Causal Inference with Bipartite Designs

A Dissertation
Submitted to
the Temple University Graduate Board

in Partial Fulfillment
of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY

by

Minzhengxiong Zhang
December, 2023

Examining Committee Members:

Dr. Edoardo M. Airoldi, Advisory Chair, The Statistics, Operations, and Data Science Department
Dr. Cheng Yong Tang, The Statistics, Operations, and Data Science Department
Dr. Donald B. Rubin, The Statistics, Operations, and Data Science Department
Dr. Jean Pouget-Abadie, Google, Algorithms and Optimization Research Group

ABSTRACT

Bipartite experiments have recently emerged as a focal point in causal inference. In these experiments, treatment is administered to one set of units, while outcomes of interest are gauged on a distinct set of units. Such experiments are especially valuable in scenarios where pronounced interference effects transpire between units on a bipartite network. For instance, in market experiments, designating treatment at the seller level and assessing outcomes at the buyer level (or vice-versa) can lead to causal models that more accurately reflect the inherent interference between buyers and sellers. Although bipartite experiments can enhance the precision of causal effect estimations in specific contexts, it's imperative to conduct the analysis judiciously to avoid introducing undue bias through the network. Drawing from the generalized propensity score literature, we demonstrate that it's feasible to achieve unbiased estimates of causal effects for bipartite experiments, given a conventional set of assumptions. Furthermore, we delve into the formulation of confidence sets with accurate coverage probabilities. By employing a bipartite graph from a publicly accessible dataset previously explored in bipartite experiment studies, we illustrate, via simulations, a notable reduction in bias and augmented coverage.

TABLE OF CONTENTS

ABSTRACT iii

LIST OF FIGURES vi

LIST OF TABLES viii

CHAPTER

1. INTRODUCTION 1

1.1 Related Work 3

2. Our Setting 5

2.1 Bipartite Network 5
2.2 Network Exposure 6

3. Estimation Theorem 8

3.1 Unconfoundedness Assumptions 8
3.2 The Generalized Propensity Score 11

4. Estimation Algorithms 15

4.1 Algorithms for Static Network 15
4.1.1 Naïve Regression (Baseline) 16
4.1.2 Two Stage Least Square (TSLS) 16
4.1.3 Model Regression 17
4.1.4 Generalized Additive Models (GAM) 18
4.1.5 Inverse Probability Weight (IPW) 19
4.1.6 IPW with Gradient Boosted Regression Tree Model (IPW Boost) 19
4.1.7 IPW Boost and TSLS (DD TSLS) 20
4.1.8 IPW Boost and GAM (DD GAM) 21
4.2 Algorithms for Dynamic Networks 22
4.2.1 Naïve Regression (Baseline) 22
4.2.2 Two Stage Least Square (TSLS) 23
4.2.3 Model Regression 24
4.2.4 Generalized Additive Models (GAM) 24

4.2.5 Inverse Probability Weight (IPW) 25
4.2.6 IPW with Gradient Boosted Regression Tree Model (IPW Boost) 26
4.2.7 IPW Boost and TSLS (DD TSLS) 27
4.2.8 IPW Boost and GAM (DD GAM) 27

5. Simulation Results 29

5.1 Buyer-Seller Simulation: Static Bipartite Network 29
5.1.1 Large Scale Simulation and ANCOVA Analysis 32
5.2 Buyer-Seller Simulation: Dynamic Bipartite Networks 34
5.2.1 S-shape Learning Curve 39
5.2.2 Dynamic Intercepts 40
5.2.3 Dynamic Bipartite Networks Simulation ANCOVA . . . 41
5.3 Practical Considerations for Unbiased Estimation 43
5.4 Additional Considerations for Observational Data 45

6. Empirical Results 46

6.1 Amazon Review Data: Ranking System to Total Costs 47
6.2 Amazon Review Data: Ranking System to Total Rating 48

7. CONCLUSION 51

BIBLIOGRAPHY 51

APPENDICES 54
A.1 Proof of Theorems 55

LIST OF FIGURES

Figure

2.1 Bipartite Network Matrix 5

2.2 Bipartite Network Example (Sellers and Buyers) 6

4.1 Dynamic Networks Simulation: Dynamic Intercepts of Baseline Algorithm 23

5.1 Static Network Simulation: Static Bipartite Network Example . . . 29

5.2 Static Network Simulation: Exposure-Outcome Scatter Plot 30

5.3 Static Network Simulation: Exposure-Outcome Effect 90% Confidence Intervals 31

5.4 Static Network Simulation: Marginal Exposure-Outcome with 90% Confidence Bands 32

5.5 Dynamic Network at Time 1 35

5.6 Dynamic Network at Time 2 36

5.7 Dynamic Network at Time 3 36

5.8 Dynamic Network at Time 4 37

5.9 Dynamic Network at Time 5 37

5.10 Dynamic Networks Simulation of Time-Exposure-Outcome 3D Scatter Plot 38

5.11 Dynamic Networks Simulation: Time-Outcome Scatter Plot (S-shape Learning Curve) 39

5.12 Dynamic Networks Simulation: Dynamic Intercepts in Baseline Algorithm 41

5.13 Dynamic Networks Simulation: Exposure-Outcome Effect 90% Confidence Intervals 42

6.1 Amazon Pet Supply: Time Series of Product Reviews 46

6.2 Amazon Pet Supply: Exposure-Cost Effect 90% Confidence Intervals 48

6.3 Amazon Pet Supply: Exposure-Rate Effect 90% Confidence Intervals 49

LIST OF TABLES

Table

4.1 Algorithm Names Table 15

5.1 Static Network Batch Simulation: $MSE(\hat{\tau})$ Averages Sample 33

5.2 Static Network Batch Simulation: P-values from ANCOVA Analysis Comparing $MSE(\hat{\tau})$ Across Simulation Hyper-parameters for Different Algorithms 33

5.3 Dynamic Networks Simulation: Estimators of Baseline Algorithm . 40

5.4 Dynamic Networks Batch Simulation: $MSE(\hat{\tau})$ Averages Sample . . 41

5.5 Dynamic Networks Batch Simulation: P-values from ANCOVA Analysis Comparing $MSE(\hat{\tau})$ Across Simulation Hyper-parameters for Different Algorithms 42

6.1 Amazon Pet Supply: Rating and Cost Summary Statistics 47

6.2 Amazon Pet Supply: Buyer's Review Proportion 47

6.3 Amazon Pet Supply: Exposure-Costs Effect Estimator and Standard Errors 47

6.4 Amazon Pet Supply: Exposure-Rate Effect Estimator and Standard Errors 49

CHAPTER 1

INTRODUCTION

Contrary to the majority of experiments employed in both academic and industry settings, which presuppose that the units receiving treatment and those with measurable outcomes impacted by the treatment are identical, bipartite experiments deviate from this assumption. In these experiments, as explored in recent causal inference literature by Papadogeorgou et al. (2019), Pouget-Abadie et al. (2019), and Zigler & Papadogeorgou (2021), two separate groups of units form a bipartite graph. One group, termed the *diversion units*, undergoes the treatment, while the other group, the *outcome units*, may be influenced by this treatment due to exposure to the treated units on the opposite side of the bipartite graph.

Consider, for instance, an experiment on a buyer-item market platform like Amazon or Airbnb, where the treatment induces a modification to the item's offer, such as a price discount or expedited delivery. Randomly assigning treatment to various buyers could present a practical dilemma: buyers might perceive discrimination if they encounter differing offers for an identical item. Implementing treatment at the item level and executing a traditional (non-bipartite) experiment would result in item-level outcome measurements, potentially leading to a different statistical challenge: the presence of substitute goods might breach the stable unit treatment value assumption (SUTVA), which is pivotal for unbiased causal effect estimation. The remedy proposed by Papadogeorgou et al. (2019), Pouget-Abadie et al. (2019), and Zigler & Papadogeorgou (2021) is to allocate treatment at the item level and gauge buyer outcomes.

Within bipartite designs, the units whose outcomes we assess—buyers, in the aforementioned example—cannot be definitively categorized as treated or controlled. To derive causal estimates, experimenters must correlate these outcomes with a measure of treatment exposure they receive, transpiring along the edges of a bipartite graph. This graph, whether weighted or unweighted, is presumed to be fully known and dictates the degree of treatment exposure a unit undergoes. In the context of the market platform experiment, buyers predominantly interacting with treated items might be deemed "highly exposed," whereas those never engaging with treated items are "never exposed." Exposure, whether real-valued or categorical, scalar or vector-valued, invariably stems from the bipartite graph and the treatment and control assignment of the diversion side of the graph. It inherently constitutes a random variable, facilitating causal assertions.

In this paper, we delve into the estimation of causal effects within a bipartite design framework. Specifically, we unveil a generalized-propensity-score-based estimator, demonstrating its unbiased nature under a set of plausible assumptions for the general bipartite graph scenario. We also touch upon the practical execution of these estimators and the statistical inferences derived from such implementations. Interference bias can be pronounced in network contexts. For example, Holtz et al. (2020) utilizes Airbnb data to juxtapose cluster-level randomized experiments, aimed at bias mitigation, with the rudimentary Bernoulli unit-level randomization design. They discern a discrepancy in estimated average treatment effects surpassing 30%, indicating significant interference bias.

Subsequent to this section, we formally delineate the setting and juxtapose our findings with prior research. In Section 2, we employ a rudimentary example to elucidate why certain estimators might falter in real-world scenarios. Section 3 introduces the prerequisites to validate the unbiased nature of our estimator reliant on the propensity score. Section 4 offers pivotal practical insights for the proposed estimation procedure's implementation. In Section 5, we ascertain that elementary bootstrap techniques yield accurate coverage under the uncorrelated error model and demonstrate that our advocated parametric bootstrap approach ensures accurate coverage under the correlated error model. Lastly, Section 6 showcases simulations on an authentic graph previously employed in bipartite experiment studies, revealing significant bias mitigation for the pertinent causal estimands.

1.1 Related Work

Bipartite randomized experiments arise from scenarios where violations of the stable unit treatment value assumption Rubin (1980) transpire, a phenomenon termed as interference. This domain of study traces its roots to early research on irrigation field contamination Kempton et al. (2012) and Struchiner et al. (1990). It has since evolved, with contributions from Hong & Raudenbush (2005), Hudgens & Halloran (2008), Tchetgen & VanderWeele (2012), Toulis & Kao (2013), Forastiere et al. (2021), Galagate (2016), Ogburn et al. (2022), Eckles et al. (2016), Saveski et al. (2017), Saint-Jacques et al. (2019), Johari et al. (2022), Fatemi & Zheleva (2020), Viviano (2020), and others. This body of literature has delved into design modifications and analyses that facilitate superior causal estimates.

The bipartite randomized experiment paradigm, pioneered by Papadogeorgou et al. (2019) and furthered by Pouget-Abadie et al. (2019), stands out for its unique approach of considering separate sets of units for treatment reception and outcome measurement. This distinction, the authors argue, is pivotal for devising more versatile and representative models of treatment responses on bipartite graphs where interference is prevalent.

Both aforementioned papers serve as foundational pillars for our current endeavor. Papadogeorgou et al. (2019) offer invaluable notation, terminology, and estimands, alongside a Horvitz-Thompson-inspired estimator for select estimands.

Pouget-Abadie et al. (2019) propose a linear exposure assumption, which we incorporate in several of our examples and simulations. Their focus leans towards optimizing the variance of prevalent estimators rather than achieving unbiased causal effect estimators. Contrasting with Papadogeorgou et al. (2019), our paper primarily zeroes in on estimating the total average treatment effect. We establish theoretical underpinnings for an unbiased estimator and its variance estimators based on bootstrap, further elucidating these findings via simulations. Our methodologies are assessed on datasets furnished by the original authors.

Our work draws inspiration from Imbens (2000) and Hirano & Imbens (2004), who extend the propensity score literature to multivalued and continuous treatment scenarios. Our proposed unbiased estimator is essentially a refinement of their contributions tailored to the bipartite experiment context. Notable distinctions persist, especially regarding treatment assignments and their independence across units. Imai & Van Dyk (2004) present an alternative estimator for continuous treatment scenarios, which we explore in Section 4.

Del Prete et al. (2020) navigate a network setting, positing that a unit's outcome is influenced by its treatment status and that of its neighbors. They employ generalized propensity score concepts to derive estimates for both direct and spillover effects. A salient aspect of Del Prete et al. (2020) is its emphasis on observational settings, a perspective we accommodate alongside our primary focus on bipartite experiments.

In conclusion, recent advancements in bipartite experiments are indebted to the foundational frameworks provided by Aronow & Samii (2017) and Sävje (2021) . These works offer a holistic perspective on treatment effect estimation on graphs, addressing interference, identification, and exposure mapping misspecification. Our contribution hones in on the specific, yet broadly applicable, bipartite graph context, introducing novel theoretical insights for estimation and inference, and validating the efficacy of our proposed methods through simulations.

CHAPTER 2

Our Setting

2.1 Bipartite Network

We designate the units that receive either treatment or control as *diversion units* and those units with measurable outcomes of interest as *outcome units*. We posit that these units are distinct and together constitute a bipartite graph, comprising N outcome units and M diversion units. Each edge (i, j), connecting outcome unit $i \in 1, \ldots, N$ and diversion unit $j \in 1, \ldots, M$, is affiliated with a weight $W_{ij} \in \mathbb{R}$. This weight is both known and remains unaffected by the treatment assignment of any diversion unit.

$$W = \overset{\text{Sellers}}{\begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1M} \\ \vdots & & \ddots & \vdots \\ W_{N1} & W_{N2} & \cdots & W_{NM} \end{bmatrix}}_{N\times M} \text{Buyers}$$

Figure 2.1: Bipartite Network Matrix

The observed outcome of the outcome unit i is represented by Y_i, while the treatment assignment of the diversion unit j is symbolized by $Z_j \in 0, 1$, where $Z_j = 1$ indicates that the diversion unit j has been treated, and 0 denotes otherwise. A visual representation of this setup is provided in Figure 1.

