# The influence of misspecified covariance on false discovery control when using posterior probabilities

Ye Liang, Joshua D. Habiger & Xiaoyi Min

Published online: 17 Oct 2017.

Submit your article to this journal

Article views: 96

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

Check for updates

# The influence of misspecified covariance on false discovery control when using posterior probabilities

Ye Liang [a], Joshua D. Habiger[a] and Xiaoyi Min[b]

[a]Department of Statistics, Oklahoma State University, Stillwater, OK, USA; [b]Department of Mathematics and Statistics, Georgia State University, Atlanta, GA, USA

**ABSTRACT**

This paper focuses on the influence of a misspecified covariance structure on false discovery rate for the large-scale multiple testing problem. Specifically, we evaluate the influence on the marginal distribution of local false discovery rate statistics, which are used in many multiple testing procedures and related to Bayesian posterior probabilities. Explicit forms of the marginal distributions under both correctly specified and incorrectly specified models are derived. The Kullback–Leibler divergence is used to quantify the influence caused by a misspecification. Several numerical examples are provided to illustrate the influence. A real spatio-temporal data on soil humidity is discussed.

## 1. Introduction

Large-scale multiple testing arises from many practical problems, from genetic studies to public health surveillance. Benjamini and Hochberg (1995) introduced the concept of the false discovery rate (FDR) and proposed a powerful testing procedure, usually referred as the BH procedure. The BH procedure relies on a positive dependence assumption (Benjamini & Yekutieli, 2001), while adaptive BH procedures (Liang & Nettleton, 2012; Storey, Taylor, & Siegmund, 2004) rely on an independence or weak dependence structure. Efron (2007) noted that correlation may result in overly liberal or overly conservative testing procedures. Though the BH procedure is valid under different dependence assumptions (Farcomeni, 2007; Wu, 2008), Sun and Cai (2009) showed that failing to model dependence can result in inefficiency. To address that problem, Sun and Cai (2009) and Sun, Reich, Cai, Guindani, and Schwartzman (2015) proposed a procedure using local significance index, which is a Bayesian posterior probability. Efron, Tibshirani, Storey, and Tusher (2001) described the connection between the FDR and Bayes procedures, where a posterior probability is referred as the local false discovery rate (Lfdr). Sun and Cai (2009)'s local significance index reduces to the Lfdr under independence. In fact, there is a rich history of using Bayesian approaches for multiplicity adjustment. Scott and Berger (2006) and Scott and Berger (2010) discussed Bayesian multiplicity adjustment in variable selections. Muller, Parmigiani, and Rice (2006) had a comprehensive discussion on the connection between the FDR, Bayesian multiple testing, and procedures using posterior probabilities.

In Sun et al. (2015), the procedure depending on unknown parameters is called the 'oracle' procedure, in the sense that we know all nuisance parameters as an oracle. Although the oracle procedure is proved to control FDR at the nominal level and be optimal in terms of false non-discovery rate (FNR), a data-driven 'adaptive' procedure relies on correctly specifying the model, including the prior specification, and/or consistent parameter estimation. For large-scale data, dependence or covariance, if in Gaussian models, is often estimated based on a structured model. The model choice or the structure choice itself may be debatable and parameter estimation remains challenging. For example, in spatial modelling, the estimation of covariance relies on structured covariance specifications and, in practice, one may have multiple choices of specifications. Intuitively, the choice of specification will influence the data-driven procedure and may eventually lead to different decisions.

In this paper, we explore the influence of a misspecified covariance structure on the testing procedure. Specifically, we study the sampling distributions of the Lfdr statistics under both correctly and incorrectly specified covariance structures. We derive explicit expressions for those distributions under a general model setting. We propose to use the Kullback–Leibler divergence as a quantitative measure for the influence. We show in both a simulation study and a real application that the influence of a misspecification leads to unappealing results. The paper is organised as follows. Section 2.1 gives a basic setup of this problem. Section 2.2 gives sampling distributions of the test statistics. Section 2.3 provides formulas for computing

---

the Kullback–Leibler divergence. Section 2.4 shows several numerical examples. Section 3 provides real data that arose from the Oklahoma soil monitoring network. Section 4 is a discussion.

## 2. Main results

### 2.1. The setup

In this paper, we consider a general model for $m$ observations $\boldsymbol{y} = (y_1, \ldots, y_m)'$:

$$y_i = \theta_i + \epsilon_i, \quad i = 1, \ldots, m, \tag{1}$$

where $\theta_i$ is the latent state and $\epsilon_i$ is the noise term which independently follows $N(0, \sigma^2)$. The dependence of observations is introduced through their latent states $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)'$. For instance, $\theta_i$ can be a realised spatial process $\theta(s)$, for which a spatial dependence structure can be specified. Consider a one-sided hypothesis

$$H_{0i} : \theta_i \geq \theta_{0i} \quad \text{versus} \quad H_{1i} : \theta_i < \theta_{0i}, \tag{2}$$

for every $i$ simultaneously. This type of one-sided hypothesis is often of interest in many practices. In spatial epidemiology, one may want to determine which regions or locations have disease rates higher than some given threshold $\theta_0$. Here, data will be spatially correlated and each hypothesis will be one-sided. Similarly, in agricultural studies, one may want to determine which locations or time periods have soil moisture levels lower than a given threshold $\theta_0$, indicating a risk of drought, and the data will be either spatially or temporally correlated and each hypothesis will be one-sided. A more general hypothesis would be $H_{0i} : \theta_i \in \Theta_{0i}$ versus $H_{1i} : \theta_i \in \Theta_{1i}$. We do not consider a precise (or two-sided) hypothesis in this paper, but some comments are given in Section 4.

The dependence of latent states $\boldsymbol{\theta}$ is usually specified through a prior model. For example, consider a normal-inverse-gamma prior on $(\boldsymbol{\theta}, \sigma^2)$,

$$\boldsymbol{\theta} \mid \sigma^2 \sim N_m(\boldsymbol{\theta}_0, g\sigma^2 \boldsymbol{\Sigma}) \quad \text{and} \quad \sigma^2 \sim IG(\alpha, \beta). \tag{3}$$

For simplicity, we assume that $\boldsymbol{\Sigma}$ is a known covariance structure and $g$ is a known scale parameter. The use of $g$ here is the same as that in Zellner's $g$-prior for Bayesian variable selection problems. The $g$ value could be fixed, estimated or have a hyperprior (Liang, Paulo, Molina, Clyde, & Berger, 2008). The prior specification (3) in fact induces a marginal probability for each hypothesis: $P(H_{0i}) = P(\theta_i \geq \theta_{0i}) = 0.5$.

Sun and Cai (2009) and Sun et al. (2015) showed that, to control FDR when data are dependent, the posterior probability $h_i = P(H_{0i} \mid \boldsymbol{y})$ is useful. The posterior probability $h_i$ is viewed as a test statistic, called local index of significance in their work. The oracle procedure orders $\boldsymbol{h}^m = (h_{(1)}, \ldots, h_{(m)})$ and rejects all $H_{(i)}$,

$i = 1, \ldots, k$ such that

$$k = \max \left\{ i : \frac{1}{i} \sum_{j=1}^{i} h_{(j)} \leq \alpha^* \right\}, \tag{4}$$

where $\alpha^*$ is the nominal level. The procedure mimics the Benjamini–Hochberg procedure, in which $p$-value is the test statistic. Sun and Cai (2009) showed that this oracle procedure controls FDR at level $\alpha^*$ and has the smallest FNR among all FDR procedures at $\alpha^*$ for a hidden Markov model. Sun et al. (2015) further showed that in a spatial random field model, this oracle procedure controls FDR at level $\alpha^*$ and has the smallest missed discovery rate (MDR). A data-driven procedure, however, depends on the estimation of other nuisance parameters. The covariance $\boldsymbol{\Sigma}$ is especially important in this case as it describes the dependence. Our objective is to determine if the procedure is sensitive when the covariance is incorrectly specified or estimated, and, if so, to quantify the sensitivity.

