Dynamic Treatment Regimes with Interference

by

Cong Jiang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2022

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:
Michael R. Kosorok
W.R. Kenan, Jr. Distinguished Professor, Department of Biostatistics,
Professor, Department of Statistics and Operations Research
University of North Carolina at Chapel Hill

Supervisor(s):
Michael P. Wallace
Associate Professor, Department of Statistics & Actuarial Science
University of Waterloo

Mary E. Thompson
Professor Emerita, Department of Statistics & Actuarial Science
University of Waterloo

Internal Member(s):
Yeying Zhu
Associate Professor, Department of Statistics & Actuarial Science
University of Waterloo

Paul Marriott
Professor, Department of Statistics & Actuarial Science
University of Waterloo

Internal-External Member:
Zahid Butt
Assistant Professor, School of Public Health Sciences
University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Precision medicine describes healthcare in which patient-level data are used to inform treatment decisions. Within this framework, dynamic treatment regimes (DTRs) are sequences of decision rules that take individual patient information as input, and then output treatment recommendations. The primary purpose of DTR research is to estimate the optimal dynamic treatment regimes: the sequence of treatment rules that will optimize some pre-defined outcomes across a population. The focus of this thesis is on developing methods for estimating optimal DTRs in the presence of interference, where one patient's outcome can be affected by others' treatment. DTR estimation methods typically rely on the assumption of no interference. In many social network contexts, such as friendship or family networks, and for many health concerns, such as infectious diseases, this assumption is questionable. Moreover, the existing doubly robust regression-based DTR estimation methods are primarily focused on continuous outcomes. DTR estimation methods for binary or ordinal outcomes are more complicated due to less information being provided by these discrete outcomes. Consequently, very few DTR estimation methods focus on binary or ordinal outcomes, let alone methods when interference is present. To address these problems, for continuous outcomes, we directly establish novel interference-aware DTR estimation methods, and for binary or ordinal outcomes, we develop methods for DTR estimation first in cases without interference and then in ones affected by it.

This thesis contains three main components: (1) a doubly robust method to estimate the optimal DTRs for individuals where the treatments of their connected neighbours in the same social network are taken into account in the decision rules; (2) a doubly robust method to estimate the optimal DTRs for binary outcomes using sequential weighted generalized linear models; (3) a doubly robust method to estimate the optimal DTRs for ordinal outcomes in the presence of household interference. In (1), we study the DTR estimation method of dynamic weighted ordinary least squares (dWOLS), which boasts easy implementation and double robustness, but relies on the no interference assumption. We define a network propensity function and build on it to establish an implementation of dWOLS that remains doubly robust under interference associated with network links. The method's properties are shown via simulation and applied to household pairs data from the Population Assessment of Tobacco and Health (PATH) Study. On the basis of the theories of dWOLS and using our interference-aware version, we focus on developing innovative DTR estimation methods for both binary and ordinal outcomes, in particular, the methods in the presence of interference. In (2), considering binary outcomes, we propose a new method for DTR estimation without interference, the dynamic weighted generalized linear model (dWGLM), which accommodates binary outcomes while offering relatively straightforward implementation and robustness to model misspecification. We introduce the method and its underlying theory, and illustrate both in an analysis of e-cigarette usage and smoking cessation,

using the observational data from the PATH study. Finally, in (3), we further extend these regression-based DTR methods to the ordinal outcome case, and also propose a robust method — the dynamic weighted proportional odds model (dWPOM). Moreover, in the presence of household interference, exploring the possible correlation between treatments in the same household, we investigate the covariate balancing weights, which rely on the joint propensity score, and methods for estimating the joint propensity score. Examining different types of balancing weights, we verify the double robustness of dWPOM with our adjusted weights via simulation studies. Lastly, we also illustrate dWPOM in the analysis of data from PATH. For each participant's household, we derive the household treatment configuration recommendations for achieving the best outcome of the pair: both individuals quit or attempt to quit smoking.

Acknowledgements

At this point, I would like to express my sincere gratitude for all the help and encouragement I have received from schoolmates, staff, instructors, and professors at UWaterloo.

Special gratitude is given to my PhD supervisors Dr. Mary Thompson and Dr. Michael Wallace. Their mentorship and support have been enormously valuable. I am fortunate because it was my supervisors who provided the choice on the topic of my PhD thesis: Dynamic Treatment Regimes with Interference, and in doing so they opened up a research area for me that I can continue to work on even after I graduate from UWaterloo.

My supervisors have been an outstanding pattern for me to follow, in both ways to conduct research and attitudes; importantly, the attitude towards challenges, difficulties, temporary setbacks and adversities, which are very common in any scientific research. My supervisors stressed my research freedom and independent thinking, and made research an enjoyable and productive experience. Their ways that lead me to conduct research are also beneficial for me in developing my statistics tastes, an ability to find an "interesting" statistical research problem and an "innovative" way to address them.

My journey at the University of Waterloo started with an email from Professor Paul Marriott who served as a member of the statistics PhD admission committee in 2017. He helped connect me with my future supervisor Dr. Mary Thompson, who provided a research project that I can work on at the beginning of my PhD program. Until now, I have always appreciated that I acquired the research opportunity that Mary provided; otherwise, I would not be able to be admitted to the statistics PhD program at UWaterloo, and not to say to finish this thesis. Mary has been instrumental in my enjoyment of the PhD program. Thanks to the weekly meeting with her, I got help with any difficulties that arose in my research. She always encouraged me when I felt disappointed about a failure in my research projects. Her research enthusiasm and attitude have always inspired me to be tenacious in addressing problems, simplifying difficult problems under certain conditions, and trying different approaches until I found one that worked. Her contributions to survey sampling and the tobacco usage area over the years have inspired me to continue to focus on my areas of interest long term.

I am also thankful to my other supervisor, Dr. Michael Wallace. I have worked with Michael since the second year of my PhD program. He has generously shared all the resources I needed to dive into the causal inference and DTR estimation areas. The Harvard causal inference workshop, in which he supported me, is one of my favourite memories. Moreover, together, we also successfully applied for an OICR Biostatistics Training Initiative Fellowship Award, which gave me more confidence to conduct my PhD thesis research in my earlier years of the program.

The suggestions and comments he provides about writing a paper for publication, presentation slides, and posts are valuable and helpful to me. Almost every presentation, he provides the details of every slide and time arrangement suggestions. He always encourages me to practice a lot before the presentations. Fortunately, these suggestions and the extra practice gave me more confidence during the presentations, allowing me to reap two presentation awards.

I also wish to express my special appreciation to Dr. Yeying Zhu and Dr. Paul Marriott for their assistance and for serving as thesis proposal and defence committee members. Yeying was my first causal inference instructor. Furthermore, I would like to express my gratitude to Dr. Michael R. Kosorok who serves as External Examiner, and to Dr. Zahid Butt who serves as Internal-External Examiner on the committee. They provided me with ample valuable and beneficial comments and suggestions for this thesis. In addition, I am also grateful for Dr. Nathaniel Stevens, who served as the chair in my thesis proposal defence, and Dr. Krista Fiolleau who served as the chair in the final thesis defence.

Finally, I want to thank my PhD classmates, in particular, my officemates in M3 4134. My officemates and I endured countless struggles throughout many nights and weekends as often witnessed by the M3 janitors. I also want to thank my friends who are both in Waterloo and Montreal.

This thesis was conducted with the support of the Ontario Institute for Cancer Research (OICR) through funding provided by the Government of Ontario. I gratefully acknowledge the support from an OICR Biostatistics Training Initiative Studentship/Fellowship Award. I also would like to express my appreciation for the use of PATH data (PATH [2020]).

Dedication

This thesis is dedicated to my parents, my father Tishan Jiang and my mother Qingwen Ma, for their unconditional and forever love.

Table of Contents

List of Tables

List of Figures

List of Symbols

The following list provides quick access to the key notation used in the thesis. Precise definitions should be referred to in the text.

Chapter 1

Symbol	Description
A	Treatment (or intervention), $A \in \mathscr{A}$, where $\mathscr{A}$ is the set of all available treatment options
Y	Observed outcome
$Y^*(a)$	Potential outcome that would be observed if treatment was set to $A = a$
$\mathbb{E}$	Expectation
δ^*	Average treatment effect
$\boldsymbol{X}_i = (X_{1i}, X_{2i}, \ldots, X_{Pi})^\top$	A vector of P covariate variables of the individual i. The corresponding observed values are $\boldsymbol{x}_i = (x_{1i}, x_{2i}, \ldots, x_{Pi})^\top$
$\mathbb{E}_{\boldsymbol{X}}$	Expectation with respect to the marginal distribution of $\boldsymbol{X}$
n	Sample size
N	Adjacency matrices, with (i, l) entry N_{il} for $i, l \in \{1, ..., n\}$
φ	Treatment saturation: the percent of individuals in the cluster who receive treatment
μ_φ	Expected potential outcome under treatment saturation φ
$\mu_{a\varphi}$	Expected potential outcome when an individual receives treatment a under treatment saturation φ for $a = 0, 1$

Chapter 2

Symbol	Description
(d^{opt}) d	(Optimal) treatment regime, or a treatment decision rule for a singe-stage decision
$\boldsymbol{Z}$	Additional covariates in the general no unmeasured confounder condition, that is, $\{Y^*(a) : a \in \mathscr{A}\} \perp\!\!\!\perp A \mid \boldsymbol{X}, \boldsymbol{Z}$ such that $\boldsymbol{X} \cup \boldsymbol{Z}$ is a sufficient adjustment set
$\mathbb{I}$	An indicator function
$\delta^*(\boldsymbol{x})$	Conditional average treatment effect
δ_i^*	Individual treatment effect
K	Total number of treatment stages
$\overline{\boldsymbol{a}}_j = (a_1, a_2, ..., a_j)$	Past treatments until j^{th} stage, where $j = 1, 2, ..., K$
$\underline{\boldsymbol{a}}_{j+1} = (a_{j+1}, ..., a_K)$	Future treatment from stage $j+1$ onwards
$\boldsymbol{h}_j = (\overline{\boldsymbol{x}}_j, \overline{\boldsymbol{a}}_{j-1})$	Covariate matrix containing patient information (history) prior to the j^{th} treatment decision. Note that when we consider $\overline{\boldsymbol{x}}_j$ that is related to history information, we suppress patient-level notation and the subscript refers to the stage.
$\boldsymbol{h}_j^{\alpha}$, $\boldsymbol{h}_j^{\beta}$ and $\boldsymbol{h}_j^{\psi}$	Subsets of covariates contained in $\boldsymbol{h}_j$
$\boldsymbol{h}_{ji}$	History matrix for j^{th} stage and i-th record. The first letter of subscript refers to the stage, and the second letter refers to the number of the record. For example, in an i.i.d. sample of n records, the i-th record consists of $(\boldsymbol{h}_{1i}, \boldsymbol{h}_{2i}, ..., \boldsymbol{h}_{Ki}; a_{Ki}, Y_{Ki})$.
$d = \{d_1(\boldsymbol{h}_1), ..., d_K(\boldsymbol{h}_K)\}$	Dynamic treatment regime for K stage decision points. *A treatment decision rule* $d_j(\boldsymbol{h}_j)$ is a function that maps an individual's history to a treatment option in $\mathscr{A}_j$, for $j = 1, ..., K$.
$\widetilde{\mathscr{Y}}_j(\boldsymbol{h}_{j+1}, \hat{\boldsymbol{\beta}}_{j+1}, \hat{\boldsymbol{\psi}}_{j+1})$	The j^{th} stage pseudo-outcome in Q-learning
$\pi(\boldsymbol{x})$	Propensity score function
$\gamma_j(\boldsymbol{h}_j, a_j)$	Blip function at j^{th} stage
$\mu_j(\boldsymbol{h}_j, a_j)$	Regret function at j^{th} stage
$f_j(\boldsymbol{h}_j)$	Treatment-free function at j^{th} stage
$\widetilde{\mathscr{Y}}_j$	The j^{th} stage pseudo-outcome in dWOLS
$X^{\otimes 2}$	$X^{\otimes 2} = XX^\top$ for column vector X

Chapter 3

Symbol	Description
G	An undirected network
$\mathcal{N}_i$	A set of neighbours of node (individual) i in the network G
$\mathbf{x}_{\mathcal{N}_i}$	The covariate values of the neighbours of node i
$A_{\mathcal{N}_i}$	Treatments of units in the neighbourhood of unit i
$Y_i^*(A_i, A_{\mathcal{N}_i})$	Potential outcome of a unit i under the neighborhood interference assumption
$\mathcal{V}\left[d(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}, a_{\mathcal{N}})\right]$	Value function: expected outcome that is achieved applying a certain treatment regime $d(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}, a_{\mathcal{N}})$
$d_j(\boldsymbol{h}_j, A_{\mathcal{N}})$	A treatment decision rule with interference at the j^{th} decision stage
$t_i = \sum_{l=1}^{n} N_{il} a_l$	Number of treated neighbours of ego i
$s_{i,a}$	Set of treated neighbours of i
$\pi_{i,a_i,s_{i,a}}(\boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i})$	Network propensity function of individual i is the joint probability of individual i receiving treatment a_i, and the treatments of his or her neighbours being in the set $s_{i,a}$.
$t(a_{\mathcal{N}})$	Network interference term that results from a exposure interference function $t : \mathbb{R}^{\lvert\mathcal{N}\rvert} \mapsto \mathbb{R}$
$\gamma^{\xi}(t(a_{\mathcal{N}}), \boldsymbol{x}^{\xi}; \boldsymbol{\xi})$	Blip functions that are constituted by interference term $t(a_{\mathcal{N}})$ and its products with the subset of variates $(\boldsymbol{x}^{\xi})$
$\mathscr{C}(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}, \mid \mathcal{N}_i \mid)$	A function that depends only on the x-variables $(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i})$ and the degree $(\mid \mathcal{N}_i \mid)$; given these variables, $\mathscr{C}(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}, \mid \mathcal{N}_i \mid)$ is a constant.
$\widehat{\mathcal{L}}[(a, a_{\mathcal{N}})]$	An estimator for the *total regret*: the total value loss arising from receiving the observed treatment regimen $(a, a_{\mathcal{N}})$ rather than the estimated optimal treatment regimen $(\hat{a}^{opt}, a_{\mathcal{N}})$

Chapter 4

Symbol	Description
g	Link function in a generalized linear model
g^{-1}	Inverse of link function
$g^{-1\prime}(\mu)$	The first derivative of the inverse link function
$\kappa(a, \boldsymbol{x})$	An "adjustment" factor in the balancing weights criterion for binary outcomes
w^d	A choice of dWOLS balancing weights that satisfy the criterion $(1 - \pi(\boldsymbol{x}))w(0, \boldsymbol{x}) = \pi(\boldsymbol{x})w(1, \boldsymbol{x})$
$\widetilde{\mathcal{Y}_j}$	The j^{th} stage binary pseudo-outcome in dWGLM
R	Number of replicates of generating the binary pseudo-outcomes
Φ	Cumulative distribution function of the standard normal distribution
O	Error term in Taylor series expansion of g^{-1} function

Chapter 5

Symbol	Description
(s, r)	A pair/couple in the same household
$(A^s, A^r) = (a^s, a^r)$	Treatment configuration for the pair (s, r)
$\pi^{a^s a^r}(\boldsymbol{x}_s, \boldsymbol{x}_r)$	Joint propensity functions for the pair (s, r)
$w(a^s, a^r)$	Covariates balancing weights
$h = 1, ..., H$	Households where H is the sample size for households
$\boldsymbol{A}_h = (A_h^s, A_h^r)^\top$	Treatment vector for the h^{th} household
$p_{hs}(\boldsymbol{\alpha})$	Marginal propensity score for individual s in the h^{th} household
$A_{hsr} := \mathbb{I}(A_h^s = 1, A_h^r = 1)$	An indicator that the couple (s, r) both are treated
$p_{hsr} := \mathbb{P}(A_{hsr} = 1)$	Joint propensity score that the couple (s, r) both are treated
τ_{hsr}	Odds ratio for the pair of correlated binary variables (A_h^s, A_h^r)
$\boldsymbol{x}_{sr}$	Pair-level covariates
$\boldsymbol{V}_h$	Working covariance matrix of $\boldsymbol{A}_h$
$Z_h = z_h$	h^{th} household "treatment and interference" levels $z_h = 1, 2, 3, 4$
$\bar{\boldsymbol{x}}_z^s = \sum_{h=1}^{N_z} \boldsymbol{x}_{s,h} w_z / \sum_{h=1}^{N_z} w_z$	Weighted mean of the individual s's covariate $\boldsymbol{x}_s$ from household h that belongs to the zth group, where w_z denotes the zth group weights.
$\boldsymbol{S}_{\boldsymbol{X}}^2 = \sum_{z=1}^4 \boldsymbol{S}_{\boldsymbol{X},z}^2 / 4$	Pooled standard deviation from groups $z = 1, 2, 3, 4$
$sd(\boldsymbol{x}_z^s), sd(\boldsymbol{x}_z^r)$	Unweighted standard deviation of covariates $(\boldsymbol{x}_s, \boldsymbol{x}_r)$ from the zth group
$\mathcal{T}(\boldsymbol{x})$	Tilting function which is a pre-specified function of covariates
$U(Y^s,\ Y^r)$	Household utility
ω_s, ω_r	Combination weights for household utilities
$d^*(\boldsymbol{x}^s, \boldsymbol{x}^r)$	Optimal household treatment decision rule
$\mathcal{X}_s, \mathcal{X}_r$	Support of $\boldsymbol{x}_s$ and $\boldsymbol{x}_r$
$c = 1, ..., C - 1$	A specific category
ζ_c	Coefficients that are category-specific intercepts in a POM
$\boldsymbol{\theta}$	Coefficients of covariates $\boldsymbol{x}_h$ in a POM
$\gamma[(A^s, A^r), \boldsymbol{x}_h; \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}]$	Household blip function
$\widetilde{\mathcal{U}}$	Ordinal pseudo-utility
$\mathcal{R}$	Total number of replicates generating the ordinal pseudo-utility
$\mathfrak{r}$	Index of the replicate for generating the ordinal pseudo-utility

Key Acronyms Throughout the Thesis

Symbol	Description
ASD	Absolute standardized differences
ATE	Average treatment effect
ATS	Adaptive treatment strategies
ATT	Average treatment effect among the treated population
ATO	Average treatment effect among the overlap population
CATE	Conditional average treatment effect
COM	Conditional outcome modeling
DTR(s)	Dynamic Treatment Regime(s)
dWOLS	Dynamic weighted ordinary least squares
DR	Doubly robust
dWGLM	Dynamic weighted generalized linear model
DAG	Directed acyclic graphs
GEE	Generalized estimating equation
HTE	Heterogeneous treatment effects
IPTW	Inverse probability of treatment weighting
i.i.d.	Independently and identically distributed
ITR	Individualized treatment rule
MARL	multi-agent reinforcement learning
MOTR	Mean optimal treatment rate
NUC	No unmeasured confounding
NIT	Neighbours' invariant treatments assumption
OWL	Outcome weighted learning
PATH	Population Assessment of Tobacco and Health study
PSD	Population standardized difference
RCT	Randomized clinical trials
RL	Reinforcement learning
SMART	Sequential multiple assignment randomized trial
SUTVA	Stable unit treatment value assumption
SNMMs	Structural nested mean models
SE	Standard error

Chapter 1

Introduction

To optimize the quality of health care, precision medicine focuses on tailoring treatments to the specific characteristics of each patient. As a formal framework of precision medicine, *dynamic treatment regime* (DTRs) are sequences of decision rules that take individual patient data as input and then output treatment recommendations. DTR estimation from observational data typically relies on the assumption of *no interference* (Cox [1958]): that is, the outcome of one individual is unaffected by the treatment assigned to others. In many social network contexts, such as friendship or family networks, this assumption is questionable. This thesis is concerned with DTR estimation methods in the presence of interference. Several innovative interference-aware DTR estimation methods are developed, providing robust estimation and easy implementation. A range of treatment assignment mechanisms is suggested for a variety of interference scenarios. These developments are illustrated in an examination of smoking cessation in household networks in which we apply the proposed methodologies in real-world data from the Population Assessment of Tobacco and Health study.

This chapter starts with a brief review of precision medicine and DTR estimation frameworks, and introduces relevant interference studies from the causal inference literature. Then it shows our real-world dataset analyzed with our methods, summarizes the contributions of the research, and ends with a brief overview of the thesis' structure.

1.1 Precision Medicine and Dynamic Treatment Regimes

Improving health outcomes by tailoring treatments to individual patient characteristics is a centuries-old approach and is a central component of medical practice. Today's efforts at tailoring

treatment grew from several past research breakthroughs. After Sir Austin Bradford Hill's 1946 success demonstrating the efficacy of streptomycin for treating tuberculosis, in what was the first randomized controlled clinical trial, the use of statistical inference as a scientific method to study medical treatment began to increase dramatically (Stusser [2010] as cited in the review paper by Kosorok and Laber [2019]). However, until both clinicians and biostatisticians became aware of patient heterogeneity, the main goal of clinical trials was to assess treatments for specific diseases. Heterogeneity between patients refers to the variation between them that is attributable to their unique characteristics (Chakraborty and Moodie [2013], Grutters et al. [2013], Kosorok and Laber [2019]). The patient characteristics that potentially explain patient heterogeneity include genetic profiles, physiological or demographic characteristics, and environment or lifestyle factors (Tsiatis [2019]). Heterogeneity in the response to treatment has been observed in a number of areas, and research provides strong evidence that the patient's characteristics should be considered in creating a specific patient's treatment regime. For example, treatment response has been observed to be heterogeneous among adults suffering from major depression (Trivedi et al. [2006]), children with attention-deficit or hyperactivity disorder, and autism spectrum disorders (Pelham Jr and Fabiano [2008], Jones et al. [2010]). Additionally, in the analysis of data on smokers in North America, the observed "patient heterogeneity" is in degree of addiction to nicotine (Gravely et al. [2020]), age group (Hyland et al. [2006]), and intention to quit smoking (Kasza et al. [2021]). In recent years, precision medicine, which refers to the principle of tailoring treatments to individual patients' characteristics in order to optimize their personal care, has grown in popularity. Specifically, precision medicine became a national priority in the U.S.A. when President Obama announced the Precision Medicine Initiative in 2015. Conceptually, precision medicine is identical to personalized medicine, as well as stratified medicine, in that it is based on identifying patients or subgroups of patients with distinct mechanisms of disease, or particular responses to treatments (Kosorok and Moodie [2015], Lonergan et al. [2017]).

One feature of precision medicine, the dynamic treatment regimes, formalizes evidence-based clinical decision-making. This formal framework is sometimes referred to as an adaptive treatment strategy, adaptive intervention, or treatment policy in different contributions to the literature. In the context of multi-stage treatment decisions, a DTR is a sequence of decision rules by which, at each stage, patient information (such as age and health status) is taken as input and used to create a treatment recommendation as output. The decision rule is primarily based on patient-specific characteristics, so DTRs are also called *individualized treatment rules* (ITRs) (Lavori and Dawson [2004]). By administering ITRs based on the characteristics of the patient, treatment strategies can be tailored to improve the patient's response. Therefore, the primary interest in this area of research is to use data to identify *optimal DTRs or ITRs*: sequences of treatment rules that yield the best expected clinical outcomes. The term "dynamic" is used to refer to multiple or single-stage rules that incorporate patient's previous history information, and is opposite to "non-dynamic"

or "static", which refer to any rules, single- or multi-stage, that do not vary with pre-treatment observations. Therefore, DTRs are also known as *adaptive treatment strategies* (ATS) (Lavori and Dawson [2000], Murphy [2005]). In many cases, static regimes are impractical because any reasonable treatment rules should take account of the patient's response status; therefore, the static regimes have less relevance to precision medicine. In addition, the word "regime" refers to any set of dynamic or static treatment rules, which are functions that map from patients' covariates to a treatment that is from a set of feasible options.

There are many real-world medical applications of DTRs. In cancer prevention programs, such as programs for smoking reduction, cessation aids, including gum or patches, e-cigarettes, and support services, can be viewed as elements of sequential treatments. Decisions to switch to different therapies can be made dependent on patient-level characteristics. From a personalized medicine perspective, Yan and Chakraborty [2018] suggested that the Singapore government consider delivering 'personalised' smoking-reduction strategies through digital means (such as mobile-based applications). Moreover, Strecher et al. [2008] conducted a two-stage randomized study of smoking cessation, and Chakraborty et al. [2010] analysed these randomized study data to identify the optimal DTR for smoking cessation. In the context of smoking cessation, an example of a DTR could be:

Decision:

Use e-cigarettes if the smoker has a plan to quit; otherwise do not use e-cigarettes.

Note that this simple DTR provides only a single treatment rule dependent on one piece of information about patients whereas, in general, a DTR provides multi-stage treatment rules based on a series of pieces of patient-level information.

A DTR can also be applied to oncology studies, for example, on acute leukaemia (Davidian and Laber [2019]). A patient's treatment often involves multiple courses of chemotherapy, and a physician could use a DTR that maximizes the patient's disease-free or overall survival time. In a setting of two decision points, for instance, the first decision is to choose induction chemotherapy for inducing a positive response, such as a partial or complete remission. Then, the second decision should be made to prescribe maintenance or salvage treatments based on the patient's response to the first decision point treatments, symptom severity, and so on. In the context of leukaemia, Wahed and Tsiatis [2004] proposed optimal estimators for the survival distribution and related quantities, for treatment policies in a two-stage clinical decision problem. As an artificial illustration, suppose that there exist only two available induction chemotherapy options (denoted as C1 and C2), two maintenance treatments for patients who respond (denoted as M1 and M2), and two salvage chemotherapies for those who do not respond (denoted as S1 and S2). Then, an example of DTRs for this acute leukaemia case could be:

Decision 1:

If the patient's age is 45+ and his or her white blood cell (WBC) count is $< 9.8 \times 10^3/\mu l$*, give the*

induction chemotherapy C1; otherwise, give C2.
Decision 2:
If the patient responds and current WBCs are at $< 11.5 \times 10^3/\mu l$*, give the maintenance treatment M1; otherwise, give M2.*
If the patient does not respond and current WBCs are at $< 11.2 \times 10^3/\mu l$*, give the salvage chemotherapy S1; otherwise, give S2.*

Before moving on to the goal and estimation of DTRs, we emphasize four crucial points that should be carefully considered for conducting DTRs research in practice. 1) The decision points or disease stage must be determined, leaving no ambiguity, thereby supporting optimal treatment decisions. The decision points may be consistent with milestones in the disease progression, diagnosis, or an evaluation of responses (e.g., in acute leukemia), or with the occurrence of events that require a treatment decision (e.g., myocardial infarction in cardiovascular disease), or they may coincide with planned clinic visits at which decisions are made according to a schedule (e.g., treating HIV infections). Therefore, the timing of decisions may be fixed or random, and different individuals can even experience different numbers of decisions. 2) The individual dynamic characteristics for treatment decision rules must be decided upon, thereby clarifying which features or covariates should be utilized to make a treatment decision. In particular, for multiple-stage decision problems, the decision rules for each stage may differ to enable the appropriate choice of covariates. Typically, these covariates on which the treatment decision rules rely are called *tailoring variables* or *prescriptive variables* (Gunter et al. [2007], see Carini et al. [2014] for definitions and details of prognostic and predictive biomarkers). 3) Feasible treatment options must be decided upon for each stage, requiring that an attainable treatment space or domain be determined as well. In practice, not all treatments are applicable to all patients at all times; for example, if patients respond well to a specific treatment in an earlier stage of the disease, this treatment may be repeated in later stages. 4) Treatment decisions that involve the chosen types of treatment rules must be specified. There are many types of treatment rules, such as those governed by deterministic or stochastic policies. Moreover, treatment rules may also differ between complicated ones such as nonparametric machine (deep) learning-based decision functions and simpler ones such as linear functions.

After reviewing the above four key points for conducting DTR research, we will introduce certain common settings in DTR estimation frameworks, and then review some standard DTR estimation methods. Finally, we end this section with a brief review of data sources for DTR estimation.

Our goal is to estimate the optimal DTR, the treatment rule, or sequence of treatment rules, which, conditional on patient information, will optimize the pre-defined outcome or outcomes. In this process, identifying the treatment or sequence of treatments corresponds to *single-decision* or *multi-stage* problem settings. In terms of an approach to estimating optimal DTRs, on the one hand,

the goal for single-decision settings is to estimate an individualized treatment rule that maximizes the expectation of the potential outcome under that rule. Under the general DTR estimation framework, both regression-based estimators, which are obtained from a regression representation of the optimal treatment regimes, and classification-based (or direct-search) estimators, which are built on importance sampling and directly identified by maximizing the regimes' value function, are widely employed estimators. *Value function* indicates the expected outcome (value) that results from following a certain treatment regimen, given the patients' covariate information. On the other hand, the multi-stage setting is more complex, because of prognostic effects and delayed treatment effects (Thall et al. [2007], Kosorok and Moodie [2015]); for example, a patient's current treatment may not only affect intermediate outcomes but also future ones. Thus, treatment decisions should be made by considering both the immediate (proximal) and delayed (distal) efficacy of a specific treatment. Multi-stage settings are classified as having either a finite time horizon or an infinite time horizon, depending on whether the number of decision points is finite (small) or indeterminate (infinite or large). Infinite time settings, in which sequential decisions need to be made over a long time, are necessary for some scenarios, for instance, in the mobile-health treatment of diabetes or other chronic diseases. Still, many realistic treatment-decision problems are typically restricted to finite time settings, such as when treating HIV infections or acute leukaemia. Different settings involve distinct analytical frameworks and methods. A Markov decision process is typically used to model decision-making over infinite time horizons; however, backwards induction is commonly used in numerous methods that are developed for making sequential treatment decisions in the finite time horizon. This thesis focuses on finite time horizon settings; certain standard DTR estimation methods in such settings are introduced next.

In finite time horizon settings, various techniques of reinforcement learning are used to estimate optimal dynamic treatment regimes, either through "indirect" methods (that have an intermediate step before the optimal decision rule is estimated), such as regression and policy-learning methods, or direct methods (focusing directly on the regimens themselves), such as value-search methods. For example, the fundamental regression-based approaches are Q-learning, A-learning and G-estimation, where G-estimation is a special A-learning (advantage learning) method developed by Robins (Robins [2004]) to estimate the parameters of interest (Schulte et al., 2014). Both Q-learning and A-learning are backwards step-wise approaches usually employed in multi-stage treatment-decision problems. A-learning, which employs a doubly robust estimating equation to estimate the contrast function, provides a more-robust way to estimate DTR than Q-learning. The doubly robust procedures provide consistent estimation of the parameters of interest even if one nuisance model is misspecified. Note that we will provide more details of these regression-based methods and rigorously define the term "doubly robust" in Chapter 2. Also used are value-search methods containing an (augmented) inverse probability of treatment weighting (IPTW, Zhang et al. [2012]) and outcome weighted learning (OWL, Zhao et al. [2012]).

Directly maximizing the value function over a restricted class of regimes, Zhang et al. [2013] estimate optimal DTRs by augmented inverse probability weighting. OWL recasts the IPTW estimator into a classification framework, and essentially changes the treatment selection problem into a two-class classification problem by considering a general class of surrogate loss functions. In particular, Zhao et al. [2015] consider the hinge loss and propose a weighted support vector machine (SVM) to estimate the decision rules. Usually, these value-search methods also face some technical challenges and obstacles, such as computational issues and products of indicators across time (Kosorok and Laber [2019]). The above-mentioned methods are fundamental approaches for estimating optimal DTRs, and they are restricted to certain basic settings, such as continuous outcomes wherein larger values are preferred, and binary treatments, which are referred to as *treatments* or *controls*. Note that this thesis primarily focuses on the regression-based indirect methods, and also starts from these basic set-ups or contexts: binary treatments and continuous outcomes. Beyond these basic set-up regression-based methods, various extensions have been proposed for lessening their drawbacks or for fitting these approaches into more-complex contexts. For example, Dynamic weighted ordinary least squares (dWOLS, Wallace and Moodie [2015]) is both an easy-to-use (like Q-learning) and doubly robust (like G-estimation) approach for estimating optimal DTRs. Building on dWOLS, many extensions are also being studied, such as multiple-value and continuous treatments in Schulz and Moodie [2021], survival outcomes in Simoneau et al. [2020] , measurement error in Spicker and Wallace [2020], and so on. Additionally, Moodie et al. [2014] first extended Q-learning to discrete utilities (i.e., both Bernoulli and Poisson utilities), and employed a generalized additive model to implement more-flexible modeling. Later, for binary outcome studies, considering cancer and graft-versus-host disease treatment cases, to maximize the binary outcome of two-year disease-free survival, Moodie and Krakow [2020] implemented Q-learning in a multistage treatment decision analysis, employing logistic regression in each stage.

Datasets for the estimation and inference of optimal dynamic treatment regimes can come from a rich range of sources, for example, experimental data from randomized clinical trials (RCTs), and non-experimental data from electronic health records, cohort studies, censuses, or surveys.

Randomized clinical trials are a gold standard for data collection because they resist unmeasured confounders and can support efficient estimation. A well-known gold standard study design for estimation or evaluation of multi-stage decision regimes is a sequential multiple assignment randomized trial (SMART) design (Lavori and Dawson [2004], Murphy [2005]). In a SMART, a patient is randomized at key decision points in the treatment process where there is clinical equipoise (i.e., no one better intervention is present for either the control or experimental group), and thus, each patient may be randomized multiple times throughout the trial. Some practical SMART designs include Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial on the major depressive disorder (Rush et al. [2004]), adaptive pharmacological and behavioral

treatments for children with ADHD (Pelham Jr et al. [2008]) and clinical antipsychotic trials of intervention effectiveness for schizophrenia (Stroup et al. [2003]). Compared with an RCT, non-experimental data collection is typically less expensive, can be done relatively quickly, and will result in a larger sample size. However, non-experimental data induce some challenges, such as in identifying the frequency and timing of data collection and unmeasured confounders for every treatment stage.

For an RCT, sometimes randomization at each decision point is impossible or nonoptimal, so various hybrid designs are studied. Hybrid designs have both randomization and observational components and often entail a pragmatic motivation. In addition, when patients' information comes from both randomized trials and real-world data, there are two main approaches for integrative analysis: meta analysis and pooled patient data analysis. See the related recent work by Hatt et al. [2022], Yang et al. [2020], Cheng and Cai [2021] and Athey et al. [2020].

1.2 Interference

The move towards personalized medicine is being affected by *interference* (Hudgens and Halloran [2008], VanderWeele and Christakis [2019]), in which outcomes from one individual's treatment are affected by the choice of treatment for another individual. Before formally introducing interference and interference structures, we introduce some notation standards used in the causal inference literature. Suppose we are interested in the causal effect of a certain treatment, or intervention, A, on some outcome Y. Let $Y^*(a)$ denote a *potential outcome* that would be observed if treatment was set to $A = a$; therefore, in binary treatment settings, i.e., A is in the set $\{0, 1\}$, each individual has two potential outcomes $Y^*(0), Y^*(1)$. *Counterfactual outcomes* are ones that would have been observed, had the treatment been different. For example, if our treatment actually was $A = 1$, then our counterfactual outcome would be $Y^*(0)$; alternatively, if our actual observed treatment was $A = 0$, then our counterfactual outcome would be $Y^*(1)$. Before the treatment decision is made, any possible outcome is a potential outcome: $Y^*(0)$ or $Y^*(1)$, whereas after the study, there is an observed outcome, $Y = Y^*(A)$, and an unobserved counterfactual outcome that is represented thus: $Y^*(1 - A)$. Typically, the terms potential outcomes and counterfactual outcomes are used interchangeably.

In principle, it is impossible to simultaneously observe both $Y^*(0)$ and $Y^*(1)$ at an individual level. Researchers, therefore, focus on the *average treatment effect* (ATE), or the causal treatment effect at the population level, which is defined as follows:

$$\delta^* := \mathbb{E}[Y^*(1) - Y^*(0)] = \mathbb{E}[Y^*(1)] - \mathbb{E}[Y^*(0)]. \tag{1.1}$$

Average treatment effect δ^* represents the difference between the average potential outcome if all individuals in the population were to receive treatment 1 and that if all of them were to receive treatment 0. Using observational data, one major and fundamental objective of causal inference then is to estimate the average treatment effect, under some conditions and certain assumptions. In much of the causal literature, the following identifiability assumptions are made: (I) *consistency*, that is, $Y = AY^*(A) + (1-A)Y^*(1-A)$; (II) *strong ignorability*, that is, denoting pre-treatment covariates by $\boldsymbol{X}$, then we have $\{Y^*(0), Y^*(1)\} \perp\!\!\!\perp A \mid \boldsymbol{X}$, where $\perp\!\!\!\perp$ denotes independence; and (III) *positivity*, $\mathbb{P}(A = a \mid \boldsymbol{X} = \boldsymbol{x}) > 0$ for any $a, \boldsymbol{x}$. Applying these three identifiability assumptions, from equation (1.1), we have

$$\begin{aligned}
\delta^* &= \mathbb{E}[Y^*(1)] - \mathbb{E}[Y^*(0)] \\
&= \mathbb{E}_{\boldsymbol{X}}\left[\mathbb{E}\{Y^*(1) \mid \boldsymbol{X}\}\right] - \mathbb{E}_{\boldsymbol{X}}\left[\mathbb{E}\{Y^*(0) \mid \boldsymbol{X}\}\right] \\
&\overset{\text{(II)}}{=} \mathbb{E}_{\boldsymbol{X}}\left[\mathbb{E}(Y^*(1) \mid \boldsymbol{X}, A = 1)\right] - \mathbb{E}_{\boldsymbol{X}}\left[\mathbb{E}(Y^*(0) \mid \boldsymbol{X}, A = 0)\right] \\
&\overset{\text{(I)}}{=} \mathbb{E}_{\boldsymbol{X}}\left[\mathbb{E}(Y \mid \boldsymbol{X}, A = 1)\right] - \mathbb{E}_{\boldsymbol{X}}\left[\mathbb{E}(Y \mid \boldsymbol{X}, A = 0)\right], \qquad (1.2)
\end{aligned}$$

where $\mathbb{E}_{\boldsymbol{X}}$ denotes that expectation is with respect to the marginal distribution of $\boldsymbol{X}$. By the positivity assumption, the conditional expectations $\mathbb{E}(Y^*(a) \mid \boldsymbol{X}, A = a)$ and $\mathbb{E}(Y \mid \boldsymbol{X}, A)$ are well defined. Therefore, the result in equation (1.2) illustrates that the ATE can be expressed as an expression of observed data when consistency, strong ignorability, and positivity hold. Especially, δ^* depends on the regression of observed outcomes on covariates and treatment received (Imbens and Rubin [2015], Hernán and Robins [2020], Tsiatis [2019]).

In addition, most causal inference literature also relies on a key assumption known as the stable unit treatment value assumption (SUTVA, Rubin [1980], and see Rubin [1990] and Imbens and Rubin [2015] for discussion). The SUTVA requires that the potential outcomes of one individual are unaffected by changes in the treatment exposures of all other individuals. Formally, in the SUTVA: (1) there are no multiple "versions" of a treatment option; (2) the probabilities (values) of an individual's potential outcomes depend only on the treatment assignment administered to that individual, not the assignments to other individuals.

The first criterion (1), also known as *stability*, means that nominally identical treatments are in fact identical. In other words, an individual's potential outcome under their treatment is the outcome that will actually be observed for that person under their treatment. This criterion may be violated by variable levels of treatment or technical errors. For instance, a surgical procedure (i.e., the treatment) can be performed by different surgeons, and an individual's outcome could depend on the surgeon. Ultimately, a judgment that there are no multiple versions of a treatment is subjective, and in many instances, it may be violated to some extent. The second criterion (2), called the *no interference assumption*, is that the observation on one unit should be unaffected by

the particular assignment of treatments to other units. While often reasonable, this assumption is violated in many cases, such as in the contexts of infectious diseases and social network contact structures (Ogburn and VanderWeele [2014], Ogburn et al. [2017]). In the real world, everyone is embedded in their own social networks, and frequently affected by others. Interference occurs when one individual's exposure affects the outcome of another individual's outcome. Therefore, identifying network interference effects is necessary for real-world problem analysis.

In the case of vaccination, for example, an unvaccinated (i.e., untreated) individual in a mainly vaccinated population may be less likely to contract an infectious disease than the same individual in a mainly unvaccinated population. In the context of school attendance, Kenyan students assigned to undergo a deworming program (i.e., treatment) may interact with others who were not assigned deworming treatment. In one study, deworming significantly improved the outcome of school participation among untreated students in both treatment schools and adjacent schools (Miguel and Kremer [2004]). In a social media example, game invitations were sent by friends; in a friendship network, one's behavior might affect one's friends, and one individual is more likely to consider the game that their friends invite them to play (Su et al. [2019]). Hong and Raudenbush [2006] found that a student's reading score is expected to be affected by the treatment assignment of the other students in their class, but not by the treatment assignment of students in other classes. Sobel [2006] considered the Moving to Opportunity (MTO) study, whereby families in housing projects were randomly assigned to three groups. Under MTO, family A, for example, may move if they are assigned to the same group as family B whose members are friendly with family A. However, if family B is assigned to a different group, family A may not move. In Hong et al. and Sobel's examples, although an individual's treatment assignment stays the same, the assignment of other units varies. Consequently, the individuals will have different results. This phenomenon is called a *spillover effect* (indirect effect, interference (Cox [1958]), peer effect (Manski [1993]) and dissemination effect (Crawford et al. [2019]) are the well-known terms in other subjects). It is distinguished from the *direct effect*, that is, the contrast between the average potential outcome when an individual receives treatment and that when the individual receives control, all other things being the same. However, spillover effects describe how individuals are affected by the treatments of others with whom they are connected.

Interference studies depend on causal mechanisms or interference structures. On one hand, in explaining causal mechanisms, Ogburn and VanderWeele [2014] used causal diagrams to illustrate three distinct causal pathways by which one individual's treatment may affect another's outcome. The first pathway is termed *direct interference*, which occurs when one individual's treatment directly affects another's outcomes, but is unmediated with respect to the first individual's outcome. The second pathway, called *interference by contagion*, is present when individuals' outcomes may have an impact on other people with whom they are in contact. This causal path is different from direct interference because it is mediated through the outcomes of the exposed individuals. The

third pathway is referred to as *allocational interference*. In this case, individuals are allocated to a group, and their interactions within that group can affect other individuals' outcomes in many ways. Thus, for allocational interference, individuals in a certain group are assigned to receive individual or group-level treatments, and their outcomes also depend on treatments of those who else are assigned in the same group. The causal estimands of interest differ according to these causal mechanisms of interference, and Ogburn and VanderWeele [2014] have described their differences and provided criteria for the identification of causal effects. As far as causal mechanisms of interference are concerned, throughout this thesis, we will primarily focus on direct interference.

On the other hand, in explorations of interference structures, *clustered interference*, *spatial interference* and *network interference* are commonly considered three of the main structures (Puelz et al. [2019]). First, clustered interference, analogous to the above-mentioned allocational interference, refers to individuals who can be divided into well-defined clusters or groups and may be affected by the treatments of others within the same cluster. For example, schools and households are frequently studied clusters in much of the literature. Therefore, many researchers assume *partial interference* (Sobel [2006], Perez-Heydrich et al. [2014]); that is, when individuals can be partitioned into clusters, it is plausible to assume that interference occurs within those clusters but not across the clusters. Second, spatial interference assumes that interactions pass through neighbouring individuals, and it is a more complicated structure than the structure of clusters. For example, an experiment in Medellin, Colombia was conducted to investigate the effects of "hot-spot policing" on crime (Collazos et al. [2021]), and the corresponding spillover effects were investigated by Puelz et al. [2019]. Third, network interference refers to interference between individuals in a network of influence (Kao [2017]), and its structure can be represented by an adjacency matrix N with (i, l) entry N_{il} for $i, l \in \{1, ..., n\}$. Under a network interference structure, many researchers assume that any spillover effects on individual i may only emanate from individual l when $N_{il} = 1$, and not from those for which $N_{il} = 0$.

Generally, under a network interference structure, an individual's response may be affected only by their *first-degree neighbours'* exposure or by more than their first-degree neighbours' exposure. These two different interference viewpoints, which distinguish the range of interfering individuals, led to the ideas of *local interference* and *global interference*. Local interference, also known as *neighbourhood interference*, restricts the interfering individuals to only first-degree neighbours. Under local interference, one individual's outcome is assumed to be affected only by their first-degree neighbours' treatments. Therefore, focusing on local interference, many researchers assume that any spillover effects on individual i may only emanate from individual l when $N_{il} = 1$, and not from those for which $N_{il} = 0$. To simplify the problem, these researchers have considered some interference measures, which are functions of neighbours' treatments, often termed *exposure mappings*, such as the number of treated neighbours or the proportion of treated

neighbours (Aronow et al. [2017], Leung [2019], Forastiere et al. [2021], and Li and Wager [2020]). Different from local interference, an individual's outcome may be affected broadly by the treatment assignment of the whole population (global treatment assignment). Under global interference, researchers allow individuals to interfere in largely arbitrary and unknown ways. For instance, Sävje et al. [2021] investigated average treatment effects in the presence of arbitrary and unknown interference. In addition, Tchetgen Tchetgen et al. [2021] considered the case where interference exists between any two individuals if there is a path of connected units linking the two. As an illustration of global and local interference, supposing a finite population of n individuals, for example, in Figure 1.1, where $n = 9$ individuals form a mini-network, individuals 2, 3, and 4 are the only first-degree neighbours of individual 1. Local interference refers to cases in which the outcome for individual 1 depends on his or her own treatment as well as that of individuals 2, 3, and 4, whereas global interference refers to cases in which the outcome of individual 1 (more generally, each individual) depends on the global treatment assignment of all n individuals. Under either a global or a local interference structure, dependence involves the relationships of a network, so it is often called network interference.

Moreover, both clustered interference and spatial interference can be expressed as special cases of network interference. That is, for clustered interference, adjacency matrix N can be expressed as a block-diagonal matrix where each block corresponds to a cluster. For spatial interference, if the spatial distance between two units is less than a creation threshold, then the corresponding adjacency-matrix entry for these two units is 1; otherwise, it is 0. In addition, in a social network, an individual of primary interest or a node whose characteristic we expect to illustrate is called an *ego*, and those to whom the ego is linked or nodes that share ties with the ego are the *alters* or *contacts*. Treating the ego has a "direct effect" on the ego, and treating an alter may have a "spillover (indirect) effect" on the ego (Ogburn et al. [2017], VanderWeele and Christakis [2019]). As shown in Figure 1.1, individual 1 represents the ego, in whom we are interested in particular, and individuals 2, 3, and 4 represent his or her alters.

Having introduced interference and its causal mechanisms and structures, let us move on to review some influential interference studies. Interference studies are conducted in a large number of scientific explorations, such as in economics (Manski [2013], Arpino et al. [2017]), education (Hong and Raudenbush [2006]), political science (Bowers et al. [2013]), and sociology (Aronow [2012]). We here primarily review the methodological developments of interference studies in the causal inference literature. Such studies can be broadly divided into two types, depending on whether or not clusters exist. Moreover, within each type, studies can be further classified based on whether they are randomized experimental studies or observational studies.

In the first type, the studies focus on clustered interference and are built on the partial interference assumption, which allows for but does not assume the existence of interference within clusters. The overall goal of the studies is to draw inferences about treatment effects that

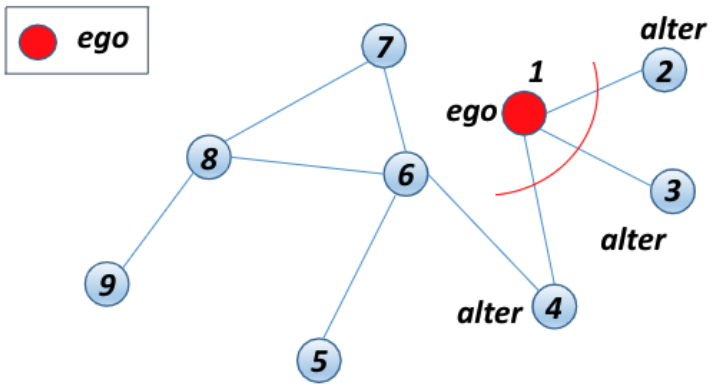


Figure 1.1: An illustration of a mini-network

quantify interference within clusters. For the experimental design studies with partial interference, two-stage randomization was proposed to estimate the treatment effects (Hayes et al. [2000], Longini Jr et al. [2002]). In the estimation, the crucial concept is treatment saturation (or policy, stochastic intervention) φ (Tchetgen and VanderWeele [2012]), which means that φ percent of individuals in a cluster receive treatment. Let μ_φ be the expected potential outcome under treatment saturation φ, and $\mu_{a\varphi}$ be the expected potential outcome when an individual receives treatment a under treatment saturation φ for $a = 0, 1$. Then, within the same treatment saturation group, with a saturation value (for example) φ_1, the direct treatment effects can be identified as $\mu_{1\varphi_1} - \mu_{0\varphi_1}$. For different treatment saturation groups with saturation values φ_1 and φ_0, the spillover effects can be defined as $\mu_{1\varphi_0} - \mu_{1\varphi_1}$ or $\mu_{0\varphi_0} - \mu_{0\varphi_1}$. To conclude, under the partial interference assumption of no spillover effects across clusters, within one cluster, one can identify the direct treatment effect using variations in the treatment status, and across clusters, one can identify the spillover effect using variations in treatment saturation and keeping the same treatment status.

Other studies have focused on the partial interference assumption and observational studies (no randomization). This area extends classic SUTVA approaches for observational studies in the presence of partial interference, and assumes strong ignorability (also known as no unmeasured confounding (NUC)). The most-related work follows Tchetgen and VanderWeele [2012]'s Bernoulli treatment allocation strategy framework, in which each individual in a cluster independently receives treatment with the same, fixed probability. Similarly, following the seminal work of Hudgens and Halloran [2008], the direct and spillover effects of treatment under policy α are the main areas of interest. For instance, various inverse probability weighted (IPW) estimators with interference are studied in Liu et al. [2016], and an IPW estimator with an extension of the work of Tchetgen and VanderWeele [2012] is found in Barkley et al. [2020]. In addition, Liu et al. [2019] constructed doubly robust (DR) estimators of direct and spillover effects, and further extended the

work of Park and Kang [2020] to study efficient semiparametric estimation. Note that, in Liu et al. [2019], the DR estimators are consistent and asymptotically normal, and employ both a propensity score model and outcome regression model, where one of these two models (not necessarily both) is correctly specified.

The second type of interference research is seen in studies where the social network does not consist of clusters, and is an extension of the partial interference assumption. Considering randomization experiments, researchers such as Bowers et al. [2013] and Bowers et al. [2016] have proposed randomization-based inference approaches to estimate both direct and spillover effects for a specified causal model. These studies rely on certain causal model assumptions. Also, when network interference or dependence is present, the independence assumption invoked in standard frequentist inference is dubious. Thus, addressing randomized network-dependent data, researchers such as Basse et al. [2019], Athey et al. [2018], Sävje et al. [2021] have conducted conditional permutation or randomization tests just for detecting interference, but without any causal model assumptions. There is also a parallel study area focusing on dealing with observational network data, with no randomization. Thus, the challenges of this area are the lack of both randomization and independent replicates. Notable works, such as by Tchetgen Tchetgen et al. [2021] and Forastiere et al. [2021], consider an arbitrary form of interference: that the outcome of an individual may depend on treatment received by other individuals with whom a network path through connected individuals exists. Tchetgen Tchetgen et al. [2021] generalized Robins' g-computation algorithm to infer causal effects on a network. Forastiere et al. [2021] developed the aggregation of spillover of treatment to construct direct and spillover effects. Recently, the causal estimands in Forastiere et al. [2021] have also been extended to capture different aspects of interference heterogeneity such as multi-valued treatments in Tortù [2020] and covariate dependence in Bargagli Stoffi et al. [2020].

1.3 Main Illustration: Smoking Cessation in Couples

Having reviewed some work on interference studies in the literature, we now present a motivating example for this thesis: that is, smoking cessation in couples. In doing so, we also introduce and illustrate a real data set — Population Assessment of Tobacco and Health (PATH), which features both a continuous outcome and a binary outcome.

As mentioned in Section 1.1, smoking cessation plays an important role in cancer prevention; therefore, smoking cessation research is common. Some studies have focused on the idea that a desire to quit smoking alone may not be a sufficient motivation in itself. For example, Hubbard et al. [2016] and Foulstone et al. [2017] have addressed smoking cessation in the couples of

families. Meanwhile, a growing body of literature suggests that e-cigarettes (vaping) can be useful as a cessation aid (e.g., Benmarhnia et al. [2018], Hajek et al. [2019]).

On the one hand, for the family-based smoking cessation studies, Foulstone et al. [2017] studied the longitudinal influence of spousal tobacco usage and relationship satisfaction on the probability of smoking cessation. They concluded that both relationship satisfaction and a partner's smoking status are important considerations in smoking cessation, and discovered that the probability that both members of a couple would quit smoking is positively affected by the current smoking cessation of one partner. Moreover, entrenched smokers in couple relationships find it easier to quit smoking if their partners are involved in the interventions. Cobb et al. [2014] analyzed data from 4,500 spouse pairs aged 45 to 64 years from the Atherosclerosis Risk in Communities Study cohort, and found that people who smoke are less likely to quit if their spouse currently smokes. They suggested that smoking cessation programs and clinical advice should consider targeting couples (families) rather than individuals. Hubbard et al. [2016] also held the opinion that family-based interventions are effective components of cessation programmes, yet the value of family-based interventions remains under-researched.

On the other hand, some evidence suggests that e-cigarette usage helps to reduce nicotine dependency or smoking. Polosa et al. [2014] firstly showed that the use of e-cigarettes substantially decreased cigarette consumption without causing significant side effects in chronic schizophrenic patients who smoke but are not intending to quit. The findings of Selya et al. [2017] support that the use of e-cigarettes may act as a smoking reduction method among highly nicotine-dependent young adult cigarette smokers. Moreover, in a randomized trial of e-cigarettes versus nicotine-replacement therapy (Hajek et al. [2019]), researchers found that, compared with nicotine-replacement therapy, e-cigarettes are more effective for smoking cessation when both products were accompanied by behavioral support.

Only a small number of studies have not only examined smoking cessation in couple or family networks, where interference may be present, but also investigated any e-cigarette usage in the participants. However, the Population Assessment of Tobacco and Health study offers a rare opportunity to investigate such contexts. Launched in 2011, the PATH study aims to inform the U.S. Food and Drug Administration's regulatory activities under the Family Smoking Prevention and Tobacco Control Act, which is a federal statute in the U.S.A. and gives the Food and Drug Administration the power to regulate the tobacco industry. The PATH Study is a collaboration between the National Institute on Drug Abuse, the National Institutes of Health, the Center for Tobacco Products, and the Food and Drug Administration.

The PATH Study, which sampled over 150,000 mailing addresses across the United States to create a national sample of tobacco users and non-users, is a longitudinal cohort study on tobacco-use behaviors, attitudes, beliefs, and tobacco-related health outcomes among the country's

adults and youth. Its primary objectives include the following: (1) to identify and explain between-person differences and within-person changes in tobacco-use patterns, including the rate and length of use by specific product type and brand, product or brand switching over time, uptake of new products; (2) to determine between-person differences and within-person changes in risk perceptions regarding harmful and potentially harmful constituents, new and emerging tobacco products, filters and other design features of tobacco products, packaging, and labeling; (3) to characterize the natural history of tobacco dependence, cessation, and relapse, including readiness and self-efficacies to quit, motivations for quitting, the number and length of quit attempts.

In the PATH study, data were gathered in *waves*, starting from 2011, with each subsequent wave beginning approximately one year after the previous one. During Wave 1, households were selected by a stratified four-stage sampling design. Within a selected household, up to two adults aged 18+ and up to two youths were selected. For each wave, data collected included demographic variables (e.g., age, sex, education, race, non-Hispanic indicator), smoking variables (e.g., frequency of smoking, cigarettes per day), e-cigarette variables (e.g., frequency of use, age when started using), predictors of cessation (e.g., plan to quit, interest in quitting, quit attempts), and cigarette dependence related variables (Strong et al., 2017).

Our purpose of analyzing PATH data is to illustrate the statistical methodology for determining an optimal DTR in a longitudinal survey setting where possible interference of treatment of one family member (the alter) may influence the result of treatment of other member (the ego). In the PATH analysis, we randomly choose one member in a household as the ego, then their spouse is automatically assigned as alter. In this case, 'treatment' is the use of e-cigarettes (vaping) by a cigarette smoker, and the outcome measure is a measure of dependence on cigarettes (continuous outcome) (Strong et al. [2017]) or cessation of smoking (binary outcome).

For the PATH data, we particularly address four research problems of interest as follows. Examining the first four waves of PATH data, and studying the cigarette dependence as a continuous outcome,

- (1) assuming no interference, we estimate the optimal DTR for smokers, in terms of a sequence of use or non-use of e-cigarettes, for reducing cigarette dependence;
- (2) allowing for interference where treatments of alters are assumed to be fixed, we estimate the optimal DTR for egos, in terms of their use or non-use of e-cigarettes, for reducing cigarette dependence in the egos.

Then, further considering the binary outcome of smoking cessation,

- (3) assuming no interference, we estimate the optimal DTR for smokers, in terms of a sequence of use or non-use of e-cigarettes, to achieve smoking cessation.

- (4) finally, allowing for interference, we estimate the optimal DTR for the individuals of interest and their spouses, in terms of their use or non-use of e-cigarettes, for achieving both individuals in the same household quit or attempt to quit smoking.

In our analysis of PATH data for continuous outcomes in the context of reducing cigarette dependence, problem (1) is addressed by the dWOLS method, proposed in Wallace and Moodie [2015], in the real data analysis in Chapter 3. Extending dWOLS in the presence with interference, we also solve problem (2) in Chapter 3 PATH analysis. Moreover, for binary outcomes in the context of achieving smoking cessation, we develop a new methodology for DTR estimation with binary outcomes to answer question (3) in Chapter 4. Building on existing interference study methodologies, we also develop novel methodologies to answer question (4) in Chapter 5.

1.4 Contributions and Organization of Thesis

Although a multitude of estimation methods have been developed to handle problems related to sequential treatment decisions, optimal DTR estimation research problems remain as this field is in its infancy. In fact, considerable research is required before these methods can be applied to real-world situations. In this thesis, on the one hand, we contribute to the current dWOLS theory, illustrated in several contexts, by providing additional support of the double robustness property's mechanism and confirming the optimal dWOLS weights. Building on this solid foundation, on the other hand, we further contribute to optimal DTR estimation with interference, in both specific household interference and general network interference. Then, we develop the doubly robust optimal DTR estimation with binary outcomes, and the doubly robust DTR estimation with ordinal outcomes in the presence of household interference. To illustrate the research focus in practice, and to emphasize its applicability to an area of critical interest, considering continuous, binary and ordinal outcomes, we have used data from the Population Assessment of Tobacco and Health Study, and applied our extended dWOLS method to obtain optimal DTRs for individuals in couples where interference is present. This thesis contains six chapters, with an Appendix offering additional materials, structured as follows:

In Chapter 2, we provide the relevant technical and theoretical background, including notations, and briefly summarizes two existing methodologies for estimating optimal DTRs in the case of no interference: Q-learning and dWOLS. Double robustness to the misspecification of treatment-free models or treatment models is explained. Crucially, we first provide two different proofs of the double robustness of dWOLS, and a theorem of the optimal dWOLS weights in the context of yielding the minimum variance of the estimators of interest. These two proofs of the double

robustness form the basis for our work on dWOLS with interference and DTR estimation for binary outcomes.

As previously stated, it is typical to assume no interference when estimating DTRs from observational data. However, considering interference in DTRs will provide a novel and practical way to investigate treatment assignment, aiming not only at the improvement of individual outcomes but also at greater treatment efficiency in terms of treatment assignment for the whole population. Concentrating on continuous outcomes, in Chapter 3, we investigate the DTR estimation method of dynamic weighted ordinary least squares, which boasts easy implementation and the so-called double robustness property, but has typically relied on the assumption of no interference. We define a network propensity function and build on it to establish an implementation of dWOLS that remains doubly robust under interference associated with network links. The method's properties are demonstrated via simulations, where different types of interference scenarios are studied. Investigating cigarette dependence within two-person household networks, and using real data analysis, we apply estimation and decision methodologies to the PATH data, with the outcome being the measure of nicotine dependence. The work in this chapter, reorganized as a research paper, Jiang et al. [2022b], has been published in *The Canadian Journal of Statistics*.

Multiple methods have been proposed for optimal DTR estimation with continuous outcomes as these outcomes are more sensitive than discrete outcomes in the implementation of a specific DTR. However, optimal DTR estimation with binary outcomes is more complicated due to less information being provided by the discrete outcomes, and thus has received comparatively little attention. Studying binary outcomes, in Chapter 4, we develop the dynamic weighted generalized linear model (dWGLM), a doubly robust and easy-to-use method for estimating an optimal DTR with binary outcomes. In our methodology, we propose two-step (or two-phase) weighted logistic regression for the consistent estimation of parameters of interest. We also provide a balancing weight criterion for any generalized linear models, and the underlying theory is suitable for any g link function. Thus, from this viewpoint, dWOLS is a special case of dWGLM, where the link function is the identity link: $g(\mu) = \mu$. Furthermore, we also offer binary pseudo-outcomes for sequential treatment decisions. In practice, we illustrate the method in an analysis of e-cigarette usage and smoking cessation, using observational data from the PATH study. We have summarized this chapter as a research paper, Jiang et al. [2022a], which has been posted to the arXiv.

Further, building on the work of Chapter 3 and Chapter 4, considering the special PATH household interference case and individuals' binary outcomes, in Chapter 5, we further investigate robust DTR estimation with ordinal outcomes in the presence of household interference. First, we study the interference, which is the scenario where not only the individuals' outcome can be affected by their neighbours' treatments, but also there exists a possible association between their treatments. We provide an analytical framework to estimate joint propensity scores in the association aware treatments case. Second, we propose a double robust DTR estimation method

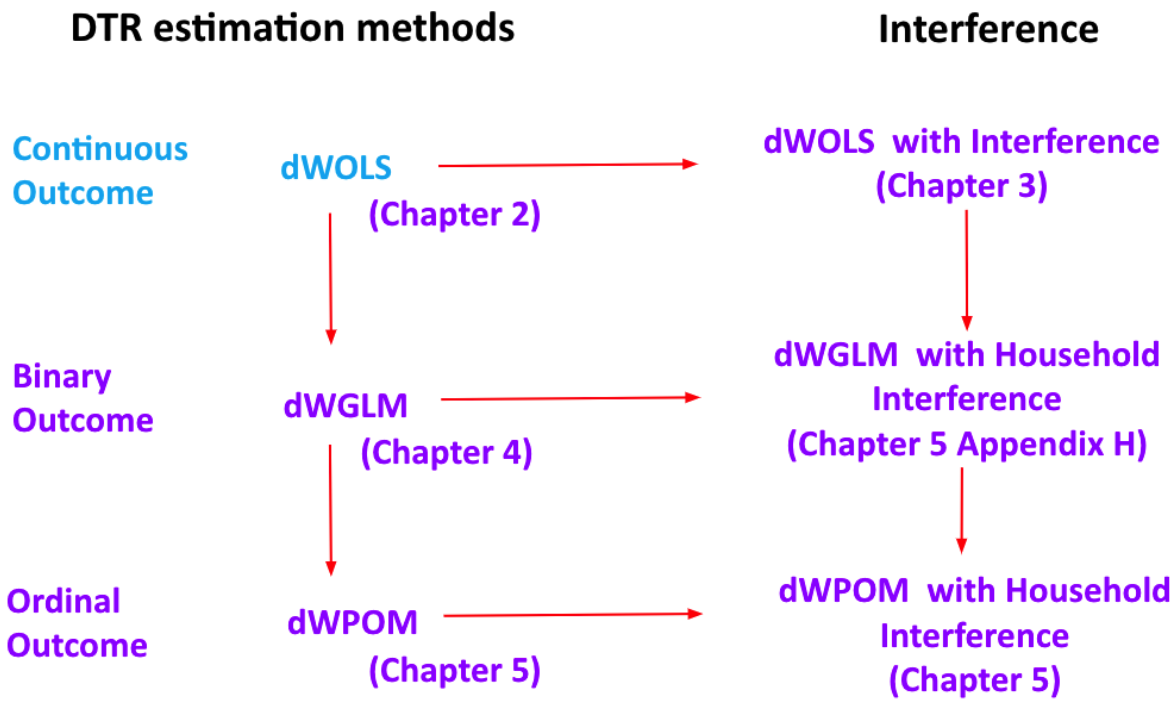


Figure 1.2: Connections among the optimal DTR estimation methods and those methods with interference. (dWOLS in blue is an existing method, and other methods in purple are proposed in this thesis).

for ordinal outcomes, namely, the dynamic weighted proportional odds model (dWPOM), by using sequential proportional odds models with ordinal outcome-specific balancing weights. Third, by modeling the household utility, which is the joint outcome of the individuals in the same household, our method can output optimal decisions for each individual in that household for maximizing the household utility. This chapter is being prepared for publication.

Finally, in Chapter 6, we summarize the whole thesis and particularly outline the methodologies that were proposed in this thesis. Furthermore, we also discuss the potential future research areas in this chapter, including heterogeneous treatment effects estimation, novel robust DTR estimation methods, and multiple-agent reinforcement learning under interference. In particular, in Appendix I, we provide some thoughts about extending X-learning (Section I.1) and outcome weighted learning (Section I.2) to cases where there is interference.

To clearly see the structure of this thesis, Figure 1.2 sets out an overview of the contributions of this research and the relations among the methodology chapters (Chapters 2, 3, 4), using different

colours to distinguish between previously existing methods and those developed in this thesis work. In blue, dWOLS, proposed by Wallace and Moodie [2015] and further explored in Chapter 2, has been the foundation for the whole thesis. Other methods, in purple, are our contributions in this thesis. From the left-hand side, vertically, dWOLS is extended to dWGLM and dWPOM, which are robust optimal DTR estimation methods for binary outcomes and ordinal outcomes, respectively. Horizontally, each of these three methods is also extended and improved to address the interference problem. Here, the "first glance" of Figure 1.2 is just a simplified overview of a more-detailed version provided in the conclusion section of Chapter 6, which summarizes the details of connections among these methods and the contributions of this thesis.

Chapter 2

A Review of Optimal DTR Estimation

In this chapter, we start by introducing single decision treatment regimes, then briefly summarize Q-learning and dWOLS for multiple-stage treatment decisions in the case of no interference. For dWOLS, using the algebra of partitioned regression, and with the use of appropriate weights to compensate for lack of balance, we show that the estimators of heterogeneous treatment effects are consistent if either the treatment-free model or the treatment model is correctly specified, that is, so-called double robustness. In addition, we also present novel insights on the double robustness of dWOLS by investigating the estimation equation systems of dWOLS. Both the algebra of partitioned regression and the estimation equation insights serve as alternative proofs of the dWOLS theorem that was presented in Wallace and Moodie [2015]. Moreover, the algebra of partitioned regression and the estimation-equation insights function as a foundation for studying dWOLS under interference (Chapter 3) and with binary outcomes (Chapter 4), respectively. Finally, in this chapter, the variance of dWOLS estimators is investigated, and we propose the formation of the optimal weights of dWOLS in the context of the most-efficient estimation.

2.1 Single Decision Treatment Regimes

In the single decision setting, the goal is to identify an optimal treatment regimen or an individualized treatment rule, d^{opt}, which satisfies $\mathbb{E}[Y^*(d^{opt})] \geq \mathbb{E}[Y^*(d)]$ for any other regime d. In order to estimate an optimal regime in practice, it must be possible to identify an optimal regime from data $\{(\boldsymbol{X}_i, A_i, Y_i)\}_{i=1}^n$, which comprise n independent and identically distributed triples $(\boldsymbol{X}, A, Y)$ where $\boldsymbol{X}$ denotes baseline patient characteristics; $A \in \mathscr{A}$ denotes the assigned treatment that is an option in $\mathscr{A}$, the set of all available treatment options; and $Y \in \mathbb{R}$ denotes a continuous

outcome variable coded such that larger values are preferred. (Kosorok and Laber [2019], Tsiatis [2019]). Typically, as noted in Chapter 1, under the causal inference analysis framework and the identifiability assumptions, namely consistency, SUTVA, strong ignorability, and positivity, we have $\mathbb{E}[Y^*(a) \mid \boldsymbol{X} = \boldsymbol{x}] = \mathbb{E}[Y^*(a) \mid A = a, \boldsymbol{X} = \boldsymbol{x}] = \mathbb{E}[Y \mid A = a, \boldsymbol{X} = \boldsymbol{x}]$. Formally, for a single-stage decision, *the optimal treatment rule* is identified from $d^{opt}(\boldsymbol{x}) = \arg\max_{a \in \mathscr{A}} \mathbb{E}[Y^*(a) \mid \boldsymbol{X} = \boldsymbol{x}]$, a function that maps an individual's covariates to an optimal treatment option, which maximizes the conditional expectation of the corresponding potential outcome. Therefore, under above assumptions, it is equivalent to $d^{opt}(\boldsymbol{x}) = \arg\max_{a \in \mathscr{A}} Q(\boldsymbol{x}, a)$, where $Q(\boldsymbol{x}, a) = \mathbb{E}[Y|\boldsymbol{X} = \boldsymbol{x}, A = a]$, which is known as the *Q-function* ("Q" represents "quality of action", and will be explained in Q-learning section). A regression of Y on $(\boldsymbol{X}, A)$ provides an estimator of $Q(\boldsymbol{x}, a)$ that is denoted $\widehat{Q_n}(\boldsymbol{x}, a)$. Then, this is optimized to obtain $\widehat{d}(\boldsymbol{x}) = \arg\max_{a \in \mathscr{A}} \widehat{Q_n}(\boldsymbol{x}, a)$. Additionally, $Q(\boldsymbol{x}, a)$ could also be estimated either semiparametrically or nonparametrically.

In randomization trials, the assumption of strong ignorability automatically holds, and so does the assumption of positivity if randomization probabilities are positive; however, in the observational studies where individuals choose (or are given) treatments in a naturalistic and unknown way, the strong ignorability assumption may not hold. Usually, to overcome this obstacle, we need to collect additional covariates (denoted as $\boldsymbol{Z}$) as much as possible, such that A is independent of $Y^*(a)$ conditional on both $\boldsymbol{X}$ and $\boldsymbol{Z}$. Some researchers called this *general no unmeasured confounding condition:* $\{Y^*(a) : a \in \mathscr{A}\} \perp\!\!\!\perp A \mid \boldsymbol{X}, \boldsymbol{Z}$ such that $\boldsymbol{X} \cup \boldsymbol{Z}$ is a sufficient adjustment set (Zeng [2021], Miao et al. [2018], Knüppel and Stang [2010]). Building on this general no unmeasured confounding condition, there are two well known approaches to estimate $\mathbb{E}[Y^*(a)|\boldsymbol{X}]$. The first estimation method is based on G-computation. Under the above-mentioned assumptions and no unmeasured confounding, one can estimate $\mathbb{E}[Y^*(a)|\boldsymbol{X}, \boldsymbol{Z}]$ using $\mathbb{E}[Y|\boldsymbol{X}, \boldsymbol{Z}, A = a]$. To identify the causal estimand of interest $\mathbb{E}[Y^*(a)|\boldsymbol{X}]$, we first estimate $\mathbb{E}[Y|\boldsymbol{X}, \boldsymbol{Z}, A = a]$ then estimate the conditional distribution of $\boldsymbol{Z}$ given $\boldsymbol{X}$; lastly, we can calculate

$$\mathbb{E}[Y^*(a)|\boldsymbol{X}] = \int \mathbb{E}[Y^*(a)|\boldsymbol{X}, \boldsymbol{Z}] f(\boldsymbol{Z}|\boldsymbol{X}) d\boldsymbol{Z} = \int \mathbb{E}[Y|\boldsymbol{X}, \boldsymbol{Z}, A = a] f(\boldsymbol{Z}|\boldsymbol{X}) d\boldsymbol{Z}.$$

Therefore, the subsequent estimation process uses data to estimate the conditional mean $\mathbb{E}[Y|\boldsymbol{X}, \boldsymbol{Z}, A = a]$ and conditional density $f(\boldsymbol{Z}|\boldsymbol{X})$. The second way to estimate $\mathbb{E}[Y^*(a) \mid \boldsymbol{X}]$ is using the inverse probability weighted expectation of Y among subjects with $A = a$. Building on

the spirit of importance sampling, this method relies on the fact that

$$\begin{aligned}\mathbb{E}\left[\frac{Y\mathbb{I}(A=a)}{\mathbb{P}(A=a \mid \boldsymbol{X}, \boldsymbol{Z})} \mid \boldsymbol{X}\right] &= \mathbb{E}\left[\frac{Y^*(a)\mathbb{I}(A=a)}{\mathbb{P}(A=a \mid \boldsymbol{X}, \boldsymbol{Z})} \mid \boldsymbol{X}\right] \\ &= \mathbb{E}\left[\mathbb{E}[Y^*(a) \mid \boldsymbol{X}, \boldsymbol{Z}] \frac{\mathbb{E}[\mathbb{I}(A=a) \mid \boldsymbol{X}, \boldsymbol{Z}]}{\mathbb{P}(A=a \mid \boldsymbol{X}, \boldsymbol{Z})} \mid \boldsymbol{X}\right] \\ &= \mathbb{E}\left[Y^*(a) \mid \boldsymbol{X}\right],\end{aligned}$$

where the first equation depends on the consistency assumption, and the second equation relies on the general no unmeasured confounding condition and the law of total expectation, with $\mathbb{I}(A=a)$ denoting an indicator function of $A=a$. For these two methods, to acquire a consistent estimator of $\mathbb{E}\left[Y^*(a) \mid \boldsymbol{X}\right]$, one needs to correctly specify either the conditional mean (for the former) or the treatment propensity model (for the latter). In light of this, some researchers have constructed inverse probability-weighted augmentation estimators to achieve double robustness and local efficiency using semiparametric efficiency theory (Rotnitzky [2008], Tsiatis [2006], Davidian [2021]).

Alternatively, from the viewpoint of the patients' heterogeneity, an essential goal of DTR estimation is to estimate *heterogeneous treatment effects* (HTEs). Rather than consider the average treatment effect $\delta^* := \mathbb{E}[Y^*(1) - Y^*(0)]$, we try to estimate individualized or personalized treatment effects. For instance, if we can observe the i^{th} patient's covariates $\boldsymbol{X}_i = \boldsymbol{x}$, we are interested in understanding how treatment effects are influenced by the observed covariates $\boldsymbol{X}_i$. Doing so brings us to the *conditional average treatment effect* (CATE), defined as

$$\delta^*(\boldsymbol{x}) := \mathbb{E}[Y_i^*(1) - Y_i^*(0) \mid \boldsymbol{X}_i = \boldsymbol{x}].$$

It is important to note that CATE is different from the – in general unknowable – *individual treatment effect* (ITE), which is defined as $\delta_i^* := Y_i^*(1) - Y_i^*(0)$ for individual i; rather, CATE is the expectation of the ITE conditional on the values of the covariates (Lei and Candès [2021]), that is, $\delta^*(\boldsymbol{x}) = \mathbb{E}[\delta_i^* \mid \boldsymbol{X}_i = \boldsymbol{x}]$. CATE is still an average treatment effect but with an average over a targeted group of samples as characterized by their covariates $\boldsymbol{X}_i$, so CATE is also known as individualized average treatment effect (IATE). It is important to note that, considering binary treatments, we can also define DTR or ITR as a function of CATE, that is,

$$d^{opt}(\boldsymbol{x}) := \mathbb{I}\left[\delta^*(\boldsymbol{x}) > 0\right] = \mathbb{I}\left(\mathbb{E}[Y_i^*(1) - Y_i^*(0) \mid \boldsymbol{X}_i = \boldsymbol{x}] > 0\right).$$

For CATE estimation, the first common method is carried out using *conditional outcome modeling* (COM), also called G-computation, parametric G-formula, or standardization in the literature. That is, we estimate the outcome model to predict $\mu(a, \boldsymbol{x}) := \mathbb{E}[Y \mid A = a, \boldsymbol{x}]$ from data; for

instance, if the estimator is $\hat{\mu}(a, \boldsymbol{x})$, then the conditional outcome modeling estimator for CATE is the difference of $\hat{\mu}(a, \boldsymbol{x})$ between inputting $a = 1$ and $a = 0$, given the covariates $\boldsymbol{x}$. The task is just to estimate the single outcome model, so the COM estimator is also referred to as S(ingle)-learner. Typically, treatment A is often a one-dimensional variable, but when the covariates $\boldsymbol{x}$ are high-dimensional, then the COM estimator of CATE $\hat{\delta}^*(\boldsymbol{x}) = \hat{\mu}(1, \boldsymbol{x}) - \hat{\mu}(0, \boldsymbol{x})$, in which only the input A is changed between terms of $\hat{\mu}(A, \boldsymbol{x})$, may be biased toward zero (Künzel et al. [2019]). It is reasonable that a learner (e.g., a neural network) could neglect A but concentrate on $\boldsymbol{x}$ as its input, if $\boldsymbol{x}$ is high-dimensional but A is one-dimensional. This may result in an CATE estimate of zero. Therefore, the so-called grouped COM comes into being, where the data are split into two groups based on the treatment value used, such that the treatment variable will not be ignored. Each collection of data for the groups is then trained for the corresponding estimator; that is, with $A = 1$ data, estimate $\mu_1(\boldsymbol{x})$, and $A = 0$ data are used to estimate $\mu_0(\boldsymbol{x})$. The corresponding estimators are $\hat{\mu}_1(\boldsymbol{x})$ and $\hat{\mu}_0(\boldsymbol{x})$, respectively. To increase the data efficiency, building on the grouped COM, Künzel et al. [2019] proposed *X-learning* and the corresponding CATE estimator, X-learner. X-learner contains three steps: first estimate $\mu_a(\boldsymbol{x})$ via appropriate regression methods for both treated ($a = 1$) and untreated ($a = 0$) groups. Second, for each one in the untreated group, specify individual treatment effect estimates, $\hat{\varphi}_{0,i} = \hat{\mu}_1(\boldsymbol{x}_i) - Y_i^*(0)$, and fit a model $\hat{\tau}_0(\boldsymbol{x})$ to predict $\hat{\varphi}_{0,i}$ from the corresponding untreated group. An estimator $\hat{\tau}_1(\boldsymbol{x})$ for the treated group is obtained analogously. Third, X-learner is constructed as the combination of the two treatment effect estimators in the weighted sense such that $w\hat{\tau}_0(\boldsymbol{x}) + (1 - w)\hat{\tau}_1(\boldsymbol{x})$, with weights (w) that are bounded in $(0, 1)$. Note that the choice of weights for the X-learner will be discussed at the end of this chapter.

In addition, building on these foundations, other popular CATE estimation methods have been proposed recently, either within a semiparametric or nonparametric framework; for example, robust strategies such as R-learner (Nie and Wager [2021]), and DR-learner (Kennedy [2020]); and deep-learning-based (SNet series) methods, such as TARNet (SNet-1, Shalit et al. [2017]), DragonNet (SNet-2, Shi et al. [2019]), and DR-CFR (SNet-3, Hassanpour and Greiner [2019]). See the work of Curth and van der Schaar [2021], which investigates the estimation of HTE using supervised learning methods and CATE estimation using meta-learners.

2.2 Multiple Decision Treatment Regimes

Analogously to the single-stage case, a statistical framework that involves concepts of causal inference is still required in the multi-stage case. However, due to the time-dependent nature of sequential decision making, making optimal decisions in multiple-decision treatment regimes is more complex than making them in single-decision settings. The sequential decision problem

is derived from reinforcement learning (RL), and the goal is to identify the best sequence of decision rules (policy) such that the average long-term reward is maximized if such rules are implemented. Note that the current decision will not only affect the current reward but also impact the future status on which further decisions will be based, and thus influence future rewards. For the multi-stage treatment decision problem, current treatment decisions will both affect the current outcomes and potentially affect the future ones. This phenomenon has been termed delaying the treatment effects. The best decision for a certain stage is not necessarily the optimal decision for the whole decision process; thus, to make sequential optimal decisions, at each decision point, not only do we need to consider immediate outcomes but also future outcomes. The result is a so-called adaptive treatment decision and requires the agent (decision maker) to be "farsighted" when making decisions.

Because a good policy on which an agent makes decisions must know how the future reward depends on the current state and action, two core value functions in RL, *state-action value function* and *state value function*, have come up. These two value functions indicate the expected return increment at a certain time given a current state and action. In addition, *the Bellman equation* (or *the dynamic programming equation*) for these two value functions is a recursive equation connecting the future values, and it indicates that the value functions at a certain stage depend only on future decision rules. Thus, the Bellman equations for the optimal policy provide a backward way to identify the optimal policy; that is, if we know the future optimal actions, we can make optimal decisions based on these optimal actions for the current stage and thus learn the optimal policy. See Sutton and Barto [2018], a seminal introductory text in RL, for more details in the context of optimal control and Bellman optimality. Building on this fundamental background in RL, in this section, we present two (regression-based) backward induction methodologies for DTR estimation in a multi-stage setting with no interference.

2.2.1 Q-learning

Q-learning, a well-known value-based method from reinforcement learning, is an approximate dynamic programming algorithm and is used to determine the optimal action-selection policy related to Q-functions. Typically, in precision medicine literature, Q-learning is a regression based approach used to identify optimal DTRs, and can be implemented in the DTR context with different variations, such as Q-learning for estimating optimal DTRs from observational data (Moodie et al. [2012]), and Q-learning with difference forms of outcomes (Moodie et al. [2014]).

Following the standard DTR literature notation, we denote that there are K treatment stages, and the beginning of each stage is one decision point. Let a_j be the j^{th} treatment decision and $\mathcal{A}_j$ denote the set of available treatment options at decision j; we study a binary treatment coded as 0

or 1, that is, $a_j \in \mathscr{A}_j = \{0, 1\}$. Let Y denote the observed patient outcome that is measured at some point after the treatment stage, i.e., after a_K has been given. Let $\boldsymbol{x}$ be the covariate value for the individual itself, and $\boldsymbol{h}_j$ be a covariate matrix containing patient information (history) prior to the j^{th} treatment decision; history can include previous treatments $a_1, ..., a_{j-1}$, and previous outcomes. Let $\boldsymbol{h}_j^{\alpha}$, $\boldsymbol{h}_j^{\beta}$ and $\boldsymbol{h}_j^{\psi}$ be three subsets of covariate contained in $\boldsymbol{h}_j$, that are employed in different three models that are respectively parameterized by $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\psi}$. Over- and underline notation denote the past and future respectively. For instance, the vector of the first j treatment decisions $(a_1, a_2, ..., a_j)$ can be written $\overline{\boldsymbol{a}}_j$, while the vector of treatment decisions from stage $j+1$ onwards $(a_{j+1}, a_{j+2}, ..., a_K)$ can be written as $\underline{\boldsymbol{a}}_{j+1}$. Then, history can be expressed as $\boldsymbol{h}_j = (\overline{\boldsymbol{x}}_j, \overline{\boldsymbol{a}}_{j-1})$. Equipped with this notation, we now formally define a dynamic treatment regime.

Definition 2.1. ***Dynamic Treatment Regime***
With K treatment decision points (where $K \geq 1$), at decision stage j, a treatment decision rule $d_j(\boldsymbol{h}_j)$ is a function that maps an individual's history to a treatment option in $\mathscr{A}_j$, for $j = 1, ..., K$. Then, a dynamic treatment regime $d \in \mathcal{D}$, where $\mathcal{D}$ denotes the class of all possible treatment regimes, is a sequence of such treatment decision rules to be applied in order; that is, with K decision points:

$$d = \{d_1(\boldsymbol{h}_1), ..., d_j(\boldsymbol{h}_j), ..., d_K(\boldsymbol{h}_K)\}.$$

As stated previously, precision medicine aims to identify an optimal treatment regimen through various DTR estimation methods. Relying on the above notations and the definition of DTRs, we now formally define the optimal treatment regime:

$$d^{opt} = \arg\max_{d \in \mathcal{D}} \mathbb{E}\{Y^*(d)\} = \arg\max_{d \in \mathcal{D}} \mathcal{V}(d),$$

where value of a treatment regime $\mathcal{V}(d) = \mathbb{E}\{Y^*(d)\}$ is expected outcome if all individuals in the population were to receive treatment according to a DTR d.

To proceed with DTR estimation for either observational or randomized data in multi-stage decision settings, we use the following identifiability assumptions: (1) consistency (Rubin [1980]): the potential outcome under some sequence of treatments is equal to the observed outcome if those treatments were those actually received; (2) there are no unmeasured confounders for any possible treatment regimes (or sequential randomization assumption Robins [1986]), i.e., conditional on current patient history, the current stage treatment is independent of future potential outcome or covariates; (3) no interference between individuals (Cox [1958]): the outcome of one patient is unaffected by the treatment assignment of other patients, and (4) positivity (Robins [2004]): at each decision point, there is a non-zero probability of being assigned to each of the treatment levels, no matter what the past treatment and covariate history.

Generally, for each stage $j \leq K$ in a K stage treatment decision problem, the fundamental concept in the Q-learning is the Q *(quality) function*, $Q_j\left(\boldsymbol{h}_j, a_j\right)$, which is the optimal outcome (reward) increment given the current state history $\boldsymbol{h}_j$ (including patients' all information collected by stage j) and treatment assignment a_j. Therefore, if we know $Q_j\left(\boldsymbol{h}_j, a_j\right)$, then the optimal DTR at stage j is $d_j^{opt}(\boldsymbol{h}_j) = \underset{a_j \in \mathcal{A}_j}{\arg\max}\, Q(\boldsymbol{h}_j, a_j)$. Moreover, the bellman equation, for the final stage K, is $Q_K\left(\boldsymbol{h}_K, a_K\right) = \mathbb{E}\left[Y | \boldsymbol{H}_K = \boldsymbol{h}_K, A_K = a_K\right]$, and for all earlier stage $j < K$, is $Q_j\left(\boldsymbol{h}_j, a_j\right) = \mathbb{E}\left[\underset{A_{j+1}}{\max} Q_{j+1}\left(\boldsymbol{H}_{j+1}, A_{j+1}\right) | \boldsymbol{H}_j = \boldsymbol{h}_j, A_j = a_j\right]$. Typically, $\underset{A_{j+1}}{\max} Q_{j+1}\left(\boldsymbol{H}_{j+1}, A_{j+1}\right)$ is defined as the $(j+1)^{th}$ pseudo-outcome and represents the potential outcome which a patient with the given history would have if they go on to receive optimal treatments from the j^{th} stage, that is receive $\underline{\boldsymbol{a}}_{j+1}^{\text{opt}} = (a_{j+1}^{opt}, a_{j+2}^{opt}, ..., a_K^{opt})$. Thus, the j^{th} Q-function is the expected value of $(j+1)^{th}$ pseudo-outcome given current history and treatment, and these quality functions could be commonly estimated by linear regressions (e.g., ordinary least square) or some supervised learning methods, such as random forest, ($\epsilon-$) support vector regression (SVR) and neural networks. In particular, if the Q-functions are linearly modeled such that $Q_j\left(\boldsymbol{h}_j, a_j; \boldsymbol{\beta}_j, \boldsymbol{\psi}_j\right) = \boldsymbol{\beta}_j^\top \boldsymbol{h}_j^\beta + \boldsymbol{\psi}_j^\top a_j \boldsymbol{h}_j^\psi$, then the optimal treatment $d_j^{opt}(\boldsymbol{h}_j) = \mathbb{I}(\boldsymbol{\psi}_j^\top \boldsymbol{h}_j^\psi > 0)$ for $j \leq K$. We note that the interactions between treatment and covariates should always be included in the regression. To enhance the prediction accuracy and interpretability of model, regularization terms, for example, least absolute shrinkage and selection operator (Lasso), could be included in the models.

Now suppose we have an i.i.d. (independently and identically distributed) sample of n records, with the i-th record consisting of $(\boldsymbol{h}_{1i}, \boldsymbol{h}_{2i}, ..., \boldsymbol{h}_{Ki}; a_{Ki}, Y_{Ki})$. Note that the first letter of subscript refers to the stage, and the second letter refers to the number of the record. The subscript of $\boldsymbol{h}_{ji}$, for example, means j^{th} stage and i-th record. Then, considering the standard ordinary least squares (OLS) for estimating Q-functions, Q-learning has following 4 steps:

- Step 1: For the stage K regression, we can compute

$$\left(\hat{\boldsymbol{\beta}}_K, \hat{\boldsymbol{\psi}}_K\right) = \underset{\boldsymbol{\beta}_K, \boldsymbol{\psi}_K}{\arg\min} \sum_{i=1}^{n} \left[Y_{Ki} - Q_K\left(\boldsymbol{h}_{Ki}, a_{Ki}; \boldsymbol{\beta}_K, \boldsymbol{\psi}_K\right)\right]^2, \tag{2.1}$$

  that is OLS regressing Y_K on history $\boldsymbol{H}_K$ and A_K. Then, in an application, the K^{th} optimal treatment rule would be to *prescribe* $A_K = 1$ *if* $\hat{\boldsymbol{\psi}}_K^\top \boldsymbol{H}_K^\psi > 0$; *and prescribe* $A_K = 0$ *otherwise*, where $\hat{\boldsymbol{\psi}}_K^\top$ comes from (2.1) and the i.i.d. observational data set.

- Step 2: We construct the j^{th} stage pseudo-outcome by using estimated parameters from the

$(j+1)^{th}$ regression step.

$$\widetilde{\mathscr{Y}}_j(\boldsymbol{H}_{j+1}, \hat{\boldsymbol{\beta}}_{j+1}, \hat{\boldsymbol{\psi}}_{j+1}) = \max_{A_{j+1}} Q_{j+1}(\boldsymbol{H}_{j+1}, A_{j+1}; \hat{\boldsymbol{\beta}}_{j+1}, \hat{\boldsymbol{\psi}}_{j+1}) = \hat{\boldsymbol{\beta}}_{j+1}^{\top} \boldsymbol{h}_{j+1}^{\beta} + (\hat{\boldsymbol{\psi}}_{j+1}^{\top} \boldsymbol{h}_{j+1}^{\psi})^{+},$$

where $x^{+} = x * \mathbb{I}(x > 0)$.

- Step 3: Stage j regression : uses OLS regression of the j^{th} pseudo-outcome on history $\boldsymbol{h}_j$ and a_j.

$$\left(\hat{\boldsymbol{\beta}}_j, \hat{\boldsymbol{\psi}}_j\right) = \arg\min_{\boldsymbol{\beta}_j, \boldsymbol{\psi}_j} \sum_{i=1}^{n} \left(\widetilde{\mathscr{Y}_{ji}} - Q_j\left(\boldsymbol{h}_{ji}, a_{ji}; \boldsymbol{\beta}_j, \boldsymbol{\psi}_j\right)\right)^2,$$

  Then, in an application, the j^{th} optimal treatment rule would be to $prescribe\ A_j = 1\ if\ \hat{\boldsymbol{\psi}}_j^{\top} \boldsymbol{H}_j^{\psi} > 0;\ and\ prescribe\ A_j = 0\ otherwise.$

- Step 4: repeat above two steps until stage 1 is reached.

In practice, a treatment regime would be applied beginning at stage 1, and using $\hat{\boldsymbol{\psi}}_1^{\top}$, then the second stage using $\hat{\boldsymbol{\psi}}_2^{\top}$, until the final stage K using $\hat{\boldsymbol{\psi}}_K^{\top}$. That is, the estimated optimal DTR is $\hat{d}^{opt} = \{\hat{d}_1(\boldsymbol{h}_1), ..., \hat{d}_j(\boldsymbol{h}_j), ..., \hat{d}_K(\boldsymbol{h}_K)\}^{opt} = \{\mathbb{I}(\hat{\boldsymbol{\psi}}_1^{\top} \boldsymbol{h}_1 > 0), ..., \mathbb{I}(\hat{\boldsymbol{\psi}}_j^{\top} \boldsymbol{h}_j > 0), ..., \mathbb{I}(\hat{\boldsymbol{\psi}}_K^{\top} \boldsymbol{h}_K > 0)\}$. An illustration of Q-learning in a three-stage example is presented in Appendix A, and this example can help clarify the whole processes of regression-based approaches. Q-learning has certain advantages. For example, it is simple to implement in standard software and easy to explain to clinical collaborators. Additionally, Q-learning can easily be combined with continuous treatments (doses) and also with many useful tools such as standard variable selection methods (e.g., Likelihood Ratio Testing) and residual diagnostics.

However, optimal DTRs estimated with Q-learning can be very poor if the outcome regression model is misspecified. Compared with the direct search approach, the objective function of Q-learning is for model fitting but not directly for value maximization. Moreover, Q-learning is also associated with certain challenges: the main one generally arises during optimal DTR inference because of the nonsmoothness of the maximum operator when pseudo-outcomes are being constructed. Identifying the optimal treatment rules necessitates using a max operation; thus, analysing all the large sample properties of corresponding estimators is subject to this requirement. Specifically, due to need for the maximum operator, the pseudo-outcomes may not be normally distributed even if all original variables are multivariate normally distributed. In addition, there exist some individuals who do not respond to latter stage treatments; in which case, the statistical inference for earlier stages' parameters from Q-learning is non-standard (Chakraborty and Moodie [2013]). Some work has been proposed to address these issues, such as robust Q-learning (Ertefaie

et al. [2021]), and inference for non-regular parameters in DTRs (Chakraborty et al. [2010], Chakraborty et al. [2013]).

