A new approach to confounder adjustment substantially increases detection power in omics association studies

Sangyoon Yi¹, Xianyang Zhang¹*, Lu Yang², Jinyan Huang³, Yuanhang Liu², Chen Wang², Daniel J. Schaid², Jun Chen²*

¹Department of Statistics, Texas A&M University, College Station, TX, 77843, U.S.A.

²Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, 55901, U.S.A.

³State Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, National Research Center for Translational Medicine, Rui-Jin Hospital, Shanghai Jiao Tong University, Shanghai, 200025, China

* Corresponding author. Email: chen.jun2@mayo.edu (J.C.), zhangxiany@stat.tamu.edu (X.Z.).
Abstract

One challenge facing omics association studies is the loss of statistical power when adjusting for confounders and multiple testing. The traditional statistical procedure involves fitting a confounder-adjusted regression model for each omics feature, followed by multiple testing correction. Here we show that the traditional procedure is not optimal and present a new approach, 2dFDR, a two-dimensional false discovery rate control procedure, for powerful confounder adjustment in multiple testing. Through extensive evaluation, we demonstrate that 2dFDR is substantially more powerful than the traditional procedure, and in the presence of strong confounding and weak signals, the power improvement could be more than 100%.

Keywords: Confounding, False discovery rate, Association testing

Background

High-throughput genomic profiling technologies enable the interrogation of the biological system at different omics levels, generating enormous amounts of omics data [1]. One central task of statistical analysis of omics data is to test the association between omics features and a covariate of interest [2]. The associated omics features, once validated, can provide biological insights into health and disease, act as potential targets for intervention and serve as biomarkers for clinical applications [3,4]. However, observational omics studies are subject to various types of confounding [5-7]. Confounding arises when the relationship between the primary variable and the omics feature is distorted by some other variable (confounder) due to its association with both.
Demographic variables like age, gender, race, and obesity are frequent confounders in omics association studies. For example, in cancer studies, cancer patients are often older than benign controls [8]. In other diseases such as rheumatoid arthritis, the prevalence could differ by gender [9]. Since these demographic variables are known to impact the omics profiles [7], their correlation with the variable of interest could confound the association analysis. Biological heterogeneity, such as the cell mixture in the tissue sample, is also a significant source of confounding. This is because the cell types have distinct omics profiles, and their composition could vary with the variable of interest. For example, the leukocyte composition in the peripheral blood often shifts in the disease condition, and therefore it could severely confound blood-based omics studies [5]. Finally, technical variation, or batch effect, which occurs when the biological samples are not processed or measured together, could strongly confound the associations of interest if the samples are not randomized into the batches [6]. Although the batch confounding can be avoided by careful study design, it is unfortunately prevalent in omics studies [6].

Confounding not only reduces the statistical power by introducing extra variability but also increases the chance of false findings if not properly accounted for. Standard statistical approaches to address confounding include stratification and regression, with the latter being most widely used due to its flexibility [10]. Although adjusting the confounders in the regression model controls false positives, it nevertheless reduces the statistical power. The need for multiple testing correction in omics association analysis further deteriorates statistical power [11]. In the presence of strong confounding, it is not unusual that no significant associations could be recovered after adjusting for
confounders and multiple testing. Therefore, improving the power under confounding and multiple testing is a topic of critical importance and could potentially rescue an underpowered study.

Previous methods separate confounder adjustment from multiple testing. The standard approach, which fits a confounder-adjusted regression model for all omics features followed by multiple testing correction such as false discovery rate (FDR) control [12,13], is used predominantly [14]. However, confounders may affect only a subset of omics features [15-18], and adjusting confounders for every omics feature will be an over-adjustment, leading to substantial power loss. To rescue the power, one naïve idea is to test the significance of the confounder first, and if it is not significant, we exclude the confounder in the regression model. Although this strategy substantially improves the power, controlling the type I error is difficult and heavily depends on the choice of the significance cutoff. We found that this strategy fails to control the type I error properly, even if we use a very lenient cutoff.

In this study, we take a different approach to this problem and integrate the confounder adjustment into multiple testing (FDR control) framework. The new approach uses the statistic from the unadjusted analysis to filter out omics features that are less likely to be associated with the covariate of interest or the confounder. FDR control is then performed based on the adjusted statistic on the remaining features. The challenge here is to account for the dependency between the unadjusted and adjusted statistic so that the FDR is controlled. We provide a robust and powerful procedure, two-dimensional false
discovery rate control procedure (2dFDR), which is proved to offer asymptotic FDR control and dominate the power of the traditional procedure.

Results

Overview of the two-dimensional false discovery rate control procedure (2dFDR)

2dFDR is based on linear models with the measurement of the omics feature as the outcome, which is one popular modeling approach for functional omics data, and assumes the confounders are known. It depends on the unadjusted and adjusted test statistics, denoted by $Z^U_i$ and $Z^A_i$, respectively, from fitting both the unadjusted and the confounder-adjusted model to the $i$th omics feature. 2dFDR proceeds in two dimensions. In the first dimension, it uses the unadjusted statistic $Z^U_i$ to screen out a large number of irrelevant features (noises) that are not associated with the covariate of interest or the confounder. In the second dimension, it uses the adjusted statistic $Z^A_i$ to identify the true signals on the remaining features and control the FDR at the desired level. Although the unadjusted statistic is biased and captures the effects from both the covariate of interest and the confounder, it can be leveraged to increase the signal density and reduce multiple testing burden in the second dimension. Thus, 2dFDR boils down to selecting features with $|Z^U_i| \geq t_1$ (first dimension) and $|Z^A_i| \geq t_2$ (second dimension). The cutoffs $t_1$ and $t_2$ are chosen to achieve maximum power while controlling the FDR at the desired level.

Fig.1A-C illustrate the idea using simulated data (Supplementary Note S1), where we plot $Z^A_i$ against $Z^U_i$ for confounded scenarios. The standard approach performs (one-dimensional) FDR control based on the adjusted statistic $Z^A_i$ only (we refer it as 1dFDR-
When there the correlation between the variable of interest and the confounder (denoted as \( \text{cor}(x, z) \)) is high, the signals (brown) and noises (blue) overlap much on \( Z^A_i \) due to loss of power with confounder adjustment (Fig. 1A). To achieve the desired FDR level, 1dFDR-A requires a high \(|Z^A_i|\) cutoff (blue line). For 2dFDR, it first uses \( Z^U_i \) to exclude a large number of irrelevant features (vertical red line). Next, a much lower \(|Z^A_i|\) cutoff (horizontal red line) is used to achieve the same FDR level. As a result, it achieves significant power improvement, and the improvement increases with the correlation between the variable of interest and the confounder (Fig. 1B-C).

A particular challenge for this new approach is to address the dependency between the two dimensions. 2dFDR simultaneously selects \( t_1 \) and \( t_2 \) and considers the selection effect in the first dimension (Methods and Supplementary Note S2). 2dFDR can be viewed as a two-dimensional generalization of the classical Benjamini-Hochberg (BH) procedure [12], where we search for the cutoff values in a two-dimensional space. An intrinsic difficulty is to estimate the expected number of false rejections at a given \( t_1 \) and \( t_2 \); this is achieved by a non-parametric Empirical Bayes method [19] (Supplementary Note S2.3). We have conducted a thorough theoretical investigation of the proposed procedure and all the theoretical results are included in Supplementary Note S3 and S4. Under suitable assumptions, we show that 2dFDR provides asymptotic FDR control (Supplementary Note S3), and the power dominates the standard 1dFDR-A (Supplementary Note S4).
Simulation studies to evaluate FDR control and power

We demonstrate the power and robustness of 2dFDR using comprehensive simulations comparing to 1dFDR-U and 1dFDR-A, two one-dimensional FDR procedures based on the unadjusted and adjusted model, respectively. A heuristic strategy (1dFDR-H), which starts with the adjusted model and uses the unadjusted model if the effect of the confounder is not significant, was also compared. We refer the omics features affected by the variable of interest as “true signals” and the omics features affected by the confounder as “confounding signals”. For both the true and confounding signals, we use “signal density” and “signal strength” to represent the percentage of features affected and their effect size, respectively.

We first study the performance of 2dFDR under varying signal density, signal strength, and cor(x, z) (Methods). Both 1dFDR-A and 2dFDR controlled the FDR at the target level across settings, while 1dFDR-U and 1dFDR-H failed to control the FDR under a medium or high cor(x, z) (Fig. 2A). 2dFDR was substantially more powerful than 1dFDR-A when cor(x, z) was high and was comparable when cor(x, z) was low (Fig. 2B). The power increase was more pronounced for weak and sparse signals, with a percent increase of more than 100% (Fig. 2B). This is particularly relevant for real applications, where weak signals and strong confounding are the most challenging situation that needs novel methodological developments.

We next study the effect of the confounding signals' strength and density by varying their magnitudes (Fig. 3) while fixing the true signals’ strength and density. Similarly,
2dFDR maintained the FDR at the target level across settings (Fig. 3A) and was significantly more powerful when $\text{cor}(x, z)$ was medium or high (Fig. 3B). The power difference, however, decreased as the confounding signals became denser (top to bottom). When the confounder affected 50% of the features, 2dFDR could be less powerful than 1dFDR-A even when $\text{cor}(x, z)$ was high (Supplementary Fig. S1). This is expected since if the confounder affects every omics feature, 1dFDR-A, which adjusts the confounder for every omics feature, is optimal. Higher strength of the confounding signals (left to right) also reduced the power difference. The results remained the same if we simulated five confounders (Supplementary Fig. S2).

We also studied the effect of colocation between the true and confounding signals, where the omics features were affected by both the variable of interest and the confounder (Supplementary Fig. S3). We found that 2dFDR was more powerful than 1dFDR-A when the density of the confounding signals was low and $\text{cor}(x, z)$ was high. However, as the confounding signals became denser, 2dFDR could be less powerful than 1dFDR-A when $\text{cor}(x, z)$ was low.

Since 2dFDR is developed based on the assumption that the omics features are independent, it is important to study the robustness of 2dFDR to the correlations among omics features. We thus simulated block and autoregressive correlation structures (Methods), which were commonly observed for omics data. We found that 2dFDR was quite robust to these two correlation structures (Supplementary Fig. S4 and S5), and the
FDR was controlled near the target level. 2dFDR maintained the power in these scenarios.

2dFDR offers asymptotic FDR control, i.e., the FDR is proved to be controlled if the sample size and feature size are large. It is interesting to study the sample size and feature size where it breaks down. We thus simulated sample sizes of 50 and 25 (Supplementary Fig. S6) and feature sizes of 500 and 100 (Supplementary Fig. S7).

We found that the performance 2dFDR remained robust and powerful at the sample size of 50 and the feature size of 500. However, it became less powerful than the traditional procedure at the sample size of 25. The FDR also started to be inflated at the feature size of 100, especially when \( \text{cor}(x, z) \) was high. We also found that increasing the sample size or feature size alone did not rescue the performance deterioration due to the other being small (Supplementary Fig. S8). Overall, 2dFDR is robust up to a moderate sample size and feature size. For a very small sample or feature size, applying 2dFDR is not recommended.

2dFDR is computationally efficient and can scale up to a large sample and feature size. It can be run in parallel on each grid point to further increase its computational speed. Supplementary Fig. S9 shows the computational time for running a simulation instance at different sample sizes and feature sizes. With \( n = 800, m = 512k \) and one confounder, 2dFDR completes the analysis in 546s using a search grid of \( 50 \times 50 \) without parallelization on a MacBook Pro laptop. The memory requirement is the same as fitting regular linear regressions and requires only accommodating a matrix multiplication of
\[ A_{p \times n} B_{n \times m} \], where \( p, n, m \) are the number of covariates, sample size and feature size, respectively.

**Evaluation of the detection power on real omics datasets**

We apply 2dFDR to three different types of omics datasets to demonstrate its empirical power on real data. We compare to the traditional adjusted procedure 1dFDR-A based on the numbers of detected omics features at the same FDR level.

The first is a hepatocellular carcinoma transcriptomics dataset from TCGA [20] \((n = 342, m = 19,329)\), which is used to detect gene expressions associated with human hepatitis B virus (HBV) infection [21]. Gender and ethnicity are confounders for this dataset and were adjusted in the model. 2dFDR detected more genes than 1dFDR-A across different FDR levels (Fig. 4A). At the standard 5% FDR level, 1dFDR-A failed to identify any HBV-associated genes, while 2dFDR successfully identified 27 genes.

The second is a metabolomics dataset [22, 23] \((n = 289, m = 1,201)\), where the aim is to identify serum metabolites associated with insulin resistance (IR), accounting for the confounding effect of body mass index (BMI). Again, 2dFDR detected more IR-associated metabolites at different FDR levels (Fig. 4B). At 5% FDR, 2dFDR and 1dFDR-A recovered 481 and 412 metabolites, respectively. 2dFDR was able to identify the majority of the metabolites by 1dFDR-A (378 out of 412) and it also recovered 103 metabolites missed by 1dFDR-A.
Finally, we benchmark 2dFDR using an extensive collection of epigenomics datasets from various epigenome-wide association studies (EWAS) using tissue samples [24] (Supplementary Table S1). The objective is to identify differentially methylated CpG positions (DMPs) associated with a condition of interest. Since a tissue sample contains a mixture of cell types, each with a distinct methylation profile, the covariation of their mixture proportion with the condition could strongly confound the associations of interest [5]. To capture the cell mixture, we used surrogate variable analysis (SVA), and the estimated surrogate variables were adjusted in the model [25]. For these EWAS datasets, 2dFDR detected significantly more DMPs than 1dFDR-A in most datasets with a median increase of 136% (Fig. 4C&D, Supplementary Table S1). Consistent with the simulations, the power improvement was more pronounced when the signals were weak (lower part of the box plot in Fig. 4C). Moreover, 2dFDR was able to detect DMPs in six datasets, where 1dFDR-A failed to identify any.

Validation of the increased detection power on EWAS datasets

To validate the additional DMPs detected by 2dFDR, we resorted to the five age-related EWAS datasets (Supplementary Table S1) to see if the additional DMPs from one age dataset had evidence of support from the other four. This was achieved by examining the confounder-adjusted p-value distribution of the DMPs detected by 2dFDR only (at 5% FDR) in the other four age datasets. If these DMPs from one age dataset were truly age-associated, we expect to see smaller p-values for them in the other age datasets, compared to the p-values of random CpG loci. Clearly, the distribution was enriched in small p-
values, indicating the plausibility of DMPs detected by 2dFDR (Fig. 5A). Validation based on the two SLE datasets reached a similar conclusion (Fig. 5B).

We further performed a downsampling analysis to validate the improved power of 2dFDR. We first curate a list of highly significant features using Bonferroni correction based on the p-values from the adjusted analysis on the full dataset. Next, we downsample the full dataset to smaller sizes and compare the ability of 2dFDR and 1dFDR-A in recovering these highly significant features as a way of power assessment.

We illustrate this strategy using one EWAS dataset (EWAS22, n=111), where the genome-wide methylation difference was compared between smokers and non-smokers in African American women using peripheral blood mononuclear cells [26]. With Bonferroni correction (alpha = 0.05) to the p-values based on the full dataset, we identified 10 differentially methylated CpG positions (DMPs) and these DMPs were treated as “gold standard” (we label them as “gDMPs”). We then subsampled the full dataset to sample sizes of 20, 40, 60, 80, and 100, and compared the power (recovery rate) of 2dFDR and 1dFDR-A in recovering these gDMPs. We observed that 2dFDR outperformed 1dFDR-A at nearly all sample sizes (Fig. 5C). The power improvement was more significant in the middle range (i.e., n = 60 and 80). At the sample size of 100, the recovery rates of both 2dFDR and 1dFDR-A both reached nearly 100%.

**Discussion**

Confounding and high-dimensionality are the two major statistical challenges in omics data analysis. Previous research separates these two problems, and methodological
developments are focused on each of them. In this study, we integrate the confounder adjustment into multiple testing by performing two-dimensional false discovery rate control based on both the adjusted and unadjusted statistics. Although the unadjusted statistic is biased, it can be leveraged to enrich signals and reduce the multiple testing burden. The resulting procedure, 2dFDR, has proven to offer asymptotic FDR control and dominate the power of the traditional procedure based on the adjusted statistic only. Through simulations and real data applications, we demonstrate that 2dFDR is substantially more powerful than the traditional procedure. We also show that 2dFDR is robust to the typical correlation structures seen in omics data and performs well at moderate sample sizes and feature sizes.

The 2dFDR procedure is the most powerful when the correlation between the variable of interest and the confounder is high, and/or the signals are weak. This makes it a practically very useful approach since existing methods have limited power in these scenarios. 2dFDR also works best when the confounder only affects a subset of omics features. This is usually a reasonable assumption for many conditions, such as age and gender [15-18]. However, there could be situations where the assumption is violated. For example, strong batch effects could possibly affect a large number of omics features. In such a case, 2dFDR has a limited power advantage or could be less powerful than the traditional procedure (Supplementary Fig. S1). One diagnostic approach is to calculate the percentage of the genomic variance explained ($R^2$) by the variable of interest and the confounder, respectively, using multivariate methods such as PERMANOVA [27]. If the $R^2$ of the confounder is substantially larger than that of the variable of interest, it
indicates that the confounder signals may be very dense. Another approach is to study
the distribution of the p-values of the confounder from the adjusted analysis. If we see a
spike on the left end of the p-value distribution, it also suggests dense confounder signals.

Since 2dFDR depends on the unadjusted statistic to filter features, when the confounder
and variable of interest have opposite effects on the omics feature with similar
magnitude, they will cancel out each other's effect, and the omics feature could be
excluded erroneously in the first dimension. The optimal cutoff of $|Z_i^U|$ is thus
determined based on the tradeoff between power reduction due to erroneously excluding
these relevant features in the first dimension and power increase due to reducing the
multiple testing burden and increasing the signal density in the second dimension. If the
true signals can only be revealed after adjusting for the confounder, for example, when
the true and confounding signals co-locate with opposite directions, unadjusted statistics
will not be informative. In this case, the best cutoff on $|Z_i^U|$ should be 0 and 2dFDR is
then reduced to the traditional 1dFDR-A. However, in finite samples, it may not always
be possible to reduce 2dFDR to 1dFDR-A exactly and 2dFDR could be less powerful
than 1dFDR-A in such situations (Supplementary Fig. S10A, 10% true and confounding
signals co-locate with opposite directions).

When the correlation between the variable of the interest and the potential confounder is
small, adjusting the confounder will lead to power improvement if the confounder has
large effects on the outcome and will not hurt the power much if the confounder has no
effects on the outcome. Therefore, in such situation, performing adjusted analyses is a
good choice. In fact, when $\text{cor}(x, z)$ is 0, the test statistic $Z_i^U$ and $Z_i^A$ are almost perfectly correlated (Supplementary Fig. 10B), and performing two-dimensional FDR control based on $Z_i^U$ and $Z_i^A$ is almost equivalent to performing one-dimensional FDR control based on $Z_i^A$. Again, in finite samples, there could be some power loss for 2dFDR (Supplementary Fig. 10B). Fortunately, $\text{cor}(x, z)$ can be known before the analysis. We thus do not recommend running 2dFDR when $\text{cor}(x, z)$ is small.

The proposed method belongs to the general topic of using auxiliary data to increase power in multiple testing, which has been an active research area recently [24,28-32]. To be effective, the auxiliary data has to be informative of the probability of the null hypothesis or the statistical power. In our case, the unadjusted statistic can be considered as a particular type of auxiliary data, which informs the prior null probability (the smaller the statistic, the more likely the null hypothesis). However, as the auxiliary data (unadjusted statistic) and the primary data (adjusted statistic) are correlated even under the null hypothesis, existing structure-adaptive multiple testing methods are not directly applicable in our problem. 2dFDR explicitly addresses such correlation to enable asymptotic FDR control.

Although developed in the linear model setting, the idea of 2dFDR can be extended to the generalized linear model or generalized linear mixed effects model [33] using the Wald $z$-statistic. Another interesting extension is to adapt 2dFDR to the setup where the omics features are treated as covariates. For example, in genome-wide association studies (GWAS), the genetic variants are usually modeled as covariates. We expect that the
same 2dFDR idea could be applied to GWAS to significantly increase its power, since
population stratification, which occurs when samples come from genetically diverse
underlying populations, is an important confounder for GWAS [34].

In summary, 2dFDR is a new approach to confounder adjustment under multiple testing.
It is powerful, robust, and scalable. As a general methodology, we envision its broad
applicability of 2dFDR in omics association studies.

Methods

The 2dFDR procedure

Here we give an overview of the method. Details could be found in Supplementary Note S2. Consider the following $m$ linear models

$$E(Y_{ij}) = x_j \alpha_i + z_j^T \beta_i, \ 1 \leq j \leq n, 1 \leq i \leq m$$

where $n$ and $m$ are the sample size and feature size, respectively, $(Y_{i1}, ..., Y_{in})^T \in \mathbb{R}^{n \times 1}$ is the measurement of the omics feature $i$, $x = (x_1, ..., x_n)^T \in \mathbb{R}^{n \times 1}$ is the vector of the covariate of interest, $(z_1, ..., z_n)^T \in \mathbb{R}^{n \times d}$ is the design matrix for the confounding factors, and $\alpha_i \in \mathbb{R}$, $\beta_i \in \mathbb{R}^{d \times 1}$ are the coefficients associated with the covariate and confounding factors respectively. Under the model, there are four different categories of features to consider: (A) Solely associated with the covariate of interest: $\alpha_i \neq 0, \beta_i = 0$; (B) Solely associated with the confounder: $\alpha_i = 0, \beta_i \neq 0$; (C) Associated with both the covariate of interest and the confounder: $\alpha_i \neq 0, \beta_i \neq 0$; (D) Not associated with either the covariate of
interest or the confounder: \( \alpha_i = 0, \beta_i = 0 \). Our goal is to develop a multiple testing procedure for simultaneously testing \( m \) hypotheses \( H_{0,i} : \alpha_i = 0 \) vs \( H_{1,i} : \alpha_i \neq 0 \) (\( i = 1, ..., m \)) in the presence of confounding effects. Let \( Z_i^A \) be the z-statistic for testing whether \( \alpha_i = 0 \) after adjusting for the confounding factors, and \( Z_i^U \) be the unadjusted version without adjusting for the confounding factors, i.e., setting \( \beta_i = 0 \) in the model.

Given the thresholds \( t_1, t_2 \geq 0 \), the 2dFDR procedure can be described as following:

**Dimension 1: Signal enrichment.** Use the unadjusted statistic \( Z_i^U \) to retain more promising features \( S_1 = \{1 \leq i \leq m : |Z_i^U| \geq t_1 \} \).

**Dimension 2: Excluding false positives.** For \( i \in S_1 \), reject \( H_{0,i} \) with \( Z_i^A \geq t_2 \). As a result, the final set of discoveries is given by \( S_2 = \{1 \leq i \leq m : |Z_i^U| \geq t_1, |Z_i^U| \geq t_2 \} \).

Given \( t_1, t_2 \geq 0 \), the FDP is defined as

\[
\text{FDP}(t_1, t_2) = \frac{\sum_{i=1}^{m} \mathbb{1}\{|Z_i^U| \geq t_1, |Z_i^A| \geq t_2, \alpha_i = 0\}}{1 \vee \sum_{i=1}^{m} \mathbb{1}\{|Z_i^U| \geq t_1, |Z_i^A| \geq t_2\}}
\]  

(2)

The key here is to select the thresholds \( t_1, t_2 \) to maximize the number of discoveries while controlling the FDR (expectation of FDP) at the desired level.

As the number of false rejections is unknown (the numerator of \( \text{FDP}(t_1, t_2) \)), we replace it by the expected number of false rejections. We can show that the expected number is given by \( \sum_{i:\alpha_i=0} L(\mu_i, t_1, t_2) = P(|V_1 + \mu_i| \geq t_1, |V_2| \geq t_2) \) with \( \mu_i \) being a nuisance parameter that depends on the confounding effect (the magnitude of \( \beta_i \) and the correlation between \( x \)).
and \( z \) and \((V_1, V_2)\) being bivariate mean-zero normal random variables, whose covariance can be calculated from the data (Supplementary Note S2.2). Note that the correlation between \( V_1 \) and \( V_2 \) is not equal to zero in general, which captures the dependence between the two dimensions.

