1-2015

Bias and variance reduction in estimating the proportion of true-null hypotheses

Yebin Cheng
Shanghai University of Finance and Economics

Dexiang Gao
University of Colorado

Tiejun Tong
Hong Kong Baptist University, tongt@hkbu.edu.hk

Link to published article: http://dx.doi.org/10.1093/biostatistics/kxu029

Bias and Variance Reduction in Estimating the Proportion of True Null Hypotheses

Yebin Cheng[1], Dexiang Gao[2] and Tiejun Tong[3*]

[1]School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, P.R.China

[2]Department of Biostatistics and Informatics, University of Colorado, Denver, Colorado, USA

[3]Department of Mathematics, Hong Kong Baptist University, Hong Kong

Email: tongt@hkbu.edu.hk

April 30, 2014

Abstract

When testing a large number of hypotheses, estimating the proportion of true nulls, denoted by π_0, becomes increasingly important. This quantity has many applications in practice. For instance, a reliable estimate of π_0 can eliminate the conservative bias of the Benjamini-Hochberg procedure on controlling the false discovery rate. It is known that most methods in the literature for estimating π_0 are conservative. Recently, some attempts have been paid to reduce such estimation bias. Nevertheless, they are either over bias-corrected or suffering from an unacceptably large estimation variance. In this paper, we propose a new method for estimating π_0 that aims to reduce the bias and variance of the estimation simultaneously. To achieve this, we first utilize the probability density functions of false-null p-values and then propose a novel algorithm to estimate the quantity of π_0. The statistical behavior of the proposed estimator is also investigated. Finally, we carry out extensive simulation studies and several real data analysis to evaluate the performance of the proposed estimator. Both simulated and real data demonstrate that the proposed method may improve the existing literature significantly.

KEY WORDS: Effect size; False-null p-value; Microarray data; Multiple testing; Probability density function; Upper tail probability

*Corresponding author. E-mail: tongt@hkbu.edu.hk

1 Introduction

When testing a large number of hypotheses, estimating the proportion of true nulls, denoted by π_0, becomes increasingly important. Studies using high-throughput techniques and microarray experiments that identify genes expressed differentially across groups, often involve testing hundreds or thousands of hypotheses simultaneously. In addition to identifying differentially expressed genes, we may also want to know the proportion of genes that are truly differentially expressed, i.e., the value of π_0. This quantity has many applications in practice. For instance, a reliable estimate of π_0 can eliminate the conservative bias of the Benjamini-Hochberg procedure (Benjamini & Hochberg 1995) on controlling the false discovery rate, and therefore increase the average power (Storey 2002, Nguyen 2004). A good estimate of π_0 can also sharpen the Bonferroni-type family-wise error controlling procedures to improve the power and reduce the false negative rate (Hochberg & Benjamini 1990, Finner & Gontscharuk 2009). Besides the broad applications, π_0 is also a quantity of interest that has its own right (Langaas, Lindqvist & Ferkingstad 2005).

The estimation of π_0 was pioneered in Schweder & Spjøtvoll (1982), where a graphical method was applied to evaluate a large number of tests on a plot of cumulative p-values using the observed significance probabilities. They claimed that the points on the graph corresponding to the true null hypotheses should fall on a straight line and that this line can then be used to estimate π_0. Their method was further studied in Storey (2002) and Storey, Taylor & Siegmund (2004). Since then, there is a rich body of literature on the estimation of π_0. For instance, Langaas et al. (2005) proposed a new method for estimating π_0 based on a nonparametric maximum likelihood estimation of the p-value density, subject to the restriction that the density is decreasing or convex decreasing. In general, their convex density estimator based on a convex decreasing density estimation outperforms other estimators with respect to the mean squared error (MSE). Other significant works in estimating

π_0 include: the smoothing spline method in Storey & Tibshirani (2003), the moment-based methods in Dalmasso, Broet & Moreau (2005) and Lai (2007), the histogram methods in Nettleton, Hwang, Caldo & Wise (2006) and Tong, Feng, Hilton & Zhao (2013), the non-parametric method in Wu, Guan & Zhao (2006), the average estimate method in Jiang & Doerge (2008), and the sliding linear model method in Wang, Tuominen & Tsai (2011), among many others.

