Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Causal Inference for Genomic Data with Multiple Heterogeneous Outcomes

Jin-Hong Du, Zhenghao Zeng, Edward H. Kennedy, Larry Wasserman & Kathryn Roeder

To cite this article: Jin-Hong Du, Zhenghao Zeng, Edward H. Kennedy, Larry Wasserman & Kathryn Roeder (2025) Causal Inference for Genomic Data with Multiple Heterogeneous Outcomes, Journal of the American Statistical Association, 120:552, 2484-2497, DOI: 10.1080/01621459.2025.2468014

To link to this article: https://doi.org/10.1080/01621459.2025.2468014

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2025, VOL. 120, NO. 552, 2484–2497: Theory and Methods
https://doi.org/10.1080/01621459.2025.2468014

Taylor & Francis
Taylor & Francis Group

Causal Inference for Genomic Data with Multiple Heterogeneous Outcomes

Jin-Hong Du[a,b], Zhenghao Zeng[a], Edward H. Kennedy[a], Larry Wasserman[a,b], and Kathryn Roeder[a,c]

[a]Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA; [b]Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA; [c]Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA

ABSTRACT

With the evolution of single-cell RNA sequencing techniques into a standard approach in genomics, it has become possible to conduct cohort-level causal inferences based on single-cell-level measurements. However, the individual gene expression levels of interest are not directly observable; instead, only repeated proxy measurements from each individual's cells are available, providing a derived outcome to estimate the underlying outcome for each of many genes. In this article, we propose a generic semiparametric inference framework for doubly robust estimation with multiple derived outcomes, which also encompasses the usual setting of multiple outcomes when the response of each unit is available. To reliably quantify the causal effects of heterogeneous outcomes, we specialize the analysis to standardized average treatment effects and quantile treatment effects. Through this, we demonstrate the use of the semiparametric inferential results for doubly robust estimators derived from both Von Mises expansions and estimating equations. A multiple testing procedure based on Gaussian multiplier bootstrap is tailored for doubly robust estimators to control the false discovery exceedance rate. Applications in single-cell CRISPR perturbation analysis and individual-level differential expression analysis demonstrate the utility of the proposed methods and offer insights into the usage of different estimands for causal inference in genomics. Supplementary materials for this article are available online, including a standardized description of the materials available for reproducing the work.

ARTICLE HISTORY
Received April 2024
Accepted February 2025

KEYWORDS
Derived outcomes; Doubly robust estimation; Multiple testing; Quantile treatment effects; scRNA-seq data

1. Introduction

In observational studies, causal inference on multiple outcomes is increasingly prevalent in scientific discoveries (Imbens and Rubin 2015). Recent advances in high-throughput techniques have enabled the collection of large-scale repeated measurements across multiple subjects in various domains. However, subject-level outcomes, such as averages or inter-correlations of measurements within each subject, are often unobserved. Instead, researchers rely on repeated measurements to construct estimates of these outcomes, referred to as *derived outcomes* (Figure 1). For example, advancements in single-cell RNA sequencing (scRNA-seq) techniques (Editorial 2023) have enabled large-scale repeated gene expression measurements across multiple cells for each individual. These measurements allow researchers to construct derived outcomes (e.g., the sample average of gene expressions for an individual) as proxies for subject-level outcomes (e.g., the true average gene expression for that individual), facilitating individual-level comparisons (Zhang et al. 2022). The goal of subject-level causal inference is to determine which unobserved outcomes are causally affected by treatment by comparing derived outcomes between treatment and control groups. However, challenges such as unobservability of outcomes, subject heterogeneity (Qiu, Sun, and Zhou 2023), and non-identical outcome distributions limit the reliability of existing causal inference methods.

One major challenge of subject-level causal inference on scRNA-seq data is the unobservability of the outcomes. Individual gene expression levels of subjects are not directly measurable; instead, repeated measurements from heterogeneous cells are aggregated to serve as proxies. Additionally, variability among subjects may arise from latent states unique to each individual that influence gene expression patterns but remain hidden from direct observation (Du, Wasserman, and Roeder 2023). Consequently, derived outcomes often violate the classic assumption of being independent and identically distributed due to biological processes, experimental conditions, and inherent cellular heterogeneity. To analyze subject-level brain functional connectivities, prior work by Qiu, Sun, and Zhou (2023) attempts to estimate average treatment effects (ATEs) using inverse probability weighting (IPW) estimators (Imbens 2004; Tsiatis 2006). However, their approach relies on accurate propensity score modeling and assumes outcome homogeneity, which may not hold for genomic data.

A second challenge arises from the heterogeneity of gene expression data, which often exhibit variable scaling and right-skewed distributions, complicating comparisons across outcomes. This heterogeneity challenges the common use of ATEs since relying solely on mean differences in counterfactual distributions can lead to misleading conclusions. To estimate the treatment effects, one can rely on propensity score modeling or outcome modeling. For instance, one common strategy

CONTACT Kathryn Roeder ✉ *roeder@andrew.cmu.edu* Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA 15213.

Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JASA*.

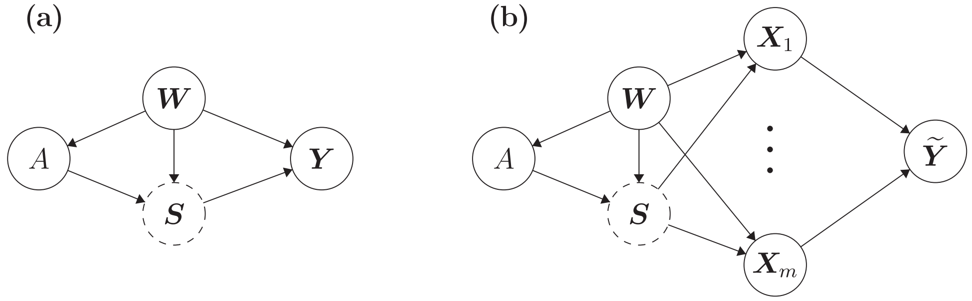

Figure 1. The causal diagram for the causal inference problems studied in this article. (a) Multiple outcomes. For a cell, its gene expression $\boldsymbol{Y} \in \mathbb{R}^p$ is causally affected by the treatment $A \in \mathbb{R}$, the latent state $\boldsymbol{S} \in \mathbb{R}^\ell$ and covariate $\boldsymbol{W} \in \mathbb{R}^q$ such as batch effects. (b) Multiple derived outcomes. In the subject-level studies, a subject's overall gene expression $\boldsymbol{Y}$ is not directly observed. Instead, repeated measurements of gene expressions $\boldsymbol{X}_1, \dots, \boldsymbol{X}_m \in \mathbb{R}^d$ of m cells from the subject provides a proxy $\widetilde{\boldsymbol{Y}}$ for $\boldsymbol{Y}$. See Section 3 for formal definitions. Note that the treatment effect of A on $\boldsymbol{Y}$ (or $\widetilde{\boldsymbol{Y}}$) is mediated by the latent state $\boldsymbol{S}$ even conditioned on the covariate $\boldsymbol{W}$. When conditioned on $\boldsymbol{W}$ and A, the outcomes $Y_1, \dots, Y_p$ within the same subject are still not independent and identically distributed.

in scRNA-seq analyses is to model outcomes directly using parameter models such as Poisson or Negative Binomial models (Sarkar and Stephens 2021), or zero-inflated models (Jiang et al. 2022). While using either IPW or regression estimators may seem intuitive, they are both sensitive to model specification.

Doubly robust (DR) estimators (Robins, Rotnitzky, and Zhao 1994; Scharfstein, Rotnitzky, and Robins 1999) offer a promising solution to mitigate sensitivity to model specification by combining IPW and outcome modeling. DR estimators are consistent as long as either of the two nuisance estimators is consistent, and $\sqrt{n}$-consistent whenever the nuisance estimators converge at only $n^{-1/4}$ rates (or more generally if the product of their errors is of the order $n^{-1/2}$). Additionally, if both the nuisance models are correctly specified, in the sense that the product of their errors is smaller order than $n^{-1/2}$, the DR estimators achieve the semiparametric efficiency bound for the unrestricted model, allowing the regression and propensity score functions to be estimated flexibly at slower than $n^{-1/2}$ rates in a wide variety of settings (Laan and Robins 2003; Tsiatis 2006). Recent work introduces a structure-agnostic framework for functional estimation, demonstrating that DR estimators are optimal for estimating ATEs when only black-box estimators of the outcome model and propensity score are available (Balakrishnan, Kennedy, and Wasserman 2023; Jin and Syrgkanis 2024). These results suggest that DR estimators cannot be improved upon without making additional structural assumptions. Given these advantages, exploring doubly robust estimation in settings with multiple heterogeneous outcomes is crucial, as it mitigates model misspecification and enables reliable statistical testing when nonparametric methods are used for outcome and propensity score estimation.

The unobservability of subject-level outcomes, heterogeneity in gene expression distributions, and sensitivity to model specification collectively challenge traditional causal inference methods in scRNA-seq studies. To address these issues, we propose a semiparametric inference framework to handle multiple derived outcomes effectively. Specifically, (i) we define causal estimands that capture meaningful counterfactual differences across multiple outcomes and establish identification conditions under hierarchical models where outcomes of interest are unobserved (Figure 1(b)); and (ii) we develop robust and efficient estimators tailored for these estimands. Additionally, we extend multiple-testing procedures to control statistical errors during simultaneous inference on high-dimensional derived outcomes. Together, these methodological advancements provide a unified approach incorporating doubly robust estimation to handle multiple derived outcomes effectively.

Focusing on multiple derived outcomes, we first establish generic results on semiparametric inference with doubly robust estimators. It also encompasses the usual setting of multiple outcomes when the response of each unit is available. By using finite-sample maximal inequality for finite maximums, we obtain interpretable conditions of uniform estimation error control for the empirical process terms and the asymptotic variances. We derive the uniform convergence rates, in terms of only finitely many moments of the influence functions' envelope and the maximum second moments of the individual estimation errors, allowing for the number of outcomes p to be potentially exponentially larger than the sample size n.

To address the challenges of outcome heterogeneity in single-cell data, we further specialize our analysis to standardized average treatment effects for comparing treatment effects across different outcomes on a common scale and quantile treatment effects for robustness against outliers. This demonstrates how generic semiparametric inferential results for DR estimators derived from von Mises expansions and estimating equations can be applied in high-dimensional settings. Furthermore, we adapt Gaussian approximation results from Chernozhukov, Chetverikov, and Kato (2013) to DR estimators and implement a step-down procedure to control false discovery (exceedance) rates with guaranteed power (Genovese and Wasserman 2006).

Our exploration includes two real-world application scenarios of the proposed causal inference methods. (a) *Single-cell CRISPR perturbation analysis*: Gene expressions from single cells are compared between perturbation and control groups in CRISPR experiments to identify target genes of individual perturbations and analyze the effects of perturbations (Dixit et al. 2016), as shown in Figure 1(a). (b) *Individual level differential expression analysis*: Aggregated gene expressions from individual subjects under two conditions (case and control) are analyzed to identify genes intrinsically affected by these conditions across subjects, corresponding to Figure 1(b). By applying our methods to these datasets, we demonstrate their practical utility while highlighting the strengths and limitations of different causal estimands. These findings emphasize the importance of suitable causal inference techniques for the accurate interpretation of genomic data.

Organization. In Section 2, we review and extend the classic semiparametric inference framework to the setting of multiple outcomes. In Section 3, we set up the problem of interest and discuss the identification conditions for the causal estimands. In Section 4, we analyze two DR estimators for standardized and quantile treatment effects and study their statistical properties. In Section 5, we study the statistical error of simultaneous inference and propose a multiple testing procedure for controlling the false discovery rate. In Sections 6 and 7, we conduct simulations and scRNA-seq data analysis to validate the proposed simultaneous causal inference method. A detailed review of related work is provided in (Section A).

Notation. Throughout our exposition, we will use the following notational conventions. We denote scalars in non-bold lower or upper case (e.g., X), vectors in bold upper case (e.g., $\boldsymbol{X}$), and matrices in non-italic bold upper case (e.g., $\mathbf{X}$). For $a, b \in \mathbb{R}$, we write $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. For any random variable X, its L_q norm is defined by $\|X\|_{L_q} = (\mathbb{E}[|X|^q])^{1/q}$ for $q = 1, \dots, \infty$. For $p \in \mathbb{N}$, $[p] := \{1, \dots, p\}$. For a set $\mathcal{A}$, let $|\mathcal{A}|$ be its cardinality. For (potentially random) measurable functions f, we denote expectations with respect to Z alone by $\mathbb{P}f(Z) = \int f \, \mathrm{d}\mathbb{P}$, and with respect to both Z and the observations where f is fitted on by $\mathbb{E}[f(Z)]$. The empirical expectation is denoted by $\mathbb{P}_n f(Z) = \frac{1}{n}\sum_{i=1}^n f(Z_i)$ for iidsamples $Z_1, \dots, Z_n$ of Z. Similarly, the population and empirical variances are denoted by $\mathbb{V}$ and $\mathbb{V}_n$, respectively. We write the (conditional) L_p norm of f as $\|f\|_{L_p} = \left[\int f(z)^p \, \mathrm{d}\mathbb{P}(z)\right]^{1/p}$ for $p \geq 1$. We use "o" and "$\mathcal{O}$" to denote the little-o and big-O notations and let "$o_{\mathbb{P}}$" and "$\mathcal{O}_{\mathbb{P}}$" be their probabilistic counterparts. For sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \ll b_n$ or $b_n \gg a_n$ if $a_n = o(b_n)$; $a_n \lesssim b_n$ or $b_n \gtrsim a_n$ if $a_n = \mathcal{O}(b_n)$; and $a_n \asymp b_n$ if $a_n = \mathcal{O}(b_n)$ and $b_n = \mathcal{O}(a_n)$. All the constants c, c_1, c_2 and C, C_1, C_2 may vary from line to line. Convergence in distribution and probability are denoted by "$\xrightarrow{\mathrm{d}}$" and "$\xrightarrow{\mathrm{P}}$".