An intrinsic difficulty here is that the expected number of false rejections depends on the effects of the confounding factors (i.e., \( \mu_i \)) on each feature. As the number of features could be huge, it thus requires estimating a large number of nuisance parameters. To tackle this challenge, we adopt a Bayesian viewpoint by assuming that the nuisance parameters \( \mu_i \) are generated from a common prior distribution \( G \). The Bayesian viewpoint allows us to express the expected number of false rejections as a functional of the prior distribution. Therefore, we can translate the task into estimating the prior distribution instead of the direct estimation of a large number of nuisance parameters. The prior distribution can be estimated via the general maximum likelihood empirical Bayes estimation [19,35], which can be cast into a convex optimization problem (Supplementary Note S2.3). Denote \( \hat{G} \) the estimate of \( G \). Then we can derive an approximate upper bound for \( \text{FDP}(t_1, t_2) \), as follows:

\[
\text{FDP}(t_1, t_2) \leq \frac{m^{-1} \sum_{i=1}^{m} L(\mu_i, t_1, t_2)}{m^{-1} \sum_{i=1}^{m} 1 \left\{ |Z^U_i| \geq t_1, |Z^A_i| \geq t_2 \right\}} \tag{3}
\]

\[
\approx \frac{\int L(\mu, t_1, t_2) d\hat{G}(\mu)}{m^{-1} \sum_{i=1}^{m} 1 \left\{ |Z^U_i| \geq t_1, |Z^A_i| \geq t_2 \right\}} := \text{FDP}(t_1, t_2).
\]

For a desired FDR level \( \alpha \in (0, 1) \), we choose the optimal threshold such that
\[
(t_1^*, t_2^*) = \arg \max_{(t_1, t_2) \in \mathcal{F}_\alpha} \sum_{i=1}^{m} 1 \{|Z^U_i| \geq t_1, |Z^A_i| \geq t_2\},
\]

(4)

where \(\mathcal{F}_\alpha = \{(t_1, t_2) \in \mathbb{R}^+ \times \mathbb{R}^+: \text{FDP}(t_1, t_2) \leq \alpha\}\). Finally, we select features with \(|Z^U_i| \geq t_1^*, |Z^A_i| \geq t_2^*\). This procedure can be viewed as two-dimensional generalization of the Benjamini-Hochberg (BH) procedure [12]. It is well known that when the number of signals is a substantial proportion of the total number of hypotheses, the BH procedure will be overly conservative. To adapt to the signal density, we develop a modification of John Storey's approach [13] in our setting (Supplementary Note S2.4).

Data simulation

We conduct comprehensive simulations to evaluate the finite-sample performance of the proposed method and compare it to competing methods. For genome-scale multiple testing, the number of hypotheses could range from thousands to millions. For demonstration purposes, we start with \(m = 10,000\) features, \(n = 100\) samples, one covariate of interest \(x\), and one confounder \(z\). To comprehensively evaluate the proposed method's performance under different scenarios, we study the following important parameters:

- The correlation between the covariate of interest (\(x\)) and the confounder (\(z\)) (denoted as “\(\text{cor}(x, z)\)’);
- The density and strength of the true signals;
- The density and strength of the confounding signals;
The degree of colocation of the true signals and confounding signals.

To induce the correlation between $\mathbf{x}$ and $\mathbf{z}$, we let $\mathbf{x}_0 \sim N(0, 1), \mathbf{x} = c\mathbf{x}_0 + N(0, 1),$
\[ \mathbf{z} = c\mathbf{x}_0 + N(0, 1). \]
We set $c = 0.5, 1.25$ and $2$, which achieves a correlation ($\rho$) about $0.2, 0.6$ and $0.8$ respectively, representing weak ('+'), moderate ('++') and strong confounding ('+++') level. For the multiple-confounder scenario, we simulate each $\mathbf{z}$ in the same way.

Next, we generate
\[ \mathbf{y}_i = \alpha_i \mathbf{x} + \beta_i \mathbf{z} + \epsilon_i, \quad (i = 1, \ldots, m) \]
where $\alpha_i, \beta_i \sim \frac{\pi}{2} \text{Unif}(-l - \delta, -l) + \frac{\pi}{2} \text{Unif}(l, l + \delta) + (1 - \pi) I_0$, $I_0$ is a mass at $0$ and $\epsilon_i \sim N(0, 1)$. For both the covariate of interest and the confounder, we simulate three levels of signal density ($\pi = 5\%, 10\%$, and $20\%$ - representing low, medium, and high density) and three levels of signal strength ($\delta = 0.2$ and $l = 0.2, 0.3$ and $0.4$, representing weak, moderate, and strong effect). In the basic setup, $\alpha_i, \beta_i$ are independently drawn from the distribution mentioned above. To study the effect of the colocation of the true and confounding signals (nonzero $\alpha_i$ and $\beta_i$), we further simulate two scenarios:

- Non-colocation: $\forall j, \alpha_j \beta_j = 0$
- $50\%$ colocation: $|S_{\alpha \beta}| = 0.5 \min\{|S_{\alpha}|, |S_{\beta}|\}$, where $S_{\alpha \beta} = \{i | \alpha_i \beta_i \neq 0\}$,
\[ S_\alpha = \{i | \alpha_i \neq 0\}, S_\beta = \{i | \beta_i \neq 0\}. \]
In addition, we investigate the robustness of the proposed method to correlated errors. Specifically, we study:

- First-order autoregressive (AR(1)) correlation structure, where $\rho(\epsilon_i, \epsilon_k) = 0.6|i-k|$.
- Block correlation structure, where we simulate 100 blocks with within-block correlation 0.6.

We compare 2dFDR to the following methods:

- 1dFDR-U: linear regression with the covariate of interest without adjusting for the confounder.
- 1dFDR-A: linear regression with the covariate of interest adjusting for the confounder, which is the traditional procedure.
- 2dFDR-H: a heuristic hybrid procedure, which first runs “1dFDR-A”, and if the confounder is not significant (nominal $p < 0.05$), “1dFDR-U” is used.

All the methods use the q-value approach for FDR control [13], following the computation of feature-wise p-values. We evaluate the performance based on FDR control (false discovery proportion) and power (true positive rate) at a target FDR level of 5%. Results are averaged over 100 simulation runs (for small feature sizes, 1,000 simulation runs are performed to reduce variability). Both the means and 95% CIs are reported in the bar plots.
Real datasets

**Transcriptomics dataset.** The dataset consists of 342 RNA-Seq samples from patients with hepatocellular carcinoma from The Cancer Genome Atlas (TCGA) [20]. We use this dataset to identify gene expressions associated with chronic infection of the Hepatitis B virus (HBV). HBV status was examined by counting the percentage of reads mapped to HBV genome[36], resulting in 103 and 239 HBV positive and negative cases. Ethnicity and gender are confounders in this analysis since they are correlated with the HBV status (OR: 0.051 and 2.67, respectively, p < 0.001) and are known to affect the gene expressions. The raw FASTQ files were processed through Mayo’s internal MAP-RSeq pipeline (Version 3.0) [37]. The gene counts were generated by FeatureCounts using the gene definitions files from Ensembl v78 [38]. Quality control was carried out using RSeqQC [39], and a total of 19,329 genes were included. Transcript per million (TPM) was calculated, quantile normalized, and log-transformed before analysis.

**Metabolomics dataset.** The dataset came from a study of the impact of the gut microbiome on host serum metabolome and insulin sensitivity in non-diabetic Danish adults [22]. It consists of measurements of 1,201 metabolites (325 serum polar metabolites and 876 serum molecular lipids) on 289 serum samples using mass spectrometry. The cleaned dataset was downloaded from https://bitbucket.org/hellekp/clinical-micro-meta-integration [23]. We use this data set to identify insulin resistance (IR)-associated metabolites. IR was estimated by the homeostatic model assessment [22]. Body mass index (BMI) is a confounder for this
dataset since it is highly correlated with IR (Spearman’s $\rho = 0.67$) and is known to affect the serum metabolome. Two samples without IR measurement were excluded. For metabolites with zero measurements, zeros were replaced by half of the minimal nonzero value. Log transformation was performed to make the data more symmetrically distributed before analysis.

**Epigenomics datasets.** The datasets came from 51 epigenome-wide association studies (EWAS) of various phenotypes using Infinium Human Methylation 450K BeadChip. They were collected and processed as previously described [24]. A total of 54 datasets with binary or continuous phenotypes and sample sizes larger than 100 were included in the evaluation (**Supplementary Table S1**). Since these EWAS studies all used tissue samples, which consist of a mixture of different cell types, each with a distinct methylation profile, the shift in the cell mixture proportions with regard to the phenotype of interest could strongly confound the association analysis [5]. For the peripheral blood sample, it consists of different leukocyte subtypes, whose composition usually changes with the onset of disease as a host immune defense mechanism. Since the cell mixture proportions were not directly available for these datasets, we used the surrogate variable analysis [25] to infer the latent factors (also called surrogate variables) that capture the cell mixtures. Specifically, the “isva” R package [40] was used to estimate the number of surrogate variables based on random matrix theory, and the "SmartSVA” R package [18] was used to compute the surrogate variables. The inferred surrogate variables were highly correlated with the phenotypes (a median $R^2 0.49$, **Supplementary Table S1**), and were adjusted in the regression analysis. All the analysis was performed on the methylation M-
values [41], and 5% FDR was used to identify differentially methylated CpG positions (DMPs).

**Abbreviations**


**Declarations**

**Ethical approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Availability of data and materials**

The R package “tdfdr” is available online (https://github.com/jchen1981/tdfdr) with documentation and a tutorial. Codes and data to reproduce the presented results can also be found in that directory.

**Competing interests**
The authors declare no competing financial interests.

Funding

This work was supported by the Center for Individualized Medicine at Mayo Clinic (Chen), National Science Foundation DMS-1830392 and NSF DMS-1811747 (Zhang), and NIH 1 R21 HG011662 (Chen&Zhang).

Author contributions

J.C. & X.Z. conceived, designed, and implemented the method together. X.Z. and S.Y. performed the theoretical analysis. J.C., X.Z. and S.Y. developed the software and performed the simulations. J.C. and L.Y. performed the real data analysis with the help from C.W., Y.L. and J.Y. D.S. contributed expertise to improve the manuscript. J.C. and X.Z led the writing of the manuscript with contributions from all other authors.

Acknowledgements

Not applicable.

Figures and Tables

Fig. 1 Illustration of the 2dFDR procedure using simulated datasets.

(A) shows the decision boundaries for 1dFDR-A (blue line) and 2dFDR (red line) at 5% FDR for a highly confounded scenario ( \( \rho \approx 0.8 \)). 1dFDR-A relies on adjusted statistic \(|Z^A|\) only (one dimension), while 2dFDR is based on both the adjusted and unadjusted
statistic $|Z^A|$ and $|Z^U|$ (two dimensions). $|Z^U|$ is used to exclude a large number of irrelevant features (red vertical line). After that, a less stringent cutoff of $|Z^A|$ (red horizontal line) can be used to achieve a higher power while maintaining the same FDR. The aim of 2dFDR is thus to find the best cutoffs on the two dimensions to maximize the power. (B, C) show that the power (true positive rate) difference between 1dFDR-A and 2dFDR increases with the correlation between the variable of interest and the confounder. When the correlation is low ('+', $\rho \approx 0.2$), $|Z^A|$ and $|Z^U|$ are highly correlated, and $|Z^U|$ provides little extra information. The power of 2dFDR is thus similar to that of 1dFDR-A. When the correlation is higher ('++', '+++', $\rho \approx 0.6, 0.8$), the signals (brown) and noises (blue) are more difficult to separate on $|Z^A|$. By using $|Z^U|$, 2dFDR excludes a large number of noises without losing many signals. The signal density on $|Z^A|$ is thus enriched, leading to a significant power gain.

**Fig. 2 Performance on simulated datasets across varying density and strength of the true signals.**

Average false discovery proportions and true positive rates are compared at 5% FDR using simulated datasets. 1dFDR-U and 1dFDR-A represent the one-dimensional FDR control procedures based on the unadjusted model and confounder-adjusted model, respectively. 1dFDA-H is a heuristic adaptive procedure that uses the adjusted or unadjusted model depending on whether the confounder effect is significant (nominal p-value < 0.05). The performance is evaluated at varying signal strength (left: weak, right: strong), signal density (top: low, bottom: high), and the correlation between the variable
of interest and the confounder (inside the panel, ‘+’, ‘++’, and ‘+++’ represent a low, medium and high correlation ($\rho \approx 0.2, 0.6, 0.8$), respectively). The density of the confounding signals is 10%, and the strength is moderate. 2dFDR and 1dFDR-A control the FDR at the target level (dashed line), while 1dFDR-U and 1dFDR-H fail to control the FDR properly when the confounding is not weak (A). 2dFDR becomes substantially more powerful than 1dFDR-A as the correlation between the variable of interest and the confounder increases (B). The difference is more pronounced when the signals are weak and sparse, as indicated by the percent increase shown on top of the bars.

**Fig. 3 Performance on simulated datasets across varying density and strength of the confounding signals.**

Average false discovery proportions (A) and true positive rates (B) are compared at 5% FDR using simulated datasets. 1dFDR-U and 1dFDR-A represent the one-dimensional FDR control procedures based on the unadjusted model and confounder-adjusted model, respectively. 1dFDA-H is a heuristic adaptive procedure that uses the adjusted or unadjusted model depending on whether the confounder effect is significant (nominal p-value < 0.05). The density of the true signals is 10%, and the strength is moderate. The performance is evaluated at varying confounding signal strength (left: weak, right: strong), confounding signal density (top: low, bottom: high), and the correlation between the variable of interest and the confounder (inside the panel, ‘+’, ‘++’, and ‘+++’ represent a low, medium and high correlation ($\rho \approx 0.2, 0.6, 0.8$), respectively). 2dFDR maintains the FDR at the target level across settings and is significantly more powerful.
when the correlation between the variable of interest and the confounder is not low (++/+++). The power difference decreased as the confounding signals become denser (top to bottom).

**Fig. 4 The empirical power of 2dFDR on real omics datasets.**

2dFDR made more discoveries than 1dFDR-A across FDR target levels for (A) TCGA hepatocellular carcinoma transcriptomics (RNA-Seq) dataset (m = 19,329, confounder: gender and ethnicity), (B) Insulin resistance metabolomics dataset pooling polar metabolites and molecular lipids (m = 1,201, confounder: BMI). (C, D) Evaluation of 2dFDR on 54 epigenomics (450K methylation array) datasets from EWAS of various phenotypes (m ≅ 450,000, confounder: cell mixtures). (C) Boxplot comparing the number of DMPs (differentially methylated positions) detected by 2dFDR and 1dFDR-A at 5% FDR over the 54 datasets. 2dFDR recovered more DMPs than 1dFDR-A in 43 datasets. Each gray dot represents a dataset, and the same dataset is connected by a line. (D) The distribution of the percent increase in detection power over the 54 datasets. 2dFDR achieves a median percent increase of 136% over 1dFDR-A.

**Fig. 5 Validation of 2dFDR using EWAS datasets.**

(A) The distribution of the confounder-adjusted p-values for those 2dFDR-exclusive DMPs in other age EWAS datasets. The diagonal parts show the densities of p-values of all loci for the five age EWAS datasets, serving as a baseline. The off-diagonal parts show the distribution of p-values of 2dFDR-exclusive DMPs in other age datasets. For
instance, the first row shows the distribution of p-values for the 2dFDR-exclusive DMPs from EWAS26 in EWAS27 (green), EWAS30 (purple), EWAS39 (brown), and EWAS45 (orange). The distribution is enriched in smaller p-values than the distribution of all p-values for the respective dataset (diagonal). (B) Similar validation using two Systemic Lupus Erythematosus (SLE) EWAS datasets (EWAS28 and EWAS29, table S1). (C) Validation using a downsampling strategy on EWAS22 (n=111). First, a list of “gold standard” DMPs (gDMPs) was created using Bonferroni correction based on the p-values from the adjusted analysis on the full dataset. Next, the full dataset was downsampled to smaller sample sizes, and the ability of 2dFDR and 1dFDR-A in recovering those gDMPs was compared. For each sample size, 100 replications were performed, and means and standard errors are plotted. 2dFDR is more powerful in identifying these gDMPs at smaller sample sizes.

**Supplementary information**

*Supplementary Information and Source Data files are available in the online version of the paper at:*

**Supplementary Table S1**
**Supplementary Figs S1 to S10**
**Supplementary Notes S1 to S4**

**References**


Fig. 1

**A**

1dFDR-A

2dFDR: Dim 1

2dFDR: Dim 2

**B**

cor(x, z):+

cor(x, z):++

cor(x, z):+++  

**C**

cor(x, z):+

cor(x, z):++

cor(x, z):+++  

|z_A| |z_U| |z_U| |z_U|  
|---|---|---|---|  
|0 5 10| 0 5 10| 0 5 10| 0 5 10|
Fig. 2

A

<table>
<thead>
<tr>
<th>Effect</th>
<th>Low density</th>
<th>Medium density</th>
<th>High density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strong effect</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>Effect</th>
<th>Low density</th>
<th>Medium density</th>
<th>High density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strong effect</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

cor(x, z)

1dFDR−U 1dFDR−A 1dFDR−H 2dFDR
Fig. 3

A

B

<table>
<thead>
<tr>
<th>Weak effect</th>
<th>Moderate effect</th>
<th>Strong effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low density</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>Medium density</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>High density</td>
<td>+</td>
<td>++</td>
</tr>
</tbody>
</table>

False Discovery Proportion

1dFDR−U  1dFDR−A  1dFDR−H

2dFDR

A

B

<table>
<thead>
<tr>
<th>Weak effect</th>
<th>Moderate effect</th>
<th>Strong effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low density</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>Medium density</td>
<td>31%</td>
<td>27%</td>
</tr>
<tr>
<td>High density</td>
<td>19%</td>
<td>14%</td>
</tr>
</tbody>
</table>

True Positive Rate

1dFDR−U  1dFDR−A  1dFDR−H

2dFDR

A

B

<table>
<thead>
<tr>
<th>Weak effect</th>
<th>Moderate effect</th>
<th>Strong effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low density</td>
<td>3%</td>
<td>1%</td>
</tr>
<tr>
<td>Medium density</td>
<td>0%</td>
<td>1%</td>
</tr>
<tr>
<td>High density</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

"cor(x, z)"
Fig. 4

A

Number of discoveries vs. FDR cutoff for 2dFDR and 1dFDR-A.

B

Number of discoveries vs. FDR cutoff for 2dFDR and 1dFDR-A.

C

Comparison of 1dFDR-A and 2dFDR using log (Number of DMPs).

D

Improvement in number of datasets based on improvement percentage.
Fig. 5

A

DMP counts

EWAS26

EWAS27

EWAS30

EWAS39

EWAS45

p-value

B

DMP counts

EWAS29

EWAS28

C

Percentage of gDMPs recovered

1dFDR-A

2dFDR

Sample size
Supplementary Notes for “A new approach to confounder adjustment substantially increases detection power in omics association studies”
Contents

Note S1: Simulation Setup for Fig. 1a-c

Note S2: Full Method Description
2.1 Basic setup ......................................................... 4
2.2 Approximation of the false discovery proportion ......................... 5
2.3 Nonparametric empirical Bayes ..................................... 6
2.4 Two-dimensional Benjamini-Hochberg procedure ....................... 7

Note S3: Asymptotic FDR Control
3.1 Statement of the theorem ........................................ 10
3.2 Full proof of the theorem ......................................... 12

Note S4: Power analysis

3
Note S1: Simulation Setup for Fig. 1a-c

To generate Fig. 1a-c, we consider the following linear models:

\[ Y_i = X\alpha_i + Z\beta_i + e_i, \quad e_i \overset{i.i.d.}{\sim} N(0, \sigma_i^2 I_n), \quad 1 \leq i \leq m, \]  

where \( Y_i \in \mathbb{R}^{n \times 1} \) is the response vector, \( X = (X_1, \ldots, X_n)^\top \in \mathbb{R}^{n \times 1} \) is the covariate of interest, \( Z = (z_1, \ldots, z_n)^\top \in \mathbb{R}^{n \times 1} \) is the confounding factor, and \( \alpha_i \in \mathbb{R} \) and \( \beta_i \in \mathbb{R} \) are the parameters associated with the covariate and confounding factor, respectively.

The 300 independent simulation runs are conducted with \( n = 100 \) and \( m = 2000 \). We simulate \( X = (X_1, \cdots, X_{100})^\top, Z = (Z_1, \cdots, Z_{100})^\top \) by

\[ X_k \sim N(\rho \varepsilon_k^*, 1), \quad Z_k \sim N(\rho \varepsilon_k^*, 1), \quad k = 1, \ldots, n. \]

where \( \rho \in \{1.0, 1.5, 2.0\} \) controls the degree of confounding, and \( \varepsilon_k^* \overset{i.i.d.}{\sim} N(0, 1) \) is the random noise. We generate \( Y_i \in \mathbb{R}^{100 \times 1} \) under (1) with \( \alpha = (\alpha_1, \ldots, \alpha_m)^\top \) and \( \beta = (\beta_1, \ldots, \beta_m)^\top \) such that

- Category A : \( \alpha_i = 0.5, \quad \beta_i = 0 \) for \( i \in I_A \),
- Category B : \( \alpha_i = 0, \quad \beta_i = 0.5 \) for \( i \in I_B \),
- Category C : \( \alpha_i = 0.5, \quad \beta_i = 0.5 \) for \( i \in I_C \),
- Category D : \( \alpha_i = 0, \quad \beta_i = 0 \) for \( i \in I_D \),

where \( I_A, I_B, I_C, I_D \subset \{1, \ldots, m\} \) denote some mutually disjoint index sets (see illustration below).

We fix \( |I_A| = 67, |I_B| = 66, |I_C| = 66 \) and randomly generate these index sets for each simulation run. We compare the proposed two-dimensional FDR control procedure (2dFDR) to the one-dimensional FDR control procedure based on the confounder-adjusted model (1dFDR-A), and the power is evaluated using true positive rate.

<table>
<thead>
<tr>
<th></th>
<th>( I_A )</th>
<th>( I_C )</th>
<th>( I_B )</th>
<th>( I_D )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>( \square )</td>
<td>( \square )</td>
<td>( \square )</td>
<td></td>
</tr>
<tr>
<td>( \beta )</td>
<td>( \square )</td>
<td>( \square )</td>
<td>( \square )</td>
<td></td>
</tr>
</tbody>
</table>

\[ \square \] nonzero    \[ \square \] zero
Note S2: Full Method Description

**Notation.** For \( x, y \in \mathbb{R} \), let \( x \lor y = \max(x, y) \) and \( x \land y = \min(x, y) \). Let \( d^2_H(f, g) = (1/2) \int (\sqrt{f(x)} - \sqrt{g(x)})^2 dx \) be the square Hellinger distance between two densities \( f \) and \( g \). For a matrix \( C \), denote by \( P_C = C(C^\top C)^{-1}C^\top \) the projection matrix associated with the column space of \( C \) and define \( P_C^\perp = I - P_C \). Let \( \|C\|_2 \) and \( \|C\|_{\max} \) be the spectral norm and the elementwise maximum norm of \( C \), respectively. Denote by \( \lambda_{\min}(C) \) and \( \lambda_{\max}(C) \) the minimum and maximum eigenvalues of \( C \). Let \( \phi(\cdot) \) and \( \Phi(\cdot) \) be the probability density function and cumulative distribution function of the standard normal distribution. Denote by \( \chi^2_k \) the chi-square distribution with \( k \) degrees of freedom.