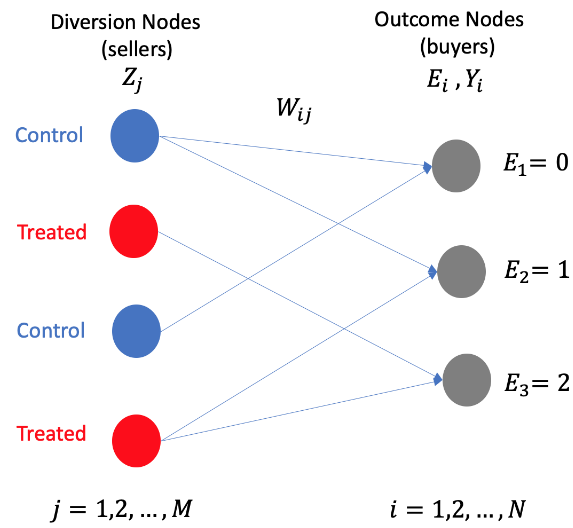

Figure 2.2: Bipartite Network Example (Sellers and Buyers)

2.2 Network Exposure

In Figure 2.2, we illustrate a simple bipartite network example. The treatment exposure E_i received by outcome unit i is a function of the bipartite graph and of the treatment assignment $\mathbf{Z} = \{Z_j\}_{j\in[1,M]}$. Because the bipartite graph is assumed to be constant—an assumption we will revisit in Section 3—we often write $E_i(\mathbf{Z})$ to represent the treatment exposure that outcome unit i has received under treatment assignment $\mathbf{Z} \in \{0,1\}^M$. The exact functional form of the treatment exposure is problem-dependent and should be determined by a domain expert. The assumption is that it is known, probabilistic, and captures all variations of potential outcomes:

$$\forall i \in [1,N], \forall \mathbf{Z} \in \{0,1\}^M, Y_i(\mathbf{Z}) = Y_i(E_i(\mathbf{Z}))$$

In the working examples of Papadogeorgou et al. (2019), the outcome of interest is influenced by a "direct effect," which is initiated by the treatment status of the nearest power plant (diversion unit) to the hospital (outcome unit). Additionally, there is an "indirect effect," which is related to the proportion of power plants upwind from the hospital that are treated. Pouget-Abadie et al. (2019) propose a slightly different functional form for the exposure, termed the *linear exposure assumption*. Under this assumption, the exposure of outcome unit i is a weighted proportion of its treated neighboring diversion units in the bipartite graph:

$$\forall i \in [1, N] \quad , \quad E_i(\mathbf{Z}) = \sum_{j=1}^{M} W_{ij} Z_j$$

While the results presented in our paper are largely agnostic to the precise functional form of the exposure, for simplicity of exposition, we often adopt the linear exposure assumption. Exposure $E = f_E(Z, W)$ is a function of the network $\{W_{ij}\}_{(N,M)}$ and sellers' treatment assignment Z.

We are interested in the marginal exposure-outcome function $\mu(e) = \mathbb{E}[Y(e)]$ at each strata $e \in \mathcal{E}$. In order to construct treatment effect estimands in a bipartite design setting, it is beneficial to consider the exposure-outcome curve. This curve maps each level of exposure to the mean of the potential outcome in the population for that specific level of exposure: $\mu : e \mapsto \mathbb{E}[Y_i(e)]$. If exposure is restricted to the interval [0, 1]—as is the case under the linear treatment exposure assumption when the graph weights are normalized appropriately—a primary estimand of interest is $\mu(1) - \mu(0)$. This represents the bipartite-design equivalent of the population average treatment effect. It measures the effect of treating all units compared to treating none and is the primary estimand discussed in Section 6. Another potential estimand of interest is the derivative of the exposure-outcome curve, which corresponds to the impact of a marginal change in exposure at a specific exposure level.

CHAPTER 3

Estimation Theorem

To produce accurate estimates based on the example from the previous section, it is essential to account for the varying exposure distributions among outcome units. In this section, we introduce estimators for causal effects. These estimators are inspired by the generalization of the propensity score in the literature on multivalued and continuous treatments Imbens (2000), Hirano & Imbens (2004), Imai & Van Dyk (2004), Robins et al. (2000) and Rosenbaum & Rubin (1983). Additionally, we draw from the literature on estimation under interference Aronow & Samii (2017) and Sävje (2021). We demonstrate their unbiasedness under a specific set of assumptions. We start by outlining the standard assumptions necessary for our results.

3.1 Unconfoundedness Assumptions

Assumption 0. Fixed Network Weights The graph weights $\{W_{ij}\}_{N,M}$ are not affected by the treatment assignment $\mathbf{Z}$. Formally, the vector of the treatment assignment $\mathbf{Z}$ is independent of all network weights:

$$\mathbf{Z} \perp \{W_{ij}\}_{N,M}$$

For example, in Papadogeorgou et al. (2019) , the bipartite graph is given by the fixed geographic distance between power plants and hospitals and is not be affected by the treatment. In the market setting of Pouget-Abadie et al. (2019), the bipartite graph is given by buyers' preferences for different item categories. It is, of course, in principle possible for items to become more or less desirable to a buyer as a function of treatment. Assumption 1 restricts our attention to the settings where the graph weights are not affected by the treatment.

In the given assumption, we assert the independence of network connections from treatment assignments, such as discounts or ranking for items. This premise emphasizes that a customer's purchasing intent or demand usually either precedes or remains independent of the formation of network connections. For illustration, a customer's decision to buy pet food typically originates from acquiring a pet, leading to the subsequent demand for pet food. When the customer engages with an online shopping platform and searches for 'pet food', bipartite network connections form during the comparison of similar items.

In some situations, especially when network connections emerge due to treatment assignment, this assumption may be compromised. A notable instance is during Amazon Prime Day when a plethora of products receive significant discounts. In these cases, the links between customers and items can be ascribed to the discount allocation on products, thereby questioning the integrity of the assumption. Although this Fixed Network Assumption might be violated in real-world data, we still think all of our theorems and algorithms are still applicable.

In numerous observational studies analyzing network data, it has been discerned that the response variable Y_i is intricately associated with the network W_{ij}. For instance: 1. Within a Seller-Buyer Network, when a prospective buyer, unfamiliar with a product, engages in the examination of multiple product pages, a notable enhancement in the confidence to finalize the purchase is observed. 2. In social network applications, establishing more friend connections is correlated with an inclination to spend increased time using the app, and a rise in the number of fans is proportional to the potential increase in the number of posts generated by the user.

Assumption 1. Strong Unconfoundedness The exposure E_i received by the outcome unit i is independent of all its potential outcomes given the graph weights $\mathbf{W} = (W_1, \ldots, W_N)^T$, where $W_i = (W_{i1}, \ldots, W_{iM})^T$.

$$E_i \perp \{Y_i(e)\}_{e \in [0,1]} | \mathbf{W}$$

Under strong unconfoundedness, the observed exposure of each outcome unit is independent of the potential outcomes of that unit, when conditioned on the bipartite graph weights. Assumption 2 is often compared with its slightly weaker version Imbens (2000).

Assumption 2. Weak Unconfoundedness For all $e \in \mathcal{E}$, this conditional independence holds for each strata of the exposure, given the pre-exposure network W.

$$Y(e) \perp E|W$$

In practice, although the slightly weaker Assumption 3 is adequate to establish most of our results, it might be challenging - as argued by Imbens (2000) - to identify examples where one assumption is valid while the other is not. It could be more straightforward to reason about the more intuitive Assumption 2 directly. In the context of the linear treatment exposure assumption presented in Pouget-Abadie et al. (2019), both Assumption 2 and Assumption 3 are confirmed for Bernoulli or Completely Randomized treatment assignments. Indeed, conditional on $\mathbf{W}$, the exposure E_i received by outcome unit i is a fixed weighted sum of random variables orthogonal to the potential outcomes of unit i.

3.2 The Generalized Propensity Score

We now introduce our proposed generalized-propensity-score-based estimator, inspired by the extension of the propensity score to multivalued and continuous treatment literature Hirano & Imbens (2004).

Definition 1. Generalized Propensity Score (GPS) Let $r(e, w)$ be the conditional density of the exposure given the network:

$$r(e, w) = f_{E|W}(e|w)$$

Then, the GPS is defined as $R = r(E, W)$. In the spirit of early results by Rosenbaum & Rubin (1983), under weak unconfoundedness, it is sufficient to condition on the generalized propensity score to achieve conditional independence of $\mathbb{I}\{E_i = e\}$ and $Y_i(e)$. Formally, this result is summarized in the following two lemmas.

Lemma 1. Balancing given the GPS For a specific stratum e, with the same value of $r(e, W)$, the probability that $E = e$ does not depend on the pre-exposure network W.

$$W \perp \mathbb{I}(E = e) \mid r(e, W) \text{ for all } e \in \mathcal{E}$$

Lemma 2. Weak Unconfoundedness Given the GPS Under Assumption 1 and Assumption 2, then:

$$\mathbb{I}(E = e) \perp Y(e) \mid r(e, W), \text{ for all } e \in \mathcal{E}$$

Lemma 2 primarily follows from Lemma 1 and is crucial for establishing the unbiasedness of our estimator. The lemma asserts that to ensure independence between the potential outcome corresponding to exposure e and the event of receiving that same exposure level, one only needs to condition on the propensity score at that exposure level. This approach obviates the need to condition on the entire vector $\boldsymbol{W}$, thereby avoiding situations where we observe few or no outcomes at a specific *conditioned* exposure level. We now introduce the primary theoretical result that facilitates unbiased estimation of the exposure-outcome function and its derived estimands.

Theorem 1. Modeling with the GPS (Estimation of Average Potential Outcomes through Adjustment for the Generalized Propensity Score) Under Assumption 1 and Assumption 2, we have the following:

1. Estimate the conditional expectation of the outcome as a function of exposure strata e and GPS R:

$$\beta(e, r) = \mathbb{E}[Y(e)|r(e, W) = r] = \mathbb{E}[Y(e)|E = e, R = r] \tag{3.2-1}$$

2. Estimate the exposure-outcome function $\mu(e)$ by averaging the above conditional expectation over the GPS at exposure strata e:

$$\mu(e) = \mathbb{E}[\beta(e, r(e, W))] \tag{3.2-2}$$

We posit that $\beta(e, r)$ lacks a causal interpretation as it merely describes the change between exposure and response. Some algorithms of Theorem 1 necessitate the adoption of a parametric assumption regarding the function $\beta(e, r)$. But we also provide semi-parametric algorithms for the function $\beta(e, r)$ in the empirical result section. The proofs of Lemmas 1 and 2, as well as Theorem 1, closely follow those in Imbens (2000) and are provided in the Appendix. We also present a result for a Horvitz-Thompson-style estimator, following the approach of Horvitz & Thompson (1952).

Theorem 2. Weighting by the GPS (Weighting and the Generalized Propensity Score) Under Assumption 1 and Assumption 2, we estimate the exposure-outcome by averaging the reweighted response at a particular stratum e:

$$\mathbb{E}\left[\frac{Y \cdot \mathbb{I}(e)}{r(E, W)}\right] = \mathbb{E}[Y(e)] \tag{3.2-3}$$

The function $r(E, W)$ needs to be 'stabilized' by the marginal probability of exposure $P_E(e)$. Theorem 2 does not need to specify the parametric function of response on exposure strata e and GPS r. It is believed that Theorem 2 also lacks a causal interpretation. Both Theorem 1 and Theorem 2 apply the information from GPS to estimate the exposure-outcome function $\mu(e)$ at particular strata e.

This theorem is proven in the Appendix. Note that under the linear exposure assumption in a simple randomized design, where each $Z_j = 1$ with probability $(0 < p_i < 1)$ independently of $\{Z_{j'}\}_{j' \neq j}$, the property $Pr(E_i = e) > 0$ for all i is satisfied for $e = 0$ and $e = 1$, but may not hold for any other value of e.

Both theorems—Modeling with the GPS and Weighting by the GPS—employ the Generalized Propensity Score (GPS) of the exposure model to mitigate the bias in estimation attributed to confounder variables, a process herein referred to as "deconfounding." More precisely, Theorem 1 incorporates the GPS as a component in modeling responses, while Theorem 2 applies the GPS to weight responses. Nonetheless, the performance of the algorithms associated with these theorems exhibits residual bias from the confounder variables. Consequently, we introduce Theorem 3, which synthesizes the deconfounding approaches from both Theorem 1 and 2. In Theorem 3, the GPS is utilized to weight the modeling results, aiming to further reduce the bias and enhance the robustness of the estimations.

Lemma 3. Weak Unconfoundedness of Response Model Given the GPS Under Assumption 1 and Assumption 2, let the assumption for $\beta(e, r)$ be defined as

$\mathbb{E}[Y(e)|r(e,W)=r]$. Then,

$$\mathbb{I}(E=e) \perp \beta(e,r)|r(e,W)$$

for all $e \in \mathcal{E}$.

Theorem 3. Weighting and Modeling with the GPS (Double Deconfounded) Under Assumption 1 and Assumption 2, we estimate the exposure-outcome by averaging the reweighted response at a particular stratum e as follows:

$$\mathbb{E}\left[\frac{\beta(e,r)\cdot\mathbb{I}(e)}{r(E,W)}\right] = \mathbb{E}[Y(e)] \tag{3.2-4}$$

Here, $\mathbb{E}$ denotes the expectation, $\beta(e,r)$ represents the parametric model of the response Y, $\mathbb{I}(e)$ is the indicator function for strata e, and $r(E,W)$ is the GPS for exposure E and covariate W. Owing to the incorporation of the Generalized Propensity Score (GPS) in both the weighting and modeling phases by Theorem 3, we designate this theorem as the Double Deconfound, abbreviated as DD.

CHAPTER 4

Estimation Algorithms

We proposed three theorems in the previous section. However, to apply these theorems, they need to be integrated with corresponding algorithms. We will illustrate the application of these algorithms through simulation examples. In the first simulation, we exhibit a seller-buyer bipartite network in a static environment. In the second simulation, we illustrate the modifications required to apply algorithms in a dynamic bipartite network environment.

4.1 Algorithms for Static Network

We have formulated three theorems to facilitate the unbiased estimation of the exposure-outcome effect. Nevertheless, the application of these theorems necessitates the development of specific algorithms. Consequently, we have devised multiple algorithms corresponding to each theorem. Table 4.1 provides a summary of the names, abbreviations, and types of all the proposed algorithms.

