Semi-parametric Panel Count Model for Drug Safety Evaluation

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Biostatistics

By

Yizhao Zhou, M.S.

Washington, DC
June 11, 2021

Semi-parametric Panel Count Model for Drug Safety Evaluation

Yizhao Zhou, M.S.

Dissertation Advisors: Ao Yuan, Ph.D. and Ming T. Tan, Ph.D.

Abstract

In this dissertation, we study issues related to drug safety evaluation. In the first part, we focus on the adverse event (AE) signal detection in post-market surveillance systems. We derive a semi-parametric panel count model to search for safety signals by accounting for background noise, issues associated with such as zero-inflated data count, and covariates information. We develop an estimating procedure with Expectation-Maximization (EM) algorithm to estimate the model. In each M-step, the maximization of the non-parametric component is reformulated as an optimization problem in isotonic regression. The strong consistency and asymptotic distributions of the model estimators are formally derived. We conduct simulation studies to evaluate the finite sample performance of the method proposed and to demonstrate the advantages of the proposed method in signal detection with high power for signal detection, high specificity, and sensitivity. The proposed method is applied to WHO VigiBase System and FAERS with several relevant covariates yielding new signals not found with standard approaches and reduced false positive rates. In the second part, we develop the doubly robust estimator in the panel count model to improve the method of inferring causal effects of medicines/vaccines on adverse events (AE) from data with Poisson outcome. Simulation studies demonstrate its robustness with respect to misspecifications of the propensity score or outcome model.

Index words: Causal effect, Covariate information, Doubly robust estimator, Drug safety, Non-randomized Experiments, Semi-parametric panel count model

Dedication

To my parents, grandparents, sister, and cousins.

Acknowledgments

I would like to express my most profound appreciation to my mentor, Dr. Ao Yuan, and Co-mentor, Dr. Ming T. Tan, for guiding me through the research journey. Dr.Yuan is very responsible, and he always makes sure I am on the right track. In addition, Dr. Yuan has a deep sense of scientific rigor, which influences me to keep pursuing advances in methodology. Dr.Tan guides me in the dissertation research with his wide range of knowledge and insightful views on things, which leads me to try creative ideas to see things from different perspectives. In addition, Dr.Tan is approachable and generous in giving help, and his encouragement helps me stay positive.

I would also like to extend my deepest gratitude to my dissertation committee for their constructive advice. I would like to extend my sincere thanks to Dr. Kepher Makambi for giving me insightful suggestions in presenting and writing articles since I was a Master's student, which helps me start a good habit of displaying my statistical work professionally. In addition, I am also grateful to Dr. Arnold L. Potosky for giving me invaluable advice about telling a story to non-statisticians, which improves my presentation and skills of communication with professionals from different fields. Next, I would like to thank Dr. Guoqing Diao for his constructive comments during committee meetings and the oral qualifying exam, which helps improve my dissertation. I would gratefully acknowledge the assistance of Max Smith, PharmD, for his collaboration and assistance in obtaining the data from WHO.

Also, I would like to thank Dr. Jaeil Ahn for influencing me to get organized and Dr. Hongbin Fang's helpful, practical suggestion when facing multiple projects and directions. I can not leave Georgetown without mentioning Drs. Valeriy R. Korostyshevskiy, Ruzong Fan, James Li, George Luta, Colin Wu, Hongkun Wang, Anca Dragomir, and Jan K. Blancato for teaching me biostatistics, epidemiology, and cancer genetics. I have greatly enjoyed and benefited from the classes I took with them over the past six years. Without the knowledge learned from them, I can not be where I am now. I would like to send thanks to all the faculty members for their unwavering support. Also, I would like to recognize the help that I received from my classmates. Especially, Tengfei Li is always generous in sharing his experience and giving invaluable suggestions. In addition, special thanks go to Minnie An and Veronica Chiu for their important administrative support and coordination of programmatic and social events. They have done more than I can list. My study here would not have been so smooth without their help. I would also like to send special thanks to my other Georgetown classmates, Tianmin Wu, Shuqi Wang, Shilan Li, and Haiou Li, for being always there for me when I need them.

When I first arrived at Georgetown, I was nervous about the unfolding new chapter of my life. During the past six years, I have received a lot of help from faculty, staff, friends, even strangers. I benefited more than I had expected from the incredible journey. I deeply appreciate the time I have spent at Georgetown University.

Last but not least, I am deeply indebted to my parents and my grandparents for their unparalleled support and unconditional love throughout my study and my life. They are always there for me and inspire me to move forward whenever I feel down.

Table of Contents

Chapter

1 Introduction 1
1.1 Post-market monitoring 1
1.2 Post-market drug safety surveillance systems 1
1.3 Inferring causal effects of medications and vaccines on adverse events 3
1.4 Drawing causal inferences of non-randomized adverse events from clinical study 5
1.5 Panel count data 5
1.6 Research aims 6
1.7 Organization of the dissertation 6

2 Signal detection in post-market drug surveillance systems using semi-parametric panel count model 8
2.1 Introduction 8
2.2 Existing statistical methods for signal detection 12
2.3 The proposed statistical model 19
2.4 Estimation procedure 20
2.5 Hypothesis testing for signal detection 25
2.6 Simulation studies: Finite sample properties 27
2.7 Analysis of cardiovascular medicines and corresponding adverse events 54
2.8 Discussion 61

3 Theoretical properties of the semi-parametric panel count model 66
3.1 Introduction 66
3.2 Regularity condition lists 66
3.3 Strong consistency of the estimators 67
3.4 Asymptotic normality and efficiency of the Euclidean component 70
3.5 Asymptotic distribution of non-parametric component 76
3.6 General asymptotic distribution of the derivatives of non-parametric component 82

4 Inferring causal effects of medications and vaccines on adverse events from observational and non-randomized studies with count outcomes . 86

4.1 Introduction 86
4.2 Inferring causal effects 88
4.3 A panel count model for count outcomes 91
4.4 Asymptotic properties 96
4.5 Simulation studies: Finite sample properties 100
4.6 Discussion 110

5 Summary and future work 112

Appendix

A Disclaimer 116

Bibliography 117

List of Figures

2.1 Estimates of $G(t)$ when $G(t)$ is convex, concave, a mixture of convex and concave, sample size is 200×10 under Case 1: Solid line: True $G_0(\cdot)$; Step line: Estimate $\hat{G}(\cdot)$. 38

2.2 Estimates of $G(t)$ when $G(t)$ is convex, concave, a mixture of convex and concave, sample size is 5000×10 under Case 2: Solid line: True $G_0(\cdot)$; Step line: Estimate $\hat{G}(\cdot)$. 39

2.3 Estimates of $G(t)$ when $G(t)$ is convex, concave, a mixture of convex and concave, sample size is 1000×100 under Case 3: Solid line: True $G_0(\cdot)$; Step line: Estimate $\hat{G}(\cdot)$. 40

List of Tables

1.1 Surveillance data for I adverse events and J drugs 3

2.1 2×2 display of distribution of the number of adverse event reports . . 12

2.2 Parameter estimation of the proposed model under Case 1-3 when sample size, dimension of $\boldsymbol{x}$ or shape of $G(t)$ varies. 32

2.3 Performance characteristic of the proposed method in signal detection when signal strength (rr) or sample size varies. 42

2.4 Performance characteristic of the proposed method without covariates information in signal detection when signal strength (rr) or sample size varies. 45

2.5 Performance characteristic of PRR, BCPNN and ZIP LRT in signal detection when signal strength (rr) or sample size varies. 48

2.6 Signals (AEs) detected by the proposed method, the proposed method without covariates, PRR, BCPNN and ZIPLRT. 51

2.7 Performance characteristic of the proposed method when the proportions of true zeros ($\rho(\%)$) and sample size varies. 52

2.8 Performance characteristic of the proposed method when rr is extreme small and the signals are extreme weak. 54

4.1 Causal effects from inverse weighting estimator, outcome models and doubly robust estimator under correct or misspecified models. 102

Chapter 1

Introduction

1.1 Post-market monitoring

According to the Food and Drug Administration (FDA), an adverse event (AE) is any undesirable experience associated with using a medical product in a patient. AE may cause emergency room visits, even hospitalization, and severe AEs may be life-threatening (FDA, 2016). Monitoring AEs is an essential aspect of drug development and public health. Pre-marketing studies generally have minimal statistical power to detect all the potential AEs (Strom, 2006; Gibbons and Amatya, 2016) due to their limited sample sizes, special patient selection criteria, and care settings. Therefore, it is essential for regulatory agencies and pharmaceutical companies to monitor post-market drug safety and identify signals in surveillance systems, which may have severe consequences, including withdrawal of the product from the market. There are thousands of reported AEs on thousands of drugs in post-market drug safety surveillance systems. It is of interest to identify AEs with high reporting rates for a specific drug or identify the drugs with high reporting rates for specific AEs.

1.2 Post-market drug safety surveillance systems

There are currently several post-market surveillance systems worldwide, including the VigiBase and the FDA Adverse Event Reporting System (FAERS). VigiBase is the

WHO (World Health Organization) global database of individual case safety reports with over 20 million suspected AEs submitted since 1968, and FAERS contains AE reports and medication errors submitted to the US FDA since 1968. More recently, in response to demands of patients and caregivers for more information about the associated benefits and potential risks of the medical products they use, the FDA, as mandated by US congress, launched the Sentinel System in 2008 for better medical product safety surveillance where more information and variables such as subgroups and long term follow-ups are included. Since Sentinel is the largest and comprehensive multi-site distributed database in the world dedicated to medical product safety, it allows further evaluation if an event of potential concern is identified in FAERS. As a component of the Sentinel System, the Active Risk Identification and Analysis (ARIA), mandated by the US Congress and established by the FDA in 2007, includes pre-defined parameterized reusable routine querying tools electronic data in the common data model. The FDA also issued a Five-Year Strategy from 2019-2023 towards the Sentinel System to augment its safety analysis capabilities using advances in data science and signal detection (FDA, 2020).

Such systems contain a large number of subjects taking different drugs over many years. If a drug on the market has shown unacceptable toxicities, the drug may have to be taken off the market for obvious public health reasons. However, it also potentially limits treatment options for some patients. A unique feature of such data is that the background is noisy with limited information collected. Moreover, after collapsing data into the form of $I \times J$ table (shown in Table 1.1), a large number of cell counts are zeros since a common AE for one drug might be rare for another drug. For example, the FAERS reports in the 4th quarter of 2019 include about 1657 drugs and 9262 types of AEs where 98.79% of cell counts in drugs by AE types are zeros, although there is a large collective count of AEs. Another feature is that such systems contain

patient demographics that could be adjusted for to remove imbalances resulting from heterogeneous patients' characteristics from one drug to another.

Table 1.1: Surveillance data for I adverse events and J drugs

	Drug 1	...	Drug j	...	Drug J
AE 1	y_{11}	...	y_{1j}	...	y_{1J}
...	...	...	...	...	...
AE i	y_{i1}	...	y_{ij}	...	y_{iJ}
...	...	...	...	...	...
AE I	y_{I1}	...	y_{Ij}	...	y_{IJ}

Although post-market drug surveillance systems play an essential role in drug safety evaluation and attract more and more attention from researchers from all over the world, the systems are limited by the data that are captured. Firstly, the information on causality assessments of the drug and event is not available in all reports. For example, FAERS collects any AEs associated with the use of a drug in human subjects, whether or not they are considered drug related (CFR, 2021). In addition, significant heterogeneity exists in the reports since they are from multiple sources. For example, the reports in VigiBase are from different countries with different reporting codes and different types of reporters. FAERS includes voluntary reports directly from health-care professionals and consumers; reports from mandatory reporters-manufacturers and importers. The multiple sources lead to heterogeneity of individual reports and duplicate reports where different reporters submitted the same report. Furthermore, systems can not verify all the information in the reports (FDA, 2018).

1.3 Inferring causal effects of medications and vaccines on adverse events

Inferring causal effects of medicines/vaccines on potential AEs is of great public health importance. Note that associations found in the post-market surveillance systems do not equal causality (Davies and Thomas, 2017). There are several reasons we can not

establish causality in post-market surveillance systems. (1) The properties of the systems mentioned in §1.2 include reporting requirements based on association instead of causality, the existence of duplicate reports, and the lack of the number of users of drugs. (2) The information of control group is not available: Only patients taking medicines are included in the systems. In addition, patients generally have a higher risk of AEs than the general population (Davies and Thomas, 2017). Thus, the data from such systems are primarily for initial signal screening but not for causal analysis. When post-market monitoring suggests patients experience specific AEs after taking medicines or receiving vaccines, regulators need to balance all the evidence from case reports, randomized trials, meta-analyses of randomized trials, and observational studies to make further decision regarding the safety of the product, including inferring causal effects of the drugs on the AEs. The gold standard in inferring causal effect is randomized experiments, which can provide the most substantial evidence of causal associations. But randomized experiments are not always feasible due to ethical or practical concerns. In this case, one may draw causal inference from observational studies by mimicking the design of the randomized trial (Hernán and Robins, 2016; Hernán and Robins, 2020).

Person-level data are the basis of inferring causal effects from observational data. Medical claims databases including Marketscan, Lifelink, Veterans Administration, and PharMetric contain complete longitudinal information on AEs before and after the drug exposure (Rubin, 2006). Also, the medical claim databases are large enough to explore rare AEs and match patients with and without prescribing drugs. For example, FDA sponsored two observational retrospective cohort studies conducted by the Department of Defense's (DoD) U.S. Army Medical Command's Pharmacovigilance Center (Meyer et al., 2013) and by the Department of Veterans Affairs' (VA) Center for Medication Safety (Cunningham et al., 2016) to infer causal effects of

varenicline on neuropsychiatric events after FAERS suggests the risk of neuropsychiatric AEs being associated with the drug varenicline. Both studies found no causality of varenicline on neuropsychiatric hospitalizations.

1.4 Drawing causal inferences of non-randomized adverse events from clinical study

Even in a randomized trial, e.g., for oral pain medication, if an AE (say, cardiovascular disease) occurs more frequently in the treatment group than in the placebo group, and it is also observed that some patients overdosed the medicine. We may want to know if overdosing instead of the medication itself caused the AE (see more details in Herson, 2009). We might perform a causal analysis of the effect of overdosing on the outcome of cardiovascular disease even though patients were not randomized based on overdosing or not. This could lead to protracted discussions if the causal analysis revealed that it was overdosing and not the experimental treatment that caused the cardiovascular disease. In this randomized trial, subjects were randomly assigned over the oral medication instead of overdosing, so the causal effect of overdosing on cardiovascular disease needs to be assessed with causal inference methodology for observational data.

1.5 Panel count data

Although most research work have focused on continuous and binary outcomes, non-negative integer outcomes for individual reports are common in a variety of pharmaceutical and epidemiological studies. For example, CD4 T counts (Hammer et al., 1996), number of hot flushes (Prague et al., 2017), and suicide attempts (Olfson et al., 2006), neuropsychiatric events (Evins et al., 2019). In this case, only the number of

occurrences of the events between subsequent observation times are available, and the exact occurrence times of the events are unknown (Sun and Zhao, 2013). For example, a patient may tell her doctors four hot flush episodes occurred between her first and second hospital visit. Poisson model is the most commonly used model or assumption in this case for analysis of panel count data (Sun and Zhao, 2013).

1.6 Research aims

In this dissertation research, we studied issues related to drug safety evaluation. Specifically, I have achieved the following aims.

Aim 1. Develop a novel semi-parametric panel count model to detect AE signals in post-market surveillance systems by adjusting for covariates information and handling zero-inflated issues.

Aim 2. Derive asymptotic properties of the proposed panel count model.

Aim 3. Infer causal effects of medicines/vaccines on potential AEs from observational/non-randomized data with Poisson outcomes using a doubly robust estimator.

1.7 Organization of the dissertation

Chapter 2 proposes a semi-parametric panel count model to detect AE signals by accounting for covariates, background AE occurrences, and excessive zero counts. This Chapter develops an estimating procedure with Expectation-Maximization (EM) algorithm to estimate the model, where in each M-step, the maximization of the non-parametric component is reformulated as an optimization problem as in isotonic regression. Also, in this Chapter, we conduct simulation studies to evaluate the finite sample performance of the method proposed and demonstrate the apparent advantage of the proposed method in signal detection with high power for signal detection, high

specificity, and sensitivity. In addition, we apply the proposed method to VigiBase datasets and an FDA dataset to detect AE signals.

Chapter 3 derives strong consistency, asymptotic distributions of the proposed semi-parametric panel count model, and proofs of all related theorems.

Chapter 4 develops the doubly robust causal estimator in a flexible semi-parametric panel count model with a non-homogeneous Poisson process to infer causal effects from observational/non-randomized studies with count outcomes. It also includes simulation studies to demonstrate the robustness of the proposed method to model misspecifications.

Finally, Chapter 5 provides a brief conclusion of the dissertation and discusses potential future work.

Simulation studies and real data analyses are coded in the R programming environment. The R codes are available upon request. It will be further enhanced in future work, for example, developing the program into more user-friendly R packages.

Chapter 2

Signal detection in post-market drug surveillance systems using semi-parametric panel count model

2.1 Introduction

Panel count data occur in a wide range of applications from scientific investigation, evidence-based policy research to business, including the number of accidents, product defects, and insurance claims. One application of significant public health importance is in post-market drug safety surveillance systems, where millions of reported AEs associated with thousands of drugs are monitored.

As mentioned earlier in the introduction of post-market surveillance systems in §1.2, such systems have three important features, which need to be accounted for in our modeling. One feature of such data is that the background noise is strong with limited information collected. Moreover, after collapsing data into the form of $I \times J$ table (shown in Table 1.1), a large number of cell counts are zeros since a common AE for one drug might be rare for another drug. Another feature is that such systems contain patient demographics that could be adjusted for to remove imbalances resulting from heterogeneous patients' characteristics from one drug to another. This motivates us to develop a novel model capable of accurately detecting true signals in such large and complex systems while incorporating essential patient characteristics, including demographic information.

Detecting AE signals in such systems has been a long-standing statistical problem for decades. For some years, the WHO drug monitoring center used the essential feature of K index, which is the root of Pearson's χ^2 test statistic with Yates's correction (Patwary, 1969; Finney, 1974). The approach is based on the principle that the most significant drug-AE combinations have the greatest K and is easily understood. It is not hard to compute all the AEs in the 1960s and 1970s when the database is not large. However, they do not have definite thresholds of the K index to define potential signals, which is only determined by previous experience and comparing between various combinations when applying this index. The Proportional Reporting Ratio (PRR) approach developed by Evans et al. (2001) reduces the vagueness in thresholds by quantifying the judgment of signals based on the PRR, the value of chi-squared, and the absolute number of reports. PRR is a straightforward approach based on the proportionate approach in a 2×2 table. PRR is simple to calculate and interpret. However, if we view the 2×2 table as one in a case-control study, which means the drug of interest as a case group and other drugs as a control group, Reporting Odds Ratio (ROR) would have been more suitable than PRR as the ORs are close to relative risk in case-control studies when the event is rare (Rothman et al., 2004; Miettinen and Wang, 1981). ROR measure does remove part of the bias in PRR. However, it suffers critical deficiencies along with PRR, such as inflated overall type I error due to multiple comparisons among large drug-event combinations.

Indeed, the likelihood ratio test (LRT) based approach proposed by Huang et al. (2011) controls the overall type I error in the signal screening process by a step-down process. LRT assumes the cell counts of the drug-AE combination tables follow Poisson distributions. However, excessive zeros in the database violate this assumption. Huang et al. (2017) improved the LRT-based approach by assuming the cell counts follow zero-inflated Poisson (ZIP) distribution to handle the zero counts issue.

Both LRT and ZIP LRT methods only consider the drug combinations for a single particular drug at one time. To monitor a class of drugs, Huang et al. (2013) extended the LRT approach for signals, including collections of drugs (or AEs), by introducing a weight matrix to indicate if drugs belong to the same drug class. Simulation studies demonstrate that LRT-based approaches control both type I error and false discovery rate (FDR). In addition, these LRT-based approaches allow stratified analysis if the number of strata is small. However, they also have several limitations: 1) their P-values are determined by empirical null distribution obtained by Monte Carlo simulation. It would not be practical and time-consuming when the number of drugs included is large and Monte Carlo simulation needs to repeat thousands of times in one process; 2) LRT based approach is still built on the PRR measure, so it shares the limitation of PRR mentioned above; 3) the handling of excessive zero counts by the ZIP model is not satisfactory: ZIP LRT dealt with the excessive zero counts, but the assumption is still questionable since excessive large collective counts of some drugs are still not accounted for by a Poisson distribution, which can be a serious problem; 4) the model and analysis are for each AE/drug separately without considering the collective roles of the AEs on the drugs; and 5) stratified analysis is not able to incorporate covariates information, especially continuous variables or when the number of covariates is large.

Furthermore, Bayesian data mining has also been introduced (DuMouchel, 1999). Bate et al. (1998) applied the Bayesian Confidence Propagation Neural Network (BCPNN) (Lansner and Holst, 1996) to the WHO database with a one-layer model based on Information Component(IC) measure. Multi-Item Gamma Poisson Shrinker (MGPS) (DuMouchel, 1999) approach assumes the prior on relative report rate (RRR) is a mixture of two gamma distributions and the number of drug-AE combination follows the Poisson distribution, then compares the Empirical Bayesian Geometric Mean (EGBM) scores with thresholds. Stratified analyses are used to control potential con-

founding factors. Norén et al. (2006) extended the BCPNN approach to allow for more complex quantitative associations by assuming cell counts of the 2×2 table following multinomial instead of binomial distribution in the original paper of Bate et al. (1998). They also propose a Mantel–Haenszel-type adjustment for the IC to control potential confounders. The biggest issue with the Bayesian approach is that it depends on prior distributions and does not address the zero-inflation issue.

Consequently, this Chapter proposes a semi-parametric panel count model to detect safety signals that capture higher reported occurrences in surveillance systems robustly (§2.3 - §2.8) after a review of existing statistical methods in §2.2. The method considers excessive zero counts, categorical and continuous covariates such as age, gender, and weights. Since the patient population in such systems is known to be very heterogeneous, such modeling has the potential to bring significant improvement in the accuracy of signal detection by alleviating the heterogeneity. We show that the new method is able to overcome the drawbacks of current methods and remove the bias from RR based approaches. In addition, this method utilizes all the data jointly, i.e., system-wide, instead of separately for each drug, and accounts for both the extra-large marginal total counts for each drug and the nonzero counts fully for which the ZIP model is not able to. We use a non-parametric component to adjust for the noisy background of the surveillance data. We formulate signal detection as hypothesis testing based on the estimated model adopting the Holm-Bonferroni approach to control the family-wise error rate (FWER) in the signal detection process.

Table 2.1: 2×2 display of distribution of the number of adverse event reports

	Drug j	Other Drugs	Row Total
AE i	y_{ij}	$y_{i\cdot} - y_{ij}$	$y_{i\cdot}$
Other AEs	$y_{\cdot j} - y_{ij}$	$y_{\cdot\cdot} - y_{i\cdot} - y_{\cdot j} + y_{ij}$	$y_{\cdot\cdot} - y_{i\cdot}$
Col Total	$y_{\cdot j}$	$y_{\cdot\cdot} - y_{\cdot j}$	$y_{\cdot\cdot}$

2.2 Existing statistical methods for signal detection

2.2.1 Proportional reporting ratio (PRR)

PRR(Evans et al., 2001) is the most straightforward measure of association between one drug j and one AE i. PRR is based on the proportionate approach in a 2×2 table (Shown in Table 2.1). PRR is defined as the ratio between two conditional probabilities,

$$PRR_{ij} = \frac{P(Drug_j|AE_i)}{P(Drug_j|AE_{\sim i})}. \tag{2.1}$$

Where $Drug_j$ denotes the j-th drug, AE_i denotes the i th AE, and $AE_{\sim i}$ denotes all other AEs except for the ith AE in the database. Then based on the Table 2.1, the estimate of (2.1) is

$$\hat{PRR}_{ij} = \frac{y_{ij}/y_{i\cdot}}{(y_{\cdot j} - y_{ij})/(y_{\cdot\cdot} - y_{i\cdot})}.$$

Based on the delta method,

$$\log(\hat{PRR}_{ij}) \sim N(\log(PRR), \sigma^2_{PRR_{ij}}),$$

where

$$\hat{\sigma}^2_{PRR_{ij}} \approx \frac{1}{y_{ij}} - \frac{1}{y_{i\cdot}} + \frac{1}{y_{\cdot j} - y_{ij}} - \frac{1}{y_{\cdot\cdot} - y_{i\cdot}}.$$

So, the 95% confidence interval (CI) for PRR can be approximated by

$$CI_{PRR} = \exp\left[\log(P\hat{R}R_{ij}) \pm 1.96 \times \sqrt{\sigma^2_{PRR_{ij}}}\right].$$

The AE i on Drug j is suggested as a signal if the lower bound of the CI is larger than 1. PRR is simple to calculate and interpret. However, this approach has its limitations. For example, the method is for a single drug-AE combination and ignores the joint distribution of drug-AE combinations. Also, no assumption is related to zero-inflated issues in the estimating process, so PRR can not handle zero-inflated problems.

2.2.2 Bayesian confidence propagation neutral network (BCPNN)

Denote p_i as the probability of AE i present on a report; p_j as the probability of Drug j present on a report; p_{ij} as the joint probability that both AE i and Drug j present on the same report. Bate et al. (1998) applied BCPNN to signal detection in post-market surveillance databases. The strength of association between drug Drug j and AE i is measured with Information Component (IC), which is defined as

$$IC_{ij} := \log_2 \frac{p_{ij}}{p_i p_j}.$$

The basic idea is if IC_{ij} is larger than 0, AE i is a positive signal of Drug j. The BCPNN method is based on the IC measure. Assume AE i and Drug j are binary variables (present or not in an individual report), with the data shown in Table 1.1, then we have

$$\begin{aligned} y_{ij} &\sim \binom{y_{..}}{p_{ij}} \\ y_{i\cdot} &\sim \binom{y_{..}}{p_i} \\ y_{\cdot j} &\sim \binom{y_{..}}{p_j}. \end{aligned}$$

Assume prior of p_i, p_j and p_{ij} are beta distributions with shape parameters $\alpha_i, \xi_i, \alpha_j, \xi_j, \alpha_{ij}, \xi_{ij}$, respectively, then we have

$$p_i \sim \text{Beta}(\alpha_i, \xi_i)$$
$$p_j \sim \text{Beta}(\alpha_j, \xi_j)$$
$$p_{ij} \sim \text{Beta}(\alpha_{ij}, \xi_{ij}).$$

Then, the expectation formula of p_i, p_j and p_{ij} from prior distributions gives us

$$E(p_i) = \frac{\alpha_i}{\alpha_i + \xi_i}$$
$$E(p_j) = \frac{\alpha_j}{\alpha_j + \xi_j}$$
$$E(p_{ij}) = \frac{\alpha_{ij}}{\alpha_{ij} + \xi_{ij}}.$$

Then, we can calculate the posterior distribution of the three probabilities with prior distribution and counts shown in Table 1.1. Because likelihood function is Binomial, and Beta is a conjugate prior distribution, the posterior distributions of p_i, p_j and p_{ij} are

$$p_i|y_{i\cdot}, y_{\cdot\cdot} \sim \text{Beta}(\alpha_i + y_{i\cdot}, \xi_i + y_{\cdot\cdot} - y_i)$$
$$p_j|y_{\cdot j}, y_{\cdot\cdot} \sim \text{Beta}(\alpha_j + y_{\cdot j}, \xi_j + y_{\cdot\cdot} - y_{\cdot j})$$
$$p_{ij}|y_{ij}, y_{\cdot\cdot} \sim \text{Beta}(\alpha_{ij} + y_{ij}, \xi_{ij} + y_{\cdot\cdot} - y_{ij}).$$

Again, expectation formula of Beta distribution gives us

$$E(p_i|y_{i\cdot}, y_{\cdot\cdot}) = \frac{\alpha_i + y_{i\cdot}}{\alpha_i + \xi_i + y_{\cdot\cdot}}$$
$$E(p_j|y_{\cdot j}, y_{\cdot\cdot}) = \frac{\alpha_j + y_{\cdot j}}{\alpha_j + \xi_j + y_{\cdot\cdot}}$$
$$E(p_{ij}|y_{\cdot\cdot}, y_{\cdot\cdot}) = \frac{\alpha_{ij} + y_{ij}}{\alpha_{ij} + \xi_{ij} + y_{\cdot\cdot}}.$$

By the properties of Beta distribution, the variances are

$$\begin{aligned}
\mathrm{Var}(p_i|y_{i\cdot}, y_{\cdot\cdot}) &= \frac{(\alpha_i + y_{i\cdot})(\xi_i + y_{\cdot\cdot} - y_{i\cdot})}{(\alpha_i + \xi_i + y_{\cdot\cdot})^2(\alpha_i + \xi_i + y_{\cdot\cdot} + 1)} \\
\mathrm{Var}(p_j|y_{\cdot j}, y_{\cdot\cdot}) &= \frac{(\alpha_j + y_{\cdot j})(\xi_j + y_{\cdot\cdot} - y_{\cdot j})}{(\alpha_j + \xi_j + y_{\cdot\cdot})(\alpha_j + \xi_j + y_{\cdot\cdot} + 1)} \\
\mathrm{Var}(p_{ij}|y_{\cdot\cdot}, y_{\cdot\cdot}) &= \frac{(\alpha_{ij} + y_{\cdot\cdot})(\xi_{ij} + y_{\cdot\cdot} - y_{\cdot\cdot})}{(\alpha_{ij} + \xi_{ij} + y_{\cdot\cdot})(\alpha_{ij} + \xi_{ij} + y_{\cdot\cdot} + 1)}.
\end{aligned}$$

By Taylor series, with updated parameters, delta method and change of base formula (Huang et al., 2014),

$$\begin{aligned}
E(IC_{ij}|y_{\cdot\cdot}, y_{\cdot\cdot}, y_{i\cdot}, y_{\cdot j}) &= \log_2 \frac{E(p_{ij}|y_{\cdot\cdot}, y_{\cdot\cdot})}{E(p_i|y_{i\cdot}, y_{\cdot\cdot})E(p_j|y_{\cdot j}, y_{\cdot\cdot})} \\
&= \frac{(\alpha_{ij} + y_{\cdot\cdot})(\alpha_i + \xi_i + y_{\cdot\cdot})(\alpha_j + \xi_j + y_{\cdot\cdot})}{(\alpha_{ij} + \xi_{ij} + y_{\cdot\cdot})(\alpha_i + y_{i\cdot})(\alpha_j + y_{\cdot j})}.
\end{aligned}$$

Then, the norm approximation of variances of IC is (Bate et al., 1998)

$$\begin{aligned}
\mathrm{Var}(IC_{ij}|y_{\cdot\cdot}, y_{\cdot\cdot}, y_{i\cdot}, y_{\cdot j}) &= \frac{1}{(\log 2)^2}\left\{\frac{\mathrm{Var}(p_{ij} - E(p_{ij}))}{[E(p_{ij})]^2} - \frac{\mathrm{Var}(p_i - E(p_i))}{[E(p_i)]^2} - \frac{\mathrm{Var}(p_j - E(p_j))}{[E(p_j)]^2}\right\} \\
&= \frac{1}{(\log 2)^2}\left\{\frac{\mathrm{Var}(p_{ij} - E(p_{ij}))}{[E(p_{ij})]^2} - \frac{\mathrm{Var}(p_i - E(p_i))}{[E(p_i)]^2} - \frac{\mathrm{Var}(p_j - E(p_j))}{[E(p_j)]^2}\right\}.
\end{aligned}$$

As more data become available, $\mathrm{Var}(IC)$ becomes smaller, the confidence interval of IC becomes narrower and the estimate of IC becomes more accurate (Gibbons and Amatya, 2016). Norén et al. (2006) extended BCPNN method to more complex quantitative associations by assuming $y_{\cdot\cdot}$, $y_{i\cdot} - y_{ij}$, $y_{\cdot j} - y_{ij}$, $y_{\cdot\cdot} - y_{\cdot j} - y_{i\cdot} + y_{ij}$ follows multinomial distribution with probability $p_{ij}, p_{i0}, p_{0j}, p_{00}$.

PRR and BCPNN are available in R package "PhViD" (Ahmed and Poncet, 2016).

2.2.3 Likelihood ratio test based method (LRT) for signal detection

Huang et al. (2011, 2017) proposed LRT-based methods for signal detection for a single drug by assuming y_{ij} follows Poisson distribution or Zero-inflated Poisson distribution. Thus, maximum likelihood ratio (MLR) test statistics can be established

based on the assumption. A Monte Carlo procedure under the null hypothesis is conducted to search the threshold for the test.

Likelihood ratio test based method (LRT)

LRT method (Huang et al., 2011) is a method based on the 2×2 table (shown in Table 2.1) by assuming

$$y_{ij} \sim \text{Poisson}(y_{i\cdot} \times p_i)$$
$$(y_{\cdot j} - y_{ij}) \sim \text{Poisson}((y_{\cdot\cdot} - y_{i\cdot}) \times q_i),$$

where p_i is the reporting rate of the jth drug for ith AE, and q_i is the reporting rate of the jth drug for any other AEs combined excluding the ith AE in the database. Then, if $p_i > q_i$, AE i is a positive signal of drug j. The test is $H_0 : p_i = q_i = p_0$ vs $H_1 : p_i > q_i$. This leads to the likelihood ratio for ith AE and jth drug as

$$LR_{ij} = \frac{L_a(\hat{p}_i, \hat{q}_i)}{L_0(\hat{p}_0)} = \left(\frac{y_{ij}}{y_{i\cdot}}\right)^{y_{ij}} \left(\frac{y_{\cdot j} - y_{ij}}{y_{\cdot\cdot} - y_{i\cdot}}\right)^{(y_{\cdot j} - y_{ij})} / \left[\left(\frac{y_{\cdot j}}{y_{\cdot\cdot}}\right)^{y_{\cdot j}}\right], \; i = 1, ..., I,$$

where $\hat{p}_i$ and $\hat{q}_i$ are the maximum likelihood estimates (MLEs) of p_i and q_i under alternative hypothesis, and $\hat{p}_0$ is the MLE of p_0 under null hypothesis.

$$MLR = \max_i (LR_{ij} I(\hat{p}_i > \hat{q}_i)) = \max_i \left(\frac{y_{ij}}{y_{i\cdot}}\right)^{y_{ij}} \left(\frac{y_{\cdot j} - y_{ij}}{y_{\cdot\cdot} - y_{i\cdot}}\right)^{(y_{\cdot j} - y_{ij})} / \left[\left(\frac{y_{\cdot j}}{y_{\cdot\cdot}}\right)^{y_{\cdot j}}\right]$$

A Monte Carlo procedure is used to get the distribution of MLR under H_0, and find the upper 5th percentile point $MLR_{0.05}$ by the procedure. A step-down process is used to control the inflated type I errors of multiple comparisons: Compare LR_{ij} from the observed dataset from the largest one with $MLR_{0.05}$.