### 2.1 Basic setup

Consider the following linear models:

\[
Y_i = 1_{n \times 1}b + X\alpha_i + Z\beta_i + e_i, \quad e_i \overset{i.i.d.}{\sim} N(0, \sigma_i^2 I_n), \quad 1 \leq i \leq m, \tag{3}
\]

where \( Y_i \in \mathbb{R}^{n \times 1} \) is the response vector, \( X = (X_1, \ldots, X_n)^\top \in \mathbb{R}^{n \times 1} \) is the covariate of interest, \( Z = (z_1, \ldots, z_n)^\top \in \mathbb{R}^{n \times d} \) is the design matrix associated with the confounding factors, and \( \alpha_i \in \mathbb{R} \) and \( \beta_i = (\beta_{i1}, \ldots, \beta_{id})^\top \in \mathbb{R}^{d \times 1} \) are the parameters associated with the covariate and confounding factors respectively. By centering the response, covariate and confounding factors, we can assume without loss of generality that \( b = 0 \) throughout the following discussions.

Under (3), there are four different categories to consider

**A.** Solely associated with the variable of interest: \( \alpha_i \neq 0, \beta_i = 0 \);

**B.** Solely associated with the confounder: \( \alpha_i = 0, \beta_i \neq 0 \);

**C.** Associated with both the variable of interest and confounder: \( \alpha_i \neq 0, \beta_i \neq 0 \);

**D.** Not associated with either the variable of interest or confounder: \( \alpha_i = 0, \beta_i = 0 \).

The goal here is to develop a multiple testing procedure for simultaneously testing \( m \) hypotheses

\[
H_{0,i} : \alpha_i = 0 \quad \text{versus} \quad H_{a,i} : \alpha_i \neq 0, \quad i = 1, 2, \ldots, m,
\]

while adjusting for the confounding effects. We let \( \hat{\alpha}_i^A \) be the estimator of \( \alpha_i \) after adjusting for the confounding effect, and \( \hat{\alpha}_i^U \) be the unadjusted version without taking into account the confounding factors. Specifically, we have

\[
\hat{\alpha}_i^A = (X^\top P_Z^\perp X)^{-1}X^\top P_Z^\perp Y_i = \alpha_i + (X^\top P_Z^\perp X)^{-1}X^\top P_Z^\perp e_i,
\]

\[
\hat{\alpha}_i^U = (X^\top X)^{-1}X^\top Y_i = \alpha_i + (X^\top X)^{-1}X^\top Z\beta_i + (X^\top X)^{-1}X^\top e_i.
\]

Under (3), the estimator of the noise level \( \sigma_i^2 \) is given by

\[
\hat{\sigma}_i^2 = \frac{1}{n - d - 1}(Y_i - X\hat{\alpha}_i^A - Z\hat{\beta}_i)^\top (Y_i - X\hat{\alpha}_i^A - Z\hat{\beta}_i) = \frac{1}{n - d - 1}Y_i^\top P_Z^\perp Y_i,
\]

where \( \hat{\beta}_i = (Z^\top P_Z^\perp Z)^{-1}Z^\top P_Z^\perp Y_i \) and \( \mathbf{W} = (X, Z) \). Let \( \Omega = X^\top X/n, \Gamma = X^\top Z/n, \Omega_{X|Z} = X^\top P_Z^\perp X/n \) and \( \Omega_{Z|X} = Z^\top P_Z^\perp Z/n \). The adjusted and unadjusted \( z \)-statistics for testing \( H_{0,i} \) can
be defined as
\[ Z_i^A = \sqrt{n} \Omega_{X|Z}^{1/2} \hat{\alpha}_i / \hat{\sigma}_i = \sqrt{n} \Omega_{X|Z}^{1/2} \alpha_i / \hat{\sigma}_i + \Omega_{X|Z}^{-1/2} X^\top P_Z \hat{e}_i / (\sqrt{n} \hat{\sigma}_i), \]
\[ Z_i^U = \sqrt{n} \Omega_{X|Z}^{1/2} \hat{\alpha}_i / \hat{\sigma}_i = \sqrt{n} \Omega_{X|Z}^{1/2} \alpha_i / \hat{\sigma}_i + \sqrt{n} \Omega_{X|Z}^{-1/2} \Gamma \beta_i / \hat{\sigma}_i + \Omega_{X|Z}^{-1/2} X^\top \hat{e}_i / (\sqrt{n} \hat{\sigma}_i), \]
where we have used the variance estimator under model (3) for both statistics. Given the thresholds \( t_1, t_2 \geq 0 \), the two-dimensional procedure can be described as follows:

Dimension 1. Use the unadjusted statistics to determine a preliminary set of features \( D_1 = \{ 1 \leq i \leq m : |Z_i^U| \geq t_1 \} \).

Dimension 2. Reject \( H_{0,i} \) for \( |Z_i^A| \geq t_2 \) and \( i \in D_1 \). As a result, the final set of discoveries is given by \( D_2 = \{ 1 \leq i \leq m : |Z_i^U| \geq t_1, |Z_i^A| \geq t_2 \} \).

The basic idea of this procedure is to use the unadjusted statistics to screen out a large number of noises in Category D and further use the adjusted statistics to identify signals from Categories A and C. Although the unadjusted statistics are unable to distinguish the noise in Category B from the signals, they can preserve or even increase the signal strength. To see this, we note that
\[ |\Omega_{X|Z}^{1/2} \alpha_i| \geq |\Omega_{X|Z}^{1/2} \alpha_i|. \]
When \( \beta_i = 0 \), the unadjusted statistics can better preserve the signal strength comparing to the adjusted one.

### 2.2 Approximation of the false discovery proportion

Recall that \( t_i \) is the threshold in Dimension \( i \) for \( i = 1, 2 \). We propose a method to simultaneously select the two thresholds. Note that the \( i \)th hypothesis will be rejected if and only if
\[ |Z_i^U| \geq t_1 \quad \text{and} \quad |Z_i^A| \geq t_2. \]

In Dimension 2, all rejections from Categories B and D (\( \alpha_i = 0 \)) will be considered as false rejections/discoveries. Therefore, the false discovery proportion (FDP) is defined as
\[
\text{FDP}(t_1, t_2) = \frac{\sum_{i: \alpha_i \neq 0} 1 \{ |Z_i^U| \geq t_1, |Z_i^A| \geq t_2 \}}{\sum_{i=1}^m 1 \{ |Z_i^U| \geq t_1, |Z_i^A| \geq t_2 \}}
\]
\[
= \frac{\sum_{i: \alpha_i = 0} 1 \{ |\sqrt{n} \Omega_{X|Z}^{-1/2} \Gamma \beta_i / \hat{\sigma}_i + \Omega_{X|Z}^{-1/2} X^\top \hat{e}_i / (\sqrt{n} \hat{\sigma}_i) | \geq t_1, |\Omega_{X|Z}^{1/2} X^\top P_Z \hat{e}_i / (\sqrt{n} \hat{\sigma}_i) | \geq t_2 \}}{\sum_{i=1}^m 1 \{ |Z_i^U| \geq t_1, |Z_i^A| \geq t_2 \}}.
\]
(4)

Let \( (V_1, V_2) \) be the bivariate normal random variables such that
\[
\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \sim N \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \Omega_{X|Z}^{1/2} \Omega_{X|Z}^{-1/2} \\ \Omega_{X|Z}^{1/2} \Omega_{X|Z}^{-1/2} & 1 \end{pmatrix}.
\]
Replacing the numerator by the corresponding expectation (conditional on $X, Z$ and $\beta_i$’s) and $\hat{\sigma}_i$ by $\sigma_i$ in (4), we obtain

$$\text{FDP}(t_1, t_2) \approx \frac{\sum_{i=1}^m L(\mu_i, t_1, t_2)}{\sum_{i=1}^m 1 \{|Z_i^U| \geq t_1, |Z_i^A| \geq t_2\}} \leq \frac{\sum_{i=1}^m L(\mu_i, t_1, t_2)}{\sum_{i=1}^m 1 \{|Z_i^U| \geq t_1, |Z_i^A| \geq t_2\}},$$

(5)

where $\mu_i = \mu_{i,n} := \sqrt{n} \Omega^{-1/2} \Gamma \beta_i / \sigma_i$ and $L(\mu_i, t_1, t_2) = P(|\mu + V_1| \geq t_1, |V_2| \geq t_2 | \mu)$.

The major challenge here is the estimation of the expected number of false rejections given by $\sum_{i=1}^m L(\mu_i, t_1, t_2)$, which involves a large number of nuisance parameters $\mu_i$’s. A natural strategy is to estimate each $\mu_i$ separately by $\hat{\mu}_i$, and replace $L(\mu_i, t_1, t_2)$ by the plug-in estimate $L(\hat{\mu}_i, t_1, t_2)$. It seems natural to use the least squares estimator given by $\hat{\mu}_i = \sqrt{n} \Omega^{-1/2} \Gamma \beta_i / \sigma_i$. However, this method does not lead to a consistent estimation of the number of false rejections when there is non-negligible proportion of weak confounding factors.

To see this, we note that $\hat{\mu}_i$ approximately follows a normal distribution with mean $\mu_i$ and variance $A^2 = \Omega^{-1/2} \Gamma \Omega^{-1} \Gamma^T \Omega^{-1/2}$. Some algebra shows that

$$L(\mu_i, t_1, t_2) \approx P(|\mu_i + \tilde{V}_1| \geq t_1, |\tilde{V}_2| \geq t_2 | \mu_i) \neq L(\mu_i, t_1, t_2),$$

where

$$\begin{pmatrix} \tilde{V}_1 \\ \tilde{V}_2 \end{pmatrix} \sim N \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 + A^2 & \Omega^{-1/2} \Omega^{-1/2} \\ \Omega^{-1/2} \Omega^{-1/2} & 1 \end{pmatrix}.$$ 

(6)

Compared to the joint distribution of $(V_1, V_2)$, we see that the least squares estimator introduces extra variation to the first component of the bivariate normal distribution. We have also considered soft and hard thresholding estimators for $\mu_i$. The consistency of these regularized estimators requires a minimum signal assumption which again rules out the case of weak confounding factors.

To overcome the difficulty, we shall adopt a Bayesian viewpoint by assuming that $\mu_i$’s are generated from a common prior distribution. The Bayesian viewpoint allows us to borrow cross-sectional information (from different linear models) to estimate the number of false rejections without estimating individual $\mu_i$ explicitly.

### 2.3 Nonparametric empirical Bayes

In this subsection, we propose a nonparametric empirical Bayes approach to estimate the number of false rejections. Define

$$\hat{a} = \frac{\Omega^{-1/2} \Gamma}{\sqrt{\Omega^{-1/2} \Gamma \Omega^{-1} \Gamma^T \Omega^{-1/2}}}, \quad \hat{\xi}_i = \sqrt{n} \beta_i / \sigma_i, \quad \xi_i = \sqrt{n} \beta_i / \sigma_i.$$ 

(7)

Under (5) and conditional on $W = (X, Z)$, we have the Gaussian location model,

$$\hat{\eta}_i = \eta_i + \epsilon_i$$

(6)

where $\hat{\eta}_i = \hat{a}^T \hat{\xi}_i$ is an estimator for $\eta_i = \hat{a}^T \xi_i$ and $\epsilon_i \overset{i.i.d.}{\sim} N(0, 1)$.

Suppose $\xi_i$’s are independently generated from some distribution, see Assumption (5). Under

\footnote{A confounding factor is said to be weak if its coefficient $\beta_i$ decays to zero at the rate $n^{-1/2}$ or faster. In this case, $\limsup_{n \to +\infty} |\mu_i, n| = \delta_i \in [0, +\infty)$.}
suitable assumptions detailed below, we can show that \( \hat{a} \to a^{\text{a.s.}} \) for a vector \( a \) defined in equation \([11]\). Denote by \( G_0 \) the (prior) distribution for \( a^T \xi_i \). The goal here is to estimate \( G_0 \) based on \( \{\eta_i\} \). It will become clear later that how the estimate of \( G_0 \) is useful in estimating the expected number of false rejections. Following Kiefer and Wolfowitz (1956) and Jiang and Zhang (2009), we consider the general maximum likelihood estimator (GMLE) \( \hat{G}_{m,n} \) for \( G_n \) defined as

\[
\hat{G}_{m,n} = \arg\max_{\hat{G} \in \mathcal{G}} \sum_{i=1}^{m} \log f_G(\hat{\eta}_i)
\]

where \( \mathcal{G} \) denotes the set of all probability distributions on \( \mathbb{R} \) and \( f_G(x) = \int \phi(x - u) dG(u) \) is the convolution between \( G \) and \( \phi \). As \( \sigma_i \)'s are generally unknown in practice, \( \hat{G}_{m,n} \) is not obtainable. To obtain a feasible estimator, we consider the GMLE \( \tilde{G}_{m,n} \) defined as

\[
\tilde{G}_{m,n} = \arg\max_{\tilde{G} \in \mathcal{G}} \sum_{i=1}^{m} \log f_G(\tilde{\eta}_i)
\]

where \( \tilde{\eta}_i = a^T \xi_i \) for \( \xi_i = \sqrt{n} \tilde{\theta}_i / \tilde{\sigma}_i \). By the Carathéodory’s theorem, there exist discrete solutions to \([7]\) and \([8]\) with no more than \( m + 1 \) support points. Thus we can write the solutions as

\[
\hat{G}_{m,n}(u) = \sum_{j=1}^{m} \hat{\pi}_j 1 \{ s_j \leq u \}, \quad \tilde{G}_{m,n}(u) = \sum_{j=1}^{m} \tilde{\pi}_j 1 \{ \tilde{s}_j \leq u \}
\]

where \( \sum_{j=1}^{m} \hat{\pi}_j = \sum_{j=1}^{m} \tilde{\pi}_j = 1 \) for \( \hat{\pi}_j, \tilde{\pi}_j \geq 0 \), and \( \{ \hat{v}_1, \ldots, \hat{v}_m \} \) and \( \{ \tilde{v}_1, \ldots, \tilde{v}_m \} \) are two sets of support points for \( \hat{G}_{m,n} \) and \( \tilde{G}_{m,n} \), respectively. From the definitions of \( \hat{G}_{m,n} \) and \( f_{\hat{G}_{m,n}} \), the support of \( \hat{G}_{m,n}(u) \) is always within the range of \( \tilde{\eta}_i \) due to the monotonicity of \( \phi(x - u) \) in \( |x - u| \). Similarly, the support of \( \tilde{G}_{m,n}(u) \) is always within the range of \( \tilde{\eta}_i \). These observations would be useful for our theoretical analysis, see Section 3.2. It is also worth noting that the optimization in \([8]\) can be cast as convex optimization problem that can be efficiently solved by modern interior point methods. The readers are referred to Koenker and Mizera (2014) for more detailed discussions.

### 2.4 Two-dimensional Benjamini-Hochberg procedure

Given the feasible estimator \( \hat{G}_{m,n} \) of the prior distribution and in view of \([5]\), we consider an approximate upper bound for \( \text{FDP}(t_1, t_2) \) defined as

\[
\hat{\text{FDP}}(t_1, t_2) := \frac{\sum_{i=1}^{m} I \{ |Z_{i}^U| \geq t_1, |Z_{i}^A| \geq t_2 \}}{\sum_{i=1}^{m} 1 \{ |Z_{i}^U| \geq t_1, |Z_{i}^A| \geq t_2 \}} = \frac{m \{ L(Ax, t_1, t_2) d\hat{G}_{m,n}(x) \}}{\sum_{i=1}^{m} 1 \{ |Z_{i}^U| \geq t_1, |Z_{i}^A| \geq t_2 \}}.
\]

For a desired FDR level \( q \in (0, 1) \), we choose the optimal threshold such that

\[
(T_1^*, T_2^*) = \arg\max_{(t_1, t_2) \in \mathcal{F}_q} \sum_{i=1}^{m} 1 \{ |Z_{i}^U| \geq t_1, |Z_{i}^A| \geq t_2 \}
\]

where \( \mathcal{F}_q = \{(t_1, t_2) \in \mathbb{R}^+ \times \mathbb{R}^+ : \hat{\text{FDP}}(t_1, t_2) \leq q \} \) with \( \mathbb{R}^+ = (0, +\infty) \). This procedure can be viewed as a variant of the Benjamini-Hochberg (BH) procedure adapted to the two-dimensional approach introduced in Section 2.1.
Remark 1. The rejection region we consider is of the form

\[ \{(z^U, z^A) : |z^U| \geq t_1, |z^A| \geq t_2 \} . \]

In particular, if \( t_1 = 0 \), it reduces to the usual rejection region \( \{z^A : |z^A| \geq t_2 \} \) from the one-dimensional approach based on the adjusted statistics. Therefore, our approach produces at least as many rejections as the one-dimensional approach as we are searching over a larger class of rejection regions to maximize the number of discoveries.

It is well known that when the number of signals is a substantial proportion of the total number of hypotheses, the BH procedure will be overly conservative. To adapt to the proportion of signals, we develop a modification of John Storey’s approach [Storey (2002)] in our setting. To illustrate the idea, we assume that \( Z_i^A \) approximately follows the mixture model:

\[ \pi_0_i N(0, 1) + (1 - \pi_0_i) N(\mu_i^A, 1) \]

where \( \mu_i^A = \sqrt{n} \lambda_i |Z| \alpha_i / \sigma_i \) and \( \pi_0_i \) denotes the prior probability that \( \alpha_i = 0 \). Notice that

\[ P(|Z_i^A| \leq \lambda) = \pi_0_i (1 - 2 \Phi(-\lambda)) + (1 - \pi_0_i) P(|N(\mu_i^A, 1)| \leq \lambda) \approx \pi_0_i (1 - 2 \Phi(-\lambda)) , \]

provided that \( P(|N(\mu_i^A, 1)| \leq \lambda) \approx 0 \). Thus \( \left\{ |Z_i^A| \leq \lambda \right\} / \{1 - 2 \Phi(-\lambda)\} \) can be viewed as a conservative estimator for the mixing probability \( \pi_0_i \). We note

\[
\frac{1}{m} \sum_{i, \alpha_i = 0} L(\mu_i, t_1, t_2) = \frac{1}{m} \sum_{i = 1}^m 1(\alpha_i = 0) L(\mu_i, t_1, t_2) \\
\approx \int L(Ax, t_1, t_2) d\widetilde{G}_{m,n}(x) \frac{1}{m} \sum_{i = 1}^m \pi_0_i \\
\leq \int L(Ax, t_1, t_2) d\widetilde{G}_{m,n}(x) \frac{1}{m} \sum_{i = 1}^m \frac{P(|Z_i^A| \leq \lambda)}{1 - 2 \Phi(-\lambda)} \\
\approx \int L(Ax, t_1, t_2) d\widetilde{G}_{m,n}(x) \frac{1}{m} \sum_{i = 1}^m \frac{1 \{ |Z_i^A| \leq \lambda \}}{1 - 2 \Phi(-\lambda)} ,
\]

where to get the first approximation, we implicitly assume that the prior distribution of \( \mu_i \) remains the same regardless of whether \( \alpha_i \) is equal to zero or not. In view of the above derivation, we consider the FDR estimate given by

\[
\widetilde{\text{FDP}}_\lambda(t_1, t_2) = \frac{\int L(Ax, t_1, t_2) d\widetilde{G}_{m,n}(x) \sum_{i = 1}^m 1 \{ |Z_i^A| \leq \lambda \}}{(1 - 2 \Phi(-\lambda)) \sum_{i = 1}^m 1 \{ |Z_i^U| \geq t_1, |Z_i^A| \geq t_2 \}} ,
\]

where \( \lambda \) is a prespecified number as in John Storey’s approach. With this modification, for a desired FDR level \( q \in (0, 1) \), we choose the optimal threshold such that

\[
(T_1^*, T_2^*) = \arg\max_{(t_1, t_2) \in \mathcal{T}_q, \lambda} \sum_{i = 1}^m 1 \{ |Z_i^U| \geq t_1, |Z_i^A| \geq t_2 \} , \tag{10}
\]

We emphasize that the validity of our procedure does not rely on the mixture model assumption which is merely used to motivate John Storey’s procedure.
where

\[ \mathcal{F}_{q, \lambda} := \left\{ (t_1, t_2) \in \mathbb{R}^+ \times \mathbb{R}^+ : \widehat{\text{FDP}}_{\lambda}(t_1, t_2) \leq q \right\}. \]
3.1 Statement of the theorem

The two-dimensional procedure is shown to provide asymptotic FDR control under suitable assumptions. Denote by \(m_0\) and \(m_1\) the number of null and alternative hypotheses among the \(m\) hypotheses respectively. Let \(\hat{Z}_i^U\) and \(\hat{Z}_i^A\) be the z-statistics by replacing \(\hat{\sigma}_i\) with \(\hat{\sigma}_i\) in \(Z_i^U\) and \(Z_i^A\), respectively. Define \(L_0(\mu, t_1, t_2) = P(|\mu + \hat{V}_1| \geq t_1, |\hat{V}_2| \geq t_2)\mu\) where

\[
\begin{pmatrix} \hat{V}_1 \\ \hat{V}_2 \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \frac{1}{\sqrt{X'}[\Omega]^{-1/2}C_{X|Z}^{-1/2}} \\ \frac{1}{\sqrt{X'}[\Omega]^{-1/2}C_{X|Z}^{-1/2}} & 1 \end{pmatrix} \right)
\]

with \(C_{X|Z} = E[\Omega] - E[\Gamma]E[\Psi]^{-1}E[\Gamma]'\) for \(\Psi = Z'Z/n\). We also let \(A_0^2 = E[\Omega]^{-1/2}E[\Gamma]C_{Z'|X}^{-1}E[\Gamma]'E[\Omega]^{-1/2}\) and

\[
a^\top = \frac{E[\Omega]^{-1/2}E[\Gamma]}{\sqrt{E[\Omega]^{-1/2}E[\Gamma]C_{Z'|X}^{-1}E[\Gamma]'E[\Omega]^{-1/2}}}
\]

(11)

where \(C_{Z'|X} = E[\Psi] - (E[\Omega])^{-1}E[\Gamma]'E[\Gamma]\). We first introduce the following definitions and assumptions.

**Definition 1.** A random variable \(X \in \mathbb{R}\) is said to be sub-gaussian with the variance proxy \(\sigma^2\) if \(E[X] = 0\) and its moment generating function satisfies

\[
E[\exp(tX)] \leq \exp \left( \frac{\sigma^2 t^2}{2} \right) \text{ for any } t \in \mathbb{R}.
\]

**Definition 2.** A random variable \(X \in \mathbb{R}\) is said to be sub-exponential with the parameter \(\theta\) if \(E[X] = 0\) and its moment generating function satisfies

\[
E[\exp(tX)] \leq \exp \left( \frac{\theta^2 t^2}{2} \right) \text{ for any } |t| \leq \frac{1}{\theta}.
\]

**Assumption 1.** Suppose \(m_0/m \rightarrow \pi_0 \in (0, 1)\).

**Assumption 2.** Assume that

\[
\sum_{i: \alpha_i = 0} \mathbb{1} \left\{ |\hat{Z}_i^U| \geq t_1, |\hat{Z}_i^A| \geq t_2 \right\} \overset{a.s.}{\rightarrow} K_0(t_1, t_2),
\]

\[
\sum_{i: \alpha_i \neq 0} \mathbb{1} \left\{ |\hat{Z}_i^U| \geq t_1, |\hat{Z}_i^A| \geq t_2 \right\} \overset{a.s.}{\rightarrow} K_1(t_1, t_2),
\]

for every \((t_1, t_2) \in \mathbb{R}^+ \times \mathbb{R}^+\), where

\[
K_0(t_1, t_2) = E_{a^\top \xi} \left[ L_0(A_0 a^\top \xi, t_1, t_2) \right] \text{ for } a^\top \xi \sim G_0
\]

(12)

and \(K_0(t_1, t_2), K_1(t_1, t_2)\) are both non-negative continuous functions of the arguments \((t_1, t_2)\).
**Assumption 3.** Assume that

\[
\lambda_{\min}(C_{Z|X}) > 0, \quad \lambda_{\min}(\mathbb{E}[\Psi]) > 0, \quad \mathbb{E}[X_1^2] > 0, \\
0 < \min_{1 \leq i \leq m} \sigma_i \leq \max_{1 \leq i \leq m'} \sigma_i < \infty.
\]

**Assumption 4.** Assume that the components of \( Z \) and \( X \) are both sub-Gaussian.

**Assumption 5.** Assume that \( \{\xi_i\}_{i=1}^d \) is a sequence of i.i.d. continuous random vectors with the density \( h \) whose support set is given by \( \{x \in \mathbb{R}^d, \|x\|_{\text{max}} \leq B(\log m)^b\} \) for some \( B, b \geq 0 \).