### 2.2. Sampling distribution of test statistics

#### 2.2.1. Known variance of noise

We now focus on how the distribution of test statistics $(h_1, \ldots, h_m)$ is influenced by a misspecified covariance structure. Assume that the data are generated from the true underlying process:

$$y_i \sim N(\theta_i, \sigma_0^2) \quad \text{and} \quad \boldsymbol{\theta} \sim N(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}_1), \tag{5}$$

for $i = 1, \ldots, m$. Assume known $\sigma_0^2$ and consider model (1) with priors $\boldsymbol{\theta} \sim N(\boldsymbol{\theta}_0, g\boldsymbol{\Sigma}_1)$ and $\boldsymbol{\theta} \sim N(\boldsymbol{\theta}_0, g\boldsymbol{\Sigma}_2)$, the latter of which has a misspecified covariance structure. The scale $g$ determines the strength of the prior. The intuition is that both $g$ and $\boldsymbol{\Sigma}_2$ will influence the test statistics $(h_1, \ldots, h_m)$ and FDR control.

**Lemma 2.1:** *Under the correct covariance $\boldsymbol{\Sigma}_1$, $h_i$ marginally has the following CDF:*

$$F(h_i) = \Phi \left[ \sqrt{\frac{a_{ii}}{b_{ii}}} \Phi^{-1}(h_i) \right], \tag{6}$$

*where $a_{ii}$ is the $i$th diagonal element in $\boldsymbol{A} = (1/\sigma_0^2 \boldsymbol{I} + 1/g\boldsymbol{\Sigma}_1^{-1})^{-1}$ and $b_{ii}$ is the $i$th diagonal element in $\boldsymbol{B} = (\boldsymbol{I} + \sigma_0^2/g\boldsymbol{\Sigma}_1^{-1})^{-1}(\sigma_0^2 \boldsymbol{I} + \boldsymbol{\Sigma}_1)(\boldsymbol{I} + \sigma_0^2/g\boldsymbol{\Sigma}_1^{-1})^{-1}$.*

*Under the misspecified covariance $\boldsymbol{\Sigma}_2$, $\boldsymbol{A} = (1/\sigma_0^2 \boldsymbol{I} + 1/g\boldsymbol{\Sigma}_2^{-1})^{-1}$ and $\boldsymbol{B} = (\boldsymbol{I} + \sigma_0^2/g\boldsymbol{\Sigma}_2^{-1})^{-1}(\sigma_0^2 \boldsymbol{I} + \boldsymbol{\Sigma}_1)(\boldsymbol{I} + \sigma_0^2/g\boldsymbol{\Sigma}_2^{-1})^{-1}$.*

Lemma 2.1 shows explicitly how the sampling distribution is altered by a misspecification. Observe that $F(h_i)$ is completely determined by the ratio $a_{ii}/b_{ii}$. Figure 1 shows different shapes of both cumulative distribution functions (CDF) and probability density functions (pdf) under different ratio values. Note that when $a_{ii} = b_{ii}$, the sampling distribution is Uniform(0, 1). Consider $g \to \infty$, for which the prior becomes non-informative, then $a_{ii}/b_{ii} \to \sigma_0^2/(\sigma_0^2 + \sigma_{1,ii}^2)$, where $\sigma_{1,ii}^2$
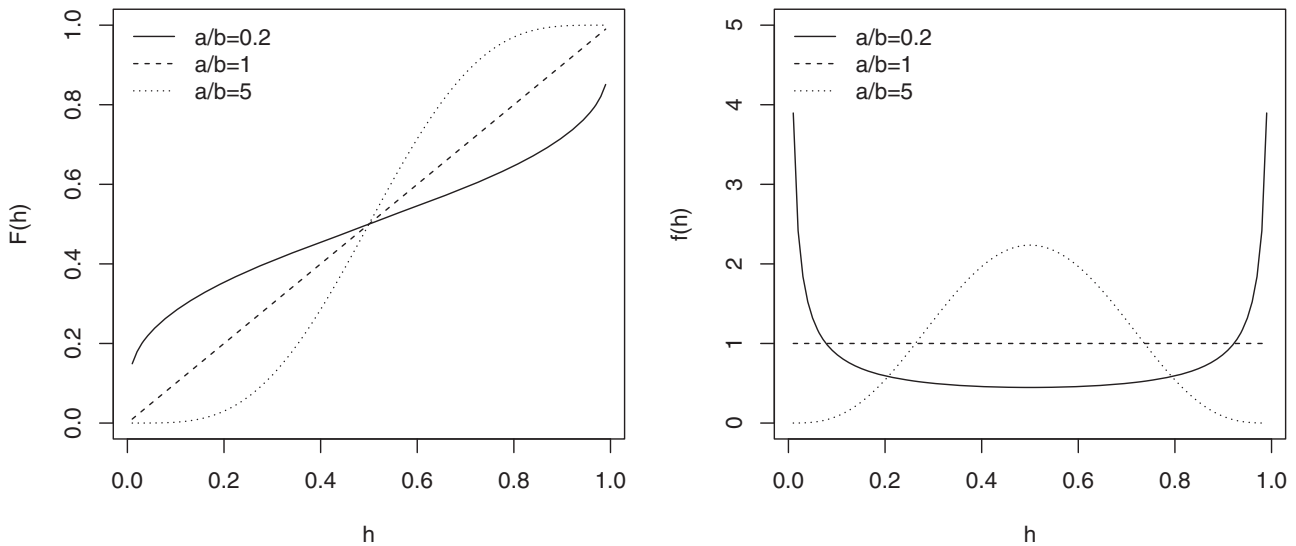
**Figure 1.** Marginal CDF and pdf for $h_i$.

is the $i$th diagonal element in $\mathbf{\Sigma}_1$. The ratio becomes irrelevant to the misspecified covariance $\mathbf{\Sigma}_2$. In other words, the behaviours of the correct specification and the incorrect specification will be similar when $g$ is large, which seems intuitive. Because FDR procedures based on $h_i$ reject $H_{0i}$ if $h_i \leq C$ for some $C$, it is important to note that $P(H_i \leq h_i) = F(h_i)$ can be substantially influenced by covariance misspecification. For example, in Figure 1, we see $F(0.2)$ ranges from 0.03 to 0.35. It should be noted that some recommend rejecting $H_{0i}$ if $h_i \leq 0.2$ (Efron, 2012). In this setting, the rejection probability ranges from 0.03 to 0.35.

Unlike the independent case, $h_1, \ldots, h_m$ are now dependent when the data are dependent. The change under a misspecified structure is revealed in their joint distribution.

**Theorem 2.1:** *Using the same definition for $\mathbf{A}$ and $\mathbf{B}$ in Lemma 2.1, $h_1, \ldots, h_m$ have a joint CDF*

$$F(h_1, \ldots, h_m) = \Phi_b^m\left[\sqrt{\frac{a_{11}}{b_{11}}}\Phi^{-1}(h_1), \ldots, \right.$$
$$\left. \sqrt{\frac{a_{mm}}{b_{mm}}}\Phi^{-1}(h_m)\right],$$

*where $\Phi_b^m$ is the CDF for a multivariate normal $N_m(\mathbf{0}, \mathbf{P}_b)$, and $\mathbf{P}_b$ is the correlation matrix of $\mathbf{B}$.*

The joint distribution of $h_1, \ldots, h_m$ represents a multivariate surface in the space $[0, 1]^m$. Notice that jointly not only the ratio $a_{ii}/b_{ii}$ plays a role but also the correlation structure of $\mathbf{B}$. Under a misspecified covariance structure, $\mathbf{P}_b$ will be altered as well. However, still, as $g \to \infty$, $\mathbf{B} \to \sigma_0^2\mathbf{I} + \mathbf{\Sigma}_1$, and there is no misspecification effect.

### 2.2.2. Unknown variance of noise

Assume the underlying data-generating process (5). Suppose $\sigma_0^2$ is unknown and consider specifying model (1) with prior (3). As before, a correct covariance structure is $\mathbf{\Sigma}_1$ and a misspecified structure is $\mathbf{\Sigma}_2$.