Zero-inflated Poisson (ZIP) LRT

ZIP LRT (Huang et al., 2017) is developed to handle the excessive zero counts in signal detection by assuming y_{ij} follows ZIP, $y_{ij} \sim \text{ZIP}(w_j, p_{ij})$, as

$$y_{ij} \sim \begin{cases} 0 & \text{with probability } w_j \\ \text{Poisson}(y_{i\cdot}p_{ij}) & \text{with probability}(1-w_j) \end{cases},$$

w_j is the probability of observing a true zero.

$$(y_{\cdot j} - y_{ij}) \sim \text{Poisson}((y_{\cdot\cdot} - y_{i\cdot}) \times q_i).$$

Define a latent random variables.

$$z_i \sim \begin{cases} 1 & \text{when } y_{ij} \text{ is a true zero} \\ 0 & \text{when } y_{ij} \text{ is from Poisson}(y_{i\cdot}p_{ij}) \end{cases},$$

and apply the EM algorithm to find the MLEs of (w, p_i, q_i) under H_0 and H_1, respectively. Similarly,

$$\begin{aligned} \tilde{LR}_{ij} &= \frac{L_a(\tilde{p}_i, \tilde{q}_i)}{L_0(\hat{w}, \hat{z}_{i0}, \hat{p}_0)} \\ &= [\hat{w}^{\hat{z}_{i0}}(1-\hat{w})^{(1-\hat{z}_{i0})}]^{(-1)}[e^{-y_{i\cdot}\hat{p}_0}(y_{i\cdot}\hat{p}_0)^{y_{ij}}/y_{ij}!]^{\hat{z}_{i0}} \\ &\times e^{(y_{\cdot\cdot}\hat{p}_0 - y_{\cdot j})}\left(\frac{y_{ij}}{y_{i\cdot}}\right)^{y_{ij}}\left(\frac{y_{\cdot j}-y_{ij}}{y_{\cdot\cdot}-y_{i\cdot}}\right)^{(y_{\cdot j}-y_{ij})}/\hat{p}_0^{y_{\cdot j}}. \qquad i = 1, ..., I. \end{aligned}$$

Let $\log \tilde{LR}_{ij} = \log LR_{ij} = 0$ if $y_{ij} = 0$. Otherwise,

$$\tilde{LR}_{ij} = (1-\hat{w})^{(-1)}e^{y_{\cdot\cdot}\hat{p}_0 - y_{\cdot j}}\left[\left(\frac{y_{\cdot j}}{y_{\cdot\cdot}}\right)/\hat{p}_0\right]^{y_{\cdot j}} LR_{ij},$$

$$MLR = \max_i(LR_{ij}I(\hat{p}_i > \hat{q}_i)) = \max_i \left(\frac{y_{ij}}{y_{i\cdot}}\right)^{y_{ij}}\left(\frac{y_{\cdot j}-y_{ij}}{y_{\cdot\cdot}-y_{i\cdot}}\right)^{(y_{\cdot j}-y_{ij})} / \left[\left(\frac{y_{\cdot j}}{y_{\cdot\cdot}}\right)^{y_{\cdot j}}\right],$$

then, like the procedures in LRT, a Monte Carlo procedure is used to get the distribution of MLR under H_0, and find the upper 5th percentile point $MLR_{0.05}$ by

the procedure. A step-down process is used to control the inflated type I errors of multiple comparisons: Compare LR_{ij} from the observed dataset from the largest one with $MLR_{0.05}$.

LRT stratified analysis

Huang et al. (2011, 2017) reparametrize the likelihood ratio to incorporate demographic information such as gender and age. LRT adjusts small number of categorical covariates by calculating the expected counts from each stratum. The expected count in stratum s is

$$E_{ij}^{(s)} = y_{i\cdot}^{s} \times \frac{y_{i\cdot}^{s}}{y_{\cdot\cdot}^{s}} \quad s = 1, ..., S. \tag{2.2}$$

The reformulated expected count is

$$E_{ij} = \sum_{s} E_{ij}^{(s)} = \sum_{s} \left(y_{i\cdot}^{s} \times \frac{y_{i\cdot}^{s}}{y_{\cdot\cdot}^{s}} \right),$$

then

$$LR_{ij}^{(s)} = \left(\frac{y_{ij}^{(s)}}{E_{ij}^{(s)}} \right)^{y_{ij}^{(s)}} \left(\frac{y_{\cdot j}^{(s)} - y_{ij}^{(s)}}{y_{\cdot j}^{(s)} - E_{ij}^{(s)}} \right)^{(y_{\cdot j}^{(s)} - y_{ij}^{(s)})},$$

then

$$L\tilde{L}R_{ij} = \log(\tilde{LR}_{ij}) = \sum_{s=1}^{S} \frac{y_{\cdot j}^{(s)}}{y_{\cdot j}} \log(LR_{ij}^{(s)}). \tag{2.3}$$

With (2.3),

$$M\tilde{L}LR = \max_{i}(L\tilde{L}R_{ij}).$$

Then, Monte Carlo samples generates the null distribution of $M\tilde{L}LR$.

Stratified ZIP LRT. To incorporate the information of categorical information, first define

$$U_0 = \frac{\sum_{i=1}^{I}[(1 - \hat{z}_{i0})y_{ij} + (y_{\cdot j} - y_{ij})]}{\sum_{i=1}^{I}[(1 - \hat{z}_{i0})E_{ij} + (y_{\cdot j} - E_{ij})]} = y_{\cdot j} / \left[y_{\cdot j} - \left(\sum_{i=1}^{I} [\hat{z}_{i0} E_{ij}] \right) / I \right].$$

Recall (2.2.3), then rewrite the log likelihood ratio statistic for non-zero counts as

$$\log(\tilde{LR}_{ij}) = -\log(1-\hat{w}) + \log\left[\frac{e^{y_{\cdot j}}(1-U_0)}{U_0^{y_{\cdot j}}}\right] + y_{ij}\log\left(\frac{y_{ij}}{E_{ij}}\right) + (y_{\cdot j} - y_{ij})\log\left(\frac{y_{\cdot j} - y_{ij}}{y_{\cdot j} - E_{ij}}\right),$$

$\log(\tilde{LR}_{ij}^{(s)})$ is the value of $\log(\tilde{LR}_{ij})$ in the stratum $s (s = 1, ..., S)$ obtained by replacing $E_{(ij)}$ andw_{ij} with $E_{(ij)}^{(s)}$ and$w_{ij}^{(s)}$. Calculation of $E_{(ij)}^{(s)}$ is shown in (2.2). Then,

$$\log\left(\tilde{LR}_{ij}^{(S)}\right) = \sum_{s=1}^{S} \frac{y_{\cdot\cdot}^{(s)}}{y_{\cdot\cdot}} \log\left(\tilde{LR}_{ij}^{(s)}\right).$$

Finally,

$$\tilde{MLR} = \max_i \left[\log\left(\tilde{LR}_{ij}^{(S)}\right)\right].$$

2.3 The proposed statistical model

To investigate J drugs or biologics (e.g., adriamycin, remdesivir, COVID-19 vaccines) and their corresponding I AEs (e.g., hives, headache, fatigue, etc.) in any surveillance system datasets, we assume there is y_{ij} reported occurrences of the i-th AE for the j-th drug and the observed data is a $I \times J$ table of counts y_{ij}'s. For each fixed drug j, y_{ij}'s can be viewed as the number of type i events in the "time" interval $(0, t_i]$. Note that we conceptually call t_i a time interval, which actually could be any variable that has a positive correlation with the response variable such as $t_i = \sum_{j=1}^{J} y_{ij} / \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}$. Throughout this Chapter, we will keep using drugs generically to refer to any agents, either biologics or drugs.

Let $\boldsymbol{y}_j$ be the length k_j vector of all the non-zero counts of $\{y_{ij} : i = 1, ..., I\}$, and $\boldsymbol{x}_j$ be the corresponding covariates (e.g. age, gender) where the order of the components in $\boldsymbol{y}_j$ match that in $\boldsymbol{x}_j$. It is well recognized that there are too many zero counts in y_{ij}'s to follow a Poisson distribution. Typically the non-zero counts $\boldsymbol{y}_j$'s are assumed to be from a Poisson distribution. For the zero counts, let δ_{ij} be the latent variable indicating if the zero is from the Poisson distribution or not, and

$\delta_{ij} = 1$ if the zero count is from the same Poisson distribution as the $\boldsymbol{y}_j$'s, otherwise, $\delta_{ij} = 0$. For fixed j, denote $\boldsymbol{\delta}_j$ be the vector of all the δ_{ij}'s, thus $\dim(\boldsymbol{\delta}_j) = I - k_j$. Existing approaches have assumed a homogeneous Poisson model for the AE rates. This assumption is questionable due to the excessive zero counts. In addition, the ZIP model can not handle this issue since Poisson distribution can not model $y_{.j} - y_{ij}$ well, which are only at high values around 10,000.

Assuming all AEs are independent, we use the following non-homogeneous Poisson process $y_{ij} = y_{ij}(t_i)$ for counts in "time" interval $(0, t_i]$ with conditional mean

$$E[y_{ij}(t_i)|\boldsymbol{x}_j] = G(t_i)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_{ij}),$$

where $\boldsymbol{x}_{ij} = (x_{ij1}, ..., x_{ijd}) \in R^d$, $y_{ij}(t_i)$ is the event count in $(0, t_i]$, and $G(\cdot) \in \mathcal{G} = \{$ monotone increasing functions on $[0, 1]$ with $0 \leq G(t_i) \leq 1.\}$. The constraint $0 \leq G(t_i) \leq 1$ is for the identifiability of the model. In this model, $G(t_i)$ is the non-parametric component to capture the background occurrence of AEs in the panel, which varies with different AEs. Denote $d = \dim(\boldsymbol{\beta})$.

2.4 Estimation procedure

To derive the algorithm to estimate the proposed model, we first construct a joint model under the "complete data" by assuming $\boldsymbol{\delta}_j$ is observable, and $D_{I,J} = \{(t_i, k_j, y_{ij}, \boldsymbol{x}_j, \boldsymbol{\delta}_j) : i = 1, ..., I; j = 1, ..., J\}$. Without loss of generality, we assume that the first k_j cells are non-zeros for a fixed j. Then conditioning on $(t_i, k_j, \boldsymbol{x}_j, \boldsymbol{\delta}_j)$, the mass function of $\boldsymbol{y}_j$ is

$$\begin{aligned} p(y_{ij}|t_i, k_j, \boldsymbol{x}, \boldsymbol{\delta}_j; \boldsymbol{\beta}, G) &= \left\{ \prod_{i=1}^{k_j} \frac{[G(t_i)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_{ij})]^{y_{ij}}}{y_{ij}!} \right\} \\ &\times \exp\left\{ - \left[\sum_{i=1}^{k_j} G(t_i)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_{ij}) + \sum_{i=k_j+1}^{I} G(t_i)\delta_{ij}\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_{ij}) \right] \right\}. \end{aligned} \tag{2.4}$$

When $\delta_{ij} = 1$, the zero is from the same Poisson distribution with non-zero cell counts, the mass function is directly from a Poisson distribution with a conditional mean $E[y_{ij}(t_i)|\boldsymbol{x}_j] = G(t_i)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_{ij})$. Otherwise, $\delta_{ij} = 0$ and the zero counts are true zeros, we ignore true zeros in the mass function.

Let $\boldsymbol{y} = (y_1, ..., y_I)'$ and $(t, k, \boldsymbol{y}, \boldsymbol{x})$ be an i.i.d. copy of the data, then the likelihood-based on the complete data $D_{I,J}$ is

$$\begin{aligned} L_{I,J}(\boldsymbol{\beta}, G) = &\prod_{j=1}^{J} \left\{ \prod_{i=1}^{k_j} \frac{[G(t_i)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_{ij})]^{y_{ij}}}{y_{ij}!} \right\} \\ &\times \exp\left\{ - \left[\sum_{i=1}^{k_j} G(t_i)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_{ij}) + \sum_{i=k_j+1}^{I} G(t_i)\delta_{ij}\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_{ij}) \right] \right\}. \end{aligned} \tag{2.5}$$

Recall $y_{\cdot j} = \sum_{i=1}^{k_j} y_{ij}$, then the log-likelihood is, omitting some constant independent of our parameter of interest,

$$\begin{aligned} \ell_{I,J}(\boldsymbol{\beta}, G) = &\sum_{j=1}^{J}\sum_{i=1}^{k_j} y_{ij}\log G(t_i) + \sum_{j=1}^{J}\sum_{i=1}^{k_j} y_{ij}(\boldsymbol{\beta}^\intercal \boldsymbol{x}_{ij}) \\ &- \sum_{j=1}^{J} \left[\sum_{i=1}^{k_j} G(t_i)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_{ij}) + \sum_{i=k_j+1}^{I} G(t_i)\delta_{ij}\exp\left(\boldsymbol{\beta}^\intercal \boldsymbol{x}_{ij}\right) \right]. \end{aligned} \tag{2.6}$$

Denote $\boldsymbol{B}$ for the space for $\boldsymbol{\beta}$, and we estimate $(\boldsymbol{\beta}, G)$ by the semi-parametric maximum likelihood estimator $(\hat{\boldsymbol{\beta}}, \hat{G})$

$$(\hat{\boldsymbol{\beta}}, \hat{G}) = \arg \max_{(\beta, G)\in(B, \mathcal{G})} \ell_{I,J}(\boldsymbol{\beta}, G).$$

However, $(\hat{\boldsymbol{\beta}}, \hat{G})$ can not be directly computed as $\{\boldsymbol{\delta}_j : j = 1, ..., J\}$ are missing, so we use an iterative version of the EM-algorithm (Dempster et al., 1977) to handle it as below.

2.4.1 Model estimation algorithm

Given starting value $(\boldsymbol{\beta}^{(0)}, G^{(0)})$, we compute the next step estimate $(\boldsymbol{\beta}^{(1)}, G^{(1)}(\cdot))$. Generally, given $(\boldsymbol{\beta}^{(r)}, G^{(r)})$, $(\boldsymbol{\beta}^{(r+1)}, G^{(r+1)})$ is updated in the M-step and E-step as follows.

E-step

Let $D^0_{I,J}$ be $D_{I,J}$ without the $\boldsymbol{\delta}_j$'s, which is the set of observed data. Given $(\boldsymbol{\beta}^{(r)}, G^{(r)})$,

$$H_{I,J}\left(\boldsymbol{\beta}, G|\boldsymbol{\beta}^{(r)}, G^{(r)}\right) = E_{(\beta^{(r)},\lambda,G^{(r)})}\left[\ell_{I,J}(\boldsymbol{\beta}, G)\Big|D^0_{I,J}, \boldsymbol{\beta}^{(r)}, G^{(r)}\right],$$

where the expectation is with respect to the missing $\boldsymbol{\delta}_j$'s conditioning on the observed data, as if $(\boldsymbol{\beta}^{(r)}, G^{(r)})$ were the true parameters. By (2.6),

$$\begin{aligned} H_{I,J}\left(\boldsymbol{\beta}, G|\boldsymbol{\beta}^{(r)}, G^{(r)}\right) &= \sum_{j=1}^{J}\sum_{i=1}^{k_j} y_{ij}\log G(t_i) + \sum_{j=1}^{J}\sum_{i=1}^{k_j} y_{ij}(\boldsymbol{\beta}^{\intercal}\boldsymbol{x}_{ij}) \\ &- \sum_{j=1}^{J}\left[\sum_{i=1}^{k_j} G(t_i)\exp(\boldsymbol{\beta}^{\intercal}\boldsymbol{x}_{ij}) + \sum_{i=k_j+1}^{I} E(\delta_{ij}|D^0_{I,J}, \boldsymbol{\beta}^{(r)}, G^{(r)})G(t_i)\exp(\boldsymbol{\beta}^{\intercal}\boldsymbol{x}_{ij})\right]. \end{aligned} \tag{2.7}$$

To compute $E(\delta_{ij}|D^0_{I,J}, \boldsymbol{\beta}^{(r)}, G^{(r)}) := \delta^{(r)}_{ij}$, let m_j be the expected number of zero counts for the observed k_j non-zero counts in the same Poisson(λ) distribution, then $k_j + m_j$ is the total number of observations in the Poisson experiment, so the number of expected zero counts $m_j \approx (k_j + m_j)\exp(-\lambda)$, or $m_j \approx k_j\exp(-\lambda)/\big(1-\exp(-\lambda)\big)$, and $E(\delta_{ij}|\lambda) = m_j/(I-k_j) \approx k_j\exp(-\lambda)/[(I-k_j)\big(1-\exp(-\lambda)\big)]$. However, in our case, $\exp(-\lambda)$ changes with individuals due to the varying covariates, so we use the

following value to approximate it,

$$\begin{aligned}\exp(-\lambda_j^{(r)}) &:= \frac{k_j}{k_j+m_j}\frac{1}{k_j}\sum_{h=1}^{k_j}\exp\left\{-G^{(r)}(t_h)\exp\left[\boldsymbol{\beta}^{(r)T}\boldsymbol{x}_{hj}\right]\right\} \\ &\quad + \frac{m_j}{k_j+m_j}\frac{1}{I-k_j}\sum_{s=1}^{I-k_j}\exp\left\{-G^{(r)}(t_s)\exp\left[\boldsymbol{\beta}^{(r)T}\boldsymbol{x}_{sj}\right]\right\} \\ &:= \frac{k_j}{k_j+m_j}A_j^{(r)} + \frac{m_j}{k_j+m_j}B_j^{(r)} = \left[1-\exp(-\lambda_j^{(r)})\right]A_j^{(r)} + \exp(-\lambda_j)B_j^{(r)}.\end{aligned} \tag{2.8}$$

The above (2.8) gives

$$\exp\left(-\lambda_j^{(r)}\right) = A_j^{(r)}/\left(1+A_j^{(r)}-B_j^{(r)}\right),$$

and

$$\begin{aligned}\delta_{ij}^{(r)} &= E(\delta_{sj}|D_{I,J}^0,\boldsymbol{\beta}^{(r)},G^{(r)}) \\ &= k_j\exp\left(-\lambda_j^{(r)}\right)/\left\{(I-k_j)\left[1-\exp\left(-\lambda_j^{(r)}\right)\right]\right\} \\ &= \frac{k_j}{I-k_j}\frac{A_j^{(r)}}{1-B_j^{(r)}}.\end{aligned} \tag{2.9}$$

Plugging the $\delta_{ij}^{(r)}$'s (see (2.9)) into the above-expected log-likelihood (2.7), we get

$$\begin{aligned}H_{I,J}\left(\boldsymbol{\beta},G|\boldsymbol{\beta}^{(r)},G^{(r)}\right) &= \sum_{j=1}^{J}\sum_{i=1}^{k_j}y_{ij}\log G(t_i) + \sum_{j=1}^{J}\sum_{i=1}^{k_j}y_{ij}(\boldsymbol{\beta}^{\intercal}\boldsymbol{x}_{ij}) \\ &\quad - \sum_{j=1}^{J}\left[\sum_{i=1}^{k_j}G(t_i)\exp(\boldsymbol{\beta}^{\intercal}\boldsymbol{x}_{ij}) + \sum_{i=k_j+1}^{I}\delta_{ij}^{(r)}G(t_i)\exp(\boldsymbol{\beta}^{\intercal}\boldsymbol{x}_{ij})\right] \\ &= \sum_{j=1}^{J}\sum_{i=1}^{I}\log G(t_i)\sum_{j=1}^{J}y_{ij} + \sum_{j=1}^{J}\sum_{j=1}^{J}y_{ij}(\boldsymbol{\beta}^{\intercal}\boldsymbol{x}_{ij}) \\ &\quad - \sum_{i=1}^{I}G(t_i)\sum_{j=1}^{J}\left[\exp(\boldsymbol{\beta}^{\intercal}\boldsymbol{x}_{ij})I(y_{ij}\neq 0) + \delta_{ij}^{(r)}\exp(\boldsymbol{\beta}^{\intercal}\boldsymbol{x}_{ij})I(y_{ij}=0)\right].\end{aligned}$$

M-step

M-step computes

$$(\boldsymbol{\beta}^{(r+1)}, G^{(r+1)}) = \arg \max_{(\beta,G)\in(\boldsymbol{B},\mathcal{G})} H_{I,J}(\boldsymbol{\beta}, G|\boldsymbol{\beta}^{(r)}, G^{(r)}).$$

In the above, the maximization over G is non-trivial. Our idea is to convert the complex optimization problem to a sequence of isotonic regression procedures. Such approach has proved successful in semi-parametric models for subgroup analysis (Yuan et al., 2020). Here our model is more complicated, and we describe the algorithm below.

2.4.2 Computation of $G^{(r+1)}(\cdot)$

In particular, we can first fix $\boldsymbol{\beta}^{(r)}$, and maximize over $G(\cdot)$ to get $G^{(r+1)}(\cdot)$, which is of the form, for some $d_i^{(r)}$'s,

$$G^{(r+1)} = \arg\max_{G\in\mathcal{G}} \sum_{i=1}^{I} \left[y_{i\cdot} \log G(t_i) - d_i^{(r)} G(t_i) \right],$$

where $d_i^{(r)} = \sum_{j=1}^{J} \left[\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_{ij}) I(y_{ij} \neq 0) + \delta_{ij}^{(r)} \exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_{ij}) I(y_{ij} = 0) \right]$. For this, without loss of generality, we assume the t_i's are arranged in increasing order, re-write the above as, with $h_i^{(r)} = y_{i\cdot}/d_i^{(r)}$,

$$\begin{aligned} G^{(r+1)} &= \arg\min_{G\in\mathcal{G}} \sum_{i=1}^{I} \left[-h_i^{(r)} \log G(t_i) + G(t_i) \right] d_i^{(r)}. \\ &= \arg\min_{G\in\mathcal{G}} \sum_{i=1}^{I} \left[h_j^{(r)} \log h_i^{(r)} - h_i^{(r)} \log G(t_i) - h_i^{(r)} + G(t_i) \right] d_i^{(r)} \qquad (2.10) \\ &= \arg\min_{G\in\mathcal{G}} \sum_{i=1}^{I} \Delta_\Phi \left[h_i^{(r)}, G(t_i) \right] d_i^{(r)}, \end{aligned}$$

where $\Delta_\Phi(u, v) = \Phi(u) - \Phi(v) - (u - v)\phi(v) = u \log u - v \log v - (u - v)(\log v + 1) = u \log u - u \log v - u + v$, and $\Phi(u) = u \log u, u \in R^+$. The first derivative of Φ is given

by $\phi(u) = \log u + 1$. $\Phi(\cdot)$ is convex on R^+, so by Theorem 1.5.1 and Example 1.5.1 in Robertson et al. (1988), the above minimization (2.10) is written as

$$G^{(r+1)} = \arg\min_{G \in \mathcal{G}} \sum_{i=1}^{I} d_i^{(r)} \left[h_i^{(r)} - G(t_i) \right]^2. \tag{2.11}$$

Then (2.11) is converted into an isotonic regression problem, and can be computed using the R-package isotonic (Best and Chakravarti, 1990; de Leeuw et al., 2009). Then, we fix $G^{(r+1)}(\cdot)$, and maximize over $\boldsymbol{\beta}$ to get $\boldsymbol{\beta}^{(r+1)}$. The iteration goes on until convergence of the sequence, and the final values are treated as the MLE $(\hat{\boldsymbol{\beta}}, \hat{G})$.

The convergence of the above iterative procedure is implied by the observation:

$$H_{I,J}(\boldsymbol{\beta}^{(r)}, G^{(r)} | \boldsymbol{\beta}^{(r)}, G^{(r)}) \leq H_{I,J}(\boldsymbol{\beta}^{(r)}, G^{(r+1)} | \boldsymbol{\beta}^{(r)}, G^{(r)}) \leq H_{I,J}(\boldsymbol{\beta}^{(r+1)}, G^{(r)} | \boldsymbol{\beta}^{(r)}, G^{(r)}),$$

and the ascending property of the EM algorithm. So the sequence $\{(\boldsymbol{\beta}^{(r)}, G^{(r)}\}$ will converges to at least some local maxima, and multiple starting points may be required to locate the global maxima, just like the EM algorithm without iteration.

2.5 Hypothesis testing for signal detection

In the surveillance problem to investigate safety signals in the $I \times J$ AE-drug combinations, let $\lambda_{ij} = E(y_{ij})$. Our interest is to test the null hypothesis $H_0 : y_{ij}$ is not a signal under Poisson $(i = 1, ..., I; j = 1, ..., J)$. vs H_1: Some of y_{ij} are signals. Recall δ_{ij} is a latent variable indicating if the zero-count is from the Poisson distribution or not, and $\hat{\delta}_{ij}$ is the conditional expectation of δ_{ij} given in the EM algorithm in §3.

For a given i-th AE, denote $\boldsymbol{x}_i = (x_{i1}, ..., x_{iJ})^\intercal$, $\boldsymbol{y}_i = (y_{i1}, ..., y_{iJ})^\intercal$ and $\hat{\boldsymbol{\delta}}_i = (\hat{\delta}_{i1}, ..., \hat{\delta}_{iJ})^\intercal$. By (2.4) and (2.5), conditioning on $(t_i, \boldsymbol{x}_i)$, without loss of generality, we assume the first v_i cells are non-zeros for a fixed i, under null hypothesis, the mass

function $\boldsymbol{y}_i$ is

$$p(\boldsymbol{y}_i|t_i, k_j, \boldsymbol{x}, \hat{\boldsymbol{\delta}}_j; \hat{\boldsymbol{\beta}}, \hat{G}) = \left\{ \prod_{j=1}^{J} \frac{\left[\hat{G}(t_i) \exp\left(\hat{\boldsymbol{\beta}}^\intercal \boldsymbol{x}_i\right)\right]^{y_{ij}}}{y_{ij}!} \right\}$$
$$\times \exp\left\{ - \left[\sum_{j=1}^{v_i} \hat{G}(t_i) \exp\left(\hat{\boldsymbol{\beta}}^\intercal \boldsymbol{x}_{ij}\right) + \sum_{j=v_i+1}^{J} \hat{G}(t_i)\hat{\delta}_{ij} \exp\left(\hat{\boldsymbol{\beta}}^\intercal \boldsymbol{x}_{ij}\right) \right] \right\},$$

where $C = \exp\left\{ - \left[\sum_{j=1}^{v_i} \hat{G}(t_i) \exp\left(\hat{\boldsymbol{\beta}}^\intercal \boldsymbol{x}_{ij}\right) + \sum_{j=v_i+1}^{J} \hat{G}(b_i)\hat{\delta}_{ij} \exp\left(\hat{\boldsymbol{\beta}}^\intercal \boldsymbol{x}_{ij}\right) \right] \right\}$ is a constant, so

$$p(\boldsymbol{y}_i|t_i, k_j, \boldsymbol{x}, \hat{\boldsymbol{\delta}}_j; \hat{\boldsymbol{\beta}}, \hat{G}) \propto \left\{ \prod_{j=1}^{J} \frac{\left[\hat{G}(t_i) \exp\left(\hat{\boldsymbol{\beta}}^\intercal \boldsymbol{x}_i\right)\right]^{y_{ij}}}{y_{ij}!} \right\}.$$

Thus, under null hypothesis, conditioning on $(t_i, \boldsymbol{x}_i, \hat{\boldsymbol{\beta}}, \hat{G}(t_i), \hat{\delta}_i)$,

$$y_{ij} \sim \text{Poisson}(\lambda_{ij}) \qquad \lambda_{ij} = \hat{G}(t_i) \exp\left(\hat{\boldsymbol{\beta}}^\intercal \boldsymbol{x}_{ij}\right).$$

And for given level of significance α, denote the $(1-\alpha)th$ upper quantile of Poisson ditribution as $Q(1-\alpha)$. If any $y_{ij} > Q(1-\alpha)$ for some (i, j), then H_0 is rejected, and there is signals in the $I \times J$ table.

Note that H_0 can be re-written as $H_0 = \cap_{i,j}^{I,J} H_{0ij}$, where H_{0ij}: y_{ij} is not a signal under Poisson for AE i - drug j combination. Various approaches for multiple comparison and significance level correction have been shown to improve the original Bonferroni correction, e.g., the Benjamini-Hochberg's correction (Benjamini and Hochberg, 1995) and the Holm-Bonferroni approach (Holm, 1979). The former is better for controlling the false discovery rate (FDR); the latter is better for controlling the family-wise error rate (FWER). We adopt the Holm-Bonferroni approach to control and decrease the number of false-positive signals because its adaptive nature. It is a sequential rejective step-down Bonferroni test, which controls the FWER in a strong sense with greater power than the Bonferroni test.

2.6 Simulation studies: Finite sample properties

We conduct simulation studies under various cases and scenarios to evaluate the finite sample properties of the statistical model, the performance of the proposed signal detection method, the associated parameter estimation in surveillance datasets and the robustness to proportions of true zeros and patterns of background noise. Datasets are generated with combinations of drug numbers J, AE types I, and proportions of zeros ρ to mimic datasets from the VigiBase database. Below we present the details of simulation settings and results of analyzing the simulated data in terms of estimation and signal detection.

2.6.1 Simulation settings

We consider different combinations of I, J, ρ to reflect the VigiBase datasets. Also, we evaluate the accuracy and precision of the estimates in different types of potential covariate information by setting different dimensions of $\boldsymbol{x}$. Moreover, we evaluate different background noise patterns by setting different shapes of $G(t)$. Design details of each simulation scenario, of sample size $N = 1000$, are as follows.

Case 1: $I = 200, J = 10, \rho \sim \text{uniform}(0.3, 0.5)$.

Scenario 1: $\boldsymbol{x}$ is sampled from a 3-dimensional multivariate normal distribution. $G(t)$ is a convex function.

Scenario 2: $\boldsymbol{x}$ is sampled from a 3-dimensional multivariate normal distribution. $G(t)$ is a concave function.

Scenario 3: $\boldsymbol{x}$ is sampled from a 3-dimensional multivariate normal distribution. $G(t)$ is a mixture of convex and concave functions: $G(t)$ is convex when $t \leq 0.5$ and $G(t)$ is concave when $t \geq 0.5$.

Scenario 4: $\boldsymbol{x}$ is sampled from a 3-dimensional multivariate normal distribution. $G(t)$ is a mixture of convex and concave functions: $G(t)$ is concave when $t \leq 0.5$ and $G(t)$ is convex when $t \geq 0.5$.

Denote $\boldsymbol{x} = (\boldsymbol{x}_1, x_2, x_3)^\intercal$, then

Scenario 5: $\boldsymbol{x}_1$ is sampled from a 2-dimensional multivariate normal distribution, x_2 is sample from a binomial distribution, and x_3 is randomly sampled from $(1, 2, 3, 4, 5, 6)$. $G(t)$ is a convex function.

Scenario 6: $\boldsymbol{x}_1$ is sampled from a 2-dimensional multivariate normal distribution, x_2 is sample from a binomial distribution, and x_3 is randomly sampled from $(1, 2, 3, 4, 5, 6)$. $G(t)$ is a concave function.

Scenario 7: $\boldsymbol{x}_1$ is sampled from a 2-dimensional multivariate normal distribution, x_2 is sample from a binomial distribution, and x_3 is randomly sampled from $(1, 2, 3, 4, 5, 6)$. $G(t)$ is a mixture of convex and concave functions: $G(t)$ is convex when $t \leq 0.5$ and $G(t)$ is concave when $t \geq 0.5$.

Scenario 8: $\boldsymbol{x}_1$ is sampled from a 2-dimensional multivariate normal distribution, x_2 is sample from a binomial distribution, and x_3 is randomly sampled from $(1, 2, 3, 4, 5, 6)$. $G(t)$ is a mixture of convex and concave functions: $G(t)$ is concave when $t \leq 0.5$ and $G(t)$ is convex when $t \geq 0.5$.

Scenario 9: $\boldsymbol{x}$ is sampled from a 5-dimensional multivariate normal distribution. $G(t)$ is a convex function.

Scenario 10: $\boldsymbol{x}$ is sampled from a 5-dimensional multivariate normal distribution. $G(t)$ is a concave function.

Scenario 11: $\boldsymbol{x}$ is sampled from a 5-dimensional multivariate normal distribution. $G(t)$ is a mixture of convex and concave functions: $G(t)$ is convex when $t \leq 0.5$ and $G(t)$ is concave when $t \geq 0.5$.

Scenario 12: $\boldsymbol{x}$ is sampled from a 5-dimensional multivariate normal distribution. $G(t)$ is a mixture of convex and concave functions: $G(t)$ is concave when $t \leq 0.5$ and $G(t)$ is convex when $t \geq 0.5$.

Case 2: $I = 5000, J = 10, \rho \sim \text{uniform}(0.2, 0.4)$. scenarios 1-12 are the same as case 1.

Case 3: $I = 1000, J = 100, \rho \sim \text{uniform}(0.3, 0.6)$. scenarios 1-12 are the same as case 1.

In data generation process, we generate true zero indicator z_{ij} from a binomial distribution. If z_{ij} is equal to 1, we set y_{ij} to be 0. Otherwise, we generate y_{ij} from a Poisson distribution with a conditional mean of $E(y_{ij}(t)|\boldsymbol{x}_j) = rr_{ij}G(t_i)\exp(\boldsymbol{\beta}^\intercal x_{ij})$. If the AE i on drug j is sampled as a signal, then $rr_{ij} = 1.5$ (or $2, 3, 4, 6$), if not, $rr_{ij} = 1$. We use different rr to represent AE signals with different strengths. We set different true numbers of signals at 0 (null hypothesis), 50, 100, 350, 700.

2.6.2 Performance evaluation measures

We evaluate the proposed method's accuracy in parameter estimation by bias which is defined as the average difference between the estimator and the true value. We consider the proposed method's precision by standard deviation. Also, we evaluate the proposed method's performance in signal detection by Type I error/Power, False Discovery Rate (FDR), Specificity, and Sensitivity.

Type I error/Power

We assume the null hypothesis is rejected if one or more signals are found in a simulated dataset. Then, type I error is defined as times of rejecting the null hypothesis in the N simulated datasets under the null hypothesis. Power is defined as times

of rejecting the null hypothesis in the N simulated datasets under the alternative hypothesis.

False discovery rate (FDR)

FDR is defined as the expected proportion of false-positive signals among all signals found. It is a common measure in multiple comparisons or searching analysis, and estimated as

$$FDR = \sum_{j=1}^{N} \frac{\text{number of false signals found in simulation j}}{\text{number of all signals found in simulation j}} / N.$$

Specificity

Specificity measures the proportion of negatives that are correctly identified, and is estimated as

$$Spe = \sum_{j=1}^{N} \frac{\text{number of true non-signals found in simulation j}}{\text{number of all non-signals in simulation j}} / N.$$

Sensitivity

Sensitivity measures the proportion of positives that are correctly identified, and is estimated by

$$Sen = \sum_{j=1}^{N} \frac{\text{number of true signals found in simulation j}}{\text{number of true signals in simulation j}} / N.$$

We compare the results from PRR, BCPNN, ZIP LRT, and the proposed approach, with and without incorporating covariates information (NHP) in the panel with signals generated with a higher than average conditional mean.

