Optimal Dynamic Treatment Regimes and Partial Welfare Ordering*

Sukjin Han
Department of Economics
University of Bristol

First Draft: August 8, 2019
This Draft: April 2, 2022

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. We relate the notion of partial ordering with a more conventional notion of partial identification using topological sorts. Practically, topological sorts can be served as a policy benchmark for a policymaker. We apply our method to understand returns to schooling and post-school training as a sequence of treatments by combining data from multiple sources. The framework of this paper can be used beyond the current context, e.g., in establishing rankings of multiple treatments or policies across different counterfactual scenarios.

*For helpful comments and discussions, the author is grateful to Donald Andrews, Isaiah Andrews, Junehyuk Jung, Yuichi Kitamura, Hiro Kaido, Shakeeb Khan, Adam McCloskey, Susan Murphy, Takuya Ura, Ed Vytlacil, Shenshen Yang, participants in the 2021 Cowles Conference, the 2021 North American Winter Meeting, the 2020 European Winter Meeting, and the 2020 World Congress of the Econometric Society, 2019 CEMMAP & WISE conference, the 2019 Asian Meeting of the Econometric Society, the Bristol Econometrics Study Group Conference, the 2019 Midwest Econometrics Group Conference, the 2019 Southern Economic Association Conference, and in seminars at UCL, U of Toronto, Simon Fraser U, Rice U, UIUC, U of Bristol, Queen Mary London, NUS, and SMU. sukjin.han@gmail.com

JEL Numbers: C14, C32, C33, C36

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

1 Introduction

Dynamic treatment regimes are dynamically personalized treatment allocations. Given that individuals are heterogeneous, allocations tailored to heterogeneity can improve overall welfare. Define a dynamic treatment regime $\boldsymbol{\delta}(\cdot)$ as a sequence of binary rules $\delta_t(\cdot)$ that map previous outcome and treatment (and possibly other covariates) onto current allocation decisions: $\delta_t(y_{t-1}, d_{t-1}) = d_t \in \{0, 1\}$ for $t = 1, ..., T$. The motivation for being adaptive to the previous outcome is that it may contain information on unobserved heterogeneity that is not captured in covariates. Then the optimal dynamic treatment regime, which is this 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 other fields of social science, despite the potentially fruitful applications of optimal dynamic regimes in various policy evaluations.

To illustrate the policy relevance of the optimal dynamic regime, consider the labor market returns to high school education and post-school training for disadvantaged individuals. A policymaker may be interested in learning a schedule of 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 a high school diploma, $\delta_2(y_1, \delta_1) \in \{0, 1\}$ assigns a job training program based on δ_1 and earlier earnings $y_1 \in \{0, 1\}$ (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 $\delta_1^* = 1$, $\delta_2^*(0, \delta_1^*) = 1$, and

$\delta_2^*(1, \delta_1^*) = 0$; that is, it turns out optimal to assign a high school diploma to all individuals and a training program to individuals with low earnings. One of the policy implications of such $\boldsymbol{\delta}^*(\cdot)$ is that the average job market performance can be improved by job trainings focusing on low performance individuals complementing with high school education. Dynamic regimes are more general than static regimes where $\delta_t(\cdot)$ is a constant function. In this sense, the optimal dynamic regime provides richer policy candidates than what can be learned from dynamic complementarity (Cunha and Heckman (2007), Cellini et al. (2010), Almond and Mazumder (2013), Johnson and Jackson (2019)). 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, in which we can at least partially learn the ranking of counterfactual welfares $W_{\boldsymbol{\delta}}$'s and hence the optimal dynamic regime $\boldsymbol{\delta}^*(\cdot)$. We view that it is important to avoid making stringent modeling assumptions in the analysis of personalized treatments, because the core motivation of the analysis is individual heterogeneity, which we want to keep intact as much as possible. Instead, we embrace the partial identification approach. Given the observed distribution of sequences of outcomes and endogenous treatments and using the instrumental variable (IV) method, we establish sharp partial ordering of welfares, and characterize the identified set of optimal regimes as a discrete subset of all possible regimes. We define welfare as a linear functional of the joint distribution of counterfactual outcomes across periods. Examples of welfare include the average counterfactual terminal 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 partial ordering and identified set proposed in this paper enable "sensitivity analyses." That is, by comparing a chosen regime (e.g., from a parametric approach) with these benchmark objects, one can determine how much the former is led by assumptions and how much is informed by data. Such a practice also allows us to gain insight into data requirements to achieve a certain level of informativeness.

The identification analysis is 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).[1] The point identification of $\boldsymbol{\delta}^*(\cdot)$ will

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

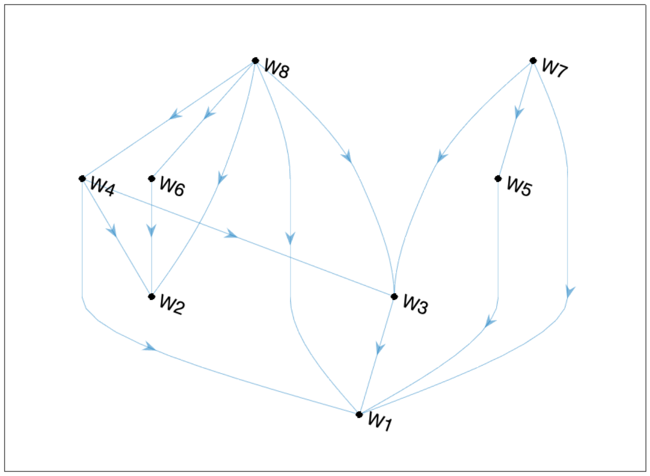


Figure 1: An Example of Sharp Partial Ordering of Welfares

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. 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 benchmark that a policymaker can be equipped with. If desired, linear programming can be solved to calculate bounds on

literature.

a small number of sorted welfares (e.g., top-tier welfares).

Given the minimal structure we impose in the data-generating process, the size of the identified set may be large in some cases. Such an identified set may still be useful in eliminating suboptimal regimes or warning about the lack of informativeness of the data. Often, however, researchers are willing to impose additional assumptions to gain identifying power. We propose identifying assumptions, such as uniformity assumptions that generalize the monotonicity assumption in Imbens and Angrist (1994), 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), Han (2020), 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 (Cui and Tchetgen Tchetgen (2020); Qiu et al. (2020)) or uniformity (Han (2020)) in a single-period setup. The dynamic effects of treatment timing (i.e., irreversible treatments) have been considered in Heckman and Navarro (2007) and Heckman et al. (2016) who utilize exclusion restrictions and infinite support assumptions. A related staggered adoption design was recently studied in multi-period difference-in-differences settings under treatment heterogeneity by Athey and Imbens (2018), Callaway and Sant'Anna (2019), and Abraham and Sun (2020).[2] This paper complements these papers by considering treatment scenarios of multiple dimensions with adaptivity as the key ingredient.

Second, this paper contributes to the literature on partial identification that utilizes linear programming approach, which 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. The advantages of this approach is that (i) bounds can be automatically obtained even when analytical derivation is not possible, (ii) the proof

[2] de Chaisemartin and d'Haultfoeuille (2020) consider a similar problem but without necessarily assuming staggered adoption.

of sharpness is straightforward and not case-by-case, and (iii) it can streamline the analysis of different identifying assumptions. The dynamic framework of this paper complicates the identification analysis and thus fully benefits from these advantages. However, a distinct feature of the present paper is that the linear programming approach is used in establishing a sharp partial ordering across counterfactual objects—a novel concept in the literature—and in such a way that separate optimizations yield a common object, namely the partial ordering. The framework of this paper can also be useful in other settings where the goal is 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.

Third, we apply our method to conduct a policy analysis with schooling and post-school training as a sequence of treatments, which is to our knowledge a novel attempt in the literature. We consider dynamic treatment regimes of allocating a high school diploma and, given pre-program earnings, a job training program for economically disadvantaged population. By combining data from the Job Training Partnership Act (JTPA), the US Census, and the National Center for Education Statistics (NCES), we construct a data set with a sequence of instruments that is used to estimate the partial ordering of expected earnings and the identified set of the optimal regime. Even though only partial orderings are recovered, we can conclude with certainty that allocating the job training program only to the low earning type is welfare optimal. We also find that more costly regimes are not necessarily welfare-improving.

The dynamic treatment regime considered in this 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. Another distinction is that we consider a dynamic setup. Finally, in order to focus on the challenge with endogeneity, we consider a simple setup where the exploration and exploitation stages are separated, unlike in the literature on bandit problems (Athey and

[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.

Imbens (2019), Kasy and Sautmann (forthcoming), Kock et al. (2021)). We believe the current setup is a good starting point.