**Assumption 6.** Assume \( m = m(n) \) such that \( m(n) \to +\infty \) as \( n \to +\infty \) and \( \limsup_{n \to +\infty} m(n) < \infty \) for some \( p_0 > 0 \).

Assumption 1 requires the asymptotic null proportion to be strictly between zero and one. Assumption 2 allows certain forms of dependence, such as \( m \)-dependence, ergodic dependence and certain mixing type dependence. A justification for equation (12) is given in Corollary 1. Assumption 3 implies that \( C_{Z|X}^{-1} \) exists and the noise level is uniformly bounded from below and above. Assumption 4 allows us to use concentration inequalities for sub-Gaussian and sub-exponential random variables in the proofs. Assumption 5 implies that \( G_0 \) has a bounded support that expands slowly with \( m \). From Assumption 6, the number of features \( m \) is some function of \( n \) and \( m \) is allowed to be polynomially larger than the sample size \( n \).

**Remark 2.** The assumption that \( \xi_i \) has a density is merely used to simplify the presentation of the proof of Lemma 9. When \( \xi_i \) has a discrete distribution, the proof of Lemma 9 can be modified to obtain a similar result. We omit the details here to conserve the space.

Before stating the main result, we introduce the following lemma which establishes the uniform convergence of \( \int L(Ax, t_1, t_2)d\tilde{G}_{m,n}(x) \).

**Lemma 1.** Let \( \tilde{G}_{m,n} \) be the estimator of \( G_0 \) as defined in (8). Under Assumptions 3, 4, for any \( t'_1, t'_2 > 0 \), we have

\[
\sup_{t_1 \leq t'_1, t_2 \leq t'_2} \left| \int L(Ax, t_1, t_2)d\tilde{G}_{m,n}(x) - \int L_0(A_0x, t_1, t_2)dG_0(x) \right| \overset{a.s.}{\longrightarrow} 0.
\]

Let

\[
V_m(t_1, t_2) = \sum_{i: \alpha_i > 0} 1 \{ |Z_i^U| \geq t_1, |Z_i^A| \geq t_2 \}, \quad S_m(t_1, t_2) = \sum_{i: \alpha_i \neq 0} 1 \{ |Z_i^U| \geq t_1, |Z_i^A| \geq t_2 \},
\]

\[
F_m(\lambda) = \frac{1}{m} \sum_{i=1}^m 1 \{ |Z_i^A| \leq \lambda \}, \quad F(\lambda) = \pi_0(1 - 2\Phi(-\lambda)) + (1 - \pi_0)(1 - K_1(0, \lambda)).
\]

The following lemma shows the almost sure convergence as in Assumption 2 with \( (\tilde{Z}_i^U, \tilde{Z}_i^A) \) replaced by \( (Z_i, Z_i) \).

**Lemma 2.** Under Assumptions 2, 3 and 6, we have

\[
\frac{1}{m_0} V_m(t_1, t_2) \overset{a.s.}{\longrightarrow} K_0(t_1, t_2), \quad \frac{1}{m_1} S_m(t_1, t_2) \overset{a.s.}{\longrightarrow} K_1(t_1, t_2), \quad F_m(\lambda) \overset{a.s.}{\longrightarrow} F(\lambda).
\]
Recall that
\[ \tilde{FDP}_\lambda(t_1, t_2) = \int L(Ax, t_1, t_2) d\tilde{G}_{m,n}(x) \sum_{i=1}^m 1 \left\{ |Z_i^A| \leq \lambda \right\} \frac{(1 - 2\Phi(-\lambda)) \sum_{i=1}^m 1 \left\{ |Z_i^U| \geq t_1, |Z_i^A| \geq t_2 \right\}}{(1 - 2\Phi(-\lambda)) K(t_1, t_2)} \]
and \((\tilde{T}_1^\ast, \tilde{T}_2^\ast)\) is the optimal threshold as defined in [10]. Define
\[ \tilde{FDP}_\infty^\lambda(t_1, t_2) := E_{a^\top \xi} \left[ L_0(A_0 a^\top \xi, t_1, t_2) \left\{ \pi_0(1 - 2\Phi(-\lambda)) + (1 - \pi_0)(1 - K_1(0, \lambda)) \right\} \right], \] (13)
where \(K(t_1, t_2) = \pi_0 K_0(t_1, t_2) + (1 - \pi_0) K_1(t_1, t_2)\). By Lemmas 1-2 it follows that
\[ \tilde{FDP}_\lambda(t_1, t_2) \xrightarrow{a.s.} \tilde{FDP}_\infty^\lambda(t_1, t_2). \]

We impose the following assumption to reduce the searching region for \((t_1, t_2)\) to a rectangle of the form \([0, t_1^\ast] \times [0, t_2^\ast]\).

**Assumption 7.** Assume that there exist \(t_1^\ast\) and \(t_2^\ast\) such that \(FDP_\infty^\lambda(t_1^\ast, 0) < q, FDP_\infty^\lambda(0, t_2^\ast) < q\) and \(K(t_1^\ast, t_2^\ast) > 0\).

Let
\[ \tilde{FDR}_m = E \left[ \frac{V_m(\tilde{T}_1^\ast, \tilde{T}_2^\ast)}{V_m(\tilde{T}_1^\ast, \tilde{T}_2^\ast) + S_m(\tilde{T}_1^\ast, \tilde{T}_2^\ast)} \right]. \]

We show the asymptotic FDR control in the following theorem.

**Theorem 1.** Under Assumptions 1-7, we have
\[ \limsup_m \tilde{FDR}_m \leq q. \]

### 3.2 Full proof of the theorem

The relationship among the theoretical results is depicted below.

![Diagram](http://example.com/diagram.png)

We first introduce some concentration inequalities for a later use.
Lemma 3. Under Assumptions \([3, 4]\), for \(x_0 > 0\), we have

\[
\begin{align*}
\mathbb{P}(\|\Omega - \mathbb{E}[\Omega]\| > x_0) & \leq C_0 \exp\left\{ -c_0 n(x_0^2/c'_0 x_0/c'_0) \right\}, \\
\mathbb{P}(\|\Psi - \mathbb{E}[\Psi]\|_{\max} > x_0) & \leq C_1 \exp\left\{ -c_1 n(x_0^2/c'_1 x_0/c'_1) \right\}, \\
\mathbb{P}(\|\Gamma - \mathbb{E}[\Gamma]\|_{\max} > x_0) & \leq C_2 \exp\left\{ -c_2 n(x_0^2/c'_2 x_0/c'_2) \right\},
\end{align*}
\]

and, for \(0 < x_1 < 3(1 + D_0^2/D_2 + D_3^2 E[\Omega]^{-1})\),

\[
\mathbb{P}(\|\Omega_{Z|X} - C_{Z|X}\|_{\max} > x_1) \leq C_3 \exp\left\{ -c_3 n(x_1^2/c'_3 x_1/c'_3) \right\},
\]

and, for \(0 < x_2 < L^{-1}(3 + D_0^2/D_2 + D_0^2 E[\Omega]^{-1}) \wedge \left( 1 + \frac{3w_0^{-1/2}}{(1\wedge D_0^{-1})} \wedge (D_1 d)^{-1}) := L' \right) \),

\[
\mathbb{P}(\|\hat{a} - a\|_{\max} > x_2) \leq C_4 \exp\left\{ -c_4 n(x_2^2/c'_4 x_2/c'_4) \right\},
\]

where \(C_1, c_1, c'_1, c''_1 > 0\) denote some absolute constants for \(l = 0, \cdots, 4\) and, for \(w_0 = \mathbb{E}[\Gamma] C_{z|X} \mathbb{E}[\Gamma]^{-1}\),

\[
L = \left( \frac{w_0^{3/2}(1 \wedge D_0^{-1})}{6 \left\{ d^2(1 + 2D_0)D_1 + D_0^2 D_2 d^3 \right\} + D_1 dw_0^{3/2}(1 \wedge D_0^{-1}) \wedge 1} \right)
\]

and \(D_0, D_1, D_2 > 0\) are some constants such that \(\|\mathbb{E}[\Gamma]\|_{\max} \leq D_0, \|C_{z|X}\|_2 \leq D_1\) and \(\mathbb{E}[X_1^2] \geq D_2\). As a consequence of the above concentration inequalities, we have

\[
\Omega_{z|X}^{-1} \overset{a.s.}{\to} C_{z|X}^{-1}, \quad \Omega_{X|Z} \overset{a.s.}{\to} C_{X|Z}, \quad A \overset{a.s.}{\to} A_0.
\]

Proof of Lemma \([3]\) Each element of \(\Psi - \mathbb{E}[\Psi], \Gamma - \mathbb{E}[\Gamma]\) and \(\Omega - \mathbb{E}[\Omega]\) is sub-exponential. The first three inequalities thus follow from the union bound and the tail bound for sum of sub-exponential random variables, see e.g. Corollary 5.17 of Vershynin (2010). For \([14]\), some algebra gives us

\[
\begin{align*}
\|\Omega_{Z|X} - C_{Z|X}\|_{\max} & \leq \|\Psi - \mathbb{E}[\Psi]\|_{\max} + \{\|\Omega^{-1} - \mathbb{E}[\Omega]^{-1}\| + D_2^{-1}\} \left\{ \|\Gamma - \mathbb{E}[\Gamma]\|_2^2 + 2D_0 \|\Gamma - \mathbb{E}[\Gamma]\|_{\max} \right\} \\
& \quad + D_0^2 \|\Omega^{-1} - \mathbb{E}[\Omega]^{-1}\|.
\end{align*}
\]

For \(0 < x_1 < 3(1 + D_0^2/D_2 + D_0^2 E[\Omega]^{-1})\), the following inclusion of events can be verified

\[
\begin{align*}
\{\|\Omega_{Z|X} - C_{Z|X}\|_{\max} > x_1\} & \subseteq \left\{\|\Psi - \mathbb{E}[\Psi]\|_{\max} > \frac{x_1}{3}\right\} \cup \left\{\|\Omega^{-1} - \mathbb{E}[\Omega]^{-1}\| \geq \frac{x_1}{3D_0^2}\right\} \cup \left\{\|\Gamma - \mathbb{E}[\Gamma]\|_{\max} > \left( \frac{D_2}{2(1 + 2D_0)} \wedge 1 \right) \frac{x_1}{3}\right\} \\
& \subseteq \left\{\|\Psi - \mathbb{E}[\Psi]\|_{\max} > \frac{x_1}{3}\right\} \cup \left\{\|\Omega^{-1} - \mathbb{E}[\Omega]^{-1}\| \geq \frac{x_1}{3D_0^2}\right\} \cup \left\{\|\Gamma - \mathbb{E}[\Gamma]\|_{\max} > \left( \frac{D_2}{2(1 + 2D_0)} \wedge 1 \right) \frac{x_1}{3}\right\}.
\end{align*}
\]

The first inclusion follows because, conditional on the events such that

\[
\left\{\|\Psi - \mathbb{E}[\Psi]\|_{\max} \leq \frac{x_1}{3}\right\} \cap \left\{\|\Omega^{-1} - \mathbb{E}[\Omega]^{-1}\| \leq \frac{x_1}{3D_0^2}\right\} \cap \left\{\|\Gamma - \mathbb{E}[\Gamma]\|_{\max} \leq \left( \frac{D_2}{2(1 + 2D_0)} \wedge 1 \right) \frac{x_1}{3}\right\},
\]
\[ (16) \text{ implies that} \]
\[ \| \Omega_{Z|X} - C_{Z|X} \|_{\text{max}} \leq \frac{x_1}{3} + \left( \frac{x_1}{3D_0^2} + D_2^{-1} \right) \left( \frac{D_2}{2(1 + 2D_0) \wedge 1} \right)^2 \frac{x_1^2}{3^2} + 2D_0 \left( \frac{D_2}{2(1 + 2D_0) \wedge 1} \right) \frac{x_1}{3} \]
\[ + \frac{x_1^2}{3D_0^2} \]
\[ \leq \frac{x_1}{3} + 2D_2^{-1}(1 + 2D_0) \left( \frac{D_2}{2(1 + 2D_0) \wedge 1} \right) \frac{x_1}{3} + \frac{x_1}{3} \leq x_1. \]

As \( x_1 < 3D_0^2\mathbb{E}[\Omega]^{-1} \) and \( \mathbb{E}[\Omega] = \mathbb{E}[X_1^2] \geq D_2 \), the second inclusion holds by noticing that
\begin{equation}
\begin{align*}
\left\{ |\Omega^{-1} - \mathbb{E}[\Omega]^{-1}| > \frac{x_1}{3D_0^2} \right\} &= \left\{ |\Omega - \mathbb{E}[\Omega]| > \frac{|\Omega\mathbb{E}[\Omega]| x_1}{3D_0^2} \right\} \\
&\subset \left\{ |\Omega - \mathbb{E}[\Omega]| > |\Omega\mathbb{E}[\Omega]| \frac{x_1}{3D_0^2} \right\} \cup \left\{ |\Omega\mathbb{E}[\Omega]| \leq \frac{E[\Omega]^2}{2} \right\} \\
&\subset \left\{ |\Omega - \mathbb{E}[\Omega]| > \frac{E[\Omega]^2 x_1}{6D_0^2} \right\} \cup \left\{ |\Omega - \mathbb{E}[\Omega]| > \frac{E[\Omega]}{2} \right\} \subset \left\{ |\Omega - \mathbb{E}[\Omega]| > \frac{D_2^2 x_1}{6D_0^2} \right\}.
\end{align*}
\end{equation}

As for the last inequality, we first observe that
\begin{equation}
\begin{align*}
\| \hat{a} - a \|_{\text{max}} &= \left\| \frac{\Gamma}{\sqrt{w}} - \frac{\mathbb{E}[\Gamma]}{\sqrt{w_0}} \right\|_{\text{max}} \\
&\leq \left| \frac{1}{\sqrt{w}} - \frac{1}{\sqrt{w_0}} \right| \left\{ \| \Gamma - \mathbb{E}[\Gamma] \|_{\text{max}} + |\mathbb{E}[\Gamma]| \right\} + \frac{1}{\sqrt{w_0}} \| \Gamma - \mathbb{E}[\Gamma] \|_{\text{max}} 
\end{align*}
\end{equation}
where \( w = \Gamma \Omega_{Z|X}^{-1} \Gamma^\top \) and \( w_0 = \mathbb{E}[\Gamma] C_{Z|X}^{-1} \mathbb{E}[\Gamma]^\top \). It also follows that, for \( 0 < x'' < w_0^{-1/2} \),
\begin{equation}
\begin{align*}
\left\{ \left| \frac{1}{\sqrt{w}} - \frac{1}{\sqrt{w_0}} \right| > x'' \right\} &\subset \left\{ \left| \sqrt{w} - \sqrt{w_0} \right| > \frac{w_0}{2} x'' \right\} \\
&= \left\{ \left| \frac{\sqrt{w}}{\sqrt{w_0}} - 1 \right| > \frac{\sqrt{w_0}}{2} x'' \right\} \\
&\subset \left\{ \left| \frac{w - w_0}{w_0} - 1 \right| > \frac{\sqrt{w_0}}{2} x'' \right\} = \left\{ |w - w_0| > \frac{w_0^{3/2} x''}{2} \right\}.
\end{align*}
\end{equation}

Combining (18) with (19) implies that, for \( 0 < x' < \left( 3 \wedge \frac{3w_0^{-1/2}}{\left( 1 \wedge D_0^{-1} \right)} \right) \),
\begin{equation}
\begin{align*}
\left\{ \| \hat{a} - a \|_{\text{max}} > x' \right\} &\subset \left\{ \left| \frac{1}{\sqrt{w}} - \frac{1}{\sqrt{w_0}} \right| > \frac{\sqrt{x'}}{\sqrt{3}} \right\} \cup \left\{ \| \Gamma - \mathbb{E}[\Gamma] \|_{\text{max}} > \sqrt{\frac{x'}{3}} \right\} \\
&\cup \left\{ \left| \frac{1}{\sqrt{w}} - \frac{1}{\sqrt{w_0}} \right| \| \mathbb{E}[\Gamma] \|_{\text{max}} > \frac{x'}{3} \right\} \cup \left\{ \frac{1}{\sqrt{w_0}} \| \Gamma - \mathbb{E}[\Gamma] \|_{\text{max}} > \frac{x'}{3} \right\} \\
&\subset \left\{ \left| \frac{1}{\sqrt{w}} - \frac{1}{\sqrt{w_0}} \right| > \frac{x'}{3} \left( 1 \wedge D_0^{-1} \right) \right\} \cup \left\{ \| \Gamma - \mathbb{E}[\Gamma] \|_{\text{max}} > \frac{x'}{3} \left( 1 \wedge \sqrt{w_0} \right) \right\} \\
&\subset \left\{ |w - w_0| > \frac{x'}{6} w_0^{3/2} \left( 1 \wedge D_0^{-1} \right) \right\} \cup \left\{ \| \Gamma - \mathbb{E}[\Gamma] \|_{\text{max}} > \frac{x'}{3} \left( 1 \wedge \sqrt{w_0} \right) \right\}. \quad (20)
\end{align*}
\end{equation}
Thus, we need to derive the tail bound of $|w - w_0|$. Note that

$$
|w - w_0| \leq |\Omega_{Z|X}^{-1} \Gamma - \mathbb{E}[\Gamma] \Omega_{Z|X}^{-1} \mathbb{E}[\Gamma]^\top + \mathbb{E}[\Gamma] \Omega_{Z|X}^{-1} \mathbb{E}[\Gamma]^\top - \mathbb{E}[\Gamma] C_{Z|X}^{-1} \mathbb{E}[\Gamma]^\top | \\
\leq \frac{\|\Gamma - \mathbb{E}[\Gamma]\|_2^2 \|\Omega_{Z|X}^{-1} - C_{Z|X}^{-1}\|_2}{(*)}
$$

and

$$(*) = \left| (\Gamma - \mathbb{E}[\Gamma]) \left( \Omega_{Z|X}^{-1} - C_{Z|X}^{-1} + C_{Z|X}^{-1} \right) \left( \Gamma - \mathbb{E}[\Gamma] + 2\mathbb{E}[\Gamma] \right) \right|^\top \\
\leq \left\{ \|\Gamma - \mathbb{E}[\Gamma]\|_2^2 + 2\|\Gamma - \mathbb{E}[\Gamma]\|_2\|\mathbb{E}[\Gamma]\|_2 \right\} \left\{ \|\Omega_{Z|X}^{-1} - C_{Z|X}^{-1}\|_2 + \|C_{Z|X}^{-1}\|_2 \right\},
$$

which implies that

$$
|w - w_0| \leq d^2 \left\{ \|\Gamma - \mathbb{E}[\Gamma]\|_2^2 + 2D_0\|\Gamma - \mathbb{E}[\Gamma]\|_2 \right\} \left\{ \|\Omega_{Z|X}^{-1} - C_{Z|X}^{-1}\|_2 + D_1 \right\} + D_0^2d^2 \|\Omega_{Z|X}^{-1} - C_{Z|X}^{-1}\|_2. 
$$

We have

$$
\|\Omega_{Z|X}^{-1} - C_{Z|X}^{-1}\|_2 = \|\Omega_{Z|X}^{-1} (C_{Z|X} - \Omega_{Z|X}) C_{Z|X}^{-1}\|_2 \\
\leq d \|\Omega_{Z|X} - C_{Z|X}\|_{\text{max}} \|\Omega_{Z|X}^{-1}\|_2 \|C_{Z|X}^{-1}\|_2 \\
\leq D_1^2d \|\Omega_{Z|X} - C_{Z|X}\|_{\text{max}} \left\{ 1 - D_1d \|\Omega_{Z|X} - C_{Z|X}\|_{\text{max}} \right\}^{-1}.
$$

The last inequality holds when $\|\Omega_{Z|X} - C_{Z|X}\|_{\text{max}} < (D_1d)^{-1}$ due to the fact that

$$
\|\Omega_{Z|X}^{-1}\|_2 \leq \|\Omega_{Z|X}^{-1} - C_{Z|X}^{-1}\|_2 + \|C_{Z|X}^{-1}\|_2 \\
\leq \|\Omega_{Z|X}^{-1}\|_2 \|C_{Z|X}^{-1}\|_2 \|C_{Z|X} - \Omega_{Z|X}\|_2 + \|C_{Z|X}^{-1}\|_2 \\
\leq D_1d \|\Omega_{Z|X}^{-1}\|_2 \|\Omega_{Z|X} - C_{Z|X}\|_{\text{max}} + \|C_{Z|X}^{-1}\|_2,
$$

which is equivalent to $\|\Omega_{Z|X}^{-1}\|_2 \left\{ 1 - D_1d \|\Omega_{Z|X} - C_{Z|X}\|_{\text{max}} \right\} \leq \|C_{Z|X}^{-1}\|_2$. For $0 < x < \left( 1 \wedge (D_1d)^{-1} \right)$ where $\kappa = d^2(1 + 2D_0)D_1 + D_0^2D_1^2d^3$ and

$$
L = \left\{ \frac{w_0^{3/2} \left( 1 \wedge D_0^{-1} \right)}{6\kappa + D_1dw_0^{3/2} \left( 1 \wedge D_0^{-1} \right)} \wedge 1 \right\},
$$

we have

$$
\left\{ \|\Omega_{Z|X} - C_{Z|X}\|_{\text{max}} \leq Lx \right\} \cap \left\{ \|\Gamma - \mathbb{E}[\Gamma]\|_{\text{max}} \leq Lx \right\} \subset \left\{ |w - w_0| \leq \frac{w_0^{3/2} \left( 1 \wedge D_0^{-1} \right)x}{6} \right\} 
$$

because, conditional on the two events $\{ \|\Omega_{Z|X} - C_{Z|X}\|_{\text{max}} \leq Lx \}$ and $\{ \|\Gamma - \mathbb{E}[\Gamma]\|_{\text{max}} \leq Lx \}$.
and (22) provide
\[ |w - w_0| \leq d^2 \left( L^2 x^2 + 2 D_0 Lx \right) \left( \frac{D_1^2 dLx}{1 - D_1 dLx} + D_1 \right) + D_0^2 \frac{D_1^2 dLx}{1 - D_1 dLx} \]
\[ \leq d^2 (1 + 2 D_0) Lx \left( \frac{D_1}{1 - D_1 dLx} \right) + \frac{D_0^2 D_1^2 d^2 Lx}{1 - D_1 dLx} = \frac{\kappa Lx}{6} \]
\[ \leq \frac{\kappa w_0^{3/2}}{6} \left( 1 \wedge D_0^{-1} \right) x \]

By (20) and (23), for \( 0 < x_2 < L^{-1} \left( 3 \wedge 3 D_0^2 / D_2 \wedge 3 D_0^2 E[\Omega]^{-1} \right) \wedge \left( 1 \wedge \frac{3 w_0^{-1/2}}{(\wedge D_0^{-1})} \wedge (D_1 d)^{-1} \right) \), we have
\[ \{ \| \hat{a} - a \|_{\max} > x_2 \} \subset \{ \| \Omega_{Z|X} - C_{Z|X} \|_{\max} > L x_2 \} \bigcup \{ \| \Gamma - E[\Gamma] \|_{\max} > \left( \frac{1 \wedge \sqrt{w_0}}{3} \wedge L \right) x_2 \}, \]
which completes the proof by applying the union bound together with the tail bounds of \( \| \Omega_{Z|X} - C_{Z|X} \|_{\max} \) and \( \| \Gamma - E[\Gamma] \|_{\max} \).