2. Semiparametric Inference with Multiple Outcomes

Prior to delving into our main topic, this section takes a brief excursion into the formulation of semiparametric inference within the context of multiple outcomes, laying the foundation for addressing our specific problem in subsequent sections.

Let $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n$ be observations of iid samples from a population $\mathcal{P}$. In the presence of multiple outcomes, we are interested in estimating p target estimands $\tau_j : \mathcal{P} \mapsto \mathbb{R}$ for $j = 1, \dots, p$. Suppose τ_j admits a von Mises expansion; that is, there exists an influence function $\varphi_j(z; \mathbb{P})$ with $\int \varphi_j(z; \mathbb{P}) \, \mathrm{d}\mathbb{P}(z) = 0$ and $\int \varphi_j(z; \mathbb{P})^2 \, \mathrm{d}\mathbb{P}(z) < \infty$, such that

$$\tau_j(\overline{\mathbb{P}}) - \tau_j(\mathbb{P}) = \int \varphi_j(z; \overline{\mathbb{P}}) \, \mathrm{d}(\overline{\mathbb{P}} - \mathbb{P}) + T_{\mathrm{R},j},$$

where $T_{\mathrm{R},j}$ is a second-order remainder term (which means it only depends on products or squares of differences between $\overline{\mathbb{P}}$ and $\mathbb{P}$). The influence function quantifies the effect of an infinitesimal contamination at the point z on the estimate, standardized by the mass of the contamination. Its historical development and definition under diverse sets of regularity conditions can be found at Hampel et al. (2011, sec. 2.1). The above expansion suggests a one-step estimator that corrects the bias of the plug-in estimator $\tau_j(\widehat{\mathbb{P}})$:

$$\widehat{\tau}_j(\mathbb{P}) := \tau_j(\widehat{\mathbb{P}}) + \mathbb{P}_n\{\varphi_j(\boldsymbol{Z}; \widehat{\mathbb{P}})\}, \tag{2.1}$$

where $\widehat{\mathbb{P}}$ is an estimate of $\mathbb{P}$. For many estimands, such as ATE and expected conditional covariance, the one-step estimator is also a DR estimator. Although for certain estimands like expected density, the standard one-step estimator is not doubly robust, it still has nuisance errors that consist of a second-order term (Kennedy, Balakrishnan, and Wasserman 2023). Then, the one-step estimator $\widehat{\tau}_j$ of the jth target estimand τ_j admits a three-term decomposition of the estimation error (Kennedy 2022, eq. (10))[1]:

$$\begin{aligned}\widehat{\tau}_j(\mathbb{P}) - \tau_j(\mathbb{P}) = &\underbrace{(\mathbb{P}_n - \mathbb{P})\{\varphi_j(\boldsymbol{Z}; \mathbb{P})\}}_{T_{\mathrm{S},j}} \\ &+ \underbrace{(\mathbb{P}_n - \mathbb{P})\{\varphi_j(\boldsymbol{Z}; \widehat{\mathbb{P}}) - \varphi_j(\boldsymbol{Z}; \mathbb{P})\}}_{T_{\mathrm{E},j}} + T_{\mathrm{R},j}.\end{aligned} \tag{2.2}$$

In the above decomposition, the first term after $\sqrt{n}$-scaling has an asymptotic normal distribution by the central limit theorem. That is, $\sqrt{n}\, T_{\mathrm{S},j} \xrightarrow{\mathrm{d}} \mathcal{N}(0, \mathbb{V}[\varphi_j(\boldsymbol{Z}; \mathbb{P})])$. The higher-order term $T_{\mathrm{R},j}$ is usually negligible and has an order of $o_{\mathbb{P}}(n^{-1/2})$ under certain conditions. On the other hand, the empirical process term $T_{\mathrm{E},j}$ will be asymptotically negligible (i.e., $o_{\mathbb{P}}(n^{-1/2})$) under Donsker assumption (van der Vaart 2000) or sample splitting (Chernozhukov et al. 2018; Kennedy, Balakrishnan, and G'Sell 2020), because it is a sample average with a shrinking variance. In our problem setting with an increasing number of outcomes, uniform control over all the empirical process terms $T_{\mathrm{E},j}$ for $j = 1, \dots, p$ is desired to facilitate the construction of simultaneous confidence intervals. Below is an extension of Lemma 2 from Kennedy, Balakrishnan, and G'Sell (2020, Appendix B) to the setting of multiple outcomes.

Lemma 1 (Uniform control of the empirical process terms). Let $\mathbb{P}_n$ denote the empirical measure over $\mathcal{D}_n = (\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n) \in \mathcal{Z}^n$, and let $g_j : \mathcal{Z} \to \mathbb{R}$ be a (possibly random) function independent of $\mathcal{D}_n$ for $j = 1, \dots, p$ with $p \geq 2$. Let $G(\cdot) := \max_{1 \leq j \leq p} |g_j(\cdot)|$ denote the envelope of $g_1, \dots, g_p$. If $\max_{1 \leq j \leq p} \|g_j\|_{L_2} < \infty$ and $\|G\|_{L_q} < \infty$ for some $q \in \mathbb{N} \cup \{\infty\}$, then the following statements hold:

$$\begin{aligned}&\mathbb{E}\left[\max_{j=1,\dots,p} |(\mathbb{P}_n - \mathbb{P})g_j| \,\middle|\, \{g_j\}_{j=1}^p\right] \\ &\quad\lesssim \left(\frac{\log p}{n}\right)^{1/2} \max_{1 \leq j \leq p} \|g_j\|_{L_2} + \left(\frac{\log p}{n}\right)^{1-1/q} \|G\|_{L_q}.\end{aligned}$$

The proof of Lemma 1 uses a finite-sample maximal inequality established in Kuchibhotla and Patra (2022) for high-dimensional estimation problems. When specified to particular target estimands, Lemma 1 suggests that $T_{\mathrm{E},j}$ in (2.2) can be

[1] For certain estimands, such as average treatment effects and expected conditional covariance functionals, it is usually more convenient to use the uncentered influence functions $\phi_j(\boldsymbol{Z}; \mathbb{P}) := \varphi_j(\boldsymbol{Z}; \mathbb{P}) + \tau_j(\mathbb{P})$ in the decomposition.

uniformly controlled over $j = 1, \dots, p$, if one can derive the uniform L_2-norm bound and the L_q-norm bound of the envelope for the estimation error of the influence functions $g_j = \varphi_j(\mathbf{Z}; \widehat{\mathbb{P}}) - \varphi_j(\mathbf{Z}; \mathbb{P})$. In particular, if g_j's are bounded, and $\log p \cdot (\max_{1\le j\le p} \|g_j\|_{L_2}^2 \vee n^{-1/2}) = o(1)$, then we have that $\max_{1\le j\le p} T_{\mathrm{E},j} = o_{\mathbb{P}}(n^{-1/2})$, which is negliable after scaled by $\sqrt{n}$. This allows the number of outcomes p to be potentially exponentially larger than the number of samples n. It is important to note that similar bounds on the empirical process term can still be derived even when the nuisance functions are not trained on an independent sample, provided certain complexity measures of the function class $\mathcal{F}_j$ that g_j belongs to are properly bounded; see (Remark B.2 in Appendix B.1) for more details.

Remark 1 (One-step estimator from estimating equations). Above, we construct the one-step estimator based on the influence function φ_j of τ_j from von Mises expansion. One can also construct efficient estimators of pathwise differentiable functionals through estimating equations, which is related to the quantile estimand as we will discuss in Section 4.2.

When $T_{\mathrm{E},j} + T_{\mathrm{R},j} = o_{\mathbb{P}}(n^{-1/2})$, by central limit theorem we have that $\sqrt{n}(\widehat{\tau}_j(\mathbb{P}) - \tau_j(\mathbb{P})) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma_j^2)$ where $\sigma_j^2 = \mathbb{V}[\varphi_j(\mathbf{Z}; \mathbb{P})]$. To construct confidence intervals, one can use the sample variance $\widehat{\sigma}_j^2 = \mathbb{V}_n[\varphi_j(\mathbf{Z}; \widehat{\mathbb{P}})]$ to consistently estimate the asymptotic variance. To derive the properties of test statistics and confidence intervals in high dimensions when $p \gg n$, it is necessary to establish strong control on the uniform convergence rate of the variance estimates over $j = 1, \dots, p$; see, for example, Qiu, Sun, and Zhou (2023, Proposition 2) and Chernozhukov, Chetverikov, and Kato (2013, Comment 2.2). In this regard, the following lemma provides general conditions for bounding the uniform estimation error.

Lemma 2 (Uniform control of the variance estimates). Denote $\varphi_j = \varphi_j(\mathbf{Z}; \mathbb{P})$, $\widehat{\varphi}_j = \varphi_j(\mathbf{Z}; \widehat{\mathbb{P}})$, $\Phi = \max_{1\le j\le p} |\widehat{\varphi}_j - \varphi_j|$, and $\Psi = \max_{1\le j\le p} |\varphi_j|$. Suppose the following conditions hold:

1. Envelope: $\max_{1\le j\le p} |\widehat{\varphi}_j + \varphi_j| \lesssim 1$ and $\|\Psi\|_{L_q} + \max_{k=1,2} \|\Phi^k\|_{L_q} \lesssim r_{1n}$ for some $q > 1$,
2. Estimation error: $\max_{1\le j\le p} \|\widehat{\varphi}_j - \varphi_j\|_{L_2} \lesssim r_{2n}$, $\max_{1\le j\le p} |\mathbb{P}[\widehat{\varphi}_j - \varphi_j]| \lesssim r_{3n}$,

with probability tending to one. Then, it holds that

$$\max_{1\le j\le p} |\widehat{\sigma}_j^2 - \sigma_j^2| \lesssim \mathcal{O}_{\mathbb{P}}\left(\left(\frac{\log p}{n}\right)^{1-1/q} r_{1n} + \left(\frac{\log p}{n}\right)^{1/2} r_{2n} + r_{3n} \right).$$

Note that r_{1n} and r_{2n} are allowed to potentially diverge. We will use Lemma 2 for the purpose of multiple testing, as demonstrated in Section 5. Typically, to accommodate an exponentially large number of outcomes p relative to n while maintaining valid statistical inference, it suffices to ensure that variance estimates are consistent at sample size n; specifically, $\max_{1\le j\le p} |\widehat{\sigma}_j^2 - \sigma_j^2| = o_{\mathbb{P}}(n^{-\alpha})$ for some constant $\alpha > 0$.

3. Subject-Level Causal Inference with Multiple Outcomes

We consider an increasingly popular study design where scRNA-seq data are collected from multiple individuals and the question of interest is to find genes that are causally differentially expressed between two groups of individuals, based on repeated single-cell measurements.

3.1. Causal Inference with Multiple Derived Outcomes

Suppose a subject can be either in the case or control group, indicated by a binary random variable $A \in \{0, 1\}$ and we sequence the expressions $\mathbf{X} = [\boldsymbol{X}_1, \dots, \boldsymbol{X}_m]^\top \in \mathbb{R}^{m\times d}$ of d genes in m cells,[2] along with subject-level covariates $\boldsymbol{W} \in \mathbb{R}^q$. Let $\mathbf{X}(a)$ denote the *potential response* of gene expressions. To characterize the complex biological processes, suppose $\boldsymbol{S}(a) \in \mathbb{R}^\ell$ is the latent potential state after receiving treatment $A = a$, which fully captures the effects of the treatment on the individual. We assume $\boldsymbol{X}_1(a), \dots, \boldsymbol{X}_m(a)$ are conditionally independent and identically distributed given $\boldsymbol{S}(a)$ and $\boldsymbol{W}$.[3] Marginally, however, they can be highly dependent because of repeated measurements from the same individual. As an example of genomics data, $\boldsymbol{S}$ can be the chromatin accessibility that governs the translation and expression of genes, while $\boldsymbol{X}_m$ is the resulting expression level of those genes.

Suppose the collection of treatment assignment, covariates, subject level parameters, and potential responses $(A, \boldsymbol{W}, \boldsymbol{S}(0), \boldsymbol{S}(1), \mathbf{X}(0), \mathbf{X}(1))$ is from some super-population $\mathcal{P}$. We require the consistency assumption on the observed response.

Assumption 1 (Consistency). The observed response is given by $\mathbf{X} = A\mathbf{X}(1) + (1 - A)\mathbf{X}(0)$.

