcal{D}_p^* = \{\boldsymbol{\delta}_1(\cdot)\}$ in panel (b).

The identified set $\mathcal{D}_p^*$ characterizes the information content of the model. Given the minimal structure we impose in the model, $\mathcal{D}_p^*$ may be large in some cases. However, we argue that an uninformative $\mathcal{D}_p^*$ still has implications for policy: (i) such set recommends the policymaker eliminate sub-optimal regimes from her options; (ii) in turn, it warns the policymaker about her lack of information (e.g., even if she has access to the experimental data); when $\mathcal{D}_p^* = \mathcal{D}$ as one extreme, "no recommendation" can be given as a non-trivial policy suggestion of the need for better data. As shown in the numerical exercise, the size of $\mathcal{D}_p^*$ is related to the strength of Z_t (i.e., the size of the complier group at t) and the strength of the dynamic treatment effects. This is reminiscent of the findings in Machado et al. (2019) for the average treatment effect in a static model. In Section 7, we list further identifying assumptions that help shrink $\mathcal{D}_p^*$.

5 Analytical Conditions in Linear Programming

In practice, a naïve approach to compute the sharp DAG and the identified set in Theorem 4.1 is to directly compute $L_{k,k'}$ by solving its linear program in (4.6) for $k, k' \in \mathcal{K}$ and $k \neq k'$. This can be computationally very costly. Note that to generate the DAG, we need to make at most "$|\mathcal{K}| = |\mathcal{D}| = 2^{2^T-1}$ choose 2" pair-wise comparisons of the welfares.[14] With the naïve approach, this amounts to solving "2^{2^T-1} choose 2" linear programs of the form (4.6), where (4.6) is a large-scale linear program. In this program, the dimension of q is $d_q = |\mathcal{Q}|+1 = |\mathcal{S}| = \prod_{t=1}^T |\mathcal{S}_t|$ where $|\mathcal{S}_t| = 2^{2^{2t-1}} \times 2^{2^{3t-2}}$, which can be immense; e.g., when $T = 2$, $d_q = (2^2 \times 2^2) \times (2^{16} \times 2^8) = 268,435,456$. In addition, the number of constraints is $d_p + d_q + 1$ where $d_p = 2^{3T} - 2^T$. Obviously, the scale of the problem becomes even larger when the regime is adaptive to other observed covariates. This computational complexity can possibly be mitigated by imposing further identifying assumptions on the data-generating process as

[13] The DAG can be conveniently represented in terms of a $|\mathcal{K}| \times |\mathcal{K}|$ adjacency matrix Ω such that its element $\Omega_{k,k'} = 1$ if $W_k \geq W_{k'}$ and $\Omega_{k,k'} = 0$ otherwise.

[14] This procedure is closely related to what is called the *bubble sort*. There are more efficient algorithms, such as the *quick sort*, although they must 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.

shown later. Nevertheless, the baseline case should still be calculated as a benchmark to conduct the sensitivity analysis and understand the identifying power of those assumptions. Moreover, the naïve approach poses non-trivial challenges in developing inference methods for $\boldsymbol{\delta}^*(\cdot)$ and other parameters. Resampling methods are commonly used for inference in partial identification, which then involve solving the set of linear programs as many times as the number of resampling repetitions. To overcome these challenges, we propose a simple analytical procedure to obtain the sharp DAG and the identified set.

This part contains general results beyond the context of optimal treatment regimes in the previous section. Therefore, we simplify the notation and consider linear programs that characterize the bounds $[L, U]$ on the parameter Δq:

$$\begin{aligned} U &= \max\nolimits_{q \in \mathcal{Q}} \Delta q, \\ L &= \min\nolimits_{q \in \mathcal{Q}} \Delta q, \end{aligned} \qquad s.t. \quad Bq = p, \tag{5.1}$$

where $\mathcal{Q} \equiv \{q : \sum_s q_s = 1 \text{ and } q_s \geq 0 \ \forall s\} \subset \mathbb{R}^{d_q}$ is a standard simplex of dimension $d_q - 1$, Δ is a $1 \times d_q$ vector, and B is a $d_p \times d_q$ matrix with $d_p < d_q$. Without loss of generality, we assume $B = (B_1 \vdots O)$ for $d_p \times d_p$ full rank matrix B_1 and $d_p \times (d_q - d_p)$ zero matrix O.[15]

We first investigate how to detect the non-identification of the sign of Δq, i.e., whether $L \leq 0 \leq U$. For example, in the context of the previous section with $W_k - W_{k'} = \Delta_{k,k'} q = \Delta q$, this problem is equivalent to detecting $W_k \sim W_{k'}$, the incomparability of W_k and $W_{k'}$. Note that $L \leq 0 \leq U$ if and only if there exists $q \in \mathcal{Q}$ such that $\Delta q = 0$ and $Bq = p$, or simply,

$$ker(\Delta) \cap \{q : Bq = p\} \cap \mathcal{Q} \neq \emptyset, \tag{5.2}$$

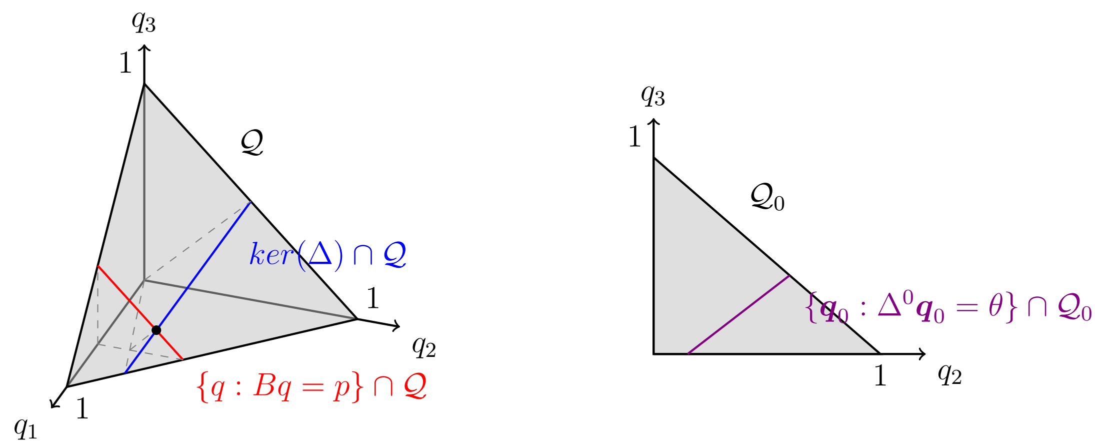
where $ker(\Delta)$ denotes the kernel (i.e., the null space) of Δ. That is, we want to find conditions under which the simplex $\mathcal{Q}$, the hyperplanes $\{q : Bq = p\}$, and $ker(\Delta)$ have a nonempty intersection. Figure 3(a) depicts this intersection for the case of $d_q = 3$.

Define partitions $\Delta = (\Delta^1 \vdots \Delta^0)$ 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

$$\boldsymbol{q}_1 = B_1^{-1} p, \tag{5.3}$$

because B_1 has full rank. That is, we can solve for the subvector of the data-generating

[15] If this 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.