A direct implication of the exponential tail bounds is
\[ \Omega \overset{a.s.}{\rightarrow} E[\Omega], \quad \Gamma \overset{a.s.}{\rightarrow} E[\Gamma], \quad \Omega_{Z|X} \overset{a.s.}{\rightarrow} C_{Z|X} \] (24)
by the Borel-Cantelli lemma. We next show that
\[ \Omega_{Z|X}^{-1} \overset{a.s.}{\rightarrow} C_{Z|X}^{-1}, \quad \Omega_{X|Z} \overset{a.s.}{\rightarrow} C_{X|Z}. \] (25)

Since \( \Omega_{Z|X} \overset{a.s.}{\rightarrow} C_{Z|X} \), we have \( \Omega_{Z|X}^{-1} \overset{a.s.}{\rightarrow} C_{Z|X}^{-1} \) by (22). Similarly, it can be shown that \( \Omega_{X|Z} \overset{a.s.}{\rightarrow} C_{X|Z} \) under the assumption that \( \lambda_{\min}(E[\Psi]) > 0 \). Thus, by the continuous mapping theorem together with (24) and (25), we conclude that \( A \overset{a.s.}{\rightarrow} A_0 \). \( \diamond \)

The following lemma shows the strong uniform consistency of \( \hat{\sigma}_i^2 \) and the tail bound for \( \hat{\sigma}_i^2 / \sigma_i^2 \).

**Lemma 4.** Under Assumptions \( \mathcal{A} \) and \( \mathcal{B} \)

\[ \max_{1 \leq i \leq m} \left| \frac{\hat{\sigma}_i^2}{\sigma_i^2} - 1 \right| \overset{a.s.}{\rightarrow} 0 \] (26)

and, for \( 0 < \delta_1 < (n - d - 1) / \log m \),
\[ P \left( \max_{1 \leq i \leq m} \left| \frac{\hat{\sigma}_i^2}{\sigma_i^2} - 1 \right| > \nu_1 \sqrt{\delta_1} \right| W \right) \leq 2 m \exp \{ -\delta_1 (\log m) \} \] (27)

where \( \nu_1 = 4 \sqrt{\log m / (n - d - 1)} \).

**Proof of Lemma 4** Since \( \hat{\sigma}_i^2 / \sigma_i^2 \ | W \sim \chi_n^2 / (n - d - 1) \), by the tail bound for chi-square random variables as in Lemma 1 of Laurent and Massart (2000), we have
\[ P \left( \left| \frac{\hat{\sigma}_i^2}{\sigma_i^2} - 1 \right| > 2 \sqrt{\frac{\delta_0}{n - d - 1}} + 2 \frac{\delta_0}{n - d - 1} \right| W \right) \leq 2 \exp(-\delta_0) \]
for $\delta_0 > 0$. By the union bound, we have
\[
\mathbb{P} \left( \max_{1 \leq i \leq m} \left| \frac{\hat{\sigma}_i^2}{\sigma_i^2} - 1 \right| > 2 \sqrt{\frac{\delta_0}{n-d-1}} \right) \leq 2m \exp(-\delta_0).
\]

Letting $\delta_0 = (n-d-1)\delta'_0$ for $0 < \delta'_0 < 1$,
\[
\mathbb{P} \left( \max_{1 \leq i \leq m} \left| \frac{\hat{\sigma}_i^2}{\sigma_i^2} - 1 \right| > 2 \sqrt{\delta'_0 + 2\delta'_0} \right) \leq \mathbb{P} \left( \max_{1 \leq i \leq m} \left| \frac{\hat{\sigma}_i^2}{\sigma_i^2} - 1 \right| > 4\delta'_0 \right) \leq 2m \exp(-(n-d-1)\delta'_0).
\]

Thus, under Assumptions 3 and 6, (26) follows by the Borel-Cantelli Lemma because
\[
\text{Lemma 3, we have}
\]

Under model (6), we have
\[
\text{for any } x > 0 \text{ by the well known result about the tail bound of a standard normal random variable.}
\]

We next derive the tail bounds for $\max_{1 \leq i \leq m} |\hat{\eta}_i|$ and $\max_{1 \leq i \leq m} |\tilde{\eta}_i|$.

**Lemma 5.** Under Assumptions 3,5, for $0 < \delta_1 < (n-d-1)/(16 \log m)$, $\delta_2 > 0$ and $x_2$ as in Lemma 3, we have
\[
\mathbb{P} \left( \max_{1 \leq i \leq m} |\hat{\eta}_i| > \nu_2(1 + \delta_2) \right) \leq C_4m \exp \left\{ -c_4n(x_2^2/c'_4 \wedge x_2/c'_4) \right\} + 2m \exp \left\{ -\nu_2^2\delta_2^2 \right\},
\]
\[
\mathbb{P} \left( \max_{1 \leq i \leq m} |\tilde{\eta}_i| > \nu_2 \left( 1 - \nu_1 \sqrt{\delta_1} \right)^{-1} (1 + \delta_2) \right) \leq 2m \exp \left\{ -\delta_1(\log m) \right\} + C_4m \exp \left\{ -c_4n(x_2^2/c'_4 \wedge x_2/c'_4) \right\} + 2m \exp \left\{ -\nu_2^2\delta_2^2 \right\},
\]
where $\nu_1 = 4\sqrt{\log m/(n-d-1)}$ and $\nu_2 = (d + ||a||_1)B(\log m)^{\nu/2}$.

**Proof of Lemma 5.** Note that
\[
\mathbb{P}(|\epsilon_i| \leq x) \geq 1 - 2\exp(-x^2)
\]
for any $x > 0$ by the well known result about the tail bound of a standard normal random variable. Under model 6, we have
\[
\mathbb{P} (|\tilde{\eta}_i| \leq \nu_2(1 + \delta_2)) \geq \mathbb{P} \left( |\eta_i| \leq \nu_2 \right) \cap \left\{ |\epsilon_i| \leq \nu_2\delta_2 \right\}
\]
\[
\geq \mathbb{P} (|\eta_i| \leq \nu_2) + \mathbb{P} (|\epsilon_i| \leq \nu_2\delta_2) - 1
\]
\[
\geq \mathbb{P} (|\eta_i| \leq \nu_2, ||\hat{\alpha} - a||_{\text{max}} \leq x_2) - 2\exp\left\{-\nu_2^2\delta_2^2\right\}
\]
\[
= \mathbb{P} (||\hat{\alpha} - a||_{\text{max}} \leq x_2) - 2\exp\left\{-\nu_2^2\delta_2^2\right\}
\]
\[
\geq 1 - C_4 \exp \left\{ -c_4n(x_2^2/c'_4 \wedge x_2/c'_4) \right\} - 2\exp\left\{-\nu_2^2\delta_2^2\right\}
\]
where the third inequality follows by choosing \( x = \nu_2 \delta_2 \) for \( \delta_2 > 0 \) in (28) and the equality holds by Lemma 3 because
\[
|\eta_i| = |\hat{a}^\top \xi_i| \leq \|\hat{a}\|_1 \|\xi_i\|_{\max} \leq (\|\hat{a} - a\|_1 + \|a\|_1)B(\log m)^b \\
\leq (d\|\hat{a} - a\|_{\max} + \|a\|_1)B(\log m)^b \\
\leq (dx_2 + \|a\|_1)B(\log m)^b \leq (d + \|a\|_1)B(\log m)^b = \nu_2
\]
conditional on the event \( \{\|\hat{a} - a\|_{\max} \leq x_2\} \) for \( x_2 \) as in Lemma 3 which is smaller than 1. Thus, it follows that
\[
P \left( \max_{1 \leq i \leq m} |\hat{\eta}_i| \leq \nu_2(1 + \delta_2) \right) = P \left( \bigcap_{i=1}^{m} \{ |\hat{\eta}_i| \leq \nu_2(1 + \delta_2) \} \right) \\
\geq 1 - C_4 m \exp \left\{ -c_4 n(x_2^2/c_4' \wedge x_2/c_4'') \right\} - 2m \exp \{ -\nu_2^2 \delta_2^2 \}.
\]
For the second inequality, we first observe that
\[
|\hat{\sigma}_i/\sigma_i - 1| \leq \left| \frac{\hat{\sigma}_i^2}{\sigma_i^2} - 1 \right| (30)
\]
and, provided that \( |\hat{\sigma}_i/\sigma_i - 1| < 1 \),
\[
|\hat{\eta}_i| = \left| \frac{\sigma_i}{\sigma_i} - 1 + 1 \right| |\hat{\eta}_i| \leq \left| \frac{|\hat{\sigma}_i/\sigma_i - 1|}{1 - |\hat{\sigma}_i/\sigma_i - 1|} \right| |\hat{\eta}_i| + |\hat{\eta}_i| = \frac{|\hat{\eta}_i|}{1 - |\hat{\sigma}_i/\sigma_i - 1|}.
\]
Thus, we have
\[
P \left( \max_{1 \leq i \leq m} |\hat{\eta}_i| \leq \nu_2 \left( 1 - \nu_1 \sqrt{\delta_1} \right)^{-1} (1 + \delta_2) \right) \\
\geq P \left( \bigcap_{1 \leq i \leq m} \left\{ \frac{\hat{\sigma}_i}{\sigma_i} - 1 \leq \nu_1 \sqrt{\delta_1} \right\} \right) \\
\geq P \left( \bigcap_{1 \leq i \leq m} \left\{ \frac{\hat{\sigma}_i^2}{\sigma_i^2} - 1 \leq \nu_1 \sqrt{\delta_1} \right\} \right) \\
\geq 1 - 2m \exp \{ -\delta_1 (\log m) \} - C_4 m \exp \left\{ -c_4 n(x_2^2/c_4' \wedge x_2/c_4'') \right\} - 2m \exp \{ -\nu_2^2 \delta_2^2 \}
\]
where the last inequality follows due to
\[
P \left( \max_{1 \leq i \leq m} \left| \frac{\hat{\sigma}_i^2}{\sigma_i^2} - 1 \right| \leq \nu_1 \sqrt{\delta_1} \right) = \mathbb{E}_W \left[ P \left( \max_{1 \leq i \leq m} \left| \frac{\hat{\sigma}_i^2}{\sigma_i^2} - 1 \right| \leq \nu_1 \sqrt{\delta_1} \right| W \right) \right] \\
\geq 1 - 2m \exp \{ -\delta_1 (\log m) \}
\]
by (27).

We derive the concentration inequalities for \( m^{-1} \sum_{i=1}^{m} \hat{\eta}_i \) and \( m^{-1} \sum_{i=1}^{m} \hat{\eta}_i^2 \) in the following lemma.

\textbf{Lemma 6.} Under Assumptions 3, 5, for \( \delta_3 > 0, 0 < \delta_4 < 2^5 \nu_3 \) and \( x_2 \) as in Lemma 3 we have
\[
P \left( \left| \frac{1}{m} \sum_{i=1}^{m} \hat{\eta}_i - \mathbb{E} [\eta_i] \right| > \delta_3 \right) \leq 2 \exp \left\{ -\frac{m \delta_3^2}{2 \nu_3} \right\} + C_4 \exp \left\{ -c_4 n(x_2^2/c_4' \wedge x_2/c_4'') \right\},
\]
and
\[
\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^{m} \hat{\eta}_i^2 - \mathbb{E}[\hat{\eta}_1^2] - 1\right| > \delta_4\right) \leq 2 \exp\left\{ - \frac{m\delta_4^2}{2^{11}\nu_3^2}\right\} + 2 \exp\left\{ - \frac{m\delta_4^2}{32(d + \|a\|_1)^2B^2(\log m)^{2b}\nu_3}\right\} \\
+ 3C_4 \exp\left\{ - c_4n(x_2^2/c'_4 \land x_2/c''_4)\right\},
\]

where \(\nu_3 = 4(d + \|a\|_1)^2B^2(\log m)^{2b} + 1\).

**Proof of Lemma 6** We have
\[
\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^{m} \hat{\eta}_i - \mathbb{E}[\eta_1]\right| > \delta_3\right) \leq \mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^{m} \hat{\eta}_i - \mathbb{E}[\eta_1]\right| > \delta_3, \|\hat{a} - a\|_{\text{max}} \leq x_2\right) \\
+ \mathbb{P}\left(\|\hat{a} - a\|_{\text{max}} > x_2\right).
\]

Since the model 6 can be rewritten as
\[
\hat{\eta}_i - \mathbb{E}[\eta_1] = \eta_i - \mathbb{E}[\eta_1] + \epsilon_i
\]
and, by (29), \(\|\eta_i - \mathbb{E}[\eta_1]\| \leq 2(d + \|a\|_1)B(\log m)^b\) conditional on the event \(\{\|\hat{a} - a\|_{\text{max}} \leq x_2\}\), it can be shown that \((\eta_i - \mathbb{E}[\eta_1])'s\) are sub-gaussian with the variance proxy \(4(d + \|a\|_1)^2B^2(\log m)^{2b}\). Thus, \((\hat{\eta}_i - \mathbb{E}[\eta_1])'s\) are also sub-gaussian with the variance proxy \(\nu_3\) as \(\epsilon_i's\) are sub-gaussian with the variance proxy one and \(\eta_i's\) and \(\epsilon_i's\) are independent. Then, the first inequality follows from Lemma 4 and Corollary 1.7 in Rigollet and Hütter (2015).

For the second inequality, we observe that
\[
(\hat{\eta}_i - \mathbb{E}[\eta_1])^2 - \mathbb{E}[(\hat{\eta}_i - \mathbb{E}[\eta_1])^2] = \hat{\eta}_i^2 - \mathbb{E}[\eta_1^2] - 2\mathbb{E}[\eta_1](\hat{\eta}_i - \mathbb{E}[\eta_1]) \\
= \hat{\eta}_i^2 - \mathbb{E}[\eta_1^2] - 1 - 2\mathbb{E}[\eta_1](\hat{\eta}_i - \mathbb{E}[\eta_1])
\]
under model 6. By Lemma 1.12 in Rigollet and Hütter (2015), it follows that conditional on the event \(\{\|\hat{a} - a\|_{\text{max}} \leq x_2\}\),
\[
\hat{\eta}_i^2 - \mathbb{E}[\eta_1^2] - 1 - 2\mathbb{E}[\eta_1](\hat{\eta}_i - \mathbb{E}[\eta_1])
\]
is sub-exponential with the parameter \(16\nu_3\). According to Theorem 1.13 in Rigollet and Hütter (2015),
\[
\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^{m} \hat{\eta}_i^2 - \mathbb{E}[\eta_1^2] - 1 - 2\mathbb{E}[\eta_1]\left(\frac{1}{m}\sum_{i=1}^{m} \hat{\eta}_i - \mathbb{E}[\eta_1]\right)\right| > \frac{\delta_4}{2}\right) \\
\leq \mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^{m} \hat{\eta}_i - \mathbb{E}[\eta_1]\right| > \frac{\delta_4}{2}, \|\hat{a} - a\|_{\text{max}} \leq x_2\right) \\
+ \mathbb{P}\left(\|\hat{a} - a\|_{\text{max}} > x_2\right) \\
\leq 2 \exp\left\{ - \frac{m}{2} \left( \frac{\delta_4^2}{2^{10}\nu_3^2} \land \frac{\delta_4^2}{2^3\nu_3}\right) \right\} + C_4 \exp\left\{ - c_4n(x_2^2/c'_4 \land x_2/c''_4)\right\}
\]
for any $\delta_4 > 0$. We have

$$
P \left( \left| \frac{1}{m} \sum_{i=1}^{m} \hat{\eta}^2_i - \mathbb{E}[\eta_i^2] \right| - 1 - 2\mathbb{E}[\eta_i] \left( \frac{1}{m} \sum_{i=1}^{m} \hat{\eta}_i - \mathbb{E}[\eta_i] \right) > \frac{\delta_4}{2} \right)
$$

\[ \geq \mathbb{P} \left( \left| \frac{1}{m} \sum_{i=1}^{m} \hat{\eta}_i - \mathbb{E}[\eta_i] \right| > \delta_4 \right) + \mathbb{P} \left( \left| 2\mathbb{E}[\eta_i] \left( \frac{1}{m} \sum_{i=1}^{m} \hat{\eta}_i - \mathbb{E}[\eta_i] \right) \right| \leq \frac{\delta_4}{2} \right) - 1
\]

\[ \geq \mathbb{P} \left( \left| \frac{1}{m} \sum_{i=1}^{m} \hat{\eta}_i - \mathbb{E}[\eta_i] \right| > \delta_4 \right) + \mathbb{P} \left( \left| \left( \frac{1}{m} \sum_{i=1}^{m} \hat{\eta}_i - \mathbb{E}[\eta_i] \right) \right| \leq \frac{\delta_4}{4\mathbb{E}[\eta_i]}, \| \hat{a} - a \|_{\text{max}} \leq x_2 \right) - 1
\]

and

$$
\mathbb{P} \left( \left| \frac{1}{m} \sum_{i=1}^{m} \hat{\eta}_i - \mathbb{E}[\eta_i] \right| \leq \frac{\delta_4}{4\mathbb{E}[\eta_i]}, \| \hat{a} - a \|_{\text{max}} \leq x_2 \right)
$$

\[ \geq \mathbb{P} \left( \left| \frac{1}{m} \sum_{i=1}^{m} \hat{\eta}_i - \mathbb{E}[\eta_i] \right| \leq \frac{\delta_4}{4(d + \| a \|_1) B \log m}, \| \hat{a} - a \|_{\text{max}} \leq x_2 \right) (\because (29))
\]

\[ \geq 1 - 2 \exp \left\{ - \frac{m \delta_4^2}{32(d + \| a \|_1)^2 B^2 \log m} \right\} - 2C_4 \exp \left\{ -c_4 n(x_2/c_4' \wedge x_2/c_4'') \right\}.
\]

The proof is completed by letting $\delta_3 = \delta_4/(4(d + \| a \|_1) B \log m)$ in (32). \hfill \Box

**Lemma 7.** For $0 < \delta_1 < (n - d - 1)/(16 \log m)$, $\delta_2, \delta_3 > 0$ and $0 < \delta_4 \leq 2^5 \nu_3$ and $x_2$ as in Lemma 3 under Assumptions 3, 5, we have

$$
P \left( \left| \frac{1}{m} \sum_{i=1}^{m} \log \frac{f_{\hat{G}_{m,n}}(\hat{\eta}_i)}{f_{\hat{G}_{m,n}}(\eta_i)} \right| > \frac{\nu_1 \sqrt{\delta_1}}{1 - \nu_1 \sqrt{\delta_1}} \left( \frac{\mathbb{E}[\eta_i^2] + 1 + \delta_1}{2} + \nu_2(1 + \delta_2) \{ \mathbb{E}[\eta_i] + \delta_3 \} \right) \right)
$$

\[ \leq 2m \exp \left\{ - \nu_2^2 \delta_2^2 \right\} + 2m \exp \left\{ - \delta_1(\log m) \right\} + 2 \exp \left\{ - \frac{m \delta_3^2}{2 \nu_3} \right\} + 2 \exp \left\{ - \frac{m \delta_4^2}{211 \nu_3^2} \right\}
\]

\[ + 2 \exp \left\{ - \frac{m \delta_4^2}{32(d + \| a \|_1)^2 B^2 \log m} \right\} + (m + 4)C_4 \exp \left\{ -c_4 n(x_2/c_4' \wedge x_2/c_4'') \right\}
\]

and

$$
P \left( \left| \frac{1}{m} \sum_{i=1}^{m} \log \frac{f_{\hat{G}_{m,n}}(\hat{\eta}_i)}{f_{\hat{G}_{m,n}}(\eta_i)} \right| > \frac{\nu_1 \sqrt{\delta_1}}{1 - \nu_1 \sqrt{\delta_1}} \left( \frac{\mathbb{E}[\eta_i^2] + 1 + \delta_1}{2} + \nu_2(1 + \delta_2) \{ \mathbb{E}[\eta_i] + \delta_3 \} \right) \right)
$$

\[ \leq 2m \exp \left\{ - \nu_2^2 \delta_2^2 \right\} + 4m \exp \left\{ - \delta_1(\log m) \right\} + 2 \exp \left\{ - \frac{m \delta_3^2}{2 \nu_3} \right\} + 2 \exp \left\{ - \frac{m \delta_4^2}{211 \nu_3^2} \right\}
\]

\[ + 2 \exp \left\{ - \frac{m \delta_4^2}{32(d + \| a \|_1)^2 B^2 \log m} \right\} + (m + 4)C_4 \exp \left\{ -c_4 n(x_2/c_4' \wedge x_2/c_4'') \right\}
\]

where $\nu_1 = 4 \sqrt{\log m/(n - d - 1)}$, $\nu_2 = (d + \| a \|_1) B(\log m)^{b+\frac{1}{2}}$ and $\nu_3 = B^2(\log m)^{2b} + 1$. 