Table 4.1: Algorithm Names Table

Abbrviation	**Theorem**	**Algorithm Name**
baseline	No	Naïve Regression
model TSLS	Model with the GPS	Two Stage Least Square
model regress	Model with the GPS	Model Regression
model GAM	Model with the GPS	Generalized Additive Models
IPW regress	Weight with the GPS	Inverse Probability Weight
IPW Boost	Weight with the GPS	Gradient Boosted Regression Model
DD TSLS	Model and Weight with the GPS	Two Stage Least Square
DD GAM	Model and Weight with the GPS	Generalized Additive Models

To determine the marginal exposure-outcome effect, we denote it as τ. In our study, we treat the exposure E_i as a continuous variable, aiming to estimate this marginal effect. As a result, the model estimator parameter $\hat{\theta}_1$ corresponding to the exposure acts as the estimator for the marginal exposure-outcome effect. It's noteworthy that all algorithms utilize the estimated parameter $\hat{\theta}_1$ in relation to the exposure E_i.

4.1.1 Naïve Regression (Baseline)

Assume the response model on exposure is given by: $Y_i|E_i \sim \mathcal{N}(\theta_0 + \theta_1 E_i, \sigma_y^2)$. Estimate the average potential outcome at each exposure strata e as follows:

$$\hat{\mathbb{E}}[Y(e)] = \frac{1}{N_e} \sum_{i=1}^{N_e} Y_i \cdot \mathbb{I}(E_i = e)$$

where $N_e = \sum_{i=1}^{N} \mathbb{I}(E_i = e)$. In this baseline algorithm, we use the model estimator parameter $\hat{\theta}_1$ for the exposure serves as the estimator for this marginal exposure-outcome effect:

$$\hat{\tau} = \hat{\theta}_1 = (E'E)^{-1}E'Y = \frac{\sum_{i=1}^{N}(E_i - \bar{E})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(E_i - \bar{E})^2}$$

4.1.2 Two Stage Least Square (TSLS)

We model with the GPS as follows: Assume the exposure model is given by:

$$E_i = \alpha_0 + \alpha_1 |W_i| + \epsilon_i$$

Since the exposure E_i is not binary, so we apply the Gaussian regression model, instead of the logistic regression model. Another potential Generalized Linear Model (GLM) is the multinomial regression model. The residuals of the exposure model are below.

$$\hat{\gamma}_i = E_i - \hat{\alpha}_0 - \hat{\alpha}_1 |W_i|$$

We then estimate the residuals and GPS using the result of the exposure model, represented by

$$\hat{R}_i(\hat{\gamma}_i) = \phi(\hat{\gamma}_i) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left[-\frac{1}{2\hat{\sigma}^2}\hat{\gamma}_i^2\right]$$

where ϕ signifies the probability density function of the normal distribution. The TSLS model is represented as follows:

$$\beta_\theta(E_i, R_i(\gamma_i)) = \theta_0 + \theta_1 E_i + \theta_2(E_i - \gamma_i) \tag{4.1-1}$$

Substitute $\hat{\gamma}_i$ into $\beta_\theta(E_i, R_i(\gamma_i))$, the expected value of $Y(e)$ is estimated as:

$$\hat{\mathbb{E}}[Y(e)] = \frac{1}{N_e} \sum_{i=2}^{N_e} \left[\hat{\theta}_0 + \hat{\theta}_1 E_i + \hat{\theta}_2(E_i - \hat{\gamma}_i)\right] \cdot I(E_i = e) \tag{4.1-2}$$

Then we estimate the marginal exposure-outcome effect as: $\hat{\tau} = \hat{\theta}_1$. This algorithm estimates the true marginal model of the simulated data.

4.1.3 Model Regression

Continue with the exposure model and residuals, we calculate the GPS in the function $\hat{R}_i(\hat{\gamma}_i) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left[-\frac{1}{2\hat{\sigma}^2}\hat{\gamma}_i^2\right]$ is used to estimate γ_i. We assume a parametric function for response on exposure E_i and estimated GPS $\hat{R}_i(\hat{\gamma}_i)$, represented as $\mathbb{E}[Y_i|E_i, R_i(\gamma_i)] = \beta_\theta(E_i, R_i(\gamma_i))$. After estimating the parameters above as $\hat{\beta}_\theta$, we then estimate the average potential outcome at each exposure strata e as:

$$\hat{\mathbb{E}}[Y(e)] = \frac{1}{N_e}\sum_{i=2}^{N_e}[\hat{\beta}_\theta(E_i, R_i(\gamma_i))] \cdot \mathbb{I}(E_i = e) \tag{4.1-3}$$

The regression model is represented as follows:

$$\beta_\theta(E_i, R_i(\gamma_i)) = \theta_0 + \theta_1 E_i + \theta_2 R_i + \theta_3 E_i^2 + \theta_4 R_i^2 + \theta_5 E_i R_i \tag{4.1-4}$$

The estimator $\hat{\theta_1}$ quantifies the marginal effect between exposure and outcome. The efficacy of this algorithm hinges on the parametric assumption of $\beta_\theta(E_i, R_i(\gamma_i))$. While alternative model assumptions can be compatible with this algorithm, it's imperative to note that relying solely on model-based algorithms for estimating the marginal exposure-outcome effect could introduce the potential for model misspecification. We will illustrate this issue in the Empirical Results section.

4.1.4 Generalized Additive Models (GAM)

Same with the Model Regression algorithm, this GAM algorithm also needs to calculate the GPS first, then apply the semi-parametric model:

$$\beta_\theta(E_i, R_i(\gamma_i)) = \theta_1 E_i + s(R_i) + s(E_i^2) + s(R_i^2) + s(E_i R_i) \tag{4.1-5}$$

Where this $s(R_i)$ adds a smoothed function of the predictor R_i, and this function $s(\cdot)$ specifies a spline smooth. Then we estimate the marginal exposure-outcome effect as: $\hat{\tau} = \hat{\theta}_1$. This algorithm assumes a semi-parametric model. It only assumes one parameter, θ_1, for exposure E_i. The rest terms are non-parametric splines. However, if applicators are only interested in the average exposure-outcome function $\mu(e) = \mathbb{E}[Y(e)]$ at a given exposure strata e, then they can also set the non-parametric $s(E_i)$ for the exposure.

4.1.5 Inverse Probability Weight (IPW)

This algorithm follows Theorem 2 (Weighting with the GPS). The algorithm first assumes under the exposure model as follows: $E_i = \alpha_0 + \alpha_1|W_i| + \epsilon_i$. The GPS is then estimated with the Gaussian linear model, represented by $\hat{\gamma}_i = E_i - \hat{\alpha}_0 - \hat{\alpha}_1|W_i|$. The resulting R_i is calculated as $\hat{R}_i = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left[-\frac{1}{2\hat{\sigma}^2}\hat{\gamma}_i^2\right]$. Subsequently, the weights are stabilized as $R'_i = \frac{\hat{P}_E(E_i)}{\hat{R}_i}$, where $\hat{P}_E(E_i)$ is the estimated marginal probability of exposure. Finally, the average potential outcome at exposure strata e is estimated as:

$$\mathbb{E}[Y(e)] = \frac{1}{N_e} \sum_{i=1}^{N_e} \frac{Y_i \cdot \mathbb{I}(E_i = e)}{R'_i} \tag{4.1-6}$$

To estimate the marginal exposure-outcome effect, we apply the Marginal Structure Model (MSM) from Robins et al. (2000) follow Theorem 2. The MSM will fit a weighted linear regression model for outcomes (Y_i) on the exposure (E_i) using the stabilized GPS (R'_i) as weights. The Objective function of the MSM to minimize is below.

$$\sum_{i=1}^{N} R'_i(Y_i - \theta_0 - \theta_1 E_i)^2 \tag{4.1-7}$$

Lastly, we estimate the marginal exposure-outcome effect as: $\hat{\tau} = \hat{\theta}_1$.

4.1.6 IPW with Gradient Boosted Regression Tree Model (IPW Boost)

Instead of fitting a Gaussian linear model for the exposure, we modify the exposure model to a gradient-boosted regression tree model for exposure:

$$E_i = \sum_{tr=1}^{Tr} \delta_{tr} \times f_{tr}(W_{i.}) \tag{4.1-8}$$

where Tr represents the number of trees or boosting rounds. $f_{tr}(\cdot)$ denotes the base tree model's prediction for network input $W_{i.}$. δ_{tr} is the weight of the tr-th base tree model $f_{tr}(\cdot)$.

In the context of this study, the Generalized Propensity Score (GPS) for the specified algorithm is denoted by $\hat{\gamma}_i = E_i - \hat{E}_i$, where $\hat{E}_i$ represents the predicted value from the exposure model. From this, we compute the resultant R_i as: $\hat{R}_i = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left[-\frac{1}{2\hat{\sigma}^2}\hat{\gamma}_i^2\right]$. Subsequent to this computation, the weights undergo stabilization, given by $R'_i = \frac{\hat{P}_E(E_i)}{\hat{R}_i}$, where $\hat{P}_E(E_i)$ signifies the estimated marginal likelihood of exposure. The ensuing step involves estimating the mean potential outcome for exposure stratum e, expressed as:

$$\mathbb{E}[Y(e)] = \frac{1}{N_e} \sum_{i=1}^{N_e} \frac{Y_i \cdot \mathbb{I}(E_i = e)}{R'_i} \tag{4.1-9}$$

To ascertain the effect of marginal exposure-outcome, we employ the Marginal Structure Model (MSM), in accordance with Theorem 2. This model facilitates the fitting of a weighted linear regression model, with outcomes (Y_i) predicated on the exposure (E_i), utilizing the stabilized GPS (R'_i) as the weighting factor. The objective function of the MSM, which is subject to minimization, is presented as:

$$\sum_{i=1}^{N} R'_i(Y_i - \theta_0 - \theta_1 E_i)^2 \tag{4.1-10}$$

Concluding our analysis, the marginal exposure-outcome effect is estimated by $\hat{\tau} = \hat{\theta}_1$.

4.1.7 IPW Boost and TSLS (DD TSLS)

In the context of this algorithm, we consider the Generalized Propensity Score (GPS) denoted as R_i and its stabilized counterpart, R'_i, as derived in the IPW Boost algorithm. Our proposed outcome model, contingent upon the Exposure and residuals

$(\hat{\gamma}_i)$), is articulated as:

$$\mathbb{E}[Y_i|E_i, R_i(\gamma_i)] = \beta_\theta(E_i, R_i(\gamma_i)) = \theta_0 + \theta_1 E_i + \theta_2(E_i - \hat{\gamma}_i)) \tag{4.1-11}$$

Subsequently, we proceed to estimate the model parameters, represented by $\hat{\theta}$. With these estimates in place, we then compute the average potential outcome for specific exposure strata e as:

$$\mathbb{E}[Y(e)] = \frac{1}{N_e}\sum_{i=2}^{N_e}\left(\frac{\hat{\theta}_0 + \hat{\theta}_1 E_i + \hat{\theta}_2(E_i - \hat{\gamma}_i)}{R'_i}\right) \cdot \mathbb{I}(E_i = e) \tag{4.1-12}$$

Concluding our analysis, the marginal effect of exposure on the outcome is estimated to be: $\hat{\tau} = \hat{\theta}_1$.

4.1.8 IPW Boost and GAM (DD GAM)

In line with the methodology of the DD TSLS algorithm, we introduce the Generalized Propensity Score (GPS), symbolized as R_i, and its stabilized version, R'_i, both of which are derived from the IPW Boost algorithm. We then employ the semi-parametric Generalized Additive Model (GAM) to represent the response model. This model can be articulated as:

$$\beta_\theta(E_i, R_i(\gamma_i)) = \theta_1 E_i + s(R_i) + s(E_i^2) + s(R_i^2) + s(E_i R_i) \tag{4.1-13}$$

Here, $s(R_i)$ embodies a smoothed function of the predictor R_i. The notation $s(\cdot)$ signifies a spline smooth, which delineates the smoothness of the predictor variables within the model framework. Following this, we undertake the estimation of the aforementioned model parameters. Upon obtaining these estimates, we compute the average potential outcome corresponding to specific exposure strata e:

$$\mathbb{E}[Y(e)] = \frac{1}{N_e}\sum_{i=1}^{N_e}\left(\frac{\theta_1 E_i + s(R_i) + s(E_i^2) + s(R_i^2) + s(E_i R_i)}{R'_i}\right)\cdot \mathbb{I}(E_i = e) \quad (4.1\text{-}14)$$

Conclusively, the marginal exposure-outcome effect is estimated as: $\hat{\tau} = \hat{\theta}_1$.

4.2 Algorithms for Dynamic Networks

To estimate the exposure-outcome effect under the dynamic networks, we modify our theorems and algorithms to adjust the time effect (S-shape learning curve). In these modified algorithms, we don't assume specific parametric models like Eqn 5.2-4. The main idea of this adjustment is to introduce dynamic intercepts for each factor time period. Therefore, the dynamic intercepts will clean the time effect from the exposure-outcome estimation. Note TSLS related algorithms are unchanged in dynamic networks. Because TSLS-related algorithms are robusted to the mis-specify or missing confounders.

4.2.1 Naïve Regression (Baseline)

In the Naïve Regression, the conditional distribution of Y_{it} given E_{it} is assumed to follow a normal distribution, denoted as:

$$Y_{it}|E_{it} \sim \text{Norm}(\theta_0 + \theta_1 E_{it} + \sum_{t\in T}\xi_t FT_{it} \;\;,\;\; \sigma_y^2) \quad (4.2\text{-}15)$$

A modification is incorporated to introduce dynamic intercepts for each factor time period, represented by $\sum_{t\in T}\xi_t FT_{it}$. By employing time factors, we abstain from assuming a specific parametric form for the learning curve, allowing for a more flexible and accurate representation of the underlying relationships.