In the next section, we introduce the dynamic regimes and related counterfactual outcomes, which define the welfare and the optimal regime. Section 3 provides a motivating example. Section 4 conducts the main identification analysis by constructing the DAG and characterizing the identified set. Sections 5–7 introduce topological sorts and additional identifying assumptions and discuss cardinality reduction for the set of regimes. Section 8 illustrates the analysis with numerical exercises, and Section 9 presents the empirical application on returns to schooling and job training. Finally, Section 10 concludes by discussing inference. In the Appendix, the analysis with binary outcomes is extended with continuous outcomes and stochastic regimes are discussed. 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 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 is a straightforward extension and thus is not considered 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)). Later, we extend the framework to allow for continuous outcome variables with discretized rules; see Appendix A.1. Whether the rule is dynamic or static, we only consider deterministic rules

[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.

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 $T = 2$

$\delta_t(\cdot) \in \{0, 1\}$. In Appendix A.2, 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] Throughout the paper, we will mostly focus on the leading case with $T = 2$ for simplicity. Also, this case already captures the essence of the dynamic features, such as adaptivity and complementarity. Table 1 lists all possible dynamic regimes $\boldsymbol{\delta}(\cdot) \equiv (\delta_1, \delta_2(\cdot))$ 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, expressing 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]:

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

[5]We can allow $\mathcal{D}$ to be a strict subset of the set of all possible regimes; see Section 7 for this relaxation.
[6]As the notation suggests, we implicitly assume the "no anticipation" condition.

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 $q_{\boldsymbol{\delta}}(\boldsymbol{y})$:

$$W_{\boldsymbol{\delta}} \equiv f(q_{\boldsymbol{\delta}}). \tag{2.4}$$

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]

$$\boldsymbol{\delta}^*(\cdot) = \arg \max_{\boldsymbol{\delta}(\cdot) \in \mathcal{D}} W_{\boldsymbol{\delta}}. \tag{2.5}$$

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.2.

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

[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.

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 high school education and post-school training for disadvantaged individuals. Let $D_{i1} = 1$ if student i has a high school diploma and $D_{i1} = 0$ otherwise; let $D_{i2} = 1$ if i participates in a job training program and $D_{i2} = 0$ if not. Also, let $Y_{i1} = 1$ if i is employed before the training program and $Y_{i1} = 0$ if not; let $Y_{i2} = 1$ if i is employed after the program and $Y_{i2} = 0$ if not. Given the data, suppose we are interested in recovering regimes that maximize the employment rate as welfare.

First, consider a static regime, which is a schedule $\boldsymbol{d} = (d_1, d_2) \in \{0,1\}^2$ of first assigning a high school diploma ($d_1 \in \{0,1\}$) and then a job training ($d_2 \in \{0,1\}$). Define associated welfare, which is the employment rate $W_{\boldsymbol{d}} = E[Y_2(\boldsymbol{d})]$. This setup is already useful in learning, for example, $E[Y_2(1,0)] - E[Y_2(0,1)]$ or complementarity (i.e., $E[Y_2(0,1)] - E[Y_2(0,0)]$ versus $E[Y_2(1,1)] - E[Y_2(1,0)]$), which cannot be learned from period-specific treatment effects. However, because d_1 and d_2 are not simultaneously given but d_1 precedes d_2, the allocation d_2 can be more informed by incorporating the knowledge about the individual's response y_1 to d_1. This motivates the dynamic regime, which is the schedule $\boldsymbol{\delta}(\cdot) = (\delta_1, \delta_2(\cdot)) \in \mathcal{D}$ of allocation *rules* that first assigns a high school diploma ($\delta_1 \in \{0,1\}$) and then a job training ($\delta_2(y_1, \delta_1) \in \{0,1\}$) depending on δ_1 and the employment status y_1. Then, the optimal regime with adaptivity $\boldsymbol{\delta}^*(\cdot)$ is the one that maximizes $W_{\boldsymbol{\delta}} = E[Y_2(\boldsymbol{\delta})]$. 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. 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). In observational settings, one can use IVs from quasi-experiments, those from RD design, or a combination of them. In the example above, we can use the distance to high schools or the number of high schools per square mile as an instrument Z_1 for D_1. Then, a random assignment of the job training in a field experiment can be used as an instrument Z_2 for the compliance decision D_2. In fact, in Section 9, we study schooling and job training as a sequence of treatments and combine IVs from experimental and observational data.

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 8 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 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

[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

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 $\boldsymbol{d}^t \mapsto y_t$ and $\boldsymbol{z}^t \mapsto d_t$ for $t = 1, ..., T$ under the exclusion restriction in Assumption SX.[10] We introduce the latent state variable $\tilde{S}_t$ whose realiza-

[10] A more primitive state of the world will determines 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 $t = 1, ..., T$, but this is not relevant to our objective as shown below.

tion represents such a state. We define $\tilde{S}_t$ as follows. For given $(\boldsymbol{d}^t, \boldsymbol{z}^t)$, let $Y_t(\boldsymbol{d}^t)$ and $D_t(\boldsymbol{z}^t)$ denote the extended counterfactual outcomes and treatments, respectively, and let $\{Y_t(\boldsymbol{d}^t)\}_{\boldsymbol{d}^t}$ and $\{D_t(\boldsymbol{z}^t)\}_{\boldsymbol{z}^t}$ denote their sequences w.r.t. $\boldsymbol{d}^t$ and $\boldsymbol{z}^t$. Then, by concatenating the two sequences, define $\tilde{S}_t \equiv (\{Y_t(\boldsymbol{d}^t)\}, \{D_t(\boldsymbol{z}^t)\}) \in \{0,1\}^{2^t} \times \{0,1\}^{2^t}$. For example, $\tilde{S}_1 = (Y_1(0), Y_1(1), D_1(0), D_1(1)) \in \{0,1\}^2 \times \{0,1\}^2$, whose realization specifies particular maps $d_1 \mapsto y_1$ and $z_1 \mapsto d_1$. It is convenient to transform $\tilde{\boldsymbol{S}} \equiv (\tilde{S}_1, ..., \tilde{S}_T)$ 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 distributions of counterfactuals can be written as linear functionals of q:

$$\begin{aligned}
\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}] &= \Pr[S \in \mathcal{S} : Y_t(\boldsymbol{d}^t) = y_t \quad \forall t] \\
&= \sum_{s \in \mathcal{S}_{\boldsymbol{y},\boldsymbol{d}}} q_s, \qquad (4.1) \\
\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}, \boldsymbol{D}(\boldsymbol{z}) = \boldsymbol{d}] &= \Pr[S \in \mathcal{S} : \boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}, \boldsymbol{D}(\boldsymbol{z}) = \boldsymbol{d}] \\
&= \Pr[S \in \mathcal{S} : Y_t(\boldsymbol{d}^t) = y_t, D_t(\boldsymbol{z}^t) = d_t \quad \forall t] \\
&= \sum_{s \in \mathcal{S}_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}}} q_s, \qquad (4.2)
\end{aligned}$$

where $\mathcal{S}_{\boldsymbol{y},\boldsymbol{d}}$ and $\mathcal{S}_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}}$ are constructed by using the definition of S; their expressions can be found in Appendix A.3.

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

$$W_k = A_k q. \qquad (4.3)$$

The formal derivation of A_k can be found in Appendix A.3, 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 (4.3) 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 4 in Table 1 yields

$$\Pr[Y_2(\boldsymbol{\delta}_4(\cdot))=1]=P[\boldsymbol{Y}(1,1)=(1,1)]+P[\boldsymbol{Y}(1,0)=(0,1)], \tag{4.4}$$

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

$$Bq=p, \tag{4.5}$$

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.3. It is worth noting that the linearity in (4.3) and (4.5) 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 $W_k=A_kq$ 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.[11]

For a given q, let $\boldsymbol{\delta}^*(\cdot;q)\equiv\arg\max_{\boldsymbol{\delta}_k(\cdot)\in\mathcal{D}}W_k=A_kq$ 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 p is*

$$\mathcal{D}_p^*\equiv\{\boldsymbol{\delta}^*(\cdot;q):Bq=p\text{ and }q\in\mathcal{Q}\}\subset\mathcal{D}, \tag{4.6}$$

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

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

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 $k > k'$.[12] Then the identified set can be characterized based on the resulting partial ordering.

The nature of the data generation induces the linear system (4.3) and (4.5). 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:

$$\begin{aligned} U_{k,k'} &= \max\nolimits_{q \in \mathcal{Q}} \Delta_{k,k'}q, \\ L_{k,k'} &= \min\nolimits_{q \in \mathcal{Q}} \Delta_{k,k'}q, \end{aligned} \qquad s.t. \quad Bq = p. \tag{4.7}$$

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 characterize the identified set using $U_{k,k'}$ and $L_{k,k'}$ for $k, k' \in \mathcal{K}$ and $k > k'$, or equivalently, $L_{k,k'}$ for $k, k' \in \mathcal{K}$ and $k \neq k'$.[13]

Theorem 4.1. *Suppose Assumptions SX and B hold. Then, (i) $G(\mathcal{K}, \mathcal{E}_p)$ with $\mathcal{E}_p \equiv \{(k, k') \in \mathcal{K} : L_{k,k'} > 0$ and $k \neq k'\}$ is sharp; (ii) $\mathcal{D}_p^*$ defined in (4.6) satisfies*

$$\begin{aligned} \mathcal{D}_p^* &= \{\boldsymbol{\delta}_{k'}(\cdot) : \nexists k \in \mathcal{K} \text{ such that } L_{k,k'} > 0 \text{ and } k \neq k'\} \qquad (4.8)\\ &= \{\boldsymbol{\delta}_{k'}(\cdot) : L_{k,k'} \leq 0 \text{ for all } k \in \mathcal{K} \text{ and } k \neq k'\}, \qquad (4.9) \end{aligned}$$

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

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 partial ordering is governed by *common* q's (each of which generates all the welfares) that are observationally equivalent; see Section 5.2 for related discussions.