2.6.3 Simulation results

Simulation results: Estimation of parameters

Table 2.2 shows that bias and standard deviation of the proposed model under Case 1-3 when sample size, dimension of covariates $\boldsymbol{x}$ and shape of $G(t)$ varies. When the sample size is 200×10, bias and standard deviation increase as the dimension of covariates increases. Bias and the standard deviation are stable as the shape of $G(t)$ varies. Similar conclusion holds for other cases (1000×100 and 5000×10). In addition, the bias of the estimates and their standard deviations (sd) decrease as the sample size increases from 200×10 to 1000×100. In general, most of the results are stable except for some extreme cases with a significant bias up to 0.048 and standard deviation up to 0.128.

Figures 2.1, 2.2 and 2.3 present the estimates of $G(\cdot)$ within different cases. Solid blue lines are true shapes of $G(\cdot)$, and black step lines are estimates of $G(\cdot)$. Figures show that the lines represent the estimates of $G(\cdot)$ and $G(\cdot)$ almost overlap, which means the estimates are accurate and stable for different patterns of background noise $G(\cdot)$.

In general, the simulation results show that the estimates of parameters and non-parametric components with the proposed approach are accurate and precise under various background noise patterns and covariates types, which later form the basis of the signal detection process.

Table 2.2: Parameter estimation of the proposed model under Case 1-3 when sample size, dimension of $\boldsymbol{x}$ or shape of $G(t)$ varies.

	Shapes of $G(\cdot)$			
	Convex	Concave	$I_{t<0.5}$Convex+$I_{t\geq 0.5}$Concave	$I_{t<0.2}$Concave+$I_{t\geq 0.2}$Convex
	Case 1: ($I = 200; J = 10$)			
$\boldsymbol{\beta}_0$[i]	(2.300,3.500,1.800)	(1.200,2.900,0.600)	(2.400,2.300,1.600)	(1.700,2.100,1.200)
$\hat{\boldsymbol{\beta}}$[ii]	(2.301,3.505,1.793)	(1.200,2.900,0.600)	(2.403,2.306,1.596)	(1.701,2.102,1.198)
sd[iii]	[0.010,0.013,0.013]	[0.002,0.002,0.003]	[0.003,0.003,0.003]	[0.005,0.006,0.004]
$\boldsymbol{\beta}_0$[i]	(-0.600,3.300,2.800)	(-1.100,2.900,1.800)	(-2.400,3.300,2.100)	(-0.900,2.300,3.400)
$\hat{\boldsymbol{\beta}}$[ii]	(-0.585,3.305,2.772)	(-1.099,2.902,1.799)	(-2.363,3.326,2.075)	(-0.895,2.304,3.386)
sd[iii]	[0.019,0.029,0.031]	[0.004,0.005,0.005]	[0.040,0.035,0.039]	[0.009,0.013,0.013]
$\boldsymbol{\beta}_0$[i]	(-1.300,2.300,2.600)	(2.200,-2.300,3.600)	(2.500,-0.300,3.200)	(3.600,-0.400,4.100)
$\hat{\boldsymbol{\beta}}$[ii]	(-1.265,2.307,2.546)	(2.174,-2.141,3.441)	(2.496,-0.280,3.159)	(3.577,-0.389,4.058)
sd[iii]	[0.034,0.029,0.043]	[0.123,0.177,0.226]	[0.036,0.039,0.070]	[0.051,0.053,0.077]
$\boldsymbol{\beta}_0$[i]	(1.400,5.300,-2.600)	(1.200,3.700,-1.800)	(2.600,1.500,-2.100)	(1.700,4.300,-2.700)
$\hat{\boldsymbol{\beta}}$[ii]	(1.404,5.308,-2.601)	(1.200,3.700,-1.799)	(2.598,1.509,-2.097)	(1.701,4.302,-2.698)
sd[iii]	[0.011,0.021,0.012]	[0.002,0.003,0.003]	[0.010,0.009,0.010]	[0.004,0.006,0.006]
$\boldsymbol{\beta}_0$[i]	(2.300,-0.600,2.400,3.100)	(-1.200,2.900,0.700,1.100)	(-0.300,1.200,2.100,2.100)	(2.800,-1.300,2.700,1.400)
$\hat{\boldsymbol{\beta}}$[ii]	(2.284,-0.601,2.383,3.097)	(-1.201,2.900,0.700.1.100)	(-0.308,1.207,2.098,2.104)	(2.821,-1.323,2.696,1.399)
sd[iii]	[0.042,0.023,0.014,0.025]	[0.008,0.004,0.004,0.002]	[0.013,0.008,0.005,0.007]	[0.025,0.016,0.015,0.010]

[i] $\boldsymbol{\beta}_0$ is for the true value of $\boldsymbol{\beta}$

[ii] $\hat{\boldsymbol{\beta}}$ is for the estimates of $\boldsymbol{\beta}$

[iii] sd is for standard deviation of $\boldsymbol{\beta}$

Table 2.2. (Cont.)

Shapes of $G(\cdot)$				
	Convex	Concave	$I_{t<0.5}$Convex+$I_{t\geq 0.5}$Concave	$I_{t<0.2}$Concave+$I_{t\geq 0.2}$Convex
$\boldsymbol{\beta}_0$[i]	(-3.900,3.600,-1.300,2.100)	(3.200,2.900,3.100,-1.100)	(3.300,-1.400,3.700,2.800)	(-4.100,3.300,2.600,2.400)
$\hat{\boldsymbol{\beta}}$[ii]	(-3.913,3.610,-1.290,2.091)	(3.166,2.923,3.104,-1.096)	(3.301,-1.405,3.699,2.812)	(-4.101,3.301,2.600,2.401)
sd[iii]	[0.034,0.027,0.014,0.019]	[0.054,0.031,0.039,0.008]	[0.032,0.018,0.012,0.014]	[0.005,0.003,0.003,0.004]
$\boldsymbol{\beta}_0$[i]	(4.700,-1.900,5.500,3.100)	(1.000,1.600,-2.700,1.100)	(1.700,2.200,-1.100,0.900)	(-3.800,2.300,1.700,2.100)
$\hat{\boldsymbol{\beta}}$[ii]	(4.699,-1.901,5.520,3.100)	(0.944,1.604,-2.697,1.101)	(1.687,2.215,-1.097,0.902)	(-3.804,2.304,1.699,2.099)
sd[iii]	[0.078,0.041,0.114,0.029]	[0.021,0.012,0.006,0.004]	[0.028,0.016,0.011,0.005]	[0.008,0.005,0.003,0.006]
$\boldsymbol{\beta}_0$[i]	(0.700,2.900,-2.500,4.100)	(2.700,-1.200,2.600,0.900)	(3.500,3.700,6.700,-0.300)	(0.500,1.300,-0.700,2.300)
$\hat{\boldsymbol{\beta}}$[ii]	(0.688,2.902,-2.493,4,121)	(2.731,-1.233,2.604,0.901)	(3.394,3.722,6.880,-0.299)	(0.499,1.301,-0.700,2.301)
sd[iii]	[0.030,0.025,0.018,0.047]	[0.015,0.010,0.007,0.002]	[0.036,0.030,0.229,0.003]	[0.002,0.002,0.001,0.004]
$\boldsymbol{\beta}_0$[i]	(2.200,-1.300,-1.400,2.200,1.600)	(2.100,-1.600,-1.200,2.900,-0.800)	(2.700,-1.300,1.600,2.100,-0.700)	(3.100,-1.000,-1.400,0.300,2.300)
$\hat{\boldsymbol{\beta}}$[ii]	(2.214,-1.297,-1.412,2.216,1.614)	(2.108,-1.604,-1.204,2.912,-0.804)	(2.725,-1.292,1.606,2.128,-0.701)	(3.097,-0.999,-1.399,0.302,2.300)
sd[iii]	[0.034,0.024,0.025,0.027,0.027]	[0.105,0.193,0.017,0.083,0.123]	[0.048,0.040,0.025,0.031,0.027]	[0.018,0.011,0.009,0.004,0.009]
$\boldsymbol{\beta}_0$[i]	(-3.100,4.300,2.400,1.200,-1.600)	(-2.100,2.700,3.100,1.900,-0.600)	(1.900,1.300,-3.600,0.500,2.300)	(-1.100,2.300,1.700,0.900,-0.300)
$\hat{\boldsymbol{\beta}}$[ii]	(-3.101,4.348,2.385,1.210,-1.608)	(-2.102,2.708,3.103,1.905,-0.600)	(1.913,1.316,-3.623,0.507,2.318)	(-1.100,2.304,1.699,0.901,-0.300)
sd[iii]	[0.071,0.093,0.048,0.029,0.046]	[0.010,0.013,0.009,0.007,0.006]	[0.018,0.026,0.026,0.009,0.021]	[0.004,0.008,0.003,0.003,0.003]
$\boldsymbol{\beta}_0$[i]	(3.600,-1.700,2.600,2.200,1.500)	(1.200,-1.500,0.300,2.400,-1.600)	(0.700,-2.100,1.000,1.500,1.300)	(-2.700,-0.700,-0.700,1.900,1.300)
$\hat{\boldsymbol{\beta}}$[ii]	(3.639,-1.703,2.609,2.230,1.530)	(1.202,-1.498,0.300,2.406,-1.603)	(0.709,-2.100,0.997,1.512,1.315)	(-2.701,-0.699,-0.702,1.903,1.303)
sd[iii]	[0.145,0.265,0.135,0.060,0.089]	[0.011,0.013,0.006,0.008,0.010]	[0.017,0.017,0.011,0.016,0.018]	[0.010,0.007,0.006,0.007,0.007]
$\boldsymbol{\beta}_0$[i]	(-2.700,2.900,1.400,2.100,-2.300)	(-1.100,1.500,3.300,0.700,3.700)	(0.700,1.900,2.400,1.300,-1.900)	(1.100,-0.900,2.400,0.300,4.100)
$\hat{\boldsymbol{\beta}}$[ii]	(-2.706,2.924,1.399,2.110,-2.307)	(-1.151,1.497,3.356,0.701,3.743)	(0.701,1.944,2.400,1.318,-1.915)	(1.110,-0.900,2.384,0.300,4.128)
sd[iii]	[0.031,0.046,0.015,0.025,0.021]	[0.802,0.374,0.776,0.183,0.419]	[0.042,0.045,0.034,0.022,0.031]	[0.075,0.060,0.061,0.030,0.079]
Case 2: ($I = 5000; J = 10$)				

Table 2.2. (Cont.)

	Shapes of $G(\cdot)$			
	Convex	Concave	$I_{t<0.5}$Convex+$I_{t\geq 0.5}$Concave	$I_{t<0.2}$Concave+$I_{t\geq 0.2}$Convex
$\boldsymbol{\beta}_0{}^{\text{i}}$	(2.300,3.500,1.800)	(1.200,2.900,0.600)	(2.400,2.300,1.600)	(1.700,2.100,1.200)
$\hat{\boldsymbol{\beta}}^{\text{ii}}$	(2.301,3.503,1.798)	(1.200,2.899,0.600)	(2.403,2.306,1.598)	(1.701,2.103,1.199)
sd$^{\text{iii}}$	[0.007,0.011,0.001]	[0.001,0.002,0.001]	[0.002,0.002,0.001]	[0.005,0.006,0.002]
$\boldsymbol{\beta}_0{}^{\text{i}}$	(-0.600,3.300,2.800)	(-1.100,2.900,1.800)	(-2.400,3.300,2.100)	(-0.900,2.300,3.400)
$\hat{\boldsymbol{\beta}}^{\text{ii}}$	(-0.596,3.302,2.790)	(-1.099,2.901,1.799)	(-2.386,3.315,2.089)	(-0.899,2.302,3.393)
sd$^{\text{iii}}$	[0.010,0.018,0.007]	[0.002,0.004,0.001]	[0.010,0.015,0.009]	[0.006,0.010,0.004]
$\boldsymbol{\beta}_0{}^{\text{i}}$	(-1.300,2.300,2.600)	(2.200,-2.300,3.600)	(2.500,-0.300,3.200)	(3.600,-0.400,4.100)
$\hat{\boldsymbol{\beta}}^{\text{ii}}$	(-1.291,2.313,2.582)	(2.201,-2.225,3.533)	(2.503,-0.294,3.189)	(3.591,-0.403,4.089)
sd$^{\text{iii}}$	[0.012,0.018,0.010]	[0.058,0.074,0.063]	[0.014,0.013,0.017]	[0.020,0.016,0.021]
$\boldsymbol{\beta}_0{}^{\text{i}}$	(1.400,5.300,-2.600)	(1.200,3.700,-1.800)	(2.600,1.500,-2.100)	(1.700,4.300,-2.700)
$\hat{\boldsymbol{\beta}}^{\text{ii}}$	(1.402,5.303,-2.601)	(1.199,3.699,-1.800)	(2.601,1.508,-2.101)	(1.701,4.302,-2.700)
sd$^{\text{iii}}$	[0.006,0.012,0.004]	[0.001,0.002,0.0005]	[0.006,0.007,0.003]	[0.003,0.005,0.002]
$\boldsymbol{\beta}_0{}^{\text{i}}$	(2.300,-0.600,2.400,3.100)	(-1.200,2.900,0.700,1.100)	(-0.300,1.200,2.100,2.100)	(2.800,-1.300,2.700,1.400)
$\hat{\boldsymbol{\beta}}^{\text{ii}}$	(2.304,-0.605,2.398,3.092)	(-1.200,2.899,0.700,1.100)	(-0.306,1.205,2.100,2.100)	(2.828,-1.323,2.699,1.398)
sd$^{\text{iii}}$	[0.005,0.003,0.003,0.019]	[0.001,0.001,0.0002,0.0005]	[0.002.0.001,0.001,0.002]	[0.009,0.006,0.010,0.008]
$\boldsymbol{\beta}_0{}^{\text{i}}$	(-3.900,3.600,-1.300,2.100)	(3.200,2.900,3.100,-1.100)	(3.300,-1.400,3.700,2.800)	(-4.100,3.300,2.600,2.400)
$\hat{\boldsymbol{\beta}}^{\text{ii}}$	(-3.915,3.610,-1.299,2.090)	(3.181,2.910,3.095,-1.099)	(3.310,-1.409,3.701,2.802)	(-4.100,3.300,2.600,2.400)
sd$^{\text{iii}}$	[0.013,0.015,0.001,0.015]	[0.011,0.008,0.006,0.001]	[0.004,0.003,0.001,0.002]	[0.001,0.001,0.001,0.002]
$\boldsymbol{\beta}_0{}^{\text{i}}$	(4.700,-1.900,5.500,3.100)	(1.000,1.600,-2.700,1.100)	(1.700,2.200,-1.100,0.900)	(-3.800,2.300,1.700,2.100)
$\hat{\boldsymbol{\beta}}^{\text{ii}}$	(4.709,-1.912,5.490,3.093)	[0.997,1.602,-2.700,1.099)	(1.692,2.210,-1.100,0.901)	(-3.804,2.303,1.700,2.099)
sd$^{\text{iii}}$	[0.009,0.005,0.035,0.017]	[0.002,0.001,0.0003,0.001]	[0.004,0.005,0.001,0.002]	[0.003,0.003,0.001,0.005]

Table 2.2. (Cont.)

	Shapes of $G(\cdot)$			
	Convex	Concave	$I_{t<0.5}$Convex+$I_{t\geq 0.5}$Concave	$I_{t<0.2}$Concave+$I_{t\geq 0.2}$Convex
$\boldsymbol{\beta}_0$[i]	(0.700,2.900,-2.500,4.100)	(2.700,-1.200,2.600,0.900)	(3.500,3.700,6.700,-0.300)	(0.500,1.300,-0.700,2.300)
$\hat{\boldsymbol{\beta}}$[ii]	(0.695,2.901,-2.498,4.100)	(2.730,-1.222,2.603,0.901)	(3.496,3.709,6.802,-0.300)	(0.499,1.300,-0.700,2.300)
sd[iii]	[0.003,0.003,0.001,0.020]	[0.009,0.006,0.004,0.002]	[0.006,0.008,0.121,0.0004]	[0.0005,0.0007,0.0001,0.004]
$\boldsymbol{\beta}_0$[i]	(2.200,-1.300,-1.400,2.200,1.600)	(2.100,-1.600,-1.200,2.900,-0.800)	(2.700,-1.300,1.600,2.100,-0.700)	(3.100,-1.000,-1.400,0.300,2.300)
$\hat{\boldsymbol{\beta}}$[ii]	(2.206,-1.298,-1.405,2.206,1.606)	(2.104,-1.599,-1.203,2.905,-0.800)	(2.714,-1.294,1.602,2.115,-0.700)	(3.100,-1.000,-1.400,0.300,2.301)
sd[iii]	[0.017,0.011,0.015,0.018,0.015]	[0.008,0.006,0.0006,0.006,0.004]	[0.019,0.016,0.009,0.017,0.009]	[0.002,0.001,0.002,0.001,0.002]
$\boldsymbol{\beta}_0$[i]	(-3.100,4.300,2.400,1.200,-1.600)	(-2.100,2.700,3.100,1.900,-0.600)	(1.900,1.300,-3.600,0.500,2.300)	(-1.100,2.300,1.700,0.900,-0.300)
$\hat{\boldsymbol{\beta}}$[ii]	(-3.099,4.316,2.394,1.203,-1.604)	(-2.102,2.706,3.102,1.903,-0.600)	(1.906,1.308,-3.610,0.503,2.309)	(-1.100,2.301,1.699,0.900,-0.300)
sd[iii]	[0.020,0.044,0.013,0.013,0.013]	[0.004,0.007,0.004,0.004,0.002]	[0.008,0.011,0.012,0.004,0.010]	[0.001,0.004,0.001,0.002,0.001]
$\boldsymbol{\beta}_0$[i]	(3.600,-1.700,2.600,2.200,1.500)	(1.2,-1.5,0.3,2.4,-1.6)	(0.7,-2.1,1.0,1.5,1.3)	(-2.7,-0.7,-0.7,1.9,1.3)
$\hat{\boldsymbol{\beta}}$[ii]	(3.623,-1.696,2.603,2.218,1.519)	(1.201,-1.499,0.300,2.403,-1.601)	(0.703,-2.100,0.999,1.504,1.306)	(-2.700,0.699,-0.701,1.901,1.301)
sd[iii]	[0.059,0.046,0.037,0.035,0.041]	[0.004,0.004,0.002,0.004,0.002]	[0.004,0.005,0.003,0.006,0.007]	[0.003,0.002,0.002,0.004,0.003]
$\boldsymbol{\beta}_0$[i]	(-2.700,2.900,1.400,2.100,-2.300)	(-1.100,1.500,3.300,0.700,3.700)	(0.700,1.900,2.400,1.300,-1.900)	(1.100,-0.900,2.400,0.300,4.100)
$\hat{\boldsymbol{\beta}}$[ii]	(-2.700,2.904,1.399,2.101,-2.301)	(-1.091,1.506,3.299,0.702,3.722)	(0.700,1.924,2.402,1.309,-1.910)	(1.105,-0.899,2.393,0.301,4.111)
sd[iii]	[0.006,0.015,0.002,0.007,0.006]	[0.107,0.128,0.067,0.041,0.119]	[0.012,0.022,0.008,0.009,0.009]	[0.018,0.025,0.022,0.006,0.029]
	Case 3: ($I = 1000; J = 100$)			
$\boldsymbol{\beta}_0$[i]	(2.300,3.500,1.800)	(1.200,2.900,0.600)	(2.400,2.300,1.600)	(1.700,2.100,1.200)
$\hat{\boldsymbol{\beta}}$[ii]	(2.302,3.504,1.798)	(1.199,2.899,0.600)	(2.402,2.305,1.598)	(1.701,2.103,1.199)
sd	[0.005,0.009,0.001]	[0.001,0.001,0.0004]	[0.001,0.002,0.0005]	[0.003,0.004,0.001]
$\boldsymbol{\beta}_0$[i]	(-0.600,3.300,2.800)	(-1.100,2.900,1.800)	(-2.400,3.300,2.100)	(-0.900,2.300,3.400)
$\hat{\boldsymbol{\beta}}$[ii]	(-0.595,3.306,2.792)	(-1.099,2.901,1.799)	(-2.390,3.312,2.091)	(-0.899,2.304,3.393)
sd[iii]	[0.007,0.013,0.005]	[0.003,0.004,0.002]	[0.008,0.014,0.007]	[0.004,0.006,0.003]

Table 2.2. (Cont.)

	Shapes of $G(\cdot)$			
	Convex	Concave	$I_{t<0.5}$Convex+$I_{t\geq 0.5}$Concave	$I_{t<0.2}$Concave+$I_{t\geq 0.2}$Convex
$\boldsymbol{\beta}_0$[i]	(-1.300,2.300,2.600)	(2.200,-2.300,3.600)	(2.500,-0.300,3.200)	(3.600,-0.400,4.100)
$\hat{\boldsymbol{\beta}}$[ii]	(-1.294,2.311,2.582)	(2.168,-2.277,3.539)	(2.496,-0.303,3.188)	(3.592,-0.404,4.088)
sd[iii iii]	[0.011,0.019,0.007]	[0.025,0.044,0.046]	[0.010,0.009,0.012]	[0.014,0.011,0.015]
$\boldsymbol{\beta}_0$[i]	(1.400,5.300,-2.600)	(1.200,3.700,-1.800)	(2.600,1.500,-2.100)	(1.700,4.300,-2.700)
$\hat{\boldsymbol{\beta}}$[ii]	(1.402,5.303,-2.601)	(1.199,3.698,-1.799)	(2.597,1.504,-2.100)	(1,701,4.302,-2.700)
sd[iii iii]	[0.004,0.008,0.003]	[0.0005,0.001,0.0002]	[0.004,0.004,0.002]	[0.002,0.003,0.001]
$\boldsymbol{\beta}_0$[i]	(2.300,-0.600,2.400,3.100)	(-1.200,2.900,0.700,1.100)	(-0.300,1.200,2.100,2.100)	(2.800,-1.300,2.700,1.400)
$\hat{\boldsymbol{\beta}}$[ii]	(2.306,-0.605,2.399,3.099)	(-1.200,2.899,0.700,1.100)	(-0.306,1.205,2.100,2.099)	(2.828,-1.322,2.702,1.400)
sd[iii]	[0.002,0.001,0.001,0.003]	[0.0002,0.001,0.0002,0.0004]	[0.001,0.001,0.0004,0.001]	[0.003,0.002,0.002,0.002]
$\boldsymbol{\beta}_0$[i]	(-3.900,3.600,-1.300,2.100)	(3.200,2.900,3.100,-1.100)	(3.300,-1.400,3.700,2.800)	(-4.100,3.300,2.600,2.400)
$\hat{\boldsymbol{\beta}}$[ii]	(-3.915,3.612,-1.299,2.097)	(3.185,2.908,3.093,-1.100)	(3.310,-1.409,3.701,2.800)	(-4.100,3.300,2.600,2.400)
sd[iii]	[0.003,0.003,0.001,0.003]	[0.005,0.003,0.003,0.001]	[0.002,0.001,0.001,0.001]	[0.0005,0.0004,0.0003,0.001]
$\boldsymbol{\beta}_0$[i]	(4.700,-1.900,5.500,3.100)	(1.000,1.600,-2.700,1.100)	(1.700,2.200,-1.100,0.900)	(-3.800,2.300,1.700,2.100)
$\hat{\boldsymbol{\beta}}$[ii]	(4.711,-1.911,5.502,3.099)	(0.997,1.602,-2.700,1.099)	(1.694,2.208,-1.100,0.901)	(-3.803,2.303,1.700,2.100)
sd[iii]	[0.004,0.002,0.009,0.004]	[0.001,0.001,0.0002,0.001]	[0.002,0.004,0.0004,0.002]	[0.001,0.001,0.0002,0.001]
$\boldsymbol{\beta}_0$	(0.700,2.900,-2.500,4.100)	(2.700,-1.200,2.600,0.900)	(3.500,3.700,6.700,-0.300)	(0.500,1.300,-0.700,2.300)
$\hat{\boldsymbol{\beta}}$[ii]	(0.697,2.901,-2.499,4.103)	(2.726,-1.220,2.601,0.900)	(3.498,3.705,6.766,-0.300)	(0.500,1.300,-0.700,2.300)
sd[iii]	[0.002,0.001,0.001,0.007]	[0.003,0.002,0.002,0.001]	[0.003,0.004,0.093,0.0003]	[0.0002,0.0002,0.0001,0.001]
$\boldsymbol{\beta}_0$[i]	(2.200,-1.300,-1.400,2.200,1.600)	(2.100,-1.600,-1.200,2.900,-0.800)	(2.700,-1.300,1.600,2.100,-0.700)	(3.100,-1.000,-1.400,0.300,2.300)
$\hat{\boldsymbol{\beta}}$[ii]	(2.205,-1.300,-1.404,2.205,1.604)	(2.103,-1.600,1.202,2.904,-0.800)	(2.711,-1.298,1.603,2.112,-0.701)	(3.101,-1.000,-1.401,0.300,2.301)
sd[iii]	[0.012,0.004,0.009,0.012,0.010]	[0.005,0.005,0.004,0.006,0.003]	[0.016,0.008,0.006,0.012,0.004]	[0.002,0.001,0.001,0.001,0.002]

Table 2.2. (Cont.)

	Shapes of $G(\cdot)$			
	Convex	Concave	$I_{t<0.5}$Convex+$I_{t\geq 0.5}$Concave	$I_{t<0.2}$Concave+$I_{t\geq 0.2}$Convex
$\boldsymbol{\beta}_0$[i]	(-3.100,4.300,2.400,1.200,-1.600)	(-2.100,2.700,3.100,1.900,-0.600)	(1.900,1.300,-3.600,0.500,2.300)	(-1.100,2.300,1.700,0.900,-0.300)
$\hat{\boldsymbol{\beta}}$[ii]	(-3.102,4.316,2.397,1.203,-1.604)	(-2.101,2.703,3.101,1.902,-0.600)	(1.904,1.304,-3.607,0.502,2.306)	(-1.100,2.301,1.700,0.900,-0.300)
sd[iii]	[0.015,0.035,0.009,0.010,0.008]	[0.004,0.005,0.003,0.003,0.002]	[0.006,0.005,0.008,0.003,0.008]	[0.001,0.002,0.001,0.001,0.0003]
$\boldsymbol{\beta}_0$[i]	(3.600,-1.700,2.600,2.200,1.500)	(1.200,-1.500,0.300,2.400,-1.600)	(0.700,-2.100,1.000,1.500,1.300)	(-2.700,-0.700,-0.700,1.900,1.300)
$\hat{\boldsymbol{\beta}}$[ii]	(3.617,-1.702,2.605,2.214,1.513)	(1.200,-1.500,0.300,2.402,-1.601)	(0.703,-2.101,0.999,1.504,1.304)	(-2.700,-0.700,-0.700,1.901,1.301)
sd[iii]	[0.043,0.029,0.029,0.025,0.024]	[0.002,0.002,0.001,0.003,0.001]	[0.004,0.004,0.003,0.005,0.005]	[0.002,0.001,0.001,0.002,0.002]
$\boldsymbol{\beta}_0$[i]	(-2.700,2.900,1.400,2.100,-2.300)	(-1.100,1.500,3.300,0.700,3.700)	(0.700,1.900,2.400,1.300,-1.900)	(1.100,-0.900,2.400,0.300,4.100)
$\hat{\boldsymbol{\beta}}$[ii]	(-2.701,2.905,1.400,2.102,-2.302)	(-1.096,1.507,3.298,0.702,3.713)	(0.701,1.915,2.402,1.306,-1.906)	(1.105,-0.901,2.396,0.300,4.111)
sd[iii]	[0.005,0.013,0.002,0.007,0.003]	[0.072,0.024,0.035,0.017,0.039]	[0.006,0.013,0.005,0.007,0.004]	[0.015,0.008,0.014,0.004,0.022]

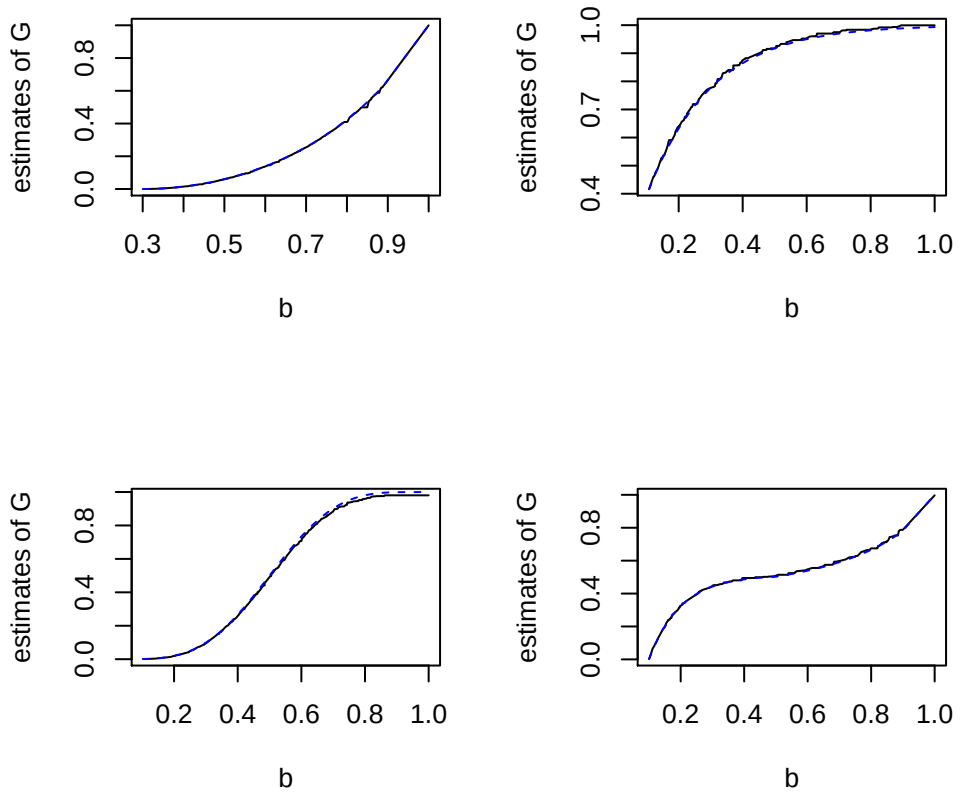

Figure 2.1: Estimates of $G(t)$ when $G(t)$ is convex, concave, a mixture of convex and concave, sample size is 200×10 under Case 1: Solid line: True $G_0(\cdot)$; Step line: Estimate $\hat{G}(\cdot)$.

Simulation results: Signal detection

The simulation findings (Table 2.3) confirm that the type I error is around 0.05 under the null hypothesis, demonstrating Holm-Bonferroni correction controls the type I error very well. Under the alternative hypothesis, the power for the proposed approach is close to 100%. Within Case 2 (1000×100), and when the number of true signals is 100, as the signal becomes strong, Table 2.3 shows that the specificity of the proposed method decreases from $> 99.99\%$ to 98.98%, the sensitivity increases from

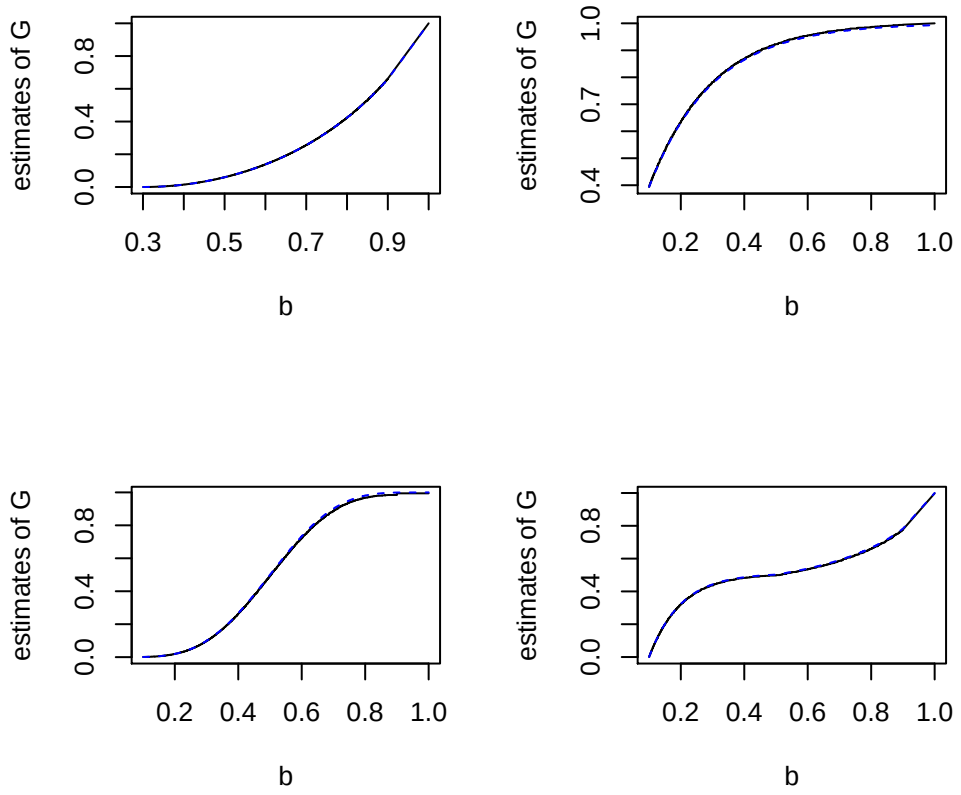

Figure 2.2: Estimates of $G(t)$ when $G(t)$ is convex, concave, a mixture of convex and concave, sample size is 5000×10 under Case 2: Solid line: True $G_0(\cdot)$; Step line: Estimate $\hat{G}(\cdot)$.