**Theorem 2.2:** *Under the correct covariance $\mathbf{\Sigma}_1$, the test statistics $h_1, \ldots, h_m$ jointly have the following CDF:*

$$F(h_1, \ldots, h_m) = \Xi_{a,b}^m\left[\sqrt{\frac{a_{11}}{b_{11}}}\Psi_{m+2\alpha}^{-1}(h_1), \ldots, \right.$$
$$\left. \sqrt{\frac{a_{mm}}{b_{mm}}}\Psi_{m+2\alpha}^{-1}(h_m)\right], \qquad (7)$$

*where $\mathbf{A} = (\mathbf{I} + 1/g\mathbf{\Sigma}_1^{-1})^{-1}$, $\mathbf{B} = (\mathbf{I} + 1/g\mathbf{\Sigma}_1^{-1})^{-1}(\sigma_0^2\mathbf{I} + \mathbf{\Sigma}_1)(\mathbf{I} + 1/g\mathbf{\Sigma}_1^{-1})^{-1}$, $\Psi_{m+2\alpha}$ is the CDF for a univariate t-distribution with degrees of freedom $m + 2\alpha$, and $\Xi_{a,b}^m$ is the CDF for the following random vector*

$$\sqrt{\frac{m + 2\alpha}{\mathbf{z}_b'\mathbf{C}\mathbf{z}_b + 2\beta}}\,\mathbf{z}_b,$$

*where $\mathbf{z}_b \sim N(\mathbf{0}, \mathbf{P}_b)$ and $\mathbf{C} = (\mathrm{diag}\,\mathbf{B})^{-1/2}(\mathbf{A}^{-2} - \mathbf{A}^{-1})(\mathrm{diag}\,\mathbf{B})^{-1/2}$, where $\mathrm{diag}$ here denotes a diagonal matrix.*

*Under the misspecified covariance $\mathbf{\Sigma}_2$, $\mathbf{A} = (\mathbf{I} + 1/g\mathbf{\Sigma}_2^{-1})^{-1}$ and $\mathbf{B} = (\mathbf{I} + 1/g\mathbf{\Sigma}_2^{-1})^{-1}(\sigma_0^2\mathbf{I} + \mathbf{\Sigma}_1)(\mathbf{I} + 1/g\mathbf{\Sigma}_2^{-1})^{-1}$.*

The main difference between Theorem 2.2 and Theorem 2.1 is that $\Xi_{a,b}^m$ has a more complicated form than $\Phi_b^m$. The ratio $a_{ii}/b_{ii}$ still plays a role in the joint distribution and $\Xi_{a,b}^m$ will be affected by a misspecified structure.

## 2.3. Kullback–Leibler divergence

Denote the sampling distribution of $\boldsymbol{h}$ under the correct covariance specification as $f_{\text{cor}}(\boldsymbol{h})$ and under the misspecified covariance as $f_{\text{mis}}(\boldsymbol{h})$. We evaluate the influence of the misspecification by the Kullback–Leibler (KL) divergence

$$D_{\text{KL}}(f_{\text{cor}} \| f_{\text{mis}}) = \int f_{\text{cor}}(\boldsymbol{h}) \log \frac{f_{\text{cor}}(\boldsymbol{h})}{f_{\text{mis}}(\boldsymbol{h})} d\boldsymbol{h}. \qquad (8)$$

The KL divergence here can be interpreted as the information loss when using $f_{\text{mis}}$ to approximate $f_{\text{cor}}$. The following two corollaries are useful to approximate the KL divergence in our cases.

**Corollary 2.1:** *As a consequence of Lemma 2.1, the marginal density function of $h_i$ is*

$$f(h_i) = \sqrt{r_i} \exp\left\{\frac{1}{2}(1 - r_i)\phi_i^2\right\},$$

*where $r_i = a_{ii}/b_{ii}$ and $\phi_i = \Phi^{-1}(h_i)$.*

**Corollary 2.2:** *As a consequence of Theorem 2.1, the joint density function of $(h_1, \ldots, h_m)'$ is*

$$f(h_1, \ldots, h_m) = \left| \boldsymbol{R}^{\frac{1}{2}} \boldsymbol{P}_b^{-1} \boldsymbol{R}^{\frac{1}{2}} \right|^{\frac{1}{2}} \exp\left\{\frac{1}{2}\boldsymbol{\phi}'(\boldsymbol{I} - \boldsymbol{R}^{\frac{1}{2}} \boldsymbol{P}_b^{-1} \boldsymbol{R}^{\frac{1}{2}})\boldsymbol{\phi}\right\},$$

*where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_m)' = (\Phi^{-1}(h_1), \ldots, \Phi^{-1}(h_m))'$ and $\boldsymbol{R} = \text{diag}(r_1, \ldots, r_m)$.*

With Corollary 2.2, the term $\log\{f_{\text{cor}}(\boldsymbol{h})/f_{\text{mis}}(\boldsymbol{h})\}$ in expression (8) is analytically available. Thus, the KL divergence $D_{\text{KL}}$ can be easily evaluated using Monte–Carlo approximation. Notice that we can draw from $f_{\text{cor}}$ exactly given the underlying model because $h_i = P(H_{0i} \mid \boldsymbol{y})$ is analytically available. To draw a sample of $\boldsymbol{h}$ from $f_{\text{cor}}$, draw $\boldsymbol{y}$ from the underlying model and then calculate $\boldsymbol{h}$. Suppose we obtain a sample $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_L$, the approximation is

$$D_{\text{KL}}(f_{\text{cor}} \| f_{\text{mis}}) \approx \frac{1}{L} \sum_{l=1}^{L} \log \frac{f_{\text{cor}}(\boldsymbol{h}_l)}{f_{\text{mis}}(\boldsymbol{h}_l)}.$$

A misspecified covariance will change the defined matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, and consequently change matrices $\boldsymbol{R}$ and $\boldsymbol{P}_b$. Notice the relationship $\boldsymbol{R}^{1/2}\boldsymbol{P}_b^{-1}\boldsymbol{R}^{1/2} = (\text{diag}^{1/2}\boldsymbol{A})\boldsymbol{B}^{-1}(\text{diag}^{1/2}\boldsymbol{A})$. Since the KL divergence will generally increase as the dimension $m$ increases (in the independent case, it is simply a sum of individual dimensions), we may also consider a relative measure of influence $D_{\text{KL}}/m$.

The KL divergence is computable under the general model (1) with $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ provided. This easy-to-compute measure can be used to quantify the influence of a misspecification. In practice, when there are multiple candidate covariances, we may assess the KL divergence between those candidates.

## 2.4. Numerical examples

We now consider two examples of misspecified covariance. In each example, without loss of generality, we set $m = 900$ and $\sigma_0^2 = 0.25$. We will numerically evaluate the KL divergence and perform a simulation study, in which we estimate FDR and FNR with Monte–Carlo replications of 1000.

**Example 2.1:** Positive spatial covariance vs. Independence.

Consider a regular spatial grid with unit distance one for generating the latent states $\boldsymbol{\theta}$. The true covariance $\boldsymbol{\Sigma}_1 = \{\sigma_{1,ij}^2\}$ has a positive decaying structure determined by an exponential covariance function $\sigma_{1,ij}^2 = \exp\{-\|s_i - s_j\|/\rho\}$ with $\rho = 5$, where $s$ represents a location. A misspecified covariance is $\boldsymbol{\Sigma}_2 = \boldsymbol{I}$. To get a rough idea of $r_i = a_{ii}/b_{ii}$, let $\text{g} = 1$ and compute $\boldsymbol{A}$ and $\boldsymbol{B}$. Under the correct specification, $r_i$ ranges from 0.12 to 0.16, and under the misspecification, $r_i = 0.25$.

Note that this misspecification is essentially to ignore the dependence and treat data as independent observations. This is quite common in practice, where domain scientists are hesitant to model complex covariance structures, though evidences suggest that data may be correlated.

**Example 2.2:** Negative AR(2) covariance vs. Positive AR(2) covariance.

Consider a time series for generating the latent states $\boldsymbol{\theta}$. The true covariance $\boldsymbol{\Sigma}_1$ is determined by an AR(2) process: $\theta_i = \rho_1\theta_{i-1} + \rho_2\theta_{i-2} + \varepsilon_i$, $\varepsilon_i \sim N(0, 1)$, and $\rho_1 = 1.5$ and $\rho_2 = -0.9$. The autocorrelation function of this specification has an oscillating pattern (mixed positive and negative values in $\boldsymbol{\Sigma}_1$). A misspecified covariance $\boldsymbol{\Sigma}_2$ is chosen to be the covariance for an AR(2) process with $\rho_1 = 0.6$ and $\rho_2 = 0.3$, whose autocorrelation is always positive. To get a rough idea of $r_i = a_{ii}/b_{ii}$, let $\text{g} = 1$ and compute $\boldsymbol{A}$ and $\boldsymbol{B}$. Under the correct specification, $r_i$ ranges from 0.088 to 0.14, and under the misspecification, $r_i$ ranges from 0.20 to 0.25.