Theorem 4.1(i) prescribes how to calculate the sharp DAG as a function of data.[14] Ac-

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

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

[14]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.

cording to (4.8) 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 may recommend the policymaker eliminate sub-optimal regimes from her options;[15] (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 6, we list further identifying assumptions that help shrink $\mathcal{D}_p^*$.

5 Set of the n-th Best Regimes, Topological Sorts, and Bounds on Sorted Welfare

In this section, we propose some ways to report results of this paper including the partial ordering. These approaches can be useful especially when the obtained partial ordering is complicated (e.g., with a longer horizon).

5.1 Set of the n-th Best Policies

When the partial ordering of welfare is the parameter of interest, the identified set of $\boldsymbol{\delta}^*(\cdot)$ can be viewed as a summary of the partial ordering. This view can be extended to introduce a set of the n-th best regimes, which further summarizes the partial ordering. With slight abuse of notation, we can formalize it as follows.

Recall $\mathcal{K}$ is the set of all regime indices. Motivated from (4.8), let $\mathcal{K}_p^{(1)} \equiv \{k' : \nexists k \in \mathcal{K} \text{ such that } L_{k,k'} > 0 \text{ and } k \neq k' \in \mathcal{K}\}$ be the set of maximal elements of the partial ordering and let $\mathcal{D}_p^{(1)} \equiv \{\boldsymbol{\delta}_{k'}(\cdot) : k' \in \mathcal{K}_p^{(1)}\}$. Theorem 4.1(ii) can be simply stated as $\mathcal{D}_p^* = \mathcal{D}_p^{(1)}$. To define the set of second-best regimes, we first remove all the elements in $\mathcal{K}_p^{(1)}$ from the set of

[15] Section 10 discusses how to do this systematically after embracing sampling uncertainty.

candidate. Accordingly, by defining

$$\mathcal{K}_p^{(2)} \equiv \{k' : \nexists k \in \mathcal{K} \backslash \mathcal{K}_p^{(1)} \text{ such that } L_{k,k'} > 0 \text{ and } k \neq k' \in \mathcal{K} \backslash \mathcal{K}_p^{(1)}\},$$

we can introduce the set of second-best regimes: $\mathcal{D}_p^{(2)} \equiv \{\boldsymbol{\delta}_{k'}(\cdot) : k' \in \mathcal{K}_p^{(2)}\}$. Iteratively, we can define the set of n-th best regimes as $\mathcal{D}_p^{(n)} \equiv \{\boldsymbol{\delta}_{k'}(\cdot) : k' \in \mathcal{K}_p^{(n)}\}$ where

$$\mathcal{K}_p^{(n)} = \left\{k' : \nexists k \in \mathcal{K} \backslash \bigcup_{j=1}^{n-1} \mathcal{K}_p^{(j)} \text{ such that } L_{k,k'} > 0 \text{ and } k \neq k' \in \mathcal{K} \backslash \bigcup_{j=1}^{n-1} \mathcal{K}_p^{(j)}\right\}.$$

The sets $\mathcal{D}_p^{(1)}, ..., \mathcal{D}_p^{(n)}$ are useful policy benchmarks. For instance, the policy maker can conduct a sensitivity analysis for her chosen regime (e.g., from a parametric model) by inspecting in which set the regime is contained.

5.2 Topological Sorts as Observational Equivalence

Another way to summarize the partial ordering 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 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 5.1. *Suppose Assumptions SX and B hold. The identified set $\mathcal{D}_p^*$ defined in (4.6) 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

$$\left(\boldsymbol{\delta}_{k_{l,1}}(\cdot), \boldsymbol{\delta}_{k_{l,2}}(\cdot), ..., \boldsymbol{d}_{k_{l,|\mathcal{D}|}}(\cdot)\right) \tag{5.1}$$

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

5.3 Bounds on Sorted Welfares

The set of n-th best regimes and topological sorts provide ordinal information about counterfactual welfares. To gain more comprehensive knowledge about the welfares, they 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, the sharp lower and upper bounds on welfare W_k can be calculated via

$$\begin{aligned} U_k &= \max\nolimits_{q \in \mathcal{Q}} A_k q, \\ L_k &= \min\nolimits_{q \in \mathcal{Q}} A_k q, \end{aligned} \qquad s.t. \quad Bq = p. \tag{5.2}$$

6 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 (4.7). These assumptions tighten the identified set $\mathcal{D}_p^*$ by reducing the dimension of simplex $\mathcal{Q}$, and thus producing a denser DAG.[16]

To incorporate these assumptions, we extend the framework introduced in Sections 4–5. 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

$$\bar{\mathcal{Q}} \equiv \{\bar{q} : \sum_s \bar{q}_s = 1 \text{ and } \bar{q}_s \geq 0 \ \forall s\}. \tag{6.1}$$

[16]Similarly, when these assumptions are incorporated in (5.2), we obtain tighter bounds on welfares.

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 (4.3) and (4.5) as

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

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

$$\bar{\mathcal{D}}_p^* \equiv \{\boldsymbol{\delta}^*(\cdot;\bar{q}) : B\bar{q} = p \text{ and } \bar{q} \in \mathcal{Q}\} \subset \mathcal{D}, \tag{6.2}$$

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, Theorem 4.1 can be modified by replacing q, B and Δ with $\bar{q}$, $\bar{B}$ 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 $\boldsymbol{D}^t$. For example, we can have $D_t(\boldsymbol{Z}^{t-1}, 1) \geq D_t(\boldsymbol{Z}^{t-1}, 0)$ for $D_{t-1} = 1$ while $D_t(\boldsymbol{Z}^{t-1}, 1) \leq D_t(\boldsymbol{Z}^{t-1}, 0)$ for $D_{t-1} = 0$. 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:

$$D_t = 1\{\pi_t(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^t) \geq \nu_t\}, \tag{6.3}$$

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

Lemma 6.1. *Suppose Assumption SX holds and $\Pr[D_t = 1|\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^t]$ is a nontrivial function of Z_t. Assumption M1 is equivalent to (6.3) being satisfied conditional on*

$(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^{t-1})$ *for each* t*.*

The dynamic selection model (6.3) 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 (6.3) 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 provided that a more granular mapping is considered as mentioned in footnote 10.

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.[17]

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* $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1})$*.*

This assumption postulates uniformity in a way that restricts heterogeneity of the contemporaneous treatment effect. However, similarly as before, without being conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1})$, there can be a general non-monotonic pattern in the way that $\boldsymbol{D}^t$ influences $\boldsymbol{Y}^t$. For example, we can have $Y_t(\boldsymbol{D}^{t-1}, 1) \geq Y_t(\boldsymbol{D}^{t-1}, 0)$ for $Y_{t-1} = 1$ while $Y_t(\boldsymbol{D}^{t-1}, 1) \leq Y_t(\boldsymbol{D}^{t-1}, 0)$ for $Y_{t-1} = 0$. It is also worth noting that Assumption M2 (and M1) does not assume the direction of monotonicity, but the direction is recovered from the data. 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

[17]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.

before, Assumption M2 is the equivalent of a dynamic version of a nonparametric triangular model:

$$Y_t = 1\{\mu_t(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^t) \geq \varepsilon_t\}, \tag{6.4}$$

$$D_t = 1\{\pi_t(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^t) \geq \nu_t\}, \tag{6.5}$$

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

Lemma 6.2. *Suppose Assumption SX holds,* $\Pr[D_t = 1|\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^t]$ *is a 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* (6.4)–(6.5) *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 (6.4)–(6.5), Assumption K implies

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

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

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).

7 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 or budget 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.

8 Numerical Studies

We conduct numerical exercises to illustrate (i) the theoretical results developed in Sections 4–5, (ii) the role of the assumptions introduced in Section 6, and (iii) the overall computational scale of the problem. For $T = 2$, we consider the following data-generating process:

$$D_{i1} = 1\{\pi_1 Z_{i1} + \alpha_i + v_{i1} \geq 0\}, \tag{8.1}$$

$$Y_{i1} = 1\{\mu_1 D_{i1} + \alpha_i + e_{i1} \geq 0\}, \tag{8.2}$$

$$D_{i2} = 1\{\pi_{21} Y_{i1} + \pi_{22} D_{i1} + \pi_{23} Z_{i2} + \alpha_i + v_{i2} \geq 0\}, \tag{8.3}$$

$$Y_{i2} = 1\{\mu_{21} Y_{i1} + \mu_{22} D_{i2} + \alpha_i + e_{i2} \geq 0\}, \tag{8.4}$$

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 (8.1)–(8.4) 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., $|\mathcal{D}| = |\mathcal{K}| = 8$. We 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.[18] As the researcher, we maintain Assumption K. Then, for each linear program, the dimension of q is $|\mathcal{Q}| + 1 = |\mathcal{S}| = |\mathcal{S}_1| \times |\mathcal{S}_2| = 2^2 \times 2^2 \times 2^8 \times 2^4 = 65,536$.[19]

[18] There are more efficient algorithms than the bubble sort, 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.