Assume that the test statistics are independent of each other. A straightforward model for the p-values is a two-component mixture model,

$$f(p) = \pi_0 + (1 - \pi_0)h(p), \qquad 0 \leq p \leq 1, \tag{1}$$

where the p-values under the null hypotheses follow the uniform distribution on $[0, 1]$, and the p-values under the false null hypotheses follow the distribution $h(p)$. Due to the unidentifiability problem (Genovese & Wasserman 2002, Genovese & Wasserman 2004), most existing methods aforementioned have targeted to estimate an identifiable upper bound of π_0, that is $\bar{\pi}_0 = \pi_0 + \inf_p h(p)$. As a consequence, those estimators always overestimate π_0 and we refer to them as conservative estimators. To obtain the identifiability in model (1), we need to make some assumptions on the density $h(p)$. For instance, if $\inf_p h(p) = 0$ or if $h(p)$ has a parametric form, the model will be identifiable and so we can estimate π_0 directly rather than the upper bound. Recently, some attempts have been made to the estimation of π_0, with a main focus on reducing the estimation bias (Pawitan, Murthy, Michiels & Ploner 2005, McLachlan, Bean & Jones 2006, Ruppert, Nettleton & Hwang 2007, Qu, Nettleton & Dekkers 2012). In particular, by assuming that absolute values of the noncentrality parameters (NCP) from the false null hypotheses follow a smooth distribution with density g, Ruppert et al. (2007) developed a new methodology that combines a parametric model for the p-values given the NCPs and a nonparametric spline model for the NCPs. The quantity π_0 and the coefficients in the spline model were then estimated by penalized least squares. In simulation studies,

the authors demonstrated that their proposed estimator has the ability to reduce the bias in estimating π_0. More recently, their method was improved by Qu et al. (2012) where the authors applied some new nonparametric and semiparametric methods to the estimation of the NCPs distribution. We refer to these estimators as bias-reduced estimators.

Though the existing bias-reduced estimators have significant merit in reducing the estimation bias, we note that the variations of these estimators are usually considerably enlarged. As reported in Tables 1 and 2 in Qu et al. (2012), the interquartile ranges of their estimators are often more than twice as large as the other competitors. In addition, we observe that their estimators only perform well when the NCPs are concentrated around zero, i.e., when a majority of false nulls have very weak signals. In the situation of microarray data analysis, to make their estimators work a large proportion of false-null genes need to be weakly differentially expressed. Otherwise, their estimators tend to be over bias-corrected (see Section 5 for more detail). In this paper, we propose a new method for estimating π_0 that aims to reduce the bias and variance of the estimation simultaneously. To achieve this, we first utilize the probability density functions of false-null p-values and then propose a novel algorithm to estimate the quantity of π_0. Simulation studies will show that the proposed method may improve the existing literature significantly.

The rest of the paper is organized as follows. In Section 2, we introduce a bias-corrected method for estimating π_0 that aims to reduce the bias and variance simultaneously. In Section 3, we derive the probability density functions of false-null p-values for testing two-sided hypotheses with unknown variances. In Section 4, we propose an algorithm for estimating π_0 and investigate the behavior of the proposed estimator. We then evaluate the performance of the proposed estimator via extensive simulation studies in Section 5 and several microarray data sets in Section 6. Finally, we conclude the paper in Section 7 and provide the technical proofs in the Appendices.

2 Main Results

Let $p_1, \ldots, p_m$ be the p-values corresponding to each of m total hypothesis tests. Let $\mathcal{M}_0$ (size m_0) denote the set of true null hypotheses and $\mathcal{M}_1$ (size m_1) the set of false null hypotheses. Then $m = m_0 + m_1$ and $\pi_0 = m_0/m$. To avoid confusion, we define the "true-null p-values" as p-values from hypothesis tests in which the null was correct, and the "false-null p-values" as p-values from hypothesis tests in which the null was false. For a given $\lambda \in (0, 1)$, define $W(\lambda) = \#\{p_i > \lambda\}$ to be the total number of p-values on $(\lambda, 1]$, $W_0(\lambda) = \#\{\text{true-null } p_i > \lambda\}$ to be the total number of true-null p-values on $(\lambda, 1]$, and $W_1(\lambda) = \#\{\text{false-null } p_i > \lambda\}$ to be the total number of false-null p-values on $(\lambda, 1]$. By definition, we have $W(\lambda) = W_0(\lambda) + W_1(\lambda)$. In addition, we have $E[W_0(\lambda)] = m\pi_0(1 - \lambda)$ since the true-null p-values are uniformly distributed in $[0, 1]$. This suggests we estimate π_0 by

$$\hat{\pi}_0(\lambda) = \frac{W_0(\lambda)}{m(1 - \lambda)}. \tag{2}$$

However, (2) is not a valid estimator as $W_0(\lambda)$ is unobservable in practice.