(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 3: Illustration of Conditions for $L \leq 0 \leq U$ (with $q = (q_1, q_2, q_3)^\top$)

process as a function of the data distribution. By plugging in (5.3), $\Delta q = \Delta^1 \boldsymbol{q}_1 + \Delta^0 \boldsymbol{q}_0 = 0$, i.e., the welfare gap being zero, can be rewritten as

$$\Delta^1 B_1^{-1} p + \Delta^0 \boldsymbol{q}_0 = 0. \tag{5.4}$$

For simplicity, let $\theta \equiv -\Delta^1 B_1^{-1} p$, 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 $\mathcal{S}_0 \subset \mathcal{S}$ is the set of indices that correspond to the subvector $\boldsymbol{q}_0$. Then, by (5.2) and (5.4), $L \leq 0 \leq U$ if and only if there exists *nonzero vector* $\boldsymbol{q}_0 \in \mathcal{Q}_0$ such that $\Delta^0 \boldsymbol{q}_0 = \theta$,[16] or

$$\{\boldsymbol{q}_0 : \Delta^0 \boldsymbol{q}_0 = \theta\} \cap \mathcal{Q}_0 \backslash \{\boldsymbol{0}\} \neq \emptyset. \tag{5.5}$$

Since $\mathcal{Q}_0$ is a *finitely generated cone* and independent of the constraints, finding conditions under which (5.5) holds is mathematically more tractable than directly analyzing (5.2); see Figure 3(b). It essentially reduces to checking whether the hyperplane $\Delta^0 \boldsymbol{q}_0 = \theta$ lies between the vertices of the cone. The next theorem states these conditions under which the sign of Δq is not identified (e.g., the incomparability $W_k \sim W_{k'}$).

Theorem 5.1 (Non-Identification of Sign or Order)**.** *Suppose Assumptions SX and B hold. Let $\theta \equiv -\Delta^1 B_1^{-1} p$, and $\underline{\Delta}^0 \in \{-1, 0, 1\}$ and $\overline{\Delta}^0 \in \{-1, 0, 1\}$ be the minimum and maximum elements of vector Δ^0, respectively. Then, $L \leq 0 \leq U$ if and only if either one of the following holds: (i) $\underline{\Delta}^0 < \theta < \overline{\Delta}^0$, (ii) $\underline{\Delta}^0 \geq \theta \geq 0$, or (iii) $\overline{\Delta}^0 \leq \theta \leq 0$.*

[16] We want to find nonzero $\boldsymbol{q}_0$, since when $\boldsymbol{q}_0 = \boldsymbol{0}$, then $\Delta q = -\theta$ for all $\boldsymbol{q}_1$ and hence we trivially point identify Δq (e.g., the welfare gap $W_k - W_{k'}$).

Since scalar θ (up to p) and $1 \times (d_q - d_p)$ vector Δ^0 are known to the researcher, we can directly detect the incomparability from the data p without solving the linear programmings.[17] Furthermore, we can show the following results, which identify the sign of Δq (e.g., whether $W_k > W_{k'}$ or $W_k < W_{k'}$):

Theorem 5.2 (Identification of Sign or Order)**.** *Suppose Assumptions SX and B hold. Let* $\theta \equiv -\Delta^1 B_1^{-1} p$*, and* $\underline{\Delta}^0 \in \{-1, 0, 1\}$ *and* $\overline{\Delta}^0 \in \{-1, 0, 1\}$ *be the minimum and maximum elements of vector* Δ^0*, respectively. Then,* $L > 0$ *if and only if*

$$\theta < \min\{0, \underline{\Delta}^0\}. \tag{5.6}$$

Similarly, $U < 0$ *if and only if* $\theta > \max\{0, \overline{\Delta}^0\}$.

Theorem 5.2 provides the basis for the systematic computation of the DAG and the identified set in the previous section. It suggests an algorithm that generates the DAG by automating the task of checking the conditions for every (k, k') pair with $\theta = \theta_{k,k'}$ and $\Delta = \Delta_{k,k'}$. Compared to directly solving the set of large-scale linear programs, finding $\underline{\Delta}^0$ and $\overline{\Delta}^0$ from a large-dimensional vector Δ^0 (e.g., $d_q - d_p = 268,435,456 - 60$ when $T = 2$) is an extremely simple computational task, especially since their values are known to be one of $\{-1, 0, 1\}$. Next, based on (4.8), $\mathcal{D}_p^*$ can be computed by negating (5.6):

$$\mathcal{D}_p^* = \{\boldsymbol{\delta}_{k'}(\cdot) : \theta_{k,k'} \geq \min\{0, \underline{\Delta}^0_{k,k'}\} \text{ for all } k \in \mathcal{K} \text{ and } k \neq k'\}. \tag{5.7}$$

Unlike Theorem 5.2, Theorem 5.1 is not directly used in the derivation of the DAG and $\mathcal{D}_p^*$, but is crucial in developing the inference procedure proposed later.

6 Topological Sorts and Bounds on Sorted Welfare

6.1 Topological Sorts as Observational Equivalence

The DAG is a useful policy benchmark. For a complicated DAG, it would be helpful to have a summary of it. In fact, the identified set $\mathcal{D}_p^*$ can be viewed as a summary of a DAG. Another way to summarize a DAG is to use topological sorts. A *topological sort* of a DAG is a linear ordering of its vertices that does not violate the order in the partial ordering given by the DAG. That is, for every directed edge $k \to k'$, k comes before k' in this linear ordering. Apparently, there can be multiple topological sorts for a DAG. Let L_G be the number of topological sorts of DAG $G(\mathcal{K}, \mathcal{E}_p)$, and let $k_{l,1} \in \mathcal{K}$ be the initial vertex of the l-th

[17] Note conditions (i)–(iii) are exclusive. These conditions and conditions $\theta < \min\{0, \underline{\Delta}^0\}$ and $\theta > \max\{0, \overline{\Delta}^0\}$ in the next theorem exhaust all possible configurations of θ, $\underline{\Delta}^0$, $\overline{\Delta}^0$, and zero.

topological sort for $1 \leq l \leq L_G$. For example, given the DAG in Figure 2(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.

In fact, topological sorts can be viewed as total orderings that are *observationally equivalent* to the true *total* ordering of welfares. That is, each q generates the total ordering of welfares via $W_k = A_k q$, and q's in $\{q : Bq = p\} \cap \mathcal{Q}$ generates observationally equivalent total orderings. This insight enables us to interpret the partial ordering we establish using the more conventional notion of partial identification: the ordering is partially identified in the sense that the set of all topological sorts is not a singleton. This insight yields an alternative way of characterizing the identified set $\mathcal{D}_p^*$ of the optimal regime.

Theorem 6.1. *Suppose Assumptions SX and B hold. The identified set $\mathcal{D}_p^*$ defined in (4.5) 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 $G(\mathcal{K}, \mathcal{E}_p)$.

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* with the partial welfare ordering. Therefore, not only $\boldsymbol{\delta}_{k_{l,1}}(\cdot) \in \mathcal{D}_p^*$ but also the full sequence of a topological sort

$$\Big(\boldsymbol{\delta}_{k_{l,1}}(\cdot), \boldsymbol{\delta}_{k_{l,2}}(\cdot), ..., \boldsymbol{d}_{k_{l,|\mathcal{D}|}}(\cdot)\Big) \tag{6.1}$$

can be useful. A policymaker can be equipped with any of such sequences as a policy menu.

6.2 Bounds on Sorted Welfares

A topological sort provides ordinal information about counterfactual welfares. To gain more comprehensive knowledge about these welfares, 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 programs. For instance, for $k \in \mathcal{K}$, the sharp lower and upper bounds on welfare W_k can be calculated via

$$\begin{array}{ll} U_k = \max_{q \in \mathcal{Q}} A_k q, & \\ L_k = \min_{q \in \mathcal{Q}} A_k q, & \end{array} \qquad s.t. \quad Bq = p. \tag{6.2}$$

This computational approach to calculating bounds is inevitable in this context. Unlike in the static case of calculating the bounds on, e.g., the average treatment effect, calculating the bounds on W_k (and $W_k - W_{k'}$) and proving their sharpness are analytically infeasible, especially when $T \geq 3$. Fortunately, as the partial order and thus the topological sort are obtained analytically by Theorem 5.2, we can focus on solving only a few linear programs.