[19] The dimension is reduced with additional identifying assumptions.

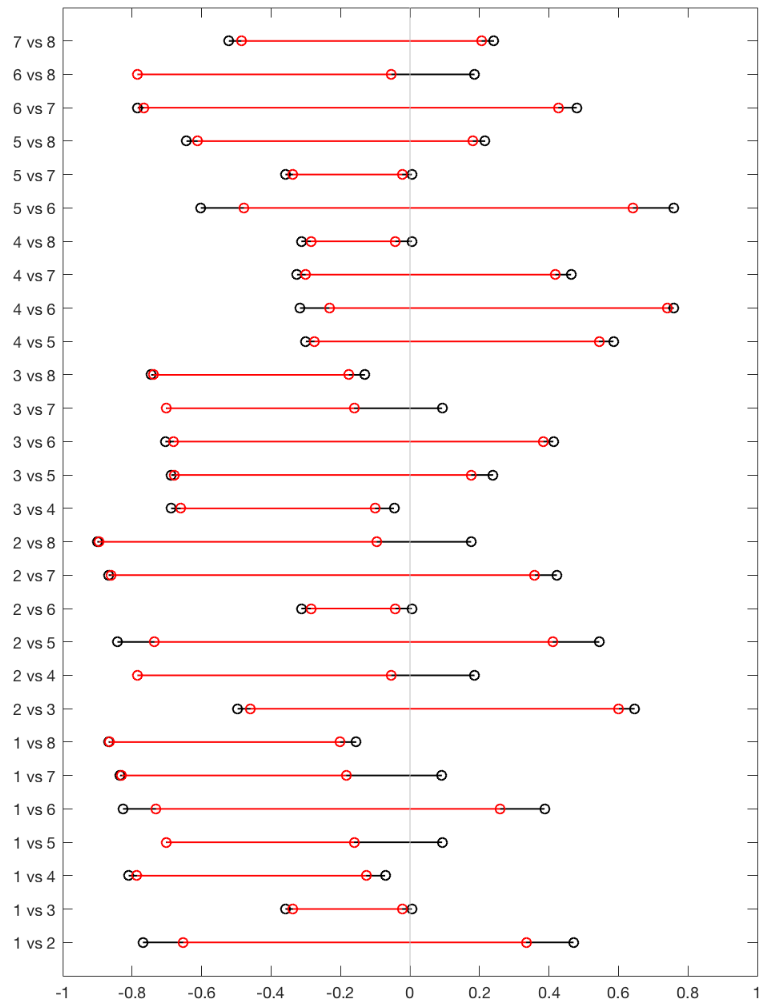


Figure 3: Sharp Bounds on Welfare Gaps under M1 (black) and M2 (red)

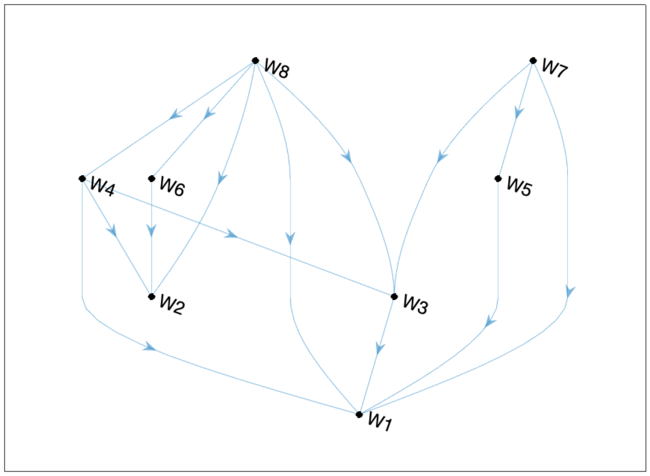


Figure 4: Sharp Directed Acyclic Graph under M2

The number of main constraints is $\dim(p) = 2^{3\times 2} - 2^2 = 60$. There are $1 + 65,536$ additional constraints that define the simplex, i.e., $\sum_s q_s = 1$ and $q_s \geq 0$ for all $s \in \mathcal{S}$. Each linear program takes less than a second to calculate $L_{k,k'}$ or $U_{k,k'}$ with a computer with a 2.2 GHz single-core processor and 16 GB memory and with a modern solver such as CPLEX, MOSEK, and GUROBI.

Figure 3 reports the bounds $(L_{k,k'}, U_{k,k'})$ on $W_k - W_{k'}$ for all $(k, k') \in \{1, ..., 8\}$ under Assumption M1 (in black) and Assumption M2 (in red). In the figure, we can determine the sign of the welfare gap for those bounds that 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. It is important to note that, because M2 does not assume the direction of monotonicity, the sign of the welfare gap is not imposed by the assumption but recovered from the data.[20] Each set of bounds generates an associated DAGs (produced as an 8×8 adjacency matrix).[21] We proceed with Assumption M2 for brevity.

Figure 4 (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

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

[20]The direction of the monotonicity in M2 can be estimated directly from the data by using the fact that $\text{sign}(E[Y_t|Z_t = 1, \boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}] - E[Y_t|Z_t = 1, \boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}]) = \text{sign}(E[Y_t(D^{t-1}, 1)|\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}] - E[Y_t(D^{t-1}, 0)|, \boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}])$ almost surely. This result is an extension of Shaikh and Vytlacil (2011) to our multi-period setting.

[21]Given the solutions of the linear programs, the adjacency matrix and thus the graph is simple to produce automatically using a standard software such as MATLAB.

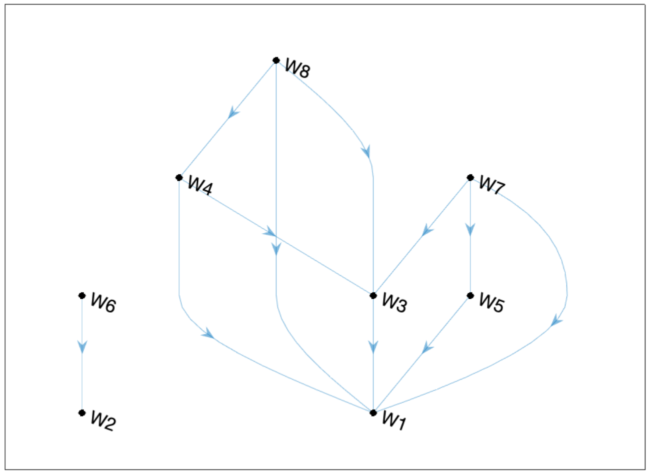


Figure 5: Sharp Directed Acyclic Graph under M2 (with only Z_1)

The common feature of the elements in $\mathcal{D}_p^*$ is that it is optimal to allocate $\delta_2 = 1$ for all $y_1 \in \{0, 1\}$. 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 (8.1)–(8.4) 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 DAG for Case (b) is shown in Figure 5. In this case, we obtain $\mathcal{D}_p^* = \{\boldsymbol{\delta}_6(\cdot), \boldsymbol{\delta}_7(\cdot), \boldsymbol{\delta}_8(\cdot)\}$.

9 Application

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

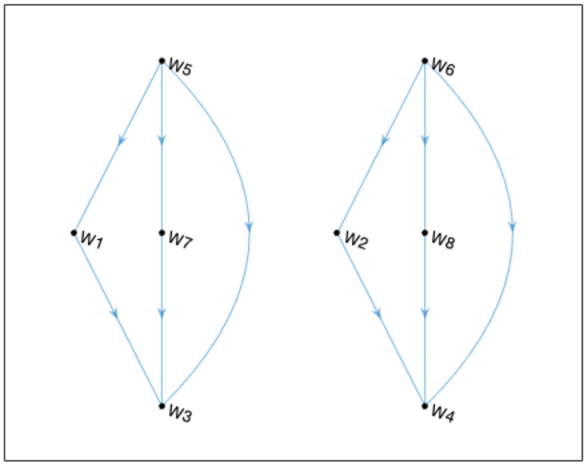
We are interested in the dynamic treatment regime $\boldsymbol{\delta}(\cdot) = (\delta_1, \delta_2(\cdot))$, where δ_1 is a HS diploma and $\delta_2(y_1)$ is the job training program given pre-program earning type y_1. The motivation of having δ_2 as a function of y_1 comes from acknowledging the dynamic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

For the purpose of our analysis, we combine the JTPA data with data from the US Census and the National Center for Education Statistics (NCES), from which we construct the following set of variables: Y_2 above or below median of 30-month earnings, D_2 the job training program, Z_2 a random assignment of the program, Y_1 above or below 80th percentile of pre-program earnings, D_1 the HS diploma or GED, and Z_1 the number of high schools per square mile.[22] The instrument Z_1 for the HS treatment appears in the literature (e.g., Neal (1997)). The number of individuals in the sample is 9,223. We impose Assumptions SX and M2 throughout the analysis.