93.54% to $>$ 99.99%, and FDR increases from 0.01 to 0.048. As the number of true signals increases from 100 to 1000, Sensitivity and Specificity do not change much, and FDR decreases. Similar conclusions hold for other cases (200×10 and 5000×10) and within each case. As the sample sizes increases from 200×10 to 1000×100, Specificity increases, Sensitivity and FDR decrease. In general, the proposed approach has high

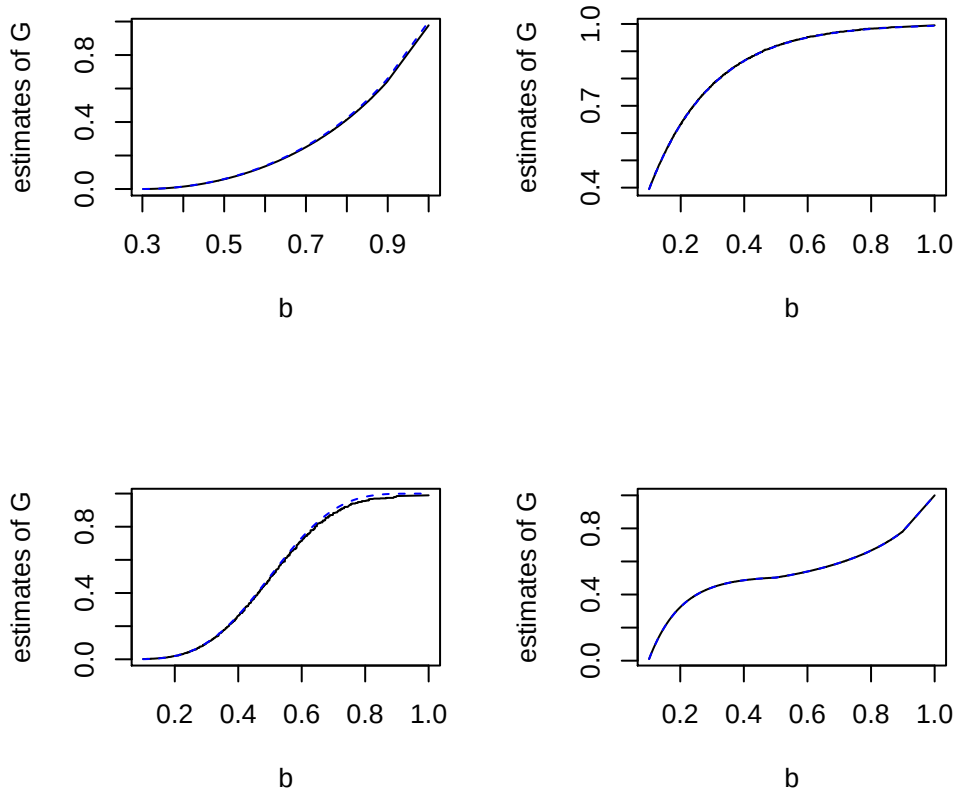

Figure 2.3: Estimates of $G(t)$ when $G(t)$ is convex, concave, a mixture of convex and concave, sample size is 1000×100 under Case 3: Solid line: True $G_0(\cdot)$; Step line: Estimate $\hat{G}(\cdot)$.

power ($> 99.99\%$), high sensitivity (92.78% -), and high specificity (91.48% -). As the signal strength increases, the specificity decreases, and the sensitivity increases.

The simulated data are generated with covariate information. Tables 2.4 and 2.5 show the performance characteristic of these methods without covariates information. Table 2.4 shows the performance characteristic of the proposed method when setting the covariates as **1**, in other words, when ignoring the covariates information. Within

Case 2 (1000×100), the power is close to 1. When the number of the true signals is 1000, as the signal strength becomes strong, the specificity increases from 74.95% to 76.36%, the sensitivity increases from 32.59% to 57.35%, and FDR decreases from 0.976 to 0.957. A similar trend holds when the number of signals is 100, 350, and 700. Specificity, Sensitivity, and FDR do not change as the number of signals changes. Similar conclusions hold for other sample size settings (200×10 and 5000×10). In general, NHP has medium specificity (73.51% - 88.83%) among all these approaches without covariates, but it also has low sensitivity (32.87% - 62.18%).

Table 2.5 shows the performance characteristic of the classic methods (PRR, BCPNN, and ZIPLRT) when ignoring the covariates information. Within Case 2 (1000×100), the power of PRR is close to 1. When the number of the true signals is 1000, as the signal strength becomes strong, the specificity increases from 63.74% to 65.49%, the sensitivity increases from 82.13% to 99.36%, and FDR is around 0.976. A similar trend holds when the number of signals is 100, 350, and 700. Specificity, Sensitivity, and FDR do not change much as the number of signals changes. Similar conclusion hold for other sample size settings (200×10 and 5000×10). In general, PRR has high sensitivity (74.94% - 99.48%), but it also has low specificity (57.50% - 77.34%). Within Case 2 (1000×100), the power of BCPNN is close to 1. When the number of the true signals is 1000, as the signal strength becomes strong, the specificity increases from 62.44% to 65.49%, the sensitivity increases from 83.72% to 99.44%, and FDR is around 0.976. A similar trend holds when the number of signals is 100, 350, and 700. Specificity, Sensitivity, and FDR do not change much as the number of signals changes. Similar conclusion hold for other sample size settings (200×10 and 5000×10). In general, BCPNN is similar to PRR, with high sensitivity (78.37% - 99.63%) and low specificity (56.12% - 76.61%). Within Case 2 (1000×100), the power of ZIP LRT is close to 1. When the number of the true signals is 1000, as

the signal strength becomes strong, the specificity increases from 67.59% to 68.99%, the sensitivity decreases from 32.66% to 31.31%, and FDR is around 0.990. A similar trend holds when the number of signals is 100, 350, and 700. Specificity, Sensitivity, and FDR do not change much as the number of signals changes. Similar conclusion hold for other sample size settings (200×10 and 5000×10). In general, ZIP LRT has low specificity (62.81% - 68.99%) and low sensitivity (31.30% - 37.65%).

Table 2.6 shows the number of signals found with different approaches. Those without incorporating covariates information (PRR, BCPNN, ZIP LRT, NHP) perform pretty poorly and tend to heavily overestimate the number of signals, except NHP that may over or underestimate. In contrast, the proposed approach with covariates adjustment finds the true number of signals with outstanding accuracy, demonstrating the additional value of the proposed method over the existing ones.

Table 2.3: Performance characteristic of the proposed method in signal detection when signal strength (rr) or sample size varies.

True #	rr	Power(%)	Specificity(%)	Sensitivity(%)	FDR
			I=200, J=10		
0	-	0.060[iv]	-	-	
50	1.5	> 99.99	99.79	95.40	0.033
50	2	> 99.99	99.39	99.66	0.071
50	4	> 99.99	93.53	> 99.99	0.327
50	6	> 99.99	91.48	> 99.99	0.382
100	1.5	> 99.99	95.91	99.92	0.009
100	2	> 99.99	99.47	99.54	0.038
100	4	> 99.99	96.49	> 99.99	0.151
100	6	> 99.99	94.88	> 99.99	0.224

[iv] when the simulated data are under null hypothesis, this column indicates type I error.

Table 2.3. (Cont.)

True #	rr	Power(%)	Specificity(%)	Sensitivity(%)	FDR
350	1.5	> 99.99	99.95	93.06	0.001
350	2	> 99.99	99.86	99.21	0.004
350	4	> 99.99	99.24	99.97	0.014
350	6	> 99.99	99.61	99.99	0.009
			I=1000, J=100		
0	-	0.02^{iv}	-	-	-
100	1.5	> 99.99	> 99.99	93.23	< 0.001
100	2	> 99.99	> 99.99	99.08	< 0.001
100	4	> 99.99	99.93	99.98	0.021
100	6	> 99.99	99.98	> 99.99	0.037
350	1.5	> 99.99	> 99.99	92.91	< 0.001
350	2	> 99.99	> 99.99	99.22	< 0.001
350	4	> 99.99	99.71	99.99	0.015
350	6	> 99.99	99.99	> 99.99	0.009
700	1.5	> 99.99	99.99	93.12	< 0.001
700	2	> 99.99	> 99.99	99.26	0.003
700	4	> 99.99	99.98	> 99.99	0.006
700	6	> 99.99	99.42	> 99.99	0.059
1000	1.5	> 99.99	> 99.99	92.78	< 0.001
1000	2	> 99.99	> 99.99	99.14	< 0.001
1000	4	> 99.99	99.90	99.99	0.013
1000	6	> 99.99	99.57	> 99.99	0.024
			I=5000, J=10		
0	-	0.04^{iv}	-	-	-

Table 2.3. (Cont.)

True #	rr	Power(%)	Specificity(%)	Sensitivity(%)	FDR
100	1.5	> 99.99	> 99.99	93.54	0.001
100	2	> 99.99	99.94	99.37	0.011
100	4	> 99.99	99.38	99.97	0.040
100	6	> 99.99	98.98	> 99.99	0.048
600	1.5	> 99.99	99.99	93.86	0.004
600	2	> 99.99	99.92	99.11	0.013
600	4	> 99.99	97.96	99.99	0.057
600	6	> 99.99	98.98	> 99.99	0.073
1000	1.5	> 99.99	> 99.99	93.50	< 0.001
1000	2	> 99.99	99.68	99.26	0.016
1000	4	> 99.99	98.09	99.99	0.056
1000	6	> 99.99	92.25	92.87	0.184

Table 2.4: Performance characteristic of the proposed method without covariates information in signal detection when signal strength (rr) or sample size varies.

True #	rr	Power(%)	Specificity(%)	Sensitivity(%)	FDR
I=200, J=10					
0	-	0.060[iv]	-	-	0.060
50	1.5	> 99.99	74.01	35.94	0.943
50	2	> 99.99	73.71	40.94	0.936
50	4	> 99.99	75.74	52.28	0.914
50	6	> 99.99	77.44	59.06	0.897
100	1.5	> 99.99	73.65	35.20	0.892
100	2	> 99.99	75.01	39.72	0.873
100	4	> 99.99	77.92	51.21	0.825
100	6	> 99.99	80.07	55.76	0.796
350	1.5	> 99.99	75.77	32.42	0.646
350	2	> 99.99	78.58	36.16	0.591
350	4	> 99.99	84.92	42.96	0.462
350	6	> 99.99	88.83	46.77	0.371
I=1000, J=100					
0	-	0.02[iv]	-	-	0.02
100	1.5	> 99.99	75.07	32.87	0.998
100	2	> 99.99	74.86	38.51	0.997
100	4	> 99.99	74.89	50.46	0.996
100	6	> 99.99	75.07	59.49	0.996

Table 2.4. (Cont.)

True #	*rr*	Power(%)	Specificity(%)	Sensitivity(%)	FDR
350	1.5	> 99.99	75.10	32.87	0.992
350	2	> 99.99	74.84	38.47	0.990
350	4	> 99.99	74.92	52.26	0.987
350	6	> 99.99	75.40	58.26	0.985
700	1.5	> 99.99	74.91	32.87	0.983
700	2	> 99.99	75.22	37.48	0.981
700	4	> 99.99	75.59	50.90	0.974
700	6	> 99.99	75.89	57.82	0.970
1000	1.5	> 99.99	74.95	32.59	0.976
1000	2	> 99.99	75.04	37.93	0.973
1000	4	> 99.99	75.87	50.46	0.963
1000	6	> 99.99	76.36	57.85	0.957
			I=5000, J=10		
0	-	0.04^{iv}	-	-	0.04
100	1.5	> 99.99	74.09	34.17	0.996
100	2	> 99.99	73.98	39.64	0.996
100	4	> 99.99	73.51	53.21	0.994
100	6	> 99.99	74.53	62.18	0.993
600	1.5	> 99.99	74.03	34.47	0.977
600	2	> 99.99	74.19	40.04	0.974
600	4	> 99.99	74.70	52.59	0.965
600	6	> 99.99	75.58	61.96	0.957

Table 2.4. (Cont.)

True #	rr	Power(%)	Specificity(%)	Sensitivity(%)	FDR
1000	1.5	> 99.99	74.48	33.95	0.962
1000	2	> 99.99	74.42	39.99	0.956
1000	4	> 99.99	75.39	53.10	0.940
1000	6	> 99.99	76.88	59.49	0.929

Table 2.5: Performance characteristic of PRR, BCPNN and ZIP LRT in signal detection when signal strength (rr) or sample size varies.

		PRR				BCPNN				ZIPLRT			
True #	rr	Power(%)[iv]	Specificity(%)	Sensitivity(%)	FDR	Power(%)	Specificity(%)	Sensitivity(%)	FDR	Power(%)	Specificity(%)	Sensitivity(%)	FDR
I=200, J=10													
0	-	1[iv]	-	-	1	1[iv]	-	-	1	1[iv]	-	-	1
50	1.5	> 99.99	60.57	80.46	0.950	> 99.99	59.35	81.78	0.951	> 99.99	63.34	36.14	0.975
50	2	> 99.99	60.95	86.38	0.946	> 99.99	59.76	87.30	0.947	> 99.99	63.35	36.10	0.975
50	4	> 99.99	61.91	95.88	0.939	> 99.99	60.73	96.12	0.941	> 99.99	64.19	35.02	0.976
50	6	> 99.99	62.57	97.38	0.937	> 99.99	61.51	97.62	0.939	> 99.99	64.45	36.34	0.974
100	1.5	> 99.99	61.70	78.36	0.903	> 99.99	60.51	79.83	0.904	> 99.99	63.52	37.25	0.949
100	2	> 99.99	62.01	85.67	0.894	> 99.99	60.83	86.80	0.896	> 99.99	63.55	36.46	0.950
100	4	> 99.99	63.91	93.98	0.879	> 99.99	62.79	94.35	0.882	> 99.99	64.70	35.87	0.949
100	6	> 99.99	65.37	96.37	0.872	> 99.99	64.28	96.60	0.875	> 99.99	65.61	33.87	0.951
350	1.5	> 99.99	67.58	77.01	0.665	> 99.99	66.60	78.37	0.668	> 99.99	63.36	36.44	0.826
350	2	> 99.99	69.30	81.75	0.639	> 99.99	68.28	82.75	0.644	> 99.99	64.09	36.09	0.824
350	4	> 99.99	74.23	88.27	0.579	> 99.99	73.42	88.86	0.585	> 99.99	66.57	33.71	0.824
350	6	> 99.99	77.34	90.25	0.542	> 99.99	76.61	90.59	0.549	> 99.99	68.12	32.02	0.824
I=1000, J=100													
0	-	1[iv]	-	-	1	1[iv]	-	-	1	1[iv]	-	-	1
100	1.5	> 99.99	63.27	82.53	0.998	> 99.99	62.00	84.15	0.998	> 99.99	67.56	31.93	0.999
100	2	> 99.99	63.32	89.67	0.998	> 99.99	62.00	90.65	0.998	> 99.99	67.55	32.36	0.999
100	4	> 99.99	63.39	98.05	0.997	> 99.99	62.11	98.36	0.997	> 99.99	67.64	32.59	0.999
100	6	> 99.99	63.53	99.56	0.997	> 99.99	62.18	99.63	0.997	> 99.99	67.70	31.68	0.999

Table 2.5. (Cont.)

True #	*rr*	Power(%)[iv]	Specificity(%)	Sensitivity(%)	FDR	Power(%)	Specificity(%)	Sensitivity(%)	FDR	Power(%)	Specificity(%)	Sensitivity(%)	FDR
350	1.5	> 99.99	63.42	82.39	0.992	> 99.99	62.14	84.04	0.992	> 99.99	67.55	32.41	0.997
350	2	> 99.99	63.55	90.03	0.991	> 99.99	62.21	91.11	0.992	> 99.99	67.59	33.16	0.996
350	4	> 99.99	63.79	98.17	0.991	> 99.99	62.49	98.42	0.991	> 99.99	67.85	32.14	0.997
350	6	> 99.99	64.07	99.48	0.990	> 99.99	62.74	99.57	0.991	> 99.99	68.10	32.06	0.996
700	1.5	> 99.99	63.57	82.14	0.984	> 99.99	62.26	83.74	0.985	> 99.99	67.60	32.77	0.993
700	2	> 99.99	63.70	89.77	0.983	> 99.99	62.45	90.85	0.983	> 99.99	67.75	32.37	0.993
700	4	> 99.99	64.34	98.11	0.981	> 99.99	63.06	98.37	0.982	> 99.99	68.12	31.84	0.993
700	6	> 99.99	64.87	99.39	0.980	> 99.99	63.55	99.47	0.981	> 99.99	68.60	31.30	0.993
1000	1.5	> 99.99	63.74	82.13	0.978	> 99.99	62.44	83.72	0.978	> 99.99	67.59	32.66	0.990
1000	2	> 99.99	63.94	89.53	0.976	> 99.99	62.68	90.60	0.976	> 99.99	67.72	32.38	0.990
1000	4	> 99.99	64.86	98.03	0.973	> 99.99	63.52	98.29	0.974	> 99.99	68.31	31.91	0.990
1000	6	> 99.99	65.49	99.36	0.972	> 99.99	64.24	99.44	0.973	> 99.99	68.99	31.31	0.990
I=5000, J=10													
0	-	1[iv]	-	-	1	1[iv]	-	-	1	1[iv]	-	-	1
100	1.5	> 99.99	57.50	74.94	0.749	> 99.99	56.12	76.43	0.997	> 99.99	63.01	36.94	0.998
100	2	> 99.99	57.57	84.13	0.996	> 99.99	56.15	85.07	0.996	> 99.99	62.98	37.04	0.998
100	4	> 99.99	57.79	95.49	0.995	> 99.99	56.33	95.84	0.996	> 99.99	62.81	37.65	0.998
100	6	> 99.99	57.81	98.06	0.995	> 99.99	56.47	98.27	0.995	> 99.99	63.17	36.89	0.998
600	1.5	> 99.99	58.03	75.19	0.979	> 99.99	56.62	76.70	0.979	> 99.99	62.99	37.04	0.988
600	2	> 99.99	58.10	83.25	0.976	> 99.99	56.75	84.37	0.977	> 99.99	62.85	37.12	0.988
600	4	> 99.99	58.69	95.12	0.973	> 99.99	57.34	95.48	0.974	> 99.99	63.18	36.83	0.988
600	6	> 99.99	58.96	97.77	0.972	> 99.99	57.61	97.93	0.973	> 99.99	63.51	36.45	0.988

Table 2.5. (Cont.)

True #	*rr*	Power(%)[iv]	Specificity(%)	Sensitivity(%)	FDR	Power(%)	Specificity(%)	Sensitivity(%)	FDR	Power(%)	Specificity(%)	Sensitivity(%)	FDR
1000	1.5	> 99.99	58.38	75.01	0.965	> 99.99	57.02	76.53	0.965	> 99.99	63.02	37.20	0.980
1000	2	> 99.99	58.46	83.13	0.961	> 99.99	57.07	84.28	0.961	> 99.99	62.94	37.06	0.980
1000	4	> 99.99	59.30	94.55	0.955	> 99.99	57.93	94.95	0.956	> 99.99	63.40	36.35	0.980
1000	6	> 99.99	59.83	97.37	0.953	> 99.99	58.51	97.56	0.954	> 99.99	64.06	35.76	0.980

Table 2.6: Signals (AEs) detected by the proposed method, the proposed method without covariates, PRR, BCPNN and ZIPLRT.

True #	*rr*	PRR	BCPNN	ZIPLRT	NHP	HP
I=200, J=10						
0	-	813	836	734	321	0.07
50	1.5	809	834	733	317	50
50	2	805	827	730	321	57
50	4	794	816	721	307	124
350	1.5	804	825	732	320	326
350	2	789	810	721	309	348
350	4	737	753	673	279	356
I=1000, J=100						
0	-	36774	38075	32472	13879	0.02
350	1.5	36744	38022	32446	13697	325
350	2	36726	37960	32212	13888	347
350	8	35911	37221	31737	13589	545
1000	1.5	36719	38018	32415	13833	928
1000	2	36577	37855	32312	13885	991
1000	8	34425	35720	30352	13023	1320
I=5000, J=10						
0	-	21302	22014	18587	9119	0.04
100	1.5	21282	21970	18494	9106	94
100	3	21248	21937	18555	8972	102
100	6	21151	21821	18413	8940	467

Table 2.6. (Cont.)

True #	rr	PRR	BCPNN	ZIPLRT	NHP	HP
600	1.5	21185	21888	18507	9162	567
600	4	20976	21647	18411	9005	1306
600	6	20860	21527	18245	8789	953

Additional simulations: Robustness to different proportions of true zeros

Above simulation settings in §2.6 are motivated by real examples in §2.7. Simulation studies in this subsection are to evaluate the robustness of the proposed method to different proportions of true zeros. We generate data with different proportions of true zeros.

Table 2.7: Performance characteristic of the proposed method when the proportions of true zeros ($\rho(\%)$) and sample size varies.

$\rho(\%)$	Power(%)	Specificity(%)	Sensitivity(%)	FDR
I=200, J=10(True #=50,$rr = 1.5$)				
40	> 99.99	99.84	95.16	0.023
50	> 99.99	99.69	95.50	0.034
60	> 99.99	99.62	95.46	0.035
70	> 99.99	99.28	96.00	0.043
80	> 99.99	99.14	95.86	0.046
90	> 99.99	97.23	94.92	0.062
I=1000, J=100(True #=350,$rr = 1.5$				

Table 2.7. (Cont.)

ρ(%)	Power(%)	Specificity(%)	Sensitivity(%)	FDR
40	> 99.99	> 99.99	92.89	0.0002
50	> 99.99	> 99.99	93.06	0.0002
60	> 99.99	> 99.99	93.11	0.0001
70	> 99.99	> 99.99	93.50	0.0003
80	> 99.99	> 99.99	93.63	0.0004
90	> 99.99	> 99.99	94.27	0.005
		I=5000, J=10		
40	> 99.99	> 99.99	93.87	0.003
50	> 99.99	99.35	94.02	0.011
60	> 99.99	> 99.99	94.45	0.001
70	> 99.99	99.98	93.74	0.007
80	> 99.99	99.96	94.62	0.009
90	> 99.99	98.82	95.23	0.053

Table 2.7 shows that FDR increases as the proportion of true zeros increases. In general, the proposed method has high power, high specificity, and high sensitivity in all cases with different proportions of true zeros with controlling FDR in most cases ($<$0.05). When the sample size is small (200×10) and the proportion of true zeros is extremely large (90%), FDR is slightly larger than 0.05. The results demonstrate the robustness of the proposed method to different proportions of true zeros.

Additional simulations: Robustness to extreme weak signals

Table 2.8 shows the results from the proposed method when signals are fragile ($rr < 1.5$). The proposed approach has high power and specificity when signals are extremely weak. However, the sensitivity is low when the signal strength is lower than 1.2. Still, it increases from 46.08% to 91.84% when the signal strength increases. One possible reason for low sensitivity, in this case, is that it is difficult to distinguish a weak signal from a noisy background.

Table 2.8: Performance characteristic of the proposed method when rr is extreme small and the signals are extreme weak.

True #	rr	Power(%)	Specificity(%)	Sensitivity(%)	FDR
I=200, J=10					
50	1.1	> 99.99	99.99	46.08	0.004
50	1.2	> 99.99	99.99	71.80	0.002
50	1.3	> 99.99	99.94	86.68	0.011
50	1.4	> 99.99	99.89	91.94	0.016

2.7 Analysis of cardiovascular medicines and corresponding adverse events

We apply the proposed modeling approach to WHO VigiBase for cases submitted to WHO International Drug Monitoring System between 2000/1 to 2020/6. The VigiBase reports include patient demographics, reported drugs, and events. We focus on common cardiovascular medicines in this analysis, which are the column names of our dataset. Furthermore, the reported events have been coded according to Medical Dictionary for Regulatory Activities (MedDRA). Therefore, we use the highest level

in MedDRA, System Organ Class(SOC), to identify the category of AEs we plan to investigate. Then, we use a lower level term Preferred Terms (PT) in MedDRA to define the single AE as the row name in the analysis, such as malaise, hepatitis, and nervousness. Moreover, the patient demographic, including individual age, gender, and country information, are used as covariates.

In the VigiBase dataset, we investigate the association between common cardiovascular medicines and their corresponding AEs. We will perform signal detection in this database and discuss the public health implications from comparisons with other approaches, including PRR, BCPNN, and ZIP LRT, which do not incorporate covariates. The structure of the datasets might be different for different investigative questions. We consider four common situations below.

2.7.1 Selected drugs and AEs (small drug-AE combinations) (VigiBase data 1)

In this case, we detect vascular disorder-related AEs signals against Acebutolol, Apixaban, and 8 other cardiovascular medicines to treat thrombosis. We extract all the vascular disorder-related AEs such as hematoma, hemorrhage, and hyperemia against one of the 10 medicines reported during the analyses period. After data processing, there are 202 vascular disorder-related AEs against at least one of the 10 drugs. After collapsing the data into the form of 202×10 table, $\rho = 68.6\%$ of the cells are zeros; for example, no occurrence of angiopathy is reported against Amiodarone, then the cell count of angiopathy-Amiodarone is zero. The total number of occurrence of AEs is $y_{..} = 30488$. Hypotension has the highest occurrence, which is 6513 in total ($\max(y_{i\cdot}) = 6513$), and aortic dissection rupture has the lowest occurrence, which is $y_{adr\cdot} = 1$. Let $t_i = y_{i\cdot} / \max(y_{i\cdot})$ to adjust for background noise $G(t_i)$, e.g., $t_i = 1/6513$

for hypotension. The results show that if we do not incorporate covariates information, PRR finds 229 signals; BCPNN finds 148 signals; the proposed approach without covariates (NHP) finds 57 signals, and ZIP LRT finds 62 signals.

It is worth noting that patients' age, gender, and country vary from drug to drug. For example, the average age for Amfetamine users is 35.25, and the average age for Apixaban users is 74.87 in this dataset. Moreover, the proportion of female users for Amiodarone is 0.41, but that for Ambrisentan users is 0.76. So it is crucial to adjust for confounders in the signal detection process to remove the imbalances resulted from patients' characteristics such as age, gender, and country information. When we incorporate such information such as country information, gender, and age, the proposed approach finds 89 signals. We compare our results with those signals found by those without incorporating demographic information.

Because the sensitivity of all the other approaches is lower than the proposed one with covariates information, PRR/BCPNN/ZIP LRT/NHP might miss some true positive signals. LRT misses 87 signals of the 89 signals found by the proposed approach, PRR misses 14 signals of the 89 signals, and BCPNN misses 29 signals of them. Even with PRR, BCPNN, and ZIP LRT combined, still 11 signals are missed among the 89 signals found by the proposed approach. For example, 809 flushing cases are reported on the Amlodipine; the PRR of the flushing on the drug and its 95%CI is 0.161 and (0.012,0.309), whose lower bound is smaller than 1. However, after incorporating covariates such as gender, based on the average level instead of PRR, the proposed approach found it as a signal. That is more likely a true signal because the proposed approach has higher specificity and sensitivity than others do. However, there are two signals found by all other approaches (PRR, BCPNN, NHP, ZIP LRT) but missed by the proposed approach. Since PRR/BCPNN/NHP/ZIP LRT do not control type

I error and FDR well when considering covariates, the two signals might be false positives.

2.7.2 Selected drugs and AEs (large drug-AE combinations) (VigiBase data 2)

Sometimes, our interest may be screening AE signals in a larger dataset. In this case, we investigate vascular or nervous system disorders-related AEs on 100 cardiovascular medicines, including the ten medicines we analyzed in VigiBase data 1. We are interested in both vascular and nervous system disorders-related AEs against at least one of the 100 cardiovascular medicines in the VigiBase. There are 1037 AE types against the 100 medicines after data processing. Then, we collapse the data into the form of 1037×100 table, and the proportion of zero counts is 78.4%, which is more than that in VigiBase data 1 as the types of AEs increase. Again, our goal is to identify AEs with high relative reporting rates for the 1037×100 drug-AE combinations. In this dataset, the total number of occurrence of AEs is $y_{..} = 627903$. Dizziness has the highest occurrence, which is 76986 in total ($\max(y_{i.}) = 76986$). We still use $t_i = y_{i.}/\max(y_{i.})$ to adjust for background noise $G(t_i)$ for AE i, e.g., $t_i = 1600/76986$ for ventricular extrasystoles, whose occurrence is 1600 during the study period. The analysis result shows that if we ignore covariates information included in the dataset, PRR finds 6006 signals; The BCPNN finds 5115 signals; NHP finds 1800 signals. Moreover, the ZIP LRT finds 1919 signals. The drug list in this dataset also includes the 10 drugs and vascular-related AEs as discussed in VigiBase data 1, so it is necessary to adjust for confounders such as age, gender, and country information. After incorporating covariates information, the proposed approach finds 2994 signals.

Furthermore, we compare the results with those signals found by approaches without incorporating demographic information since we do not want to miss true

positive signals because of the lower sensitivity with those approaches without covariates. LRT misses 2875 signals among 2994 signals found by the proposed approach, and PRR misses 1052 signals of them. Furthermore, BCPNN misses 908 signals. There are 854 signals that can only be found by the proposed approach. For example, palpitations is found as a signal on the Carvedilol by the proposed approach with covariates incorporated, while for the PRR, its estimated value and 95%CI is 0.849 and (0.736, 0.692), whose lower bound is smaller than one. Thus, it is not a signal by PRR. But after adjusting for the covariates, the proposed approach identifies the palpitations as an AE signal on the Carvedilol, which may draw additional attention. That is more likely a true signal because the proposed approach has higher specificity and sensitivity than do others. However, there are three signals found by all other approaches (PRR, BCPNN, NHP, ZIP LRT) but missed by the proposed approach. Again, since PRR/BCPNN/NHP/ZIP LRT do not control type I error and FDR well when considering covariates, the three signals might be false positives.

2.7.3 Selected drugs and their corresponding AEs (VigiBase data 3)

We consider a more general case, where the researchers aim to detect all AE signals against a certain number of medicines. In this analysis, we search signals in all reports submitted to WHO against at least one of the ten cardiovascular medicines we discussed in VigiBase data 1. There are 5312 AE types against at least one of the 10 medicines after data cleaning. So we collapse the data into a 5312×10 table, and 65.7% of the cell counts are zeros. Our goal is to detect AE signals in the panel after removing the background noise from all the other AEs. The total number of occurrences of AEs is $y_{\cdot\cdot} = 686661$. Dyspnoea has the highest occurrence ($\max(y_{i\cdot}) = 19731$). Then, we use $t_i = y_{i\cdot}/\max(y_{i\cdot})$ to adjust for background noise $G(t_i)$ for AE i. Obviously, t_i varies from AE to AE, which is reasonable for adjusting for different AEs, e.g.,

$t_i = 16191/19731$ for edema peripheral, whose occurrence is 16191 during the study period, while $t_i = 5627/19731$ for asthenia, whose occurrence is 5627 during the study period. The analysis result shows that if we do not adjust for existing demographic information, PRR finds 6193 signals; BCPNN finds 4721 signals; NHP finds 1554 signals, and ZIP LRT finds 1515 signals. We consider demographic information, including the mean age, the proportion of Americans, and the proportion of females as the covariates of each drug-AE combination in this case. After incorporating subjects' covariates, the proposed approach finds 2648 signals.

Among the 2648 signals found by the proposed approach, 2555 signals are missed by LRT, 643 by PRR, and 646 by BCPNN. For example, gastric infection is an AE signal of Amiodarone found by the proposed approach after incorporating region, gender, and age. 31 occurrences of gastric infection to various agents were reported on Amiodarone. With the PRR approach, its estimated value and 95%CI of this "signal" is 1.018 and (0.734,1.412), respectively, with a lower bound being smaller than one, and it is not considered a signal by the PRR approach. In addition, PRR/BCPNN/ZIP LRT together miss 539 of the 2648 signals, e.g., "blood potassium decreased" on Amlodipine, found by the proposed approach. However, there are ten signals found by all other approaches (PRR, BCPNN, NHP, ZIP LRT) but missed by the proposed approach. Again, since PRR/BCPNN/NHP/ZIP LRT do not control type I error and FDR well when considering covariates, the ten signals might be false positives.

2.7.4 Selected drugs and their corresponding AEs (VigiBase data 4)

In this case, we detect signals in a panel of drugs against all the AEs submitted, similar to example 3. There are 8379 AEs submitted against at least one of 146 drugs. After collapsing the data into an 8379×146 table, 79.1% of the cell counts are

zeros. The analysis aims to identify AEs with high relative reporting rates for the 10 drugs, respectively. If we do not have covariates information incorporated, PRR finds 65061 signals, BCPNN finds 60346 signals, NHP finds 15132 signals, and ZIP LRT finds 16796 signals. The proposed method finds 26041 signals adjusting for covariates. LRT misses 25164 signals among the 26041 signals found by the proposed method, PRR misses 9610 signals; BCPNN misses 7999 signals. For example, muscle contractions involuntary is an AE signal of Furosemide found by the proposed method after incorporating region, gender, and age. Only 176 occurrences of muscle contractions involuntary were reported on the Furosemide. the PRR and its 95%CI of this "signal" is 0.770 and (0.665,0.891), whose lower bound is smaller than 1. The signal would be missed if we apply PRR here. In addition, PRR/BCPNN/ZIP LRT combined miss 7692 signals found by the proposed method, such as chills on Clopidogrel. That is more likely a true signal because the proposed approach has higher specificity and sensitivity than others do. However, there are 27 signals found by all other approaches (PRR, BCPNN, NHP, ZIP LRT) but missed by the proposed approach. Again, since PRR/BCPNN/NHP/ZIP LRT do not control type I error and FDR well when considering covariates, the 27 signals might be false positives.

2.7.5 FDA data

In this FDA FAERs data, we aim to search for signals in all the reports submitted in quarter 4 of 2019. This dataset contains basic demographic information such as age, weights, and the place where the report was submitted. So we use weighted mean weights, weighted mean age, and females proportion as confounders. Here are 9262 AEs submitted against at least one of the 1657 drugs. After collapsing the data into a 9262×1657 table, 98.8% of the cell counts are zeros. The analysis aims to identify AEs with high relative reporting rates for the 10 drugs, respectively. If we

do not incorporate covariates information in the searching process, PRR finds 111664 signals; BCPNN finds 37845 signals; NHP finds 5778 signals, and ZIP LRT finds 13402 signals. If we consider the covariates we mention, the proposed method finds 7401 signals adjusting for covariates. Among these 7401 signals found by the proposed method, ZIP LRT misses 7296 signals, PRR misses 926 signals, and BCPNN misses 815 signals. PRR/BCPNN/ZIP LRT as an ensemble miss 791 signals. For example, pain in extremity on Apremilast can only be found by the proposed method. That is more likely a true signal because the proposed approach has higher specificity and sensitivity than others do. However, there are two signals found by all other approaches (PRR, BCPNN, NHP, ZIP LRT) but missed by the proposed approach. Again, since PRR/BCPNN/NHP/ZIP LRT do not control type I error and FDR well when considering covariates, the two signals might be false positives.

Finally, we want to reiterate (Huang et al., 2011) that caution needs to be exercised in terms of clinical interpretation because there may be a lack of biological plausibility for some drug-AE associations and a heterogeneous patient population even after adjusting for available covariates. However, it is recognized that it is useful to have a comparison within the same drug class or for the same or similar indication for class labeling and comparative effectiveness.

2.8 Discussion

We have proposed a general semi-parametric panel count model that can be applied to inferring associations between AEs and drugs while accounting for other available patient level information in post-market surveillance systems. A non-homogeneous Poisson process has been utilized to model the cell counts of the $I \times J$ AE-drug combinations. The non-parametric component $G(\cdot)$ of the model adjusts for back-

ground occurrence of AE, and parametric components incorporate covariates information while accounting for the excessive zeros with a latent variable. In addition, the approach finds signals by comparing the AE-drug combinations of interest with the expected values, which removes the bias that RR-based approaches share (Rothman et al., 2004). The proposed model is flexible and robust with respect to model assumptions, and it handles the excessive zero counts more accurately than does the ZIP model, which has been the main concern in detecting signals in surveillance systems.