2.2.2 Dynamic Weighted Ordinary Least Squares

A major limitation of Q-learning is that it requires correct specification of all the Q-functions; however, other methodologies have been developed that allow more flexibility in modeling. One example is dynamic weighted ordinary least squares. In the following parts, we first introduce the dWOLS model in the single decision problem, and then present dWOLS in a multi-stage case.

For the single-stage model, we let $\boldsymbol{x}_i = (x_{1i}, x_{2i}, \ldots, x_{Pi})$ be a vector of P covariate variables of the individual i. Suppose that the true outcome model can be split into two components and be semiparametric such as $\mathbb{E}\left[Y_i^*(a_i) \mid \boldsymbol{x}_i\right] = f\left(\boldsymbol{x}_i^{\beta}; \boldsymbol{\beta}\right) + \boldsymbol{\psi}^{\top} a_i \boldsymbol{x}_i^{\psi}$, for some function f and unknown parameter $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$, where i denotes individual. One is the f function, which we usually call the *treatment-free function*, and it is the expected outcome under no treatment. The second is the gamma function, which we commonly call the *blip function*, and it is the impact of treatment on outcome. $\boldsymbol{x}_i^{\beta}$ and $\boldsymbol{x}_i^{\psi}$ are subsets of $\boldsymbol{x}_i$, and $\boldsymbol{x}_i^{\psi}$ in the blip function often termed the *prescriptive* or *tailoring* variables. Note that the treatment-free function has no relationship with treatment; therefore, to choose the optimal treatment, we really need to choose a treatment to maximize the blip function.

Dynamic weighted ordinary least squares (Wallace and Moodie [2015]) is applied by specifying three models: (1) the treatment-free model: $f(\boldsymbol{x}^{\beta}; \boldsymbol{\beta})$; (2) the blip model: $\gamma(\boldsymbol{x}^{\psi}, a; \boldsymbol{\psi}) = a\boldsymbol{\psi}^{\top}\boldsymbol{x}^{\psi}$; (3) the treatment model: $\mathbb{E}[A|\boldsymbol{x}; \boldsymbol{\alpha}]$, which is the propensity score model (Rosenbaum and Rubin [1983]) when the treatment is a binary variable.

Under dWOLS, we estimate $\boldsymbol{\psi}$ via weighted ordinary least square of outcome on covariates in treatment-free and blip models, with balancing weights, for example, $w = |a - \mathbb{E}[A|\boldsymbol{x}; \hat{\boldsymbol{\alpha}}]|$; (the balancing weights will be discussed later in this section). The fundamental property of dWOLS is *double robustness* (Wallace and Moodie [2015]); that is, parameters $\boldsymbol{\psi}$ can be consistently estimated if at least one of the treatment or treatment-free models is correctly specified.

dWOLS draws heavily on the method of G-estimation (Robins [2004]), with both approaches grounded in structural nested mean models (SNMMs, Robins [1994]). In multi-stage settings, an SNMM is used to model the effect of a treatment at each stage as a function of the available patient history. This is called the blip function (Robins [2004], Chakraborty and Moodie [2013]):

Definition 2.2. ***Blip function***
A function describing the expected difference in potential outcome when receiving a treatment a_j

at the j^{th} stage ($j < K$) instead of the reference treatment a_j^{ref} (such as a control or standard care, taking the value 0 in our setting) at j^{th} stage and subsequently receiving optimal treatment from the $(j+1)^{th}$ stage, given covariate history $\boldsymbol{h}_j$.

Mathematically, the blip function is defined at each stage $j < K$ as

$$\gamma_j\left(\boldsymbol{h}_j, a_j\right) = \mathbb{E}\left[Y^*(\overline{\boldsymbol{a}}_{j-1},\ a_j,\ \underline{\boldsymbol{a}}_{j+1}^{opt}) - Y^*(\overline{\boldsymbol{a}}_{j-1},\ a_j^{ref},\ \underline{\boldsymbol{a}}_{j+1}^{opt})|\boldsymbol{H}_j = \boldsymbol{h}_j\right],$$

where $Y^*(\overline{\boldsymbol{a}})$ is the potential outcome under treatment sequence $\overline{\boldsymbol{a}}$. A related concept is the *regret* function (Murphy [2003]): $\mu_j\left(\boldsymbol{h}_j, a_j\right) = \mathbb{E}\left[Y^*(\overline{\boldsymbol{a}}_{j-1}\ ,\ \underline{\boldsymbol{a}}_j^{opt}) - Y^*(\overline{\boldsymbol{a}}_j\ ,\ \underline{\boldsymbol{a}}_{j+1}^{opt})|\boldsymbol{H}_j = \boldsymbol{h}_j\right]$, with $\mu_K(\boldsymbol{h}_K, a_K) = \mathbb{E}[Y^*(\overline{\boldsymbol{a}}_{K-1}, a_K^{opt}) - Y^*(\overline{\boldsymbol{a}}_K) \mid \boldsymbol{H}_K = \boldsymbol{h}_K]$, which is the expected loss arising from receiving treatment a_j at the j^{th} stage rather than the optimal treatment a_j^{opt}. Then, the relationship between the blip function and regret function is $\mu_j(\boldsymbol{h}_j, a_j) = \gamma_j(\boldsymbol{h}_j, a_j^{opt}) - \gamma_j(\boldsymbol{h}_j, a_j)$, so the regret function will equal 0 if the individual was optimally treated, and be positive otherwise (representing the expected increase in outcome had the patient received optimal treatment).

Models for the blip and the regret functions are examples of SNMMs and as such are amenable to analysis by several established methods. Typically, the blip model depends on a relatively small subset of the variables in $\boldsymbol{h}_j$, denoted as $\boldsymbol{h}_j^{\psi}$, which is the tailoring variables. The blip model forms one component of the outcome model ($\mathbb{E}\left[Y^*(\boldsymbol{a})|\boldsymbol{H} = \boldsymbol{h}\right] = \sum_{j=1}^{K}[f_j\left(\boldsymbol{h}_j; \boldsymbol{\beta}\right) + \gamma_j(\boldsymbol{h}_j^{\psi}, a_j; \boldsymbol{\psi}_j)]$, where f_j denotes the expected contribution of covariates in the absence of treatment (called the treatment-free model) for stage j. The optimal treatment rule at each decision point is that which maximizes the expected outcome. As the treatment-free model does not depend on treatment, the optimal treatment is therefore that which maximizes the blip (or minimizes the regret), namely: "$a_j^{opt} = 1,\ if\ \gamma_j(\boldsymbol{h}_j^{\psi}, 1; \boldsymbol{\psi}_j) > 0;\ a_j^{opt} = 0,\ otherwise$." Similar to Q-learning, the estimated optimal DTR is $\hat{d}^{opt} = \{\hat{d}_1(\boldsymbol{h}_1), ..., \hat{d}_j(\boldsymbol{h}_j), ..., \hat{d}_K(\boldsymbol{h}_K)\}^{opt} = \{\mathbb{I}(\hat{\boldsymbol{\psi}}_1^{\top}\boldsymbol{h}_1 > 0), ..., \mathbb{I}(\hat{\boldsymbol{\psi}}_j^{\top}\boldsymbol{h}_j > 0), ..., \mathbb{I}(\hat{\boldsymbol{\psi}}_K^{\top}\boldsymbol{h}_K > 0)\}$. It is therefore sufficient to estimate $\boldsymbol{\psi}_j$ (the parameters in the blip); the parameters of the treatment-free model are nuisance parameters.

If the blip and treatment-free models are correctly specified, their parameters may be estimated in the manner of a typical regression problem. Standard methods may also be applied if the treatment and tailoring covariates are independent. The latter condition, though common in randomized trials, is typically problematic in observational studies, for example, when sicker patients are prescribed more aggressive therapies. However, various methods - including dWOLS and G-estimation - offer robustness to mis-specification of the treatment-free component even when treatments are covariate dependent. In addition to the blip and treatment-free models, these methods require specification of a third model - the treatment model. This establishes the

relationship between treatment and covariates: $\mathbb{E}\left[A_j|\boldsymbol{H}_j = \boldsymbol{h}_j; \boldsymbol{\alpha}_j\right]$.

Estimation via dWOLS follows a recursive procedure, one stage at a time. Beginning at the final stage (stage K), and working backwards to stage 1, at each stage (e.g., stage j), we specify the blip model such as a linear model $\gamma_j\left(\boldsymbol{h}_j, a_j; \boldsymbol{\psi}_j\right) = \boldsymbol{\psi}_j^\top a_j \boldsymbol{h}_j^\psi$, and the treatment-free model (such as a linear model again, which could also be a more general function of the parameter vector $\boldsymbol{\beta}_j$) : $f_j(\boldsymbol{h}_j; \boldsymbol{\beta}_j) = \boldsymbol{\beta}_j^\top \boldsymbol{h}_j^\beta$, where vector $\boldsymbol{h}_j^\beta$ (often termed *predictive variables*) is a subset of the history $\boldsymbol{h}_j$. Then, dWOLS proceeds as follows:

- Step 1: Construct the stage j *pseudo-outcome*: set $\widetilde{\mathcal{Y}_j} = y$ if $j = K$, otherwise, use prior estimates $\underline{\hat{\boldsymbol{\psi}}}_{j+1} = (\hat{\boldsymbol{\psi}}_{j+1}, ..., \hat{\boldsymbol{\psi}}_K)$ to form $\widetilde{\mathcal{Y}_j} = y + \sum_{k=j+1}^K \mu_k(\boldsymbol{h}_k, a_k; \hat{\boldsymbol{\psi}}_k)$.

- Step 2: Specify the stage j treatment model $\mathbb{E}\left[A_j|\boldsymbol{h}_j; \boldsymbol{\alpha}_j\right]$. The treatment model parameters $\boldsymbol{\alpha}_j$ (estimated, for example, via logistic regression) are used to compute a weight w_j, such as $w_j = \left|a_j - \mathbb{E}\left[A_j|\boldsymbol{h}_j; \hat{\boldsymbol{\alpha}}_j\right]\right|$.

- Step 3: Specify the stage j treatment-free and blip models, and conduct WOLS to estimate $\hat{\boldsymbol{\psi}}_j$ and form treatment-rules: regress $\widetilde{\mathcal{Y}_j}$ on $(\boldsymbol{h}_j^\beta, \boldsymbol{h}_j^\psi)$ with weights w_j to find parameter estimators $\hat{\boldsymbol{\psi}}_j$ to construct the j^{th} stage optimal treatment rule, which is $prescribe\ a_j = 1\ if\ \hat{\boldsymbol{\psi}}_j^\top \boldsymbol{H}_j^\psi > 0;\ then\ prescribe\ a_j = 0\ otherwise.$

- Step 4: Return to Step 1 and analyze stage $j-1$ if there are more stages to analyze.

Assuming that the blip model is linearly set and correctly (or at worst, over-) specified, dWOLS leads to *doubly robust* blip parameter estimators: if at least one of the treatment or treatment-free models is correctly specified, the blip parameter estimators will be consistent. This property is a consequence of the weight function establishing covariate balance: in effect removing any dependence between treatment and covariates in the weighted dataset (Wallace and Moodie [2015]). As a sequence of weighted regressions, dWOLS is easily implemented in most software environments (with a dedicated package in R, Wallace et al. [2017b]).

The double robustness property of dWOLS is appealing, particularly in the case of randomized trials data, where the treatment model will often be known precisely. We illustrate this property with the single decision case. To overcome the misspecification of the treatment-free model, the crucial part of dWOLS is to identify the weights for the weighted least squares. Wallace and Moodie [2015] has proposed a covariate balancing theorem for the weights. We denote the propensity score as $\pi(\boldsymbol{x}) := \mathbb{P}(A = 1 \mid \boldsymbol{x})$, and suppose the weight $w(a, \boldsymbol{x})$ is a function of $A = a$

and $\boldsymbol{x}$. Then, a weighted least squares regression of Y on $(\boldsymbol{x}^{\beta}, \boldsymbol{a}\boldsymbol{x}^{\psi})$ will yield consistent estimates of $\boldsymbol{\psi}$ if the weights satisfy

$$\pi(\boldsymbol{x})w(1,\boldsymbol{x}) = (1-\pi(\boldsymbol{x}))w(0,\boldsymbol{x}). \tag{2.2}$$

As indicated previously, the double robustness of dWOLS relies on that of G-estimation, which was originally developed in causal inference for longitudinal data where the focus was on SNMM. One crucial condition of the consistency estimation of $\boldsymbol{\psi}$ from G-estimation is no unmeasured confounding (NUC). NUC guarantees the independence of treatment and potential outcomes; thus, from the generalized estimating equation (GEE) in G-estimation, because of the (conditional) independence under UNC, the expectation of the whole GEE is the product of two independent parts of GEE, one part related to a (possibly counterfactual) treatment-free outcome, the other related to the treatment. Therefore, the GEE in the G-estimation is unbiased if either the treatment-free or treatment model is correctly specified. The rationale behind this is that the estimating functions are obtained from semiparametric efficiency theory and they turn out to be some influence function for parameters of interest when either model is correctly specified (Tsiatis [2006], Kennedy [2017]). Moreover, note that, in G-estimation, the treatment (propensity score) model does not necessarily need to be correctly specified in the sense of capturing the data-generating mechanism, but rather it must contain and model correctly the impact of all confounding variables (Chakraborty and Moodie [2013]). Relying on the NUC condition and recasting G-estimation in a WOLS manner, Wallace and Moodie [2015] provided a class of weights, which are associated with the treatment part in the GEE of G-estimation, and these weights can balance the covariate. Wallace and Moodie [2015] presented a heuristic proof for this theorem. Building on this proof, in Chapter 3, we extend the theorem and proof in the presence of interference.

Our first new proof for the balancing weights used in dWOLS is based on the algebra of the partition regression. Firstly, we note that if the treatment-free model, where $f(\boldsymbol{x}^{\beta};\boldsymbol{\beta}) = \boldsymbol{\beta}^{\top}\boldsymbol{x}^{\beta}$ in the linear case, is correctly identified, we can acquire unbiased estimators of $\boldsymbol{\psi}$ (i.e., $\mathbb{E}[\hat{\boldsymbol{\psi}}] - \boldsymbol{\psi} = \mathbf{0}$) via weighted linear regression for any set of weights. We then show that if we apply weights that satisfy $\pi(\boldsymbol{x})w(1,\boldsymbol{x}) = (1-\pi(\boldsymbol{x}))w(0,\boldsymbol{x})$, we also have $\mathbb{E}[\hat{\boldsymbol{\psi}}] - \boldsymbol{\psi} = \mathbf{0}$, even if the treatment-free model is incorrectly specified. During this process, considering weighted OLS system equations, we first derive conditional expectation of $\hat{\boldsymbol{\beta}}$ (i.e., $\mathbb{E}[\hat{\boldsymbol{\beta}} \mid a, \boldsymbol{x}]$), and then $\mathbb{E}[\hat{\boldsymbol{\psi}} \mid a, \boldsymbol{x}]$. Further, taking expectation of $\mathbb{E}[\hat{\boldsymbol{\psi}} \mid a, \boldsymbol{x}]$ with respect to the distribution of A given $\boldsymbol{x}$, we have the expectation of $\hat{\boldsymbol{\psi}}$, conditional on $\boldsymbol{x}$; thus, we derive the bias of $\hat{\boldsymbol{\psi}}$ (i.e., $\mathbb{E}[\hat{\boldsymbol{\psi}}] - \boldsymbol{\psi}$). Finally, the key factor in the bias of $\hat{\boldsymbol{\psi}}$ is denoted as $\mathcal{E} = \{\mathcal{B}(\boldsymbol{x})(I - \boldsymbol{x}^{\beta}[\mathcal{A}(\boldsymbol{x})\boldsymbol{x}^{\beta}]^{-1}\mathcal{A}(\boldsymbol{x}))\}f(\boldsymbol{x}^{\beta},\boldsymbol{\beta})$, that is, (B.7) in Appendix B, where $\mathcal{B}(\boldsymbol{x}) = (\boldsymbol{x}^{\psi})^{\top}R(\boldsymbol{x})$, $\mathcal{A}(\boldsymbol{x}) = (\boldsymbol{x}^{\beta})^{\top}R(\boldsymbol{x})[2I - \mathcal{P}^{\psi}R(\boldsymbol{x})]$ and $\mathcal{P}^{\psi} = \boldsymbol{x}^{\psi}[(\boldsymbol{x}^{\psi})^{\top}R(\boldsymbol{x})\boldsymbol{x}^{\psi}]^{-1}(\boldsymbol{x}^{\psi})^{\top}$ are expressions that are related to $R(\boldsymbol{x})$, so that $\hat{\boldsymbol{\psi}}$ is unbiased if and only if $\mathcal{E} = \mathbf{0}$. The details of this proof are provided in Appendix B. There we apply the spirit of the Frisch-Waugh-Lovell Theorem that is often employed in econometrics, and use an expectation of ratio-approximation techniques. We note that this proof lays the foundation for

studying dWOLS in the presence of interference (i.e., as in Chapter 3), in particular, a foundation for the Invalid dWOLS Theorem (E.1) in Appendix E.

In addition, in Appendix C, we also demonstrate the double robustness of dWOLS from the estimation equation perspective. In the proof, we study the estimation equations systems of dWOLS, and through solving the system of estimating equations, we solve for estimators $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}})$. We show that $\hat{\boldsymbol{\psi}}$ solving the system of estimating equations of dWOLS is a consistent estimator for the true value $\boldsymbol{\psi}$, if at least one of either the treatment-free model or treatment model is correctly specified. This perspective and thought process inspire us to propose doubly robust estimation methods for extending dWOLS to the case of binary outcomes (i.e., as in Chapter 4). Note that, based on the strong heredity principle (Chipman [1996]), dWOLS requires that the treatment-free model must include the main effects for all covariates in the blip model, that is, the tailoring variables should be the subset of the predictive variables ($\boldsymbol{x}^{\psi} \subseteq \boldsymbol{x}^{\beta}$). In Appendix C, we separately consider two different cases: (1) $\boldsymbol{x}^{\psi} = \boldsymbol{x}^{\beta} = \boldsymbol{x}$, and (2) $\boldsymbol{x}^{\psi} \subset \boldsymbol{x}^{\beta} = \boldsymbol{x}$.

Furthermore, before finishing this chapter, we introduce a theorem that identifies the optimal dWOLS weights that yield the minimum variance of blip-parameter (i.e., $\boldsymbol{\psi}$) estimators. Consistent with the theorem framework and formats in Wallace and Moodie [2015], we have

Theorem 2.1. *The optimal dWOLS balancing weights*
A weighted least squares regression of Y on $(\boldsymbol{x}^{\beta}, a\boldsymbol{x}^{\psi})$ will yield the most efficient consistent estimates of $\boldsymbol{\psi}$ if the weights satisfy

$$w(1, \boldsymbol{x}) \propto [1 - \pi(\boldsymbol{x})], \textit{and } w(0, \boldsymbol{x}) \propto \pi(\boldsymbol{x}). \tag{2.3}$$

Proof of Theorem 2.1: See Appendix D.

One option for the optimal weight in (2.3), $w(1, \boldsymbol{x}) = [1 - \pi(\boldsymbol{x})], \textit{and } w(0, \boldsymbol{x}) = \pi(\boldsymbol{x})$, called the "overlap weight", has been extensively discussed by Li et al. [2018]. Overlap weights satisfy the dWOLS balancing weights criterion (2.2), and have three main advantages that are presented in Li et al. [2018]: (1) they include relatively few extreme weights; (2) there is no need to choose an artificial cutoff point when the weight values have to be trimmed; and (3) they provide the minimum variance of the sample estimator of weighted ATE among all balancing weights (Theorem 2 and Corollary 1 in Li et al. [2018]). Demonstrating via simulation studies, Wallace and Moodie [2015] suggested the use of "absolute value" weights for efficient estimation. These weights, in the form $w = |A - \mathbb{E}(A \mid \boldsymbol{x})|$, are identical to the overlap weights. That is, when $A = 1$, $w = |A - \mathbb{E}(A \mid \boldsymbol{x})| = 1 - \pi(\boldsymbol{x})$, and when $A = 0$, $w = |A - \mathbb{E}(A \mid \boldsymbol{x})| = |0 - \pi(\boldsymbol{x})| = \pi(\boldsymbol{x})$. Moreover, in the X-learning, which was introduced as a CATE method at the beginning of this chapter, Künzel et al. [2019] also observed that the propensity score worked fairly well as a weight if the number of untreated individuals is much larger or smaller than the number of treated

individuals. Recall that the X-learner is a combination of the two different estimators for the CATE, and includes a weight of w: $w\hat{\tau}_0(\boldsymbol{x}) + (1-w)\hat{\tau}_1(\boldsymbol{x})$, where $\hat{\tau}_0(\boldsymbol{x})$ is one CATE estimator that corresponds to the untreated group, and $\hat{\tau}_1(\boldsymbol{x})$ is the other estimator and is based on the treated group. If $w = \pi(\boldsymbol{x})$, as suggested by Künzel et al. [2019], then the X-learning CATE estimator is $\pi(\boldsymbol{x})\hat{\tau}_0(\boldsymbol{x}) + [1-\pi(\boldsymbol{x})]\hat{\tau}_1(\boldsymbol{x})$. That is, the X-learner that employs the overlap weights is recommended by Künzel et al. [2019].

Theorem 2.1 provides the rationale behind the choice of optimal weights for dWOLS and guarantees that the variance of $\hat{\boldsymbol{\psi}}$ is the smallest, that is, corresponding to the asymptotically efficient estimator. When we employ the dWOLS method for DTR estimation, we can simply analyze the overlap weights instead of considering a family of dWOLS weights. In the proof of Theorem 2.1, the variance of $\hat{\boldsymbol{\psi}}$ was derived using a matrix formulation of the weighted OLS. The weights satisfy weight criterion (2.2), the biases of these $\hat{\boldsymbol{\psi}}$ are zero, and thus the optimal weights corresponds to the one that involves the smallest variance of $\hat{\boldsymbol{\psi}}$. Accordingly, to find the optimal weights, we use the Cauchy-Schwarz inequality to find the lower bound of the variance of $\hat{\boldsymbol{\psi}}$, and then identify the corresponding weights.

In this chapter, we have provided a general review of optimal DTR estimation, starting from single-stage treatment decisions and moving to multiple-stage ones. In particular, this chapter presented the Q-learning and dWOLS methods and discussed the theoretical foundations of dWOLS from two different perspectives: direct proof of dWOLS in Appendix B, and estimation equation proof of dWOLS in Appendix C. We provided a theorem that identifies the optimal dWOLS weights, and showed that the overlap weights (or absolute value weights) are one choice of the optimal dWOLS weights. If at least one of the treatment-free or treatment model is correctly specified, then with the optimal weights, dWOLS provides consistent estimation with lowest variance among all the standard dWOLS weights. These contributions build a solid basis for investigating dWOLS in the presence of interference.

Chapter 3

Optimal DTR Estimation in the Presence of Interference

Building on the Q-learning and dWOLS methods introduced in Chapter 2, in Chapter 3, we explore the estimation of DTRs in the presence of interference. In the presence of interference, our first step is to examine Su et al. [2019]'s Q-learning methods and to propose a weighted version of their methods. Then, focusing on dWOLS, we investigate the impact of interference on analyses where it is ignored, and demonstrate how dWOLS may be used to account for it. We also develop a modification to dWOLS based on network propensity functions, which can offer modeling flexibility and efficiency gains compared to the original dWOLS methodology in this setting. We then apply our work to a real-world study of smoking cessation: the Population Assessment of Tobacco and Health study.

This chapter is organized as follows. In Section 3.1, we introduce recent advances in precision medicine under interference, as well as causal inference assumptions under interference. Next, in Section 3.2, we examine the Q-learning method under interference. In Section 3.3, we highlight the impact of interference on estimation using dWOLS and propose various solutions, including our extension based on network propensity functions. Three simulation studies, described in Section 3.4, investigate the consequences of failing to account for interference and demonstrate how interference may be accommodated in dWOLS analysis. In Section 3.5, we illustrate our methodology using data from the PATH study. We conclude this chapter with a discussion in Section 3.6.

3.1 Preliminaries

Various researchers have studied violations of the no-interference assumption in broader contexts (Hayes et al., 2000; Sobel, 2006; Tchetgen and VanderWeele, 2012; Leung, 2020), such as diverse interference studies in Section 1.2. However, despite some recent increased awareness of interference issues within the specific context of precision medicine or DTRs, for example, Su et al., 2019 and Sherman et al., 2020, there is little methodological work addressing such problems. In the interference-aware DTRs framework, treatments are already dependent on patient-level characteristics, and so accounting for the treatments of those *connected* to each patient provides an additional - but related - challenge. Su et al. [2019] assume treatments are independently assigned between individuals. To characterize the network interference, they consider neighbours' covariates, treatment, and covariate-treatment interactions at a single decision point of the Q-learning and A-learning models; the outcome considered is a network mean. Their result, using a double summation switching technique, is that under their model the optimal treatment at the individual level is dependent only on that individual's own covariates and the degree of that individual (number of neighbours). Because their models do not include interactions between others' treatment assignments and one's own treatment assignment, the optimal treatment rules also do not rely on the neighbours' treatments. Su et al. [2019]'s Q-learning model will be discussed in Section 3.2 and a weighted version of their model will be investigated in the same section. In another approach, assuming a treatment is randomized, and depends on the unit's own and neighbour's covariates, and assuming network exogeneity, Viviano [2020] proposed a methodology for estimating treatment allocation policy under network interference. This work considered a semi-parametric model to estimate a network's empirical welfare, which is a function of policy. Additionally, to identify the optimal policy, an exact optimization algorithm was employed to solve optimization questions with capacity constraints.

Following standard network analysis convention, a network is a set of nodes connected by edges and is usually represented by a graph. The total number of nodes is the size of the network, and the degree of a node is the number of connections it has to other nodes. Building on this framework, we will establish new approaches within an undirected network G in which individuals (or patients) whose covariates are independent and identically distributed are indexed by i. We denote the set of neighbours of node (individual) i in the network G by $\mathcal{N}_i$ and let $\mathbf{x}_{\mathcal{N}_i}$ consist of the covariate values of the neighbours of node i. We also establish four network interference assumptions for $i = 1, ..., n$:

Interference Assumption 1. *The distribution of potential outcomes of a unit i, conditional on the treatments on all units, depends only on its treatment and the treatments of units in its neighbourhood $\mathcal{N}_i$ (Forastiere et al., 2021), that is, $Y_i^*(\boldsymbol{A} = \boldsymbol{a}) = Y_i^*(\boldsymbol{A} = \boldsymbol{a}^\dagger)$ for all $(\boldsymbol{a}, \boldsymbol{a}^\dagger)$, such that $a_i = a_i^\dagger$ and $a_l = a_l^\dagger$ for all $l \in \mathcal{N}_i$. Under this neighborhood interference assumption, the*

potential outcome of a unit i can be denoted as $Y_i^*(\boldsymbol{A}) = Y_i^*(A_i, A_{\mathcal{N}_i})$*, a function of both a scalar* A_i *and a vector* $A_{\mathcal{N}_i}$*, where* A_i *is the treatment of unit i, and* $A_{\mathcal{N}_i}$ *represents treatments of units in the neighbourhood of unit i.*

Interference Assumption 2. *No Unmeasured Confounding (NUC) of individual and neighborhood treatment:* $Y_i^*(a_i, a_{\mathcal{N}_i}) \perp\!\!\!\perp (A_i, A_{\mathcal{N}_i}) \mid (\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i})$.

Interference Assumption 3. *Stability (or Consistency):* $Y_i^{obs}(a_i, a_{\mathcal{N}_i}) = \mathbb{I}(A_i = a_i, A_{\mathcal{N}_i} = a_{\mathcal{N}_i}) Y_i^*(A_i, A_{\mathcal{N}_i})$.

Interference Assumption 4. *Positivity:* $\mathbb{P}[A_i = a_i, A_{\mathcal{N}_i} = a_{\mathcal{N}_i} \mid (\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i})] > 0$, *for all* $(A_i, A_{\mathcal{N}_i})$ *such that* $\mathbb{P}[(\boldsymbol{X}_i, \mathbf{X}_{\mathcal{N}_i}) = (\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i})] > 0$.

Under these four network interference assumptions, we have $\mathbb{E}[Y^*(a, a_{\mathcal{N}}) \mid (\boldsymbol{x}, \mathbf{x}_{\mathcal{N}})] = \mathbb{E}[Y \mid (\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}), (A, A_{\mathcal{N}}) = (a, a_{\mathcal{N}})]$. When potential outcomes depend on neighbours' treatment, then the optimal DTR that maximizes the outcomes will rely on neighbours' treatment. Thus, the *value function* – the expected outcome that is achieved applying a certain treatment regime, namely, in a single-stage decision setting, $d(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}, a_{\mathcal{N}})$ that maps covariates $(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}})$ and neighbours' treatments $(a_{\mathcal{N}})$ to a treatment configuration $(a, a_{\mathcal{N}})$ – is $\mathcal{V}\left[d(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}, a_{\mathcal{N}})\right] = \mathbb{E}\left(Y^*[d(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}, a_{\mathcal{N}})]\right) = \mathbb{E}\left[Y^*(a, a_{\mathcal{N}})\right]$. Again, our primary goal is identifying the optimal DTRs, that is an optimal regime $d^{opt} \in \mathcal{D}$ satisfies that (e.g., for a single-stage case, just a piece of decision rule)

$$d^{opt}(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}, a_{\mathcal{N}}) = \arg\max_{d \in \mathcal{D}} \mathbb{E}\left\{Y^*[d(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}, a_{\mathcal{N}})]\right\} := \arg\max_{d \in \mathcal{D}} \mathcal{V}[d(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}, a_{\mathcal{N}})].$$

That is, we want to identify a treatment regimen that maximizes the objective function, which is the expectation of potential outcome and denoted as $\mathcal{V}(d)$. Therefore, an optimal treatment regime is a function that maps the covariates and neighbours' treatments, that is, $(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}, a_{\mathcal{N}})$, to an optimal treatment configuration $(a, a_{\mathcal{N}})^{opt}$: $d^{opt}(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}) = (a, a_{\mathcal{N}})^{opt}$. The corresponding value function of d^{opt} is $\mathcal{V}[(a, a_{\mathcal{N}})^{opt}] = \mathbb{E}[Y^*[(a, a_{\mathcal{N}})^{opt}]] = \mathbb{E}\left[\mathbb{E}[Y^*[(a, a_{\mathcal{N}})^{opt}] \mid (\boldsymbol{x}, \mathbf{x}_{\mathcal{N}})]\right] = \mathbb{E}\left[\max_{(a, a_{\mathcal{N}})} \mathbb{E}[Y^*(a, a_{\mathcal{N}}) \mid (\boldsymbol{x}, \mathbf{x}_{\mathcal{N}})]\right] = \mathbb{E}\left[\max_{(a, a_{\mathcal{N}})} \mathbb{E}[Y \mid (\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}), \mathbf{A} = (a, a_{\mathcal{N}})]\right]$.

It is simpler to assume that we are optimizing for the individual (the ego), rather than for both the ego and the neighbours. That is, the goal is to identify the optimal treatment for the ego only, and we assume that the neighbours' treatments are pre-determined and thus fixed throughout the study period. We term this assumption as the *neighbours' invariant treatments (NIT)* assumption. Under NIT, the corresponding optimal value function is then $\mathcal{V}[(a, a_{\mathcal{N}})^{opt}] = \mathbb{E}\left[\max_{(a, a_{\mathcal{N}})} \mathbb{E}[Y \mid (\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}), \mathbf{A} = (a, a_{\mathcal{N}})]\right] = \mathbb{E}\left[\max_{a} \mathbb{E}[Y \mid (\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}), \mathbf{A} = (a, a_{\mathcal{N}})]\right]$. As a result, in this single-stage case, treatment decision rules for egos will depend both on their own covariates,

and their neighbours' covariates as well as the treatments. Looking ahead to the multi-stage case, each decision rule of a DTR will also have such covariates and neighbours' treatment information as an input. Building on these important background details, we now formally define a dynamic treatment regime in the presence of interference. We adopt notation that generalizes the covariate to a "history" variable $\boldsymbol{h}$, which also possibly include neighbours' history information, and continue to suppose that there are two treatment options in set $\mathscr{A} = \{0, 1\}$. In the presence of interference, an treatment regimen d are sequence decision rules that each of them maps an individual's history and their neighbours' history and treatments to a treatment option in a set of available treatments. Formally, under the invariant treatments of neighbours assumption, a dynamic treatment regime in the presence of interference is defined as follows.

Definition 3.1. ***Dynamic Treatment Regime with Interference***
With K treatment decision points, where $K \geq 1$, at decision stage j, a treatment decision rule $d_j(\boldsymbol{h}_j, A_{\mathcal{N}})$ is a function that maps an individual's, or possibly their neighbours', history and their neighbours' treatments to a treatment option in $\mathscr{A}_j$, for $j = 1, ..., K$. Then, a dynamic treatment regime $d \in \mathcal{D}$ is a sequence of such treatment decision rules to be applied in order; that is, with K decision points:

$$d = \{d_1(\boldsymbol{h}_1, A_{\mathcal{N}}), ..., d_j(\boldsymbol{h}_j, A_{\mathcal{N}}), ..., d_K(\boldsymbol{h}_K, A_{\mathcal{N}})\}.$$

We will study interference in the following part of Chapter 3 in a simpler case where we optimize for the ego, but not for the ego and its neighbours. Thus we assume that the neighbours' treatments are fixed throughout the study period. Building on this invariant treatments of neighbours assumption, we will devise interference-aware DTR estimation approaches. Extending this assumption, furthermore, in Chapter 5, we will examine household interference and optimize for a pair in the same household, that is, optimizing for both ego and alter.

3.2 Q-learning with Interference

In the previous section, we have built extended causal assumptions under interference, and have discussed that, under the interference scenario, the optimal treatment regimes depend on both the individual's information and their neighbours' treatments. We also rigorously defined DTR in the presence of interference. In this section, we will investigate Q-learning in the presence of interference.

3.2.1 Methodology

We suppose independent treatments with the same treatment rule applying to all individuals, and suppose that our goal is to identify an optimal regime $d^{opt} \in \mathcal{D}$ that satisfies $d^{opt} = \arg\max_{d\in\mathcal{D}} \mathbb{E}\{Y^*(d)\}$, and this goal is equivalent to finding the optimal treatment that maximizes $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left\{Y_i^*\left[d\left(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}, a_{\mathcal{N}_i}\right), a_{\mathcal{N}_i}\right] \mid \boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}\right\}$, where $i = 1, 2, ...n$ are egos. For a single-decision problem, under four network interference assumptions in Section 3.1, our goal is to identify the treatment that maximizes

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left\{Y_i^*\left[a_i, a_{\mathcal{N}_i}\right] \mid \boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}, a_i = d\left(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}, a_{\mathcal{N}_i}\right), A_{\mathcal{N}_i} = a_{\mathcal{N}_i}\right\}. \tag{3.1}$$

Generally, for the K- stage treatment decisions, the Q functions of Q-learning with network interference are defined as:

$$Q_K\left(\boldsymbol{h}_K, a_K, a_{\mathcal{N}}^K\right) = \mathbb{E}\left[Y \mid \boldsymbol{h}_K, a_K, a_{\mathcal{N}}\right];$$

and for $j < K$,

$$Q_j\left(\boldsymbol{h}_j, a_j, a_{\mathcal{N}}^j\right) = \mathbb{E}\left[\max_{A_{j+1}} Q_{j+1}\left(\boldsymbol{h}_{j+1}, A_{j+1}, A_{\mathcal{N}}^{j+1}\right) \mid \boldsymbol{h}_j, a_j, a_{\mathcal{N}}\right],$$

where $a_{\mathcal{N}}^j$ denotes the neighbours' treatments at j^{th} stage for $j = 1, 2, ...K$, and $a_{\mathcal{N}}^j = a_{\mathcal{N}}$ is based on the assumption that the neighbours' treatments are fixed throughout the study period. The use of network-based regression models in a single-decision setting of Q-learning with interference was pioneered by Su et al. [2019]. Their model can be expressed as:

$$\begin{aligned}\mathbb{E}[Y_i \mid \boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}, a_i, a_{\mathcal{N}_i}] =& \alpha + \boldsymbol{\beta}^\top \boldsymbol{x}_i^\beta + \gamma_1 \sum_{l\neq i}^{n} N_{il}\boldsymbol{\beta}^\top \boldsymbol{x}_l^\beta + \eta a_i + \gamma_2 \sum_{l\neq i}^{n} N_{il}\eta a_l \\ &+ a_i \boldsymbol{\psi}^\top \boldsymbol{x}_i^\psi + \gamma_3 \sum_{l\neq i}^{n} N_{il} a_l \boldsymbol{\psi}^\top \boldsymbol{x}_l^\psi,\end{aligned} \tag{3.2}$$

where $N_{il} = 1$ if i and l are connected in a social network and $N_{il} = 0$ otherwise, and $\boldsymbol{x}^\beta$ and $\boldsymbol{x}^\psi$ are sub-vectors of $\boldsymbol{x}$, of lengths p_β and p_ψ, respectively.

According to Su et al.'s model (3.2), after some algebra for the objective function (3.1), the goal is to identify an optimal treatment that maximizes:

$$\frac{1}{n}\sum_{i=1}^{n}\left(\alpha+\boldsymbol{\beta}^{\top}\boldsymbol{x}_i^{\beta}+\gamma_1\sum_{l}^{n}N_{il}\boldsymbol{\beta}^{\top}\boldsymbol{x}_l^{\beta}\right)+\frac{1}{n}\sum_{i=1}^{n}d\left(\boldsymbol{x}_i,\mathbf{x}_{\mathcal{N}_i},a_{\mathcal{N}_i}\right)\left[\left(1+\gamma_2\sum_{l}^{n}N_{il}\right)\eta+\left(1+\gamma_3\sum_{l}^{n}N_{il}\right)\boldsymbol{\psi}^{\top}\boldsymbol{x}_i^{\psi}\right].$$

Then, for the binary treatments, the optimal treatment rule for subject i is:

$$d^{opt}\left(\boldsymbol{x}_i,\mathbf{x}_{\mathcal{N}_i},a_{\mathcal{N}_i}\right)=\mathbb{I}\left\{\left[\left(1+\gamma_2\sum_{l}^{n}N_{il}\right)\eta+\left(1+\gamma_3\sum_{l}^{n}N_{il}\right)\boldsymbol{\psi}^{\top}\boldsymbol{x}_i^{\psi}\right]>0\right\}. \tag{3.3}$$

Note that the optimal treatment rule in equation (3.3) depends on $\boldsymbol{x}_i$, and $\mathbf{x}_{\mathcal{N}_i}$, but not $a_{\mathcal{N}_i}$. Thus, for Su et al.'s model (3.2), the optimal decision rule is $d^{opt}\left(\boldsymbol{x}_i,\mathbf{x}_{\mathcal{N}_i},a_{\mathcal{N}_i}\right)=d^{opt}\left(\boldsymbol{x}_i,\mathbf{x}_{\mathcal{N}_i}\right)$, with parameters γ_2, η, γ_3, and $\boldsymbol{\psi}$. Then, the estimation of these parameters is as follows.

For model (3.2), we can also write

$$\epsilon_i=Y_i-\alpha-\boldsymbol{\beta}^{\top}\boldsymbol{x}_i^{\beta}-\gamma_1\boldsymbol{\beta}^{\top}\sum_{l\neq i}N_{il}\boldsymbol{x}_l^{\beta}-\eta a_i-\gamma_2\eta t_i-a_i\boldsymbol{\psi}^{\top}\boldsymbol{x}_i^{\psi}-\gamma_3\boldsymbol{\psi}^{\top}\sum_{l\neq i}N_{il}a_l\boldsymbol{x}_l^{\psi},$$

where $t_i=\sum_{l=1}^{n}N_{il}a_l$, the number of treated neighbours of the ego. With weighted least squares, the normal equations are proportional to the partial derivatives of $\sum_{i=1}^{n}w_i\epsilon_i^2$ with respect to the parameters or parameter vectors α, $\boldsymbol{\beta}$, η, $\boldsymbol{\psi}$, γ_1, γ_2, and γ_3. In the model, we suggest use the "absolute value" weights, that is, $w=|A-\mathbb{P}(A=1\mid\boldsymbol{x})|$, proposed by Wallace and Moodie [2015]. Nevertheless, any standard dWOLS balancing weight is an option.

Based on Su et al.'s iterative approach, we now propose a weighted iterative method, with the standard dWOLS weights in each least squares regression. From the model (3.2), equivalently, the error, $\epsilon_i=Y_i-\alpha+\boldsymbol{\beta}^{\top}\boldsymbol{x}_i^{\beta}+\gamma_1\sum_{l\neq i}^{n}N_{il}\boldsymbol{\beta}^{\top}\boldsymbol{x}_l^{\beta}+\eta a_i+\gamma_2\sum_{l\neq i}^{n}N_{il}\eta a_l+a_i\boldsymbol{\psi}^{\top}\boldsymbol{x}_i^{\psi}+\gamma_3\sum_{l\neq i}^{n}N_{il}a_l\boldsymbol{\psi}^{\top}\boldsymbol{x}_l^{\psi}$, is not linear in the parameters. It is convenient to solve the normal equations by an iterative method, as follows.

- **Step 1:** Estimate the treatment model parameters $\boldsymbol{\alpha}$, and compute "absolute value" weight:

$$w_i=\left|A_i-\mathbb{E}\left[A_i|\boldsymbol{x}_i;\hat{\boldsymbol{\alpha}}\right]\right|.$$

- **Step 2:** Fix values of γ_1, γ_2, and γ_3, and let

$$u_i^{\beta}=\boldsymbol{x}_i^{\beta}+\gamma_1\sum_{l\neq i}N_{il}\boldsymbol{x}_l^{\beta},\ z_i^{\eta}=a_i+\gamma_2\sum_{l\neq i}N_{il}a_l,\ \textit{and}\ v_i^{\psi}=a_i\boldsymbol{x}_i^{\psi}+\gamma_3\sum_{l\neq i}N_{il}a_l\boldsymbol{x}_l^{\psi}.$$

Write $\epsilon_i = Y_i - \alpha - \boldsymbol{\beta}^\top u_i^\beta - \eta z_i^\eta - \boldsymbol{\psi}^\top v_i^\psi$, and by weighted least squares find $\hat{\alpha}(\gamma)$, $\hat{\boldsymbol{\beta}}(\gamma)$, $\hat{\eta}(\gamma)$ and $\hat{\boldsymbol{\psi}}(\gamma)$.

- **Step 3:** Fix the values of α, $\boldsymbol{\beta}$, η and $\boldsymbol{\psi}$ at these estimated values, and let

$$\mathcal{A}_i = \alpha + \boldsymbol{\beta}^\top \boldsymbol{x}_i^\beta + \eta a_i + a_i \boldsymbol{\psi}^\top \boldsymbol{x}_i^\psi, \ \mathcal{B}_i = \sum_{l \neq i} N_{il} \boldsymbol{\beta}^\top \boldsymbol{x}_l^\beta,$$

$$C_i = \sum_{l \neq i} N_{il} \eta a_l, \ \textit{and} \ \ \mathcal{D}_i = \sum_{l \neq i} N_{il} a_l \boldsymbol{\psi}^\top \boldsymbol{x}_l^\psi.$$

  Write $\epsilon_i = Y_i - \mathcal{A}_i - \gamma_1 \mathcal{B}_i - \gamma_2 C_i - \gamma_3 \mathcal{D}_i$, and by weighted least squares find $\hat{\gamma}_1$, $\hat{\gamma}_2$, $\hat{\gamma}_3$, which are new estimates of parameters γ_1, γ_2, and γ_3, respectively.

- **Step 4:** Repeat these two steps until the estimates converge.

In addition to our weighted iterative method for Su et al.'s model (3.2), we have now proposed Q-learning for pairs case in the presence of household interference.

For the K-stage process, the quality function for a pair case, that is, an ego and alter case, could be defined as

$$Q_K\left(\boldsymbol{h}_K, a_K^{ego}, a_K^{alter}\right) = \mathbb{E}\left[Y \mid \boldsymbol{h}_K, a_K^{ego}, a_K^{alter}\right],$$

and for $j < K$,

$$Q_j\left(\boldsymbol{h}_j, a_j^{ego}, a_j^{alter}\right) = \mathbb{E}\left[\max_{A_{j+1}} Q_{j+1}\left(\boldsymbol{h}_{j+1}, A_{j+1}^{ego}, A_{j+1}^{alter}\right) \mid \boldsymbol{h}_j, A_j^{ego} = a_j^{ego}, A_j^{alter} = a_j^{alter}\right].$$

In the presence of household interference, we propose a linear model for the Q-functions such that:

$$Q_j\left(\boldsymbol{h}_j, a_j^{ego}, a_j^{alter}; \boldsymbol{\beta}_j, \boldsymbol{\psi}_j\right) = \boldsymbol{\beta}_j^\top \boldsymbol{h}_j^\beta + \boldsymbol{\psi}_j^\top a_j^{ego} \begin{pmatrix} \boldsymbol{h}_j^\psi \\ a_j^{alter} \end{pmatrix}, \tag{3.4}$$

a similar process can be defined for the case of the model (3.2), where a_j^{alter} is replaced by t_j, the number of treated neighbours of the ego at j^{th} stage. Following the assumption of invariant treatments of neighbours introduced (i.e., NIT assumption) in Section 3.1, then we have $a_j^{alter} = a^{alter}$ for $j = 1, 2, \ldots K$. It is important to note that, compared with Su et al.'s model (3.2), model (3.4) includes a^{alter} in the interaction terms (i.e., model contains term $a^{ego} a^{alter}$), and thus the decision rules will also depend on a^{alter}.

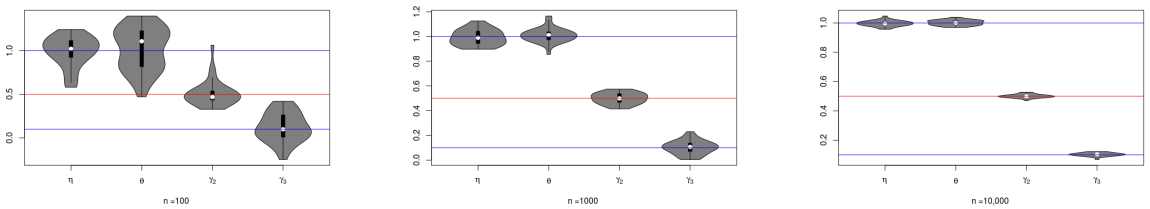


Figure 3.1: Treatment decision parameter (η, ψ, γ_2, and γ_3) estimates for $n = 100$, $n = 1,000$, and $n = 1,000$ simulated datasets under the BA model network.

3.2.2 Simulation Studies

In the previous subsection, Q-learning with network interference and household interference are investigated. In this subsection, we demonstrate Q-learning, for example, model (3.2) under network interference and household interference via simulation. We first consider network interference with a general fixed network, generated from the Barabási-Albert model, then, we examine pairs' (ego and alter) case. Note that in this simulation, we consider the pairs' household interference case as a special case of network interference, that is, each individual only has one neighbour in the network.

Barabási-Albert model generates random scale-free networks, that is, networks for which the degree distribution follows a power law for large k. These scale-free networks are characterized by the presence of large hubs, or nodes that are highly connected to other nodes in the network. In particular, the degree distribution resulting from the BA model is a power law of the form $\mathbb{P}(k) \sim k^{-3}$ (Albert and Barabási, 2002). Many observed networks, such as the world wide web, citation networks, and some social networks are approximated by scale-free networks. Thus, we employ the BA model in this social network simulation studies. In this simulation, we set the power of the BA model as 1, that is, linear preferential attachment, and set $m = 6$, which refers to the number (m) of edges to add in each time step of generating the graph. We generate data as follows:
Patient Information: $X_i \sim U(0, 1.5)$; treatment $\mathbb{P}(A_i = 1 \mid X_i = x_i) = \alpha_0 + \alpha_1 x_i$, where true $\alpha_0 = 0.5,\ and\ \alpha_1 = 0$; outcome $Y_i \sim N(\mu_i, 1)$ where $\mu_i = \alpha + \boldsymbol{\beta}^\top \boldsymbol{x}_i^\beta + \gamma_1 \sum_{l \neq i} N_{il} \boldsymbol{\beta}^\top \boldsymbol{x}_l^\beta + \eta a_i + \gamma_2 \sum_{l \neq i} N_{il} \eta a_l + a_i \boldsymbol{\psi}^\top \boldsymbol{x}_i^\psi + \gamma_3 \sum_{l \neq i} N_{il} a_l \boldsymbol{\psi}^\top \boldsymbol{x}_l^\psi$, where our interest is in estimating the decision parameters (from equation 3.3) $\eta = \psi = 1, \gamma_2 = 0.5, and\ \gamma_3 = 0.1$.

The results of Figure 3.1, under a BA modeled network interference, illustrate that the decision parameters, that is $\eta, \psi, \gamma_2, and\ \gamma_3$, are consistently estimated.

For the pair case, Figure 3.2 shows the Q-learning results under interference. The results in Figure 3.2 show that the decision parameters, that is $\eta, \psi, \gamma_2, and\ \gamma_3$, are consistently estimated.

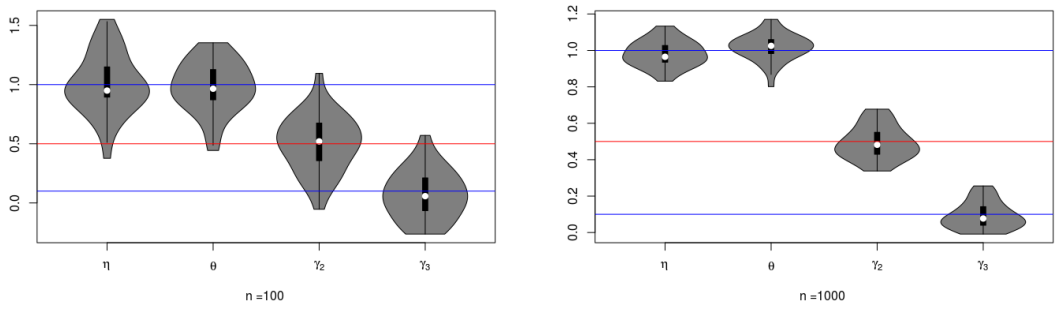


Figure 3.2: Treatment decision parameter estimates for $n = 100$ and $n = 1000$ simulated datasets under the pair case.

Considering network interference, Su et al. proposed a novel single-stage Q-learning method. Building on their models and iterative approach to estimating the parameters in the Q-learning model, we have studied the weighted version of their methods and demonstrated them in our simulation studies. We employed the standard dWOLS weights in each least squares regression. Due to the limitation of Su et al.'s modeling, that is, their model does not take into account interactions between an individual's own treatment assignment and that of their neighbours, the optimal DTR depends only on the individuals and their neighbours' covariates, not on their neighbours' treatments. In addition, Su et al. provide only a single-stage Q-learning method, and do not address the more-complicated multi-stage version. However, in this section, we have proposed a multi-stage Q-learning approach with both network interference and household interference. In particular, examining household interference where there is a pair (ego and alter) case, and aiming to include neighbours' treatment in the optimal treatment rules, we have proposed a model (3.4) for interference-aware $Q-$learning. Moreover, in equation (2) in their paper, Su et al. consider maximizing the expectation of the network means, but we focus on the egos' outcomes. Note that when we have interference between ego and alters, we will not in general be able to optimize the outcome of each one individually. When there are spillover effects, the best treatment for the ego, from the ego's point of view, may not be the best treatment for the ego from the point of view of the alters (neighbours of the ego).

As an "appetizer" or "starter course" to DTR estimation with interference, the study of Q-learning with interference in this section is intended to present how complications and challenges arise when the dynamic treatment regimes are in the presence of interference, and how they are addressed in the methods recently proposed by various researchers. In the following section, we will further explore dWOLS methods with interference, in particular, the balancing property when we incorporate the neighbour's treatment in the outcome model.

3.3 dWOLS and Interference

After exploring Q-learning with interference, in this section, we will investigate the dWOLS method with interference.

As previously discussed, dWOLS, like most DTR analysis tools, relies on several assumptions common to causal inference methodology: positivity, no unmeasured confounders, and SUTVA (Kosorok and Laber, 2019). If the assumption of no interference does not hold, dWOLS may result in biased estimators of treatment parameters, leading to non-optimal treatment regimes. For example, the optimal treatment decision for an individual may depend on the treatment of those they are connected with. Within our modeling framework, this would necessitate the inclusion of the treatments received by the ego's neighbours in the blip model. Under an assumption of no interference, this term will be omitted, and no guarantees of consistency of the remaining blip parameters can be made.

3.3.1 Analytical framework

Before introducing our proposed estimation method, we introduce the following analytical framework. A fixed network of a certain size is established first, being either an observed network, such as a set of family groups or a set of spatially located units connected by proximity, or a single realization of a random graph (a network generated by a stochastic model). There is no restriction on the pattern of neighbour sharing, such as the ruling out of isolated nodes, triangles or other forms. For our purposes we suppose that each node represents an individual. The covariate values for each of the individuals are assumed independent and identically distributed, independently of the network configuration. The individual treatments are also assumed independent from individual to individual, and depend on the individual's covariate values but are independent of the network configuration. Thus, the propensity for the individual treatment and treatments of neighbours is calculable, conditional on the network and all covariate values.

Because of neighbour sharing, the treatments of neighbours are clearly not necessarily independent from individual to individual, conditional on the network and all covariate values. The outcomes, however, are assumed to rely not only on own covariates and treatment but also on those of neighbours. Thus, if the model for the outcomes is correctly specified, the coefficient estimating function terms for individuals are unbiased conditionally on the network, and all covariate and treatment values. When the treatment-free part of the outcome model is not correctly specified, but the dWOLS balancing weights are applied to the estimating function terms, the combination of estimating function terms used for the coefficients in the (correctly specified) linear treatment part (the blip function Definition 2.2 in Page 28) is unbiased when the expectation is taken over the

distribution of the covariate values, the individual treatments and the outcome, while the network remains fixed. This unbiasedness is what guarantees the consistency of estimators of the blip function parameters (Yuan and Jennrich, 1998) and hence the optimal rule, under appropriate regularity conditions, if the asymptotic framework is an increasing number of replications (on the same network) of covariates, treatments and outcomes.

We will show via simulation that a failure to account for interference can lead to badly mis-identified treatment rules, but when interference is suspected, this can be addressed at the analysis stage. We discuss two methods to accomplish this. First, the most direct (and intuitive) approach, which we will refer to as 'standard' dWOLS, is to add the spillover effect term - that is, the treatment covariates for any alters - to the blip model. However, these spillover terms must also be added to the treatment-free component, regardless of whether they are truly present in the underlying model. (For a specific illustration of this in the case of single decision treatment regimes, see the Invalid dWOLS Theorem and an example in Appendix E.) As an alternative, we propose a new weighting criterion (or balancing condition). This new approach, which we refer to as dWOLS with *network weights*, restores the double robustness of dWOLS without the need to include superfluous terms in the treatment-free model. In addition, we will show via simulation that this approach can also offer more flexible modeling and provides greater robustness for estimation of spillover effects.

3.3.2 Network Propensity Function and Two Methods to Account for Interference

We assume that data are observed from a network where each node is treated independently according to the same randomization rule. We will relax this assumption for pairs in Theorems 3.1 and 3.2, and in Appendix F.1, we extend our methods to cases in which the treatments of ego and alter are not independent. First, we denote $s_{i,a}$ as the set of treated, that is, having $a = 1$, neighbours of i, and then define the "network propensity function" or "network propensity scores" under network interference:

Definition 3.2. *The network propensity function of individual i is the joint probability of individual i receiving treatment a_i, and the treatments of his or her neighbours are in the set $s_{i,a}$. That is,*

$$\pi_{i,a_i,s_{i,a}}(\boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i}) = \mathbb{P}(a_i \cap s_{i,a} \mid \boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i}).$$

Furthermore, we define the standard propensity as $\pi_i = \pi_i(\boldsymbol{x}_i) = \mathbb{P}(A_i = 1 \mid \boldsymbol{x}_i)$, where all $A_1 \ldots, A_n$ are independent, given the covariates $\boldsymbol{x}_i$, for $i = 1 \ldots, n$, and $\boldsymbol{x}_i$ represents the

covariate values for node i itself. We conclude that the network propensity function under network interference can be written as:

$$\pi_{i,a_i,s_{i,a}}(\boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i}) = \mathbb{P}(a_i \ \cap \ s_{i,a} \mid \boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i})$$

$$= \pi_i(\boldsymbol{x}_i)^{a_i}(1-\pi_i(\boldsymbol{x}_i))^{(1-a_i)} \times \prod_{l \in s_{i,a}} \pi_l(\boldsymbol{x}_l) \prod_{l \in \mathcal{N}_i \setminus s_{i,a}} (1-\pi_l(\boldsymbol{x}_l)).$$

Having established these new network propensity functions, we recover the double robustness of dWOLS in the presence of interference in two ways in the following theorems. First, we define network interference term:

Definition 3.3. *The network interference term results from a exposure interference function t that maps from $a_{\mathcal{N}}$ to a scalar ($t : \{0,1\}^{|\mathcal{N}|} \mapsto t(a_{\mathcal{N}}) \in \mathbb{R}$), where $a_{\mathcal{N}}$ are the treatments of units in the neighbourhood.*

That is, network interference term maps the configuration of the neighbour's treatment onto an exposure interference term, e.g., number or proportion of treated neighbours. Suppose that the network interference term or this mapping function is correctly specified (Aronow et al., 2017). We define weights $w_i = w_i(\boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i}, a_i, s_{i,a})$, and assume that the treatment-free model contains a linear term in $t(a_{\mathcal{N}})$, and the blip function is linearly and correctly specified as $\gamma(\boldsymbol{x}^{\psi}, a, t(a_{\mathcal{N}}); \boldsymbol{\psi}) = (\boldsymbol{\psi}_{(\mathbf{1})}^{\mathsf{T}}, \psi_{(2)})(a\boldsymbol{x}^{\psi}, at(a_{\mathcal{N}}))^{\top}$; then, we have the following theorem:

Theorem 3.1. *Balancing Property 1 with Network Interference*

When the true outcome model satisfies $\mathbb{E}[Y|\boldsymbol{x}, a, t(a_{\mathcal{N}})] = f(\boldsymbol{x}^{\beta}, t(a_{\mathcal{N}}); \boldsymbol{\beta}) + \gamma(\boldsymbol{x}^{\psi}, a, t(a_{\mathcal{N}}); \boldsymbol{\psi})$, a weighted ordinary least squares regression based on the corresponding linear model will yield consistent estimators of $\boldsymbol{\psi}$ if at least one of the treatment and treatment-free models is correctly specified, and the weights satisfy $\pi(\boldsymbol{x})w(1,\boldsymbol{x}) = (1-\pi(\boldsymbol{x}))w(0,\boldsymbol{x})$ or its sufficient but not necessary condition

$$w_i(\boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i}, a_i, s_{i,a})\pi_{i,a_i,s_{i,a}}(\boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i}) = \mathscr{C}(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}, \mid \mathcal{N}_i \mid), \tag{3.5}$$

where $\mathscr{C}$ is a function that depends only on the x-variables ($\boldsymbol{x}_i$, $\mathbf{x}_{\mathcal{N}_i}$) and the degree ($\mid \mathcal{N}_i \mid$); given these variables, $\mathscr{C}(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}, \mid \mathcal{N}_i \mid)$ is a constant.

Proof of Theorem 3.1: See Appendix F.2.

Theorem 3.1 expresses that if the true outcome model includes two parts: treatment-free and blip part, and both include the interference term, then both standard dWOLS weights and new

proposed network propensity weights will give us a consistent estimator of $\boldsymbol{\psi}$. Equation (3.5) is our proposed network weights criterion. The product of network weights and the network propensity function is constant, given the covariates and degree of a network. In addition, in this theorem, we showed that the network propensity weights criterion is the sufficient but not necessary condition of the standard dWOLS weights criterion. Any network weights also satisfy the standard dWOLS weight criterion, but not the other way around.

For The second modeling method, we build up a new estimation framework. We consider two blip functions in terms of the network interference term. Suppose that the treatment-free model does not contain a linear term in $t(a_{\mathcal{N}})$, considering this interference term $t(a_{\mathcal{N}})$ and its products with the subset of variates ($\boldsymbol{x}^{\xi}$) to constitute a part of the treatment explanatory variable space, so that the blip functions are: $\gamma^{\xi}(t(a_{\mathcal{N}}), \boldsymbol{x}^{\xi}; \boldsymbol{\xi}) = \boldsymbol{\xi}^{\top} t(a_{\mathcal{N}}) \boldsymbol{x}^{\xi}$, and $\gamma^{\psi}(a, t(a_{\mathcal{N}}), \boldsymbol{x}^{\psi}; \boldsymbol{\psi}) = (\boldsymbol{\psi}_{(\mathbf{1})}^{\mathsf{T}}, \psi_{(2)})(a\boldsymbol{x}^{\psi}, at(a_{\mathcal{N}}))^{\top}$, and assume these two blip functions are correctly specified. The γ^{ξ} model is only related to the spillover effects, basically, it is the interference term (function of neighbours' treatments) and tailoring variables $\boldsymbol{x}^{\xi}$, and γ^{ψ} model is the original blip functions, depending on interference term and tailoring variables $\boldsymbol{x}^{\psi}$. Further, we propose the second theorem of balancing property with network interference:

Theorem 3.2. *Balancing Property 2 with Network Interference*

When the true outcome model satisfies $\mathbb{E}[Y \mid \boldsymbol{x}, a, t(a_{\mathcal{N}})] = f(\boldsymbol{x}^{\beta}; \boldsymbol{\beta}) + \gamma^{\xi}(t(a_{\mathcal{N}}), \boldsymbol{x}^{\xi}; \boldsymbol{\xi}) + \gamma^{\psi}(a, t(a_{\mathcal{N}}), \boldsymbol{x}^{\psi}; \boldsymbol{\psi})$*, a weighted ordinary least squares regression based on the corresponding linear model will yield consistent estimators of* $\boldsymbol{\psi}$ *as well as* $\boldsymbol{\xi}$ *if at least one of the treatment and treatment-free models is correctly specified, and the weights satisfy*

$$w_i(\boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i}, a_i, s_{i,a}) \pi_{i,a_i,s_{i,a}}(\boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i}) = \mathscr{C}(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}, \mid \mathcal{N}_i \mid). \tag{3.6}$$

Proof. Let $\boldsymbol{A} = (A_i, A_{\mathcal{N}_i})$ be the treatment assignment of unit i's group, where A_i is the treatment of unit i and $A_{\mathcal{N}_i}$ are treatments of the units in the neighborhood of unit i. Therefore, there are $2^{|\mathcal{N}_i|+1}$ possibilities for the treatments assignments of $\boldsymbol{A}$; for each unique assignment of treatments to i's group, there is one corresponding weight. It is sufficient to perform a standard regression on a weighted dataset $(y^w, \boldsymbol{x}^w, a^w)$, wherein

$$(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i})^w \perp\!\!\!\perp (a_i, a_{\mathcal{N}_i})^w \tag{3.7}$$

and to satisfy (3.7) it is sufficient to find weights such that:

$$\frac{\mathbb{P}[(A_i, A_{\mathcal{N}_i})^w = (a_i, a_{\mathcal{N}_i})^w | \mathcal{X}_i^w = (\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i})^w]}{\mathbb{P}[(A_i, A_{\mathcal{N}_i})^w = (a_i, a_{\mathcal{N}_i})^w]} = \mathscr{C}(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}, \mid \mathcal{N}_i \mid). \tag{3.8}$$

We follow the strategy of first considering the numerators and denominators separately, then observing that weights could guarantee, for any given treatment assignments (e.g., $(a_i, a_{\mathcal{N}_i}) \in \{0,1\}^{(|\mathcal{N}_i|+1)}$), both numerators and denominators are separately equal.

Let k be the summation of all the products of weights and corresponding propensity scores, which is $k = \sum_{(a_i, a_{\mathcal{N}_i})} w_i(\boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i}, a_i, s_{i,a}) \pi_{i,a_i,s_{i,a}}(\boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i}) = 2^{|\mathcal{N}_i|+1} \mathscr{C}(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}, |\ \mathcal{N}_i\ |)$, where summation is over possible values of $(a_i, a_{\mathcal{N}_i})$, i fixed. For $\forall (a_i, a_{\mathcal{N}_i}) \in \{0,1\}^{(|\mathcal{N}_i|+1)}$,

$$
\begin{aligned}
&\mathbb{P}[(A_i, A_{\mathcal{N}_i})^w = (a_i, a_{\mathcal{N}_i})^w \mid \mathcal{X}_i^w = (\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i})^w] \\
&= \frac{1}{k} \mathbb{P}[(A_i, A_{\mathcal{N}_i}) = (a_i, a_{\mathcal{N}_i}) \mid \mathcal{X}_i = (\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i})] w_i(\boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i}, a_i, s_{i,a}) \\
&= \frac{1}{k} \mathscr{C}(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}, |\ \mathcal{N}_i\ |) = \frac{1}{2^{|\mathcal{N}_i|+1}}.
\end{aligned}
$$

Let $l = \oiint f_{\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}}(\mathcal{X}) k \, d\mathcal{X}$, where $f_{\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}}(\mathcal{X})$ is the joint probability distribution function of $\mathcal{X}_i = (\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i})$.

$$
\begin{aligned}
&\mathbb{P}[(A_i, A_{\mathcal{N}_i})^w = (a_i, a_{\mathcal{N}_i})^w] \\
&= \frac{1}{l} \oiint f_{\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}}(\mathcal{X}) \mathbb{P}[(A_i, A_{\mathcal{N}_i}) = (a_i, a_{\mathcal{N}_i})] w_i(\boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i}, a_i, s_{i,a}) \, d\mathcal{X} \\
&= \frac{1}{l} \oiint f_{\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}}(\mathcal{X}) \mathscr{C}(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}, |\ \mathcal{N}_i\ |) \, d\mathcal{X}
\end{aligned}
$$

The equation holds if weights are of the specified form.

We also remark that the balance property and consistency results hold when the explanatory variable in γ^{ξ} is not the full configuration of neighbours' treatments but a reduced variable that is a function of the neighbours' treatments $t(a_{\mathcal{N}})$, such as the number of treated neighbours. Asymptotically, in the weighted distribution, $t(a_{\mathcal{N}})$, as a function of the treatment configuration, is independent of the ego's explanatory variables $(\boldsymbol{x})$, and therefore as a scalar variable it is uncorrelated with those explanatory variables; the coefficients of $t(a_{\mathcal{N}})$ and its products with variables in $\boldsymbol{x}$ will be estimated consistently. □

The double robustness of dWOLS relies critically on the specification of weights that are included in a linear regression model, and Theorems 3.1 and 3.2 provide details with respect to the balancing condition of the weights in the presence of interference. Therefore, dWOLS is still doubly robust in the presence of interference if either the treatment-free or treatment models are correctly specified when the balancing condition weights in Theorems 3.1 and 3.2

are employed. Furthermore, the network weights proposed in Theorem 3.2 will also guarantee to provide consistent spillover estimators ($\hat{\boldsymbol{\xi}}$). In addition, we reemphasise that only if the blip functions are correctly specified will the blip parameters' estimators be meaningful (see Wallace et al., 2017a for assessing the blip model specifications). However, even if the blip functions are not correctly specified, the balancing weights would still achieve a kind of balancing similar to what randomization testing provides, as long as the treatment model is correct. See the proof of Theorem 3.2, where the balancing condition does not depend on the exposure mapping function.

Note that these theorems apply to a wide variety of network contexts, as indicated by the flexibility of the exposure interference function (t), which maps configuration of the neighbour's treatment onto an exposure interference term. For example, $t(a_{\mathcal{N}})$ can be the number of treated neighbours, i.e., $t_i(a_{\mathcal{N}_i}) = \sum_{l=1}^{n} N_{il}a_l$, where N denotes the adjacency matrix of the network. Alternatively, t could simply be an indicator function (indicating the existence of a treated neighbour, i.e., $t_i = \mathbb{I}(\sum_{l=1}^{n} N_{il}a_l > 0)$) or represent the proportion of neighbours who are treated ($t_i = \sum_{l=1}^{n} N_{il}a_l / \sum_{l=1}^{n} N_{il}$). This t function is analogous to the "exposure mapping" in Aronow et al. [2017], Leung [2019], and Forastiere et al. [2021], in which the exposure mapping function illustrates the idea of dimension reduction, i.e., $t(a_{\mathcal{N}})$ is a low-dimensional summary of neighbours' treatment assignments. Additionally, the exposure mapping function is application-specific, and needs to be identified by consideration of the specific interference structure. To estimate the full blip model and guarantee identifiability of the interference term, the exposure mapping function t cannot take only one value for every configuration of neighbours' treatments, that is, it cannot be true that $t(a_{\mathcal{N}}) \equiv c$ for every $a_{\mathcal{N}}$, where c is a fixed value. An illustrative example of the balancing weights conditions (3.5) and (3.6) and the demonstration of choice and computation of the network weights are provided in Appendix F.3.

3.3.3 Implementation and Choice of Network Weights

In the presence of interference, our proposed methods, namely standard dWOLS and dWOLS with network weights, are implemented by carrying out a weighted ordinary least squares where the weights of these two methods are, respectively, the standard dWOLS weights and weights formed from the network propensity scores. Furthermore, we propose using the "absolute value" weights (or overlap weights in Li et al., 2018; Schulz and Moodie, 2021) so that the network weights for individual i in a network are:

$$w_i = |a_i - \mathbb{E}[A_i|X_i = x_i]\,| \times \prod_{l} |a_l - \mathbb{E}[A_l|X_l = x_l]\,|, \tag{3.9}$$

where $l \in \mathcal{N}_i$. Moreover, examples of both the standard absolute value weights and the network absolute value weights will be presented in the simulation section.

Under the network weight definition, the larger the degree of the node, the smaller the weights. For a network in which the nodes all have the same degree, the weights would be comparable, but for variable degree distributions they would not. The principle of using network weights is to balance the treatment distribution with respect to the covariates. Therefore, for estimation we choose a set of egos with the same degree. Within that set, when the joint distribution of the treatment of the ego and the number of treated neighbours is balanced with respect to their covariates, then the weights are comparable. A possible extension to this approach would be to consider the network as the union of subsets of constant degree, and to combine the estimating function systems corresponding to each degree value in an efficient manner. Accordingly, in the different ego-degree mixture cases, we suggest using stabilized (rescaled) weights - dividing the network weight by the average value of the network weight over all nodes of the same degree - to reduce the variability of weights overall and boost estimation efficiency.

Just as standard dWOLS establishes covariate balance of patient-level covariates between treatments, the network weights extend this idea to provide balance of patient-level covariates under the treatments of their neighbours. In this sense, the network weights function as a full factorial design, which is (normally) expensive but more efficient than one-factor-at-a-time experiments (Montgomery, 2017). Compared with standard dWOLS, the inclusion of network weights affords two key advantages. First, incorporating the alter's treatment model into the analysis can offer considerable gains in efficiency of our blip parameter estimators. We note this does not necessarily require additional modeling, as in practice it will often be reasonable to assume the same treatment model for all individuals. Second, the use of network weights offers more flexibility in modeling in the case where the treatment-free model contains interference terms, i.e., covariates relating to the alter. In this scenario, the treatment-free model can itself be decomposed into two sub-models: one containing only terms related to the ego, and one containing terms relating to the alter. We require that either the part of the treatment-free model associated with the ego or the treatment model is correctly specified; then the blip parameter ($\boldsymbol{\psi}$) estimators are consistent. Furthermore, if these two models are correct, but the part of the treatment-free model related to the alter sub-model is incorrect, then the spillover effects estimators can also be consistently estimated from dWOLS with network weights, whereas standard dWOLS cannot accomplish this.

3.3.4 Value and Regret Functions in the Presence of Interference

Finally, if the neighbours' treatments are pre-determined, that is, alters' treatments are fixed throughout the study period, then the optimal value function that was

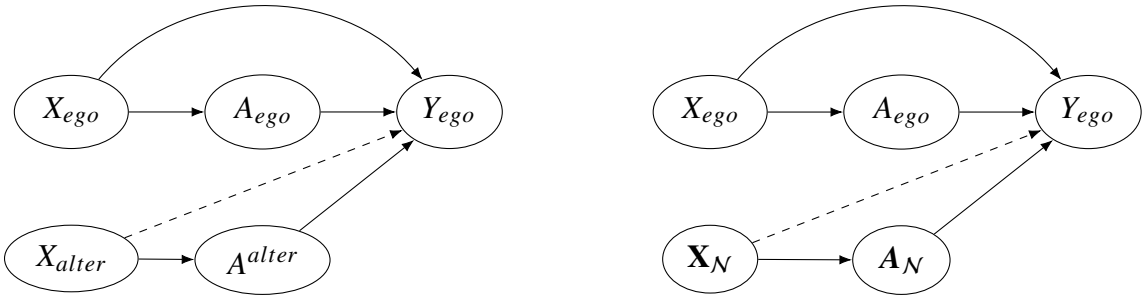
mentioned in Section 3.1 is $\mathcal{V}[(a, a_{\mathcal{N}})^{opt}] = \mathbb{E}\left[\max_{(a,a_{\mathcal{N}})} \mathbb{E}[Y \mid (\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}), \mathbf{A} = (a, a_{\mathcal{N}})]\right] = \mathbb{E}\left[\max_{a} \mathbb{E}[Y \mid (\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}), \mathbf{A} = (a, a_{\mathcal{N}})]\right]$. Building on the regression model in Theorem 3.1 or 3.2, an outcome regression estimator for $\mathcal{V}[(a, a_{\mathcal{N}})^{opt}]$ is $\widehat{\mathcal{V}}[(a, a_{\mathcal{N}})^{opt}] = \widehat{\mathcal{V}}[(\hat{a}^{opt}, a_{\mathcal{N}})] = n^{-1} \sum_{i=1}^{n} \left(Y_i + \gamma^{\psi}(a_i, t(a_{\mathcal{N}_i}), \boldsymbol{x}_i^{\psi}; \hat{\boldsymbol{\psi}}) \left[\hat{a}_i^{opt} - a_i\right]\right)$, where $\hat{a}_i^{opt} = \mathbb{I}\left(\gamma^{\psi}(1, t(a_{\mathcal{N}_i}), \boldsymbol{x}_i^{\psi}; \hat{\boldsymbol{\psi}}) > 0\right)$ and $i = 1, ..., n$ are egos in the network. Consequently, a consistent estimator for the *total regret*, the total value loss arising from receiving the observed treatment regimen $(a, a_{\mathcal{N}})$ rather than the estimated optimal treatment regimen $(\hat{a}^{opt}, a_{\mathcal{N}})$, is $\widehat{\mathcal{L}}[(a, a_{\mathcal{N}})] = \widehat{\mathcal{V}}[(a, a_{\mathcal{N}})^{opt}] - \mathcal{V}[(a, a_{\mathcal{N}})] = n^{-1} \sum_{i=1}^{n} \left(\gamma^{\psi}(a_i, t(a_{\mathcal{N}_i}), \boldsymbol{x}_i^{\psi}; \hat{\boldsymbol{\psi}}) \left[\mathbb{I}\left(\gamma^{\psi}(1, t(a_{\mathcal{N}_i}), \boldsymbol{x}_i^{\psi}; \hat{\boldsymbol{\psi}}) > 0\right) - a_i\right]\right)$, where $\mathcal{V}[(a, a_{\mathcal{N}})] = n^{-1} \sum_{i=1}^{n} Y_i$. Note that the total regret aims to compare the difference between the estimated optimal treatment regimes and originally observed treatment regimes, and can be computed from data even though we do not know the true optimal treatment regimes. Therefore, the larger the total regret, the greater the improvement of the estimated optimal treatment regime over the originally observed treatment regime.

3.4 Simulation Studies

We now investigate the impact of interference and demonstrate the use of both standard and network weighted dWOLS via a series of simulation studies. Five studies are for single-stage decision problems: three with respect to a network of couples (*Studies 1a - 1c*) and two with respect to broader network scenarios (*Studies 2a and 2b*). Two further simulation studies are for multiple-stage decision problems: one for a network of couples (*Study 3a*) and the other for a spatial network (*Study 3b*).

3.4.1 Single-stage Decision for a Couples Case

Our first simulations (*Study 1a*) consider the case of a single-stage treatment decision with interference for isolated couples (that is, each individual is connected to precisely one other person). Following the causal diagram in Figure 3.3a (where Y_{ego} does not depend on X_{alter}), we generate data as follows. Patient information: $X_i \sim N(2, 1)$; $X_{alter_i} \sim N(2, 1)$; treatment $\mathbb{P}(A_i = 1 \mid X_i) = \text{expit}[-0.25X_i + \sin(X_i)]$ and $\mathbb{P}(A_i^{alter} = 1 \mid X_{alter_i}) = \text{expit}[-0.25X_{alter_i} + \sin(X_{alter_i})]$, where $\text{expit}(x) = [1 + exp(-x)]^{-1}$; outcome $Y_i \sim N(\mu_i, 1)$ where $\mu_i = X_i + \log(|X_i|) + X_i^3 + A_i^{alter} + A_i^{alter} X_i + A_i(\psi_0 + \psi_1 X_i + \psi_2 A_i^{alter} + \psi_3 X_i A_i^{alter})$, where we note that both the treatment-free and blip components depend on the spillover term A_i^{alter}. Our interest is in estimating the blip



(a) DAG 1 (b) DAG 2

Figure 3.3: Directed Acyclic Graphs (DAGs) of simulations in 3.1 and 3.2. For the dashed line (from X_{alter} to Y_{ego}), it is supposed that a possible interference path is due to Y_{ego} depending on X_{alter}. This interference path does not exist in *Studies 1a* and *2a* but it is present in *Studies 1b, 2b* and *1c*.

parameters $\psi_0 = -3.5, \psi_1 = 1.5, \psi_2 = 3$, and $\psi_3 = 2$. We consider three methods. In *Method 1* (a 'naive' analysis) we study the impact of interference on an analysis that fails to take it into account, omitting the a^{alter} terms entirely from the working models. In *Method 2* (standard dWOLS), we consider the original dWOLS methodology (with weights based only on the treatment model) but with a^{alter} terms included in the blip and treatment-free models. Finally, in *Method 3*, we demonstrate dWOLS with network weights, constructed from the standard dWOLS weights (those that depend on the ego's treatment model) and the corresponding network factor for the alter.

For each simulation, we conduct analyses in the following four scenarios: a) both the treatment and treatment-free models are mis-specified; b) the treatment-free model is mis-specified but the treatment model is correctly specified; c) the treatment model is mis-specified but the treatment-free model is correctly specified; and d) both the treatment and treatment-free models are correctly specified. Model mis-specification is implemented via the omission of non-linear terms in the treatment and treatment-free models. Note that 'treatment model' refers to the model used for both the ego and alter's treatment. Model set-ups of Methods 1 - 3 are: (1) Method 1, a 'naive' analysis, where we ignore the interference entirely; (2) standard dWOLS, and (3) dWOLS with network weights. For Scenarios a) and b) where the treatment-free model is mis-specified, we consider outcome models:

- Method 1:
  $\mathbb{E}[Y|\boldsymbol{X} = \boldsymbol{x}] = \beta_0 + \beta_1 x + a(\psi_0 + \psi_1 x)$, with weight $w = |a - \mathbb{E}(A \mid \boldsymbol{x})|$.
- Method 2:

Table 3.1: Mean optimal treatment rate and its standard errors of Methods 1, 2, and 3 when neither (scenario a), one (scenarios b and c) or both (scenario d) treatment and treatment-free outcome models are correct in *Study 1a* (ME: Method).

		Mean optimal treatment rate (MOTR) %			Standard errors of MOTR		
Scenario		$n = 200$	$n = 500$	$n = 1000$	$n = 200$	$n = 500$	$n = 1000$
a	ME1	63.34	67.67	69.53	0.1515	0.0836	0.0471
	ME2	78.64	80.93	80.11	0.1265	0.0864	0.0683
	ME3	78.20	80.11	80.76	0.1248	0.0829	0.0724
b	ME1	68.84	70.31	70.75	0.0850	0.0413	0.0287
	ME2	81.62	86.84	88.64	0.1301	0.1052	0.0838
	ME3	80.89	85.03	88.09	0.1372	0.1040	0.0876
c	ME1	70.47	71.03	70.81	0.0637	0.0397	0.0284
	ME2	97.44	98.51	98.84	0.0280	0.0152	0.0103
	ME3	97.32	98.47	98.79	0.0290	0.0156	0.0110
d	ME1	70.29	70.63	70.93	0.0642	0.0413	0.0285
	ME2	97.35	98.30	98.93	0.0289	0.0155	0.0098
	ME3	97.30	98.18	98.89	0.0293	0.0166	0.0098

$\mathbb{E}[Y|\boldsymbol{X} = \boldsymbol{x}] = \beta_0 + \beta_1 x + \beta_2 a^{alter} + \beta_3 x a^{alter} + a_{ego}(\psi_0 + \psi_1 x + \psi_2 a^{alter} + \psi_3 x a^{alter})$, and with weight $w = |a_{ego} - \mathbb{P}(A_{ego} = 1 \mid \boldsymbol{x}_{ego})|$.

- Method 3:
  $\mathbb{E}[Y|\boldsymbol{X} = \boldsymbol{x}] = \beta_0 + \beta_1 x + \beta_2 a^{alter} + \beta_3 x a^{alter} + a_{ego}(\psi_0 + \psi_1 x + \psi_2 a^{alter} + \psi_3 x a^{alter})$, and with network weight $w_{net}(x) = |a_{ego} - \mathbb{E}(A_{ego} \mid x_{ego})| * |a^{alter} - \mathbb{E}(A^{alter} \mid \boldsymbol{x}_{alter})|$.

Our initial results (Table 3.1) summarize the impact of interference on treatment recommendations. The total regret values for each scenario and method, which result from the value functions, are indicated in Table 3.2. As expected, when the analysis does not account for interference in any way (Method 1), the rates at which recommended treatments align with those that are truly optimal (i.e., mean optimal treatment rate $n^{-1} \sum_{i=1}^{n} \mathbb{I}\left(a_i^{opt} = \hat{a}_i\right)$) are poor. Both proposed methods, meanwhile, demonstrate marked improvements. The improvement in treatment recommendation stems from accurate parameter estimation: both standard dWOLS and dWOLS with network weights demonstrate the expected double robustness (Figure 3.4).

Further to these initial results, we extend the preceding simulations to more complex scenarios. These include the case where the treatment-free model contains additional interference terms

Table 3.2: Total regret and its standard errors of Methods 1, 2, and 3 when neither (scenario a), one (scenarios b and c) or both (scenario d) treatment and treatment-free outcome models are correct in Study 1a (ME: Method).

		Total Regret			Standard Errors of Total Regret		
Scenario		$n = 200$	$n = 500$	$n = 1000$	$n = 200$	$n = 500$	$n = 1000$
a	ME1	2.35	0.23	1.11	0.19	0.14	0.11
	ME2	3.31	1.87	2.27	0.28	0.12	0.11
	ME3	2.29	2.01	2.01	0.24	0.12	0.10
b	ME1	0.91	0.22	0.18	0.28	0.13	0.10
	ME2	1.75	1.25	0.60	0.34	0.15	0.09
	ME3	1.90	1.23	1.04	0.21	0.14	0.09
c	ME1	0.82	0.39	0.56	0.21	0.14	0.09
	ME2	2.00	1.65	2.22	0.24	0.12	0.11
	ME3	2.01	1.67	2.25	0.24	0.11	0.09
d	ME1	0.97	1.35	1.38	0.21	0.14	0.10
	ME2	1.79	2.67	1.81	0.20	0.16	0.09
	ME3	1.85	2.64	1.80	0.20	0.15	0.09

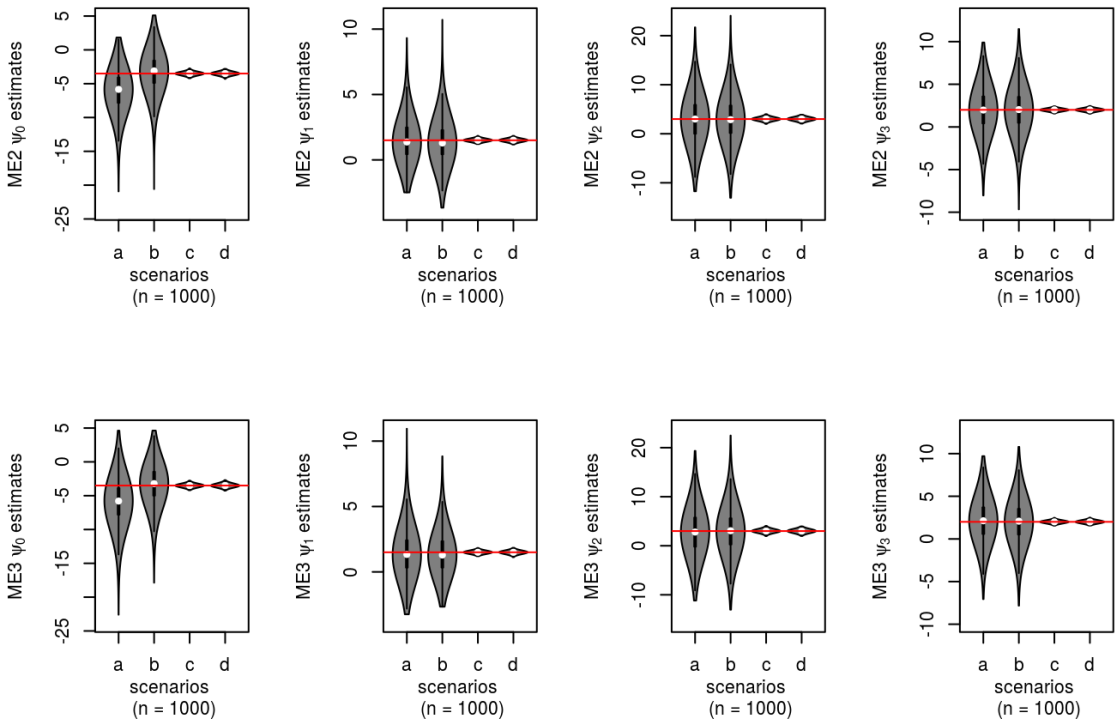


Figure 3.4: Blip function parameter estimates via Method 2 (standard dWOLS, top row) and Method 3 (dWOLS with network weights, bottom row) when neither model (scenario a), treatment model only (scenario b), treatment-free model only (scenario c), or both models are correctly specified.

(*Study 1b*), and where the true ego and alter treatment models are not identical (*Study 1c*). These additional simulations demonstrate the potential for further efficiency gains (Figure F.1 in Appendix Section F.5), and greater modeling flexibility (consistent estimation of spillover parameters, Figures F.2 and F.4 in Appendix Section F.5), when using network weights compared with standard dWOLS. Full details are included in Appendix Section F.5.

3.4.2 Single-stage Decision for Egos in a Social Network

We next consider treatment decisions for egos in a social network (that is, where each individual may have more than one neighbour). In particular, we consider five types of fixed network: (i) a ring consisting of points on a circle, (ii) a square lattice, (iii) an Erdős-Rényi (ER) network realization, (iv) a Barabási-Albert (BA) network realization and (v) the Longleaf Pines (Rathbun and Cressie, 1994) spatial network. The ring and the square lattice are simple network structures with constant degree. An Erdős-Rényi network (Erdős and Rényi, 1960) is a random graph $G(n, p)$ with n vertices, where each pair of vertices is independently connected with a fixed probability p. The degree distribution is binomial $(n-1, p)$, and if $\lambda = p/n$ this leads in the limiting case to the Poisson random graph model with mean degree λ. Similar to our approach in simulation Section 3.2.2, the BA network is also generated from the Barabási-Albert model. Finally, we consider interference (e.g., Puelz et al., 2019) in a spatial network, i.e. one in which each node has a spatial location. Let $d(i, j)$ denote the spatial distance between nodes i and j. We assume that $j \in \mathcal{N}_i$ if $d(i, j) < \rho$. Thus the entries of the adjacency matrix of the spatial network are $N_{ij} = \mathbb{I}(d(i, j) < \rho)$. For illustration, we analyse a real point process dataset of the Longleaf Pines in the R **spatstat** package (Baddeley and Turner, 2005), where an edge in the spatial network exists if the spatial distance is less than 5 meters.

For the five types of fixed network, we consider an interference term of the form $t_i(a_{\mathcal{N}_i}) = \sum_{l=1}^{n} N_{il}a_l$, that is, the number of treated neighbours of an ego i. Further, we consider a model such that: $\mathbb{E}[Y|\boldsymbol{x}, a, t] = f(\boldsymbol{x}^{\beta}, t; \boldsymbol{\beta}) + \gamma(\boldsymbol{x}^{\psi}, a, t; \boldsymbol{\psi})$, and apply weights from (3.9) to guarantee the robustness of the blip estimates when one of the treatment-free or treatment model is misspecified.

The data generated for *Study 2a* follow the causal diagram in Figure 3.3b, in a case where Y_{ego} does not depend on $\mathbf{X}_{\mathcal{N}}$. Data are generated as follows. Patient information $X_i \sim N(2, 1)$; the treatment model is $\mathbb{P}(A_i = 1 \mid X_i = x_i) = \text{expit}[-0.25x_i + \sin(x_i)]$; the outcome $Y_i \sim N(\mu_i, 1)$ where $\mu_i = 1 + X_i + X_i^2 + t_i(A_{\mathcal{N}_i}) + t_i(A_{\mathcal{N}_i})X_i + A_i[\psi_0 + \psi_1 X_i + \psi_2 t_i(A_{\mathcal{N}_i}) + \psi_3 X_i t_i(A_{\mathcal{N}_i})]$. Our interest is in estimating the blip parameters $\psi_0 = -5, \psi_1 = 0.9, \psi_2 = 1, and\ \psi_3 = 0.3$. We set $n = 1000$ nodes in the ring. For the square lattice, we set $n = 900$, and 784 of the egos have degree 4 (i.e., 4 neighbours). In the degree distribution of the 2000-node Erdős-Rényi network (generated from the $G(n = 2000, p = 1/400)$ model), 339 nodes have degree 4. These are set as egos.

Next, for the 1000-node Barabási-Albert network (generated from the model with parameters $n = 1000, m = 5$, where m refers to the number of edges to add in each time step of generating the graph), the 806 nodes with the highest degree (i.e., 5) are designated as egos. Lastly, for the 584-node Longleaf Pines spatial network, the spatial figures and degree summary are shown in Appendix F.6, and 98 nodes with 2-degree are designated as egos.

Similar to our previous simulation *Study 1a*, we conduct three analyses, where *Method 1* denotes a failure to account for interference entirely, *Method 2* employs standard dWOLS, and *Method 3* uses dWOLS with network weights. The three working models are therefore:

- (I) Method 1: $\mathbb{E}[Y|x, a; \boldsymbol{\beta}, \boldsymbol{\psi}] = \beta_0 + \beta_1 x + a(\psi_0 + \psi_1 x)$, with weight $w(x_i) = |a_i - \mathbb{E}[A|X_i = x_i]\,|$.
- (II) Method 2: $\mathbb{E}[Y|x, a, t; \boldsymbol{\beta}, \boldsymbol{\psi}] = \beta_0 + \beta_1 x + \beta_2 t + \beta_3 xt + a(\psi_0 + \psi_1 x + \psi_2 t + \psi_3 xt)$, with weight $w(x_i) = |a_i - \mathbb{E}[A|X_i = x_i]\,|$.
- (III) Method 3: $\mathbb{E}[Y|x, a, t; \boldsymbol{\beta}, \boldsymbol{\psi}] = \beta_0 + \beta_1 x + \beta_2 t + \beta_3 xt + a(\psi_0 + \psi_1 x + \psi_2 t + \psi_3 xt)$, with network weight $w_i = |a_i - \mathbb{E}[A_i|X_i = x_i]\,| * \prod_l |a_l - \mathbb{E}[A_l|X_l = x_l]\,|$, where $l \in \mathcal{N}_i$.

Although the studied network structures are different (i.e., rings, square lattices, Erdős-Rényi, Barabási-Albert, Longleaf Pines spatial network), if we select egos that have the same degree, the results are similar. We present the conclusions in the general case but show the corresponding figures in the case of Erdős-Rényi network. Similar to the results of Section 3.4.1, we find that a failure to account for interference is severely damaging to optimal treatment assignment rates and the total regrets (values), while the two dWOLS implementations offer considerable improvements. The results have the same pattern as those in Table 3.1: the optimal treatment rates of Methods 2 and 3 are similar, and are markedly superior to those of Method 1. Again, these improvements are attributable to the double robustness of our blip parameter estimators when using standard weights and network weights.