When comparing gene expressions between two groups of individuals, the p-dimensional subject-level parameter of interest, is the *potential outcome* $\boldsymbol{Y}(a) \in \mathbb{R}^p$, a functional that maps the conditional distribution of $\boldsymbol{X}_1(a)$ given $\boldsymbol{S}(a)$ and $\boldsymbol{W}$ to $\mathbb{R}^p$:

$$\boldsymbol{Y}(a) = \mathbb{E}\left[f(\boldsymbol{X}_1(a)) \,\middle|\, \boldsymbol{S}(a), \boldsymbol{W} \right], \tag{3.1}$$

for some prespecified function $f : \mathbb{R}^d \to \mathbb{R}^p$. The choice of f should align with the user's research goals and the specific aspects of the data they intend to capture. For example, when f is the identity map and $p = d$, the potential outcome $\boldsymbol{Y}(a)$ represents the conditional mean; when $f(\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{X}^\top - \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\top \mid \boldsymbol{S}(a), \boldsymbol{W}]$ and $p = d$, it represents the intrasubject covariance matrix. When considering conditional means among the potential responses $X_{1j}(a)$'s, it can also be nodewise regression coefficients as considered in Qiu, Sun, and Zhou (2023). Intuitively, $\boldsymbol{Y}(a)$ is an individual / within-group characteristic that depends on the conditional distribution of $\boldsymbol{X}_1(a)$ given $\boldsymbol{S}(a), \boldsymbol{W}$. From Assumption 1, we also have $\boldsymbol{Y} = A\boldsymbol{Y}(1) + (1 - A)\boldsymbol{Y}(0)$. Compared to the classical causal inference setting,

[2] For notational simplicity, we treat the number of cell m as fixed across subjects, though the method also applies when the number of cell m_i varies for subjects $i = 1, \dots, n$.

[3] Technically, the potential response can be denoted as $\boldsymbol{X}_m(\boldsymbol{S}(a))$; however, because $\boldsymbol{S}(a)$ is unobservable and the variable to intervene is A, we use a simplified notation $\boldsymbol{X}_m(a)$ to denote the potential response.

the subject-level outcome $\boldsymbol{Y}$ is not observed for each subject, while only the repeated measurements of gene expressions $\mathbf{X}$ from multiple cells are available and can be used to construct a derived outcome $\widetilde{\boldsymbol{Y}}$. For a given f, we consider a statistic $\widetilde{\boldsymbol{Y}} := g(\mathbf{X})$ for some function $g : (\boldsymbol{X}_1, \dots, \boldsymbol{X}_m) \mapsto \widetilde{\boldsymbol{Y}}$. There can be different choices of g to estimate $\boldsymbol{Y}(a)$ by $\widetilde{\boldsymbol{Y}}(a)$. For instance, if f is a linear function for the potential outcome $\boldsymbol{Y}(a)$ in (3.1), then g can be a simple sample average as a natural choice of the derived outcome; alternatively, g can also be the median-of-means estimator as the derived outcomes.

Under the derived outcomes framework, Qiu, Sun, and Zhou (2023) studied the IPW estimator for ATE:

$$\tau_j^{\mathrm{ATE}} = \mathbb{E}[Y_j(1) - Y_j(0)], \qquad j = 1, \dots, p. \tag{3.2}$$

By focusing on the expected potential outcomes, we next describe the identification condition and semiparametric inferential results based on derived outcomes $\widetilde{\boldsymbol{Y}}$.

Identification. We require two extra classic causal assumptions for observational studies.

Assumption 2 (Positivity). The propensity score $\pi_a(\boldsymbol{W}) := \mathbb{P}(A = a \mid \boldsymbol{W}) \in (0, 1)$.

Assumption 3 (No unmeasured confounders). $A \perp\!\!\!\perp \mathbf{X}(a) \mid \boldsymbol{W}$, for all $a \in \{0, 1\}$.

The above assumptions on the propensity score and the potential responses are standard for observational studies in the causal inference literature (Imbens and Rubin 2015; Kennedy 2022). Assumption 2 suggests that both treated and control units of each subject can be found for any value of the covariate with a positive probability. Assumption 3 ensures that the treatment assignment is fully determined by the observed covariate $\boldsymbol{W}$. These assumptions are required to estimate functionals of $\mathbf{X}(a)$ with observed variables $(A, \boldsymbol{W}, \mathbf{X})$.

Let $\boldsymbol{Z} = (A, \boldsymbol{W}, \mathbf{X}, \boldsymbol{Y})$ denote the tuple of observed random variables and unobserved outcomes $\boldsymbol{Y}$. Because the outcome $\boldsymbol{Y}$ is not observed for each subject, we are interested in constructing a proxy $\widetilde{\boldsymbol{Y}} = g(\mathbf{X})$ of $\boldsymbol{Y}$ from repeated measurements $\boldsymbol{X}_1, \dots, \boldsymbol{X}_m$ from the same subject. Analogously, we denote $\widetilde{\boldsymbol{Y}}(a) := g(\mathbf{X}(a))$ for the potential outcomes and quantify the bias as $\boldsymbol{\Delta}_m(a) := \mathbb{E}[\widetilde{\boldsymbol{Y}}(a) \mid \boldsymbol{W}, \boldsymbol{S}(a)] - \boldsymbol{Y}(a)$. Below, we introduce a notion of asymptotic unbiased estimate in Definition 1, where the expected bias is negligible uniformly over multiple outcomes.

Definition 1 (Asymptotic unbiased estimate). For $a \in \{0, 1\}$, the derived outcome $\widetilde{\boldsymbol{Y}}(a)$ is asymptotic unbiased to $\boldsymbol{Y}(a)$ if the bias tends to zero: $\max_{1 \le j \le p} |\mathbb{E}[\Delta_{mj}(a)]| = o(1)$ as $m \to \infty$.

Note that when $\widetilde{\boldsymbol{Y}}(a)$ is marginally unbiased, that is, $\mathbb{E}[\boldsymbol{Y}(a)] = \mathbb{E}[\widetilde{\boldsymbol{Y}}(a)]$, it also implies that $\widetilde{\boldsymbol{Y}}(a)$ is asymptotically unbiased to $\boldsymbol{Y}(a)$. Therefore, our framework also includes the common setting where all the outcomes $\boldsymbol{Y}(a) = \widetilde{\boldsymbol{Y}}(a) = \boldsymbol{X}_1(a)$ (with $m = 1$) are observed. When $\widetilde{\boldsymbol{Y}}(a)$ is an asymptotic unbiased estimate of $\boldsymbol{Y}(a)$, Lemma 1 from Qiu, Sun, and Zhou (2023) suggests that the counterfactual of unobserved outcomes can be identified asymptotically, as detailed in Proposition 3.

Proposition 3 (Identification of linear functionals). Under Assumptions 1–3, if $\widetilde{\boldsymbol{Y}}(a)$ is asymptotically unbiased to $\boldsymbol{Y}(a)$, then $\mathbb{E}[\boldsymbol{Y}(a)]$ can be identified by $\mathbb{E}[\mathbb{E}[\widetilde{\boldsymbol{Y}} \mid \boldsymbol{W}, A = a]]$ as $m \to \infty$.

Semiparametric inference. When the target causal estimands are the expectation of the potential outcomes $\tau_j = \mathbb{E}[Y_j(a)]$ for $j = 1, \dots, p$, one can adopt results from Section 2 to establish the asymptotic normality of certain estimators under proper assumptions on the convergence rate of the nuisance function estimates. However, because $\boldsymbol{Y}$ is unobservable, we are not able to directly estimate its influence function and hence its influence-function-based one-step estimator (2.1). Instead, we can rely on the influence function of $\widetilde{\tau}_j = \mathbb{E}[\widetilde{Y}_j(a)]$: $\widetilde{\varphi}_j(\boldsymbol{Z}; \mathbb{P}) = \mathbb{1}\{A = a\}\pi_a(\boldsymbol{W})^{-1}(\widetilde{Y}_j - \mu_{aj}(\boldsymbol{W})) + \mu_{aj}(\boldsymbol{W}) - \widetilde{\tau}_j(\mathbb{P})$, where $\pi_a(\boldsymbol{W}) = \mathbb{P}(A = a \mid \boldsymbol{W})$ and $\mu_{aj}(\boldsymbol{W}) = \mathbb{E}[\widetilde{Y}_j \mid A = a, \boldsymbol{W}]$ for $j = 1, \dots, p$, are the propensity score and regression functions, respectively. This, in turn, yields an analog of the one-step estimator (2.1):

$$\widehat{\widetilde{\tau}}_j(\mathbb{P}) := \widetilde{\tau}_j(\widehat{\mathbb{P}}) + \mathbb{P}_n\{\widetilde{\varphi}(\boldsymbol{Z}; \widehat{\mathbb{P}})\},$$

which further implies the decomposition of the estimation error for the causal estimand τ_j:

$$\widehat{\widetilde{\tau}}_j(\mathbb{P}) - \tau_j(\mathbb{P}) = T_{\mathrm{S},j} + T_{\mathrm{E},j} + T_{\mathrm{R},j} + \mathbb{E}[\Delta_{mj}], \tag{3.3}$$

where the sample average term $T_{\mathrm{S},j}$, the empirical process term $T_{\mathrm{E},j}$ and the reminder term $T_{\mathrm{R},j}$ are as in (2.2) with φ_j replaced by $\widetilde{\varphi}_j$. The asymptotic variance $\sigma_j^2 = \mathbb{V}[\widetilde{\varphi}_j(\boldsymbol{Z}; \mathbb{P})]$ can be estimated by the empirical variance $\widehat{\sigma}_j^2 = \mathbb{V}_n[\widetilde{\varphi}_j(\boldsymbol{Z}; \widehat{\mathbb{P}})]$ analogously. However, the application of Lemmas 1 and 2 would require the verification of conditions for the perturbed influenced function $\widetilde{\varphi}_j$ instead of φ_j. Similar ideas apply to the one-step and doubly robust estimators of other target estimands.

3.2. Beyond Average Treatment Effects

For single-cell gene expressions exhibiting different scales and skew-distributed, simply comparing the average treatment effects (3.2) may not be reliable. One approach to improve on the naive estimand is to consider standardized average treatment effects (STE):

$$\tau_j^{\mathrm{STE}} = \frac{\mathbb{E}[Y_j(1) - Y_j(0)]}{\sqrt{\mathbb{V}[Y_j(0)]}}, \qquad j = 1, \dots, p \tag{3.4}$$

which allows for consistent and comparative analysis across different scales and variances, enhancing the interpretability and comparability of treatment effects in diverse and complex datasets (Kennedy, Kangovi, and Mitra 2019). Another approach is to consider quantile effects (QTE):

$$\tau_j^{\mathrm{QTE}_\varrho} = Q_\varrho[Y_j(1)] - Q_\varrho[Y_j(0)], \qquad j = 1, \dots, p, \tag{3.5}$$

where $Q_\varrho[U]$ denote the ϱ-quantile of random variable U. In particular, when $\varrho = 0.5$, the ϱ-quantile equals the median $Q_\varrho(U) = \mathrm{Med}(U)$, and we reveal one of the commonly used robust estimand $\tau_j^{\mathrm{QTE}} = \mathrm{Med}[Y_j(1)] - \mathrm{Med}[Y_j(0)]$ for location-shift hypotheses. QTE may be more robust and less affected by the outliers of gene expressions (Chakrabortty, Dai, and Tchetgen 2022; Kallus, Mao, and Uehara 2024).

Note that the identification condition and semiparametric inferential results in Section 3.1 do not apply directly to target estimands other than ATE. Therefore, efforts are required to generalize the results to include STE and QTE for multiple derived outcomes. This demonstrates the utility and validity of our semiparametric inferential framework on one-step estimators defined through the von Mises expansion and the formulation of estimating equations, respectively, as investigated next.

4. Doubly Robust Estimation

In this section, we analyze the DR estimators for STE and QTE, which exemplify the application of general theoretical results in Section 2 to specific target estimands.

4.1. Standardized Average Effects

Recall the standardized average treatment effects $\tau_j^{\rm STE}$ defined in (3.4), for $j = 1,\dots,p$. The following lemma provides the identified forms of STE based on observational data.