Moreover, we have shown that the hypothesis testing based on the proposed model estimates is very effective in detecting signals while accounting for the joint effect of a large number of AEs. For example, LRT-based approaches (Huang et al., 2011) rely on Monto Carlo simulation to find the empirical distribution of the test statistics, which is time-consuming. Furthermore, the joint analysis allows us to consider the AEs collective roles on the drug. The proposed model also outperforms statistical learning approaches such as lasso to detect smaller effects because lasso usually omits small effects, which may yield collectively large effects. Simulation studies in this Chapter have demonstrated high power, sensitivity and specificity of the proposed approach. It is shown that existing methods such as PRR, BCPNN, and even the new approach ZIP LRT may miss signals stronger than average level (§2.6 & §2.7), demonstrating the proposed approach brings new insight into the safety of a drug. Finally and importantly, the approach with its ability to adjust for covariates may provide a powerful approach to explore data in the Sentinel System, which is being developed and where more information is collected and planned in the Sentinel System 5-Year Strategy.

In hypothesis testing, we have adopted the Holm-Bonferroni procedure to control FWER with greater power than the Bonferroni procedure. If the safety of a drug is well established, and a more conservative approach for monitoring further signals may be

warranted. Then, we may consider the classic Bonferroni procedure. We expect fewer signals with the classic Bonferroni correction. Some other approaches control FDR instead of FWER in the multiple comparison process, such as Benjamini-Hochberg's correction (Benjamini and Hochberg, 1995). It controls FDR, which can also control FWER but in a relatively weak sense, but it requires tests to be independent, which may be a questionable assumption.

Furthermore, the small effects in hypothesis testing have a remarkable resemblance to testing rare alleles in genetics study. We can translate tests in genetics to detecting signals. More specifically, as an example, recall the conditional mean under the alternative hypothesis in §2.5 is

$$E(y_{ij}(t)|\boldsymbol{x}_j) = G(t)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_{ij} + \sigma_{ij}).$$

Then, if σ_{ij} is significantly greater than zero, the ith AE is considered a signal on the jth drug. We denote $\boldsymbol{\sigma}$ be a $I \times J$ matrix with σ_{ij} as the element for row i column j. Obviously, $\boldsymbol{\sigma}$ is of high dimension. In reality, $\boldsymbol{\sigma}$ is very small componentwise as AEs are generally rare events, and our main interest is to test if elements of $\boldsymbol{\sigma}$ are significantly greater than zero or not. So we adopt the similar idea of SCARVA (simultaneous common and rare variants analysis) (Morris and Zeggini, 2010; Yuan et al., 2012) to test the null hypothesis $\sigma_{ij} = 0$. The method is originally applied in genetic analysis to test the significance of coefficients of each allele separately based on the likelihood ratio test. We can regard the AEs as a series of common alleles for a specific drug in our case. Then, we can use a similar procedure to test if the coefficients of AEs for a fixed drug are equal to zero. Denote the hypothesis of no signal for Drug j on AE i as $H_{0ij} : \sigma_{ij} = 0$. For a fixed j, let D_j be the jth column of the data and $D_{-i,j}$ be the ith row after removing the ith AE. Let $(\hat{\boldsymbol{\beta}}, \hat{G}, \hat{\delta}_{ij})$ be the same in §2.4. Under $(D_j, \hat{\boldsymbol{\beta}}, \hat{G}, \hat{\boldsymbol{\delta}})$, we calculate the MLE $\hat{\sigma}_{ij}$ of σ_{ij} for each AE i

with non-zero occurrence separately. The log-likelihood is, omitting some constant c independent of our parameter of interest,

$$\ell_J(\sigma_{ij}) = \sum_{i=1}^{k_j} y_{ij}\sigma_{ij} - \left[\sum_{i=1}^{k_j} \hat{G}(t_i)\exp\left(\hat{\boldsymbol{\beta}}^\intercal \boldsymbol{x}_{ij} + \sigma_{ij}\right) + \sum_{i=k_j+1}^{I} \hat{G}(t_i)\hat{\delta}_{ij}\exp\left(\hat{\boldsymbol{\beta}}^\intercal \boldsymbol{x}_{ij} + \sigma_{ij}\right)\right].$$

Therefore, we can estimate σ_{ij} by the MLE $\hat{\sigma}_{ij}$,

$$\hat{\sigma}_{ij} = \arg\max \ell_J(\sigma_{ij}).$$

Similarly, under $D_{-i,j}$, without loss of generality, we can estimate σ_{-ij} by the MLE $\hat{\sigma}_{-ij}$,

$$\hat{\sigma}_{-ij} = \arg\max \ell_{-i,J}(\sigma_{ij}).$$

Then, $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{G}(b_i), \hat{\sigma}_{ij})$. Under D_{-j}, we have $\hat{\boldsymbol{\theta}}_{-ij} = (\hat{\boldsymbol{\beta}}, \hat{G}(t_i), \hat{\sigma}_{i,-j})$. Let $\ell_j(\beta)$ be the log-likelihood for drug j, and let χ_1^2 be the centered chi-squared distribution with degree of freedom 1. If H_{0ij} is true, then

$$2\left[\ell(\hat{\sigma}_{ij}) - \ell(\hat{\sigma}_{-ij})\right] \sim \chi_1^2.$$

Given a significance level α, if

$$2\left[\ell(\hat{\sigma}_{ij}) - \ell(\hat{\sigma}_{-ij})\right] > \chi_1^2(1-\alpha),$$

we reject H_{0ij}, where $\chi_1^2(1-\alpha)$ is the $(1-\alpha)$-th upper bound of χ_1^2.

If H_{0ij} above is rejected and $\sigma_{-ij} > \sigma_{ij}$, the ith AE is a signal for the jth drug. We can also consider some correction to control FWER in this case.

In an initial simple simulation study, which is available upon request, the method shows high specificity with an acceptable sensitivity, which we will report in future work.

On the other hand, there are limitations to the proposed method in signal detection. One is that when there are a large number of potential signals in the same panel,

the expected counts of signal increase, which may give more false negatives. However, most existing approaches share the same problem. Another limitation, which in fact applies to most current methods, is that the correlation among AEs is ignored. Potentially such dependence can be modeled through shared frailty in the proposed model. Signal detection aims to bring our attention to potential signals and decide the next step towards the signal. If a confirmatory analysis is needed, the model may need to be further extended to allow causal inference on whether certain AEs have resulted from patients taking certain drugs. These possibilities will be explored in our future research.

Chapter 3

Theoretical properties of the semi-parametric panel count model

3.1 Introduction

In this Chapter, we derive the asymptotic properties of the proposed panel count model. Theorem 3.1 below demonstrates strong consistency of the estimators $(\hat{\boldsymbol{\beta}}, \hat{G})$; Theorem 3.2 below gives asymptotic normality and efficiency of the Euclidean component $\hat{\boldsymbol{\beta}}$; Theorem 3.3 gives the asymptotic distribution of $\hat{G}(\cdot)$ in a closed-form; and Theorem 3.4 displays the asymptotic distribution of $\hat{G}(t)$ under non-varnishing condition of its $m-th$ derivative for each point t.

Let $(\boldsymbol{\beta}_0, G_0)$ be the "true" parameters giving rise to the observed data, $\ell(\boldsymbol{\beta}, G|t, \boldsymbol{y}, \boldsymbol{x}) = \log f(t, \boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\beta}, \gamma, G)$, $\dot{\ell}_\beta(\boldsymbol{\beta}, G) = \partial\ell(\boldsymbol{\beta}, G|t, \boldsymbol{y}, \boldsymbol{x})/\partial\boldsymbol{\beta}$, and $\dot{\ell}_G(\boldsymbol{\beta}, G)[h]$ be the Hadamard derivative of $\ell(\boldsymbol{\beta}, G|t, \boldsymbol{y}, \boldsymbol{x})$ with respect to G in the direction h.

3.2 Regularity condition lists

To derive asymptotic properties of parametric component and non-parametric component of the proposed model, we first list the following regularity conditions:
(C3.1). $\boldsymbol{\beta}_0$ belongs to some bounded set, and $\gamma_0 \in [c, C]$ for some $0 < c < C < \infty$.
(C3.2). The support of $\boldsymbol{X}$ is bounded, and the support of T is $[0, B]$ for some $0 < B < \infty$.

(C3.3). $G_0 \in \mathcal{G}$ and $G_0(\cdot)$ is continuous.

(C3.4). $\sup_{G\in\mathcal{G}} \sup_{t\in[0,B]} G(t) < \infty$.

(C3.5). For all $G \in \mathcal{G}$, $G(\cdot)$ has derivative $\dot{G}(\cdot)$ which is uniformly bounded over $\mathcal{G}$.

(C3.6). $\sup_{(\beta,G)\in(\beta,\mathcal{G}):d(\beta-\beta_0,G-G_0)<\delta}(||\dot{\ell}_\beta(\boldsymbol{\beta},G)||_{L_2(P)} + ||\dot{\ell}_G(\boldsymbol{\beta},G)[1]||_{L_2(P)}) \leq C$.

(C3.7). I^*_β given in Theorem 3.2 is invertible.

(C3.8). Let $f(\cdot)$ be the density function of t, $\dot{G}_0(t) = dG_0(t)/dt$. Assume $f(t) > 0$ and $\dot{G}_0(t) > 0$ at the point t in Theorem 3.3 below.

(C3.9). $D(\cdot)$ and $r(\cdot)$ given in Theorem 3.3 below are continuous at t.

3.3 Strong consistency of the estimators

Theorem 3.1. *Assume (C3.1)-(C3.6), then for fixed J, as $I \to \infty$,*

$$\hat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}_0$$

$$\textit{and} \quad \sup_{t\in R^+} |\hat{G}(t) - G_0(t)| \xrightarrow{a.s.} 0.$$

Proof of Theorem 3.1. Let $h(\cdot)$ be the density of $\boldsymbol{X}$, and $r(\cdot)$ be the density of $\boldsymbol{y}$. Denote $S(\cdot|\boldsymbol{\beta},G) = p(\cdot|\boldsymbol{\beta},G)h(\cdot)r(\cdot)$ for the joint density of $(\boldsymbol{Y},\boldsymbol{X},\boldsymbol{t})$, and

$$q(\cdot|\boldsymbol{\beta},G) = \log\frac{S(\cdot|\boldsymbol{\beta},G)}{S(\cdot|\boldsymbol{\beta}_0,G_0)} = \log\frac{p(\cdot|\boldsymbol{\beta},G)}{p(\cdot|\boldsymbol{\beta}_0,G_0)}$$

for the log-likelihood ratio. Let P be the probability measure of $p(\cdot|\boldsymbol{\beta}_0,G_0)$.

The true mean of q is

$$Pq(\cdot|\boldsymbol{\beta},G) = \int q(\cdot|\boldsymbol{\beta},G)dP(\cdot),$$

and the empirical mean of q based on the data $\{(\boldsymbol{y}_i,\boldsymbol{x}_i,t_i); i = 1,...,I\}$ is

$$P_I q(\cdot|\boldsymbol{\beta},G) = I^{-1}\sum_{i=1}^{I} q(\cdot|\boldsymbol{\beta},G).$$

Note that $Pq(\cdot|\boldsymbol{\beta},G)$ is the negative Kullback-Leiber divergence between $S(\cdot|\boldsymbol{\beta},G)$ and $S(\cdot|\boldsymbol{\beta}_0,G_0)$, and as a function of $(\boldsymbol{\beta},G)$, $Pq(\cdot|\boldsymbol{\beta},G) \leq 0$. Since the model is

identifiable, $Pq(\cdot|\boldsymbol{\beta}, G) = 0$ only if $(\boldsymbol{\beta}, G) = (\boldsymbol{\beta}_0, G_0)$. Since $(\hat{\boldsymbol{\beta}}, \hat{G})$ is the MLE of $(\boldsymbol{\beta}_0, G_0)$, so

$$(\boldsymbol{\beta}_0, G_0) = \arg \max_{(\beta,G)\in(\beta,\mathcal{G})} Pq(\cdot|\boldsymbol{\beta}, G)$$

and

$$(\hat{\boldsymbol{\beta}}, \hat{G}) = \arg \max_{(\beta,G)\in(\beta,\mathcal{G})} P_I q(\cdot|\boldsymbol{\beta}, G).$$

Also, $Pq(\cdot|\boldsymbol{\beta}, G)$ is continuous with respect to $(\boldsymbol{\beta}, G)$ and the model is identifiable, so $(\boldsymbol{\beta}_0, G_0)$ is the unique maximizer of $Pq(\cdot|\boldsymbol{\beta}, G)$. Thus for all $\delta > 0$,

$$\sup_{(\beta,G)\in(\beta,\mathcal{G}):d(\beta-\beta_0,G-G_0)>\delta} Pq(\cdot|\boldsymbol{\beta}, G) < Pq(\cdot|\boldsymbol{\beta}_0, G_0).$$

By definition of MLE,

$$P_I q(\cdot|\hat{\boldsymbol{\beta}}, \hat{G}) \geq P_I q(\cdot|\boldsymbol{\beta}_0, G_0).$$

So by Theorem 5.8 in van der Vaart (2002), if $\mathcal{Q} = \{q(\cdot|\boldsymbol{\beta}, G) : (\boldsymbol{\beta}, G) \in (\boldsymbol{\beta}, \mathcal{G})\}$ is a Glivenko-Cantelli class, then

$$d[(\hat{\beta}, \hat{G}), (\boldsymbol{\beta}_0, G_0)] \xrightarrow{a.s.} 0.$$

Now we check $\mathcal{Q} = \{q(\cdot|\boldsymbol{\beta}, G) : (\boldsymbol{\beta}, G) \in (\boldsymbol{\beta}, \mathcal{G})\}$ is a Glivenko-Cantelli class.

Let $\dot{\ell}_\beta(\boldsymbol{\beta}, G) = \partial \ell_\beta(\boldsymbol{\beta}, G)/\partial \boldsymbol{\beta}$, $\dot{\ell}_G(\boldsymbol{\beta}, G)[h]$ be the Hadamard derivative for G at direction h, and $N_{[]}(\epsilon, \mathcal{Q}, L_2(P))$ be the minimum number of ϵ-brackets needed to

cover $\mathcal{Q}$ under the norm $||\cdot||_{L_2(P)}$. Note that

$$||q(\cdot|\boldsymbol{\beta}_1, G_1) - q(\cdot|\boldsymbol{\beta}_2, G_2)||_{L_1(P)} = ||\log S(\cdot|\boldsymbol{\beta}_1, G_1) - \log S(\cdot|\boldsymbol{\beta}_2, G_2)||_{L_1(P)}$$

$$=||(\dot{\ell}_\beta(\boldsymbol{\beta}, G)(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) + \dot{\ell}_G(\boldsymbol{\beta}, G)[G_1 - G_2]||_{L_1(P)}$$

(by Minkowski inequality)

$$\leq ||(\dot{\ell}_\beta(\boldsymbol{\beta}, G)(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)||_{L_1(P)} + ||\dot{\ell}_G(\boldsymbol{\beta}, G)[G_1 - G_2]||_{L_1(P)}$$

(by Höder's inequality)

$$\leq ||(\dot{\ell}_\beta(\boldsymbol{\beta}, G)||_{L_2(P)}||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2||_{L_2(P)} + ||\dot{\ell}_G(\boldsymbol{\beta}, G)||_{L_2(P)}||G_1 - G_2||_{L_2(P)}$$

(by (C3.6), $\exists\ 0 < C_1, C_2 < \infty$ such that)

$$\leq C_1||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2||_{L_2(P)} + C_2||G_1 - G_2||_{L_2(P)}.$$

Thus,

$$N_{[]}(\epsilon, \mathcal{Q}, ||\cdot||_{L_1(P)} \leq N_{[]}\left(\frac{\epsilon}{2C_1}, \beta, ||\cdot||_{L_2(P)}\right) N_{[]}\left(\frac{\epsilon}{2C_2}, \mathcal{G}, ||\cdot||_{L_2(P)}\right).$$

With (C3.4) and that $\mathcal{G}$ is a set of bounded and monotone increasing functions on $[0, B]$, by Theorem 2.7.5 in van der Vaart and Wellner (1996),

$$N_{[]}\left(\frac{\epsilon}{2C_2}, \mathcal{G}, ||\cdot||_{L_2(P)}\right) \leq \exp\left(\frac{C}{\epsilon}\right), \quad \forall \epsilon > 0.$$

With (C3.2) and $\boldsymbol{\beta}$ is Euclidean space, $N_{[]}(\frac{\epsilon}{2C_1}, \boldsymbol{\beta}, ||\cdot||_{L_2(P)}) = O(\epsilon^{-d})$, with $d = \dim(\boldsymbol{\beta})$,

$$N_{[]}(\epsilon, \mathcal{Q}, ||\cdot||_{L_1(P)}) \leq \frac{C}{\epsilon^d} \exp\left(\frac{C}{\epsilon}\right) < \infty, \quad \forall \epsilon > 0.$$

By Theorem 2.4.1 in van der Vaart and Wellner (1996), $\mathcal{Q} = \{q(\cdot|\boldsymbol{\beta}, G) : (\boldsymbol{\beta}, G) \in (\boldsymbol{\beta}, \mathcal{G})\}$ is a Glivenko-Cantelli class.

3.4 Asymptotic normality and efficiency of the Euclidean component

Denote $\xrightarrow{D}$ for convergence in distribution. Theorem 3.2 below gives asymptotic normality and efficiency of the Euclidean component $\hat{\boldsymbol{\beta}}$.

Theorem 3.2. *Assume (C3.1)-(C3.7), then for fixed J, as $I \to \infty$,*

$$\sqrt{I}\big(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\big) \xrightarrow{D} N(\mathbf{0}, (I^*_\beta)^{-1})$$

$$I^*_\beta = E_{(\beta_0, G_0)}[\ell^*_\beta (\ell^*_\beta)^\intercal],$$

where, under the "complete data" $D = \{(t, y_j, \boldsymbol{x}_j, z_j, \delta_j) : j = 1, ..., k\}$ with the δ_j's as missing data, the efficient score ℓ^*_β for $\boldsymbol{\beta}$ is given by

$$\ell^*_\beta = \ell^*_\beta(\boldsymbol{\beta}_0, G_0|D) = \sum_{j=1}^{J} \dot{\ell}_j \left\{ \boldsymbol{x}_j - \frac{E\left[\left(\sum_{j=1}^J \dot{\ell}_j \boldsymbol{x}_j\right)\left(\sum_{j=1}^J \dot{\ell}_j\right)|t\right]}{E\left[\left(\sum_{j=1}^J \dot{\ell}_j\right)^2 |t\right]} \right\},$$

where

$$\dot{\ell}_j = y_j - G_0(t)\big[\exp(\boldsymbol{\beta}_0^\intercal \boldsymbol{x}_j) I(y_j > 0) + \delta_j \exp(\boldsymbol{\beta}_0^\intercal \boldsymbol{x}_j) I(y_j = 0)\big].$$

Proof of Theorem 3.2. *Lemma 3.1 for proof of Theorem 3.2. Let $d(\boldsymbol{\beta}, G; \boldsymbol{\beta}_0, G) = ||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| + \sup_{t\in[0,B]} |G(t) - G_0(t)|$. Under conditions of Theorem 3.1,*

$$d(\boldsymbol{\beta}, G; \boldsymbol{\beta}_0, G_0) = O_p(I^{-1/3}).$$

Proof of Lemma 3.1. Define $\mathbb{M}_I(\boldsymbol{\beta}, G) = \mathbb{P}_I \ell(\boldsymbol{\beta}, G)$ and $\mathbb{M}(\boldsymbol{\beta}, G) = \mathbb{P}\ell(\boldsymbol{\beta}, G)$. Since $(\hat{\boldsymbol{\beta}}, \hat{G})$ is the MLE of $(\boldsymbol{\beta}_0, G_0)$, so

$$\mathbb{M}_I(\hat{\boldsymbol{\beta}}, \hat{G}) \geq \mathbb{M}_I(\beta, G) \geq \mathbb{M}_I(\beta, G) - O_p(r_I^{-2})$$

for any $(\boldsymbol{\beta}, G) \in (\boldsymbol{\Theta}, \mathcal{G})$ and any positive sequence $r_I \to \infty$. From Theorem 3.1, we can get

$$d(\boldsymbol{\beta}, G; \boldsymbol{\beta}_0, G) \xrightarrow{P} 0.$$

Note that

$$(\boldsymbol{\beta}_0, G_0) = \arg \sup_{(\beta,G)\in(\boldsymbol{\Theta},G)} \mathbb{M}(\boldsymbol{\beta}, G).$$

Denote $\dot{\ell}_\beta(\boldsymbol{\beta}_0, G_0)$, $\dot{\ell}_G(\boldsymbol{\beta}_0, G_0)[G - G_0]$, $\ddot{\ell}_{\beta,\beta}(\boldsymbol{\beta}, G)$, $\ddot{\ell}_{\beta,G}(\boldsymbol{\beta}, G)[G - G_0]$, $\ddot{\ell}_{G,\beta}(\boldsymbol{\beta}, G)[h]$ and $\ddot{\ell}_{G,G}(\boldsymbol{\beta}, G)[G - G_0, G - G_0]$ be the partial derivatives of $\ell(\boldsymbol{\beta}, G)$ with respect to $(\boldsymbol{\beta}, G)$. That for G is in the Hadamard sense.

Note that $E_{(\beta_0,G_0)}\dot{\ell}_\beta(\boldsymbol{\beta}_0, G_0) = \mathbf{0}$, $E_{(\beta_0,G_0)}\dot{\ell}_G(\boldsymbol{\beta}_0, G_0)[G - G_0] = \mathbf{0}$. Denote $\ddot{Q}(\boldsymbol{\beta}, G)[G - G_0, G - G_0]$ for the $(d + 1) \times (d - 1)$ matrix of expected all the second order partial derivatives, then by Taylor expansion,

$$\mathbb{M}(\boldsymbol{\beta}, G) - \mathbb{M}(\boldsymbol{\beta}_0, G_0) = \frac{1}{2}E_{(\beta_0,G_0)}\big[(\boldsymbol{\beta} - \boldsymbol{\beta}_0, 1)^\intercal \ddot{Q}(\tilde{\boldsymbol{\beta}}, \tilde{G})[G - G_0; g - g_0](\boldsymbol{\beta} - \boldsymbol{\beta}_0, 1)\big].$$

It can be proved readily that the above is of order $O(d^2(\boldsymbol{\beta}, G; \boldsymbol{\beta}_0, G_0))$, where $(\tilde{\boldsymbol{\beta}}, \tilde{G})$ is between $(\boldsymbol{\beta}, G)$ and $(\boldsymbol{\beta}_0, G_0)$.

So for any $0 < \eta_I \to 0$ and any η and $\eta < \infty$ with $\eta_I < \tau \leq \eta$, for some $0 < C < \infty$,

$$\sup_{\tau/2<d(\beta-\beta_0,G-G_0)\leq\tau,(\beta,G)\in(\beta,G)} \mathbb{M}(\boldsymbol{\beta}, G) - \mathbb{M}(\boldsymbol{\beta}_0, G_0) \leq -C\tau^2.$$

Let $\mathcal{M}_1 = \{\mathbb{M}(\boldsymbol{\beta}, G) : (\boldsymbol{\beta}, G) \in (\boldsymbol{\beta}, \mathcal{G})\}$. Similarly following the proof of Theorem 3.1, we can have $N_{[]}(\epsilon, \mathcal{M}_1, ||\cdot||_{L_2(P)} \leq Ce^\epsilon$ for some $0 < C < \infty$, and so

$$J_{[]}(\tau, \mathcal{M}_1, ||\cdot||_{L_2(P)} := \int_0^\tau \sqrt{1 + N_{[]}(\epsilon, \mathcal{M}_1, ||\cdot||_{L_2(P)}}d\epsilon = \int_{-\infty}^{\log\tau} e^s\sqrt{1 + Ce^{-s}}ds \leq C\tau^{1/2}.$$

In the above, we used the fact that for small $\tau > 0$, $\log\tau < 0$, so $e^{-s} > 1$ and $1 + Ce^{-s} \leq (1 + C)e^{-s}$ on $(-\infty, \log\tau)$.

Let $q(\cdot|\boldsymbol{\beta}, G)$ be as defined in the proof of Theorem 3.1, then

$$\mathcal{Q}(\tau) = \{q(\cdot|\boldsymbol{\beta}, G) : (\boldsymbol{\beta}, G) \in (\boldsymbol{\beta}, \mathcal{G}), \tau/2 < d(\boldsymbol{\beta}, G; \boldsymbol{\beta}_0, G_0) \leq \tau\},$$

then again we have $J_{[]}(\tau, \mathcal{Q}(\tau), ||\cdot||_{L_2(P)}$. Denote

$$\mathbb{W}_I q = \sqrt{I}(P_I - P)q$$

$$||\mathbb{G}_I||_{\mathcal{Q}(\tau)} = \sup_{q \in \mathcal{Q}(\tau)} |\mathbb{W}_I q|.$$

From the proof of Theorem 3.1, we have that $Pq^2 < C\tau^2$ and $||q||_\infty < C$ for all $q \in \mathcal{Q}(\tau)$ for some generic $0 < C < \infty$. By Lemma 3.4.2 in van der Vaart and Wellner (1996),

$$E^*||\mathbb{W}_I||_{\mathcal{Q}(\tau)} \leq CJ_{[]}(\tau, \mathcal{Q}(\tau), ||\cdot||_{L_2(P)}) \left[1 + \frac{CJ_{[]}(\tau, \mathcal{Q}(\tau), ||\cdot||_{L_2(P)})}{\tau^2\sqrt{I}} C\right]$$
$$\leq C\tau^{1/2}(1 + \tau^{-3/2}I^{-1/2}).$$

So with $\phi(\tau) = C\tau^{1/2}(1 + \tau^{-3/2}I^{-1/2})$ and $r_I = I^{1/3}$, we have

$$r_I^2 \phi_n \left(\frac{1}{r_I}\right) = CI^{1/3} \leq \sqrt{I}.$$

Now by the Theorem 3.4.1 in van der Vaart and Wellner (1996),

$$I^{1/3} d(\hat{\boldsymbol{\beta}}, \hat{G}; \boldsymbol{\beta}_0, G_0) = O_p(1).$$

Lemma 3.2 for proof of Theorem 3.2. Under the "complete data" $D = \{(t, y_j, \boldsymbol{x}_j, \delta_j) : j = 1, ...J\}$, the efficient score for estimating $\boldsymbol{\beta}_0$ is

$$\ell^*_\beta(\boldsymbol{\beta}, G|D) = \sum_{j=1}^J \dot{\ell}_j \left\{ \boldsymbol{x}_j - \frac{E\left[\left(\sum_{j=1}^J \dot{\ell}_j \boldsymbol{x}_j\right)\left(\sum_{j=1}^J \dot{\ell}_j\right)|T = t\right]}{E\left[\left(\sum_{j=1}^J \dot{\ell}_j\right)^2 |T = t\right]} \right\}$$

and

$$\dot{\ell}_j = y_j - G(t)\big[\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_j) I(y_j > 0) + \delta_j \exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_j) I(y_j = 0)\big].$$

Proof of Lemma 3.2. The log-likelihood under data $D = \{(t, y_j, \boldsymbol{x}_j, z_j, \delta_j) : j = 1, ..., k\}$ is

$$\ell(\boldsymbol{\beta}, G|D) = \log G(t) \sum_{j=1}^J y_{ij} + \sum_{j=1}^J y_j(\boldsymbol{\beta}^\intercal \boldsymbol{x}_j)$$

$$-G(t) \sum_{j=1}^J \left[\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_j) I(y_j \neq 0) + \delta_j^{(r)} \exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_j) I(y_j = 0)\right].$$

Denote $\dot{\ell}_\beta(\boldsymbol{\beta}, G|D) = \partial\ell(\boldsymbol{\beta}, G|D)/\partial\boldsymbol{\beta}$, and $\dot{\ell}_G(\boldsymbol{\beta}, G|D)[h]$ be the Hadamard derivative of $\ell(\boldsymbol{\beta}, G|D)$ with respect to G in the direction h. Then

$$\begin{aligned}\dot{\ell}_\beta(\boldsymbol{\beta}, G|D) &= \sum_{j=1}^{J} \left\{ y_j - G(t) \left[\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_j) I(y_j > 0) + \delta_j \exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_j) I(y_j = 0) \right] \right\} \boldsymbol{x}_j \\ &= \sum_{j=1}^{J} \dot{\ell}_j(\boldsymbol{\beta}, G|D) \boldsymbol{x}_j,\end{aligned}$$

where

$$\begin{aligned}\dot{\ell}_G(\boldsymbol{\beta}, G|D)[h] &= \frac{h(t)}{G(t)} \sum_{j=1}^{J} y_j - h(t) \sum_{j=1}^{J} \left[\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_j) I(y_j > 0) + \delta_j \exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_j) I(y_j = 0) \right] \\ &= \frac{h(t)}{G(t)} \sum_{j=1}^{J} \dot{\ell}_j(\boldsymbol{\beta}, G|D).\end{aligned}$$

The efficient score for estimating $\boldsymbol{\beta}_0$ is

$$\ell_\beta^{*\intercal}(\boldsymbol{\beta}_0, G_0|D) = \dot{\ell}_\beta(\boldsymbol{\beta}_0, G_0|D) - \dot{\ell}_G(\boldsymbol{\beta}_0, G_0|D)[\boldsymbol{h}^*],$$

where $\boldsymbol{h}^* = (h_1^*, ..., h_d^*)^\intercal$, $d = \dim(\boldsymbol{\beta})$, $\dot{\ell}_G(\boldsymbol{\beta}_0, G_0|D)[\boldsymbol{h}^*] = (\dot{\ell}_G(\boldsymbol{\beta}_0, G_0|D)[h_1^*], ..., \dot{\ell}_G(\boldsymbol{\beta}_0, G_0|D)[h_d^*])^\intercal$, and the least favorable direction $\boldsymbol{h}^*$ is determined by

$$\begin{aligned}\forall\ h, \quad \mathbf{0} &= E\left[\left(\dot{\ell}_\beta(\boldsymbol{\beta}_0, G_0|D) - \dot{\ell}_G(\boldsymbol{\beta}_0, G_0|D)[\boldsymbol{h}^*] \right) \dot{\ell}_G(\boldsymbol{\beta}_0, G_0|D)[h] \right] \\ &= E\left\{ E\left[\sum_{j=1}^{J} \dot{\ell}_i(\boldsymbol{\beta}_0, G_0|D) \left(\boldsymbol{x}_j - \frac{\boldsymbol{h}^*(t)}{G(t)} \right) \left(\sum_{j=1}^{J} \dot{\ell}_i(\boldsymbol{\beta}_0, G_0|D) \right) \right] \frac{h(t)}{G(t)} \Big| T = t \right\}.\end{aligned}$$

The above gives

$$E\left\{ \sum_{j=1}^{J} \dot{\ell}_j(\boldsymbol{\beta}_0, G_0|D) \left(\boldsymbol{x}_j - \frac{\boldsymbol{h}^*(t)}{G(t)} \right) \left[\sum_{j=1}^{J} \dot{\ell}_i(\boldsymbol{\beta}_0, G_0|D) \right] \right\} \equiv \mathbf{0}$$

or

$$\boldsymbol{h}^*(t) = G(t) \frac{E\left\{ \left[\sum_{j=1}^{J} \dot{\ell}_j(\boldsymbol{\beta}_0, G_0|D) \boldsymbol{x}_j \right] \left[\sum_{j=1}^{J} \dot{\ell}_j(\boldsymbol{\beta}_0, G_0|D) \right] \Big| T = t \right\}}{E\left\{ \left[\sum_{j=1}^{J} \dot{\ell}_j(\boldsymbol{\beta}_0, G_0|D) \right]^2 \Big| T = t \right\}},$$

thus,

$$\ell_\beta^*(\boldsymbol{\beta}, G|D) = \sum_{j=1}^{J} \dot{\ell}_j \left\{ \boldsymbol{x}_j - \frac{E\left[\left(\sum_{i=j}^{J} \dot{\ell}_j \boldsymbol{x}_j \right) \left(\sum_{j=1}^{J} \dot{\ell}_j \right) | T = t \right]}{E\left[\left(\sum_{j=1}^{J} \dot{\ell}_j \right)^2 | T = t \right]} \right\}.$$

The Hadamard derivative for G at the least favorable direction h^* is denoted as $\dot{\ell}_G(\beta, G)[h^*]$. Recall that $(\hat{\boldsymbol{\beta}}, \hat{G})$ is the MLE of $(\boldsymbol{\beta}_0, G_0)$, then $\dot{\ell}_\beta(\hat{\boldsymbol{\beta}}, \hat{G}) = 0$ and $\dot{\ell}_G(\hat{\beta}, \hat{G})[h^*] = 0$. Also, $P\dot{\ell}_\beta(\boldsymbol{\beta}_0, G_0) = 0$ and $P\dot{\ell}_G(\boldsymbol{\beta}_0, G_0)[h^*] = 0$. Let

$$\mathcal{M}_1 = \{\dot{\ell}_\beta(\boldsymbol{\beta}, G) : (\boldsymbol{\beta}, G) \in (\boldsymbol{\beta}, \mathcal{G})\}$$
$$\mathcal{M}_2 = \{\dot{\ell}_G(\boldsymbol{\beta}, G)[\boldsymbol{h}^*] : (\boldsymbol{\beta}, G) \in (\boldsymbol{\beta}, \mathcal{G})\}.$$

For some generic constant $0 < C < \infty$ and as in the proof of Theorem 3.1,

$$N_{[]}(\epsilon, \mathcal{M}_1, ||\cdot||_{L_1(P)} \leq N_{[]}(\frac{\epsilon}{2C_1}, \beta, ||\cdot||_{L_2(P)}) N_{[]}\left(\frac{\epsilon}{2C_2}, \mathcal{G}, ||\cdot||_{L_2(P)}\right)$$

and

$$N_{[]}(\epsilon, \mathcal{M}_1, ||\cdot||_{L_1(P)} \leq \exp\left(\frac{C}{\epsilon}\right)$$
$$J_{[]}(1, \mathcal{M}_1, ||\cdot||_{L_1(P)} \leq \int_0^1 \sqrt{\frac{C}{\epsilon}} d\epsilon < \infty,$$

so, by Theorem 6.8 in van der Vaart (2002), $\mathcal{M}_1$ is a Donsker class. Similarly, $\mathcal{M}_2$ is also a Donsker Class. Then, by Corollary 2.3.12 in van der Vaart and Wellner (1996),

$$|\sqrt{I}(P_I - P)\dot{\ell}_\beta(\hat{\boldsymbol{\beta}}, \hat{G}) - \sqrt{I}(P - P)\dot{\ell}_\beta(\boldsymbol{\beta}_0, G_0)| = o_p(1)$$

$$|\sqrt{I}(P_I - P)\dot{\ell}_G(\hat{\boldsymbol{\beta}}, \hat{G})[\boldsymbol{h}^*] - \sqrt{I}(P_I - P)\dot{\ell}_G(\boldsymbol{\beta}_0, G_0)[\boldsymbol{h}^*]| = o_p(1).$$