Note that the false-null p-values are more likely to be small. Thus for a reasonably large λ, the majority of p-values on $(\lambda, 1]$ should correspond to true-null p-values and so $W_0(\lambda) \approx W(\lambda)$. By this, Storey (2002) proposed to estimate π_0 by

$$\hat{\pi}_0^S(\lambda) = \frac{W(\lambda)}{m(1 - \lambda)}, \tag{3}$$

where λ is the tuning parameter. We refer to $\hat{\pi}_0^S(\lambda)$ as the Storey estimator. For any $0 < \lambda < 1$, it is easy to verify that

$$E[\hat{\pi}_0^S(\lambda)] = \frac{E(W_0(\lambda))}{m(1 - \lambda)} + \frac{E(W_1(\lambda))}{m(1 - \lambda)} = \pi_0 + \frac{E(W_1(\lambda))}{m(1 - \lambda)} \geq \pi_0.$$

This shows that $\hat{\pi}_0^S(\lambda)$ always overestimates π_0, and therefore, is a conservative estimator of π_0. The conservativeness of $\hat{\pi}_0^S(\lambda)$ can be rather significant when the sample size and/or the effect sizes of false-null hypotheses are small.

2.1 New Methodology

We now propose a bias-corrected method for estimating π_0. For each false-null p-value with effect size δ_i, let $f_{\delta_i}(p)$ be the probability density function and $Q_{\delta_i}(\lambda) = P(p_i > \lambda) = \int_\lambda^1 f_{\delta_i}(p)dp$ be the upper tail probability on $(\lambda, 1]$. By the definition of $W_1(\lambda)$, we have

$$E[W_1(\lambda)] = \sum_{i \in \mathcal{M}_1} Q_{\delta_i}(\lambda) = m(1-\pi_0)Q(\lambda), \tag{4}$$

where $Q(\lambda) = \sum_{i \in \mathcal{M}_1} Q_{\delta_i}(\lambda)/m_1 = \sum_{i \in \mathcal{M}_1} Q_{\delta_i}(\lambda)/[m(1-\pi_0)]$ is the average upper tail probability for all false-null p-values. By (4) and the fact that $E[W_0(\lambda)] = m\pi_0(1-\lambda)$, we have $E[W(\lambda)] = E[W_0(\lambda)] + E[W_1(\lambda)] = m\pi_0(1-\lambda) + m(1-\pi_0)Q(\lambda)$. This leads to

$$\pi_0 = \frac{E[W(\lambda)] - mQ(\lambda)}{m(1-\lambda) - mQ(\lambda)}. \tag{5}$$

Let $\widehat{Q}(\lambda)$ be an estimate of $Q(\lambda)$. By (5), we propose a new estimator of π_0 as

$$\hat{\pi}_0^U(\lambda) = \frac{W(\lambda) - m\widehat{Q}(\lambda)}{m(1-\lambda) - m\widehat{Q}(\lambda)}. \tag{6}$$

Note that $\hat{\pi}_0^U(\lambda)$ is not guaranteed to be within $[0, 1]$ in practice. As in Storey (2002), we truncate $\hat{\pi}_0^U(\lambda)$ to 1 if $\hat{\pi}_0^U(\lambda) > 1$, and round $\hat{\pi}_0^U(\lambda)$ to 0 if $\hat{\pi}_0^U(\lambda) < 0$. This leads to the estimator to be $\min\{1, \max\{0, \hat{\pi}_0^U(\lambda)\}\}$.

The term $\widehat{Q}(\lambda)$ serves as a regularization parameter of the proposed estimator. When $\widehat{Q}(\lambda) = 0$, $\hat{\pi}_0^U(\lambda)$ reduces to $\hat{\pi}_0^S(\lambda)$. When $\widehat{Q}(\lambda) > 0$, in Appendix C we show that

$$\min\left\{1, \max\left\{0, \frac{W(\lambda) - m\widehat{Q}(\lambda)}{m(1-\lambda) - m\widehat{Q}(\lambda)}\right\}\right\} < \min\left\{1, \frac{W(\lambda)}{m(1-\lambda)}\right\}$$

for any $\lambda \in (0, 1)$. That is, the proposed estimator is always less conservative than Storey's estimator for any λ. More discussion on $\widehat{Q}(\lambda)$ is given in Sections 3 and 4.

Finally, in addition to the bias elimination, we apply the average estimate method in Jiang & Doerge (2008) to further reduce the estimation variance. Specifically, let $\Lambda =$

$\{a + k(b-a)/\tau,\ k = 0, \ldots, \tau\}$ where $0 < a < b < 1$ and τ is an integer value. We then compute $\hat{\pi}_0^U(\lambda)$ for each $\lambda \in \Lambda$ and take their average as the final estimate,

$$\hat{\pi}_0^U = \frac{1}{J} \sum_{\lambda_j \in \Lambda} \min\{1, \max\{0, \hat{\pi}_0^U(\lambda_j)\}\}, \tag{7}$$

where $J = \tau + 1$ is the number of λ contained in the set Λ. We note that the average estimate method is very robust when the independence assumption is violated.