7 Additional Assumptions

Often, researchers are willing to impose more assumptions based on priors about the data-generating process, e.g., agent's behaviors. Examples are uniformity, agent's learning, Markovian structure, and stationarity. These assumptions are easy to incorporate within the linear programming (4.6), and thus the conditions in Theorems 5.1 and 5.2. These assumptions tighten the identified set $\mathcal{D}_p^*$ by reducing the dimension of simplex $\mathcal{Q}$, and thus producing a denser DAG.[18]

To incorporate these assumptions, we extend the framework introduced in Sections 4–6. Suppose h is a $d_q \times 1$ 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 simplex for $\bar{q} \equiv Hq$ as

$$\bar{\mathcal{Q}} \equiv \{\bar{q} : \sum_s \bar{q}_s = 1 \text{ and } \bar{q}_s \geq 0 \ \forall s\}. \tag{7.1}$$

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 (4.2) and (4.4) as

$$\begin{aligned} B\bar{q} &= p, \\ W_k &= A_k \bar{q}, \end{aligned}$$

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 the identifying assumptions coded in h is defined as

$$\bar{\mathcal{D}}_p^* \equiv \{\boldsymbol{\delta}^*(\cdot;\bar{q}) : B\bar{q} = p \text{ and } \bar{q} \in \mathcal{Q}\} \subset \mathcal{D}, \tag{7.2}$$

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, Theorems 4.1, 5.1, and 5.2 can be modified by replacing q, B and Δ with

[18] Similarly, when these assumptions are incorporated in (6.2), we obtain tighter bounds on welfares.

$\bar{q}$, $\bar{B}$ and $\bar{\Delta}$, respectively.

We now list examples of identifying assumptions. This list is far from complete, and there may be other assumptions on how $(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z})$ are generated. The first assumption is a sequential version of the uniformity assumption (i.e., the monotonicity 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 postulates that there is no defying (or complying) behavior in decision D_t conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^{t-1})$. Without being 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 h by having $h_s = 0$ 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 the equivalent of imposing a threshold-crossing model for D_t:

$$D_t = 1\{\pi_t(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^t) \geq \nu_t\}, \tag{7.3}$$

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

Lemma 7.1. *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 (7.3) being satisfied conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^{t-1})$ for each t.*

The dynamic selection model (7.3) should not be confused with the dynamic regime (2.1). Compared to the dynamic regime $d_t = \delta_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1})$, which is a hypothetical quantity, equation (7.3) 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 policymaker has no access to $\boldsymbol{\nu} \equiv (\nu_1, ..., \nu_T)$. The functional dependence of D_t on $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1})$ and $\boldsymbol{Z}^{t-1}$ 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, an agent has the ability to revise her next period's decision based on her memory. To illustrate, consider the second period decision, $D_2(y_1, d_1)$. Under

Assumption L, an agent who would switch her treatment decision at $t = 2$ had she 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 she experienced good health, i.e., $D_2(1, 1) = 1$. Moreover, if an agent has not switched even after bad health, i.e., $D_2(0, 1) = 1$, it should be because of her unobserved preference, and thus $D_2(1, 1) = 1$, not because she cannot learn from the past, i.e., $D_2(1, 1) = 0$ cannot happen.[19]

Sometimes, we want to further impose uniformity in the formation of Y_t on top of Assumption M1:

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

This assumption postulates uniformity in a way that restricts heterogeneity of the contemporaneous treatment effect. As before, without being 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 worth noting that Assumption M2 (and M1) does not assume the direction of monotonicity. This is in contrast to the monotone treatment response assumption in, e.g., Manski (1997) and Manski and Pepper (2000), which assume the direction. Using a similar argument as before, Assumption M2 is the equivalent of a dynamic version of a nonparametric triangular model:

$$Y_t = 1\{\mu_t(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^t) \geq \varepsilon_t\}, \tag{7.4}$$

$$D_t = 1\{\pi_t(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^t) \geq \nu_t\}, \tag{7.5}$$

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

Lemma 7.2. *Suppose Assumption SX holds, $\Pr[D_t = 1|\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^t]$ is a non-trivial function of Z_t, and $\Pr[Y_t = 1|\boldsymbol{Y}^{t-1}, \boldsymbol{D}^t]$ is a non-trivial function of D_t. Assumption M2 is equivalent to (7.4)–(7.5) being satisfied conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^{t-1})$ for each t.*

The next 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 (7.4)–(7.5), Assumption K implies

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

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

[19] As suggested in this example, Assumption L makes the most sense when Y_t and D_t are the same (or at least similar) types over time, which is not generally required for the analysis of this paper.

which yields the familiar structure of dynamic discrete choice models found in the literature. Lastly, when there are more than two periods, an assumption that imposes stationarity can be helpful for identification. Such an assumption can be found in Torgovitsky (2019).

8 Cardinality Reduction

The typical time horizons we consider in this paper are short. 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 $\mathcal{D}$ 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 $\delta_t : \{0,1\} \times \{0,1\} \to \{0,1\}$ that maps only the lagged outcome and treatment onto a treatment allocation $d_t \in \{0,1\}$:

$$\delta_t(y_{t-1}, d_{t-1}) = d_t.$$

In this case, we have $|\mathcal{D}| = 2^{2T-1}$ instead of 2^{2^T-1}. An even simpler rule, $\delta_t(y_{t-1})$, appears in Murphy et al. (2001).

Another possibility is to be 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 considerations. For example, suppose that the policymaker decides to introduce the adaptive rule at $t = T$ while maintaining static rules for $t \leq T - 1$. Finally, $\mathcal{D}$ can be restricted by budget or policy constraints that, e.g., the treatment is allocated to each individual at most once.

9 Numerical Studies

We conduct numerical exercises to illustrate (i) the theoretical results developed in Sections 4–6, (ii) the role of the assumptions introduced in Section 7, and (iii) the overall computational scale of the problem. For $T = 2$, we consider the following data-generating process:

$$D_{i1} = 1\{\pi_1 Z_{i1} + \alpha_i + v_{i1} \geq 0\}, \tag{9.1}$$

$$Y_{i1} = 1\{\mu_1 D_{i1} + \alpha_i + e_{i1} \geq 0\}, \tag{9.2}$$

$$D_{i2} = 1\{\pi_{21} Y_{i1} + \pi_{22} D_{i1} + \pi_{23} Z_{i2} + \alpha_i + v_{i2} \geq 0\}, \tag{9.3}$$

$$Y_{i2} = 1\{\mu_{21} Y_{i1} + \mu_{22} D_{i2} + \alpha_i + e_{i2} \geq 0\}, \tag{9.4}$$

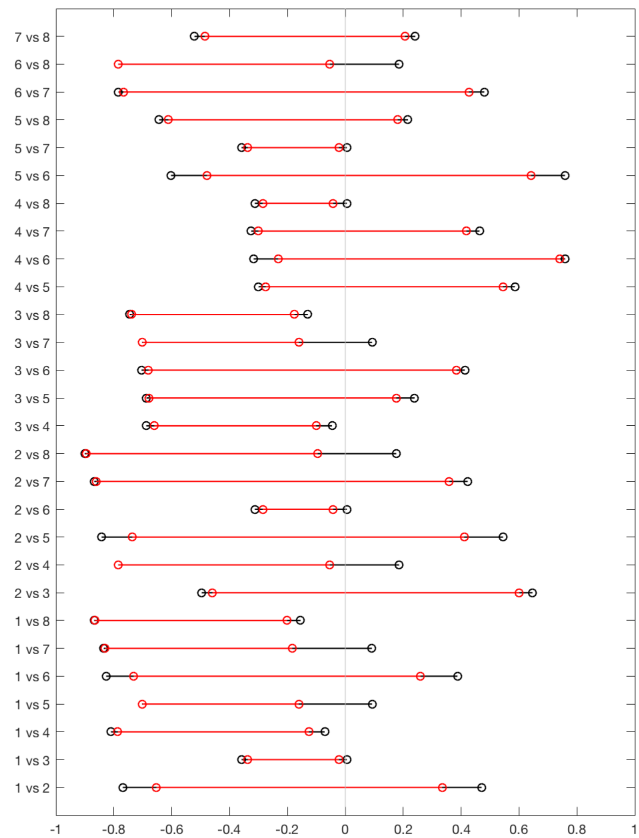


Figure 4: Sharp Bounds on Welfare Gaps under M1 (black) and M2 (red)