To highlight the advantages of network weights in particular, we also extend the preceding simulations to a more complex scenario, namely the case where the treatment-free model contains additional interference terms (*Study 2b*). These additional simulations illustrate the potential for efficiency gains (Figure F.6 in Appendix F.7), and greater modeling flexibility (Figure F.7 in Appendix F.7), when using network weights compared with standard dWOLS. That is, compared with Method 2, Method 3 appears to have better results: efficiency gains of blip estimators of $\boldsymbol{\psi}$ and consistent spillover estimators of $\boldsymbol{\xi}$. Full details are included in Appendix F.7. Finally, to check the variability of the balancing weights, based on these simulations, the histograms of both the standard absolute value weights and the network absolute value weights are also shown in Appendix F.7.

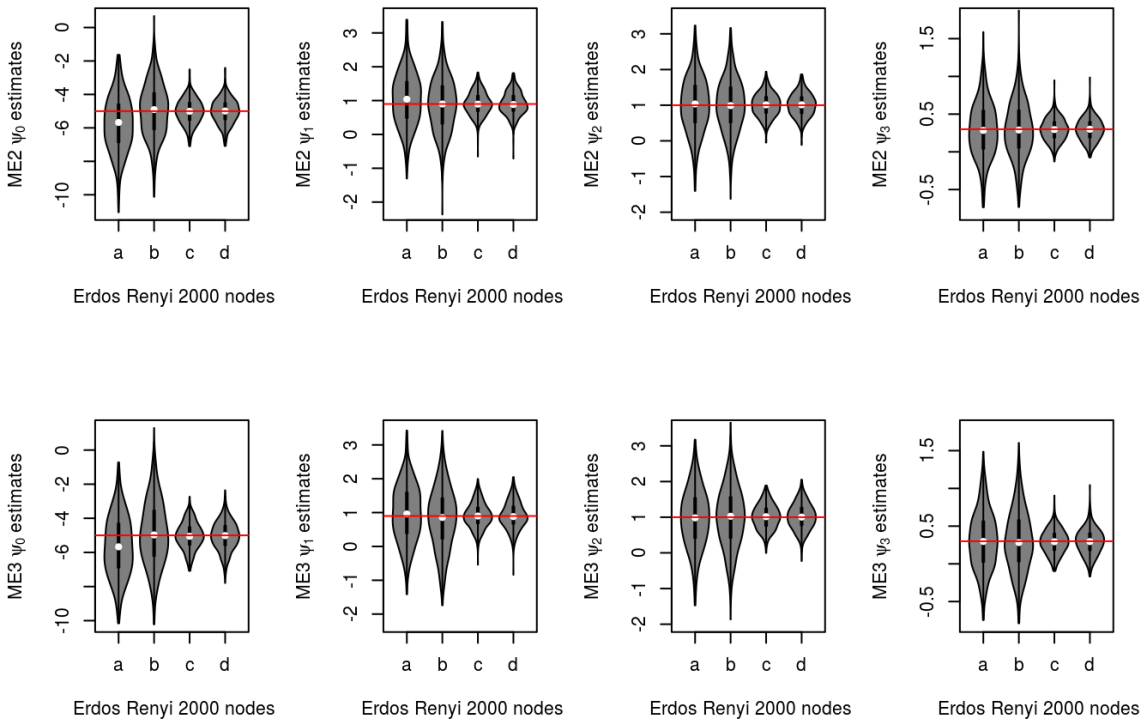


Figure 3.5: Erdős-Rényi 2000 nodes. Blip function parameter estimates via Method 2 (standard dWOLS, top row) and Method 3 (dWOLS with network weights, bottom row) when neither model (scenario a), treatment model only (scenario b), treatment-free model only (scenario c), or both models are correctly specified.

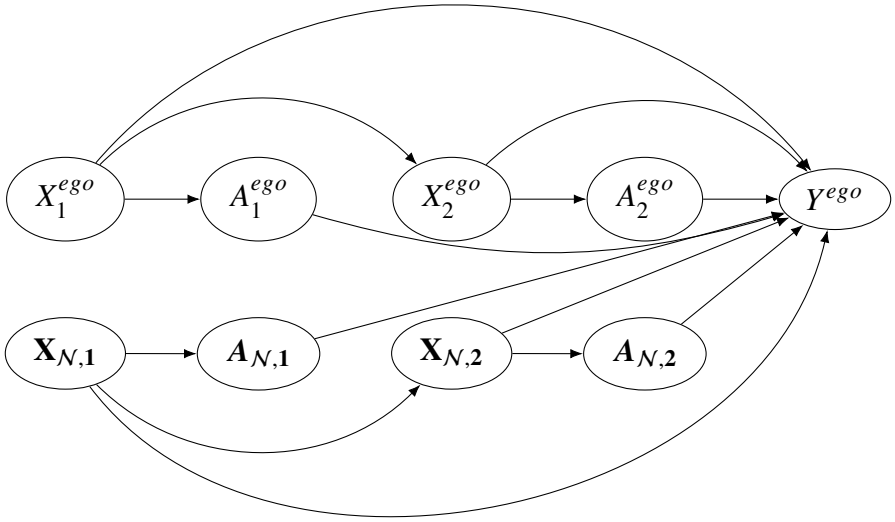


Figure 3.6: DAGs of Simulations in 3.3 (*Studies 3a* and *3b*): dWOLS with interference two-stage decision problems. Note that $\mathbf{X}_{\mathcal{N},\mathbf{1}}$ and $\mathbf{X}_{\mathcal{N},\mathbf{2}}$ (or $\boldsymbol{A}_{\mathcal{N},\mathbf{1}}$ and $\boldsymbol{A}_{\mathcal{N},\mathbf{2}}$) denote the first and second stage covariates (or treatments) of ego's neighbours.

3.4.3 Multiple-stage Decisions

For the multiple-stage decision problem, dWOLS uses backward induction, solving the multiple-stage problems recursively beginning with the final stage. As we move backwards through the stages we construct pseudo-outcomes, which are the potential outcomes if the patients were treated - possibly contrary to fact - optimally at subsequent stages. In expressing the optimal sequence for an individual, we assume that the network (structure) is fixed, and the alters' treatments are known and predetermined. Formally, for a $K-$ stage decision problem in the present of interference, starting from the K^{th} (last) stage, one can estimate the blip parameters $\boldsymbol{\psi}_K$ via WOLS with our proposed network weights, and establish the corresponding K^{th} stage optimal treatment decision rule $\hat{a}_K^{opt}$. Then moving to the previous stages $j = K-1, ..., 1$, dWOLS with interference is applied to the pseudo-outcome: $\widetilde{\mathcal{Y}_j} = y + \sum_{k=j+1}^{K} \left[\gamma_k(\boldsymbol{h}_k^{\psi}, \hat{a}_k^{opt}, t_k(a_{\mathcal{N}}); \hat{\boldsymbol{\psi}}_k) - \gamma_k(\boldsymbol{h}_k^{\psi}, a_k, t_k(a_{\mathcal{N}}); \hat{\boldsymbol{\psi}}_k)\right]$.

In simulation *Study 3*, we illustrate a multistage decision problem via two-stage simulations for both a network of couples case (*Study 3a*) and a spatial network case (*Study 3b*). The data generating

process of this two-stage problem follows the causal diagram Figure 3.6. Patient information: $X_1 \sim N(2,1), X_2 \sim N(1{+}0.5X_1, 1)$; treatment: $\mathbb{P}(A_1 = 1 \mid x_1) = \text{expit}[-0.25x_1{+}\sin(x_1)], \mathbb{P}(A_2 = 1 \mid x_2) = \text{expit}[-0.25x_2 + \sin(x_2)]$; the blip functions: $\gamma_j^{\psi}\left(\boldsymbol{h}_j^{\psi}, a_j, t(a_{\mathcal{N},j}); \boldsymbol{\psi}_j\right) = a_j \boldsymbol{\psi}_j^{\top} \boldsymbol{x}_j^{\psi}$, with $\boldsymbol{x}_j^{\psi} = (1, x_j, t(a_{\mathcal{N},j}), x_j t(a_{\mathcal{N},j}))^{\top}$, where $\boldsymbol{\psi}_j = (\psi_{0j}, \psi_{1j}, \psi_{2j}, \psi_{3j})^{\top} = (-2, 1, 1, 0.5)^{\top}$, and $\gamma_j^{\xi}\left(\boldsymbol{h}_j^{\xi}, t(a_{\mathcal{N},j}); \boldsymbol{\xi}_j\right) = t(a_{\mathcal{N},j}) \boldsymbol{\xi}_j^{\top} \boldsymbol{x}_j^{\xi}$, with $\boldsymbol{x}_j^{\xi} = (1, x_j)^{\top}$, where $\boldsymbol{\xi}_j = (\xi_{0j}, \xi_{1j})^{\top} = (-1, 0.5)^{\top}$ for $j = 1, 2$. Outcome: $Y \sim N(\mu, 2)$, where $\mu = \log(|X_1|){+}\sin(X_2){+}\log(|\tau(\mathbf{X}_{\mathcal{N},\mathbf{1}})|){+}\log(|\tau(\mathbf{X}_{\mathcal{N},\mathbf{2}})|){+}\sum_{j=1}^{2} \gamma_j^{\xi}\left(\boldsymbol{h}_j^{\xi}, t(a_{\mathcal{N},j}); \boldsymbol{\xi}_j\right) + \sum_{j=1}^{2} \gamma_j^{\psi}\left(\boldsymbol{h}_j^{\psi}, a_j, t(a_{\mathcal{N},j}); \boldsymbol{\psi}_j\right)$. Note that the treatment-free model contains non-linear terms of both egos' covariates and the summary of their neighbours' covariates. In particular, in *Study 3b*, a general spatial network case, we consider the fixed 584 nodes Longleaf Pines spatial network in Section 3.4.2, and designate 98 nodes with 2-degree as egos; for an ego i in the network, we consider $\tau(\mathbf{X}_{\mathcal{N}_i}) = \sum_{l=1}^{n} N_{il} X_l / \sum_{l=1}^{n} N_{il}$, and let $\mathbf{X}_{\mathcal{N}_i,\mathbf{1}}$ and $\mathbf{X}_{\mathcal{N}_i,\mathbf{2}}$ (respectively $\boldsymbol{A}_{\mathcal{N}_i,\mathbf{1}}$ and $\boldsymbol{A}_{\mathcal{N}_i,\mathbf{2}}$) denote the first and second stage covariates (respectively treatments) of the neighbours of ego i, while the interference term is determined by exposure mapping $t(a_{\mathcal{N}_i}) = \sum_{l=1}^{n} N_{il} a_l$, i.e., the number of treated neighbours. In *Study 3a* we consider a network of couples where each ego has only one neighbour; therefore, both $\mathbf{X}_{\mathcal{N}_i}$ and $a_{\mathcal{N}_i}$ are scalars, and $\tau(\mathbf{X}_{\mathcal{N}_i})$ and $t(a_{\mathcal{N}_i})$ are the only alter covariate and treatment of the ego i, respectively. We generate 1000 pairs in this study.

The results of *Study 3a* are presented in Figure 3.7 (See Appendix F.8 for the results of *Study 3b*). Although the treatment-free models are misspecified in both stages, dWOLS with network weights provides apparently consistent estimators of blip parameters as anticipated.

3.5 Real Data Analysis from Population Assessment of Tobacco and Health Study

In Chapter 1, we have introduced the dataset from the PATH study, and in this section, we will apply our interference-aware dWOLS methods in the analysis of PATH data.

3.5.1 Variables Selection and Models Specification

At the beginning of the study, e-cigarettes were beginning to see widespread use in the United States, and usage increased considerably between Waves 1 and 2, and between Waves 2 and 3. Wave 4, meanwhile, coincided with the emergence of products that saw widespread use particularly among younger people (Huang et al., 2019). A growing body of literature suggests that e-cigarettes

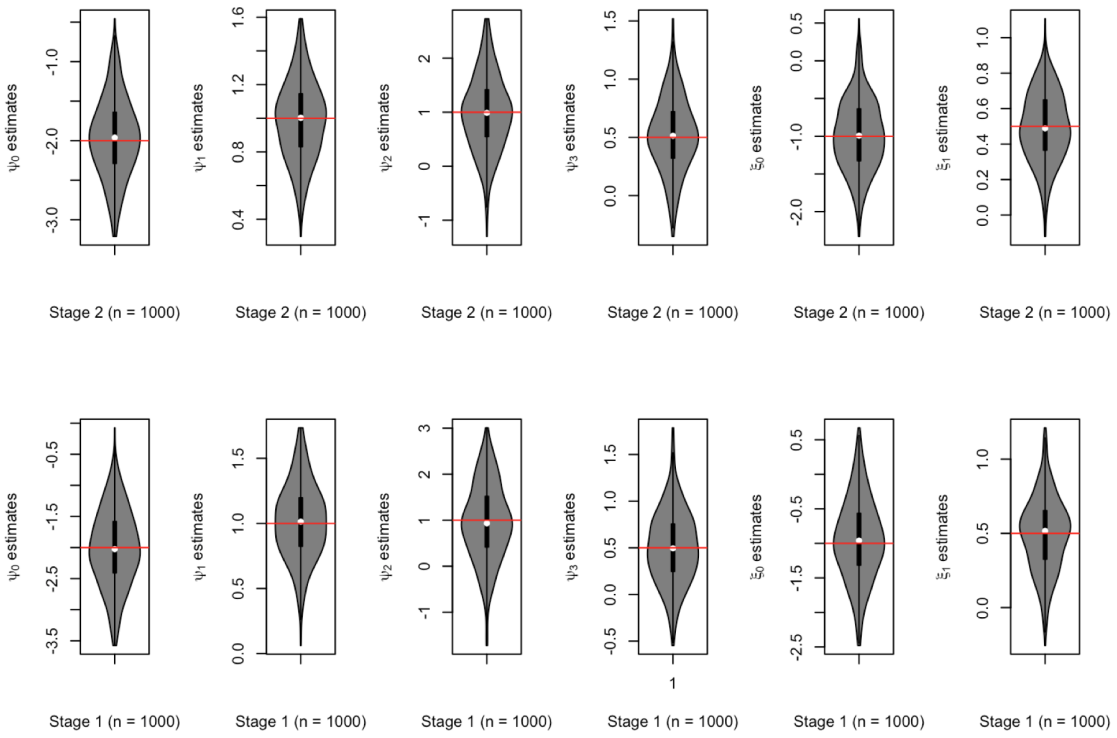


Figure 3.7: Two-stage decision blip parameter estimates for the 1000 couples case (*Study 3a*). Second stage blip estimates are presented in the top row, and those of the first stage are shown in the bottom row. In both rows, the plots display estimates of $\psi_0, \psi_1, \psi_2, \psi_3, \xi_0, \xi_1$ from left to right.

(vaping) can be a useful cessation aid (Villanti et al. [2018]; Hajek et al. [2019]); thus, we set the treatment variable as the use of e-cigarettes by a cigarette smoker. Due to the long follow-up, with approximately one year between the start of each wave, we define e-cigarette use reported at the wave of the measured outcome as indicative of the pre-wave treatment (i.e., e-cigarette use). In addition, the e-cigarette usage variable is determined by participants being asked during the study whether they "now use e-cigarettes (a) Every day (b) Some days (c) Not at all." If participants answer either "Every day" or "Some days," then they are deemed to use e-cigarettes and are coded $A = 1$; otherwise, if they respond "Not at all," they are coded $A = 0$.

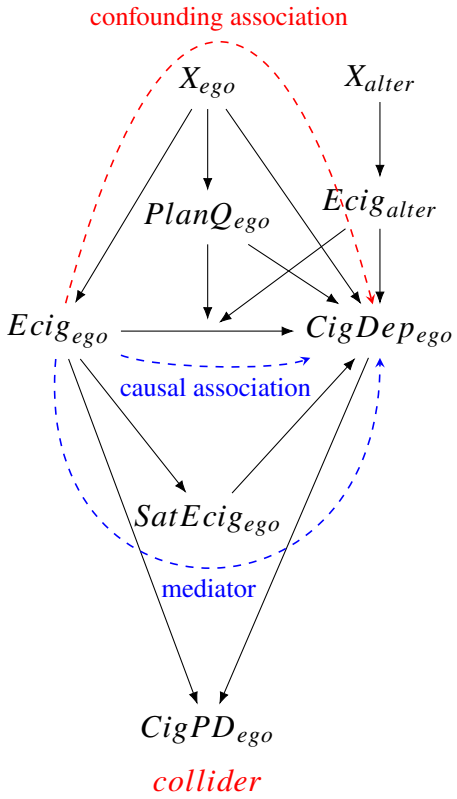
Our outcome of interest is a measure of participants' dependence on cigarettes (Strong et al., 2017). Cigarette dependence may be measured in both smokers and recent former smokers via a scale constructed by combining the responses to "level of agreement" questions, with scores ranging from 15 (lowest dependence) to 76 (highest dependence). Someone who has quit smoking more than 12 months earlier and is no longer eligible for the questionnaire is assigned a score of 7. The cigarette dependence scale is treated as a continuous cigarette-dependence outcome. Furthermore, we constructed a four-wave outcome measure $DiffDep14$, which is cigarette dependence, Wave 1, minus cigarette dependence, Wave 4; therefore, the larger value of $DiffDep14$ is preferred.

We restrict attention to households where, in Wave 1, two adults were interviewed, both were cigarette smokers, and both were followed up at Waves 2, 3 and 4. Each pair has been given a unique identification number and, for our purposes, one of each pair is randomly designated the ego and the other the alter, with the ego the unit of analysis. The survey weight for the ego was taken to be the adult all-waves longitudinal weight for Wave 1 to Wave 4. There are 769 pairs, and thus 769 egos, each randomly selected from a couple. We assume for the sake of illustration that drop-out is uninformative.

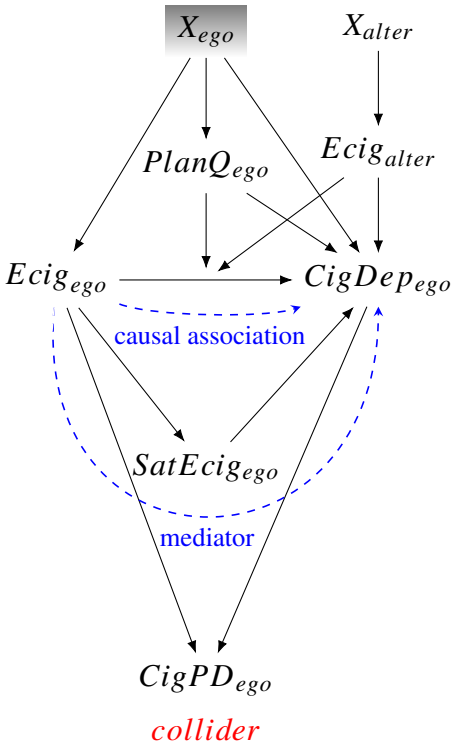
Analyzing the PATH data for pairs of adults who are smokers, we study the effects of the alter adopting e-cigarettes on the success of the ego in reducing cigarette dependence through their own adoption of e-cigarettes. Additionally, we investigate the optimal treatment regime for the ego and alter, in terms of their use or non-use of e-cigarettes, for reducing cigarette dependence in the ego.

We formulate our analysis as a three-stage decision problem and define the j^{th} stage ($j = 1, 2, 3$) to be the time from Wave j up to but not including Wave $j + 1$. We first define the final-stage observed outcome Y as the decrease in cigarette dependence between Wave 1 and Wave 4 (with a larger change being preferred).

For the j^{th} stage, we select the Wave j variables: age, education, sex, non-Hispanic, race and "plan to quit" denoted, respectively, as the components of $\boldsymbol{x}_j^{\beta} = (x_{j1}, x_{j2}, x_3, x_4, x_5, x_{j6})^{\top}$ (Benmarhnia et al., 2018). The non-Hispanic, race, and "plan to quit" variables are indicator functions for the related yes or no questions. The age variable is an indicator of "less than 35" or



(a) PATH causal graph depicting an example of how confounding and causal association flow. E-cigarettes use ($Ecig$) is defined as the administered treatment; Cigarettes dependence ($CigDep$) is outcome; $SatEcig$ and $CigPD$ refer to "satisfaction" variables from e-cigarettes and cigarettes per day, respectively.



(b) Confounding association is blocked by conditioning on full confounders (X_{ego}), causal association (blue dashed lines) flows from $Ecig$ to its children - $CigDep$ and $SatEcig$ (mediator), and non-causal association is blocked by (not conditioning on) $CigPD$ (collider).

Figure 3.8: Directed acyclic graphs of PATH analysis. Note: Subscript ego or alter stands for the variables are from ego or alter respectively. $PlanQ_{ego}$ ("plan to quit") and $Ecig_{alter}$ as moderators are selected as tailoring variables.

not, and the sex variable is set as an indicator of "Male". In particular, the education covariates are set as a three-category variable that is determined by the education question, that is, designating "less than high school" as 1, "between high school and bachelor degrees" as 2, then "bachelor or advanced degree" as 3. We propose a dependency structure as depicted in the directed acyclic graph of Figure 3.8. Here, we note the existence of e-cigarettes "satisfaction" variables that are dependent on e-cigarette use (the treatment) and in turn influence cigarette dependence (the outcome) and thus act as mediators. In addition, some smoking variables that are descendants of both e-cigarette use and cigarette dependence function as colliders. To identify the total effect of e-cigarette use on cigarette dependence we should not condition on these mediators and colliders or any descendants of the treatment (Pearl, 2000). We adjust for all measured potential confounders $\boldsymbol{x}_j^{\beta}$ to block the confounding association path.

The tailoring variables are those associated with the efficacy of the treatment and are to be selected from the set of moderator variables (Almirall et al., 2014). Le Grande et al. [2021] suggest the importance of age as a moderator (or a stratification variable) in the relationships of prior wave predictors of quitting or reducing smoking. In addition, "plan to quit" - a variable that measures motivation to quit or reduce consumption - may also influence the efficacy of the treatment of the ego. Finally, the use of e-cigarettes by the alter (denoted a^{alter}) is naturally of interest due to the possibility of interference. We therefore select, at each stage, the variables age ("less than 35" or "35+"), "plan to quit" and the alter's treatment, as tailoring variables $\boldsymbol{x}_j^{\psi} = (x_{j1}, x_{j6}, a_{j+1}^{alter})^{\top}$.

Accordingly, we set up the blip model as $\gamma\left(a_{j+1}, \boldsymbol{x}_j^{\psi}; \boldsymbol{\psi}_j\right) = a_{j+1}\left(\psi_{j0} + \boldsymbol{\psi}_{j1}^{\top}\boldsymbol{x}_j^{\psi}\right)$, and the treatment-free model as $f_j(\boldsymbol{x}_j^{\beta}; \boldsymbol{\beta}_j, \boldsymbol{\xi}_j) = \beta_{j0} + \boldsymbol{\beta}_{j1}^{\top}\boldsymbol{x}_j^{\beta} + \xi_{j0}a_{j+1}^{alter}$.

We conduct three sets of analyses corresponding to those carried out in the simulations of Section 3.4.1. For Method 2 (M2) and Method 3 (M3), the outcome model is employed with both standard dWOLS weights and the network weights. That is, in M2, for the standard dWOLS weights, $w_j = \left|a_{j+1} - \mathbb{E}\left[A_{j+1}|\boldsymbol{x}_j^{\alpha}\right]\right|$, where the propensity scores are estimated by maximum likelihood under a logistic regression model, $logit(\mathbb{E}\left[A_{j+1}|\boldsymbol{x}_j^{\alpha}\right]) = \alpha_{j0} + \boldsymbol{\alpha}_{j1}^{\top}\boldsymbol{x}_j^{\alpha}$, for $\boldsymbol{x}_j^{\alpha} = (x_{j1}, x_{j2}, x_3, x_4, x_5, x_{j6})^{\top}$, and in M3, for the network weights, $w_j^{net} = \left|a_{j+1} - \mathbb{E}\left[A_{j+1}|\boldsymbol{x}_j^{\alpha}\right]\right| * \left|a_{j+1}^{alter} - \mathbb{E}\left[A_{j+1}^{alter}|\boldsymbol{x}_j^{\alpha,alter}\right]\right|$. Crucially, we note that the outcome model with weight $w_j = \left|a_{j+1} - \mathbb{E}\left[A_{j+1}|\boldsymbol{x}_j^{\alpha}\right]\right|$ corresponds to that of M2 in simulation Study 1 in Section 3.4.1, while the outcome model using network weights corresponds to that of M3. In addition, for Method 1 (M1), in which we do not consider interference, we exclude the interference term (a^{alter}) in both the treatment-free model and the blip model, and employ the standard dWOLS weights. Finally, we note that because survey weights multiply the estimating function terms for the regression models, the sample estimating functions are assumed to be sampling-design-unbiased estimates of their population counterparts.

Under the conditions for double robustness, the dWOLS or network weights are constructed to ensure unbiased population-level estimating functions for estimation of the blip parameters, and the design-unbiasedness of the corresponding sample estimating functions makes the sample estimating functions unbiased with respect to the combination of model and sampling design. Accordingly, in the terminology of survey methods, the estimators of the blip function parameters will be model-design consistent (Lumley, 2004).

3.5.2 Assessing Balance

Before moving on to present the PATH data analysis results, to make sure sufficient overlap when the alter's treatment is fixed, we also checked the overlap of the ego's propensity distributions for treated and untreated groups, given the alter's treatment.

To check the balancing performance of the weights, we measure covariate mean balance for each covariate by the absolute standardized mean difference:

$$\left| \frac{\sum_{i=1}^{N} \boldsymbol{x}_i A_i w_i}{\sum_{i=1}^{N} A_i w_i} - \frac{\sum_{i=1}^{N} \boldsymbol{x}_i \left(1 - A_i\right) w_i}{\sum_{i=1}^{N} \left(1 - A_i\right) w_i} \right| / \sqrt{s_1^2/N_1 + s_0^2/N_0},$$

where s_a^2 is the variance of the unweighted covariate in group $a = 0, 1$ and N_a is the sample size in group a, that is, Equation 12 in Li et al. [2018]. For example, the distributions of the fitted propensity scores for Stage 3 (i.e., Wave 3 to Wave 4), by values of the ego's treatments when the alter's treatment is fixed at 1, are given in Figure 3.9 (left plot), and by values of the ego's treatments when the alter's treatment is fixed at 0 are given in Figure 3.10 (left plot). In addition, the propensities of the cases for Stages 1 and 2 appear the same degree of overlap. Based on three weighting schemes in the PSweight package (Zhou et al. [2020]), in Figures 3.9 and 3.10 (right plot) we present the standardized mean difference for each covariate, and assess the weighted covariate balance in a single plot.

For Figures 3.9 and 3.10 (right plots), the covariates are imbalanced across the treated and untreated groups prior to any weighting, conditional on alter's treatment being 1 and 0, respectively. Inverse probability weighting can generally improve covariate balance, but overlap weights (absolute value weights) provide the best balance (in terms of overlap) across all covariates. For Stages 1 and 2, similarly, the balancing results appear the same pattern.

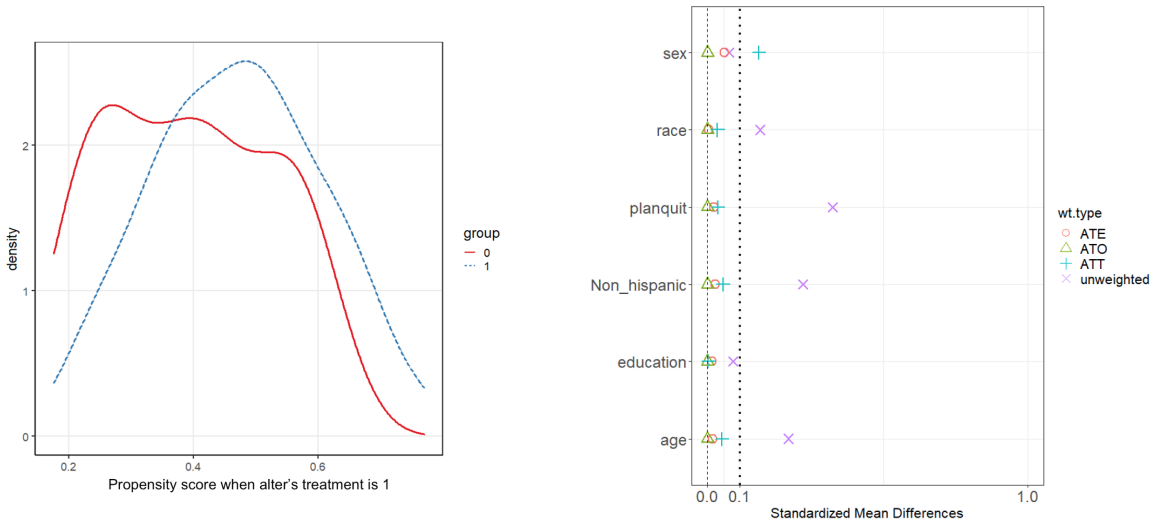


Figure 3.9: Conditional on alter's treatment is 1, the left plot is the density of estimated propensity scores with respect to the treatment variable for decision Stage 3. Group 0 stands for the untreated group and Group 1 for the treated group. The right plot presents the maximum pairwise absolute standardized difference plot using different weights for Stage 3, conditional on alter's treatment is 1. (ATE specifies the inverse probability weights for estimating the average treatment effect among the combined population. ATT specifies the weights for estimating the average treatment effect among the treated. ATO specifies the overlap weights for the average treatment effect among the overlap population.)

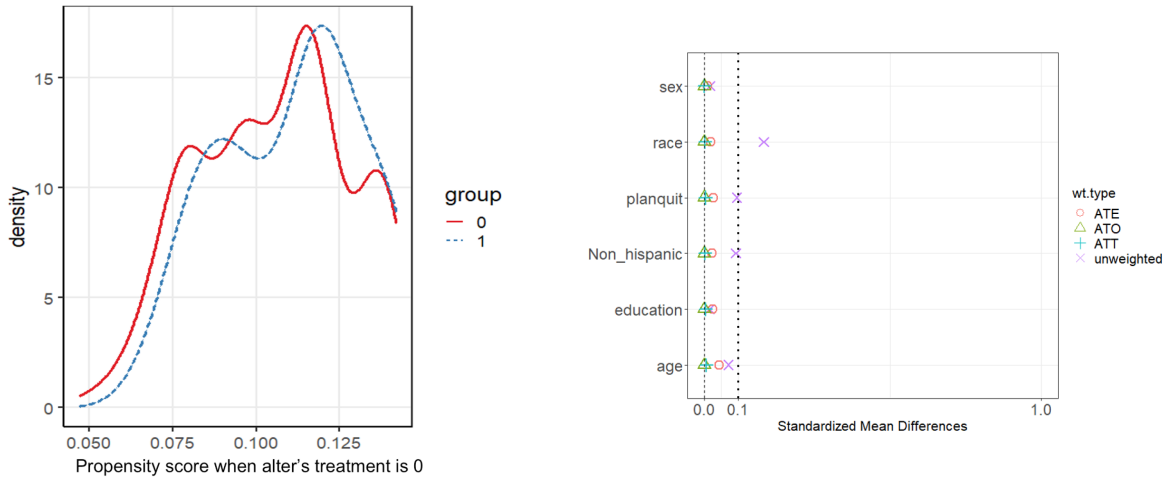


Figure 3.10: Conditional on alter's treatment is 0, the left plot is the density of estimated propensity scores with respect to the treatment variable for decision Stage 3. Group 0 stands for the untreated group and Group 1 for the treated group. The right plot presents the maximum pairwise absolute standardized difference plot using different weights for Stage 3, conditional on alter's treatment is 0. (ATE specifies the inverse probability weights for estimating the average treatment effect among the combined population. ATT specifies the weights for estimating the average treatment effect among the treated. ATO specifies the overlap weights for the average treatment effect among the overlap population.)

Table 3.3: Analysis and optimal DTRs of PATH data. Optimal DTRs are indicator functions of $\boldsymbol{\psi}$ estimates. If a^{alter} is not pre-determined, optimal DTRs also rely on spillover $\boldsymbol{\xi}$ estimates: first by determination of A^{alter}, and then by determination of the ego treatment, with a view to optimization for the ego.

Wave	Methods	Optimal DTRs	Spillover effects estimates $\hat{\xi}$
$1 \sim 2$	M1	$\mathbb{I}[3.46 - 5.81x_{11} - 0.20x_{16}]$	$\times$
	M2	$\mathbb{I}[7.18 + (-6.16, 1.87, -7.56) \cdot \boldsymbol{x}_1^{\psi}]$	$\mathbb{I}[3.72]$
	M3	$\mathbb{I}[11.32 + (-19.12, -1.17, -5.04) \cdot \boldsymbol{x}_1^{\psi}]$	$\mathbb{I}[3.00]$
$2 \sim 3$	M1	$\mathbb{I}[-2.01 - 0.96x_{21} - 7.01x_{26}]$	$\times$
	M2	$\mathbb{I}[-3.47 + (-0.16, 7.62, 4.55) \cdot \boldsymbol{x}_2^{\psi}]$	$\mathbb{I}[1.72]$
	M3	$\mathbb{I}[2.34 + (1.85, 0.30, 4.96) \cdot \boldsymbol{x}_2^{\psi}]$	$\mathbb{I}[1.14]$
$3 \sim 4$	M1	$\mathbb{I}[-0.45 + 0.86x_{31} + 3.08x_{36}]$	$\times$
	M2	$\mathbb{I}[-3.87 + (2.16, 2.64, 7.93) \cdot \boldsymbol{x}_3^{\psi}]$	$\mathbb{I}[-0.30]$
	M3	$\mathbb{I}[-9.23 + (3.25, 7.28, 8.42) \cdot \boldsymbol{x}_3^{\psi}]$	$\mathbb{I}[0.17]$

Note: Method 1 (M1) does not include an interference term, so has no ($\times$) ξ estimates, and its decision rules only depend on age (x_{j1}) and plan to quit (x_{j6}). $\mathbb{I}(x) = \mathbb{I}_{(0,\infty)}(x)$ is the indicator that x is in the set $(0, \infty)$, and $u \cdot v$ means the dot product of vectors u and v.

3.5.3 PATH Analysis Results

Following the balancing checking of ego's covariates, we will now present the results of the PATH analysis. The blip function and spillover effect estimates from Methods 1, 2 and 3, and corresponding optimal DTRs are summarized in Table 3.3. The blip function coefficient estimates from Methods 2 and 3 exhibit some difference, whereas the spillover estimates are largely similar. We summarize the recommended use of e-cigarettes based on the results of Method 3 (network weights). If a^{alter} is pre-determined; in the first stage, if the ego is 35 years or older, do not use e-cigarettes; otherwise, use e-cigarettes. In the second stage, use e-cigarettes. In the last stage, if $a^{alter} = 1$ and $planquit = 1$, use e-cigarettes; if $a^{alter} = 1$, $planquit = 0$, and the ego is 35 years or older, use e-cigarettes; if $a^{alter} = 0$, $planquit = 1$ and the ego is 35 years or older, use e-cigarettes; otherwise do not. The value function captures the performance of the treatment regimes. For example, while the mean of the observed value or outcome (i.e., $DiffDep14$) is 4.75, the means of the outcomes of DTRs from Methods 1, 2, and 3 are 10.68, 12.04 and 14.83, respectively.

We note that this recommendation presumes that the alter's treatment is known, whereas in

practice we may wish to make a recommendation to both the ego and alter simultaneously. If desired, the results of our analysis may be applied in this setting by comparing the four expected outcomes that arise from both ego and alter having two treatment options.

3.6 Conclusion and Discussion

Dynamic treatment regimes are an essential tool in the analysis and implementation of precision medicine. However, little work has been done to study, or address, the obstacles interference can pose to DTR estimation. Even though many current methods such as G-estimation and dWOLS have focused on model robustness, none have sought to directly investigate interference and how it may be accounted for.

In this chapter, using the method of dynamic weighted ordinary least squares, we have demonstrated the impact a failure to account for interference can have on optimal DTR estimation. We have shown how interference may be accommodated using the dWOLS method, first, through the explicit inclusion of interference terms in dWOLS' constituent models, and second, by extending dWOLS through the use of network weights to more directly account for interference.

While both standard dWOLS and dWOLS with network weights afford consistent parameter estimation in the presence of interference, we have shown through simulation that the latter can afford greater flexibility in modeling, in addition to providing efficiency gains. Although this estimation does require the correct identification of the network interference term $t(a_N)$, which is application-specific, we note that any approach that accounts for interference must incorporate some form of additional information or modeling. It is important to recognize that we have only considered a setting where there is a specific form for the function t. Future work could conduct sensitivity analyses that investigate the effects of misspecification of this function.

Future work should also consider alternative approaches to estimation of the treatment model. Here, we have used a logistic model for two primary reasons. First, absolute value weights (i.e., overlap weights) based on a logistic propensity score model have the small-sample exact balance property (Theorem 3 in Li et al. [2018]). Second, logistic regression is a method with widespread recognition and understanding, and comparably straightforward to implement (which itself is one of the motivations for using the dWOLS method). However, extensions that employ fancier machine learning tools such as BART (Bayesian Additive Regression Tree, Chipman et al., 2010), Super Learner [Van der Laan et al., 2007] or deep neural networks for estimating propensity scores (see the discussion in Herren and Hahn, 2020) could be studied.

Our treatment models have also been grounded in some key assumptions, in particular that treatment is conditionally independent of the treatments of other individuals, given the covariates

of all. In observational studies, the treatment assignment could depend on the covariates, treatment and outcomes of neighbours. If so, the blip function should therefore also rely on this information, through interaction terms such as $a\tau(\mathbf{x}_N)$, where τ has to be equivalent to a fixed, known function. We study interference where the treatments of neighbours can affect only the outcome; consequently, relaxing of the independent treatments assumption should be addressed in future work. In addition, we also assume that individuals' propensities depend on their own covariate values, which could depend to a limited degree on network characteristics.

In the estimation of our simulations, the data used belong to egos that have the same number of neighbours in the network; however, we again note that we can consider the network as the union of subsets of constant degree, and combine the estimating function systems corresponding to different degree values in an efficient manner, for example, by using stabilized (rescaled) weights to reduce the variability of weights overall. In addition, an important next step is the multiple-stage decision problem with time-varying of the network structure. In such a problem, the network structure information at each stage (such as number of neighbours) could be different.

Through our analysis of the PATH study, we have demonstrated how our proposed methods may be applied in practice. We acknowledge that our analysis, intended primarily as an illustration, has been limited by a number of simplifying assumptions. In addition to the assumption of non-informative dropouts (which may, in principle, be accommodated through an analysis using the inverse probability of censoring weights), our results should be set in the broader context of a changing e-cigarette landscape as new products emerge and trends develop. More fundamentally, we note that in this context a prescription has not been clearly identified [Villanti et al., 2018]. For all these reasons, we do not intend the results to be seen as an actual recommendation for a treatment strategy. Nevertheless, the analysis serves as an informative demonstration of both the underlying principles of interference in such contexts, and of our proposed methodology.

A more in-depth analysis of these, and related, data within the context of interference is certainly well-warranted. In the estimation process, various checks of robustness such as analyses of sensitivity to model and estimation uncertainty would have to be carried out in order for the results to be useful. For example, one possible robustness check is to specify two pairs of adults as one network and see if they obtain, up to statistical uncertainty, the same DTR.

We observe that, as illustrated in the PATH analysis, our strategy for constructing treatment decisions may depend on whether we may assume that the alters' treatments are pre-determined or whether they should be considered part of the treatment allocation process. Even though we have largely operated within the former framework, it is certainly possible to apply our methods to the latter. In the couples case, for example, we may seek to maximize the average outcome across a couple, comparing the four possible treatment combinations (both treated, neither treated, and one or the other treated). How this would extend to more complex networks is a more challenging, but

interesting, problem.

Finally, we note that, as this is one of the first pieces of work on the specific topic of interference in DTRs, there are many questions that remain unanswered. In addition to the limitations outlined above, we also observe in particular that extensions to outcomes and treatments of different types (such as binary for the former and continuous for the latter) would be a natural next step for this research. With recent extensions to dWOLS for continuous treatments (Schulz and Moodie, 2021) and survival outcomes (Simoneau et al., 2020), there is already a strong foundation upon which this work can build.

Chapter 4

Optimal DTR Estimation for Binary Outcomes

In Chapter 3, considering continuous outcomes, we developed the methods for DTR estimation in the presence of network interference. From here on, we investigate optimal DTR estimation in the context of binary outcomes. In Chapter 4, we study DTR estimation for binary outcomes without interference. Building on Chapter 4, Chapter 5 studies DTR estimation in situations of binary outcomes with interference.

4.1 Introduction

Most optimal DTR estimation methodologies focus on continuous outcomes. These include regression-based methods such as Q-learning (Sutton and Barto [2018]), G-estimation (Robins [2004]) and dynamic weighted ordinary least squares (dWOLS, Wallace and Moodie [2015]), and value-search methods such as (augmented) inverse probability of treatment weighting (Zhang et al. [2013]) and outcome weighted learning (Zhao et al. [2012]). As a continuous-outcome problem counterpart, the discrete-outcome problem is more challenging, yet valuable in real-world applications. Binary outcomes, such as treatment failure or success, are crucial measures in many medical or health studies. However, to date, optimal DTR estimation for binary outcomes has received little attention.

Although most methodologies target continuous outcomes, there are some existing studies regarding non-continuous outcomes. Some theoretical developments in the DTR literature have focused on discrete-outcome settings, including Q-learning with discrete outcomes (Bernoulli and

Poisson) utilities (Moodie et al. [2014]), an extension of G-estimation to the case of non-additive treatment effects for discrete outcomes (Wallace et al. [2019]), and the extension of dWOLS to time-to-event data with survival outcomes subject to right-censoring (Simoneau et al. [2020]).

For binary outcomes, the recently proposed DTR estimation approaches are reliant on either Q-learning, which offers relatively straightforward implementation, or G-estimation, which is *doubly robust* in the sense of offering a consistent estimator of a treatment effect if at least one of two nuisance models is correctly specified. For example, considering cases of cancer and graft-versus-host disease treatment, to maximize the probability of the binary outcome of two-year disease-free survival, Moodie and Krakow [2020] implemented Q-learning in a multi-stage treatment decision analysis, employing logistic regression at each stage. This method was shown to be easy to implement, but suffered from problems of sensitivity to misspecification of the outcome model. Wang et al. [2017], meanwhile, proposed a G-estimation based method for binary multiplicative structural nested mean models. They constructed a locally semiparametric efficient estimator, and demonstrated that such estimators boast the aforementioned double robustness property. Recently, analyzing micro-randomized trial data with binary outcomes, Qian et al. [2019] defined the causal excursion effect, which refers to a log relative risk between two excursions from a treatment protocol, and also provided a semiparametric and locally efficient estimator of the causal excursion effect. Based on the research of Robins (Robins et al. [1994], Robins [2004]), these two semiparametric methods are doubly robust. However, their drawbacks include complexity of theory and implementation which may prove challenging for many practitioners.

In contrast to Q-learning and G-estimation, dynamic weighted ordinary least squares offers an approach to DTR estimation that is doubly robust while also relatively straigthforward to implement. In the case of identifying a multi-stage DTR (that is, a sequence of treatment decisions at fixed time points), dWOLS proceeds via a sequence of weighted ordinary least squares regressions. Building on this methodology, we propose the dynamic weighted generalized linear model (dWGLM); an extension of dWOLS to the case of binary outcomes that offers similar properties in terms of robustness and ease of implementation.

Chapter 4 is organized as follows: Section 4.2 introduces the proposed doubly robust regression-based DTR estimation framework with binary outcomes, where we take the term doubly robust to include cases of approximate consistency of the estimator; Section 4.3 describes simulation studies, demonstrating the double robustness of our methods; Section 4.4 illustrates our methodology using observational data from the Population Assessment of Tobacco and Health study, and Section 4.5 concludes this chapter with a discussion.

4.2 Methodology

4.2.1 Preliminaries

Using the same necessary notation and settings in multi-stage decision estimation in the previous section, we assume that a DTR contains a total of K treatment stages in a multiple-stage treatment decision problem. Let Y denote the patient outcome, a binary variable that takes values in $\{0, 1\}$; it is observed after the assignment of all the treatments. We also assume that $Y = 1$ is preferred. Following the notations that were introduced in Section 2.2, again, over- and under-bar are used to represent the past and future, respectively. $\bar{a}_j$ indicates a vector of the first j treatment decisions and $\underline{a}_{j+1}$ denotes the vector of treatment decisions from Stage $j + 1$ onwards.

Consistent with regression-based DTR estimation approach in Section 2.2, a fundamental component is the blip function. Denoting $Y^*(\boldsymbol{a})$ as the potential (or counterfactual) binary outcome under treatment regime $\boldsymbol{a}$, we then define the optimal blip-to-reference function for Stage j as:

$$\gamma_j\left(\boldsymbol{h}_j, a_j\right) = g\left(\mathbb{P}\left[Y^*\left(\bar{a}_j, \underline{a}_{j+1}^{opt}\right) = 1 \mid \boldsymbol{H}_j = \boldsymbol{h}_j\right]\right) - g\left(\mathbb{P}\left[Y^*\left(\bar{a}_{j-1}, a_j^{ref}, \underline{a}_{j+1}^{opt}\right) = 1 \mid \boldsymbol{H}_j = \boldsymbol{h}_j\right]\right), \tag{4.1}$$

which is the difference in the g link function transformation of the mean of the binary outcome when using a reference treatment a_j^{ref} (such as a control) instead of a_j at Stage j, in individuals with history $\boldsymbol{h}_j$ who are subsequently optimally treated (i.e., receiving $\underline{a}_{j+1}^{opt}$). Note that, for binary outcomes, there are several options for the link function $g : (0, 1) \rightarrow (-\infty, +\infty)$, and if g is the canonical (logit) link for the binomial, i.e., $g(p) = log(p/(1 - p))$, the equation (4.1) blip function then denotes the log odds ratio of expected potential binary outcomes for patients who are treated versus untreated. When g is the identity link $g(p) = p$, the blip function represents the risk difference of expected potential binary outcomes for patients who are treated versus untreated. Other choices of link function could be the probit link $g(p) = \Phi^{-1}(p)$, where Φ is the cumulative distribution function (C.D.F.) of the standard normal distribution, and the complementary log-log function $g(p) = log(-log(1 - p))$. Further, the *robit link*, $g(p) = F_\nu^{-1}(p)$ where $F_\nu(x)$ is the C.D.F. of the Student's **t**-distribution with center zero, scale parameter one, and ν degrees of freedom, could be considered. Some appealing properties of the robit link have been identified. For example, the robit link can be considered as a generalization of probit link and an approximate generalization of logit link, and it provides a robust estimation in that the coefficient estimates are less influenced by individual outlying data points (Liu [2004]).

The outcome probabilities can be framed in the context of blip functions such that $g[\mathbb{P}(Y = 1)] = g[\mathbb{P}(Y^{opt} = 1)] - \sum_{j=1}^{K}\left[\gamma_j\left(\boldsymbol{h}_j, a_j^{opt}\right) - \gamma_j\left(\boldsymbol{h}_j, a_j\right)\right]$. A corresponding concept to blips is regrets (Murphy [2003]). In the generalized linear model (GLM) framework, the regret function

(for Stage j) can be defined as

$$\mu_j\left(\boldsymbol{h}_j, a_j\right) = g\left(\mathbb{P}\left[Y^*\left(\bar{a}_{j-1}, \underline{a}_j^{opt}\right) = 1 \mid \boldsymbol{H}_j = \boldsymbol{h}_j\right]\right) - g\left(\mathbb{P}\left[Y^*\left(\bar{a}_j, \underline{a}_{j+1}^{opt}\right) = 1 \mid \boldsymbol{H}_j = \boldsymbol{h}_j\right]\right),$$

which is the g (link) function transformation of expected loss or regret arising from prescribing treatment a_j at Stage j instead of the optimal treatment a_j^{opt}, assuming optimal treatment is received in the later stages. Thus, similar to the continuous outcome case, the blip and regret functions can be related such that $\mu_j\left(\boldsymbol{h}_j, a_j\right) = \gamma_j\left(\boldsymbol{h}_j, a_j^{opt}\right) - \gamma_j\left(\boldsymbol{h}_j, a_j\right)$, showing that, if the individual was optimally treated, the regret function will equal 0. Otherwise, the regret function will be positive. The positive value of regrets represents the expected improvement in outcome had the patient received optimal treatment compared to the observed treatment. The concept of regrets provide a distinct view to consider the effects of the treatment in terms of the optimal treatment, and it also simplifies some expressions in later sections.

Building on the concept of the blip and regret functions, the main goal of DTR estimation is to identify the optimal treatment decision that maximizes the blip function or equivalently minimizes the regret function. For instance, we consider the outcome model that can be decomposed into two components: $g(\mathbb{E}\left[Y^*(\boldsymbol{a})|\boldsymbol{H} = \boldsymbol{h}\right]) = \sum_{j=1}^{K}\left[f_j\left(\boldsymbol{h}_j^{\beta}; \boldsymbol{\beta}\right) + \gamma_j(\boldsymbol{h}_j^{\psi}, a_j; \boldsymbol{\psi}_j)\right]$, where $f_j\left(\boldsymbol{h}_j^{\beta}; \boldsymbol{\beta}\right)$ and $\gamma_j(\boldsymbol{h}_j^{\psi}, a_j; \boldsymbol{\psi}_j)$ are treatment-free and blip models, respectively, and $\boldsymbol{h}_j^{\beta}$ and $\boldsymbol{h}_j^{\psi}$ are subsets of covariates in $\boldsymbol{h}_j$. It is important to note that, analogously to the continuous outcome modeling in Section 2.2, our GLM modeling for binary outcomes follows the same mechanism: the optimal treatment is that which maximizes the blip or minimizes the regrets. That is, given the tailoring variable $\boldsymbol{h}_j^{\psi}$, typically a small subset of $\boldsymbol{h}_j$, the optimal treatment decision for Stage j is "$a_j^{opt} = 1,\ if\ \gamma_j(\boldsymbol{h}_j^{\psi}, 1; \boldsymbol{\psi}_j) > 0;\ a_j^{opt} = 0,\ otherwise$." Therefore, to make the optimal treatment decision, it is sufficient to estimate the parameters in the blip model, and consider the parameters in the treatment-free model as nuisance parameters.

4.2.2 Q-learning with Binary Outcomes

We first introduce Q-learning for binary outcomes as motivation and elucidation, then provide our proposed method. To identify the optimal DTRs, Q-learning recursively solves treatment decision problems starting from the last stage, and at each stage, the Q-function is defined as follows (Moodie et al. [2014], Moodie and Krakow [2020]):

$$Q_K\left(\boldsymbol{h}_K, a_K\right) = g\left(\mathbb{P}\left[Y^*(a_K) = 1 \mid \boldsymbol{H}_K = \boldsymbol{h}_K, A_K = a_K\right]\right);$$

and

$$Q_j\left(\boldsymbol{h}_j, a_j\right) = g\left(\mathbb{E}\left[\max_{A_{j+1}} g^{-1}\left[Q_{j+1}\left(\boldsymbol{H}_{j+1}, A_{j+1}\right)\right] \mid \boldsymbol{H}_j = \boldsymbol{h}_j, A_j = a_j\right]\right) \text{ for } j < K.$$

Suppose the Q-functions are modeled linearly such that $Q_j\left(\boldsymbol{h}_j, a_j; \boldsymbol{\beta}_j, \boldsymbol{\psi}_j\right) = \boldsymbol{\beta}_j^\top \boldsymbol{h}_j^\beta + \boldsymbol{\psi}_j^\top a_j \boldsymbol{h}_j^\psi$, where $\boldsymbol{h}_j^\beta$ and $\boldsymbol{h}_j^\psi$ are subsets of covariates in $\boldsymbol{h}_j$. Note that the pseudo-outcome-probability, because of the monotone increasing property of the g^{-1} function,

$$\max_{A_{j+1}} g^{-1}\left[Q_{j+1}\left(\boldsymbol{H}_{j+1}, A_{j+1}\right)\right] = g^{-1}\left[\max_{A_{j+1}} Q_{j+1}\left(\boldsymbol{H}_{j+1}, A_{j+1}\right)\right],$$

refers to the "best possible" probability of the outcome a patient could have based on the proposed outcome models in the preceding stage. Then the treatment decisions are made according to the estimates of $\boldsymbol{\psi}$ in each stage. For example, in Stage j, "$If\ \hat{\boldsymbol{\psi}}_j^\top \boldsymbol{h}_j^\psi > 0,\ treat; otherwise,\ leave\ untreated$." However, we note that we should correctly specify all Q-function models, including the treatment-free models, to acquire consistent estimators of $\boldsymbol{\psi}$.

4.2.3 Balancing Property

In the previous subsection, Q-learning provides a comparatively simple to follow method via a generalized linear model with binary outcomes, such as logistic regression, but it lacks robustness to misspecification of treatment-free models. Our proposed approach, which will be described in this section, inspired by dWOLS, employs balancing weights to overcome the possible misspecification of these models.

dWOLS employs a series of sequential weighted regressions to consistently estimate the parameters of interest in the outcome model. The double robustness of dWOLS relies on balancing weights, which are a function of the propensity score and thus determined by the underlying treatment model. Consistent with the notation in Section 2.2, we denote the standard propensity score as $\pi(\boldsymbol{x}) := \mathbb{P}(A = 1 \mid \boldsymbol{x})$. Then the balancing weights criterion introduced in dWOLS establish independence between the covariates and treatment in the weighted dataset. Thus, the bias in estimating the blip parameter, introduced due to the dependence between covariates and treatment is removed. We denote by w^d a choice of dWOLS balancing weights that satisfy the criterion $(1 - \pi(\boldsymbol{x}))w(0, \boldsymbol{x}) = \pi(\boldsymbol{x})w(1, \boldsymbol{x})$ as proposed in Theorem 1 of Wallace and Moodie [2015].

Suppose that $\boldsymbol{x}^\beta$ and $\boldsymbol{x}^\psi$ are two subsets of the covariates included in $\boldsymbol{x}$, and that the true

outcome model is $g\left(\mathbb{P}[Y=1|\boldsymbol{x},a]\right)=f(\boldsymbol{x}^{\beta};\boldsymbol{\beta})+\gamma(\boldsymbol{x}^{\psi},a;\boldsymbol{\psi})$, where $\gamma(\boldsymbol{x}^{\psi},a;\boldsymbol{\psi})$ is in the linear form $\boldsymbol{\psi}^{\top}a\boldsymbol{x}^{\psi}$ but f is an arbitrary function, and g is the link function for binary outcomes. Then, the following theorem holds, where the term *approximately consistent* is explained below.

Theorem 4.1. *A technique approximating balance for GLM with binary outcomes*

When the true outcome model satisfies $g\left(\mathbb{P}[Y=1|\boldsymbol{x},a]\right)=f(\boldsymbol{x}^{\beta};\boldsymbol{\beta})+\gamma(\boldsymbol{x}^{\psi},a;\boldsymbol{\psi})$, *where* g *is the link function, a weighted generalized linear model based on the corresponding linear predictor will yield approximately consistent estimators of* $\boldsymbol{\psi}$ *if the weights satisfy*

$$(1-\pi(\boldsymbol{x}))w(0,\boldsymbol{x})\kappa(0,\boldsymbol{x})=\pi(\boldsymbol{x})w(1,\boldsymbol{x})\kappa(1,\boldsymbol{x}), \tag{4.2}$$

where $\kappa(a,\boldsymbol{x})=g^{-1\prime}(\boldsymbol{\beta}^{*\top}\boldsymbol{x}^{\beta}+\boldsymbol{\psi}^{*\top}a\boldsymbol{x}^{\psi})$ *and* $g^{-1\prime}(\mu)$ *is the first derivative of the inverse link function (i.e.,* $g^{-1\prime}(\mu)=\frac{dg^{-1}(\mu)}{d\mu}$*), and* $\boldsymbol{\beta}^{*}$ *and* $\boldsymbol{\psi}^{*}$ *are defined through* $\mathbb{E}\left((1-\pi(\boldsymbol{x}))\boldsymbol{x}w^{d}(0,\boldsymbol{x})\left[g^{-1}(f(\boldsymbol{x}^{\beta}))-g^{-1}(\boldsymbol{\beta}^{*\top}\boldsymbol{x}^{\beta})\right]\right)=\mathbf{0}$ *and* $\mathbb{E}\left(\pi(\boldsymbol{x})\boldsymbol{x}w^{d}(1,\boldsymbol{x})\left[g^{-1}(f(\boldsymbol{x}^{\beta})+\boldsymbol{\psi}^{*\top}\boldsymbol{x})-g^{-1}(\boldsymbol{\beta}^{*\top}\boldsymbol{x}^{\beta}+\boldsymbol{\psi}^{*\top}\boldsymbol{x})\right]\right)=\mathbf{0}$*, respectively.*

Proof of Theorem 4.1: See Appendix G.1. In addition, building on the systems of estimating equations in this proof, Appendix G.2 presents the proof of the uniqueness of $\boldsymbol{\beta}^{*}$.

The balancing weights criterion (equation 4.2) is similar to that used in dWOLS, both being built on propensity scores. However, equation (4.2) contains extra factors $\kappa(a,\boldsymbol{x})$. This $\kappa(a,\boldsymbol{x})$ is related to the g link function, and it is derived, as shown in the proof G.1, from the system of estimation functions of the GLMs for the purpose of addressing the misspecification of the treatment-free model. Table 4.1 provides some examples of link functions and their corresponding κ functions. It is important to emphasize that one may employ simply a linear predictor for any form of the true treatment-free model in a GLM. Even if the true treatment-free model is non-linear, the estimator of $\boldsymbol{\psi}$ is guaranteed to be "approximately consistent", where the use of "approximately" originates from the error term in the Taylor series approximation of the inverse link function being close to zero. (see proof of Theorem 4.1 in Appendix G.1).

We use estimating function theory to assess the consistency of the blip estimators. If the estimating function is unbiased, the estimator will be consistent in an appropriate framework, i.e., an appropriate rule for the growth of the sample size. If an estimating function is approximately unbiased with a small quantifiable bias, the estimate in the limiting case will be approximately unbiased with a small quantifiable bias as well. Thus, our term "approximately consistent" means the second case where the estimators are derived from the estimating functions which are approximately unbiased with a small quantifiable bias. That is, in the proof of Theorem 4.1 in Appendix G.1, by the Taylor expansion of the inverse link function (g^{-1}), we can connect the two estimation equations of two different treatment groups, and thus construct the corresponding

Table 4.1: Examples of link functions and their corresponding κ functions (ϕ and Φ denote the PDF and CDF of the standard Gaussian random variable, and f_ν and F_ν denote the PDF and CDF of the Student's **t**-distribution with center zero, scale parameter one, and ν degrees of freedom).

Link function	$g(\mu)$	$g^{-1}(\mu)$	$\kappa(\mu) = g^{-1\prime}(\mu)$
Identity	μ	μ	1
Logit	$\log(\frac{\mu}{1-\mu})$	$\frac{\exp(\mu)}{1+\exp(\mu)}$	$\frac{\exp(\mu)}{1+\exp(\mu)}[1 - \frac{\exp(\mu)}{1+\exp(\mu)}]$
Probit	$\Phi^{-1}(\mu)$	$\Phi(\mu)$	$\phi(\mu)$
Complementary log-log	$\log[-\log(1-\mu)]$	$1 - \exp[-\exp(\mu)]$	$\exp[\mu - \exp(\mu)]$
Robit	$F_\nu^{-1}(\mu)$	$F_\nu(\mu)$	$f_\nu(\mu)$

balancing weights criterion. Accordingly, the linear Taylor expansion induces an error term, and omitting this error term leads to an estimating function that is approximately unbiased. However, this error term will be small when a linear predictor tends to vary in an interval where g^{-1} is approximately linear. In some applications, to acquire a more accurate estimation, it may be possible to choose the range of covariates so that the linear predictor varies in such an interval.

Therefore, from the standpoint of robust estimation of the GLM with g link, we call this $\kappa(a, \boldsymbol{x})$ an "adjustment" factor in the balancing weights criterion for binary outcomes. The "adjustment" factor is a κ function of the linear predictor $\boldsymbol{\beta}^{*\top}\boldsymbol{x}^{\beta} + \boldsymbol{\psi}^{*\top}a\boldsymbol{x}^{\psi}$, where $\boldsymbol{\beta}^*$ and $\boldsymbol{\psi}^*$ are roots of the estimating functions of the GLMs with standard dWOLS weights w^d, and the κ function is the the first derivative of the inverse link function, that is, $\kappa(\mu) = g^{-1\prime}(\mu) = \frac{dg^{-1}(\mu)}{d\mu}$. Consequently, to construct the balancing weights for GLM with binary outcomes, two crucial steps are required: (1) identify the "adjustment" factor by conducting a weighted GLM (e.g., logistic regression) with the standard dWOLS weights; (2) compute the balancing weights based on the estimated "adjustment" factor, propensity score, and weights criterion 4.2. These two steps are illustrated in our proposed method of optimal DTR estimation with binary outcomes in the next section.

4.2.4 Dynamic Weighted Generalized Linear Model

Inspired by the easy implementation of Q-learning and the double robustness property of G-estimation and dWOLS, our proposed method, the dynamic weighted generalized linear model (dWGLM), estimates the blip parameters in terms of binary treatments and outcomes. Note that, for the multiple-stage decision problems, because of prognostic effects and delayed treatment effects (Kosorok and Laber [2019]), the current treatment decisions will not only affect the intermediate outcomes but also affect the future ones; thus, the decisions should be "farsighted". Backward induction is used in Q-learning for sequential decision problems. Thus the sequential

decision problems can be divided into a set of single-stage problems, each of which aims to optimize the stage specific *pseudo-outcomes*: the potential outcomes if the patients were treated - possibly contrary to fact - optimally at subsequent stages. Similar to the process of Q-learning with binary outcomes, dWGLM involves a series of weighted generalized linear models of either the observed outcome y (at Stage K) or *binary pseudo-outcomes* $\widetilde{\mathcal{Y}_j}$ (for stages $j < K$) on subject histories. These binary pseudo-outcomes are random variables from the Bernoulli distribution with success probability $\mathbb{P}(\widetilde{\mathcal{Y}_j} = 1) = g^{-1}\left[g[\mathbb{P}(Y = 1 \mid \boldsymbol{h}_K, a_K; \hat{\boldsymbol{\beta}}_K, \hat{\boldsymbol{\psi}}_K)] + \sum_{k=j+1}^{K} \mu_k\left(\boldsymbol{h}_k, a_k; \hat{\boldsymbol{\psi}}_k\right)\right]$. In continuous outcome G-estimation and dWOLS settings, the pseudo-outcome definition relies on the final observed outcome y, that is, $\widetilde{\mathcal{Y}_j} = y + \sum_{k=j+1}^{K} \mu_k\left(\boldsymbol{h}_k, a_k; \hat{\boldsymbol{\psi}}_k\right)$. For the binary case, however, we concentrate on probabilities of the outcome being one rather than directly using the observed binary outcome y. Therefore, we employ the last stage model to estimate $\mathbb{P}(Y = 1 \mid \boldsymbol{h}_K, a_K)$, and then combine the regrets to acquire the pseudo-outcome probabilities. We also emphasize that, in keeping with the goal of GLM, we mainly focus on modeling the probability that the pseudo-outcome equals one, i.e., $\mathbb{P}(\widetilde{\mathcal{Y}_j} = 1 \mid \boldsymbol{h}_j, a_j)$, for each stage ($j = 1, 2, ..., K$). Moreover, to improve the efficiency of $\boldsymbol{\psi}$ estimators, we construct the $\widetilde{\mathcal{Y}}$ multiple times (say R times) in each stage, and implement the estimation R times in each stage. Therefore, for the multistage decision analysis, the dWGLM procedure could be implemented by the following steps at each stage of the analysis, starting from the last stage K and working backwards towards the first stage:

- **Step 1:** Construct the stage j *pseudo-outcome*: set $\widetilde{\mathcal{Y}_j} = y$ if $j = K$. Otherwise, use prior estimates $\hat{\boldsymbol{\beta}}_K$ and $\underline{\hat{\boldsymbol{\psi}}}_{j+1} = (\hat{\boldsymbol{\psi}}_{j+1}, ..., \hat{\boldsymbol{\psi}}_K)$ to randomly generate $\widetilde{\mathcal{Y}_j}$, which takes the value 1 with probability $\mathbb{P}(\widetilde{\mathcal{Y}_j} = 1) = g^{-1}\left[g[\mathbb{P}(Y = 1 \mid \boldsymbol{h}_K, a_K; \hat{\boldsymbol{\beta}}_K, \hat{\boldsymbol{\psi}}_K)] + \sum_{k=j+1}^{K} \mu_k(\boldsymbol{h}_k, a_k; \hat{\boldsymbol{\psi}}_k)\right]$, R times, to yield $\widetilde{\mathcal{Y}_j^1}, \widetilde{\mathcal{Y}_j^2}, ..., \widetilde{\mathcal{Y}_j^R}$.

- **Step 2:** Specify the stage j treatment model $\mathbb{E}\left[A_j | \boldsymbol{h}_j^{\alpha}; \boldsymbol{\alpha}_j\right]$. The treatment model parameters $\boldsymbol{\alpha}_j$ (estimated, for example, via logistic regression) are used to compute a weight w_j, such as $w_j = \left|a_j - \mathbb{E}\left[A_j | \boldsymbol{h}_j^{\alpha}; \hat{\boldsymbol{\alpha}}_j\right]\right|$.

- **Step 3:** Specify the stage j treatment-free and blip models, and perform a weighted generalized linear model of $\widetilde{\mathcal{Y}_j^r}$ on the terms in the treatment-free and blip models, using weights from Step 2 to get estimates $\hat{\boldsymbol{\beta}}_j^{old,r}, \hat{\boldsymbol{\psi}}_j^{old,r}$ for $r = 1, ..., R$; that is, for each $r = 1, ..., R$, use the model

$$g(\mathbb{E}[\widetilde{\mathcal{Y}_j^r} \mid a_j, \boldsymbol{h}_j; \boldsymbol{\beta}_j, \boldsymbol{\psi}_j]) = \boldsymbol{\beta}_j^{\top} \boldsymbol{h}_j^{\beta} + \boldsymbol{\psi}_j^{\top} a_j \boldsymbol{h}_j^{\psi}. \tag{4.3}$$

- **Step 4:** Use $\hat{\boldsymbol{\beta}}_j^{old,r}$, $\hat{\boldsymbol{\psi}}_j^{old,r}$ from Step 3 to compute

$$\kappa^r(a_j, \boldsymbol{h}_j) = g^{-1\prime}(\hat{\boldsymbol{\beta}}_j^{old,r^\top} \boldsymbol{h}_j^\beta + \hat{\boldsymbol{\psi}}_j^{old,r^\top} a_j \boldsymbol{h}_j^\psi),$$

  where g^{-1} is identified based on the link function in Step 3. Then, construct the new weights

$$w_j^{new,r}(a_j; \boldsymbol{h}_j) = |a_j - \mathbb{E}[A_j|\boldsymbol{h}_j^\alpha]| * \kappa^r(1 - a_j, \boldsymbol{h}_j). \tag{4.4}$$

- **Step 5:** Perform a weighted GLM with the new weights (i.e., $w_j^{new,r}(a_j; \boldsymbol{h}_j)$) to get revised estimates $\hat{\boldsymbol{\beta}}_j^r$, $\hat{\boldsymbol{\psi}}_j^r$ for each r. Estimate $\boldsymbol{\psi}_j$ by $\hat{\boldsymbol{\psi}}_j = R^{-1} \sum_r \hat{\boldsymbol{\psi}}_j^r$, then use parameter estimators $\hat{\boldsymbol{\psi}}_j$ to construct the j^{th} stage optimal treatment rule, which is $prescribe\ a_j = 1\ if\ \hat{\boldsymbol{\psi}}_j^\top \boldsymbol{H}_j^\psi > 0;\ then\ prescribe\ a_j = 0\ otherwise.$

- **Step 6:** Return to Step 1 and analyze Stage $j - 1$ if there are more stages to analyze.

Our proposed dWGLM approach thus contains at each stage a two-step GLM estimation process for binary outcomes. Each step uses GLM for binary outcomes (e.g., logistic regression) to estimate the parameters of interest. The first step could employ logistic regression with the dWOLS balancing weights, and acquire estimates ($\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\psi}}$). Building on these estimates and the weights function (equation 4.4) which satisfies weights criterion (4.2), we can obtain new balancing weights for binary outcomes. Thus, the second step will utilize the logistic regression again with the new balancing weights to estimate the parameter of interest.

dWGLM is doubly robust against misspecification of either the treatment or the treatment-free model. If we misspecify the treatment model but correctly specify the treatment-free model, the estimator of blip parameters will be consistent. Alternatively, if the treatment-free model is misspecified, but we employ the balancing weights that are derived from a correct treatment model, the approximate consistency of the blip parameters will also be ensured. In addition, we note that the blip parameters are only meaningful if the blip model is correctly defined. Similar to our approach in Chapter 3, dWOLS with interference, to specify the optimal treatment strategy, we also need to correctly specify the blip model. For continuous outcomes, Wallace et al. [2017a] develop methods for assessing the blip model specification, and similar problems for binary outcomes can be further investigated.

4.3 Simulation Studies

We now demonstrate the implementation and double robustness of dWGLM via two simulation studies that address problems in both single-stage decision and multi-stage decision settings. In the single-stage setting (Study 1), we consider four different scenarios to verify the double robustness property of our method. In each scenario, we also consider two different link functions and compare out dWGLM with proposed new weights with two alternatives: Q-learning and GLM with standard "absolute value" weights. To test the robust estimation ability of our methods, in Study 2, with its multi-stage decision settings, we examine two different data-generating processes that can be employed in different real situations. One (Study 2a) is analogous to Wallace and Moodie [2015]'s two-stage setting, while the other (Study 2b), which follows Moodie et al. [2012]'s setting, distinguishes between the components that are tailoring variables and those that are predictive variables including potential confounders. In each of Study 2a and 2b, we also test different misspecification cases to demonstrate the double robustness of dWGLM.

4.3.1 Single-stage Decision for Binary Outcomes

Our first simulations (*Study 1*) consider the case of a single-stage treatment decision with binary outcomes. The data-generating process is as follows. Patient information: $X_i \sim U(0, 2)$, where subscript i indicates patient-level data; treatment $\mathbb{P}(A_i = 1 \mid X_i) = \text{expit}[-2X_i + \sin(X_i) + X_i^2]$, where $\text{expit}(x) = [1 + exp(-x)]^{-1}$; outcome $g[\mathbb{P}(Y_i = 1)] = X_i + \log(|X_i|) + \cos(\pi X_i) + X_i^3 + A_i(\psi_0 + \psi_1 X_i)$, where both probit ($g(x) = \Phi^{-1}(x)$) and logit ($g(x) = x/(1 - x)$) links are considered. Note that the treatment-free function is set as a nonlinear function that $f(x) = x + \log(|x|) + \cos(\pi x) + x^3$, and its plot against x is shown in Figure 4.1. The blip function is set in the form $\gamma(x, a; \boldsymbol{\psi}) = a(\psi_0 + \psi_1 x)$ with $\psi_0 = -1, \psi_1 = 2$, so that the optimal treatment is given by $a^{opt} = \mathbb{I}(\psi_0 + \psi_1 x > 0)$ (or $a^{opt} = \mathbb{I}(x > 0.5)$). Our interest is then in estimating the blip parameters ψ_0, ψ_1. In this study, we consider three estimation methods. In *Method 0* (Q-learning in Moodie and Krakow [2020]) we propose GLM with no weights for binary outcomes. In *Method 1* (GLM with standard dWOLS weights), we consider GLM but with the original dWOLS "absolute value" weights (e.g., $w = |a - \mathbb{E}(A|\boldsymbol{x})|$). Finally, in *Method 2*, which is our proposed method dWGLM, we consider GLM with the proposed weights (4.4), constructed from the standard dWOLS weights and the estimates from the model of (4.3).

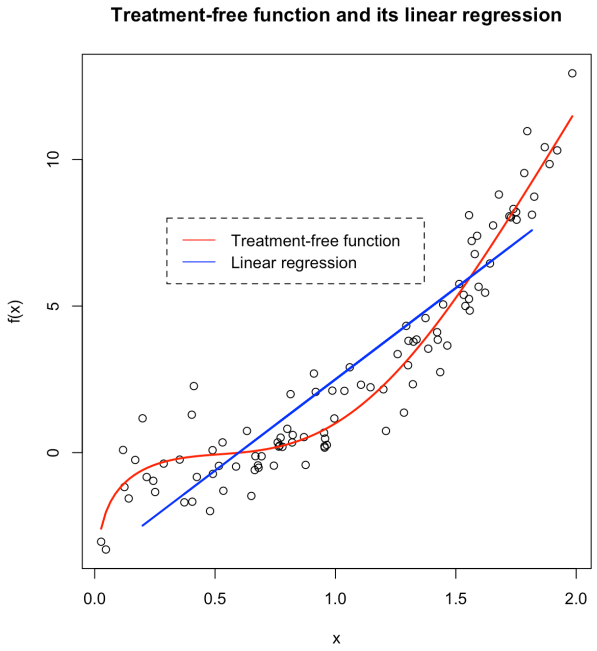


Figure 4.1: Nonlinear treatment-free function and its linear approximation

For each simulation, we conduct analyses in the following four scenarios: 1) both the treatment and treatment-free models are mis-specified; 2) the treatment-free model is mis-specified but the treatment model is correctly specified; 3) the treatment model is mis-specified but the treatment-free model is correctly specified; and 4) both the treatment and treatment-free models are correctly specified. Model mis-specification is implemented via the omission of non-linear terms in the treatment and treatment-free models.

Our simulation demonstrates the expected results as shown in Figure 2, which presents the results of the GLM with logit link, and those of the GLM with probit link appears a similar pattern. In the first two scenarios, where the treatment-free model is incorrectly specified, both Methods 0 and 1 provide biased estimators of blip function parameters. However, Method 2, the proposed dWGLM method with new balancing weights (4.4) offers blip estimators that are close to unbiased (and therefore likely to be close to consistent) in Scenario 2 and blip estimators with a small bias in Scenario 1. For the last two scenarios (Scenarios 3 and 4), because the treatment-free models were correctly specified, all of these three methods perform well in estimating the blip parameters; that is, they all provide close to consistent blip function parameter estimators. We note that, for dWGLM, comparing Scenario 1 with Scenario 2 (or Scenario 3 with Scenario 4), we observe a gain in efficiency due to the correct specification of the treatment model.

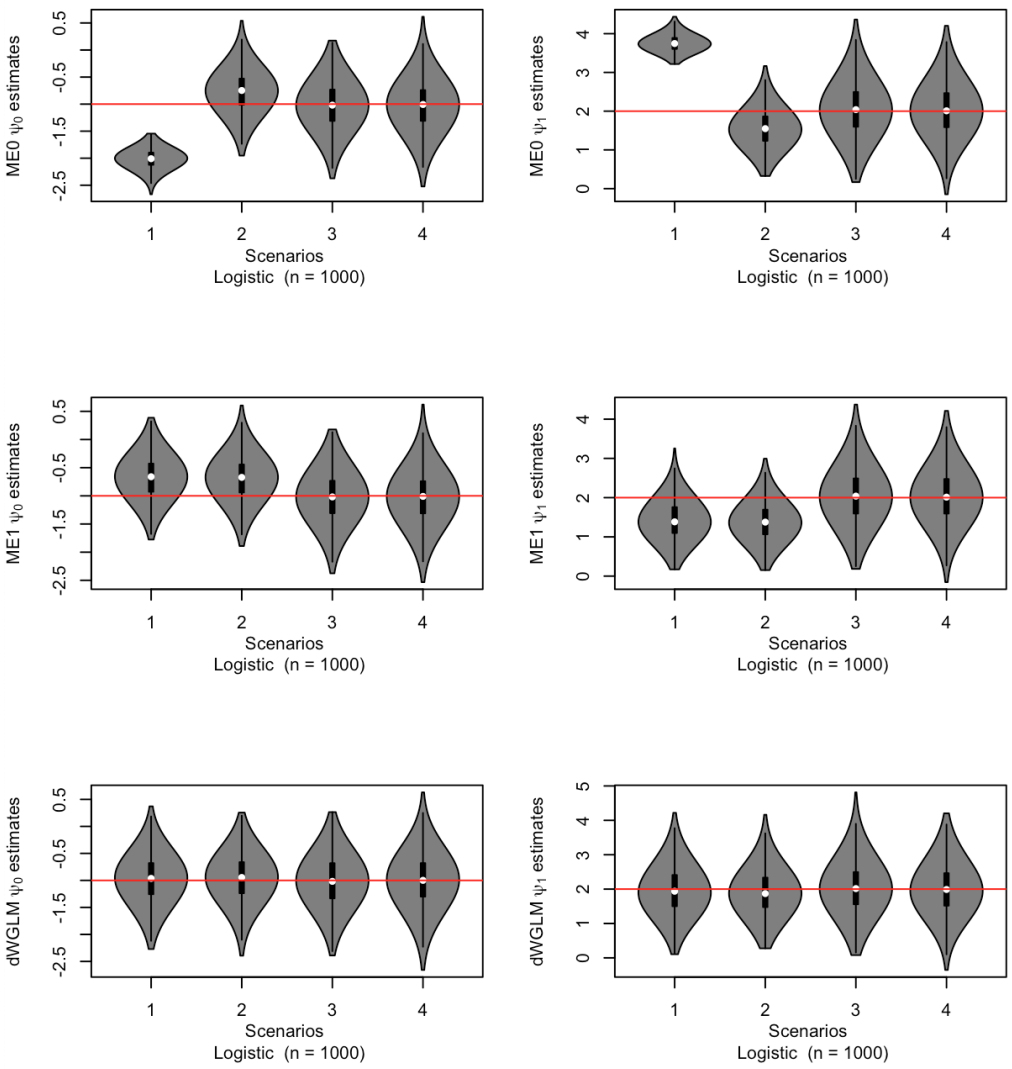


Figure 4.2: Blip function parameter estimates via Method 0 (M0 Q-learning, top row), Method 1 (M1 GLM model with standard dWOLS weights, middle row) and Method 2 (dWGLM, bottom row) with logit link when neither model (scenario 1), treatment model only (scenario 2), treatment-free model only (scenario 3), or both models are correctly specified (scenario 4).

4.3.2 Two-stage Decision for Binary Outcomes

Our second set of studies will demonstrate the implementation of our strategies in simulated datasets for a two-stage treatment decision process. In *Study 2a*, we consider the outcome model of form: $\text{logit}[\mathbb{P}(Y=1)] = \text{logit}[\mathbb{P}(Y^{opt}=1)] - \sum_{j=1}^{K}\left[\gamma_j\left(\boldsymbol{h}_j, a_j^{opt}\right) - \gamma_j\left(\boldsymbol{h}_j, a_j\right)\right]$, and examine the double robustness of the proposed dWGLM in the two-stage ($K=2$) decision problem. A causal diagram of this two-stage decision is shown in Figure 4.3. Writing the column vector $\boldsymbol{\psi}_j = (\psi_{0j}, \psi_{1j})^\top$, the data-generating process is as follows.

- Patient information: $X_1 \sim N(2,1)$, $X_2 \sim N(1+0.5X_1, 2)$;
- Treatment: $\mathbb{P}(A_1 = 1 \mid x_1) = \text{expit}[-5 + x_1 + x_1^2]$, $\mathbb{P}(A_2 = 1 \mid x_2) = \text{expit}[-2.5x_2 + x_2^2 + \sin(x_2)]$;
- Blip functions: $\gamma_j\left(\boldsymbol{h}_j^{\psi}; \boldsymbol{\psi}_j\right) = a_j \boldsymbol{\psi}_j^\top \boldsymbol{x}_j^{\psi}$, with $\boldsymbol{x}_j^{\psi} = (1, x_j)^\top$, $\psi_{0j} = -2$ and $\psi_{1j} = -1$ for $j = 1, 2$. The regret function is thus $\mu_j\left(\boldsymbol{h}_j^{\psi}; \boldsymbol{\psi}_j\right) = \left(a_j^{opt} - a_j\right)\boldsymbol{\psi}_j^\top \boldsymbol{x}_j^{\psi}$, where $a_j^{opt} = \mathbb{I}(\boldsymbol{\psi}_j^\top \boldsymbol{x}_j^{\psi} > 0)$;
- Outcome: $\text{logit}[\mathbb{P}(Y=1)] = \text{logit}[\mathbb{P}(Y^{opt}=1)] - \mu_1\left(\boldsymbol{h}_1^{\psi}; \boldsymbol{\psi}_1\right) - \mu_2\left(\boldsymbol{h}_2^{\psi}; \boldsymbol{\psi}_2\right)$, where $\text{logit}[\mathbb{P}(Y^{opt}=1)] = x_1 + \log(|x_1|) + \cos(\pi x_1)$.

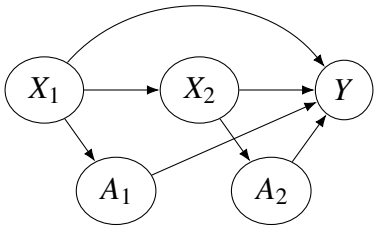
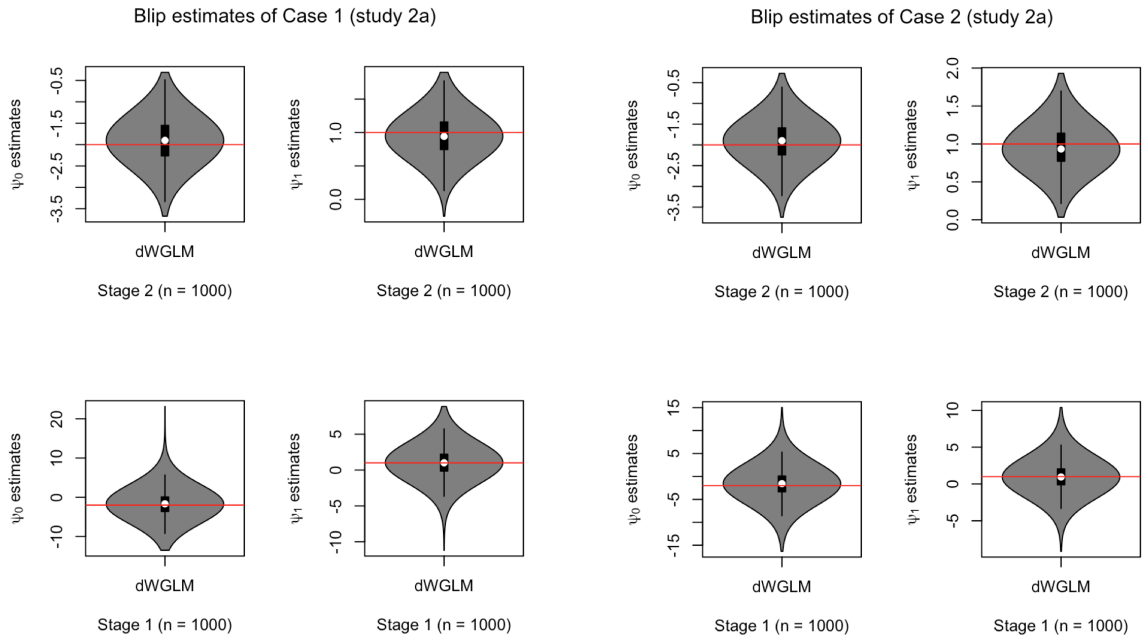


Figure 4.3: Directed acyclic graph of Simulations 3.2 (*Study 2a*).

In this two-stage decision problem, to evaluate the double robustness property of our dWGLM approach, we consider various forms of model misspecification. In particular, we emphasize the following two cases: 1) the treatment-free models were misspecified for both stages by only considering linear terms, but the treatment models were specified correctly; 2) the treatment-free model was misspecified for the second stage, but the treatment model was specified correctly; in



(a) Case 1 (b) Case 2

Figure 4.4: Two-stage decision simulation Study 2a. The top row is the blip parameter estimates via dWGLM for Stage 2, and the bottom is the blip parameter estimates for Stage 1. The left two columns (a) and the right two columns (b) are estimates from Case 1 and 2, respectively.

contrast, the treatment model was misspecified for the first stage, yet the treatment-free model was correctly identified.

For the two-stage binary-outcome problem where the binary outcome was generated by the model of the form $\text{logit}\left[\mathbb{P}\left(Y=1\right)\right] = \text{logit}\left[\mathbb{P}\left(Y^{opt}=1\right)\right] - \sum_{j=1}^{K}\left[\mu_j\left(\boldsymbol{h}_j, a_j\right)\right]$, the simulation results are as expected. The blip function parameter estimates are shown in Figure 4.4. For Case 1, in both Stage 1 and Stage 2, the treatment-free models are incorrectly specified, but the treatment models are all correctly specified. The blip parameter estimates from both stages of Case 1 are plotted in Figure 4.4a. The top row shows the blip parameter estimates via dWGLM for Stage 2, and the bottom gives the blip parameter estimates for Stage 1. Both stages' blip parameter estimates appear to be consistent. In Case 2, where only the treatment model is correctly specified in Stage 2 and only the treatment-free model is correctly specified in Stage 1, our results show that the blip parameters (plotted in Figure 4.4b) are also consistently estimated. Therefore, these results are as expected: the blip estimators appear consistent, and the double robustness of dWGLM in this study is verified.

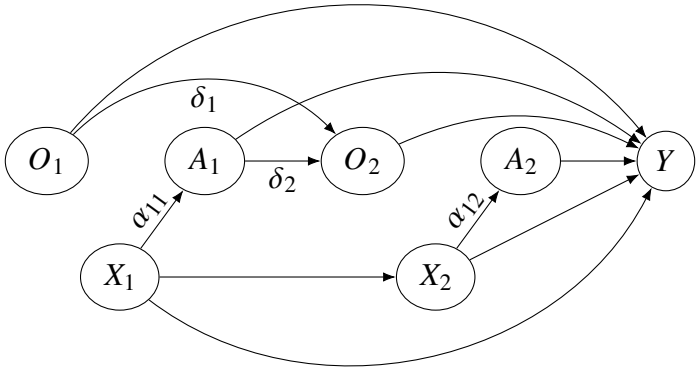


Figure 4.5: DAG of Simulations 3.2 (*Study 2b*).

In *Study 2b*, motivated by real observational data where treatment assignment is confounded by covariates, and based on the simulation study in Moodie et al. [2012], we distinguish between tailoring variables and predictive variables that include potential confounders, where the tailoring variables are denoted as O_j, and the predictive variables are denoted as X_j. The datasets feature some covariates recorded at each stage, where the second stage covariates are potentially affected by first stage variables (both treatment and non-treatment covariates). A causal diagram of this two-stage decision is shown in Figure 4.5; therefore, the data generating process is as follows.

The covariates are $X_1 \sim N(3,1)$, $X_2 \sim N(-0.5 + 0.5X_1, 1)$, and the treatment model is $\mathbb{P}(A_j = 1 \mid X_j) = \text{expit}(\alpha_{0j} + \alpha_{1j}X_j)$ for $j = 1,2$. The binary tailoring variables satisfy $\mathbb{P}(O_1 = 1) = 0.5$, and $\mathbb{P}(O_2 = 1 \mid O_1, A_1) = \text{expit}(\delta_1 O_1 + \delta_2 A_1)$. The outcome variable satisfies

$$\mathbb{P}(Y = 1 \mid X_1, O_1, A_1, X_2, O_2, A_2; \boldsymbol{\theta}) = \text{expit}[m(x_1, x_2, o_1, o_2, a_1, a_2)],$$

where $m(x_1, x_2, o_1, o_2, a_1, a_2) = \theta_0 + \theta_1 X_1 + \theta_2 O_1 + \theta_3 A_1 + \theta_4 O_1 A_1 + \theta_5 X_2 + \theta_6 A_2 + \theta_7 O_2 A_2 + \theta_8 A_1 A_2 + \varphi_1(X_1) + \varphi_2(X_2)$, and φ_1 and φ_2 may be non-linear functions such as $\varphi_1(X_1) = X_1^3$ and $\varphi_2(X_2) = \log(|X_2|)$. We concentrated on the setting where $\boldsymbol{\theta} = (0, 1, 0, -0.5, -0.1, 1, 0.25, 0.5, 0.35)^\top$, $\boldsymbol{\delta} = (0.5, 0.6)^\top$, and $\boldsymbol{\alpha}_1 = (\alpha_{01}, \alpha_{11})^\top = (-2.5, 1.25)^\top$, $\boldsymbol{\alpha}_2 = (\alpha_{02}, \alpha_{12})^\top = (-0.5, 1.25)^\top$. We note that these choices of the parameters pertain to regular settings in the sense of Chakraborty et al. [2010], but other choices that correspond to the non-regular settings can be further studied.

For the second stage, the true treatment-free function is $f_2(x_1, o_1, a_1, x_2, o_2, a_2) = \theta_0 + \theta_1 X_1 + \theta_2 O_1 + \theta_3 A_1 + \theta_4 O_1 A_1 + \theta_5 X_2 + \varphi_1(X_1) + \varphi_2(X_2)$, and true blip function is $\gamma_2(o_2, a_2, a_1) =$

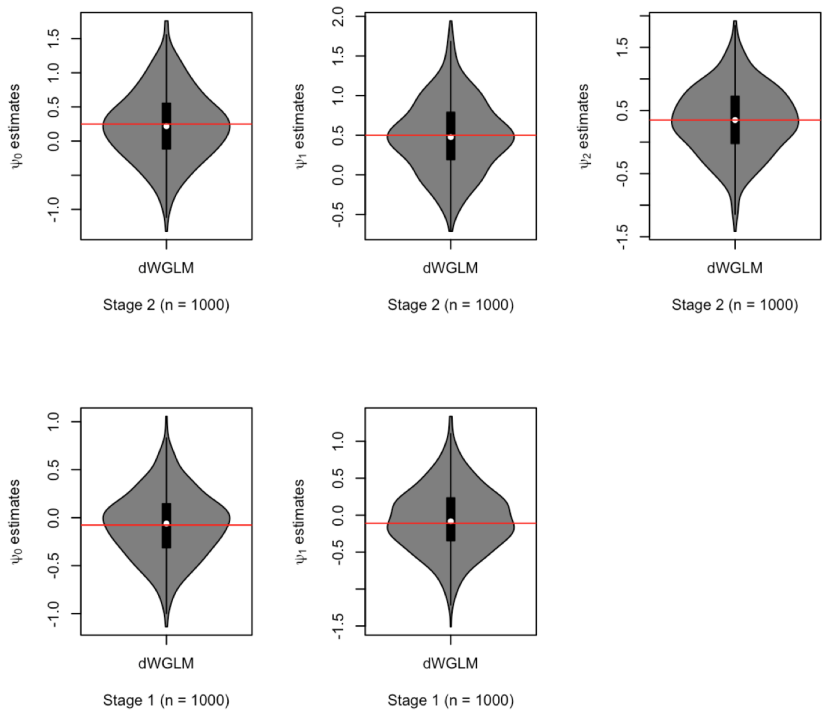


Figure 4.6: Estimates of blip parameters in Study2b. Two-stage decision (Stage 2, top row) blip function parameter estimates via dWGLM with logit link (left to right columns are for ψ_0, ψ_1, and ψ_2, respectively) when only the treatment model is correctly specified.

$\theta_6 A_2 + \theta_7 O_2 A_2 + \theta_8 A_1 A_2$. Thus, for the second stage, the true blip parameters are $\boldsymbol{\psi}_2 = (\theta_6, \theta_7, \theta_8)^\top = (0.25, 0.5, 0.35)^\top$. However, the true first stage decision rule parameters are more complicated because O_2 depends on A_1. Building on the work of Moodie et al. [2014], in Appendix G.3, we derive the true first-stage decision rule parameters (i.e., $\boldsymbol{\psi}_1 = (\psi_{10}, \psi_{11})^\top$) as a function of the data-generating parameters. That is, for the true blip parameters $\boldsymbol{\psi}_1 = (\psi_{10}, \psi_{11})^\top$, we have the coefficient of A_1 as

$$\psi_{10} = \theta_3 + |\phi_3|^+ - |\phi_4|^+ + k_3 \left(|\phi_1|^+ - |\phi_3|^+\right) - k_4 \left(|\phi_2|^+ - |\phi_4|^+\right),$$

and the coefficient of $O_1 A_1$ as

$$\psi_{11} = \theta_4 + (k_1 - k_3) \left(|\phi_1|^+ - |\phi_3|^+\right) - (k_2 - k_4) \left(|\phi_2|^+ - |\phi_4|^+\right).$$

Figure 4.6 shows the blip estimates from the simulation Study 2b, and we can conclude that

the blip parameters appear to be consistently estimated; therefore, the results from Study 2b, with a parameterization that resembles that of a real dataset, are also as expected.

4.4 Real Data Analysis from PATH Study

In Chapter 3, we have applied our interference-aware dWOLS methods in the analysis of PATH data. In this section, assuming no interference, we will examine the PATH dataset again to apply our dWGLM, which is proposed for DTR estimation with binary outcomes.