2.2 Choice of the set Λ

Needless to say, the set Λ may play an important role for the proposed estimator. In what follows we investigate the choice of an appropriate set Λ in practice. Recall that for the estimator $\hat{\pi}_0^S(\lambda)$ in (3), there is a severe bias-variance trade-off on the tuning parameter λ. Specifically, (i) when $\lambda \to 0$, the variance of $\hat{\pi}_0^S(\lambda)$ is smaller but the bias increases; and (ii) when $\lambda \to 1$, the bias of $\hat{\pi}_0^S(\lambda)$ is smaller but the variance increases. In practice, the optimal λ is suggested to be the one that minimizes the MSE and is implemented by a bootstrap procedure in Storey et al. (2004).

We note that, unlike the Storey estimator $\hat{\pi}_0^S(\lambda)$, the proposed estimator $\hat{\pi}_0^U(\lambda)$ in (6) has little bias and does not suffer a severe bias-variance trade-off along with the choice of λ. Thus, to choose an appropriate λ value, we can aim to minimize the variance of the estimator only. Simulations (not shown) indicate that the variance of $\hat{\pi}_0^U(\lambda)$ is usually larger when λ is near 0 or 1 than when it is near the middle of the range. In addition, from a theoretical point of view, we found that $\lim_{\lambda \to 0}[m(1-\lambda) - m\widehat{Q}(\lambda)] = \lim_{\lambda \to 1}[m(1-\lambda) - m\widehat{Q}(\lambda)] = 0$ for the proposed $\widehat{Q}(\lambda)$ in Section 4.1. Recall that $m(1-\lambda) - m\widehat{Q}(\lambda)$ is the denominator of (6). This implies that $\hat{\pi}_0^U(\lambda)$ may not be stable and may have a large variation when λ is near 0 or 1. That is, to make $\hat{\pi}_0^U(\lambda)$ a good estimate the λ value should not be too small or large. In Appendix D, a simulation study is conducted that investigates how sensitive the method is to the choice of boundaries a, b and τ. According to the simulation results, we

apply the set $\Lambda = \{0.20, 0.25, \ldots, 0.50\}$ throughout the paper.

3 Probability density function of false-null p-values

Given the set Λ, to implement the estimator (7) we need to have an appropriate estimate for the unknown quantity $Q(\lambda)$. To achieve this, we need to have the probability density functions $f_{\delta_i}(p)$ for each false-null p-value with effect size δ_i, where $i \in \mathcal{M}_1$. For ease of notation, in this section we will not specify the subscript i in effect sizes unless otherwise specified. Our aim is then to determine the probability density function $f_\delta(p)$ of a false-null p-value with effect size δ.

For the one-sample comparison, let $X_1, \ldots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Let $\bar{X} = \sum_{i=1}^n X_i/n$ be the sample mean, $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ be the sample variance, and $\delta = \mu/\sigma$ be the effect size. For testing the one-sided hypothesis

$$H_0 : \mu = 0 \qquad \text{versus} \qquad H_1 : \mu > 0, \tag{8}$$

Hung, O'nell, Bauer & Köhne (1997) assumed a known σ^2 and considered the test statistic $T = \sqrt{n}\bar{X}/\sigma$. Under H_0, the test statistic T follows a standard normal distribution. This yields a p-value of $p = 1 - \Phi(t)$, where t is the realization of T and $\Phi(\cdot)$ is the probability function of the standard normal distribution. Under H_1, T is normally distributed with mean $\sqrt{n}\delta$ and variance 1. Then, by Jacobian transformation, for given n and δ the probability density function of p is

$$f_\delta(p) = \frac{\phi(z_p - \sqrt{n}\delta)}{\phi(z_p)}, \qquad 0 < p < 1, \tag{9}$$

where z_p is the $(1-p)$th percentile of the standard normal distribution. Further, we have

$$Q_\delta(\lambda) = \int_\lambda^1 \frac{\phi(z_p - \sqrt{n}\delta)}{\phi(z_p)} dp = \Phi(z_\lambda - \sqrt{n}\delta), \qquad 0 < \lambda < 1. \tag{10}$$

Needless to say, the assumption of known variances and also the restriction to one-sided tests in Hung et al. (1997) limited its application in testing the differential expression of genes. The small sample size in such studies can be another concern. Hence, to accommodate the needs of microarray studies, we extend their method to the two-sided testing problems with unknown variances.