Since $(\hat{\boldsymbol{\beta}}, \hat{G})$ is the MLE, $P_I\dot{\ell}_\beta(\hat{\boldsymbol{\beta}}, \hat{G}) = 0$ and $P_I\dot{\ell}_G(\hat{\boldsymbol{\beta}}_0, \hat{G})[\boldsymbol{h}^*] = 0$, also, we have $P\dot{\ell}_\beta(\boldsymbol{\beta}_0, G_0) = 0$ and $P\dot{\ell}_G(\boldsymbol{\beta}_0, G_0)[\boldsymbol{h}^*] = 0$. From above, we can get

$$-\sqrt{I}P\dot{\ell}_\beta(\hat{\boldsymbol{\beta}}, \hat{G}) - \sqrt{I}P_I\dot{\ell}_\beta(\boldsymbol{\beta}_0, G_0) = o_p(1)$$

$$-\sqrt{I}P\dot{\ell}_G(\hat{\boldsymbol{\beta}}, \hat{G})[\boldsymbol{h}^*] - \sqrt{I}P_I\dot{\ell}_G(\boldsymbol{\beta}_0, G_0)[\boldsymbol{h}^*] = o_p(1).$$

By Lemma 1 and dominated convergence,

$$E(||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||^2 + ||\hat{G} - G_0||^2) = O(I^{-2/3}),$$

so, by Taylor expansion,

$$\dot{\ell}_\beta(\hat{\boldsymbol{\beta}},\hat{G})-\dot{\ell}_\beta(\boldsymbol{\beta}_0,G_0)=\ddot{\ell}_{\beta\beta}(\boldsymbol{\beta}_0,G_0)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+\ddot{\ell}_{\beta,G}(\boldsymbol{\beta}_0,G_0)[\hat{G}-G_0]+O(||\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}||^2+||\hat{G}-G_0||^2),$$

$$P\dot{\ell}_\beta(\hat{\boldsymbol{\beta}},\hat{G})-P\dot{\ell}_\beta(\boldsymbol{\beta}_0,G_0)-P\ddot{\ell}_{\beta\beta}(\boldsymbol{\beta}_0,G_0)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)-P\ddot{\ell}_{\beta,G}(\boldsymbol{\beta}_0,G_0)[\hat{G}-G_0]=O(I^{-2/3}).$$

From the above, we get

$$-P_I\dot{\ell}_\beta(\boldsymbol{\beta}_0,G_0)-P\ddot{\ell}_{\beta\beta}(\boldsymbol{\beta}_0,G_0)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)-P\ddot{\ell}_{\beta,G}(\boldsymbol{\beta}_0,G_0)[\hat{G}-G_0]=o_p(I^{-1/2}). \quad (3.1)$$

Similarly,

$$-P_I\dot{\ell}_G(\boldsymbol{\beta}_0,G_0)[\boldsymbol{h}^*]-P\ddot{\ell}_{\beta,G}(\boldsymbol{\beta}_0,G_0)[\boldsymbol{h}^*](\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)-P\ddot{\ell}_{G,G}(\boldsymbol{\beta}_0,G_0)[\boldsymbol{h}^*,\hat{G}-G_0]=o_p(I^{-1/2}). \quad (3.2)$$

Also, by definition of $\boldsymbol{h}^*$,

$$\begin{aligned}
&P\ddot{\ell}_{\beta,G}(\boldsymbol{\beta}_0,G_0)[\hat{G}-G_0]-P\ddot{\ell}_{G,G}(\boldsymbol{\beta}_0,G_0)[\boldsymbol{h}^*,\hat{G}-G_0]\\
&=-P\left[\dot{\ell}_\beta(\boldsymbol{\beta}_0,G_0)\dot{\ell}_G(\boldsymbol{\beta}_0,G_0)[\hat{G}-G_0]\right]+P\left[\dot{\ell}_G(\boldsymbol{\beta}_0,G_0)[\boldsymbol{h}^*]\dot{\ell}_G(\boldsymbol{\beta}_0,G_0)[\hat{G}-G_0]\right]\\
&=-P\left\{[\dot{\ell}_\beta(\boldsymbol{\beta}_0,G_0)-\dot{\ell}_G(\boldsymbol{\beta}_0,G_0)[\boldsymbol{h}^*]]\dot{\ell}_G(\boldsymbol{\beta}_0,G_0)[\hat{G}-G_0]\right\}\\
&=0.
\end{aligned}$$

Now (3.2)-(3.1) gives

$$\sqrt{I}P\left[\ddot{\ell}_{\beta\beta}(\boldsymbol{\beta}_0,G_0)-\ddot{\ell}_{\beta,G}(\boldsymbol{\beta}_0,G_0)[\boldsymbol{h}^*]\right]\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)=-\sqrt{I}P_I\left[\dot{\ell}_\beta(\boldsymbol{\beta}_0,G_0)-\dot{\ell}_G(\boldsymbol{\beta}_0,G_0)[\boldsymbol{h}^*]\right]+o_p(1).$$

Thus,

$$\begin{aligned}
&P\left[\ddot{\ell}_{\beta\beta}(\boldsymbol{\beta}_0,G_0)-\ddot{\ell}_{\beta,G}(\boldsymbol{\beta}_0,G_0)[\boldsymbol{h}^*]\right]\\
&=-P\left\{\left[\dot{\ell}_\beta(\boldsymbol{\beta}_0,G_0)(\dot{\ell}_\beta(\boldsymbol{\beta}_0,G_0)-\dot{\ell}_G(\boldsymbol{\beta}_0,G_0)[\boldsymbol{h}^*]\right]^\intercal\right\}\\
&=-P\left\{\left[\dot{\ell}_\beta(\boldsymbol{\beta}_0,G_0)-\dot{\ell}_G(\boldsymbol{\beta}_0,G_0)[\boldsymbol{h}^*]\right]\left[\dot{\ell}_\beta(\boldsymbol{\beta}_0,G_0)-\dot{\ell}_G(\boldsymbol{\beta}_0,G_0)[\boldsymbol{h}^*]\right]^\intercal\right\}\\
&=P\left\{\left[\dot{\ell}^*_\beta(\boldsymbol{\beta}_0,G_0)\right]\left[\dot{\ell}^*_\beta(\boldsymbol{\beta}_0,G_0)\right]^\intercal\right\}\\
&:=-I^*(\boldsymbol{\beta}_0,G_0).
\end{aligned}$$

Then

$$\sqrt{I}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = (I^*)^{-1}\sqrt{I}P_J\left[\dot{\ell}^*_{\beta}(\boldsymbol{\beta}_0, G_0)\right] + o_p(1).$$

This gives the desired result.

3.5 Asymptotic distribution of non-parametric component

Let $\mathbb{B}(\cdot)$ be the two-sided Brownian motion originating from zero: A mean zero Gaussian process on R with $\mathbb{B}(0) = 0$, and $E\big[\mathbb{B}(s) - \mathbb{B}(h)\big]^2 = |s - h|$ for all $s, h \in R$. Denote $\xrightarrow{D}$ for convergence in distribution. Theorem 3.3 gives the asymptotic distribution of $\hat{G}(t)$ under non-varnishing condition of its derivative for each point t.

Theorem 3.3. *Assume (C3.1)-(C3.9), then for fixed J, as $I \to \infty$,*

$$I^{1/3}\left[\hat{G}(t) - G_0(t)\right] \xrightarrow{D} \left[\frac{4\eta^2(t)D^2(t)\dot{G}_0(t)}{d^2(t)f(t)}\right]^{1/3} \arg\min_h \left[\mathbb{B}(h) + h^2\right],$$

where

$$\eta^2(t) = E(\epsilon_i^2|T = t)$$

$$\epsilon_i = y_{i\cdot}/d_i - G_0(t)$$

$$d_i = \sum_{j=1}^{J}\left[\exp\left(\boldsymbol{\beta}_0^{\intercal}\boldsymbol{x}_{ij}\right) I(y_{ij} \neq 0) + \delta_{ij}^{(r)} \exp\left(\boldsymbol{\beta}_0^{\intercal}\boldsymbol{x}_{ij}\right) I(y_{ij} = 0)\right]$$

$$d(t) = E(d_i|T = t)$$

$$D^2(t) = E(d_i^2|T = t).$$

Proof of Theorem 3.3. Note that

$$\begin{aligned}
\hat{G}(\cdot) &= \arg\max_{G\in\mathcal{G}} \ell(\hat{\boldsymbol{\beta}}, G) \\
&= \arg\max_{G\in\mathcal{G}} \left[\frac{1}{I}\sum_{i=1}^{I} y_{i\cdot}\log G(t_i) - \hat{d}_i G(t_i)\right] \\
&= \arg\max_{G\in\mathcal{G}} \left[\frac{1}{I}\sum_{i=1}^{I} y_{i\cdot}\log G(t_i) - d_i G(b_i)\right] + O(I^{-1/2}) \\
&= \tilde{G}(t) + O(I^{-1/2})
\end{aligned}$$

$$I^{1/3}(\hat{G}(t) - G_0(t)) = I^{1/3}[\tilde{G}(t) - G_0(t)] + O(I^{-1/6}),$$

where $d_i = \sum_{j=1}^{J}\left[\exp(\boldsymbol{\beta}_0^\intercal \boldsymbol{x}_{ij}) I(y_{ij}\neq 0) + \delta_{ij}^{(r)}\exp(\boldsymbol{\beta}_0^\intercal \boldsymbol{x}_{ij}) I(y_{ij}=0)\right]$, and $\hat{d}_i$ is d_i with $\boldsymbol{\beta}_0$ replaced by $\hat{\boldsymbol{\beta}}$. Then, without loss of generality, we assume

$$\hat{G}(t) = \arg\min_{G\in\mathcal{G}} \left[\frac{1}{I}\sum_{i=1}^{I} y_{i\cdot}\log G(t_i) - d_i G(t_i)\right].$$

By Theorem 1.5.1 and Example 1.5.1 in Robertson et al. (1988), with similar procedure in estimation of $G(\cdot)$, the above is

$$\begin{aligned}
&\arg\min_{G\in\mathcal{G}} \frac{1}{I}\sum_{i=1}^{I} d_i \left[\frac{y_{i\cdot}}{d_i} - G(t_i)\right]^2 \\
&= \arg\min_{G\in\mathcal{G}} \frac{1}{I}\sum_{i=1}^{I} d_i \left[G(t_i) - G_0(t_i) + G_0(t_i) - \frac{y_{i\cdot}}{d_i}\right]^2 \\
&= \arg\min_{G\in\mathcal{G}} \frac{1}{I}\sum_{i=1}^{I} d_i \left[G(t_i) - G_0(t_i) - \epsilon_i\right]^2 \\
&= \arg\min_{G\in\mathcal{G}} \sum_{i=1}^{I} d_i \left[G(t_i) - G_0(t_i) - \epsilon_i\right]^2.
\end{aligned}$$

Note that

$$\begin{aligned} E(\epsilon_i) &= E\left[E(\epsilon_i|\boldsymbol{x}_i, t_i)\right] \\ &= E\left[y_{\cdot i}/d_i - G_0(t_i)\right] \\ &= \sum_{j=1}^{J} E(y_{ij})/d_i - G_0(t_i) = G_0(t_i)d_i/d_i - G_0(t_i) \\ &= 0. \end{aligned}$$

Following convention, upper letters stand for random variables and lower letters for the observations. Define

$$U_I(t) = P_I\left[DI(T \le t)\right] = \frac{1}{I}\sum_{i=1}^{I} d_i I(t_i \le t)$$

$$V_I(t) = P_I\left[D(G_0(T) + \epsilon)I(T \le t)\right] = \frac{1}{I}\sum_{i=1}^{I} d_i\left[G_0(t_i) + \epsilon_i\right] I(t_i \le t).$$

By Example 3.2.15 in van der Vaart and Wellner (1996), for all $t, a \in R$, $\hat{G}(t) \le a$ iff $\arg\min_s\{V_I(s) - aU_I(s)\} \ge t$.

Note that $\left\{I^{1/3}\left[\hat{G}_0(t) - G_0(t)\right] \le c\right\} = \left\{\hat{G}_0(t) \le G_0(t) + I^{-1/3}c\right\}$, so, with $a = G_0(t) + cI^{-1/3}$,

$$P\left[\hat{G}(t) \le G_0(t) + cI^{-1/3}\right] = P\left[\arg\min_s\{V_I(s) - (G_0(t) + cI^{-1/3})U_J(s)\} - t \ge 0\right].$$

By change of variables $s = t + I^{-1/3}h$, we get

$$\begin{aligned} &\arg\min_s\left\{V_I(s) - [G_0(t) + cI^{-1/3}]U_I(s)\right\} - t \\ &= I^{-1/3}\arg\min_h\left\{V_I(t + I^{-1/3}h) - [G_0(t) + cI^{-1/3}]U_I(t + I^{-1/3}h)\right\}. \end{aligned}$$

Let $\hat{h}_I$ be the above minimizer. From above, we need to compute the limit of $P(\hat{h}_I \geq 0)$. Note that

$$\begin{aligned}
\hat{h}_I &= \arg\min_h \left\{V_I(t+I^{-1/3}h) - (G_0(t)+cI^{-1/3})U_I(t+I^{-1/3}h)\right\} \\
&= \arg\min_h \{P_I[D(G_0(T)+\epsilon)I(T \leq t+I^{-1/3}h)] \\
&- (G_0(t)+cI^{-1/3})P_I[DI(\boldsymbol{T} \leq t+I^{-1/3}h)]\} \\
&= \arg\min_h \{P_I[D(G_0(T)-G_0(t)+\epsilon)I(T \leq t+I^{-1/3}h)] \\
&- cI^{-1/3}P_I[DI(T \leq t+I^{-1/3}h)]\} \\
&= \arg\min_h \{P_I[D(G_0(T)-G_0(t)+\epsilon)I(t < T \leq t+I^{-1/3}h)] \\
&- cI^{-1/3}P_I[DI(t < T \leq t+I^{-1/3}h)]\} \\
&= \arg\min_h \{(P_I-P)[D(G_0(T)-G_0(t)+\epsilon)I(t < C \leq t+I^{-1/3}h)] \\
&- cI^{-1/3}P_I[DI(t < T \leq t+I^{-1/3}h)] \\
&+ P[D(G_0(T)-G_0(t)+\epsilon)I(t < T \leq t+I^{-1/3}h)]\} \\
&= \arg\min_h \{I^{2/3}(P_I-P)[D(G_0(T)-G_0(t)+\epsilon)I(t < T \leq t+I^{-1/3}h)] \\
&- cI^{1/3}P_I[DI(t < T \leq t+I^{-1/3}h)] \\
&+ I^{2/3}P[D(G_0(T)-G_0(t)+\epsilon)I(t < T \leq t+I^{-1/3}h)]\} \\
&:= \arg\min_h \{B_{1,I}(h)+B_{2,I}(h)+B_{3,I}(h)\}, \\
B_{1,I}(h) &= I^{2/3}(P_I-P)[D(G_0(T)-G_0(t)+\epsilon)I(t < T \leq t+I^{-1/3}h)] \\
&= I^{1/2}I^{1/6}(P_J-P)[D(G_0(T)-G_0(t)+\epsilon)I(t < T \leq t+I^{-1/3}h)] \\
&= I^{1/2}(P_I-P)S_{I,h}(X),
\end{aligned}$$

where

$$S_{I,h}(X) = I^{1/6}[D(G_0(T)-G_0(t)+\epsilon)I(t < T \leq t+I^{-1/3}h)].$$

Let $\mathcal{S} = \{S_{I,h} : h \in [-K, K]\}$ for some $0 < K < \infty$. Then $\mathcal{S}$ has an envelop $S_I(c) = I^{1/6}C(|1+\epsilon|)|(I[t - KI^{-1/3} < T \leq t] + I[t < T \leq t + KI^{-1/3}])$ with some $0 < C < \infty$. Below we check the conditions of Theorem 2.11.23 in van der Vaart and Wellner (1996). Note that

$$\begin{aligned} PS_I^2 &= I^{1/3}E\{C^2[I(t - KI^{-1/3} < T \leq t) + I(t < T \leq t + KI^{-1/3})]\} \\ &= I^{1/3}E\{C^2[I(t - KI^{-1/3} < T \leq t + KI^{-1/3})]\} \\ &= I^{1/3}C^2 \int_{t-KI^{-1/3}}^{t+KI^{-1/3}} f(T)d \sim 2Kf(t), \end{aligned}$$

so,

$$PS_I^2 = O(1).$$

Note that $S_I(t) \leq I^{1/6}$, so for each $\eta > 0$, $I(S_I > \eta\sqrt{I}) = 0$ for each $I \leq N$, with $N \geq 1/(2\eta)^3$, we have $PS_I^2 I(S_I > \eta\sqrt{I}) \to 0$. Also,

$$PS_{I,h}(t) = I^{1/6}E[E[D|T = t](G_0(T) - G_0(t) + \epsilon)I(t < T \leq t + I^{-1/3}h)] = O(I^{-1/6}) \to 0,$$

and with $D^2(t) := E[D^2|T = t]$, and $\eta^2(t) = E(\epsilon^2|T = t)$,

$$\begin{aligned} &P[S_{I,h_1}(t)S_{I,h_2}(t)] = I^{1/3}E\left\{D^2(G_0(T) - G_0(t) + \epsilon)^2 I(t < T \leq t + I^{-1/3}(h_1 \wedge h_2)\right\} \\ &= I^{1/3}E\left\{E\left(D^2(G_0(T) - G_0(t) + \epsilon)^2 I(t < T \leq t + I^{-1/3}(h_1 \wedge h_2)\middle|T\right)\right\} \\ &= I^{1/3}E\left\{E\left(D^2\left[(G_0(T) - G_0(t))^2 + 2(G_0(T) - G_0(t))\epsilon + \epsilon^2\right] I(t < T \leq t + I^{-1/3}(h_1 \wedge h_2)\middle|T\right)\right\} \\ &\sim I^{1/3}E\left\{E\left(D^2\left[\dot{G}_0^2(t)(T-t)^2 + 2\dot{G}_0(t)(T-t)\epsilon + \epsilon^2\right] I(t < T \leq t + I^{-1/3}(h_1 \wedge h_2)\middle|T\right)\right\} \\ &\sim I^{1/3}E\left\{E\left(D^2\epsilon^2 I(t < T \leq t + I^{-1/3}(h_1 \wedge h_2)\middle|T\right)\right\} \\ &= I^{1/3}\int_t^{t+I^{-1/3}(h_1\wedge h_2)} E\left(D^2\epsilon^2|T = s\right) f(s)ds. \end{aligned} \tag{3.3}$$

It is typical to assume that conditioning on T, D and ϵ are independent, so with condition (C3.9), the above (3.3) is

$$= I^{1/3}\int_t^{t+I^{-1/3}(h_1\wedge h_2)} D^2(s)\epsilon^2(s)f(s)ds \sim (h_1 \wedge h_2)D^2(t)f(t)\eta^2(t).$$

Thus as $I \to \infty$, $P[S_{I,h_1}(t)S_{I,h_2}(t)] \sim (h_1 \wedge h_2)D^2(t)f(t)\eta^2(t)$. This shows that $cov(S_{I,h_1,t}, S_{J,h_2,t})$ converges to the covariance function of the process $\eta(t)D(t)f^{1/2}(t)\mathbb{B}(\cdot)$, where $\mathbb{B}(h)$ is a zero mean Gaussian process on R with covariance function $E[\mathbb{B}(h_1)\mathbb{B}(h_2)] = h_1 \wedge h_2$.

Also, similarly as in the proof of Lemma 1, it can be derived that

$$\int_0^{\delta_n} \sqrt{N_{[]}(\epsilon||G_n||_{P,2}, \mathcal{S}, L_2(P))}d\epsilon < \infty.$$

Thus by Theorem 2.11.23 in van der Vaart and Wellner (1996),

$$B_{1,I}(h) \xrightarrow{D} \eta(t)D(t)f^{1/2}(t)\mathbb{B}(h), \quad \text{in } \ell^\infty[-K, K].$$

It can shown, as in van der Vaart and Wellner (1996), $\hat{h} = \arg\min_h\{B_{1,I}(h)+B_{2,I}(h)+B_{3,I}(h)\}$ is bounded in probability, and so the above weak convergence can be regarded as in $\ell^\infty(R)$.

Now we handle $B_{2,I}(h)$ and $B_{3,I}(h)$. With $d(t) = E(D|T = t]$,

$$\begin{aligned} B_{2,I}(h) &= -cI^{1/3}P_I[DI(t < T \le t + I^{-1/3}h)] \\ &= -cI^{1/3}[d(t) + o(1)] \int_t^{t+I^{-1/3}h} f(s)ds \to -cd(t)f(t)h, \end{aligned}$$

$$\begin{aligned} B_{3,I}(h) &= I^{2/3}Pd(T)[G_0(T) - G_0(t) + \boldsymbol{\epsilon}]I(t < T \le t + I^{-1/3}h) \\ &= I^{2/3}[d(t) + o(1)] \int_t^{t+I^{-1/3}h} (G_0(s) - G_0(t))f(s)ds \\ &= I^{2/3}[d(t) + o(1)][f(t) + o(1)] \int_t^{t+I^{-1/3}h} \dot{G}(t)(s - t)ds \\ &\to \frac{1}{2}d(t)f(t)\dot{G}(t)h^2. \end{aligned}$$

Now combine the asymptotic results of $B_{1,I}(h)$, $B_{2,I}(h)$ and $B_{3,I}(h)$, we have

$$\begin{aligned} &\arg\min_h\{B_{1,I}(h) + B_{2,I}(h) + B_{3,I}(h)\} \\ &\xrightarrow{D} \arg\min_h\{\eta(t)D(t)f^{1/2}(t)\mathbb{B}(h) - cd(t)f(t)h + \frac{1}{2}d(t)f(t)\dot{G}(t)h^2\}. \end{aligned}$$

Using Problem 3.2.5 in van der Vaart and Wellner (1996), the above is re-written as

$$\left\{\frac{4\eta^2(t)D^2(t)f(t)}{[d(t)f(t)\dot{G}(t)]^2}\right\}^{1/3}\arg\min_h[\mathbb{B}(h)-h^2]+\frac{c}{\dot{G}(t)}.$$

Then,

$$\begin{aligned}
&P[I^{1/3}(\hat{G}(t)-G(t))\le c]\\
&=P\left\{\left\{\frac{4\eta^2(t)D^2(t)f(t)}{[d(t)f(t)\dot{G}(t)]^2}\right\}^{1/3}\arg\min_h[\mathbb{B}(h)-h^2]+\frac{c}{\dot{G}(t)}\ge 0\right\}\\
&=P\left\{\left\{\frac{4\eta^2(t)D^2(t)f(t)\dot{G}(t)}{[d(t)f(t)]^2}\right\}^{1/3}\arg\max_h[-\mathbb{B}(h)+h^2]\le c\right\}\\
&=P\left\{\left\{\frac{4\eta^2(t)D^2(t)f(t)\dot{G}(t)}{(d(t)f(t))^2}\right\}^{1/3}\arg\min_h\{\mathbb{B}(h)+h^2\}\le c\right\}\\
&=P\left\{\left\{\frac{4\eta^2(t)D^2(t)\dot{G}(t)}{d^2(t)f(t)}\right\}^{1/3}\arg\min_h[\mathbb{B}(h)+h^2]\le c\right\},
\end{aligned}$$

which completes the proof.

3.6 General asymptotic distribution of the derivatives of non-parametric component

In Theorem 3.3, it is assumed that $\dot{G}_0(t)>0$, and it is interesting to know what happens when $\dot{G}_0(t)=0$. Theorem 3.4 states that if $G_0(\cdot)$ has k-th derivative $G_0^{(k)}(t)\neq 0$ and $G_0^{(m)}(t)=0$ for $m=1,...k-1$, then $\hat{G}(t)$ has convergence rate $I^{k/(2k+1)}$ faster than the rate $I^{1/3}$ in Theorem 3.3. This result is new for this type of estimators.

Theorem 3.4. *Under conditions of Theorem 3.3, with the condition* $\dot{G}_0(t)>0$ *replaced by* $G_0^{(m)}(t)=0$ *for* $m=1,...,k-1$, *and* $G_0^{(k)}(t)\neq 0$. *Then for fixed* J, *as* $I\to\infty$,

$$I^{k/(2k+1)}\left[\hat{G}(t)-G_0(t)\right]\xrightarrow{D}Z,$$

where the distribution function of Z is given by, $\forall x \in R$,

$$P(Z \le x) = P\left\{ \arg\min_h \left[\eta(t) D(t) f(t)^{1/2} \mathbb{B}(h) - x d(t) f(t) h + d(t) f(t) G_0^{(k)}(t) h^{k+1}/(k+1)! \right] \ge 0 \right\},$$

where $\eta(t), d(t)$ and $D(t)$ are given in Theorem 3.3.

Proof of Theorem 3.4. By the proof of Theorem 3.3, and using notations given in there, for $c \in R$,

$$[I^{k/(2k+1)}(\hat{G}_0(t) - G_0(t)) \le c] = [\hat{G}_0(t) \le G_0(t) + I^{-k/(2k+1)} c.$$

So, with $a = G_0(t) + cI^{-k/(2k+1)}$,

$$P[\hat{G}(t) \le G_0(t) + cI^{-k/(2k+1)}] = P\{\arg\min_s [V_I(s) - (G_0(t) + cI^{-k/(2k+1)}) U_J(s)] - t \ge 0\}.$$

By change of variables $s = t + I^{-1/(2k+1)} h$, we get

$$\arg\min_s \{V_I(s) - [G_0(t) + cI^{-k/(2k+1)}] U_I(s)\} - t$$
$$= I^{-1/(2k+1)} \arg\min_h \{V_I(t + I^{-1/(2k+1)} h) - [G_0(t) + cI^{-k/(2k+1)}] U_I(t + I^{-1/(2k+1)} h)\}.$$

Let $\hat{h}_I$ be the above minimizer. From above, we need to compute the limit of $P(\hat{h}_I \ge 0)$. Similarly as in the proof of Theorem 3.3, we have

$$
\begin{aligned}
\hat{h}_I &= \arg\min_h\{V_I(t+I^{-1/(2k+1)}h) - (G_0(t)+cI^{-k/(2k+1)})U_I(t+I^{-1/(2k+1)}h)\} \\
&= \arg\min_h\{P_I[D(T_i)(G_0(T_i)+\epsilon)I(T \le t+I^{-1/(2k+1)}h)] \\
&- (G_0(t)+cI^{-k/(2k+1)})P_I[DI(T \le t+I^{-1/(2k+1)}h)]\} \\
&= \arg\min_h\{P_I[D(G_0(T)-G_0(t)+\epsilon)I(T \le t+I^{-1/(2k+1)}h)] \\
&- cI^{-k/(2k+1)}P_I[DI(T \le t+I^{-1/(2k+1)}h)]\} \\
&= \arg\min_h\{P_I[D(G_0(T)-G_0(t)+\epsilon)I(t < T \le t+I^{-1/(2k+1)}h)] \\
&- cI^{-k/(2k+1)}P_I[DI(t < T \le t+I^{-1/(2k+1)}h)]\} \\
&= \arg\min_h\{(P_I-P)[D(G_0(T)-G_0(t)+\epsilon)I(t < C \le t+I^{-1/(2k+1)}h)] \\
&- cI^{-k/(2k+1)}P_I[DI(t < T \le t+I^{-1/(2k+1)}h)] \\
&+ P[w(T)(G_0(T)-G_0(t)+\epsilon)I(t < T \le t+I^{-1/(2k+1)}h)]\} \\
&= \arg\min_h\{I^{(k+1)/(2k+1)}(P_I-P)[D(T)(G_0(T)-G_0(t)+\epsilon)I(t < T \le t+I^{-1/(2k+1)}h)] \\
&- cI^{1/(2k+1)}P_I[DI(t < T \le t+I^{-1/(2k+1)}h)] \\
&+ I^{(k+1)/(2k+1)}P[D(G_0(T)-G_0(t)+\epsilon)I(t < T \le t+I^{-1/(2k+1)}h)]\} \\
&:= \arg\min_h\{B_{1,I}(h)+B_{2,I}(h)+B_{3,I}(h)\}.
\end{aligned}
$$

Furthermore, we have

$$
\begin{aligned}
B_{3,I}(h) &= I^{(k+1)/(2k+1)}P[D(G_0(T)-G_0(t)+\epsilon)I(t < T \le t+I^{-1/(2k+1)}h)] \\
&= I^{(k+1)/(2k+1)}(d(t)+o(1))\int_t^{t+I^{-1/(2k+1)}h}(G_0(s)-G_0(t))f(s)ds \\
&= I^{(k+1)/(2k+1)}(d(t)+o(1))(f(t)+o(1))\int_t^{t+I^{-1/(2k+1)}h}G^{(k)}(t)(s-t)^k/k!ds \\
&\sim d(t)f(t)G^{(k)}(t)h^{k+1}/(k+1)!
\end{aligned}
$$

$$
\begin{aligned}
B_{2,I}(h) &= -cI^{1/(2k+1)}P_I[DI(t < T \le t+I^{-1/(2k+1)}h)] \\
&\sim -cI^{1/(2k+1)}(d(t)+o(1))\int_t^{t+I^{-1/(2k+1)}h}f(s)ds \sim -cd(t)f(t)h
\end{aligned}
$$

$$B_{1,I}(h) = I^{(k+1)/(2k+1)}(P_I - P)[D(G_0(T) - G_0(t) + \epsilon)I(t < T \le t + I^{-1/(2k+1)}h)].$$

Define

$$S_{I,h}(t) = I^{1/2(2k+1)}D[G_0(T) - G_0(t) + \epsilon]I[t < x \le t + I^{-1/(2k+1)}h],$$

then $B_{1,I}(h) = I^{1/2}(P_I - P)S_{I,h}(t)$.

Let $\mathcal{S} = \{S_{I,h} : h \in [-K, K]\}$ for some $0 < K < \infty$. Then $\mathcal{S}$ has an envelop $S_I(c) = I^{1/2(2k+1)}C(|1 + \boldsymbol{\epsilon}|)|(I[t - KI^{-1/(2k+1)} < T \le t] + I[t < T \le t + KI^{-1/(2k+1)}])$ with some $0 < C < \infty$. Similarly as in the proof of Theorem 3.3,

$$\begin{aligned} P[S_{I,s}(t)S_{I,r}(t)] &\sim n^{1/(2k+1)}E\int_t^{t+(r\wedge s)I^{-1/(2k+1)}} D^2(G_0(s) - G_0(t) + \epsilon)^2 f(s)ds \\ &\sim D^2(t)\eta^2(t)f(t)(r \wedge s) \end{aligned}$$

and

$$B_{1,n}(h) \xrightarrow{D} D(t)\eta(t)f^{1/2}(t)\mathbb{B}(h).$$

Thus again,

$$\begin{aligned} &\arg\min_h\{B_{1,I}(h) + B_{2,I}(h) + B_{3,I}(h)\} \\ &\xrightarrow{D} \arg\min_h\{\eta(t)D(t)f^{1/2}(t))\mathbb{B}(h) - cd(t)f(t)h + d(t)f(t)G^{(k)}(t)h^{k+1}/(k+1)!\}. \end{aligned}$$

From the above we obtain, for all $c \in R$,

$$\begin{aligned} &\{I^{k/(2k+1)}(\hat{G}_0(t) - G_0(t)) \le c\} \\ &\to P\big\{\arg\min_h[\eta(t)D(t)f(t)^{1/2}\mathbb{B}(h) - cd(t)f(t)h + d(t)f(t)G^{(k)}(t)h^{k+1}/(k+1)!] \ge 0\big\}. \end{aligned}$$

This completes the proof.

Chapter 4

Inferring causal effects of medications and vaccines on adverse events from observational and non-randomized studies with count outcomes

4.1 Introduction

Inferring causal effects of medicines/vaccines on potential AEs is of great public health importance. When post-market monitoring suggests patients experience specific AEs after prescribing medicines or receiving vaccines, regulators need to balance evidence from case reports, randomized trials, meta-analyses of randomized trials, and observational studies to make further decision regarding the safety of the product because associations found in the post-market surveillance systems do not equal causality (Davies and Thomas, 2017). The gold standard in inferring causal effect is randomized experiments, which can provide the most substantial evidence to causal associations. But randomized experiments are not always feasible due to ethical or practical concerns. In this case, one may draw causal inference from observational studies with non-random group assignment.

In observational studies without treatment randomization, treatment assignment might be correlated with potential study outcomes. Thus, the basis of inferring causal effects from observational studies is that all the possible confounders that are related to both the potential study outcome and the treatment exposure have to be adjusted

for in the analysis (Hernán et al., 2008). Assuming all the confounders are measured, the most straightforward estimation of causal effect is adjusting for the confounders by a regression model (Rawlings et al., 1998) such as a linear and logistic regression model. However, the estimators are biased when the regression model is not correctly specified. The inverse propensity score weighted estimator (IPSWE) proposed by Horvitz and Thompson (1952) and Rosenbaum and Rubin (1983) give unbiased estimators of the treatment effects if the propensity score model is correctly specified. Otherwise, it is still biased. The more recent doubly robust estimator (DRE) for continuous outcomes (Robins et al., 1994, 2000; Scharfstein et al., 1999; Rotnitzky et al., 2012) represents a significant advancement in the field and is unbiased if either the propensity score model or the outcome model is correctly specified, namely, as long as one of the two models is correct.

Despite the progress with the continuous outcome, little attention has been given to non-negative integer outcomes. However, non-negative integer outcomes occur in a variety of pharmaceutical and epidemiological studies, for example, CD4 T counts (Hammer et al., 1996), number of hot flushes (Prague et al., 2017), and suicide attempts (Olfson et al., 2006), neuropsychiatric events (Evins et al., 2019). To draw causal inference from the studies with a non-negative integer outcome, some researchers regarded the count outcome as a binary variable (Cunningham et al., 2016), or a continuous variable (Rubin, 2006). However, this is an oversimplification and may lead to misleading results as count data have a different probabilistic distribution from continuous data. Some Bayesian methods estimate the causal effect by assuming that the potential count outcome is a Poisson distribution (Gutman et al., 2018), or a Zero-Inflated Poisson Distribution (Choi et al., 2020). However, in practice, observational data are not always measured at the regularly spaced time points. Built upon the studies with count outcomes at an irregular time point (Meyer et al., 2013;

Cunningham et al., 2016) and the need for causal estimators for count outcomes, we develop the doubly robust estimator for the panel count model proposed in Chapter 2, which is a flexible semi-parametric model for count data via a non-homogeneous Poisson process.

4.2 Inferring causal effects

4.2.1 Rubin causal model

We adopt the Rubin causal model (Rubin, 1974; Holland, 1986) and the counterfactual framework. We denote the observed data by $D_n = \{(y_i, \boldsymbol{x}_i, z_i, t_i) : i = 1, ..., n\}$, where $y_i \in R$ is the number of AEs observed in the time interval $(0, t_i)$ for the i-th subject, z_i is the subject's observed treatment indicator, $z_i = 1($ or $0)$ if the individual receives treatment (or control), $\boldsymbol{x}_i = (x_{i1}, ..., x_{id})' \in R^d$ is the vector of confounders of the i-th subject. For each subject i, there are two potential outcomes $(y_i(0), y_i(1))$, under treatment and control, but at most one of which is observed.