In analyzing the PATH data, our interest in this chapter is in estimating the optimal DTR for each smoker, in terms of a sequence of use or non-use of e-cigarettes to achieve smoking cessation. Analogously to the PATH analysis in Section 3.5, we now consider the subset of respondents who are smokers in Wave 1. Using the first four waves of data, we formulate our analysis as a three-stage decision problem and define the j^{th} stage ($j = 1, 2, 3$) to be the time from Wave j up to but not including Wave $j + 1$. As in Chapter 3, we continue to set the treatment variable as the use of e-cigarettes by a cigarette smoker. As previously discussed, due to the long participant-follow-up of approximately one year, we define e-cigarette use reported at the wave of the measured outcome as indicative of the pre-wave treatment. In addition, the binary outcome in our analysis is the indicator of smoking cessation (of traditional cigarettes only) or not based on the question "Do you now smoke cigarettes (a) Every day (b) Some days (c) Not at all?" in the study. If participants respond (c), then their binary outcomes are coded as $Y = 1$; if they respond (a) or (b), then their binary outcomes are coded as $Y = 0$.

Building on previous PATH analyses such as that in Benmarhnia et al. [2018], for the j^{th} stage, we also select the Wave j variables age ("less than 35" or "35+"), education, sex, non-Hispanic, race and "plan to quit", denoted, respectively, as the covariates $\boldsymbol{x}_j^{\beta} = (x_{j1}, x_{j2}, x_3, x_4, x_5, x_{j6})^{\top}$ in the treatment-free model. We note that the questionnaire will no longer ask the question regarding "plan to quit" to participants who have already quit smoking; thus, we assign the value of that question to 1 (i.e., they have a plan to quit smoking) for those participants. In addition, as discussed in Section 3.5, the tailoring variables that are related to the efficacy of the treatment should be selected from a set of moderator variables (Almirall et al. [2014]). Building on previous work of studying moderators in the relationships of prior wave predictors of quitting smoking, we select at each stage the variables age and "plan to quit" as tailoring variables, i.e., $\boldsymbol{x}_j^{\psi} = (x_{j1}, x_{j6})^{\top}$ (Le Grande et al. [2021]). The covariates in the treatment propensity models are chosen based on the work of Benmarhnia et al. [2018], and $\boldsymbol{x}_j^{\alpha} = \boldsymbol{x}_j^{\beta} = (x_{j1}, x_{j2}, x_3, x_4, x_5, x_{j6})^{\top}$. Therefore, in estimation, the blip model is set up as $\gamma\left(\boldsymbol{x}_j^{\psi}; \boldsymbol{\psi}_j\right) = a_{j+1}\left(\psi_{j0} + \boldsymbol{\psi}_{j1}^{\top}\boldsymbol{x}_j^{\psi}\right)$, and the treatment-free model as $f_j(\boldsymbol{x}_j^{\beta}; \boldsymbol{\beta}_j) = \beta_{j0} + \boldsymbol{\beta}_{j1}^{\top}\boldsymbol{x}_j^{\beta}$. Four sets of analyses corresponding to those carried out in the

Table 4.2: Analysis and optimal DTRs of PATH data. Optimal DTRs are indicator functions of $\boldsymbol{\psi}$ estimates.

Wave	Estimates $\hat{\boldsymbol{\psi}}$	Methods			
		M0	M1	M2	M3
$1 \sim 2$	$\hat{\psi}_0$	0.0188	0.0236	-0.0013	-0.0041
	$\hat{\psi}_1$	-0.0841	-0.0710	-0.0250	-0.0055
	$\hat{\psi}_2$	0.1142	0.0428	-0.0219	0.0014
$2 \sim 3$	$\hat{\psi}_0$	0.0602	0.0975	0.1244	0.0380
	$\hat{\psi}_1$	-0.0714	-0.1114	-0.1229	-0.0345
	$\hat{\psi}_2$	0.0315	0.0693	0.1198	0.0150
$3 \sim 4$	$\hat{\psi}_0$	-0.0469	0.0035	-0.1553	0.0088
	$\hat{\psi}_1$	0.2535	0.2478	0.4581	0.1547
	$\hat{\psi}_2$	0.2937	0.2621	0.1745	0.1261

simulations of Section 4.3.1 are conducted. At each stage of Method 0 (M0), i.e., the Q-learning approach, a logistic regression is implemented. Method 1 is similar to M0 but uses the dWOLS balancing weights in each logistic regression estimation. Then, Method 2 (M2), our proposed doubly robust method, uses the logistic regression model with the new weights based on the equation (4.4). To perform a sensitivity analysis for the link function, we also consider Method 3 (M3), which is analogous to M2 as it includes two-step robust estimation in each stage, but which uses the probit link function. We note that only M0 does not use any balancing weights in the estimation process, but M1, M2, and M3 do use weights for the purpose of balancing.

Our proposed new weights (equation 4.4) are built to provide an unbiased estimator of the blip parameters through estimating weighted population-level estimating equations. Similar to the PATH data analysis in Chapter 3, we also employ the sampling design weights in each stage. Regarding the combination of model and sampling design, the sample estimating functions are unbiased with respect to the sample design of the population-level estimating functions; thus, the estimators of the blip function parameters are model-design consistent (Lumley [2004]).

As shown in Section 4.2.3, our use of balancing weights is to ensure consistent blip estimators. M2 and M3, which employ the proposed balancing weights, are expected to provide consistent blip estimators, but M1, which uses standard dWOLS weights, and M0, which does not use any balancing weights, are not. The blip parameter estimates from Methods 0, 1, 2, and 3 are summarized in Table 4.2. The results of M0 and M1 are similar, especially $\hat{\psi}_0$ and $\hat{\psi}_1$ in Stage 1 (Wave $1 \sim 2$), and $\hat{\psi}_1$ and $\hat{\psi}_2$ in Stage 3 ($3 \sim 4$). Both M2 and M3 employ the proposed balancing weights, but they use different link functions and return different estimates. The difference between estimates from M1 and M2 shows the difference in results obtained with our method

when employing two different weights: the original dWOLS weights that are for the continuous outcome model and our proposed balancing weights that are for the binary outcomes. We also see differences between M2 (using the logit link) and M3 (using the probit link). Compared with the estimates of M2, those of M3 are attenuated; this is expected because the standard normal distribution has a lighter tail than the logistic distribution. For example, given a certain probability larger than 0.5, the value of the inverse of the logistic function is greater than that of the inverse of the standard normal distribution; thus the coefficient estimates obtained with the probit link function will tend to be smaller compared with those obtained with the logit link function if the covariates are the same. Although the estimates of M2 and M3 are different, their patterns that decide the treatment recommendations are similar.

Building on the blip parameter estimates, the corresponding optimal treatment regime will be $\hat{a}_j^{opt} = \mathbb{I}(\hat{\boldsymbol{\psi}}_j^\top \boldsymbol{x}_j^\psi > 0)$, for $j = 1, 2, 3$; for instance, for Wave $1 \sim 2$, Method 2 (i.e., dWGLM) outputs $\mathbb{I}[-0.0013 - 0.0250x_{11} - 0.0219x_{16} > 0]$; for Wave $2 \sim 3$, Method 2 outputs $\mathbb{I}[0.1244 - 0.1229x_{11} + 0.1198x_{16} > 0]$, and for Wave $3 \sim 4$, it outputs $\mathbb{I}[-0.1553 + 0.4581x_{11} + 0.1745x_{16} > 0]$. These results from Method 2 can be interpreted as the following treatment recommendations about the use of e-cigarettes. In the first stage, do not use e-cigarettes. In the second stage, use e-cigarettes. In the last stage, if a smoker's age is less than 35 and he or she has no plan to quit, do not use e-cigarettes; otherwise, use e-cigarettes. Finally, we note that, through the three-stage treatment decision analysis of the PATH data, our intention is mainly to illustrate that our dWGLM can be applied in practice, but not to put forward the results as authentic recommendations for a treatment strategy.

4.5 Conclusion and Discussion

Dynamic treatment regimes are a mechanism by which treatment decisions are made based upon individual-level information, used in optimizing long-term expected outcomes. Many approaches for optimal DTR estimation are limited to continuous outcomes. The few used to address binary outcomes are limited in their robustness to model mis-specification or complexity of implementation. Our dWGLM method, motivated by its continuous-outcome predecessor dWOLS, provides double robustness to model misspecification while being comparatively easy to implement. Our method can be viewed as a series of weighted GLM analyses. Meanwhile, to make optimal sequential decisions, some care in constructing stage-specific binary pseudo-outcomes is also needed. We offer a new balancing weight criterion to overcome the misspecification of treatment-free models, and the method for each decision stage involves just a two-step or two-phase logistic regression. Note that our new balancing weight criterion is feasible for any GLM, and the corresponding theory is also suitable for any g link function. From this viewpoint, dWOLS is a

special case of dWGLM, where the link function is an identify link: $g(\mu) = \mu$. In practice, using observed data from the Population Assessment of Tobacco and Health study, we illustrated our dWGLM in an analysis of e-cigarette usage and smoking cessation.

Our dWGLM approach is doubly robust for estimating the parameter of interest, a property demonstrated via simulation. It is important to acknowledge that our approach relies on the suitability of the local linear approximation to the inverse link function (g^{-1}). We utilize Taylor expansion of g^{-1} about $\boldsymbol{\beta}^\top \boldsymbol{X}$ evaluated at $f(\boldsymbol{X})$ (Equation G.6 in Appendix G.1) in the proof of Theorem 4.1, and the error term is close to zero if a linear predictor tends to vary in an interval where g^{-1} is close to linear. At the end of Appendix G.1, we discuss the accuracy of approximation that can be decided by the range of covariates in our method and the inverse link function. Therefore, a possible extension would be to conduct sensitivity analyses to study different link functions as well as treatment-free functions in GLM.

In future work of this chapter, we note that some machine learning (ML) methods can be employed in our dWGLM analytical framework. For example, tree-based methods (e.g., Bayesian additive regression trees, Chipman et al. [2010]) are commonly used in estimating the treatment model, and some ensemble methods (e.g., Super Learner, Van der Laan et al. [2007]) can be used for last stage estimation to provide a more accurate prediction of $\mathbb{P}(Y = 1 \mid \boldsymbol{h}_K, a_K)$, and thus to produce accurate pseudo outcome prediction. For another example, Moodie et al. [2012] employs a generalized additive model in Q-learning. The ML methods, of course, are chosen based on the purpose of the estimation. For the last stage outcome model, the ML method is required for precise prediction; however, for the treatment model, correct modeling of the data generating mechanism is not necessary, but rather all confounders must be included to correctly model the impact of the treatment (Ertefaie et al. [2013]). Further, due to regularization and overfitting, the "prediction-focused" ML estimators may be biased (Chernozhukov et al. [2018]); thus, orthogonalization and data splitting should be carefully investigated to control the regularization and the overfitting bias. Therefore, one important extension to our work is to employ different ML models in the corresponding process to produce accurate optimal treatment regimes.

Chapter 5

Optimal DTR Estimation with Household Interference for Ordinal Outcomes

Considering continuous outcomes in Chapter 3, we have developed a doubly robust method for estimating DTRs in the presence of network interference. As a follow-up to the interference-aware DTR method in Chapter 3, with the aim of examining cases where each individual has a binary outcome, our goal is to develop a corresponding optimal DTR estimation method. To achieve this, as stated in Chapter 4, our first task was to examine DTR estimation for binary outcomes without interference. Then, building on Chapter 4, this chapter (Chapter 5) continues the study of DTR estimation with interference for binary outcomes. We propose two ways to examine this DTR estimation with interference problem by the case of pairs where ego and alter have binary outcomes. First, as in Chapter 3, we assume that all alters' treatments are fixed, and taking those treatments as tailoring variables, we aim to estimate optimal DTRs for the egos. Second, assuming that all alters' treatments also need to be determined by decision rules, we estimate optimal DTRs for couples in the same household. For this goal, we consider a composite *ordinal outcome* for a household based on the binary outcomes for the ego and the alter. In particular, motivated by the PATH data, we will emphasize household interference in which there is an association between the treatments of connected individuals and household utility, defined by combining the binary outcomes of the two individuals of interest in each household. We also outline the corresponding application for assigning treatment strategies for a pair in the same household, based on the individual-level and household-level covariates.

5.1 Introduction

Since COVID-19 appeared, research on this contagious disease has become popular, causing the number of interference studies in the causal inference area to gradually expand. Meanwhile, it has gradually become necessary to account for interference in optimal DTR estimation. Recently, some researchers, such as Jiang et al. [2022b], Su et al. [2019], have focused on optimal DTR estimation in the presence of interference, mainly considering how other patients' treatments may affect the outcome of individuals of interest. In such cases, treatment-decision rules should involve others' information such as treatments and covariates. To conduct robust optimal DTR estimation, in Chapter 3, we proposed network balancing weights to apply dWOLS in cases of interference where there was an ego and alters, and the covariates or treatments of the alters could affect the treatment or outcome of the ego. We considered the interference in different network scenarios such as the Erdős-Rényi network, the Barabási-Albert network and the Longleaf Pines spatial network. In particular, we examined two-person household networks in the PATH study, where a couple in the same family is regarded as a household network. These recently developed interference-aware DTR estimation methods, and even the standard DTR estimation methods, however, focus extensively on continuous outcomes. To the best of our knowledge, few publications have considered optimal DTR estimation for discrete outcomes, such as binary and ordinal outcomes, in the presence of interference.

Our work described in this chapter (Chapter 5) is intended to address this knowledge gap, and builds on the material in Chapters 3 and 4. In Chapter 4, we proposed a novel doubly robust and easily-implemented DTR estimation method, dWGLM, in preparation for the work in Chapter 5. To study DTR estimation under binary outcomes and household interference, the straightforward way is similar to the approaches in Chapter 3. That is, aiming to optimize outcomes of egos, and building on interference-aware dWOLS presented in Chapter 3 and the dWGLM method described in Chapter 4, we have developed an approach to making treatment decisions for egos, taking into account the alters' treatment information (see Appendix H.1.2). Alternatively, we can also treat the problem in a different way. That is, motivated by the PATH data where interference exists in the household of interest, we target making treatment decisions for both individuals of a pair in the same household, and we concentrate on optimizing household utility models. This novel viewpoint is one of the main contributions of this chapter (Chapter 5).

Considering the binary outcomes for both individuals of a pair in the same household, we build a combination of the binary outcomes as a utility for the household. The household utility that depends on both members' binary outcomes is constructed as an ordinal outcome. Then, corresponding *proportional odds models* (POMs) are employed to estimate the parameters of interest. To further study the interference in the household case, we also extend the independent treatment assignment assumption to cases in which a correlation exists between the treatments of a

pair in a household. We provide an estimation method for estimating a household's joint propensity score. Building on the household joint propensity score, in the single-stage setting, we then propose adjusted balancing weights for POMs in the case of household ordinal outcomes under household interference. We also explain the corresponding estimation process for multi-stage decision problems. In addition, to maximize the utility of the whole household, we propose the optimal treatment decision rules for a couple in the household, which map from the space of the couple's covariates into the space of feasible treatment combinations.

This chapter is organized as follows. In Section 5.2, we propose the estimation process of the joint propensity functions when correlations between the treatments of individuals in the same household are present, and construct the corresponding balancing weights. Further, we provide some measures in Section 5.3 to assess the balance of the proposed weights. In Section 5.4, we introduce the proposed doubly robust regression-based DTR estimation framework for the case of a household with ordinal utilities under household interference. In addition, we explain the analytical framework of optimal treatment decision rules for a couple in the household. We demonstrate in Section 5.5, through simulations of both single- and multi-stage treatment decisions, that our method is doubly robust against misspecification of either treatment-free or treatment models. Section 5.6 presents the implementation of our methods on PATH data. Section 5.7 concludes with a discussion of future research.

5.2 Correlated Treatments in a Household

In this section, we first recall the dWOLS balancing weights in the presence of household interference. Next, we propose novel balancing weights that take account of the correlations between treatments of individuals in the same household. In addition, considering the case where there exist correlations between treatments of individuals, we demonstrate the possible data-generation process and propose an estimation process based on the methods currently regarded as the most efficient — alternating logistic regressions.

Considering Chapter 3's network propensity function, we define the joint propensity function for a household case:

Definition 5.1. *In the household case where there is a pair* (s, r)*, for* $a^s = 0, 1$ *and* $a^r = 0, 1$ *the joint propensity function is the joint probability of one individual in the household receiving treatment* a^s *and the other receiving treatment* a^r*, that is,*

$$\pi^{a^s a^r}(\boldsymbol{x}_s, \boldsymbol{x}_r) = \mathbb{P}(A^s = a^s, A^r = a^r \mid \boldsymbol{x}_s, \boldsymbol{x}_r).$$

Moreover, building on Chapter 3's network interference weights criterion, we have the *standard balancing weights criterion* that

$$\pi^{00}w(0,0,\boldsymbol{x}) = \pi^{01}w(0,1,\boldsymbol{x}) = \pi^{10}w(1,0,\boldsymbol{x}) = \pi^{11}w(1,1,\boldsymbol{x}), \tag{5.1}$$

where $\boldsymbol{x}$ represents the household-level covariates for determining the balancing weights. For simplicity and without loss of generality, $\boldsymbol{x}$ represents $(\boldsymbol{x}_s, \boldsymbol{x}_r)$. (We are not assuming $\boldsymbol{x}_s$ and $\boldsymbol{x}_r$ to be independent.) In terms of a weighted dataset $(a^s, a^r)^w$ and $(\boldsymbol{x}_s, \boldsymbol{x}_r)^w$, the balancing weights criterion balances the weights in the sense that $(a^s, a^r)^w \perp\!\!\!\perp (\boldsymbol{x}_s, \boldsymbol{x}_r)^w$, that is, the weighted treatment configuration is independent of the weighted covariate configuration. From the balancing weights criterion (5.1), balancing weights are constructed by resorting to joint propensity functions $\pi^{a^s a^r}(\boldsymbol{x}_s, \boldsymbol{x}_r)$. In Chapter 3, we assumed that the treatments of s and r are conditionally independent, that is, that the joint propensity functions are equal to the product of marginal propensities; mathematically, $\pi^{a^s a^r}(\boldsymbol{x}_s, \boldsymbol{x}_r) = \mathbb{P}(A^s = a^s, A^r = a^r \mid \boldsymbol{x}_s, \boldsymbol{x}_r) = \mathbb{P}(A^s = a^s \mid \boldsymbol{x}_s)\mathbb{P}(A^r = a^r \mid \boldsymbol{x}_r)$. In this chapter, however, we assume that there possibly exists an association between A^s and A^r, and we need to focus on estimating the joint propensity functions. Building on the interference weights criterion and the idea of constructing (so-called) overlap weights for multiple treatments in Li [2019], we might propose the inverse probability based interference balancing weights for the correlated treatment in a household:

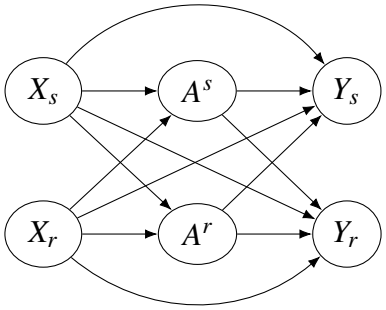
$$w(a^s, a^r) \propto \frac{1}{\pi^{a^s a^r}} \times \frac{1}{\sum_{a^s, a^r} 1/\pi^{a^s a^r}}, \quad \text{for } a^s = 0, 1; a^r = 0, 1. \tag{5.2}$$

The proposed weights satisfy the weights criterion (5.1) and thus provide a balancing guarantee between treatments and covariates configurations. The weight is proportional to the inverse of the joint propensities and divided by a "normalization" factor $\sum_{a^s, a^r} 1/\pi^{a^s a^r}$.

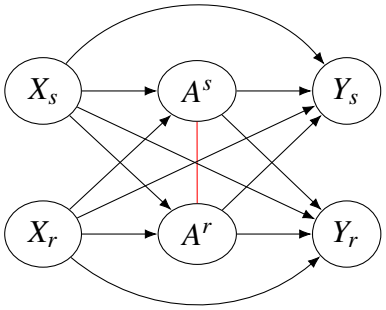
On the other hand, considering the balancing weights criterion in the form $\pi^{00}w(0,0,\boldsymbol{x}) = \pi^{01}w(0,1,\boldsymbol{x}) = \pi^{10}w(1,0,\boldsymbol{x}) = \pi^{11}w(1,1,\boldsymbol{x}) = \pi^{00}\pi^{10}\pi^{01}\pi^{11}$, we might also propose overlap-type balancing weights:

$$w(a^s, a^r) \propto \frac{\pi^{00}\pi^{10}\pi^{01}\pi^{11}}{\pi^{a^s a^r}}, \quad \text{for } a^s = 0, 1; a^r = 0, 1. \tag{5.3}$$

The overlap-type weight for one treatment pair realization is proportional to the product of the joint propensities for the other possible realizations. To estimate the balancing weights, from either (5.2) or (5.3), the key is to precisely estimate the joint propensity function $\pi^{a^s a^r}(\boldsymbol{x}_s, \boldsymbol{x}_r)$. Again, it is important to note that, if the treatments of members of the household are mutually independent, that is, each individual independently receives treatment, then the joint propensity is the product of marginal propensities. However, assuming the independence of the treatment is



(a) DAG 1



(b) DAG 2

Figure 5.1: Directed acyclic graphs of interference analysis, with and without association between A^s and A^r conditional on $(\boldsymbol{x}_s, \boldsymbol{x}_r)$.

not reasonable in many cases; for example, an individual could receive treatment based on the treatment status of the other household member. In a case where the treatments are correlated, the joint propensity functions are not equal to the product of the marginal propensities. To build accurate balancing weights and thus make robust estimations of optimal DTRs, we take into account the dependence among treatments observed in the same household. As presented in Figure 5.1, the directed acyclic graphs for the data-generating process, the red path between A^s and A^r in DAG 5.1 indicates an association (conditional on $\boldsymbol{x}_s$ and $\boldsymbol{x}_r$) of the treatments between the individuals who are interfering with each other in the same household.

With regard to estimation methods for joint propensity, for our pairs case, one straightforward way is to model three probabilities (Baker [1995]); for example, those for $(A^s, A^r) = (0,1), (1,0)$ and $(1,1)$. Thus, the remaining probability (i.e., $\mathbb{P}(A^s = 0, A^r = 0)$) equals to one minus the probabilities of the three other possibilities. However, this method would not easily be extended to a larger number of individuals in the same cluster. Therefore, we focus on an alternative general way to model the marginal probabilities and the associated structures. For these structures, the odds ratio, which has some desirable properties and is employed as the correlation coefficient, is studied as a measure of association between pairs of binary variables (Lipsitz et al. [1991], Yi and Thompson [2005]). To generate the correlated treatments for a household, the idea is to specify the joint distribution (or joint propensity) of the two binary variables by specifying the marginal distributions of those two binary variables and the odds ratio. Then, building on the joint propensity, we can generate the treatment configuration as $(0,0), (0,1), (1,0)$ or $(1,1)$. Formally, suppose there are $h = 1, ...H$ households, and one pair is in each household, that is, the pair $(s,r)^h$ belongs to the h^{th} household. Based on Lipsitz et al. [1991]'s parameterization of correlated binary data, letting $\boldsymbol{A}_h = (A_h^s, A_h^r)^\top$ be the treatment vector

A_h^s \ A_h^r	1	0
1	p_{hsr}	$p_{hs} - p_{hsr}$
0	$p_{hr} - p_{hsr}$	$1 - p_{hs} - p_{hr} + p_{hsr}$

Table 5.1: Contingency table for binary treatment variables (A_h^s, A_h^r) in the h^{th} household.

for the h^{th} household, we define $p_{hs}(\boldsymbol{\alpha}) := \mathbb{P}(A_h^s = 1 \mid \boldsymbol{x}_{hs}, \boldsymbol{\alpha})$ and $p_{hr}(\boldsymbol{\alpha}) := \mathbb{P}(A_h^r = 1 \mid \boldsymbol{x}_{hr}, \boldsymbol{\alpha})$ and $A_{hsr} := \mathbb{I}(A_h^s = 1, A_h^r = 1) = A_h^s A_h^r$, where $\mathbb{I}(x)$ is an indicator function. Also, we define $p_{hsr} := \mathbb{P}(A_{hsr} = 1) = \mathbb{P}(A_h^s = 1, A_h^r = 1)$. Through cross-classifying each individual's treatment at household h, we can form a 2×2 contingency table (Table 5.1), and the odds ratio for the pair of correlated binary variables (A_h^s, A_h^r) is

$$\tau_{hsr} := \frac{\mathbb{P}\left(A_h^s = 1, A_h^r = 1\right) \mathbb{P}\left(A_h^s = 0, A_h^r = 0\right)}{\mathbb{P}\left(A_h^s = 1, A_h^r = 0\right) \mathbb{P}\left(A_h^s = 0, A_h^r = 1\right)} = \frac{p_{hsr}\left(1 - p_{hs} - p_{hr} + p_{hsr}\right)}{\left(p_{hs} - p_{hsr}\right)\left(p_{hr} - p_{hsr}\right)}. \tag{5.4}$$

If the odds ratio is known, based on equation (5.4), then we can then solve for p_{hsr} in terms of the two marginal probabilities (p_{hs} and p_{hr}) and the odds ratio (τ_{hsr}) such that:

$$p_{hsr} = \begin{cases} \dfrac{b_{hsr} - \sqrt{b_{hsr}^2 - 4\tau_{hsr}\left(\tau_{hsr} - 1\right) p_{hs} p_{hr}}}{2\left(\tau_{hsr} - 1\right)} & \left(\tau_{hsr} \neq 1\right), \\ \tau_{hsr} p_{hs} p_{hr} & \left(\tau_{hsr} = 1\right), \end{cases} \tag{5.5}$$

where $b_{hsr} = [1 - (1 - \tau_{hsr})\left(p_{hr} + p_{hs}\right)]$. Note that p_{hsr} is always bounded in $[0, 1]$, so there is only one feasible choice between the two solutions of the quadratic equation leading to (5.5) (Mardia [1967]). Therefore, in the dependent binary treatment-generating process, first, we generate the marginal propensity based on the individual covariates. Second, we generate the odds ratio, which is typically based on a log-linear regression model (Lipsitz et al. [1991]), such that $log\tau_{hsr}(\boldsymbol{o}) = \boldsymbol{o}^\top \boldsymbol{x}_{hsr}$, where $\boldsymbol{x}_{sr}$ suppressing the h are some pair-level covariates that may influence the odds-ratio between A^s and A^r, and $\boldsymbol{o}$ represents the corresponding coefficients. We can consider these pair-level covariates $\boldsymbol{x}_{sr}$ to be functions of $\boldsymbol{x}_s$, $\boldsymbol{x}_r$. Third, based on the marginal propensity and odds ratio, we generate the joint probability (joint propensity) of the binary treatments configuration. Finally, we generate a treatment configuration for each household according to its joint propensity.

After seeing the data-generating process for the correlated treatments, let us now review some

notable methods for estimating joint propensity score. In the estimation process, the first step is to identify marginal models for multivariate binary data; Liang and Zeger [1986] proposed a first-order generalized estimating equation method to provide efficient estimates of regression coefficients. Suppose that there are $h = 1, 2, .., H$ households and household h contains a treatment record of a pair $\boldsymbol{A}_h = (A_h^s, A_h^r)^\top$, and we are interested in inference about the parameters of the marginal probabilities such that $p_{ht}(\boldsymbol{\alpha}) = \mathbb{P}(A_h^t = 1 \mid \boldsymbol{x}_{ht}, \boldsymbol{\alpha})$ in terms of covariates $\boldsymbol{x}_h$ and marginal parameter $\boldsymbol{\alpha}$ for $t = s, r$. The optimal estimating equation for parameter $\boldsymbol{\alpha}$ is

$$U(\boldsymbol{\alpha}) = \sum_{h=1}^{H} \boldsymbol{D}_h^\top \boldsymbol{V}_h^{-1} \{\boldsymbol{A}_h - p_h(\boldsymbol{\alpha})\} = 0 \tag{5.6}$$

where $\boldsymbol{D}_h = \partial p_h(\boldsymbol{\alpha})/\partial \boldsymbol{\alpha}^\top$, and $\boldsymbol{V}_h$ is the working covariance matrix of $\boldsymbol{A}_h$. The working covariance matrix $\boldsymbol{V}_h$ has the form as $\boldsymbol{V}_h = \Lambda_h^{1/2} corr(\boldsymbol{A}_h) \Lambda_h^{1/2}$, where the diagonal matrix $\Lambda_h = diag\{p_{ht}(1 - p_{ht})\}$, and correlation matrix $corr(\boldsymbol{A}_h)$.

Then the second step is to model the association between pairs of responses. Prentice [1988] utilized second-order estimating equations, which also offered efficient estimates of association parameters; however, as the cluster size grows, the computation cost becomes huge. Lipsitz et al. [1991] considered the odds ratio to model the association between binary responses, and then modified the estimating equations of Prentice to estimate the pairwise odds ratios. Define $\tau_{sr} := \frac{\mathbb{P}(A^s=1,A^r=1)\mathbb{P}(A^s=0,A^r=0)}{\mathbb{P}(A^s=1,A^r=0)\mathbb{P}(A^s=0,A^r=1)}$, as in Lipsitz et al.'s odds ratio model mentioned in the above generating process, i.e., $log\tau_{sr}(\boldsymbol{o}) = \boldsymbol{o}^\top \boldsymbol{x}_{sr}$. The pairwise odds ratios are assumed to be non-negative and are modeled through a generalized linear model with the log link and parameters $\boldsymbol{o}$.

Carey et al. [1993] proposed alternating logistic regressions (ALR), simultaneously regressing the response on covariates and modeling the association among responses in terms of odds ratios. Two logistic regressions are iterated: one to estimate regression coefficients using first-order generalized estimating equations (5.6), the other to update the odds ratio parameters $\boldsymbol{o}$ using an offset.

These estimation methods provide the estimators of both marginal probabilities ($p_{hs}(\hat{\boldsymbol{\alpha}})$ and $p_{hr}(\hat{\boldsymbol{\alpha}})$) and the odds ratios ($\hat{\tau}_{hsr}$). Then the estimator of joint propensity $\pi^{11}(\boldsymbol{x}_{hs}, \boldsymbol{x}_{hr}) = p_{hsr}$ can be built based on $p_{hs}(\hat{\boldsymbol{\alpha}})$, $p_{hr}(\hat{\boldsymbol{\alpha}})$, and $\hat{\tau}_{hsr}$ by equation (5.5). Further, we have other estimators: $\hat{\pi}^{10}(\boldsymbol{x}_{hs}, \boldsymbol{x}_{hr}) = p_{hs}(\hat{\boldsymbol{\alpha}}) - \hat{\pi}^{11}(\boldsymbol{x}_{hs}, \boldsymbol{x}_{hr})$, $\hat{\pi}^{01}(\boldsymbol{x}_{hs}, \boldsymbol{x}_{hr}) = p_{hr}(\hat{\boldsymbol{\alpha}}) - \hat{\pi}^{11}(\boldsymbol{x}_{hs}, \boldsymbol{x}_{hr})$, and $\hat{\pi}^{00}(\boldsymbol{x}_{hs}, \boldsymbol{x}_{hr}) = 1 - p_{hs}(\hat{\boldsymbol{\alpha}}) - p_{hr}(\hat{\boldsymbol{\alpha}}) + \hat{\pi}^{11}(\boldsymbol{x}_{hs}, \boldsymbol{x}_{hr})$. Therefore, building on equation 5.2 and equation (5.3), we have IPW-type estimator of weights $\hat{w}(a^s, a^r) = \frac{1/\hat{\pi}^{a^s a^r}}{\sum_{a^s,a^r} 1/\hat{\pi}^{a^s a^r}}$, and overlap-type estimator of weights $\hat{w}(a^s, a^r) = \frac{\hat{\pi}^{00}\hat{\pi}^{10}\hat{\pi}^{01}\hat{\pi}^{11}}{\hat{\pi}^{a^s a^r}}$, for $a^s = 0, 1; a^r = 0, 1$.

In this section, new balancing weights in the presence of interference were proposed. New weights are constructed according to the joint propensity function. Thus, considering the possible correlations between the treatments of individuals, we have reviewed a possible data-generation process of the correlated treatments, and estimation methods for estimating joint propensity function. In the next section, we will assess the balance of the proposed weights, and in particular compare them with the weights that do not consider the correlations between individuals, that is, weights that assume individuals independently receive the treatment.

5.3 Balance Assessment

In the previous section, new balancing weights that satisfy the interference weights criterion (5.1) were proposed. In the presence of interference, the balancing weights are employed to balance the covariates and treatment configurations. In the non-interference literature, numerous recommendations have been made for balance assessment; however, very limited research has been done to access balance in the presence of interference. In this section, considering household interference, we will describe how to assess the performance of the balancing weights.

As mentioned at the beginning of the previous section, the balancing weights criterion (5.1) ensures the balance such that $(a^s, a^r)^w \perp\!\!\!\perp (\boldsymbol{x}_s, \boldsymbol{x}_r)^w$. Equivalently, conditional on the balancing score, as a function of covariates, treatments are independent of those covariates, that is, $(a^s, a^r) \perp\!\!\!\perp (\boldsymbol{x}_s, \boldsymbol{x}_r) \mid \pi^{a^s a^r}$ (Rosenbaum and Rubin [1983]). Specifically, as shown in Chapter 3 in the proof of Theorem 3.2, the balancing weights criterion (5.1) is equivalent to

$$\begin{aligned}\frac{\mathbb{P}[(A^s, A^r)^w = (1,1)^w \mid (\boldsymbol{x}_s, \boldsymbol{x}_r)^w]}{\mathbb{P}[(A^s, A^r)^w = (1,1)^w]} &= \frac{\mathbb{P}[(A^s, A^r)^w = (1,0)^w \mid (\boldsymbol{x}_s, \boldsymbol{x}_r)^w]}{\mathbb{P}[(A^s, A^r)^w = (1,0)^w]} \\ &= \frac{\mathbb{P}[(A^s, A^r)^w = (0,1)^w \mid (\boldsymbol{x}_s, \boldsymbol{x}_r)^w]}{\mathbb{P}[(A^s, A^r)^w = (0,1)^w]} \\ &= \frac{\mathbb{P}[(A^s, A^r)^w = (0,0)^w \mid (\boldsymbol{x}_s, \boldsymbol{x}_r)^w]}{\mathbb{P}[(A^s, A^r)^w = (0,0)^w]},\end{aligned}$$

and this implies $(a^s, a^r)^w \perp\!\!\!\perp (\boldsymbol{x}_s, \boldsymbol{x}_r)^w$.

Motivated by the balancing check work in the settings of multiple treatments in Li [2019], and the balance that is derived from the interference weights criterion (5.1), we propose two methods to check for balance or overlap in assessing the performance of balancing weights. In particular, we compare weights that are assumed to have been independently assigned to each individual with the pairwise weights corresponding to correlation within the household. Denote the treatment configuration of (a_h^s, a_h^r) as $Z_h = z_h$ having four possible "treatment and interference" levels

$z_h = 1, 2, 3, 4$. Consider the pair sharing the same household ID to have the same "treatment and interference" levels. Note that the goals for reassigning the levels of both treatment and interference are (1) to fit the balancing check analytical framework in Li [2019]; (2) to simplify some notation, such that just one $H-$vector of Z_h, for example, can represent an $H \times 2$ matrix of the treatment configurations $(A^s, A^r)_h$, for $h = 1, 2, ..., H$. First, the interference weights criterion (5.1) directly implies pairwise balance, that is, $f_z(\boldsymbol{x}_s, \boldsymbol{x}_r)w_z(\boldsymbol{x}_s, \boldsymbol{x}_r) = f_{z'}(\boldsymbol{x}_s, \boldsymbol{x}_r)w_{z'}(\boldsymbol{x}_s, \boldsymbol{x}_r)$ for different z and z' in $\{1, 2, 3, 4\}$, where $f(\boldsymbol{x}_s, \boldsymbol{x}_r)$, the joint density of the covariates, is assumed to exist with respect to a base measure, and $f_z(\boldsymbol{x}_s, \boldsymbol{x}_r)$ and $f_{z'}(\boldsymbol{x}_s, \boldsymbol{x}_r)$ are the group specific densities multiplied by the pairwise treatment propensities. Thus, we can assess balance by checking the pairwise absolute standardized differences (ASD), i.e., $ASD(z, z') = ||(\bar{\boldsymbol{x}}^s_z, \bar{\boldsymbol{x}}^r_z) - (\bar{\boldsymbol{x}}^s_{z'}, \bar{\boldsymbol{x}}^r_{z'})||_2/\boldsymbol{S}^2_{\boldsymbol{X}}$, where $\bar{\boldsymbol{x}}^s_z = \sum_{h=1}^{N_z} \boldsymbol{x}_{s,h}w_z / \sum_{h=1}^{N_z} w_z$ denotes the weighted mean of the individual s's covariate $\boldsymbol{x}_s$ from household h that belongs to the zth group, and $\boldsymbol{S}^2_{\boldsymbol{X}} = \sum_{z=1}^{4} \boldsymbol{S}^2_{\boldsymbol{X},z}/4$ denotes the pooled standard deviation with $\boldsymbol{S}^2_{\boldsymbol{X},z} = sd(\boldsymbol{x}^s_z)^2 + sd(\boldsymbol{x}^r_z)^2$, with $sd(\boldsymbol{x}^s_z)$, $sd(\boldsymbol{x}^r_z)$ denoting the unweighted standard deviation of covariates $(\boldsymbol{x}_s, \boldsymbol{x}_r)$ from the zth group. In addition, Euclidean distance $\|\mathfrak{p} - \mathfrak{q}\|_2 = \sqrt{\sum_{i=1}^{n} \sum_{p=1}^{Q} \left(\mathfrak{p}_{ip} - \mathfrak{q}_{ip}\right)^2}$ for two points $(\mathfrak{p}, \mathfrak{q})$ in Euclidean $n \times Q$-space. Then, we can specify $max_{z \neq z'}|ASD(z, z')|$ as the balance metric for the whole set of covariates. Alternatively, the interference weights criterion (5.1) also indicates the weighted covariate balance between each group and the target population, $f_z(\boldsymbol{x}_s, \boldsymbol{x}_r)w_z(\boldsymbol{x}_s, \boldsymbol{x}_r) = f_{z'}(\boldsymbol{x}_s, \boldsymbol{x}_r)w_{z'}(\boldsymbol{x}_s, \boldsymbol{x}_r) = f(\boldsymbol{x}_s, \boldsymbol{x}_r)\mathcal{T}(\boldsymbol{x})$, in the context of a tilting function $\mathcal{T}(\boldsymbol{x})$, which is a pre-specified function of covariates. For example, $\mathcal{T}(\boldsymbol{x}) = 1$ corresponds to the inverse probability based balancing weights (5.2), and $\mathcal{T}(\boldsymbol{x}) = \pi^{00}\pi^{10}\pi^{01}\pi^{11}$ corresponds to the overlap-type balancing weights (5.3). Thus, we define the population standardized difference (PSD) for each covariate pair and each treatment interference level (z) as $PSD(z) = ||(\bar{\boldsymbol{x}}^s_z, \bar{\boldsymbol{x}}^r_z) - (\bar{\boldsymbol{x}}^s, \bar{\boldsymbol{x}}^r)||_2/\boldsymbol{S}^2_{\boldsymbol{X}}$, where $\bar{\boldsymbol{x}}^s = \sum_{h=1}^{H} \boldsymbol{x}^s_h \mathcal{T}(\boldsymbol{x}_h) / \sum_{h=1}^{H} \mathcal{T}(\boldsymbol{x}_h)$ is defined as the average value of covariate in the target population. Similarly, we consider the balance metric $max_z|PSD(z)|$ for the overall covariate.

In summary, in this section, we provide some methods to assess the performance of the proposed weights in terms of balance in covariates, in the presence of household interference. Two different measures, $ASD(z, z')$ and $PSD(z)$, are proposed to assess balance in the presence of interference, and thus test the performance of the weights from the cases where we have assumed both independence of treatment and the correlation between treatments in the same household.

5.4 Proportional Odds Model for Household Ordinal Utilities

5.4.1 Single-stage Decisions with Household Ordinal Utilities

From another perspective, in the presence of interference, the proper outcomes should be related to both individuals of a couple in the same household (Lewis et al. [2006]); thus, both covariates and treatments of individuals in the same household need to be considered in a household outcome model. First, let us examine the household utility function, which can be defined as a combination of the individuals' outcomes in the same household. For example, for a pair (s, r), we have the utility that $U(Y^s,\ Y^r)$ is equivalent to $\omega_s Y^s + \omega_r Y^r$, where the combination weights (ω_s and ω_r) can be set based on the specific analytical goals. For instance, we might set $\omega_s = \omega_r$ if outcomes of both s and r are equally considered. Additionally, the ego and alter case, where we aim to optimize the outcome of egos, is another example, which corresponds to ego and alter having weights of 1 and 0, respectively. For the binary outcome pairs (Y^s, Y^r), where $(Y^s, Y^r) \in \{(0,0), (0,1), (1,0), (1,1)\}$, for simplicity and the goal of studying DTR with ordinal outcomes, we will specify that all the combination weights are equal to one ($\omega_s = \omega_r = 1$), that is, the utility is summation of outcomes from a pair ($Y^s + Y^r$). Adding 1 to each sum, there are three possibilities 1, 2, or 3 of $U(Y^s,\ Y^r)$ for a pair in the same household, and these can be considered ordinal outcomes for the household. That is, in this setting of a household's utility, the utilities of households, $U(Y^s,\ Y^r) = 1, 2, 3$, can be interpreted in an ordered way: for a pair in a household, (1) neither, (2) one, or (3) both of them incur a benefit such as smoking cessation, and the largest value (i.e., 3) is preferred. Building on household utility, we consider that the utility model can be captured in the form of a function of

$$f(\boldsymbol{x}^{\beta}) + d_{\xi}(a^s, \boldsymbol{x}^{\xi}) + d_{\psi}(a^r, \boldsymbol{x}^{\psi}) + d_{int}(a^s a^r, \boldsymbol{x}^{\phi}), \tag{5.7}$$

where $\boldsymbol{x}^{\beta}$, often termed predictive variables, function to increase the precision of estimates, and $\boldsymbol{x}^{\xi}$, $\boldsymbol{x}^{\psi}$, $\boldsymbol{x}^{\phi}$, the so-called prescriptive or tailoring variables, are used to adapt treatment decisions to pairs in a household. That is, the model has a treatment-free function $f(\boldsymbol{x}^{\beta})$ and some decision functions $d_{\xi}(a^s, \boldsymbol{x}^{\xi})$, $d_{\psi}(a^r, \boldsymbol{x}^{\psi})$, and $d_{int}(a^s a^r, \boldsymbol{x}^{\phi})$. In practice, in the household-level model (5.7), covariates $\boldsymbol{x}^{\xi}$ and $\boldsymbol{x}^{\psi}$ can be individual-level covariates from each individual in the same household. These two covariates that contain individuals' characteristics indicate the "personalized" side of the model (5.7). Covariates $\boldsymbol{x}^{\phi}$ can be household-level covariates, and thus they represent households' characteristics. $\boldsymbol{x}^{\phi}$ are special tailoring variables for our household-interference treatment decisions case. Given the above utility model, the goal is to identify an *optimal household treatment decision rule* $d^*(\boldsymbol{x}_s, \boldsymbol{x}_r) : \mathcal{X}_s \times \mathcal{X}_r \rightarrow \mathcal{A}^s \times \mathcal{A}^r$ that maximizes the utility $U(Y^s,\ Y^r)$, for binary outcomes Y^s and Y^r, where $\mathcal{X}_s$ and $\mathcal{X}_r$ denote the support of $\boldsymbol{x}_s$ and $\boldsymbol{x}_r$, respectively. In our household case, the treatment decision rule $d(\boldsymbol{x}_s, \boldsymbol{x}_r)$ takes as input both individuals' covariates

and outputs a treatment configuration for a couple in the same household.

Let U be an ordinal outcome with $C = 3$ categories. Then $\mathbb{P}(U \leq c)$ is the cumulative probability of U less than or equal to a specific category $c = 1, ..., C-1$. The log-odds of being less than or equal to a particular c category can be defined as

$$log\frac{\mathbb{P}(U \leq c)}{\mathbb{P}(U > c)} = logit[\mathbb{P}(U \leq c)], \quad \text{for } c = 1, ..., C-1,$$

where the logit link function is defined as $logit(p) = p/(1-p)$. Note that the denominator $\mathbb{P}(U > c)$ of the above equation will be zero if $c = C$; thus, $c = 1, ..., C-1$. The *proportional odds model* (POM) that specifies the cumulative log-odds for a particular category assumes that each explanatory variable exerts the same effect on each cumulative logit regardless of the cutoff c, and is proposed by McCullagh [1980] to be $logit[\mathbb{P}(U_h \leq c \mid \boldsymbol{x}_h)] = \zeta_c - \boldsymbol{\theta}^\top \boldsymbol{x}_h$, where coefficients ζ_c are category-specific intercepts and $\boldsymbol{\theta}$ are coefficients of covariates $\boldsymbol{x}_h$. The intercepts ζ_c are the only part that varies across the equations, and the effects of covariates $\boldsymbol{x}_h$ are assumed to be constant for all c, i.e., $\boldsymbol{\theta}_c = \boldsymbol{\theta}$. In the POM ζ_c represents the cumulative log-odds of the utility being in category c or lower for those in the reference exposure group, whereas $\boldsymbol{\theta}$ represents a cumulative log-odds ratio: the odds of "at least c" under two different conditions of $\boldsymbol{x}_h$. Because $\boldsymbol{\theta}$ values are not dependent on the cutoff c, the odds ratio is constant across each split of outcome domain; hence $logit[\mathbb{P}(U_h \leq c \mid \boldsymbol{x}_h)] = \zeta_c - \boldsymbol{\theta}^\top \boldsymbol{x}_h$ is named the "proportional odds model". The sign of $\boldsymbol{\theta}$ in the model fit (i.e., in $\zeta_c - \boldsymbol{\theta}^\top \boldsymbol{x}_h$, $-\boldsymbol{\theta}$ is used) provides a certain relationship between covariates and outcome that, if $\boldsymbol{\theta} > 0$, an increase in $\boldsymbol{x}_h$ is associated with an increase in U_h. Based on the POM, we can calculate $\mathbb{P}(U_h = c)$ for each c. When $c = 1$, for instance, $\mathbb{P}(U_h = 1 \mid \boldsymbol{x}_h) = \mathbb{P}(U_h \leq 1) = \text{expit}(\zeta_1 - \boldsymbol{\theta}^\top \boldsymbol{x}_h)$, then, for $c = 2, ..., C-1$, $\mathbb{P}(U_h = c \mid \boldsymbol{x}_h) = \mathbb{P}(U_h \leq c \mid \boldsymbol{x}_h) - \mathbb{P}(U_h \leq c-1 \mid \boldsymbol{x}_h)$. Building on the typical POM and our treatment decision set-up, we propose a proportional odds model for household ordinal utilities as follows, for $c = 1, 2;\ \ h = 1, 2, ...H$,

$$logit[\mathbb{P}(U_h \leq c \mid a_h^s, a_h^r, \boldsymbol{x}_h)] = \zeta_c - \boldsymbol{\beta}^\top \boldsymbol{x}_h^\beta - a_h^s \boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi - a_h^r \boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi - a_h^s a_h^r \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi. \tag{5.8}$$

According to the general utility model (5.7), the treatment-free functions are identified in the above POM model as a linear form $f(\boldsymbol{x}^\beta) = \zeta_c - \boldsymbol{\beta}^\top \boldsymbol{x}^\beta$, and the decision functions are specified as $d_\xi(a^s, \boldsymbol{x}^\xi) = -a^s \boldsymbol{\xi}^\top \boldsymbol{x}^\xi$, $d_\psi(a^r, \boldsymbol{x}^\psi) = -a^r \boldsymbol{\psi}^\top \boldsymbol{x}^\psi$ and $d_{int}(a^s a^r, \boldsymbol{x}^\phi) = -a^s a^r \boldsymbol{\phi}^\top \boldsymbol{x}^\phi$, respectively.

Building on the household ordinal outcome (5.8), we define the *household blip function* as $\gamma[(A^s, A^r), \boldsymbol{x}_h; \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}] = A^s \boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi + A^r \boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi + A^s A^r \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi$, which represents the effects of the treatment configuration (A^s, A^r) for a household compared with the null treatment configuration $(0, 0)$. The goal is obviously to estimate blip parameters $\boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}$; building on these blip-parameter estimates and given the household tailoring variables, the optimal treatment decisions for a pair in

A_h^s \ A_h^r	1	0
1	$\boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi + \boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi + \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi$	$\boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi$
0	$\boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi$	0

Table 5.2: Blip values for different treatment configurations (A_h^s, A_h^r)

the household can be made. Given the four choices of $(A^s, A^r) = (1,1), (1,0), (0,1)$ or $(0,0)$, as Table 5.2 presents, the corresponding blip value $\gamma[(A^s, A^r), \boldsymbol{x}_h; \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}]$ is $\boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi + \boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi + \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi$, $\boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi$, $\boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi$, and 0, respectively. The goal is to maximize the outcome across a couple, which is equivalent to maximising the blip function. Taking into account the blip values of all possible treatment configurations, an optimal treatment rule must choose the configuration that corresponds to the maximum blip value. Therefore, we have the following treatment decision rules for a household:

Decision 1. The optimal household decision rules:

- Rule 1: $d^*(\boldsymbol{x}^\xi, \boldsymbol{x}^\psi, \boldsymbol{x}^\phi) = (1,1)$, if $\boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi + \boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi + \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi > 0$ and $\boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi + \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi > 0$, and $\boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi + \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi > 0$.
- Rule 2: $d^*(\boldsymbol{x}^\xi, \boldsymbol{x}^\psi, \boldsymbol{x}^\phi) = (1,0)$, if $\boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi + \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi < 0$ and $\boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi > \boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi$ and $\boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi > 0$.
- Rule 3: $d^*(\boldsymbol{x}^\xi, \boldsymbol{x}^\psi, \boldsymbol{x}^\phi) = (0,1)$, if $\boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi + \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi < 0$ and $\boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi > \boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi$ and $\boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi > 0$.
- Rule 4: $d^*(\boldsymbol{x}^\xi, \boldsymbol{x}^\psi, \boldsymbol{x}^\phi) = (0,0)$, if $\boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi + \boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi + \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi < 0$ and $\boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi < 0$ and $\boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi < 0$.

Further, if we know the blip parameters $\boldsymbol{\xi}, \boldsymbol{\psi}$, and $\boldsymbol{\phi}$, then we have $\gamma^*[d^*(\boldsymbol{x}^\xi, \boldsymbol{x}^\psi, \boldsymbol{x}^\phi); \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}] = A^{s*} \boldsymbol{\xi}^\top \boldsymbol{x}^\xi + A^{r*} \boldsymbol{\psi}^\top \boldsymbol{x}^\psi + A^{s*} A^{r*} \boldsymbol{\phi}^\top \boldsymbol{x}^\phi$, where γ^* means the arguments (A^s, A^r) in the γ function follow the optimal household decision rules Decision 1, and $d^*(\boldsymbol{x}^\xi, \boldsymbol{x}^\psi, \boldsymbol{x}^\phi)$ and (A^{s*}, A^{r*}) are the corresponding optimal treatments for the pair (i, r). Thus, we can define the regret function, which is the expected loss or regret a medical practitioner experiences when forced to prescribe treatments (A^s, A^r) instead of the optimal treatments (A^{s*}, A^{r*}), as $\mu[d^*, (A^s, A^r); \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}] = \gamma^*[d^*(\boldsymbol{x}^\xi, \boldsymbol{x}^\psi, \boldsymbol{x}^\phi); \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}] - \gamma[(A^s, A^r), \boldsymbol{x}; \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}] = (A^{s*} - A^s) \boldsymbol{\xi}^\top \boldsymbol{x}^\xi + (A^{r*} - A^r) \boldsymbol{\psi}^\top \boldsymbol{x}^\psi + (A^{s*} A^{r*} - A^s A^r) \boldsymbol{\phi}^\top \boldsymbol{x}^\phi$. By using the regret function, the notations in the multistage treatment decision will become simpler, and we will make use of the function to develop the DTR key concepts in the household interference case.

As a result, based on the proposed cumulative probability model (5.8), we can also calculate each category's utility probability:

$$\begin{aligned}
\mathbb{P}(U_h = 1 \mid a_h^s, a_h^r, \boldsymbol{x}_h) &= \text{expit}[\zeta_1 - \boldsymbol{\beta}^\top \boldsymbol{x}_h^\beta - a_h^s \boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi - a_h^r \boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi - a_h^s a_h^r \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi] \\
\mathbb{P}(U_h = 2 \mid a_h^s, a_h^r, \boldsymbol{x}_h) &= \text{expit}[\zeta_2 - \boldsymbol{\beta}^\top \boldsymbol{x}_h^\beta - a_h^s \boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi - a_h^r \boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi - a_h^s a_h^r \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi] \\
&\quad - \text{expit}[\zeta_1 - \boldsymbol{\beta}^\top \boldsymbol{x}_h^\beta - a_h^s \boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi - a_h^r \boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi - a_h^s a_h^r \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi] \\
\mathbb{P}(U_h = 3 \mid a_h^s, a_h^r, \boldsymbol{x}_h) &= 1 - \text{expit}[\zeta_2 - \boldsymbol{\beta}^\top \boldsymbol{x}_h^\beta - a_h^s \boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi - a_h^r \boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi - a_h^s a_h^r \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi].
\end{aligned} \tag{5.9}$$

5.4.2 Balancing Property

In the previous section, we investigated the POM for the household ordinal utilities and proposed the corresponding decision rules. To achieve the robust blip parameter estimation on which the decision rules depend, we need to develop the balancing weights. The motivation behind the balancing weights or balancing properties comes from the estimation equation systems of the POM. The key concept is the propensity score function, but in our household interference case, it is the joint propensities function. Following and extending the balancing-weight derivation process introduced in Chapter 4, we denote by w^{sd} a choice of "standard" interference-aware balancing weights that satisfy the criterion, that is,

$$\pi^{00} w^{sd}(0,0,\boldsymbol{x}) = \pi^{01} w^{sd}(0,1,\boldsymbol{x}) = \pi^{10} w^{sd}(1,0,\boldsymbol{x}) = \pi^{11} w^{sd}(1,1,\boldsymbol{x}). \tag{5.10}$$

One option of the "standard" interference-aware balancing weights is the overlap weights (5.3). Then we have the balance property as follows:

Theorem 5.1. *Balancing Property with Household Interference for Ordinal Outcomes*

When the true ordinal-outcome model satisfies $logit[\mathbb{P}(U \le c \mid a^s, a^r, \boldsymbol{x})] = \zeta_c - f(\boldsymbol{x}^\beta; \boldsymbol{\beta}) - a^s \boldsymbol{\xi}^\top \boldsymbol{x}^\xi - a^r \boldsymbol{\psi}^\top \boldsymbol{x}^\psi - a^s a^r \boldsymbol{\phi}^\top \boldsymbol{x}^\phi$, *for* $c = 1, 2$, *a weighted proportional odds model based on the corresponding linear model will yield approximately consistent estimators of* $\boldsymbol{\xi}$, $\boldsymbol{\psi}$ *as well as* $\boldsymbol{\phi}$ *if at least one of the treatment and treatment-free models is correctly specified, and the weights satisfy*

$$\pi^{00} w(0,0,\boldsymbol{x})\kappa(0,0) = \pi^{01} w(0,1,\boldsymbol{x})\kappa(0,1) = \pi^{10} w(1,0,\boldsymbol{x})\kappa(1,0) = \pi^{11} w(1,1,\boldsymbol{x})\kappa(1,1), \tag{5.11}$$

where

$$\kappa(a^s, a^r) = expit\,(\eta_2)\,[1 - expit\,(\eta_1)]\,[1 - expit\,(\eta_2) + expit\,(\eta_1)]\,, \tag{5.12}$$

with $\eta_1(a^s, a^r, \boldsymbol{x}) = \zeta_1^* + \boldsymbol{\beta}^{*\top} \boldsymbol{x}^\beta + \boldsymbol{\xi}^{*\top} a^s \boldsymbol{x}^\xi + \boldsymbol{\psi}^{*\top} a^r \boldsymbol{x}^\psi + \boldsymbol{\phi}^{*\top} a^s a^r \boldsymbol{x}^\phi$, $\eta_2(a^s, a^r, \boldsymbol{x}) = \zeta_2^* + \boldsymbol{\beta}^{*\top} \boldsymbol{x}^\beta + \boldsymbol{\xi}^{*\top} a^s \boldsymbol{x}^\xi + \boldsymbol{\psi}^{*\top} a^r \boldsymbol{x}^\psi + \boldsymbol{\phi}^{*\top} a^s a^r \boldsymbol{x}^\phi$, *and* ζ_1^*, ζ_2^*, $\boldsymbol{\beta}^*$, $\boldsymbol{\xi}^*$, $\boldsymbol{\psi}^*$ *and* $\boldsymbol{\phi}^*$ *are the solutions of the estimation functions of the POM (5.8) with standard interference-aware balancing weight satisfying (5.10).*

Proof: See Appendix H.2.

Similar to our balancing properties in dWOLS and dWGLM, the balancing properties of POM rely on the propensity score; however, the inference of household interference depends on the joint propensity functions, which have been discussed in Section 5.2, in terms of estimation and construction of the balancing weights. The key factor of the balancing criterion (5.11) is the κ, as presented in Chapter 4, called the "adjustment" factor. It adjusts for the nonlinearity of the link function, and it is special for the POM. Based on (5.12), we can conclude that the adjustment factor is the product of three terms: $\text{expit}(\eta_2)$, $1 - \text{expit}(\eta_1)$, and $1 - \text{expit}(\eta_2) + \text{expit}(\eta_1)$, where the first term $\text{expit}(\eta_2)$ represents the estimated cumulative probabilities of categorical utilities 1 and 2, the second $1 - \text{expit}(\eta_1)$ represents the estimated cumulative probabilities of categorical utilities 2 and 3, and the third term $1 - \text{expit}(\eta_2) + \text{expit}(\eta_1)$ represents the estimated cumulative probabilities of categorical utilities 1 and 3. Alternatively, these three terms of κ (5.12) also can be expressed as $\text{expit}(\eta_2) = 1 - \mathbb{P}(U = 3)$, $1 - \text{expit}(\eta_1) = 1 - \mathbb{P}(U = 1)$, and $1 - \text{expit}(\eta_2) + \text{expit}(\eta_1) = 1 - \mathbb{P}(U = 2)$. Then the "adjustment" factor of (5.12) can be written as $\kappa(a^s, a^r, \boldsymbol{x}) = \Pi_{c=1}^{3}[1 - \mathbb{P}(U = c)]$. Note that the "adjustment" factor of (5.12) is specific to the POM with our $C = 3$ case, but for $C > 3$, the corresponding "adjustment" factor can be derived as $\kappa(a^s, a^r, \boldsymbol{x}) = \Pi_{c=1}^{C}[1 - \mathbb{P}(U = c)]$. Again, as stressed in Section 4.2.3 in the context of dWGLM, the term "approximately consistent" refers to a case where the estimators are derived from the estimating functions which are approximately unbiased with a small quantifiable bias. This quantifiable bias will be small when a linear predictor tends to vary in an interval where g^{-1} is approximately linear.

In this section, considering household interference, for a single-stage decision problem, we have proposed a proportional odds model for household ordinal utilities, that is model (5.8). Based on our model (5.8), we have also provided the treatment assignment mechanisms for the household interference case, for the purpose of optimising the combined outcomes of a pair in the same household. Further, based on the nonlinearity of the link function, we developed adjusted balancing weights for the model (5.8), for which the parameters of interest will be consistently estimated even if the treatment-free model is misspecified.

5.4.3 Multiple-stage Decisions with Household Ordinal Utilities

Having investigated the process for estimating the single-stage optimal treatment configurations for household ordinal utilities, we now continue to the sequential treatment decision in the case of optimal treatment regimes estimation with interference for household ordinal utilities.

For the multistage treatment decision setting, backward induction is utilized in most methods for sequential decision problems. Therefore, the multistage treatment decision problems can be

broken down into a group of single-stage decision problems. Then, for each stage, we employ a weighted POM to consistently estimate the blip parameters, i.e., $\boldsymbol{\xi}$, $\boldsymbol{\psi}$, and $\boldsymbol{\phi}$. Accordingly, we name our novel approach for DTR estimation with ordinal outcomes the dynamic weighted proportional odds model, namely dWPOM.

If we acquire parameter estimates $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\psi}}$, and $\hat{\boldsymbol{\phi}}$, then we have estimated optimal treatment blip $\hat{\gamma}[\hat{d}^*(\boldsymbol{x}^{\xi}, \boldsymbol{x}^{\psi}, \boldsymbol{x}^{\phi}); \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\phi}}] = \hat{A}^{s*}\hat{\boldsymbol{\xi}}^{\top}\boldsymbol{x}^{\xi} + \hat{A}^{r*}\hat{\boldsymbol{\psi}}^{\top}\boldsymbol{x}^{\psi} + \hat{A}^{s*}\hat{A}^{r*}\hat{\boldsymbol{\phi}}^{\top}\boldsymbol{x}^{\phi}$, where the estimated optimal decisions $\hat{d}^*(\boldsymbol{x}^{\xi}, \boldsymbol{x}^{\psi}, \boldsymbol{x}^{\phi}) = (\hat{A}^{s*}, \hat{A}^{r*})$ also depend on estimates $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\psi}}$, and $\hat{\boldsymbol{\phi}}$, and can be calculated by decision rules in Decision 1. Further, we can generate household level *ordinal pseudo-utility* based on the ordinal pseudo-utility probability that:

$$
\begin{aligned}
&\mathbb{P}(\widetilde{\mathcal{U}_h} = 1 \mid \hat{d}_h^*, \boldsymbol{x}_h) = \operatorname{expit}\left(\hat{\zeta}_1 - \hat{\boldsymbol{\beta}}^{\top}\boldsymbol{x}_h^{\beta} - \hat{\gamma}_h[\hat{d}_h^*; \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\phi}}]\right),\\
&\mathbb{P}(\widetilde{\mathcal{U}_h} = 2 \mid \hat{d}_h^*, \boldsymbol{x}_h) = \operatorname{expit}\left(\hat{\zeta}_2 - \hat{\boldsymbol{\beta}}^{\top}\boldsymbol{x}_h^{\beta} - \hat{\gamma}_h[\hat{d}_h^*; \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\phi}}]\right) - \operatorname{expit}\left(\hat{\zeta}_1 - \hat{\boldsymbol{\beta}}^{\top}\boldsymbol{x}_h^{\beta} - \hat{\gamma}_h[\hat{d}_h^*; \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\phi}}]\right),\\
&\mathbb{P}(\widetilde{\mathcal{U}_h} = 3 \mid \hat{d}_h^*, \boldsymbol{x}_h) = 1 - \operatorname{expit}\left(\hat{\zeta}_2 - \hat{\boldsymbol{\beta}}^{\top}\boldsymbol{x}_h^{\beta} - \hat{\gamma}_h[\hat{d}_h^*; \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\phi}}]\right).
\end{aligned}
\tag{5.13}
$$

Therefore, building on equation (5.13) and the estimates, we can compute the ordinal pseudo-utility probability, which is employed in the multiple-stage treatment decision settings. This ordinal pseudo-utility probability represents the probability of the potential outcome that a household with the given history would have if they go on to receive the optimal treatment configuration in the current stage. Similar to our approach in Chapter 4, to increase the estimation efficiency, we still generate the ordinal pseudo-utility probability $\mathcal{R}$ times, and conduct $\mathcal{R}$ times estimation for the parameters of interest. Then, the final estimates of parameters are the averages of these $\mathcal{R}$ estimates.

For the multistage decision analysis, the dWPOM procedure could be implemented by the following steps at each stage of the analysis, starting from the last stage K and working backwards towards the first stage (subscript j indicates the number of stages):

- **Step 1:** Construct the stage j ordinal *pseudo-utility*: set $\widetilde{\mathcal{U}_j} = u_K$, where u_K is the observed value of U_K, if $j = K$. Otherwise, use prior estimates $\hat{\boldsymbol{\beta}}_K$, $\underline{\hat{\boldsymbol{\xi}}}_{j+1} = (\hat{\boldsymbol{\xi}}_{j+1}, ..., \hat{\boldsymbol{\xi}}_K)$, $\underline{\hat{\boldsymbol{\psi}}}_{j+1} = (\hat{\boldsymbol{\psi}}_{j+1}, ..., \hat{\boldsymbol{\psi}}_K)$, and $\underline{\hat{\boldsymbol{\phi}}}_{j+1} = (\hat{\boldsymbol{\phi}}_{j+1}, ..., \hat{\boldsymbol{\phi}}_K)$ to randomly generate $\widetilde{\mathcal{U}_j}$, which takes the ordinal value c with the ordinal probability $\mathbb{P}(\widetilde{\mathcal{U}_j} = c)$, $\mathcal{R}$ times, to yield $\widetilde{\mathcal{U}_j^1}, \widetilde{\mathcal{U}_j^2}, ..., \widetilde{\mathcal{U}_j^{\mathcal{R}}}$.

- **Step 2:** Estimate the stage j joint propensity model $\pi^{a^s a^r}(\boldsymbol{h}_{js}, \boldsymbol{h}_{jr})$ (e.g., via alternating logistic regression), then compute the corresponding the association-concerned interference

balancing weights, such as

$$w(a_j^s, a_j^r) = \frac{\pi^{00}\pi^{10}\pi^{01}\pi^{11}}{\pi^{a_j^s a_j^r}} \text{ for } a_j^s = 0, 1; a_j^r = 0, 1.$$

- **Step 3:** Specify the stage j treatment-free and blip models, and perform a weighted cumulative link mixed model of $\widetilde{\mathcal{U}_j^{\mathfrak{r}}}$ on the terms in the treatment-free and blip models, using weights from Step 2 to get estimates $\hat{\zeta}_{1j}^{*\,\mathfrak{r}}$, $\hat{\zeta}_{2j}^{*\,\mathfrak{r}}$, $\hat{\boldsymbol{\beta}}^{*\mathfrak{r}}{}_j$, $\hat{\boldsymbol{\xi}}^{*\mathfrak{r}}{}_j$, $\hat{\boldsymbol{\psi}}^{*\mathfrak{r}}{}_j$, and $\hat{\boldsymbol{\phi}}^{*\mathfrak{r}}{}_j$ for $\mathfrak{r} = 1, ..., \mathcal{R}$; for example, for each $\mathfrak{r} = 1, ..., \mathcal{R}$, use the POM, for $c = 1, 2$,

$$logit[\mathbb{P}(\widetilde{\mathcal{U}_j^{\mathfrak{r}}} \leq c \mid a_j^s, a_j^r, \boldsymbol{h}_j; \boldsymbol{\xi}_j, \boldsymbol{\psi}_j, \boldsymbol{\phi}_j)] = \zeta_{cj} - \boldsymbol{\beta}_j^\top \boldsymbol{h}_j^\beta - a_j^s \boldsymbol{\xi}_j^\top \boldsymbol{h}_j^\xi - a_j^r \boldsymbol{\psi}_j^\top \boldsymbol{h}_j^\psi - a_j^s a_j^r \boldsymbol{\phi}_j^\top \boldsymbol{h}_j^\phi.$$

- **Step 4:** Use estimates from Step 3 (i.e., $\hat{\zeta}_{1j}^{*\,\mathfrak{r}}$, $\hat{\zeta}_{2j}^{*\,\mathfrak{r}}$, $\hat{\boldsymbol{\beta}}^{*\mathfrak{r}}{}_j$, $\hat{\boldsymbol{\xi}}^{*\mathfrak{r}}{}_j$, $\hat{\boldsymbol{\psi}}^{*\mathfrak{r}}{}_j$, and $\hat{\boldsymbol{\phi}}^{*\mathfrak{r}}{}_j$) to compute

$$\kappa^{\mathfrak{r}}(a_j^s, a_j^r, \boldsymbol{h}_j) = \text{expit}\,(\hat{\eta}_2)\,[1 - \text{expit}\,(\hat{\eta}_1)]\,[1 - \text{expit}\,(\hat{\eta}_2) + \text{expit}\,(\hat{\eta}_1)]\,,$$

where $\hat{\eta}_c = \hat{\eta}_c(a_j^s, a_j^r, \boldsymbol{h}_j) = \hat{\zeta}_c - \hat{\boldsymbol{\beta}}^\top \boldsymbol{h}_j^\beta - \hat{\boldsymbol{\xi}}^\top a_j^s \boldsymbol{h}_j^\xi - \hat{\boldsymbol{\psi}}^\top a_j^r \boldsymbol{h}_j^\psi - \hat{\boldsymbol{\phi}}^\top a_j^s a_j^r \boldsymbol{h}_j^\phi$, with values $(\hat{\zeta}_c, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\phi}}) = (\hat{\zeta}_{cj}^{*\,\mathfrak{r}}, \hat{\boldsymbol{\beta}}^{*\mathfrak{r}}{}_j, \hat{\boldsymbol{\xi}}^{*\mathfrak{r}}{}_j, \hat{\boldsymbol{\psi}}^{*\mathfrak{r}}{}_j, \hat{\boldsymbol{\phi}}^{*\mathfrak{r}}{}_j)$, for $c = 1, 2$. Then, for $a_j^s,\ a_j^r = 0, 1$, construct the new weights

$$w^{new,\mathfrak{r}}(a_j^s, a_j^r) = \frac{\pi^{00}\pi^{10}\pi^{01}\pi^{11}}{\pi^{a_j^s a_j^r}} \times \frac{\kappa^{\mathfrak{r}}(0, 0, \boldsymbol{h}_j)\kappa^{\mathfrak{r}}(1, 0, \boldsymbol{h}_j)\kappa^{\mathfrak{r}}(0, 1, \boldsymbol{h}_j)\kappa^{\mathfrak{r}}(1, 1, \boldsymbol{h}_j)}{\kappa^{\mathfrak{r}}(a_j^s, a_j^r, \boldsymbol{h}_j)}. \quad (5.14)$$

- **Step 5:** Perform a weighted POM with the new weights (i.e., $w^{new,\mathfrak{r}}(a_j^s, a_j^r)$) to get revised estimates $\hat{\boldsymbol{\xi}}_j^{\mathfrak{r}}$, $\hat{\boldsymbol{\psi}}_j^{\mathfrak{r}}$, and $\hat{\boldsymbol{\phi}}_j^{\mathfrak{r}}$ for each $\mathfrak{r}$, and estimate $\boldsymbol{\xi}_j$, $\boldsymbol{\psi}_j$, and $\boldsymbol{\phi}_j$ by $\hat{\boldsymbol{\xi}}_j = \mathcal{R}^{-1}\sum_r \hat{\boldsymbol{\xi}}_j^{\mathfrak{r}}$, $\hat{\boldsymbol{\psi}}_j = \mathcal{R}^{-1}\sum_r \hat{\boldsymbol{\psi}}_j^{\mathfrak{r}}$, and $\hat{\boldsymbol{\phi}}_j = \mathcal{R}^{-1}\sum_r \hat{\boldsymbol{\phi}}_j^{\mathfrak{r}}$, respectively, then use parameter estimators $\hat{\boldsymbol{\xi}}_j$, $\hat{\boldsymbol{\psi}}_j$, and $\hat{\boldsymbol{\phi}}_j$ to construct the j^{th} stage optimal treatment rule, which is based on Decision 1.
- Step 6: Return to Step 1 and analyze stage $j - 1$ if there are more stages to analyze.

5.5 Simulation Studies

In this section, we provide two simulation studies (Study 1 and 2) to illustrate our proposed methods for estimating optimal DTR with ordinal outcomes under household interference. Furthermore, we also demonstrate and verify the double robustness of our method in both studies. In Study 1,

we consider single-stage treatment decision problems, and in Study 2, we investigate a multi-stage decision problem by means of a two-stage case.

To assess the performance of methods, we construct three measures: (1) mean optimal treatment rate, (2) regret value, and (3) value functions for ordinal outcomes. First, based on the data-generating parameters, we can calculate the authentic optimal treatments for each household. Then, we can construct the recommended treatments from the estimated rules based on the estimated decision parameters. The mean optimal treatment rate is then the percentage of the estimated recommended treatments that are in accord with the authentic optimal treatments. Second, the regret value measures the difference between the blip value under the true optimal regime and under the estimated regime. Finally, we construct value functions for ordinal outcomes, which mainly compare the estimated optimal treatments with the observed treatments. In the later section, we will give the formal definition of the value functions for ordinal outcomes building on the concept of the odds ratio. It is important to note that, because of the specific nature of ordinal outcomes, for the single-stage settings in Study 1, we mainly focus on the consistent estimation. Therefore, we consider both the mean optimal treatment rate and regret value in Study 1. For the multi-stage settings in Study 2, we primarily concentrate on the long-term treatment effects of DTR. Then, we examine value functions for ordinal outcomes to compare different methods.

5.5.1 Single-stage Treatment Decision for a Couples Case

In this single-stage decision setting (*Study 1*), to examine the double robustness of the proposed method, we specifically examine four scenarios:

- Scenario 1: neither the treatment-free model nor the treatment model is correctly specified.
- Scenario 2: the treatment-free model is correctly specified but the treatment model is misspecified.
- Scenario 3: the treatment model is correctly specified but the treatment-free model is misspecified.
- Scenario 4: both treatment-free model and treatment model are correctly specified.

Scenario 1 fails to specify a correct model, so consistent estimation of the blip parameters cannot be guaranteed. However, Scenarios 2, 3, and 4 correctly specify at least one of the treatment-free and treatment models, so the estimator of blip parameters should be close to consistent based on our theory. In addition, note that we have only linear terms in our POM, while

the true models can contain non-linear terms. If the true models contain non-linear terms, then we have misspecified the model. Moreover, in a real application, it is more challenging to correctly specify the treatment-free model than the treatment model, so we particularly highlight the results of Scenario 3.

In each scenario, five different methods are investigated. Method 0 (M0) employs the proposed proportional odds model (5.8) without any balancing weights. Method 1 (M1) considers the same POM and uses the interference balancing weights in Chapter 3, but assumes independence between the treatments. That is, $w = |A^s - \mathbb{P}(A^s = 1 \mid \boldsymbol{x}_s)| * |A^r - \mathbb{P}(A^r = 1 \mid \boldsymbol{x}_r)|$. However, Methods 2 and 3 (M2 and M3) both consider the same POM, yet use the proposed interference balancing weights, which allow dependence between the treatments within the same household. In particular, M2 employs the inverse probability-based weights (5.2) and M3 uses the overlap-type weights (5.3). Furthermore, to contrast the performance of the weights in M2 and M3 with the weights that include the adjustment factor, i.e., (5.12), we also consider Method 4 (M4), that is, using the same POM (5.8) with the adjusted overlap weights (5.14).

Note that, from the perspective of methodology, M0 is Q-learning, and M1, M2, M3, and M4 belong to our proposed dWPOM yet with different balancing weights. M1 uses a no-treatment-association dWPOM, but M2, M3, and M4 use treatment-association aware dWPOMs. Methods M2 and M3 employ inverse probability type and overlap type weights, respectively. However, M4 utilizes adjusted overlap type weights. The adjusted weights (5.14) satisfy the weight criterion in Theorem 5.1, so M4 is expected to provide close to consistent blip parameter estimators in Scenarios 2, 3, and 4.

In the generation of ordinal outcomes for the households, based on the mixed cumulative logit model (5.8), the ordinal outcome is a function of household treatment assignments and covariates $\boldsymbol{x}^\beta$, $\boldsymbol{x}^\xi$, $\boldsymbol{x}^\psi$ and $\boldsymbol{x}^\phi$. In this study, we take $H = 3000$ households with $B = 500$ Monte Carlo replicates. For each individual, covariates are generated as $x_1 \sim U[0,1]$ that is uniformly distributed on $[0,1]$, $x_2 \sim N(0,1)$ that is normally distributed with mean 0 and variance 1, and $x_3 \sim Ber(0.5)$, $x_4 \sim Ber(0.75)$ that are Bernoulli distributed with success probability 0.5 and 0.75, respectively. Supposing each household contains two individuals denoted as (s,r), we let $\boldsymbol{x}^\beta$ and $\boldsymbol{x}^\phi$ include both individuals' covariates, but $\boldsymbol{x}^\xi$ and $\boldsymbol{x}^\psi$ each include only the single individual's covariates. For example, in this simulation, denoting $\boldsymbol{x}^s$ and $\boldsymbol{x}^r$ as the covariates of (s,r), respectively, we let $\boldsymbol{x}^\beta = (1, x_1^s, x_2^s, x_3^s, x_4^s, 1, x_1^r, x_2^r, x_3^r, x_4^r)$, $\boldsymbol{x}^\xi = (1, x_1^s)$, $\boldsymbol{x}^\psi = (1, x_1^r)$, and $\boldsymbol{x}^\phi = (1, x_3^s + x_3^r)$, where $x^s + x^r$ are the household-level data which is the sum of two individuals' information. In Study 1, the true treatment-free function is set as $f(\boldsymbol{x}^\beta) = \cos(x_1^s + x_1^r) - (x_1^s + x_1^r)^3 - \log|1/x_1^s| - 2 * (x_1^r)^2 + (x_3^s + x_3^r)^3$. It is important to note that this true treatment-free setting relies on various non-linear functions, including both even and odd functions. It also consists of both household-level information, such as $x_1^s + x_1^r$ and $x_3^s + x_3^r$, and individual-level information, including x_1^s and x_1^r. In addition, the true blip function is

$\gamma[(A^s, A^r), \boldsymbol{x}; \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}] = A^s * (-0.5 + x_1^s) + A^r * (-0.5 + x_1^r) + A^s * A^r * [-1 + 0.5 * (x_3^s + x_3^r)]$, where the tailoring variables are set as $\boldsymbol{x}^{\xi} = (1, x_1^s)^\top$, $\boldsymbol{x}^{\psi} = (1, x_1^r)^\top$, and $\boldsymbol{x}^{\phi} = (1, x_3^s + x_3^r)^\top$. The corresponding parameters of interest, that is, the true tailoring parameters, are set as follows: $\boldsymbol{\xi} = (-0.5, 1)^\top$, $\boldsymbol{\psi} = (-0.5, 1)^\top$, and $\boldsymbol{\phi} = (-1, 0.5)^\top$. Therefore, the true mean of the latent household outcome is $\mu = f(\boldsymbol{x}^{\beta}) + A^s * (-0.5 + x_1^s) + A^r * (-0.5 + x_1^r) + A^s * A^r * [-1 + 0.5 * (x_3^s + x_3^r)]$.

Further, as introduced in the method section 5.4.1, ζ_c represents the cumulative log odds of being in category c or lower and corresponds to the threshold of the latent continuous data generating process. Our three categorical outcomes case is similar to the case that generates binary outcomes in a latent continuous process, where the observed binary outcome simply indicates whether or not the unseen continuous outcome has exceeded a specified threshold. Our present case needs two thresholds (ζ_1 and ζ_2) to classify the latent continuous outcomes into three categories. The following is the process to generate the three categorical outcomes in our study. In Step 1, we set reference probabilities for the three-category outcome $\boldsymbol{p} = (p_1, p_2, p_3) = (0.65, 0.25, 0.1)$, and compute the cumulative reference probabilities. In Step 2, with the logistic distribution, we map cumulative reference probabilities to thresholds, which are $\zeta_1 = 0.619$, $\zeta_2 = 2.197$. In Step 3, based on individual and household covariates and treatments, we calculate household-level thresholds such that $(\zeta_1 - \mu, \zeta_2 - \mu)$. In Step 4, for each household, based on household-level thresholds, we can compute cumulative probabilities and thus acquire the category probabilities. Finally, in Step 5, we generate household-level category outcomes based on the category probabilities in Step 4.

The treatments generation process follows the methods that are introduced in Section 5.2. The marginal propensity model for each individual is set as $\mathbb{P}(A = 1 \mid \boldsymbol{x}) = \text{expit}(-1.15 + 0.5 * \exp(x_1) - 0.25 * x_2^2 + 0.25 * x_3 + 0.6 * x_4)$, and the odds ratio model is set as $\tau = \exp[-0.25 + 0.25 * (x_1^s + x_1^r) + 0.5 * (x_3^s + x_3^r)]$. Based on the marginal propensity and odds ratio models, the joint probabilities can be generated. Therefore, the treatment configuration (A^s, A^r) can be generated by the corresponding joint propensities.

The treatment decision rules in Decision 1 rely on the estimates of blip parameters, that is, $\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\psi}}$, and $\hat{\boldsymbol{\phi}}$. Figure 5.2 presents the distribution of blip parameter estimates from Methods 0, 1, 2, 3, and 4 in Scenario 3, where the treatment model is correctly specified but the treatment-free model is misspecified. Moreover, the distributions of the blip parameter estimates in Scenarios 1, 2 and 4 are presented in Appendix H.3. From these result figures, in particular Figure 5.2, the approximate consistent estimation of blip parameters ($\boldsymbol{\xi}, \boldsymbol{\psi}$, and $\boldsymbol{\phi}$) from M4 is as expected. That is, the estimates of Method 4 from Scenarios 2, 3 and 4 appear consistent, and this verifies the double robustness of our proposed adjusted weights (5.14) in the simulation setting. However, in Scenario 3, M0, M1, M2, and M3 offer biased blip parameter estimators. Even though M1, M2, and M3 estimators are biased, the bias is relatively smaller than for the M0 estimator which does not employ any balancing weights. Moreover, in this Scenario 3, compared with M1, where independence of the treatments is assumed, M2 and M3, which address the association between

treatments, provide less biased estimators. This result confirms that if a correlation exists between treatments in the same household in truth, failing to take that into account will lead to biased estimation. Furthermore, as Figures H.3 and H.4 in Appendix H.3 indicate, in both Scenarios 2 and 4, where treatment-free models are correctly specified, all the methods, even for M1, provide unbiased estimators of blip parameters. Thus, we find that all the methods in the case where the treatment-free model is correctly specified do not have much distinction.

Furthermore, as mentioned at the beginning of this section, we construct two main measures, the mean optimal treatment rate (MOTR) and value functions for ordinal outcomes, to evaluate the performance of each method that corresponds to different balancing weights. MOTR measures the rate at which the recommended treatment configurations align with those that are truly optimal. The agreement between the estimated optimal treatment regimes and the actual optimal treatment regimes is measured with this metric. In practice, however, despite not knowing the true optimal treatment regimes, value functions aim to compare the difference between the estimated optimal treatment regimes and the observed treatment regimens.

First, we consider the mean optimal treatment rate. The estimated recommended treatment configuration for a pair could be that both, one or neither of the treatments are the same as the true optimal treatments; therefore, there are two different quantities: one is the mean optimal treatment rate for the household, and the other is the mean optimal treatment rate for the individual. The household MOTR represents the average decision accuracy for the pair in the same household. Formally, if the true treatment decision for h^{th} household is (a_h^{s*}, a_h^{r*}), the optimal treatment rate for the household can be expressed as $H^{-1}\sum_{h=1}^{H}\mathbb{I}\left(a_h^{s*}=\hat{a}_h^{s}, a_h^{r*}=\hat{a}_h^{r}\right)$. In this case, decision accuracy requires estimated pairs' treatment configurations to be consistent with the true optimal treatment configuration, i.e., both individuals' treatments are in accordance with the true treatments. Alternatively, the individual MOTR describes the average decision accuracy for individuals, and can be denoted as $(2H)^{-1}\sum_{h=1}^{H}\left[\mathbb{I}\left(a_h^{s*}=\hat{a}_h^{s}\right)+\mathbb{I}\left(a_h^{r*}=\hat{a}_h^{r}\right)\right]$. In both cases, the higher MOTR value indicates greater decision accuracy and thus the superiority of the corresponding method.

Second, in a single-stage setting, to compare the effects of the optimal treatment configuration with those of the estimated one, we consider the value of the regret function when the estimated treatment configuration is implemented. According to the definition of the regret function which is defined in Section 5.4.1, with the estimated treatment configuration input, the regret value is as follows:

$$\begin{aligned}\mu[d^*,\hat{d}^*;\boldsymbol{\xi},\boldsymbol{\psi},\boldsymbol{\phi}] &= \gamma^*[d^*(\boldsymbol{x}^{\xi},\boldsymbol{x}^{\psi},\boldsymbol{x}^{\phi});\boldsymbol{\xi},\boldsymbol{\psi},\boldsymbol{\phi}]-\gamma^*[\hat{d}^*(\boldsymbol{x}^{\xi},\boldsymbol{x}^{\psi},\boldsymbol{x}^{\phi});\boldsymbol{\xi},\boldsymbol{\psi},\boldsymbol{\phi}]\\ &= (A^{s*}-\hat{A}^{s*})\boldsymbol{\xi}^{\top}\boldsymbol{x}^{\xi}+(A^{r*}-\hat{A}^{r*})\boldsymbol{\psi}^{\top}\boldsymbol{x}^{\psi}+(A^{s*}A^{r*}-\hat{A}^{s*}\hat{A}^{r*})\boldsymbol{\phi}^{\top}\boldsymbol{x}^{\phi}.\end{aligned}$$

The regret value measures the difference between the value under the optimal regime (i.e.,

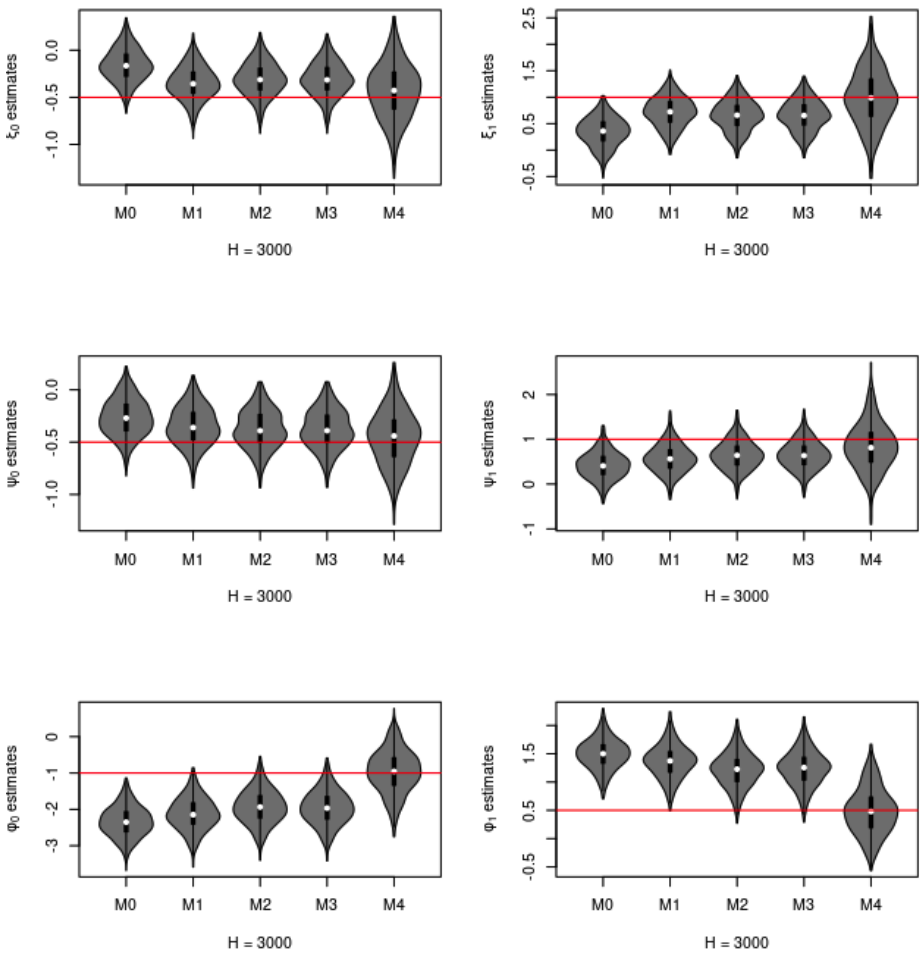


Figure 5.2: Blip function parameter estimates, $\hat{\boldsymbol{\xi}}$ (top row), $\hat{\boldsymbol{\psi}}$ (middle row), and $\hat{\boldsymbol{\phi}}$ (bottom row) via Method 0 (M0, Q-learning), Method 1 (M1, no treatment-association dWPOM), Method 2 (M2, treatment-association aware dWPOM with IPW-type weights), Method 3 (M3, treatment-association aware dWPOM with overlap-type weights) and Method 4 (M4, treatment-association aware dWPOM with adjusted overlap-type weights), when the treatment model is correctly specified but the treatment-free model is misspecified (Scenario 3).

Table 5.3: Methods' performance measure estimates and their standard errors (in parenthesis) from Methods 0, 1, 2, 3, and 4, when the treatment model is correctly specified but the treatment-free model is misspecified (Scenario 3) in Study 1. H denotes the number of households.

H	Performance	Method				
		M0	M1	M2	M3	M4
500	Household MOTR	0.383 (0.18)	0.438 (0.18)	0.446 (0.18)	0.445 (0.18)	0.421 (0.17)
	Individual MOTR	0.584 (0.12)	0.628 (0.11)	0.635 (0.11)	0.634 (0.11)	0.617 (0.10)
	Regret values	0.188 (0.09)	0.164 (0.09)	0.165 (0.10)	0.165 (0.10)	0.174 (0.14)
1000	Household MOTR	0.432 (0.17)	0.517 (0.17)	0.520 (0.17)	0.520 (0.17)	0.508 (0.18)
	Individual MOTR	0.620 (0.11)	0.683 (0.10)	0.688 (0.10)	0.688 (0.10)	0.679 (0.10)
	Regret values	0.162 (0.10)	0.126 (0.10)	0.127 (0.10)	0.126 (0.10)	0.138 (0.12)
3000	Household MOTR	0.475 (0.14)	0.591 (0.13)	0.608 (0.14)	0.609 (0.14)	0.610 (0.16)
	Individual MOTR	0.654 (0.08)	0.738 (0.07)	0.747 (0.07)	0.747 (0.07)	0.753 (0.07)
	Regret values	0.142 (0.06)	0.094 (0.05)	0.086 (0.05)	0.087 (0.05)	0.095 (0.07)
5000	Household MOTR	0.517 (0.13)	0.634 (0.11)	0.654 (0.11)	0.652 (0.10)	0.684 (0.12)
	Individual MOTR	0.680(0.07)	0.759 (0.06)	0.774 (0.06)	0.772 (0.05)	0.802 (0.06)
	Regret values	0.126 (0.05)	0.081 (0.03)	0.073 (0.03)	0.072 (0.03)	0.062 (0.04)
10000	Household MOTR	0.555 (0.13)	0.667 (0.09)	0.691 (0.09)	0.689 (0.09)	0.746 (0.11)
	Individual MOTR	0.708 (0.07)	0.780 (0.05)	0.799 (0.05)	0.797 (0.05)	0.837 (0.06)
	Regret values	0.111 (0.04)	0.072 (0.03)	0.062 (0.03)	0.064 (0.03)	0.038 (0.03)

$d^*(\boldsymbol{x}^{\xi}, \boldsymbol{x}^{\psi}, \boldsymbol{x}^{\phi}) = (A^{s*}, A^{r*})$) and that under the estimated regime (i.e., $\hat{d}^*(\boldsymbol{x}^{\xi}, \boldsymbol{x}^{\psi}, \boldsymbol{x}^{\phi}) = (\hat{A}^{s*}, \hat{A}^{r*})$). A smaller regret value indicates that the estimated regime is closer to the optimal regime; therefore, the smaller the regret value corresponds to the better method.

Setting different numbers of household H as 500, 1000, 3000, 5000, and 10000 in Scenario 3, Table 5.3 presents above mentioned three performance measures for all methods: (1) household MOTR, (2) individual MOTR, and (3) the regret value. As expected, in all the cases, compared with the Q-learning (M0), dWPOM (M1 — M4) methods provide higher both household and individual MOTR, and less regret value. From Table 5.3, in the larger household sample cases, that is, $H \geq 3000$, compared with either M0 (Q-learning) or M1, M2, and M3 (dWPOM with different types of weights), M4 which is dWPOM with adjusted weights provides the highest in both household and individual MOTRs, and the lowest regret values. As well, M0 which does not use any balancing weights outputs the lowest MOTRs and the highest regret values. These results verify that the estimated treatment configuration from M4 is the closest to the optimal treatment configuration. Thus, in these large household sample cases, M4 outperforms among all these methods, and M0 performs the worst. In the smaller household sample cases, $H < 3000$, M0 still

provides the worst MOTRs and regret values. M4 does not always provide the highest MOTRs and the lowest regret values; however, in these cases, either M2 or M3, which are treatment-association aware dWPOM, offer the best MOTR or regret values, compared with M1 which assumes independent treatment. Note that M2, M3, and M4 all belong to treatment-association aware dWPOM; therefore, these results indicate that treatment-association aware dWPOM performs better than no treatment-association dWPOM if an association exists between treatments.

To conclude, in Study 1, we investigated a single-stage DTR estimation problem by simulations. The double robust property of the adjusted weights proposed in Theorem 5.1 was verified in the case of the simulation settings. In addition, we also demonstrated that the estimation will be biased if there is a treatment association that is not taken into account.