3.1 Two-sided tests with unknown variances

We first consider the one-sample, two-sided comparison. For testing the hypothesis

$$H_0 : \mu = 0 \qquad \text{versus} \qquad H_1 : \mu \neq 0. \tag{11}$$

We consider the test statistic $T = \sqrt{n}\bar{X}/S$, where $\bar{X} = \sum_{i=1}^{n} X_i/n$ and $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$ are the sample mean and sample variance respectively. Let $\delta = \mu/\sigma$ be the effect size. Under H_0, the test statistic T follows a Student's t distribution with $\nu = n - 1$ degrees of freedom.

The p-value for testing (11) is given as $p = 2 - 2F_\nu(|t|)$, where t is the realization of T, $\nu = n - 1$ and $F_\nu(\cdot)$ is the probability function of Student's t distribution with ν degrees of freedom. Under H_1, it is easy to verify that T follows a non-central t distribution with ν degrees of freedom and non-centrality parameter $\sqrt{n}\delta$. Let $t_\nu(p)$ be the $(1-p)$th percentile of Student's t distribution with ν degrees of freedom. In Appendix B, for any given n and δ, we show that the probability density function of p is

$$f_\delta(p) = \frac{f_{\nu,\sqrt{n}\delta}(t_\nu(p/2))}{2f_\nu(t_\nu(p/2))} + \frac{f_{\nu,\sqrt{n}\delta}(-t_\nu(p/2))}{2f_\nu(t_\nu(p/2))}, \qquad 0 < p < 1. \tag{12}$$

where $f_\nu(\cdot)$ is the probability density function of Student's t distribution with ν degrees of freedom, and $f_{\nu,\sqrt{n}\delta}(\cdot)$ is the probability density function of the non-central t distribution with ν degrees of freedom and non-centrality parameter $\sqrt{n}\delta$. When $\delta = 0$, both $f_{\nu,\sqrt{n}\delta}(t_\nu(p))$ and $f_{\nu,\sqrt{n}\delta}(-t_\nu(p))$ reduce to $f_\nu(t_\nu(p))$ so that p follows a uniform distribution in $[0, 1]$. When

$\delta \neq 0$, we have

$$Q_\delta(\lambda) = F_{\nu,\sqrt{n}\delta}(t_\nu(\lambda/2)) - F_{\nu,\sqrt{n}\delta}(-t_\nu(\lambda/2)), \qquad 0 < \lambda < 1, \tag{13}$$

where $F_{\nu,\sqrt{n}\delta}(\cdot)$ is the probability function of the non-central t distribution with ν degrees of freedom and non-centrality parameter $\sqrt{n}\delta$.

Now we consider the two-sample, two-sided comparison. Let $X_{11}, \ldots, X_{1n_1}$ be a random sample of size n_1 from the normal distribution with mean μ_1 and variance σ_1^2, and $X_{21}, \ldots, X_{2n_2}$ be a random sample of size n_2 from the normal distribution with mean μ_2 and variance σ_2^2. Let also $\bar{X}_1 = \sum_{i=1}^{n_1} X_{1i}/n_1$ and $\bar{X}_2 = \sum_{i=1}^{n_2} X_{2i}/n_2$ be the sample means for the two samples, respectively. For testing the hypothesis

$$H_0 : \mu_1 = \mu_2 \qquad \text{versus} \qquad H_1 : \mu_1 \neq \mu_2, \tag{14}$$

we consider the test statistic $T = (\bar{X}_1 - \bar{X}_2)/\sqrt{S_{pool}^2(1/n_1 + 1/n_2)}$, where $S_{pool}^2 = [(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2]/(n_1 + n_2 - 2)$ is the pooled sample variance with $S_1^2 = \sum_{i=1}^{n_1}(X_{1i} - \bar{X}_1)^2/(n_1 - 1)$ and $S_2^2 = \sum_{i=1}^{n_2}(X_{2i} - \bar{X}_2)^2/(n_2 - 1)$. Under H_0, T follows a Student's t distribution with $n_1 + n_2 - 2$ degrees of freedom. Under H_1, T follows a non-central t distribution with $n_1 + n_2 - 2$ degrees of freedom and non-centrality parameter $(\mu_1 - \mu_2)\sigma^{-1}\sqrt{n_1 n_2/(n_1 + n_2)}$. Thus to make formulas (12) and (13) applicable to the two-sample comparison, we only need to redefine ν, n and δ as follows: $\nu = n_1 + n_2 - 2$, $n = n_1 n_2/(n_1 + n_2)$ and $\delta = (\mu_1 - \mu_2)/\sigma$. Finally, if a common variance in (14) is not assumed, we may apply Welch's t test statistic and it follows an approximate t distribution.