Then, potential outcomes and assignments jointly determine the values of the observed outcomes y_i.

$$y_i = y_i(1)z_i + y_i(0)(1 - z_i).$$

Let $\boldsymbol{y} = (y_1...y_n)$, then the goal is to estimate the average treatment effect

$$\Delta := E[\boldsymbol{y}(1)) - E(\boldsymbol{y}(0)]. \tag{4.1}$$

Under the assumption of strong ignorability of treatment (Rosenbaum and Rubin, 1983), the goal is to estimate the average treatment effect

$$\mu_1 := E(\boldsymbol{y}|z = 1, \boldsymbol{x}) \tag{4.2a}$$

$$\mu_0 := E(\boldsymbol{y}|z = 0, \boldsymbol{x}) \tag{4.2b}$$

$$\Delta := E(\boldsymbol{y}|z = 1, \boldsymbol{x}) - E(\boldsymbol{y}|z = 0, \boldsymbol{x}). \tag{4.2c}$$

4.2.2 Existing statistical methods for causal inference

Adjustment by regression models

The basic strategy behind regression analysis is to remove the correlation between the treatment variable $\boldsymbol{z}$ and the error term by including a set of covariates $\boldsymbol{x}$ related to both the outcome and the treatment exposure in the regression model (Winship and Morgan, 1999). If the outcome $\boldsymbol{y}$ is a continuous variable, then the linear regression of $\boldsymbol{y}$ on $\boldsymbol{x}$ and $\boldsymbol{z}$ can be identified as

$$E(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{\beta}_r^\intercal \boldsymbol{x} + \boldsymbol{z}\alpha_r. \tag{4.3}$$

Under the same conditions mentioned in §4.3.1,

$$\Delta_{lr} = E(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{z} = 1) - E(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{z} = 0) = \alpha_r.$$

Ordinary least squares (OLS) can be used to estimate (4.3). Let $\hat{\alpha}_r$ be the estimator of α_r

$$\hat{\Delta}_{lr} = \hat{\alpha}_r.$$

When $\boldsymbol{y}$ is not continuous, we can extend (4.3) to more general cases. For example, if $\boldsymbol{y}$ is a binary variable, we can use a logistic regression model, then

$$\Delta_{br} = \frac{\exp(\boldsymbol{\beta}_{lr}^\intercal \boldsymbol{x} + \boldsymbol{z}\alpha_{lr})}{1 + \exp(\boldsymbol{\beta}_{lr}^\intercal \boldsymbol{x} + \boldsymbol{z}\alpha_{lr})} - \frac{\exp(\boldsymbol{\beta}_{lr}^\intercal \boldsymbol{x})}{1 + \exp(\boldsymbol{\beta}_{lr}^\intercal \boldsymbol{x})}.$$

One problem with the adjustment by regression models is that $\hat{\Delta}$ is biased if the regression model is not correctly specified.

The propensity score

The propensity score is the conditional probability of a subject being assigned to a particular treatment given a vector of observed covariates, which is sufficient to

remove bias due to the observed covariates (Rosenbaum and Rubin, 1983). So we can use the propensity score to remove the imbalance between the treatment and control groups resulting from all the observed confounders (Rosenbaum and Rubin, 1983) including constructing matched samples from treatment groups with the propensity score, and creating weights with the inverse of the propensity score (Hernán and Robins, 2020).

Inverse weighting estimator. Let

$$\pi(\boldsymbol{x}_i) = P(z_i = 1|\boldsymbol{x}_i, \gamma) \tag{4.4}$$

be the propensity score. Then the treated individual receives the weight

$$\frac{1}{\pi(z_i = 1|\boldsymbol{x}_i, \boldsymbol{\gamma})},$$

and the untreated individual receives the weight

$$\frac{1}{1 - \pi(z_i = 1|\boldsymbol{x}_i, \boldsymbol{\gamma})}.$$

Let $\hat{\boldsymbol{\gamma}}$ be the estimator of $\boldsymbol{\gamma}$, then with the Horvitz-Thompson estimator (Horvitz and Thompson, 1952), the causal effect estimator is

$$\hat{\Delta}_{iw} = n^{-1} \sum_i^n \frac{z_i y_i}{\pi(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})} - n^{-1} \sum_i^n \frac{(1 - z_i) y_i}{1 - \pi(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})}. \tag{4.5}$$

The estimator is biased when the postulated propensity score model is not correctly specified.

Doubly robust estimator

Robins et al. (1994) developed a more robust estimator named doubly robust estimator. The doubly robust estimator is constructed by specifying the treatment assignment mechanism and the regression model. We specify the propensity score of treatment assignment as (4.4),

$$\pi(\boldsymbol{x}_i) = P(z_i = 1|\boldsymbol{x}_i, \boldsymbol{\gamma}),$$

and outcome regression model as (4.3)

$$E(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{\beta}_r^{\intercal}\boldsymbol{x} + \boldsymbol{z}\alpha_r.$$

More specifically, the outcome models for treatment group and control group are

$$m_1 = E(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{z} = 1) = \boldsymbol{\beta}_r^{\intercal}\boldsymbol{x} + \alpha_r$$
$$m_0 = E(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{z} = 0) = \boldsymbol{\beta}_r^{\intercal}\boldsymbol{x}.$$

Let $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\beta}}_r$ and $\hat{\alpha}_r$ be the consistent estimators of $\boldsymbol{\gamma}$, $\boldsymbol{\beta}_r$ and α_r, then the doubly robust estimator of the causal effect is

$$\hat{\Delta}_{dr} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{z_i y_i}{\pi(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})} - \frac{z_i - \pi(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})}{\pi(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})}m_1\right]$$
$$-\frac{1}{n}\sum_{i=1}^{n}\left[\frac{(1 - z_i)y_i}{1 - \pi(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})} + \frac{z_i - \pi(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})}{1 - \pi(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})}m_0\right].$$

The estimator remains unbiased when the propensity score or outcome model is correctly specified.

4.3 A panel count model for count outcomes

To infer the causal effect of the medicine/vaccine on the AEs from data with a count outcome in the time interval $(0, t)$, recall the model proposed for the panel count

outcome in Chapter 2, then we use the following non-homogeneous Poisson process $y_i = y_i(t_i)$ for counts in the time interval $(0, t_i]$ with conditional mean

$$E[y_i(t_i)|\boldsymbol{x}_i, z_i] = G(t_i)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_i + \boldsymbol{z}_i\alpha).$$

Recall $\boldsymbol{x}_i = (x_{i1}, ..., x_{id})' \in R^d$ and $y_i(t_i)$ is the event count in $(0, t_i]$ for individual i, $G(\cdot) \in \mathcal{G} = \{$monotone increasing functions on $[0, 1]$ with $0 \leq G(t_i) \leq 1.\}$. The constraint $0 \leq G(t_i) \leq 1$ is for model identifiability. In this model, $G(t_i)$ is the non-parametric component that adjusts for the irregularly spaced time for different individuals in the observed data. Denote $d = \dim(\boldsymbol{\beta})$. α is the treatment effect. Here it is assumed that the larger value of α represents the better treatment effect.

4.3.1 Estimation procedure of panel count model

The likelihood-based on the observed data D_n is

$$L(\boldsymbol{\beta}, \alpha, G) = \left\{\prod_{i=1}^{n} \frac{[G(t_i)\exp\left(\boldsymbol{\beta}^\intercal \boldsymbol{x}_i + z_i\alpha\right)]^{y_i}}{y_i!}\right\} \exp\left[-\sum_{i=1}^{n} G(t_i)\exp\left(\boldsymbol{\beta}^\intercal \boldsymbol{x}_i + z_i\alpha\right)\right].$$

Then, the log-likelihood is, omitting some constant independent of our parameter of interest,

$$\ell(\boldsymbol{\beta}, \alpha, G|D_n) = \sum_{i=1}^{n} \left[y_i \log G(t_i) + y_i\boldsymbol{\beta}_0^\intercal \boldsymbol{x}_i + y_i z_i\alpha - G(t_i)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_i + z_i\alpha)\right].$$

Denote $\boldsymbol{\Theta}$ for the space for $(\boldsymbol{\beta}, \alpha)$ and $\mathcal{G}$ is the space for G, and we estimate $(\boldsymbol{\beta}, \alpha, G)$ by the semi-parametric maximum likelihood estimator $(\hat{\boldsymbol{\beta}}, \hat{\alpha}, \hat{G})$,

$$(\hat{\boldsymbol{\beta}}, \hat{\alpha}, \hat{G}) = \arg \max_{(\boldsymbol{\beta}, \alpha, G) \in (\boldsymbol{\Theta}, \mathcal{G})} \ell(\boldsymbol{\beta}, \alpha, G|D_n).$$

The estimation of $(\boldsymbol{\beta}, \alpha)$ is trivial, but $\hat{G}$ is not easy to compute. I will use a similar algorithm and technique derived in Chapter 2 to estimate $\hat{G}$.

The algorithm. We use an iterative algorithm to estimate parameters. For this, given starting value $(\boldsymbol{\beta}^{(0)}, \alpha^{(0)}, G^{(0)}(\cdot))$, we get the next step estimate $(\boldsymbol{\beta}^{(1)}, \alpha^{(1)}, G^{(1)}(\cdot))$.... Generally the updating is as follows.

Estimation of $G(\cdot)$. In particular, we can first fix $(\boldsymbol{\beta}^{(r)}, \alpha^{(r)})$, and maximize over $G(\cdot)$ to get $G^{(r+1)}(\cdot)$, which is of the form, for some $d_i^{(r)}$'s,

$$G^{(r+1)} = \arg\max_{G \in \mathcal{G}} \sum_{i=1}^{n} \left[y_{i\cdot} \log G(t_i) - d_i^{(r)} G(t_i) \right].$$

The above can be solved by the isotonic regression technique, as in Chapter 2. Without loss of generality, we assume the t_i's are arranged in increasing order, re-write the above as, with $h_i^{(r)} = y_{i\cdot}/d_i^{(r)}$,

$$\begin{aligned} G^{(r+1)} &= \arg\min_{G \in \mathcal{G}} \sum_{i=1}^{I} \left[-h_i^{(r)} \log G(t_i) + G(t_i) \right] d_i^{(r)}. \\ &= \arg\min_{G \in \mathcal{G}} \sum_{i=1}^{I} \left[h_i^{(r)} \log h_i^{(r)} - h_i^{(r)} \log G(t_i) - h_i^{(r)} + G(t_i) \right] d_i^{(r)} \\ &= \arg\min_{G \in \mathcal{G}} \sum_{i=1}^{I} \Delta_\Phi \left[h_i^{(r)}, G(t_i) \right] d_i^{(r)}, \end{aligned}$$

where $\Delta_\Phi(u, v) = \Phi(u) - \Phi(v) - (u - v)\phi(v) = u \log u - v \log v - (u - v)(\log v + 1) = u \log u - u \log v - u + v$, and $\Phi(u) = u \log u, u \in R^+$. The first derivative of Φ is given by $\phi(u) = \log u + 1$. $\Phi(\cdot)$ is convex on R^+, so by Theorem 1.5.1 and Example 1.5.1 in Robertson et al. (1988), δ_j^* of δ_j with weights $w = d_i$ provide the maximum likelihood estimates of $G_1, ...G_I$ subject to the constraint that the estimates be non-decreasing $G(\cdot)$. The above minimization is the same the following isotonic regression solution

$$G^{(r+1)} = \arg\min_{G \in \mathcal{G}} \sum_{i=1}^{I} d_i^{(r)} \left[h_i^{(r)} - G(t_i) \right]^2.$$

The R-package isotonic (Best and Chakravarti, 1990; de Leeuw et al., 2009) can compute the above minimizer.

Then, we can fix $G^{(r+1)}(\cdot)$, and maximize over $\boldsymbol{\beta}$ and α to get $\boldsymbol{\beta}^{(r+1)}$ and $\alpha^{(r+1)}$. The iteration goes on until convergence of the sequence, and the final values are the MLE $(\hat{\boldsymbol{\theta}}, \hat{G})$.

4.3.2 Estimating the average treatment effect

Recall (4.1) and (4.2). Under certain conditions, the aim is to estimate

$$\Delta = E(\boldsymbol{y}|z = 1, \boldsymbol{x}) - E(\boldsymbol{y}|z = 0, \boldsymbol{x}).$$

Adjustment in the panel count model

The most straightforward way to estimate Δ is adjusting for counfounders $\boldsymbol{x}_i$ by the outcome model. Let $\hat{\boldsymbol{\beta}}$ and $\hat{\alpha}$ be the consistent estimators of $\boldsymbol{\beta}$ and α, then

$$\hat{\mu}_{1out} = n^{-1} \sum_{i=1}^{n} \left[\hat{G}(t_i) \exp\left(\hat{\boldsymbol{\beta}}^{\intercal} x_i + \hat{\alpha} \right) \right]$$

$$\hat{\mu}_{0out} = n^{-1} \sum_{i=1}^{n} \left[\hat{G}(t_i) \exp\left(\hat{\boldsymbol{\beta}}^{\intercal} x_i \right) \right].$$

Then,

$$\hat{\Delta}_{out} = n^{-1} \sum_{i=1}^{n} \left[\hat{G}(t_i) \exp\left(\hat{\boldsymbol{\beta}}^{\intercal} x_i + \hat{\alpha} \right) - \hat{G}(t_i) \exp\left(\hat{\boldsymbol{\beta}}^{\intercal} x_i \right) \right].$$

$\hat{\Delta}_{out}$ is an unbiased estimator if the panel count model is correctly specified.

Inverse weighting estimator

Recall (4.5), then

$$\hat{\Delta}_{iw} = n^{-1} \sum_{i}^{n} \frac{z_i y_i}{\pi(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})} - n^{-1} \sum_{i}^{n} \frac{(1 - z_i) y_i}{1 - \pi(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})}.$$

Doubly robust estimator

With the proposed panel count model, now we develop the doubly robust estimator. We specify the propensity score model and the outcome model as:

$$\pi(\boldsymbol{x}_i, \boldsymbol{\gamma}) = P(z_i = 1|\boldsymbol{x}_i, \boldsymbol{\gamma}), \quad E(y_i(t_i)|\boldsymbol{x}_i, z_i = j) = m_j(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\beta}, \alpha) \quad (j = 0, 1).$$

More specifically, $\pi(\boldsymbol{x}_i, t_i) = P(z_i = 1|\boldsymbol{x}_i, t_i)$ is specified as the logistic model

$$\pi(\boldsymbol{x}) = \frac{\exp(\boldsymbol{x}^\intercal \boldsymbol{\gamma})}{1 + \exp(\boldsymbol{x}^\intercal \boldsymbol{\gamma})}.$$

The outcome model $m_j(\boldsymbol{x}, \boldsymbol{\beta})$ is specified as follows,

$$m_1(\boldsymbol{x}, \boldsymbol{\beta}, \alpha) = G(t_i) \exp\left(\boldsymbol{\beta}^\intercal \boldsymbol{x}_i + \alpha\right)$$

$$m_0(\boldsymbol{x}, \boldsymbol{\beta}) = G(t_i) \exp\left(\boldsymbol{\beta}^\intercal \boldsymbol{x}_i\right).$$

Recall $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\beta}}$ and $\hat{\alpha}$ are the consistent estimators of $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$ and α, and

$$\hat{\mu}_{1dr} = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{z_i y_i}{\hat{\pi}(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})} - \frac{z_i - \hat{\pi}(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})}{\hat{\pi}(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})} m_1(\boldsymbol{x}, \hat{\boldsymbol{\beta}}, \hat{\alpha}) \right].$$

It is easy to see that if either $\hat{\pi}(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}}) = \pi(\boldsymbol{x}_i)$ (propensity model correctly specified), or $m_1(\boldsymbol{x}, \hat{\boldsymbol{\beta}}, \hat{\alpha}) = m_1(\boldsymbol{x})$ (outcome model correctly specified), then $\hat{\mu}_{1dr} \to \mu_{1dr}$.

Similarly, the doubly robust estimate of μ_0 is

$$\hat{\mu}_{0dr} = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{(1 - z_i) y_i}{1 - \hat{\pi}(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})} + \frac{z_i - \hat{\pi}(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})}{1 - \hat{\pi}(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})} m_0(\boldsymbol{x}, \hat{\boldsymbol{\beta}}) \right].$$

Then, the estimate of causal effect is

$$\begin{aligned} \hat{\Delta}_{dr} = & \frac{1}{n} \sum_{i=1}^{n} \left[\frac{z_i y_i}{\hat{\pi}(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})} - \frac{z_i - \hat{\pi}(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})}{\hat{\pi}(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})} \hat{G}(t_i) \exp\left(\hat{\boldsymbol{\beta}}^\intercal \boldsymbol{x}_i + \hat{\alpha}\right) \right] \\ & - \frac{1}{n} \sum_{i=1}^{n} \left[\frac{(1 - z_i) y_i}{1 - \hat{\pi}(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})} + \frac{z_i - \hat{\pi}(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})}{1 - \hat{\pi}(\boldsymbol{x}_i, \hat{\boldsymbol{\gamma}})} \hat{G}(t_i) \exp\left(\hat{\boldsymbol{\beta}}^\intercal \boldsymbol{x}_i\right) \right]. \end{aligned}$$

4.4 Asymptotic properties

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \alpha)$ and $(\boldsymbol{\theta}_0, G_0)$ be the 'true' parameters generating the data, $\ell(\boldsymbol{\theta}, G|z, \boldsymbol{y}, \boldsymbol{x}, t) = \log f(\boldsymbol{y}|\boldsymbol{x}, z, t, ; \boldsymbol{\beta}, \alpha, G)$, $\dot{l}_\theta(\boldsymbol{\theta}, G) = \partial\ell(\boldsymbol{\theta}, G|z, \boldsymbol{y}, \boldsymbol{x}, t)/\partial\boldsymbol{\theta}$, and $\dot{l}_G(\boldsymbol{\theta}, G)[h]$ be the Hadamard derivative of $\ell(\boldsymbol{\theta}, G|z, \boldsymbol{y}, \boldsymbol{x})$ with respect to G in the direction h. We need the following regularity conditions:

(C4.1). $\boldsymbol{\theta}$ belongs to some bounded set.

(C4.2). The support of $\boldsymbol{X}$ is bounded, and the support of T is $[0, B]$ for some $0 < B < \infty$.

(C4.3). $G_0 \in \mathcal{G}$ and $G_0(\cdot)$ is continuous.

(C4.4). $\sup_{G\in\mathcal{G}} \sup_{t\in[0,B]} G(t) < \infty$.

(C4.5). For all $G \in \mathcal{G}$, $G(\cdot)$ has derivative $\dot{G}(\cdot)$ which is uniformly bounded over $\mathcal{G}$.

(C4.6). $\sup_{(\theta,G)\in(\Theta,\mathcal{G}):d(\theta-\theta_0,G-G_0)<\delta}(||\dot{l}_\theta(\boldsymbol{\theta}, G)||_{L_2(P)} + ||\dot{l}_G(\boldsymbol{\theta}, G)[1]||_{L_2(P)}) \leq C$.

We only give outlines in this Chapter, and the details of the theorems and proofs are similar to those given in Chapter 3. Theorem 4.1 below gives strong consistency of the estimators $(\hat{\boldsymbol{\theta}}, \hat{G})$.

Theorem 4.1. *Assume (C4.1)-(C4.6), then as $n \to \infty$,*

$$\hat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}_0,$$

$$\textit{and} \qquad \sup_{z\in R^+} |\hat{G}(t) - G_0(t)| \xrightarrow{a.s.} 0$$

(C4.7). I^*_θ given in Theorem 4.2 is invertible.

Proof of Theorem 4.1. See proof of Theorem 3.1.

Denote $\xrightarrow{D}$ for convergence in distribution. Theorem 4.2 below gives asymptotic normality and efficiency of the Euclidean component $\hat{\boldsymbol{\theta}}$.

Theorem 4.2. *Let $\tilde{\boldsymbol{x}} = (\boldsymbol{x}, z)$, and assume (C4.1)-(C4.7), then as $n \to \infty$,*

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) \xrightarrow{D} N(\mathbf{0}, (I_\theta^*)^{-1})$$

$$I_\theta^* = E_{(\theta_0, G_0)}[\ell_\theta^*(\ell_\theta^*)^\intercal],$$

where, under the observed data $D_n = \{(\boldsymbol{y}_i, \boldsymbol{x}_i, z_i, t_i) : i = 1, ...n\}$, the efficient score $\ell_{\boldsymbol{\theta}}^*$ for $\boldsymbol{\theta}$ is given by

$$\ell_{\boldsymbol{\theta}}^* = \ell_{\boldsymbol{\theta}}^*(\boldsymbol{\theta}_0, G_0 | D_n) = \sum_{i=1}^n \dot{\ell}_i \left\{ \tilde{\boldsymbol{x}}_i - \frac{E\left[\left(\sum_{i=1}^n \dot{\ell}_i \tilde{\boldsymbol{x}}_i\right)\left(\sum_{i=1}^n \dot{\ell}_i\right)|t\right]}{E\left[\left(\sum_{i=1}^n \dot{\ell}\right)^2 |t\right]} \right\},$$

where

$$\dot{\ell}_i = y_i - G(t_i)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_i + z_i\alpha).$$

Proof of Theorem 4.2. *Lemma 4.1 for proof of Theorem 4.2. Let $d(\boldsymbol{\theta}, G; \boldsymbol{\theta}_0, G) = ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| + \sup_{t\in[0,B]} |G(t) - G_0(t)|$. Under conditions of Theorem 4.1,*

$$d(\boldsymbol{\theta}, G; \boldsymbol{\theta}_0, G_0) = O_p(I^{-1/3}).$$

Proof of Lemma 4.1. See proof of Lemma 3.1.

Lemma 4.2 for proof of Theorem 4.2. Under the observed data $D_n = \{(\boldsymbol{y}_i, \boldsymbol{x}_i, z_i, t_i) : i = 1, ...n\}$, the efficient score for estimating $\boldsymbol{\theta}_0$ is

$$\ell_{\boldsymbol{\theta}}^*(\boldsymbol{\theta}, G | D_n) = \sum_{i=1}^n \dot{\ell}_i \left\{ \tilde{\boldsymbol{x}}_i - \frac{E\left[(\sum_{i=1}^n \dot{\ell}_i \tilde{\boldsymbol{x}}_i)(\sum_{i=1}^n \dot{\ell}_i)|T = t\right]}{E\left[\left(\sum_{i=1}^n \dot{\ell}_i\right)^2 |T = t\right]} \right\},$$

and

$$\dot{\ell}_i = y_i - G(t_i)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_i + z_i\alpha).$$

Proof of Lemma 4.2. The log-likelihood under data $D_n = \{(\boldsymbol{y}_i, \boldsymbol{x}_i, z_i, t_i) : i = 1, ...n\}$ is

$$\ell(\boldsymbol{\theta}, G | D_n) = \sum_{i=1}^n \left[y_i \log G(t_i) + y_i \boldsymbol{\beta}_0^\intercal \boldsymbol{x}_i + y_i z_i \alpha - G(t_i)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_i + z_i\alpha) \right].$$

Denote $\dot{\ell}_{\boldsymbol{\theta}}(\boldsymbol{\theta}, G|D_n) = \partial\ell(\boldsymbol{\theta}, G|D_n)/\partial\boldsymbol{\theta}$, and $\dot{\ell}_G(\boldsymbol{\theta}, G|D_n)[h]$ be the Hadamard derivative of $\ell(\boldsymbol{\theta}, G|D_n)$ with respect to G in the direction h. Then

$$\begin{aligned}\dot{\ell}_\theta(\boldsymbol{\theta}, G|D_n) &= \sum_{i=1}^{n} \left[y_i - G(t)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_i + z_i\alpha)\right]\tilde{\boldsymbol{x}}_i \\ &= \sum_{i=1}^{n} \dot{\ell}_i(\boldsymbol{\theta}, G|D_n)\tilde{\boldsymbol{x}}_i,\end{aligned}$$

and it can be checked that

$$\begin{aligned}\dot{\ell}_G(\boldsymbol{\theta}, G|D)[h] &= \sum_{i=1}^{n} \left[\frac{h(t)}{G(t)}y_i - h(t)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_i + z_i\alpha)\right] \\ &= \sum_{i=1}^{n} \frac{h(t)}{G(t)}\dot{\ell}_i(\boldsymbol{\theta}, G|D).\end{aligned}$$

The efficient score for estimating $\boldsymbol{\theta}_0$ is

$$\ell^{*\intercal}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, G_0|D_n) = \dot{\ell}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, G_0|D_n) - \dot{\ell}_G(\boldsymbol{\theta}_0, G_0|D_n)[\boldsymbol{h}^*],$$

where $\boldsymbol{h}^* = (h_1^*, ..., h_d^*)^\intercal$, $d = \dim(\boldsymbol{\theta})$, $\dot{\ell}_G(\boldsymbol{\theta}_0, G_0|D)[\boldsymbol{h}^*] = (\dot{\ell}_G(\boldsymbol{\theta}_0, G_0|D_n)[h_1^*], ..., \dot{\ell}_G(\boldsymbol{\theta}_0, G_0|D_n)[h_d^*])^\intercal$, and the least favorable direction $\boldsymbol{h}^*$ is determined by

$$\begin{aligned}\forall\ h, \quad \mathbf{0} &= E\left\{\left[\dot{\ell}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, G_0|D_n) - \dot{\ell}_G(\boldsymbol{\theta}_0, G_0|D_n)[\boldsymbol{h}^*]\right]\dot{\ell}_G(\boldsymbol{\theta}_0, G_0|D_n)[h]\right\} \\ &= E\left\{E\left[\sum_{i=1}^{n}\dot{\ell}_i(\boldsymbol{\theta}_0, G_0|D_n)\left(\tilde{\boldsymbol{x}}_i - \frac{\boldsymbol{h}^*(t)}{G(t)}\right)\left(\sum_{i=1}^{n}\dot{\ell}_i(\boldsymbol{\theta}_0, G_0|D_n)\right)\right]\frac{h(t)}{G(t)}|T = t\right\}.\end{aligned}$$

The above gives

$$E\left\{\sum_{i=1}^{n}\dot{\ell}_i(\boldsymbol{\theta}_0, G_0|D_n)\left(\tilde{\boldsymbol{x}}_i - \frac{\boldsymbol{h}^*(t)}{G(t)}\right)\left[\sum_{i=1}^{n}\dot{\ell}_i(\boldsymbol{\theta}_0, G_0|D_n)\right]\right\} \equiv \mathbf{0},$$

or

$$\boldsymbol{h}^*(t) = G(t)\frac{E\left\{\left[\sum_{i=1}^{n}\dot{\ell}_i(\boldsymbol{\theta}_0, G_0|D_n)\tilde{\boldsymbol{x}}_i\right]\left[\sum_{i=1}^{n}\dot{\ell}_i(\boldsymbol{\theta}_0, G_0|D_n)\right]|T = t\right\}}{E\left\{\left[\sum_{i=1}^{n}\dot{\ell}_i(\boldsymbol{\theta}_0, G_0|D_n)\right]^2|T = t\right\}}.$$

Thus,

$$\ell^*_{\boldsymbol{\theta}}(\boldsymbol{\theta}, G|D_n) = \sum_{i=1}^{n}\dot{\ell}_i\left\{\boldsymbol{x}_i - \frac{E\left[(\sum_{i=1}^{n}\dot{\ell}_i\tilde{\boldsymbol{x}}_i)(\sum_{i=1}^{n}\dot{\ell}_i)|T = t\right]}{E\left[\left(\sum_{i=1}^{n}\dot{\ell}_i\right)^2|T = t\right]}\right\}.$$

With Lemma 4.1 and 4.2, see proof of Theorem 3.2.

Let $\mathbb{B}(\cdot)$ be the two-sided Brownian motion originating from zero: A mean zero Gaussian process on R with $\mathbb{B}(0) = 0$, and $E\big(\mathbb{B}(s) - \mathbb{B}(h)\big)^2 = |s - h|$ for all $s, h \in R$. Denote $\xrightarrow{D}$ for convergence in distribution. Theorem 4.3 gives the asymptotic distribution of $\hat{G}(t)$ for each point t.

(C4.8). Let $f(\cdot)$ be the density function of z, $\dot{G}_0(t) = dG_0(t)/dt$. Assume $f(t) > 0$ and

$\dot{G}_0(t) > 0$ at the point t in Theorem 4.3 below.

(C4.9). $D(\cdot)$ and $\eta(\cdot)$ given in Theorem 4.3 below are continuous at z.

Theorem 4.3. *Assume (C4.1)-(C4.9), then as* $n \to \infty$,

$$n^{1/3}\left(\hat{G}(t) - G_0(t)\right) \xrightarrow{D} \left[\frac{4\eta^2(t)D^2(t)\dot{G}_0(t)}{d^2(t)f(t)}\right]^{1/3} \arg\min_h \left[\mathbb{B}(h) + h^2\right],$$

where

$$\eta^2(t) = E(\epsilon_i^2|T = t)$$

$$\epsilon_i = y_{i\cdot}/d_i - G_0(t)$$

$$d_i = \exp(\boldsymbol{\beta}_0^\intercal \boldsymbol{x}_i + z_i\alpha)$$

$$d(t) = E(d_i|T = t)$$

$$D^2(t) = E(d_i^2|T = t).$$

Proof of Theorem 4.3. See proof of Theorem 3.3.

In Theorem 4.3, it is assumed that $\dot{G}_0(t) > 0$, and it is interesting to know what happens when $\dot{G}_0(t) = 0$, which states that if $G_0(\cdot)$ has k-th derivative $G_0^{(k)}(t) \neq 0$ and $G_0^{(m)}(t) = 0$ for $m = 1, ...k - 1$. Then, $\hat{G}(t)$ has convergence rate $n^{k/(2k+1)}$, which is faster than the rate $n^{1/3}$ in Theorem 4.3, as given in the following Theorem 4.4.

Theorem 4.4. *Assume conditions of Theorem 4.3, with the condition* $\dot{G}_0(t) > 0$ *replaced by* $G_0^{(m)}(t) = 0$ *for* $m = 1, ..., k-1$, *and* $G_0^{(k)}(t) \neq 0$. *Then as* $n \to \infty$,

$$n^{k/(2k+1)} \left(\hat{G}(t) - G_0(t) \right) \xrightarrow{D} W,$$

where the distribution function of W *is given by,* $\forall x \in R$,

$$P(W \leq x) = P \left\{ \arg\min_h \left[\eta(t) D(t) f(z)^{1/2} \mathbb{B}(h) - x d(t) f(t) h + d(t) f(t) G_0^{(k)}(t) h^{k+1}/(k+1)! \right] \geq 0 \right\},$$

where $\eta(t), d(t)$ *and* $D(t)$ *are given in Theorem 4.3.*

Proof of Theorem 4.4. See proof of Theorem 3.4.

4.5 Simulation studies: Finite sample properties

We conduct simulation studies to evaluate the finite sample behavior of the doubly robust estimator. The results are summarized in terms of bias, standard deviation(SD), and 95% confidence interval (95% CI). The sample size is n=1000, 2000, and 3000. We consider different true $\Delta = 0.56, 1.28, 2.41$ to represent small causal effect, medium causal effect, and large causal effect. We also consider false models for propensity score and outcome models to investigate the robustness to misspecifications. Two patterns of misspecifications are considered: 1) models are wrong; 2) some confounders are missing. Below are lists of the details of the true and wrong models. For each scenario, the simulation size is N=1000.

"True" propensity score model giving rise to the observed data: The model is with a logit link function.

$$\pi(z_i; \boldsymbol{x}_i, t_i) = \frac{\exp(\gamma^\intercal \boldsymbol{x}_i)}{1 + \exp(\gamma^\intercal \boldsymbol{x}_i)}.$$

"True" outcome model giving rise to the observed data: The model is the proposed panel count model.

$$E(y_i(t_i) | \boldsymbol{x}_i, z_i) = G(t_i) \exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}_i + \boldsymbol{z}_i \alpha).$$

First pattern of wrong models: Models are misspecified.

Misspecified propensity score model: The model is with a complementary log-log (cloglog) link function instead of a logit link function.

$$\pi(z_i; \boldsymbol{x}_i, t_i)_{wrong1} = 1 - \exp[-\exp(\gamma^\intercal \boldsymbol{x}_i)].$$

Misspecified outcome model: It is a simple linear regression model instead of the panel count model.

$$E(y_i(t_i)|\boldsymbol{x}_i, z_i)_{wrong1} = \beta^\intercal \boldsymbol{x}_i + z_i\alpha.$$

Second pattern of wrong models: Some covariates are missing.

Let

$$\boldsymbol{x}' = (\boldsymbol{x}_1, \boldsymbol{x}_2).$$

Misspecified propensity score model:

$$\pi(z_i; \boldsymbol{x}'_i, t_i)_{wrong2} = \frac{\exp(\gamma^\intercal \boldsymbol{x}'_i)}{1 + \exp(\gamma^\intercal \boldsymbol{x}'_i)}.$$

Misspecified outcome model:

$$E(y_i(t_i)|\boldsymbol{x}'_i, z_i) = G(t_i)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{x}'_i + \boldsymbol{z}_i\alpha).$$

Table 4.1 shows the causal effects from inverse weighting estimator, outcome models and doubly robust estimator under correct or misspecified models.

Table 4.1: Causal effects from inverse weighting estimator, outcome models and doubly robust estimator under correct or misspecified models.

True Δ	Estimator	Estimates	SD	95% CI
		$n = 1000$		
0.56	$\hat{\Delta}_{out}$ [i]	0.589	0.134	(0.396,0.845)
	$\hat{\Delta}_{iw}$ [ii]	0.561	0.141	(0.268,0.841)
	$\hat{\Delta}_{dr}$ [iii]	0.565	0.098	(0.384,0.748)
	$\hat{\Delta}_{out.f1}$ [iv]	0.567	0.140	(0.285,0.841)
	$\hat{\Delta}_{dr.of1}$ [v]	0.560	0.138	(0.286,0.839)
	$\hat{\Delta}_{iw.f1}$ [vi]	0.604	0.143	(0.314,0.895)
	$\hat{\Delta}_{dr.pif1}$ [vii]	0.559	0.098	(0.360,0.745)
	$\hat{\Delta}_{dr.bothf1}$ [viii]	0.531	0.136	(0.271,0.800)
	$\hat{\Delta}_{out.f2}$ [ix]	0.675	0.209	(0.368,1.162)
	$\hat{\Delta}_{dr.of2}$ [x]	0.569	0.107	(0.365,0.799)
	$\hat{\Delta}_{iw.f2}$ [xi]	0.544	0.133	(0.276,0.816)

[i] $\hat{\Delta}_{out}$ is the estimator with adjustment by correct outcome model.

[ii] $\hat{\Delta}_{iw}$ is the inverse weighting estimator with adjustment by the correct propensity score model.

[iii] $\hat{\Delta}_{dr}$ is the doubly robust estimator with the correct propensity score and outcome models.