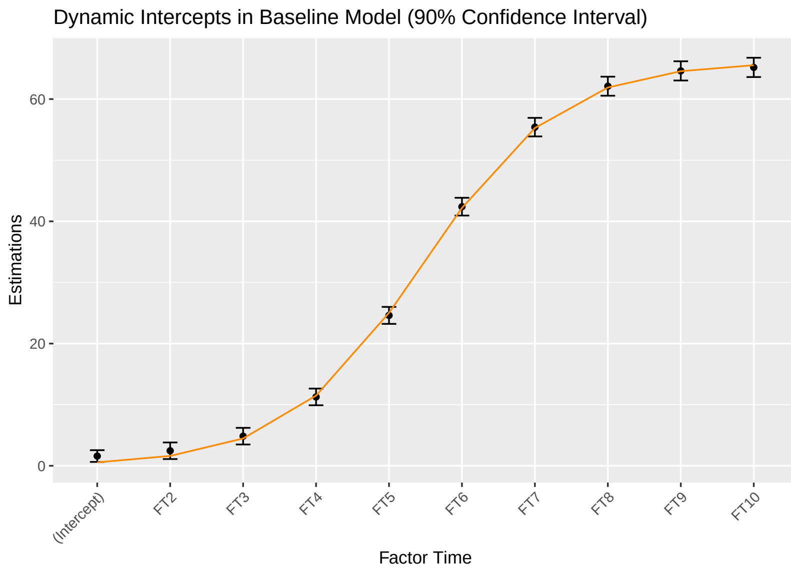

Figure 4.1: Dynamic Networks Simulation: Dynamic Intercepts of Baseline Algorithm

In the analysis of dynamic intercepts, represented by $\sum_{t \in T} \xi_t FT_{it}$, we observe that the orange curve corresponds to a fitted S-shape curve. This curve is mathematically described by the equation: $\kappa/[1 + e^{-\lambda(\tau - \tau_0)}]$, where κ denotes the carrying capacity, λ signifies the growth rate, and τ_0 is the time of median growth.

4.2.2 Two Stage Least Square (TSLS)

In the context of Dynamic Networks, the TSLS algorithm remains consistent with its application in Static Networks. The algorithm leverages the residuals from the exposure model as an input covariate, aiming to mitigate the bias associated with the exposure-outcome effect estimator. Given that the dynamic temporal information is inherently encapsulated within the residuals of the exposure model, there's no necessity to introduce an additional dynamic intercept term, such as $\sum_{t \in T} \xi_t FT_{it}$, to account for the temporal influence.

4.2.3 Model Regression

In alignment with the algorithm used for static networks, we employ the GPS in the following manner: Let's consider the exposure model as $E_i = \alpha_0 + \alpha_1|W_i| + \epsilon_i$. Given that the exposure E_i is continuous, we opt for the Gaussian regression model over the logistic regression model. An alternative Generalized Linear Model (GLM) could be the multinomial regression model. Subsequently, we compute the residuals and GPS based on the exposure model, denoted by $\hat{\gamma}_i = E_i - \hat{\alpha}_0 - \hat{\alpha}1|W_i|$ and $Ri(\hat{\gamma}i) = \phi(\hat{\gamma}i) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left[-\frac{1}{2\hat{\sigma}^2}\hat{\gamma}i^2\right]$, where ϕ signifies the probability density function of the normal distribution. The outcome regression model is modified as:

$$\beta_\theta(E_i, R_i(\gamma_i)) = \theta_0 + \theta_1 E_i + \theta_2 R_i + \theta_2 E_i^2 + \theta_4 R_i^2 + \theta_5 E_i R_i + \sum_{t \in T} \xi_t FT_{it} \qquad (4.2\text{-}16)$$

with $\sum_{t \in T} \xi_t FT_{it}$ representing the dynamic intercept terms.

We then estimate the average potential outcome for each exposure stratum e, expressed as

$$\mathbb{E}[Y(e)] = \frac{1}{N_e} \sum_{i=1}^{N_e} [\theta_0 + \theta_1 E_i + \theta_2 R_i + \theta_2 E_i^2 + \theta_4 R_i^2 + \theta_5 E_i R_i + \sum_{t \in T} \xi_t FT_{it}] \cdot \mathbb{I}(E_i = e) \qquad (4.2\text{-}17)$$

Lastly, we estimate the marginal exposure-outcome effect as: $\hat{\tau} = \hat{\theta}_1$.

4.2.4 Generalized Additive Models (GAM)

In accordance with the methodology applied to static networks, we utilize the Generalized Propensity Score (GPS). Specifically, the exposure model is defined as: $E_i = \alpha_0 + \alpha_1|W_i| + \epsilon_i$. From this, we derive the residuals and the GPS based on the exposure model, represented as: $\hat{\gamma}_i = E_i - \hat{\alpha}_0 - \hat{\alpha}_1|W_i|$ and $R_i(\hat{\gamma}_i) = \phi(\hat{\gamma}_i) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left[-\frac{\hat{\gamma}_i^2}{2\hat{\sigma}^2}\right]$. Subsequently, we reformulate the outcome model into a semi-parametric form:

$$\beta_\theta(E_i, R_i(\gamma_i)) = \theta_1 E_i + s(R_i) + s(E_i^2) + s(R_i^2) + s(E_i R_i) + \sum_{t \in T} \xi_t FT_{it} \tag{4.2-18}$$

The average potential outcome for each exposure stratum e is then given by:

$$\mathbb{E}[Y(e)] = \frac{1}{N_e} \sum_{i=1}^{N_e} \left[\theta_1 E_i + s(R_i) + s(E_i^2) + s(R_i^2) + s(E_i R_i) + \sum_{t \in T} \xi_t FT_{it} \right] \cdot \mathbb{I}(E_i = e) \tag{4.2-19}$$

Finally, the marginal exposure-outcome effect is estimated as: $\hat{\tau} = \hat{\theta}_1$.

In the dynamic network setting, we add the dynamic intercepts represented by $\sum_{t \in T} \xi_t FT_{it}$ for the outcome model. Therefore, all algorithms for Theorem 2 (Weighting by the GPS) are modified to Theorem 3 (Weighting and Modeling with the GPS) to include the dynamic intercepts.

4.2.5 Inverse Probability Weight (IPW)

In our analysis, we employ the Generalized Propensity Score (GPS). The exposure model is articulated as: $E_i = \alpha_0 + \alpha_1 |W_i| + \epsilon_i$. From this model, we extract the residuals and compute the GPS as: $\hat{\gamma}_i = E_i - \hat{\alpha}_0 - \hat{\alpha}_1 |W_i|$ and $R_i(\hat{\gamma}_i) = \phi(\hat{\gamma}_i) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left[-\frac{\hat{\gamma}_i^2}{2\hat{\sigma}^2}\right]$. The stabilization yields $R'_i = \frac{\hat{P}_E(E_i)}{\hat{R}_i}$.

For each time range in the factor set $FT_t \in (t \in T)$, we estimate the dynamic intercept terms, ξ_t. Under the assumption of a parametric function for the response Y_{it} relative to the Factor Time, FT_{it}, the conditional expectation $E[Y_{it}|FT_{it}]$ is defined as:

$$\beta_\xi(FT_{it}) = \sum_{t \in T} \xi_t FT_{it} \tag{4.2-20}$$

Our focus is on estimating the mean response, ξ_t, for each FT_{it} instance. Upon parameter estimation, denoted as $\hat{\xi}_t$, we calculate the average potential outcome for exposure strata e as:

$$\mathbb{E}[Y(e)] = \frac{1}{N_e} \sum_{i=1}^{N_e} \frac{\beta_\xi(FT_{it}) \cdot I(E_{it} = e)}{R'_{it}} \tag{4.2-21}$$

To determine the marginal exposure-outcome effect, we apply the Marginal Structure Model (MSM) in line with Theorem 2. This model is adept at fitting a weighted linear regression, where outcomes (Y_i) are based on the exposure (E_i), using the stabilized GPS (${R'}_i$) for weighting. The MSM's objective function, which is minimized, is given by:

$$\sum_{i=1}^{N} {R'}_i (Y_i - \sum_{t \in T} \xi_t FT_{it} - \theta_0 - \theta_1 E_i)^2 \tag{4.2-22}$$

To conclude, the estimated marginal exposure-outcome effect is $\hat{\tau} = \hat{\theta}_1$.

4.2.6 IPW with Gradient Boosted Regression Tree Model (IPW Boost)

In the context of the Inverse Probability Weighting (IPW) methodology, our approach introduces a modification. Specifically, we replace the Gaussian linear model traditionally used for exposure with a Gradient Boosted Regression Tree Model. The exposure, E_i, is then defined as:

$$E_i = \sum_{tr=1}^{Tr} \delta_{tr} \times f_{tr}(W_{i.}) \tag{4.2-23}$$

Residuals are derived from the fitted exposure values, denoted as $\hat{E}_i$. We introduce the Generalized Propensity Score (GPS) and articulate the exposure model as: $R_i(\hat{\gamma}_i) = \phi(\hat{\gamma}_i) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left[-\frac{\hat{\gamma}_i^2}{2\hat{\sigma}^2}\right]$. Stabilization of the model results in $R'_i = \frac{\hat{P}_E(E_i)}{\hat{R}_i}$. The subsequent potential outcome model, $\beta_\xi(FT_{it}) = \sum_{t \in T} \xi_t FT_{it}$, and the computation for the average potential outcome for exposure strata, $\mathbb{E}[Y(e)]$, remain consistent with the traditional IPW approach. For the estimation of the marginal exposure-outcome effect, we employ the Marginal Structural Model (MSM) and determine that $\hat{\tau} = \hat{\theta}_1$.

4.2.7 IPW Boost and TSLS (DD TSLS)

Given that the residuals from the exposure model encapsulate dynamic temporal information, the DD TSLS algorithm for dynamic networks remains consistent with that for static networks. It is unnecessary to incorporate dynamic intercepts, denoted as $\sum_{t \in T} \xi_t FT_{it}$, in the outcome model to account for temporal effects.

We employ a Gradient Boosted Regression Tree Model for the exposure model. The exposure, E_i, is thus expressed as: $E_i = \sum_{tr=1}^{Tr} \delta_{tr} \times f_{tr}(W_{i.})$. The residuals are computed from the fitted exposure values and are represented as $\hat{E}_i$. We define the Generalized Propensity Score (GPS) and formulate the exposure model as: $R_i(\hat{\gamma}_i) = \phi(\hat{\gamma}_i) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left[-\frac{\hat{\gamma}_i^2}{2\hat{\sigma}^2}\right]$. The stabilization yields $R'_i = \frac{\hat{P}_E(E_i)}{\hat{R}_i}$.

The proposed outcome model, contingent upon the Exposure and residuals $(\hat{\gamma}_i)$, is given by:

$$\mathbb{E}[Y_i|E_i, R_i(\gamma_i)] = \beta_\theta(E_i, R_i(\gamma_i)) = \theta_0 + \theta_1 E_i + \theta_2(E_i - \hat{\gamma}_i) \tag{4.2-24}$$

Following this, we estimate the model parameters, denoted by $\hat{\theta}$. Using these estimates, the average potential outcome for a specific exposure stratum e is:

$$\mathbb{E}[Y(e)] = \frac{1}{N_e} \sum_{i=2}^{N_e} \left(\frac{\hat{\theta}_0 + \hat{\theta}_1 E_i + \hat{\theta}_2(E_i - \hat{\gamma}_i)}{R'_i} \right) \cdot \mathbb{I}(E_i = e) \tag{4.2-25}$$

To conclude, the estimated marginal effect of exposure on the outcome is: $\hat{\tau} = \hat{\theta}_1$.

4.2.8 IPW Boost and GAM (DD GAM)

In a manner consistent with the DD TSLS algorithm, we integrate the semi-parametric Generalized Additive Model (GAM) for the outcome model in the concluding phase.

For the exposure model, we utilize the Gradient Boosted Regression Tree Model. The exposure, denoted by E_i, is articulated as: $E_i = \sum_{tr=1}^{Tr} \delta_{tr} \times f_{tr}(W_{i.})$. From the fitted exposure values, we compute the residuals, denoted as $\hat{E}_i$. We introduce the Generalized Propensity Score (GPS) and express the exposure model as: $R_i(\hat{\gamma}_i) = \phi(\hat{\gamma}_i) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left[-\frac{\hat{\gamma}_i^2}{2\hat{\sigma}^2}\right]$. Stabilization results in $R'_i = \frac{\hat{P}_E(E_i)}{\hat{R}_i}$. We then put forth the outcome model as:

$$\beta_\theta(E_i, R_i(\gamma_i)) = \theta_1 E_i + s(R_i) + s(E_i^2) + s(R_i^2) + s(E_i R_i) + \sum_{t \in T} \xi_t FT_{it} \tag{4.2-26}$$

After this, we derive the model parameters, represented by $\hat{\theta}$. Employing these estimates, the average potential outcome for a designated exposure stratum e is articulated as:

$$\mathbb{E}[Y(e)] = \frac{1}{N_e} \sum_{i=1}^{N_e} \left(\frac{\theta_1 E_i + s(R_i) + s(E_i^2) + s(R_i^2) + s(E_i R_i) + \sum_{t \in T} \xi_t FT_{it}}{R'_i} \right) \cdot \mathbb{I}(E_i = e) \tag{4.2-27}$$

In summation, the inferred marginal influence of exposure on the outcome is denoted as: $\hat{\tau} = \hat{\theta}_1$. The utilization of dynamic networks is pivotal in understanding customer behavior, particularly in scenarios involving the dissemination of new product information, advertisements, or discounts. This approach acknowledges the temporal aspect of customer education, recognizing that there is an inherent delay between the introduction of new information and its assimilation by customers. As customers gradually learn and assimilate this new information, they develop new demands, which in turn lead to the formation of new network connections. This dynamic interplay between information dissemination, customer learning, and network formation underscores the importance of dynamic networks in capturing the evolving nature of customer interactions and preferences.

CHAPTER 5

Simulation Results

5.1 Buyer-Seller Simulation: Static Bipartite Network

In our study, we construct a Data Generating Process to analyze the interaction between buyers and sellers. We assume there are $N = 1000$ buyers and $M = 50$ sellers or products in the market. The randomized Bernoulli design is represented as $Z_j \sim$ IID Ber$(p = 0.5)$, where each seller or product is independently and identically distributed.

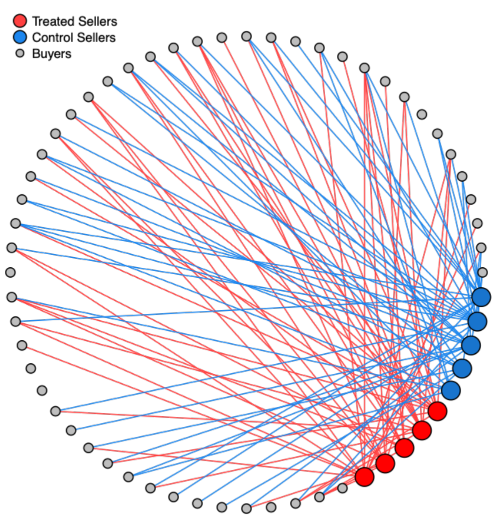

Figure 5.1: Static Network Simulation: Static Bipartite Network Example

The demand or intention to connect for each buyer is denoted by $S_i \sim$ IID Pois$(\lambda = 10)$, indicating that the buyers' intention to connect follows an independent and identical Poisson distribution with a mean of 10.