Note that this example shows a scenario where model is correctly specified but parameter estimates are wrong. This example also compares a covariance matrix containing negative values with a covariance matrix containing all positive values.

Results of Examples 2.1 and 2.2 are shown in Figures 2 and Figure 3. We choose different $g$ values representing different strengths of information brought in by the prior dependence. Four plots are shown in each result: the estimated FDR, the estimated FNR, the difference between the rejection rate of the correct specification and that of the misspecification, i.e., $(\# \, discovery_{\text{cor}} - \# \, discovery_{\text{mis}})/m$, and $D_{\text{KL}}/m$. We can reach the following conclusions from these plots. First, as $g$ increases, both the KL divergence and the difference between rejection rates decrease, and also both the FDR
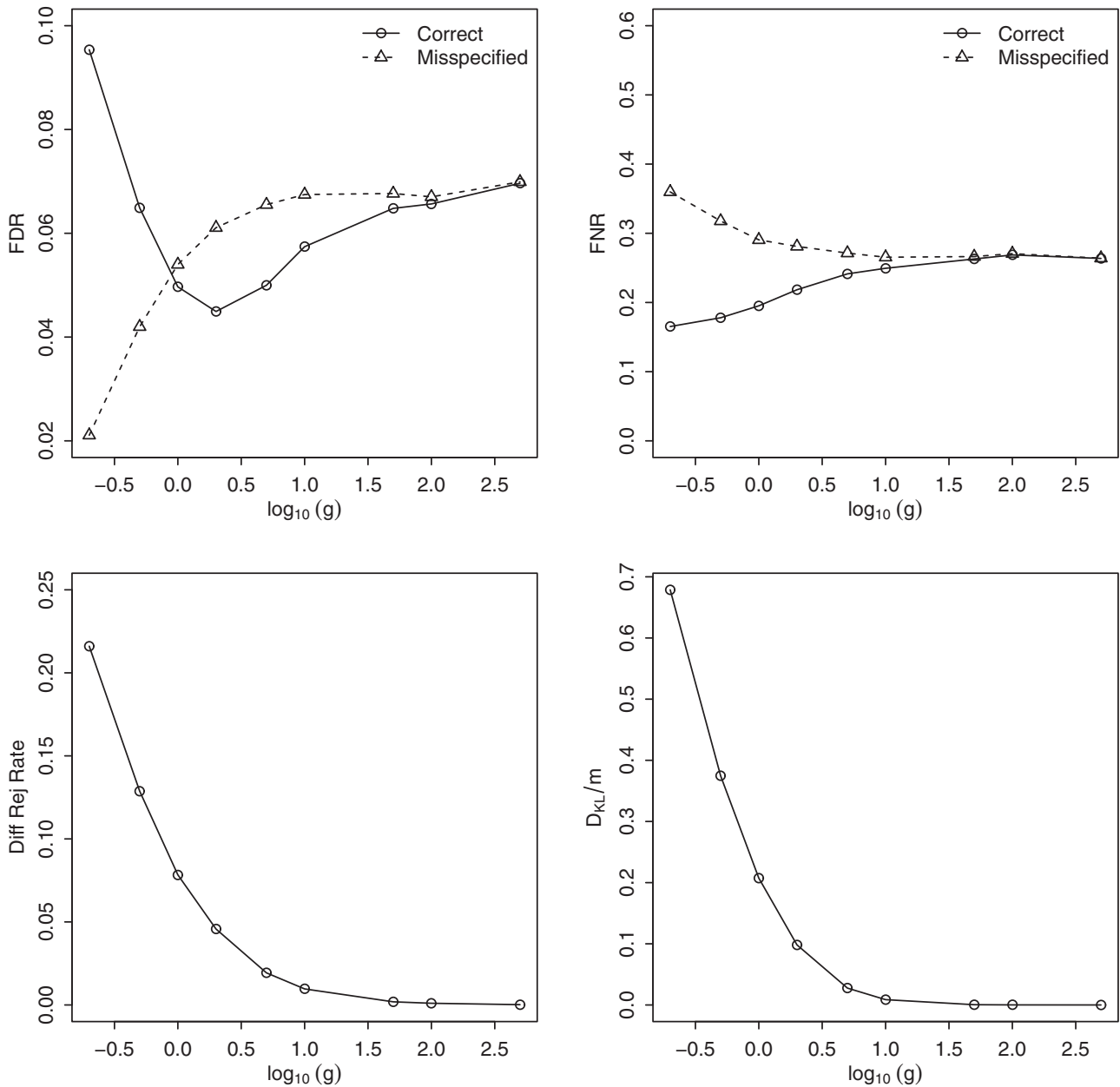
**Figure 2.** Results for Example 2.1. FDR, FNR, difference of rejection rate ($\text{rate}_{cor} - \text{rate}_{mis}$) and $D_{KL}/m$. The sequence of $g$ is from 0.2 to 500. The nominal level is 0.05.

and the FNR become closer, all of which are as expected, indicating a decreasing misspecification influence. Second, the FNR under the misspecification is universally higher than under the correct specification. Hence, misspecification results in an inefficient procedure. Also notice that the procedure tends to give less discoveries under the misspecification than under the correct specification. Last but not least, the FDR change is not monotonic and the comparison between the two specifications is profound. Notice that the nominal level is 0.05 and $g = 1$ (or $\log_{10} g = 0$) represents a 'true' scale. When both the structure $\Sigma_1$ and the scale $g$ are correct, from the top left plot in both results, we can see that the FDR is controlled at the nominal level. This seems to suggest that a correctly estimated scale of prior dependence is desired, which should be neither too strong

nor too weak. And this correct scale will only work as expected if the covariance structure is correct as well.

**Example 2.3:** Positive spatial covariance revisit.

Revisit Example 2.1. Consider to fix $g = 1$. The correct specification is exactly the same as the underlying model with $\rho = 5$. For the misspecification, let the spatial range parameter $\rho$ change from 0.1 to 20, representing the strength of dependence.