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 of the unobservables are 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, K, M1, and M2. We consider a data-generating process where all the coefficients in (9.1)–(9.4) 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., $|\mathcal{D}| = |\mathcal{K}| = 8$. Because the current exercise is small scale in generating the DAG, instead of using the analytical algorithm proposed in Theorems 5.1 and 5.2, we directly calculate the lower and upper bounds $(L_{k,k'}, U_{k,k'})$ on the welfare gap $W_k - W_{k'}$ for all pairs $k, k' \in \{1, ..., 8\}$ $(k < k')$. This is to illustrate the role of assumptions in improving the bounds. We conduct the bubble sort, which makes $\binom{8}{2} = 28$ pair-wise comparisons, resulting in 28×2 linear programs to run. As the researcher, we maintain Assumption K.

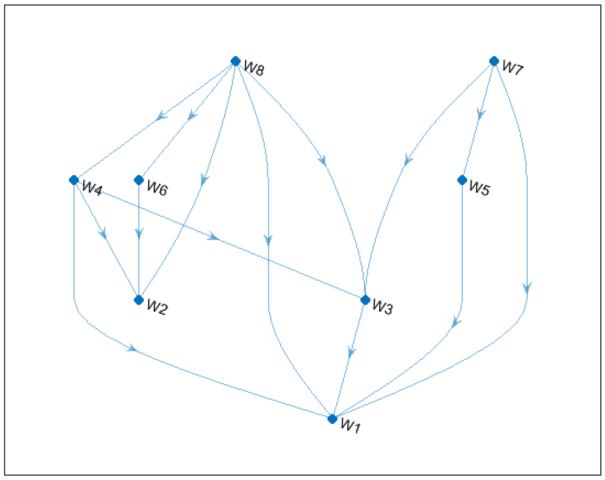


Figure 5: Sharp Directed Acyclic Graph under M2

Figure 4 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 exclude zero. The difference between the 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 with M2, which is consistent with the theory. Each set of bounds generates an associated DAGs (produced as an 8×8 adjacency matrix). We proceed with Assumption M2 for brevity.

Figure 5 (identical to Figure 1 in the Introduction) depicts the sharp DAG generated from $(L_{k,k'}, U_{k,k'})$'s under Assumption M2, based on Theorem 4.1(i). Then, by Theorem 4.1(ii), 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:

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

We also conducted a parallel analysis but with a slightly different data-generating process, where (a) all the coefficients in (9.1)–(9.4) are positive except $\mu_{22} < 0$ and (b) Z_2 does not exist. In Case (a), we obtain $\mathcal{D}_p^* = \{\boldsymbol{\delta}_2(\cdot)\}$ as a singleton, i.e., we point identify $\boldsymbol{\delta}^*(\cdot) = \boldsymbol{\delta}_2(\cdot)$. The results for Case (b) is shown in Figures 6 and 7. In this case, we obtain $\mathcal{D}_p^* = \{\boldsymbol{\delta}_6(\cdot), \boldsymbol{\delta}_7(\cdot), \boldsymbol{\delta}_8(\cdot)\}$.

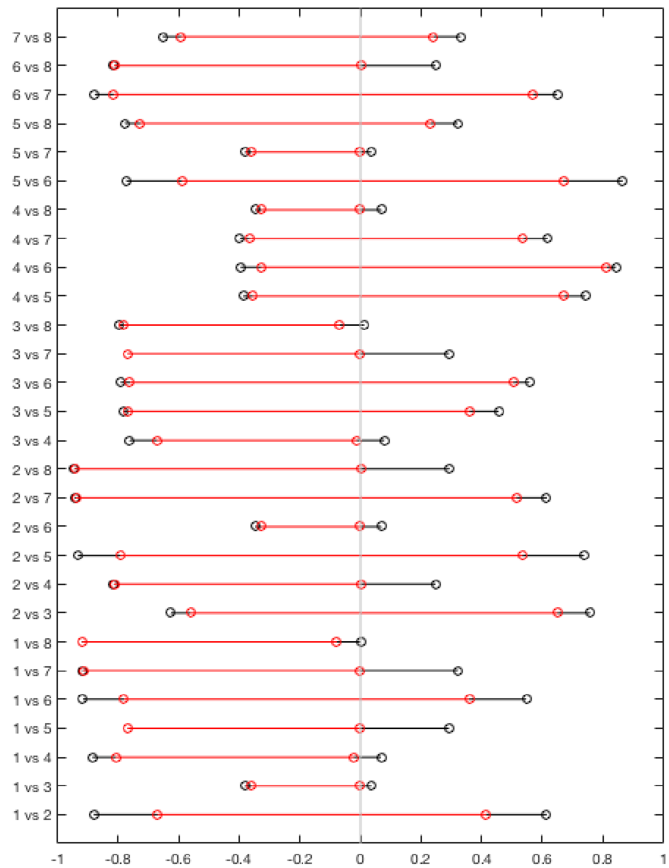


Figure 6: Sharp Bounds on Welfare Gaps under M1 (black) and M2 (red) (with only Z_1)

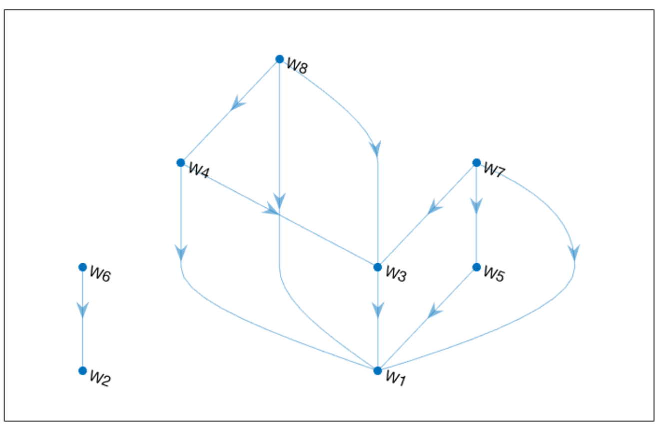


Figure 7: Sharp Directed Acyclic Graph under M2 (with only Z_1)

10 Estimation and Inference

The estimation of the DAG and the identified set $\mathcal{D}_p^*$ is straightforward given the conditions in Theorem 5.2. Recall that B_1, $\Delta^1_{k,k'}$, $\underline{\Delta}^0_{k,k'}$, and $\overline{\Delta}^0_{k,k'}$ are known objects to the researcher. The only unknown object in the condition is p, the joint distribution of $(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z})$, which can be estimated as $\hat{p}$, 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 $\hat{\theta}_{k,k'} \equiv -\Delta^1_{k,k'} B_1^{-1}\hat{p}$, the estimated DAG is $G(\mathcal{K}, \hat{\mathcal{E}}_p)$, where

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

Then, $\mathcal{D}_p^*$ can be estimated as

$$\widehat{\mathcal{D}}^* = \{\boldsymbol{\delta}_{k'}(\cdot) : \hat{\theta}_{k,k'} \geq \min\{0, \underline{\Delta}^0_{k,k'}\} \text{ for all } k \in \mathcal{K} \text{ and } k \neq k'\}.$$