The estimation of the DAG and the identified set $\mathcal{D}_p^*$ is straightforward given the conditions in Theorem 4.1 and the linear programs (4.7). The only unknown object is p, the joint distribution of $(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z})$, which can be estimated as $\hat{p}$, a vector of $\hat{p}_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}} = \sum_{i=1}^N 1\{\boldsymbol{Y}_i = \boldsymbol{y}, \boldsymbol{D}_i = \boldsymbol{d}, \boldsymbol{Z}_i = \boldsymbol{z}\} / \sum_{i=1}^N 1\{\boldsymbol{Z}_i = \boldsymbol{z}\}$.

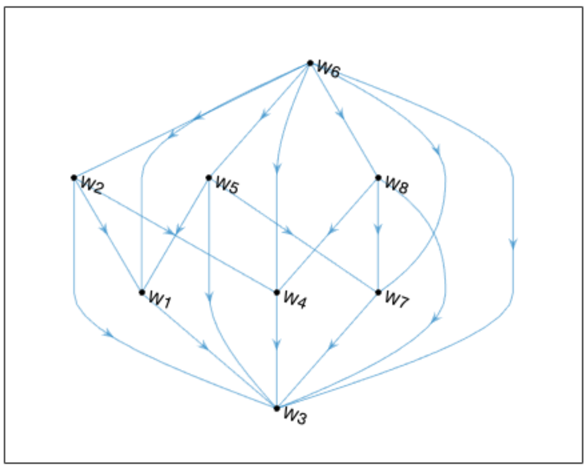
Figure 6 reports the estimated partial ordering of welfare $W_{\boldsymbol{\delta}} = E[Y_2(\boldsymbol{\delta}(\cdot))]$ (left) and the resulting estimated set $\hat{\mathcal{D}}$ (right, highlighted in red) that we estimate using $\{(\boldsymbol{Y}_i, \boldsymbol{D}_i, \boldsymbol{Z}_i)\}_{i=1}^{9,223}$. Although there exist welfares that cannot be ordered, we can conclude *with certainty* that allocating the program only to the low earning type ($Y_2 = 0$) is welfare optimal, as it is the common implication of Regimes 5 and 6 in $\hat{\mathcal{D}}$. Also, the second best policy is to either

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



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

Figure 6: Estimated DAG of $W_{\boldsymbol{\delta}} = E[Y_2(\boldsymbol{\delta}(\cdot))]$ and Estimated Set for $\boldsymbol{\delta}^*$ (red)



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

Figure 7: Estimated DAG of $W_{\boldsymbol{\delta}} = E[Y_1(\delta_1)] + E[Y_2(\boldsymbol{\delta}(\cdot))]$ and Estimated Set for $\boldsymbol{\delta}^*$ (red)

allocate the program to the entire population or none, while allocating it only to the high earning type ($Y_2 = 1$) produces the lowest welfare. This result is consistent with the eligibility of Title II of the JTPA, which concerns individuals with "barriers to employment" where the most common barriers are unemployment spells and high-school dropout status (Abadie et al. (2002)). Possibly due to the fact that the first-stage instrument Z_1 is not strong enough, we have the two disconnected sub-DAGs and thus the two elements in $\hat{\mathcal{D}}$, which are agnostic about the optimal allocation in the first stage or the complementarity between the first- and second- stage allocations.

Figure 7 reports the estimated partial ordering and the estimated set with $W_{\boldsymbol{\delta}} = E[Y_1(\delta_1)] + E[Y_2(\boldsymbol{\delta}(\cdot))]$. Despite the partial ordering, $\hat{\mathcal{D}}$ is a singleton for this welfare and $\boldsymbol{\delta}^*$ is estimated to be Regime 6. According to this regime, the average lifetime earning is maximized by allocating HS education to all individuals and the training program to individuals with low pre-program earnings. As discussed earlier, additional policy implications can be obtained by inspecting suboptimal regimes. Interestingly, Regime 8, which allocates the treatments regardless, is inferior to Regime 6. This can be useful knowledge for policy makers especially because Regime 8 is the most "expensive" regime. Similarly, Regime 1, which does not allocate any treatments regardless and thus is the least expensive regime, is superior to Regime 3, which allocates the program to high-earning individuals. The estimated DAG shows how more expensive policies do not necessarily achieve greater welfare. Moreover, these conclusions can be compelling as they are drawn without making arbitrary parametric restrictions nor strong identifying assumptions.

Finally, as an alternative approach, we use $\{(\boldsymbol{Y}_i, \boldsymbol{D}_i, Z_{2i})\}_{i=1}^{9,223}$ for estimation, that is, we drop Z_1 and only use the exogenous variation from Z_2. This reflects a possible concern that Z_1 may not be as valid as Z_2. Then, the estimated DAG looks identical to the left panel of Figure 6 whether the targeted welfare is $E[Y_2(\boldsymbol{\delta}(\cdot))]$ or $E[Y_1(\delta_1)] + E[Y_2(\boldsymbol{\delta}(\cdot))]$. Clearly, without Z_1, the procedure lacks the ability to determine the first stage's best treatment. Note that, even though the DAG for $E[Y_2(\boldsymbol{\delta}(\cdot))]$ is identical for the case of one versus two instruments, the inference results will reflect such difference by producing a larger confidence set for the former case.

10 Inference

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 (4.8) or (4.9). For each test given $\tilde{\mathcal{K}} \subset \mathcal{K}$, we construct a null hypothesis that W_k and $W_{k'}$ are not comparable for all $k, k' \in \tilde{\mathcal{K}}$. Given (4.7), the incomparability of W_k and $W_{k'}$ is equivalent to $L_{k,k'} \le 0 \le U_{k,k'}$. In constructing the null hypothesis, it is helpful to invoke strong duality for the primal programs (4.7) and write the following dual programs:

$$U_{k,k'} = \min_{\lambda} \tilde{p}'\lambda, \qquad s.t. \quad \tilde{B}'\lambda \ge \Delta'_{k,k'} \tag{10.1}$$

$$L_{k,k'} = \max_{\lambda} -\tilde{p}'\lambda, \qquad s.t. \quad \tilde{B}'\lambda \ge -\Delta'_{k,k'} \tag{10.2}$$

where $\tilde{B} \equiv \begin{bmatrix} B \\ \mathbf{1}' \end{bmatrix}$ is a $(d_p+1) \times d_q$ matrix with $\mathbf{1}$ being a $d_q \times 1$ vector of ones and $\tilde{p} \equiv \begin{bmatrix} p \\ 1 \end{bmatrix}$ is a $(d_p + 1) \times 1$ vector. By using a vertex enumeration algorithm (e.g., Avis and Fukuda

(1992)), one can find all (or a relevant subset) of vertices of the polyhedra $\{\lambda : \tilde{B}'\lambda \geq \Delta'_{k,k'}\}$ and $\{\lambda : \tilde{B}'\lambda \geq -\Delta'_{k,k'}\}$. Let $\Lambda_{U,k,k'} \equiv \{\lambda_1, ..., \lambda_{J_{U,k,k'}}\}$ and $\Lambda_{L,k,k'} \equiv \{\lambda_1, ..., \lambda_{J_{L,k,k'}}\}$ be the sets that collect such vertices, respectively. Then, it is easy to see that $U_{k,k'} = \min_{\lambda \in \Lambda_{U,k,k'}} \tilde{p}'\lambda$ and $L_{k,k'} = \max_{\lambda \in \Lambda_{L,k,k'}} -\tilde{p}'\lambda$. Therefore, the null hypothesis that $L_{k,k'} \leq 0 \leq U_{k,k'}$ can be written as

$$H_{0,\tilde{\mathcal{K}}} : \tilde{p}'\lambda > 0 \text{ for all } \lambda \in \Lambda_{\tilde{\mathcal{K}}}. \tag{10.3}$$

where $\Lambda_{\tilde{\mathcal{K}}} \equiv \bigcup_{k,k' \in \tilde{\mathcal{K}}} \Lambda_{k,k'}$ with $\Lambda_{k,k'} \equiv \Lambda_{U,k,k'} \cup \Lambda_{L,k,k'}$.

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, $T_{\tilde{\mathcal{K}}} \equiv \min_{k,k' \in \tilde{\mathcal{K}}} t_{k,k'}$ can be used as the test statistic for $H_{0,\tilde{\mathcal{K}}}$ where $t_{k,k'} \equiv \min_{\lambda \in \Lambda_{k,k'}} t_\lambda$ and t_λ is a standard t-statistic. The distribution of $T_{\tilde{\mathcal{K}}}$ can be estimated using bootstrap. In Step 2, a candidate for $k_{\tilde{\mathcal{K}}}$ is $k_{\tilde{\mathcal{K}}} \equiv \arg\min_{k \in \tilde{\mathcal{K}}} \min_{k' \in \tilde{\mathcal{K}}} t_{k,k'}$.