Results of Example 2.3 are shown in Figure 4. The FDR is maintained at 0.05 only when $\rho$ has the correct value. The pattern of change in FDR is also reflected in the KL divergence plot. In this example, using $D_{KL}/m$ as a measure of influence seems to be reasonable. The FNR, on the other hand, monotonically decreases as
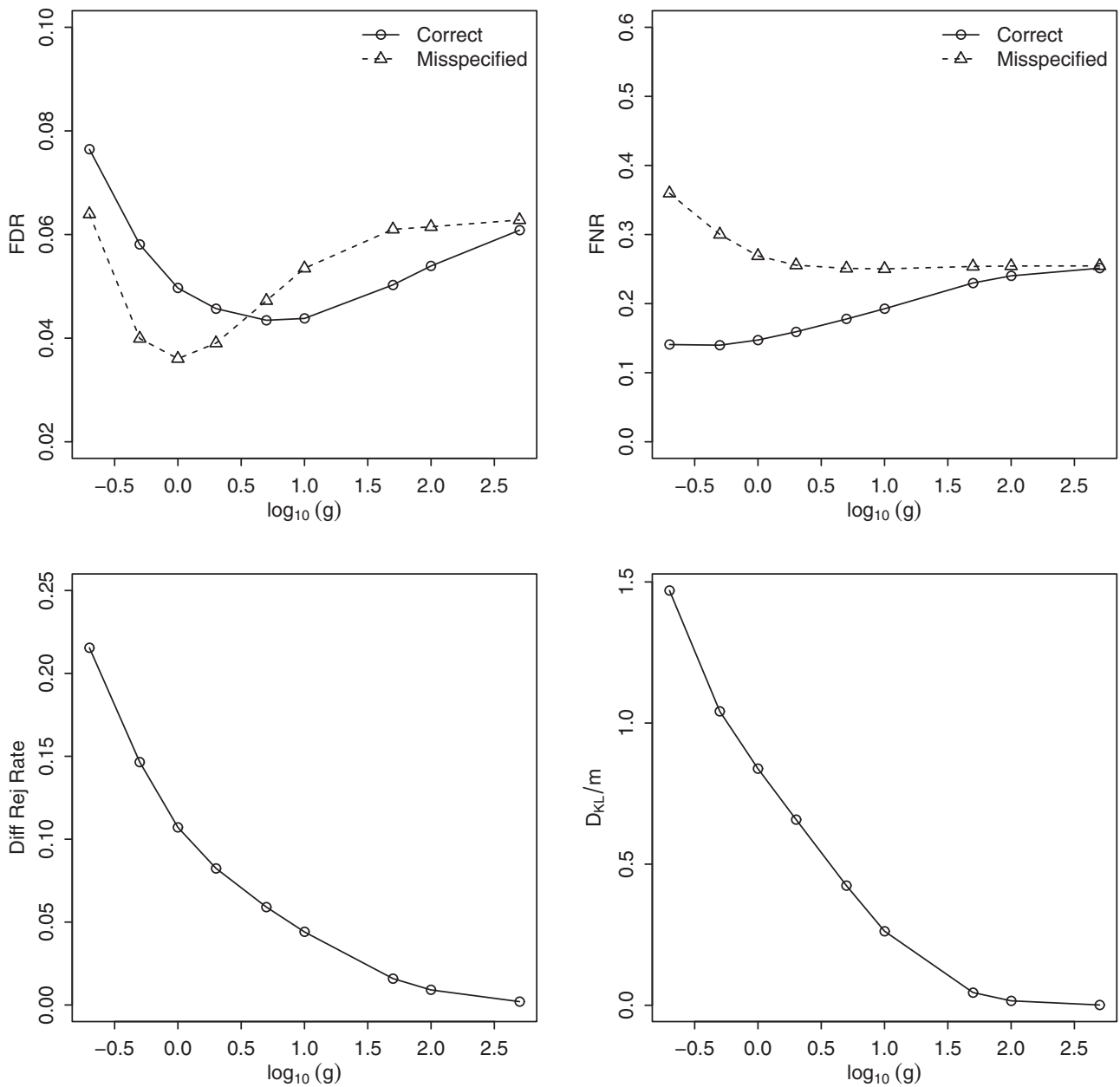
**Figure 3.** Results for Example 2.2. FDR, FNR, difference of rejection rate ($\text{rate}_{\text{cor}} - \text{rate}_{\text{mis}}$) and $D_{\text{KL}}/m$. The sequence of $g$ is from 0.2 to 500. The nominal level is 0.05.
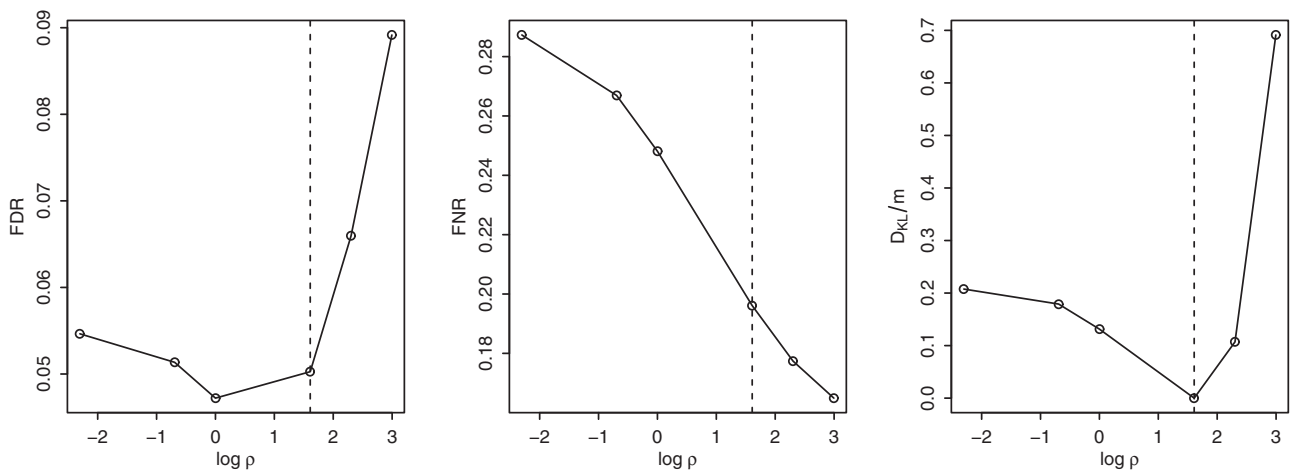


**Figure 4.** Results for Example 2.3. FDR, FNR and $D_{\text{KL}}/m$. The dashed vertical line is the correct value $\rho = 5$. The sequence of $\rho$ is (0.1, 0.5, 1, 5, 10, 20). The nominal level is 0.05.
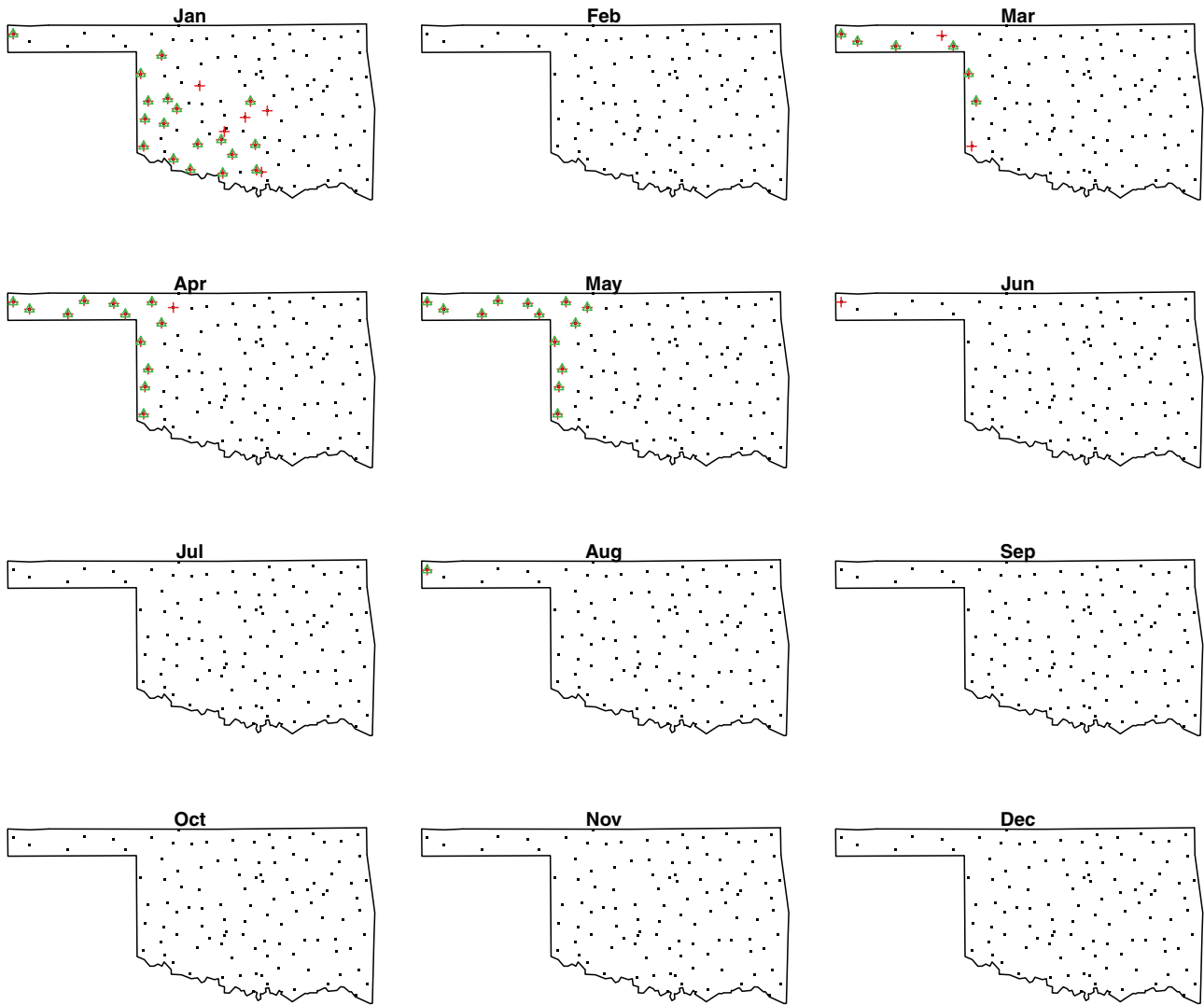
**Figure 5.** Rejected sites for the year of 2014 in Oklahoma, from January to December. The dots (·) are the observations, the crosses (+) are rejections using Model 1, and the triangles (△) are rejections using Model 2.

the dependence goes stronger. We shall note here that, when $g$ is large, the KL divergence becomes sensitive to detect a misspecification as the measure approaches zero. However, in that case, the prior is vague and the influence on FDR is negligible.