Lemma 4 (Identification of standardized average effects). Under Assumptions 1–3, if $\mathbb{V}[Y_j(0)] > 0$ and $\widetilde{Y}_j(a)^k$ is asymptotically unbiased to $Y_j(a)^k$ for $k = 1,2$ and $a = 0,1$ such that $k + a \le 2$, that is, the bias of the derived outcomes $\Delta_{mkj}(a) := \mathbb{E}[\widetilde{Y}_j(a)^k \mid \boldsymbol{W}, \boldsymbol{S}(a)] - Y_j(a)^k$ satisfies that $\delta_m := \max_{k,a}\max_{1\le j\le p}|\mathbb{E}[\Delta_{mkj}(a)]| = o(1)$, then as $m \to \infty$, the standardized average treatment effect $\tau_j^{\rm STE}$ can be identified by $\tau_j^{\rm STE} = \widetilde{\tau}_j^{\rm STE} + o_{\mathbb{P}}(1)$ where

$$\widetilde{\tau}_j^{\rm STE} := \frac{\mathbb{E}[\mathbb{E}(\widetilde{Y}_j \mid A = 1, \boldsymbol{W})] - \mathbb{E}[\mathbb{E}(\widetilde{Y}_j \mid A = 0, \boldsymbol{W})]}{\sqrt{\mathbb{E}[\mathbb{E}(\widetilde{Y}_j^2 \mid A = 0, \boldsymbol{W})] - \mathbb{E}[\mathbb{E}(\widetilde{Y}_j \mid A = 0, \boldsymbol{W})]^2}}. \tag{4.1}$$

As suggested by Lemma 4, estimating STE requires estimating the conditional expectation of $\widetilde{Y}_j^k$ given $A = a$ and $\boldsymbol{W}$. For this purpose, we consider the DR estimator of treatment effect $\mathbb{E}[\widetilde{Y}_j(1)]$:

$$\widetilde{\phi}_{akj}(\boldsymbol{Z};\pi_a,\boldsymbol{\mu}_a) := \frac{\mathbb{1}\{A = a\}}{\pi_a(\boldsymbol{W})}(\widetilde{Y}_j^k - \mu_{akj}(\boldsymbol{W})) + \mu_{akj}(\boldsymbol{W}),$$

where $\boldsymbol{\mu}_a : \mathbb{R}^d \to \mathbb{R}^{2\times p}$ is the mean regression function with entry $\mu_{akj}(\boldsymbol{W}) = \mathbb{E}[\widetilde{Y}_j^k \mid \boldsymbol{W}, A = a]$ and $\pi_a(\boldsymbol{W}) = \mathbb{P}(A = a \mid \boldsymbol{W})$ is the propensity score function. By plugging in the DR estimators for individual counterfactual expectations consisting in (4.1), we obtain a natural estimator for the STE:

$$\widehat{\tau}_j^{\rm STE} = \frac{\mathbb{P}_n\{\widetilde{\phi}_{11j}(\boldsymbol{Z};\widehat{\pi}_1,\widehat{\boldsymbol{\mu}}_1) - \widetilde{\phi}_{01j}(\boldsymbol{Z};\widehat{\pi}_0,\widehat{\boldsymbol{\mu}}_0)\}}{\sqrt{\mathbb{P}_n\{\widetilde{\phi}_{02j}(\boldsymbol{Z};\widehat{\pi}_0,\widehat{\boldsymbol{\mu}}_0)\} - \mathbb{P}_n\{\widetilde{\phi}_{01j}(\boldsymbol{Z};\widehat{\pi}_0,\widehat{\boldsymbol{\mu}}_0)\}^2}}, \tag{4.2}$$

which is also the DR estimator of $\widetilde{\tau}_j^{\rm STE}$. The following theorem shows that under mild conditions, the above estimator $\widehat{\tau}_j^{\rm STE}$ is doubly robust for estimating $\tau_j^{\rm STE}$, with the remainder terms uniformly controlled over all outcomes.

Theorem 5 (Linear expansion of STE). Under Assumptions 1–3 and the identification condition in Lemma 4, consider the one-step estimator (4.2), where $\mathbb{P}_n$ is the empirical measure over $\mathcal{D} = \{\boldsymbol{Z}_1,\dots,\boldsymbol{Z}_n\}$ and $(\widehat{\pi}_a,\widehat{\boldsymbol{\mu}}_a)$ is an estimate of $(\pi_a,\boldsymbol{\mu}_a)$ for $a = 0,1$ from samples independent of $\mathcal{D}$. Suppose the following hold for $k = 1,2$ and $a = 0,1$ with probability tending to one:

1. Boundedness: There exists $c, C > 0$ and $\epsilon \in (0,1)$ such that $\max\{|Y_j|,|\widetilde{Y}_j|\} < C$, $\max\{\|\mu_{akj}\|_{L_\infty},\|\widehat{\mu}_{akj}\|_{L_\infty}\} < C$, $\mathbb{V}[Y_j(0)] > c$ for all $j \in [p]$, and $\pi_a,\widehat{\pi}_a \in [\epsilon, 1-\epsilon]$.
2. Nuisance: The rates of nuisance estimates are $\max_{j\in[p]}\|\widehat{\mu}_{akj} - \mu_{akj}\|_{L_2} = \mathcal{O}(n^{-\alpha})$ and $\|\widehat{\pi}_a - \pi_a\|_{L_2} = \mathcal{O}(n^{-\beta})$ for some $\alpha,\beta \in (0,1/2)$ such that $\alpha + \beta > 1/2$.

Then as $m,n,p \to \infty$, it holds that $\widehat{\tau}_j^{\rm STE} - \tau_j^{\rm STE} = \mathbb{P}_n\{\widetilde{\varphi}_j^{\rm STE}\} + \varepsilon_j$, $j = 1,\dots,p$, where the residual terms satisfy $\max_{j\in[p]}|\varepsilon_j| = \mathcal{O}_{\mathbb{P}}(n^{-(\alpha+\beta)} + \vartheta^{\rm STE}\sqrt{(\log p)/n} + (\log p)/n + \delta_m)$ with $\vartheta^{\rm STE} := n^{-(\alpha\wedge\beta)}$ and the influence function is given by

$$\widetilde{\varphi}_j^{\rm STE} = \frac{\widetilde{\phi}_{11j} - \widetilde{\phi}_{01j}}{\sqrt{\mathbb{V}[\widetilde{Y}_j(0)]}} - \widetilde{\tau}_j^{\rm STE}\left[\frac{\widetilde{\phi}_{02j} + \mathbb{E}[\widetilde{Y}_j(0)^2] - 2\mathbb{E}[\widetilde{Y}_j(0)]\widetilde{\phi}_{01j}}{2\,\mathbb{V}[\widetilde{Y}_j(0)]}\right]. \tag{4.3}$$

The proof of Theorem 5 requires the analysis of the linear expansions for the individual counterfactual expectations $\mathbb{E}[Y_j^k(a)]$ (see (Lemma D.1)). It then requires the application of the delta method to derive the uniform convergence rates of the residuals over multiple outcomes. For the residuals' rate, the term $n^{-(\alpha+\beta)}$ is the product of the two nuisance estimation errors, which shows the benefit of the double robustness property, while the term $\vartheta^{\rm STE}\sqrt{(\log p)/n} + (\log p)/n$ is related to the empirical process terms of individual counterfactuals by applying Lemma 1. From triangular-array central limit theorem (see (Lemma D.4)), a direct consequence of Theorem 5 is the asymptotic normality of individual STE estimators, as presented in Corollary 6.

Corollary 6 (Asymptotic normality). Under conditions in Theorem 5, when $(\vartheta^{\rm STE} \vee n^{-1/4})\sqrt{\log p} = o(1)$, $\delta_m = o(n^{-1/2})$ and $\sigma_j^2 := \mathbb{V}(\widetilde{\varphi}_j^{\rm STE}(\boldsymbol{Z};\pi,\boldsymbol{\mu})) \ge c$ for some constant $c > 0$, it holds that,

$$\sqrt{n}(\widehat{\tau}_j^{\rm STE} - \tau_j^{\rm STE}) \xrightarrow{\rm d} \mathcal{N}(0,\sigma_j^2), \qquad j = 1,\dots,p.$$

Compared to Definition 1, Corollary 6 requires a stronger condition on the rate of the bias δ_m, which is mild. For instance, when $\boldsymbol{X}_1(a),\dots,\boldsymbol{X}_m(a)$ are iid conditional on $(\boldsymbol{W},\boldsymbol{S}(a))$, the bias is zero, that is, $\mathbb{E}[\Delta_{mkj}(a)] \equiv 0$; when they are weakly dependent, for example, Qiu, Sun, and Zhou (2023, Proposition S1) show that the bias is of order $o(n^{-1/2})$ under the β-mixing condition when $n^{1/2}\log p = o(m)$.

Because the influence function of STE (4.2) is a complicated function of all the nuisances and the observations in $\mathcal{D}$, it is hard to show usual sample variance of the estimated influence function $\widehat{\varphi}_j^{\rm STE} = \widetilde{\varphi}_j^{\rm STE}(\boldsymbol{Z};\{\widehat{\pi}_a,\widehat{\boldsymbol{\mu}}_a\}_{a\in\{0,1\}})$ provides a consistent estimate. In the following proposition, we thus rely on extra independent observations to estimate the asymptotic variance.

However, one can employ the cross-fitting procedure (Chernozhukov et al. 2018) on $\mathcal{D}$ to decouple the dependency of $\widehat{\varphi}_j^{\mathrm{STE}}$ and the observations used to compute the empirical variance. This ensures that the variance estimation errors are of polynomial rates of n uniformly in p outcomes when $\log(p)/n \le Cn^{-c}$.

Proposition 7 (Consistent variance estimates). Under the same conditions in Theorem 5, let $\widehat{\varphi}_j^{\mathrm{STE}}$ be the estimated influence function (4.3) with $(\mathbb{E}[\widetilde{Y}_j(0)], \mathbb{E}[\widetilde{Y}_j(0)^2], \widetilde{\tau}_j^{\mathrm{STE}})$ estimated by the doubly robust estimators on $\mathcal{D}$, and $\mathbb{P}_n'$ be the empirical measure over a separate independent sample $\mathcal{D}' = \{\boldsymbol{Z}_{n+1}, \dots, \boldsymbol{Z}_{2n}\}$. Define the sample variance on $\mathcal{D}'$ as $\widehat{\sigma}_j^2 = \mathbb{V}_n'(\widehat{\varphi}_j^{\mathrm{STE}})$. It holds that $\max_{j\in[p]} |\widehat{\sigma}_j^2 - \sigma_j^2| = \mathcal{O}_{\mathbb{P}}(r_\sigma^{\mathrm{STE}})$ where $r_\sigma^{\mathrm{STE}} = \sqrt{\log p/n} + \vartheta^{\mathrm{STE}}$.

4.2. Quantile Effects

In practice, examining quantile effects offers a robust alternative to mean-based analysis, particularly when confronted with highly variable treatment assignment probabilities and heavy-tailed outcomes. Estimating causal effects on the mean is a challenging problem in such scenarios because the signal-noise ratio is generally small. In cases where the mean is undefined but the median exists (such as the Cauchy distribution), using the median may result in more powerful tests for the location-shift hypothesis (Díaz 2017).

We first introduce the DR estimator for the median effect (3.5) when $\varrho = 0.5$, while the proposal naturally extends to other quantile levels ϱ as well. For $j \in [p]$, let θ_{aj} be the ϱ-quantile of the counterfactual response $Y_j(a)$, which solves the following equation:

$$0 = \mathbb{E}[\psi(Y_j(a), \theta)], \text{ where } \psi(y, \theta) := \mathbb{1}\{y \le \theta\} - \varrho. \tag{4.4}$$

Since the potential outcome $Y_j(a)$ is not directly observed, we need to rely on the counterfactual derived outcomes $\widetilde{Y}_j(a)$ to identify the quantile of $Y_j(a)$. The following lemma summarizes the identification results of general M-estimators for functionals of $Y_j(a)$ using the derived outcomes $\widetilde{Y}_j(a)$.

Lemma 8 (Identification of M-estimators). Under Assumptions 1–3, consider the causal estimand θ_{aj} as the solution to the estimation equations:

$$M_j(\theta) = \mathbb{E}[F_j(Y_j(a), \theta)] = 0, \qquad j = 1, \dots, p,$$

where M_j is differentiable and the magnitude of its derivative $|M_j'|$ is uniformly lower bounded around θ_{aj}: $\min_{1\le j\le p} \inf_{\theta\in\mathcal{B}(\theta_{aj},\delta)} |M_j'(\theta)| \ge c > 0$ for some constant $\delta > 0$. Suppose $\widetilde{Y}_j(a)$'s are derived outcomes such that $F_j(\widetilde{Y}_j(a), \theta)$ is asymptotically unbiased to $F_j(Y_j(a), \theta)$, that is as $m \to \infty$, $\Delta_{mj}(a, \theta) = \mathbb{E}[F_j(\widetilde{Y}_j(a), \theta) \mid \boldsymbol{S}(a), \boldsymbol{W}] - F_j(Y_j(a), \theta)$ satisfies that $\delta_m := \max_{j\in[p]} \sup_{\theta\in\mathcal{B}(\theta_{aj},\delta)} |\mathbb{E}[\Delta_{mj}(a, \theta)]| = o(1)$. Let $\widetilde{\theta}_{aj} \in \mathcal{B}(\theta_{aj}, \delta)$ be the solution to the estimating equation

$$\mathbb{E}[\mathbb{E}[F_j(\widetilde{Y}_j, \theta) \mid A = a, \boldsymbol{W}]] = 0,$$

then θ_{aj} can be identified by $\widetilde{\theta}_{aj}$ as $m \to \infty$.

Under conditions in Lemma 8 with $F_j(Y_j(a), \theta) = \psi(Y_j(a), \theta)$, we can focus on estimating the quantile of $\widetilde{Y}_j(a)$ to approximate the quantile of $Y_j(a)$. Specifically, consider a doubly robust expansion of the above question: $0 = \mathbb{E}[\psi(\widetilde{Y}_j, \theta)] = -\mathbb{E}[\omega_{aj}(\boldsymbol{Z}, \theta)]$, where the estimating function is given by

$$\omega_{aj}(\boldsymbol{Z}, \theta) = \frac{\mathbb{1}\{A = a\}}{\pi_a(\boldsymbol{W})}(\nu_{aj}(\boldsymbol{W}, \theta) - \psi(\widetilde{Y}_j, \theta)) - \nu_{aj}(\boldsymbol{W}, \theta).$$

Here, $\nu_{aj}(\boldsymbol{W}, \theta) = \mathbb{E}[\psi(\widetilde{Y}_j, \theta) \mid \boldsymbol{W}, A = a] = \mathbb{P}(\widetilde{Y}_j \le \theta \mid \boldsymbol{W}, A = a) - \varrho$ is the excess (conditional) cumulative distribution functions (cdfs) of $\widetilde{Y}_j(a)$, and π_a is the propensity score function as before. One may then expect to obtain an estimator of θ_{aj} by solving the empirical version of (4.4) for θ:

$$0 = \mathbb{P}_n[\widehat{\omega}_{aj}(\boldsymbol{Z}, \theta)], \tag{4.5}$$

where $\widehat{\omega}_{aj}(\boldsymbol{Z}, \theta) = \frac{\mathbb{1}\{A=a\}}{\widehat{\pi}_a(\boldsymbol{W})}(\widehat{\nu}_{aj}(\boldsymbol{W}, \theta) - \psi(\widetilde{Y}_j, \theta)) - \widehat{\nu}_{aj}(\boldsymbol{W}, \theta)$ is the estimated influence function and $(\widehat{\pi}_a, \widehat{\nu}_{aj} + \varrho)$ are the estimated propensity score and cdf functions with range $[0, 1]$.