Although we do not fully investigate inference in the current paper, we briefly discuss it. To conduct inference on the optimal regime $\boldsymbol{\delta}^*(\cdot)$, we can construct a confidence set (CS) for $\mathcal{D}_p^*$ with the following procedure. We consider a sequence of hypothesis tests, in which we eliminate regimes that are (statistically) significantly inferior to others. This is a statistical analog of the elimination procedure encoded in (4.7) or (4.8). This inference procedure extends Hansen et al. (2011)'s approach for the model confidence set, but in this novel context. For each test given $\tilde{\mathcal{K}} \subset \mathcal{K}$, we construct a null hypothesis that $W_k \sim W_{k'}$ for all $k, k' \in \tilde{\mathcal{K}}$. Geometrically, according to (5.5), this hypothesis restricts the range of $\theta_{k,k'}$ so that the hyperplane $\theta_{k,k'} = \Delta^0_{k,k'}\boldsymbol{q}_0$ lies within the cone $\mathcal{Q}_2$. Based on conditions (i)–(iii) in Theorem 5.1, this results in a one-sided test for

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

where l_1 and l_2 satisfy (i) $l_1(\Delta^0_{k,k'}) = (\overline{\Delta}^0_{k,k'} + \underline{\Delta}^0_{k,k'})/2$ and $l_2(\Delta^0_{k,k'}) = (\overline{\Delta}^0_{k,k'} - \underline{\Delta}^0_{k,k'})/2$ if $\underline{\Delta}^0_{k,k'} < 0 < \overline{\Delta}^0_{k,k'}$; (ii) $l_1(\Delta^0_{k,k'}) = \underline{\Delta}^0_{k,k'}/2$ and $l_2(\Delta^0_{k,k'}) = \underline{\Delta}^0_{k,k'}/2$ if $\underline{\Delta}^0_{k,k'} \geq 0$; (iii) $l_1(\Delta^0_{k,k'}) = \overline{\Delta}^0_{k,k'}/2$ and $l_2(\Delta^0_{k,k'}) = -\overline{\Delta}^0_{k,k'}/2$ if $\overline{\Delta}^0_{k,k'} \leq 0$, corresponding to conditions (i)–(iii) in Theorem 5.1, respectively.

Then, the procedure of constructing the CS, denoted as $\widehat{\mathcal{D}}_{CS}$, is as follows: *Step 0. Initially set $\tilde{\mathcal{K}} = \mathcal{K}$. Step 1. Test $H_{0,\tilde{\mathcal{K}}}$ at level α with test function $\phi_{\tilde{\mathcal{K}}} \in \{0,1\}$. Step 2. If $H_{0,\tilde{\mathcal{K}}}$ is not rejected, define $\widehat{\mathcal{D}}_{CS} = \{\boldsymbol{\delta}_k(\cdot) : k \in \tilde{\mathcal{K}}\}$; otherwise eliminate vertex $k_{\tilde{\mathcal{K}}}$ from $\tilde{\mathcal{K}}$ and repeat from Step 1.* In Step 1, $T_{\tilde{\mathcal{K}}} \equiv \max_{k,k' \in \tilde{\mathcal{K}}} t_{k,k'}$ can be used as the test statistic for $H_{0,\tilde{\mathcal{K}}}$ 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 $T_{\tilde{\mathcal{K}}}$ can be estimated using bootstrap. In Step 2, a candidate for $k_{\tilde{\mathcal{K}}}$ is $k_{\tilde{\mathcal{K}}} \equiv \arg\max_{k \in \tilde{\mathcal{K}}} \max_{k' \in \tilde{\mathcal{K}}} t_{k,k'}$. Following Hansen et al. (2011), we can show that the resulting CS has desirable properties. Let $H_{A,\tilde{\mathcal{K}}}$ be the alternative hypothesis.

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

Proposition 10.1. *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 constructing the CS 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 refute the assumption when the CS under that assumption is empty.

Inference on the welfare bounds in (6.2) 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) consider 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. Extending the framework to the current setting with partially identified welfare and dynamic regimes under treatment endogeneity would also be interesting future work.

A Appendix

A.1 Stochastic Regimes

For each $t = 1, ..., T$, define an adaptive *stochastic* treatment rule $\tilde{\delta}_t : \{0,1\}^{t-1} \times \{0,1\}^{t-1} \to [0,1]$ that allocates the probability of treatment:

$$\tilde{\delta}_t(\boldsymbol{y}^{t-1}, \tilde{\boldsymbol{d}}^{t-1}) = \tilde{d}_t \in [0,1]. \tag{A.1}$$

Then, the vector of these $\tilde{\delta}_t$'s is a *dynamic stochastic regime* $\tilde{\boldsymbol{\delta}}(\cdot) \equiv \tilde{\boldsymbol{\delta}}^T(\cdot) \in \mathcal{D}_{stoch}$ where $\mathcal{D}_{stoch}$ is the set of all possible stochastic regimes.[20] A deterministic regime is a special case where $\tilde{\delta}_t(\cdot)$ takes the extreme values of 1 and 0. Therefore, $\mathcal{D} \subset \mathcal{D}_{stoch}$ where $\mathcal{D}$ is the set of deterministic regimes. We define $Y_T(\tilde{\boldsymbol{\delta}}(\cdot))$ with $\tilde{\boldsymbol{\delta}}(\cdot) \in \mathcal{D}_{stoch}$ as the counterfactual outcome

[20] Dynamic stochastic regimes are considered in, e.g., Murphy et al. (2001), Murphy (2003), and Manski (2004).

$Y_T(\boldsymbol{\delta}(\cdot))$ where the deterministic rule $\delta_t(\cdot) = 1$ is randomly assigned with probability $\tilde{\delta}_t(\cdot)$ and $\delta_t(\cdot) = 0$ otherwise for all $t \leq T$. Finally, define

$$W_{\tilde{\boldsymbol{\delta}}} \equiv \mathbb{E}[Y_T(\tilde{\boldsymbol{\delta}}(\cdot))],$$

where $\mathbb{E}$ denotes an expectation over the counterfactual outcome and the random mechanism defining a rule, and define $\tilde{\boldsymbol{\delta}}^*(\cdot) \equiv \arg\max_{\tilde{\boldsymbol{\delta}}(\cdot)\in\mathcal{D}_{stoch}} W_{\tilde{\boldsymbol{\delta}}}$. The following theorem show that a deterministic regime is achieved as being optimal even though stochastic regimes are allow.

Theorem A.1. *Suppose* $W_{\tilde{\boldsymbol{\delta}}} \equiv \mathbb{E}[Y_T(\tilde{\boldsymbol{\delta}}(\cdot))]$ *for* $\tilde{\boldsymbol{\delta}}(\cdot) \in \mathcal{D}_{stoch}$ *and* $W_{\boldsymbol{\delta}} \equiv E[Y_T(\boldsymbol{\delta}(\cdot))]$ *for* $\boldsymbol{\delta}(\cdot) \in \mathcal{D}$*. It satisfies that*

$$\boldsymbol{\delta}^*(\cdot) \equiv \arg\max_{\boldsymbol{\delta}(\cdot)\in\mathcal{D}} W_{\boldsymbol{\delta}} = \arg\max_{\tilde{\boldsymbol{\delta}}(\cdot)\in\mathcal{D}_{stoch}} W_{\tilde{\boldsymbol{\delta}}}.$$