## 3. Real data: soil relative humidity

Oklahoma Mesonet (Illston et al., 2008) is a comprehensive observatory network monitoring environmental variables across the state. One of the focuses of the network is soil moisture. Extreme weather conditions, especially drought, severely impact Oklahoma's agriculture, which is a leading economy of the state. Soil moisture is fundamentally important to many hydrological, biological and biogeochemical processes. The information is valuable to a wide range of government agencies and private companies. We take a small dataset from their data warehouse as an example of real application. Consider only one variable here: the *relative humidity*, ranging from 0% to 100%. The dataset consists of

monthly averages in 2014 for 108 monitoring stations, which is in total 1296 measurements. Consider each hypothesis being $H_{0i} : \theta_i \geq 50$ versus $H_{0i} : \theta_i < 50$ for detecting low humidity times and locations.

We consider two different specifications for the dependence structure. Consider a spatio-temporal process, for a spatial location $s$ and a time point $t$, $y(s; t) = \theta(s; t) + \epsilon(s; t)$, where $\epsilon(s; t)$ is pure error process with $N(0, \sigma^2)$, and $\theta(s; t)$ is a stationary Gaussian process with a constant mean $\mu$ and a separable covariance function:

$$C(h; \tau) = \delta C^{(s)}(h) C^{(t)}(\tau),$$

where $h = \|s - s'\|$ and $\tau = |t - t'|$ are both Euclidean distances. Specify $C^{(s)}(h) = \exp\{-h/\rho\}$ and $C^{(t)}(\tau) = \alpha^\tau$. Specify priors for parameters: $\sigma^2 \sim IG(1, 1)$, $\delta \sim IG(1, 1)$, $\rho \sim Uni(0, +\infty)$ and $\alpha \sim Uni(0, 1)$. Posterior distributions are obtained through standard Markov chain Monte–Carlo (MCMC). We ensure that the chain is long enough to converge and take 10,000 MCMC

**Table 1.** Soil data: disagreed decisions under Models 1 and 2, for hypothesis $H_{0i} : \theta_i \geq 50$ versus $H_{0i} : \theta_i < 50$.

| Time and location | Observed value | Upper C.I. under Model 1 | Upper C.I. under Model 2 |
|---|---|---|---|
| January, Site 6 | 49.49 | 50.42[a] | 51.03 |
| January, Site 70 | 49.55 | 50.52[a] | 51.30 |
| January, Site 72 | 49.62 | 50.37[a] | 50.89 |
| January, Site 87 | 49.70 | 50.67[a] | 51.07 |
| January, Site 101 | 49.62 | 50.70[a] | 51.21 |
| March, Site 8 | 49.94 | 50.78[a] | 51.10 |
| March, Site 47 | 49.58 | 50.36[a] | 51.22 |
| April, Site 38 | 49.53 | 50.19[a] | 50.90 |
| June, Site 53 | 49.63 | 50.85[a] | 51.83 |

[a] A rejection under the model. The nominal level is 0.05.

samples. Posterior probabilities $P(H_{0i} \mid \boldsymbol{y})$ are approximated with posterior samples.

The second model is specified that the process is independent over time but the variance is not stationary over time:

$$C(h; t, t') = \begin{cases} \delta_t C^{(s)}(h), & \text{if } t = t'; \\ 0, & \text{otherwise.} \end{cases}$$

This is a different specification from Model 1 and neither model is a reduced case of the other. We use the same specification for $C^{(s)}(h)$ as in Model 1 and also the same prior distributions for $\sigma^2$, $\delta_t$ and $\rho$. We take 10,000 MCMC samples and posterior probabilities are approximated with MCMC samples.

For both models, we follow the FDR control procedure (4) at a nominal level 0.05. Results of both are shown in Figure 5. Note that drought impacts Oklahoma mostly in the western areas. Both results seem reasonable and meaningful for practitioners and they overlap on most decisions. However, we do observe that, at nine time/location points, they do not agree with each other. Those nine points are all rejected in Model 1 but neither in Model 2. Table 1 shows the observed values for the nine disagreed sites, along with their upper credible intervals inferred from each model. We can see that all nine points are boundary cases and Model 2 results in higher upper bounds than Model 1, causing the disagreed decisions. Such disagreed decisions will likely cause confusions in practice. As one must assume normality in the first place before performing a small sample $t$ test, we believe that, in a good practice, it is necessary to clearly assume and carefully check the model specification before using posterior probabilities from the model for testing.

## 4. Discussion

In this paper, we explore the influence of a misspecified covariance structure on the multiple testing procedure using Bayesian posterior probabilities. We explicitly show the influence on the test statistics and discuss the KL divergence as a measure of that influence. We see from a simulation study that both the correct strength of

dependence and structure of dependence are necessary to ensure control of the FDR at the nominal level. We also see that misspecified covariance can significantly impact efficiency, in terms of FNR. From a real application, we see that different covariance specifications can result in different decisions.

This paper does not cover any discussion on a precise (or two-sided) hypothesis: $H_{0i} : \theta_i = 0$ versus $H_{1i} : \theta_i \neq 0$. In that scenario, a mixture model is often assumed: $f(y) = p_0 f_0(y) + p_1 f_1(y)$. The Lfdr by Efron et al. (2001) is $p_0 f_0(y)/f(y)$ under independence. When data are dependent, it is unclear how to properly incorporate the dependence into the mixture model. One practical example given by Brown, Lazar, Datta, Jang, and McDowell (2014) specifies: $y_i \sim N(\gamma_i \mu_i, \sigma^2)$, $\gamma_i \sim$ Bern$(1 - p)$ and $\boldsymbol{\mu} \sim$ CAR$(\rho, \tau^2)$, which mimics the independent model in Scott and Berger (2006). In the Bayesian framework, to compute $P(H_{0i} \mid \boldsymbol{y})$, we would need $P(H_{0i})$, $P(H_{1i})$ and priors $\pi_{i0}(\theta_i)$ under $H_{0i}$ and $\pi_{i1}(\theta_i)$ under $H_{1i}$. Moreover, $\pi_{i0}(\theta_i)$ and $\pi_{i1}(\theta_i)$ should have a dependence structure for $i = 1, \ldots, m$, in some way. A misspecified covariance (or model) would be worth further investigation in this setting.

## Notes on contributors

*Ye Liang* is an assistant professor in the Department of Statistics at Oklahoma State University, USA. His research areas include Bayesian statistics and spatial statistics. He primarily focuses on applications in ecological, agricultural and environmental studies, and biomedical and healthcare studies.

*Joshua D. Habiger* is an associate professor in the Department of Statistics at Oklahoma State University, USA. His research interests include high dimensional statistical inference and categorical data analysis.

*Xiaoyi Min* is an assistant professor in the Department of Mathematics and Statistics at Georgia State University. His research interests include Bayesian statistics, statistical genetics and meta-analysis.

## ORCID

*Ye Liang* 🟢 http://orcid.org/0000-0002-6513-8962

## References

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

Benjamini, Y., & Yekutieli, (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics, 29*, 1165–1188.

Brown, A., Lazar, N., Datta, G., Jang, W., & McDowell, J. (2014). Incorporating spatial dependence into Bayesian multiple testing of statistical parametric maps in functional neuroimaging. *Neuroimage, 84*, 97–112.