However, directly solving (4.5) is not straightforward due to its non-smoothness and non-linearity in θ. A reasonable strategy to adopt instead is a one-step update approach (van der Vaart 2000; Tsiatis 2006) using the influence function:

$$\widehat{\theta}_{aj} = \widehat{\theta}_{aj}^{\mathrm{init}} + \frac{1}{\widehat{f}_{aj}(\widehat{\theta}_{aj}^{\mathrm{init}})}\mathbb{P}_n[\widehat{\omega}_{aj}(\boldsymbol{Z}, \widehat{\theta}_{aj}^{\mathrm{init}})], \tag{4.6}$$

where $\widehat{\theta}_{aj}^{\mathrm{init}}$ is an initial estimator of θ_{aj} and $\widehat{f}_{aj}(\widehat{\theta}_{aj}^{\mathrm{init}})$ is the estimated density of $\widetilde{Y}_j(a)$ at $\widehat{\theta}_{aj}^{\mathrm{init}}$.

For $a = 0, 1$, let $\boldsymbol{f}_a = (f_{aj})_{j\in[p]}$, $\widetilde{\boldsymbol{\theta}}_a = (\widetilde{\theta}_{aj})_{j\in[p]}$, and $\boldsymbol{\nu}_a = (\nu_{aj})_{j\in[p]}$ be the vectors of true density functions, the ϱ-quantiles, and the excess cdf functions cdfs of $\widetilde{Y}_j(a)$, respectively. Moreover, let $\widehat{\boldsymbol{f}}_a = (\widehat{f}_{aj})_{j\in[p]}$, $\widehat{\boldsymbol{\theta}}_a = (\widehat{\theta}_{aj})_{j\in[p]}$, and $\widehat{\boldsymbol{\nu}}_a = (\widehat{\nu}_{aj})_{j\in[p]}$ be the corresponding vectors of estimated nuisances. Based on (4.6), an estimator for τ_j^{QTE} is given by

$$\widehat{\tau}_j^{\mathrm{QTE}} = \widehat{\theta}_{1j} - \widehat{\theta}_{0j}. \tag{4.7}$$

The following theorem provides the asymptotic normality of the one-step estimator (4.7).

Theorem 9 (Linear expansion of QTE). Under Assumptions 1–3, suppose the identification conditions in Lemma 8 hold with $F_j = \psi$ defined in (4.4). Consider the one-step estimator (4.7) for the median treatment effect, where $\mathbb{P}_n$ is the empirical measure over $\mathcal{D} = \{\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n\}$ and $(\widehat{\boldsymbol{\theta}}_a^{\mathrm{init}}, \widehat{\boldsymbol{f}}_a, \widehat{\pi}_a, \widehat{\boldsymbol{\nu}}_a)$ is an estimate of $(\widetilde{\boldsymbol{\theta}}_a, \boldsymbol{f}_a, \pi_a, \boldsymbol{\nu}_a)$ from samples independent of $\mathcal{D}$ for $a = 0, 1$. Suppose the following conditions hold for $a = 0, 1$ with probability tending to one:

1. Boundedness: The quantile θ_{aj} is in the interior of its parameter space. There exists $C, c > 0$ and $\epsilon, \delta \in (0, 1)$ such that $\max_{1\le j\le p} \max\{|Y_j|, |\widetilde{Y}_j|\} < C$, and $\pi_a, \widehat{\pi}_a \in [\epsilon, 1 - \epsilon]$, and f_{aj} is uniformly bounded : $c \le f_{aj} \le C$ for all j and has a bounded derivative in a neighborhood $\mathcal{B}(\widetilde{\theta}_{aj}, \delta)$ for all $j \in [p]$: $\max_{1\le j\le p} \max_{\theta\in\mathcal{B}(\widetilde{\theta}_{aj},\delta)} |f_{aj}'(\theta)| \le C$.
2. Initial estimation: The initial quantile and density estimators satisfy that $\max_{j\in[p]} |\widehat{\theta}_{aj}^{\mathrm{init}} - \widetilde{\theta}_{aj}| = \mathcal{O}(n^{-\gamma})$ and $\max_{j\in[p]} |\widehat{f}_{aj}(\widehat{\theta}_{aj}^{\mathrm{init}}) - f_{aj}(\widetilde{\theta}_{aj})| = \mathcal{O}(n^{-\kappa})$ with $\gamma > 1/4, \kappa > 0$ such that $\gamma + \kappa > 1/2$.

3. Nuisance: The rates of nuisance estimates satisfy $\max_{j\in[p]} \sup_{\theta\in\mathcal{B}(\widetilde{\theta}_{aj},\delta)} \|\widehat{\nu}_{aj}(\cdot,\theta) - \nu_{aj}(\cdot,\theta)\|_{L_2} = \mathcal{O}(n^{-\alpha})$ and $\|\widehat{\pi}_a - \pi_a\|_{L_2} = \mathcal{O}(n^{-\beta})$ for some $\alpha, \beta \in (0, 1/2)$ such that $\alpha + \beta > 1/2$.

Then as $m, n, p \to \infty$, it holds that $\widehat{\tau}_j^{\text{QTE}} - \tau_j^{\text{QTE}} = \mathbb{P}_n\{\widetilde{\varphi}_j^{\text{QTE}}\} + \varepsilon_j, j = 1, \dots, p$, where the residual term satisfy $\max_{j\in[p]} |\varepsilon_j| = \mathcal{O}_{\mathbb{P}}(\vartheta^{\text{QTE}}\sqrt{(\log p)/n} + (\log p)/n + n^{-(\alpha+\beta)\wedge(\gamma+\kappa)\wedge(2\gamma)} + \delta_m)$ with $\vartheta^{\text{QTE}} := n^{-(\alpha\wedge\beta\wedge\kappa\wedge\frac{\gamma}{2})}$ and the influence function is given by

$$\widetilde{\varphi}_j^{\text{QTE}}(\mathbf{Z}; \{\widetilde{\boldsymbol{\theta}}_a, \boldsymbol{f}_a, \pi_a, \boldsymbol{\nu}_a\}_{a\in\{0,1\}}) \\ = [f_{1j}(\widetilde{\theta}_{1j})]^{-1}\omega_{1j}(\mathbf{Z}, \widetilde{\theta}_{1j}) - [f_{0j}(\widetilde{\theta}_{0j})]^{-1}\omega_{0j}(\mathbf{Z}, \widetilde{\theta}_{0j}).$$

(Section F.1) provides details for obtaining initial estimators for the quantiles and the corresponding densities. Similar to STE, we can also obtain individual asymptotic normality for the DR estimator (4.7) of QTE and consistently estimate its variance.

Proposition 10 (Asymptotic normality of QTE). Under the conditions in Theorem 9, when $(\vartheta^{\text{QTE}} \vee n^{-1/4})\sqrt{\log p} = o(1)$, $\delta_m = o(n^{-1/2})$ and $\sigma_j^2 := \mathbb{V}(\widetilde{\varphi}_j^{\text{QTE}}) \geq c$ for some constant $c > 0$, it holds

$$\sqrt{n}(\widehat{\tau}_j^{\text{QTE}} - \tau_j^{\text{QTE}}) \xrightarrow{\text{d}} \mathcal{N}(0, \sigma_j^2), \qquad j = 1, \dots, p.$$

Define the sample variance $\widehat{\sigma}_j^2 = \mathbb{V}_n(\widehat{\varphi}_j^{\text{QTE}})$ for the estimated influence function $\widehat{\varphi}_j^{\text{QTE}} := \widetilde{\varphi}_j^{\text{QTE}}(\mathbf{Z}; \{\widehat{\boldsymbol{\theta}}_a^{\text{init}}, \widehat{\boldsymbol{f}}_a, \widehat{\pi}_a, \widehat{\boldsymbol{\nu}}_a\}_{a\in\{0,1\}})$. It further holds that $\max_{j\in[p]} |\widehat{\sigma}_j^2 - \sigma_j^2| = \mathcal{O}_{\mathbb{P}}(r_\sigma^{\text{QTE}})$ where $r_\sigma^{\text{QTE}} = (\log p)/n + \sqrt{(\log p)/n}\vartheta^{\text{QTE}} + \vartheta^{\text{QTE}}$.

Apart from the mild rate requirement on the nuisance functions, no metric entropy conditions are assumed in Theorem 9 and Corollary 10. This allows one to estimate nuisances with machine learning methods and achieve asymptotical normality with cross-fitting. While the doubly-robust estimators for QTE have also been considered by Chakrabortty, Dai, and Tchetgen (2022); Kallus, Mao, and Uehara (2024) for a single outcome ($p = 1$), they both require metric entropy or Donsker class conditions.

5. Simultaneous Inference

5.1. Large-Scale Multiple Testing

For a target estimand $\tau_j \in \{\tau_j^{\text{STE}}, \tau_j^{\text{QTE}}\}$, the asymptotic normality established in Corollary 6 and Proposition 10 can be used to test the null hypotheses $H_{0j} : \tau_j = \tau_j^*$ for $j = 1, \dots, p$. This implies that one can control the Type-I error of the individual tests using the statistics $t_j = \sqrt{n}(\widehat{\tau}_j - \tau_j^*)/\widehat{\sigma}_j$, with empirical variance given in Propositions 7 and 10. The confidence intervals for individual causal estimates can also be constructed. To conduct simultaneous inference, however, the tests above are too optimistic when multiple tests are of interest. Therefore, to obtain valid inferential statements, we must perform a multiplicity adjustment to control the family-wise error explicitly. This subsection provides simultaneous tests and confidence intervals for causal effects with multiple outcomes.

For $j \in [p]$, let $\varphi_{ij} = \widetilde{\varphi}_j(\mathbf{Z}_i)$ and $\widehat{\varphi}_{ij} = \widehat{\varphi}_j(\mathbf{Z}_i)$ be the influence function value and its estimate evaluated at the ith observation $\mathbf{Z}_i = (A_i, \mathbf{W}_i, \mathbf{X}_i)$, as defined in Propositions 7 and 10 for τ_j being τ_j^{STE} and τ_j^{QTE}, respectively. We require a condition from Chernozhukov, Chetverikov, and Kato (2013, Theorem J.1) for feasible inference.

Assumption 4 (Bounded variances and covariances). There exist a constant $a, c_1 \in (0, 1)$ and a set of informative hypotheses $\mathcal{A}^* \subseteq [p]$ such that $|\mathcal{A}^*| \geq ap$, $\max_{j\in\mathcal{A}^{*c}} \sigma_j^2 = o(1)$, $\min_{j\in\mathcal{A}^*} \sigma_j^2 \geq c_1$ and $\max_{j_1\neq j_2\in\mathcal{A}^*} |\text{corr}(\varphi_{1j_1}, \varphi_{1j_2})| \leq 1 - c_1$.

When the value of σ_j is 0, the population distribution of the jth influence function is degenerated and has no variability. In Assumption 4, the first condition precludes the existence of such super-efficient estimators over $\mathcal{A}^*$, which is commonly required even in classical settings where the number of variables p is small compared to the sample size n (Belloni et al. 2018). In practice, one can use a small threshold c_n to screen out outcomes that have small variations and obtain a set of informative outcomes $\mathcal{A}_1 = \{j \in [p] \mid \widehat{\sigma}_j \geq c_n\}$.

For DR estimators derived in the previous section, the following Gaussian approximation result over a family of null hypotheses allows for data-dependent choices of the set of hypotheses and suggests a multiplier bootstrap procedure (Chernozhukov, Chetverikov, and Kato 2013) for simultaneous inference.

Lemma 11 (Gaussian approximation for nested hypotheses). For $\tau_j = \tau_j^{\text{STE}}$ and $\vartheta = \vartheta^{\text{STE}}$, suppose conditions in Proposition 7 and Assumption 4 hold. Futher assume that there exist some constants $c_2, C_2 > 0$ such that $\max\{\log(pn)^7/n, \log(pn)^2\vartheta, \sqrt{n\log(pn)\delta_m}\} \leq C_2 n^{-c_2}$. For all $\mathcal{S} \subseteq \mathcal{A}^* \subseteq [p]$, define $M_{\mathcal{S}} = \max_{j\in\mathcal{S}} |\sqrt{n}(\widehat{\tau}_j - \tau_j)/\widehat{\sigma}_j|$, $\widehat{\boldsymbol{\varphi}}_i = (\widehat{\varphi}_{ij})_{j\in\mathcal{S}}$, $\widehat{\boldsymbol{E}}_{\mathcal{S}} = n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{\varphi}}_i\widehat{\boldsymbol{\varphi}}_i^\top$, and $\widehat{\boldsymbol{D}}_{\mathcal{S}} = \text{diag}((\widehat{\sigma}_j)_{j\in\mathcal{S}})$. Consider null hypotheses $H_0^{\mathcal{S}}$ indexed by $\mathcal{S}$ that $\forall\, j \in \mathcal{S}, \tau_j = \tau_j^*$. As $m, n, p \to \infty$, it holds that

$$\sup_{H_0^{\mathcal{S}}:\mathcal{S}\subseteq\mathcal{A}^*} \sup_{x\in\mathbb{R}} |\mathbb{P}(\overline{M}_{\mathcal{S}} > x) - \mathbb{P}(\|\boldsymbol{g}_{\mathcal{S}}\|_\infty > x \mid \{\mathbf{Z}_i\}_{i=1}^n)| \xrightarrow{\text{p}} 0,$$

where $\boldsymbol{g}_{\mathcal{S}} \sim \mathcal{N}(\mathbf{0}, \widehat{\boldsymbol{D}}_{\mathcal{S}}^{-1}\widehat{\boldsymbol{E}}_{\mathcal{S}}\widehat{\boldsymbol{D}}_{\mathcal{S}}^{-1})$. The conclusion also holds for $\tau_j = \tau_j^{\text{QTE}}$ and $\vartheta = \vartheta^{\text{QTE}}$ under conditions in Proposition 10 and Assumption 4.