The eliminated vertices (i.e., regimes) are statistically suboptimal regimes, which are already policy-relevant outputs of the procedure. Note that the null hypothesis (10.3) consists of multiple inequalities. This incurs the issue of uniformity in that the null distribution depends on binding inequalities, whose identities are unknown. Such a problem has been studied in the literature, as in Hansen (2005), Andrews and Soares (2010), and Chen and Szroeter (2014). Hansen et al. (2011)'s bootstrap approach for constructing the model confidence set builds on Hansen (2005). We apply a similar inference method as in Hansen et al. (2011), but in this novel context and by being conscious about the computational challenge of our problem. In particular, the dual problem (10.1)–(10.2) and the vertex enumeration algorithm are introduced to ease the computational burden in simulating the distribution of $T_{\tilde{\mathcal{K}}}$. That is, the calculation of $\Lambda_{\tilde{\mathcal{K}}}$, the computationally intensive step, occurs only once, and then for each bootstrap sample, it suffices to calculate $\hat{p}$ instead of solving the linear programs (4.7) for all $k, k' \in \tilde{\mathcal{K}}$.

Analogous to 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)* $\limsup_{n\to\infty} \Pr[\phi_{\tilde{\mathcal{K}}} = 1 | H_{0,\tilde{\mathcal{K}}}] \leq \alpha$, *(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* $\liminf_{n\to\infty} \Pr[\mathcal{D}_p^* \subset \widehat{\mathcal{D}}_{CS}] \geq 1-\alpha$ *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. 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 (5.2) 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 Extension: Continuous Y_t

Suppose the outcomes Y_t's are continuously distributed on $[y_l, y_u]$. Consider the treatment allocation $\tilde{\delta}_t$ with continuous $y_t \in [y_l, y_u]$ and binary $d_t \in \{0,1\}$:

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

This rule may not be a feasible or practical strategy considering the cost of incrementally customizing the allocation based on continuous characteristics. Instead, the planner may want to employ a regime that is only discretely adaptive to the continuous outcomes. This can be achieved by a threshold-crossing allocation rule: for each $t = 1, ..., T$,

$$\tilde{\delta}_t(\boldsymbol{y}^{t-1}, \boldsymbol{d}^{t-1}; \boldsymbol{\gamma}^{t-1}) = \delta_t(1\{y_1 \geq \gamma_1\}, ..., 1\{y_{t-1} \geq \gamma_{t-1}\}, \boldsymbol{d}^{t-1}), \tag{A.2}$$

where $\boldsymbol{\gamma}^{t-1} \equiv (\gamma_1, ..., \gamma_{t-1})$ is a threshold vector and $\delta_t(\cdot)$ is the original treatment allocation rule based on discrete outcomes. This threshold-crossing rule is a popular decision rule in practice due to its intuitive form and is considered in earlier theoretical studies such as in Murphy (2003). Note that the full regime $(\tilde{\delta}_1, \tilde{\delta}_2(\cdot;\gamma_1), ..., \tilde{\delta}_T(\cdot;\boldsymbol{\gamma}))$ where $\boldsymbol{\gamma} \equiv (\gamma_1, ..., \gamma_{T-1})$ can be characterized by $(\boldsymbol{\delta}(\cdot), \boldsymbol{\gamma})$ where $\boldsymbol{\delta}(\cdot)$ the original regime with discrete outcomes. Therefore, we proceed with latter in the following analysis.

Based on $(\boldsymbol{\delta}, \boldsymbol{\gamma}) \in \mathcal{D} \times \Gamma$, we define the welfare $W_{\boldsymbol{\delta},\boldsymbol{\gamma}}$ analogous to (2.4). For example, $W_{\boldsymbol{\delta},\boldsymbol{\gamma}} = E[Y_T(\boldsymbol{\delta}, \boldsymbol{\gamma})]$ where $Y_T(\boldsymbol{\delta}, \boldsymbol{\gamma})$ is defined as (2.2)–(2.3) but each $\delta_t(\cdot)$ replaced with

$\tilde{\delta}_t(\cdot;\boldsymbol{\gamma}^{t-1})$ defined above. We wish to find $(\boldsymbol{\delta}^*,\boldsymbol{\gamma}^*)$ that maximize welfare $W_{\boldsymbol{\delta},\boldsymbol{\gamma}}$:

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

This maximization problem, equivalently the identification of $(\boldsymbol{\delta}^*,\boldsymbol{\gamma}^*)$, is challenging because $W_{\boldsymbol{\delta},\boldsymbol{\gamma}}$ may not be point identified. Therefore, analogous to the partial identification approach in the main text, we proceed as follows. For a given pair $(\boldsymbol{\delta},\boldsymbol{\gamma})$ and $(\boldsymbol{\delta}',\boldsymbol{\gamma}')$ in $\mathcal{D}\times\Gamma$, let $L(\boldsymbol{\delta},\boldsymbol{\gamma},\boldsymbol{\delta}',\boldsymbol{\gamma}')$ be the lower bound on the welfare gap

$$W_{\boldsymbol{\delta},\boldsymbol{\gamma}} - W_{\boldsymbol{\delta}',\boldsymbol{\gamma}'}.$$

Then, the identified set for $(\boldsymbol{\delta}^*,\boldsymbol{\gamma}^*)$ can be characterized as

$$\{(\boldsymbol{\delta}',\boldsymbol{\gamma}') : L(\boldsymbol{\delta},\boldsymbol{\gamma},\boldsymbol{\delta}',\boldsymbol{\gamma}') \leq 0 \text{ for all } (\boldsymbol{\delta},\boldsymbol{\gamma}) \in \mathcal{D}\times\Gamma \text{ and } (\boldsymbol{\delta},\boldsymbol{\gamma}) \neq (\boldsymbol{\delta}',\boldsymbol{\gamma}')\}. \tag{A.3}$$

Note that, for given $\boldsymbol{\gamma}\in\Gamma$, the maximization of $W_{\boldsymbol{\delta},\boldsymbol{\gamma}}$ with respect to $\boldsymbol{\delta}$ can be solved by establishing the partial ordering of $W_{\boldsymbol{\delta},\boldsymbol{\gamma}}$ with respect to $\boldsymbol{\delta}$. Therefore, for policy, it would also be useful to inspect the partial ordering of $W_{\boldsymbol{\delta},\boldsymbol{\gamma}}$ for any given $\boldsymbol{\gamma}$. This analysis can be done by constructing the DAG for $W_{\boldsymbol{\delta},\boldsymbol{\gamma}}$ using the lower bound $L(\boldsymbol{\delta},\boldsymbol{\delta}';\boldsymbol{\gamma})$ on the welfare gap $W_{\boldsymbol{\delta},\boldsymbol{\gamma}} - W_{\boldsymbol{\delta}',\boldsymbol{\gamma}}$.

We first consider the calculation of $L(\boldsymbol{\delta},\boldsymbol{\delta}';\boldsymbol{\gamma})$ for given $\boldsymbol{\gamma}$, which can be done by solving a sequence of LPs. The challenge is that the continuous outcome variables generate infinite-dimensional programs, which are infeasible to solve in practice. We overcome this challenge by means of approximation. Let $\tilde{\boldsymbol{Y}}_t \equiv \{Y_t(\boldsymbol{d}^t)\}_{\boldsymbol{d}^t} \in [y_l,y_u]^{2^t}$ and $\tilde{\boldsymbol{D}}_t \equiv \{D_t(\boldsymbol{z}^t)\}_{\boldsymbol{z}^t} \in \{0,1\}^{2^t}$ be vectors that constitute $\tilde{S}_t \equiv (\tilde{\boldsymbol{Y}}_t,\tilde{\boldsymbol{D}}_t)$, which is defined analogous to that in the text, and let $y_t(\boldsymbol{d}^t)$ and $d_t(\boldsymbol{z}^t)$ be the realized mappings of $Y_t(\boldsymbol{d}^t)$ and $D_t(\boldsymbol{z}^t)$. Also, define $\tilde{\boldsymbol{Y}} \equiv (\tilde{\boldsymbol{Y}}_1,...,\tilde{\boldsymbol{Y}}_T)$ and $\tilde{\boldsymbol{D}} \equiv (\tilde{\boldsymbol{D}}_1,...,\tilde{\boldsymbol{D}}_T)$. The key element in the formulation is the following conditional cumulative distribution function:

$$\begin{aligned} q(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{d}}) &\equiv \Pr[\tilde{\boldsymbol{Y}} \leq \tilde{\boldsymbol{y}} | \tilde{\boldsymbol{D}} = \tilde{\boldsymbol{d}}] \\ &\equiv \Pr[Y_t(\boldsymbol{d}^t) \leq y_t(\boldsymbol{d}^t)\ \forall \boldsymbol{d}^t \text{ and } t | D_t(\boldsymbol{z}^t) = d_t(\boldsymbol{z}^t)\ \forall \boldsymbol{z}^t \text{ and } t], \end{aligned}$$

where "$\leq$" between vectors is understood as element-wise inequalities. The object $q(\cdot)$ is the decision variable in the optimization. Let $\mathcal{Q}$ be the space of all $q(\cdot)$'s.