4 The proposed algorithm for estimating π_0

For the one-sample comparison, an intuitive estimator of $\delta_i = \mu_i/\sigma_i$ is given as $\tilde{\delta}_i = \bar{X}_i/S_i$, where $\bar{X}_i$ is the sample mean and S_i is the sample standard deviation. However, $\tilde{\delta}_i$ is suboptimal as it is biased. Alternatively, because $\bar{X}_i$ and S_i are independent of each other,

we have

$$E(\tilde{\delta}_i) = \frac{\sqrt{n-1}}{\sigma_i} E(\bar{X}_i) E(U_i^{1/2}) = \frac{\sqrt{n-1}}{\sqrt{2}} \frac{\Gamma(n/2-1)}{\Gamma((n-1)/2)} \delta_i, \tag{15}$$

where $U_i = \sigma_i^2/[(n-1)S_i^2]$ follows an inverse-chi-square distribution with $n-1$ degrees of freedom, $E(U_i^{1/2}) = \Gamma[n/2-1]/[\sqrt{2}\,\Gamma((n-1)/2)]$, and $\Gamma(\cdot)$ is the gamma function. By (15), an unbiased estimator of $\delta_i = \mu_i/\sigma_i$ is given as

$$\hat{\delta}_i = \frac{\sqrt{2}}{\sqrt{n-1}} \frac{\Gamma((n-1)/2)}{\Gamma(n/2-1)} \frac{\bar{X}_i}{S_i}. \tag{16}$$

Similarly, for the two-sample comparison, an unbiased estimator of $\delta_i = (\mu_{i1} - \mu_{i2})/\sigma_i$ is

$$\hat{\delta}_i = \frac{\sqrt{2}}{\sqrt{n_1+n_2-2}} \frac{\Gamma((n_1+n_2-2)/2)}{\Gamma((n_1+n_2-3)/2)} \frac{\bar{X}_i - \bar{Y}_i}{S_{i,pool}}, \tag{17}$$

where $S_{i,pool}^2 = [(n_1-1)S_{i1}^2 + (n_2-1)S_{i2}^2]/(n_1+n_2-2)$ is the pooled sample variance.

4.1 Algorithm for estimating π_0

For the sake of brevity, we present in this section the estimation procedure for the one-sample, two-sided comparison only. Note that the procedure is generally applicable when estimating π_0 in other settings. The proposed algorithm for estimating π_0 is as follows.

(a) For each $i = 1, \ldots, m$, we estimate δ_i by the unbiased estimator $\hat{\delta}_i$ in (16).

(b) For each $i = 1, \ldots, m$ and $\lambda_j \in \Lambda$, we estimate the upper tail probability $Q_{\delta_i}(\lambda_j)$ by

$$\widehat{Q}_{\delta_i}(\lambda_j) = F_{\nu,\sqrt{n}\hat{\delta}_i}(t_\nu(\lambda_j/2)) - F_{\nu,\sqrt{n}\hat{\delta}_i}(-t_\nu(\lambda_j/2)), \qquad \nu = n-1. \tag{18}$$

We then order the values of $\widehat{Q}_{\delta_1}(\lambda_j)$, $\ldots$, $\widehat{Q}_{\delta_m}(\lambda_j)$ for each λ_j such that

$$\widehat{Q}_{(1)}(\lambda_j) \le \cdots \le \widehat{Q}_{(m)}(\lambda_j).$$

(c) Let $d = [m * (1 - \hat{\pi}_0^I)]$, where $\hat{\pi}_0^I$ is an initial estimate of π_0 and $[x]$ is the integral part of x. Then for each $\lambda_j \in \Lambda$, we estimate the average upper tail probability $Q(\lambda_j)$ by

$$\widehat{Q}(\lambda_j) = \frac{1}{d} \sum_{i=1}^{d} \widehat{Q}_{(i)}(\lambda_j), \tag{19}$$

(d) Given the estimates $\widehat{Q}(\lambda_j)$ for all $\lambda_j \in \Lambda$, we estimate π_0 by

$$\hat{\pi}_0^U = \frac{1}{\#(\Lambda)} \sum_{\lambda_j \in \Lambda} \min \left\{ 1, \max \left\{ 0, \frac{W(\lambda_j) - m\widehat{Q}(\lambda_j)}{m(1-\lambda_j) - m\widehat{Q}(\lambda_j)} \right\} \right\}. \tag{20}$$

We note that the initial estimate of π_0 may play an important role for the proposed estimation procedure. When the initial estimate $\hat{\pi}_0^I$ is too large, $d = [m * (1 - \hat{\pi}_0^I)]$ tends to be small and so is $\widehat{Q}(\lambda_j)$. As a consequence, the bias-correction of $\hat{\pi}_0^U(\lambda_j)$ over $\hat{\pi}_0^I(\lambda_j)$ may not be observable. On the other hand, when the initial estimate $\hat{\pi}_0^I$ is too small, it may result in an over bias-corrected estimate. In Appendix E, a simulation study is conducted that investigates how sensitive the method is to the choice of the initial estimator of π_0. According to the simulation results, we adopt the bootstrap estimator $\hat{\pi}_0^B$ in Storey et al. (2004) as the initial estimate of π_0 in the proposed algorithm.