[iv] $\hat{\Delta}_{out.f1}$ is the estimator with adjustment by the misspecified simple linear regression model instead of the true panel count model (the first pattern of wrong model)

[v] $\hat{\Delta}_{dr.of1}$ is the doubly robust estimator with adjustment by the correct propensity score model and the misspecified simple linear regression model instead of the true panel count model (the first pattern of wrong model)

[vi] $\hat{\Delta}_{iw.f1}$ is the inverse weighting estimator with adjustment by the misspecified propensity score model with a cloglog link function instead of a true logit link function (the first pattern of wrong models).

[vii] $\hat{\Delta}_{dr.pif1}$ is the inverse weighting estimator with adjustment by the misspecified propensity score model with a cloglog link function instead of a true logit link function (the first pattern of wrong models) and the correct outcome model.

[viii] $\hat{\Delta}_{dr.bothf1}$ is the doubly robust estimator with adjustment by the misspecified propensity score and outcome models (the first pattern of wrong models).

[ix] $\hat{\Delta}_{out.f2}$ is the estimator with adjustment by the misspecified outcome model when some covariates of the outcome model are missing (the second pattern of wrong model).

[x] $\hat{\Delta}_{dr.of2}$ is the doubly robust estimator with adjustment by the correct propensity score model and the misspecified outcome model when some covariates of the outcome model are missing (the second pattern of wrong model)

[xi] $\hat{\Delta}_{iw.f2}$ is the inverse weighting estimator with adjustment by the misspecified propensity score model when some covariates of the propensity score model are missing (the second pattern of wrong model).

Table 4.1. (Cont.)

True Δ	Estimator	Estimates	SD	95% CI
	$\hat{\Delta}_{dr.pif2}$ [xii]	0.565	0.099	(0.387,0.751)
	$\hat{\Delta}_{dr.bothf2}$ [xiii]	0.510	0.113	(0.284,0.747)
1.28	$\hat{\Delta}_{out}$ [i]	1.297	0.136	(1.075,1.606)
	$\hat{\Delta}_{iw}$ [ii]	1.272	0.170	(0.951,1.618)
	$\hat{\Delta}_{dr}$ [iii]	1.276	0.112	(1.061,1.505)
	$\hat{\Delta}_{out.f1}$ [iv]	1.291	0.169	(0.996,1.630)
	$\hat{\Delta}_{dr.of1}$ [v]	1.272	0.167	(0.958,1.604)
	$\hat{\Delta}_{iw.f1}$ [vi]	1.319	0.174	(0.995,1.656)
	$\hat{\Delta}_{dr.pif1}$ [vii]	1.272	0.112	(1.062,1.493)
	$\hat{\Delta}_{dr.bothf1}$ [viii]	1.225	0.164	(0.917,1.550)
	$\hat{\Delta}_{out.f2}$ [ix]	1.393	0.263	(1.064,2.017)
	$\hat{\Delta}_{dr.of2}$ [x]	1.284	0.137	(1.053,1.556)
	$\hat{\Delta}_{iw.f2}$ [xi]	1.251	0.136	(1.075,1.606)
	$\hat{\Delta}_{dr.pif2}$ [xii]	1.276	0.112	(1.062,1.505)
	$\hat{\Delta}_{dr.bothf2}$ [xiii]	1.217	0.140	(0.970,1.486)
2.41	$\hat{\Delta}_{out}$ [i]	2.438	0.192	(2.148,2.827)
	$\hat{\Delta}_{iw}$ [ii]	2.406	0.197	(2.009,2.787)
	$\hat{\Delta}_{dr}$ [iii]	2.413	0.139	(2.138,2.663)
	$\hat{\Delta}_{out.f1}$ [iv]	2.443	0.197	(2.042,2.815)
	$\hat{\Delta}_{dr.of1}$ [v]	2.405	0.194	(2.010,2.771)

[xii] $\hat{\Delta}_{dr.pif2}$ is the inverse weighting estimator with adjustment the misspecified propensity score model when some covariates of the propensity score model are missing (the second pattern of wrong model) and the correct outcome model.

[xiii] $\hat{\Delta}_{dr.bothf2}$ is the doubly robust estimator with adjustment by the misspecified propensity score and outcome models when some covariates of both models are missing (the second pattern of wrong models).

Table 4.1. (Cont.)

True Δ	Estimator	Estimates	SD	95% CI
	$\hat{\Delta}_{iw.f1}$ [vi]	2.461	0.204	(2.061,2.859)
	$\hat{\Delta}_{dr.pif1}$ [vii]	2.408	0.135	(2.139,2.672)
	$\hat{\Delta}_{dr.bothf1}$ [viii]	2.329	0.191	(1.946,2.692)
	$\hat{\Delta}_{out.f2}$ [ix]	2.548	0.290	(2.155,3.369)
	$\hat{\Delta}_{dr.of2}$ [x]	2.419	0.146	(2.145,2.699)
	$\hat{\Delta}_{iw.f2}$ [xi]	2.379	0.186	(2.013,2.731)
	$\hat{\Delta}_{dr.pif2}$ [xii]	2.413	0.140	(2.139,2.662)
	$\hat{\Delta}_{dr.bothf2}$ [xiii]	2.341	0.152	(2.055,2.646)
		$n = 2000$		
0.56	$\hat{\Delta}_{out}$ [i]	0.583	0.094	(0.427,0.792)
	$\hat{\Delta}_{iw}$ [ii]	0.558	0.103	(0.360,0.749)
	$\hat{\Delta}_{dr}$ [iii]	0.563	0.072	(0.422,0.701)
	$\hat{\Delta}_{out.f1}$ [iv]	0.565	0.102	(0.369,0.757)
	$\hat{\Delta}_{dr.of1}$ [v]	0.559	0.101	(0.364,0.754)
	$\hat{\Delta}_{iw.f1}$ [vi]	0.601	0.105	(0.396,0.794)
	$\hat{\Delta}_{dr.pif1}$ [vii]	0.559	0.073	(0.409,0.697)
	$\hat{\Delta}_{dr.bothf1}$ [viii]	0.529	0.100	(0.338,0.721)
	$\hat{\Delta}_{out.f2}$ [ix]	0.677	0.212	(0.418,0.200)
	$\hat{\Delta}_{dr.of2}$ [x]	0.570	0.117	(0.419,0.725)
	$\hat{\Delta}_{iw.f2}$ [xi]	0.541	0.098	(0.355,0.730)
	$\hat{\Delta}_{dr.pif2}$ [xii]	0.564	0.072	(0.422,0.701)
	$\hat{\Delta}_{dr.bothf2}$ [xiii]	0.511	0.119	(0.349,0.677)
1.28	$\hat{\Delta}_{out}$ [i]	1.300	0.138	(1.125,1.561)

Table 4.1. (Cont.)

True Δ	Estimator	Estimates	SD	95% CI
	$\hat{\Delta}_{iw}$[ii]	1.275	0.120	(1.061,1.508)
	$\hat{\Delta}_{dr}$[iii]	1.277	0.088	(1.109,1.443)
	$\hat{\Delta}_{out.f1}$[iv]	1.296	0.120	(1.072,1.532)
	$\hat{\Delta}_{dr.of1}$[v]	1.276	0.119	(1.061,1.519)
	$\hat{\Delta}_{iw.f1}$[vi]	1.324	0.122	(1.104,1.573)
	$\hat{\Delta}_{dr.pif1}$[vii]	1.272	0.088	(1.098,1.435)
	$\hat{\Delta}_{dr.bothf1}$[viii]	1.228	0.116	(1.009,1.468)
	$\hat{\Delta}_{out.f2}$[ix]	1.392	0.244	(1.112,2.001)
	$\hat{\Delta}_{dr.of2}$[x]	1.280	0.094	(1.102,1.458)
	$\hat{\Delta}_{iw.f2}$[xi]	1.253	0.112	(1.043,1.476)
	$\hat{\Delta}_{dr.pif2}$[xii]	1.277	0.089	(1.111,1.443)
	$\hat{\Delta}_{dr.bothf2}$[xiii]	1.214	0.096	(1.025,1.408)
2.41	$\hat{\Delta}_{out}$[i]	2.441	0.147	(2.225,2.734)
	$\hat{\Delta}_{iw}$[ii]	2.412	0.146	(2.127,2.700)
	$\hat{\Delta}_{dr}$[iii]	2.416	0.099	(2.220,2.623)
	$\hat{\Delta}_{out.f1}$[iv]	2.452	0.146	(2.157,2.738)
	$\hat{\Delta}_{dr.of1}$[v]	2.414	0.145	(2.127,2.707)
	$\hat{\Delta}_{iw.f1}$[vi]	2.467	0.153	(2.174,2.766)
	$\hat{\Delta}_{dr.pif1}$[vii]	2.411	0.101	(2.211,2.616)
	$\hat{\Delta}_{dr.bothf1}$[viii]	2.336	0.144	(2.055,2.621)
	$\hat{\Delta}_{out.f2}$[ix]	2.560	0.279	(2.198,3.332)
	$\hat{\Delta}_{dr.of2}$[x]	2.422	0.109	(2.207,2.642)
	$\hat{\Delta}_{iw.f2}$[xi]	2.384	0.138	(2.115,2.647)

Table 4.1. (Cont.)

True Δ	Estimator	Estimates	SD	95% CI
	$\hat{\Delta}_{dr.pif2}$ [xii]	2.417	0.099	(2.220,2.623)
	$\hat{\Delta}_{dr.bothf2}$ [xiii]	2.344	0.115	(2.117,2.587)
		$n = 3000$		
0.56	$\hat{\Delta}_{out}$ [i]	0.578	0.097	(0.449,0.843)
	$\hat{\Delta}_{iw}$ [ii]	0.554	0.084	(0.388,0.717)
	$\hat{\Delta}_{dr}$ [iii]	0.556	0.057	(0.444,0.666)
	$\hat{\Delta}_{out.f1}$ [iv]	0.561	0.083	(0.392,0.723)
	$\hat{\Delta}_{dr.of1}$ [v]	0.554	0.082	(0.385,0.712)
	$\hat{\Delta}_{iw.f1}$ [vi]	0.597	0.085	(0.431,0.765)
	$\hat{\Delta}_{dr.pif1}$ [vii]	0.551	0.059	(0.430,0.658)
	$\hat{\Delta}_{dr.bothf1}$ [viii]	0.524	0.081	(0.360,0.681)
	$\hat{\Delta}_{out.f2}$ [ix]	0.661	0.189	(0.433,1.164)
	$\hat{\Delta}_{dr.of2}$ [x]	0.558	0.063	(0.436,0.683)
	$\hat{\Delta}_{iw.f2}$ [xi]	0.536	0.078	(0.383,0.692)
	$\hat{\Delta}_{dr.pif2}$ [xii]	0.556	0.057	(0.445,0.666)
	$\hat{\Delta}_{dr.bothf2}$ [xiii]	0.497	0.065	(0.376,0.619)
1.28	$\hat{\Delta}_{out}$ [i]	1.300	0.102	(1.146,1.511)
	$\hat{\Delta}_{iw}$ [ii]	1.272	0.097	(1.084,1.477)
	$\hat{\Delta}_{dr}$ [iii]	1.279	0.065	(1.141,1.404)
	$\hat{\Delta}_{out.f1}$ [iv]	1.291	0.096	(1.098,1.476)
	$\hat{\Delta}_{dr.of1}$ [v]	1.272	0.095	(1.086,1.458)
	$\hat{\Delta}_{iw.f1}$ [vi]	1.318	0.099	(1.116,1.528)
	$\hat{\Delta}_{dr.pif1}$ [vii]	1.274	0.066	(1.136,1.402)

Table 4.1. (Cont.)

True Δ	Estimator	Estimates	SD	95% CI
	$\hat{\Delta}_{dr.bothf1}$[viii]	1.224	0.093	(1.037,1.406)
	$\hat{\Delta}_{out.f2}$[ix]	1.391	0.209	(1.134,1.997)
	$\hat{\Delta}_{dr.of2}$[x]	1.280	0.069	(1.139,1.429)
	$\hat{\Delta}_{iw.f2}$[xi]	1.249	0.092	(1.067,1.437)
	$\hat{\Delta}_{dr.pif2}$[xii]	1.279	0.066	(1.140,1.405)
	$\hat{\Delta}_{dr.bothf2}$[xiii]	1.212	0.075	(1.069,1.362)
2.41	$\hat{\Delta}_{out}$[i]	2.444	0.139	(2.264,2.790)
	$\hat{\Delta}_{iw}$[ii]	2.412	0.119	(2.180,2.642)
	$\hat{\Delta}_{dr}$[iii]	2.415	0.083	(2.257,2.583)
	$\hat{\Delta}_{out.f1}$[iv]	2.450	0.119	(2.226,2.671)
	$\hat{\Delta}_{dr.of1}$[v]	2.411	0.117	(2.185,2.630)
	$\hat{\Delta}_{iw.f1}$[vi]	2.466	0.124	(2.225,2.701)
	$\hat{\Delta}_{dr.pif1}$[vii]	2.408	0.085	(2.238,2.571)
	$\hat{\Delta}_{dr.bothf1}$[viii]	2.333	0.116	(2.111,2.554)
	$\hat{\Delta}_{out.f2}$[ix]	2.555	0.277	(2.253,3.371)
	$\hat{\Delta}_{dr.of2}$[x]	2.421	0.043	(2.328,2.498)
	$\hat{\Delta}_{iw.f2}$[xi]	2.382	0.111	(2.173,2.596)
	$\hat{\Delta}_{dr.pif2}$[xii]	2.415	0.084	(2.258,2.584)
	$\hat{\Delta}_{dr.bothf2}$[xiii]	2.340	0.093	(2.168,2.533)

In the case of $n = 1000$ and $\Delta = 0.56$, when the outcome and propensity score models are correctly specified, the inverse weighting estimator and doubly robust estimator have small biases (0.001 and 0.005). The estimator from the outcome model

has more significant biases (≤ 0.029). All of the 95% Confidence Intervals include the true Δ. When the propensity score model is correctly specified and the outcome model is a simple linear regression model instead of the true panel count model, the doubly robust estimator remains accurate (bias is smaller than 0.001). In contrast, the estimator from the outcome model has a more considerable bias of 0.007. When the propensity score model is correctly specified, and some covariates in the outcome model are missing, the bias of the doubly robust estimator is 0.009. However, the estimator from the outcome model has a large bias of 0.115. When the outcome model is correctly specified and the propensity score model is misspecified with a cloglog link function instead of a true logit link function, the doubly robust estimator remains accurate with a bias of 0.001. In contrast, the estimator from the inverse weighting estimator has a more considerable bias of 0.044. When the outcome model is correctly specified, and some covariates in the propensity score model are missing, the bias of the doubly robust estimator is 0.005. However, the estimator from the outcome model has a significant bias of 0.021. The same trend holds for other scenarios with different Δ.

In the case of $n = 1000$ and $\Delta = 0.56$, when the outcome and propensity score models are correctly specified, the inverse weighting estimator and doubly robust estimator have small biases (0.002 and 0.003). The estimator from the outcome model has a more significant bias (≤ 0.023). All of the 95% Confidence Intervals include the true Δ. When the propensity score model is correctly specified, and the outcome model is a simple linear regression model instead of the true panel count model, the doubly robust estimator remains accurate (bias is 0.001). In contrast, the estimator from the outcome model has a more considerable bias of 0.005. When the propensity score model is correctly specified, and some covariates in the outcome model are missing, the bias of the doubly robust estimator is 0.010. However, the estimator

from the outcome model has a large bias, which is 0.117. When the outcome model is correctly specified and the propensity score model is misspecified with a cloglog link function instead of a true logit link function, the doubly robust estimator remains accurate with a bias of 0.001. In contrast, the estimator from the inverse weighting estimator has a more considerable bias of 0.039. When the outcome model is correctly specified, and some covariates in the propensity score model are missing, the bias of the doubly robust estimator is 0.004. However, the estimator from the outcome model has a significant bias, which is 0.019. The same trend holds for other scenarios with different Δ.

In the case of $n = 3000$ and $\Delta = 0.56$, when the outcome and propensity score models are correctly specified, the inverse weighting estimator and doubly robust estimator have small biases (0.006 and 0.004). The estimator from the outcome model has a more significant bias (≤ 0.018). All of the 95% Confidence Intervals include the true Δ. When the propensity score model is correctly specified, and the outcome model is a simple linear regression model instead of the true panel count model, the doubly robust estimator and the outcome estimator remain accurate with biases of 0.006 and 0.001. When the propensity score model is correctly specified, and some covariates in the outcome model are missing, the bias of the doubly robust estimator is 0.002. However, the estimator from the outcome model has a large bias of 0.101. When the outcome model is correctly specified, and the propensity score model is misspecified with a cloglog link function instead of a true logit link function, the doubly robust estimator remains accurate with a bias of 0.009. In contrast, the estimator from the inverse weighting estimator has a more considerable bias, which is 0.037. When the outcome model is correctly specified, and some covariates in the propensity score model are missing, the bias of the doubly robust estimator is 0.004. However, the

estimator from the outcome model has a significant bias of 0.024. The same trend holds for other scenarios with different Δ.

In general, when both outcome and propensity score models are correctly specified, inverse weighting estimator and doubly robust estimator give more stable and accurate estimates. When either outcome or propensity score models is misspecified, the doubly robust estimator is more robust to the misspecifications.

4.6 Discussion

I have developed the doubly robust estimator in a semi-parametric panel count model to infer causal effects in observational data with a non-integer outcome. Simulation studies demonstrated the robustness of doubly robust estimators to model misspecifications. The model was motivated by an observational study. The method can be extended to infer causal effects of non-randomized AEs in clinical research. Even in a randomized trial, if an AE in the treatment group occurs more frequently than that in the placebo group, and some patients are overdosed, we may want to explore if overdosing instead of the medication itself caused the AE. In this case, subjects in this study are not randomly assigned to overdosed vs non-overdosed groups.

The proposed method for count variables is applicable to a variety of outcomes including rare AEs such as suicidal thoughts/attempts, or AEs that only occur in certain subgroups such as infants, who are not usually allowed to be enrolled in a clinical trial because of ethical issues. For the rare and severe type of AEs, FDA and CDC make further decision by reviewing and assessing reports submitted to the post-market surveillance system. For example, the FDA and CDC recommended that Johnson & Johnson's Janssen (J & J&J/Janssen) COVID-19 Vaccine resume in the United States after assessing the reports submitted to the Vaccine Adverse Event

Reporting System (VAERS). Sometimes, randomized experiments are required to be conducted for further evaluation. Occasionally, FDA reviews observational studies for further decision (Meyer et al., 2013; Cunningham et al., 2016).

The method also has some limitations. For example, when both the propensity score and outcome models are not correctly specified, the estimator may be biased. One potential solution is to consider using a more flexible non-parametric propensity model as did in Yuan et al. (2021) for continuous outcomes to enhance the robustness, which we plan to report in a future manuscript.

Chapter 5

Summary and future work

Drug safety evaluation is one of the critical topics in public health. Researchers have put a lot of effort into this area for decades to improve public health and solve drug safety issues. Post-market monitoring helps decrease the occurrence of AEs, including withdrawal of the product with potential association with severe AEs from the market. It is the subject of pharmacoepidemiology when dealing with a specific population. This area also attracts a lot of attention from biostatisticians/statisticians/data scientists, who have been developing different methods to detect AE signals in post-market surveillance systems. However, statistical methods for monitoring drug safety are still evolving because no single approach has been demonstrated absolute superiority (Gibbons and Amatya, 2016). The unique features of the surveillance data include noisy background and zero-inflated issues, which are caused by the numerous reports collected over the years. In addition, the systems have patients' demographic information, such as age, gender, and weights. If this information can be accounted for in the signal detection process, imbalances resulting from heterogeneous patients' characteristics from one drug to another can be removed, resulting in improved detection procedure.

The most current methods are the RR-based approaches that are not able to account for the unique features of surveillance data when detecting signals in post-market surveillance systems. Thus, in Chapter 2, we proposed a semi-parametric panel count model to search AE signals by accounting for background noises, zero-inflated problems, and covariates information. A non-homogeneous process has been utilized to model the cell counts of the $I \times J$ AE-drug combinations. The non-parametric component $G(\cdot)$ of the model adjusts the background occurrence of AE, and parametric components incorporate covariates information while accounting for the excessive zeros with a latent variable. In addition, the approach finds signals by comparing the AE-drug combinations of interest with the expected values, which removes the bias that RR-based approaches share (Rothman et al., 2004). Simulation studies demonstrate its high performance and robustness to different proportions of true zeros. We applied the proposed methods to FDA and WHO surveillance data. New safety signals were found in these databases with the proposed method, demonstrating the proposed method can provide unique insight into post-market surveillance systems and has the potential to improve the accuracy of the current monitoring systems. In Chapter 2, we searched signals by assuming the cell counts are Poisson distributions and comparing the cell counts with the upper quantile of Poisson distribution. Furthermore, the small effects in hypothesis testing have a remarkable resemblance to the test of rare alleles in genetics studies. So we can translate tests in genetics to detect signals called simultaneous common and rare variants analysis (SCARVA). Chapter 2 gives the related preliminary idea, and my future research will assess the performance of the procedure with rare events.

Although the proposed new methods have demonstrated their high performance in searching for signals in post-market surveillance systems, there are limitations to the proposed method in signal detection, as we discussed in §2.8. One is that a large

number of signals will increase the expected counts of the whole panel, thus leading to more false negatives. Also, the proposed approach does not account for potential dependence among AEs, although the two limitations also apply to most current methods. The correlation among AEs can be modeled through shared frailty in the proposed model potentially. These possibilities will be explored in my future research.

As we all know, consistency and asymptotic distributions of the estimators are fundamental to estimating parameters and establishing the hypothesis test statistics. Otherwise, the estimators make no sense. In Chapter 3, we derived the asymptotic properties of the panel count model, including asymptotic consistency and distributions. Theorem 3 gives the asymptotic distribution of the non-parametric component of $G(t)$ when the first derivative of G is non-zero for each point. When this assumption of non-varnishing condition is violated, we can use the more general Theorem 4. The convergence rate is going to be faster. This Chapter provides theoretical support to the broader application of the proposed model in other fields.

Also, we should be aware that the data from post-marketing drug safety monitoring systems are primarily for initial signal screening but not for causal analysis. Safety signals found in surveillance systems do not imply causality. Researchers need to conduct further research to infer the causal effect of medicines/vaccines on AEs if spontaneous reports suggest a potential association between AEs and medicines. Regulators need to balance all the evidence from case reports, randomized trials, meta-analyses of randomized trials, and observational studies to make further decision regarding the safety of the product. To infer causal effects of medicines/vaccines on AEs from observational studies, a doubly robust estimator has been shown to perform better than the traditional methods (Robins et al., 1994). Despite the progress with the continuous outcome, little attention has been given to panel count data which are common in various pharmaceutical and epidemiological studies. So, in Chapter 4,

we linked the panel count model and doubly robust estimator to infer the causal effect for the Poisson outcome. Simulation studies have demonstrated its robustness to the misspecification of the propensity score model or the outcome model. We will apply the doubly robust estimator to an extensive, substantial study to infer the causal effects of varenicline on neuropsychiatric events and report the results. The proposed doubly robust estimator is robust to misspecifications, but it also has limitations. When both the propensity score and outcome models are not correctly specified, the estimator may be biased. One potential solution is to consider using a more flexible non-parametric propensity model in Yuan et al. (2021).

This dissertation provides new insight into drug safety monitoring and research and addresses the issues related to public health. In addition, although the proposed semi-parametric model is motivated by issues related to drug safety, it has the potential for broader applications to other studies with counting dependent variables, such as the number of accidents, product defects, and insurance claims. Furthermore, the modified EM algorithm procedure with isotonic regression technique in each M step for parameter estimation provides a basis for estimation algorithm developments in other areas.

Appendix A

Disclaimer

Uppsala Monitoring Centre, National Centres(UMC) provided the data extracted from VigiBase, but the information and the study results do not represent the UMC or WHO's opinion. Also, the information comes from various sources, and the suspected adverse effects are drug-related and not the same in all cases.

Bibliography

Ahmed, I. and Poncet, A. (2016), *PhViD: an R package for PharmacoVigilance signal Detection*. R package version 1.0.8.

Bate, A., Lindquist, M., Edwards, I. et al. (1998), 'A bayesian neural network method for adverse drug reaction signal generation.', *European Journal of Clinical Pharmacology* **54**, 315–321.

Benjamini, Y. and Hochberg, Y. (1995), 'Controlling the false discovery rate: A practical and powerful approach to multiple testing.', *Journal of the Royal Statistical Society. Series B. Methodological* **57**(1), 289–300.

Best, M. J. and Chakravarti, N. (1990), 'Active set algorithms for isotonic regression; a unifying framework.', *Mathematical Programming* **47**, 425–39.

CFR (2021), 'Postmarketing Reporting of Adverse Drug Experiences.', `https://www.ecfr.gov/cgi-bin/text-idx?SID=77821c5f7ae204cd5251ed82639d876a&mc=true&node=se21.5.314_180&rgn=div8.` Accessed: 2021-05-06.

Choi, J., Chapkin, R. and Ni, Y. (2020), Bayesian causal structural learning with zero-inflated poisson bayesian networks, *in* H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan and H. Lin, eds, 'Advances in Neural Information Processing Systems', Vol. 33, Curran Associates, Inc., pp. 5887–5897.

Cunningham, F. E., Hur, K., Dong, D., Miller, D. R., Zhang, R., Wei, X., McCarren, M. et al. (2016), 'A comparison of neuropsychiatric adverse events during early treatment with varenicline or a nicotine patch', *Addiction (Abingdon, England)* **111**(7), 1283–92.

Davies, N. M. and Thomas, K. H. (2017), 'The Food and Drug Administration and varenicline: Should risk communication be improved?', *Addiction (Abingdon, England)* **112**(4), 555–58.

de Leeuw, J., Hornik, K. and Mair, P. (2009), 'Isotone optimization in R: Pool-adjacentviolators (PAVA) and active set methods', *Journal of Statistical Software* **32**(5), 1–24.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society. Series B. Methodological* **39**, 1–38.

DuMouchel, W. (1999), 'Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system.', *The American Statistician* **53**, 177–90.

Evans, S. J., Waller, P. C. and Davis, S. (2001), 'Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports.', *Pharmacoepidemiology and Drug Safety* **10**(6), 483–6.

Evins, A. E., Benowitz, N. L., West, R., Russ, C., McRae, T., Lawrence, D., Krishen, A., St Aubin, L., Maravic, M. C. and Anthenelli, R. M. (2019), 'Neuropsychiatric Safety and Efficacy of Varenicline, Bupropion, and Nicotine Patch in Smokers With Psychotic, Anxiety, and Mood Disorders in the EAGLES Trial', *Journal of Clinical Psychopharmacology* **39**(2), 108–16.

FDA (2016), 'Fda: What is a Serious Adverse Event', https://www.fda.gov/safety/reporting-serious-problems-fda/what-serious-adverse-event. Accessed: 2016-02-01.

FDA (2018), 'FDA: Questions and Answers on FDA's Adverse Event Reporting System (faers)', https://www.fda.gov/drugs/surveillance/questions-and-answers-fdas-adverse-event-reporting-system-faers. Accessed:2018-06-04.

FDA (2020), 'FDA's sentinel initiative', https://www.fda.gov/safety/fdas-sentinel-initiative/fdas-sentinel-initiative-background. Accessed:2020-03-09.

Finney, D. J. (1974), 'Systemic signalling of adverse reactions to drugs', *Methods of Information in Medicine* **13**(1), 1–10.

Gibbons, R. and Amatya, A. (2016), *Statistical Methods for Drug Safety*, Chapman and Hall/CRC, New York.

Gutman, R., Intrator, O. and Lancaster, T. (2018), 'A Bayesian procedure for estimating the causal effects of nursing home bed-hold policy', *Biostatistics (Oxford, England)* **19**(4), 444–60.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K. et al. (1996), 'A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. AIDS Clinical Trials Group Study 175 Study Team', *The New England Journal of Medicine* **335**(15), 1081–90.

Hernán, M. A., Alonso, A., Logan, R., Grodstein, F., Michels, K. B., Willett, W. C., Manson, J. E. and Robins, J. M. (2008), 'Observational studies analyzed like randomized experiments: An application to postmenopausal hormone therapy and coronary heart disease', *Epidemiology (Cambridge, Mass.)* **19**(6), 766–79.

Hernán, M. A. and Robins, J. M. (2016), 'Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available', *American Journal of Epidemiology* **183**(8), 758–64.

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

Herson, J. (2009), *Data and Safety Monitoring Committees in Clinical Trials*, CRC Press, Ukraine.

Holland, P. W. (1986), 'Statistics and Causal Inference', *Journal of the American Statistical Association* **81**(396), 945–60.

Holm, S. (1979), 'A simple sequentially rejective multiple test procedure.', *Scandinavian Journal of Statistics* **6**, 65–70.

Horvitz, D. G. and Thompson, D. J. (1952), 'A Generalization of Sampling Without Replacement From a Finite Universe', *Journal of the American Statistical Association* **47**(260), 663–85.

Huang, L., Guo, T., Zalkikar, J. N. and Tiwari, R. C. (2014), 'A Review of Statistical Methods for Safety Surveillance', *Therapeutic Innovation & Regulatory Science* **48**(1), 98–108.

Huang, L., Zalkikar, J. and Tiwari, R. (2011), 'A likelihood ratio test based method for signal detection with application to FDA's drug safety data.', *Journal of the American Statistical Association* **106**, 1230–41.

Huang, L., Zalkikar, J. and Tiwari, R. C. (2013), 'Likelihood ratio test-based method for signal detection in drug classes using FDA's AERS database.', *Journal of Biopharmaceutical Statistics* **23**(1), 178–200.

Huang, L., Zheng, D., Zalkikar, J. and Tiwari, R. (2017), 'Zero-inflated Poisson model based likelihood ratio test for drug safety signal detection.', *Statistical Methods in Medical Research* **26**(1), 471–88.

Lansner, A. and Holst, A. (1996), 'A higher order Bayesian neural network with spiking units.', *International Journal of Neural Systems* **7**(2), 115–28.

Meyer, T. E., Taylor, L. G., Xie, S., Graham, D. J., Mosholder, A. D., Williams, J. R., Moeny, D., Ouellet-Hellstrom, R. P. and Coster, T. S. (2013), 'Neuropsychiatric events in varenicline and nicotine replacement patch users in the Military Health System', *Addiction (Abingdon, England)* **108**(1), 203–10.

Miettinen, O. S. and Wang, J. D. (1981), 'An alternative to the proportionate mortality ratio.', *American Journal of Epidemiology* **114**(1), 144–8.

Morris, A. P. and Zeggini, E. (2010), 'An evaluation of statistical approaches to rare variant analysis in genetic association studies.', *Genetic Epidemiology* **34**(2), 188–93.

Norén, G. N., Bate, A., Orre, R. and Edwards, I. R. (2006), 'Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events.', *Statistics in Medicine* **25**(21), 3740–57.

Olfson, M., Marcus, S. C. and Shaffer, D. (2006), 'Antidepressant drug therapy and suicide in severely depressed children and adults: A case-control study', *Archives of General Psychiatry* **63**(8), 865–72.

Patwary, K. W. (1969), Report on statistical aspects of the pilot research project for international drug monitoring., *in* 'Confidential Report prepared for World Health Organization', Geneva.

Prague, J. K., Roberts, R. E., Comninos, A. N., Clarke, S., Jayasena, C. N., Nash, Z., Doyle, C. et al. (2017), 'Neurokinin 3 receptor antagonism as a novel treatment for menopausal hot flushes: A phase 2, randomised, double-blind, placebo-controlled trial', *Lancet* **389**(10081), 1809–20.

Rawlings, J. O., Pantula, S. G. and Dickey, D. A., eds (1998), *Applied Regression Analysis*, Springer Texts in Statistics.

Robertson, T., Wright, F. T. and Dykstra, R. (1988), 'Order Restricted Statistical Inference, John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore. Rothman, K.J., Lanes, S., Sacks, S.T. (2004) The reporting odds ratio and its advantages over the proportional reporting ratio.', *Pharmacoepidemiology and Drug Safety* **13**(8), 519–23.

Robins, J. M., Hernán, M. A. and Brumback, B. (2000), 'Marginal structural models and causal inference in epidemiology', *Epidemiology (Cambridge, Mass.)* **11**(5), 550–60.

Robins, J. M., Rotnitzky, A. and Lue, P. Z. (1994), 'Estimation of Regression Coefficients When Some Regressors Are Not Always Observed', *Journal of the American Statistical Association* **89**(427), 846–66.

Rosenbaum, P. R. and Rubin, D. B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* **70**(1), 41–55.
URL: *https://doi.org/10.1093/biomet/70.1.41*

Rothman, K. J., Lanes, S. and Sacks, S. T. (2004), 'The reporting odds ratio and its advantages over the proportional reporting ratio.', *Pharmacoepidemiology and Drug Safety* **13**(8), 519–23.

Rotnitzky, A., Lei, Q., Sued, M. and Robins, J. M. (2012), 'Improved double-robust estimation in missing data and causal inference models', *Biometrika* **99**(2), 439–56.

Rubin, D. B. (1974), 'Estimating causal effects of treatments in randomized and nonrandomized studies', *Journal of Educational Psychology* **66**(5), 688–701.

Rubin, D. B. (2006), 'Causal inference through potential outcomes and principal stratification: Application to studies with "censoring" due to death', *Statistical Science* **21**(3), 299–309.

Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. (1999), 'Adjusting for Non-ignorable Drop-Out Using Semiparametric Nonresponse Models', *Journal of the American Statistical Association* **94**(448), 1096–120.

Strom, B. L. (2006), 'How the US drug safety system should be changed.', *Journal of the American Medical Association* **295**(17), 2072–5.

Sun, J. and Zhao, X. (2013), *Statistical Analysis of Panel Count Data*, Springer.

van der Vaart, A. (2002), Semiparametric statistics, *in* 'Lecture Notes in Mathematics, Ed. Cachan, J.M.M., Groningen, F.T., Paris, B.T.', Springer.

van der Vaart, A. and Wellner, J. (1996), *Weak Convergence and Empirical Processes*, Springer.

Winship, C. and Morgan, S. L. (1999), 'The estimation of causal effects from observational data', *Annual Review of Sociology* **25**(1), 659–706.
URL: *https://doi.org/10.1146/annurev.soc.25.1.659*

Yuan, A., Chen, G., Zhou, Y., Bentley, A. and Rotimi, C. (2012), 'A novel approach for the simultaneous analysis of common and rare variants in complex traits.', *Bioinformatics and Biology Insights* **6**, 1–9.

Yuan, A., Yin, A. and Tan, M. T. (2021), 'Enhanced Doubly Robust Procedure for Causal Inference', *Statistics in Biosciences* .

Yuan, A., Zhou, Y. and Tan, M. T. (2020), 'Subgroup analysis with a nonparametric unimodal symmetric error distribution', *Communications in Statistics - Theory and Methods* **0**(0), 1–22.