Efron, B. (2007). Correlation and large-scale simultaneous significance testing. *Journal of the American Statistical Association, 102*, 93–103.

Efron, B. (2012). *Large-scale inference: Empirical Bayes methods for estimation, testing and prediction*. Cambridge, UK: Cambridge University Press.

Efron, B., Tibshirani, R., Storey, J., & Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association, 96*, 1151–1160.

Farcomeni, A. (2007). Some results on the control of the false discovery rate under dependence. *Scandinavian Journal of Statistics, 34*, 275–297.

Illston, B., Basara, J., Fisher, D., Elliott, R., Fiebrich, C., Crawford, K., … Hunt, E. (2008). Mesoscale monitoring of soil moisture across a statewide network. *Journal of Atmospheric and Oceanic Technology, 25*, 167–182.

Liang, F., Paulo, R., Molina, G., Clyde, M., & Berger, J. (2008). Mixtures of *g* priors for Bayesian variable selection. *Journal of the American Statistical Association, 103*, 410–423.

Liang, K., & Nettleton, D. (2012). Adaptive and dynamic adaptive procedures for false discovery rate control and estimation. *Journal of the Royal Statistical Society, Series B, 74*, 163–182.

Muller, P., Parmigiani, G., & Rice, K. (2006). FDR and Bayesian multiple comparisons rules. *Proceedings of Valencia/ISBA 8th world meeting on Bayesian statistics*, Benidorm, Alicante, Spain.

Scott, J., & Berger, J. (2006). An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference, 136*, 2144–2162.

Scott, J., & Berger, J. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics, 38*, 2587–2619.

Storey, J., Taylor, J., & Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. *Journal of the Royal Statistical Society, Series B, 66*, 187–205.

Sun, W., & Cai, T. (2009). Large-scale multiple testing under dependence. *Journal of the Royal Statistical Society, Series B, 71*, 393–424.

Sun, W., Reich, B., Cai, T., Guindani, M., & Schwartzman, A. (2015). False discovery control in large-scale spatial multiple testing. *Journal of the Royal Statistical Society, Series B, 77*, 59–83.

Wu, W. (2008). On false discovery control under dependence. *Annals of Statistics, 36*, 364–380.

## Appendix

### A.1 Proof of Lemma 2.1

**Proof:** First, given the underlying true model (5), it is straightforward to derive the true marginal distribution for $y$:

$$p(y) = \int_{\theta} p(y \mid \theta) p(\theta) d\theta = \mathrm{N}(\theta_0, \sigma_0^2 I + \Sigma_1). \quad (A1)$$

If we estimate the posterior using the correct covariance structure, we will have the following posterior distribution: $\theta \mid y \sim \mathrm{N}(\theta^{(y)}, \Sigma^{(y)})$, where $\theta^{(y)} = (1/\sigma_0^2 I + 1/g\Sigma_1^{-1})^{-1}(1/\sigma_0^2 y + 1/g\Sigma_1^{-1}\theta_0)$ and $\Sigma^{(y)} = (1/\sigma_0^2 I + 1/g\Sigma_1^{-1})^{-1}$. Marginally, $\theta_i \mid y \sim \mathrm{N}(\theta_i^{(y)}, \Sigma_{ii}^{(y)})$. Then,

$$H_i = P(H_{0i} \mid y) = P(\theta_i \geq \theta_{0i} \mid y) = \Phi\left(\frac{\theta_i^{(y)} - \theta_{0i}}{\sqrt{\Sigma_{ii}^{(y)}}}\right).$$

Using (A1), we have the marginal distribution: $\theta^{(y)} \sim \mathrm{N}(\theta_0, B)$, where $B = (I + \sigma_0^2/g\Sigma_1^{-1})^{-1}(\sigma_0^2 I + \Sigma_1)(I + \sigma_0^2/g\Sigma_1^{-1})^{-1}$. Note that $\Sigma^{(y)}$ is free of $y$, so let $A = \Sigma^{(y)}$. Marginally, $\theta_i^{(y)} \sim \mathrm{N}(\theta_{0i}, b_{ii})$ which leads to

$$\frac{\theta_i^{(y)} - \theta_{0i}}{\sqrt{\Sigma_{ii}^{(y)}}} \sim \mathrm{N}(0, b_{ii}/a_{ii}).$$

Now, under the true covariance, the CDF for $H_i$ is given by

$$\begin{aligned} F(h_i) &= P\left[\Phi\left(\frac{\theta_i^{(y)} - \theta_{0i}}{\sqrt{\Sigma_{ii}^{(y)}}}\right) \leq h_i\right] \\ &= P\left[\frac{\theta_i^{(y)} - \theta_{0i}}{\sqrt{\Sigma_{ii}^{(y)}}} \leq \Phi^{-1}(h_i)\right] \\ &= \Phi\left[\sqrt{\frac{a_{ii}}{b_{ii}}}\Phi^{-1}(h_i)\right]. \end{aligned}$$

If a misspecified covariance $\Sigma_2$ is used to estimate the posterior, then in the posterior distribution $\theta \mid y \sim \mathrm{N}(\theta^{(y)}, \Sigma^{(y)})$, we will have $\Sigma_2$, instead of $\Sigma_1$ in both $\theta^{(y)}$ and $\Sigma^{(y)}$. As a consequence, $A = (1/\sigma_0^2 I + 1/g\Sigma_2^{-1})^{-1}$ and $B = (I + \sigma_0^2/g\Sigma_2^{-1})^{-1}(\sigma_0^2 I + \Sigma_1)(I + \sigma_0^2/g\Sigma_2^{-1})^{-1}$. The rest remains the same. □

### A.2 Proof of Theorem 2.1

**Proof:** According to the proof in Lemma 2.1, $\theta^{(y)} \sim \mathrm{N}(\theta_0, B)$. Hence, $\mathrm{diag}^{-1/2}(B)(\theta^{(y)} - \theta_0) \sim \mathrm{N}(0, P_b)$, where $P_b$ is the correlation matrix of $B$. Equivalently,

$$\left(\frac{\theta_1^{(y)} - \theta_{01}}{\sqrt{b_{11}}}, \ldots, \frac{\theta_m^{(y)} - \theta_{0m}}{\sqrt{b_{mm}}}\right)' \sim \mathrm{N}(0, P_b).$$

Then, the joint CDF of $(H_1, \ldots, H_m)'$ is

$$
\begin{aligned}
F(h_1, \ldots, h_m) &= P(H_1 \leq h_1, \ldots, H_m \leq h_m) \\
&= P\left[ \Phi\left( \frac{\theta_1^{(y)} - \theta_{01}}{\sqrt{a_{11}}} \right) \leq h_1, \ldots, \right. \\
&\qquad \left. \Phi\left( \frac{\theta_m^{(y)} - \theta_{0m}}{\sqrt{a_{mm}}} \right) \leq h_m \right] \\
&= P\left[ \frac{\theta_1^{(y)} - \theta_{01}}{\sqrt{b_{11}}} \leq \sqrt{\frac{a_{11}}{b_{11}}} \Phi^{-1}(h_1), \ldots, \right. \\
&\qquad \left. \frac{\theta_m^{(y)} - \theta_{0m}}{\sqrt{b_{mm}}} \leq \sqrt{\frac{a_{mm}}{b_{mm}}} \Phi^{-1}(h_m) \right] \\
&= \Phi_b^m\left[ \sqrt{\frac{a_{11}}{b_{11}}} \Phi^{-1}(h_1), \ldots, \right. \\
&\qquad \left. \sqrt{\frac{a_{mm}}{b_{mm}}} \Phi^{-1}(h_m) \right].
\end{aligned}
$$
□