4.2 Behavior of the proposed estimator

The following result shows that the proposed estimator is always less conservative than the estimator of Jiang & Doerge (2008).

Theorem 1. *For any given λ set Λ, the proposed $\hat{\pi}_0^U$ in (20) is a less conservative estimator of π_0 than the average estimate $\hat{\pi}_0^A$ in Jiang & Doerge (2008), where*

$$\hat{\pi}_0^A = \frac{1}{\#(\Lambda)} \sum_{\lambda_j \in \Lambda} \min \left\{ 1, \frac{W(\lambda_j)}{m(1-\lambda_j)} \right\}. \tag{21}$$

The proof of Theorem 1 is given in Appendix C. In addition, under certain conditions we can show that $E(\hat{\pi}_0^U)$ is asymptotically larger than π_0 so that the bias of $\hat{\pi}_0^U$ is not over corrected. Specifically, we assume that (i) the initial estimate $\hat{\pi}_0^U > \pi_0$; and (ii) $\{\hat{\delta}_i, i \in \mathcal{M}_1\}$ is a random sample from a certain distribution with a finite second moment. By (i), we have $d = [m * (1 - \hat{\pi}_0^U)] \leq m * (1 - \pi_0) = m_1$ and so

$$\widehat{Q}(\lambda_j) = \frac{1}{d} \sum_{i=1}^{d} \widehat{Q}_{(i)}(\lambda_j) \leq \frac{1}{m_1} \sum_{i=1}^{m_1} \widehat{Q}_{(i)}(\lambda_j) \leq \frac{1}{m_1} \sum_{i \in \mathcal{M}_1} \widehat{Q}_{\delta_i}(\lambda_j).$$

By (ii) and by the strong law of large numbers, we have $\sum_{i\in\mathcal{M}_1}\widehat{Q}_{\delta_i}(\lambda_j)/m_1 \overset{a.s.}{\to} Q(\lambda_j)$ as $m_1 \to \infty$. Alternatively if the sample size $n \to \infty$, by (16) we have $\hat{\delta}_i \overset{a.s.}{\to} \delta_i$ and $\widehat{Q}_{\delta_i}(\lambda_j) \overset{a.s.}{\to} Q_{\delta_i}(\lambda_j)$. Then for any fixed m_1, $\sum_{i\in\mathcal{M}_1}\widehat{Q}_{\delta_i}(\lambda_j)/m_1 \overset{a.s.}{\to} Q(\lambda_j)$ as $n \to \infty$. This shows that the proposed estimator protects from over bias-correction and so is an asymptotically conservative estimator. In this sense, the proposed estimator improved the bias-reduced estimators in Ruppert et al. (2007) and Qu et al. (2012). Finally, we hope to clarify that the assumptions made above are very strong and may not hold in practice. Further research is warranted to investigate the statistical properties of the proposed estimator.

5 Simulation studies

In this section we conduct simulation studies to assess the performance of the proposed estimator under various simulation settings. The five estimators we adopt for comparison are (1) the bootstrap estimator $\hat{\pi}_0^B$ in Storey et al. (2004), (2) the average estimate estimator $\hat{\pi}_0^A$ in Jiang & Doerge (2008), (3) the convex estimator $\hat{\pi}_0^C$ in Langaas et al. (2005), (4) the parametric estimator $\hat{\pi}_0^P$ in Qu et al. (2012) and (5) the proposed estimator $\hat{\pi}_0^U$.

5.1 Simulation setup

Consider a microarray experiment with m genes and n arrays. In this study, we set $m = 1000$ and consider $n = 5$ and 10. The m-dimensional arrays are generated from a multivariate normal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)^T$ and covariance matrix Σ. To mimic a realistic scenario, we assume the covariance matrix is a block diagonal matrix such that

$$\Sigma = \begin{pmatrix} \sigma_1^2\Sigma_\rho & 0 & \cdots & 0 \\ 0 & \sigma_2^2\Sigma_\rho & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r^2\Sigma_\rho \end{pmatrix}_{m\times m},$$

where $r = m/b$ and $\Sigma_\rho = (\rho^{|i-j|})_{b\times b}$ follows an auto-regressive structure. Let $b = 50$ throughout the simulation studies. We consider four different values of ρ, ranging from 0, 0.4 to

0.8, to represent different levels of dependence. Note that $\rho = 0$ corresponds to a diagonal matrix and so is the situation where all the genes are independent of each other. Finally, we simulate $\sigma_1^2, \ldots, \sigma_r^2$ independent and identically distributed (i.i.d) from the chi-square distribution $\chi_{10}^2/10$ to account for the heterogeneity of variance in genes.