5.5.2 Multiple-stage Treatment Decision for a Couples Case

For multi-stage settings with ordinal outcomes, the technique of dWPOM is explained in Section 5.4. We now conduct a simulation study (Study 2) to illustrate our estimation of a two-stage treatment decision problem with ordinal outcomes under household interference. The causal diagram for two-stage treatment decisions with household ordinal outcomes in the presence of interference is presented in Figure 5.3, and the simulation's data-generating process is according to this causal diagram. Again, the red path between A^s and A^r in DAG 5.3 indicates that the individuals who are interfering with each other in the same household receive the correlated treatments.

To generate data, we denote covariates as x_{jpt}, where $j = 1, 2$ indicates the j^{th} stage, $p = 1, 2$ represents the dimension of covariates, and $t = s, r$ corresponds to individual s or r in the same household. In Stage 1, the covariates of s and r are $x_{11s} \sim N(0, 1)$, $x_{11r} \sim N(0, 1)$, $x_{12s} \sim Ber(0.5)$, $x_{12r} \sim Ber(0.5)$, that is, x_{11} of s and r are normally distributed with mean 0 and variance 1, and x_{12} of s and r are Bernoulli distributed with success probability 0.5 in Stage 1. Similarly, in Stage 2, the covariates of s and r are $x_{21s} \sim 0.5 * N(x_{11s}, 1)$, $x_{21r} \sim 0.5 * N(x_{11r}, 1)$, $x_{22s} \sim Ber(0.1 + 0.5 * x_{12s})$, $x_{22r} \sim Ber(0.1 + 0.5 * x_{12r})$. To generate correlated treatments of a pair, for the stage $j = 1, 2$, the marginal propensity model for each individual is set as $\mathbb{P}(A_j = 1 \mid \boldsymbol{x}_j) = \text{expit}(-1 + 1.15 * \exp(x_{j1}) - 0.5 * x_{j2})$, and the odds ratio model is set as $\tau = \exp[-0.15 + 0.25 * (x_{j2s} + x_{j2r})]$. The true treatment-free function is set as $f(\boldsymbol{x}^\beta) = \sum_{j=1}^{2} f_j(\boldsymbol{x}_j^\beta) = \cos[\pi * (x_{11s} + x_{11r})] + 0.5 * \exp(x_{21s} + x_{21r}) + 0.2 * (x_{12s} + x_{12r})^3$, and for $j = 1, 2$, the true blip functions $(\gamma_j[(A_j^s, A_j^r), \boldsymbol{x}_j; \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}])$ are set as:

$$A_j^s * (-0.25 + 0.5 * x_{j1s}) + A_j^r * (-0.25 + 0.5 * x_{j1r}) + A_j^s * A_j^r * [-0.5 + 0.25 * (x_{j2s} + x_{j2r})],$$

where true blip parameters are $\boldsymbol{\xi} = (-0.25, 0.5)^\top$, $\boldsymbol{\psi} = (-0.25, 0.5)^\top$, and $\boldsymbol{\phi} = (-0.5, 0.25)^\top$.

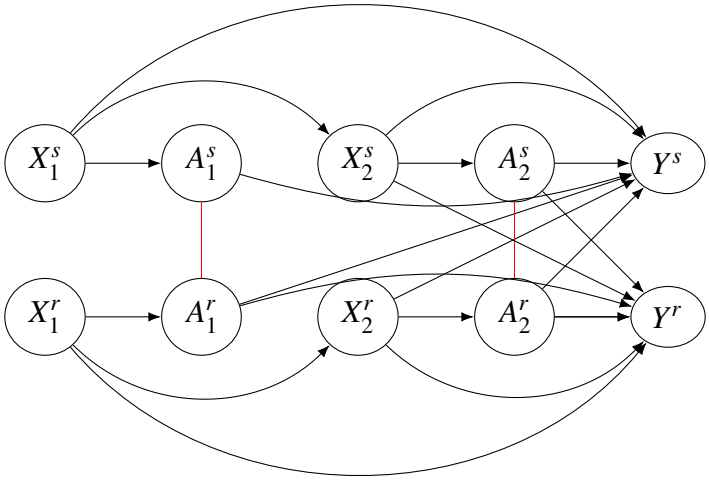


Figure 5.3: The DAG of interference analysis: dWPOM for household ordinal outcomes two-stage decision problems.

To generate household-level ordinal outcomes, following same outcome-generating procedure in Study 1, we set $\zeta_1 = 0.405, \zeta_2 = 1.735$. Building on individual and household covariates and treatments, we calculate household-level thresholds such that $(\zeta_1 - \mu, \zeta_2 - \mu)$, where $\mu = f(\boldsymbol{x}^\beta) + \sum_{j=1}^{2} \gamma_j[(A_j^s, A_j^r), \boldsymbol{x}_j; \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}]$. Then, we can compute cumulative probabilities and thus acquire the category probabilities.

In this simulation study, examining two-person household interference, we focus only on interference-aware Q-learning and our proposed dWPOM with adjusted weights. As previously stated, our value functions for ordinal outcomes are evaluated by computing the odds ratio. We now define value functions for ordinal outcomes by first introducing the odds of a particular outcome. From the proposed POM, we can predict ordinal outcomes when the estimated optimal treatments have been implemented. Building on these predicted ordinal outcomes, we can compute the odds: the probability that the preferred outcome will occur is divided by the probability that the preferred outcome will not occur. That is, we calculate the ratio of the number of the preferred outcome to the number of the outcome that is not preferred. For instance, in our three ordinal outcome case ($U = 1, 2, 3$), the preferred outcome is 3, and the odds of $U = 3$ being among the predicted ordinal outcomes is the ratio of the number of $U = 3$ to the number of $U = 1$ and $U = 2$. Now, we have odds of $U = 3$ among the predicted ordinal outcomes, estimated by implementing the estimated DTRs. Then, from among the observed ordinal outcomes, we can compute the other odds of $U = 3$, which correspond to implementing the observed DTRs. Finally, we can define the

value functions for ordinal outcomes: the ratio of the odds of $U = 3$ among the predicted ordinal outcomes to the odds of $U = 3$ among the observed ordinal outcomes. Similarly, if our preferred outcome is "$U = 2$ or $U = 3$", meaning that at least one individual quits smoking, we can also compute the corresponding value functions.

In addition, we highlight the approximate consistent estimation of the proposed dWPOM method if at least either the treatment-free or treatment model is correct, by considering various types of model misspecification at the two stages. As stated previously, the treatment-free model is harder to correctly specify; thus, in this study, we specifically consider two cases: (1) in both Stages 2 and 1, the treatment-free models are misspecified, but the treatment models are correctly specified; (2) in Stage 2, the treatment-free model is misspecified, but the treatment model is correctly specified, while in Stage 1, alternatively the treatment model is misspecified, but the treatment-free model is correctly specified. Note that the true treatment-free models and treatment models contain non-linear covariate terms, and if we consider only the linear covariate terms in a model, we will mis-specify that model.

For Case (1), the distributions of the blip estimates (i.e., $\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\phi}}$) are presented in Figures 5.4 and 5.5, which correspond to Stage 1 and Stage 2, respectively. The distributions of the blip estimates from Case (2) are presented in Appendix H.4 Figures H.5 and H.6. In Figure 5.5, which depicts blip estimates from Case (1) in Stage 2, all the blip estimates from our dWPOM appear to be symmetrically distributed and centred at the true blip parameter values, but Q-learning provides biased estimators. From Figure 5.4, which corresponds to Case (1) in Stage 1, for our dWPOM, blip estimates $\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\phi}}$ are also symmetrically distributed and centred at the true parameters' values, but the blip estimates $\hat{\phi}_0, \hat{\phi}_1$ appear to be slightly off the true values. We suspect that this misalignment results from the fact that the estimation is only approximately consistent, because of the omission of remainder terms in the Taylor expansion (see proof of Theorem 5.1 in Appendix H.2).

Moreover, the value function for ordinal outcomes, which is the odds ratio of outcome $U = 3$ from interference-aware Q-learning in Case (1), is 1.25, but that from interference-aware dWPOM is 1.36, which is greater than Q-learning's. This result indicates that the proposed interference-aware dWPOM performs better than interference-aware Q-learning.

5.6 Real Data Analysis from Population Assessment of Tobacco and Health Study

In accordance with real-world data discussed in previous chapters, investigating household ordinal outcomes we now apply our approach, dWPOM, to the data in the Population Assessment of

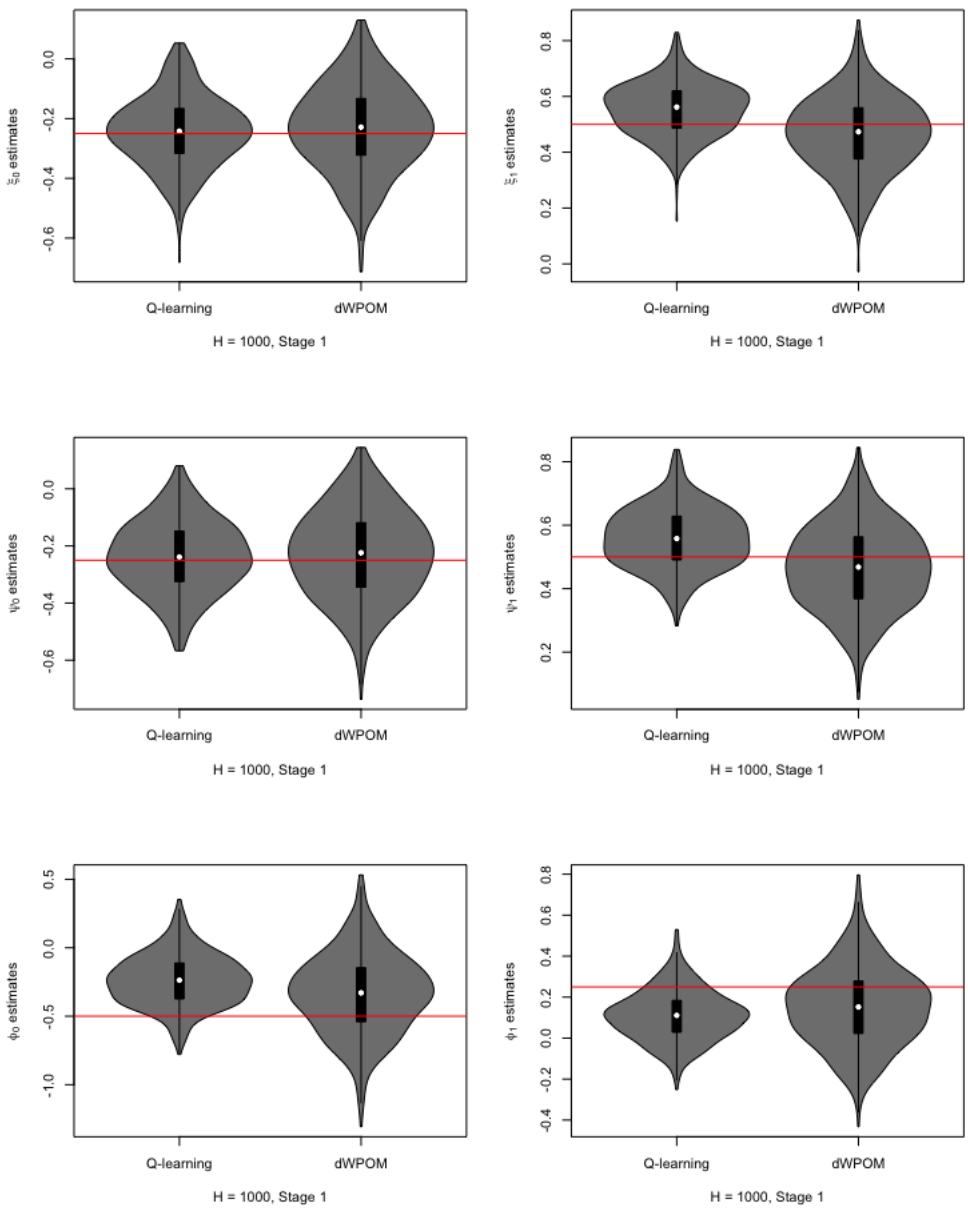


Figure 5.4: Blip function parameter estimates in Stage 1 of Study 2, $\hat{\boldsymbol{\xi}}$ (top row), $\hat{\boldsymbol{\psi}}$ (middle row), and $\hat{\boldsymbol{\phi}}$ (bottom row) via Q-learning and treatment-association aware dWPOM with adjusted overlap-type weights in Case (1), where the treatment-free models are misspecified, but the treatment models are correctly specified in both Stages 2 and 1.

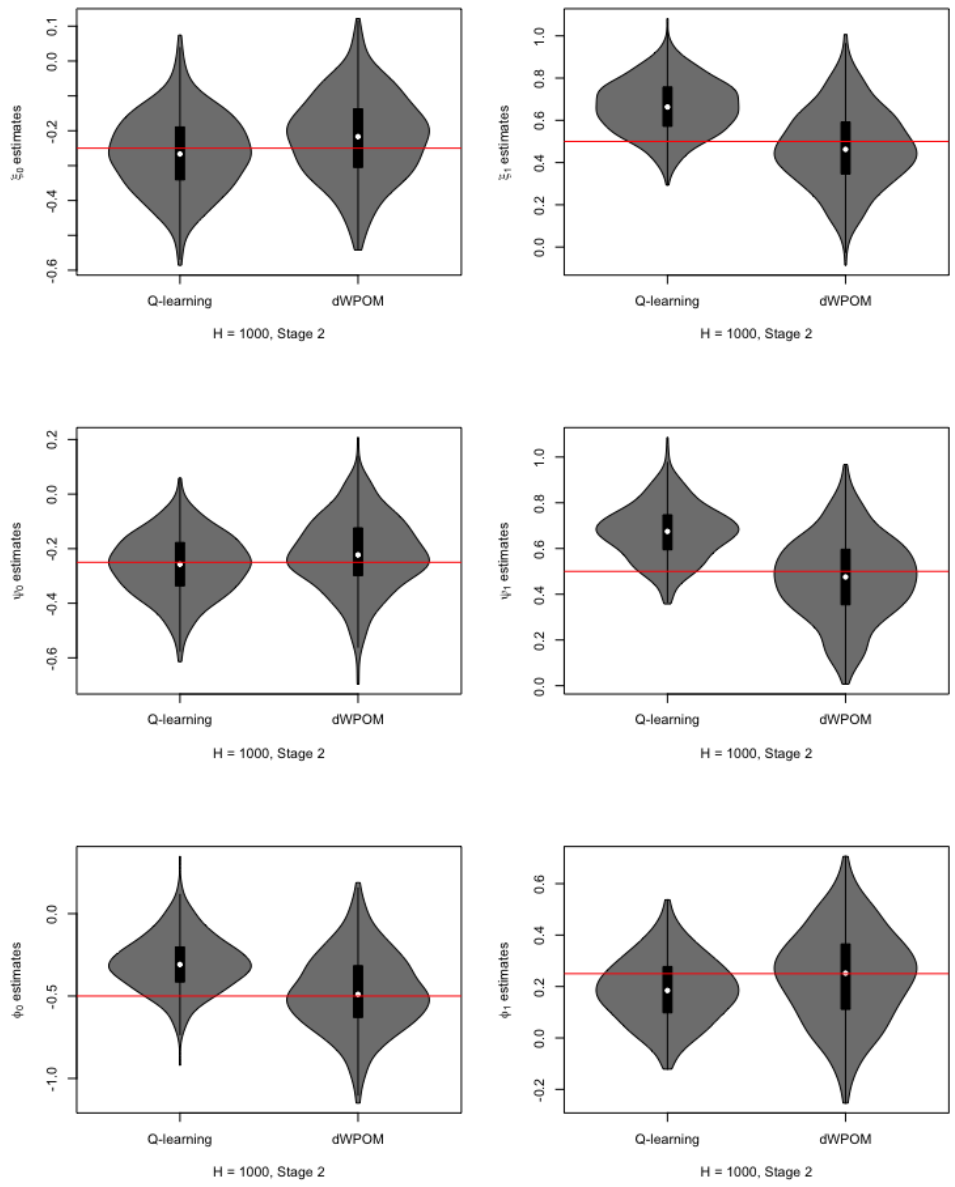


Figure 5.5: Blip function parameter estimates in Stage 2 of Study 2, $\hat{\boldsymbol{\xi}}$ (top row), $\hat{\boldsymbol{\psi}}$ (middle row), and $\hat{\boldsymbol{\phi}}$ (bottom row) via Q-learning and treatment-association aware dWPOM with adjusted overlap-type weights in Case (1), where the treatment-free models are misspecified, but the treatment models are correctly specified in both Stages 2 and 1.

Tobacco and Health study. As in both Chapters 3 and 4, in this chapter, it is our aim to estimate the optimal DTR for a pair in the same household, based on a sequence of rules of e-cigarette use or non-use, for achieving the smoking cessation of the pair in the household. We maintain the basic setup introduced in Chapter 3 and consider the subset of participants who smoke in Wave 1. Similarly, our analysis makes use of the first four waves of data with a three-stage decision problem, defining the j^{th} stage, for $j = 1, 2, 3$, as the time from Wave j to but not including Wave $j + 1$.

In this analysis, again, the treatment variable is the use of e-cigarettes by cigarette smokers. Because the participants were followed for approximately one year, we define e-cigarette use reported at the wave of the measured outcome as indicative of the pre-wave treatment. Consistent with the e-cigarette usage variable in PATH analysis of Chapters 3 and 4, the e-cigarette usage variable is determined by the question "now use e-cigarettes (a) Every day (b) Some days (c) Not at all." Answers of either "Every day" or "Some days" imply $A = 1$, and answers of "Not at all" indicate $A = 0$.

Based on our analysis purpose, the binary outcome variable is an indicator for whether participants have given up smoking (traditional cigarettes) or have tried to quit smoking or using tobacco product(s). The binary outcome variable of each participant is decided by two questionnaire questions in the study: one is based on the question "Do you now smoke cigarettes (a) Every day (b) Some days (c) Not at all ?", and the other is "In the past 12 months have you tried to quit smoking or using tobacco product(s)? (1) Yes (2) No ". For the former question, if participants respond (c), then their first preliminary binary outcome is coded as $Y_1 = 1$; if they respond (a) or (b), then their first preliminary binary outcome is coded as $Y_1 = 0$. For the latter question, if participants respond (1), then their second preliminary binary outcome is coded as $Y_2 = 1$; if they respond (2), then their first preliminary binary outcome is coded as $Y_2 = 0$. Participants' final binary outcome is coded as $Y = 1$ as long as either their first outcome is 1, i.e., $Y_1 = 1$, or their second outcome is 1, i.e., $Y_2 = 1$. Accordingly, the household utility is the sum of the final binary outcomes of a pair in the same household; then adding 1 to the sum, the corresponding household ordinal outcome, which could be either 1, 2, or 3, can be constructed.

As for the household covariates choice in our POM, similarly to previous PATH analyses (e.g., Section 3.5), for the j^{th} stage, we first select the individual-level Wave j variables: age ("less than 35" or "35+"), education, non-Hispanic, race and "plan to quit". Then, building on these individual-level covariates, we can construct household or joint covariates for our household POM. For instance, the individual-level age variable is an indicator of "less than 35"; for the household-level age variable, we thus have three possibilities for a pair in the same household: both, one of them, or neither of them is less than 35. Therefore, we thus construct the single age variable with three categories for the household model. Similarly, with three possible values for each variable, we construct the non-Hispanic, race, and "plan to quit" variables at the

household level. In addition, the individual-level education covariates are set as variables with three categories : 1 = "less than high school", 2 = "between high school and bachelor degrees", then 3 = "bachelor or advanced degree". Building on individual-level education covariates, we can construct household-level education covariates as three category variables, that is, both, one of them, or neither of them hold a bachelor's or advanced degree.

Again, because the questionnaire will no longer ask the question regarding "plan to quit" to participants who have already quit smoking, we assign the value of that question to 1, that is, they have a plan to quit smoking, for those participants. In addition, the participants who have quit smoking also will not be asked the quit attempt question in the following wave. Thus, the second binary outcome Y_2 that is mentioned above is also affected. To address this problem, we assume that those who have quit smoking in the previous wave made a quit attempt between these two waves, but we called them "quit maintenance", rather than calling them "quit attempt". For example, someone who has quit still uses e-cigarettes to maintain smoking quitting. By assuming this, on one hand, we follow the same rules as the "plan to quit" question. On the other hand, we construct a "novel" second binary outcome Y_2 that encompasses the outcomes for the two participant situations: participants made a quit attempt if they need to quit, and they have maintained the quit if they need to maintain quitting.

Therefore, for the household covariates, we have denoted, age, education, non-Hispanic, race, and "plan to quit", as the covariates $\boldsymbol{x}_j^{\beta} = (x_{j1}, x_{j2}, x_3, x_4, x_{j5})^{\top}$ in the Stage j treatment-free model, respectively. We note that, compared with the previous chapters' PATH analysis, we omit the sex variable in the household covariates, because it has not been very significant when we focus on the household-level model. Individuals s and r are respectively the first and second listed members of the household pair in the data set. In addition, building on previous work of studying moderators in the relationships of prior wave predictors of quitting smoking, we select at each stage the variables age and "plan to quit" as tailoring variables, that is, $\boldsymbol{x}_j^{\xi} = (1, x_{j1}^s)^{\top}$, $\boldsymbol{x}_j^{\psi} = (1, x_{j1}^r)^{\top}$, and $\boldsymbol{x}_j^{\phi} = (1, x_{j5}^s + x_{j5}^r)^{\top}$. Therefore, in estimation, the blip model is set up as $\gamma[(a_{j+1}^s, a_{j+1}^r), \boldsymbol{x}_j; \boldsymbol{\xi}_j, \boldsymbol{\psi}_j, \boldsymbol{\phi}_j] = a_{j+1}^s \boldsymbol{\xi}_j^{\top} \boldsymbol{x}_j^{\xi} + a_{j+1}^r \boldsymbol{\psi}_j^{\top} \boldsymbol{x}_j^{\psi} + a_{j+1}^s a_{j+1}^r \boldsymbol{\phi}_j^{\top} \boldsymbol{x}_j^{\phi}$, and the treatment-free model as $f(\boldsymbol{x}_j^{\beta}; \boldsymbol{\beta}_j) = \boldsymbol{\beta}_j^{\top} \boldsymbol{x}_j^{\beta}$. Accordingly, solving the sequential decision problem by backward induction, for the Stage $j = 3, 2, 1$ and $c = 1, 2$, we have the POM

$$logit[\mathbb{P}(\widetilde{\mathcal{U}}_j^r \leq c \mid a_j^s, a_j^r, \boldsymbol{x}_j; \boldsymbol{\xi}_j, \boldsymbol{\psi}_j, \boldsymbol{\phi}_j)] = \zeta_{cj} - \boldsymbol{\beta}_j^{\top} \boldsymbol{x}_j^{\beta} - a_j^s \boldsymbol{\xi}_j^{\top} \boldsymbol{x}_j^{\xi} - a_j^r \boldsymbol{\psi}_j^{\top} \boldsymbol{x}_j^{\psi} - a_j^s a_j^r \boldsymbol{\phi}_j^{\top} \boldsymbol{x}_j^{\phi}. \tag{5.15}$$

To construct the balancing weights for the proposed POM, as introduced in Section 5.2, we primarily estimate the marginal propensity scores, the pairwise odds ratios (τ_{sr}), and the joint propensity scores ($\pi^{11}(\boldsymbol{x}_s, \boldsymbol{x}_r) = \mathbb{P}(A^s = 1, A^r = 1 \mid \boldsymbol{x}_s, \boldsymbol{x}_r)$). We first choose covariates at the

individual level for the marginal treatment propensity models, based on the previous PATH studies of Benmarhnia et al. [2018] and Jiang et al. [2022b], as $\boldsymbol{x}_j^{\alpha} = (x_{j1}, x_{j2}, x_3, x_4, x_{5j}, x_6)^{\top}$, namely, age, education, non-Hispanic, race, "plan to quit", and sex. Then, we employ logistic regression to acquire marginal treatment propensity scores. The pairwise odds ratios are modeled through a generalized linear model with the log link and covariates $\boldsymbol{x}_{sr} = [1, (x_{j5}^{s} + x_{j5}^{r})]^{\top}$, which represents the number of individuals in the same household who have a plan to quit. That is, with parameter $\boldsymbol{o}$, $log\tau_{sr}(\boldsymbol{o}) = \boldsymbol{o}^{\top}\boldsymbol{x}_{sr}$.

Following the methods that were introduced in Section 5.2, we can further estimate the joint propensity score, and the corresponding weights. In particular, we compare four different weights, which are (I) no balancing weights, (II) no-association overlap weights, where the joint propensity functions are equal to the product of marginal propensities, (III) association-aware overlap weights (5.3), and (IV) adjusted association-aware overlap weights (5.14). These four different weights correspond to M0, M1, M3, and M4 in simulation Study 1, respectively, and we omit M2, which is dWPOM with inverse probability type weights and similar to M3. In this PATH analysis, we call them Methods I (Q-learning), II, III, and IV. It is important to note that Method IV employs the adjusted balancing weights, and is our desired treatment-association aware dWPOM.

Building on weighted population-level estimating equations, our adjusted association-aware overlap weights (Method IV), in theory, guarantee the consistent estimators of the blip parameters. Consistent with the previous PATH analysis, in this PATH households analysis, we also utilise the sampling design weights in each stage. Because the sampling design weights of households are not available, and our purposes are illustrative, we have used the average of the two weights of the individuals in the same household as the household weights. The rationale behind this is that the individual level weights are modifications of the household level weights obtained by multiplying by a within-household raising factor (often equal to 1) followed by adjustment for response propensity and calibration to population benchmarks; if (say) individual s is male and individual r is female, their adjustment factors will be similar.

Table 5.4 summarizes the blip estimates and their bootstrap standard errors (in parenthesis) from Methods I, II, III, and IV in this PATH analysis. It is important to note that, for both members of a couple to either quit or attempt to quit smoking, the optimal DTRs for the household are functions of blip parameter estimates and the couple's tailoring variables, that is, the decision rules in Decision 1. Method IV, which employs the adjusted balancing weights, is expected to provide consistent estimation of these blip parameters. Thus, we particularly focus on the results from Method IV, while accounting for those from other methods.

Because the household case with four treatment configurations is more complicated than in the previous individual-level analysis, based on Rule 1, we give several examples of how the results may be interpreted. For Method IV, in Stage 3 (Wave 3 $\sim$ 4), for example, the blip estimate is

Table 5.4: Blip estimates and their bootstrap standard errors (in parenthesis) from the analysis of PATH data. Optimal DTRs are functions of blip parameter estimates based on decision rules in Decision 1. Est. stands for the blip parameters' estimates.

Wave	Est.	Methods			
		I (Q-learning)	II	III	IV
$1 \sim 2$	$\hat{\xi}_0$	-0.100 (0.037)	-0.124 (0.044)	0.218 (0.072)	0.107 (0.047)
	$\hat{\xi}_1$	0.180 (0.043)	0.130 (0.055)	-0.188 (0.052)	-0.150 (0.068)
	$\hat{\psi}_0$	-0.167 (0.040)	-0.321 (0.049)	-0.140 (0.047)	-0.070 (0.050)
	$\hat{\psi}_1$	0.251 (0.044)	0.257 (0.058)	0.017 (0.052)	-0.273 (0.059)
	$\hat{\phi}_0$	0.102 (0.079)	0.376 (0.078)	0.009 (0.081)	-0.001 (0.099)
	$\hat{\phi}_1$	0.001 (0.047)	0.106 (0.048)	0.087 (0.050)	0.045 (0.058)
$2 \sim 3$	$\hat{\xi}_0$	0.341 (0.036)	0.116 (0.046)	0.081 (0.045)	-0.031 (0.045)
	$\hat{\xi}_1$	-0.276 (0.043)	0.188 (0.062)	-0.044 (0.056)	0.200 (0.058)
	$\hat{\psi}_0$	-0.078 (0.037)	0.052 (0.052)	0.205 (0.048)	0.004 (0.048)
	$\hat{\psi}_1$	0.068 (0.045)	0.101 (0.067)	0.054 (0.058)	0.064 (0.062)
	$\hat{\phi}_0$	0.377 (0.097)	0.663 (0.110)	0.360 (0.108)	0.358 (0.126)
	$\hat{\phi}_1$	-0.317 (0.060)	-0.568 (0.065)	-0.414 (0.067)	-0.507 (0.076)
$3 \sim 4$	$\hat{\xi}_0$	0.966 (0.040)	1.067 (0.041)	1.233 (0.040)	0.785 (0.047)
	$\hat{\xi}_1$	-0.419 (0.055)	-0.268 (0.055)	-0.391 (0.054)	0.304 (0.060)
	$\hat{\psi}_0$	0.808 (0.038)	1.527 (0.047)	1.217 (0.043)	0.690 (0.044)
	$\hat{\psi}_1$	0.448 (0.044)	-0.507 (0.054)	0.138 (0.049)	0.987 (0.052)
	$\hat{\phi}_0$	-1.612 (0.137)	-0.150 (0.150)	-1.479 (0.147)	-1.151 (0.228)
	$\hat{\phi}_1$	0.331 (0.069)	0.085 (0.076)	0.004 (0.076)	-0.091 (0.115)

$A^s(0.785+0.304*age_s)+A^r(0.690+0.987*age_r)+A^s*A^r(-1.151-0.091*PQ)$, where PQ represents the plans of quitting for a couple in the same household, and age_s and age_r are ages of s and r. When we plug in four possibilities of $(A^s, A^r) = (1,1), (1,0), (0,1), (0,0)$, the blip estimates are $0.785+0.304*age_s+0.690+0.987*age_r-1.151-0.091*PQ$, $0.785+0.304*age_s$, $0.690+0.987*age_r$, and 0, respectively.

Example 1: If we have household tailoring variables such that $age_s = 1$ $age_r = 0$, and $PQ = 0$, the blip estimates are $0.785+0.304+0.690-1.151 = 0.628$, $0.785+0.304 = 1.089$, 0.690, and 0, respectively. The largest blip estimate is 1.089, and corresponds to the treatment configuration $(A^s = 1, A^r = 0)$. Therefore, in Stage 3 (Wave $3 \sim 4$), if individual s is less than 35 but r is not, and both of them have no plan to quit, then the treatment recommendation for this household should be $(A^s = 1, A^r = 0)$.

Example 2: If we have household tailoring variables such that $age_s = 0$, $age_r = 1$, and $PQ = 2$ (individual s is over 35 but r is not, and both of them have plans to quit), the blip estimates are $0.785+0.690+0.987-1.151-0.091*2 = 1.192$, 0.785, $0.690+0.997 = 1.687$, and 0, respectively; then the treatment recommendation should be $(A^s = 0, A^r = 1)$.

Example 3: If we have household tailoring variables such that $age_s = 0$, $age_r = 0$, and $PQ = 2$, the blip estimates are $0.785+0.690-1.151-0.091*2 = 0.142$, 0.785, 0.690, and 0; then the treatment recommendation should be $(A^s = 1, A^r = 0)$.

5.7 Conclusion and Discussion

In this chapter, considering household interference and household utility, we proposed a double robust DTR estimation method for ordinal outcomes to consistently estimate optimal DTRs. This method, namely, dWPOM, uses sequential weighted POM with adjusted balancing weights. Through simulation studies, where we investigated different forms of balancing weights, we demonstrated the double robustness property of our dWPOM approach, which utilizes the proposed adjusted balancing weights. In the presence of household interference, our dWPOM addresses household ordinal utility problems in this chapter and provides optimal treatment recommendations for both individuals in the same household. We emphasize that our dWPOM can also be employed in the standard DTR estimation case with individual-level ordinal outcomes, and enable the consistent estimation of individual's blip parameters if the adjusted balancing weights are used. To address the ordinal outcomes challenge, we consider a POM, which employs the logit link in a cumulative link mixed model (CLMM, Tutz and Hennevogl [1996]). We note that any POM related tools or techniques, such as those for variable selections or model diagnosis of POMs, can be employed in our dWPOM.

Moreover, we also made a methodological contribution to the interference study. In addition to considering the effects of neighbours' treatments on an individual's outcome, we also considered a possible association between their treatments. Building on this, we presented the estimation process for joint propensity scores in the case where there exists an association between treatments of individuals in the same household, then estimated the corresponding balancing weights that satisfy the balancing criterion. Our simulation studies revealed that if there exists an association between treatments but we fail to consider it, then the DTR estimation will lead to bias. It is straightforward to extend our household interference case to the partial interference one, where treatments of individuals blocked by clusters can affect outcomes of the individuals in the same cluster, while also accounting for the association between these treatments of individuals in the same cluster. However, the association-aware estimation has an extra cost: modeling association between pairs of binary treatments, such as through a pairwise odds ratio model in our household case. For the cluster partial-interference case, to extend our work, we suggest considering the log-linear model to estimate the "higher-order" odds ratio association (Yi and Thompson [2005]). Note that, in these association cases, the final purpose is to estimate the joint propensity scores; therefore, we recommend employing machine learning methods, such as random forest or deep neural network, to directly train the model for the joint propensity scores. In addition, in the presence of household interference, balancing check methods are proposed in Section 5.3. Some projects about balancing checks can be further investigated in the interference case, such as conducting related simulation studies and extending theory to partial or general interference cases.

Chapter 6

Summary and Future Work

6.1 Summary

Precision medicine and interference are the major topics covered in this thesis. Precision medicine leads health professionals to concentrate on individual patients' specific characteristics. The primary focus of precision medicine is on decision support, often in the form of dynamic treatment regimes, which are sequences of decision rules. At each decision point, the decision rules determine the next treatment according to the patient's baseline characteristics and accrued information available up to that point. Implementing DTRs based on patient-specific characteristics provides an avenue for improving the outcomes for all patients being treated or untreated for the same condition. Considering interference, whereby one patient's outcome can be affected by others' treatment, will provide a novel and practical way to investigate treatment assignment, aiming not only at the improvement of individual outcomes but also at greater treatment efficiency in terms of treatment assignment for the whole population. Interference reflects the complicated nature of real-world scientific problems, so it is necessary to develop novel methods to address complex health scenarios.

By mapping individual patient characteristics to recommended treatments, dynamic treatment regimes operationalize clinical decision-making for precision medicine. For the finite time horizon DTR estimation, we have demonstrated the use of a backward induction (dynamic programming) approach to estimating optimal DTRs, a type of sequential decision-making approach which provides treatment decision rules for optimizing all the patient's outcomes. In particular, focusing on regression-based estimation methods, we studied the dWOLS method, which is straightforward to implement and provides interpretability and accessibility to users across a wide range of fields. dWOLS affords a simpler way to estimate optimal DTRs compared with G-estimation and also

guarantees a kind of robustness of the estimation. We have proved the double robustness property of dWOLS in different ways. Built on the theory of dWOLS, this thesis develops the theoretical aspects of a dWOLS framework that accounts for varying types of outcomes and for interference.

Much of this thesis is dedicated to estimating the optimal DTRs in the presence of interference. In Chapter 1, building on the causal inference analytical framework, two crucial topics of the thesis are introduced: DTR estimation and interference estimation. Chapter 2 continues to dive into the regression-based DTR estimation methods in terms of single-stage and multiple-stage decision estimation. In particular, we further investigated the dWOLS theory, which is a "foundation stone" of the whole thesis, paving a solid base for the subsequent studies of DTR estimation with interference. For the investigation of DTR estimation with interference, Chapter 3 is a "pillar" of the whole thesis. It provides our general extension of dWOLS to either a household interference case in which a pair in a family influences each other's outcome, or the network interference case where any possible neighbours can affect an individual's outcome. We have offered a balancing weights criterion to ensure the double robustness of dWOLS in the presence of network interference. To study DTR estimation for binary outcomes with interference, we first proposed an extension of dWOLS to the case of binary outcomes in Chapter 4. Then, in Chapter 5, we have addressed the challenge of estimating DTR for individual binary outcomes in the presence of interference, and have proposed a robust DTR estimation method under household interference, where ordinal outcomes (household utility) are constructed from individuals' binary outcomes. Chapter 4 presents a novel method of DTR estimation for binary outcomes, a dynamic weighted generalized linear model (dWGLM), and the simulation studies to verify the double robustness of dWGLM. In Chapter 5, three contributions are noteworthy: (1) the dWGLM method is extended to optimize individuals' binary outcomes in the presence of interference; (2) the presented method, namely, a dynamic weighted proportional odds model, can also be considered as the generalization of dWOLS to ordinal outcomes; (3) in the presence of household interference, new covariates balancing weights which account for treatment association are proposed. Moreover, as part of the PATH analysis, we also provided corresponding answers to the PATH questions that are posed in the introduction chapter (Section 1.3 in Chapter 1). That is, taking interference into account, we illustrated how to offer treatment recommendations for individuals or households, in terms of a sequence of use or non-use of e-cigarettes, for either reducing cigarette dependence or smoking cessation.

As mentioned in Chapter 1, we provided a brief look at Figure 1.2 to summarize our proposed optimal DTR estimation methods and their extensions that address the interference problem. Here, we present in detail the same figure, labelled Figure 6.1, to see the structure and contributions of the whole thesis. Starting from the left, and by examining the continuous outcomes, we note that dWOLS is a fundamental doubly robust method for estimating optimal DTRs in this thesis. Following the arrow down from dWOLS, we extended it to binary and ordinal outcomes cases

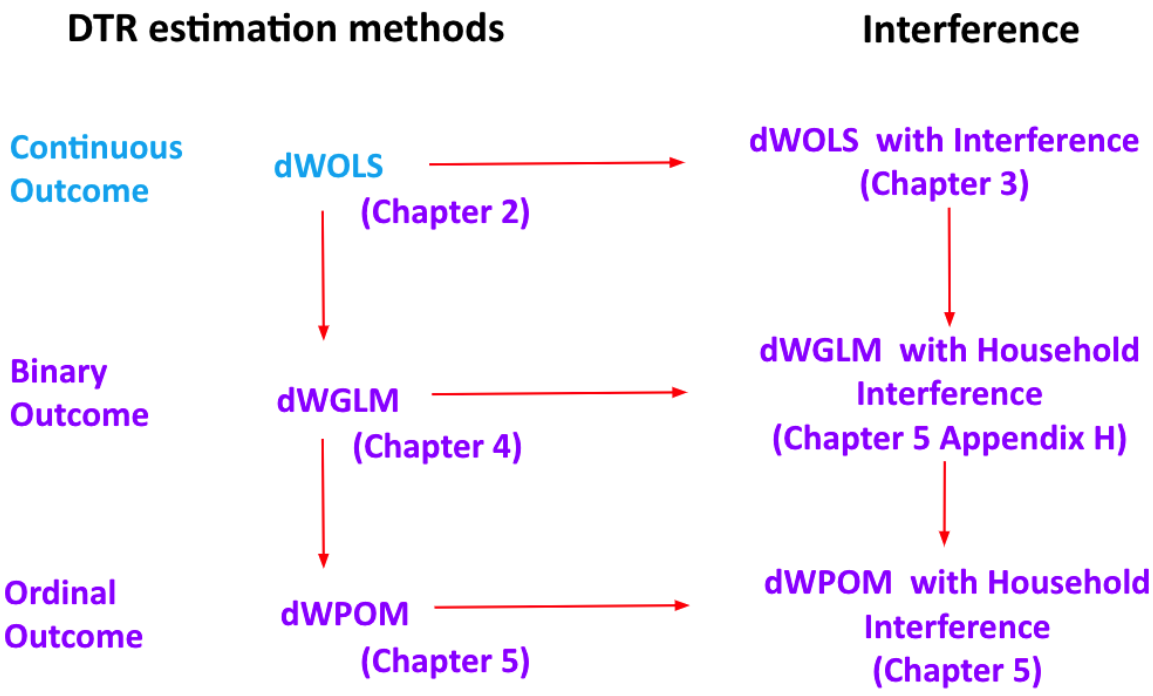


Figure 6.1: Connections among the optimal DTR estimation methods and those methods with interference. (dWOLS in blue is an existing method, and other methods in purple are proposed in this thesis).

and developed dWGLM and dWPOM, which retain the two main advantages of dWOLS: its ease of implementation and its robustness to the misspecification of nuisance models. In Chapter 5, even though dWPOM was presented in the case of household interference, the use of dWPOM in the case of non-interference is also straightforward. From dWOLS to the right, we extended the dWOLS theory to take into account network interference in Chapter 3. In Chapter 5 Appendix H.1, considering the household interference case, we presented covariate balancing theorems of dWOLS and dWGLM to consistently estimate the parameters of interest. These two cases aim to optimize individual outcomes, and the treatment decisions for individuals, while taking into account their neighbour's treatments, are provided. In the main context of Chapter 5, we demonstrated the theory of dWPOM for optimizing the household utility in the presence of household interference. The corresponding treatment decision rules take a couple's covariates as input, and output a treatment configuration for the couple.

From the perspective of DTR estimation, dWOLS, dWGLM, and dWPOM are a series of doubly robust DTR estimation methods, which are also comparatively easy to implement, to deal with continuous, binary, and ordinal outcomes, respectively. For making sequential treatment decisions, they rely on pseudo-outcomes and use backwards induction. However, to conduct a robust estimation of the parameters of interest, dWGLM and dWPOM employ a "two-phase" estimation: the first phase is standard WGLM or WPOM with standard covariates balancing weights to construct an "adjustment" factor, such as in equations (4.2) and (5.12); then, the second phase is a similar WGLM or WPOM but with the new weights, which are standard covariates balancing weights multiplied by the "adjustment" factor. In fact, dWOLS is a special case of dWGLM, where the link function is the identity function, and dWGLM, or precisely, dynamic weighted logistic regression can also be considered a special case of dWPOM, where the ordinal outcomes contain just two categories. All these "two-phase" estimation inspirations are from the investigation of the estimation equation systems of the dWOLS in Appendix C, which was also mentioned as a foundation in Chapter 2.

Furthermore, from the perspective of the interference study, in Chapter 3, assuming the independence between treatments, we provided the general network interference covariate balancing weights, which depend on the joint propensity function. Motivated by our real dataset from the PATH study and considering the household interference case, in Chapter 5, we examined the correlations between treatments of individuals in the same household and provided methods for estimating the weights that are suitable for this case. In addition, to evaluate the performance of these balancing weights, we proposed some empirical measures for assessing covariate balancing in Section 5.3.

Briefly, the core of dWOLS is the weights, which are employed in ordinary least squares, and the key to this thesis can be simply considered as to how to construct a "right" weight for distinct realistic scenarios, such as ordinal outcomes, network interference or both, to robustly

estimate optimal DTRs. In the contribution of regression-based DTR estimation methods, for conducting robust estimation, we provide balancing weights criteria for generalized linear models and proportional odds models, which especially deal with binary and ordinal outcomes, respectively. As previously discussed, these weight criteria are built on the standard dWOLS weight criterion and incorporate an extra "adjustment" factor. In the presence of interference, these weight criteria rely on the joint propensity functions. Therefore, in the contribution of interference studies, we provided an estimation framework for network propensity functions in terms of network interference. In particular, for household interference, we also presented an estimation process for the joint propensity score for a couple whose treatments are possibly correlated with each other.

6.2 Future Work

Some possible future extensions of each project are detailed in each chapter's discussion section. However, as a conclusion to the thesis, we would like to put forward some thoughts about further exploration directions in the area of DTR estimation and interference. These include the possibility of novel DTR estimation methods as well as DTR approaches with interference.

First, in estimating the optimal DTRs, at each stage, the key task is to estimate heterogeneous treatment effects–that is, to estimate the conditional average treatment effect (CATE) function, which has been introduced in Chapter 2. One possible direction is to relax Robinson [1988]'s *partially linear* treatment effect model for robust estimation of CATE, wherein a linear specification of CATE is applied. We can study more flexible modeling of both outcome and treatment models, that is, ML methods that can adaptively discover a good representation for the CATE, for example, Hassanpour and Greiner [2019]'s DR-CFR (SNet-3) deep learning approaches to estimating heterogeneous treatment effects.

It is crucial to note that the ML methods that are usually effective in the prediction do not necessarily perform well in the estimation or inference of causal parameters of interest. The so-called "naive" or "prediction-focused" ML estimators are biased, with the bias having two main sources: regularization and overfitting (Chernozhukov et al. [2018]). However, using orthogonalization can overcome regularization bias, and data splitting can control the overfitting bias. For recent CATE estimation methods, R-learner (Nie and Wager [2021]), based on orthogonalization, can provide a CATE estimator whose error bounds match the best available bounds for the oracle method. In contrast, another popular estimator, X-learner, which is not orthogonalization-type, has been proven efficient when the numbers of individuals in distinct treatment groups are significantly different (Künzel et al. [2019]). On one hand, we can develop X-learner based Q-learning methods for robust DTR estimation. On the other hand, we can also apply the orthogonalization technique in estimating causal effects with interference, and

extend X-learner in the presence of household interference. In addition, within the proposed ML estimator-based approaches, one area to study is the relationship between overfitting (or data splitting) and cross-fitting.

Then, regarding dynamic treatment regimes with interference, we have provided the optimal DTR estimation in Chapter 3. However, the following areas remain to be investigated: (1) how to determine the order in which patients are prescribed treatments; (2) how to balance the needs of the individual against the needs of the population. (3) how to assign treatment if resources are limited. These research directions require proposing the corresponding method to address the problem. One possible way is multi-agent reinforcement learning (MARL, Busoniu et al. [2008]). Many DTR methods have originated from reinforcement learning approaches that address sequential decision problems, such as Q-learning. Recently, MARL has gradually become popular in the multi-agent sequential decision case. In a stochastic environment in which multiple agents are simultaneously learning, multi-agent reinforcement learning solves the problem of how each agent should act optimally.

Therefore, one area is to consider multi-DTR estimation, where treatment decision rules are addressed in a network interference setting. Combining *game theory* knowledge (stochastic games) with multiple DTRs could help achieve equilibrium in *centralized training with decentralized execution* settings (Kraemer and Banerjee [2016]). In this context, equilibrium means roughly that no player wants to unilaterally deviate to an alternative strategy. In multiple DTR problems, one decision from a DTR will typically affect decisions from other DTRs, so the communications between DTRs need to be examined. Two main ways of communicating in common multi-agent reinforcement learning: centralized and decentralized, will be studied. Considering the interference between individuals, the projects can improve DTRs for public health by considering multi-agent sequential decisions, such as COVID-19 vaccination decisions.

Moreover, motivated by PATH data household interference, we have considered a pair (s, r) in the same household, four treatment configurations (a^s, a^r) for $a^s = 0, 1$ and $a^r = 0, 1$, and the corresponding outcome models $\mu_{a^s a^r}(\boldsymbol{x}) := \mathbb{E}[Y^*(a^s, a^r)|\boldsymbol{x}]$. In this project, we may in future extend the X-learner method to account for household interference. To date, we have constructed estimators $\hat{\mu}_{11}(\boldsymbol{x})$, $\hat{\mu}_{10}(\boldsymbol{x})$, $\hat{\mu}_{01}(\boldsymbol{x})$, and $\hat{\mu}_{00}(\boldsymbol{x})$ of the regression functions $\mu_{11}(\boldsymbol{x})$, $\mu_{10}(\boldsymbol{x})$, $\mu_{01}(\boldsymbol{x})$, and $\mu_{00}(\boldsymbol{x})$, respectively, using corresponding $(A^s, A^r) = (1, 1), (1, 0), (0, 1), (1, 1)$ datasets. Second, building on the estimators for four treatment configurations, we have imputed the unobserved potential outcomes for each individual in a certain treatment combination group. Third, relying on estimated unobserved potential outcomes, for each treatment configuration, we have established the estimates of interest and fit the corresponding models for the estimator of interest. Finally, the estimators of interest have been constructed by weighted estimators of each treatment configuration, where the household interference weights proposed in Jiang et al. [2022b] or treatment-association aware weights in Chapter 5 are suggested. X-learner performs well when

the number of units in one treatment group is much larger than another. The extended method is expected to fit the case in PATH, where the numbers of units in the four treatment configurations are quite different. This project contributes to the CATE estimation with interference, and is the basis of my DTR estimation with interference research direction. Appendix I.1 presents our proposed Net-learner algorithm, which summarizes the above ideas and addresses the household interference problem.

Additionally, further investigations could include work following the estimation of DTRs, such as sensitivity analysis for estimated treatment regimes. If treatment regimes are sensitive to small changes in estimated parameters, optimization by point estimation is not sufficient. Therefore, additional studies can be conducted to include conduction sensitivity analysis and the estimation of intervals for optimal treatment regimes. Furthermore, dWOLS and its extended methods with interference in our thesis are predicated on a key assumption; that is, there is no unmeasured confounding. Observational data used to estimate a dynamic treatment regime can introduce bias due to unmeasured confounding. Utilizing instrumental variables or proxy variables (Miao et al. [2018]; Tchetgen et al. [2020]) is a typical way to address unmeasured confounding problems. In addition, sensitivity analyses are also useful for evaluating the impact of unmeasured confounders on the study's conclusions. Some researchers have conducted fundamental work for unmeasured confoundings, such as Rose et al. [2022]'s sensitivity analysis for unmeasured confounding in DTRs. Stensrud and Sarvet [2022] have proposed *superoptimal* regimes, which use natural treatment values as extra inputs of the treatment decision functions. Because superoptimal regimes use this extra observed information (natural treatment values) more than conventional regimes do, the performance of superoptimal regimes has proved superior to that of conventional optimal regimes. More examples can be found in recent work by Cui and Tchetgen [2021a] and Cui and Tchetgen [2021b].

With regard to theoretical results in this thesis, the main contribution is to develop methods for consistently estimating the blip parameters. These doubly robust methods provide the point estimates of the parameters of interest. Nevertheless, theoretical properties will still need to be established in the future, such as identifying (1) appropriate uncertainty estimates and their properties; (2) the properties of the point estimate of the blip parameters: the possible asymptotic frameworks and the corresponding consistency and asymptotic normality results.

Note that variance estimation of blip parameters resulting from dWOLS is provided in the appendix of Wallace and Moodie [2015]. In the presence of interference, derivation of an estimator of the blip parameter variance is the next step of work. In addition, for complete statistical inference, we must derive asymptotic variance for dWGLM and dWPOM estimators, and propose methods for constructing valid hypothesis tests in the presence of a nuisance parameter.

In Appendix D, we identify the optimal dWOLS weights in terms of minimizing the blip

parameter variance. A further important step would be to identify the optimal network weights to minimize blip parameter variance when the egos' degrees vary.

The above-mentioned areas are potential next explorations that may afford more-accurate treatment recommendations in precision medicine. Current methods should also be closely examined and if possible, improved so that they provide greater accuracy in recommending treatment regimes. As accounting for interference has become part of optimal DTR estimation research, we hope that the field of DTR estimation research will gradually expand, and treatment assignment in practice will become more accurate, optimal, and efficient.

References

R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47, 2002.

D. Almirall, I. Nahum-Shani, N. E. Sherwood, and S. A. Murphy. Introduction to smart designs for the development of adaptive interventions: with application to weight loss research. *Translational Behavioral Medicine*, 4(3):260–274, 2014.

W. An and T. J. VanderWeele. Opening the blackbox of treatment interference: Tracing treatment diffusion through network analysis. *Sociological Methods & Research*, 51(1):141–164, 2022.

P. M. Aronow. A general method for detecting interference between units in randomized experiments. *Sociological Methods & Research*, 41(1):3–16, 2012.

P. M. Aronow, C. Samii, et al. Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4):1912–1947, 2017.

B. Arpino, L. D. Benedictis, and A. Mattei. Implementing propensity score matching with network data: the effect of the general agreement on tariffs and trade on bilateral trade. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(3):537–554, 2017.

S. Athey, D. Eckles, and G. W. Imbens. Exact p-values for network interference. *Journal of the American Statistical Association*, 113(521):230–240, 2018.

S. Athey, R. Chetty, and G. Imbens. Combining experimental and observational data to estimate treatment effects on long term outcomes. *arXiv preprint, arXiv:2006.09676*, 2020.

A. Baddeley and R. Turner. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12:1–42, 2005.

S. Baird, J. De Hoop, and B. Özler. Income shocks and adolescent mental health. *Journal of Human Resources*, 48(2):370–403, 2013.

S. G. Baker. Marginal regression for repeated binary data with outcome subject to non-ignorable non-response. *Biometrics*, 51:1042–1052, 1995.

F. Bargagli Stoffi, C. Tortú, and L. Forastiere. Heterogeneous treatment and spillover effects under clustered network interference. *arXiv preprint, arXiv:2008.00707*, 2020.

B. G. Barkley, M. G. Hudgens, J. D. Clemens, M. Ali, and M. E. Emch. Causal inference from observational studies with clustered interference, with application to a cholera vaccine study. *The Annals of Applied Statistics*, 14(3):1432–1448, 2020.

G. Basse, A. Feller, and P. Toulis. Conditional randomization tests of causal effects with interference between units. *arXiv preprint, arXiv:1709.08036*, 2017.

G. W. Basse, A. Feller, and P. Toulis. Randomization tests of causal effects under interference. *Biometrika*, 106(2):487–494, 2019.

A. Belloni, V. Chernozhukov, and C. Hansen. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650, 2014.

T. Benmarhnia, J. P. Pierce, E. Leas, M. M. White, D. R. Strong, M. L. Noble, and D. R. Trinidad. Can e-cigarettes and pharmaceutical aids increase smoking cessation and reduce cigarette consumption? Findings from a nationally representative cohort of American smokers. *American Journal of Epidemiology*, 187(11):2397–2404, 2018.

B. Bollobás and B. Béla. *Random Graphs*. Cambridge University Press, 2001.

J. Bowers, M. M. Fredrickson, and C. Panagopoulos. Reasoning about interference between units: A general framework. *Political Analysis*, 21(1):97–124, 2013.

J. Bowers, M. M. Fredrickson, and P. M. Aronow. Research note: a more powerful test statistic for reasoning about interference between units. *Political Analysis*, 24(3):395–403, 2016.

L. Busoniu, R. Babuška, and B. De Schutter. Multi-agent reinforcement learning: An overview. *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, 38 (2):156–172, 2008.

J. Cai, A. De Janvry, and E. Sadoulet. Social networks and the decision to insure. *American Economic Journal: Applied Economics*, 7(2):81–108, 2015.

V. Carey, S. L. Zeger, and P. Diggle. Modelling multivariate binary data with alternating logistic regressions. *Biometrika*, 80(3):517–526, 1993.

C. Carini, S. M. Menon, and M. Chang. *Clinical and Statistical Considerations in Personalized Medicine*. CRC Press, 2014.

C. Cava, G. Bertoli, A. Colaprico, C. Olsen, G. Bontempi, and I. Castiglioni. Integration of multiple networks and pathways identifies cancer driver genes in pan-cancer analysis. *BMC Genomics*, 19(1):25, 2018.

B. Chakraborty and E. E. Moodie. *Statistical Methods for Dynamic Treatment Regimes*. Springer, 2013.

B. Chakraborty, S. Murphy, and V. Strecher. Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research*, 19(3):317–343, 2010.

B. Chakraborty, E. B. Laber, and Y. Zhao. Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. *Biometrics*, 69(3):714–723, 2013.

B. Chakraborty, P. Ghosh, E. E. Moodie, and A. J. Rush. Estimating optimal shared-parameter dynamic regimens with application to a multistage depression clinical trial. *Biometrics*, 72(3): 865–876, 2016.

D. Cheng and T. Cai. Adaptive combination of randomized and observational data. *arXiv preprint, arXiv:2111.15012*, 2021.

V. Chernozhukov, I. Fernández-Val, and Y. Luo. The sorted effects method: discovering heterogeneous effects beyond their averages. *Econometrica*, 86(6):1911–1938, 2018.

H. Chipman. Bayesian variable selection with related predictors. *Canadian Journal of Statistics*, 24(1):17–36, 1996.

H. A. Chipman, E. I. George, and R. E. McCulloch. Bart: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266–298, 2010.

L. K. Cobb, M. A. McAdams-DeMarco, R. R. Huxley, M. Woodward, S. Koton, J. Coresh, and C. A. Anderson. The association of spousal smoking status with the ability to quit smoking: the atherosclerosis risk in communities study. *American Journal of Epidemiology*, 179(10): 1182–1187, 2014.

D. Collazos, E. García, D. Mejía, D. Ortega, and S. Tobón. Hot spots policing in a high-crime environment: An experimental evaluation in medellin. *Journal of Experimental Criminology*, 17(3):473–506, 2021.

D. R. Cox. *Planning of Experiments.* Wiley, 1958.

F. W. Crawford, O. Morozova, A. L. Buchanan, and D. Spiegelman. Interpretation of the individual effect under treatment spillover. *arXiv preprint, arXiv:1902.01377*, 2019.

Y. Cui and E. T. Tchetgen. On a necessary and sufficient identification condition of optimal treatment regimes with an instrumental variable. *Statistics & Probability Letters*, 178:109180, 2021a.

Y. Cui and E. T. Tchetgen. A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity. *Journal of the American Statistical Association*, 116(533):162–173, 2021b.

A. Curth and M. van der Schaar. Nonparametric estimation of heterogeneous treatment effects: From theory to learning algorithms. In A. Banerjee and K. Fukumizu, editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 1810–1818. PMLR, 13–15 Apr 2021. URL https://proceedings.mlr.press/v130/curth21a.html.

M. Davidian. Methods based on semiparametric theory for analysis in the presence of missing data. *Annual Review of Statistics and Its Application*, 9, 2021.

M. Davidian and E. Laber. Lecture notes “Introduction to Dynamic Treatment Regimes". https://dtrcoursespring2019.wordpress.ncsu.edu/, 2019.

S. Davis, B. Abbasi, S. Shah, S. Telfer, and M. Begon. Spatial analyses of wildlife contact networks. *Journal of the Royal Society Interface*, 12(102):20141004, 2015.

F. Emmert-Streib, M. Dehmer, and B. Haibe-Kains. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Frontiers in Cell and Developmental Biology*, 2:38, 2014.

P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5 (1):17–60, 1960.

A. Ertefaie, M. Asgharian, and D. A. Stephens. Variable selection in causal inference using penalization. *arXiv preprint, arXiv:1311.1283*, 2013.

A. Ertefaie, J. R. McKay, D. Oslin, and R. L. Strawderman. Robust Q-learning. *Journal of the American Statistical Association*, 116(533):368–381, 2021.

J. P. Fine and R. J. Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509, 1999.

L. Forastiere, E. M. Airoldi, and F. Mealli. Identification and estimation of treatment and interference effects in observational studies on networks. *arXiv preprint, arXiv:1609.06245*, 2016.

L. Forastiere, E. M. Airoldi, and F. Mealli. Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association*, 116(534):901–918, 2021.

A. R. Foulstone, A. B. Kelly, and T. Kifle. Partner influences on smoking cessation: A longitudinal study of couple relationships. *Journal of Substance Use*, 22(5):501–506, 2017.

S. A. Geer and S. van de Geer. *Empirical Processes in M-estimation*. Cambridge University Press, 2000.

P. Ghosh and B. Chakraborty. Comparison of dynamic treatment regimes with an ordinal outcome. *arXiv preprint, arXiv:1808.07287*, 2018.

V. P. Godambe. *Estimating Functions*. Oxford University Press, 1991.

M. Goossens, F. Mittelbach, and A. Samarin. *The LaTeX Companion*. Addison-Wesley, Reading, Massachusetts, 1994.

D. J. Graham, E. J. McCoy, and D. A. Stephens. Quantifying the effect of area deprivation on child pedestrian casualties by using longitudinal mixed models to adjust for confounding, interference and spatial dependence. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(4):931–950, 2013.

S. Gravely, G. Meng, K. M. Cummings, A. Hyland, R. Borland, D. Hammond, R. J. O'Connor, M. L. Goniewicz, K. A. Kasza, A. McNeill, et al. Changes in smoking and vaping over 18 months among smokers and recent ex-smokers: Longitudinal findings from the 2016 and 2018 ITC Four Country smoking and vaping surveys. *International Journal of Environmental Research and Public Health*, 17(19):7084, 2020.

J. P. Grutters, M. Sculpher, A. H. Briggs, J. L. Severens, M. J. Candel, J. E. Stahl, D. De Ruysscher, A. Boer, B. L. Ramaekers, and M. A. Joore. Acknowledging patient heterogeneity in economic evaluation. *Pharmacoeconomics*, 31(2):111–123, 2013.

L. Gunter, J. Zhu, and S. Murphy. Variable selection for optimal decision making. In *Conference on Artificial Intelligence in Medicine in Europe*, pages 149–154. Springer, 2007.

P. Hajek, A. Phillips-Waller, D. Przulj, F. Pesola, K. Myers Smith, N. Bisal, J. Li, S. Parrott, P. Sasieni, L. Dawkins, et al. A randomized trial of e-cigarettes versus nicotine-replacement therapy. *New England Journal of Medicine*, 380(7):629–637, 2019.

M. E. Halloran and C. J. Struchiner. Causal inference in infectious diseases. *Epidemiology*, 6: 142–151, 1995.

N. Hassanpour and R. Greiner. Learning disentangled representations for counterfactual regression. In *International Conference on Learning Representations*, 2019.

T. Hastie and R. Tibshirani. Generalized additive models for medical research. *Statistical Methods in Medical Research*, 4(3):187–196, 1995.

T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models*. CRC Press, 1990.

T. Hatt, J. Berrevoets, A. Curth, S. Feuerriegel, and M. van der Schaar. Combining observational and randomized data for estimating heterogeneous treatment effects. *arXiv preprint, arXiv:2202.12891*, 2022.

R. Hayes, N. D. Alexander, S. Bennett, and S. Cousens. Design and analysis issues in cluster-randomized trials of interventions against infectious diseases. *Statistical Methods in Medical Research*, 9(2):95–116, 2000.

M. A. Hernán and J. M. Robins. *Causal Inference: What If*. Boca Raton: Chapman and Hall/CRC., 2020.

A. Herren and P. R. Hahn. Semi-supervised learning and the question of true versus estimated propensity scores. *arXiv preprint, arXiv:2009.06183*, 2020.

S. J. T. Hidalgo, M. T. Lawson, D. J. Luckett, M. Chaudhari, J. Chen, A. Choudhury, A. D. Florio, X. Jiang, C. T. Nguyen, and M. R. Kosorok. A master pipeline for discovery and validation of biomarkers. In *Machine Learning for Health Informatics*, pages 259–288. Springer, 2016.

G. Hong and S. W. Raudenbush. Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*, 101 (475):901–910, 2006.

J. Huang, Z. Duan, J. Kwok, S. Binns, L. E. Vera, Y. Kim, G. Szczypka, and S. L. Emery. Vaping versus JUULing: how the extraordinary growth and marketing of JUUL transformed the US retail e-cigarette market. *Tobacco Control*, 28(2):146–151, 2019.

Y. Huang and Y. Fong. Identifying optimal biomarker combinations for treatment selection via a robust kernel method. *Biometrics*, 70(4):891–901, 2014.

G. Hubbard, T. Gorely, G. Ozakinci, R. Polson, and L. Forbat. A systematic review and narrative summary of family-based smoking cessation interventions to help adults quit smoking. *BMC Family Practice*, 17(1):73, 2016.

P. J. Huber. *Robust Statistics*. John Wiley & Sons, 2004.

M. G. Hudgens and M. E. Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, 2008.

A. Hyland, R. Borland, Q. Li, H. H. Yong, A. McNeill, G. T. Fong, R. J. O'Connor, and K. M. Cummings. Individual-level predictors of cessation behaviours among participants in the International Tobacco Control (ITC) Four Country Survey. *Tobacco Control*, 15(suppl 3): iii83–iii94, 2006.

A. Hyland, B. K. Ambrose, K. P. Conway, N. Borek, E. Lambert, C. Carusi, K. Taylor, S. Crosse, G. T. Fong, K. M. Cummings, et al. Design and methods of the Population Assessment of Tobacco and Health (PATH) Study. *Tobacco Control*, 26(4):371–378, 2017.

G. W. Imbens and D. B. Rubin. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015.

C. Jiang, M. Wallace, and M. Thompson. Doubly-robust dynamic treatment regimen estimation for binary outcomes. *arXiv preprint, arXiv:2203.08269*, 2022a.

C. Jiang, M. P. Wallace, and M. E. Thompson. Dynamic treatment regimes with interference. *Canadian Journal of Statistics*, 2022b.

A. P. Jones, F. G. Happé, F. Gilbert, S. Burnett, and E. Viding. Feeling, caring, knowing: different types of empathy deficit in boys with psychopathic tendencies and autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, 51(11):1188–1197, 2010.

E. K. Kao. Causal inference under network interference: A framework for experiments on social networks. *arXiv preprint, arXiv:1708.08522*, 2017.

K. A. Kasza, K. C. Edwards, H. L. Kimmel, A. Anesetti-Rothermel, K. M. Cummings, R. S. Niaura, A. Sharma, E. M. Ellis, R. Jackson, C. Blanco, et al. Association of e-cigarette use with discontinuation of cigarette smoking among adult smokers who were initially never planning to quit. *JAMA Network Open*, 4(12):e2140880–e2140880, 2021.

E. H. Kennedy. Semiparametric theory. *arXiv preprint, arXiv:1709.06418*, 2017.

E. H. Kennedy. Optimal doubly robust estimation of heterogeneous causal effects. *arXiv preprint, arXiv:2004.14497*, 2020.

S. Knüppel and A. Stang. DAG Program: Identifying minimal sufficient adjustment sets. *Epidemiology*, 21(1):159, 2010.

D. Knuth. *The TEXbook*. Addison-Wesley, Reading, Massachusetts, 1986.

M. R. Kosorok and E. B. Laber. Precision medicine. *Annual Review of Statistics and its Application*, 6:263–286, 2019.

M. R. Kosorok and E. E. Moodie. *Adaptive treatment strategies in practice: planning trials and analyzing data for personalized medicine*. SIAM, 2015.

L. Kraemer and B. Banerjee. Multi-agent reinforcement learning as a rehearsal for decentralized planning. *Neurocomputing*, 190:82–94, 2016.

S. R. Künzel, J. S. Sekhon, P. J. Bickel, and B. Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10):4156–4165, 2019.

E. B. Laber, D. J. Lizotte, and B. Ferguson. Set-valued dynamic treatment regimes for competing outcomes. *Biometrics*, 70(1):53–61, 2014.

L. Lamport. *LATEX — A Document Preparation System*. Addison-Wesley, Reading, Massachusetts, Second edition, 1994.

P. W. Lavori and R. Dawson. A design for testing clinical strategies: biased adaptive within-subject randomization. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 163(1): 29–38, 2000.

P. W. Lavori and R. Dawson. Dynamic treatment regimes: practical design considerations. *Clinical Trials*, 1(1):9–20, 2004.

M. Le Grande, R. Borland, H.-H. Yong, K. M. Cummings, A. McNeill, M. E. Thompson, and G. T. Fong. Predictive power of dependence measures for quitting smoking. findings from the 2016 to 2018 ITC Four Country Smoking and Vaping Surveys. *Nicotine and Tobacco Research*, 23(2):276–285, 2021.

L. Lei and E. J. Candès. Conformal inference of counterfactuals and individual treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2021.

M. P. Leung. Treatment and spillover effects under network interference. *Review of Economics and Statistics*, 2016.

M. P. Leung. Causal inference under approximate neighborhood interference. *Available at SSRN 3479902*, 2019.

M. P. Leung. Treatment and spillover effects under network interference. *Review of Economics and Statistics*, 102(2):368–380, 2020.

M. A. Lewis, C. M. McBride, K. I. Pollak, E. Puleo, R. M. Butterfield, and K. M. Emmons. Understanding health behavior change among couples: An interdependence and communal coping approach. *Social Science & Medicine*, 62(6):1369–1380, 2006.

F. Li. Propensity score weighting for causal inference with multiple treatments. *The Annals of Applied Statistics*, 13(4):2389–2415, 2019.

F. Li, K. L. Morgan, and A. M. Zaslavsky. Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, 113(521):390–400, 2018.

S. Li and S. Wager. Random graph asymptotics for treatment effect estimation under network interference. *arXiv preprint, arXiv:2007.13302*, 2020.

K.-Y. Liang and S. L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

S. R. Lipsitz, N. M. Laird, and D. P. Harrington. Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika*, 78(1):153–160, 1991.

C. Liu. Robit regression: A simple robust alternative to logistic and probit regression. *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*, pages 227–238, 2004.

L. Liu, M. G. Hudgens, and S. Becker-Dreps. On inverse probability-weighted estimators in the presence of interference. *Biometrika*, 103(4):829–842, 2016.

L. Liu, M. G. Hudgens, B. Saul, J. D. Clemens, M. Ali, and M. E. Emch. Doubly robust estimation in observational studies with partial interference. *Stat*, 8(1):e214, 2019.

D. J. Lizotte, M. H. Bowling, and S. A. Murphy. Efficient reinforcement learning with multiple reward functions for randomized controlled trial analysis. In *ICML*, pages 695–702, 2010. URL https://icml.cc/Conferences/2010/papers/464.pdf.

M. Lonergan, S. J. Senn, C. McNamee, A. K. Daly, R. Sutton, A. Hattersley, E. Pearson, and M. Pirmohamed. Defining drug response for stratified medicine. *Drug discovery today*, 22(1): 173–179, 2017.

I. M. Longini Jr, M. E. Halloran, and A. Nizam. Model-based estimation of vaccine effects from community vaccine trials. *Statistics in Medicine*, 21(4):481–495, 2002.

M. C. Lovell. Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, 58(304):993–1010, 1963.

T. Lumley. Analysis of complex survey samples. *Journal of Statistical Software*, 9:1–19, 2004.

T. Lumley, R. Kronmal, and S. Ma. Relative risk regression in medical research: models, contrasts, estimators, and algorithms. *UW Biostatistics Working Paper Series*, 2006.

C. F. Manski. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542, 1993.

C. F. Manski. Identification of treatment response with social interactions. *The Econometrics Journal*, 16(1):S1–S23, 2013.

K. V. Mardia. Some contributions to contingency-type bivariate distributions. *Biometrika*, 54 (1-2):235–249, 1967.

P. McCullagh. Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):109–127, 1980.

C. W. Mendo, M. Maurel, I. Doré, J. O'Loughlin, and M.-P. Sylvestre. Depressive symptoms and cigarette smoking in adolescents and young adults: Mediating role of friends smoking. *Nicotine and Tobacco Research*, 23(10):1771–1778, 2021.

W. Miao, Z. Geng, and E. J. Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993, 2018.

E. Miguel and M. Kremer. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217, 2004.

C. H. Miles, M. Petersen, and M. J. van der Laan. Causal inference when counterfactuals depend on the proportion of all subjects exposed. *Biometrics*, 2019.

D. C. Montgomery. *Design and Analysis of Experiments*. John Wiley & Sons, 2017.

E. E. Moodie and E. F. Krakow. Precision medicine: Statistical methods for estimating adaptive treatment strategies. *Bone Marrow Transplantation*, 55(10):1890–1896, 2020.

E. E. Moodie, T. S. Richardson, and D. A. Stephens. Demystifying optimal dynamic treatment regimes. *Biometrics*, 63(2):447–455, 2007.

E. E. Moodie, B. Chakraborty, and M. S. Kramer. Q-learning for estimating optimal dynamic treatment rules from observational data. *Canadian Journal of Statistics*, 40(4):629–645, 2012.

E. E. Moodie, N. Dean, and Y. R. Sun. Q-learning: Flexible learning about useful utilities. *Statistics in Biosciences*, 6(2):223–243, 2014.

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

S. A. Murphy. An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*, 24(10):1455–1481, 2005.

X. Nie and S. Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108 (2):299–319, 2021.

E. L. Ogburn and T. J. VanderWeele. Causal diagrams for interference. *Statistical science*, 29(4): 559–578, 2014.

E. L. Ogburn, T. J. VanderWeele, et al. Vaccines, contagion, and social networks. *The Annals of Applied Statistics*, 11(2):919–948, 2017.

J. O'Loughlin, E. N. Dugas, J. Brunet, J. DiFranza, J. C. Engert, A. Gervais, K. Gray-Donald, I. Karp, N. C. Low, C. Sabiston, et al. Cohort profile: the nicotine dependence in teens (ndit) study. *International Journal of Epidemiology*, 44(5):1537–1546, 2015.

C. Park and H. Kang. Efficient semiparametric estimation of network treatment effects under partial interference. *arXiv preprint, arXiv:2004.08950*, 2020.

PATH. *Population Assessment of Tobacco and Health (PATH) Study [United States] Restricted-Use Files*. United States Department of Health and Human Services. National Institutes of Health. National Institute on Drug Abuse, and United States Department of Health and Human Services. Food and Drug Administration. Center for Tobacco Products. Inter-university Consortium for Political and Social Research [distributor], 2020-06-24. https://doi.org/10.3886/ICPSR36231.v25, 2020. doi: 10.3886/ICPSR36231.v25.

J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2000.

W. E. Pelham Jr and G. A. Fabiano. Evidence-based psychosocial treatments for attention-deficit/hyperactivity disorder. *Journal of Clinical Child & Adolescent Psychology*, 37(1): 184–214, 2008.

W. E. Pelham Jr, G. Fabiano, J. Waxmonsky, A. Greiner, M. Hoffman, S. Murphy, E. M. Foster, J. Yu, E. Gnagy, I. Bhatia, et al. Adaptive pharmacological and behavioral treatments for children with ADHD: Sequencing, combining, and escalating doses. In *Poster presented at the Institute of Educational Sciences' Third Annual Research Conference, Washington, DC*, 2008.

C. Perez-Heydrich, M. G. Hudgens, M. E. Halloran, J. D. Clemens, M. Ali, and M. E. Emch. Assessing effects of cholera vaccination in the presence of interference. *Biometrics*, 70(3): 731–741, 2014.

W. J. Platt, G. W. Evans, and S. L. Rathbun. The population dynamics of a long-lived conifer (pinus palustris). *The American Naturalist*, 131(4):491–525, 1988.

R. Polosa, P. Caponnetto, M. Maglia, J. B. Morjaria, and C. Russo. Success rates with nicotine personal vaporizers: a prospective 6-month pilot study of smokers not intending to quit. *BMC Public Health*, 14(1):1159, 2014.

R. L. Prentice. Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 44:1033–1048, 1988.

D. Puelz, G. Basse, A. Feller, and P. Toulis. A graph-theoretic approach to randomization tests of causal effects under general interference. *arXiv preprint, arXiv:1910.10862*, 2019.

T. Qian, H. Yoo, P. Klasnja, D. Almirall, and S. A. Murphy. Estimating time-varying causal excursion effect in mobile health with binary outcomes. *arXiv preprint, arXiv:1906.00528*, 2019.

S. L. Rathbun and N. Cressie. A space-time survival point process for a longleaf pine forest in southern Georgia. *Journal of the American Statistical Association*, 89(428):1164–1174, 1994.

T. S. Richardson, J. M. Robins, and L. Wang. On modeling and estimation for the relative risk and risk difference. *Journal of the American Statistical Association*, 112(519):1121–1130, 2017.

J. Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9-12):1393–1512, 1986.

J. M. Robins. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics - Theory and Methods*, 23(8):2379–2412, 1994.

J. M. Robins. Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics*, pages 189–326. Springer, 2004.

J. M. Robins, A. Rotnitzky, and L. P. Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427): 846–866, 1994.

P. M. Robinson. Root-N-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, 56:931–954, 1988.

M. J. Rohrbaugh, V. Shoham, S. Trost, M. Muramoto, R. M. Cate, and S. Leischow. Couple dynamics of change-resistant smoking: Toward a family consultation model. *Family Process*, 40(1):15–31, 2001.

E. J. Rose, E. E. Moodie, and S. Shortreed. Monte Carlo sensitivity analysis for unmeasured confounding in dynamic treatment regimes. *arXiv preprint, arXiv:2202.09448*, 2022.

P. R. Rosenbaum. Interference between units in randomized experiments. *Journal of the American Statistical Association*, 102(477):191–200, 2007.

P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

A. Rotnitzky. Inverse probability weighted methods. In *Longitudinal Data Analysis*, pages 467–490. Chapman and Hall/CRC, 2008.

D. B. Rubin. Comment: Randomization analysis of experimental data: The Fisher randomization test. *Journal of the American Statistical Association*, 75(371):591–593, 1980.

D. B. Rubin. Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science*, 5(4):472–480, 1990.

A. J. Rush, M. Fava, S. R. Wisniewski, P. W. Lavori, M. H. Trivedi, H. A. Sackeim, M. E. Thase, A. A. Nierenberg, F. M. Quitkin, T. M. Kashner, et al. Sequenced treatment alternatives to relieve depression (STAR* D): rationale and design. *Controlled Clinical Trials*, 25(1):119–142, 2004.

F. Sävje. Causal inference with misspecified exposure mappings. *arXiv preprint, arXiv:2103.06471*, 2021.

F. Sävje, P. M. Aronow, and M. G. Hudgens. Average treatment effects in the presence of unknown interference. *The Annals of Statistics*, 49(2):673–701, 2021.

M. E. Schnitzer, M. J. van der Laan, E. E. Moodie, and R. W. Platt. Effect of breastfeeding on gastrointestinal infection in infants: a targeted maximum likelihood approach for clustered longitudinal data. *The Annals of Applied Statistics*, 8(2):703–725, 2014.

P. J. Schulte, A. A. Tsiatis, E. B. Laber, and M. Davidian. Q-and A-learning methods for estimating optimal dynamic treatment regimes. *Statistical Science*, 29(4):640–661, 2014.

J. Schulz and E. E. Moodie. Doubly robust estimation of optimal dosing strategies. *Journal of the American Statistical Association*, 116(533):256–268, 2021.

A. S. Selya, L. Dierker, J. S. Rose, D. Hedeker, and R. J. Mermelstein. The role of nicotine dependence in e-cigarettes' potential for smoking reduction. *Nicotine and Tobacco Research*, 20(10):1272–1277, 2017.

U. Shalit, F. D. Johansson, and D. Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In D. Precup and Y. W. Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 3076–3085. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/shalit17a.html.

E. Sherman, D. Arbour, and I. Shpitser. General identification of dynamic treatment regimes under interference. *arXiv preprint, arXiv:2004.01218*, 2020.

C. Shi, D. Blei, and V. Veitch. Adapting neural networks for the estimation of treatment effects. *Advances in Neural Information Processing Systems*, 32, 2019.

G. Simoneau, E. E. Moodie, J. S. Nijjar, R. W. Platt, S. E. R. A. I. C. Investigators, et al. Estimating optimal dynamic treatment regimes with survival outcomes. *Journal of the American Statistical Association*, 115(531):1531–1539, 2020.

M. E. Sobel. What do randomized studies of housing mobility demonstrate? causal inference in the face of interference. *Journal of the American Statistical Association*, 101(476):1398–1407, 2006.

O. Sofrygin, R. Neugebauer, and M. J. van der Laan. Conducting simulations in causal inference with networks-based structural equation models. *arXiv preprint, arXiv:1705.10376*, 2017.

D. Spicker and M. P. Wallace. Measurement error and precision medicine: Error-prone tailoring covariates in dynamic treatment regimes. *Statistics in Medicine*, 39(26):3732–3755, 2020.

M. J. Stensrud and A. L. Sarvet. Optimal regimes for algorithm-assisted human decision-making. *arXiv preprint, arXiv:2203.03020*, 2022.

V. J. Strecher, J. B. McClure, G. L. Alexander, B. Chakraborty, V. N. Nair, J. M. Konkel, S. M. Greene, L. M. Collins, C. C. Carlier, C. J. Wiese, et al. Web-based smoking-cessation programs: results of a randomized trial. *American Journal of Preventive Medicine*, 34(5):373–381, 2008.

D. R. Strong, J. Pearson, S. Ehlke, T. Kirchner, D. Abrams, K. Taylor, W. M. Compton, K. P. Conway, E. Lambert, V. R. Green, et al. Indicators of dependence for different types of tobacco product users: Descriptive findings from Wave 1 (2013–2014) of the Population Assessment of Tobacco and Health (PATH) study. *Drug and Alcohol Dependence*, 178:257–266, 2017.

T. S. Stroup, J. P. McEvoy, M. S. Swartz, M. J. Byerly, I. D. Glick, J. M. Canive, M. F. McGee, G. M. Simpson, M. C. Stevens, and J. A. Lieberman. The national institute of mental health clinical antipsychotic trials of intervention effectiveness (catie) project: schizophrenia trial design and protocol development. *Schizophrenia bulletin*, 29(1):15, 2003.

R. Stusser. Reflections on the scientific method in medicine. *Medical and Health Sciences*, 8: 53–77, 2010.

L. Su, W. Lu, and R. Song. Modelling and estimation for optimal treatment decision with interference. *Stat*, 8(1):e219, 2019.

R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT press, 2018.

E. J. T. Tchetgen and T. J. VanderWeele. On causal inference in the presence of interference. *Statistical Methods in Medical Research*, 21(1):55–75, 2012.

E. J. T. Tchetgen, A. Ying, Y. Cui, X. Shi, and W. Miao. An introduction to proximal causal learning. *arXiv preprint, arXiv:2009.10982*, 2020.

E. J. Tchetgen Tchetgen, J. M. Robins, and A. Rotnitzky. On doubly robust estimation in a semiparametric odds ratio model. *Biometrika*, 97(1):171–180, 2009.

E. J. Tchetgen Tchetgen, I. R. Fulcher, and I. Shpitser. Auto-g-computation of causal effects on a network. *Journal of the American Statistical Association*, 116(534):833–844, 2021.

P. F. Thall, L. H. Wooten, C. J. Logothetis, R. E. Millikan, and N. M. Tannir. Bayesian and frequentist two-stage treatment strategies based on sequential failure times subject to interval censoring. *Statistics in medicine*, 26(26):4687–4702, 2007.

C. Tortù. *Essays on causal inference and complex networks*. PhD dissertation, IMT School for Advanced Studies Lucca, 2020. URL http://e-theses.imtlucca.it/id/eprint/323.

M. H. Trivedi, A. J. Rush, S. R. Wisniewski, A. A. Nierenberg, D. Warden, L. Ritz, G. Norquist, R. H. Howland, B. Lebowitz, P. J. McGrath, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR* D: implications for clinical practice. *American Journal of Psychiatry*, 163(1):28–40, 2006.

A. A. Tsiatis. *Semiparametric Theory and Missing Data*. Springer, 2006.

A. A. Tsiatis. *Dynamic Treatment Regimes: Statistical Methods for Precision Medicine*. CRC Press, 2019.

G. Tutz and W. Hennevogl. Random effects in ordinal regression models. *Computational Statistics & Data Analysis*, 22(5):537–557, 1996.

M. J. Van der Laan. Causal inference for a population of causally connected units. *Journal of Causal Inference*, 2(1):13–74, 2014.

M. J. Van der Laan, E. C. Polley, and A. E. Hubbard. Super learner. *Statistical Applications in Genetics and Molecular Biology*, 6(1), 2007.

T. J. VanderWeele and N. A. Christakis. Network multipliers and public health. *International Journal of Epidemiology*, 2019.

A. C. Villanti, S. P. Feirman, R. S. Niaura, J. L. Pearson, A. M. Glasser, L. K. Collins, and D. B. Abrams. How do we determine the impact of e-cigarettes on cigarette smoking cessation or reduction? Review and recommendations for answering the research question with scientific rigor. *Addiction*, 113(3):391–404, 2018.

D. Viviano. Experimental design under network interference. *arXiv preprint, arXiv:2003.08421*, 2020.

A. S. Wahed and A. A. Tsiatis. Optimal estimator for the survival distribution and related quantities for treatment policies in two-stage randomization designs in clinical trials. *Biometrics*, 60(1): 124–133, 2004.

M. P. Wallace and E. E. M. Moodie. Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics*, 71(3):636–644, 2015.

M. P. Wallace, E. E. Moodie, and D. A. Stephens. Model validation and selection for personalized medicine using dynamic-weighted ordinary least squares. *Statistical Methods in Medical Research*, 26(4):1641–1653, 2017a.

M. P. Wallace, E. E. Moodie, D. A. Stephens, et al. Dynamic treatment regimen estimation via regression-based techniques: Introducing R package dtrreg. *Journal of Statistical Software*, 80 (2):1–20, 2017b.

M. P. Wallace, E. E. Moodie, and D. A. Stephens. Model selection for g-estimation of dynamic treatment regimes. *Biometrics*, 75(4):1205–1215, 2019.

L. Wang, T. S. Richardson, and J. M. Robins. Congenial causal inference with binary structural nested mean models. *arXiv preprint, arXiv:1709.08281*, 2017.

H. White. Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pages 1–25, 1982.

S. N. Wood. *Generalized Additive Models: An Introduction with R*. CRC press, 2017.

Y. Xu, P. Müller, A. S. Wahed, and P. F. Thall. Bayesian nonparametric estimation for dynamic treatment regimes with sequential transition times. *Journal of the American Statistical Association*, 111(515):921–950, 2016.

X. Yan and B. Chakraborty. Time for smoking reduction strategies to get more personal, 2018. URL https://www.todayonline.com/commentary/time-smoking-reduction-strategies-get-more-personal.

S. Yang, D. Zeng, and X. Wang. Elastic integrative analysis of randomized trial and real-world data for treatment heterogeneity estimation. *arXiv preprint, arXiv:2005.10579*, 2020.

G. Y. Yi and M. E. Thompson. Marginal and association regression models for longitudinal binary data with drop-outs: A likelihood-based approach. *Canadian Journal of Statistics*, 33(1):3–20, 2005.

J. Yin, T. S. Richardson, and L. Wang. Multiplicative effect modeling: The general case. *arXiv preprint, arXiv:1906.00558*, 2019.

K.-H. Yuan and R. I. Jennrich. Asymptotics of estimating equations under natural conditions. *Journal of Multivariate Analysis*, 65(2):245–260, 1998.

D. Zeng. Lecture notes "Introduction to Statistical Learning and Personalized Medicine". https://www.bios.unc.edu/~dzeng/Bios740.html, 2021.

B. Zhang, A. A. Tsiatis, E. B. Laber, and M. Davidian. A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018, 2012.

B. Zhang, A. A. Tsiatis, E. B. Laber, and M. Davidian. Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika*, 100(3):681–694, 2013.

Y. Zhao, D. Zeng, A. J. Rush, and M. R. Kosorok. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499): 1106–1118, 2012.

Y.-Q. Zhao, D. Zeng, E. B. Laber, and M. R. Kosorok. New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association*, 110(510):583–598, 2015.

T. Zhou, G. Tong, F. Li, and L. E. Thomas. Psweight: An R package for propensity score weighting analysis. *arXiv preprint, arXiv:2010.08893*, 2020.

X. Zhou, N. Mayer-Hamblett, U. Khan, and M. R. Kosorok. Residual weighted learning for estimating individualized treatment rules. *Journal of the American Statistical Association*, 112 (517):169–187, 2017.

APPENDICES

Appendix A

Q-learning and dWOLS Three-stage examples

Example A.1. Assuming that the observed outcome is Y with history $(\boldsymbol{H}_3, A_3)$, the Q-learning focused mainly on the quality functions (Q-functions) which are defined as:

$$Q_3\left(\boldsymbol{h}_3, a_3\right) = \mathbb{E}\left[Y|\boldsymbol{H}_3 = \boldsymbol{h}_3\right];$$

$$Q_2\left(\boldsymbol{h}_2, a_2\right) = \mathbb{E}\left[\max_{A_3} Q_3\left(\boldsymbol{H}_3, A_3\right)|\boldsymbol{H}_2 = \boldsymbol{h}_2, A_2 = a_2\right];$$

$$Q_1\left(\boldsymbol{h}_1, a_1\right) = \mathbb{E}\left[\max_{A_2} Q_2\left(\boldsymbol{H}_2, A_2\right)|\boldsymbol{H}_1 = \boldsymbol{h}_1, A_1 = a_1\right].$$

Stage 3: When the true Q-functions are not known, one has to estimate them from the data. If we propose a linear working model for the quality functions such as:

$$Q_3\left(\boldsymbol{h}_3, a_3\right) = \mathbb{E}\left[Y|\boldsymbol{H}_3 = \boldsymbol{h}_3\right] = \boldsymbol{\beta}_3^\top \boldsymbol{h}_3^\beta + \boldsymbol{\psi}_3^\top a_3 \boldsymbol{h}_3^\psi.$$

We can regress the observed outcome Y on the history $\boldsymbol{H}_3$ to acquire the estimated parameters $\hat{\boldsymbol{\beta}}_3$ and $\hat{\boldsymbol{\psi}}_3$. Then the optimal decision rule could be decided by estimated parameters $\hat{\boldsymbol{\psi}}_3$ and history

$\boldsymbol{H}_3$; it is to *prescribe* $A_3 = 1$ *if* $\hat{\boldsymbol{\psi}}_3^\top \boldsymbol{H}_3^\psi > 0$; *and prescribe* $A_3 = 0$ *otherwise*.

Stage 2: For the second stage, we use estimated parameters $\hat{\boldsymbol{\beta}}_3$ and $\hat{\boldsymbol{\psi}}_3$ to construct a pseudo-outcome $\widetilde{\mathcal{Y}_2}(\boldsymbol{H}_3, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\psi}}_3) = \max_{A_3} Q_3(\boldsymbol{H}_3, A_3; \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\psi}}_3)$, this pseudo-outcome means the expected potential outcome that a patient with the given history would have if he/she goes on to receive optimal third-stage treatment. Then, we still propose the linear model for stage two such that:

$$Q_2\left(\boldsymbol{h}_2, a_2\right) = \mathbb{E}\left[\max_{A_3} Q_3\left(\boldsymbol{H}_3, A_3\right) | \boldsymbol{H}_2 = \boldsymbol{h}_2, A_2 = a_2\right] = \boldsymbol{\beta}_2^\top \boldsymbol{h}_2^\beta + \boldsymbol{\psi}_2^\top a_2 \boldsymbol{h}_2^\psi.$$

Through regressing the pseudo-outcome $\widetilde{\mathcal{Y}_2} = \max_{A_3} Q_3(\boldsymbol{H}_3, A_3; \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\psi}}_3)$ on history $\boldsymbol{H}_2$, we can obtain estimated parameters $\hat{\boldsymbol{\beta}}_2$ and $\hat{\boldsymbol{\psi}}_2$, and the second stage optimal treatment rule is to *prescribe* $A_2 = 1$ *if* $\hat{\boldsymbol{\psi}}_2^\top \boldsymbol{H}_2^\psi > 0$; *and prescribe* $A_2 = 0$ *otherwise*.

Stage 1: Following the same logic as in the second stage, we build a pseudo-outcome (i.e., $\widetilde{\mathcal{Y}_1}(\boldsymbol{H}_2, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\psi}}_2) = \max_{A_2} Q_2(\boldsymbol{H}_2, A_2; \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\psi}}_2)$), then propose the linear model as follows:

$$Q_1\left(\boldsymbol{h}_1, a_1\right) = \mathbb{E}\left[\max_{A_2} Q_2\left(\boldsymbol{H}_2, A_2\right) | \boldsymbol{H}_1 = \boldsymbol{h}_1, A_1 = a_1\right] = \boldsymbol{\beta}_1^\top \boldsymbol{h}_1^\beta + \boldsymbol{\psi}_1^\top a_1 \boldsymbol{h}_1^\psi,$$

using a regression and acquiring estimated parameters, then we have the first stage optimal treatment rule which is to *prescribe* $A_1 = 1$ *if* $\hat{\boldsymbol{\psi}}_1^\top \boldsymbol{H}_1^\psi > 0$; *and prescribe* $A_1 = 0$ *otherwise*.

Example A.2. We now illustrate the dWOLS in a three-stage example.

We assume that the observed outcome is $Y = y$ with history $(\boldsymbol{H}_3, A_3)$. Like Q-learning, dWOLS also uses backward induction; therefore, we start with the third (last) stage.

Stage 3: The pseudo-outcome of the last stage is identical to the observed outcome $\widetilde{\mathcal{Y}_3} = Y$. Also, suppose we propose the linear model for both the blip model and the treatment-free model as

$$\mathbb{E}\left[\widetilde{\mathcal{Y}_3} | \boldsymbol{H}_3 = \boldsymbol{h}_3; \boldsymbol{\beta}_3, \boldsymbol{\psi}_3\right] = \boldsymbol{\beta}_3^\top \boldsymbol{h}_3^\beta + \boldsymbol{\psi}_3^\top a_3 \boldsymbol{h}_3^\psi,$$

and employ the weight w_3 that satisfy the criterion (2.2) such as $w_3 = |a_3 - \mathbb{E}\left[A_3 | \boldsymbol{H}_3\right]|$.

We can compute

$$\left(\hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\psi}}_3\right) = \underset{\boldsymbol{\beta}_3, \boldsymbol{\psi}_3}{\arg\min} \sum_{i=1}^{n} w_3 \left(\widetilde{\mathcal{Y}_{3i}} - \boldsymbol{\beta}_3^\top \boldsymbol{h}_{3i}^{\beta} + \boldsymbol{\psi}_3^\top a_3 \boldsymbol{h}_{3i}^{\psi}\right)^2,$$

using weighted ordinary least squares (WOLS) regressing $\widetilde{\mathcal{Y}_3}$ on $\left(\boldsymbol{h}_3^{\beta}, \boldsymbol{h}_3^{\psi}\right)$ with weights $w_3 = |a_3 - \mathbb{E}\left[A_3|\boldsymbol{H}_3\right]|$. Therefore, the third stage optimal treatment rule is to *prescribe* $A_3 = 1$ *if* $\hat{\boldsymbol{\psi}}_3^\top \boldsymbol{H}_3^{\psi} > 0$; *then prescribe* $A_3 = 0$ *otherwise*.