By the law of iterative expectation, we have

$$\mathbb{E}[Y_T(\tilde{\boldsymbol{\delta}}(\cdot))] = \mathbb{E}\left[\mathbb{E}\left[\cdots\mathbb{E}\left[\mathbb{E}[Y_T(\tilde{\boldsymbol{d}})|\boldsymbol{Y}^{T-1}(\tilde{\boldsymbol{d}}^{T-1})]\middle|\boldsymbol{Y}^{T-2}(\tilde{\boldsymbol{d}}^{T-2})\right]\cdots\middle|Y_1(\tilde{d}_1)\right]\right], \tag{A.2}$$

where the bridge variables $\tilde{\boldsymbol{d}} = (\tilde{d}_1, ..., \tilde{d}_T)$ satisfy

$$\begin{aligned}
\tilde{d}_1 &= \tilde{\delta}_1, \\
\tilde{d}_2 &= \tilde{\delta}_2(Y_1(\tilde{d}_1), \tilde{d}_1), \\
\tilde{d}_3 &= \tilde{\delta}_3(\boldsymbol{Y}^2(\tilde{\boldsymbol{d}}^2), \tilde{\boldsymbol{d}}^2), \\
&\vdots \\
\tilde{d}_T &= \tilde{\delta}_T(\boldsymbol{Y}^{T-1}(\tilde{\boldsymbol{d}}^{T-1}), \tilde{\boldsymbol{d}}^{T-1}).
\end{aligned}$$

Given (A.2), we prove the theorem by showing that the solution $\tilde{\boldsymbol{\delta}}^*(\cdot)$ can be justified by backward induction in a finite-horizon dynamic programming. To illustrate this with deterministic regimes when $T = 2$, we have

$$\delta_2^*(y_1, d_1) = \arg\max_{d_2} E[Y_2(\boldsymbol{d})|Y_1(d_1) = y_1], \tag{A.3}$$

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

$$\delta_1^* = \arg\max_{d_1} E[V_2(Y_1(d_1), d_1)]. \tag{A.4}$$

Then, $\boldsymbol{\delta}^*(\cdot)$ is equal to the collection of these solutions: $\boldsymbol{\delta}^*(\cdot) = (\delta_1^*, \delta_2^*(\cdot))$.

Proof. First, given (A.2), the optimal stochastic rule in the final period can be defined as

$$\tilde{\delta}_T^*(\boldsymbol{y}^{T-1}, \tilde{\boldsymbol{d}}^{T-1}) \equiv \arg\max_{\tilde{d}_T \in [0,1]} \mathbb{E}[Y_T(\tilde{\boldsymbol{d}})|\boldsymbol{Y}^{T-1}(\tilde{\boldsymbol{d}}^{T-1}) = \boldsymbol{y}^{T-1}].$$

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

$$\tilde{\delta}_t^*(\boldsymbol{y}^{t-1}, \tilde{\boldsymbol{d}}^{t-1}) \equiv \arg\max_{\tilde{d}_t \in [0,1]} \mathbb{E}[V_{t+1}(\boldsymbol{Y}^t(\tilde{\boldsymbol{d}}^t), \tilde{\boldsymbol{d}}^t)|\boldsymbol{Y}^{t-1}(\tilde{\boldsymbol{d}}^{t-1}) = \boldsymbol{y}^{t-1}]$$

and $V_t(\boldsymbol{y}^{t-1}, \tilde{\boldsymbol{d}}^{t-1}) \equiv \max_{\tilde{d}_t} \mathbb{E}[V_{t+1}(\boldsymbol{Y}^t(\tilde{\boldsymbol{d}}^t), \tilde{\boldsymbol{d}}^t)|\boldsymbol{Y}^{t-1}(\tilde{\boldsymbol{d}}^{t-1}) = \boldsymbol{y}^{t-1}]$. Then, $\tilde{\boldsymbol{\delta}}^*(\cdot) = (\tilde{\delta}_1^*, ..., \tilde{\delta}_T^*(\cdot))$. Since $\{0,1\} \subset [0,1]$, the same argument can apply for the deterministic regime using the current framework but each maximization domain being $\{0,1\}$. This analogously defines $\delta_t^*(\cdot) \in \{0,1\}$ for all t, and then $\boldsymbol{\delta}^*(\cdot) = (\delta_1^*, ..., \delta_T^*(\cdot))$, similarly as in Murphy (2003).

Now, for the maximization problems above, let $\tilde{W}_t(\tilde{\boldsymbol{d}}^t, \boldsymbol{y}^{t-1})$ represent the objective function at t for $2 \leq t \leq T$ with $\tilde{W}_1(\tilde{d}_1)$ for $t = 1$. By the definition of the stochastic regime, it satisfies that

$$\begin{aligned}
\tilde{W}_t(\tilde{\boldsymbol{d}}^t, \boldsymbol{y}^{t-1}) &= \tilde{d}_t W_t(1, \tilde{\boldsymbol{d}}^{t-1}, \boldsymbol{y}^{t-1}) + (1 - \tilde{d}_t) W_t(0, \tilde{\boldsymbol{d}}^{t-1}, \boldsymbol{y}^{t-1}) \\
&= \tilde{d}_t \left\{ W_t(1, \tilde{\boldsymbol{d}}^{t-1}, \boldsymbol{y}^{t-1}) - W_t(0, \tilde{\boldsymbol{d}}^{t-1}, \boldsymbol{y}^{t-1}) \right\} + W_t(0, \tilde{\boldsymbol{d}}^{t-1}, \boldsymbol{y}^{t-1}).
\end{aligned}$$

Therefore, $W_t(1, \tilde{\boldsymbol{d}}^{t-1}, \boldsymbol{y}^{t-1}) \geq W_t(0, \tilde{\boldsymbol{d}}^{t-1}, \boldsymbol{y}^{t-1})$ or $1 = \arg\max_{\tilde{d}_t \in \{0,1\}} \tilde{W}_t(\tilde{\boldsymbol{d}}^t, \boldsymbol{y}^{t-1})$ if and only if $1 = \arg\max_{\tilde{d}_t \in [0,1]} \tilde{W}_t(\tilde{\boldsymbol{d}}^t, \boldsymbol{y}^{t-1})$. Symmetrically, $0 = \arg\max_{\tilde{d}_t \in \{0,1\}} \tilde{W}_t(\tilde{\boldsymbol{d}}^t, \boldsymbol{y}^{t-1})$ if and only if $0 = \arg\max_{\tilde{d}_t \in [0,1]} \tilde{W}_t(\tilde{\boldsymbol{d}}^t, \boldsymbol{y}^{t-1})$. This implies that $\tilde{\delta}_t^*(\cdot) = \delta_t^*(\cdot)$ for all $t = 1, ..., T$, which proves the theorem. □

A.2 Matrices in Section 4.3

We show how to construct matrices A_k and B in (4.2) and (4.4) for the linear programming (4.6). 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, we have

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

where $\mathcal{S}_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}} \equiv \{S = \beta(\tilde{\boldsymbol{S}}) : Y_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^t) = y_t, D_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^t) = d_t \ \forall t\}$, $\tilde{\boldsymbol{S}} \equiv (\tilde{S}_1, ..., \tilde{S}_T)$ with $\tilde{S}_t \equiv (\{Y_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^t)\}, \{D_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^t)\})$, and $\beta(\cdot)$ is a one-to-one map that transforms a binary sequence into a decimal value. Then, for a $d_q \times 1$ vector $B_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}}$,

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

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 $q_{\boldsymbol{\delta}_k}(\boldsymbol{y}) \equiv \Pr[\boldsymbol{Y}(\boldsymbol{\delta}_k(\cdot)) = \boldsymbol{y}]$. For given $\boldsymbol{\delta}(\cdot)$, by repetitively applying the law of iterated expectation, we can show

$$\begin{aligned}
&\Pr[\boldsymbol{Y}(\boldsymbol{\delta}(\cdot)) = \boldsymbol{y}] \\
=&\Pr[Y_T(\boldsymbol{d}) = y_T | \boldsymbol{Y}^{T-1}(\boldsymbol{d}^{T-1}) = \boldsymbol{y}^{T-1}] \\
&\times \Pr[Y_{T-1}(\boldsymbol{d}^{T-1}) = y_{T-1} | \boldsymbol{Y}^{T-2}(\boldsymbol{d}^{T-2}) = \boldsymbol{y}^{T-2}] \times \cdots \times \Pr[Y_1(d_1) = y_1],
\end{aligned} \tag{A.6}$$

where, because of the appropriate conditioning in (A.6), the bridge variables $\boldsymbol{d} = (d_1, ..., d_T)$ satisfies