20
**Proof of Lemma 7** When $|\hat{\sigma}_i/\sigma_i - 1| < 1$, by (30), we have

$$
\left| \frac{1}{m} \sum_{i=1}^{m} (\hat{n}_i^2 - n_i^2) \right| \leq \max_{1 \leq i \leq m} \left| \frac{\hat{n}_i^2}{\sigma_i^2} - 1 \right| \left| \frac{1}{m} \sum_{i=1}^{m} n_i^2 \right| \leq \frac{\max_{1 \leq i \leq m} \left| \frac{\hat{n}_i^2}{\sigma_i^2} - 1 \right|}{1 - \max_{1 \leq i \leq m} \left| \frac{\hat{n}_i^2}{\sigma_i^2} - 1 \right|} \left| \frac{1}{m} \sum_{i=1}^{m} \hat{n}_i^2 \right| ,
$$

$$
\left| \frac{1}{m} \sum_{i=1}^{m} (\hat{n}_i - n_i) \right| \leq \frac{\max_{1 \leq i \leq m} \left| \frac{\hat{n}_i}{\sigma_i} - 1 \right|}{1 - \max_{1 \leq i \leq m} \left| \frac{\hat{n}_i}{\sigma_i} - 1 \right|} \left| \frac{1}{m} \sum_{i=1}^{m} \hat{n}_i \right| \leq \frac{\max_{1 \leq i \leq m} \left| \frac{\hat{n}_i}{\sigma_i} - 1 \right|}{1 - \max_{1 \leq i \leq m} \left| \frac{\hat{n}_i}{\sigma_i} - 1 \right|} \left| \frac{1}{m} \sum_{i=1}^{m} \hat{n}_i \right| .
$$

Recall that $\hat{G}_{m,n}(u) = \sum_{j=1}^{m} \hat{\pi}_j 1 \{ \hat{v}_j \leq u \}$ denotes the solution of (7), where $\hat{\pi}_j \geq 0$ with $\sum_{j=1}^{m} \hat{\pi}_j = 1$ and $\{ \hat{v}_1, \ldots, \hat{v}_m \}$ is the set of supporting points for $\hat{G}_{m,n}$. It follows that

$$
\left| \frac{1}{m} \sum_{i=1}^{m} \log f_{\hat{G}_{m,n}}(\hat{n}_i) - \frac{1}{m} \sum_{i=1}^{m} \log f_{\hat{G}_{m,n}}(\hat{n}_i) \right| = \left| \frac{1}{m} \sum_{i=1}^{m} \log \left[ \exp \left\{ -(\hat{n}_i^2 - n_i^2)/2 \right\} \sum_{j=1}^{m} \exp \left\{ \hat{n}_i \hat{v}_j + (\hat{n}_i - \hat{n}_i) \hat{v}_j - \hat{v}_j^2/2 \right\} \hat{\pi}_j \right] / \sum_{j=1}^{m} \exp \left\{ \hat{n}_i \hat{v}_j - \hat{v}_j^2/2 \right\} \hat{\pi}_j \right| \leq \frac{1}{m} \sum_{i=1}^{m} \left| \frac{\hat{n}_i^2}{\sigma_i^2} - 1 \right| \frac{1}{m} \sum_{i=1}^{m} \hat{n}_i^2 \right| + \max_{1 \leq i \leq m} |\hat{n}_i| \frac{\max_{1 \leq i \leq m} \left| \frac{\hat{n}_i^2}{\sigma_i^2} - 1 \right|}{1 - \max_{1 \leq i \leq m} \left| \frac{\hat{n}_i^2}{\sigma_i^2} - 1 \right|} \left| \frac{1}{m} \sum_{i=1}^{m} \hat{n}_i \right| ,
$$

where the second inequality follows from the fact that the support of $\hat{G}_{m,n}(u)$ is always within the range of $\hat{n}_i$ as noticed in Section 2.3 Let

$$
U := \frac{\max_{1 \leq i \leq m} \left| \frac{\hat{n}_i^2}{\sigma_i^2} - 1 \right|}{2 \left( 1 - \max_{1 \leq i \leq m} \left| \frac{\hat{n}_i^2}{\sigma_i^2} - 1 \right| \right)} \left| \frac{1}{m} \sum_{i=1}^{m} \hat{n}_i^2 \right| \max_{1 \leq i \leq m} |\hat{n}_i| \frac{\max_{1 \leq i \leq m} \left| \frac{\hat{n}_i^2}{\sigma_i^2} - 1 \right|}{1 - \max_{1 \leq i \leq m} \left| \frac{\hat{n}_i^2}{\sigma_i^2} - 1 \right|} \left| \frac{1}{m} \sum_{i=1}^{m} \hat{n}_i \right| ,
$$

$$
u_{m,n} := \frac{\nu_1 \sqrt{\delta_1}}{1 - \nu_1 \sqrt{\delta_1}} \left( \frac{\mathbb{E}[\hat{n}_i^2]}{2} + 1 + \delta_4 \right) + \nu_2 (1 + \delta_2) (\mathbb{E}[\hat{n}_i] + \delta_3) .
$$

We have the following inclusions of the events

$$
\left\{ \max_{1 \leq i \leq m} |\hat{n}_i| \leq \nu_2 (1 + \delta_2) \right\} \cap \left\{ \max_{1 \leq i \leq m} \left| \frac{\hat{n}_i^2}{\sigma_i^2} - 1 \right| \leq \nu_1 \sqrt{\delta_1} \right\} \cap \left\{ \frac{1}{m} \sum_{i=1}^{m} \hat{n}_i - \mathbb{E}[\hat{n}_i] \leq \delta_3 \right\} \cap \left\{ \frac{1}{m} \sum_{i=1}^{m} \hat{n}_i^2 - \mathbb{E}[\hat{n}_i^2] - 1 \leq \delta_4 \right\} \subseteq \left\{ \max_{1 \leq i \leq m} |\hat{n}_i| \leq \nu_2 (1 + \delta_2) \right\} \cap \left\{ \max_{1 \leq i \leq m} \left| \frac{\hat{n}_i^2}{\sigma_i^2} - 1 \right| \leq \nu_1 \sqrt{\delta_1} \right\} \cap \left\{ \frac{1}{m} \sum_{i=1}^{m} \hat{n}_i \leq \mathbb{E}[\hat{n}_i] + \delta_3 \right\} \cap \left\{ \frac{1}{m} \sum_{i=1}^{m} \hat{n}_i^2 \leq \mathbb{E}[\hat{n}_i^2] + 1 + \delta_4 \right\} \subseteq \{ U \leq \nu_{m,n} \} .
$$
Thus, by (33)-(34), we have

\[
\mathbb{P}\left(\left|\frac{1}{m} \sum_{i=1}^{m} \log \frac{f_{\tilde{G}_{m,n}}(\tilde{\eta}_i)}{f_{G_n}(\tilde{\eta}_i)}\right| \leq u_{m,n}\right) \geq \mathbb{P}(U \leq u_{m,n})
\]

\[
\geq \mathbb{P}\left(\max_{1 \leq i \leq m} |\tilde{\eta}_i| \leq \nu_2(1 + \delta_2)\right) + \mathbb{P}\left(\max_{1 \leq i \leq m} \left|\frac{\sigma_i^2}{\sigma_1^2} - 1\right| \leq \nu_1 \sqrt{\delta_1}\right)
\]

\[
+ \mathbb{P}\left(\left|\frac{1}{m} \sum_{i=1}^{m} \tilde{\eta}_i - \mathbb{E}[\eta_1]\right| \leq \delta_3\right) + \mathbb{P}\left(\left|\frac{1}{m} \sum_{i=1}^{m} \tilde{\eta}_i^2 - \mathbb{E}[\eta_1^2] - 1\right| \leq \delta_4\right) - 3,
\]

which completes the proof of the first inequality by Lemmas 5-6 and (31). Similar argument can be used to verify the second inequality.

We introduce the large deviation inequality for \(d_H(f_{\tilde{G}_{m,n}}, f_{G_n})\) which can be proved using the arguments for Theorem 1 of Zhang (2009) and (29).

**Lemma 8.** Under Assumptions 3-5 and the event \(\{\|\tilde{a} - a\|_{\max} \leq x_2\}\) for \(x_2\) as in Lemma 3, \(G_{m,n}\) satisfies

\[
\prod_{i=1}^{m} \left\{ \frac{f_{\tilde{G}_{m,n}}(\tilde{\eta}_i)}{f_{G_n}(\tilde{\eta}_i)} \right\} \geq e^{-2t^2m^2c_m/15} \tag{35}
\]

where

\[
c_m = \sqrt{\log(m)} \left[ m^{1/p} \sqrt{\log m} \left\{ (d + \|a\|_1)B \vee 1 \right\} (\log m)^p \right]^{2/(2 - 2p)} \tag{36}
\]

for some \(p > 0\), then there exists an universal constant \(t^*\) such that for all \(t \geq t^*\) and \(\log m \geq 4/p\),

\[
\mathbb{P}\left(d_H(f_{\tilde{G}_{m,n}}, f_{G_n}) \geq tc_m\right) \leq \exp\left\{ -\frac{t^2m^2c_m^2}{2\log m} \right\} \leq e^{-t^2\log m}.
\]

The following lemma shows that \(d_H(f_{\tilde{G}_{m,n}}, f_{G_n}) = o_{a.s.}(1)\).

**Lemma 9.** Under Assumptions 3-6, \(d_H(f_{\tilde{G}_{m,n}}, f_{G_n}) = o_{a.s.}(1)\).

**Proof of Lemma 9.** Define \(d_H := d_H(f_{\tilde{G}_{m,n}}, f_{G_n})\) and note that \(d_H\) is indexed by both \(n\) and \(m\). Since \(d_H\) is indexed by only \(n\) under Assumption 6, for any \(\varepsilon > 0\), it suffices to show that \(\sum_{n=1}^{\infty} \mathbb{P}(d_H \geq \varepsilon) < \infty\) by the Borel-Cantelli Lemma.

Under Assumption 6, there exists a \(p > 0\) such that \((p + 1)/(2p) > p_0\). Define \(c_m\) as in (36) based on such \(p\). For the same \(x_2\) in Lemma 3 and large enough \(n\), it follows that

\[
\mathbb{P}(d_H \geq \varepsilon) \leq \mathbb{P}(d_H \geq tc_m) \leq \mathbb{P}(d_H \geq tc_m, \|\tilde{a} - a\|_{\max} \leq x_2) + \mathbb{P}(\|\tilde{a} - a\|_{\max} > x_2) \tag{37}
\]

where the first inequality follows because \(c_m\) can be made arbitrarily small for large enough \(n\) under
for some constant $d_H \geq t c_m, \| \hat{a} - a \|_{\text{max}} \leq x_2$)

$P \left( d_H \geq t c_m, \| \hat{a} - a \|_{\text{max}} \leq x_2, \frac{1}{m} \sum_{i=1}^{m} \log \frac{f_{\hat{G}_{m,n}}(\hat{\eta}_i)}{f_{G_n}(\hat{\eta}_i)} \geq - \frac{2t^2 c_m^2}{15} \right)$

$= \mathbb{E}_W \left[ P \left( d_H \geq t c_m, \frac{1}{m} \sum_{i=1}^{m} \log \frac{f_{\hat{G}_{m,n}}(\hat{\eta}_i)}{f_{G_n}(\hat{\eta}_i)} \geq - \frac{2t^2 c_m^2}{15} \mid W \right) \right] \leq D_3 m^{-2}$

for some constant $D_3 > 0$ and large enough $t$. Moreover, we have

$P \left( \frac{1}{m} \sum_{i=1}^{m} \log \frac{f_{\hat{G}_{m,n}}(\hat{\eta}_i)}{f_{G_n}(\hat{\eta}_i)} < - \frac{2t^2 c_m^2}{15}, \| \hat{a} - a \|_{\text{max}} \leq x_2 \right)$

$= P \left( \frac{1}{m} \sum_{i=1}^{m} \log \frac{f_{\hat{G}_{m,n}}(\hat{\eta}_i)}{f_{G_n}(\hat{\eta}_i)} \leq \frac{1}{m} \sum_{i=1}^{m} \log \frac{f_{\hat{G}_{m,n}}(\hat{\eta}_i)}{f_{G_n}(\hat{\eta}_i)} - \frac{1}{m} \sum_{i=1}^{m} \log \frac{f_{\hat{G}_{m,n}}(\hat{\eta}_i)}{f_{G_n}(\hat{\eta}_i)} \right.$

$- \frac{1}{m} \sum_{i=1}^{m} \log \frac{f_{\hat{G}_{m,n}}(\hat{\eta}_i)}{f_{G_n}(\hat{\eta}_i)} < - \frac{2t^2 c_m^2}{15}, \| \hat{a} - a \|_{\text{max}} \leq x_2 \left. \right)$

$\leq P \left( \frac{2t^2 c_m^2}{15} \leq \frac{1}{m} \sum_{i=1}^{m} \log \frac{f_{\hat{G}_{m,n}}(\hat{\eta}_i)}{f_{G_{m,n}}(\hat{\eta}_i)} \mid \frac{1}{m} \sum_{i=1}^{m} \log \frac{f_{\hat{G}_{m,n}}(\hat{\eta}_i)}{f_{G_{m,n}}(\hat{\eta}_i)} > \| \hat{a} - a \|_{\text{max}} \leq x_2 \right)$

$\leq P \left( 2u_{m,n} \leq \frac{1}{m} \sum_{i=1}^{m} \log \frac{f_{\hat{G}_{m,n}}(\hat{\eta}_i)}{f_{G_{m,n}}(\hat{\eta}_i)} \mid \| \hat{a} - a \|_{\text{max}} \leq x_2 \right)$

$\leq P \left( u_{m,n} \leq \frac{1}{m} \sum_{i=1}^{m} \log \frac{f_{\hat{G}_{m,n}}(\hat{\eta}_i)}{f_{G_{m,n}}(\hat{\eta}_i)} \right) + P \left( u_{m,n} \leq \frac{1}{m} \sum_{i=1}^{m} \log \frac{f_{\hat{G}_{m,n}}(\hat{\eta}_i)}{f_{G_{m,n}}(\hat{\eta}_i)} \right)$

where the second inequality follows due to the fact that, conditional on the event $\{ \| \hat{a} - a \|_{\text{max}} \leq x_2 \}$,

$u_{m,n} = \frac{\nu_1^2 \delta_1}{1 - \nu_1^2 \delta_1} \left( \mathbb{E}[\eta_i^2] + 1 + \delta_2 \right) + \frac{1}{2} + \nu_2 (1 + \delta_2) (||\mathbb{E}[\eta_1]|| + \delta_3) = O \left( \frac{(\log m)^{b+\frac{1}{2}+\frac{b}{2}}}{\sqrt{n}} \right)$

by (29) and

$\frac{2t^2 c_m^2}{15} \leq \frac{2t^2}{15} \frac{1}{m^p/(1+p)} \geq \frac{2t^2}{15} \frac{(\log m)^{b+\frac{1}{2}+\frac{b}{2}}}{\sqrt{n}} \geq \frac{2t^2}{15} \frac{u_{m,n}}{\sqrt{n}} \geq 2u_{m,n}$

for large enough $n$ and $t \geq (t^* \vee \sqrt{15D_4D_5})$, where $D_4, D_5 \geq 0$ denote some constants such that
\[ u_{m,n} \leq D_4 \frac{(\log m)^{b + \frac{1}{2} + (6\sqrt{d})}}{\sqrt{n}} \] and

\[ m^{p/(1+p)} = m^{1/(2p_0) - \epsilon} \leq D_5 \frac{\sqrt{n}}{(\log m)^{b + \frac{1}{2} + (6\sqrt{d})}} \]

for \( \epsilon = 1/(2p_0) - p/(1 + p) > 0 \) under Assumptions 5.6. Thus, for large enough \( n \), combining (37) together with (38) and (39) implies that

\[ \prod_{\epsilon} \text{for } \epsilon \leq u \]

Lemma 10.

Under Assumptions .3-.6, we have

Proof of Lemma 10

We first note that

\[ \sum \text{side can be shown to be finite with the choice of } \delta \]

by Lemma 7. Then, under Assumption .6, each infinite series with respect to \( n \) on the right hand side can be shown to be finite with the choice of \( \delta_1 > 2 \) and \( \delta_2 > \sqrt{2}/(d + \|a\|_1)B \). This implies

\[ \sum_{n=1}^{\infty} \mathbb{P}(d_H \geq \epsilon) < \infty, \text{ which completes the proof.} \]

We next show \( d_H (f_{G_n}, f_{G_0}) = o_{a.s.}(1) \) in the following lemma.

Lemma 10. Under Assumptions .3-.6, we have \( d_H (f_{G_n}, f_{G_0}) = o_{a.s.}(1) \).

Proof of Lemma 10

We first note that \( d_H^2 (f_{G_n}, f_{G_0}) \leq d_{TV} (f_{G_n}, f_{G_0}) \) where \( d_{TV} \) denotes the total variation distance such that

\[ d_{TV} (f_{G_n}, f_{G_0}) = \frac{1}{2} \int \left| \int \phi(x - y) \left\{ dP(\widehat{a}^\top \xi_i \leq y|\mathbf{W}) - dP(a^\top \xi_i \leq y) \right\} \right| dx. \]

The main idea of the proof is to find an upper bound of \( d_{TV} (f_{G_n}, f_{G_0}) \) in terms of \( \|\widehat{a} - a\|_{\text{max}} \).

Let \( a = (a_1, \ldots, a_d)^\top \). Without loss of generality, we suppose the first \( l \) elements of \( a \) are zero and the rest is non-zero for some \( 0 \leq l \leq d \). Then, the density of \( a^\top \xi \) can be written as

\[ f_{a^\top \xi}(y) = \begin{cases} \frac{1}{|a_d|} \int h \left( z_1, \ldots, z_{d-1}, \frac{y - \sum_{i=l+1}^{d-1} a_i z_i}{\delta_d} \right) dz_1 \cdots dz_{d-1}, & \text{for } l < d, \\ \delta_y = 0, & \text{for } l = d, \end{cases} \]

where \( \sum_{i=l+1}^{d-1} a_i z_i \) is defined to be 0 if \( l = d - 1 \). We prove the result when \( l < d \) as it is straightforward if \( l = d \). When \( l < d \), the joint density of \( (a_1 \xi_1, \ldots, a_d \xi_d) \) is

\[ g_a(z) = \left( \prod_{i=1}^{l} \delta_{z_i = 0} \right) \frac{1}{\prod_{i=l+1}^{d} |a_i|} h_{-l} \left( z_{l+1}/a_{l+1}, \ldots, z_d/a_d \right) \]

where \( \prod_{i=1}^{l} \delta_{z_i = 0} \) is defined to be 1 if \( l = 0 \) and \( h_{-l} \) denotes the density of \( (\xi_{l+1}, \ldots, \xi_d) \). Then, the
density of $a^\top \xi$ is

\[
f_{a^\top \xi}(y) = \frac{1}{\prod_{i=l+1}^{d} a_i} \int h_{-l} \left( \frac{z_{l+1}}{a_{l+1}}, \ldots, \frac{y - \sum_{i=l+1}^{d-1} a_i z_i}{a_d} \right) dz_{l+1} \ldots dz_{d-1}
\]

\[
= \frac{1}{|a_d|} \int h_{-l} \left( \frac{z_{l+1}, \ldots, z_{d-1}, y - \sum_{i=l+1}^{d-1} a_i z_i}{a_d} \right) dz_{l+1} \ldots dz_{d-1}
\]

\[
= \frac{1}{|a_d|} \int h \left( \frac{z_1, \ldots, z_{d-1}, y - \sum_{i=l+1}^{d-1} a_i z_i}{a_d} \right) dz_1 \ldots dz_{d-1}.
\]

Suppose the first $k$ elements of $\hat{a}$ are zero and the rest are non-zero for some $0 \leq k \leq d$. Similarly, the density of $\hat{a}^\top \xi$ is given by

\[
f_{\hat{a}^\top \xi}(y) = \begin{cases} \frac{1}{|a_d|} \int h \left( \frac{z_1, \ldots, z_{d-1}, y - \sum_{i=k+1}^{d-1} \hat{a}_i z_i}{a_d} \right) dz_1 \ldots dz_{d-1}, & \text{for } k < d, \\ \delta_{y=0}, & \text{for } k = d. \end{cases}
\]

When $l < d$ and $k < d$, we have

\[
d_{TV} (f_{G_n}, f_{G_0}) = \frac{1}{2} \int \left| \int \frac{\phi(x - y)}{|a_d|} \int h \left( \frac{z_1, \ldots, z_{d-1}, y - \sum_{i=k+1}^{d-1} \hat{a}_i z_i}{a_d} \right) dz_1 \ldots dz_{d-1} dy \right. \\
\left. - \int \frac{\phi(x - y)}{|a_d|} \int h \left( \frac{z_1, \ldots, z_{d-1}, y - \sum_{i=k+1}^{d-1} a_i z_i}{a_d} \right) dz_1 \ldots dz_{d-1} dy \right| dx \\
\leq \frac{1}{2} \int \left| \int \phi(x - \hat{a}^\top z) h(z) dz - \phi(x - a^\top z) h(z) dz \right| dx \\
\leq \frac{1}{2} \int \left\{ \exp \left\{ \left| (\hat{a} - a)^\top z \right|^2 \right\} - 1 \right\}^{1/2} h(z) dz \\
\leq \frac{1}{\sqrt{2}} \int \sqrt{\left| (\hat{a} - a)^\top z \right|^2} h(z) dz \\
\leq \frac{dB(\log m)^b \|\hat{a} - a\|_{\max}}{\sqrt{2}}
\]

under the event that $\|\hat{a} - a\|_{\max} \leq (\log (\log m))^{-1} \epsilon$ for $0 < \epsilon < 1$ because $e^x \leq 1 + 2x$ for any $0 < x < 1$ and

$$(\hat{a} - a)^\top z \leq \|\hat{a} - a\|_{\max}^2 \|z\|_1^2 \leq (\log (\log m))^2 \|\hat{a} - a\|_{\max}^2$$

under Assumption 5. The same bound holds if $l = d$ or $k = d$. Then, for $0 < \epsilon < \sqrt{2} dB(\log m)^b L'$ where $L'$ is the same as in Lemma 3, it follows that

\[
\left\{ d_{H}^2 (f_{G_n}, f_{G_0}) > \epsilon/2 \right\} \\
\subset \left\{ d_{TV} (f_{G_n}, f_{G_0}) > \epsilon/2, \|\hat{a} - a\|_{\max} \leq (\log (\log m))^b \|\hat{a} - a\|_{\max}^2 \right\} \\
\cup \left\{ \|\hat{a} - a\|_{\max} > (\log (\log m))^b \|\hat{a} - a\|_{\max} \right\} \\
\subset \left\{ \frac{dB(\log m)^b \|\hat{a} - a\|_{\max}}{\sqrt{2}} > \epsilon/2, \|\hat{a} - a\|_{\max} \leq (\log (\log m))^b \|\hat{a} - a\|_{\max} \right\} \\
\cup \left\{ \|\hat{a} - a\|_{\max} > (\log (\log m))^b \|\hat{a} - a\|_{\max} \right\} \\
\subset \left\{ \|\hat{a} - a\|_{\max} > (\log (\log m))^b \|\hat{a} - a\|_{\max} \right\}.
\]
Under Assumption [6] for large enough $n$, we have
\[
\sum_{n=1}^{\infty} P \left( d_H^2(f_{G_n}, f_{G_0}) > \varepsilon/2 \right) \leq \sum_{n=1}^{\infty} P \left( \| \hat{a} - a \|_{\max} > (dB(\log m)^b)^{-1} \varepsilon/\sqrt{2} \right) \\
\leq \sum_{n=1}^{\infty} P \left( \| \hat{a} - a \|_{\max} > \frac{\varepsilon}{\sqrt{2} dB D_6 (\log n)^b} \right),
\]
where $D_6 > 0$ is some constant such that $(\log m)^b \leq D_6 (\log n)^b$ for large enough $n$ which must exist under Assumption [6]. By Lemma [3], the summation above is finite. Thus we conclude that $d_H(f_{G_n}, f_{G_0}) = o_{a.s.}(1)$ by the Borel-Cantelli lemma.

Now we provide the proofs of Lemmas [1] and [2].

**Proof of Lemma [7]** Note that
\[
\left| \int L(Ax, t_1, t_2) dG_{m,n}(x) - \int L_0(A_0x, t_1, t_2) dG_0(x) \right| \\
\leq \left| \int \{L(Ax, t_1, t_2) - L_0(A_0x, t_1, t_2)\} dG_{m,n}(x) \right| + \left| \int L_0(A_0x, t_1, t_2)(dG_{m,n}(x) - dG_0(x)) \right|.
\]
It thus suffices to show that
\[
\sup_{t_1 \leq t'_1 \leq t_2 \leq t'_2, x \in \mathbb{R}} |L(Ax, t_1, t_2) - L_0(A_0x, t_1, t_2)| \overset{a.s.}{\to} 0, \quad (41)
\]
and
\[
\sup_{t_1 \leq t'_1 \leq t_2 \leq t'_2} \left| \int L_0(A_0x, t_1, t_2)(dG_{m,n}(x) - dG_0(x)) \right| \overset{a.s.}{\to} 0. \quad (42)
\]
We first verify the pointwise convergence in (41)-(42) and proceed to show the uniform convergence as stated. As $G_{m,n} = G_{m(n),n}$ under Assumption [6] throughout the proof, we denote $G_{m,n}$ by $G_n$ for notational simplicity.