When m is sufficiently large such that the error of derived outcomes δ_m is ignorable, the rate conditions in Lemma 11 can be satisfied if the *logarithm* of the numbers of hypotheses grows slower than $n^{\frac{1}{7}} \wedge \vartheta^{-\frac{1}{2}}$ for at least a polynomial factor of n. Lemma 11 suggests that if $\mathcal{A}_1 \subseteq \mathcal{A}^*$ only contains informative hypotheses, then the distribution of the maximal statistic $M_1 = \max_{j\in\mathcal{A}_1} |t_j|$ can be well approximated by $\boldsymbol{g}_1 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}_{n1}^{-1}\boldsymbol{E}_{n1}\boldsymbol{D}_{n1}^{-1})$, where $\boldsymbol{E}_{n1} = n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{\varphi}}_{i1}\widehat{\boldsymbol{\varphi}}_{i1}^\top$ is the sample covariance matrix with $\widehat{\boldsymbol{\varphi}}_{i1} = (\widehat{\varphi}_{ij})_{j\in\mathcal{A}_1}$ and $\boldsymbol{D}_{n1} = \text{diag}((\widehat{\sigma}_j)_{j\in\mathcal{A}_1})$ is the diagonal matrix of the estimated standard deviations. This allows us to simulate the null distribution efficiently using the multiplier bootstrap procedure. To generate B bootstrap samples, for all $b = 1, \dots, B$, we first sample n standard normal variables $\varepsilon_{11}^{(b)}, \dots, \varepsilon_{n1}^{(b)} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$ and then apply a linear transformation to obtain the

Algorithm 1 Multiple testing on doubly robust estimation of treatment effects

```
Input: The estimated centered influence function values $\widehat{\varphi}_{ij}$, the estimated variance $\widehat{\sigma}_j^2$ for $i = 1, \dots, n$ and $j = 1, \dots, p$. The FDP exceedance threshold $c$ and probability $\alpha$, and the number of bootstrap samples $B$.
Initialize the iteration number $\ell = 1$, the candidate set $\mathcal{A}_1 = \{j \in [p] \mid \widehat{\sigma}_j^2 \geq c_n\}$, the set of discoveries $\mathcal{V}_1 = \varnothing$, and the statistic $t_j = \sqrt{n}(\widehat{\tau}_j - \tau_j^*)/\widehat{\sigma}_j$ for $j \in [p]$.
while not converge do
    Draw multiplier bootstrap samples $\boldsymbol{g}_\ell^{(b)} = (\sqrt{n}\boldsymbol{D}_{n\ell})^{-1} \sum_{i=1}^n \varepsilon_{i\ell}^{(b)} \widehat{\boldsymbol{\varphi}}_{i\ell}$, where $\varepsilon_{i\ell}^{(b)}$'s are independent samples from $\mathcal{N}(0, 1)$ for $i = 1, \dots, n$ and $b = 1, \dots, B$.
    Compute the maximal statistic $M_\ell = \max_{j \in \mathcal{A}_\ell} |t_j|$.
    Estimate the upper $\alpha$-quantile of $M_\ell$ under $H_0^{(\ell)} : \forall\, j \in \mathcal{A}_\ell,\ \tau_j = \tau_j^*$ by
        $$\widehat{q}_\ell(\alpha) = \inf\left\{x \,\middle|\, \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\{\|\boldsymbol{g}_\ell^{(b)}\|_\infty \leq x\} \geq 1 - \alpha\right\}.$$
    if $M_\ell > \widehat{q}_\ell(\alpha)$ then
        Set $j_\ell = \operatorname{argmax}_{j \in \mathcal{A}_\ell} |t_j|$, $\mathcal{A}_{\ell+1} = \mathcal{A}_\ell \setminus \{j_\ell\}$, and $\mathcal{V}_{\ell+1} = \mathcal{V}_\ell \cup \{j_\ell\}$.
    else
        Declare the treatment effects in $\mathcal{A}_\ell$ are not significant and stop the step-down process.
    end if
    $\ell \leftarrow \ell + 1$.
end while
Augmentation: Set $\mathcal{V}$ to be the union of $\mathcal{V}_\ell$ and the $\lfloor |\mathcal{V}_\ell| \cdot c/(1-c) \rfloor$ elements from $\mathcal{A}_\ell$ with largest magnitudes of $|t_j|$.
Output: The set of discoveries $\mathcal{V}$.
```

multivariate normal vectors $\boldsymbol{g}_1^{(b)} = (\sqrt{n}\boldsymbol{D}_{n1})^{-1} \sum_{i=1}^n \varepsilon_{i1}^{(b)} \widehat{\boldsymbol{\varphi}}_{i1}$. It is easy to verify that $\boldsymbol{g}_1^{(1)}, \dots, \boldsymbol{g}_1^{(B)} \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}_{n1}^{-1}\boldsymbol{E}_{n1}\boldsymbol{D}_{n1}^{-1})$ conditioned on $\{\boldsymbol{Z}_i\}_{i=1}^n$. Based on the bootstrap samples, we can estimate the upper α quantile of M_1 by $\widehat{q}_1(\alpha) = \inf\left\{x \,\middle|\, B^{-1}\sum_{b=1}^B \mathbb{1}\{\|\boldsymbol{g}_1^{(b)}\|_\infty \leq x\} \geq 1 - \alpha\right\}$. To test multiple hypotheses $H_{0j} : \tau_j = \tau_j^*$ for $j \in \mathcal{A}_1$, we reject those are in the set $\widehat{\mathcal{A}} = \{j \in \mathcal{A}_1 \mid |t_j| > \widehat{q}_1(\alpha)\}$. The next proposition shows that the informative hypotheses can be identified, and the family-wise error rate (FWER) can be controlled.

Proposition 12 (Type-I error control). For (τ_j, r_σ) being $(\tau_j^{\text{STE}}, r_\sigma^{\text{STE}})$ or $(\tau_j^{\text{QTE}}, r_\sigma^{\text{QTE}})$, suppose conditions in Lemma 11 hold . Let $\mathcal{V}^* = \{j \mid H_{0j} \text{ is false}, j = 1, \dots, p\} \cap \mathcal{A}^*$ be the set of informative non-null hypotheses. If $\max\{r_\sigma, \max_{j \in \mathcal{A}^{*c}} \sigma_j^2\} = o(c_n)$, then as $m, n, p \to \infty$, it holds that $\lim \mathbb{P}(\mathcal{A}^* = \mathcal{A}_1) = 1$ and $\limsup \mathbb{P}(\widehat{\mathcal{A}} \cap \mathcal{V}^{*c} \neq \varnothing) \leq \alpha$.

As suggested by Proposition 12, because the lower bound of informative variance in Assumption 4 is unknown, a slowly shrinking threshold c_n is needed to recover the true candidate set $\mathcal{A}^*$ and control the FWER. In practice, one can set c_n as a small value, such as 0.01, to exclude uninformative tests. If lowly expressed genes have already been excluded, thresholding may not be necessary.

5.2. False Discovery Rate Control

When p is large, controlling for the false discovery proportion (FDP) or the false discovery rate (FDR) is more desirable to improve the powers when performing simultaneous testing. The FDP is the ratio of false positives to total discoveries, while the FDR is the expected value of the FDP. To control the FDP, we adopt the step-down procedure (Genovese and Wasserman 2006) to test the sequential hypotheses,

$$\begin{aligned} &H_0^{(\ell)} : \forall\, j \in \mathcal{A}_\ell,\ \tau_j = \tau_j^*, && \text{versus} \\ &H_a^{(\ell)} : \exists\, j \in \mathcal{A}_\ell,\ \tau_j \neq \tau_j^*, && \ell = 1, 2, \dots \end{aligned}$$

where $\mathcal{A}_1, \mathcal{A}_2, \dots$ is a sequence of nested sets. The proposed multiple testing method in Algorithm 1 incorporates both the Gaussian multiplier bootstrap and step-down procedure, which aims to control the FDP exceedance rate $\text{FDX} := \mathbb{P}(\text{FDP} > c)$, the probability that FDP surpasses a given threshold c at a confidence level α. This provides a strengthened control on FDP and is asymptotically powerful, as shown in the following theorem.

Theorem 13 (Multiple testing). Under the conditions of Proposition 12, consider testing multiple hypotheses $H_{0j} : \tau_j = 0$ versus $H_{aj} : \tau_j \neq 0$ for $j = 1, \dots, p$ based on the step-down procedure with augmentation. As $m, n, p \to \infty$, the set of discoveries $\mathcal{V}$ returned by Algorithm 1 satisfies that

- (FDX) $\limsup \mathbb{P}(\text{FDP} > c) \leq \alpha$ where $\text{FDP} = |\mathcal{V} \cap \mathcal{V}^{*c}|/|\mathcal{V}|$.
- (Power) $\mathbb{P}(\mathcal{V}^* \subset \mathcal{V}) \to 1$ if $\min_{j \in \{j \in [p] \mid \tau_j \neq 0\}} |\tau_j| \geq c'\sqrt{\log(p)/n}$ for some constant $c' > 0$.

Theorem 13 extend previous results by Belloni et al. (2018) for many approximate means and by Qiu, Sun, and Zhou (2023) for IPW estimators to DR estimators. On the one hand, Belloni et al. (2018) directly imposes assumptions on the influence functions and linearization errors, while we need to analyze the effect of nuisance functions estimation for the doubly robust

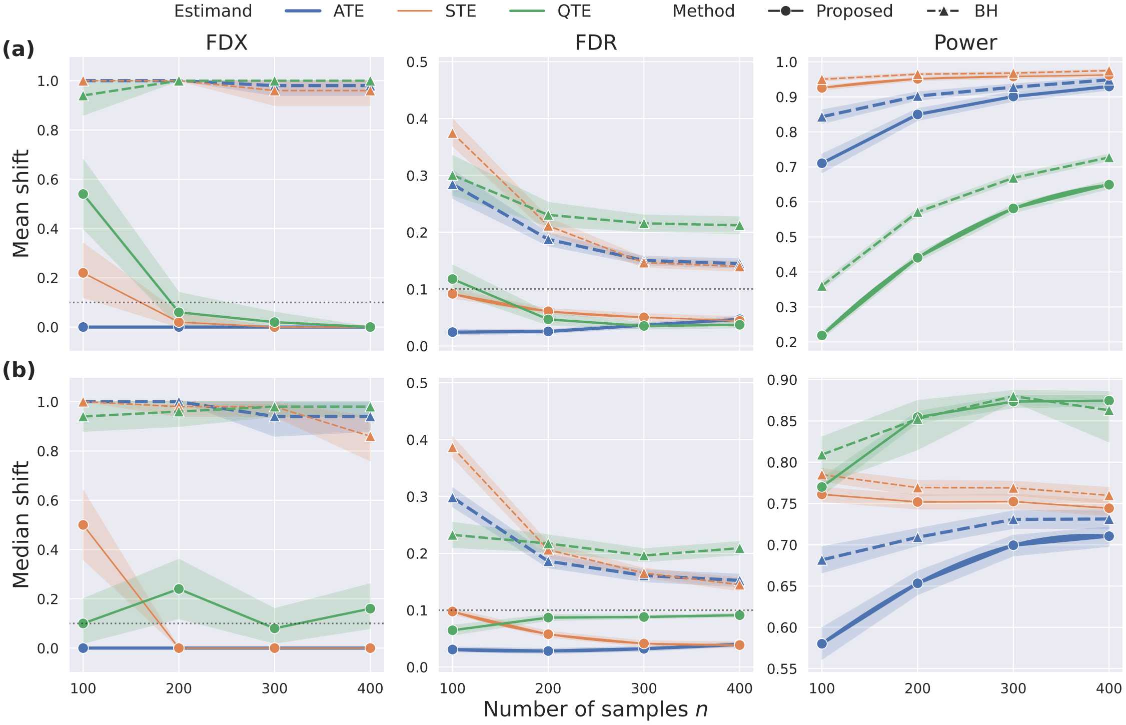

Figure 2. Simulation results of the hypothesis testing of $p = 8000$ outcomes based on different causal estimands and FDP control methods for detecting differential signals under (a) mean shifts and (b) median shifts averaged over 50 randomly simulated datasets without sample splitting. The gray dotted lines denote the nominal level of 0.1.

estimators. On the other hand, Qiu, Sun, and Zhou (2023) requires sub-Gaussian assumptions and $\sqrt{n}$-consistency of the maximum likelihood estimation for the propensity score to establish Gaussian approximation for their proposed statistics, which does not apply to our problem setups.