To generate the bipartite network, we consider the degree of buyer i, represented by $|W_{i\cdot}| \sim$ IID Pois(S_i), which implies the number of connections of buyer i is independently and identically distributed following a Poisson distribution with a mean of S_i. The connection index is given by $\text{Index}_i = \text{sample}(\{1, 2, \ldots, M\}, |W_{i\cdot}|)$, and the network of buyer i is represented as $W_{i\cdot}[\text{Index}_i] = 1$, with the rest set as 0. This process is repeated for $i = 1, 2, \ldots, N$.

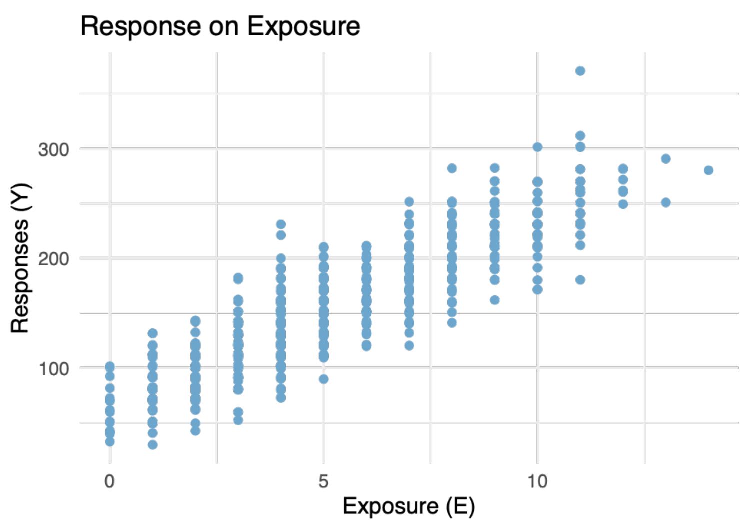

Figure 5.2: Static Network Simulation: Exposure-Outcome Scatter Plot

The exposure for each buyer is calculated as $E_i = W_{(i.)}Z = \sum_{j \in W_{(i.)}} W_{ij}Z_j$. It is important to note that we do not observe the intention or demand (S_i), but we estimate the intention by the network information (degree) $|W_{i\cdot}|$.

To generate outcomes, we assume the true Exposure Effect: $\tau = 10$, and the true Intention Effect: $\psi = 10$. The potential outcome function of $Y(e)$ given S can be represented as:

$$Y(e)|S = 1 + \tau e + \psi S + \epsilon,$$

where $\epsilon \sim \text{Norm}(0, 1)$, indicating that ϵ follows a normal distribution with mean 0 and standard deviation 1.

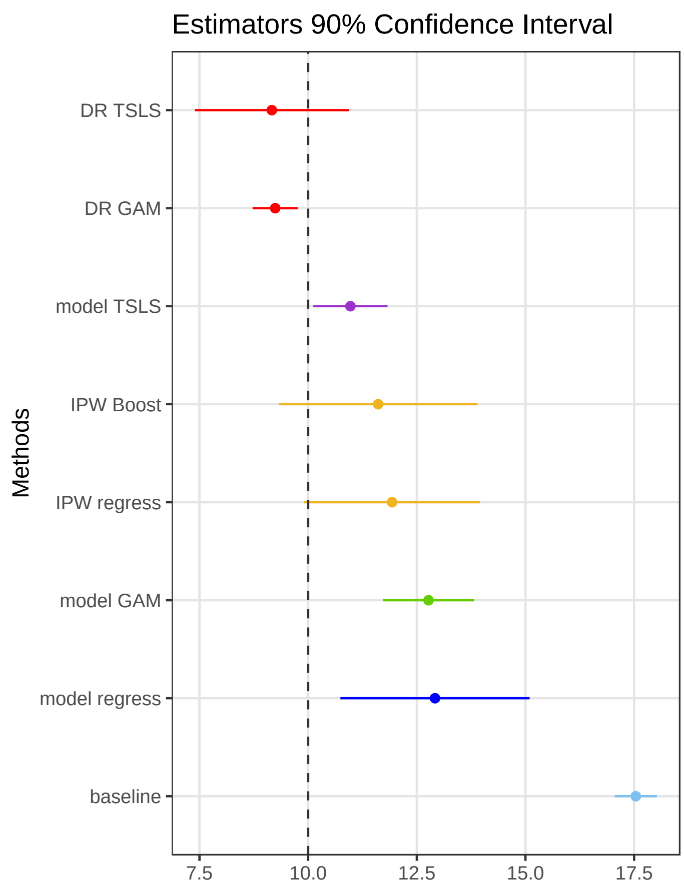

Figure 5.3: Static Network Simulation: Exposure-Outcome Effect 90% Confidence Intervals

Based on the above data-generating process, the true marginal exposure-outcome function ($\mu(e)$) is calculated by integrate out the variation of buyer demand (S_i). The true marginal exposure-response function ($\mu(e)$) is the expected value of $Y(e)$ only one the exposure, and can be represented as:

$$\mu(e) = \mathbb{E}[Y(e)] = 101 + \tau e \tag{5.1-1}$$

5.1.1 Large Scale Simulation and ANCOVA Analysis

In this batch simulation, the relationship between hyper-parameters and the mean squared error of the marginal exposure-outcome effect $MSE(\hat{\tau})$, represented as the sum of bias squared and variance, was examined.

$$MSE(\hat{\tau}) = Bias(\hat{\tau})^2 + Var(\hat{\tau})$$

where $Bias(\hat{\tau}) = \hat{\tau} - \tau$, τ is the known true parameter of the simulation, $Var(\hat{\tau})$ is the variance of the $\hat{\theta_1}$ of each algorithm.

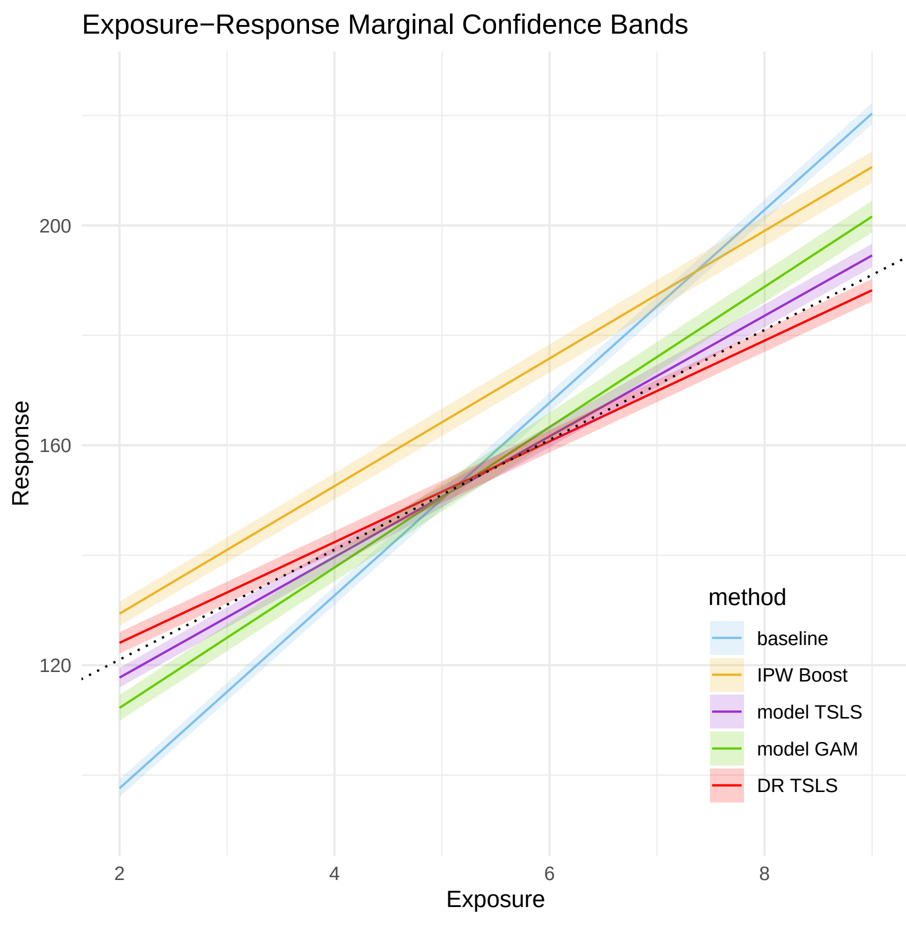

Figure 5.4: Static Network Simulation: Marginal Exposure-Outcome with 90% Confidence Bands

For the hyper-parameters of the batch simulation, this study considered the number of buyers, N, and the number of sellers or products, M. The demand or intention to connect for each buyer, S_i, was independently and identically distributed (IID) following a Poisson distribution with parameter λ.

The randomized Bernoulli design for each seller, Z_j, was also IID with a Bernoulli distribution parameterized by p. Various combinations of hyper-parameters were simulated, with M values of 50, 75, and 100; N values of 500, 1000, and 1500; λ values of 7, 10, and 13; and p values of 0.3, 0.5, and 0.7. Each combination was iterated 100 times.

Table 5.1: Static Network Batch Simulation: $MSE(\hat{\tau})$ Averages Sample

λ	p	$M \times N$	baseline	IPW.regress	IPW.Boost	model.TSLS	model.regress	model.GAM	DD.TSLS	DD.GAM
7	0.3	75x1000	67	10.4	6.1	0.5	80.2	6.3	3.2	1.1
10	0.3	75x1000	67.4	13.2	5.8	0.6	37.8	3.6	2.6	1.1
13	0.3	75x1000	74.2	11.3	7.7	0.6	30.8	3.5	3.2	1.5
7	0.5	75x1000	49.7	16.3	9.7	0.5	5	3	5.8	1.3
10	0.5	75x1000	51.2	17.8	11.5	0.5	5.6	3.9	4.3	1.5
13	0.5	75x1000	51.1	14.7	11.9	0.5	6.9	4.9	3.9	1.7
7	0.7	75x1000	39.2	32.2	19.1	0.5	10.9	7.3	14.5	2.7
10	0.7	75x1000	37.5	22.7	16.1	0.7	13.3	9.1	7.6	2.2
13	0.7	75x1000	38.1	20.2	15.2	0.6	16.1	10.4	6.8	2.6

Significant sensitivity ($P - value < 0.1$): the $MSE(\hat{\tau})$ might be influenced by simulation hyper-parameters under these algorithms.

Table 5.2: Static Network Batch Simulation: P-values from ANCOVA Analysis Comparing $MSE(\hat{\tau})$ Across Simulation Hyper-parameters for Different Algorithms

Algorithms	M	N	λ	p
baseline	0	0.096	0	0
IPW regress	0.058	0	0.622	0
IPW Boost	0	0	0.019	0
model TSLS	0	0	0	0.15
model regress	0.04	0	0	0
model GAM	0	0	0.002	0
DD TSLS	0.523	0	0	0
DD GAM	0.572	0	0.634	0

The exposure of a buyer is defined as $E_i = W_{i.}Z = \sum_{j \in W_{i.}} W_{ij} Z_j$. It is observed that an augmentation in the variability of the network, W_{ij}, coupled with the variability in the treatment assignment probability, $Z_j \sim IID\, Ber(p)$, amplifies the variability in exposure. This escalation in variability subsequently culminates in a diminution of the $MSE(\hat{\tau})$ in the estimators. Notably, certain algorithms exhibit insensitivity to a limited set of hyper-parameters.

5.2 Buyer-Seller Simulation: Dynamic Bipartite Networks

We initialize the model with a set of parameters where the number of buyers, N, is 1000, and the number of sellers or products, M, is 50. The exposure effect, τ, and the intention effect, ψ, are both set to 10. It is postulated that both demand and connection increase in an S-shaped curve over time, T. This is reflective of scenarios where customers, upon learning about new products, generate new demands, exemplified by the release of the iPhone 1 on June 29, 2007. Additionally, the time-dependent nature of the model accounts for the duration required to educate customers using new advertisements or discount information.

The study explores dynamic network growth through the lens of a logistic growth function, represented as:

$$\eta(t) = \frac{K}{1 + \exp(-b(t - t_0))} \tag{5.2-2}$$

where $K = 200$ denotes the carrying capacity, $b = 0.1$ signifies the growth rate, and $t_0 = 5$ is the time of median growth. This function is pivotal for illustrating the mean demand. To elucidate the dynamic alterations within the network, a bipartite network is drawn, incorporating 50 buyers and 10 sellers.

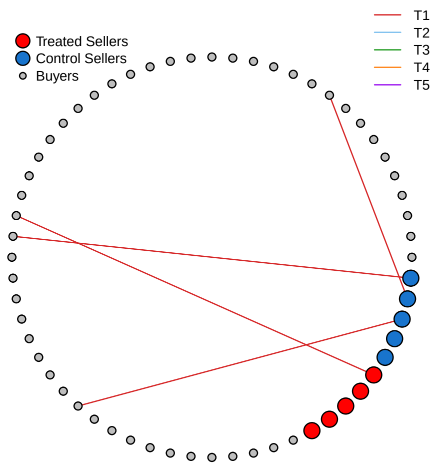

Figure 5.5: Dynamic Network at Time 1

The bipartite network is analyzed across five distinct periods, each representing different stages of network evolution. 1. $T1$ the Lag Phase: Represents the initial stage of network evolution where adoption is minimal. 2. $T2$ the Acceleration Phase: Characterizes the stage where the adoption rate begins to increase. 3. $T3$ the Exponential Adoption Phase: Denotes the stage where the network experiences rapid growth in adoption. 4. $T4$ the Deceleration Phase: Represents the stage where the rate of adoption starts to slow down. 5. $T5$ the Saturation Phase: Signifies the final stage where the network reaches its maximum adoption capacity.