### A.3 Proof of Theorem 2.2

**Proof:** First, given the underlying true process (5), marginally, $y \sim N(\theta_0, \sigma_0^2 I + \Sigma_1)$.

If the correct covariance is used, we have the posterior distribution

$$
\begin{aligned}
p(\theta \mid y) &\propto \int_0^\infty p(y \mid \theta) p(\theta \mid \sigma^2) p(\sigma^2) d\sigma^2 \\
&\propto \left[ (y - \theta)'(y - \theta) + 1/g(\theta - \theta_0)'\Sigma^{-1}(\theta - \theta_0) + 2\beta \right]^{-m-\alpha} \\
&\propto \left[ (\theta - \theta^{(y)})'(I + 1/g\Sigma^{-1})(\theta - \theta^{(y)}) \right. \\
&\qquad \left. + (y - \theta_0)'(I + g\Sigma)^{-1}(y - \theta_0) + 2\beta \right]^{-m-\alpha} \\
&\propto \left[ 1 + \frac{1}{m + 2\alpha} (\theta - \theta^{(y)})'(V^{(y)})^{-1}(\theta - \theta^{(y)}) \right]^{-\frac{m + (m+2\alpha)}{2}},
\end{aligned}
$$

which is $t_m(m + 2\alpha, \theta^{(y)}, V^{(y)})$, with location $\theta^{(y)} = (I + 1/g\Sigma^{-1})^{-1}(y + 1/g\Sigma^{-1}\theta_0)$ and scale $V^{(y)} = (m + 2\alpha)^{-1}[2\beta + (y - \theta_0)'(I + g\Sigma)^{-1}(y - \theta_0)](I + 1/g\Sigma^{-1})^{-1}$. Similarly as Lemma 2.1, define $A = (I + 1/g\Sigma_1^{-1})^{-1}$ and $B = (I + 1/g\Sigma_1^{-1})^{-1}(\sigma_0^2 I + \Sigma_1)(I + 1/g\Sigma_1^{-1})^{-1}$.

Marginally, $\theta_i \mid y \sim t_1(m + 2\alpha, \theta_i^{(y)}, V_{ii}^{(y)})$, then each test statistic is

$$
H_i = P(H_{0i} \mid y) = P(\theta_i \geq \theta_{i0} \mid y) = \Psi_{m+2\alpha}\left( \frac{\theta_i^{(y)} - \theta_{i0}}{\sqrt{V_{ii}^{(y)}}} \right).
$$

In order to find the joint CDF, we need the joint distribution of $((\theta_1^{(y)} - \theta_{10})/\sqrt{V_{11}^{(y)}}, \ldots, (\theta_m^{(y)} - \theta_{m0})/\sqrt{V_{mm}^{(y)}})'$, which can be re-written as

$$
\begin{aligned}
&(\operatorname{diag} V^{(y)})^{-\frac{1}{2}} (\theta^{(y)} - \theta_0) \\
&= \sqrt{\frac{m + 2\alpha}{(y - \theta_0)'(I + g\Sigma_1)^{-1}(y - \theta_0) + 2\beta}} \\
&\qquad \times (\operatorname{diag} A)^{-\frac{1}{2}} (\theta^{(y)} - \theta_0). \quad (A2)
\end{aligned}
$$

Given the marginal distribution of $y$, we have $\theta^{(y)} - \theta_0 = A(y - \theta_0) \sim N(\theta_0, B)$, or, $\theta^{(y)} - \theta_0 =$

$(\operatorname{diag} B)^{1/2} z_b$. The quadratic term in (A2) is

$$
\begin{aligned}
&(y - \theta_0)'(I + g\Sigma_1)^{-1}(y - \theta_0) \\
&= (y - \theta_0)'(I - A)(y - \theta_0) \\
&= z_b'(\operatorname{diag} B)^{-1/2}(A^{-2} - A^{-1})(\operatorname{diag} B)^{-1/2} z_b.
\end{aligned}
$$

Equation (A2) is then

$$
\begin{aligned}
(\operatorname{diag} V^{(y)})^{-\frac{1}{2}} (\theta^{(y)} - \theta_0) &= \sqrt{\frac{m + 2\alpha}{z_b' C z_b + 2\beta}} (\operatorname{diag} A)^{-\frac{1}{2}} \\
&\quad \times (\operatorname{diag} B)^{\frac{1}{2}} z_b.
\end{aligned}
$$

Or, equivalently,

$$
\begin{aligned}
&\left( \sqrt{\frac{a_{11}}{b_{11}}} \frac{\theta_1^{(y)} - \theta_{10}}{\sqrt{V_{11}^{(y)}}}, \ldots, \sqrt{\frac{a_{mm}}{b_{mm}}} \frac{\theta_m^{(y)} - \theta_{m0}}{\sqrt{V_{mm}^{(y)}}} \right)' \\
&= \sqrt{\frac{m + 2\alpha}{z_b' C z_b + 2\beta}} z_b.
\end{aligned}
$$

The joint CDF of $H_1, \ldots, H_m$ is given by

$$
\begin{aligned}
F(h_1, \ldots, h_m) &= P(H_1 \leq h_1, \ldots, H_m \leq h_m) \\
&= P\left[ \Psi_{m+2\alpha}\left( \frac{\theta_1^{(y)} - \theta_{10}}{\sqrt{V_{11}^{(y)}}} \right) \leq h_1, \ldots, \right. \\
&\qquad \left. \Psi_{m+2\alpha}\left( \frac{\theta_m^{(y)} - \theta_{m0}}{\sqrt{V_{mm}^{(y)}}} \right) \leq h_m \right] \\
&= P\left[ \frac{\theta_1^{(y)} - \theta_{10}}{\sqrt{V_{11}^{(y)}}} \leq \Psi_{m+2\alpha}^{-1}(h_1), \ldots, \right. \\
&\qquad \left. \frac{\theta_m^{(y)} - \theta_{m0}}{\sqrt{V_{mm}^{(y)}}} \leq \Psi_{m+2\alpha}^{-1}(h_m) \right] \\
&= \Xi_{a,b}^m\left[ \sqrt{\frac{a_{11}}{b_{11}}} \Psi_{m+2\alpha}^{-1}(h_1), \ldots, \right. \\
&\qquad \left. \sqrt{\frac{a_{mm}}{b_{mm}}} \Psi_{m+2\alpha}^{-1}(h_m) \right].
\end{aligned}
$$

If the misspecified covariance is used, follow the same argument in Lemma 2.1. □

### A.4 Proof of Corollary 2.1

**Proof:** Notice that $d\Phi^{-1}(x)/dx = 1/\varphi(\Phi^{-1}(x))$, where $\varphi(x) = 1/\sqrt{2\pi} e^{-x^2/2}$. Then,

$$
\begin{aligned}
f(h_i) &= \varphi\left( \sqrt{r_i} \Phi^{-1}(h_i) \right) \frac{\sqrt{r_i}}{\varphi(\Phi^{-1}(h_i))} \\
&= \sqrt{r_i} \exp\left\{ \frac{1}{2} (1 - r_i)[\Phi^{-1}(h_i)]^2 \right\}.
\end{aligned}
$$
□

### A.5 Proof of Corollary 2.2

**Proof:** Let $u_i = \sqrt{r_i}\Phi^{-1}(h_i)$. Using the chain rule,

$$f(h_1, \ldots, h_m) = \frac{\partial^m F}{\partial h_1 \cdots \partial h_m} = \frac{\partial^m \Phi_b^m}{\partial u_1 \cdots \partial u_m} \prod_{i=1}^m \frac{\partial u_i}{\partial h_i}$$

$$= \varphi_b^m(u_1, \ldots, u_m) \prod_{i=1}^m \frac{\sqrt{r_i}}{\varphi(\Phi^{-1}(h_i))}$$

$$= |P_b|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} u' P_b^{-1} u\right\} \prod_{i=1}^m \sqrt{r_i}$$

$$\times \exp\left\{\frac{1}{2}[\Phi^{-1}(h_i)]^2\right\}$$

$$= \left\{\prod_{i=1}^m \sqrt{r_i}\right\} |P_b|^{-\frac{1}{2}}$$

$$\times \exp\left\{-\frac{1}{2}\phi' R^{\frac{1}{2}} P_b^{-1} R^{\frac{1}{2}} \phi + \frac{1}{2}\phi'\phi\right\}$$

$$= \left|R^{\frac{1}{2}} P_b^{-1} R^{\frac{1}{2}}\right|^{\frac{1}{2}}$$

$$\times \exp\left\{\frac{1}{2}\phi'(I - R^{\frac{1}{2}} P_b^{-1} R^{\frac{1}{2}})\phi\right\}. \qquad \square$$