1. **Pointwise convergence**

Let $a_n = a_n(t_1, t_2) = \left| \int L_0(A_0x, t_1, t_2)(dG_n(x) - dG_0(x)) \right|$ under Assumption [6]. By Lemmas [9] and [10]
\[
d_H(f_{G_n}(x), f_{G_0}(x)) \leq d_H(f_{G_n}(x), f_{G_n}(x)) + d_H(f_{G_n}(x), f_{G_0}(x)) = o_{a.s.}(1). \quad (43)
\]
Since $\int |f(x) - g(x)|dx \leq 2d_H(f, g)\sqrt{2 - d_H^2(f, g)}$ and $f_G(x) \leq 1/\sqrt{2\pi}$ for any $G$, we have
\[
\int |f_{G_n}(x) - f_{G_0}(x)|^2dx \leq \frac{2}{\sqrt{2\pi}} \int |f_{G_n}(x) - f_{G_0}(x)|dx = o_{a.s.}(1).
\]
For a density function or distribution function $f$, denote by $f^*$ its Fourier transformation. By the Parseval’s identity, we have
\[
\int |f_{G_n}(x) - f_{G_0}(x)|^2dx = \int |\phi^*(t)G_n^*(t) - \phi^*(t)G_0^*(t)|^2 dt = \int \phi^*(t)^2 \left| G_n^*(t) - G_0^*(t) \right|^2 dt = o_{a.s.}(1).
\]
Consider any convergent subsequence of \( a_n \), say \( a_{n_j} \). As 
\[
\int \phi^*(t)^2 \left| \tilde{G}_{n_{j}}^* (t) - G_0^* (t) \right|^2 dt = o.a.s. (1),
\]
there exists a further subsequence \( \tilde{G}_{n_{j_k}}^* (t) \) such that
\[
\tilde{G}_{n_{j_k}}^* (t) \xrightarrow{a.s.} G_0^* (t)
\]
for almost every \( t \) with respect to the measure \( \phi^*(t)dt \) (and thus also with respect to the Lebesgue measure). By the continuity theorem, \( \tilde{G}_{n_{j_k}}^* (t) \xrightarrow{a.s.} G_0^* (t) \) for any continuous point \( t \) of \( G_0(t) \). Then, we have \( a_{n_{j_k}} \to 0 \) by the Portmanteau theorem. As \( a_{n_j} \) is convergent, it must converge to zero as well. Since \( a_{n_j} \) is an arbitrary convergent subsequence, we have
\[
a_n = \left| \int L_0(A_0x, t_1, t_2)(d\tilde{G}_n(x) - dG_0(x)) \right| \xrightarrow{a.s.} 0 \tag{44}
\]
for any \( t_1, t_2 \). To show that \( |L(Ax, t_1, t_2) - L_0(A_0x, t_1, t_2)| \xrightarrow{a.s.} 0 \) for given \( x, t_1 \) and \( t_2 \), we note that the covariance matrix of \((V_1, V_2)\) converges almost surely to the covariance matrix of \((\tilde{V}_1, \tilde{V}_2)\) by Lemma 3. It follows from the continuous mapping theorem that
\[
|L(Ax, t_1, t_2) - L_0(A_0x, t_1, t_2)| \xrightarrow{a.s.} 0.
\]

2. Uniform convergence

We first show (41). For any \( \delta > 0 \), we can choose a large enough \( M' = M'(\delta) \) such that
\[
\int_{|x| > M'} dG_0(x) < \delta. \tag{45}
\]
For such \( M' \), using similar argument as above, we can verify that
\[
\left| \int_{|x| > M'} (d\tilde{G}_n(x) - dG_0(x)) \right| \xrightarrow{a.s.} 0. \tag{46}
\]
We partition the rectangular region \([0, t_1'] \times [0, t_2']\) into finite disjoint sets \( \bigcup_{1 \leq j \leq B} V_j \) such that uniformly over \( j \),
\[
\sup_{(t_1, t_2), (\tilde{t}_1, \tilde{t}_2) \in V_j} \left| \int (L_0(A_0x, t_1, t_2) - L_0(A_0x, \tilde{t}_1, \tilde{t}_2)) dG_0(x) \right| < \delta, \tag{47}
\]
and also for \( |x| \leq M' \)
\[
\sup_{(t_1, t_2), (\tilde{t}_1, \tilde{t}_2) \in V_j} \left| L_0(A_0x, t_1, t_2) - L_0(A_0x, \tilde{t}_1, \tilde{t}_2) \right| < \delta. \tag{48}
\]
We have for large enough \( m \) and uniformly over \( j \),

\[
\sup_{(t_1,t_2), \tilde{t}_1, \tilde{t}_2 \in V_j} \left| \int (L_0(A_0x, t_1, t_2) - L_0(A_0x, \tilde{t}_1, \tilde{t}_2) \right) d\tilde{G}_n(x) \right| \\
\leq \sup_{(t_1,t_2), \tilde{t}_1, \tilde{t}_2 \in V_j} \left| \int (L_0(A_0x, t_1, t_2) - L_0(A_0x, \tilde{t}_1, \tilde{t}_2) \right) d\tilde{G}_n(x) \right| \\
+ \sup_{(t_1,t_2), \tilde{t}_1, \tilde{t}_2 \in V_j} \left| \int (L_0(A_0x, t_1, t_2) - L_0(A_0x, \tilde{t}_1, \tilde{t}_2) \right) d\tilde{G}_n(x) \right| \\
\leq \sup_{(t_1,t_2), \tilde{t}_1, \tilde{t}_2 \in V_j} \left| \int (L_0(A_0x, t_1, t_2) - L_0(A_0x, \tilde{t}_1, \tilde{t}_2) \right) d\tilde{G}_n(x) \right| \\
+ 2 \int |\tilde{G}_n(x) - dG_0(x) + dG_0(x)| \leq 5\delta
\]

almost surely from (45), (46) and (48). Choosing \((\tilde{t}_1, \tilde{t}_2) \in V_j \) for \( 1 \leq j \leq B_1 \), we have

\[
\sup_{t_1 \leq t_1', t_2 \leq t_2'} \left| \int L_0(A_0x, t_1, t_2) (d\tilde{G}_n(x) - dG_0(x)) \right| \\
= \sup_{(t_1, t_2) \in U_{1 \leq j \leq B} V_j} \left| \int L_0(A_0x, t_1, t_2) (d\tilde{G}_n(x) - dG_0(x)) \right| \\
= \max_{1 \leq j \leq B} \sup_{(t_1, t_2) \in V_j} \left| \int \left\{ L_0(A_0x, t_1, t_2) - L_0(A_0x, \tilde{t}_1, \tilde{t}_2) \right\} (d\tilde{G}_n(x) - dG_0(x)) \right| \\
\leq 6\delta + \max_{1 \leq j \leq B_1} \left| \int L_0(A_0x, \tilde{t}_1, \tilde{t}_2) (d\tilde{G}_n(x) - dG_0(x)) \right|
\]

by (47) and (49). The proof for (41) is completed in view of (44).

To show (42), by the triangle inequality, it suffices to show that

\[
\sup_{t_1 \leq t_1', t_2 \leq t_2'} \left| L(Ax, t_1, t_2) - L_0(Ax, t_1, t_2) \right| \xrightarrow{a.s.} 0, \tag{50}
\]

\[
\sup_{t_1 \leq t_1', t_2 \leq t_2'} \left| L_0(Ax, t_1, t_2) - L_0(A_0x, t_1, t_2) \right| \xrightarrow{a.s.} 0. \tag{51}
\]

Then, (50) follows by applying Scheffe’s Lemma together with the pointwise almost sure convergence of the probability density of \((V_1, V_2)\) to that of \((\tilde{V}_1, \tilde{V}_2)\) by Lemma 3. We next show (51). We first observe that, for any \( a, t_1, t_2 \in \mathbb{R} \),

\[
L_0(a, t_1, t_2) = \mathbb{P}(a + \tilde{V}_1 \leq t_1, \tilde{V}_2 \geq t_2) + \mathbb{P}(a + \tilde{V}_1 \leq -t_1, \tilde{V}_2 \geq t_2) \\
+ \mathbb{P}(a + \tilde{V}_1 \geq t_1, \tilde{V}_2 \leq -t_2) + \mathbb{P}(a + \tilde{V}_1 \leq -t_1, \tilde{V}_2 \leq -t_2) \\
= \{ \Phi(-t_2) - \Phi(t_1 - a) + \Psi_1(t_1 - a, t_2) \} \\
+ \{ \Phi(-t_1 - a) - \Psi_1(-t_1 - a, t_2) \} + \{ \Phi(-t_2) - \Psi_1(t_1 - a, -t_2) \} \\
+ \Psi_1(-t_1 - a, -t_2), \tag{52}
\]

where \( \Psi_1(t_1, t_2) = \mathbb{P}(\tilde{V}_1 \leq t_1, \tilde{V}_2 \leq t_2). \) For any \( \delta > 0 \), we can choose some large enough
\( M''(\delta, t'_1, t'_2, A_0) > 0 \) such that
\[
\{ \Phi(t'_1 - A_0 M'') \lor \Psi_1(t'_1 - A_0 M'', t'_2) \} \leq \delta,
\]
\[
\{ \Phi(-t'_1 + A_0 M'') \land \Psi_1(-t'_1 + A_0 M'', -t'_2) \} \geq 1 - \delta
\]
which implies that for any \( x \geq M'' \) and \( t_1 \leq t'_1, t_2 \leq t'_2, \)
\[
\Phi(-t_1 - A_0 x) \leq \delta,
\]
\[
\Psi_1(-t_1 - A_0 x, -t_2) \leq \delta,
\]
\[
\Psi_1(t_1 - A_0 x, -t_2) \leq \delta,
\]
and, for any \( x \leq -M'' \),
\[
1 - \delta \leq \Phi(t_1 - A_0 x),
\]
\[
1 - \delta \leq \Psi_1(-t_1 - A_0 x, t_2),
\]
\[
1 - \delta \leq \Psi_1(t_1 - A_0 x, t_2).
\]
Here we use the fact that \( \Phi \) and \( \Psi_1 \) are non-decreasing continuous functions. Thus, to show (51), it suffices to show that
\[
\sup_{t_1 \leq t'_1, t_2 \leq t'_2, x \geq M''} |L_0(Ax, t_1, t_2) - L_0(A_0 x, t_1, t_2)| \xrightarrow{a.s.} 0,
\]
(56)
\[
\sup_{t_1 \leq t'_1, t_2 \leq t'_2, x < M''} |L_0(Ax, t_1, t_2) - L_0(A_0 x, t_1, t_2)| \xrightarrow{a.s.} 0.
\]
(57)
By (52),
\[
\sup_{t_1 \leq t'_1, t_2 \leq t'_2, x \geq M''} |L_0(Ax, t_1, t_2) - L_0(A_0 x, t_1, t_2)|
\]
\[
\leq \sup_{t_1 \leq t'_1, t_2 \leq t'_2, x \geq M''} |\Phi(-t_1 - A x) - \Phi(-t_1 - A_0 x)|
\]
\[
+ \sup_{t_1 \leq t'_1, t_2 \leq t'_2, x \geq M''} |\Phi(t_1 - A x) - \Phi(t_1 - A_0 x)|
\]
\[
+ \sup_{t_1 \leq t'_1, t_2 \leq t'_2, x \geq M''} |\Psi_1(-t_1 - A x, -t_2) - \Psi_1(-t_1 - A_0 x, -t_2)|
\]
\[
+ \sup_{t_1 \leq t'_1, t_2 \leq t'_2, x \geq M''} |\Psi_1(-t_1 - A x, t_2) - \Psi_1(-t_1 - A_0 x, t_2)|
\]
\[
+ \sup_{t_1 \leq t'_1, t_2 \leq t'_2, x \geq M''} |\Psi_1(t_1 - A x, -t_2) - \Psi_1(t_1 - A_0 x, -t_2)|
\]
\[
+ \sup_{t_1 \leq t'_1, t_2 \leq t'_2, x \geq M''} |\Psi_1(t_1 - A x, t_2) - \Psi_1(t_1 - A_0 x, t_2)|.
\]
By the fact that \( A \xrightarrow{a.s.} A_0 \) and (53), there exists some \( N(M''(\delta, t'_1, t'_2, A_0, \omega)) > 0 \) such that when \( n \geq N(M''(\delta, t'_1, t'_2, A_0, \omega)) \),
\[
\{ \Phi(t'_1 - A M'') \lor \Psi_1(t'_1 - A M'', t'_2) \} \leq \delta,
\]
\[
\{ \Phi(-t'_1 + A M'') \land \Psi_1(-t'_1 + A M'', -t'_2) \} \geq 1 - \delta.
\]
Since \( \Phi \) and \( \Psi_1 \) are both non-decreasing continuous functions, for \( |x| \geq M'' \),
\[
\Phi(-t_1 - Ax) \leq \delta, \\
\Psi_1(-t_1 - Ax, -t_2) \leq \delta, \\
\Psi_1(t_1 - Ax, -t_2) \leq \delta,
\]
(58)
or
\[
1 - \delta \leq \Phi(t_1 - Ax), \\
1 - \delta \leq \Psi_1(-t_1 - Ax, t_2), \\
1 - \delta \leq \Psi_1(t_1 - Ax, t_2).
\]
(59)
Therefore, (56) can be verified by (54)-(55) and (58)-(59) as \( \delta \) is arbitrary. For (57), as \( L_0 \) is Lipschitz with respect to the first argument, it follows that
\[
|L_0(Ax, t_1, t_2) - L_0(A_0x, t_1, t_2)| \leq \frac{6}{\sqrt{2\pi}} |Ax - A_0x| < \frac{6M''}{\sqrt{2\pi}} |A - A_0|
\]
(60)
which completes the proof because \( A \overset{a.s.}{\to} A_0 \).

\[\Box\]

**Proof of Lemma 2** We first note that
\[
V_m(t_1, t_2) = \sum_{i: \alpha_i = 0} 1 \left\{ |\tilde{Z}_i^U| \geq \tilde{r}_i t_1, |\tilde{Z}_i^A| \geq \tilde{r}_i t_2 \right\} = \tilde{V}_m(\tilde{r}_i t_1, \tilde{r}_i t_2)
\]
where \( \tilde{r}_i = \tilde{\sigma}_i / \sigma_i \). Define an event
\[
A = \left\{ \max_{1 \leq i \leq m} |\tilde{\sigma}_i^2 / \sigma_i^2 - 1| \to 0 \right\} \cap \left\{ \sup_{1 \leq i \leq m} \left| \frac{m_0^{-1}}{m_0} \sum_{i: \alpha_i = 0} 1 \left\{ |\tilde{Z}_i^U| \geq t_1, |\tilde{Z}_i^A| \geq t_2 \right\} - K_0(t_1, t_2) \right| \to 0 \right\}.
\]
Then, \( \mathbb{P}(A) = 1 \) by Assumption 2 and Lemma 1. For any \( \omega \in A \) and \( \delta > 0 \), we have \( 1 - \delta < \tilde{r}_i < 1 + \delta \) for any \( i \) when \( m \) is large enough. For this \( \omega \), we have
\[
\tilde{V}_m((1 + \delta)t_1, (1 + \delta)t_2) \leq V_m(t_1, t_2) \leq \tilde{V}_m((1 - \delta)t_1, (1 - \delta)t_2)
\]
for large enough \( m \). By Assumption 2 we get
\[
K_0((1 + \delta)t_1, (1 + \delta)t_2) \leq \liminf_m \frac{1}{m_0} V_m(t_1, t_2) \leq \limsup_m \frac{1}{m_0} V_m(t_1, t_2) \leq K_0((1 - \delta)t_1, (1 - \delta)t_2).
\]
As \( K_0 \) is continuous and \( \delta \) is arbitrary, we have \( V_m(t_1, t_2)/m_0 \to K_0(t_1, t_2) \) for any \( \omega \in A \). Similar arguments can be used to show the rest.

\[\Box\]

**Proof of Theorem 7** By the Glivenko-Cantelli lemma, we can show that the convergence in Lemma 2 holds uniformly, i.e.,
\[
\sup_{t_1 \leq i \leq t_2 \leq t_2} \left| \frac{V_m(t_1, t_2)}{m_0} - K_0(t_1, t_2) \right| \overset{a.s.}{\to} 0 \quad \text{and} \quad \sup_{t_1 \leq i \leq t_2 \leq t_2} \left| \frac{S_m(t_1, t_2)}{m_1} - K_1(t_1, t_2) \right| \overset{a.s.}{\to} 0.
\]
This implies, under Assumption 1

\[
\sup_{t_1 \leq t_1^*, t_2 \leq t_2^*} |\mathcal{K}_m(t_1, t_2) - K(t_1, t_2)| \xrightarrow{a.s.} 0 \tag{61}
\]

where \( \mathcal{K}_m(t_1, t_2) = m^{-1} \{ V_m(t_1, t_2) + S_m(t_1, t_2) \} \). We next show that

\[
\sup_{t_1 \leq t_1^*, t_2 \leq t_2^*} |\widehat{\text{FDP}}_\lambda(t_1, t_2) - \text{FDP}_\lambda^\infty(t_1, t_2)| \xrightarrow{a.s.} 0, \tag{62}
\]

\[
\sup_{t_1 \leq t_1^*, t_2 \leq t_2^*} \left| \frac{V_m(t_1, t_2)}{V_m(t_1, t_2) + S_m(t_1, t_2)} - \frac{\pi_0 K_0(t_1, t_2)}{K(t_1, t_2)} \right| \xrightarrow{a.s.} 0. \tag{63}
\]

To show (62), we first observe that, for large enough \( m \),

\[
|\mathcal{K}_m(t_1, t_2) - K(t_1, t_2)| \leq \frac{|K(t_1, t_2)|}{2}
\]

which implies

\[
|\mathcal{K}_m(t_1, t_2)| \geq \frac{|K(t_1, t_2)|}{2} \geq \frac{K(t_1^*, t_2^*)}{2} > 0
\]

because \( \inf_{t_1 \leq t_1^*, t_2 \leq t_2^*} |K(t_1, t_2)| \geq K(t_1^*, t_2^*) > 0 \). For large enough \( m \), it follows that

\[
|\widehat{\text{FDP}}_\lambda(t_1, t_2) - \text{FDP}_\lambda^\infty(t_1, t_2)|
\]

\[
= \frac{1}{(1 - 2\Phi(-\lambda))} \left| \int L(Ax, t_1, t_2)d\hat{G}_{m,n}(x)F_m(\lambda)K(t_1, t_2) - \mathbb{E}_{a\sim\xi}[L_0(A_0a^T\xi, t_1, t_2)]F(\lambda)\mathcal{K}_m(t_1, t_2) \right|
\]

\[
\leq \frac{2}{(1 - 2\Phi(-\lambda))} \left| \int L(Ax, t_1, t_2)d\hat{G}_{m}(x)F_m(\lambda)K(t_1, t_2) - \mathbb{E}_{a\sim\xi}[L_0(A_0a^T\xi, t_1, t_2)]F(\lambda)\mathcal{K}_m(t_1, t_2) \right|. \]

Thus, for (62), it suffices to verify that

\[
\sup_{t_1 \leq t_1^*, t_2 \leq t_2^*} \left| \int L(Ax, t_1, t_2)d\hat{G}_{m}(x)F_m(\lambda)K(t_1, t_2) - \mathbb{E}_{a\sim\xi}[L_0(A_0a^T\xi, t_1, t_2)]F(\lambda)\mathcal{K}_m(t_1, t_2) \right| \xrightarrow{a.s.} 0,
\]

which can be shown by (61) and Lemmas 1 and 2. We can also show (63) by using similar argument. As \( \text{FDP}_\lambda^\infty(t_1^*, 0) < q \) and \( \text{FDP}_\lambda^\infty(0, t_2^*) < q \), we have for large enough \( m \),

\[
\widehat{\text{FDP}}_\lambda(t_1^*, 0) < q \quad \text{and} \quad \text{FDP}_\lambda(0, t_2^*) < q,
\]
which implies that $\tilde{T}_1^* \leq t_1^*$ and $\tilde{T}_2^* \leq t_2^*$. Thus we have

$$
\liminf_m \left\{ \hat{\text{FDP}}(\tilde{T}_1^*, \tilde{T}_2^*) - \frac{V_m(\tilde{T}_1^*, \tilde{T}_2^*)}{V_m(\tilde{T}_1^*, \tilde{T}_2^*) + S_m(\tilde{T}_1^*, \tilde{T}_2^*)} \right\} \\
\geq \lim_m \inf_{t_1 \leq t_1', t_2 \leq t_2'} \left\{ \hat{\text{FDP}}(t_1, t_2) - \frac{V_m(t_1, t_2)}{V_m(t_1, t_2) + S_m(t_1, t_2)} \right\} \\
= \lim_m \inf_{t_1 \leq t_1', t_2 \leq t_2'} \left\{ \hat{\text{FDP}}(t_1, t_2) - \text{FDP}_\lambda^\infty(t_1, t_2) + \frac{\pi_0 K_0(t_1, t_2)}{K(t_1, t_2)} - \frac{V_m(t_1, t_2)}{V_m(t_1, t_2) + S_m(t_1, t_2)} \right\} \geq 0
$$

by (62), (63) and the fact that

$$
\text{FDP}_\lambda^\infty(t_1, t_2) - \frac{\pi_0 K_0(t_1, t_2)}{K(t_1, t_2)} \geq 0
$$

from (12). As $\hat{\text{FDP}}(\tilde{T}_1^*, \tilde{T}_2^*) \leq q$, we obtain

$$
\limsup_m \frac{V_m(\tilde{T}_1^*, \tilde{T}_2^*)}{V_m(\tilde{T}_1^*, \tilde{T}_2^*) + S_m(\tilde{T}_1^*, \tilde{T}_2^*)} \leq q.
$$

Finally by Fatou’s lemma, we get

$$
\limsup_m \hat{\text{FDR}}_m \leq \mathbb{E} \left[ \limsup_m \frac{V_m(\tilde{T}_1^*, \tilde{T}_2^*)}{V_m(\tilde{T}_1^*, \tilde{T}_2^*) + S_m(\tilde{T}_1^*, \tilde{T}_2^*)} \right] \leq q.
$$

We justify equation (12) in the following.

**Corollary 1.** Under Assumptions 3, 6 for every $t_1, t_2 > 0$, we have

$$
\left| \int L(Ax, t_1, t_2) dG_n(x) - \int L_0(A_0 x, t_1, t_2) dG_0(x) \right| \buildrel {a.s.} \over \rightarrow 0.
$$

**Proof of Corollary 1**

Note that

$$
\left| \int L(Ax, t_1, t_2) dG_n(x) - \int L_0(A_0 x, t_1, t_2) dG_0(x) \right| \\
= \left| \int L(Ax, t_1, t_2) dG_n(x) - \int L_0(A_0 x, t_1, t_2) dG_n(x) \right| \\
+ \left| \int L_0(A_0 x, t_1, t_2) dG_n(x) - \int L_0(A_0 x, t_1, t_2) dG_0(x) \right|.
$$

Thus, it suffices to show that

$$
\left| L(Ax, t_1, t_2) - L_0(A_0 x, t_1, t_2) \right| \buildrel {a.s.} \over \rightarrow 0,
$$

$$
\left| \int L_0(A_0 x, t_1, t_2) (dG_n(x) - dG_0(x)) \right| \buildrel {a.s.} \over \rightarrow 0,
$$

32
both of which can be verified by using similar arguments as those for Lemma [1]
Note S4: Power analysis

We derive the asymptotic power of the two-dimensional John Storey procedure. Our derivation is heuristic but can be made rigorous under suitable assumptions. Define

$$\left( T_{1,\text{Two}}, T_{2,\text{Two}} \right) = \arg \max_{(t_1, t_2) \in \mathcal{F}_{q,\text{Two}}} K(t_1, t_2)$$

where $K(t_1, t_2) = \pi_0 K_0(t_1, t_2) + (1 - \pi_0) K_1(t_1, t_2)$ with $K_1$ and $K_2$ defined in Assumption 2 and

$$\mathcal{F}_{q,\text{Two}} = \{(t_1, t_2) \in \mathbb{R}^+ \times \mathbb{R}^+ : \text{FDP}_X^\infty(t_1, t_2) \leq q\}.$$

Then, the asymptotic power of the two-dimensional procedure is given by

$$\text{Power}_{\text{Two}} = \lim_{m \to \infty} \sum_{i: \alpha_i \neq 0} 1 \left\{ |Z_i^U| \geq T_{1,\text{Two}}, |Z_i^A| \geq T_{2,\text{Two}} \right\} = K_1(T_{1,\text{Two}}, T_{2,\text{Two}}).$$

As a comparison the asymptotic power of the one-dimensional procedure is equal to

$$\text{Power}_{\text{One}} = K_1(0, T_{2,\text{One}})$$

where $T_{2,\text{One}} = \arg \max_{t_2 \in \mathcal{F}_{q,\text{One}}} K(t_1, t_2)$ with $\mathcal{F}_{q,\text{One}} = \{t_2 \in \mathbb{R}^+ : \text{FDP}_X^\infty(0, t_2) \leq q\}$. Since $\mathcal{F}_{q,\text{One}} \subset \mathcal{F}_{q,\text{Two}}$, we have

$$K(T_{1,\text{Two}}, T_{2,\text{Two}}) \geq K(0, T_{2,\text{One}}),$$

that is, the two-dimensional procedure delivers more rejections.

**Lemma 11.** Let $M(\lambda) = (1 - 2 \Phi(-\lambda))^{-1}(1 - \pi_0)(1 - K_1(0, \lambda))$ and Suppose $\text{FDP}_X^\infty(t_1, t_2)$ is a continuous function of $(t_1, t_2)$. Then we have $\text{Power}_{\text{Two}} \geq \text{Power}_{\text{One}}$.