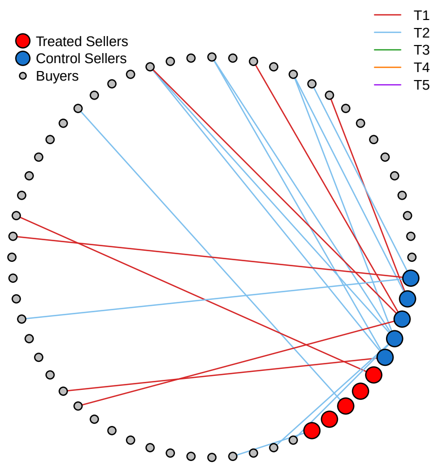

Figure 5.6: Dynamic Network at Time 2

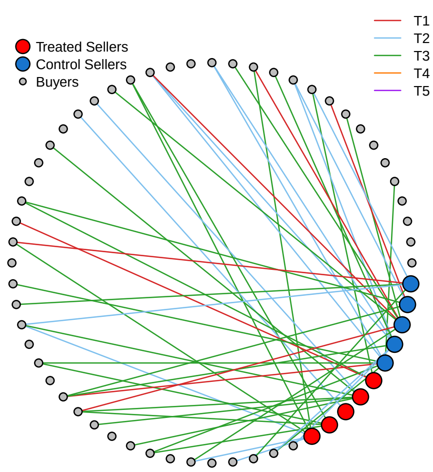

Figure 5.7: Dynamic Network at Time 3

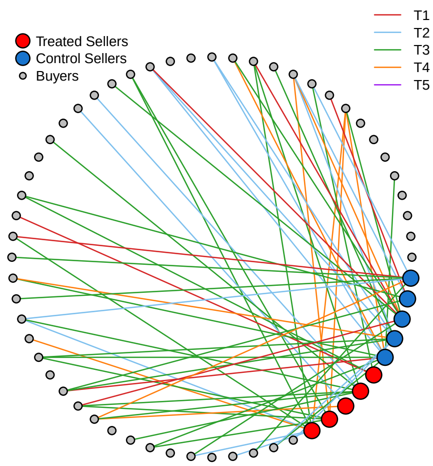

Figure 5.8: Dynamic Network at Time 4

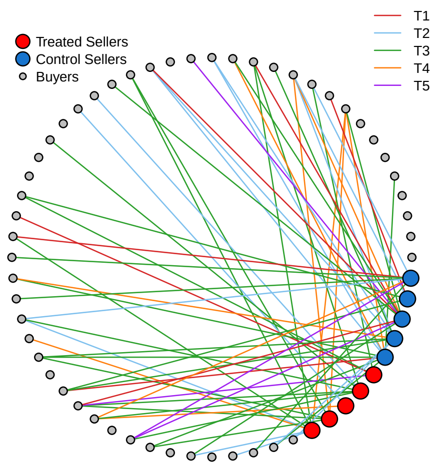

Figure 5.9: Dynamic Network at Time 5

In the exploration of Static Randomized Bernoulli Design, we denote $Z_j \sim IID$ $Ber(p = 0.5)$, illustrating that each Z_j is independently and identically distributed following a Bernoulli distribution with a probability of 0.5.

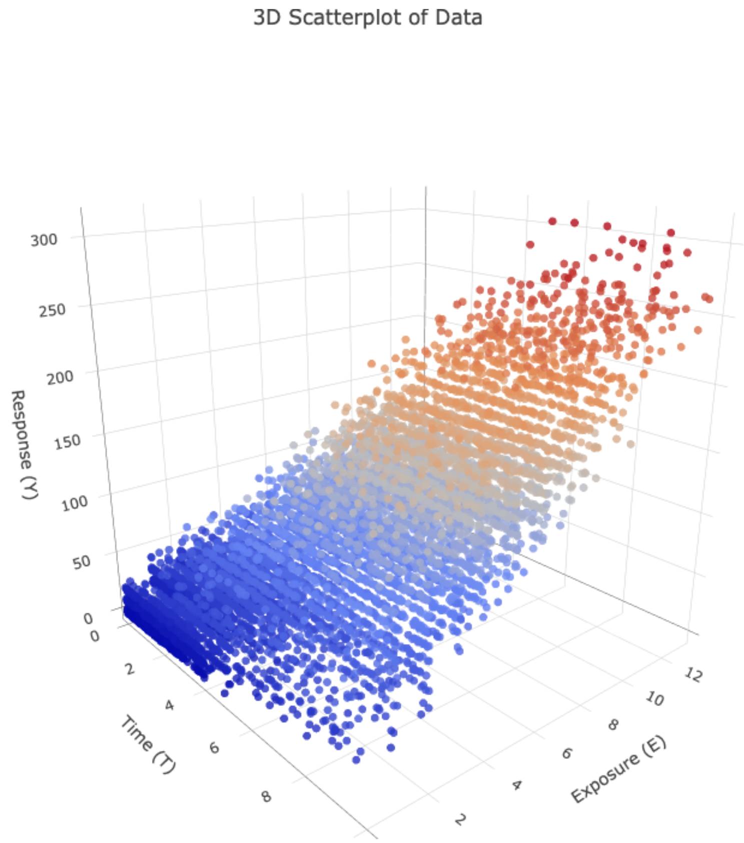

Figure 5.10: Dynamic Networks Simulation of Time-Exposure-Outcome 3D Scatter Plot

For Dynamic Exposure, we represent it as $E_{it} = (W_{i.})_t Z = \sum_{j \in (W_{i.})_t} (W_{ij})_t Z_j$, where the summation is over all possible values in the set $(W_{i.})_t$. Furthermore, we define the potential outcome function of $Y(e)$ given S as:

$$Y_{it}(E_{it})|S_{it} = 1 + \tau E_{it} + \psi S_{it} + \epsilon_i \tag{5.2-3}$$

with $\epsilon_i \sim IID\, Norm(0, 1)$, indicating that the error term ϵ_i follows an independent and identically distributed normal distribution with a mean of 0 and a standard deviation of 1. This formulation succinctly encapsulates the essence of the studied design and exposure dynamics, providing a coherent representation of the underlying mathematical structures and relationships.

5.2.1 S-shape Learning Curve

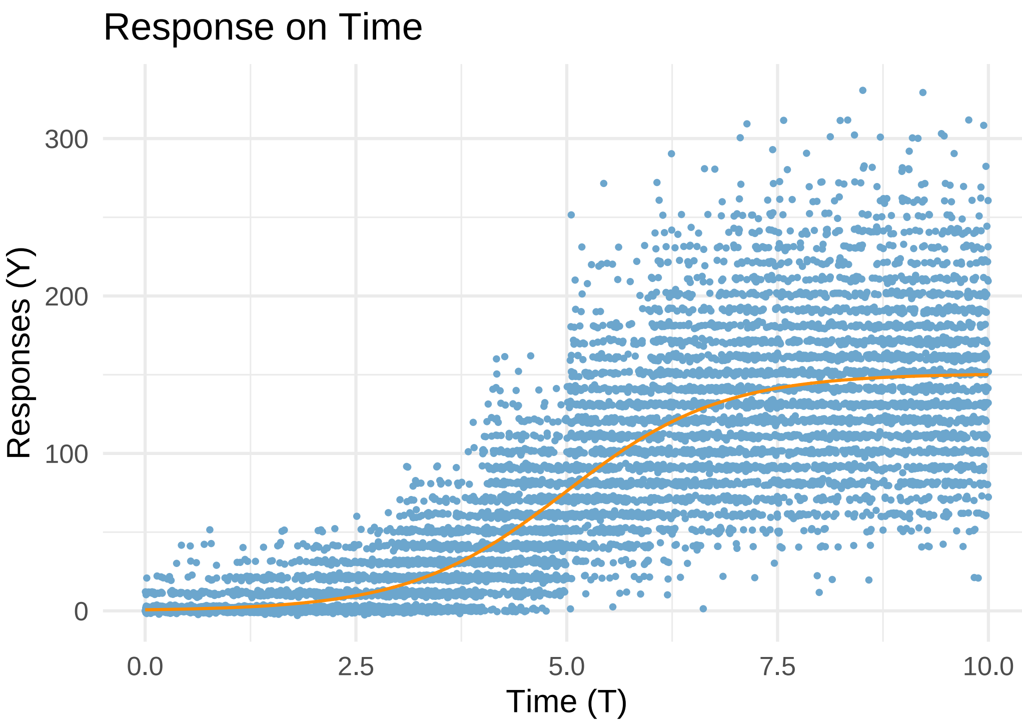

Figure 5.11: Dynamic Networks Simulation: Time-Outcome Scatter Plot (S-shape Learning Curve)

In this subsection, we use the model-based algorithm. The response over time can be modeled by the S-shaped curve, represented by the equation:

$$Y(T) = \frac{\hat{K}}{1 + \exp(-\hat{b}(T - \hat{t}_0))} \tag{5.2-4}$$

where $Y(T)$ is the response at time T. In this equation, $\hat{K} = 150.7$ denotes the maximum of the response mean, indicating the upper asymptotic of the curve. The parameter $\hat{b} = 0.925$ represents the growth or learning rate, defining the steepness of the curve at its midpoint. Lastly, $\hat{t}_0 = 4.98$ signifies the time of median growth, representing the point in time where the curve transitions from the initial phase of slow growth to the subsequent phase of rapid growth. The dispersion of responses also increases with time.

5.2.2 Dynamic Intercepts

In the Naïve Regression, the conditional distribution of Y_{it} given E_{it} is assumed to follow a normal distribution, denoted as:

$$Y_{it}|E_{it} \sim \text{Norm}(\xi_0 + \xi_1 E_{it} + \sum_{t \in T} \xi_t FT_{it} \quad , \ \sigma_y^2) \tag{5.2-5}$$

A modification is incorporated to introduce dynamic intercepts for each factor time period, represented by $\sum_{t \in T} \xi_t FT_{it}$. By employing time factors, we abstain from assuming a specific parametric form for the learning curve, allowing for a more flexible and accurate representation of the underlying relationships.

Table 5.3: Dynamic Networks Simulation: Estimators of Baseline Algorithm

Name	Estimation	S.E.
FT1	1.59	0.58
FT2	2.46	0.83
FT3	4.85	0.83
FT4	11.27	0.83
FT5	24.61	0.84
FT6	42.4	0.88
FT7	55.41	0.92
FT8	62.11	0.95
FT9	64.62	0.96
FT10	65.19	0.96
E	17.31	0.1

Compare with the result without dynamic intercepts.

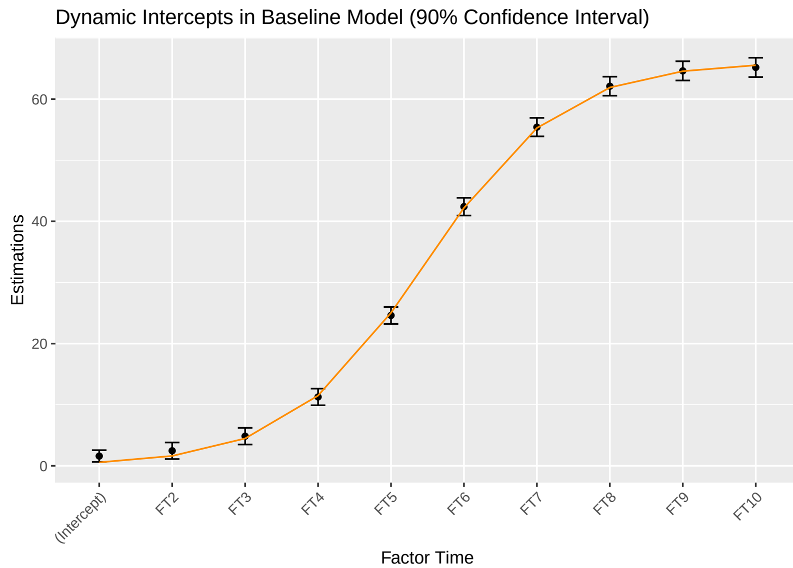

Figure 5.12: Dynamic Networks Simulation: Dynamic Intercepts in Baseline Algorithm

5.2.3 Dynamic Bipartite Networks Simulation ANCOVA

In our study, we examined the relationship between hyper-parameters and the mean squared error (MSE), represented as $MSE = \text{Bias}^2 + \text{Var}$, with a fixed network size of 1000×50. The maximum dynamic network degree is denoted by K, the growth or learning rate of the network by b, and the time of median growth of the network by t_0. We employed a randomized Bernoulli design where Z_j follows an IID Bernoulli distribution with parameter p. The simulation hyper-parameters were set as follows: p values of 0.4, 0.5, and 0.6; K values of 100, 200, and 300; b values of 0.02, 0.05, and 0.10; and t_0 values of 30, 50, and 70. The iterations were performed 30 times.