Stage 2: The pseudo-outcome of the second stage is $\widetilde{\mathcal{Y}_2} = Y + \mu_3\left(\boldsymbol{h}_3, a_3; \hat{\boldsymbol{\psi}}_3\right) = Y + (\hat{\boldsymbol{\psi}}_3^\top a_3 \boldsymbol{h}_3^{\psi})^+ - \hat{\boldsymbol{\psi}}_3^\top a_3 \boldsymbol{h}_3^{\psi}$. We propose the linear model for both the blip model and the treatment-free model such as

$$\mathbb{E}\left[\widetilde{\mathcal{Y}_2}|\boldsymbol{H}_2 = \boldsymbol{h}_2; \boldsymbol{\beta}_2, \boldsymbol{\psi}_2\right] = \boldsymbol{\beta}_2^\top \boldsymbol{h}_2^{\beta} + \boldsymbol{\psi}_2^\top a_2 \boldsymbol{h}_2^{\psi},$$

and employing the weight $w_2 = |a_2 - \mathbb{E}\left[A_2|\boldsymbol{H}_2\right]|$ compute

$$\left(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\psi}}_2\right) = \underset{\boldsymbol{\beta}_2, \boldsymbol{\psi}_2}{\arg\min} \sum_{i=1}^{n} w_2 \left(\widetilde{\mathcal{Y}_{2i}} - \boldsymbol{\beta}_2^\top \boldsymbol{h}_{2i}^{\beta} + \boldsymbol{\psi}_2^\top a_2 \boldsymbol{h}_{2i}^{\psi}\right)^2,$$

using weighted ordinary least squares (WOLS) regressing $\widetilde{\mathcal{Y}_2}$ on $\left(\boldsymbol{h}_2^{\beta}, \boldsymbol{h}_2^{\psi}\right)$ with weights $w_2 = |a_2 - \mathbb{E}\left[A_2|\boldsymbol{H}_2\right]|$. Thus, the second stage optimal treatment rule is to *prescribe* $A_2 = 1$ *if* $\hat{\boldsymbol{\psi}}_2^\top \boldsymbol{H}_2^{\psi} > 0$; *then prescribe* $A_2 = 0$ *otherwise*.

Stage 1: For the first stage, the process is similar. Note that the the pseudo-outcome $\widetilde{\mathcal{Y}_1} = Y + \sum_{k=2}^{2} \mu_k\left(\boldsymbol{h}_k, a_k; \boldsymbol{\psi}_k\right) = Y + [(\hat{\boldsymbol{\psi}}_2^\top a_2 \boldsymbol{h}_2^{\psi})^+ - \hat{\boldsymbol{\psi}}_2^\top a_2 \boldsymbol{h}_2^{\psi}] + [(\hat{\boldsymbol{\psi}}_3^\top a_3 \boldsymbol{h}_3^{\psi})^+ - \hat{\boldsymbol{\psi}}_3^\top a_3 \boldsymbol{h}_3^{\psi}]$. Apply WOLS regressing $\widetilde{\mathcal{Y}_1}$ on $\left(\boldsymbol{h}_1^{\beta}, \boldsymbol{h}_1^{\psi}\right)$ with weights $w_1 = |a_1 - \mathbb{E}\left[A_1|\boldsymbol{H}_1\right]|$. We finally obtain that the first stage optimal treatment rule is to *prescribe* $A_1 = 1$ *if* $\hat{\boldsymbol{\psi}}_1^\top \boldsymbol{H}_1^{\psi} > 0$; *then prescribe* $A_1 = 0$ *otherwise*.

Appendix B

dWOLS alternative proof — single stage

This section provides one alternative proof of dWOLS. We refer to this proof as a direct proof of dWOLS because it employs the algebra of partitioned regression directly.

We have an i.i.d. model: X is a vector of p covariates (Note that the covariates X in appendix sections are not in boldface, but are nevertheless the same as the covariates in boldface $\boldsymbol{X}$ in the body of the thesis); A is the treatment assignment, taking values 1 or 0; Y^A is the potential outcome under treatment A (same as the notation $Y^*(A)$ in the body of the thesis); the values of these are independent from individual to individual; realized values are x, a and y^a; we can think of x being determined first, then a conditional on x, then y^a conditional on x and a. For a straightforward consistency argument, we assume not only that X_i for individuals are i.i.d. but also that the components have finite variance. We assume that, conditional on X and A, the Y^A have constant finite variance σ^2. We write

$$\mathbb{P}(A = 1 \mid x) = \pi(x);$$

for each x, $\pi(x)$ is between 0 and 1.

For individual i, where $i = 1, 2, \ldots, n$ we can write

$$\mathbb{E}[Y_i^a \mid x_i, a_i] = f(x_i^\beta, \beta) + a_i x_i^\psi \psi,$$

where x_i^β and x_i^ψ are the row vectors of covariate values for the i-th individual.

Suppose the values of x and a are determined for all individuals, and let n_1 and n_0 be the numbers of individuals with $a_i = 1$ and $a_i = 0$, respectively. Let their x and y matrices/vectors be $(x_1^\beta, x_1^\psi, y_1^1)$, dimensions $(n_1 \times p_\beta, n_1 \times p_\psi, n_1 \times 1)$ and $(x_0^\beta, x_0^\psi, y_0^0)$, dimensions $(n_0 \times p_\beta, n_0 \times p_\psi, n_0 \times 1)$,

respectively. Note that for the unbiased estimator of ψ to hold, X need not be random.

If we were applying ordinary least squares, we would use the following system of normal equations:

$$(x_1^\psi)^\top (y_1^1 - x_1^\beta \beta - x_1^\psi \psi) = 0;$$
$$(x_1^\beta)^\top (y_1^1 - x_1^\beta \beta - x_1^\psi \psi) + (x_0^\beta)^\top (y_0^0 - x_0^\beta \beta) = 0.$$

However, if we are applying weighted least squares where the weights are dependent on a and x, we would write:

$$(x_1^\psi)^\top W_1 (y_1^1 - x_1^\beta \beta - x_1^\psi \psi) = 0; \tag{B.1}$$

and

$$(x_1^\beta)^\top W_1 (y_1^1 - x_1^\beta \beta - x_1^\psi \psi) + (x_0^\beta)^\top W_0 (y_0^0 - x_0^\beta \beta) = 0. \tag{B.2}$$

where W_1 is the $n_1 \times n_1$ diagonal matrix of the weights $w(1, x_i)$ for the individuals with $a_i = 1$, and W_0 is the $n_0 \times n_0$ diagonal matrix of the weights $w(0, x_i)$ for the individuals with $a_i = 0$.

From (B.1),

$$\hat{\psi} = [(x_1^\psi)^\top W_1 x_i^\psi]^{-1} (x_1^\psi)^\top W_1 (y_1^1 - x_1^\beta \beta). \tag{B.3}$$

Plugging $\hat{\psi}$ into (B.2) gives

$$(x_1^\beta)^\top W_1 (y_1^1 - x_1^\beta \beta - P_1^\psi W_1 (y_1^1 - x_1^\beta \beta)) + (x_0^\beta)^\top W_0 (y_0^0 - x_0^\beta \beta) = 0,$$

where

$$P_1^\psi = x_1^\psi [(x_1^\psi)^\top W_1 x_i^\psi]^{-1} (x_1^\psi)^\top,$$

of dimension $n_1 \times n_1$. Solving for $\hat{\beta}$ gives

$$\hat{\beta} = [(x_1^\beta)^\top W_1 (I - P_1^\psi W_1) x_1^\beta + (x_0^\beta)^\top W_0 x_0^\beta]^{-1} [(x_1^\beta)^\top W_1 (I - P_1^\psi W_1) y_1^1 + (x_0^\beta)^\top W_0 y_0^0]. \tag{B.4}$$

Let the "denominator" of (B.4) be denoted by D.

Taking expectation in (B.4) with respect to the conditional distribution of Y^a, given $X = x$ and $A = a$, and noting that $(I - P_1^\psi W_1) x_1^\psi$ is the 0 vector, gives

$$\mathbb{E}(\hat{\beta} \mid x, a) = D^{-1} [(x_1^\beta)^\top W_1 (I - P_1^\psi W_1) f(x_1^\beta, \beta) + (x_0^\beta)^\top W_0 f(x_0^\beta, \beta)]. \tag{B.5}$$

This is not equal to β unless $f(x_1^\beta, \beta)$ is $x_1^\beta \beta$ and $f(x_0^\beta, \beta)$ is $x_0^\beta \beta$.

Taking expectation in (B.3) with respect to the conditional distribution of Y^a, given $X = x$ and $A = a$, gives

$$\mathbb{E}(\hat{\psi} \mid x, a) = \psi + [(x_1^{\psi})^{\top} W_1 x_1^{\psi}]^{-1} (x_1^{\psi})^{\top} W_1 (f(x_1^{\beta}, \beta) - x_1^{\beta} D^{-1} C), \tag{B.6}$$

where C is the "numerator" of (B.5). This is equal to ψ if $f(x_1^{\beta}, \beta)$ is $x_1^{\beta}\beta$ and $f(x_0^{\beta}, \beta)$ is $x_0^{\beta}\beta$, or if the weights can be defined in such a way as to make the second term of (B.6) disappear when expectation with respect to the distribution of A, given x, is taken.

A vector z_1 which has the elements of the vector z for the individuals with $a_i = 1$ and 0 entries otherwise has expectation

$$(z_1\pi(x_1), z_2\pi(x_2), \ldots, z_n\pi(x_n))^{\top}.$$

Similarly, a vector z_0 which has the elements of the vector z for the individuals with $a_i = 0$ and 0 entries otherwise has expectation

$$(z_1(1 - \pi(x_1)), z_2(1 - \pi(x_2)), \ldots, z_n(1 - \pi(x_n)))^{\top}.$$

Using these facts, one can show that, with respect to the distribution of A given x:

$$\mathbb{E}((x_0^{\beta})^{\top} W_0 x_0^{\beta}) = (x^{\beta})^{\top} \mathcal{W}_0 (I - \Pi) x^{\beta},$$

where Π is the $n \times n$ diagonal matrix of the $\pi(x_i)$ values and $\mathcal{W}_0$ is the $n \times n$ diagonal matrix of the $w(0, x_i)$ values.

$$\mathbb{E}((x_1^{\beta})^{\top} W_1 x_1^{\beta}) = (x^{\beta})^{\top} \mathcal{W}_1 \Pi x^{\beta},$$

where $\mathcal{W}_1$ is the $n \times n$ diagonal matrix of the $w(1, x_i)$ values. Similarly,

$$\mathbb{E}((x_0^{\beta})^{\top} W_0 f(x_0^{\beta}, \beta)) = (x^{\beta})^{\top} \mathcal{W}_0 (I - \Pi) f(x^{\beta}, \beta);$$

$$\mathbb{E}((x_1^{\beta})^{\top} W_1 f(x_1^{\beta}, \beta)) = (x^{\beta})^{\top} \mathcal{W}_1 \Pi f(x^{\beta}, \beta);$$

$$\mathbb{E}((x_1^{\psi})^{\top} W_1 x_1^{\psi}) = (x^{\psi})^{\top} \mathcal{W}_1 \Pi x^{\psi};$$

$$\mathbb{E}((x_1^{\psi})^{\top} W_1 f(x_1^{\beta}, \beta)) = (x^{\psi})^{\top} \mathcal{W}_1 \Pi f(x^{\beta}, \beta);$$

$$\mathbb{E}((x_1^\beta)^\top W_1 x_1^\psi) = (x^\beta)^\top \mathcal{W}_1 \Pi x^\psi.$$

$$\mathbb{E}((x_1^\beta)^\top W_1 P_1^\psi W_1 x_1^\beta) = \mathbb{E}((x_1^\beta)^\top W_1 x_1^\psi [(x_1^\psi)^\top W_1 x_1^\psi]^{-1} (x_1^\psi)^\top W_1 x_1^\beta),$$

which is approximately

$$(x^\beta)^\top \mathcal{W}_1 \Pi x^\psi [(x^\psi)^\top \mathcal{W}_1 \Pi x^\psi]^{-1} (x^\psi)^\top \mathcal{W}_1 \Pi x^\beta.$$

$$\mathbb{E}((x_1^\beta)^\top W_1 P_1^\psi W_1 f(x_1^\beta, \beta)) = \mathbb{E}((x_1^\beta)^\top W_1 x_1^\psi [(x_1^\psi)^\top W_1 x_1^\psi]^{-1} (x_1^\psi)^\top W_1 f(x_1^\beta, \beta)),$$

which is approximately

$$(x^\beta)^\top \mathcal{W}_1 \Pi \mathcal{P}^\psi \mathcal{W}_1 \Pi f(x^\beta, \beta).$$

where

$$\mathcal{P}^\psi = x^\psi [(x^\psi)^\top \mathcal{W}_1 \Pi x^\psi]^{-1} (x^\psi)^\top.$$

Now suppose $\mathcal{W}_0(I - \Pi)$ is identically equal to $\mathcal{W}_1 \Pi = R(x)$ (a diagonal matrix) for all x. Then

$$\mathcal{P}^\psi = x^\psi [(x^\psi)^\top R(x) x^\psi]^{-1} (x^\psi)^\top,$$

and from (B.6), the expectation of $\hat{\psi}$, conditional on x, is approximately ψ plus

$$[(x^\psi)^\top R(x) x^\psi]^{-1} (x^\psi)^\top R(x) (f(x^\beta, \beta) - x^\beta \mathcal{D}^{-1} C),$$

where the difference between the true expectation and its approximation will be of order $O_p(1/n)$; and where $\mathcal{D} = (x^\beta)^\top R(x)(I - \mathcal{P}^\psi R(x)) x^\beta + (x^\beta)^\top R(x) x^\beta$, and $C = (x^\beta)^\top R(x)(I - \mathcal{P}^\psi R(x)) f(x^\beta, \beta) + (x^\beta)^\top R(x) f(x^\beta, \beta)$.

The factor $(x^\psi)^\top R(x)(f(x^\beta, \beta) - x^\beta \mathcal{D}^{-1} C)$ is equal to

$$\{\mathcal{B}(x)(I - x^\beta [\mathcal{A}(x) x^\beta]^{-1} \mathcal{A}(x))\} f(x^\beta, \beta) \tag{B.7}$$

where $\mathcal{B}(x) = (x^\psi)^\top R(x)$ and $\mathcal{A}(x) = (x^\beta)^\top R(x)(2I - \mathcal{P}^\psi R(x))$, and we denote expression (B.7) as $\mathcal{E}$, that is, $\mathcal{E} := \{\mathcal{B}(x)(I - x^\beta [\mathcal{A}(x) x^\beta]^{-1} \mathcal{A}(x))\} f(x^\beta, \beta)$.

Now suppose $x^\beta = x^\psi = x$. Then $P^\psi = x(x^\top R(x)x)^{-1}x^\top$ and $\mathcal{A}(x) = 2x^\top R(x) - x^\top R(x) = x^\top R(x)$. The part of $\mathcal{E}$ (B.7) in brace brackets is

$$x^\top R(x) - x^\top R(x)x[x^\top R(x)x]^{-1}x^\top R(x) = 0.$$

What if it is not true that $x^\beta = x^\psi = x$? Consider the case where x^ψ is a subset of x^β, and without much loss of generality, suppose that the rest of x^β is orthogonal in the weighted sense to x^ψ, as would be approximately true if, in the model for x, x^ψ and the rest of x^β were uncorrelated in an appropriately weighted sense.

Accordingly, suppose $x^\beta = (x^\psi \mid z)$, where z is of dimension $n \times (p^\beta - p^\psi)$ and is orthogonal in the weighted sense to x^ψ, so that

$$\mathbb{E}[z^\top R(x)x^\psi] = \mathbf{0}, \tag{B.8}$$

the right-hand side being a $(p^\beta - p^\psi) \times p^\psi$ matrix of zeros.

Now we have Lemma:

Lemma B.1. *Suppose x^ψ is a subset of x^β, and $x^\beta = (x^\psi \mid z)$ where z is of dimension $n \times (p^\beta - p^\psi)$ and is orthogonal in the weighted sense to x^ψ, so that $\mathbb{E}[z^\top R(x)x^\psi] = \mathbf{0}$. Then expression $\mathcal{E}$ (B.7)*

$$\{\mathcal{B}(x)(I - x^\beta[\mathcal{A}(x)x^\beta]^{-1}\mathcal{A}(x))\}f(x^\beta, \beta)$$

is $\mathbf{0}$.

Proof. First, for

$$\mathcal{A}(x) = \begin{pmatrix} (x^\psi)^\top \\ z^\top \end{pmatrix} R(x)(2I - P^\psi R(x)),$$

we use the specific form of $\mathcal{A}$, which is

$$\mathcal{A}(x) = 2\begin{pmatrix} (x^\psi)^\top \\ z^\top \end{pmatrix} R(x) - \begin{pmatrix} (x^\psi)^\top R(x) \\ z^\top R(x)P^\psi R(x) \end{pmatrix}.$$

Because of (B.8), $z^\top R(x)P^\psi = 0$, and therefore

$$\mathcal{A}(x) = \begin{pmatrix} (x^\psi)^\top R(x) \\ 2z^\top R(x) \end{pmatrix},$$

and

$$\mathcal{A}(x)x^{\beta} = \begin{pmatrix} (x^{\psi})^{\top}R(x)x^{\psi} & (x^{\psi})^{\top}R(x)z \\ 2z^{\top}R(x)x^{\psi} & 2z^{\top}R(x)z \end{pmatrix} = \begin{pmatrix} (x^{\psi})^{\top}R(x)x^{\psi} & 0 \\ 0 & 2z^{\top}R(x)z \end{pmatrix}.$$

It follows that

$$\begin{aligned} &\mathcal{B}(x)x^{\beta}[\mathcal{A}(x)x^{\beta}]^{-1}\mathcal{A}(x) = \\ &(x^{\psi})^{\top}R(x)(x^{\psi} \mid z) \begin{pmatrix} [(x^{\psi})^{\top}R(x)x^{\psi}]^{-1} & 0 \\ 0 & [2z^{\top}R(x)z]^{-1} \end{pmatrix} \begin{pmatrix} (x^{\psi})^{\top}R(x) \\ 2z^{\top}R(x) \end{pmatrix}. \end{aligned}$$

Again using (B.8), it can be shown that the right-hand side reduces to $\mathcal{B}(x)$, and hence that the expression in $\mathcal{E}$ (B.7) is $\mathbf{0}$. □

Therefore, we have shown that the expression in $\mathcal{E}$ (B.7) is $\mathbf{0}$ under a couple of assumptions, namely that $x^{\beta} = x^{\psi} = x$ or more generally that x^{ψ} is a subset of x^{β}, and the rest of x^{β} is orthogonal to x^{ψ}.

Appendix C

Study the double robustness of dWOLS from estimation equation perspective

In this section, we demonstrate the double robustness of dWOLS from an estimation equation perspective in two different cases.

Let us consider the true model: $\mathbb{E}[Y|A, X] = f(X^\beta) + \gamma(A, X^\psi; \boldsymbol{\psi})$, where $Y \in \mathbb{R}$, and $X \in \mathbb{R}^p$ is a vector of p covariates (again, the covariates X in appendix sections are not in boldface, but are nevertheless the same as the covariates in boldface $\boldsymbol{X}$ in the body of the thesis); $X^\beta \in \mathbb{R}^{p^\beta}$ and $X^\psi \in \mathbb{R}^{p^\psi}$ are two (potentially identical) subsets of the variables contained in X; $A \in \{0, 1\}$ is the treatment assignment. Their realized values are y, x, x^β , x^ψ and a, and we write the propensity score as $\mathbb{P}(A = 1 \mid x) = \pi(x)$. We assume that $\gamma(A, X^\psi; \boldsymbol{\psi})$ is correctly specified, and set $\gamma(A, X^\psi; \boldsymbol{\psi}) = \boldsymbol{\psi}^\top A X^\psi$.

Given the posited outcome regression model $Q(X, A; \boldsymbol{\beta}, \boldsymbol{\psi}) = \boldsymbol{\beta}^\top X^\beta + \boldsymbol{\psi}^\top A X^\psi$, the weighted OLS estimator for $(\boldsymbol{\beta}^\top, \boldsymbol{\psi}^\top)$ is obtained by solving the system of estimating equations: $\sum_i^n U_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}}; A_i, X_i) = \mathbf{0}$, that is,

$$\sum_i^n \begin{pmatrix} X_i^\beta \\ A_i X_i^\psi \end{pmatrix} w(A_i, X_i) \left(Y_i - \boldsymbol{\psi}^\top A_i X_i^\psi - \boldsymbol{\beta}^\top X_i^\beta \right) = \sum_i^n \begin{bmatrix} U_{1i}(\boldsymbol{\beta}, \psi; A_i, X_i) \\ U_{2i}(\boldsymbol{\beta}, \psi; A_i, X_i) \end{bmatrix} = \mathbf{0}. \tag{C.1}$$

C.0.1 Case 1 of dWOLS

We first assume that $X^\beta = X^\psi = X$; that is, the covariates in the treatment free and blip part are the same. Then, we assume that for each i independently, X_i for individuals is i.i.d. and is generated

first; thereafter, A_i followed by Y_i. On the one hand, we consider

$$\sum_i^n (U_{1i} - U_{2i}) = \sum_i^n (1 - A_i) X_i w(A_i, X_i) \left(Y_i - \boldsymbol{\psi}^\top A_i X_i - \boldsymbol{\beta}^\top X_i\right) \tag{C.2}$$

$$= \sum_i^n (1 - A_i) X_i w(A_i, X_i) \left(Y_i - \boldsymbol{\beta}^\top X_i\right) = \mathbf{0}, \tag{C.3}$$

where the second equality follows because the only non-zero terms (in equation (C.2)) will be those for which $A_i = 0$. Thus, equation (C.3) (or $\sum_i^n (U_{1i} - U_{2i})$) does not depend on $\boldsymbol{\psi}$. We note that $\sum_i^n (U_{1i} - U_{2i}) = \mathbf{0}$ can be solved for $\boldsymbol{\beta}$, and its solution is

$$\hat{\boldsymbol{\beta}} = (\sum_i^n (1 - A_i) w(A_i, X_i) X_i X_i^\top)^{-1} \sum_i^n (1 - A_i) w(A_i, X_i) X_i Y_i.$$

The expectation of $\sum_i^n (U_{1i} - U_{2i})$ conditional on $(A_1, \ldots A_n)$ and $(X_1, \ldots X_n)$, that is, $\sum_i^n (1 - A_i) X_i w(A_i, X_i) \left(f(X_i) - \boldsymbol{\beta}^\top X_i\right)$, is not zero unless the true treatment-free model $f(X; \boldsymbol{\beta})$ is linear in X with true coefficient $\boldsymbol{\beta}$. However, the expectation of $\sum_i^n (U_{1i} - U_{2i})$ conditional on $(X_1, \ldots X_n)$ is $\sum_i^n (1 - \pi(X_i)) X_i w(0, X_i) \left(f(X_i) - \boldsymbol{\beta}^\top X_i\right)$, and if the unconditional expectation of $\sum_i^n (1 - \pi(X_i)) X_i w(0, X_i) \left(f(X_i) - \boldsymbol{\beta}^\top X_i\right)$, $n\mathbb{E}\left[(1 - \pi(X)) X w(0, X) \left(f(X) - \boldsymbol{\beta}^\top X\right)\right]$ is $\mathbf{0}$ for $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, then, according to the large sample theory approximation, $\hat{\boldsymbol{\beta}}$ tends to $\boldsymbol{\beta}^*$ as $n \to \infty$. White [1982].

On the other hand, we consider $\sum_i^n U_{2i} = \sum_i^n A_i X_i w(A_i, X_i) \left(Y_i - \boldsymbol{\psi}^\top A_i X_i - \boldsymbol{\beta}^\top X_i\right)$, where the only non-zero terms are those for which $A_i = 1$, then $\hat{\boldsymbol{\psi}}$ can be solved in terms of $\hat{\boldsymbol{\beta}}$ from $\sum_i^n U_{2i} = \mathbf{0}$. We particularly note that, from the dWOLS weights criterion $(1 - \pi(X)) w(0, X) = \pi(X) w(1, X)$, the expectation of $\sum_i^n U_{2i}$ conditional on $(X_1, \ldots X_n)$, that is, $\sum_i^n \pi(X_i) X_i w(1, X_i) \left(f(X_i) - \boldsymbol{\beta}^\top X_i\right)$, equals $\sum_i^n (1 - \pi(X_i)) X_i w(0, X_i) \left(f(X_i) - \boldsymbol{\beta}^\top X_i\right)$, where the unconditional expectation of $\sum_i^n (1 - \pi(X_i)) X_i w(0, X_i) \left(f(X_i) - \boldsymbol{\beta}^\top X_i\right)$ is $\mathbf{0}$ for $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ and for any $\boldsymbol{\psi}$. The equation $\sum_i^n U_{2i} = \mathbf{0}$ with $\boldsymbol{\beta}$ set equal to $\hat{\boldsymbol{\beta}}$ will approach the equation $\sum_i^n U_{2i} = \mathbf{0}$ with $\boldsymbol{\beta}$ set equal to $\boldsymbol{\beta}^*$ as $n \to \infty$.

We find estimators $(\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\psi}}^\top)$ through solving the system of estimating equations $\sum_i^n U_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}}; A_i, X_i) = \mathbf{0}$. Supposing the treatment-free model $f(X^\beta; \boldsymbol{\beta})$ is mis-specified, we now show that $(\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\psi}}^\top)$ solving the estimating equations (C.1) converges in probability to $(\boldsymbol{\beta}^{*\top}, \boldsymbol{\psi}^\top)$, where $\boldsymbol{\beta}^*$ is the solution to (unconditional expectation) $\mathbb{E}\left[(1 - \pi(X)) X w(0, X) \left(f(X) - \boldsymbol{\beta}^{*\top} X\right)\right] = \mathbf{0}$. Because $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}^*$, then from the theory of M-estimation (Huber [2004], Geer and van de Geer [2000], and Tsiatis [2019]), under regularity

conditions, this follows if the system of estimating functions corresponding to estimating equations (C.1) has mean zero when evaluated at $(\boldsymbol{\beta}^{*\top}, \boldsymbol{\psi}^\top)$; that is,

$$\mathbb{E}[U(\boldsymbol{\beta}^*, \boldsymbol{\psi}; A, X)] = \mathbb{E}\left[\begin{pmatrix} X \\ AX \end{pmatrix} w(A,X)\left(Y - \boldsymbol{\beta}^{*\top}X - \boldsymbol{\psi}^\top AX\right)\right] = \mathbb{E}\begin{bmatrix} U_1(\boldsymbol{\beta}^*, \psi; A, X) \\ U_2(\boldsymbol{\beta}^*, \psi; A, X) \end{bmatrix} = \mathbf{0};$$

equivalently,

$$\mathbb{E}\begin{bmatrix} U_1(\boldsymbol{\beta}^*, \psi; A, X) - U_2(\boldsymbol{\beta}^*, \psi; A, X) \\ U_2(\boldsymbol{\beta}^*, \psi; A, X) \end{bmatrix} = \mathbb{E}\left[\begin{pmatrix} 1-A \\ A \end{pmatrix} Xw(A,X)\left(Y - \boldsymbol{\beta}^{*\top}X - \boldsymbol{\psi}^\top AX\right)\right] = \mathbf{0}. \tag{C.4}$$

Applying iterated conditional expectations conditioning on (A, X), $\mathbb{E}\left[\begin{pmatrix} 1-A \\ A \end{pmatrix} Xw(A,X)\left(Y - \boldsymbol{\beta}^{*\top}X - \boldsymbol{\psi}^\top AX\right)\right]$ in (C.4) is $\mathbb{E}\left[\begin{pmatrix} 1-A \\ A \end{pmatrix} Xw(A,X)\left(f(X) - \boldsymbol{\beta}^{*\top}X\right)\right]$. Then, taking conditional expectation with respect to A given X, gives

$$\mathbb{E}\begin{bmatrix} (1-\pi(X))Xw(0,X)\left(f(X) - \boldsymbol{\beta}^{*\top}X\right) \\ \pi(X)Xw(1,X)\left(f(X) - \boldsymbol{\beta}^{*\top}X\right) \end{bmatrix}. \tag{C.5}$$

By definition of $\boldsymbol{\beta}^*$, the first expectation in (C.5) is zero; then the second expectation in (C.5) is also zero if the dWOLS weights criteria ($(1-\pi(X))w(0,X) = \pi(X)w(1,X)$) are fulfilled and the propensity score is correctly specified.

In the case of $X = X^\beta = X^\psi$, we conclude that $\hat{\boldsymbol{\psi}}$ solving the system of estimating equations (C.1) is a consistent estimator for the true value $\boldsymbol{\psi}$, if either the treatment-free model or treatment model is correctly specified.

C.0.2 Case 2 of dWOLS

Suppose X^ψ is a subset of $X = X^\beta$, and without loss of generality that $X = X^\beta = (X^{\psi\top}, Z^\top)^\top$, where Z is a vector with length $(p^\beta - p^\psi)$ that is orthogonal to X^ψ in the weighted sense that $\mathbb{E}[ZR(X)X^{\psi\top}] = \mathbf{0}_{(p^\beta - p^\psi)\times p^\psi}$ where $R(X) := (1-\pi(X))w(0,X) = \pi(X)w(1,X)$. Let the treatment-free part of the expectation of Y conditional on X be $f(X)$, and let the working model for this quantity be $\boldsymbol{\beta}^\top X = \boldsymbol{\beta}^\top X^\beta$. Let the blip function be $\gamma(A, X^\psi; \boldsymbol{\psi}) = A(\boldsymbol{\psi}^\top X^\psi + \boldsymbol{\delta}^\top Z)$,

where the true value of $\boldsymbol{\delta}^\top$ is known to be $\mathbf{0}$. Then,

$$\sum_i^n U_{1i} = \sum_i^n \begin{pmatrix} X_i^\psi \\ Z_i \end{pmatrix} w(A_i, X_i) \left(Y_i - \boldsymbol{\psi}^\top A_i X_i^\psi - \boldsymbol{\beta}^\top X_i\right) = \mathbf{0}, \tag{C.6}$$

$$\sum_i^n U_{2i} = \sum_i^n A_i X_i^\psi w(A_i, X_i) \left(Y_i - \boldsymbol{\psi}^\top A_i X_i^\psi - \boldsymbol{\beta}^\top X_i\right) = \mathbf{0}, \tag{C.7}$$

is the system to be solved to find $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\psi}}$.

If we replace (C.7) by

$$\sum_i^n U_{2i}^\dagger = \sum_i^n A_i \begin{pmatrix} X_i^\psi \\ Z_i \end{pmatrix} w(A_i, X_i) \left(Y_i - \boldsymbol{\psi}^\top A_i X_i^\psi - \boldsymbol{\delta}^\top A_i Z_i - \boldsymbol{\beta}^\top X_i\right) = \mathbf{0} \tag{C.8}$$

and (C.6) by

$$\sum_i^n U_{1i}^\dagger = \sum_i^n \begin{pmatrix} X_i^\psi \\ Z_i \end{pmatrix} w(A_i, X_i) \left(Y_i - \boldsymbol{\psi}^\top A_i X_i^\psi - \boldsymbol{\delta}^\top A_i Z_i - \boldsymbol{\beta}^\top X_i\right) = \mathbf{0} \tag{C.9}$$

the system (C.8), (C.9) is equivalent to (C.9) and

$$\sum_i^n (U_{1i}^\dagger - U_{2i}^\dagger) = \sum_i^n (1 - A_i) \begin{pmatrix} X_i^\psi \\ Z_i \end{pmatrix} w(A_i, X_i) \left(Y_i - \boldsymbol{\psi}^\top A_i X_i^\psi - \boldsymbol{\delta}^\top A_i Z_i - \boldsymbol{\beta}^\top X_i\right) = \mathbf{0}. \tag{C.10}$$

Now $\sum_i^n (U_{1i}^\dagger - U_{2i}^\dagger) = \sum_i^n (1 - A_i) \begin{pmatrix} X_i^\psi \\ Z_i \end{pmatrix} w(A_i, X_i) \left(Y_i - \boldsymbol{\beta}^\top X_i\right) = \mathbf{0}$ is solvable for $\hat{\boldsymbol{\beta}}$. It is unbiased for $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ where $\boldsymbol{\beta}^*$ solves $\mathbb{E}[R(X)X(f(X) - \boldsymbol{\beta}^\top X)] = \mathbf{0}$. On the other hand, $\sum_i^n U_{2i}^\dagger = \mathbf{0}$, with $\boldsymbol{\beta}$ replaced by $\hat{\boldsymbol{\beta}}$, is solvable for $\hat{\boldsymbol{\psi}}^\dagger$, $\hat{\boldsymbol{\delta}}^\dagger$. It is unbiased for any $\boldsymbol{\psi}$, $\boldsymbol{\delta}$ as long as $\boldsymbol{\beta} = \boldsymbol{\beta}^*$. Thus, we have $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}}^\dagger, \hat{\boldsymbol{\delta}}^\dagger) \xrightarrow{p} (\boldsymbol{\beta}^*, \boldsymbol{\psi}, \boldsymbol{\delta})$, where $\boldsymbol{\delta} = \mathbf{0}$. Now we just need to show that $\hat{\boldsymbol{\psi}}^\dagger$ and $\hat{\boldsymbol{\psi}}$, respectively, obtained from $\sum_i^n U_{2i}^\dagger = \mathbf{0}$ and $\sum_i^n U_{2i} = \mathbf{0}$, are asymptotically equivalent. To see this, we write

$$\hat{\boldsymbol{\psi}} = \left[\sum_i^n A_i w(A_i, X_i) X_i^\psi X_i^{\psi\top}\right]^{-1} \left[\sum_i^n A_i w(A_i, X_i) X_i^\psi (Y_i - \hat{\boldsymbol{\beta}}^\top X_i)\right],$$

and

$$\begin{pmatrix} \hat{\boldsymbol{\psi}}^{\dagger} \\ \hat{\boldsymbol{\delta}}^{\dagger} \end{pmatrix} = \left[\sum_{i}^{n} A_i w(A_i, X_i) \begin{pmatrix} X_i^{\psi} \\ Z_i \end{pmatrix} X_i^{\psi\top} \right]^{-1} \left[\sum_{i}^{n} A_i w(A_i, X_i) \begin{pmatrix} X_i^{\psi} \\ Z_i \end{pmatrix} (Y_i - \hat{\boldsymbol{\beta}}^{\top} X_i) \right].$$

Dividing top and bottom by n in each case and letting $n \to \infty$, the weak law of large numbers and continuous mapping theorem imply that the right-hand sides approach (respectively)

$$\left(\mathbb{E}\left[R(X) X^{\psi} X^{\psi\top} \right] \right)^{-1} \mathbb{E}\left[R(X) X^{\psi} (Y - \boldsymbol{\beta}^{*\top} X) \right],$$

and

$$\begin{aligned} &\begin{pmatrix} \mathbb{E}\left[R(X) X^{\psi} X^{\psi\top} \right] & \mathbf{0} \\ \mathbf{0} & \mathbb{E}\left[R(X) Z Z^{\top} \right] \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{E}\left[R(X) X^{\psi} (Y - \boldsymbol{\beta}^{*\top} X) \right] \\ \mathbb{E}\left[R(X) Z (Y - \boldsymbol{\beta}^{*\top} X) \right] \end{pmatrix} \\ &= \begin{pmatrix} \left(\mathbb{E}\left[R(X) X^{\psi} X^{\psi\top} \right] \right)^{-1} \mathbb{E}\left[R(X) X^{\psi} (Y - \boldsymbol{\beta}^{*\top} X) \right] \\ \left(\mathbb{E}\left[R(X) Z Z^{\top} \right] \right)^{-1} \mathbb{E}\left[R(X) Z (Y - \boldsymbol{\beta}^{*\top} X) \right] \end{pmatrix}. \end{aligned}$$

Therefore, $\hat{\boldsymbol{\psi}}^{\dagger}$ and $\hat{\boldsymbol{\psi}}$ are asymptotically equivalent, and this finishes the whole proof.

Appendix D

Proof of dWOLS optimal weights

In this section, we will prove the theorem (Theorem 2.1) of optimal dWOLS balancing weights which is presented at the end of Chapter 2.

Denote $X^{(0)} = X(I - A)$, which is a $p \times n$ matrix, where X is a $p \times n$ matrix and $I - A$ is a $n \times n$ diagonal matrix; $X^{(1)} = XA$ with a $n \times n$ diagonal matrix A; $Y^{(w)} = w(A, X)Y$ with $n \times n$ diagonal matrix $w(A, X)$; and $X^{(0w)} = X(I - A)w(A, X)$ and $X^{(1w)} = XA(A, X)$.

Denote $V_1 = U_1 - U_2$ and $V_2 = U_2$ where U_1 and U_2 are matrix versions of U_1 and U_2 in Appendix C. Then we have:

$$V(\boldsymbol{\beta}, \boldsymbol{\psi}; A, X) = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{pmatrix} X^{(0)} \\ X^{(1)} \end{pmatrix} w(A, X) \left(Y - AX^\top \boldsymbol{\psi} - X^\top \boldsymbol{\beta}\right) = \begin{pmatrix} 0 \\ 0 \end{pmatrix};$$

or

$$\begin{pmatrix} X^{(0)} \\ X^{(1)} \end{pmatrix} \left(Y^{(w)} - X^{(1w)}\boldsymbol{\psi} - \left[X^{(0w)} + X^{(1w)}\right]^\top \boldsymbol{\beta}\right) = \begin{pmatrix} X^{(0)} \\ X^{(1)} \end{pmatrix} \epsilon^{(w)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

where $\epsilon^{(w)} = Y^{(w)} - X^{(1w)}\boldsymbol{\psi} - \left[X^{(0w)} + X^{(1w)}\right]^\top \boldsymbol{\beta}$.

Note that $X^{(0)}X^{(1w)} = 0$ *and* $X^{(1)}X^{(0w)} = 0$. Then

$$V_1 = X^{(0)} \left(Y^{(w)} - X^{(0w)\top}\boldsymbol{\beta}\right) = 0,$$

and we can solve $\hat{\boldsymbol{\beta}} = (X^{(0)}X^{(0w)\top})^{-1}X^{(0)}Y^{(w)}$. Also,

$$X^{(1)} \left(Y^{(w)} - X^{(1w)\top}\boldsymbol{\psi} - X^{(1w)\top}\boldsymbol{\beta}\right) = 0,$$

and we can solve $\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\psi}} = \left(X^{(1)}X^{(1w)\top}\right)^{-1}\left(X^{(0)}Y^{(w)}\right)$. Thus we have

$$\hat{\boldsymbol{\psi}} - \boldsymbol{\psi} = \left[\left(X^{(1)}X^{(1w)\top}\right)X^{(1)} - X^{(0)}X^{(0w)\top}X^{(0)}\right]\varepsilon^{(w)}.$$

Further, by law of total expectation or the law of iterated expectations, we have

$$\mathrm{Var}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) = \mathbb{E}\left(\mathrm{Var}[(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \mid A, X]\right),$$

where $\mathrm{Var}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi} \mid A, X) = QW^2(A, X)Q^\top\sigma_\varepsilon^2$ and $Q = \left[\left(X^{(1)}X^{(1w)\top}\right)X^{(1)} - X^{(0)}X^{(0w)\top}X^{(0)}\right]$.

Let $\mathbb{E}\left[X^{(0w)} \mid X\right] = X((1-\pi(X))w(0, X) = XR(X)$ and $\mathbb{E}\left[X^{(1w)} \mid X\right] = XR(X)$, then we have

$$\begin{aligned}
QW^2(A, X)Q^\top/\sigma_\varepsilon^2 = &\left[\left(X^{(1)}X^{(1w)\top}\right)^{-1}X^{(1)}W^2(A, X)X^{(1)\top}\left(X^{(1)}X^{(1w)\top}\right)^{-1}\right] \\
&- \left[\left(X^{(1)}X^{(1w)\top}\right)^{-1}X^{(1)}W^2(A, X)X^{(0)\top}\left(X^{(0)}X^{(0w)\top}\right)^{-1}\right] \\
&- \left[\left(X^{(0)}X^{(0w)\top}\right)^{-1}X^{(0)}W^2(A, X)X^{(1)\top}\left(X^{(1)}X^{(1w)\top}\right)^{-1}\right] \\
&+ \left[\left(X^{(0)}X^{(0w)\top}\right)^{-1}X^{(0)}W^2(A, X)X^{(0)\top}\left(X^{(0)}X^{(0w)\top}\right)^{-1}\right].
\end{aligned}$$

Note that $X^{(1)}W^2(A, X)X^{(0)\top} = XAW^2(A, X)(I - A)X^\top = 0$, so the middle terms are 0 and we have:

$$\begin{aligned}
QW^2(A, X)Q^\top/\sigma_\varepsilon^2 = &\left[\left(X^{(1)}X^{(1w)\top}\right)^{-1}X^{(1)}W^2(A, X)X^{(1)\top}\left(X^{(1)}X^{(1w)\top}\right)^{-1}\right] \\
&+ \left[\left(X^{(0)}X^{(0w)\top}\right)^{-1}X^{(0)}W^2(A, X)X^{(0)\top}\left(X^{(0)}X^{(0w)\top}\right)^{-1}\right]. \qquad \text{(D.1)}
\end{aligned}$$

To evaluate two terms in (D.1) in the limiting case, note for example that $X^{(1)}W^2(A, X)X^{(1)\top}$ is $XAW^2(A, X)AX^\top$, a $p \times p$ matrix whose $l-m$th entry is a sum $\sum_{i=1}^n X_i^{(l)}A_iW^2\left(A_i, X_i\right)A_iX_i^{(m)}$, and similarly for the other terms in (D.1). If the $p \times 1$ vectors X_i are i.i.d., then by the LLN we

have a limiting result

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^{(l)} A_i W^2\left(A_i, X_i\right) A_i X_i^{(m)} \mid X_i\right) = \frac{1}{n}\sum_{i}^{n} X_i^{(l)} R\left(X_i\right) W\left(1, X_i\right) X_i^{(m)}, \tag{D.2}$$

converges to $l-m$th element of $\mathbb{E}\left[XR(X)W(1,X)X^{\top}\right]$, where $R(X)$ is $diag(R(X_i))$, $W(1,X)$ is $diag(W(1,X_i))$ and similarly for the other factors in (D.1). Therefore, it can be shown that in the limit the $p \times p$ matrix $nQW^2(A,X)Q^{\top}/\sigma_{\epsilon}^2$ converges to

$$\mathbb{E}\left(X_1 R\left(X_1\right) X_1^{\top}\right)^{-1} \mathbb{E}\left(X_1\left[W\left(1, X_1\right) + W(0, X_1)\right] R(X_1) X_1\right) \mathbb{E}\left(X_1 R\left(X_1\right) X_1^{\top}\right)^{-1}, \tag{D.3}$$

where X_1 is the $p \times 1$ matrix of covariates for the first subject.

To optimize this, we note that $W\left(1, X_1\right) + W\left(0, X_1\right)$ is equal to $R\left(X_1\right) / [\pi\left(X_1\right)\left(1-\pi\left(X_1\right)\right)]$, and limiting result of equation (D.3) is

$$\left[\mathbb{E}(R(X_1)X_1X_1^{\top})\right]^{-1} \mathbb{E}\{[(R^2(X_1)/(\pi(X_1)(1-\pi(X_1))]X_1X_1^{\top}\} \left[\mathbb{E}(R(X_1)X_1X_1^{\top})\right]^{-1}. \tag{D.4}$$

Denoting $X_1^{\otimes 2} = X_1X_1^{\top}$, by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\{\mathbb{E}[R(X_1)X_1^{\otimes 2}]\}^2 &= \left[\mathbb{E}\left\{\frac{R(X_1)X_1}{\sqrt{\pi(X_1)(1-\pi(X_1))}}\sqrt{\pi(X_1)(1-\pi(X_1))}X_1^{\top}\right\}\right]^2 \\ &\leq \mathbb{E}\left\{\frac{R^2(X_1)}{\pi(X_1)(1-\pi(X_1))}X_1^{\otimes 2}\right\}\mathbb{E}[\pi(X_1)(1-\pi(X_1))X_1^{\otimes 2}],\end{aligned}$$

or equivalently,

$$\{\mathbb{E}[R(X_1)X_1^2]\}^{\otimes 2}(\mathbb{E}[\pi(X_1)(1-\pi(X_1))X_1^{\otimes 2}])^{-1} \leq \mathbb{E}\left\{\frac{R^2(X_1)}{\pi(X_1)(1-\pi(X_1))}X_1^{\otimes 2}\right\},$$

with equality when

$$\frac{R(X_1)}{\sqrt{\pi(X_1)(1-\pi(X_1))}} \propto \sqrt{\pi(X_1)(1-\pi(X_1))}.$$

Then the lower bound of equation (D.4) is attained by $(\mathbb{E}[\pi(X_1)(1-\pi(X_1))X_1^{\otimes 2}])^{-1}$.

Therefore, the optimal $R(X)$ is $\propto \pi(x)(1-\pi(x))$, and it corresponds to the optimal weights, which are $w(1,x) \propto (1-\pi(x))$, and $w(0,x) \propto \pi(x)$.

Appendix E

Invalid dWOLS Theorem

In a different approach to the proof of the consistency result for dWOLS presented in Wallace and Moodie [2015], we first consider the theorem in the case where $\boldsymbol{x}^{\psi}$ is a subset of $\boldsymbol{x}^{\beta}$: *Suppose* $\boldsymbol{x}^{\psi}$ *is a subset of* $\boldsymbol{x}^{\beta}$*, and* $\boldsymbol{x}^{\beta} = (\boldsymbol{x}^{\psi} \mid z)$ *where* z *is of dimension* $n \times (p^{\beta} - p^{\psi})$ *and is orthogonal to* $\boldsymbol{x}^{\psi}$*, in the sense that* $\mathbb{E}[z^{\top} R(\boldsymbol{x}) \boldsymbol{x}^{\psi}] = \mathbf{0}$*, where* $R(\boldsymbol{x}) := \pi(\boldsymbol{x}) w(1, \boldsymbol{x}) = (1 - \pi(\boldsymbol{x})) w(0, \boldsymbol{x})$*. Then a weighted least squares regression of* Y *on* $\left(\boldsymbol{x}^{\beta}, \boldsymbol{a}\boldsymbol{x}^{\psi}\right)$ *using the weights* $w(\boldsymbol{a}, \boldsymbol{x})$ *will yield consistent estimators of* $\boldsymbol{\psi}$ *if either the treatment-free model or the treatment model is correctly specified.*

The entire proof is presented in Appendix B. To summarize, first, we note that if the treatment-free model, where $f(\boldsymbol{x}^{\beta}; \boldsymbol{\beta}) = \boldsymbol{\beta}^{\top} \boldsymbol{x}^{\beta}$ in the linear case, is correctly identified, we can acquire unbiased estimators of $\boldsymbol{\psi}$ (i.e., $\mathbb{E}[\hat{\boldsymbol{\psi}}] - \boldsymbol{\psi} = \mathbf{0}$) via weighted linear regression for any set of weights. We then show that if we apply weights that satisfy $\pi(\boldsymbol{x}) w(1, \boldsymbol{x}) = (1 - \pi(\boldsymbol{x})) w(0, \boldsymbol{x})$, we also have $\mathbb{E}[\hat{\boldsymbol{\psi}}] - \boldsymbol{\psi} = \mathbf{0}$, even if the treatment-free model is incorrectly specified. During this process, considering weighted OLS system equations, we first derived conditional expectation of $\hat{\boldsymbol{\beta}}$ (i.e., $\mathbb{E}[\hat{\boldsymbol{\beta}} | a, \boldsymbol{x}]$), and then $\mathbb{E}[\hat{\boldsymbol{\psi}} | a, \boldsymbol{x}]$. Further, taking expectation of $\mathbb{E}[\hat{\boldsymbol{\psi}} | a, \boldsymbol{x}]$ with respect to the distribution of A given $\boldsymbol{x}$, we have the expectation of $\hat{\boldsymbol{\psi}}$, conditional on $\boldsymbol{x}$; thus, we derived the bias of $\hat{\boldsymbol{\psi}}$ (i.e., $\mathbb{E}[\hat{\boldsymbol{\psi}}] - \boldsymbol{\psi}$). Finally, the key factor in the bias of $\hat{\boldsymbol{\psi}}$ is denoted as $\mathcal{E} = \{\mathcal{B}(\boldsymbol{x})(I - \boldsymbol{x}^{\beta}[\mathcal{A}(\boldsymbol{x})\boldsymbol{x}^{\beta}]^{-1}\mathcal{A}(\boldsymbol{x}))\} f(\boldsymbol{x}^{\beta}, \boldsymbol{\beta})$, that is, B.7 in Appendix B, where $\mathcal{B}(\boldsymbol{x}) = (\boldsymbol{x}^{\psi})^{\top} R(\boldsymbol{x})$, $\mathcal{A}(\boldsymbol{x}) = (\boldsymbol{x}^{\beta})^{\top} R(\boldsymbol{x})[2I - \mathcal{P}^{\psi} R(\boldsymbol{x})]$ and $\mathcal{P}^{\psi} = \boldsymbol{x}^{\psi}[(\boldsymbol{x}^{\psi})^{\top} R(\boldsymbol{x}) \boldsymbol{x}^{\psi}]^{-1} (\boldsymbol{x}^{\psi})^{\top}$ are expressions that are related to $R(\boldsymbol{x})$, so that $\hat{\boldsymbol{\psi}}$ is unbiased if and only if $\mathcal{E} = \mathbf{0}$. Either (i) $\boldsymbol{x}^{\beta} = \boldsymbol{x}^{\psi} = \boldsymbol{x}$ or (ii) $\boldsymbol{x}^{\beta} = (\boldsymbol{x}^{\psi} \mid z)$ where $\mathbb{E}[z^{\top} R(\boldsymbol{x}) \boldsymbol{x}^{\psi}] = \mathbf{0}$ can be shown to imply that $\mathcal{E} = \mathbf{0}$ if we apply weights that are satisfied by $\pi(\boldsymbol{x}) w(1, \boldsymbol{x}) = (1 - \pi(\boldsymbol{x})) w(0, \boldsymbol{x}) = R(\boldsymbol{x})$.

Another scenario (iii), in which $\boldsymbol{x}^{\psi} = (\boldsymbol{x}^{\beta} \mid h)$ where $\mathbb{E}[h^{\top} R(\boldsymbol{x}) \boldsymbol{x}^{\beta}] = \mathbf{0}$, is addressed in the following Invalid dWOLS Theorem:

Theorem E.1. Invalid dWOLS Theorem *Suppose* $\boldsymbol{x}^{\beta}$ *is a subset of* $\boldsymbol{x}^{\psi}$*, and* $\boldsymbol{x}^{\psi} = (\boldsymbol{x}^{\beta} \mid h)$ *where the dimension of* h *is* $n \times (p^{\psi} - p^{\beta})$ *and where* h *is orthogonal to* $\boldsymbol{x}^{\beta}$ *in the sense that* $\mathbb{E}[h^{\top} R(\boldsymbol{x}) \boldsymbol{x}^{\beta}] = \mathbf{0}$*. With weight criteria that* $\pi(\boldsymbol{x}) w(1, \boldsymbol{x}) = (1 - \pi(\boldsymbol{x})) w(0, \boldsymbol{x}) = R(\boldsymbol{x})$ *, a weighted least squares regression of* Y *on* $\left\{\boldsymbol{x}^{\beta}, \boldsymbol{a}\boldsymbol{x}^{\psi}\right\}$ *will yield consistent estimates of* $\boldsymbol{\psi}$ *if and only if* $\mathbb{E}[h^{\top} R(\boldsymbol{x}) f(\boldsymbol{x}^{\beta}, \boldsymbol{\beta})] = \mathbf{0}$*.*

The proof contains two steps. First, we prove that $\mathbb{E}[h^{\top} R(\boldsymbol{x}) \boldsymbol{x}^{\beta}] = \mathbf{0}$, expression $\mathcal{E}$ (B.7) is $\mathbf{0}$ if and only if $h^{\top} R(\boldsymbol{x}) f(\boldsymbol{x}^{\beta}, \beta) = \mathbf{0}$. That is the case in which $\boldsymbol{x}^{\beta} = \boldsymbol{x}^{\psi} = \boldsymbol{x}$. Thus for the proof of asymptotic unbiasedness of $\hat{\boldsymbol{\psi}}$ when the treatment-free part is of general form (Section 2.2), it must be true that $\boldsymbol{x}^{\beta}$ includes all components of $\boldsymbol{x}^{\psi}$. In the second step, we prove the theorem using the argument presented in the first step.

Proof. **Step 1:** For $\boldsymbol{x}^{\psi} = (\boldsymbol{x}^{\beta} \mid h)$, where h dimension is $n \times (p^{\psi} - p^{\beta})$ then

$$\mathcal{P}^{\psi} = \boldsymbol{x}^{\psi} [(\boldsymbol{x}^{\psi})^{\top} R(\boldsymbol{x}) \boldsymbol{x}^{\psi}]^{-1} (\boldsymbol{x}^{\psi})^{\top} := (\boldsymbol{x}^{\beta}, h) E^{-1} \begin{pmatrix} (\boldsymbol{x}^{\beta})^{\top} \\ h^{\top} \end{pmatrix},$$

where $E^{-1} = \left[\begin{pmatrix} (\boldsymbol{x}^{\beta})^{\top} \\ h^{\top} \end{pmatrix} R(\boldsymbol{x}) (\boldsymbol{x}^{\beta}, h) \right]^{-1} = \begin{pmatrix} [(\boldsymbol{x}^{\beta})^{\top} R(\boldsymbol{x}) \boldsymbol{x}^{\beta}]^{-1} & \mathbf{0} \\ \mathbf{0} & [h^{\top} R(\boldsymbol{x}) h]^{-1} \end{pmatrix}.$

Thus

$$\begin{aligned} \mathcal{A}(x) &= (\boldsymbol{x}^{\beta})^{\top} R(\boldsymbol{x}) (2I - \mathcal{P}^{\psi} R(\boldsymbol{x})) \\ &= 2(\boldsymbol{x}^{\beta})^{\top} R(\boldsymbol{x}) - [(\boldsymbol{x}^{\beta})^{\top} R(\boldsymbol{x}) \boldsymbol{x}^{\beta}, \mathbf{0}_{p^{\beta} \times (p^{\psi} - p^{\beta})}] E^{-1} \begin{pmatrix} (\boldsymbol{x}^{\beta})^{\top} \\ h^{\top} \end{pmatrix} R(\boldsymbol{x}). \end{aligned}$$

For the second term of $\mathcal{A}(x)$, denoted as $\mathcal{E}(x)$,

$$\begin{aligned} \mathcal{E}(x) &= [(\boldsymbol{x}^{\beta})^{\top} R(\boldsymbol{x}) \boldsymbol{x}^{\beta}, \mathbf{0}_{p^{\beta} \times (p^{\psi} - p^{\beta})}] \begin{pmatrix} [(\boldsymbol{x}^{\beta})^{\top} R(\boldsymbol{x}) \boldsymbol{x}^{\beta}]^{-1} & \mathbf{0} \\ \mathbf{0} & [h^{\top} R(\boldsymbol{x}) h]^{-1} \end{pmatrix} \begin{pmatrix} (\boldsymbol{x}^{\beta})^{\top} \\ h^{\top} \end{pmatrix} R(\boldsymbol{x}) \\ &= [I_{p^{\beta} \times p^{\beta}}, \mathbf{0}_{p^{\beta} \times (p^{\psi} - p^{\beta})}] \begin{pmatrix} (\boldsymbol{x}^{\beta})^{\top} \\ h^{\top} \end{pmatrix} R(\boldsymbol{x}) \\ &= (\boldsymbol{x}^{\beta})^{\top} R(\boldsymbol{x}). \end{aligned}$$

Then, $\mathcal{E}(x)\boldsymbol{x}^{\beta} = (\boldsymbol{x}^{\beta})^{\top}R(\boldsymbol{x})\boldsymbol{x}^{\beta}$ and $\mathcal{A}(x)\boldsymbol{x}^{\beta} = 2(\boldsymbol{x}^{\beta})^{\top}R(\boldsymbol{x})\boldsymbol{x}^{\beta} - \mathcal{E}(x)\boldsymbol{x}^{\beta} = (\boldsymbol{x}^{\beta})^{\top}R(\boldsymbol{x})\boldsymbol{x}^{\beta}$.

$$\begin{aligned}\mathcal{B}(x)\boldsymbol{x}^{\beta}[\mathcal{A}(x)\boldsymbol{x}^{\beta}]^{-1}\mathcal{A}(x) &= \begin{pmatrix} (\boldsymbol{x}^{\beta})^{\top} \\ h^{\top} \end{pmatrix} R(\boldsymbol{x})\boldsymbol{x}^{\beta}[(\boldsymbol{x}^{\beta})^{\top}R(\boldsymbol{x})\boldsymbol{x}^{\beta}]^{-1}(\boldsymbol{x}^{\beta})^{\top}R(\boldsymbol{x}) \\ &= \begin{pmatrix} (\boldsymbol{x}^{\beta})^{\top}R(\boldsymbol{x})\boldsymbol{x}^{\beta} \\ \mathbf{0}_{(p^{\psi}-p^{\beta})\times p^{\beta}} \end{pmatrix} [(\boldsymbol{x}^{\beta})^{\top}R(\boldsymbol{x})\boldsymbol{x}^{\beta}]^{-1}(\boldsymbol{x}^{\beta})^{\top}R(\boldsymbol{x}) \\ &= \begin{pmatrix} (\boldsymbol{x}^{\beta})^{\top}R(\boldsymbol{x}) \\ \mathbf{0}_{(p^{\psi}-p^{\beta})\times n} \end{pmatrix}.\end{aligned}$$

The sufficient and necessary condition for $\mathcal{B}(x)\boldsymbol{x}^{\beta}[\mathcal{A}(x)\boldsymbol{x}^{\beta}]^{-1}\mathcal{A}(x)f(\boldsymbol{x}^{\beta},\beta) = \mathcal{B}(x)f(\boldsymbol{x}^{\beta},\beta) = (\boldsymbol{x}^{\psi})^{\top}R(\boldsymbol{x})f(\boldsymbol{x}^{\beta},\beta)$, that is $\mathcal{E}$ (B.7) $= \mathbf{0}$, is $h^{\top}R(\boldsymbol{x})f(\boldsymbol{x}^{\beta},\beta) = \mathbf{0}$.

Step 2: Appendix B shows that with weight criteria that $\pi(\boldsymbol{x})w(1,\boldsymbol{x}) = (1-\pi(\boldsymbol{x}))w(0,\boldsymbol{x})$ a weighted least squares regression of Y on $\left(\boldsymbol{x}^{\beta}, \boldsymbol{a}\boldsymbol{x}^{\psi}\right)$ will yield consistent estimates of ψ if expression $\mathcal{E}$ (B.7) is $\mathbf{0}$; then apply the conclusion in **Step 1**, to show that $\mathcal{E}$ (B.7) is $\mathbf{0}$ if and only if $h^{\top}R(\boldsymbol{x})f(\boldsymbol{x}^{\beta},\beta) = \mathbf{0}$. □

To prove the Invalid dWOLS Theorem, we show that $\mathcal{E} = \mathbf{0}$ iff $h^{\top}R(\boldsymbol{x})f(\boldsymbol{x}^{\beta},\boldsymbol{\beta}) = \mathbf{0}$. To illustrate this, we use the following example: for a single-stage treatment decision, suppose covariates $\boldsymbol{x}_1$ are common to the treatment-free and blip models, and we extend the blip model by correctly incorporating a covariate (A^{alter}) that indicates the alter's treatment. The true outcome model can therefore be expressed as: $\mathbb{E}[Y|\boldsymbol{x}_1, a, a^{alter}] = f(\boldsymbol{x}_1;\boldsymbol{\beta}) + \gamma(\boldsymbol{x}_1, a, a^{alter};\boldsymbol{\psi})$, where f is nonlinear. For illustration, we propose the following model:

$$\mathbb{E}[Y|\boldsymbol{x}_1, a, a^{alter}] = \beta_0 + \boldsymbol{\beta}_1^{\top}\boldsymbol{x}_1 + a\left(\psi_0 + \boldsymbol{\psi}_1^{\top}\boldsymbol{x}_1 + \psi_2 a^{alter}\right), \tag{E.1}$$

where the treatment-free model is given by $\beta_0 + \boldsymbol{\beta}_1^{\top}\boldsymbol{x}_1$, and the blip function $\gamma(\boldsymbol{x}_1, a, a^{alter})$ is modeled as $a\left(\psi_0 + \boldsymbol{\psi}_1^{\top}\boldsymbol{x}_1 + \psi_2 a^{alter}\right)$.

First, suppose A^{alter} depends only on the covariate for the alter, and that covariate is independent of $\boldsymbol{x}_1$ for the ego. For a model such as (E.1), we can consistently estimate the blip parameters of $\boldsymbol{\psi}$. Second, suppose that A^{alter} also depends on the covariate for the ego (the possible interference path is due to A^{alter} depending on $\boldsymbol{x}_1$). The Invalid dWOLS Theorem shows that estimators of $\boldsymbol{\psi}$ will not be consistent, and the inconsistency will affect the estimation of optimal treatment decision rules. That is, if a component of $\boldsymbol{x}^{\psi}$ is missing from $\boldsymbol{x}^{\beta}$, and this component depends on $\boldsymbol{x}^{\beta}$, we have inconsistency. In a model such as (E.1), it is the dependence of the interference term a^{alter} on $\boldsymbol{x}_1$ for the ego that invalidates the double robustness of dWOLS.

Appendix F

Appendix of Chapter 3

F.1 Chapter 3 Appendix A1: Extension to the pair case where the treatments of ego and alter are not independent.

For pairs, our method can be extended to the case where the treatments of ego and alter are not independent. To see this, define the network propensities for the pairs as, for $a^r = 0, 1$ and $a^s = 0, 1$, $\pi^{a^r,a^s}(\boldsymbol{x}_{ego}, \boldsymbol{x}_{alter}) = \mathbb{P}(A^{ego} = a^r, A^{alter} = a^s \mid \boldsymbol{x}_{ego}, \boldsymbol{x}_{alter}) = \pi_1^{(a^r)}\pi_2^{(a^s)}\lambda^{(a^r,a^s)}$ where $\pi_1^{(a^r)}$, $\pi_2^{(a^s)}$, and $\lambda^{(a^r,a^s)}$ all depend on $(\boldsymbol{x}_{ego}, \boldsymbol{x}_{alter})$, and $\pi_1^{(a^r)}$ and $\pi_2^{(a^s)}$ are marginal probabilities that $A^{ego} = a^r$ and $A^{alter} = a^s$, respectively. Then the $\lambda^{(a^r,a^s)}$ functions describe the dependence between treatments of ego and alter, and this is exchangeable if $\lambda^{(a^r,a^s)} = \lambda^{(a^s,a^r)}$. Balancing weights (w^{a^r,a^s}) that depend on $(\boldsymbol{x}_{ego}, \boldsymbol{x}_{alter})$ are given by:

$$
\begin{aligned}
a^{ego} = 0, a^{alter} = 0 &: w^{0,0} = \pi_1^{(1)}\pi_2^{(1)}/\lambda^{(0,0)} \\
a^{ego} = 0, a^{alter} = 1 &: w^{0,1} = \pi_1^{(1)}\pi_2^{(0)}/\lambda^{(0,1)} \\
a^{ego} = 1, a^{alter} = 0 &: w^{1,0} = \pi_1^{(0)}\pi_2^{(1)}/\lambda^{(1,0)} \\
a^{ego} = 1, a^{alter} = 1 &: w^{1,1} = \pi_1^{(0)}\pi_2^{(0)}/\lambda^{(1,1)}
\end{aligned}
$$

The function $\lambda^{(a^r,a^s)}$ is a factor that accounts for the association, and $\lambda^{(a^r,a^s)}$ is identically equal to 1 if the treatments of ego and alter are independent for each value of the covariates. The balancing weights w^{a^r,a^s} satisfy the balancing property required in Theorem 3.2, i.e., $\pi^{0,0}w^{0,0} = \pi^{0,1}w^{0,1} = \pi^{1,0}w^{1,0} = \pi^{1,1}w^{1,1}$.

F.2 Chapter 3 Appendix A2: Proof of Theorem 3.1

We now prove Theorem 3.1. First, the terms related to $t(a_{\mathcal{N}})$ are treated as explanatory variables, and the explanatory variables in the blip component are the same as the explanatory variables in the treatment-free component ($\boldsymbol{x}^{\beta} = \boldsymbol{x}^{\psi}$). According to the argument outlined in Chapter 3 Appendix A1, the standard dWOLS weights provide consistent estimation.

Second, we show that the network balancing condition (3.5) implies the appropriate analogue of $\pi(\boldsymbol{x})w(1,\boldsymbol{x}) = (1-\pi(\boldsymbol{x}))w(0,\boldsymbol{x})$.

We add $\mathbf{x}_{\mathcal{N}_i}$, $s_{i,a}$ to $\boldsymbol{x}_i$, so that the new full set of covariates is considered to be $(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}, s_a)$, then define $\pi_a(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}, s_a) = \pi_{a,s_a}(\boldsymbol{x}, \mathcal{N}, \mathbf{x}_{\mathcal{N}})$. The network weights satisfy network balancing conditions (i.e., $w(\boldsymbol{x}, \mathcal{N}, \mathbf{x}_{\mathcal{N}}, a, s_a)\pi_{a,s_a}(\boldsymbol{x}, \mathcal{N}, \mathbf{x}_{\mathcal{N}})$ are the same $\mathscr{C}(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}, |\mathcal{N}|)$ for $a = 1, 0$). Therefore, we conclude that $w(\boldsymbol{x}, \mathcal{N}, \mathbf{x}_{\mathcal{N}}, a, s_a)$ can be written as $w_a(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}, s_a)$ for $a = 0, 1$, and that $w_a(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}, s_a)\pi_a(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}}, s_a)$ are the same for w and π subscripts $a = 0, 1$.

Lastly, we show that the network weights have the balancing property for the propensities of a_i if the dependence is on a dimension-reducing function of those explanatory variables (i.e., $t(a_{\mathcal{N}_i})$). For any network weights that satisfy network balancing conditions (3.5), if $\pi_{i,a_i,s_{i,a}}(\boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i})$ is actually a function of $s_{i,a}$ through $t(a_{\mathcal{N}_i})$, then $w_i(\boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i}, a_i, s_{i,a})$ also will be a function of $s_{i,a}$ through $t(a_{\mathcal{N}_i})$.

We conclude that, therefore, if the balance condition is satisfied for the configuration of neighbours' treatments it would be satisfied for the function of the neighbours' treatments, because the constant term in the network balancing conditions (3.5) should depend only on the $(\boldsymbol{x}, \mathbf{x}_{\mathcal{N}})$ values and the degree.

F.3 Chapter 3 Appendix A3: An illustrative example of the balancing weights conditions

This section provides an example illustrating the balancing weights conditions (3.5) and (3.6) and the choice and computation of the network weights.

Suppose $|\mathcal{N}_i| = 1$ for all i; this is the scenario of isolated pairs or couples. There are four possibilities for treatments of one pair: $(a_i, a_{\mathcal{N}_i}) \in \{(1,1),(1,0),(0,1),(0,0)\}$. For $\mathcal{X} = (\boldsymbol{x}, \mathbf{x}_{\mathcal{N}})$, let $\pi_1(\mathcal{X}) = \mathbb{P}(A = 1|\mathcal{X})$, $\eta_1(\mathcal{X}) = \mathbb{P}(A_{\mathcal{N}} = 1|\mathcal{X})$, thus $\pi_0(\mathcal{X}) = \mathbb{P}(A = 0|\mathcal{X}) = 1 - \pi_1(\mathcal{X})$, and $\eta_0(\mathcal{X}) = \mathbb{P}(A_{\mathcal{N}} = 0|\mathcal{X}) = 1 - \eta_1(\mathcal{X})$. Let π_{rs} denote the network propensity in this pairs case such that $\pi_{rs} = \mathbb{P}(A = r, A_{\mathcal{N}_i} = s|\mathcal{X})$ where $r, s = 0, 1$. Therefore, we can create weights that satisfy, $\pi_{01}w_{01}(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}) = \pi_{00}w_{00}(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}) = \pi_{10}w_{10}(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i}) = \pi_{11}w_{11}(\boldsymbol{x}_i, \mathbf{x}_{\mathcal{N}_i})$; if the

two treatment assignments (for ego and alter) are independent, then, $[\pi_0(X)\eta_1(X)]w_{01}(X) = [\pi_0(X)\eta_0(X)]w_{00}(X) = [\pi_1(X)\eta_0(X)]w_{10}(X) = [\pi_1(X)\eta_1(X)]w_{11}(X)$.

We now demonstrate the choice and computation of the network weights. In the case of no interference, Wallace and Moodie [2015] recommend the "absolute value" weights, that is $w(a;x) = |a - \mathbb{E}[A|\boldsymbol{X} = \boldsymbol{x}]|$. For binary treatments, these weights are identical to overlap weights (Li et al. [2018]; Schulz and Moodie [2021]), which are bounded, yield minimum variance of the nonparametric estimator among all balancing weights, and lead to exact balance for means of included covariates in a logistic propensity score model. Let $\boldsymbol{v} = (|a_1 - \mathbb{E}[A_1|\boldsymbol{X}_1 = \boldsymbol{x}_1]|, ..., |a_n - \mathbb{E}[A_n|\boldsymbol{X}_n = \boldsymbol{x}_n]|)^\top$; thus, each entry is the dWOLS weight under no interference, and so the network weight for a network of size n is $\boldsymbol{w} = \boldsymbol{v} \odot exp[\boldsymbol{N} \cdot log(\boldsymbol{v})]$, where $\odot$ denotes component-wise multiplication, and $\boldsymbol{N}$ is the adjacency matrix. Therefore, for $i = 1, 2, ..., n$, the i^{th} component of $\boldsymbol{w}$ is w_i that is expressed in equation (3.9).

F.4 Chapter 3 Appendix A4a: Model set up of Methods 1 - 3 and regrets values of each scenario in *Study 1*

In Simulation 3.1 *Study 1*, we investigated three methods: (1) Method 1, a 'naive' analysis, where we ignore the interference entirely; (2) standard dWOLS, and (3) dWOLS with network weights. For Scenarios a) and b) where the treatment-free model is mis-specified, we consider outcome models:

- Method 1:
  $\mathbb{E}[Y|X = x] = \beta_0 + \beta_1 x + a(\psi_0 + \psi_1 x)$, with weight $w = |a - \mathbb{E}(A \mid x)|$.
- Method 2:
  $\mathbb{E}[Y|X = x] = \beta_0 + \beta_1 x + \beta_2 a^{alter} + \beta_3 x a^{alter} + a_{ego}(\psi_0 + \psi_1 x + \psi_2 a^{alter} + \psi_3 x a^{alter})$, and with weight $w = |a_{ego} - \mathbb{P}(A_{ego} = 1 \mid x_{ego})|$.
- Method 3:
  $\mathbb{E}[Y|X = x] = \beta_0 + \beta_1 x + \beta_2 a^{alter} + \beta_3 x a^{alter} + a_{ego}(\psi_0 + \psi_1 x + \psi_2 a^{alter} + \psi_3 x a^{alter})$, and with network weight $w_{net}(x) = |a_{ego} - \mathbb{E}(A_{ego} \mid x_{ego})| * |a^{alter} - \mathbb{E}(A^{alter} \mid x_{alter})|$.

Table F.1: Total regret and its standard errors of Methods 1, 2, and 3 when neither (scenario a), one (scenarios b and c) or both (scenario d) treatment and treatment-free outcome models are correct in Study 1a (ME: Method).

		Total Regret			Standard Errors of Total Regret		
Scenario		$n = 200$	$n = 500$	$n = 1000$	$n = 200$	$n = 500$	$n = 1000$
a	ME1	2.35	0.23	1.11	0.19	0.14	0.11
	ME2	3.31	1.87	2.27	0.28	0.12	0.11
	ME3	2.29	2.01	2.01	0.24	0.12	0.10
b	ME1	0.91	0.22	0.18	0.28	0.13	0.10
	ME2	1.75	1.25	0.60	0.34	0.15	0.09
	ME3	1.90	1.23	1.04	0.21	0.14	0.09
c	ME1	0.82	0.39	0.56	0.21	0.14	0.09
	ME2	2.00	1.65	2.22	0.24	0.12	0.11
	ME3	2.01	1.67	2.25	0.24	0.11	0.09
d	ME1	0.97	1.35	1.38	0.21	0.14	0.10
	ME2	1.79	2.67	1.81	0.20	0.16	0.09
	ME3	1.85	2.64	1.80	0.20	0.15	0.09

F.5 Chapter 3 Appendix A4b: *Study 1b* and *Study 1c* and their results

Study 1a demonstrates the dangers of failing to account for interference, and suggests that while standard dWOLS can account for it, our proposed network weights provide a competitive alternative. We now demonstrate some of the advantages of network weights over standard dWOLS. First, we consider scenarios where the treatment-free model is more complex in terms of covariates relating to the alter (*Studies 1b* and *1c*). We consider an outcome model of the form $\mathbb{E}[Y|X, X_{alter}, A, A^{alter}] = X + \log(|X|) + X^3 + \exp(X_{alter}) + A^{alter}(1 + X) + A(-3 + 1.2X + 3A^{alter} + 2XA^{alter})$, where we note that we now have a non-linear term in X_{alter} (In Figure 3.3a DAG 1, Y_{ego} depends on X_{alter}). Like the treatment models setting in *Study 1a*, both the ego's and the alter's treatment models are correctly specified (or not) at the same time (*Study 1b*). Building on the idea of splitting the treatment-free model into ego and alter components, we divide scenarios (a) and (b) into six new scenarios. The resulting eight scenarios are summarized in Table F.3. Note that, if one of either TF-ego or TF-alter is mis-specified, then the treatment-free model is mis-specified; therefore, scenarios 1, 3 and 5 belong to scenario (a), and scenarios 2, 4 and 6 belong to scenario (b). We surmise that (I) Methods 2 and 3 offer consistent $\boldsymbol{\psi}$ estimators in scenarios 2, 4 - 8, but the variances of $\hat{\boldsymbol{\psi}}$ from Method 3 are smaller than those from Method 2 (Figure F.1). (II) An efficiency gain of Method 3 (network weights) produces much higher optimal treatment rates, specifically in scenarios 2, 5, and 6 (Table F.2). (III) The network weights provide greater robustness in modeling: Method 3 appears to offer consistent estimators of spillover effect terms (A^{alter}, $A^{alter}X$) in scenarios 2, 6 (where the treatment model is correct), but Method 2 does not in these two scenarios (Figure F.2).

Study 1c We also consider the scenario where the ego's treatment model is correctly specified but the alter's is not, or vice-versa (*Study 1c*). In this scenario, the true outcome model is the same as in *Study 1b* where the non-linear term $\exp(X_{alter})$ is included. However, the true treatment models are set as: $\mathbb{P}(A_i = 1 \mid X_i) = \text{expit}[-0.25X_i + \sin(X_i)]$, and $\mathbb{P}(A_{alter_i} = 1 \mid X_{alter_i}) = \text{expit}[-0.1X_{alter_i} + 0.2\sin(X_{alter_i}) + \log(|X_{alter_i}|)]$. We note that the true $\mathbb{P}(A^{alter} = 1|X_{alter})$ model is more complex than the ego's. In real analysis, we propose the same treatment model for both ego and alter. Therefore, for the case where the $\mathbb{P}(A^{alter} = 1|X_{alter})$ model is incorrect but the $\mathbb{P}(A_{ego} = 1|X_{ego})$ is correct, we propose models that only include $x, \sin(x)$ terms; for the case where both treatment models are misspecified, we just include the linear term of covariates in the model. Results are summarized in the following, where we observe that Methods 2 and 3 also offer consistent $\boldsymbol{\psi}$ estimators in scenarios 2, 4 - 8 (Figure F.3). The efficiency gain of Method 3 (network weights) still appears to provide higher (but not by much) optimal treatment rates, specifically in scenarios 2,5, and 6. The greater robustness property of network weights still exists in scenarios 2 - 8, but also appears in scenario 1: Method 3 offers consistent estimators of spillover

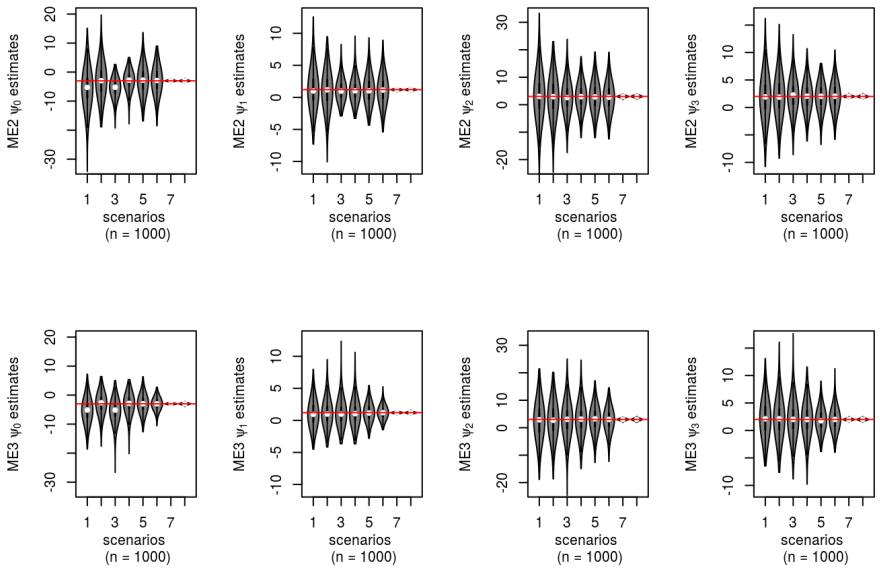


Figure F.1: *Study 1b*. Blip function parameter ($\boldsymbol{\psi}$) estimates for 1000 simulated datasets via Methods 2 and 3 when none (scenario 1), one (scenarios 2,3,4,5 and 6), two (scenario 7) or three (scenario 8) of treatment-free ego (TF-ego), treatment-free alter (TF-alter), and treatment models are correct. Note that Method 2 (ME2) is above and Method 3 (ME3) is below.

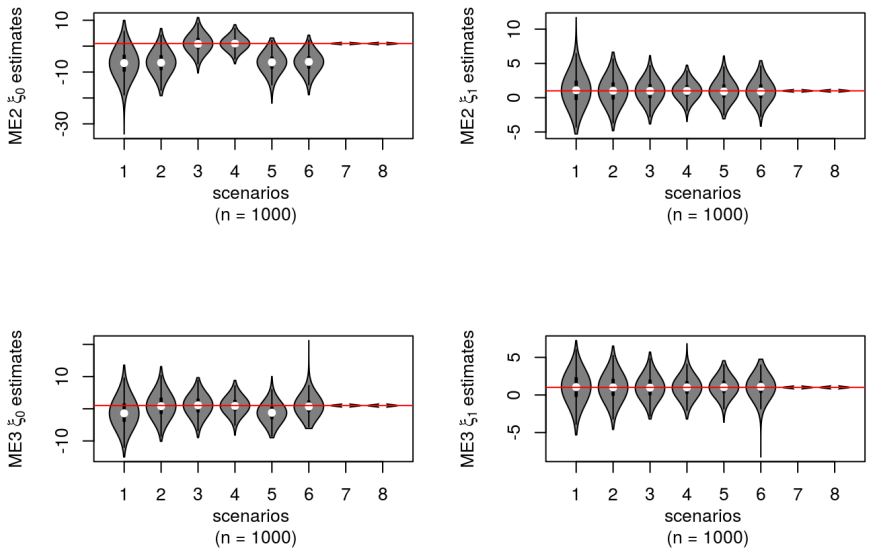


Figure F.2: *Study 1b*. Spillover effects parameter (denoted as $\boldsymbol{\xi}$) estimates for 1000 simulated datasets via Methods 2 and 3 when no (scenario 1), one (scenarios 2,3,4,5 and 6), two (scenario 7) or three (scenario 8) treatment-free ego (TF-ego), treatment-free alter (TF-alter), and treatment models are correct. Note that Method 2 (ME2) is above and Method 3 (ME3) is below.

Table F.2: Mean optimal treatment rate and total regret and their standard errors of Methods 1, 2, and 3 in *Study 1b* (ME: Method, MOTR: Mean optimal treatment rate, SE: Standard errors).

Scenario (n = 1000)			MOTR %	SE of MOTR	Total regret	SE of total regret
a		ME1	63.91	0.0841	-0.24	0.11
	1	ME2	81.50	0.0922	**2.57**	0.14
		ME3	81.53	0.0841	1.41	0.16
	3	ME2	87.70	0.0746	2.75	0.10
		ME3	87.19	0.0756	2.43	0.10
	5	ME2	81.33	0.1208	1.45	0.12
		ME3	**87.40**	0.0889	1.79	0.12
b		ME1	66.78	0.0324	2.84	0.12
	2	ME2	81.02	0.1200	1.83	0.13
		ME3	**83.92**	0.1027	**3.07**	0.13
	4	ME2	87.93	0.0759	2.75	0.11
		ME3	87.43	0.0790	2.43	0.10
	6	ME2	82.05	0.1177	3.25	0.14
		ME3	**87.28**	0.0916	3.64	0.13
c		ME1	67.00	0.0311	-0.63	0.12
	7	ME2	98.68	0.0117	2.18	0.11
		ME3	98.62	0.0124	2.19	0.11
d		ME1	66.80	0.0315	2.36	0.14
	8	ME2	98.59	0.0116	2.61	0.12
		ME3	98.53	0.0124	2.62	0.11

Table F.3: Summary of analyses 1 to 8 in *Study 1b*. (*Note:* TF-ego refers to the treatment-free model related to ego; TF-alter means the treatment-free model related to the alter, and Treatment stands for the treatment model. Scenarios 1, 3 and 5 belong to scenario (a), and scenarios 2, 4 and 6 belong to scenario (b))

Scenarios	1 (a)	2 (b)	3 (a)	4 (b)	5 (a)	6 (b)	7(c)	8 (d)
TF-ego	✗	✗	✗	✗	✓	✓	✓	✓
TF-alter	✗	✗	✓	✓	✗	✗	✓	✓
Treatment	✗	✓	✗	✓	✗	✓	✗	✓
	✗: incorrectly specified				✓: correctly specified			

Table F.4: Summary of analyses 1 to 8 in *Study 1c*. (*Note:* Treatment stands for the ego's treatment model. Treatment-alter stands for the alter's treatment model. We consider the case where alter's treatment models are always mis-specified)

Scenarios	1 (a)	2 (b)	3 (a)	4 (b)	5 (a)	6 (b)	7(c)	8 (d)
TF-ego	✗	✗	✗	✗	✓	✓	✓	✓
TF-alter	✗	✗	✓	✓	✗	✗	✓	✓
Treatment	✗	✓	✗	✓	✗	✓	✗	✓
Treatment-alter	✗	✗	✗	✗	✗	✗	✗	✗
	✗: incorrectly specified				✓: correctly specified			

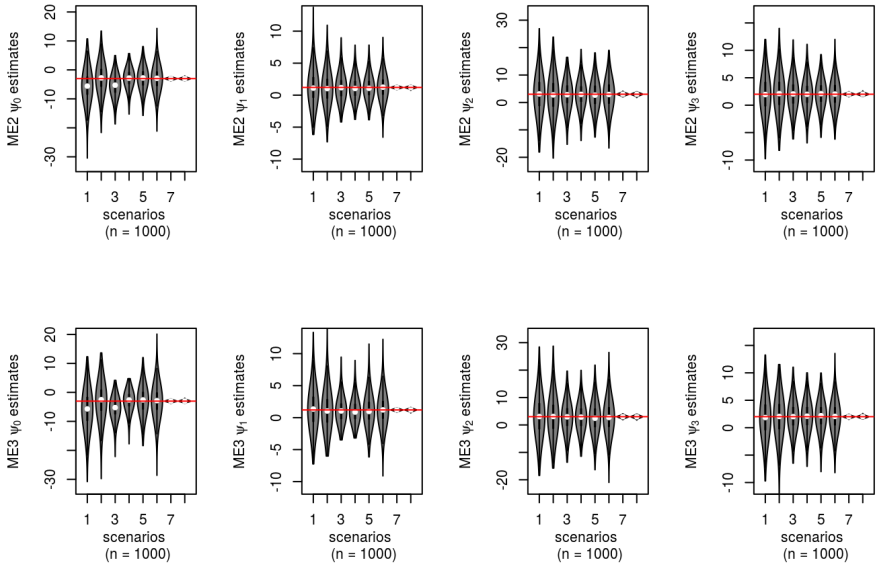


Figure F.3: *Study 1c* (the alter's treatment models are always mis-specified). Blip function parameter ($\boldsymbol{\psi}$) estimates for 1000 simulated datasets via Methods 2 and 3 when no (scenario 1), one (scenarios 2,3,4,5 and 6), two (scenario 7) or three (scenario 8) of the treatment-free ego (TF-ego), treatment-free alter (TF-alter), and treatment models are correct. Note that Method 2 (ME2) is above and Method 3 (ME3) is below.

effect terms in scenarios 2, 6 (where the ego's treatment model is correct); however, Method 2 does not in these two scenarios (Figure F.4).

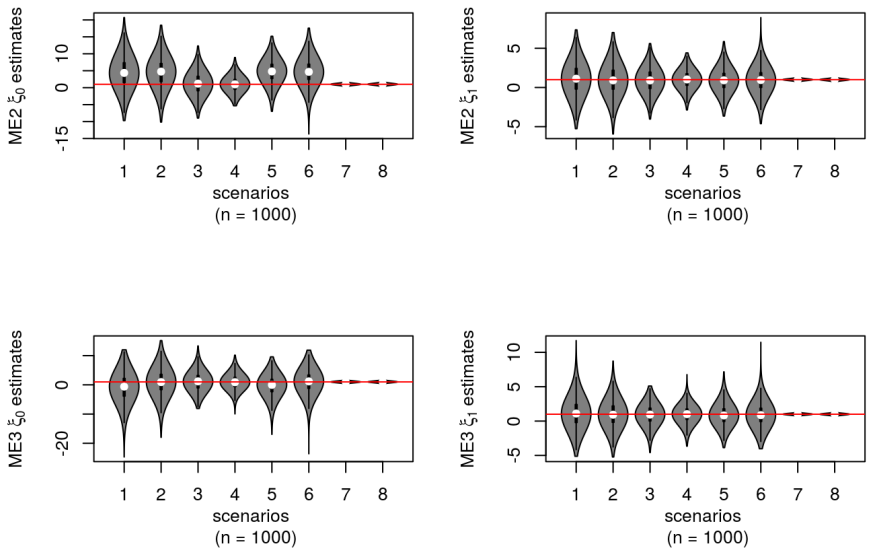


Figure F.4: *Study 1c* (the alter's treatment models are always mis-specified). Spillover effects parameter (denoted as $\boldsymbol{\xi}$) estimates for 1000 simulated datasets via Methods 2 and 3 when no (scenario 1), one (scenarios 2,3,4,5 and 6), two (scenario 7) or three (scenario 8) of the treatment-free ego (TF-ego), treatment-free alter (TF-alter), and treatment models are correct. Note that Method 2 (ME2) is above and Method 3 (ME3) is below.

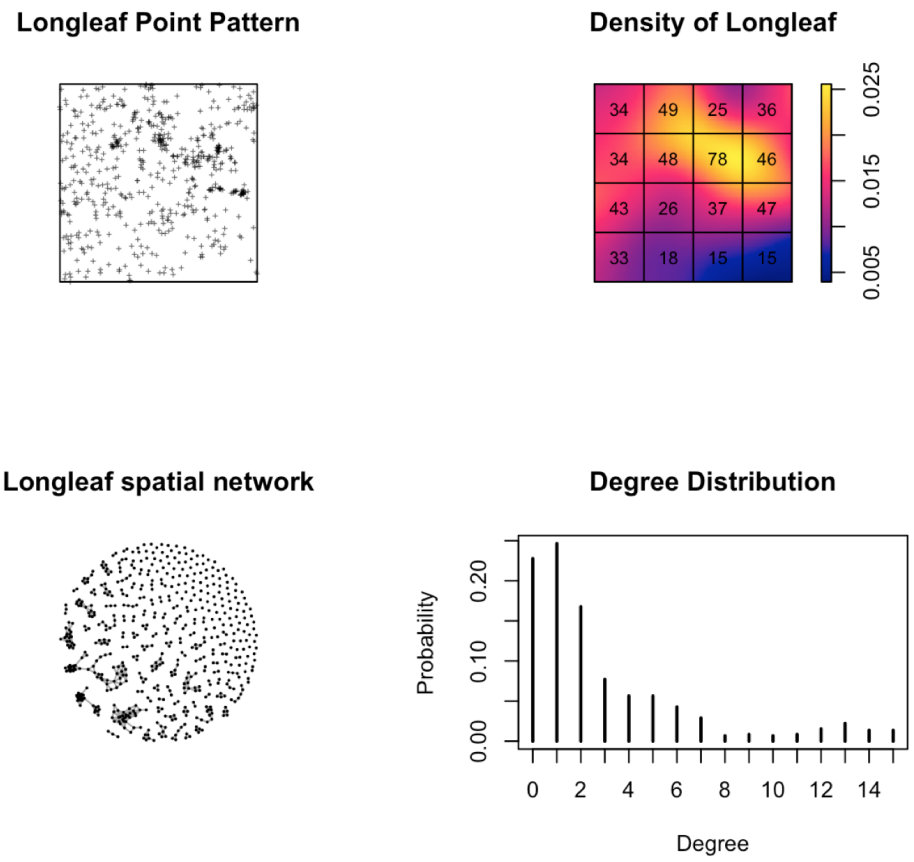


Figure F.5: Longleaf Pines spatial figures and degree summary. The first row presents the Longleaf Pines point pattern in the real world space, and the density of the Longleaf Pines (the numbers correspond to the numbers of pines). The second row is the Longleaf pines spatial network where any two pines less than 5 meters apart are linked, and the degree distribution of the Longleaf Pines spatial network.

F.6 Chapter 3 Appendix A5a: Longleaf Pines spatial networks.

The Longleaf Pines spatial figures and degree summary are shown in Figure F.5. In total 584 nodes (pines) are in the Longleaf Pines spatial network; in simulation *Study 2a* and *Study 2b*, 98 nodes with degree equal to 2 are designated as egos.

F.7 Chapter 3 Appendix A5b: *Study 2b* and its results, and additional simulations for extreme or sparse network weights.

Analogously to *Study 2a*, we highlight advantages of network weights over standard dWOLS weights by considering the case where the treatment-free model is more complex in terms of covariates relating to the alter. In *Study 2b*, we consider an interference term $t_i(a_{\mathcal{N}_i}) = \sum_{l=1}^{n} N_{il} a_l$, and an outcome model of the form $\mathbb{E}[Y|X, \mathbf{X}_{\mathcal{N}}, A, A_{\mathcal{N}}] = 1 + X + X^2 + \exp(\tau(\mathbf{X}_{\mathcal{N}})) + t(A_{\mathcal{N}}) + t(A_{\mathcal{N}})X + A[\psi_0 + \psi_1 X + \psi_2 t(A_{\mathcal{N}}) + \psi_3 X t(A_{\mathcal{N}})]$), where we note that we now have a non-linear term in $\tau_i(\mathbf{X}_{\mathcal{N}_i}) = \sum_{l=1}^{n} N_{il} X_l / \sum_{l=1}^{n} N_{il}$ (i.e., in Figure 3.3b DAG 2, Y_{ego} depends on $\mathbf{X}_{\mathcal{N}}$). Building on the idea of splitting the treatment-free model into ego and alter components, we divide scenarios (a) and (b) into six new scenarios. The resulting eight scenarios are summarized in Table F.3. It appears that (I) Methods 2 and 3 offer consistent $\boldsymbol{\psi}$ estimators in scenarios 2, 4 - 8, but that the variances of $\hat{\boldsymbol{\psi}}$ from Method 3 are smaller than those from Method 2 (Figure F.6). (II) The efficiency gain of Method 3 (network weights) produces much higher optimal treatment rates, specifically in scenarios 2, 5, and 6 (Table F.2). (III) The greater robustness in modeling property of network weights is seen: Method 3 always offers consistent estimators of spillover effect terms $(t(A_{\mathcal{N}}), t(A_{\mathcal{N}})X)$ in scenarios 2, 6 (where the treatment model is correct), but Method 2 does not in these two scenarios (Figure F.7).

Histograms of network absolute value weights and the standard absolute value weights
We now demonstrate the distributions of the network absolute value weights and the standard absolute value weights, which depends on the degree of the egos. From Figure F.8, for the standard absolute value weights, more values are distributed between 0.4 and 0.7. However, the network absolute value weights (with ego's degree equal to 5 in a BA network) are centred around 0.015.

Additional simulations for extreme or sparse network weights
This section presents the additional simulation figures from the case of the same setting as in Study 2b — the Erdős-Rényi network with 2000 nodes but not restricted to a narrow band of degrees of egos (the network weights are extreme and/or sparse). Similarly, Table F.3 indicates what is correctly specified in each of Scenarios 1 to 8.