To construct the constraints of the program, consider the distribution of the data:

$$
\begin{aligned}
&\Pr[\boldsymbol{Y} \leq \boldsymbol{y}, \boldsymbol{D} = \boldsymbol{d} | \boldsymbol{Z} = \boldsymbol{z}] \\
&= \Pr[Y_t(\boldsymbol{d}^t) \leq y_t, D_t(\boldsymbol{z}^t) = d_t \quad \forall t] \\
&= \Pr[D_t(\boldsymbol{z}^t) = d_t \quad \forall t] \Pr[Y_t(\boldsymbol{d}^t) \leq y_t \quad \forall t | D_t(\boldsymbol{z}^t) = d_t \quad \forall t] \\
&= \sum_{\tilde{\boldsymbol{d}}: d_t(\boldsymbol{z}^t) = d_t \forall t} \Pr[\tilde{\boldsymbol{D}} = \tilde{\boldsymbol{d}}] \int_{y_l}^{y_u} \cdots \int_{y_l}^{y_u} \left\{ \int_{y_l}^{y_1} \cdots \int_{y_l}^{y_T} q(\tilde{\boldsymbol{y}}, \tilde{\boldsymbol{d}}) dy_1(d_1) \cdots d\boldsymbol{y}_T(\boldsymbol{d}^T) \right\} d\tilde{\boldsymbol{y}}_1^- \cdots d\tilde{\boldsymbol{y}}_T^- \\
&\equiv T_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}} \circ q,
\end{aligned}
$$

where $\tilde{\boldsymbol{y}}_t^-$ is $\tilde{\boldsymbol{y}}_t$ without $y_t(\boldsymbol{d}^t)$ (with some ambiguity of notation), $\int_{y_l}^{y_u} (\cdot) d\tilde{\boldsymbol{y}}_t^-$ is the corresponding multivariate intergral, and $T_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}} : \mathcal{Q} \to \mathbb{R}$ is the operator of $q(\cdot)$. Then, the continuum of constraints can be written as

$$
(T \circ q)(\boldsymbol{y}) = p(\boldsymbol{y}) \qquad \forall \boldsymbol{y} \in [y_l, y_u]^T,
$$

where T is a vector of operators $T_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}}$'s across $(\boldsymbol{d}, \boldsymbol{z})$ and $p(\boldsymbol{y})$ is a d_p-vector of $\Pr[\boldsymbol{Y} \leq \boldsymbol{y}, \boldsymbol{D} = \boldsymbol{d} | \boldsymbol{Z} = \boldsymbol{z}]$'s across $(\boldsymbol{d}, \boldsymbol{z})$. Fix $\boldsymbol{\gamma} \in \Gamma$. Since the welfare is also an integral of $q(\cdot)$, we can write

$$
W_{\boldsymbol{\delta},\boldsymbol{\gamma}} = T_{\boldsymbol{\delta},\boldsymbol{\gamma}} \circ q
$$

for an appropriate operator $T_{\boldsymbol{\delta},\boldsymbol{\gamma}} : \mathcal{Q} \to \mathbb{R}$. Consequently, for $\boldsymbol{\delta}, \boldsymbol{\delta}' \in \mathcal{D}$, we have the following program:

$$
\begin{array}{r} U(\boldsymbol{\delta}, \boldsymbol{\delta}'; \boldsymbol{\gamma}) = \max_{q \in \mathcal{Q}} (T_{\boldsymbol{\delta},\boldsymbol{\gamma}} - T_{\boldsymbol{\delta}',\boldsymbol{\gamma}}) \circ q, \\ L(\boldsymbol{\delta}, \boldsymbol{\delta}'; \boldsymbol{\gamma}) = \min_{q \in \mathcal{Q}} (T_{\boldsymbol{\delta},\boldsymbol{\gamma}} - T_{\boldsymbol{\delta}',\boldsymbol{\gamma}}) \circ q, \end{array} \quad s.t. \quad (T \circ q)(\boldsymbol{y}) = p(\boldsymbol{y}) \quad \forall \boldsymbol{y} \in [y_l, y_u]^T. \tag{A.4}
$$

Note that $q(\cdot) \in \mathcal{Q}$ is an infinite-dimensional object (unlike q in the case of discrete Y_t) and the constraints are also infinite dimensional, which makes the program (A.4) infinite-dimensional. To gain feasibility, we transform this infinite-dimensional program into a (finite-dimensional) linear program as follows. First, we approximate $q(\cdot)$ using the method of sieve. In particular, the Bernstein polynomial is a suitable choice for seive basis, because equality and inequality constraints on $q(\cdot)$ can be easily imposed as equality and inequality constraints on the coeffecients of the basis functions. Second, we combine the continuum of constraints using the following result: for any measurable function $h : [y_l, y_u]^T \to \mathbb{R}^{d_p}$, $E\|h(\boldsymbol{Y})\| = 0$ if and only if $h(\boldsymbol{y}) = 0$ almost everywhere in $[y_l, y_u]^T$. Therefore, the contraints can be replaced with $E\|(T \circ q)(\boldsymbol{Y}) - p(\boldsymbol{Y})\| = 0$. We refer the reader to Section 7 of Han and Yang (2020) for details of the Bernstein approximation and the transformation of constraints. Finally, an analogous approach can be used to calculate $L(\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\delta}', \boldsymbol{\gamma}')$ for each pair of $(\boldsymbol{\delta}, \boldsymbol{\gamma})$ and $(\boldsymbol{\delta}', \boldsymbol{\gamma}')$ in $\mathcal{D} \times \Gamma$. In practice, we can use grid $\bar{\Gamma} \subseteq \Gamma$ for Γ to characterize the identified set (A.3).

A.2 Stochastic Regimes

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

$$\rho_t(\boldsymbol{y}^{t-1}, \boldsymbol{r}^{t-1}) = r_t \in [0,1]. \tag{A.5}$$

Then, the vector of these ρ_t's is a *dynamic stochastic regime* $\boldsymbol{\rho}(\cdot) \equiv \boldsymbol{\rho}^T(\cdot) \in \mathcal{D}_{stoch}$ where $\mathcal{D}_{stoch}$ is the set of all possible stochastic regimes.[23] A deterministic regime is a special case where $\rho_t(\cdot)$ 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 $Y_T(\boldsymbol{\rho}(\cdot))$ with $\boldsymbol{\rho}(\cdot) \in \mathcal{D}_{stoch}$ as the counterfactual outcome $Y_T(\boldsymbol{\delta}(\cdot))$ where the deterministic rule $\delta_t(\cdot) = 1$ is randomly assigned with probability $\rho_t(\cdot)$ and $\delta_t(\cdot) = 0$ otherwise for all $t \leq T$. Finally, define

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

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

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

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

By the law of iterative expectation, we have

$$\mathbb{E}[Y_T(\boldsymbol{\rho}(\cdot))] = \mathbb{E}\left[\mathbb{E}\left[\cdots \mathbb{E}\left[\mathbb{E}[Y_T(\boldsymbol{r})|\boldsymbol{Y}^{T-1}(\boldsymbol{r}^{T-1})]\middle| \boldsymbol{Y}^{T-2}(\boldsymbol{r}^{T-2})\right] \cdots \middle| Y_1(r_1)\right]\right], \tag{A.6}$$

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

where the bridge variables $\boldsymbol{r} = (r_1, ..., r_T)$ satisfy

$$
\begin{aligned}
r_1 &= \rho_1, \\
r_2 &= \rho_2(Y_1(\rho_1), \rho_1), \\
r_3 &= \rho_3(\boldsymbol{Y}^2(\boldsymbol{\rho}^2), \boldsymbol{\rho}^2), \\
&\vdots \\
r_T &= \rho_T(\boldsymbol{Y}^{T-1}(\boldsymbol{\rho}^{T-1}), \boldsymbol{\rho}^{T-1}).
\end{aligned}
$$

Given (A.6), we prove the theorem by showing that the solution $\boldsymbol{\rho}^*(\cdot)$ can be justified by backward induction in a finite-horizon dynamic programming. To illustrate this with deterministic regimes when $T = 2$, we have

$$\delta_2^*(y_1, d_1) = \arg\max_{d_2} E[Y_2(\boldsymbol{d})|Y_1(d_1) = y_1], \tag{A.7}$$

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

$$\delta_1^* = \arg\max_{d_1} E[V_2(Y_1(d_1), d_1)]. \tag{A.8}$$

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

Proof. First, given (A.6), the optimal stochastic rule in the final period can be defined as

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

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

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

and $V_t(\boldsymbol{y}^{t-1}, \boldsymbol{r}^{t-1}) \equiv \max_{r_t} \mathbb{E}[V_{t+1}(\boldsymbol{Y}^t(\boldsymbol{r}^t), \boldsymbol{r}^t)|\boldsymbol{Y}^{t-1}(\boldsymbol{r}^{t-1}) = \boldsymbol{y}^{t-1}]$. Then, $\boldsymbol{\rho}^*(\cdot) = (\rho_1^*, ..., \rho_T^*(\cdot))$. 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 $\tilde{W}_t(\boldsymbol{r}^t, \boldsymbol{y}^{t-1})$ represent the objective function at t for $2 \leq t \leq T$ with $\tilde{W}_1(r_1)$ for $t = 1$. By the definition of the stochastic regime, it satisfies that