Table 5.4: Dynamic Networks Batch Simulation: $MSE(\hat{\tau})$ Averages Sample

p	K	b	t0	baseline	IPW.regress	IPW.Boost	model.TSLS	model.regress	model.GAM	DD.TSLS	DD.GAM
0.5	200	0.02	30	43	42.7	10.8	0.4	1.6	4.9	15	6.5
0.5	200	0.02	50	56.8	95.4	7.9	0.3	4	4	3.9	1.8
0.5	200	0.02	70	42.4	27.7	9.5	0.5	3	5.7	10.1	7.8
0.5	200	0.05	30	54.1	19	8.1	0.2	1.3	5.9	15.2	7.2
0.5	200	0.05	50	45.8	18.9	7.1	0.6	4.8	14.1	1.5	0.3
0.5	200	0.05	70	54.5	50.3	2	0.4	3.8	7.1	11	2.9
0.5	200	0.1	30	48.6	13.1	10.6	0.5	4.3	9.6	11.5	4.3
0.5	200	0.1	50	43.8	8	4.8	0.1	5.1	13.5	11.2	24.1
0.5	200	0.1	70	41.2	33.1	0.4	0.1	14.8	20.3	8.7	13.1

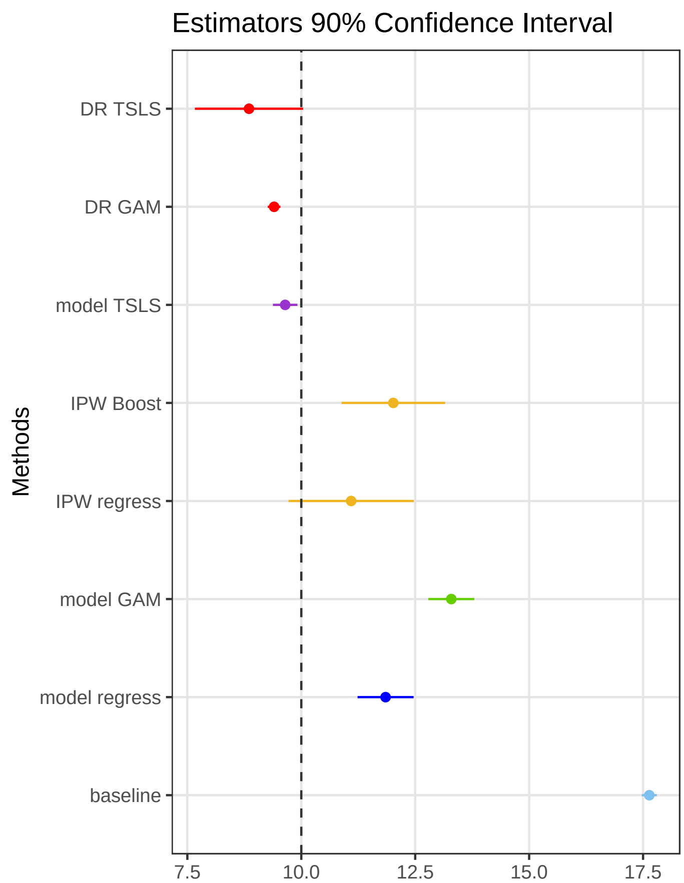

Figure 5.13: Dynamic Networks Simulation: Exposure-Outcome Effect 90% Confidence Intervals

Significant sensitivity ($P < 0.1$): the MSE might be influenced by simulation hyper-parameters under these algorithms.

Table 5.5: Dynamic Networks Batch Simulation: P-values from ANCOVA Analysis Comparing $MSE(\hat{\tau})$ Across Simulation Hyper-parameters for Different Algorithms

Algorithms	**K**	**b**	**t0**	**p**
baseline	0.898	0	0	0
IPW regress	0.919	0	0	0.101
IPW Boost	0.981	0.001	0.011	0
model TSLS	0.053	0	0	0.738
model regress	0.181	0.006	0.001	0
model GAM	0.595	0	0	0
DD TSLS	0.953	0.571	0	0
DD GAM	0.843	0.001	0	0

The exposure of a buyer is defined as $E_i = W_{i.}Z = \sum_{j \in W_{i.}} W_{ij}Z_j$. It is observed that an augmentation in the variability of the network, W_{ij}, coupled with the variability in the treatment assignment probability, $Z_j \sim IID\, Ber(p)$, amplifies the variability in exposure. This escalation in variability subsequently culminates in a diminution of the mean squared error (MSE) in the estimators. Notably, certain algorithms exhibit insensitivity to a limited set of hyper-parameters.

In conclusion, our findings indicate that an increase in the variability of the network, denoted as W_{ij_t}, along with temporal variations represented by time (t), is associated with a corresponding increase in the variability of exposure, expressed as E_{it}. Furthermore, it was observed that the effectiveness of all algorithms under consideration did not show significant dependence on the parameter K.

When to use static and dynamic algorithms. What affects the $MSE(\hat{\tau})$ in ANCOVA. network with higher exposure variation. Connect this point to Jean's paper for future joint papers.

5.3 Practical Considerations for Unbiased Estimation

Provably unbiased estimates of the exposure-outcome function can be obtained only at exposure levels where every outcome unit has a positive probability of receiving. Depending on the nature of the bipartite graph, the weights assigned to the edges, and the treatment assignment design, this might eliminate most, if not all, exposure levels from consideration. Fortunately, practitioners often assume some form of regularity for potential outcomes. By bucketing exposure levels with appropriate granularity, we can accurately represent the exposure-outcome curve, ensuring that each outcome unit has a chance to receive exposure within every exposure bucket with a positive probability. To compute the probability of an outcome unit receiving an exposure level within a specific bucket, simulating a sufficient number of treatment assignments and computing a histogram approximation of each outcome unit's exposure distribution might be a more straightforward approach.

Moreover, while the generalized propensity score methodology starts by estimating the exposure-level-cross-propensity-score function $\beta(e, r)$, estimating it nonparametrically can be challenging if the data is too sparse to obtain meaningful estimates. This remains true even when bucketing exposure levels and propensity scores as suggested in the previous paragraph. Practitioners might achieve better results using a parametric form for $\beta(e, r)$. In their paper on propensity scores for the continuous treatment case, Hirano & Imbens (2004) suggest using a second-degree polynomial of the exposure, E_i, and the generalized propensity score, R_i. This involves running a regression of Y_i on a constant, E_i, E_i^2, R_i, R_i^2, and the interaction term, $E_i \cdot R_i$, and using the resulting approximation $\hat{\beta}(e, r)$ in the second step of the unbiased estimation methodology: $\mu(e) = N^{-1}\hat{\beta}$ that captures the nonlinearity of β. In Section 6, we present results based on using kernel $\hat{\beta}(e, \mathbf{W}_i)$. Another alternative is to use a flexible machine learning approach, ridge regression (see, for example, Friedman et al. (2001)).

Moreover, while our estimator is provably unbiased under a standard set of assumptions, it may suffer from having a large variance in practice, a common problem of propensity-score-based methods. One suggestion is to impute the exposure-response curve at many different levels of exposure and fit a parametric form to "smooth out" the imputed curve. For example, under the linear exposure assumption with normalized weights considered in Pouget-Abadie et al. (2019), as the number of outgoing edges of an outcome unit i grows, the variance of its received exposure shrinks towards its expectation $\mathbb{E}[E_i] = p$, leaving the experimenter with few observations at exposures $e = \{0, 1\}$.

Finally, an alternative to the suggested generalized-propensity-score-based estimator is to stratify based on characteristics of each unit's exposure distribution (e.g., moments of that distribution). Such a stratified estimator computes the average observed outcomes for all units receiving a given exposure coupled with those characteristics within a specified range. The estimates from each stratum are then pooled together to estimate the exposure-outcome function. A similar method was proposed by Imai et al. (2009) for the continuous treatment case in classically run randomized experiments. Although this method does not guarantee unbiased estimates, it might be simpler to compute than generalized propensity scores. In certain scenarios, it can still reduce bias compared to more *naïve* estimators.

5.4 Additional Considerations for Observational Data

While we are primarily concerned with experimental settings, the results of Section 3 are formulated in a way that makes them valid in observational settings as long as the unconfoundedness assumptions are satisfied. In practical terms, working with observational data usually entails the following: 1. The functional form of the generalized propensity scores is unknown, and the generalized propensity scores must be estimated. 2. There is a variety of potential estimands of interest.

The first point is self-explanatory and is expanded upon in Section 6.3. However, the second point merits further discussion. In many experimental settings, researchers are primarily interested in estimating $\mu(1) - \mu(0)$, the average effect of treating the entire population versus not treating anyone. When dealing with observational data, treating the entire population may not be feasible, and researchers might be interested in evaluating the cost-effectiveness of a marginal intervention, as is the case in, for example, Papadogeorgou et al. (2019).

CHAPTER 6

Empirical Results

Using simulations on synthetic graphs and the Amazon buyer-item graph as presented by Pouget-Abadie et al. (2019) and Ni et al. (2019), we analyzed data from the 2018 Amazon Pet Supplies product review. Although the dataset includes 29 categories, only the "Pet Supplies" subcategory was chosen for analysis due to its extensive size. This data provides both product details and reviews from October 1, 1998, to July 1, 2014. Our findings indicate that the proposed estimators and bootstrap variance estimation techniques outperform the baseline methods.

Data summary of the Pet Supply. Number of reviewers: 740985. Number of products: 103288.

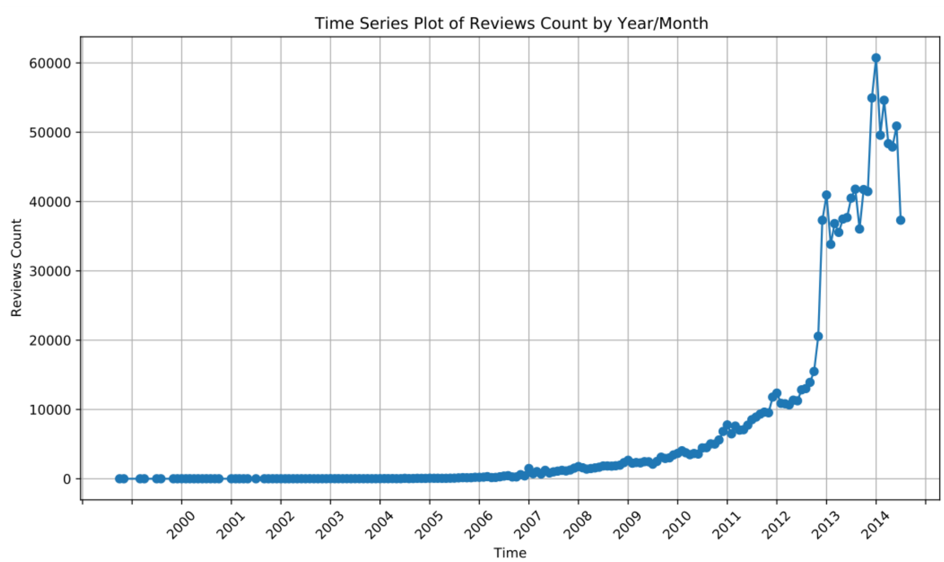

Figure 6.1: Amazon Pet Supply: Time Series of Product Reviews

Table 6.1: Amazon Pet Supply: Rating and Cost Summary Statistics

Statistic	Min	25%Q	50%Q	75%Q	Max	Mean	S.D.
Rating	1	4	5	5	5	4.11	1.33
Cost	0.01	8.99	15.48	30.39	999	29.83	49.31

Table 6.2: Amazon Pet Supply: Buyer's Review Proportion

Number of Reviews	**Reviewer Count**	**Proportion**
1	538197	72.63%
2	109452	14.77%
3	40474	5.46%
4∼10	46897	6.33%
11∼20	4816	0.65%
>20	1149	0.16%

6.1 Amazon Review Data: Ranking System to Total Costs

In the study of Amazon reviews data, the relationship between product rankings and customers' costs was investigated. The response variable was defined as the summation of the reviewer's total cost, while the exposure was the summation of the inverse of the products' rank.

Table 6.3: Amazon Pet Supply: Exposure-Costs Effect Estimator and Standard Errors

Algorithms	Estimator	S.E.
baseline	4.4193	0.0213
IPW regress	9.9250	1.1504
IPW Boost	7.0815	0.5120
model TSLS	7.0424	0.0362
model regress	-7.6924	0.2040
model GAM	10.5282	0.0703
DD TSLS	7.9395	0.6182
DD GAM	8.4485	0.0300

A network confounder was introduced, which was the summation of the number of items also viewed by the customer in the Pet Supply subcategory. Owing to the significant variation observed in the summation ranking and the confounders, quantile normalization was employed to equalize the size of each quantile stratum. The exposure strata were denoted as $E_1, E_2, \ldots, E_{10}$ and the network confounder strata as $D_1, D_2, \ldots, D_{10}$.

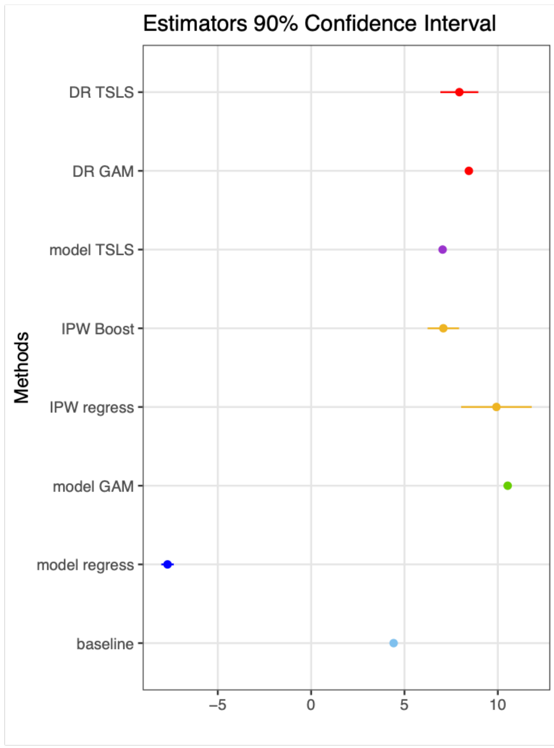

Figure 6.2: Amazon Pet Supply: Exposure-Cost Effect 90% Confidence Intervals

6.2 Amazon Review Data: Ranking System to Total Rating

In the analysis of Amazon reviews data, the response variable is defined as the summation of the reviewer's rating, while the exposure is characterized by the summation of the inverse of the products' rank. Additionally, a network confounder is introduced, which is the summation of the number of items also viewed by the customer in the Pet Supply subcategory. Given the significant variation observed in the summation ranking (E) and the confounders (D), quantile normalization is employed to equalize the size of each quantile stratum. The exposure strata are denoted as $E_1, E_2, \ldots, E_{10}$ and the network confounder strata as $D_1, D_2, \ldots, D_{10}$.

Table 6.4: Amazon Pet Supply: Exposure-Rate Effect Estimator and Standard Errors

Algorithms	Estimator	S.E.
baseline	0.8647	0.0015
IPW regress	0.2760	0.0063
IPW Boost	0.2197	0.0138
model TSLS	0.3785	0.0020
model regress	-2.2335	0.0057
model GAM	0.3954	0.0061
DD TSLS	0.3551	0.0052
DD GAM	0.1914	0.0012

In Figure 6.2 and Table 6.4, we present the 90% Confidence Intervals (C.I.s) and estimator with standard errors (S.E.s) for the Marginal Exposure-Cost Effect across various algorithms. The baseline algorithm yields an estimator of 0.8646(0.0015).