To study the case where the network weights are extreme and/or sparse, we chose the Erdős-Rényi network with 2000 nodes, and did not restrict to a narrow band of degrees of egos. The network weights distributions are shown in Figure F.9, where two cases were investigated: (I) the egos' degrees are either 1 or 4, and (II) the egos' degrees are either 2 or 5.

In simulation Study 2b, which highlights advantages of network weights over standard dWOLS weights by considering the case where the treatment-free model is more complex in terms of

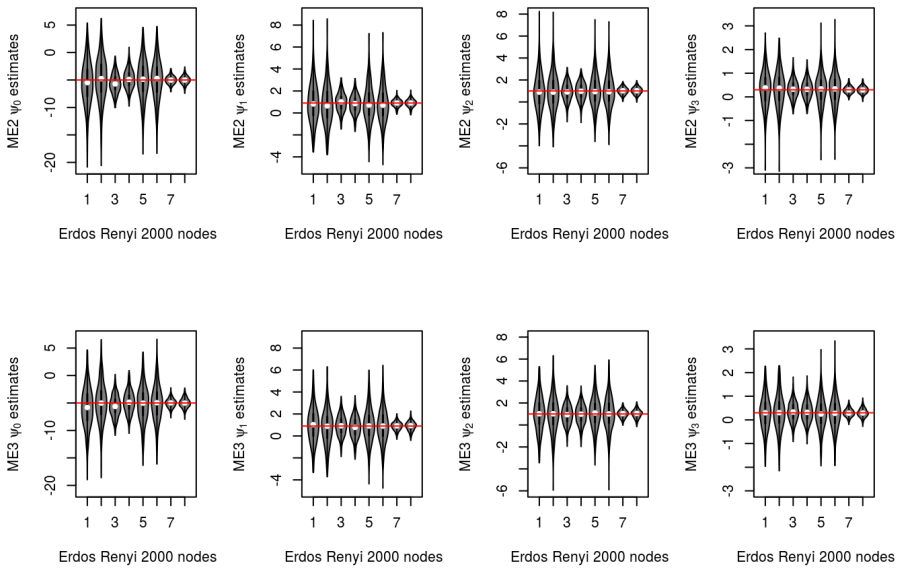


Figure F.6: *Study 2b* (Erdős-Rényi 2000 nodes). Blip function parameter ($\boldsymbol{\psi}$) estimates for 1000 simulated datasets via Methods 2 and 3 when no (scenario 1), one (scenarios 2,3,4,5 and 6), two (scenario 7) or three (scenario 8) of the treatment-free ego (TF-ego), treatment-free alter (TF-alter), and treatment models are correct. Note that Method 2 (ME2) is above and Method 3 (ME3) is below.

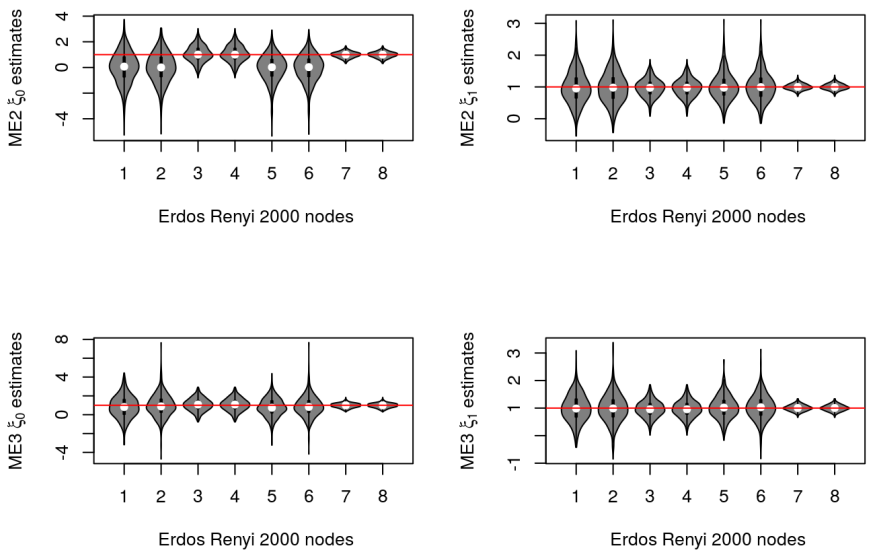


Figure F.7: *Study 2b* (Erdős-Rényi 2000 nodes). Spillover effects parameter (denoted as $\boldsymbol{\xi}$) estimates for 1000 simulated datasets via Methods 2 and 3 when no (scenario 1), one (scenarios 2,3,4,5 and 6), two (scenario 7) or three (scenario 8) of the treatment-free ego (TF-ego), treatment-free alter (TF-alter), and treatment models are correct. Note that Method 2 (ME2) is above and Method 3 (ME3) is below.

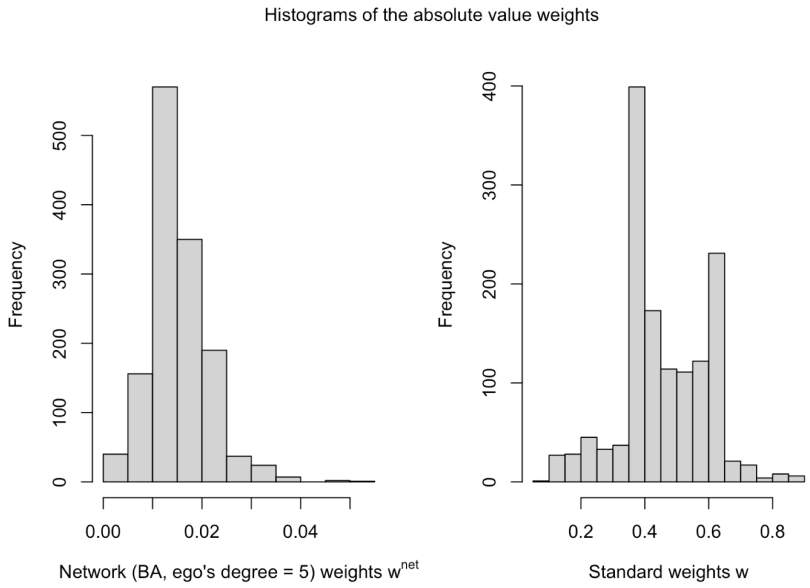


Figure F.8: Histograms of network absolute value weights from BA network with ego's degree is 5 (left) and the standard absolute value weights (right).

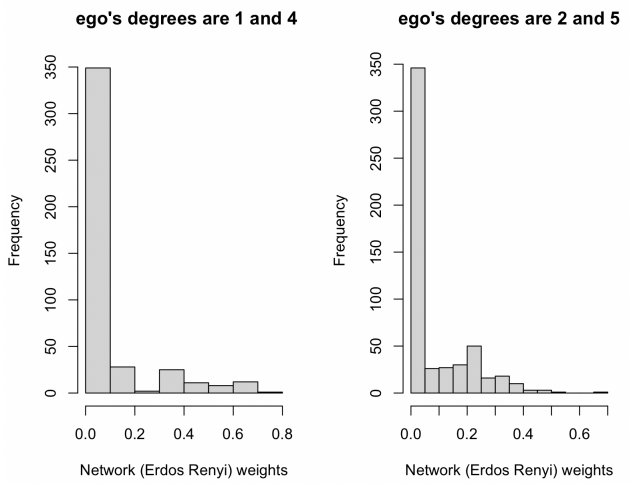


Figure F.9: Histograms of network absolute value weights from ER network (2000 nodes) where the ego's degrees are 1 and 4 (left) and 2 and 5 (right).

covariates relating to the alter, we had investigated eight scenarios (Table F.3) where the egos' degrees are the same and thus the network weights are well-behaved (are centred around a constant value). However, in this simulation, the egos' degrees vary and the network weights have more extreme values and more near-zero values.

Most results from these scenarios concerning apparently consistent estimation appear similar to those of Study 2b. That is, for both cases (Cases I and II), both Method 2 and Method 3 provide apparently consistent estimators of blip parameters $\boldsymbol{\psi}$ and Method 3 provides apparently consistent estimators of spillover parameters $\boldsymbol{\xi}$ as expected. However, in scenarios where the alter's treatment-free component $\exp(\tau(\mathbf{X}_{\mathcal{N}}))$ is misspecified, the more highly variable network weights are associated with low efficiency of the estimators of $\boldsymbol{\psi}$ from both Methods 2 and 3 (Figures F.10 and F.11). Nevertheless, using rescaled weights (dividing the network weight by the average value of the network weight over all nodes of the same degree) in these mixture cases, we observe that Method 3 performs well, and better than Method 2 (Figures F.12 and F.13) in terms of efficiency. Note that these rescaled weights still satisfy the balancing property and reduce the variability of weights overall, making for more efficient estimation.

Figure F.10 shows $\boldsymbol{\psi}$ estimates and Figure F.11 shows $\boldsymbol{\xi}$ estimates from the simulation where the egos' degrees are either 1 or 4. When the treatment model is correct, in both Methods 2 and 3 the $\boldsymbol{\psi}$ parameters appear consistently estimated, even when the treatment-free models are incorrectly specified. However, variances of the estimators of $\boldsymbol{\psi}$ from both Methods 2 and 3 become larger because the egos' degrees can vary. Figure F.12 and Figure F.13 show the blip estimates from the simulation where the egos' degrees are either 1 or 4 and the network weights are rescaled by dividing the network weight by the average value of the network weight over all nodes of the same degree). Both the $\boldsymbol{\psi}$ and $\boldsymbol{\xi}$ parameters appear consistently estimated when this is predicted by the theory. The variances of the estimators of $\boldsymbol{\psi}$ from Method 2 are larger compared with those of the estimators of $\boldsymbol{\psi}$ from Method 3.

F.8 Chapter 3 Appendix A5c: Results of *Study 4b*.

In this section, Figure F.14 shows the blip estimates from simulation *Study 3b*, the two-stage decision estimation for the Longleaf Pine spatial network case. Both the $\boldsymbol{\psi}$ and $\boldsymbol{\xi}$ parameters are consistently estimated, even though the treatment-free models are incorrectly specified in both stages.

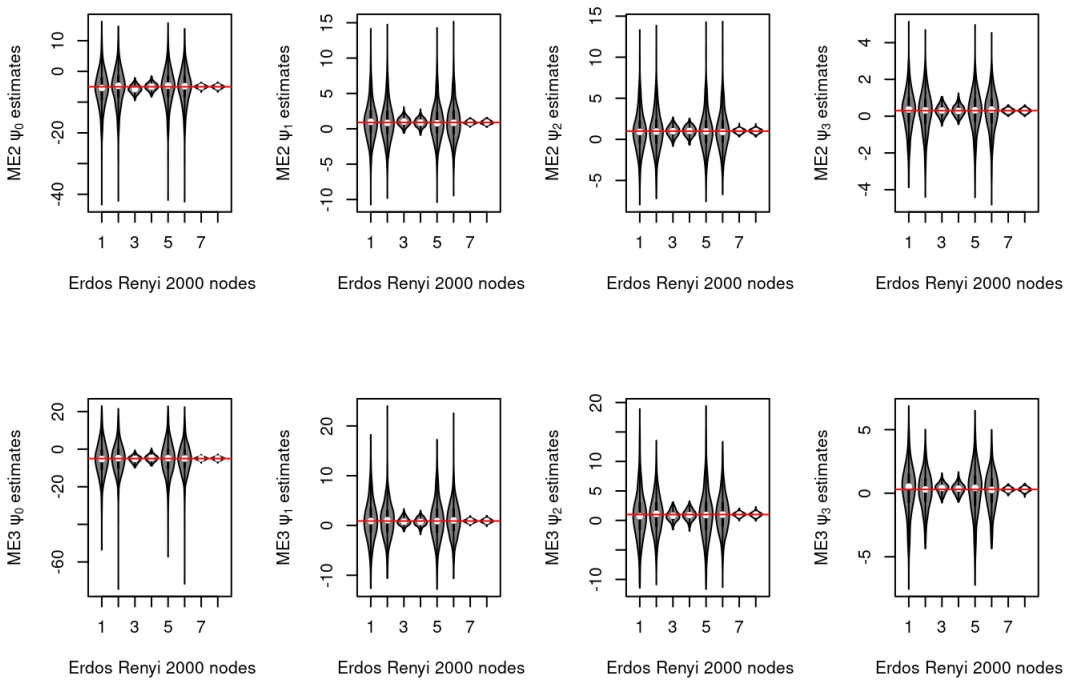


Figure F.10: (Erdős-Rényi 2000 nodes with egos' degrees either 1 or 4). Blip function parameter ($\boldsymbol{\psi}$) estimates for 1000 simulated datasets via Methods 2 and 3 when no (scenario 1), one (scenarios 2,3,4,5 and 6), two (scenario 7) or three (scenario 8) of the treatment-free ego (TF-ego), treatment-free alter (TF-alter), and treatment models are correct. Note that Method 2 (ME2) is above and Method 3 (ME3) is below.

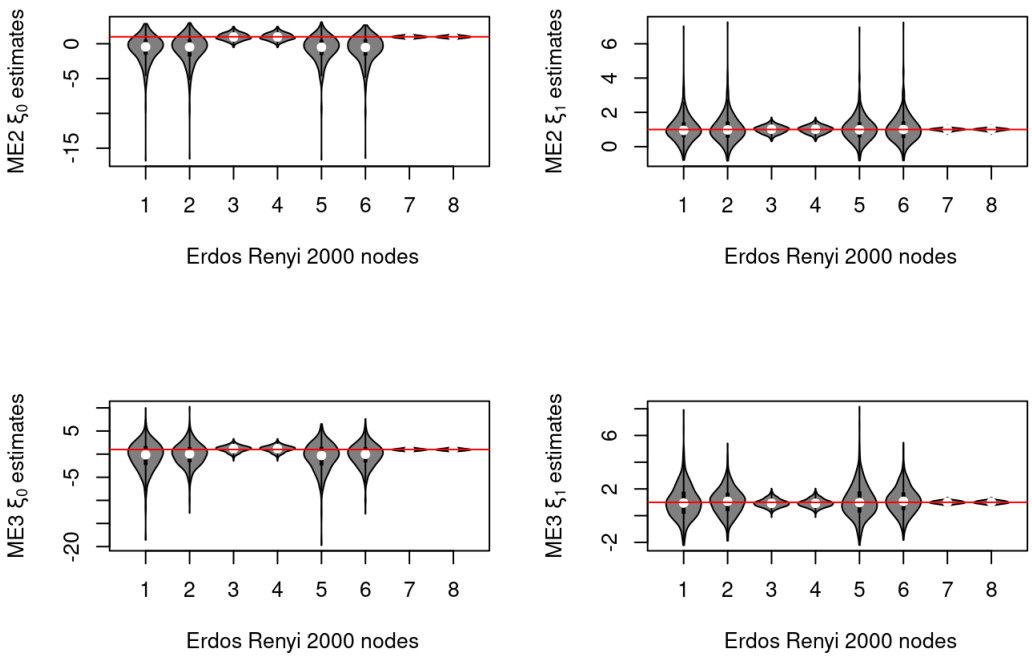


Figure F.11: (Erdős-Rényi 2000 nodes with egos' degrees are either 1 or 4). Spillover effects parameter (denoted as $\boldsymbol{\xi}$) estimates for 1000 simulated datasets via Methods 2 and 3 when no (scenario 1), one (scenarios 2,3,4,5 and 6), two (scenario 7) or three (scenario 8) of the treatment-free ego (TF-ego), treatment-free alter (TF-alter), and treatment models are correct. Note that Method 2 (ME2) is above and Method 3 (ME3) is below.

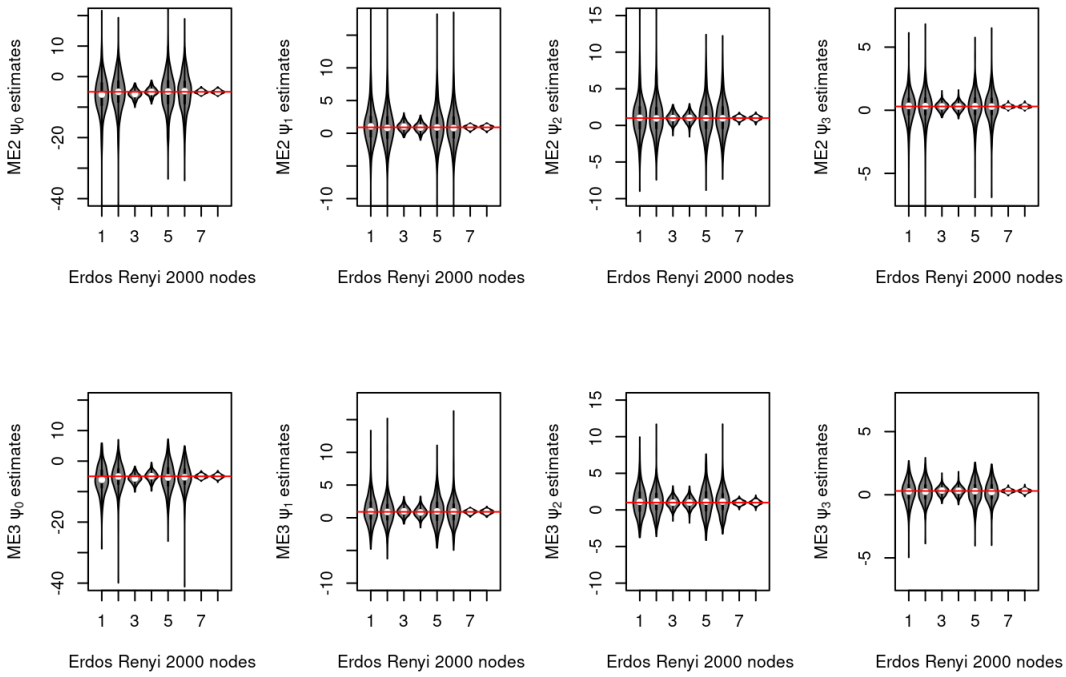


Figure F.12: (Erdős-Rényi 2000 nodes with egos' degrees are either 1 or 4, and using rescaled weight). Blip function parameter ($\boldsymbol{\psi}$) estimates for 1000 simulated datasets via Methods 2 and 3 when no (scenario 1), one (scenarios 2,3,4,5 and 6), two (scenario 7) or three (scenario 8) of the treatment-free ego (TF-ego), treatment-free alter (TF-alter), and treatment models are correct. Note that Method 2 (ME2) is above and Method 3 (ME3) is below.

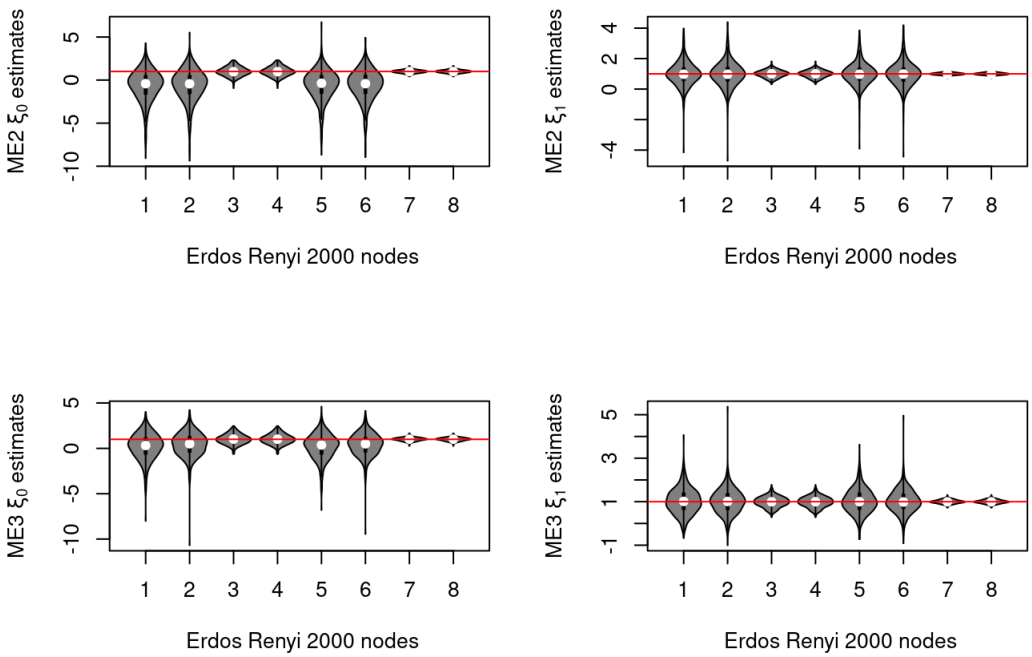


Figure F.13: (Erdős-Rényi 2000 nodes with egos' degrees are either 1 or 4, and using rescaled weight). Spillover effects parameter (denoted as $\boldsymbol{\xi}$) estimates for 1000 simulated datasets via Methods 2 and 3 when no (scenario 1), one (scenarios 2,3,4,5 and 6), two (scenario 7) or three (scenario 8) of the treatment-free ego (TF-ego), treatment-free alter (TF-alter), and treatment models are correct. Note that Method 2 (ME2) is above and Method 3 (ME3) is below.

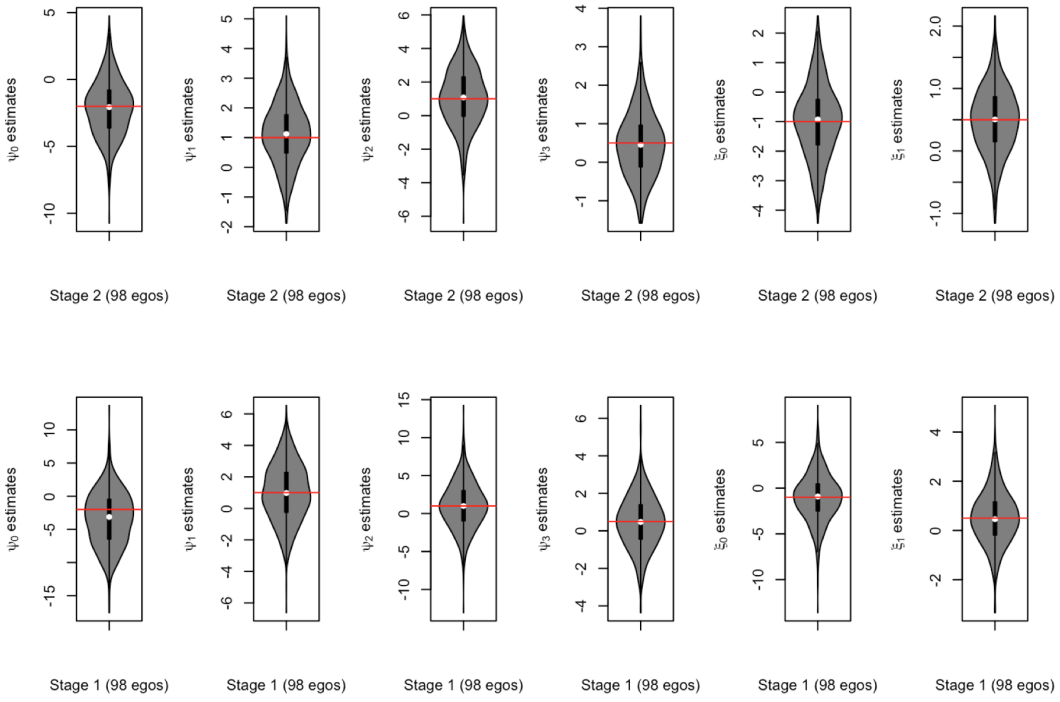


Figure F.14: Two-stage decision blip parameter estimates for the Longleaf Pine network with 98 egos (*Study 3b*). Second stage blip estimates are presented in the top row, and those of the first stage are shown in the bottom row. In both rows, the plots display estimates of $\psi_0, \psi_1, \psi_2, \psi_3, \xi_0, \xi_1$ from left to right.

Appendix G

Appendix of Chapter 4

G.1 Proof of Theorem 4.2.1: Balancing property for binary outcomes

This section presents the proof of Theorem 4.2.1. Building on the proof, the two-step estimation process of the proposed methods is also shown at the end.

Proof. Let us consider the true model: $g(\mathbb{E}[Y|A, X]) = f(X^\beta) + \gamma(A, X^\psi; \boldsymbol{\psi})$, where $Y \in \{0, 1\}$, and $X \in \mathbb{R}^p$ is a vector of p covariates; $X^\beta \in \mathbb{R}^{p^\beta}$ and $X^\psi \in \mathbb{R}^{p^\psi}$ are two (potentially identical) subsets of the variables contained in X; $A \in \{0, 1\}$ is the treatment assignment. Their realized values are y, x, x^β, x^ψ and a, respectively, and we write the propensity score as $\mathbb{P}(A = 1 \mid x) = \pi(x)$. Note that the link function that relates the predictor to the expected value of the random variable Y is denoted as $g(*)$. We assume that $\gamma(A, X^\psi; \boldsymbol{\psi})$ is correctly specified, and set $\gamma(A, X^\psi; \boldsymbol{\psi}) = \boldsymbol{\psi}^\top A X^\psi$.

In the case that $f(X^\beta)$ is linear, as the estimation model assumes, let a latent continuous outcome defining the observed outcome be $Y_c = \boldsymbol{\beta}^\top X^\beta + \boldsymbol{\psi}^\top A X^\psi + \epsilon$, where the error term ϵ could be assumed to follow a logistic distribution conditional on the explanatory variables. This generates the standard logistic model. However, it is not necessary that ϵ has a logistic distribution. It could have a standard normal distribution, yielding a probit model, or another reasonable distribution. The cumulative distribution function of ϵ is the inverse link function, i.e., $g^{-1}(*)$. Define the binary outcome Y as a dichotomization of the latent continuous outcome Y_c, such that $Y = \mathbb{I}(Y_c \geq 0) = \mathbb{I}\left(\epsilon \geq -\boldsymbol{\beta}^\top X^\beta - \boldsymbol{\psi}^\top A X^\psi\right)$, where $\mathbb{I}$ is the indicator function.

Thus $\mathbb{P}(Y = 1 \mid x) = \mathbb{P}(Y_c \geq 0 \mid x) = 1 - g^{-1}(-\boldsymbol{\beta}^\top X^\beta - \boldsymbol{\psi}^\top A X^\psi)$. For example, the logistic model is $\mathbb{P}(Y = 1 \mid x) = \mathbb{P}(Y_c \geq 0 \mid x) = 1 - \frac{\exp\left(0 - \boldsymbol{\beta}^\top X^\beta - \boldsymbol{\psi}^\top A X^\psi\right)}{1 + \exp\left(0 - \boldsymbol{\beta}^\top X^\beta - \boldsymbol{\psi}^\top A X^\psi\right)} = \frac{\exp\left(\boldsymbol{\beta}^\top X^\beta + \boldsymbol{\psi}^\top A X^\psi\right)}{1 + \exp\left(\boldsymbol{\beta}^\top X^\beta + \boldsymbol{\psi}^\top A X^\psi\right)}$. Therefore,

$$\mathbb{P}(Y = y \mid x) = \frac{\exp\left[y(\boldsymbol{\beta}^\top X^\beta + \boldsymbol{\psi}^\top A X^\psi)\right]}{1 + \exp\left(\boldsymbol{\beta}^\top X^\beta + \boldsymbol{\psi}^\top A X^\psi\right)} \text{ for } y = 0, 1,$$

and the log-likelihood for logistic regression is

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\psi} \mid y, x) &= \log \prod_i \frac{\exp\left[y_i(\boldsymbol{\beta}^\top X_i^\beta + \boldsymbol{\psi}^\top A_i X_i^\psi)\right]}{1 + \exp\left(\boldsymbol{\beta}^\top X_i^\beta + \boldsymbol{\psi}^\top A_i X_i^\psi\right)} \\ &= \sum_{i:y_i=1} (\boldsymbol{\beta}^\top X_i^\beta + \boldsymbol{\psi}^\top A_i X_i^\psi) - \sum_i \log\left(1 + \exp\left(\boldsymbol{\beta}^\top X_i^\beta + \boldsymbol{\psi}^\top A_i X_i^\psi\right)\right). \end{aligned}$$

Thus, the score function system components are

$$\sum_i \left(y_i - \frac{\exp\left(\boldsymbol{\beta}^\top X_i^\beta + \boldsymbol{\psi}^\top A_i X_i^\psi\right)}{1 + \exp\left(\boldsymbol{\beta}^\top X_i^\beta + \boldsymbol{\psi}^\top A_i X_i^\psi\right)} \right) X_i^\beta \tag{G.1}$$

and

$$\sum_i \left(y_i - \frac{\exp\left(\boldsymbol{\beta}^\top X_i^\beta + \boldsymbol{\psi}^\top A_i X_i^\psi\right)}{1 + \exp\left(\boldsymbol{\beta}^\top X_i^\beta + \boldsymbol{\psi}^\top A_i X_i^\psi\right)} \right) A_i X_i^\psi. \tag{G.2}$$

Given posited outcome regression model $Q(X, A; \boldsymbol{\beta}, \boldsymbol{\psi}) = g^{-1}(\boldsymbol{\beta}^\top X^\beta + \boldsymbol{\psi}^\top A X^\psi)$, the weighted GLM estimator for $(\boldsymbol{\beta}^\top, \boldsymbol{\psi}^\top)^\top$ is obtained from solving the system of estimating equations: $\sum_i^n U_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}}; A_i, X_i) = \mathbf{0}$, that is,

$$\sum_i^n \begin{pmatrix} X_i^\beta \\ A_i X_i^\psi \end{pmatrix} w(A_i, X_i) \left[Y_i - g^{-1}(\boldsymbol{\psi}^\top A_i X_i^\psi + \boldsymbol{\beta}^\top X_i^\beta) \right] = \sum_i^n \begin{bmatrix} U_{1i}(\boldsymbol{\beta}, \psi; A_i, X_i) \\ U_{2i}(\boldsymbol{\beta}, \psi; A_i, X_i) \end{bmatrix} = \mathbf{0} \tag{G.3}$$

Based on the strong heredity principle (Chipman [1996]), it is required that the treatment-free model must include the main effects for all covariates in the blip model, that is, the tailoring

variables should be a subset of the predictive variables ($X^{\psi} \subseteq X^{\beta}$). For simplicity, and without loss of generality, we assume that $X^{\psi} = X^{\beta} = X$, that is, the covariates in the treatment free and blip components are the same. Then, we assume that for each i independently, X_i for individuals are independent and identically distributed and are generated first, thereafter A_i followed by Y_i. On the one hand, we consider

$$\sum_i^n (U_{1i} - U_{2i}) = \sum_i^n (1 - A_i) X_i w(A_i, X_i) \left[Y_i - g^{-1}(\boldsymbol{\psi}^\top A_i X_i^{\psi} + \boldsymbol{\beta}^\top X_i^{\beta}) \right] \tag{G.4}$$

$$= \sum_i^n (1 - A_i) X_i w(A_i, X_i) \left[Y_i - g^{-1}(\boldsymbol{\beta}^\top X_i^{\beta}) \right] = \mathbf{0} \tag{G.5}$$

where the second equality follows because the only non-zero terms (in equation (G.4)) will be those for which $A_i = 0$. Thus, the left hand side of equation (G.5) (or $\sum_i^n (U_{1i} - U_{2i})$) does not depend on $\boldsymbol{\psi}$. We note that $\sum_i^n (U_{1i} - U_{2i}) = \mathbf{0}$ can be solved for $\boldsymbol{\beta}$, and its solution is denoted by $\hat{\boldsymbol{\beta}}$.

The expectation of $\sum_i^n (U_{1i} - U_{2i})$ conditional on $(A_1, \ldots . A_n)$ and $(X_1, \ldots X_n)$, that is, $\sum_i^n (1 - A_i) X_i w(A_i, X_i) \left[g^{-1}(f(X_i)) - g^{-1}(\boldsymbol{\beta}^\top X_i^{\beta}) \right]$, is not zero unless the true treatment-free model $f(X; \boldsymbol{\beta})$ is linear in X with true coefficient $\boldsymbol{\beta}$. However, the expectation of $\sum_i^n (U_{1i} - U_{2i})$ conditional on $(X_1, \ldots X_n)$ is $\sum_i^n (1 - \pi(X_i)) X_i w(0, X_i) \left[g^{-1}(f(X_i)) - g^{-1}(\boldsymbol{\beta}^\top X_i^{\beta}) \right]$, and if the (unconditional) expectation of $\sum_i^n (1 - \pi(X_i)) X_i w(0, X_i) \left[g^{-1}(f(X_i)) - g^{-1}(\boldsymbol{\beta}^\top X_i^{\beta}) \right]$, i.e., $n\mathbb{E}\left[(1 - \pi(X)) X w(0, X) \left[g^{-1}(f(X)) - g^{-1}(\boldsymbol{\beta}^\top X^{\beta}) \right] \right]$, is $\mathbf{0}$ for $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ and if $\boldsymbol{\beta}^*$ is unique, then, according to large sample theory, $\hat{\boldsymbol{\beta}}$ tends to $\boldsymbol{\beta}^*$ as $n \to \infty$ (White [1982]). Note that the uniqueness of $\boldsymbol{\beta}^*$ is shown in Appendix G.2.

On the other hand, if we consider

$$\sum_i^n U_{2i} = \sum_i^n A_i X_i w(A_i, X_i) \left[Y_i - g^{-1}(\boldsymbol{\psi}^\top A_i X_i + \boldsymbol{\beta}^\top X_i) \right], \tag{G.6}$$

where the only non-zero terms are those for which $A_i = 1$, then $\hat{\boldsymbol{\psi}}$ can be solved in terms of $\hat{\boldsymbol{\beta}}$ from $\sum_i^n U_{2i} = \mathbf{0}$. In order to show that $\hat{\boldsymbol{\psi}}$ is consistent, we would need to show that the expectation of

$$\sum_i^n A_i X_i w(A_i, X_i) \left[g^{-1}(\boldsymbol{\psi}^\top A_i X_i + \boldsymbol{\beta}^{*\top} X_i) - g^{-1}(\boldsymbol{\psi}^\top A_i X_i + f(X_i)) \right]$$

equals or is close to $\mathbf{0}$ for general $\boldsymbol{\psi}$. If this is not the case, then the expectation of the equation $\sum_i^n U_{2i} = \mathbf{0}$ with $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ and $\boldsymbol{\psi} = \hat{\boldsymbol{\psi}}$ may approach the equation $\sum_i^n U_{2i} = \mathbf{0}$ with $\boldsymbol{\beta}$ set equal to $\boldsymbol{\beta}^*$ and $\boldsymbol{\psi}$ set equal to a similar limiting value $\boldsymbol{\psi}^*$ as $n \to \infty$. The vector $\boldsymbol{\psi}^*$ will satisfy the condition that the expectation of $\sum_i^n A_i X_i w(A_i, X_i) \left[g^{-1}(\boldsymbol{\psi}^{*\top} A_i X_i + \boldsymbol{\beta}^{*\top} X_i) - g^{-1}(\boldsymbol{\psi}^{*\top} A_i X_i + f(X_i)) \right]$ equals or is close to $\mathbf{0}$, but $\boldsymbol{\psi}^*$ will in general be different from the true $\boldsymbol{\psi}$.

Let $g^{-1\prime}$ denote the derivative of the inverse link function g. Then the expectation of $\sum_i^n (U_{1i} - U_{2i})$ conditional on $(X_1, \ldots X_n)$, that is, $\sum_i^n (1 - \pi(X_i)) X_i w(0, X_i) \left[g^{-1}(f(X_i)) - g^{-1}(\boldsymbol{\beta}^\top X_i) \right]$, can be written using a Taylor series expansion (function $g^{-1}(f(X_i))$ at the point $\boldsymbol{\beta}^\top X_i$) as

$$\sum_i^n (1 - \pi(X_i)) X_i w(0, X_i) \left[g^{-1\prime}(\boldsymbol{\beta}^\top X_i)(f(X_i) - \boldsymbol{\beta}^\top X_i) + O[(f(X_i) - \boldsymbol{\beta}^\top X_i)^2] \right], \tag{G.7}$$

where the big O describes the error term in an approximation to the g^{-1} function. The notation $O[(f(X_i) - \boldsymbol{\beta}^\top X_i)^2]$ means the absolute-value of the error of $g^{-1}(f(X_i)) - g^{-1}(\boldsymbol{\beta}^\top X_i) - g^{-1\prime}(\boldsymbol{\beta}^\top X_i)(f(X_i) - \boldsymbol{\beta}^\top X_i)$ is at most some constant times $(f(X_i) - \boldsymbol{\beta}^\top X_i)^2$ when $f(X_i) - \boldsymbol{\beta}^\top X_i$ is close enough to 0. Further, the expectation of $\sum_i^n U_{2i}$ conditional on $(X_1, \ldots X_n)$ is

$$\sum_i^n \pi(X_i) X_i w(1, X_i) \left[g^{-1}(\boldsymbol{\psi}^\top X_i + f(X_i)) - g^{-1}(\boldsymbol{\psi}^\top X_i + \boldsymbol{\beta}^\top X_i) \right],$$

which can be written as

$$\sum_i^n \pi(X_i) X_i w(1, X_i) \left[g^{-1\prime}(\boldsymbol{\beta}^\top X_i + \boldsymbol{\psi}^\top X_i)(f(X_i) - \boldsymbol{\beta}^\top X_i) + O[(f(X_i) - \boldsymbol{\beta}^\top X_i)^2] \right]. \tag{G.8}$$

Define $\kappa^*(A, X) = g^{-1\prime}(\boldsymbol{\beta}^{*\top} X + \boldsymbol{\psi}^{*\top} A X)$, where $\boldsymbol{\psi}^*$ is an assumed limiting value for $\hat{\boldsymbol{\psi}}$. Then if weights are defined to satisfy a new balancing criterion $(1 - \pi(X)) w(0, X) \kappa^*(0, X) = \pi(X) w(1, X) \kappa^*(1, X)$, and if the distribution of X is such that the inverse link function is close to linear for the range of $f(X) - \boldsymbol{\beta}^{*\top} X$ (so that the Taylor expansion error term is small), the fact that the expectation of G.7 is $\mathbf{0}$ for $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ means that the expectation of G.8 is close to $\mathbf{0}$ for $\boldsymbol{\beta} = \boldsymbol{\beta}^*$. This argument is what was needed to establish the approximate consistency of the corresponding new estimator of $\boldsymbol{\psi}$.

Therefore, in single-stage decision settings, the algorithm for estimation of $\boldsymbol{\psi}$ is concluded as follows:

Step 1: Conduct a weighted GLM (e.g., logistic regression) to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\psi}}$. Here the weights are from standard dWOLS weights, such as $w(a; x) = |a - \mathbb{E}[A|X = x]|$.

Step 2: Compute the new weights that satisfy $(1 - \pi(X)) w(0, X) \kappa(0, X) = \pi(X) w(1, X) \kappa(1, X)$,

where $\kappa(A, X) = g^{-1\prime}(\hat{\boldsymbol{\beta}}^\top X + \hat{\boldsymbol{\psi}}^\top AX)$, and $g^{-1\prime}$ is identified based on the link function in Step 1. For example, the weights can be

$$w^{new}(a; x) = |a - \mathbb{E}[A|X = x]| * \kappa(1 - A, X).$$

Step 3: Use the new weights from Step 2, and conduct weighted GLM again, to get a new estimator $\widetilde{\boldsymbol{\beta}}$, and using this, an approximately consistent estimator $\widetilde{\boldsymbol{\psi}}$ of $\boldsymbol{\psi}$.

□

Remark: In the proof of Theorem 4.2.1, we show that the consistency of $\widetilde{\boldsymbol{\psi}}$ depends on the small error term of the first order Taylor expansion of g^{-1} about $\boldsymbol{\beta}^\top X$, evaluated at $f(X)$. This error term will be small when $\boldsymbol{\beta}^\top X$ tends to vary in the range where g^{-1} is approximately linear. In practice, as mentioned in the Methodology section, it may be possible to choose the range of X so that $\boldsymbol{\beta}^\top X$ varies in such range, and our estimation will be more precise. Moreover, the constraint of covariates may have an advantage in terms of the overlap (positivity) assumption for the treatment distributions, and thus the possibility of designing the study to constrain the distribution of $\boldsymbol{\beta}^\top X$ is worth exploring more.

G.2 Proof of the uniqueness of $\boldsymbol{\beta}^*$

In Appendix G.1, in the proof of Theorem 4.2.1, we rely on the uniqueness of $\boldsymbol{\beta}^*$ and large sample theory in White [1982], and conclude that $\hat{\boldsymbol{\beta}}$ tends to the unique $\boldsymbol{\beta}^*$ as $n \to \infty$. Now we are going to prove the uniqueness of $\boldsymbol{\beta}^*$.

Proof. Note that $\boldsymbol{\beta}^*$ is defined as

$$\mathbb{E}\left[(1 - \pi(X))Xw^d(0, X)\left[g^{-1}(f(X)) - g^{-1}(\boldsymbol{\beta}^{*\top}X)\right]\right] = \mathbf{0},$$

and is the root of an analogue of a (vector) score function:

$$\boldsymbol{S}(X; \boldsymbol{\beta}) = \mathbb{E}\left[(1 - \pi(X))Xw^d(0, X)\left[g^{-1}(f(X)) - g^{-1}(\boldsymbol{\beta}^\top X)\right]\right].$$

That is, $\boldsymbol{S}(X; \boldsymbol{\beta}^*) = \mathbf{0}$ with the dWOLS weights that satisfy $(1 - \pi(X))w^d(0, X) = \pi(X)w^d(1, X)$. For the continuous outcome where the link function is $g(\mu) = \mu$, the parameter $\boldsymbol{\beta}^*$ from $\boldsymbol{S}(X; \boldsymbol{\beta}^*) = \mathbf{0}$ can be solved explicitly. In the general case, assuming that differentiation with

respect to $\boldsymbol{\beta}$ can be carried through the expectation, then the analogue of the Hessian matrix is

$$\begin{aligned}\boldsymbol{H}(\boldsymbol{\beta}) = \frac{\partial \boldsymbol{S}(X;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= -\mathbb{E}\left[\frac{\partial\left\{(1-\pi(X))Xw^d(0,X)\left[g^{-1}(\boldsymbol{\beta}^\top X)\right]\right\}}{\partial \boldsymbol{\beta}}\right] \\ &= -\mathbb{E}\left[R(0,X)X\kappa(0,X)X^\top\right],\end{aligned}$$

where $\kappa(A,X) = g^{-1\prime}(\boldsymbol{\beta}^\top X + \boldsymbol{\psi}^\top AX)$ and $R(0,X) = (1-\pi(X))w^d(0,X)$. Then we have the following key corollary to show the uniqueness of $\boldsymbol{\beta}^*$.

The matrix $\boldsymbol{H}(\boldsymbol{\beta})$ is negative semi-definite for any $\boldsymbol{\beta} \in \mathbb{R}^p$ if the link function is monotone increasing, that is, $\kappa(0,X) = g^{-1\prime}(\boldsymbol{\beta}^\top X) > 0$ for any $\boldsymbol{\beta}^\top X$.

Proof: for any $\boldsymbol{u} \in \mathbb{R}^p$, we have

$$\boldsymbol{u}^\top \boldsymbol{H}(\boldsymbol{\beta})\boldsymbol{u} = -\mathbb{E}[\sum_r^p (x_r u_r)^2 R(0,X)\kappa(0,X)],$$

where $R(0,X)$ is positive. Further, if the first derivative of g^{-1} is always positive, then the above expression $\boldsymbol{u}^\top \boldsymbol{H}(\boldsymbol{\beta})\boldsymbol{u} \leq \mathbf{0}$ for all $\boldsymbol{u} \in \mathbb{R}^p$ and $\boldsymbol{\beta} \in \mathbb{R}^p$ with equality holding when $\boldsymbol{u} = 0$. Thus, the matrix $\boldsymbol{H}(\boldsymbol{\beta})$ is negative definite for all $\boldsymbol{\beta}$ and the corresponding log-likelihood function analogue is strictly concave.

Therefore $\boldsymbol{\beta}^*$ satisfying $\boldsymbol{S}(X;\boldsymbol{\beta}^*) = \mathbf{0}$ would be the unique root. Note that for the linear treatment-free case, i.e., $f(X) = \boldsymbol{\beta}_0^\top X$, then we have $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$.

Note that our dWGLM approach involves two-step regression to consistently estimate the parameter of interest. Each step uses GLM for binary outcomes (e.g., logistic regression) to estimate the parameter. The above argument can be used to show that the first step (logistic) regression provides unique first stage estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\psi}}$ for the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$. Building on these first stage estimates and weights equation (4.4), we can get new weights, which satisfy weights criterion (4.2) and are always positive. Using the same argument as for the uniqueness of $\boldsymbol{\beta}^*$, we can prove the uniqueness of the limiting value $\widetilde{\boldsymbol{\beta}^*}$ of the second step parameter denoted as $\widetilde{\boldsymbol{\beta}^*}$ when the new weights w^{new} are used. Note that, $\widetilde{\boldsymbol{\beta}^*}$ is defined through $\mathbb{E}\left((1-\pi(X))Xw^{new}(0,X)\left[g^{-1}(f(X)) - g^{-1}(\widetilde{\boldsymbol{\beta}^*}^\top X)\right]\right) = \mathbf{0}$.

□

G.3 Derivation of the First-stage True DTR Parameters

In this section, we derive the true values of the first-stage decision rule parameters in terms of the data generating parameters (Moodie et al. [2014]). Following the notations in Section 4.3.2 (Study 2b), let $f_2 = \theta_0 + \theta_1 X_1 + \theta_2 O_1 + \theta_3 A_1 + \theta_4 O_1 A_1 + \theta_5 X_2 + \varphi_1(X_1) + \varphi_2(X_2)$, then we have $\mathbb{P}(\widetilde{\mathcal{Y}_1} = 1) = expit\left(f_2 + |\theta_6 + \theta_7 O_2 + \theta_8 A_1|^+\right)$, which equals to

$$expit\left(f_2 + O_2 A_1 |\phi_1|^+ + O_2(1-A_1)|\phi_2|^+ + (1-O_2)A_1|\phi_3|^+ + (1-O_2)(1-A_1)|\phi_4|^+\right),$$

where $|x|^+ = x * \mathbb{I}(x > 0)$, and $\phi_1 = \theta_6 + \theta_7 + \theta_8$, $\phi_2 = \theta_6 + \theta_7$, $\phi_3 = \theta_6 + \theta_8$, $\phi_4 = \theta_6$ for binary variables O_2 and A_1 in $\{0, 1\}$. Furthermore, $\mathbb{E}(O_2 \mid O_1, A_1) = expit(\delta_1 O_1 + \delta_2 A_1) = 1 - \mathbb{E}(1 - O_2 \mid O_1, A_1)$, thus, $Q_1(H_1, A_1) = \mathbb{E}(\widetilde{\mathcal{Y}_1} \mid H_1, A_1)$, which equals to

$$\begin{aligned} expit\big(&f_2 + \mathbb{E}(O_2 \mid H_1, A_1)[A_1|\phi_1|^+ + (1-A_1)|\phi_2|^+] \\ &+ \mathbb{E}(1-O_2 \mid O_1, A_1)[A_1|\phi_3|^+ + (1-A_1)|\phi_4|^+]\big). \end{aligned}$$

Moreover,

$$\begin{aligned} expit(\delta_1 O_1 + \delta_2 A_1) &= O_1 A_1 expit(\delta_1 + \delta_2) + O_1(1-A_1)expit(\delta_1) \\ &\quad + (1-O_1)A_1 expit(\delta_2) + (1-O_1)(1-A_1)expit(0) \\ &:= O_1 A_1 k_1 + O_1(1-A_1)k_2 + (1-O_1)A_1 k_3 + (1-O_1)(1-A_1)k_4, \end{aligned}$$

where $k_1 = expit(\delta_1 + \delta_2), k_2 = expit(\delta_1), k_3 = expit(\delta_2), k_4 = expit(0)$, and $A_1^2 = A_1, (1 - A_1)^2 = 1 - A_1, A_1(1 - A_1) = 0$. Therefore, we have

$$\begin{aligned} logit[\mathbb{E}(\widetilde{\mathcal{Y}_1} \mid h_1, a_1)] = f_2 &+ |\phi_4|^+ + \left(|\phi_3|^+ - |\phi_4|^+\right) A_1 \\ &+ O_1 A_1 k_1 \left(|\phi_1|^+ - |\phi_3|^+\right) + O_1(1-A_1)k_2\left(|\phi_2|^+ - |\phi_4|^+\right) \\ &+ (1-O_1)A_1 k_3\left(|\phi_1|^+ - |\phi_3|^+\right) + (1-O_1)(1-A_1)k_4\left(|\phi_2|^+ - |\phi_4|^+\right). \end{aligned}$$

Therefore, for the true blip parameters $\boldsymbol{\psi}_1 = (\psi_{10}, \psi_{11})^\top$, the above equation gives the coefficient of A_1 as

$$\psi_{10} = \theta_3 + |\phi_3|^+ - |\phi_4|^+ + k_3\left(|\phi_1|^+ - |\phi_3|^+\right) - k_4\left(|\phi_2|^+ - |\phi_4|^+\right),$$

and the coefficient of $O_1 A_1$ as

$$\psi_{11} = \theta_4 + (k_1 - k_3)\left(|\phi_1|^+ - |\phi_3|^+\right) - (k_2 - k_4)\left(|\phi_2|^+ - |\phi_4|^+\right).$$

Appendix H

Appendix of Chapter 5

H.1 The double robustness of dWOLS and dWGLM in the presence of household interference

In this section, considering two-person household interference and aiming to optimize outcomes of egos, we propose an approach to making treatment decisions for egos, taking into account the alters' treatment information. However, we emphasize that our model can also easily be extended to the household utility case where the optimal decisions for both individuals in the same household are to be made (See the method proposed in Chapter 5). Note that, in the following subsections, we demonstrate the models and proof with the notation of a pair or couple (s, r). For the ego and alter case, without loss of generality, we just need to consider $s = ego$, $r = alter$, and $a^s = a, a^r = a^{alter}$; then the essence and results are the same. This section is divided into two subsections: one is for the ego's continuous outcome and the other is for the ego's binary outcome. The double robustness of our methods relies on the balancing property; therefore, for simplicity, in this section, we primarily present the balancing property theorem for each case and provide the corresponding proof. In particular, in the second subsection (binary outcome case), we also present a dWGLM in the presence of household interference.

H.1.1 The double robustness of dWOLS in the presence of household interference

In this section, considering continuous outcomes and household interference, which is illustrated in DAGs 5.1, we demonstrate the double robustness of dWOLS in a new outcome model scenario.

Let us consider a pair (s, r) in the same household, and the true continuous-outcome model:

$$\mathbb{E}[Y \mid A^s = a^s, A^r = a^r, \boldsymbol{x}] = f(\boldsymbol{x}^\beta; \boldsymbol{\beta}) + a^s \boldsymbol{\xi}^\top \boldsymbol{x}^\xi + a^r \boldsymbol{\psi}^\top \boldsymbol{x}^\psi + a^s a^r \boldsymbol{\phi}^\top \boldsymbol{x}^\phi, \tag{H.1}$$

where, as previously discussed, $\boldsymbol{x}^\beta$ are often termed predictive variables to function as increasing the precision of estimates, and $\boldsymbol{x}^\xi$, $\boldsymbol{x}^\psi$, $\boldsymbol{x}^\phi$ are so-called tailoring variables for adapting treatment decisions to individuals in the same household. They are four (potentially identical) subsets of the variables contained in $\boldsymbol{X}_h$, covariates that contain individual information of both s and r in the same household h; (A^s, A^r) where $A^s, A^r \in \{0, 1\}$ is the treatment assignment. Here and subsequently we assume that the variables in $\boldsymbol{x}^\xi$, $\boldsymbol{x}^\psi$, and $\boldsymbol{x}^\phi$ are contained in those of $\boldsymbol{x}^\beta$. The realized values are y, $\boldsymbol{x}_h$, a^s and a^r, and we write the joint propensity score as $\mathbb{P}(A^s = a^s, A^r = a^s \mid \boldsymbol{x}) = \pi^{a^s a^r}(\boldsymbol{x})$. We assume that the interference-based blip function $\gamma[(A^s, A^r), \boldsymbol{x}]$ is correctly specified, and set $\gamma[(A^s, A^r), \boldsymbol{x}] = A^s \boldsymbol{\xi}^\top \boldsymbol{x}^\xi + A^r \boldsymbol{\psi}^\top \boldsymbol{x}^\psi + A^s A^r \boldsymbol{\phi}^\top \boldsymbol{x}^\phi$.

Theorem H.1. *Balancing Property with Household Interference for Continuous Outcomes*

When the true outcome model satisfies $\mathbb{E}[Y \mid A^s = a^s, A^r = a^r, \boldsymbol{x}] = f(\boldsymbol{x}^\beta; \boldsymbol{\beta}) + a^s \boldsymbol{\xi}^\top \boldsymbol{x}^\xi + a^r \boldsymbol{\psi}^\top \boldsymbol{x}^\psi + a^s a^r \boldsymbol{\phi}^\top \boldsymbol{x}^\phi$*, a weighted ordinary least squares regression based on the corresponding linear model will yield consistent estimators of* $\boldsymbol{\xi}$*,* $\boldsymbol{\psi}$ *as well as* $\boldsymbol{\phi}$ *if at least one of the treatment and treatment-free models is correctly specified, and the weights satisfy*

$$\pi^{00} w(0, 0, \boldsymbol{x}) = \pi^{01} w(0, 1, \boldsymbol{x}) = \pi^{10} w(1, 0, \boldsymbol{x}) = \pi^{11} w(1, 1, \boldsymbol{x}).$$

Proof: Given the posited interference-aware outcome regression model $Q(X, A^s, A^r; \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}) = \boldsymbol{\beta}^\top X^\beta + A^s \boldsymbol{\xi}^\top \boldsymbol{x}^\xi + A^r \boldsymbol{\psi}^\top \boldsymbol{x}^\psi + A^s A^r \boldsymbol{\phi}^\top \boldsymbol{x}^\phi$, the weighted OLS estimator for $(\boldsymbol{\beta}^\top, \boldsymbol{\xi}^\top, \boldsymbol{\psi}^\top, \boldsymbol{\phi}^\top)$ is obtained by solving the system of estimating equations: $\sum_h^H U_h(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\phi}}; A_h^s, A_h^r, X_h) = \mathbf{0}$, that is,

$$\begin{aligned}
&\sum_h^H \begin{pmatrix} X_h^\beta \\ A_h^s X_h^\xi \\ A_h^r X_h^\psi \\ A_h^s A_h^r X_h^\phi \end{pmatrix} w(A_h^s, A_h^r, X_h) \left(Y_h - \boldsymbol{\beta}^\top X_h^\beta - A^s \boldsymbol{\xi}^\top X_h^\xi - A^r \boldsymbol{\psi}^\top X_h^\psi - A^s A^r \boldsymbol{\phi}^\top X_h^\phi \right) \\
&= \sum_h^H \begin{bmatrix} U_{1h}(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}; A_h^s, A_h^r, X_h) \\ U_{2h}(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}; A_h^s, A_h^r, X_h) \\ U_{3h}(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}; A_h^s, A_h^r, X_h) \\ U_{4h}(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}; A_h^s, A_h^r, X_h) \end{bmatrix} = \mathbf{0}
\end{aligned} \tag{H.2}$$

We first assume that $X^\beta = X^\xi = X^\psi = X^\phi = X$; that is, the covariates in the treatment-free and

blip part are the same. Then, we assume that for each household h independently, X_h for each household is generated first; thereafter, $A_h = (A_h^s, A_h^r)$ is generated followed by Y_h. On one hand, denoting $\epsilon_h = Y_h - \boldsymbol{\beta}^\top X_h - A^s \boldsymbol{\xi}^\top X_h - A^r \boldsymbol{\psi}^\top X_h - A^s A^r \boldsymbol{\phi}^\top X_h$, we consider

$$\begin{aligned}\sum_h^H (U_{1h} - U_{2h} - U_{3h} + U_{4h}) &= \sum_h^H (1 - A_h^s)(1 - A_h^r) X_h w(A_h^s, A_h^r, X_h) \epsilon_h \\ &= \sum_h^H (1 - A_h^s)(1 - A_h^r) X_h w(A_h^s, A_h^r, X_h) \left(Y_h - \boldsymbol{\beta}^\top X_h\right) = \mathbf{0}, \quad \text{(H.3)}\end{aligned}$$

where the second equality follows because the only non-zero terms (in equation (H.3)) will be those for which $(A_h^s, A_h^r) = (0, 0)$. Thus, equation (H.3) (i.e., $\sum_h^H (U_{1h} - U_{2h} - U_{3h} + U_{4h})$), which only involves with $\boldsymbol{\beta}$, can be solved for $\boldsymbol{\beta}$, and its solution is

$$\hat{\boldsymbol{\beta}} = \left(\sum_h^H (1 - A_h^s)(1 - A_h^r) w(A_h^s, A_h^r, X_h) X_h X_h^\top \right)^{-1} \sum_h^H (1 - A_h^s)(1 - A_h^r) w(A_h^s, A_h^r, X_h) X_h Y_h.$$

Similarly, we have both

$$\begin{aligned}\sum_h^H (U_{2h} - U_{4h}) &= \sum_h^H (1 - A_h^r) A_h^s X_h w(A_h^s, A_h^r, X_h) \epsilon_h \\ &= \sum_h^H (1 - A_h^r) A_h^s X_h w(A_h^s, A_h^r, X_h) \left(Y_h - \boldsymbol{\xi}^\top A_h^s X_h - \boldsymbol{\beta}^\top X_h\right) = \mathbf{0}, \quad \text{(H.4)}\end{aligned}$$

and

$$\begin{aligned}\sum_h^H (U_{3h} - U_{4h}) &= \sum_h^H (1 - A_h^s) A_h^r X_h w(A_h^s, A_h^r, X_h) \epsilon_h \\ &= \sum_h^H (1 - A_h^s) A_h^r X_h w(A_h^s, A_h^r, X_h) \left(Y_h - \boldsymbol{\psi}^\top A_h^r X_h - \boldsymbol{\beta}^\top X_h\right) = \mathbf{0}, \quad \text{(H.5)}\end{aligned}$$

where the second equality follows because the only non-zero terms in Equations (H.4) and (H.5) will be those for which $(A_h^s, A_h^r) = (1, 0)$ and $(A_h^s, A_h^r) = (0, 1)$, respectively.

Conditional on both treatments and covariates, i.e., $(A_1^s, A_1^r, ..., A_H^s, A_H^r)$ and $(X_1, ...X_H)$, the expectation of $\sum_h^H (U_{1h} - U_{2h} - U_{3h} + U_{4h})$ is $\sum_h^H (1-A_h^s)(1-A_h^r) X_h w(A_h^s, A_h^r, X_h) \left(f(X_h) - \boldsymbol{\beta}^\top X_h\right)$, and it is not zero unless the true treatment-free model $f(X; \boldsymbol{\beta})$ is linear in X with true coefficient $\boldsymbol{\beta}$. However, the expectation of $\sum_h^H (U_{1h} - U_{2h} - U_{3h} + U_{4h})$ conditional on $(X_1, ...X_n)$ is $\sum_h^H \pi^{00}(X_h) X_h w(0,0,X_h) \left(f(X_h) - \boldsymbol{\beta}^\top X_h\right)$, and if the unconditional expectation of $\sum_h^H \pi^{00}(X_h) X_h w(0,0,X_h) \left(f(X_h) - \boldsymbol{\beta}^\top X_h\right)$, $H\mathbb{E}_X \left[\pi^{00}(X) X w(0,0,X) \left(f(X) - \boldsymbol{\beta}^\top X\right)\right]$, is $\mathbf{0}$ for $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, then, according to the large sample theory approximation, $\hat{\boldsymbol{\beta}}$ tends to converge to $\boldsymbol{\beta}^*$ as $H \to \infty$ (White [1982]).

In addition, we consider $\sum_h^H (U_{2h} - U_{4h}) = \sum_h^H (1 - A_h^r) A_h^s X_h w(A_h^s, A_h^r, X_h) \left(Y_h - \boldsymbol{\xi}^\top A_h^s X_h - \boldsymbol{\beta}^\top X_h\right)$, where the only non-zero terms are those for which $(A_h^s, A_h^r) = (1,0)$, then $\hat{\boldsymbol{\xi}}$ can be solved in terms of $\hat{\boldsymbol{\beta}}$ from $\sum_h^H (U_{2h} - U_{4h}) = \mathbf{0}$. Similarly, considering $\sum_h^H (U_{3h} - U_{4h}) = \sum_h^H (1 - A_h^s) A_h^r X_h w(A_h^s, A_h^r, X_h) \left(Y_h - \boldsymbol{\psi}^\top A_h^r X_h - \boldsymbol{\beta}^\top X_h\right)$, we can solve $\hat{\boldsymbol{\psi}}$ in terms of $\hat{\boldsymbol{\beta}}$ from $\sum_h^H (U_{3h} - U_{4h}) = \mathbf{0}$.

Note that, from the dWOLS weights criterion $\pi^{00} w(0,0,X_h) = \pi^{10} w(1,0,X_h)$, the expectation of $\sum_h^H (U_{2h} - U_{4h})$ conditional on $(X_1, ...X_H)$, that is, $\sum_h^H \pi^{10}(X_h) X_h w(1,0,X_h) \left(f(X_h) - \boldsymbol{\beta}^\top X_h\right)$, equals $\sum_h^H \pi^{00}(X_h) X_h w(0,0,X_h) \left(f(X_h) - \boldsymbol{\beta}^\top X_h\right)$, where the unconditional expectation of $\sum_h^H \pi^{00}(X_h) X_h w(0,0,X_h) \left(f(X_h) - \boldsymbol{\beta}^\top X_h\right)$ is $\mathbf{0}$ for $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ and for any $\boldsymbol{\xi}$. The equation $\sum_h^H (U_{2h} - U_{4h}) = \mathbf{0}$ with $\boldsymbol{\beta}$ set equal to $\hat{\boldsymbol{\beta}}$ will approach the equation $\sum_h^H (U_{2h} - U_{4h}) = \mathbf{0}$ with $\boldsymbol{\beta}$ set equal to $\boldsymbol{\beta}^*$ as $H \to \infty$. Correspondingly, building on weights criterion $\pi^{00} w(0,0,X_h) = \pi^{01} w(0,1,X_h)$, the equation $\sum_h^H (U_{3h} - U_{4h}) = \mathbf{0}$ with $\boldsymbol{\beta}$ set equal to $\hat{\boldsymbol{\beta}}$ will approach the equation $\sum_h^H (U_{3h} - U_{4h}) = \mathbf{0}$ with $\boldsymbol{\beta}$ set equal to $\boldsymbol{\beta}^*$ as $H \to \infty$.

Finally, we consider

$$\sum_h^H (U_{4h}) = \sum_h^H A_h^s A_h^r X_h w(A_h^s, A_h^r, X_h) \left(Y_h - \boldsymbol{\beta}^\top X_h^\beta - A^s \boldsymbol{\xi}^\top X_h^\xi - A^r \boldsymbol{\psi}^\top X_h^\psi - A^s A^r \boldsymbol{\phi}^\top X_h^\phi\right),$$

where the only non-zero terms are those for which $(A_h^s, A_h^r) = (1,1)$. Then $\hat{\boldsymbol{\phi}}$ can be solved in terms of $\hat{\boldsymbol{\xi}}$, $\hat{\boldsymbol{\psi}}$, and $\hat{\boldsymbol{\beta}}$ from $\sum_h^H U_{2h} = \mathbf{0}$. Analogously to previous analysis, from the dWOLS weights criterion $\pi^{00} w(0,0,X_h) = \pi^{11} w(1,1,X_h)$, the expectation of $\sum_h^H U_{4h}$ conditional on $(X_1, ...X_h)$, that is, $\sum_h^H \pi^{11}(X_h) X_h w(1,1,X_h) \left(f(X_h) - \boldsymbol{\beta}^\top X_h\right)$, equals $\sum_h^H \pi^{00}(X_h) X_h w(0,0,X_h) \left(f(X_h) - \boldsymbol{\beta}^\top X_h\right)$, where the unconditional expectation of $\sum_h^H \pi^{00}(X_h) X_h w(0,0,X_h) \left(f(X_h) - \boldsymbol{\beta}^\top X_h\right)$ is $\mathbf{0}$ for $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ and for any $\boldsymbol{\phi}$. The equation $\sum_h^H U_{4h} = \mathbf{0}$ with $\boldsymbol{\beta}$ set equal to $\hat{\boldsymbol{\beta}}$ will approach the equation $\sum_h^H U_{4h} = \mathbf{0}$ with $\boldsymbol{\beta}$ set equal to $\boldsymbol{\beta}^*$ as $H \to \infty$.

We find estimators $(\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\xi}}^\top, \hat{\boldsymbol{\psi}}^\top, \hat{\boldsymbol{\phi}}^\top)$ through solving the system of estimating equations $\sum_h^H U_h(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}}; A_h, X_h) = \mathbf{0}$. Supposing the treatment-free model $f(X^\beta; \boldsymbol{\beta})$ is mis-specified, we now show that $(\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\xi}}^\top, \hat{\boldsymbol{\psi}}^\top, \hat{\boldsymbol{\phi}}^\top)$ solving the estimating equations (H.2) converges in probability to $(\boldsymbol{\beta}^{*\top}, \boldsymbol{\xi}^\top, \boldsymbol{\psi}^\top, \boldsymbol{\phi}^\top)$, where $\boldsymbol{\beta}^*$ is the solution to (unconditional expectation) $\mathbb{E}\left[\pi^{00}(X))Xw(0,0,X)\left(f(X) - \boldsymbol{\beta}^{*\top}X\right)\right] = \mathbf{0}$. Because $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}^*$, then from the theory of M-estimation (Huber [2004], Geer and van de Geer [2000], and Tsiatis [2019]), under regularity conditions, this follows if the system of estimating functions corresponding to estimating equations (H.2) has a mean of zero when evaluated at $(\boldsymbol{\beta}^{*\top}, \boldsymbol{\xi}^\top, \boldsymbol{\psi}^\top, \boldsymbol{\phi}^\top)$; that is,

$$\begin{aligned}\mathbb{E}[U(\boldsymbol{\beta}^*, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}; A^s, A^r, X)] &= \mathbb{E}\left[\begin{pmatrix} X \\ A^sX \\ A^rX \\ A^sA^rX \end{pmatrix} w(A^s, A^r, X)\epsilon\right] \\ &= \mathbb{E}\begin{bmatrix} U_1(\boldsymbol{\beta}^*, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}; A^s, A^r, X) \\ U_2(\boldsymbol{\beta}^*, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}; A^s, A^r, X) \\ U_3(\boldsymbol{\beta}^*, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}; A^s, A^r, X) \\ U_4(\boldsymbol{\beta}^*, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}; A^s, A^r, X) \end{bmatrix} = \mathbf{0};\end{aligned}$$

where $\epsilon = \left(Y - \boldsymbol{\beta}^{*\top}X - \boldsymbol{\xi}^\top A^sX - \boldsymbol{\psi}^\top A^rX - \boldsymbol{\phi}^\top A^sA^rX\right)$.

Equivalently,

$$\mathbb{E}\begin{bmatrix} U_1 - U_2 - U_3 + U_4 \\ U_2 - U_4 \\ U_3 - U_4 \\ U_4 \end{bmatrix} = \mathbb{E}\left[\begin{pmatrix} (1-A^s)(1-A^r) \\ (1-A^r)A^s \\ (1-A^s)A^r \\ A^rA^s \end{pmatrix} Xw(A^s, A^r, X)\epsilon\right] = \mathbf{0}. \tag{H.6}$$

Applying iterated conditional expectations, conditioning on (A^s, A^r, X), $\mathbb{E}\left[\begin{pmatrix} (1-A^s)(1-A^r) \\ (1-A^r)A^s \\ (1-A^s)A^r \\ A^rA^s \end{pmatrix} Xw(A^s, A^r, X)\epsilon\right]$ in (H.6) is $\mathbb{E}\left[\begin{pmatrix} (1-A^s)(1-A^r) \\ (1-A^r)A^s \\ (1-A^s)A^r \\ A^rA^s \end{pmatrix} Xw(A^s, A^r, X)\left(f(X) - \boldsymbol{\beta}^{*\top}X\right)\right]$. Then, taking conditional expectation

with respect to (A^s, A^r) given X, we have

$$\mathbb{E}\begin{bmatrix} \pi^{00}(X)Xw(0,0,X)\left(f(X)-\boldsymbol{\beta}^{*\top}X\right) \\ \pi^{10}(X)Xw(1,0,X)\left(f(X)-\boldsymbol{\beta}^{*\top}X\right) \\ \pi^{01}(X)Xw(0,1,X)\left(f(X)-\boldsymbol{\beta}^{*\top}X\right) \\ \pi^{11}(X)Xw(1,1,X)\left(f(X)-\boldsymbol{\beta}^{*\top}X\right) \end{bmatrix}. \tag{H.7}$$

By definition of $\boldsymbol{\beta}^*$, the first expectation in (H.7) is zero; then the second to the fourth expectations in (H.7) are also zeros if the interference-aware dWOLS weights criterion, that is, $\pi^{00}w(0,0,X) = \pi^{01}w(0,1,X) = \pi^{10}w(1,0,X) = \pi^{11}w(1,1,X)$, is fulfilled and the joint propensity score is correctly specified.

In the case of $X^\beta = X^\xi = X^\psi = X^\phi = X$, we conclude that $\hat{\boldsymbol{\xi}}$, $\hat{\boldsymbol{\psi}}$, $\hat{\boldsymbol{\phi}}$ solving the system of estimating equations (H.2) is a consistent estimator for the true values $\boldsymbol{\xi}$, $\boldsymbol{\psi}$ $\boldsymbol{\phi}$, respectively, if either the treatment-free model or the joint propensity model is correctly specified. The case where X^ξ, X^ψ, and X^ϕ are subsets of X^β can be handled analogously to the case where $\boldsymbol{x}^\psi$ is a subset of $\boldsymbol{x}^\beta$ in Chapter 3.

H.1.2 The double robustness of dWGLM in the presence of household interference

In the case of binary outcomes, we assume that the true model is

$$g\left[\mathbb{P}(Y=1 \mid \boldsymbol{x}, A^s, A^r; \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi})\right] = f(\boldsymbol{x}^\beta) + A^s\boldsymbol{\xi}^\top\boldsymbol{x}^\xi + A^r\boldsymbol{\psi}^\top\boldsymbol{x}^\psi + A^sA^r\boldsymbol{\phi}^\top\boldsymbol{x}^\phi, \tag{H.8}$$

for an arbitrary treatment-free function $f(\boldsymbol{x}^\beta)$. However, for estimation we employ an interference-aware generalized linear model in which:

$$g\left[\mathbb{P}(Y=1 \mid \boldsymbol{x}, A^s, A^r; \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi})\right] = \boldsymbol{\beta}^\top\boldsymbol{x}^\beta + A^s\boldsymbol{\xi}^\top\boldsymbol{x}^\xi + A^r\boldsymbol{\psi}^\top\boldsymbol{x}^\psi + A^sA^r\boldsymbol{\phi}^\top\boldsymbol{x}^\phi,$$

where g is the link function for the binary outcomes.

Analogously to the non-interference binary outcome analysis of Chapter 4, we have the following theorem:

Theorem H.2. *Balancing Property with Household Interference for Binary Outcomes*

When the true outcome model satisfies $\mathbb{P}[Y=1 \mid A^s=a^s, A^r=a^r, \boldsymbol{x}] = f(\boldsymbol{x}^\beta;\boldsymbol{\beta}) + a^s\boldsymbol{\xi}^\top\boldsymbol{x}^\xi + a^r\boldsymbol{\psi}^\top\boldsymbol{x}^\psi + a^sa^r\boldsymbol{\phi}^\top\boldsymbol{x}^\phi$*, a weighted ordinary least squares regression based on the corresponding*

linear model will yield (approximately) consistent estimators of $\boldsymbol{\xi}$, $\boldsymbol{\psi}$ *as well as* $\boldsymbol{\phi}$ *if at least one of the treatment and treatment-free models is correctly specified, and the weights satisfy*

$$\pi^{00}w(0,0,\boldsymbol{x})\kappa(0,0) = \pi^{01}w(0,1,\boldsymbol{x})\kappa(0,1) = \pi^{10}w(1,0,\boldsymbol{x})\kappa(1,0) = \pi^{11}w(1,1,\boldsymbol{x})\kappa(1,1),$$

where $\kappa(a^s, a^r) = g^{-1\prime}(\eta)$ *and* $g^{-1\prime}(\eta)$ *is the first derivative of the inverse link function (i.e.,* $g^{-1\prime}(\eta) = \frac{dg^{-1}(\eta)}{d\eta}$*),* $\eta(a^s, a^r) = \boldsymbol{\beta}^{*\top}\boldsymbol{x}^{\beta} + \boldsymbol{\xi}^{*\top}a^s\boldsymbol{x}^{\xi} + \boldsymbol{\psi}^{*\top}a^r\boldsymbol{x}^{\psi} + \boldsymbol{\phi}^{*\top}a^sa^r\boldsymbol{x}^{\phi}$*, and* $\boldsymbol{\beta}^*$, $\boldsymbol{\xi}^*$, $\boldsymbol{\psi}^*$ *and* $\boldsymbol{\phi}^*$ *are defined through*

$$\mathbb{E}\begin{bmatrix} \pi^{00}(\boldsymbol{x})\boldsymbol{x}w^{iad}(0,0,\boldsymbol{x})\left[g^{-1}(f(\boldsymbol{x})) - g^{-1}(\eta(0,0))\right] \\ \pi^{10}(\boldsymbol{x})\boldsymbol{x}w^{iad}(1,0,\boldsymbol{x})\left[g^{-1}(f(\boldsymbol{x}) + \boldsymbol{\xi}^{*\top}\boldsymbol{x}^{\xi}) - g^{-1}(\eta(1,0))\right] \\ \pi^{01}(\boldsymbol{x})\boldsymbol{x}w^{iad}(0,1,\boldsymbol{x})\left[g^{-1}(f(\boldsymbol{x}) + \boldsymbol{\psi}^{*\top}\boldsymbol{x}^{\psi}) - g^{-1}(\eta(0,1))\right] \\ \pi^{11}(\boldsymbol{x})\boldsymbol{x}w^{iad}(1,1,\boldsymbol{x})\left[g^{-1}(f(\boldsymbol{x}) + \boldsymbol{\xi}^{*\top}\boldsymbol{x}^{\xi} + \boldsymbol{\psi}^{*\top}\boldsymbol{x}^{\psi} + \boldsymbol{\phi}^{*\top}\boldsymbol{x}^{\phi}) - g^{-1}(\eta(1,1))\right] \end{bmatrix} = \mathbf{0},$$

where w^{iad} *represents interference-aware dWOLS weights, which satisfy* $\pi^{00}w(0,0,\boldsymbol{x}) = \pi^{01}w(0,1,\boldsymbol{x}) = \pi^{10}w(1,0,\boldsymbol{x}) = \pi^{11}w(1,1,\boldsymbol{x})$.

Based on the weights derivation of the binary outcomes in Chapter 4 (Theorem 4.1), we can conclude that the corresponding interference-aware binary-outcome overlap-type weights needed for (approximate) consistency are:

$$w(a^s, a^r, \boldsymbol{x}) \propto \frac{\pi^{00}\pi^{10}\pi^{01}\pi^{11}}{\pi^{a^sa^r}} \times \frac{\kappa(0,0,\boldsymbol{x})\kappa(1,0,\boldsymbol{x})\kappa(0,1,\boldsymbol{x})\kappa(1,1,\boldsymbol{x})}{\kappa(a^s, a^r, \boldsymbol{x})}, \quad \text{for } a^s = 0,1; a^r = 0,1.$$

If there is no association between treatments, then the overlap-type weights can be simplified as: $w(a^s, a^r; \boldsymbol{x}) = |a^s - \mathbb{E}[A|\boldsymbol{x}_s; \hat{\boldsymbol{\alpha}}]| * |a^r - \mathbb{E}[A^r|\boldsymbol{x}_r; \hat{\boldsymbol{\alpha}}]| * \kappa^r(1-a^s, \boldsymbol{x}, 1-a^r)$.

Proof: To prove Theorem H.2, the whole process is similar to the proof of Theorem H.1, but the spirit depends on the proof of Theorem 4.1 in Appendix G.1. Given the posited outcome regression model $Q(X, A^s, A^r; \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}) = g^{-1}(\boldsymbol{\beta}^\top X^{\beta} + A^s\boldsymbol{\xi}^\top X^{\xi} + A^r\boldsymbol{\psi}^\top X^{\psi} + A^sA^r\boldsymbol{\phi}^\top X^{\phi})$, the weighted GLM estimator for $(\boldsymbol{\beta}^\top, \boldsymbol{\xi}^\top, \boldsymbol{\psi}^\top, \boldsymbol{\phi}^\top)^\top$ is obtained by solving the system of estimating equations: $\sum_h^H U_h(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\phi}}; A_h^s, A_h^r, X_h) = \mathbf{0}$, that is,

$$\sum_h^H \begin{pmatrix} X_h^\beta \\ A_h^s X_h^\xi \\ A_h^r X_h^\psi \\ A_h^s A_h^r X_h^\phi \end{pmatrix} w(A_h^s, A_h^r, X_h) \left[Y_h - g^{-1}\left(\boldsymbol{\beta}^\top X_h^\beta - A^s \boldsymbol{\xi}^\top X_h^\xi - A^r \boldsymbol{\psi}^\top X_h^\psi - A^s A^r \boldsymbol{\phi}^\top X_h^\phi \right) \right]$$
$$= \sum_h^H \begin{bmatrix} U_{1h}(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}; A_h^s, A_h^r, X_h) \\ U_{2h}(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}; A_h^s, A_h^r, X_h) \\ U_{3h}(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}; A_h^s, A_h^r, X_h) \\ U_{4h}(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi}; A_h^s, A_h^r, X_h) \end{bmatrix} = \mathbf{0}$$

First, denoting $\epsilon_h = Y_h - g^{-1}\left(\boldsymbol{\beta}^\top X_h^\beta - A^s \boldsymbol{\xi}^\top X_h^\xi - A^r \boldsymbol{\psi}^\top X_h^\psi - A^s A^r \boldsymbol{\phi}^\top X_h^\phi \right)$ and again for simplicity assuming $X^\beta = X^\xi = X^\psi = X^\phi = X$, we consider

$$\begin{aligned} \sum_h^H (U_{1h} - U_{2h} - U_{3h} + U_{4h}) &= \sum_h^H (1 - A_h^s)(1 - A_h^r) X_h w(A_h^s, A_h^r, X_h) \epsilon_h \\ &= \sum_h^H (1 - A_h^s)(1 - A_h^r) X_h w(A_h^s, A_h^r, X_h) \left[Y_h - g^{-1}\left(\boldsymbol{\beta}^\top X_h \right) \right] = \mathbf{0}, \end{aligned}$$

which is only related to $\boldsymbol{\beta}$. Thus, it can be solved for $\hat{\boldsymbol{\beta}}$. According to large sample theory, $\hat{\boldsymbol{\beta}}$ tends to converge to $\boldsymbol{\beta}^*$ as $H \to \infty$ (White [1982]), where $\boldsymbol{\beta}^*$ is defined as $H\mathbb{E}_X \left[\pi^{00}(X) X w(0,0,X) \left(g^{-1}(f(X)) - g^{-1}(\boldsymbol{\beta}^{*\top} X) \right) \right] = \mathbf{0}$.

Then, we consider

$$\begin{aligned} \sum_h^H (U_{2h} - U_{4h}) &= \sum_h^H (1 - A_h^r) A_h^s X_h w(A_h^s, A_h^r, X_h) \epsilon_h \\ &= \sum_h^H (1 - A_h^r) A_h^s X_h w(A_h^s, A_h^r, X_h) \left[Y_h - g^{-1}\left(\boldsymbol{\xi}^\top A_h^s X_h + \boldsymbol{\beta}^\top X_h \right) \right] = \mathbf{0}, \end{aligned}$$

and

$$\sum_h^H (U_{3h} - U_{4h}) = \sum_h^H (1 - A_h^s) A_h^r X_h w(A_h^s, A_h^r, X_h) \epsilon_h$$
$$= \sum_h^H (1 - A_h^s) A_h^r X_h w(A_h^s, A_h^r, X_h) \left[Y_h - g^{-1} \left(\boldsymbol{\psi}^\top A_h^r X_h + \boldsymbol{\beta}^\top X_h \right) \right] = \mathbf{0},$$

which can be solved for $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\psi}}$, respectively, in terms of $\hat{\boldsymbol{\beta}}$.

To show that $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\psi}}$ are (approximately) consistent, we would need to show that the expectation of

$$\sum_h^H (1 - A_h^r) A_h^s X_h w(A_h^s, A_h^r, X_h) \left[g^{-1}(\boldsymbol{\xi}^\top A_h^s X_h + \boldsymbol{\beta}^{*\top} X_h) - g^{-1}(\boldsymbol{\xi}^\top A_h^s X_h + f(X_h)) \right],$$

and

$$\sum_h^H (1 - A_h^s) A_h^r X_h w(A_h^s, A_h^r, X_h) \left[g^{-1}(\boldsymbol{\psi}^\top A_h^r X_h + \boldsymbol{\beta}^{*\top} X_h) - g^{-1}(\boldsymbol{\psi}^\top A_h^r X_h + f(X_h)) \right],$$

equals or is close to $\mathbf{0}$ for general $\boldsymbol{\xi}$ and $\boldsymbol{\psi}$.

Again, let $g^{-1\prime}$ denote the derivative of the inverse link function g. Then the expectation of $\sum_h^H (U_{1h} - U_{2h} - U_{3h} + U_{4h})$ conditional on $(X_1, ...X_H)$, that is, $\sum_h^H \pi^{00} X_h w(0, 0, X_h) \left[g^{-1}(f(X_h)) - g^{-1}\left(\boldsymbol{\beta}^\top X_h\right) \right]$, can be written using a Taylor series expansion (function $g^{-1}(f(X_h))$ at the point $\boldsymbol{\beta}^\top X_h$) as

$$\sum_h^H \pi^{00} X_h w(0, 0, X_h) \left[g^{-1\prime}(\boldsymbol{\beta}^\top X_h)(f(X_h) - \boldsymbol{\beta}^\top X_h) + O[(f(X_h) - \boldsymbol{\beta}^\top X_h)^2] \right], \tag{H.9}$$

where the big O describes the error term in an approximation to the g^{-1} function. The notation $O[(f(X_h) - \boldsymbol{\beta}^\top X_h)^2]$ means the absolute-value of the error of $g^{-1}(f(X_h)) - g^{-1}(\boldsymbol{\beta}^\top X_h) - g^{-1\prime}(\boldsymbol{\beta}^\top X_h)(f(X_h) - \boldsymbol{\beta}^\top X_h)$ is at most some constant times $(f(X_h) - \boldsymbol{\beta}^\top X_h)^2$ when $f(X_h) - \boldsymbol{\beta}^\top X_h$ is close enough to 0. Further, the expectations of $\sum_h^H (U_{2h} - U_{4h})$ and $\sum_h^H (U_{3h} - U_{4h})$ conditional on $(X_1, ..., X_H)$ are

$$\sum_h^n \pi^{10} X_h w(1, 0, X_h) \left[g^{-1}(\boldsymbol{\xi}^\top X_h + f(X_h)) - g^{-1}(\boldsymbol{\xi}^\top X_h + \boldsymbol{\beta}^\top X_h) \right],$$

and

$$\sum_{h}^{n} \pi^{01} X_h w(0,1,X_h) \left[g^{-1}(\boldsymbol{\psi}^\top X_h + f(X_h)) - g^{-1}(\boldsymbol{\psi}^\top X_h + \boldsymbol{\beta}^\top X_h) \right],$$

which can be written as

$$\sum_{h}^{n} \pi^{10} X_h w(1,0,X_h) \left[g^{-1\prime}(\boldsymbol{\beta}^\top X_h + \boldsymbol{\xi}^\top X_h)(f(X_h) - \boldsymbol{\beta}^\top X_h) + O[(f(X_h) - \boldsymbol{\beta}^\top X_h)^2] \right]. \quad \text{(H.10)}$$

and

$$\sum_{h}^{n} \pi^{01} X_h w(0,1,X_h) \left[g^{-1\prime}(\boldsymbol{\beta}^\top X_h + \boldsymbol{\psi}^\top X_h)(f(X_h) - \boldsymbol{\beta}^\top X_h) + O[(f(X_h) - \boldsymbol{\beta}^\top X_h)^2] \right]. \quad \text{(H.11)}$$

A similar process is followed for $\sum_h^H U_{4h}$, and its expectation conditional on $(X_1, \ldots X_H)$ is

$$\sum_{h}^{n} \pi^{11} X_h w(1,1,X_h) \left[g^{-1}(\boldsymbol{\xi}^\top X_h + \boldsymbol{\psi}^\top X_h + \boldsymbol{\phi}^\top X_h + f(X_h)) - g^{-1}(\boldsymbol{\xi}^\top X_h + \boldsymbol{\psi}^\top X_h + \boldsymbol{\phi}^\top X_h + \boldsymbol{\beta}^\top X_h) \right],$$

which can be written as

$$\sum_{h}^{n} \pi^{11} X_h w(1,1,X_h) \left[g^{-1\prime}(\boldsymbol{\beta}^\top X_h + \boldsymbol{\xi}^\top X_h + \boldsymbol{\psi}^\top X_h + \boldsymbol{\phi}^\top X_h)(f(X_h) - \boldsymbol{\beta}^\top X_h) + O[(f(X_h) - \boldsymbol{\beta}^\top X_h)^2] \right]. \quad \text{(H.12)}$$

Therefore, we can define $\kappa^*(A^s, A^r, X) = g^{-1\prime}(\boldsymbol{\beta}^{*\top} X + \boldsymbol{\xi}^{*\top} A^s X + \boldsymbol{\psi}^{*\top} A^r X + \boldsymbol{\phi}^{*\top} A^s A^r X)$, where $\boldsymbol{\xi}^*$, $\boldsymbol{\psi}^*$, and $\boldsymbol{\phi}^*$ are assumed to be limiting values for $\hat{\boldsymbol{\xi}}$, $\hat{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\phi}}$, respectively. Then if weights are defined to satisfy a new balancing criterion $\pi^{00} w(0,0,X)\kappa^*(0,0,X) = \pi^{10} w(1,0,X)\kappa^*(1,0,X) = \pi^{01} w(0,1,X)\kappa^*(0,1,X) = \pi^{00} w(1,1,X)\kappa^*(1,1,X)$, and if the distribution of X is such that the inverse link function is close to linear for the range of $f(X) - \boldsymbol{\beta}^{*\top} X$ (so that the Taylor expansion error term is small), the fact that the expectation of (H.9) is $\mathbf{0}$ for $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ means that the expectations of (H.10), (H.11), and (H.12) are also close to $\mathbf{0}$ for $\boldsymbol{\beta} = \boldsymbol{\beta}^*$. This argument establishes the approximate consistency of the corresponding new estimator of $\boldsymbol{\xi}$, $\boldsymbol{\psi}$, and $\boldsymbol{\phi}$.

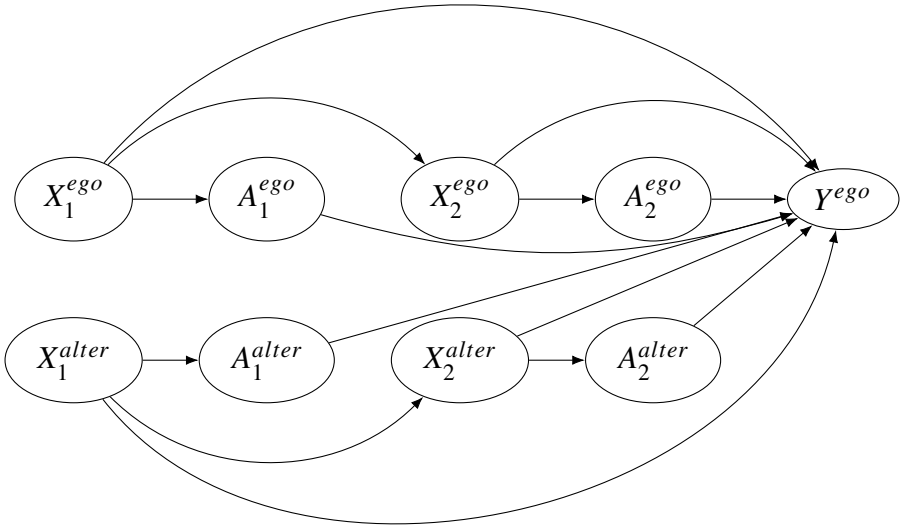


Figure H.1: Directed acyclic graphs of interference analysis: dWGLM with interference two-stage decision problems.

Multiple-stage decisions

For the multiple-stage decision problem, like many DTR estimation methods, we use backward induction, solving the multiple-stage problems recursively beginning with the final stage. The causal diagram for the two-stage treatment decisions with binary outcomes in the presence of interference is presented in Figure (H.1).

As we move backwards through the stages we construct pseudo-outcomes, which are the potential outcomes if the patients were treated - possibly contrary to fact - optimally at subsequent stages. Therefore, for the multistage decision analysis, in the presence of household interference, the dWGLM procedure could be implemented by the following steps at each stage of the analysis, starting from the last stage K and working backwards towards the first stage (without loss of generality, we set $s = ego$, $r = alter$, and $a^s = a, a^r = a^{alter}$):

- Step 1: Construct the stage j *pseudo-outcome*: set $\widetilde{\mathcal{Y}_j} = y$ if $j = K$. Otherwise, use prior estimates $\hat{\boldsymbol{\beta}}_K$ and $\underline{\hat{\boldsymbol{\psi}}}_{j+1} = (\hat{\boldsymbol{\psi}}_{j+1}, ..., \hat{\boldsymbol{\psi}}_K)$ to randomly generate $\widetilde{\mathcal{Y}_j}$, which takes the value 1

with probability:

$$\mathbb{P}(\widetilde{\mathcal{Y}_j} = 1) = g^{-1}\left(g[\mathbb{P}(Y = 1 \mid \boldsymbol{h}_K, a_K, a_K^{alter}; \hat{\boldsymbol{\beta}}_K, \hat{\boldsymbol{\psi}}_K)] + \sum_{k=j+1}^{K}\left[\gamma_k(\boldsymbol{h}_k^{\psi}, \hat{a}_k^{opt}, a_k^{alter}; \hat{\boldsymbol{\psi}}_k) - \gamma_k(\boldsymbol{h}_k^{\psi}, a_k, a_k^{alter}; \hat{\boldsymbol{\psi}}_k)\right]\right),$$

R times, namely, $\widetilde{\mathcal{Y}_j^1}, \widetilde{\mathcal{Y}_j^2}, ..., \widetilde{\mathcal{Y}_j^R}$.

- Step 2: Specify the stage j treatment model $\mathbb{E}\left[A_j|\boldsymbol{h}_j^{\alpha}; \boldsymbol{\alpha}_j\right]$. The treatment model parameters $\boldsymbol{\alpha}_j$ (estimated, for example, via logistic regression) are used to compute a weight w_j, such as $w_j = \left|a_j - \mathbb{E}\left[A_j|\boldsymbol{h}_j^{\alpha}; \hat{\boldsymbol{\alpha}}_j\right]\right| * \left|a_j^{alter} - \mathbb{E}\left[A_j^{alter}|\boldsymbol{h}_j^{\alpha,alter}; \hat{\boldsymbol{\alpha}}_j^{alter}\right]\right|$.
- Step 3: Specify the stage j treatment-free and blip models, and perform a weighted generalized linear model of $\widetilde{\mathcal{Y}_j^r}$ on the terms in the treatment-free and blip models, using weights from Step 2 to get estimates $\hat{\boldsymbol{\beta}}_j^{old,r}$, $\hat{\boldsymbol{\psi}}_j^{old,r}$, $\hat{\beta}_{j,al}^{old,r}$ and $\hat{\psi}_{j,al}^{old,r}$ for $r = 1, ..., R$; that is, for each time of the total R, use the model

$$g(\mathbb{E}[\widetilde{\mathcal{Y}_j^r} \mid a_j, a_j^{alter}, \boldsymbol{h}_j; \boldsymbol{\beta}_j, \boldsymbol{\psi}_j]) = \boldsymbol{\beta}_j^{\top}\boldsymbol{h}_j^{\beta} + \beta_{j,al}a_j^{alter} + \boldsymbol{\psi}_j^{\top}a_j\boldsymbol{h}_j^{\psi} + a_j\psi_{j,al}a_j^{alter}.$$

- Step 4: Use $\hat{\boldsymbol{\beta}}_j^{old,r}$, $\hat{\boldsymbol{\psi}}_j^{old,r}$, $\hat{\beta}_{j,al}^{old,r}$ and $\hat{\psi}_{j,al}^{old,r}$ from Step 3 to compute $\kappa^r(a_j, \boldsymbol{h}_j, a_j^{alter}) = g^{-1'}(\hat{\boldsymbol{\beta}}_j^{old,r^{\top}}\boldsymbol{h}_j^{\beta} + \hat{\beta}_{j,al}^{old,r}a_j^{alter} + \hat{\boldsymbol{\psi}}_j^{old,r^{\top}}a_j\boldsymbol{h}_j^{\psi} + a_j\hat{\psi}_{j,al}^{old,r}a_j^{alter})$, where $g^{-1'}$ is identified based on the link function in Step 3. Then, construct the new weights $w_j^{new,r}(a_j, a_j^{alter}; \boldsymbol{h}_j) = |a_j - \mathbb{E}[A_j|\boldsymbol{h}_j^{\alpha}]| * \left|a_j^{alter} - \mathbb{E}\left[A_j^{alter}|\boldsymbol{h}_j^{\alpha,alter}; \hat{\boldsymbol{\alpha}}_j^{alter}\right]\right| * \kappa^r(1 - a_j, \boldsymbol{h}_j, 1 - a_j^{alter})$.
- Step 5: Perform a weighted GLM with the new weights (i.e., $w_j^{new,r}(a_j, a_j^{alter}; \boldsymbol{h}_j)$) to get revised estimates $\hat{\boldsymbol{\beta}}_j^r$, $\hat{\boldsymbol{\psi}}_j^r$, $\hat{\beta}_{j,al}^r$ and $\hat{\psi}_{j,al}^r$. Estimate $\boldsymbol{\psi}_j$ by $\hat{\boldsymbol{\psi}}_j = R^{-1}\sum_r \hat{\boldsymbol{\psi}}_j^r$ and $\psi_{j,al}$ by $\hat{\psi}_{j,al} = R^{-1}\sum_r \hat{\psi}_{j,al}^r$, then use parameter estimators $\hat{\boldsymbol{\psi}}_j$ and $\hat{\psi}_{j,al}$ to construct the j^{th} stage optimal treatment rule, which is $prescribe\ a_j = 1\ if\ \hat{\boldsymbol{\psi}}_j^{\top}\boldsymbol{H}_j^{\psi} + \hat{\psi}_{j,al}a_j^{alter} > 0$; $then\ prescribe\ a_j = 0\ otherwise$.
- Step 6: Return to Step 1 and analyze stage $j - 1$ if there are more stages to analyze.