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

Therefore, (A.6) can be viewed as a linear functional of $\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}]$. To illustrate, when

$T = 2$, the welfare defined as the average counterfactual terminal outcome satisfies

$$
\begin{aligned}
E[Y_T(\boldsymbol{\delta}(\cdot))] &= \sum_{y_1} \Pr[Y_2(\delta_1, \delta_2(Y_1(\delta_1), \delta_1)) = 1 | Y_1(\delta_1) = y_1] \Pr[Y_1(\delta_1) = y_1] \\
&= \sum_{y_1} \Pr[Y_2(\delta_1, \delta_2(y_1, \delta_1)) = 1, Y_1(\delta_1) = y_1]. \qquad \text{(A.7)}
\end{aligned}
$$

Then, for a chosen $\boldsymbol{\delta}(\cdot)$, the values $\delta_1 = d_1$ and $\delta_2(y_1, \delta_1) = d_2$ at which $Y_2(\delta_1, \delta_2(y_1, \delta_1))$ and $Y_1(\delta_1)$ are defined is given in Table 1 as shown in the main text. Therefore, $E[Y_2(\boldsymbol{\delta}(\cdot))]$ can be written as a linear functional of $\Pr[Y_2(d_1, d_2) = y_2, Y_1(d_1) = y_1]$.

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 (A.6). 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 (A.5). 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 (4.3) 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 4.1

Let $\mathcal{Q}_p \equiv \{q : Bq = p\} \cap \mathcal{Q}$ be the feasible set. To prove part (i), first note that the sharp DAG can be explicitly defined as $G(\mathcal{K}, \mathcal{E}_p)$ with

$$
\mathcal{E}_p \equiv \{(k, k') \in \mathcal{K} : A_k q > A_{k'} q \text{ for all } q \in \mathcal{Q}_p\}.
$$

Here, $A_k q > A_{k'} q$ for all $q \in \mathcal{Q}_p$ if and only if $L_{k,k'} > 0$ as $L_{k,k'}$ is the sharp lower bound of $(A_k - A_{k'})q$ in (4.6). The latter is because the feasible set $\{q : Bq = p \text{ and } q \in \mathcal{Q}\}$ 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.

To prove part (ii), it is helpful to note that $\mathcal{D}_p^*$ in (4.5) can be equivalently defined as

$$\begin{aligned}\mathcal{D}_p^* &\equiv \{\boldsymbol{\delta}_{k'}(\cdot) : \nexists k \in \mathcal{K} \text{ such that } A_k q > A_{k'} q \text{ for all } q \in \mathcal{Q}_p\} \\ &= \{\boldsymbol{\delta}_{k'}(\cdot) : A_k q \leq A_{k'} q \text{ for all } k \in \mathcal{K} \text{ and some } q \in \mathcal{Q}_p\}.\end{aligned}$$

Let $\tilde{\mathcal{D}}_p^* \equiv \{\boldsymbol{\delta}_{k'}(\cdot) : \nexists k \in \mathcal{K} \text{ such that } L_{k,k'} > 0 \text{ and } k \neq k'\}$. First, we prove that $\mathcal{D}_p^* \subset \tilde{\mathcal{D}}_p^*$. Note that

$$\mathcal{D} \backslash \tilde{\mathcal{D}}_p^* = \{\boldsymbol{\delta}_{k'} : L_{k,k'} > 0 \text{ for some } k \neq k'\}.$$

Suppose $\boldsymbol{\delta}_{k'} \in \mathcal{D} \backslash \tilde{\mathcal{D}}_p^*$. Then, for some $k \neq k'$, $(A_k - A_{k'})q \geq L_{k,k'} > 0$ for all $q \in \mathcal{Q}_p$. Therefore, for such k, $A_k q > A_{k'} q$ for all $q \in \mathcal{Q}_p$, and thus $\boldsymbol{\delta}_{k'} \notin \mathcal{D}_p^* \equiv \{\arg\max_{\boldsymbol{\delta}_k} A_k q : q \in \mathcal{Q}_p\}$.

Now, we prove that $\tilde{\mathcal{D}}_p^* \subset \mathcal{D}_p^*$. Suppose $\boldsymbol{\delta}_{k'} \in \tilde{\mathcal{D}}_p^*$. Then $\nexists k \neq k'$ such that $L_{k,k'} > 0$. Equivalently, for any given $k \neq k'$, either (a) $U_{k,k'} \leq 0$ or (b) $L_{k,k'} < 0 < U_{k,k'}$. Consider (a), which is equivalent to $\max_{q \in \mathcal{Q}_p}(A_k - A_{k'})q \leq 0$. This implies that $A_k q \leq A_{k'} q$ for all $q \in \mathcal{Q}_p$. Consider (b), which is equivalent to $\min_{q \in \mathcal{Q}_p}(A_k - A_{k'})q < 0 < \max_{q \in \mathcal{Q}_p}(A_k - A_{k'})q$. This implies that $\exists q \in \mathcal{Q}_p$ such that $A_k q = A_{k'} q$. Combining these implications of (a) and (b), it should be the case that $\exists q \in \mathcal{Q}_p$ such that, for all $k \neq k'$, $A_{k'} q \geq A_k q$. Therefore, $\boldsymbol{\delta}_k \in \mathcal{D}_p^*$. $\square$

A.4 Proof of Theorem 5.1

Since $\mathcal{Q}_0$ is a finitely generated cone, finding conditions under which (5.5) holds is equivalent to finding conditions under which $\Delta^0 \boldsymbol{q}_0 = \theta$ 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 $\{\boldsymbol{q}_0 : q_s \in [0,1] \text{ for } s \in \mathcal{S}_0 \text{ and other elements are zero}\}$. For $s \in \mathcal{S}_0$, let γ_s denote the element of Δ^0. 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,

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

if and only if

$$t = \frac{\theta - \gamma_{s'}}{\gamma_s - \gamma_{s'}}.$$

But then $t \in [0,1]$ by (i), and thus such $\boldsymbol{q}_0 \in \mathcal{Q}_0$. Therefore, (5.5) 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,

$$\theta = \Delta^0 \boldsymbol{q}_0 = \gamma_s q_s$$

if and only if $q_s = \theta/\gamma_s$ (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, (5.5) holds. The proof with condition (iii) is symmetric, so omitted. $\square$

A.5 Proof of Theorem 5.2

For $s \in \mathcal{S}_0$, let γ_s denote the element of Δ^0. We prove the case with $L > 0$; the case with $U < 0$ can be symmetrically proved, so omitted. Suppose (5.6) holds, i.e., $\underline{\Delta}^0 > \theta$ and $0 > \theta$. If $\underline{\Delta}^0 > \theta$, i.e., $\gamma_s > \theta$ for all $s \in \mathcal{S}_0$, then $\sum_{s \in \mathcal{S}_0} \gamma_s q_s > \theta \sum_{s \in \mathcal{S}_0} q_s$ since $q_s \geq 0$ for all $s \in \mathcal{S}_0$ and $q_s > 0$ for some $s \in \mathcal{S}_0$ (since $\boldsymbol{q}_0$ is a nonzero vector). But $\theta \sum_{s \in \mathcal{S}_0} q_s \geq \theta$ since $\sum_{s \in \mathcal{S}_0} q_s \leq 1$ and $\theta < 0$. Combining these results, $\Delta q = \Delta^0 \boldsymbol{q}_0 - \theta > 0$ for any $q \in \mathcal{Q}$, or equivalently, $L > 0$. Conversely, when (5.6) is violated, the case falls into either one of the three conditions in Theorem 5.1 or $\theta > \max\{0, \overline{\Delta}^0\}$. The former case implies incomparability which contradicts $L > 0$. The latter case implies $U < 0$ (by an argument symmetric to above) which is contradiction since $L \leq U$. This proves necessity and sufficiency of the condition. $\square$