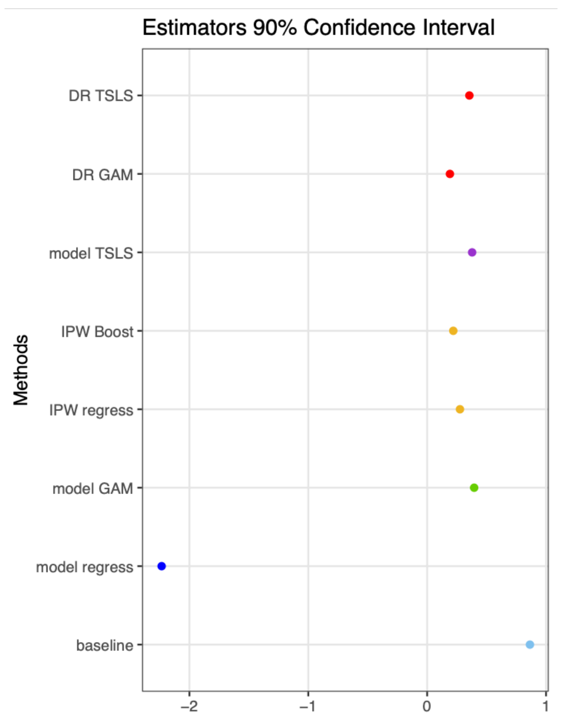

Figure 6.3: Amazon Pet Supply: Exposure-Rate Effect 90% Confidence Intervals

Notably, estimators from alternative algorithms, with the exception of the model regress, range between 0.1914 and 0.3954. Intriguingly, the estimator derived from the model regress algorithm, which operates under the assumption of a comprehensive parametric model, is $-2.2335(0.0057)$. We think this anomaly is due to the misspecification of the parametric terms within the outcome model, denoted as $\beta(e, r)$. While the other proposed algorithms also incorporate at least one parameter (θ_1) associated with the exposure, their primary emphasis is on either weighting mechanisms or non-parametric models to mitigate network confounding effects.

The implemented ranking system may exert an influence on customers, potentially leading to increased expenditure during shopping. However, it is noteworthy that when composing reviews, customers predominantly concentrate on the intrinsic quality of the product, rather than the influence of the ranking system.

CHAPTER 7

CONCLUSION

In this study, we underscore the importance for researchers employing bipartite designs to recognize potential biases in inferential methods that overlook network effects and exposure distributions. Drawing inspiration from seminal works by Imbens (2000), Hirano & Imbens (2004), and Robins et al. (2000), we advocate for the use of propensity score adjustments and discuss practical implications for their application. We present novel theoretical insights, suggesting that our advanced theorems provide precise coverage probabilities for the marginal exposure-outcome function. Our algorithms, as proposed, effectively mitigate the estimator bias associated with the marginal exposure-outcome effect. These theoretical assertions are validated through synthetic simulations and empirical graph analysis. In addition, we investigate an observational setting, contrasting results obtained from both conventional and our recommended algorithms. We highlight the inherent risks associated with relying solely on model-based approaches for causal effect estimation. Future research directions include an in-depth exploration of estimation and inference methodologies that cater to experiments that might alter the structure of the bipartite graph, as well as other observational studies influenced by network confounders.

BIBLIOGRAPHY

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

Del Prete, D., Forastiere, L., & Sciabolazza, V. L. (2020). Causal inference on networks under continuous treatment interference. *arXiv preprint arXiv:2004.13459*.

Eckles, D., Karrer, B., & Ugander, J. (2016). Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference*, *5*(1), 20150021.

Fatemi, Z., & Zheleva, E. (2020). Minimizing interference and selection bias in network experiment design. In *Proceedings of the international aaai conference on web and social media* (Vol. 14, pp. 176–186).

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

Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning. vol. 1 springer series in statistics. *New York*.

Galagate, D. (2016). *Causal inference with a continuous treatment and outcome: Alternative estimators for parametric dose-response functions with applications* (Unpublished doctoral dissertation). University of Maryland, College Park.

Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, *226164*, 73–84.

Holtz, D., Lobel, R., Liskovich, I., & Aral, S. (2020). Reducing interference bias in online marketplace pricing experiments. *arXiv preprint arXiv:2004.12489*.

Hong, G., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational evaluation and policy analysis*, *27*(3), 205–224.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, *47*(260), 663–685.

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

Imai, K., King, G., & Nall, C. (2009). The essential role of pair matching in cluster-randomized experiments, with application to the mexican universal health insurance evaluation. *Statistical Science*, *24*(1), 29–53.

Imai, K., & Van Dyk, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, *99*(467), 854–866.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, *87*(3), 706–710.

Johari, R., Li, H., Liskovich, I., & Weintraub, G. Y. (2022). Experimental design in two-sided platforms: An analysis of bias. *Management Science*, *68*(10), 7069–7089.

Kempton, R. A., Fox, P. N., & Cerezo, M. (2012). *Statistical methods for plant variety evaluation*. Springer Science & Business Media.

Ni, J., Li, J., & McAuley, J. (2019). Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (emnlp-ijcnlp)* (pp. 188–197).

Ogburn, E. L., Sofrygin, O., Diaz, I., & Van der Laan, M. J. (2022). Causal inference for social network data. *Journal of the American Statistical Association*, 1–15.

Papadogeorgou, G., Mealli, F., & Zigler, C. M. (2019). Causal inference with interfering units for cluster and population level treatment allocation programs. *Biometrics*, *75*(3), 778–787.

Pouget-Abadie, J., Aydin, K., Schudy, W., Brodersen, K., & Mirrokni, V. (2019). Variance reduction in bipartite experiments through correlation clustering. *Advances in Neural Information Processing Systems*, *32*.

Robins, J. M., Hernan, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 550–560.

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

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American statistical association*, *75*(371), 591–593.

Saint-Jacques, G., Sorenson, J. E., Chen, N., & Xu, Y. (2019). A method for measuring network effects of one-to-one communication features in online a/b tests. *arXiv preprint arXiv:1903.08766*.

Saveski, M., Pouget-Abadie, J., Saint-Jacques, G., Duan, W., Ghosh, S., Xu, Y., & Airoldi, E. M. (2017). Detecting network effects: Randomizing over randomized experiments. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining* (pp. 1027–1035).

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

Struchiner, C. J., Halloran, M. E., Robins, J. M., & Spielman, A. (1990). The behaviour of common measures of association used to assess a vaccination programme under complex disease transmission patterns—a computer simulation study of malaria vaccines. *International journal of epidemiology*, *19*(1), 187–196.

Tchetgen, E. J. T., & VanderWeele, T. J. (2012). On causal inference in the presence of interference. *Statistical methods in medical research*, *21*(1), 55–75.

Toulis, P., & Kao, E. (2013). Estimation of causal peer influence effects. In *International conference on machine learning* (pp. 1489–1497).

Viviano, D. (2020). Experimental design under network interference. *arXiv preprint arXiv:2003.08421*.

Zigler, C. M., & Papadogeorgou, G. (2021). Bipartite causal inference with interference. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *36*(1), 109.

APPENDIX A

Appendix A

A.1 Proof of Theorems

Proof of Lemma 1 (Balancing Given the Generalized Propensity Score)

First,

$$\begin{aligned}
\Pr(\mathbb{I}(e) = 1|W, r(e,W)) &= \mathbb{E}[\mathbb{I}(e)|W, r(e,W)] \\
&= \mathbb{E}[\mathbb{I}(e)|W] \\
&= r(e,W)
\end{aligned}$$

because by definition, $r(e,W) = \mathbb{E}[\mathbb{I}(e)|W]$.

Second,

$$\begin{aligned}
\Pr(\mathbb{I}(e) = 1|r(e,W)) &= \mathbb{E}[\mathbb{I}(e)|r(e,W)] \\
&= \mathbb{E}\left[\mathbb{E}[\mathbb{I}(e)|W, r(e,W)]\middle| r(e,W)\right] \\
&= \mathbb{E}[r(e,W)|r(e,W)] \\
&= r(e,W)
\end{aligned}$$

Hence, $\Pr(\mathbb{I}(e) = 1|W, r(e,W)) = \Pr(\mathbb{I}(e) = 1|r(e,W))$ and conditionally on

$r(e, W)$, the treatment indicator $\mathbb{I}(e)$ and the pre-treatment variables W are independent. **QED.**

Proof of Lemma 2 (Weak Unconfoundedness Given the Generalized Propensity Score)

First,

$$
\begin{aligned}
\Pr(\mathbb{I}(e) = 1 | Y(e), r(e, W)) &= \mathbb{E}[\mathbb{I}(e) | Y(e), r(e, W)] \\
&= \mathbb{E}\left[\mathbb{E}[\mathbb{I}(e) | Y(e), W, r(e, W)] \middle| Y(e), r(e, W) \right] \\
&= \mathbb{E}[r(e, W) | Y(e), r(e, W)] \\
&= r(e, W)
\end{aligned}
$$

Second, as shown before in the proof for Lemma 1,

$$
\Pr(\mathbb{I}(e) = 1 | r(e, W)) = r(e, W).
$$

Hence,

$$
\Pr(\mathbb{I}(e) = 1 | Y(e), r(e, W)) = \Pr(\mathbb{I}(e) = 1 | r(e, W)),
$$

and conditionally on $r(e, W)$ the treatment indicator $\mathbb{I}(e)$ and the potential outcome $Y(e)$ are independent. **QED**.

Proof of Theorem 1: Modeling with the GPS (Estimation of Average Potential Outcomes through Adjustment for the Generalized Propensity Score)

Under Assumption 1 and Assumption 2, then:

Part (i). First,

$$\begin{aligned}\mathbb{E}[Y|E=e, r(E,W)=r] &= \mathbb{E}[Y(e)|E=e, r(E,W)=r] \\ &= \mathbb{E}[Y(e)|E=e; r(e,W)=r] \\ &= \mathbb{E}[Y(e)|\mathbb{I}(e)=1; r(e,W)=r]\end{aligned}$$

which by unconfoundedness is equal to

$$\mathbb{E}[Y(e)|r(e,W)=r]$$

Part (ii) follows directly by applying iterated expectations.

Proof of Theorem 2: Weighting by the GPS (Weighting and the Generalized Propensity Score)

First, rewrite the expectation as an iterated expectation with the inner expectation conditional on W:

$$\mathbb{E}\left[\frac{Y \cdot \mathbb{I}(e)}{r(E,W)}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{Y \cdot \mathbb{I}(e)}{r(E,W)}\middle| W\right]\right]$$

Next, by conditioning on $\mathbb{I}(e)=1$ and multiplying by the probability of $\mathbb{I}(e)=1$ this is equal to

$$\mathbb{E}\left[\mathbb{E}\left[\frac{Y}{r(E,W)}\middle|\mathbb{I}(e)=1, W\right] \cdot \Pr(\mathbb{I}(e)=1|W)\right]$$

Conditional on $\mathbb{I}(e)=1$, $Y=Y(e)$ and $r(E,W)=r(e,W)$, so this can be rewritten as:

$$\mathbb{E}\left[\mathbb{E}\left[\frac{Y(e)}{r(e,W)}\middle|\mathbb{I}(e)=1,W\right]\cdot \Pr(\mathbb{I}(e)=1|W)\right]$$

Now the conditioning on $\mathbb{I}(e)$ is irrelevant by the weak unconfoundedness assumption, so this is equal to:

$$\mathbb{E}\left[\mathbb{E}\left[\frac{Y(e)}{r(e,W)}\middle|W\right]\cdot r(e,W)\right]=\mathbb{E}\left[\mathbb{E}[Y(e)|W]\right]=\mathbb{E}[Y(e)]$$

QED.

Proof of Lemma 3 (Weak Unconfoundedness of Outcome Model Given the GPS)

$$\begin{aligned}\Pr(\mathbb{I}(e)=1|\beta(e,r),r(e,W)) &= \mathbb{E}[\mathbb{I}(e)|\beta(e,r),r(e,W)]\\ &= \mathbb{E}\left[\mathbb{E}[\mathbb{I}(e)|\beta(e,r),W,r(e,W)]\middle|\beta(e,r),r(e,W)\right]\\ &= \mathbb{E}[r(e,W)|\beta(e,r),r(e,W)]\\ &= r(e,W)\end{aligned}$$

$$\Pr(\mathbb{I}(e)=1|r(e,W))=r(e,W).$$

Hence,

$$\Pr(\mathbb{I}(e)=1|\beta(e,r),r(e,W))=\Pr(\mathbb{I}(e)=1|r(e,W)),$$

and conditionally on $r(e,W)$ the treatment indicator $\mathbb{I}(e)$ and the potential outcome model $\beta(e, r)$ are independent. **QED**.

Proof of Theorem 3: Weighting and Modeling with the GPS (Double Deconfounded)

$$\mathbb{E}\left[\frac{\beta(e,r)\cdot\mathbb{I}(e)}{r(E,W)}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{\beta(e,r)\cdot\mathbb{I}(e)}{r(E,W)}\middle|W\right]\right]$$

Next, by conditioning on $\mathbb{I}(e) = 1$ and multiplying by the probability of $\mathbb{I}(e) = 1$.

$$\mathbb{E}\left[\mathbb{E}\left[\frac{\beta(e,r)}{r(E,W)}\middle|\mathbb{I}(e) = 1, W\right]\cdot\Pr(\mathbb{I}(e) = 1|W)\right]$$

Based on the Lemma 3, Conditional on given strata $\mathbb{I}(e) = 1$, $\beta(e, r)$ and $r(E, W) = r(e, W)$, then rewritten as:

$$\mathbb{E}\left[\mathbb{E}\left[\frac{\beta(e,r)}{r(e,W)}\middle|W\right]\cdot\Pr(\mathbb{I}(e) = 1|W)\right]$$

Now the conditioning on $\mathbb{I}(e)$ is irrelevant by the weak unconfoundedness assumption, so this is equal to:

$$\mathbb{E}\left[\mathbb{E}\left[\frac{\beta(e,r)}{r(e,W)}\middle|W\right]\cdot r(e,W)\right] = \mathbb{E}\left[\mathbb{E}\left[\beta(e,r)|W\right]\right]$$

We know $\beta(e, r)$ is the outcome model function for the outcome $Y(e)$. by Lemma 1, given the network information W, then $\mathbb{E}[\beta(e, r)|W] = Y(e)$. Then above double expectation is equivalent to:

$$\mathbb{E}\left[\mathbb{E}\left[\beta(e,r)|W\right]\right] = \mathbb{E}\left[Y(e)\right] = \mu(e)$$

QED.