6. Simulation

We consider a simulation setting with $p = 8000$ genes and generate an active set of genes $\mathcal{V}^* = \mathcal{A}^* \subset [p]$ with size 200. We draw covariates $\boldsymbol{W} \in \mathbb{R}^d$ with iid $\mathcal{N}(0,1)$ entries and the treatment A follows a logistic regression model with probability $\mathbb{P}(A = 1 \mid \boldsymbol{W}) = 1/(1 + \exp(\mathbf{1}_d^\top \boldsymbol{W}/(d+1)))$. Then, we generate the counterfactual gene expressions. For a gene j, the single-cell gene expression $X_j(0)$ is drawn from a Poisson distribution with mean $\lambda_j = \exp(\boldsymbol{W}^\top \boldsymbol{b}_j) \in \mathbb{R}$ where the entries of both the coefficients $\boldsymbol{b}_j \in \mathbb{R}^d$ with 1 as the first entry and the remaining entries independently drawn from $\mathcal{N}(0, 1/4)$. The gene expressions $X_j(1)$ for $j \notin \mathcal{V}^*$ are generated from the same model, while for gene $j \in \mathcal{V}^*$, we consider two treatment mechanisms that favor the mean-based and quantile-based tests, respectively; see (Section F.2.1) for more details about the data generating processes.

Next, we draw m observations $\boldsymbol{X}_1(A), \dots, \boldsymbol{X}_m(A)$ independently, which are summed up as the overall gene expression $\widetilde{\boldsymbol{Y}}(A)$. Then, the observed gene expression matrix is given by $\mathbf{X} = A\mathbf{X}(1) + (1-A)\mathbf{X}(0)$ and analogously $\widetilde{\boldsymbol{Y}} = A\widetilde{\boldsymbol{Y}}(1) + (1-A)\widetilde{\boldsymbol{Y}}(0)$. We then draw n independent observed samples $\{(A_i, \boldsymbol{W}_i, \mathbf{X}_i, \widetilde{\boldsymbol{Y}}_i)\}_{i=1}^n$. The parameters are set to be $d = 5$, $m = 100$, $n \in \{100, 200, 300, 400\}$. For nuisance function estimation, we employ Logistic regression to estimate the propensity score and Poisson generalized linear model (GLM) with the log link to estimate the mean regression functions. For quantile-based methods, the initial estimators of the quantile and density are described in (Section F.1).

To quantify the performance of different test statistics and multiple testing procedures, we compare the empirical FDX, FDR, and power. We aimed to control FDX over 0.1 at 0.05, namely $\mathbb{P}(\mathrm{FDP} > 0.1) \leq 0.05$. We also compared with the Benjamini-Hochberg (BH) procedure with targeting FDR controlled at 0.05. The experiment results are summarized in Figure 2 without sample splitting and in (Figure F2) with cross-fitting. As shown in Figure 2, in the high signal-noise ratio (SNR) setting, the proposed method controls both FDX and FDR at the desired level for all three causal estimands when the sample size is relatively large, that is, $n > 100$. On the other hand, the BH procedure fails to control the FDX and FDP at any sample sizes because the p-values are not close to uniform distribution (see (Figure F1)), though the gaps of FDP become smaller when the sample size gets larger. Though the BH procedure has lower FDP with sample splitting (see (Figure F2)), it still fails to control FDX for all estimands. This indicates that the proposed multiple testing procedure consistently outperforms the BH procedure by correctly accounting for the dependencies among the test statistics and providing valid statistical error control.

In the low SNR setting, we see that the quantile-based estimand has larger powers than mean-based tests, which is expected because of the designed data-generating process. Such a low SNR scenario is often encountered with scRNA-seq data. In this case, the proposed method still has better control of both FDX and FDR compared to the BH procedure. Although the QTE test is slightly anti-conservative regarding FDX, it still controls the FDR well. Furthermore, standardized tests are more powerful than unstandardized estimands, especially when the sample size is small. Overall, the results in Figure 2 demonstrate the valid FDP control of the proposed multiple testing procedure for various causal estimands and suggest that testing based on different estimands could be helpful in different scenarios. In contrast, the commonly used BH procedure in genomics may be substantially biased due to the complex dependency among tests.

7. Real Data

7.1. LUHMES Data with CRISPR Repression

In this subsection, we aim to validate the proposed causal inference procedure with multiple outcomes on the single-cell CRISPR experiments. In this case, we directly measure one observation $\boldsymbol{Y}$ instead of many observations $\mathbf{X}$ per subject, simplifying the model. As a concrete data example, Lalli et al. (2020) employed a comprehensive single-cell functional genomics approach to unravel the molecular underpinnings of genes associated with neurodevelopmental disorders. Using a modified CRISPR-Cas9 system, they conducted gene suppression experiments on 13 genes linked to Autism Spectrum Disorder (ASD). The experimental setup involved 14 groups of LUHMES neural progenitor cells, encompassing 13 treatment groups where each group had a specific single gene knockdown and one control group with no targeted gene suppression. Single-cell RNA sequencing was applied to assess the resultant gene expression changes in response to each gene knockdown. At varying stages of maturation, the cells classified as "early" or "late" resulted in 28 unique cell groups (14 groups and 2 maturation stages). A critical scientific objective was to analyze and compare the transcriptional profiles across these diverse groups of neuron cells, thereby gaining insights into the transcriptional dynamics influenced by ASD-related gene modifications.

We focus on the late-stage cells and perturbations related to 4 target genes *PTEN*, *CHD2*, *ASH1L*, and *ADNP* that are shown to be influential for neural development at this stage (Lalli et al. 2020, Figure 4E). We include six covariates: the intercept, the logarithm of the library size, two cell cycle scores ("S.Score" and "G2M.Score"), the processing batch, and the pseudotime states (categorical with values in $\{4, 5, 6\}$). We compare gene expressions of each perturbation to the control group, and filter out genes expressed in less than 50 cells, where the resulting sample sizes are shown in Table F1. We aimed to control FDX over 0.01 at 0.05, namely $\mathbb{P}(\text{FDP} > 0.01) \leq 0.05$. The stringent threshold gives a reasonable number of discoveries that we can visually compare across ATE and STE tests. We didn't evaluate quantile-based tests in the current data analysis because of the sparse counts, which makes the quantile estimates of most of the genes exactly zero.

For perturbations targeting *PTEN*, the ATE tests identify 32 genes as differentially expressed, while the STE tests identify 26 genes, and 23 genes are determined by both tests. The detailed results of all perturbations are summarized in (Table F2). Next, we inspect the covariate-adjusted residuals of the gene expressions from parts of these discoveries. As shown in Figure 3, the significant genes for STE tests show a clear separation between the control and the perturbed groups, compared to those only significant for ATE tests. However, most of the significant genes

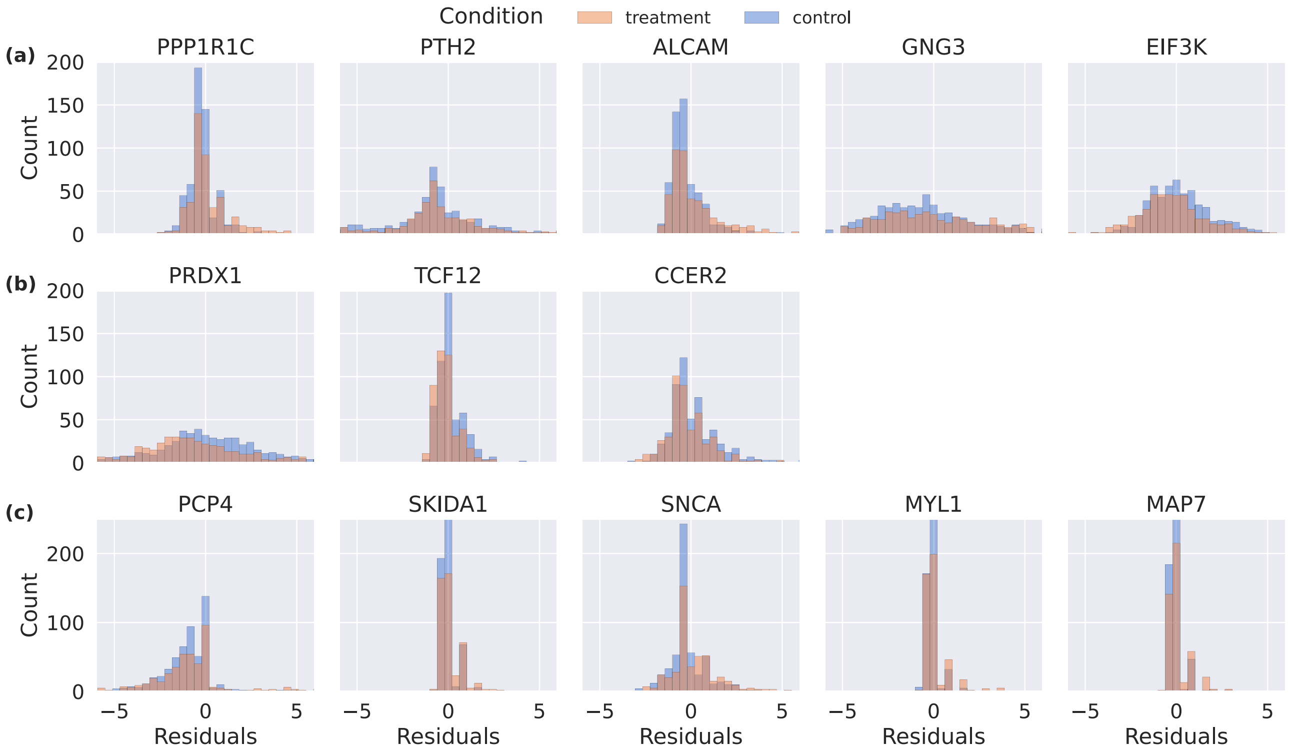

Figure 3. Histogram of covariate-adjusted residuals of gene expressions under different conditions (treatment with target gene *PTEN* knockdown and control/non-targeting) in the late-stage samples of the LUHMES dataset. The residuals are obtained by fitting NB GLMs of the gene expressions on all covariates except the condition. Each row shows several examples of genes that are (a) significant for both ATE and STE tests; (b) significant for STE tests but insignificant for ATE tests; (c) significant for ATE tests but insignificant for STE tests.

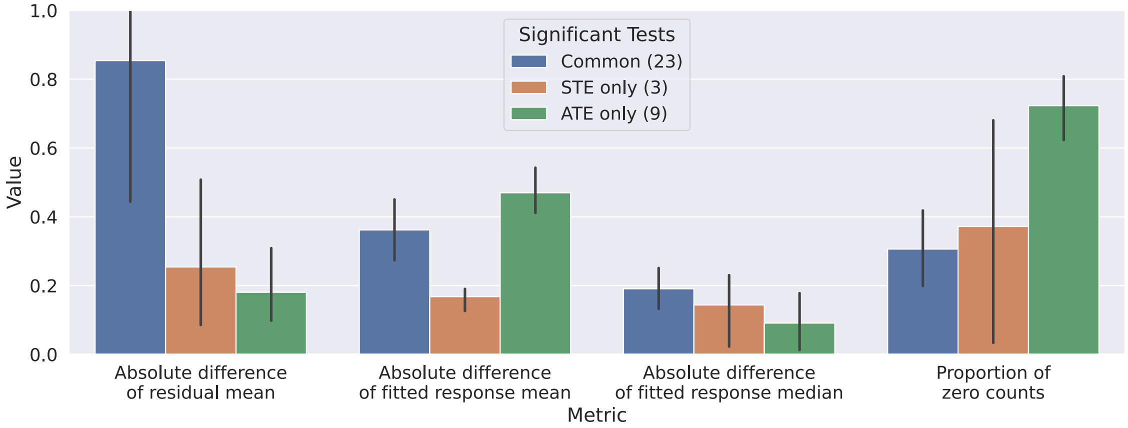

Figure 4. Summarized statistics for significant genes from ATE and STE tests under different conditions (treatment with target gene *PTEN* knockdown and control/non-targeting) in the late-stage samples of the LUHMES dataset. The first three metrics show the absolute differences (between the treatment and the control group) of the covariate-adjusted residual mean and the mean and median of the GLM-fitted response, respectively. The last metric shows the proportion of zero counts of the significant genes.

for ATE tests exhibit zero-inflated expression in Figure 3(c). The ATE tests of these genes may be biased because of the bimodal and skew distributions.

To further inspect the conjecture, we summarize the absolute differences of the covariate-adjusted residual mean and the mean/median of the GLM-fitted response, as well as the proportion of zero counts of the significant genes in Figure 4. The common discoveries have a significant difference in the residual mean, which is expected. The ATE-only discoveries have a larger difference in the fitted mean, although the difference in the median of the fitted responses is negligible. Because the proportion of zero counts for ATE-only discovered genes is also higher, it is evident that the ATE tests are biased because of the zero-inflated expressions. This suggests that STE tests provide more robust and reliable results than ATE tests in the presence of zero-inflated expressions.