**Proof of Lemma 11.** Since $\text{FDP}_X^\infty(t_1, t_2)$ is a continuous function of $(t_1, t_2)$, we must have

$$\frac{\pi_0 K_0(T_{1,\text{Two}}, T_{2,\text{Two}}) + M(\lambda) K_0(T_{1,\text{Two}}, T_{2,\text{Two}})}{K(T_{1,\text{Two}}, T_{2,\text{Two}})} = q, \quad \frac{\pi_0 K_0(0, T_{2,\text{One}}) + M(\lambda) K_0(0, T_{2,\text{One}})}{K(0, T_{2,\text{One}})} = q.$$  \hspace{1cm} (64)

The fact that $K(T_{1,\text{Two}}, T_{2,\text{Two}}) \geq K(0, T_{2,\text{One}})$ implies both $K_0(T_{1,\text{Two}}, T_{2,\text{Two}}) \geq K_0(0, T_{2,\text{One}})$ from (64) and

$$\begin{align*}
(1 - \pi_0) K_1(T_{1,\text{Two}}, T_{2,\text{Two}}) - M(\lambda) K_0(T_{1,\text{Two}}, T_{2,\text{Two}}) \\
= (1 - q) K_1(T_{1,\text{Two}}, T_{2,\text{Two}}) \\
\geq (1 - q) K(0, T_{2,\text{One}}) = (1 - \pi_0) K_1(0, T_{2,\text{One}}) - M(\lambda) K_0(0, T_{2,\text{One}})
\end{align*}$$

after rearranging the terms in (64). As $K_0(T_{1,\text{Two}}, T_{2,\text{Two}}) \geq K_0(0, T_{2,\text{One}})$, it follows that

$$\begin{align*}
(1 - \pi_0) K_1(T_{1,\text{Two}}, T_{2,\text{Two}}) \geq (1 - \pi_0) K_1(0, T_{2,\text{One}}) + M(\lambda) \{ K_0(T_{1,\text{Two}}, T_{2,\text{Two}}) - K_0(0, T_{2,\text{One}}) \} \\
\geq (1 - \pi_0) K_1(0, T_{2,\text{One}})
\end{align*}$$

which completes the proof. ☐
References


<table>
<thead>
<tr>
<th>ID</th>
<th>Number of discoveries by 2dFDR</th>
<th>Number of discoveries by 1dFDR-A</th>
<th>% improvement</th>
<th>R²</th>
<th>GEO accession</th>
<th>Phenotype</th>
<th>PMID</th>
<th>Tissue</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWAS1</td>
<td>301</td>
<td>155</td>
<td>94%</td>
<td>88.9%</td>
<td>GSE101764</td>
<td>Colorectal cancer</td>
<td>28791728</td>
<td>colorectal</td>
<td>261</td>
</tr>
<tr>
<td>EWAS10</td>
<td>1</td>
<td>1</td>
<td>0%</td>
<td>14.3%</td>
<td>GSE111223</td>
<td>Parkinson's Disease</td>
<td>26655927</td>
<td>saliva</td>
<td>259</td>
</tr>
<tr>
<td>EWAS11</td>
<td>2</td>
<td>1</td>
<td>50%</td>
<td>23.3%</td>
<td>GSE111629</td>
<td>Parkinson's Disease</td>
<td>26655927</td>
<td>blood</td>
<td>571</td>
</tr>
<tr>
<td>EWAS12</td>
<td>41</td>
<td>22</td>
<td>83%</td>
<td>31.0%</td>
<td>GSE114753</td>
<td>Smoking</td>
<td>28950428</td>
<td>esophagus</td>
<td>156</td>
</tr>
<tr>
<td>EWAS13</td>
<td>48</td>
<td>2</td>
<td>1533%</td>
<td>67.9%</td>
<td>GSE39279</td>
<td>Non-small Cell Lung Cancer</td>
<td>24081945</td>
<td>lung</td>
<td>444</td>
</tr>
<tr>
<td>EWAS14</td>
<td>75466</td>
<td>80224</td>
<td>-6%</td>
<td>30.4%</td>
<td>GSE40279</td>
<td>Age</td>
<td>23177740</td>
<td>blood</td>
<td>656</td>
</tr>
<tr>
<td>EWAS15</td>
<td>5</td>
<td>1</td>
<td>200%</td>
<td>21.9%</td>
<td>GSE40576</td>
<td>Asthma</td>
<td>25769910</td>
<td>blood</td>
<td>194</td>
</tr>
<tr>
<td>EWAS16</td>
<td>73</td>
<td>40</td>
<td>80%</td>
<td>49.8%</td>
<td>GSE42861</td>
<td>Children Born Small For Rheumatoid Arthritis</td>
<td>23334450</td>
<td>blood</td>
<td>689</td>
</tr>
<tr>
<td>EWAS17</td>
<td>3872</td>
<td>4361</td>
<td>-11%</td>
<td>91.8%</td>
<td>GSE47880</td>
<td>Gestational Age</td>
<td>24793120</td>
<td>blood</td>
<td>105</td>
</tr>
<tr>
<td>EWAS18</td>
<td>862</td>
<td>227</td>
<td>279%</td>
<td>83.0%</td>
<td>GSE48684</td>
<td>Colorectal Cancer</td>
<td>24500968</td>
<td>pancreatic</td>
<td>196</td>
</tr>
<tr>
<td>EWAS19</td>
<td>2786</td>
<td>1759</td>
<td>58%</td>
<td>79.2%</td>
<td>GSE49149</td>
<td>Pancreatic Ductal Adenocarcinoma</td>
<td>25008768</td>
<td>brain</td>
<td>157</td>
</tr>
<tr>
<td>EWAS2</td>
<td>35</td>
<td>4</td>
<td>620%</td>
<td>23.8%</td>
<td>GSE102468</td>
<td>Panic Disorder</td>
<td>29249830</td>
<td>blood</td>
<td>165</td>
</tr>
<tr>
<td>EWAS21</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>6.6%</td>
<td>GSE51057</td>
<td>Menarcheal Age</td>
<td>24278132</td>
<td>blood</td>
<td>329</td>
</tr>
<tr>
<td>EWAS22</td>
<td>170</td>
<td>75</td>
<td>125%</td>
<td>75.9%</td>
<td>GSE53045</td>
<td>Smoking</td>
<td>24559495</td>
<td>blood</td>
<td>111</td>
</tr>
<tr>
<td>EWAS23</td>
<td>30</td>
<td>19</td>
<td>55%</td>
<td>11.4%</td>
<td>GSE53740</td>
<td>Dementia (FTD)</td>
<td>25589773</td>
<td>blood</td>
<td>349</td>
</tr>
<tr>
<td>EWAS24</td>
<td>23303</td>
<td>22580</td>
<td>3%</td>
<td>72.6%</td>
<td>GSE45403</td>
<td>Hepatocellular Carcinoma</td>
<td>23208076</td>
<td>liver</td>
<td>132</td>
</tr>
<tr>
<td>EWAS25</td>
<td>81</td>
<td>31</td>
<td>156%</td>
<td>89.2%</td>
<td>GSE58218</td>
<td>Anaplastic Gliomas</td>
<td>25008768</td>
<td>brain</td>
<td>157</td>
</tr>
<tr>
<td>EWAS26</td>
<td>80</td>
<td>39</td>
<td>103%</td>
<td>93.9%</td>
<td>GSE58888</td>
<td>Age</td>
<td>25888029</td>
<td>blood</td>
<td>143</td>
</tr>
<tr>
<td>EWAS27</td>
<td>1150</td>
<td>776</td>
<td>48%</td>
<td>71.0%</td>
<td>GSE59065</td>
<td>Age</td>
<td>26286994</td>
<td>blood</td>
<td>100</td>
</tr>
<tr>
<td>EWAS28</td>
<td>822</td>
<td>539</td>
<td>52%</td>
<td>41.5%</td>
<td>GSE59250</td>
<td>Systemic Lupus Erythematosus</td>
<td>23950730</td>
<td>blood</td>
<td>104</td>
</tr>
<tr>
<td>EWAS29</td>
<td>87</td>
<td>26</td>
<td>226%</td>
<td>54.4%</td>
<td>GSE59250</td>
<td>Systemic Lupus Erythematosus</td>
<td>23950730</td>
<td>blood</td>
<td>149</td>
</tr>
<tr>
<td>EWAS3</td>
<td>4</td>
<td>0</td>
<td>400%</td>
<td>77.4%</td>
<td>GSE103186</td>
<td>Intestinal Metaplasia</td>
<td>29290541</td>
<td>gastric antrum</td>
<td>191</td>
</tr>
<tr>
<td>EWAS30</td>
<td>115</td>
<td>6</td>
<td>1557%</td>
<td>89.9%</td>
<td>GSE60132</td>
<td>Age</td>
<td>25806089</td>
<td>blood</td>
<td>192</td>
</tr>
<tr>
<td>EWAS31</td>
<td>554</td>
<td>161</td>
<td>243%</td>
<td>83.3%</td>
<td>GSE60185</td>
<td>In Situ And Invasive Carcinoma</td>
<td>25146004</td>
<td>breast</td>
<td>285</td>
</tr>
<tr>
<td>EWAS32</td>
<td>1207</td>
<td>1418</td>
<td>-15%</td>
<td>99.5%</td>
<td>GSE61278</td>
<td>Fetal And Adult Liver</td>
<td>25282492</td>
<td>liver</td>
<td>110</td>
</tr>
<tr>
<td>EWAS33</td>
<td>19</td>
<td>10</td>
<td>82%</td>
<td>47.7%</td>
<td>GSE64380</td>
<td>Developmental Delay/Intellectual Disability</td>
<td>26003415</td>
<td>blood</td>
<td>100</td>
</tr>
<tr>
<td>EWAS34</td>
<td>25</td>
<td>134</td>
<td>167%</td>
<td>19.4%</td>
<td>GSE64495</td>
<td>Syndrome X</td>
<td>25991677</td>
<td>blood</td>
<td>113</td>
</tr>
<tr>
<td>EWAS35</td>
<td>20</td>
<td>0</td>
<td>2000%</td>
<td>21.6%</td>
<td>GSE65183</td>
<td>Melanoma</td>
<td>26359985</td>
<td>melanoma</td>
<td>108</td>
</tr>
<tr>
<td>EWAS36</td>
<td>4538</td>
<td>4430</td>
<td>2%</td>
<td>26.4%</td>
<td>GSE67393</td>
<td>Sex Differences Of Leukocytes</td>
<td>26113971</td>
<td>blood</td>
<td>117</td>
</tr>
<tr>
<td>EWAS37</td>
<td>31</td>
<td>23</td>
<td>33%</td>
<td>24.6%</td>
<td>GSE69270</td>
<td>Age</td>
<td>26861258</td>
<td>blood</td>
<td>184</td>
</tr>
<tr>
<td>EWAS39</td>
<td>184</td>
<td>90</td>
<td>103%</td>
<td>44.9%</td>
<td>GSE70977</td>
<td>Oral And Pharyngeal Carcinoma</td>
<td>26653906</td>
<td>oral rinse</td>
<td>223</td>
</tr>
<tr>
<td>EWAS4</td>
<td>4937</td>
<td>4563</td>
<td>8%</td>
<td>74.7%</td>
<td>GSE72872</td>
<td>Esophageal Adenocarcinoma</td>
<td>26905591</td>
<td>esophagus</td>
<td>104</td>
</tr>
<tr>
<td>EWAS44</td>
<td>1087</td>
<td>425</td>
<td>155%</td>
<td>29.5%</td>
<td>GSE73115</td>
<td>Obesity</td>
<td>26449484</td>
<td>blood</td>
<td>355</td>
</tr>
<tr>
<td>EWAS45</td>
<td>383</td>
<td>155</td>
<td>4%</td>
<td>40.7%</td>
<td>GSE73115</td>
<td>Age</td>
<td>27498152</td>
<td>blood</td>
<td>180</td>
</tr>
</tbody>
</table>
**SUPPLEMENTARY TABLE**

<table>
<thead>
<tr>
<th>EWAS</th>
<th>#TSFDR</th>
<th>#OSFDR</th>
<th>#TSFDR - #OSFDR</th>
<th>#OSFDR + 1</th>
<th>% Improvement</th>
<th>% RSquare</th>
<th>GSE</th>
<th>Disease</th>
<th>Platform</th>
<th>Tissue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWAS47</td>
<td>22</td>
<td>17</td>
<td>28%</td>
<td>800%</td>
<td>48.0%</td>
<td>25.25%</td>
<td>GSE73894</td>
<td>Psoriatic Skin</td>
<td>skin</td>
<td>217</td>
<td>1277537</td>
</tr>
<tr>
<td>EWAS48</td>
<td>17</td>
<td>1</td>
<td>800%</td>
<td>63.1%</td>
<td>200%</td>
<td>48.3%</td>
<td>GSE74104</td>
<td>Testicular Germ Cell Tumors</td>
<td>testicular</td>
<td>267</td>
<td>29618</td>
</tr>
<tr>
<td>EWAS49</td>
<td>44</td>
<td>15</td>
<td>181%</td>
<td>57.6%</td>
<td>28.0%</td>
<td>20.9%</td>
<td>GSE74193</td>
<td>Schizophrenia</td>
<td>brain</td>
<td>526</td>
<td>85717</td>
</tr>
<tr>
<td>EWAS5</td>
<td>8943</td>
<td>6426</td>
<td>39%</td>
<td>85.6%</td>
<td>46.0%</td>
<td>4.56%</td>
<td>GSE104707</td>
<td>Esophageal Adenocarcinoma</td>
<td>esophagus</td>
<td>105</td>
<td>17600</td>
</tr>
<tr>
<td>EWAS50</td>
<td>35025</td>
<td>35607</td>
<td>-2%</td>
<td>12.6%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>GSE74845</td>
<td>Early Ovarian Cancer</td>
<td>fallopian</td>
<td>216</td>
<td>237800</td>
</tr>
<tr>
<td>EWAS51</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>37.6%</td>
<td>37.6%</td>
<td>37.6%</td>
<td>GSE80261</td>
<td>Human Fetal Alcohol</td>
<td>buccal</td>
<td>216</td>
<td>2735853</td>
</tr>
<tr>
<td>EWAS52</td>
<td>2</td>
<td>0</td>
<td>0%</td>
<td>200%</td>
<td>200%</td>
<td>200%</td>
<td>GSE80970</td>
<td>Alzheimer's Disease</td>
<td>brain</td>
<td>144</td>
<td>2955019</td>
</tr>
<tr>
<td>EWAS53</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>32.0%</td>
<td>32.0%</td>
<td>32.0%</td>
<td>GSE80970</td>
<td>Alzheimer's Disease</td>
<td>brain</td>
<td>142</td>
<td>2955019</td>
</tr>
<tr>
<td>EWAS54</td>
<td>113</td>
<td>32</td>
<td>245%</td>
<td>81.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>GSE84207</td>
<td>Breast Cancer</td>
<td>breast</td>
<td>279</td>
<td>29123100</td>
</tr>
<tr>
<td>EWAS55</td>
<td>13</td>
<td>2</td>
<td>367%</td>
<td>27.9%</td>
<td>399.0%</td>
<td>399.0%</td>
<td>GSE87095</td>
<td>Rheumatoid Arthritis</td>
<td>blood</td>
<td>122</td>
<td>28475762</td>
</tr>
<tr>
<td>EWAS56</td>
<td>892</td>
<td>637</td>
<td>40%</td>
<td>76.9%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>GSE88883</td>
<td>Breast Cancer</td>
<td>breast</td>
<td>100</td>
<td>28693600</td>
</tr>
<tr>
<td>EWAS58</td>
<td>36</td>
<td>0</td>
<td>3600%</td>
<td>81.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>GSE93933</td>
<td>Non Syndromic Cleft Lip</td>
<td>blood</td>
<td>126</td>
<td>28475860</td>
</tr>
<tr>
<td>EWAS59</td>
<td>38</td>
<td>0</td>
<td>3800%</td>
<td>72.7%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>GSE97362</td>
<td>CHARGE And Kabuki</td>
<td>blood</td>
<td>125</td>
<td>28475860</td>
</tr>
<tr>
<td>EWAS6</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>12.9%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>GSE104942</td>
<td>Susceptibility To Breast</td>
<td>blood</td>
<td>136</td>
<td>28475860</td>
</tr>
<tr>
<td>EWAS60</td>
<td>9</td>
<td>2</td>
<td>233%</td>
<td>76.3%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>GSE97362</td>
<td>Syndromes</td>
<td>blood</td>
<td>101</td>
<td>28475860</td>
</tr>
<tr>
<td>EWAS61</td>
<td>16</td>
<td>0</td>
<td>1600%</td>
<td>75.7%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>GSE97466</td>
<td>Thyroid Tumors</td>
<td>thyroid</td>
<td>110</td>
<td>28938429</td>
</tr>
<tr>
<td>EWAS67</td>
<td>27</td>
<td>9</td>
<td>180%</td>
<td>24.6%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>GSE106648</td>
<td>Multiple Sclerosis</td>
<td>blood</td>
<td>279</td>
<td>29921915</td>
</tr>
<tr>
<td>EWAS68</td>
<td>58</td>
<td>47</td>
<td>23%</td>
<td>24.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>GSE107351</td>
<td>Epimutations</td>
<td>blood</td>
<td>113</td>
<td>29982128</td>
</tr>
<tr>
<td>EWAS69</td>
<td>16</td>
<td>1</td>
<td>750%</td>
<td>12.2%</td>
<td>737.8%</td>
<td>737.8%</td>
<td>GSE109914</td>
<td>Arsenic Exposure</td>
<td>blood</td>
<td>119</td>
<td>29982128</td>
</tr>
</tbody>
</table>

* % improvement is defined as (#TSFDR - #OSFDR) / (#OSFDR + 1), where the addition of 1 is to avoid being divided by 0.

* R² (percent explained variance) is calculated by regressing the phenotype on the surrogate variables. It measures the association between the phenotype and the surrogate variables.
Fig. S1  Performance comparison when 50% of the features are affected by the confounder. From left to right, we increase the strength (effect size) of the confounding signals. False discovery proportions (A) and true positive rates (B) were averaged over 100 simulation runs. Error bars represent the 95% CIs and the dashed horizontal line indicates the target FDR level of 0.05. The density of the true signals is 10% and the strength is moderate. ‘+’, ‘++’ and ‘+++’ represent a low, medium and high correlation between the variable of interest and the confounder ($\rho = 0.2, 0.6, 0.8$), respectively. 2dFDR maintains FDR at the target level but the power is slightly lower than 1dFDR-A.
Fig. S2  Performance on simulated datasets across varying density (top to bottom) and strength (left to right) of the confounding signals when there are five confounders. False discovery proportions (A) and true positive rates (B) were averaged over 100 simulation runs. Error bars represent the 95% CIs and the dashed horizontal line indicates the target FDR level of 0.05. The density of the true signals is 10% and the strength is moderate. ‘+’, ‘++’ and ‘+++’ represent a low, medium and high correlation between the variable of interest and the confounders ($\rho = 0.2, 0.6, 0.8$), respectively. The performance is similar to the setting with one confounder.
Fig. S3  Performance across varying density (top to bottom) and strength (left to right) of the confounding signals when the confounding and true signals do not overlap (“NoCoLoc”) and when the true and confounding signals have extensive overlap (“CoLoc”). False discovery proportions (A, B) and true positive rates (C,D) were averaged over 100 simulation runs. Error bars represent the 95% CIs and the dashed horizontal line indicates the target FDR level of 0.05. The density of the true signals is 10% and the strength is moderate. ‘+’, ‘++’ and ‘+++’ represent a low, medium and high correlation between the variable of interest and the confounder ($\rho = 0.2, 0.6, 0.8$), respectively. 2dFDR is more powerful than 1dFDR-A when the density of the confounding signals is low and the correlation between the variable of interest and the confounder is high. However, as the confounding signals become denser, the power improvement decreases and 2dFDR is less powerful than 1dFDR-A when the correlation between the variable of interest and the confounder is low.
Fig. S4  Performance across varying density (top to bottom) and strength (left to right) of the confounding signals when the errors have a block correlation structure. False discovery proportions (A) and true positive rates (B) were averaged over 100 simulation runs. Error bars represent the 95% CIs and the dashed horizontal line indicates the target FDR level of 0.05. The density of the true signals is 10% and the strength is moderate. ‘+’, ‘++’ and ‘+++’ represent a low, medium and high correlation between the variable of interest and the confounder ($\rho = 0.2, 0.6, 0.8$), respectively. Here we simulate the block correlation structure with positive within-block correlation, which is commonly encountered in genomics data such as gene expression data. Under the block correlation structure, 2dFDR controls the FDR around the target level and the power is similar to that of the independent case.
Fig. S5  Performance across varying density (top to bottom) and strength (left to right) of the confounding signals when the errors have the first-order auto-regressive (AR(1)) correlation structure. False discovery proportions (A) and true positive rates (B) were averaged over 100 simulation runs. Error bars represent the 95% CIs and the dashed horizontal line indicates the target FDR level of 0.05. The density of the true signal is 10% and the strength is moderate. ‘+’, ‘++’ and ‘+++’ represent a low, medium and high correlation between the variable of interest and the confounder ($\rho = 0.2, 0.6, 0.8$), respectively. Here we simulate the AR(1) structure (correlation decays with the distance between the genomic features), which is commonly encountered in genomics data such as DNA methylation data. Under the AR(1) correlation structure, 2dFDR controls the FDR around the target level and the power is similar to that of the independent case.
Fig. S6  Performance comparison across varying density (top to bottom) and strength (left to right) of the confounding signals under smaller sample sizes. The density of the true signals is 10% and the strength is moderate. ‘+’, ‘++’ and ‘+++’ represent a low, medium and high correlation between the variable of interest and the confounder ($\rho \approx 0.2, 0.6, 0.8$), respectively. False discovery proportions (A, B) and true positive rates (C, D) were averaged over 100 simulation runs. We observed that the performance at $n = 50$ was similar to that at $n = 100$. At $n = 25$, 2dFDR becomes conservative in many settings.
Fig. S7  Performance comparison across varying density (top to bottom) and strength (left to right) of the confounding signals under smaller feature sizes. False discovery proportions (A, B) and true positive rates (C, D) were averaged over 1,000 simulation runs. Error bars represent the 95% CIs and the dashed horizontal line indicates the target FDR level of 0.05. The density of the true signals is 10% and the strength is moderate. ‘+’, ‘++’ and ‘+++’ represent a low, medium and high correlation between the variable of interest and the confounder ($\rho = 0.2, 0.6, 0.8$), respectively. We observe that the performance at $m = 500$ is similar to that at $m = 10000$. However, FDR is inflated at $m = 100$. 
**Fig. S8** Performance comparison across varying sample size (top to bottom) and feature size (left to right). False discovery proportions (A) and true positive rates (B) were averaged over 1,000 simulation runs. Error bars represent the 95% CIs and the dashed horizontal line indicates the target FDR level of 0.05. The density of the true/confounding signals is 10% and the strength is moderate. ‘+’, ‘++’ and ‘+++’ represent a low, medium and high correlation between the variable of interest and the confounder ($\rho = 0.2, 0.6, 0.8$), respectively. When both the sample size and feature size are small, FDR is slightly inflated for 2dFDR. Increasing the sample size or feature size does not rescue the degraded performance due to the other being small.
Fig. S9  The computation time (in seconds) under different sample sizes and feature sizes based on simulated datasets with one confounder, medium density and strength of the true and confounding signals, and a medium confounding level. A search grid of $50 \times 50$ was used without parallelization. The computation was performed under R 3.6.3 and macOS Catalina (v10.15.6) on a MacBook Pro with 2.2 GHz Quad-Core Intel Core i7 and 16 GB 1600 MHz DDR3.
Fig. S10  The decision boundaries of 2dFDR and 1dFDR-A under two unfavorable scenarios for 2dFDR. (A) Colocation of the confounding and true signals with opposite effect sizes so the signals can be revealed only by adjustment. Moderate correlation between the variable of interest and the confounder was simulated ($\rho = 0.6$); (B) Independence between the variable of interest and the confounder. Both data examples were simulated with $n = 100$, $m = 10,000$ and 10% true and confounding signals with moderate strength.