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

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

A.3 Matrices in Section 4.3

We show how to construct matrices A_k and B in (4.3) and (4.5) for the linear programming (4.7). The construction of A_k and B uses the fact that any linear functional of $\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}]$ or $\Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}, \boldsymbol{D}(\boldsymbol{z}) = \boldsymbol{d}]$ 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

$$\begin{aligned}p_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}} &= \Pr[\boldsymbol{Y}(\boldsymbol{d}) = \boldsymbol{y}, \boldsymbol{D}(\boldsymbol{z}) = \boldsymbol{d}] \\ &= \Pr[S : Y_t(\boldsymbol{d}^t) = y_t, D_t(\boldsymbol{z}^t) = d_t \quad \forall t] \\ &= \sum_{s \in \mathcal{S}_{\boldsymbol{y},\boldsymbol{d}|\boldsymbol{z}}} q_s,\end{aligned} \tag{A.9}$$

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

$$\begin{aligned}
&\Pr[\boldsymbol{Y}(\boldsymbol{\delta}(\cdot))=\boldsymbol{y}]\\
=&\Pr[Y_T(\boldsymbol{d})=y_T|\boldsymbol{Y}^{T-1}(\boldsymbol{d}^{T-1})=\boldsymbol{y}^{T-1}]\\
&\times\Pr[Y_{T-1}(\boldsymbol{d}^{T-1})=y_{T-1}|\boldsymbol{Y}^{T-2}(\boldsymbol{d}^{T-2})=\boldsymbol{y}^{T-2}]\times\cdots\times\Pr[Y_1(d_1)=y_1],
\end{aligned}\tag{A.10}$$

where, because of the appropriate conditioning in (A.10), 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.10) 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

$$\begin{aligned}
E[Y_T(\boldsymbol{\delta}(\cdot))]&=\sum_{y_1}\Pr[Y_2(\delta_1,\delta_2(Y_1(\delta_1),\delta_1))=1|Y_1(\delta_1)=y_1]\Pr[Y_1(\delta_1)=y_1]\\
&=\sum_{y_1}\Pr[Y_2(\delta_1,\delta_2(y_1,\delta_1))=1,Y_1(\delta_1)=y_1].
\end{aligned}\tag{A.11}$$

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 maps $\Pr[\boldsymbol{Y}(\boldsymbol{d})=\boldsymbol{y}]$ into $\Pr[\boldsymbol{Y}(\boldsymbol{\delta}_k(\cdot))=\boldsymbol{y}]$ according to (A.10). But note that $\Pr[\boldsymbol{Y}(\boldsymbol{d})=\boldsymbol{y}]=\sum_{s\in\mathcal{S}_{\boldsymbol{y},\boldsymbol{d}}}q_s$ by

$$\begin{aligned}
&\Pr[\boldsymbol{Y}(\boldsymbol{d})=\boldsymbol{y}]\\
&=\Pr[S:Y_t(\boldsymbol{d}^t)=y_t\quad\forall t]\\
&=\sum_{s\in\mathcal{S}_{\boldsymbol{y},\boldsymbol{d}}}q_s,
\end{aligned}\tag{A.12}$$

where $\mathcal{S}_{\boldsymbol{y},\boldsymbol{d}} \equiv \{S = \beta(\tilde{\boldsymbol{S}}) : Y_t(\boldsymbol{d}^t) = y_t \ \forall t\}$. 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 (4.4) 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.4 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 (4.7). 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 (4.6) 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.5 Alternative Characterization of the Identified Set

Given the DAG, the identified set of $\boldsymbol{\delta}^*(\cdot)$ can also be obtained as the collection of initial vertices of all the directed paths of the DAG. For a DAG $G(\mathcal{K}, \mathcal{E})$, 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 $\tilde{k}_{j,1}$.[24] 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 (4.6) satisfies*

$$\mathcal{D}_p^* = \{\boldsymbol{\delta}_{\tilde{k}_{j,1}}(\cdot) \in \mathcal{D} : 1 \leq j \leq J\}, \tag{A.13}$$

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

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, $\mathcal{D}_p^* = \tilde{\mathcal{D}}^*$. $\square$

[24] 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 $V_2 = \{2, 3, 4\}$ ($\tilde{k}_{2,1} = 4$).

A.6 Proof of Theorem 5.1

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.7 Proof of Lemma 6.1

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 (6.3) 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 (6.3) 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:

$$Y_t(\boldsymbol{d}^t), D_t(\boldsymbol{z}^t) \perp \boldsymbol{Z}^t | (\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^{t-1}). \tag{A.14}$$

Conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1}, \boldsymbol{Z}^{t-1})$, (6.3) and (A.14) correspond to Assumption S-1 in Vytlacil (2002). Assumption R(i) and (A.14) 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.8 Proof of Lemma 6.2

We are remained to prove that, conditional on $(\boldsymbol{Y}^{t-1}, \boldsymbol{D}^{t-1})$, (6.4) is equivalent to the second part of Assumption M2. But this proof is analogous to the proof of Lemma 6.1 by replacing the roles of D_t and Z_t with those of Y_t and D_t, respectively. Therefore, we have the desired result. □

References

Abadie, A., J. Angrist, and G. Imbens (2002): "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings," *Econometrica*, 70, 91–117. 9, 9

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, D. W. and G. Soares (2010): "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78, 119–157. 10

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

Ashenfelter, O. and D. Card (2010): *Handbook of labor economics*, Elsevier. 9

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

——— (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

Avis, D. and K. Fukuda (1992): "A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra," *Discrete & Computational Geometry*, 8, 295–313. 10

Backlund, E., P. D. Sorlie, and N. J. Johnson (1996): "The shape of the relationship between income and mortality in the United States: evidence from the National Longitudinal Mortality Study," *Annals of epidemiology*, 6, 12–20. 9

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

Bloom, H. S., L. L. Orr, S. H. Bell, G. Cave, F. Doolittle, W. Lin, and J. M. Bos (1997): "The benefits and costs of JTPA Title II-A programs: Key findings from the National Job Training Partnership Act study," *Journal of human resources*, 549–576. 9

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

Case, A., D. Lubotsky, and C. Paxson (2002): "Economic status and health in childhood: The origins of the gradient," *American Economic Review*, 92, 1308–1334. 9

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

Chen, L.-Y. and J. Szroeter (2014): "Testing multiple inequality hypotheses: A smoothed indicator approach," *Journal of Econometrics*, 178, 678–693. 10

Conduct Problems Prevention Research Group (1992): "A developmental and clinical model for the prevention of conduct disorder: The FAST Track Program," *Development and Psychopathology*, 4, 509–527. 3, 7

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

de Chaisemartin, C. and X. d'Haultfoeuille (2020): "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 110, 2964–96. 2

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. (2020): "Comment: Individualized Treatment Rules Under Endogeneity," *Journal of the American Statistical Association*. 1

——— (forthcoming): "Nonparametric Identification in Models for Dynamic Treatment Effects," *Journal of Econometrics*. 1, 2.1

Han, S. and S. Yang (2020): "Sharp Bounds on Treatment Effects for Policy Evaluation," *arXiv preprint arXiv:2009.13861*. A.1

Hansen, P. R. (2005): "A test for superior predictive ability," *Journal of Business & Economic Statistics*, 23, 365–380. 10

Hansen, P. R., A. Lunde, and J. M. Nason (2011): "The model confidence set," *Econometrica*, 79, 453–497. 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, 6

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. 5.2

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, 9, 10

KITAMURA, Y. AND J. STOYE (2019): "Nonparametric Counterfactuals in Random Utility Models," *arXiv preprint arXiv:1902.08350*. 1

KOCK, A. B., D. PREINERSTORFER, AND B. VELIYEV (2021): "Functional sequential treatment allocation," *Journal of the American Statistical Association*, 1–13. 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. 6

——— (2004): "Statistical treatment rules for heterogeneous populations," *Econometrica*, 72, 1221–1246. 1, 23

——— (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. 6

McDonough, P., G. J. Duncan, D. Williams, and J. House (1997): "Income dynamics and adult mortality in the United States, 1972 through 1989." *American journal of public health*, 87, 1476–1483. 9

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, A.1, 23, A.2

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, 7, 23

Neal, D. (1997): "The effects of Catholic secondary schooling on educational achievement," *Journal of Labor Economics*, 15, 98–123. 9

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

Shaikh, A. M. and E. J. Vytlacil (2011): "Partial identification in triangular systems of equations with binary dependent variables," *Econometrica*, 79, 949–955. 20

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, 6

Vytlacil, E. (2002): "Independence, monotonicity, and latent index models: An equivalence result," *Econometrica*, 70, 331–341. 6, A.7

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