7.2. Lupus Data

In this subsection, we consider a subject-level analysis of systemic lupus erythematosus (SLE) multiplexed single-cell RNA sequencing (mux-seq) dataset from Perez et al. (2022). The dataset includes 1.2 million peripheral blood mononuclear cells, spanning eight major cell types from 261 participants, of whom 162 are SLE patients, and 99 are healthy individuals of either Asian or European descent. SLE is an autoimmune condition that primarily affects women and people with Asian, African, and Hispanic backgrounds. The objective of this cell-type-specific differential expression (DE) analysis is to enhance our understanding of SLE's diagnosis and treatment. We follow the preprocessing procedure in Du, Wasserman, and Roeder (2023) to aggregate the gene expressions of cells of each cell type in each subject, which serve as the derived outcomes for each cell type of each subject. The resulting pseudo-bulk dataset contains five cell types T4, cM, B, T8, and NK, with numbers of subjects and genes (n, p) being (256,1255), (256,1208), (254,1269), (256,1281), and (256,1178), respectively. For each subject, we also measure the SLE status (condition) and $d = 6$ covariates: the logarithm of the library size, sex, population, and processing cohorts (4 levels).

Similar to Section 7.1, we aim to control the FDX rate over 0.1 at 0.05. The inference results are compared based on causal estimands. Besides ATE, STE, and QTE, we also examine the standardized quantile treatment effects (SQTE), which equals the QTE normalized by the interquartile range of counterfactuals $\boldsymbol{Y}(0)$; see (Section F.1.3) for precise definitions. For the quantile-based methods, we exclude genes that have zero medians. The discoveries of different tests in the current pseudo-bulk analysis are summarized in Figure F3. We observed that the standardized tests are generally more conservative than the unstandardized tests, as expected from our analysis in Section 7.1. Most of the former discoveries are also parts of the latter. Furthermore, the ATE tests are anti-conservative, yielding dozens of extra discoveries that are insignificant for other estimands. Overall, the mean-based tests are much more powerful than the quantile-based tests.

We visualize the summarized statistics in Figure 5 to further compare the mean-based and quantile-based tests. As expected, the quantile-based tests are more sensitive to the median differences between the two counterfactual distributions. The results suggest that different causal estimands focus on different characteristics of the distributions, and one may get more reliable inferential results by focusing on the common discoveries by different tests.

8. Discussion

This article investigates semiparametric causal inference approaches for analyzing multiple derived outcomes arising from the increasingly popular study design of subject-level scRNA-seq data analysis. The doubly robust estimators for standardized and quantile treatment effects are proposed and analyzed to overcome the challenge of heterogeneous subjects and outcomes. Building on the Gaussian approximation results for the doubly robust estimators, we propose a multiple testing procedure that provably controls the false discovery rate and is asymptotically powerful. In simulation and real data analysis with single-cell count data, we use simple parametric and nonparametric models to estimate the nuisance functions and evaluate the multiple testing procedures. Notably, the proposed

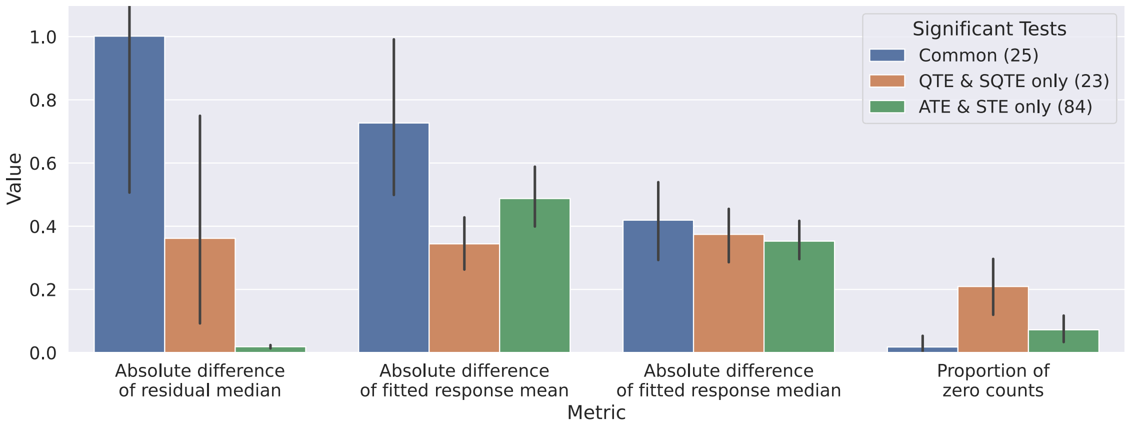

Figure 5. Summarized statistics for significant genes from mean-based tests (ATE and STE) and quantile-based tests (QTE and SQTE) under different conditions (case and control) in the T4 cell type of the lupus dataset. The first three metrics show the absolute differences (between the treatment and the control group) of the covariate-adjusted residual mean and the mean and median of the GLM-fitted response, respectively. The last metric shows the proportion of zero counts of the significant genes.

semiparametric inference framework allows one to incorporate more advanced machine learning and deep learning methods to obtain valid inference results.

The present study, while contributing valuable insights, is not devoid of limitations. First, the doubly robust estimation can lead to bias if both nuisance functions are estimated using data-adaptive methods (e.g., machine learning) and only one is consistent (Van der Laan 2014; Benkeser et al. 2017). Model misspecification is one of the main causes of the inconsistency, especially in the presence of high-dimensional conditional outcome models. The issue of misspecification itself is challenging for any statistical problem. In theory, if we have prior knowledge that the outcome models lie in a certain function class (e.g., the α-Hölder functions), then our estimator allows a wide range of nonparametric methods to model nuisance functions and hence avoids model misspecification. In practice, the estimation of the nuisance functions requires prior knowledge. For genomics data, the gene expression counts are usually modeled by practitioners using generalized linear models, provided that the low-quality data are filtered beforehand (Sarkar and Stephens 2021). For this reason, we mitigate the risk of misspecification of outcome models in the application by incorporating such prior knowledge.

Second, we only consider STE and QTE as two specific examples, which may not be appropriate in all cases. For example, for lowly expressed genes of single-cell data with zero-inflation patterns, the null distribution is not unimodal, so any test that compares the location statistics is not ideal. When more than half of the samples are zero for a specific gene, the mean/quantile regressions for computing the test statistics may be inaccurate, and this will affect the test results. To mitigate these challenges, one approach is to prescreen and exclude genes with low expression levels, thereby reducing the impact of zero inflation. Alternatively, one can also quantify treatment effects using other measures of the distribution. For instance, the fold change is the ratio of expected expressions under two conditions, whose sign indicates whether a gene is up-regulated or down-regulated. The proposed method naturally applies to these kinds of target estimands. Yet, the utility of different causal estimands in scientific discoveries is worth further exploration.

Supplementary Materials

Supplementary Notes include the proof, computational details, and extra experiment results. These include Lemmas D.1 and D.4, Figures E3 and F1-F3, and Tables F1-F2.

Acknowledgments

We are grateful for comments from the editors and anonymous reviewers. We thank Arun K. Kuchibhotla for insightful discussions of pivotal maximal inequality for finite maximums. This work used the Bridges-2 system at the Pittsburgh Supercomputing Center (PSC) through allocation MTH230011P from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

This project was funded by the National Institute of Mental Health (NIMH) grant R01MH123184.

ORCID

Jin-Hong Du http://orcid.org/0000-0001-9683-4146
Zhenghao Zeng http://orcid.org/0000-0003-1430-6118

References

Balakrishnan, S., Kennedy, E. H., and Wasserman, L. (2023), “The Fundamental Limits of Structure-Agnostic Functional Estimation,” arXiv preprint arXiv:2305.04116. [2485]

Belloni, A., Chernozhukov, V., Chetverikov, D., Hansen, C., and Kato, K. (2018), “High-Dimensional Econometrics and Regularized GMM,” arXiv preprint arXiv:1806.01888. [2491,2492]

Benkeser, D., Carone, M., Laan, M. V. D., and Gilbert, P. B. (2017), “Doubly Robust Nonparametric Inference on the Average Treatment Effect,” *Biometrika*, 104, 863–880. [2496]

Chakrabortty, A., Dai, G., and Tchetgen, E. T. (2022), “A General Framework for Treatment Effect Estimation in Semi-Supervised and High Dimensional Settings,” arXiv preprint arXiv:2201.00468. [2489,2491]

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018), “Double/Debiased Machine Learning for Treatment and Structural Parameters: Double/Debiased Machine Learning," *The Econometrics Journal*, 21, C1–C68. [2486,2490]

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013), "Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors," *The Annals of Statistics*, 41, 2786–2819. [2485,2487,2491]

Díaz, I. (2017), "Efficient Estimation of Quantiles in Missing Data Models," *Journal of Statistical Planning and Inference*, 190, 39–51. [2490]

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., Marjanovic, N. D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016), "Perturb-seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens," *cell*, 167, 1853–1866. [2485]

Du, J.-H., Wasserman, L., and Roeder, K. (2023), "Simultaneous Inference for Generalized Linear Models with Unmeasured Confounders," arXiv preprint arXiv:2309.07261. [2484,2495]

Editorial. (2023), "A Focus on Single-Cell Omics," *Nature Reviews Genetics*, 24, 485. [2484]

Genovese, C. R., and Wasserman, L. (2006), "Exceedance Control of the False Discovery Proportion," *Journal of the American Statistical Association*, 101, 1408–1417. [2485,2492]

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (2011), *Robust Statistics: The Approach based on Influence Functions*, New York: Wiley. [2486]

Imbens, G. W. (2004), "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *Review of Economics and statistics*, 86, 4–29. [2484]

Imbens, G. W., and Rubin, D. B. (2015), *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge: Cambridge University Press. [2484,2488]

Jiang, R., Sun, T., Song, D., and Li, J. J. (2022), "Statistics or Biology: The Zero-Inflation Controversy about scRNA-seq Data," *Genome Biology*, 23, 1–24. [2485]

Jin, J., and Syrgkanis, V. (2024), "Structure-Agnostic Optimality of Doubly Robust Learning for Treatment Effect Estimation," arXiv preprint arXiv:2402.14264. [2485]

Kallus, N., Mao, X., and Uehara, M. (2024), "Localized Debiased Machine Learning: Efficient Inference on Quantile Treatment Effects and Beyond," *Journal of Machine Learning Research*, 25, 1–59. [2489,2491]

Kennedy, E. H. (2022), "Semiparametric Doubly Robust Targeted Double Machine Learning: A Review," arXiv preprint arXiv:2203.06469. [2486,2488]

Kennedy, E. H., Balakrishnan, S., and G'Sell, M. (2020), "Sharp Instruments for Classifying Compliers and Generalizing Causal Effects," *The Annals of Statistics*, 48, 2008–2030. [2486]

Kennedy, E. H., Balakrishnan, S., and Wasserman, L. (2023), "Semiparametric Counterfactual Density Estimation," *Biometrika*, 110, 875–896. [2486]

Kennedy, E. H., Kangovi, S., and Mitra, N. (2019), "Estimating Scaled Treatment Effects with Multiple Outcomes," *Statistical Methods in Medical Research*, 28, 1094–1104. [2488]

Kuchibhotla, A. K., and Patra, R. K. (2022), "On Least Squares Estimation Under Heteroscedastic and Heavy-Tailed Errors," *The Annals of Statistics*, 50, 277–302. [2486]

Laan, M. J., and Robins, J. M. (2003), *Unified Methods for Censored Longitudinal Data and Causality*, New York: Springer. [2485]

Lalli, M. A., Avey, D., Dougherty, J. D., Milbrandt, J., and Mitra, R. D. (2020), "High-Throughput Single-Cell Functional Elucidation of Neurodevelopmental Disease–Associated Genes Reveals Convergent Mechanisms Altering Neuronal Differentiation," *Genome Research*, 30, 1317–1331. [2494]

Perez, R. K., Gordon, M. G., Subramaniam, M., Kim, M. C., Hartoularos, G. C., Targ, S., Sun, Y., Ogorodnikov, A., Bueno, R., Lu, A., et al. (2022), "Single-Cell RNA-seq Reveals Cell Type-Specific Molecular and Genetic Associations to Lupus," *Science*, 376, eabf1970. [2495]

Qiu, Y., Sun, J., and Zhou, X.-H. (2023), "Unveiling the Unobservable: Causal Inference on Multiple Derived Outcomes," *Journal of the American Statistical Association*, 119, 2178–2189. [2484,2487,2488,2489,2492,2493]

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994), "Estimation of Regression Coefficients When Some Regressors are not Always Observed," *Journal of the American Statistical Association*, 89, 846–866. [2485]

Sarkar, A., and Stephens, M. (2021), "Separating Measurement and Expression Models Clarifies Confusion in Single-Cell RNA Sequencing Analysis," *Nature Genetics*, 53, 770–777. [2485,2496]

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999), "Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models," *Journal of the American Statistical Association*, 94, 1096–1120. [2485]

Tsiatis, A. A. (2006), *Semiparametric Theory and Missing Data* (Vol. 4), New York: Springer. [2484,2485,2490]

Van der Laan, M. J. (2014), "Targeted Estimation of Nuisance Parameters to Obtain Valid Statistical Inference," *The International Journal of Biostatistics*, 10, 29–57. [2496]

van der Vaart, A. W. (2000), *Asymptotic Statistics* (Vol. 3), Cambridge: Cambridge University Press. [2486,2490]

Zhang, M., Liu, S., Miao, Z., Han, F., Gottardo, R., and Sun, W. (2022), "Ideas: Individual Level Differential Expression Analysis for Single-Cell RNA-Seq Data," *Genome Biology*, 23, 1–17. [2484]