A.6 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 $G(\mathcal{K}, \mathcal{E})$, a *directed path* is a subgraph $G(\mathcal{K}_j, \mathcal{E}_j)$ $(1 \leq j \leq J \leq 2^{|\mathcal{K}|})$ where $\mathcal{K}_j \subset \mathcal{K}$ is a totally ordered set with initial vertex $\tilde{k}_{j,1}$.[21] 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.2. *Suppose Assumptions SX and B hold. Then, $\mathcal{D}_p^*$ defined in (4.5) satisfies*

$$\mathcal{D}_p^* = \{\boldsymbol{\delta}_{\tilde{k}_{j,1}}(\cdot) \in \mathcal{D} : 1 \leq j \leq J\}, \tag{A.8}$$

where $\tilde{k}_{j,1}$ is the initial vertex of the directed path $G(\mathcal{K}_{p,j}, \mathcal{E}_{p,j})$ of $G(\mathcal{K}, \mathcal{E}_p)$.

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

Proof. 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 $\boldsymbol{\delta}^*(\cdot; q)$ 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 6.1

Given Theorem A.2, 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 $G(\mathcal{K}_{p,j}, \mathcal{E}_{p,j})$. 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 7.1

Conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^{t-1}) = (\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^{t-1})$, it is easy to show that (7.3) implies Assumption M1. Suppose $\pi_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^{t-1}, 1) > \pi_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{z}^{t-1}, 1)$ as $\pi_t(\cdot)$ is a nontrivial function of Z_t. Then, we have

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

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

Now, we show that Assumption M1 implies (7.3) conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^{t-1})$. For each t, Assumption SX implies $Y_t(\boldsymbol{d}^t), D_t(\boldsymbol{z}^t) \perp \boldsymbol{Z}^t | (\boldsymbol{Y}^{t-1}(\boldsymbol{d}^{t-1}), \boldsymbol{D}^{t-1}(\boldsymbol{z}^{t-1}), \boldsymbol{Z}^{t-1})$, which in turn implies the following conditional independence:

$$Y_t(\boldsymbol{d}^t), D_t(\boldsymbol{z}^t) \perp \boldsymbol{Z}^t | (\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^{t-1}). \tag{A.9}$$

Conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^{t-1})$, (7.3) and (A.9) correspond to Assumption S-1 in Vytlacil (2002). Assumption R(i) and (A.9) 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 7.2

We are remained to prove that, conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^{t-1})$, (7.4) is equivalent to the second part of Assumption M2. But this proof is analogous to the proof of Lemma 7.1 by replacing the roles of D_t and Z_t with those of Y_t and D_t, respectively. Therefore, we have the desired result. □

References

Abraham, S. and L. Sun (2020): "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *MIT*. 1

Almond, D. and B. Mazumder (2013): "Fetal origins and parental responses," *Annu. Rev. Econ.*, 5, 37–56. 1

Andrews, I., T. Kitagawa, and A. McCloskey (2019): "Inference on winners," Tech. rep., National Bureau of Economic Research. 10

Athey, S. and G. W. Imbens (2018): "Design-based Analysis in Difference-In-Differences Settings with Staggered Adoption," *National Bureau of Economic Research*. 1

——— (2019): "Machine learning methods that economists should know about," *Annual Review of Economics*, 11, 685–725. 1

Athey, S. and S. Wager (2017): "Efficient policy learning," *arXiv preprint arXiv:1702.02896*. 1, 3

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

Bhattacharya, D. and P. Dupas (2012): "Inferring welfare maximizing treatment assignment under budget constraints," *Journal of Econometrics*, 167, 168–196. 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

Cellini, S. R., F. Ferreira, and J. Rothstein (2010): "The value of school facility investments: Evidence from a dynamic regression discontinuity design," *The Quarterly Journal of Economics*, 125, 215–261. 1

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, 8

Cui, Y. and E. Tchetgen Tchetgen (2020): "A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity," *Journal of the American Statistical Association*, just–accepted. 1, 2.1

Cunha, F. and J. Heckman (2007): "The technology of skill formation," *American Economic Review*, 97, 31–47. 1

Deb, R., Y. Kitamura, J. K.-H. Quah, and J. Stoye (2017): "Revealed price preference: Theory and stochastic testing," *Cowles Foundation Discussion Paper*. 1, 10

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

Hansen, P. R., A. Lunde, and J. M. Nason (2011): "The model confidence set," *Econometrica*, 79, 453–497. 10, 10

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

Hirano, K. and J. R. Porter (2009): "Asymptotics for statistical treatment rules," *Econometrica*, 77, 1683–1701. 1

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475. 1, 7

Johnson, R. C. and C. K. Jackson (2019): "Reducing inequality through dynamic complementarity: Evidence from Head Start and public school spending," *American Economic Journal: Economic Policy*, 11, 310–49. 1, 3

Kahn, A. B. (1962): "Topological sorting of large networks," *Communications of the ACM*, 5, 558–562. 6.1

Kamat, V. (2019): "Identification with latent choice sets: The case of the head start impact study," *arXiv preprint arXiv:1711.02048*. 1

Kasy, M. (2016): "Partial identification, distributional preferences, and the welfare ranking of policies," *Review of Economics and Statistics*, 98, 111–131. 1

Kasy, M. and A. Sautmann (forthcoming): "Adaptive treatment assignment in experiments for policy choice," *Econometrica*. 1

Kitagawa, T. and A. Tetenov (2018): "Who should be treated? empirical welfare maximization methods for treatment choice," *Econometrica*, 86, 591–616. 1, 10

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

Kock, A. B. and M. Thyrsgaard (2017): "Optimal sequential treatment allocation," *arXiv preprint arXiv:1705.09952*. 1

Kramer, M. S., B. Chalmers, E. D. Hodnett, Z. Sevkovskaya, I. Dzikovich, S. Shapiro, J.-P. Collet, I. Vanilovich, I. Mezen, T. Ducruet, et al. (2001): "Promotion of Breastfeeding Intervention Trial (PROBIT): a randomized trial in the Republic of Belarus," *Jama*, 285, 413–420. 3

Machado, C., A. Shaikh, and E. Vytlacil (2019): "Instrumental variables and the sign of the average treatment effect," *Journal of Econometrics*, 212, 522–555. 1, 4.4

Manski, C. F. (1997): "Monotone treatment response," *Econometrica: Journal of the Econometric Society*, 1311–1334. 7

——— (2004): "Statistical treatment rules for heterogeneous populations," *Econometrica*, 72, 1221–1246. 1, 20

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

Manski, C. F. and J. V. Pepper (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010. 7

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, 1, 2.1, 20, A.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, 8, 20

Qiu, H., M. Carone, E. Sadikova, M. Petukhova, R. C. Kessler, and A. Luedtke (2020): "Optimal individualized decision rules using instrumental variable methods," *Journal of the American Statistical Association*, just–accepted. 1

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

Stoye, J. (2012): "Minimax regret treatment choice with covariates or with limited validity of experiments," *Journal of Econometrics*, 166, 138–156. 1

Tebaldi, P., A. Torgovitsky, and H. Yang (2019): "Nonparametric estimates of demand in the california health insurance exchange," Tech. rep., National Bureau of Economic Research. 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, 7

Vytlacil, E. (2002): "Independence, monotonicity, and latent index models: An equivalence result," *Econometrica*, 70, 331–341. 7, A.8

Zhang, Y., E. B. Laber, A. Tsiatis, and M. Davidian (2015): "Using decision lists to construct interpretable and parsimonious treatment regimes," *Biometrics*, 71, 895–904. 2.1