In this section, considering continuous and binary outcomes, and the corresponding household interference outcome models (H.1) and (H.8), we provided two theorems for the doubly robust

estimation of the blip parameters in these two outcome models. These models can be utilized to make decisions not only for the ego (individual s in the pair (s, r) case) given the fixed alter (individual r), but also for both of them, i.e., for a pair in the same household.

H.2 Proof of Theorem 5.1

In this Appendix section, we will prove Theorem 5.1. We assume that the ordinal outcome U_h that takes the value c follows a multinomial distribution that $U_h^* \sim multinom(\boldsymbol{\pi}_h, 1)$, where U^* is "one-hot" encoded as a $C-$vector with a 1 at the cth entry and 0 otherwise. Then, the likelihood function is:

$$\mathbb{P}\left(\boldsymbol{U}_h^* = \boldsymbol{u}_h^*\right) = \prod_c \pi_{hc}^{u_{hc}^*}, \ for\ c = 1, 2, ..., C,$$

where $\pi_{hc} = \mathbb{P}(U_h = c)$. In our $C = 3$ case, building on the assumed POM (5.8) and denoting $\eta_{1h} = \zeta_1 - \boldsymbol{\beta}^\top \boldsymbol{x}_h^\beta - a^s \boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi - a^r \boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi - a^s a^r \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi$ and $\eta_{2h} = \zeta_2 - \boldsymbol{\beta}^\top \boldsymbol{x}_h^\beta - a^s \boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi - a^r \boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi - a^s a^r \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi$, then we have

$$\begin{aligned}
\pi_{h1} &= \mathbb{P}(U_h = 1 \mid a^s, a^r, \boldsymbol{x}_h) = g^{-1}[\eta_{1h}] \\
\pi_{h2} &= \mathbb{P}(U_h = 2 \mid a^s, a^r, \boldsymbol{x}_h) = g^{-1}[\eta_{2h}] - g^{-1}[\eta_{1h}] \\
\pi_{h3} &= \mathbb{P}(U_h = 3 \mid a^s, a^r, \boldsymbol{x}_h) = 1 - g^{-1}[\eta_{2h}],
\end{aligned}$$

where g^{-1} is the inverse of the link function. However, we assume that the true POM model has $\eta_{1h} = \zeta_1 - f(\boldsymbol{x}_h^\beta) - a^s \boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi - a^r \boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi - a^s a^r \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi$ and $\eta_{2h} = \zeta_2 - f(\boldsymbol{x}_h^\beta) - a^s \boldsymbol{\xi}^\top \boldsymbol{x}_h^\xi - a^r \boldsymbol{\psi}^\top \boldsymbol{x}_h^\psi - a^s a^r \boldsymbol{\phi}^\top \boldsymbol{x}_h^\phi$, for an arbitrary treatment-free function $f(\boldsymbol{x}_h^\beta)$. Denoting the nuisance parameters as $\boldsymbol{\theta_1} = (\boldsymbol{\zeta}, \boldsymbol{\beta})$ and the parameters of interest as $\boldsymbol{\theta_2} = (\boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi})$, the log-likelihood function is

$$\begin{aligned}
\ell\left(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}; \boldsymbol{u}^*\right) &= \sum_h \sum_c u_{hc}^* \log \pi_{hc} \\
&= \sum_h \sum_c I(u_h = c) \log \pi_{hc}.
\end{aligned}$$

Denoting $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$ as the whole parameter in POM (5.8), then the corresponding score function system components are

$$\sum_h \sum_c I(u_h = c)(\pi_{hc})^{-1} \partial \pi_{hc} / \partial \boldsymbol{\theta}.$$

We first give the unweighted score function system. In particular, for our case we have three

components of the score estimation equation system, which are

$$\sum_h I(u_h = 1)(g^{-1}[\eta_{1h}])^{-1} g^{-1\prime}[\eta_{1h}] \begin{pmatrix} 1 \\ 0 \\ -X_h^\beta \\ -A_h^s X_h^\xi \\ -A_h^r X_h^\psi \\ -A_h^s A_h^r X_h^\phi \end{pmatrix}$$

$$\sum_h I(u_h = 2)(g^{-1}[\eta_{2h}] - g^{-1}[\eta_{1h}])^{-1} \left[g^{-1\prime}[\eta_{2h}] \begin{pmatrix} 0 \\ 1 \\ -X_h^\beta \\ -A_h^s X_h^\xi \\ -A_h^r X_h^\psi \\ -A_h^s A_h^r X_h^\phi \end{pmatrix} - g^{-1\prime}[\eta_{1h}] \begin{pmatrix} 1 \\ 0 \\ -X_h^\beta \\ -A_h^s X_h^\xi \\ -A_h^r X_h^\psi \\ -A_h^s A_h^r X_h^\phi \end{pmatrix} \right],$$

$$- \sum_h I(u_h = 3)(1 - g^{-1}[\eta_{2h}])^{-1} g^{-1\prime}[\eta_{2h}] \begin{pmatrix} 0 \\ 1 \\ -X_h^\beta \\ -A_h^s X_h^\xi \\ -A_h^r X_h^\psi \\ -A_h^s A_h^r X_h^\phi \end{pmatrix},$$

respectively. That is, the score estimation equation system is

$$\sum_h S_{1h} \begin{pmatrix} 1 \\ 0 \\ -X_h^\beta \\ -A_h^s X_h^\xi \\ -A_h^r X_h^\psi \\ -A_h^s A_h^r X_h^\phi \end{pmatrix} + \sum_h S_{2h} \begin{pmatrix} 0 \\ 1 \\ -X_h^\beta \\ -A_h^s X_h^\xi \\ -A_h^r X_h^\psi \\ -A_h^s A_h^r X_h^\phi \end{pmatrix} = \mathbf{0}, \tag{H.13}$$

where S_{1h} and S_{2h} are defined as

$$\begin{cases} S_{1h} := \left[I(u_h = 1)(g^{-1}[\eta_{1h}])^{-1} - I(u_h = 2)(g^{-1}[\eta_{2h}] - g^{-1}[\eta_{1h}])^{-1} \right] g^{-1'}[\eta_{1h}] \\ S_{2h} := \left[I(u_h = 2)(g^{-1}[\eta_{2h}] - g^{-1}[\eta_{1h}])^{-1} - I(u_h = 3)(1 - g^{-1}[\eta_{2h}])^{-1} \right] g^{-1'}[\eta_{2h}]. \end{cases}$$

Vertically, there are two major components of the POM score equation system in (H.13). One of the components is the top two equations associated with ζ_1 and ζ_2, and the other is the bottom equations related to $(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{\phi})$. The top two rows yield $\sum_h S_{1h} = 0$ and $\sum_h S_{2h} = 0$.

Further, for convenience, we denote by $b_h = (b_{1h}, b_{2h})^\top$ the vector $(-S_{1h}/(S_{1h} + S_{2h}), -S_{2h}/(S_{1h} + S_{2h}))^\top$. In addition, considering the balancing weights, we can write the weighted analogues of (H.13) as:

$$\sum_h (S_{1h} + S_{2h})\, w(A_h^s, A_h^r, X_h) \begin{pmatrix} b_h \\ X_h^\beta \\ A_h^s X_h^\xi \\ A_h^r X_h^\psi \\ A_h^s A_h^r X_h^\phi \end{pmatrix} = \sum_h \begin{pmatrix} V_{0h} \\ V_{1h} \\ V_{2h} \\ V_{3h} \\ V_{4h} \end{pmatrix} = \mathbf{0};$$

where

$$\sum_h V_{0h} = \sum_h \begin{pmatrix} S_{1h} \\ S_{2h} \end{pmatrix} w(A_h^s, A_h^r, X_h) = \mathbf{0}$$

are additional equations that are used for solving ζ_1, ζ_2. Now, for simplicity, we assume that ζ_1 and ζ_2 are known. (In the implementation of the overall estimation procedure, these "known" values would be replaced by preliminary estimates.)

If we focus on the logit link, that is, $g^{-1}(t) = [1 + \exp(-t)]^{-1}$, then we have some properties, such as $g^{-1'}(t) = g^{-1}(t)[1 - g^{-1}(t)]$ or $g^{-1'}(t) = g^{-1}(t)g^{-1}(-t)$, where $[1 - g^{-1}(t)] = g^{-1}(-t)$. Thus, we have $I(u_h = 1)(g^{-1}[\eta_{1h}])^{-1} g^{-1'}[\eta_{1h}] = I(u_h = 1)(1 - g^{-1}[\eta_{1h}])$, and $I(u_h = 3)(1 -$

$g^{-1}[\eta_{2h}])^{-1}g^{-1\prime}[\eta_{2h}] = I(u_h = 3)g^{-1}[\eta_{2h}]$. In addition, we also have

$$
\begin{aligned}
& I(u_h = 2)\left(g^{-1}[\eta_{2h}] - g^{-1}[\eta_{1h}]\right)^{-1}\left(g^{-1\prime}[\eta_{2h}] - g^{-1\prime}[\eta_{1h}]\right) \\
=& I(u_h = 2)\left(g^{-1}[\eta_{2h}] - g^{-1}[\eta_{1h}]\right)^{-1}\left(g^{-1}[\eta_{2h}](1 - g^{-1}[\eta_{2h}]) - g^{-1}[\eta_{1h}](1 - g^{-1}[\eta_{1h}])\right) \\
=& I(u_h = 2)\left(g^{-1}[\eta_{2h}] - g^{-1}[\eta_{1h}]\right)^{-1}\left(g^{-1}[\eta_{2h}] - g^{-1^2}[\eta_{2h}] - g^{-1}[\eta_{1h}] + g^{-1^2}[\eta_{1h}]\right) \\
=& I(u_h = 2)\left(g^{-1}[\eta_{2h}] - g^{-1}[\eta_{1h}]\right)^{-1}\left(g^{-1}[\eta_{2h}] - g^{-1}[\eta_{1h}] + g^{-1^2}[\eta_{1h}] - g^{-1^2}[\eta_{2h}]\right) \\
=& I(u_h = 2)\left(g^{-1}[\eta_{2h}] - g^{-1}[\eta_{1h}]\right)^{-1}\left(g^{-1}[\eta_{2h}] - g^{-1}[\eta_{1h}]\right)\left(1 - g^{-1}[\eta_{1h}] - g^{-1}[\eta_{2h}]\right) \\
=& I(u_h = 2)\left(1 - g^{-1}[\eta_{1h}] - g^{-1}[\eta_{2h}]\right).
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
S_{1h} + S_{2h} &= I(u_h = 1)(1 - g^{-1}[\eta_{1h}]) + I(u_h = 2)\left(1 - g^{-1}[\eta_{1h}] - g^{-1}[\eta_{2h}]\right) - I(u_h = 3)g^{-1}[\eta_{2h}] \\
&= [I(u_h = 1) + I(u_h = 2)](1 - g^{-1}[\eta_{1h}]) - [I(u_h = 2) + I(u_h = 3)]g^{-1}[\eta_{2h}]
\end{aligned}
$$

Therefore, going through the same process in interference-aware dWOLS section, without loss of generality, we assume that $X^\beta = X^\xi = X^\psi = X^\phi = X$; thus, we have

$$
\sum_h \begin{bmatrix} V_{1h} - V_{2h} - V_{3h} + V_{4h} \\ V_{2h} - V_{4h} \\ V_{3h} - V_{4h} \\ V_{4h} \end{bmatrix} = \sum_h \left[\begin{pmatrix} (1 - A_h^s)(1 - A_h^r) \\ (1 - A_h^r)A_h^s \\ (1 - A_h^s)A_h^r \\ A_h^r A_h^s \end{pmatrix} X w(A_h^s, A_h^r, X_h)\,(S_{1h} + S_{2h}) \right] = \mathbf{0}.
$$

$\sum_h (V_{1h} - V_{2h} - V_{3h} + V_{4h})$ uses the untreated part of the sample, and can be used to estimate $\boldsymbol{\beta}$. That is, $\hat{\boldsymbol{\beta}}$ can be solved by $\sum_h (V_{1h} - V_{2h} - V_{3h} + V_{4h}) = \sum_h^H (1 - A_h^s)(1 - A_h^r) X_h w(A_h^s, A_h^r, X_h)(S_{1h}^{00} + S_{2h}^{00}) = \mathbf{0}$, where $S_{1h}^{00} + S_{2h}^{00}$ is defined as

$$
S_{1h}^{00} + S_{2h}^{00} := [I(u_h = 1) + I(u_h = 2)](1 - g^{-1}[\zeta_1 - \boldsymbol{\beta}^\top X_h]) - [I(u_h = 2) + I(u_h = 3)]g^{-1}[\zeta_2 - \boldsymbol{\beta}^\top X_h].
$$

If the treatment-free model is not correct, the estimates will actually estimate $\boldsymbol{\beta}^*$, which is defined as the solution of $\mathbb{E}_X\left[\pi^{00}(X) X w(0, 0, X)(\mathcal{S}_1^{00} + \mathcal{S}_2^{00})\right] = \mathbf{0}$, that is,

$\mathbb{E}_X\left[\pi^{00}(X)Xw(0,0,X)(\mathcal{S}_1^{00}+\mathcal{S}_2^{00})\right]\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}=\mathbf{0}$, where $\mathcal{S}_1^{00}+\mathcal{S}_2^{00}$ is defined as

$$\begin{aligned}\mathcal{S}_1^{00}+\mathcal{S}_2^{00}:=&g^{-1}[\zeta_2-f(X)](1-g^{-1}[\zeta_1-\boldsymbol{\beta}^\top X])-(1-g^{-1}[\zeta_1-f(X)])g^{-1}[\zeta_2-\boldsymbol{\beta}^\top X]\\ =&g^{-1}[\zeta_2-f(X)]-g^{-1}[\zeta_2-\boldsymbol{\beta}^\top X]+g^{-1}[\zeta_1-f(X)]g^{-1}[\zeta_2-\boldsymbol{\beta}^\top X]\\ &-g^{-1}[\zeta_2-f(X)]g^{-1}[\zeta_1-\boldsymbol{\beta}^\top X].\end{aligned}$$

Note that, conditional on treatments and covariates, $\mathbb{E}[I(u=1)]=\pi_1=g^{-1}[\eta_1]=g^{-1}[\zeta_1-f(X)]$, which depends on the true outcome models in terms of the treatment-free function $f(X)$. Similar results are derived from $\mathbb{E}[I(u=2)]$ and $\mathbb{E}[I(u=3)]$. That is, $\mathbb{E}[I(u=2)]=\pi_2=g^{-1}[\eta_2]-g^{-1}[\eta_1]=g^{-1}[\zeta_2-f(X)]-g^{-1}[\zeta_1-f(X)]$, and $\mathbb{E}[I(u=3)]=\pi_3=1-g^{-1}[\eta_2]=1-g^{-1}[\zeta_2-f(X)]$. Thus, conditional on treatments and covariates, we have $\mathbb{E}[I(u=1)]+\mathbb{E}[I(u=2)]=g^{-1}[\zeta_2-f(X)]$, and $\mathbb{E}[I(u=2)]+\mathbb{E}[I(u=3)]=1-g^{-1}[\zeta_1-f(X)]$.

To sum up, $\sum_h(V_{1h}-V_{2h}-V_{3h}+V_{4h})=\mathbf{0}$ can be solved for $\hat{\boldsymbol{\beta}}$. According to large sample theory, $\hat{\boldsymbol{\beta}}$ tends to converge to $\boldsymbol{\beta}^*$, as $H\to\infty$ (White [1982]), where $\boldsymbol{\beta}^*$ is the solution of $H\mathbb{E}_X\left[\pi^{00}(X)Xw(0,0,X)(\mathcal{S}_1^{00}+\mathcal{S}_2^{00})\right]=\mathbf{0}$, with $\mathcal{S}_1^{00}+\mathcal{S}_2^{00}:=g^{-1}[\zeta_2-f(X)](1-g^{-1}[\zeta_1-\boldsymbol{\beta}^\top X])-(1-g^{-1}[\zeta_1-f(X)])g^{-1}[\zeta_2-\boldsymbol{\beta}^\top X]$.

Then the expectation of $\sum_h^H(V_{1h}-V_{2h}-V_{3h}+V_{4h})$ conditional on $(X_1,\ldots X_H)$, that is,

$$\sum_h^H\pi^{00}X_hw(0,0,X_h)\left[(\mathcal{S}_{1h}^{00}+\mathcal{S}_{2h}^{00})(\boldsymbol{\zeta},\boldsymbol{\beta})\right], \tag{H.14}$$

where, in terms of parameters of $\boldsymbol{\zeta}$, and $\boldsymbol{\beta}$, $(\mathcal{S}_{1h}^{00}+\mathcal{S}_{2h}^{00})(\boldsymbol{\zeta},\boldsymbol{\beta})$ is defined as:

$$\begin{aligned}(\mathcal{S}_{1h}^{00}+\mathcal{S}_{2h}^{00})(\boldsymbol{\zeta},\boldsymbol{\beta}):=&g^{-1}\left[\zeta_2-f(X_h)\right]-g^{-1}\left(\zeta_2-\boldsymbol{\beta}^\top X_h\right)+g^{-1}[\zeta_1-f(X_h)]g^{-1}[\zeta_2-\boldsymbol{\beta}^\top X_h]\\ &-g^{-1}[\zeta_2-f(X_h)]g^{-1}[\zeta_1-\boldsymbol{\beta}^\top X_h],\end{aligned}$$

can be written using a Taylor series expansion, function $g^{-1}[\zeta-f(X_h)]$ at the points $\zeta_1-\boldsymbol{\beta}^\top X_h$ and $\zeta_2-\boldsymbol{\beta}^\top X_h$, respectively, as

$$\sum_h^H\pi^{00}X_hw(0,0,X_h)\left[\mathcal{K}\Delta+O[\Delta^2]\right], \tag{H.15}$$

where

$$\mathcal{K}:=g^{-1}(\zeta_2-\boldsymbol{\beta}^\top X_h)\left[1-g^{-1}(\zeta_1-\boldsymbol{\beta}^\top X_h)\right]\left[1-g^{-1}(\zeta_2-\boldsymbol{\beta}^\top X_h)+g^{-1}(\zeta_1-\boldsymbol{\beta}^\top X_h)\right],$$

and $\Delta := \boldsymbol{\beta}^\top X_h - f(X_h)$, and the big O describes the error term in an approximation to the g^{-1} function. To prove this, we initially focus on the first two terms of $\mathcal{S}^{00}_{1h} + \mathcal{S}^{00}_{2h}$ in expression (H.14), that is, $g^{-1}\left[\zeta_2 - f(X_h)\right] - g^{-1}\left(\zeta_2 - \boldsymbol{\beta}^\top X_h\right) = g^{-1\prime}(\zeta_2 - \boldsymbol{\beta}^\top X_h)(\Delta) + O[\Delta^2]$. Then the third and fourth terms are

$$\begin{cases} g^{-1}[\zeta_1 - f(X_h)]g^{-1}[\zeta_2 - \boldsymbol{\beta}^\top X_h] = \left(g^{-1}\left(\zeta_1 - \boldsymbol{\beta}^\top X_h\right) + g^{-1\prime}(\zeta_1 - \boldsymbol{\beta}^\top X_h)(\Delta) + O[\Delta^2]\right) g^{-1}[\zeta_2 - \boldsymbol{\beta}^\top X_h], \\ g^{-1}[\zeta_2 - f(X_h)]g^{-1}[\zeta_1 - \boldsymbol{\beta}^\top X_h] = \left(g^{-1}\left(\zeta_2 - \boldsymbol{\beta}^\top X_h\right) + g^{-1\prime}(\zeta_2 - \boldsymbol{\beta}^\top X_h)(\Delta) + O[\Delta^2]\right) g^{-1}[\zeta_1 - \boldsymbol{\beta}^\top X_h]. \end{cases}$$

Thus, the Taylor series expansion of $\mathcal{S}^{00}_{1h} + \mathcal{S}^{00}_{2h}$ in expression (H.14) is

$$\left(g^{-1\prime}(\zeta_2 - \boldsymbol{\beta}^\top X_h) + g^{-1\prime}(\zeta_1 - \boldsymbol{\beta}^\top X_h)g^{-1}(\zeta_2 - \boldsymbol{\beta}^\top X_h) - g^{-1\prime}(\zeta_2 - \boldsymbol{\beta}^\top X_h)g^{-1}(\zeta_1 - \boldsymbol{\beta}^\top X_h)\right)(\Delta) + O[\Delta^2]. \tag{H.16}$$

Based on the property of the g^{-1} that $g^{-1\prime}(t) = g^{-1}(t)[1 - g^{-1}(t)]$, then expression (H.16) equals $g^{-1}(\zeta_2 - \boldsymbol{\beta}^\top X_h)\left[1 - g^{-1}(\zeta_1 - \boldsymbol{\beta}^\top X_h)\right]\left[1 - g^{-1}(\zeta_2 - \boldsymbol{\beta}^\top X_h) + g^{-1}(\zeta_1 - \boldsymbol{\beta}^\top X_h)\right](\Delta) + O[\Delta^2]$. That is, the Taylor series expansion of $(\mathcal{S}^{00}_{1h} + \mathcal{S}^{00}_{2h})(\boldsymbol{\zeta}, \boldsymbol{\beta})$ is

$$(\mathcal{S}^{00}_{1h} + \mathcal{S}^{00}_{2h})(\boldsymbol{\zeta}, \boldsymbol{\beta}) = \mathcal{K}\Delta + O[\Delta^2].$$

Therefore, we finally have expression (H.15).

Then, we consider $\sum_h^H (V_{2h} - V_{4h}) = \mathbf{0}$ and $\sum_h^H (V_{3h} - V_{4h}) = \mathbf{0}$ which can be solved for $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\psi}}$, respectively, in terms of $\hat{\zeta}_1, \hat{\zeta}_2$, and $\hat{\boldsymbol{\beta}}$. First, we study $\sum_h^H (V_{2h} - V_{4h}) = \mathbf{0}$, that is, $\sum_h^H (1 - A^r_h) A^s_h X w(1, 0, X_h)\ (\mathcal{S}_{1h} + \mathcal{S}_{2h}) = \mathbf{0}$, which provides $\hat{\boldsymbol{\xi}}$ in terms of $\hat{\zeta}_1, \hat{\zeta}_2$, and $\hat{\boldsymbol{\beta}}$. To show that $\hat{\boldsymbol{\xi}}$ is (approximately) consistent, we would need to show that the expectation of $\sum_h^H (1 - A^r_h) A^s_h X_h w(1, 0, X_h)\ (\mathcal{S}^{10}_{1h} + \mathcal{S}^{10}_{2h})(\boldsymbol{\zeta}, \boldsymbol{\beta}^*, \boldsymbol{\xi})$ equals or is close to $\mathbf{0}$ for general $\boldsymbol{\xi}$, where $(\mathcal{S}^{10}_{1h} + \mathcal{S}^{10}_{2h})(\boldsymbol{\zeta}, \boldsymbol{\beta}^*, \boldsymbol{\xi})$ is defined as

$$\begin{aligned} (\mathcal{S}^{10}_{1h} + \mathcal{S}^{10}_{2h})(\boldsymbol{\zeta}, \boldsymbol{\beta}^*, \boldsymbol{\xi}) &:= g^{-1}[\zeta_2 - f(X_h) - \boldsymbol{\xi}^\top X_h](1 - g^{-1}[\zeta_1 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\xi}^\top X_h]) \\ &\quad - (1 - g^{-1}[\zeta_1 - f(X_h) - \boldsymbol{\xi}^\top X_h])g^{-1}[\zeta_2 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\xi}^\top X_h] \\ &= g^{-1}[\zeta_2 - f(X_h) - \boldsymbol{\xi}^\top X_h] - g^{-1}[\zeta_2 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\xi}^\top X_h] \\ &\quad + g^{-1}[\zeta_1 - f(X_h) - \boldsymbol{\xi}^\top X_h]g^{-1}[\zeta_2 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\xi}^\top X_h] \\ &\quad - g^{-1}[\zeta_2 - f(X_h) - \boldsymbol{\xi}^\top X_h]g^{-1}[\zeta_1 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\xi}^\top X_h]. \end{aligned}$$

Similar to the Taylor series expansion of $\mathcal{S}_{1h}^{00} + \mathcal{S}_{2h}^{00}$ in expression (H.14), we have

$$(\mathcal{S}_{1h}^{10} + \mathcal{S}_{2h}^{10})(\boldsymbol{\zeta}, \boldsymbol{\beta}^*, \boldsymbol{\xi}) = g^{-1}(\zeta_2 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\xi}^\top X_h) \left[1 - g^{-1}(\zeta_1 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\xi}^\top X_h)\right] \left[1 - g^{-1}(\zeta_2 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\xi}^\top X_h) + g^{-1}(\zeta_1 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\xi}^\top X_h)\right] (\Delta^*) + O[\Delta^{*2}], \tag{H.17}$$

where $\Delta^* := \boldsymbol{\beta}^{*\top} X_h - f(X_h)$. If we denote $\eta_1^{a^s, a^r}(\boldsymbol{\beta}^*) = \zeta_1 + \boldsymbol{\beta}^\top \boldsymbol{x}^\beta + \boldsymbol{\xi}^\top a^s \boldsymbol{x}^\xi + \boldsymbol{\psi}^\top a^r \boldsymbol{x}^\psi + \boldsymbol{\phi}^\top a^s a^r \boldsymbol{x}^\phi$ and $\eta_2^{a^s, a^r}(\boldsymbol{\beta}^*) = \zeta_2 + \boldsymbol{\beta}^\top \boldsymbol{x}^\beta + \boldsymbol{\xi}^\top a^s \boldsymbol{x}^\xi + \boldsymbol{\psi}^\top a^r \boldsymbol{x}^\psi + \boldsymbol{\phi}^\top a^s a^r \boldsymbol{x}^\phi$, then, from equation (H.17), we have

$$(\mathcal{S}_{1h}^{10} + \mathcal{S}_{2h}^{10})(\boldsymbol{\zeta}, \boldsymbol{\beta}^*, \boldsymbol{\xi}) = g^{-1}[\eta_2^{1,0}(\boldsymbol{\beta}^*)] \left[1 - g^{-1}[\eta_1^{1,0}(\boldsymbol{\beta}^*)]\right] \left[1 - g^{-1}[\eta_2^{1,0}(\boldsymbol{\beta}^*)] + g^{-1}[\eta_1^{1,0}(\boldsymbol{\beta}^*)]\right] (\Delta^*) + O[\Delta^{*2}].$$

If this is not the case, then the expectation of equation $\sum_i^n (V_{2h} - V_{4h}) = \mathbf{0}$ with $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ and $\boldsymbol{\xi} = \hat{\boldsymbol{\xi}}$ may approach the expectation of equation $\sum_h^n (V_{2h} - V_{4h}) = \mathbf{0}$ with $\boldsymbol{\beta}$ set equal to $\boldsymbol{\beta}^*$ and $\boldsymbol{\xi}$ set equal to a similar limiting value $\boldsymbol{\xi}^*$ as $n \to \infty$. The vector $\boldsymbol{\xi}^*$ will satisfy the condition that the expectation of $\sum_h^n A_h^s (1 - A_h^r) X_h w(1, 0, X_h) \left[g^{-1}(-\boldsymbol{\xi}^{*\top} X_h - \boldsymbol{\beta}^{*\top} X_h) - g^{-1}(-\boldsymbol{\xi}^{*\top} X_h - f(X_h))\right]$ equals or is close to $\mathbf{0}$, but $\boldsymbol{\xi}^*$ will in general be different from the true $\boldsymbol{\xi}$.

Next, we study $\sum_h^H (V_{3h} - V_{4h}) = \mathbf{0}$, that is, $\sum_h^H (1 - A^s) A^r X_h w(0, 1, X_h) (\mathcal{S}_{1h} + \mathcal{S}_{2h}) = \mathbf{0}$, which offers $\hat{\boldsymbol{\psi}}$ in terms of $\hat{\zeta}_1, \hat{\zeta}_2$, and $\hat{\boldsymbol{\beta}}$. To show that $\hat{\boldsymbol{\psi}}$ is (approximately) consistent, we would need to show that the expectation of $\sum_h^H (1 - A_h^s) A_h^r X_h w(0, 1, X_h) (\mathcal{S}_{1h}^{01} + \mathcal{S}_{2h}^{01})(\boldsymbol{\zeta}, \boldsymbol{\beta}^*, \boldsymbol{\psi})$ equals or is close to $\mathbf{0}$ for general $\boldsymbol{\psi}$, where $(\mathcal{S}_{1h}^{01} + \mathcal{S}_{2h}^{01})(\boldsymbol{\zeta}, \boldsymbol{\beta}^*, \boldsymbol{\psi})$ is defined as

$$\begin{aligned}
(\mathcal{S}_{1h}^{01} + \mathcal{S}_{2h}^{01})(\boldsymbol{\zeta}, \boldsymbol{\beta}^*, \boldsymbol{\psi}) &:= g^{-1}[\zeta_2 - f(X_h) - \boldsymbol{\psi}^\top X_h](1 - g^{-1}[\zeta_1 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\psi}^\top X_h]) \\
&\quad - (1 - g^{-1}[\zeta_1 - f(X_h) - \boldsymbol{\psi}^\top X_h]) g^{-1}[\zeta_2 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\psi}^\top X_h] \\
&= g^{-1}[\zeta_2 - f(X_h) - \boldsymbol{\psi}^\top X_h] - g^{-1}[\zeta_2 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\psi}^\top X_h] \\
&\quad + g^{-1}[\zeta_1 - f(X_h) - \boldsymbol{\psi}^\top X_h] g^{-1}[\zeta_2 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\psi}^\top X_h] \\
&\quad - g^{-1}[\zeta_2 - f(X_h) - \boldsymbol{\psi}^\top X_h] g^{-1}[\zeta_1 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\psi}^\top X_h].
\end{aligned}$$

Again, consistent with the Taylor series expansion of $(\mathcal{S}_{1h}^{00} + \mathcal{S}_{2h}^{00})(\boldsymbol{\zeta}, \boldsymbol{\beta})$ in expression (H.14), we have

$$(\mathcal{S}_{1h}^{01} + \mathcal{S}_{2h}^{01})(\boldsymbol{\zeta}, \boldsymbol{\beta}^*, \boldsymbol{\psi}) = g^{-1}(\zeta_2 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\psi}^\top X_h) \left[1 - g^{-1}(\zeta_1 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\psi}^\top X_h)\right] \left[1 - g^{-1}(\zeta_2 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\psi}^\top X_h) + g^{-1}(\zeta_1 - \boldsymbol{\beta}^{*\top} X_h - \boldsymbol{\psi}^\top X_h)\right] (\Delta^*) + O[\Delta^{*2}],$$

or equivalently,

$$(\mathcal{S}_{1h}^{01}+\mathcal{S}_{2h}^{01})(\boldsymbol{\zeta},\boldsymbol{\beta}^*,\boldsymbol{\psi}) = g^{-1}[\eta_2^{0,1}(\boldsymbol{\beta}^*)]\left[1-g^{-1}[\eta_1^{0,1}(\boldsymbol{\beta}^*)]\right] \left[1-g^{-1}[\eta_2^{0,1}(\boldsymbol{\beta}^*)]+g^{-1}[\eta_1^{0,1}(\boldsymbol{\beta}^*)]\right](\Delta^*)+O[\Delta^{*2}],$$

Similar arguments are applied to $\sum_h^H V_{4h}=\mathbf{0}$, that is, $\sum_h^H A^s A^r X_h w(1,1,X_h)\,(\mathcal{S}_{1h}+\mathcal{S}_{2h})=\mathbf{0}$, which offers $\hat{\boldsymbol{\phi}}$ in terms of $\hat{\zeta}_1,\hat{\zeta}_2,\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\xi}}$, and $\hat{\boldsymbol{\phi}}$. To show that $\hat{\boldsymbol{\phi}}$ is (approximately) consistent, we would need to show that the expectation of $\sum_h^H A_h^s A_h^r X_h w(1,1,X_h)(\mathcal{S}_{1h}^{11}+\mathcal{S}_{2h}^{11})(\boldsymbol{\zeta},\boldsymbol{\beta}^*,\boldsymbol{\theta}_2)$ equals or is close to $\mathbf{0}$ for general $\boldsymbol{\phi}$, where $(\mathcal{S}_{1h}^{11}+\mathcal{S}_{2h}^{11})(\boldsymbol{\zeta},\boldsymbol{\beta}^*,\boldsymbol{\theta}_2)$ can be written as by Taylor series expansion

$$(\mathcal{S}_{1h}^{11}+\mathcal{S}_{2h}^{11})(\boldsymbol{\zeta},\boldsymbol{\beta}^*,\boldsymbol{\theta}_2) := g^{-1}[\eta_2^{1,1}(\boldsymbol{\beta}^*)]\left[1-g^{-1}[\eta_1^{1,1}(\boldsymbol{\beta}^*)]\right]\left[1-g^{-1}[\eta_2^{1,1}(\boldsymbol{\beta}^*)] + g^{-1}[\eta_1^{1,1}(\boldsymbol{\beta}^*)]\right](\Delta^*)+O[\Delta^{*2}].$$

Therefore, the expectation of $\sum_h \begin{bmatrix} V_{1h}-V_{2h}-V_{3h}+V_{4h} \\ V_{2h}-V_{4h} \\ V_{3h}-V_{4h} \\ V_{4h} \end{bmatrix}$ conditional on $(X_1,\ldots X_n)$, that is, $\sum_h \left\{ \begin{pmatrix} \pi^{00} \\ \pi^{10} \\ \pi^{01} \\ \pi^{11} \end{pmatrix} X_h \begin{pmatrix} w(0,0,X_h) \\ w(1,0,X_h) \\ w(0,1,X_h) \\ w(1,1,X_h) \end{pmatrix} \begin{bmatrix} (\mathcal{S}_{1h}^{00}+\mathcal{S}_{2h}^{00})(\boldsymbol{\zeta},\boldsymbol{\beta}) \\ (\mathcal{S}_{1h}^{10}+\mathcal{S}_{2h}^{10})(\boldsymbol{\zeta},\boldsymbol{\beta},\boldsymbol{\xi}) \\ (\mathcal{S}_{1h}^{01}+\mathcal{S}_{2h}^{01})(\boldsymbol{\zeta},\boldsymbol{\beta},\boldsymbol{\psi}) \\ (\mathcal{S}_{1h}^{11}+\mathcal{S}_{2h}^{11})(\boldsymbol{\zeta},\boldsymbol{\beta},\boldsymbol{\theta}_2) \end{bmatrix} \right\}$, can be written using a Taylor series expansion as

$$\sum_h \left\{ \begin{pmatrix} \pi^{00} \\ \pi^{10} \\ \pi^{01} \\ \pi^{11} \end{pmatrix} X_h \begin{pmatrix} w(0,0) \\ w(1,0) \\ w(0,1) \\ w(1,1) \end{pmatrix} \begin{bmatrix} g^{-1}(\eta_2^{00})\left[1-g^{-1}(\eta_1^{00})\right]\left[1-g^{-1}(\eta_2^{00})+g^{-1}(\eta_1^{00})\right](\Delta)+O[\Delta^2] \\ g^{-1}(\eta_2^{10})\left[1-g^{-1}(\eta_1^{10})\right]\left[1-g^{-1}(\eta_2^{10})+g^{-1}(\eta_1^{10})\right](\Delta)+O[\Delta^2] \\ g^{-1}(\eta_2^{01})\left[1-g^{-1}(\eta_1^{01})\right]\left[1-g^{-1}(\eta_2^{01})+g^{-1}(\eta_1^{01})\right](\Delta)+O[\Delta^2] \\ g^{-1}(\eta_2^{11})\left[1-g^{-1}(\eta_1^{11})\right]\left[1-g^{-1}(\eta_2^{11})+g^{-1}(\eta_1^{11})\right](\Delta)+O[\Delta^2] \end{bmatrix} \right\}, \tag{H.18}$$

where $\eta_1^{a^s a^r}=\zeta_1+\boldsymbol{\beta}^\top \boldsymbol{x}^\beta+\boldsymbol{\xi}^\top a^s \boldsymbol{x}^\xi+\boldsymbol{\psi}^\top a^r \boldsymbol{x}^\psi+\boldsymbol{\phi}^\top a^s a^r \boldsymbol{x}^\phi$ and $\eta_2^{a^s a^r}=\zeta_2+\boldsymbol{\beta}^\top \boldsymbol{x}^\beta+\boldsymbol{\xi}^\top a^s \boldsymbol{x}^\xi+\boldsymbol{\psi}^\top a^r \boldsymbol{x}^\psi+\boldsymbol{\phi}^\top a^s a^r \boldsymbol{x}^\phi$.

Define $\kappa^*(A^s,A^r,X)=g^{-1}(\eta_2)\left[1-g^{-1}(\eta_1)\right]\left[1-g^{-1}(\eta_2)+g^{-1}(\eta_1)\right]$ with $\eta_1(a^s,a^r,X)=\zeta_1^*+\boldsymbol{\beta}^{*\top}X+\boldsymbol{\xi}^{*\top}a^s X+\boldsymbol{\psi}^{*\top}a^r X+\boldsymbol{\phi}^{*\top}a^s a^r X$, $\eta_2(a^s,a^r,X)=\zeta_2^*+\boldsymbol{\beta}^{*\top}X+\boldsymbol{\xi}^{*\top}a^s X+\boldsymbol{\psi}^{*\top}a^r X+\boldsymbol{\phi}^{*\top}a^s a^r X$, and $\boldsymbol{\xi}^*$, $\boldsymbol{\psi}^*$, and $\boldsymbol{\phi}^*$ are assumed to be limiting values for $\hat{\boldsymbol{\xi}}$, $\hat{\boldsymbol{\psi}}$, and $\hat{\boldsymbol{\phi}}$, respectively.

Then if weights are defined to satisfy a new balancing criterion

$$\pi^{00}w(0,0,X)\kappa(0,0,X) = \pi^{01}w(0,1,X)\kappa(0,1,X) = \pi^{10}w(1,0,X)\kappa(1,0,X) = \pi^{11}w(1,1,X)\kappa(1,1,X),$$

and if the distribution of X is such that the inverse link function is close to linear for the range of $f(X) - \boldsymbol{\beta}^{*\top}X$ (so that the Taylor expansion error term is small), the fact that the expectation of the first (top) equation in (H.18) is $\mathbf{0}$ for $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ means that the expectation of the remaining (the second to the forth) equations (H.18) are close to $\mathbf{0}$ too for $\boldsymbol{\beta} = \boldsymbol{\beta}^*$. Again, this argument establishes the approximate consistency of the corresponding new estimators of $\boldsymbol{\xi}$, $\boldsymbol{\psi}$, and $\boldsymbol{\phi}$.

To conclude, the corresponding overlap-type weights for POM with ordinal outcomes are:

$$w(a^s, a^r) \propto \frac{\pi^{00}\pi^{10}\pi^{01}\pi^{11}}{\pi^{a^s a^r}} \times \frac{\kappa(0,0,\boldsymbol{x})\kappa(1,0,\boldsymbol{x})\kappa(0,1,\boldsymbol{x})\kappa(1,1,\boldsymbol{x})}{\kappa(a^s, a^r, \boldsymbol{x})}, \quad \text{for } a^s,\ a^r = 0, 1.$$

where $\kappa(a^s, a^r, \boldsymbol{x}) = g^{-1}(\eta_2)\left[1 - g^{-1}(\eta_1)\right]\left[1 - g^{-1}(\eta_2) + g^{-1}(\eta_1)\right]$ with $\eta_1(a^s, a^r, \boldsymbol{x}) = \zeta_1^* + \boldsymbol{\beta}^{*\top}\boldsymbol{x}^\beta + \boldsymbol{\xi}^{*\top}a^s\boldsymbol{x}^\xi + \boldsymbol{\psi}^{*\top}a^r\boldsymbol{x}^\psi + \boldsymbol{\phi}^{*\top}a^s a^r\boldsymbol{x}^\phi$, $\eta_2(a^s, a^r, \boldsymbol{x}) = \zeta_2^* + \boldsymbol{\beta}^{*\top}\boldsymbol{x}^\beta + \boldsymbol{\xi}^{*\top}a^s\boldsymbol{x}^\xi + \boldsymbol{\psi}^{*\top}a^r\boldsymbol{x}^\psi + \boldsymbol{\phi}^{*\top}a^s a^r\boldsymbol{x}^\phi$, and ζ_1^*, ζ_2^*, $\boldsymbol{\beta}^*$, $\boldsymbol{\xi}^*$, $\boldsymbol{\psi}^*$ and $\boldsymbol{\phi}^*$ are the solutions of the estimation functions of POM (5.8) with standard overlap weights (5.3).

Consequently, in single-stage decision settings, the following algorithm for estimation of $\boldsymbol{\xi}$, $\boldsymbol{\psi}$, and $\boldsymbol{\phi}$ applies:
Step 1: Conduct a weighted POM to obtain $\hat{\zeta}$, $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\xi}}$, $\hat{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\phi}}$ with standard overlap weights $w(a^s, a^r) = \frac{\pi^{00}\pi^{10}\pi^{01}\pi^{11}}{\pi^{a^s a^r}}$.
Step 2: Compute the new weights that satisfy $\pi^{00}w(0,0,\boldsymbol{x})\kappa(0,0,\boldsymbol{x}) = \pi^{01}w(0,1,\boldsymbol{x})\kappa(0,1,\boldsymbol{x}) = \pi^{10}w(1,0,\boldsymbol{x})\kappa(1,0,\boldsymbol{x}) = \pi^{11}w(1,1,\boldsymbol{x})\kappa(1,1,\boldsymbol{x})$. For example, the weights can be

$$w(a^s, a^r) = \frac{\pi^{00}\pi^{10}\pi^{01}\pi^{11}}{\pi^{a^s a^r}} \times \frac{\kappa(0,0,\boldsymbol{x})\kappa(1,0,\boldsymbol{x})\kappa(0,1,\boldsymbol{x})\kappa(1,1,\boldsymbol{x})}{\kappa(a^s, a^r, \boldsymbol{x})}, \quad for\ a^s,\ a^r = 0, 1,$$

where $\kappa(a^s, a^r, \boldsymbol{x}) = g^{-1}(\hat{\eta}_2)\left[1 - g^{-1}(\hat{\eta}_1)\right]\left[1 - g^{-1}(\hat{\eta}_2) + g^{-1}(\hat{\eta}_1)\right]$ with $\hat{\eta}_1(a^s, a^r, \boldsymbol{x}) = \hat{\zeta}_1 + \hat{\boldsymbol{\beta}}^\top\boldsymbol{x}^\beta + \hat{\boldsymbol{\xi}}^\top a^s\boldsymbol{x}^\xi + \hat{\boldsymbol{\psi}}^\top a^r\boldsymbol{x}^\psi + \hat{\boldsymbol{\phi}}^\top a^s a^r\boldsymbol{x}^\phi$, $\hat{\eta}_2(a^s, a^r, \boldsymbol{x}) = \hat{\zeta}_2 + \hat{\boldsymbol{\beta}}^\top\boldsymbol{x}^\beta + \hat{\boldsymbol{\xi}}^\top a^s\boldsymbol{x}^\xi + \hat{\boldsymbol{\psi}}^\top a^r\boldsymbol{x}^\psi + \hat{\boldsymbol{\phi}}^\top a^s a^r\boldsymbol{x}^\phi$.
Step 3: Use the new weights from Step 2, and conduct weighted POM again, to get new approximately consistent estimators $\widetilde{\boldsymbol{\xi}}$, $\widetilde{\boldsymbol{\psi}}$, $\widetilde{\boldsymbol{\phi}}$ for treatment decisions.

H.3 Simulation Study 1 figures

In this section, we present the simulation (Study 1) figures, which depict the distribution of the blip parameter estimates in Scenarios 1, 2, and 4. Figure H.2 corresponds to Scenario 1 where both the

treatment model and treatment-free model are misspecified; Figure H.3 correspond to Scenario 2 where the treatment-free model is correctly specified but the treatment model is misspecified, and Figure H.3 present the result from Scenario 4 in which both treatment model and treatment-free model are correctly specified.

From Figure H.2, even though both the treatment model and treatment-free model are incorrectly specified, M4 which utilizes adjusted overlap-type weights can still provide less biased blip parameters' estimators. Except for estimator $\hat{\boldsymbol{\psi}}$, which displays little bias, estimators $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\phi}}$ appear unbiased. From Figures H.3 and H.4, because of the correct identification of the treatment-free model, all the methods provide consistent estimators of the blip parameters.

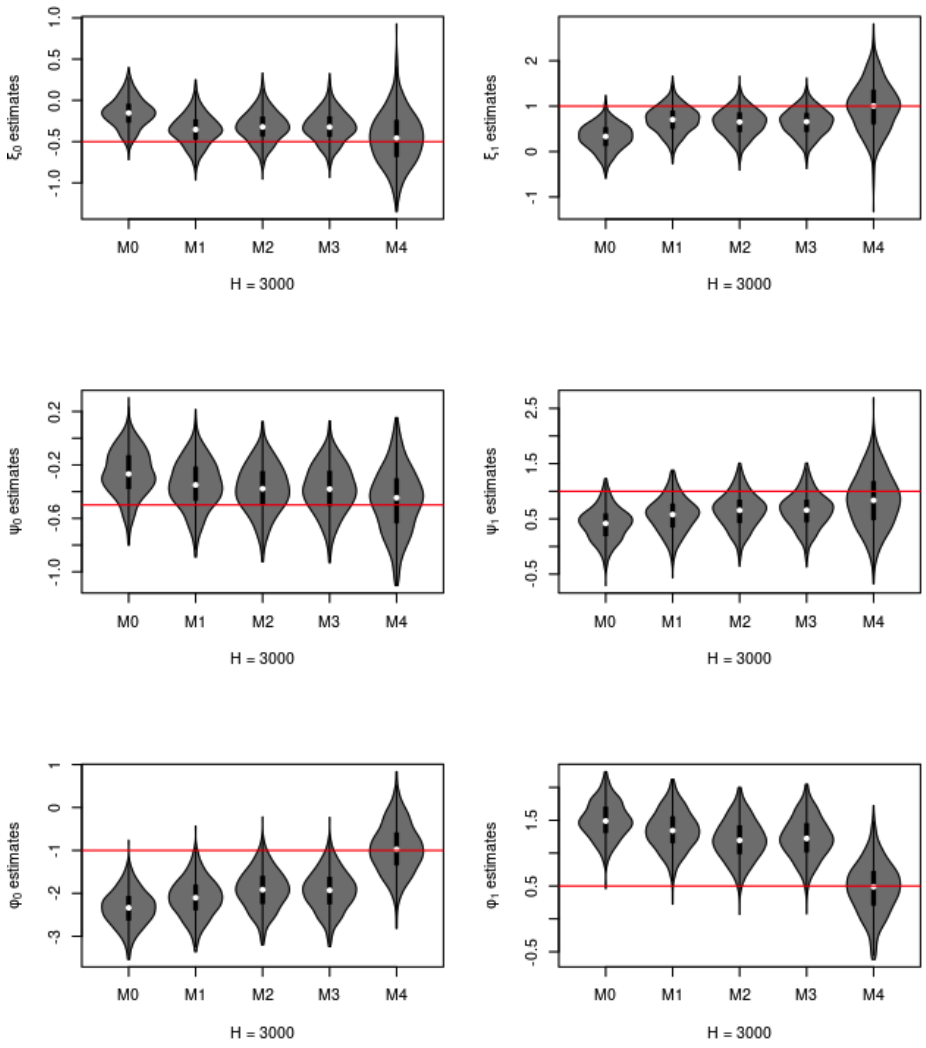


Figure H.2: Blip function parameter estimates, $\hat{\boldsymbol{\xi}}$ (top row), $\hat{\boldsymbol{\psi}}$ (middle row), and $\hat{\boldsymbol{\phi}}$ (bottom row) via Method 0 (M0, Q-learning), Method 1 (M1, no treatment-association dWPOM), Method 2 (M2, treatment-association aware dWPOM with IPW-type weights), Method 3 (M3, treatment-association aware dWPOM with overlap-type weights) and Method 4 (M4, treatment-association aware dWPOM with adjusted overlap-type weights), when both treatment model and treatment-free model are misspecified (Scenario 1).

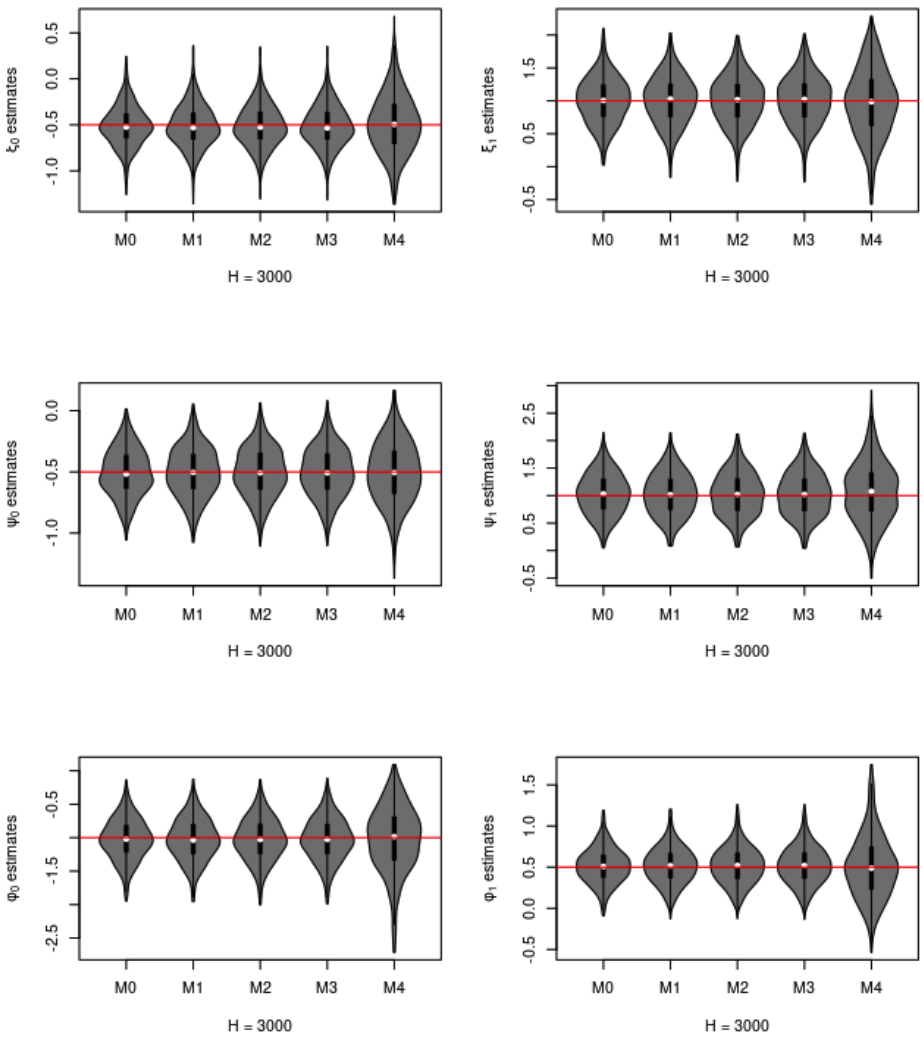


Figure H.3: Blip function parameter estimates, $\hat{\boldsymbol{\xi}}$ (top row), $\hat{\boldsymbol{\psi}}$ (middle row), and $\hat{\boldsymbol{\phi}}$ (bottom row) via Method 0 (M0, Q-learning), Method 1 (M1, no treatment-association dWPOM), Method 2 (M2, treatment-association aware dWPOM with IPW-type weights), Method 3 (M3, treatment-association aware dWPOM with overlap-type weights) and Method 4 (M4, treatment-association aware dWPOM with adjusted overlap-type weights), when the treatment-free model is correctly specified but the treatment model is misspecified (Scenario 2).

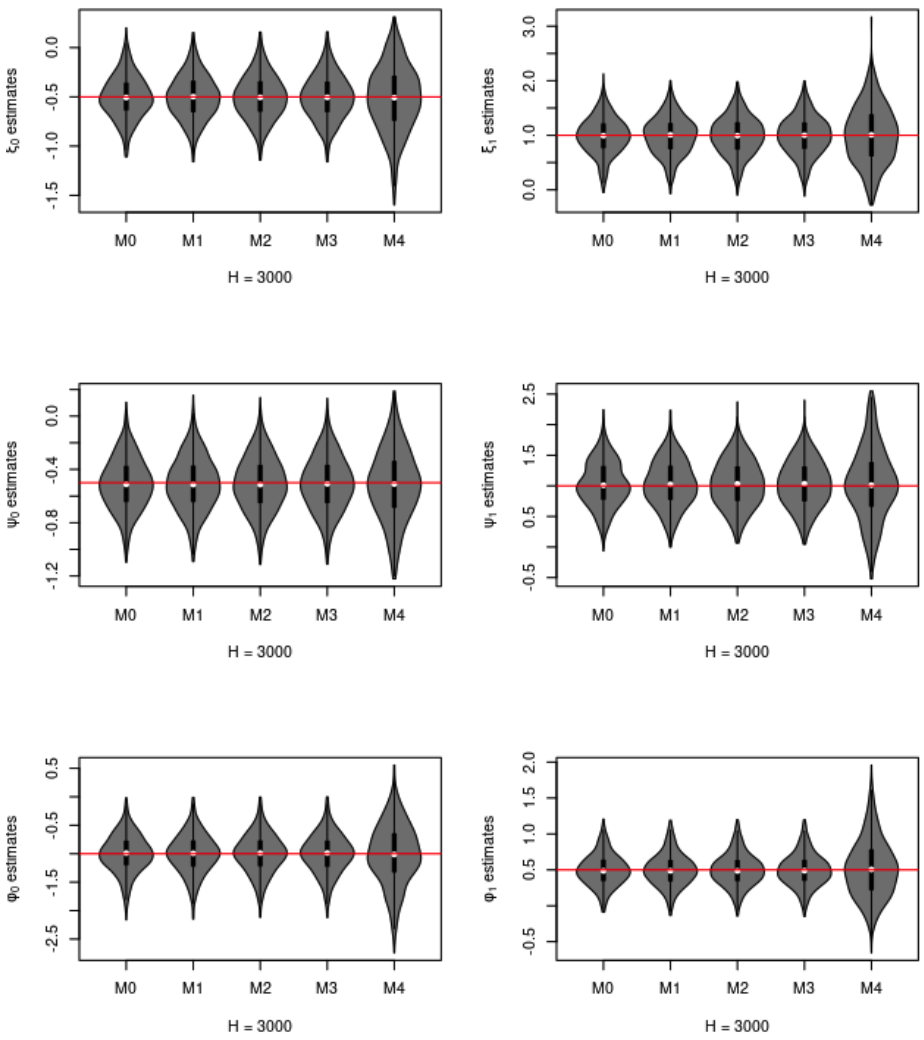


Figure H.4: Blip function parameter estimates, $\hat{\boldsymbol{\xi}}$ (top row), $\hat{\boldsymbol{\psi}}$ (middle row), and $\hat{\boldsymbol{\phi}}$ (bottom row) via Method 0 (M0, Q-learning), Method 1 (M1, no treatment-association dWPOM), Method 2 (M2, treatment-association aware dWPOM with IPW-type weights), Method 3 (M3, treatment-association aware dWPOM with overlap-type weights) and Method 4 (M4, treatment-association aware dWPOM with adjusted overlap-type weights), when both treatment model and treatment-free model are correctly specified (Scenario 4).

H.4 Additional simulation results for the two-stage decision problem (Study 2)

In this section, we present the additional simulation results from simulation Study 2, a two-stage DTR estimation with ordinal outcomes under interference. Consistent with the same data-generating process in Case (1) Study 2, Case (2), in Stage 2, misspecifies the treatment-free model but correctly specifies the treatment model. Stage 1, Case (2) misspecifies the treatment model, but correctly specifies the treatment-free model.

The distributions of the blip estimates (i.e., $\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\phi}}$) from Case (2) are presented in Figures H.5 and H.6, which correspond to Stage 1 and Stage 2, respectively. Similar to the results for Case (1), in Figure H.6, which depicts blip estimates from Cases (2) in Stage 2, all the blip estimates from our dWPOM appear to be normally distributed and centred by the true blip parameters' values, but Q-learning provides biased estimators. From Figure H.5, which corresponds to Case (1) in Stage 1, for our dWPOM, blip estimates $\hat{\xi}_0, \hat{\psi}_0, \hat{\boldsymbol{\phi}}$ are also normally distributed and centred by the true parameters' values, but the blip estimates $\hat{\xi}_1, \hat{\psi}_1$ appear to be slightly off the true values. Again, we suspect that this misalignment results from the approximately consistent estimation, essentially caused by the omission of remainder terms in the Taylor expansion.

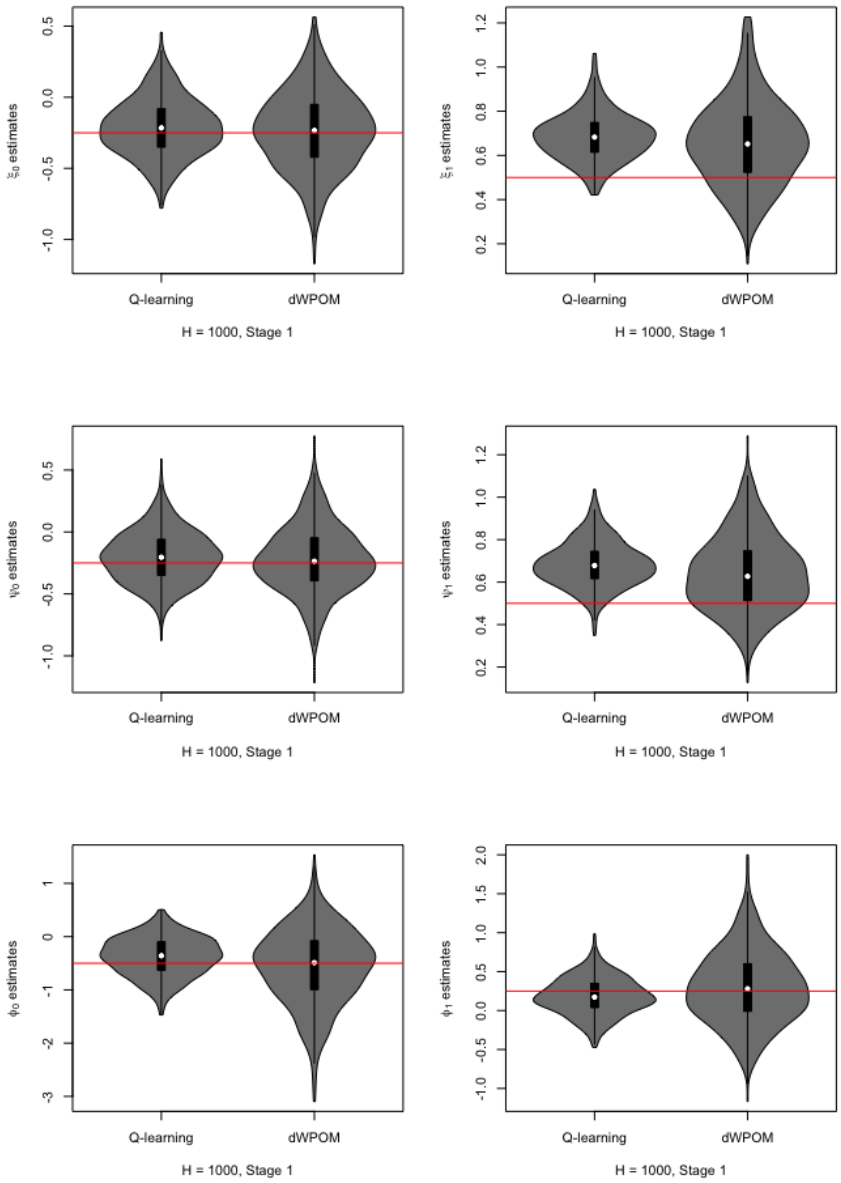


Figure H.5: Blip function parameter estimates in Stage 1 of Study 2, $\hat{\boldsymbol{\xi}}$ (top row), $\hat{\boldsymbol{\psi}}$ (middle row), and $\hat{\boldsymbol{\phi}}$ (bottom row) via Q-learning and treatment-association aware dWPOM with adjusted overlap-type weights in Case (2), where the treatment-free model is misspecified, but the treatment model is correctly specified in Stage 2, and the treatment model is misspecified, but the treatment-free model is correctly specified in Stage 1.

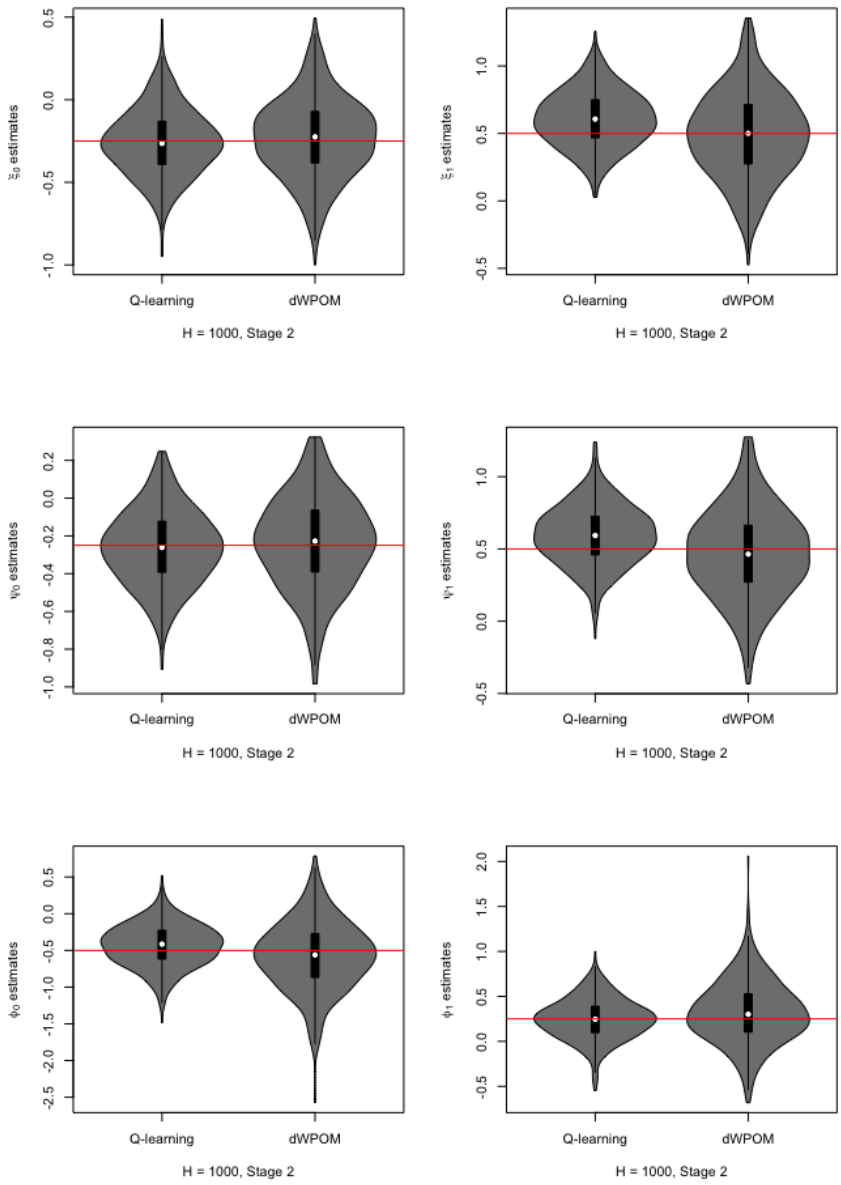


Figure H.6: Blip function parameter estimates in Stage 2 of Study 2, $\hat{\boldsymbol{\xi}}$ (top row), $\hat{\boldsymbol{\psi}}$ (middle row), and $\hat{\boldsymbol{\phi}}$ (bottom row) via Q-learning and treatment-association aware dWPOM with adjusted overlap-type weights in Case (2), where the treatment-free model is misspecified, but the treatment model is correctly specified in Stage 2, and the treatment model is misspecified, but the treatment-free model is correctly specified in Stage 1.

Appendix I

Appendix of Chapter 6

This appendix contains two sections. One is related to the extension of X-leaner for household interference, the other is the extension of the outcome weighted learning for the interference.

I.1 Net-learning for treatment decision with household interference

In this appendix section, we extend the X-learner method to the case of household interference, and our new method is referred to as Net-learning. Considering four potential outcome for the pair in the same household: $Y^*(1,1), Y^*(1,0), Y^*(0,1), Y^*(0,0)$, we assume that no unmeasured confounding, such that $Y^*(a^s, a^r) \perp\!\!\!\perp (a^s, a^r) \mid \boldsymbol{x}$. Then, we define that:

$$\begin{aligned}
\mu_{11}(\boldsymbol{x}) &:= \mathbb{E}[Y^*(1,1)|\boldsymbol{x}] = \mathbb{E}[Y|A^s = 1, A^r = 1, \boldsymbol{x}], \\
\mu_{10}(\boldsymbol{x}) &:= \mathbb{E}[Y^*(1,0)|\boldsymbol{x}] = \mathbb{E}[Y|A^s = 1, A^r = 0, \boldsymbol{x}], \\
\mu_{01}(\boldsymbol{x}) &:= \mathbb{E}[Y^*(0,1)|\boldsymbol{x}] = \mathbb{E}[Y|A^s = 0, A^r = 1, \boldsymbol{x}], \\
\mu_{00}(\boldsymbol{x}) &:= \mathbb{E}[Y^*(0,0)|\boldsymbol{x}] = \mathbb{E}[Y|A^s = 0, A^r = 0, \boldsymbol{x}].
\end{aligned}$$

Suppose the true outcome model can be set as the form of

$$\mathbb{E}[Y|A^s = a^s, A^r = a^r, \boldsymbol{x}] = f(\boldsymbol{x}^\beta; \boldsymbol{\beta}) + a^s d_\xi(\boldsymbol{\xi}, \boldsymbol{x}^\xi) + a^r d_\psi(\boldsymbol{\psi}, \boldsymbol{x}^\psi) + a^s a^r d_{int}(\boldsymbol{\phi}, \boldsymbol{x}^\phi). \quad \text{(I.1)}$$

where $\boldsymbol{x}^\beta$ are predictive variables, and $\boldsymbol{x}^\xi$, $\boldsymbol{x}^\psi$, $\boldsymbol{x}^\phi$ are tailoring variables for adapting treatment decisions to patients. We also suppose that $f(\boldsymbol{x}^\beta; \boldsymbol{\beta})$, $d_\xi(\boldsymbol{\xi}, \boldsymbol{x}^\xi)$, $d_\psi(\boldsymbol{\psi}, \boldsymbol{x}^\psi)$ and $d_{int}(\boldsymbol{\phi}, \boldsymbol{x}^\phi)$ can be

any arbitrary function. Thus, we have:

$$\begin{aligned}
\mu_{11}(\boldsymbol{x}) &= f(\boldsymbol{x}^{\beta};\boldsymbol{\beta}) + d_{\xi}(\boldsymbol{\xi},\boldsymbol{x}^{\xi}) + d_{\psi}(\boldsymbol{\psi},\boldsymbol{x}^{\psi}) + d_{int}(\boldsymbol{\phi},\boldsymbol{x}^{\phi}),\\
\mu_{10}(\boldsymbol{x}) &= f(\boldsymbol{x}^{\beta};\boldsymbol{\beta}) + d_{\xi}(\boldsymbol{\xi},\boldsymbol{x}^{\xi}),\\
\mu_{01}(\boldsymbol{x}) &= f(\boldsymbol{x}^{\beta};\boldsymbol{\beta}) + d_{\psi}(\boldsymbol{\psi},\boldsymbol{x}^{\psi}),\\
\mu_{00}(\boldsymbol{x}) &= f(\boldsymbol{x}^{\beta};\boldsymbol{\beta}).
\end{aligned}$$

After some algebra, we have different effects functions, we called it *decision functions* on which we make decisions based:

$$\begin{aligned}
d_{\xi}(\boldsymbol{\xi},\boldsymbol{x}^{\xi}) &= \mu_{10}(\boldsymbol{x}) - \mu_{00}(\boldsymbol{x}) := \mathbb{E}[Y^*(1,0) - Y^*(0,0)|\boldsymbol{x}],\\
d_{\psi}(\boldsymbol{\psi},\boldsymbol{x}^{\psi}) &= \mu_{01}(\boldsymbol{x}) - \mu_{00}(\boldsymbol{x}) := \mathbb{E}[Y^*(0,1) - Y^*(0,0)|\boldsymbol{x}],\\
d_{int}(\boldsymbol{\phi},\boldsymbol{x}^{\phi}) &= \mu_{11}(\boldsymbol{x}) - \mu_{10}(\boldsymbol{x}) - \mu_{01}(\boldsymbol{x}) + \mu_{00}(\boldsymbol{x})\\
&:= \mathbb{E}[Y^*(1,1) - Y^*(1,0) - Y^*(0,1) + Y^*(0,0)|\boldsymbol{x}].
\end{aligned} \tag{I.2}$$

Based on two distinct household decision scenes, we suggest two distinct decision rules: 1. Making a decision only for one individual while the other's treatment is assumed to be fixed. 2. Making a decision for both individuals in the same household. First, we have decision rule 1:

Decision 2. The optimal decision rules only for one individual:
The treatment of one individual (say, s of the pair (s,r) in the household) is already predetermined (i.e., treatment is fixed throughout the study period), and the objective is to make the optimal decision for the other (i.e., r).

- If $A^s = 1$, and if $d_{\psi}(\boldsymbol{x}^{\psi}) + d_{int}(\boldsymbol{x}^{\phi}) > 0$, then $A^{r*} = 1$; otherwise $A^{r*} = 0$;
- If $A^s = 0$, and if $d_{\psi}(\boldsymbol{x}^{\psi}) > 0$, then $A^{r*} = 1$; otherwise $A^{r*} = 0$.

Then, in the second case, we are allowed to determine both the treatment of the ego and the treatment of the alter. That is, we can choose $(A^s, A^r) = (1,1), (1,0), (0,1)$ or $(0,0)$ to maximize the outcome. These treatment combinations correspond to γ being $d_{\xi}(\boldsymbol{x}^{\xi}) + d_{\psi}(\boldsymbol{x}^{\psi}) + d_{int}(\boldsymbol{x}^{\phi})$, $d_{\xi}(\boldsymbol{x}^{\xi})$, $d_{\psi}(\boldsymbol{x}^{\psi})$ and 0. Therefore, we have four treatment rules in the decsion rule 2:

Decision 3. The optimal household decision rules:

- Rule 1: $(A^s, A^r)^{opt} = (1,1)$ if $d_{\xi}(\boldsymbol{x}^{\xi}) + d_{\psi}(\boldsymbol{x}^{\psi}) + d_{int}(\boldsymbol{x}^{\phi}) > 0$ and $d_{\psi}(\boldsymbol{x}^{\psi}) + d_{int}(\boldsymbol{x}^{\phi}) > 0$, and $d_{\xi}(\boldsymbol{x}^{\xi}) + d_{int}(\boldsymbol{x}^{\phi}) > 0$.

- Rule 2: $(A^s, A^r)^{opt} = (1,0)$ if $d_{\psi}(\boldsymbol{x}^{\psi}) + d_{int}(\boldsymbol{x}^{\phi}) < 0$ and $d_{\xi}(\boldsymbol{x}^{\xi}) > d_{\psi}(\boldsymbol{x}^{\psi})$ and $d_{\xi}(\boldsymbol{x}^{\xi}) > 0$.
- Rule 3: $(A^s, A^r)^{opt} = (0,1)$ if $d_{\xi}(\boldsymbol{x}^{\xi}) + d_{int}(\boldsymbol{x}^{\phi}) < 0$ and $d_{\xi}(\boldsymbol{x}^{\xi}) < d_{\psi}(\boldsymbol{x}^{\psi})$ and $d_{\psi}(\boldsymbol{x}^{\psi}) > 0$.
- Rule 4: $(A^s, A^r)^{opt} = (0,0)$ if $d_{\xi}(\boldsymbol{x}^{\xi}) + d_{\psi}(\boldsymbol{x}^{\psi}) + d_{int}(\boldsymbol{x}^{\phi}) < 0$ and $d_{\xi}(\boldsymbol{x}^{\xi}) < 0$ and $d_{\psi}(\boldsymbol{x}^{\psi}) < 0$.

For the estimation process, based on different treatment combinations, $(A^s, A^r) = (1,1), (1,0), (0,1)$ or $(0,0)$, data can be divided into four groups. Building on these subgroup dataset, we can estimate $\mu_{11}(\boldsymbol{x}), \mu_{10}(\boldsymbol{x}), \mu_{01}(\boldsymbol{x}), \mu_{00}(\boldsymbol{x})$ by any supervised learning algorithm and the estimators are denoted as $\hat{\mu}_{11}(\boldsymbol{x}), \hat{\mu}_{10}(\boldsymbol{x}), \hat{\mu}_{01}(\boldsymbol{x})$ and $\hat{\mu}_{00}(\boldsymbol{x})$, respectively. Building on the estimators, we can impute the unobserved potential outcomes for each individual in a certain treatment combination group, and the results are shown in Table I.1.

Table I.1: Individuals' potential outcomes and the treatment configuration of a pair in the household.

Groups \ Outcomes	$Y^*(1,1)$	$Y^*(1,0)$	$Y^*(0,1)$	$Y^*(0,0)$
$(A^s, A^r) = (1,1)$	Y_{11}	$\hat{\mu}_{10}$	$\hat{\mu}_{01}$	$\hat{\mu}_{00}$
$(A^s, A^r) = (1,0)$	$\hat{\mu}_{11}$	Y_{10}	$\hat{\mu}_{01}$	$\hat{\mu}_{00}$
$(A^s, A^r) = (0,1)$	$\hat{\mu}_{11}$	$\hat{\mu}_{10}$	Y_{01}	$\hat{\mu}_{00}$
$(A^s, A^r) = (0,0)$	$\hat{\mu}_{11}$	$\hat{\mu}_{10}$	$\hat{\mu}_{01}$	Y_{00}

Note that, in Table I.1, $Y_{a^s a^r}$ refers to the observed (potential) outcomes for the (a^s, a^r) treatment group, where $a^s = 0, 1$ and $a^r = 0, 1$. This follows the so-called consistency assumption that $Y_{a^s a^r} = Y^*(a^s, a^r)\mathbb{I}(A^s = a^s, A^r = a^r)$. $\hat{\mu}_{a^s a^r}$ are the corresponding estimated outcomes based on the covariate in a certain treatment group of interest. For a certain treatment group (each row in Table I.1), according to equations (I.2), one can compute individual treatment effect estimates that $\hat{\varphi}_{\xi,i}(\boldsymbol{x}), \hat{\varphi}_{\psi,i}(\boldsymbol{x})$ and $\hat{\varphi}_{\phi,i}(\boldsymbol{x})$ for each individual i in that group, then, respectively, fit models $\hat{d}_{\xi}(\boldsymbol{x}), \hat{d}_{\psi}(\boldsymbol{x})$ and $\hat{d}_{\phi}(\boldsymbol{x})$ to predict $\hat{\varphi}_{\xi,i}(\boldsymbol{x}), \hat{\varphi}_{\psi,i}(\boldsymbol{x})$ and $\hat{\varphi}_{\phi,i}(\boldsymbol{x})$ from the corresponding treatment combination group (a^s, a^r). For example, for individuals with covariates $\boldsymbol{x}_{11}$ in the $(A^s, A^r) = (1,1)$ group (i.e., consider the first row in Table I.1),

$$\begin{aligned}
\hat{\varphi}^{11}_{\xi,i}(\boldsymbol{\xi}, \boldsymbol{x}^{\xi}) &= \hat{\mu}_{10}(\boldsymbol{x}_{11,i}) - \hat{\mu}_{00}(\boldsymbol{x}_{11,i}), \\
\hat{\varphi}^{11}_{\psi,i}(\boldsymbol{\psi}, \boldsymbol{x}^{\psi}) &= \hat{\mu}_{01}(\boldsymbol{x}_{11,i}) - \hat{\mu}_{00}(\boldsymbol{x}_{11,i}), \\
\hat{\varphi}^{11}_{int,i}(\boldsymbol{\phi}, \boldsymbol{x}^{\phi}) &= Y_{11} - \hat{\mu}_{10}(\boldsymbol{x}_{11,i}) - \hat{\mu}_{01}(\boldsymbol{x}_{11,i}) + \hat{\mu}_{00}(\boldsymbol{x}_{11,i}),
\end{aligned} \tag{I.3}$$

where the subscript of x means the individual i's data is from the $(A^s, A^r) = (1, 1)$ dataset, and the superscript of $\hat{\varphi}$ functions refers to the calculation for the treatment combination group $(A^s, A^r) = (1, 1)$. Then, based on individuals data one can fit a model $\hat{d}_{\xi}^{11}(\boldsymbol{x})$ to predict $\hat{\varphi}_{\xi,i}^{11}(\boldsymbol{x})$, similarly, acquiring models $\hat{d}_{\psi}^{11}(\boldsymbol{x})$ and $\hat{d}_{int}^{11}(\boldsymbol{x})$ for the treatment combination group $(A^s, A^r) = (1, 1)$. Therefore, we totally have four estimators for each decision function, and each one is from one treatment combination group. Finally, employing the balancing interference weights in Jiang's, which depending on the joint propensity score $\mathbb{P}(A^s, A^r \mid \boldsymbol{x})$, we can acquire the final combined decision functions:

$$\begin{aligned}
\hat{d}_{\xi}(\boldsymbol{x}^{\xi}) &= w_{11}\hat{d}_{\xi}^{11}(\boldsymbol{x}^{\xi}) + w_{10}\hat{d}_{\xi}^{10}(\boldsymbol{x}^{\xi}) + w_{01}\hat{d}_{\xi}^{01}(\boldsymbol{x}^{\xi}) + w_{00}\hat{d}_{\xi}^{00}(\boldsymbol{x}^{\xi}) \\
\hat{d}_{\psi}(\boldsymbol{x}^{\psi}) &= w_{11}\hat{d}_{\psi}^{11}(\boldsymbol{x}^{\psi}) + w_{10}\hat{d}_{\psi}^{10}(\boldsymbol{x}^{\psi}) + w_{01}\hat{d}_{\psi}^{01}(\boldsymbol{x}^{\psi}) + w_{00}\hat{d}_{\psi}^{00}(\boldsymbol{x}^{\psi}), \\
\hat{d}_{int}(\boldsymbol{x}^{\phi}) &= w_{11}\hat{d}_{int}^{11}(\boldsymbol{x}^{\phi}) + w_{10}\hat{d}_{int}^{10}(\boldsymbol{x}^{\phi}) + w_{01}\hat{d}_{int}^{01}(\boldsymbol{x}^{\phi}) + w_{00}\hat{d}_{int}^{00}(\boldsymbol{x}^{\phi}).
\end{aligned}$$

Note that, for the above $(A^s, A^r) = (1, 1)$ group case, when computing $\hat{\varphi}_{int,i}^{11}(\boldsymbol{\phi}, \boldsymbol{x}^{\phi})$ in equation I.3, we use Y_{11} that is observed outcome rather than $\hat{\mu}_{11}$. Therefore, when computing individual treatment effects $\hat{\varphi}^{11}$ for $(A^s, A^r) = (1, 1)$ group, we use the information from all these four treatment combinations groups: Y_{11} is from $(A^s, A^r) = (1, 1)$ group, and $\hat{\mu}$s are from other groups. Similarly, other treatment groups' $\hat{\varphi}$ is also estimated with data from all four treatment groups. This is the spirit of X-learning. If we draw a line between the observed outcomes (one diagonal line from top left to bottom right) and also draw lines between imputed potential outcomes (five parallel lines from the bottom left to top right), we will see that these lines are intertwined together like a net, which is centred by the diagonal line from top left to bottom right. Thus, the name of our Net-learner is derived from this. Now, we conclude this section by outlining the Net-learner algorithm with some discussions (see Algorithm 1).

It is important to note that X-learner is suitable for cases where the propensity score varies significantly, because it explicitly accounts for variation in the propensities. In the presence of interference, treatment configurations are rarely uniformly distributed in the observational data, and extreme treatment configuration distribution is common. Accordingly, the balancing weights that are suggested in Net-learner will function similarly in X-leaner. Because of the proposed interference balancing weights, Net-learner can examine the variation of the joint propensity and address extreme treatment configuration distribution cases.

This section has provided an extension of X-learning in household interference. Our proposed method, termed net-learner, inherits the spirit of X-learning and uses all the treatment-configuration data for an efficient estimation.

Algorithm 1: Net-learner

Input: Outcome, patient history information, and weights

Output: Estimated decision functions

Step 1. Nuisance training:

Construct estimates $\hat{\mu}_{11}(\boldsymbol{x}), \hat{\mu}_{10}(\boldsymbol{x}), \hat{\mu}_{01}(\boldsymbol{x})$, and $\hat{\mu}_{00}(\boldsymbol{x})$ of the regression functions $\mu_{11}(\boldsymbol{x}), \mu_{10}(\boldsymbol{x}), \mu_{01}(\boldsymbol{x})$, and $\mu_{00}(\boldsymbol{x})$, respectively, using the corresponding $(A^s, A^r) = (1,1), (1,0), (0,1), (1,1)$ dataset.

Step 2. Decision functions regression:

(2a) Construct individual treatment effects: for each treatment combination sub dataset, building on equations (I.2), construct each individual's treatment effects, such that, for $a^s = 0, 1$ and $a^r = 0, 1$

$$\begin{aligned}
\hat{\varphi}_{\xi}^{a^s a^r}(\boldsymbol{x}^{\xi}) &= \hat{\mu}_{10}(\boldsymbol{x}) - \hat{\mu}_{00}(\boldsymbol{x}), \\
\hat{\varphi}_{\psi}^{a^s a^r}(\boldsymbol{x}^{\psi}) &= \hat{\mu}_{01}(\boldsymbol{x}) - \hat{\mu}_{00}(\boldsymbol{x}), \\
\hat{\varphi}_{int}^{a^s a^r}(\boldsymbol{x}^{\phi}) &= \hat{\mu}_{11}(\boldsymbol{x}) - \hat{\mu}_{10}(\boldsymbol{x}) - \hat{\mu}_{01}(\boldsymbol{x}) + \hat{\mu}_{00}(\boldsymbol{x}).
\end{aligned} \tag{I.4}$$

Note that $\hat{\mu}_{a^s a^r}(\boldsymbol{x})$ will be substituted by the observed outcome ($Y_{a^s a^r}$) when computing for each corresponding $(A^s, A^r) = (a^s, a^r)$ group. E.g., for $(A^s, A^r) = (0,0)$ group, all $\hat{\mu}_{00}(\boldsymbol{x})$ in equations I.4 should be substituted by observed outcomes, i.e., Y_{00} in Table I.1.

(2b) Regress individual treatment effects on the corresponding covariates, yielding

$$\begin{aligned}
\hat{d}_{\xi}^{a^s a^r}(\boldsymbol{x}^{\xi}) &= \widehat{\mathbb{E}}_n\left[\hat{\varphi}_{\xi}^{a^s a^r}(\boldsymbol{x}^{\xi}) \mid \boldsymbol{X}^{\xi} = \boldsymbol{x}^{\xi}\right], \\
\hat{d}_{\psi}^{a^s a^r}(\boldsymbol{x}^{\psi}) &= \widehat{\mathbb{E}}_n\left[\hat{\varphi}_{\psi}^{a^s a^r}(\boldsymbol{x}^{\psi}) \mid \boldsymbol{X}^{\psi} = \boldsymbol{x}^{\psi}\right], \\
\hat{d}_{int}^{a^s a^r}(\boldsymbol{x}^{\phi}) &= \widehat{\mathbb{E}}_n\left[\hat{\varphi}_{\phi}^{a^s a^r}(\boldsymbol{x}^{\phi}) \mid \boldsymbol{X}^{\phi} = \boldsymbol{x}^{\phi}\right].
\end{aligned}$$

Step 3. Weighted decision functions outputting:

$$\hat{d}_{\xi}(\boldsymbol{x}^{\xi}) = \sum_{a^s, a^r} w_{a^s a^r} \hat{d}_{\xi}^{a^s a^r}; \quad \hat{d}_{\psi}(\boldsymbol{x}^{\psi}) = \sum_{a^s, a^r} w_{a^s a^r} \hat{d}_{\psi}^{a^s a^r}; \quad \hat{d}_{int}(\boldsymbol{x}^{\phi}) = \sum_{a^s, a^r} w_{a^s a^r} \hat{d}_{\phi}^{a^s a^r}.$$

I.2 Interference-aware outcome weighted learning

As stated in Chapter 1, outcome weighted learning directly maximises the value function over a restricted class of regimes, and recasts the IPTW estimation for the treatment rule into a classification problem. In this section, we extend outcome weighted learning to the case in which there is network interference, and in particular, household interference.

Let $\mathbb{P}$ denote the distribution of $(\boldsymbol{X}, A, Y)$, where treatments are randomized, and $\mathbb{P}^d$ denote the distribution of $(\boldsymbol{X}, A, Y)$, where treatments are chosen according to rule $d(\boldsymbol{X})$. Building on the interference assumptions which are presented in Chapter 3, the marginal mean outcome under a regime d is

$$\mathcal{V}(d) = \mathbb{E}^d(Y) = \int Y d\mathbb{P}^d = \int Y \frac{d\mathbb{P}^d}{d\mathbb{P}} d\mathbb{P} = \mathbb{E}\left\{ \frac{\mathbb{I}\left[(A, A_{\mathcal{N}}) = d(\boldsymbol{X})\right]}{\mathbb{P}(A, A_{\mathcal{N}} \mid \boldsymbol{X})} Y \right\}, \tag{I.5}$$

where $\mathbb{P}(A, A_{\mathcal{N}} \mid \boldsymbol{X})$ is the network propensity function, which is defined in Chapter 3.

Optimal Individualized Treatment Rule

$$d^* = \underset{d \in \mathcal{D}}{\operatorname{argmax}} \mathcal{V}(d),$$

or equivalently,

$$d^* = \underset{(A, A_{\mathcal{N}})}{\operatorname{argmax}} \mathbb{E}(Y \mid X, A, A_{\mathcal{N}})$$

Maximizing the value of equations (I.5) is equivalently to minimize the risk

$$\mathbb{E}\left\{ \frac{\mathbb{I}\left[(A, A_{\mathcal{N}}) \neq d(\boldsymbol{X})\right]}{\mathbb{P}(A, A_{\mathcal{N}} \mid \boldsymbol{X})} Y \right\}. \tag{I.6}$$

For simplicity, if we only focus on the egos that are connected with the same number of alters, that is, for all egos, $|\mathcal{N}|$ are equal. Then, the question in Equation (I.6) is a multi-class ($2^{|\mathcal{N}|+1}$ class) classification problem. Various methods of multi-class classifications, such as support vector machines, decision trees, and neural networks, can be employed in this problem. In addition, a key problem is to estimate the network propensity function, $\mathbb{P}(A, A_{\mathcal{N}} \mid \boldsymbol{X})$. In the following part, we concentrate on the household interference. First, we consider household utility, which is defined in Chapter 5, and the mean household utility under a treatment regime d is defined as:

$$\mathcal{V}(d) = \mathbb{E}^d(U) = \int U d\mathbb{P}^d = \int U \frac{d\mathbb{P}^d}{d\mathbb{P}} d\mathbb{P} = \mathbb{E}\left[\frac{\mathbb{I}[(A^s, A^r) = d(\boldsymbol{X})]}{\mathbb{P}(A^s, A^r \mid \boldsymbol{X})} U\right], \tag{I.7}$$

where $\mathbb{P}$ denote the distribution of $(\boldsymbol{X}, A, Y)$, $\mathbb{P}^d$ is the distribution under $(A^s, A^r) = d(\boldsymbol{X})$, and $\mathbb{P}(A^s, A^r \mid \boldsymbol{X})$ is the joint propensity score. Again, the goal is to identify the optimal treatment regime that

$$d^* = \underset{(A^s, A^r)}{\operatorname{argmax}} \mathbb{E}(U \mid X, A^s, A^r),$$

or equivalently minimizing the risk

$$\mathbb{E}\left\{\frac{\mathbb{I}\left[(A^s, A^r) \neq d(\boldsymbol{X})\right]}{\mathbb{P}(A^s, A^r \mid \boldsymbol{X})} U\right\}. \tag{I.8}$$

The question in equation (I.8) is a $2^2 = 4$-class classification problem, and one crucial problem is the estimation of the joint propensity score $\mathbb{P}(A^s, A^r \mid \boldsymbol{X})$.

In this section, we have provided some thoughts on outcome weighted learning in the presence of interference. In the presence of network interference, the DTR estimations for a creation network become a multi-class classification problem, and the key is to estimate the joint propensity functions. For household interference, the DTR estimation for a household is a four-class classification problem. Also, the joint propensity score of the pair is the crucial component of the weight in the outcome weighted learning.

Glossary

"Adjustment" factor: a κ function ($\kappa(\mu) = g^{-1\prime}(\mu) = \frac{dg^{-1}(\mu)}{d\mu}$) of the linear predictor of the GLMs with standard dWOLS weights, used in the balancing weights criterion for binary outcomes.

Blip function: the term or component $\gamma(\boldsymbol{x}^{\psi}, a)$ in the outcome model (single stage), expressing the impact of the treatment on the outcome. (See Section 2.2.2 and Definition 2.2.)

Dynamic treatment regime: sequence of decision rules taking patient information as input and outputting treatment recommendations. (See Definitions 2.1 and 3.1.)

Double robustness of dWOLS: consistency of estimators of $\boldsymbol{\psi}$ if either treatment and/or treatment-free model is correctly specified.

Ego and Alter: In a network, an individual of primary interest is called an ego, and those to whom the ego is linked are the alters.

Household utility function: a combination of the outcomes of individuals sharing a household.

Interference: one individual's outcome being affected by the treatment of others, conditional on that individual's own treatment.

Interference terms: The network interference term maps the configuration of the neighbours' treatments onto an exposure interference term, e.g., number or proportion of treated neighbours. (See Definition 3.3.)

Joint propensity function for a household case: the joint probability of individuals sharing a household receiving a certain treatment configuration. (See Definition 5.1.)

Network propensity function: The joint probability of individual i receiving treatment a_i, and the treatments of his or her neighbours being in the set $s_{i,a}$. That is, $\pi_{i,a_i,s_{i,a}}(\boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i}) = \mathbb{P}(a_i \cap s_{i,a} \mid \boldsymbol{x}_i, \mathcal{N}_i, \mathbf{x}_{\mathcal{N}_i})$. (See Definition 3.2.)

Optimal dynamic treatment regime: sequence of treatment rules that optimizes the expected clinical outcome across all patients.

Optimal household decision rules: decision rules that optimize a chosen household utility function $U(Y^s, Y^r)$. (See Decision 1.)

Treatment-free function: the term or component $f(\boldsymbol{x}^{\beta}; \boldsymbol{\beta})$ in the outcome model (single stage), expressing the expected outcome under no treatment. (See Section 2.2.2.)

Treatment model: $\mathbb{E}[A|\boldsymbol{x}; \boldsymbol{\alpha}]$, which is the propensity score model when the treatment is a binary variable. (See Section 2.2.2.)