The next step is to split the m genes with $m_0 = m\pi_0$ constant genes corresponding to the true null hypotheses, and $m_1 = m(1 - \pi_0)$ differential expressed genes corresponding to the false null hypotheses. To achieve this, we first randomly sample a set of m_0 numbers, denoted by $\mathcal{I}_0$, from the integer set $S = \{1, \ldots, m\}$. Let $\mathcal{I}_1$ be the complement set so that $\mathcal{I}_0 \cup \mathcal{I}_1 = S$. We then assign $\mu_i = 0$ for each $i \in \mathcal{I}_0$, and simulate μ_i i.i.d. from the uniform distribution on the interval $[0.5, 1.5]$ for each $i \in \mathcal{I}_1$. In other words, we specify the mean vector as $\boldsymbol{\mu} = \{\mu_i = 0, \forall i \in \mathcal{I}_0\} \cup \{\mu_i \neq 0, \forall i \in \mathcal{I}_1\}$. For a complete comparison, we consider 9 values of π_0, ranging from 0.1, 0.2 to 0.9, to represent different levels of proportion of true null hypotheses.

For each combination of ρ and π_0, we first generate $\boldsymbol{\mu}$ and Σ using the algorithm specified above. We then simulate the n arrays $\boldsymbol{X}_i = (X_{1i}, \ldots, X_{mi})^T$, $i = 1, \ldots, n$, independently from the multivariate normal distribution with the generated mean $\boldsymbol{\mu}$ and covariance matrix Σ. To test the hypotheses $H_0 : \mu_i = 0$ versus $H_1 : \mu_i \neq 0$, we let $T_i = \sqrt{n}\bar{X}_i/S_i$, where $\bar{X}_i$ and S_i are the sample mean and sample standard deviation of gene $i = 1, \ldots, m$, respectively. We then compute the p-values as $p_i = 2 - 2F_{n-1}(|t_i|)$, with t_i the realization of T_i, and estimate the estimators $\hat{\pi}_0^B$, $\hat{\pi}_0^C$, $\hat{\pi}_0^A$ and $\hat{\pi}_0^U$ using the computed p-values.

5.2 Simulation results

Following the above procedure, we simulate $N = 1000$ sets of independent data for each combination setting of (n, ρ, π_0). For each method, we compute the MSE as

$$\text{MSE}(\hat{\pi}_0) = (\hat{\pi}_{\bar{0}} - \pi_0)^2 + \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\pi}_0^{(i)} - \hat{\pi}_{\bar{0}}\right)^2,$$

where $\hat{\pi}_0^{(i)}$ is the estimated π_0 for the ith simulated data set and $\hat{\pi}_{\bar{0}} = \sum_{i=1}^{N} \hat{\pi}_0^{(i)}/N$ is the sample average. We report the MSEs of the five estimators as functions of the true π_0 in Figure 1 for $n = 5$ and 10 and $\rho = 0$, 0.4 and 0.8, respectively. It is evident that the proposed $\hat{\pi}_0^U$ provides a smaller MSE than the other four estimators in most settings. Specifically, we note that (i) for small and moderate π_0 values, the proposed $\hat{\pi}_0^U$ is always the best estimator; and (ii) for large π_0 values, the proposed $\hat{\pi}_0^U$ is in a league with $\hat{\pi}_0^A$ and $\hat{\pi}_0^C$ that provides the best performance. We note that the comparison results among $\hat{\pi}_0^B$, $\hat{\pi}_0^A$ and $\hat{\pi}_0^C$ remain similar to those reported in Langaas et al. (2005) and Jiang & Doerge (2008). In addition, the estimator $\hat{\pi}_0^P$ is always suboptimal throughout the simulations.

To visualize how the proposed method improves the existing methods, we plot the density estimates of the distributions of the estimators in Figure 2 for $n = 5$ and in Figure 3 for $n = 10$. To save space, we only present the results for $\pi_0 = 0.3$, 0.6 and 0.9 and $\rho = 0$ and 0.8; the comparison patterns for other combination settings remain similar. From the densities, we note that (i) for small π_0 values such as $\pi_0 = 0.3$, the proposed $\hat{\pi}_0^U$ provides to be an unbiased estimator or a slightly underestimated estimator, whereas $\hat{\pi}_0^P$ underestimates π_0 and the other three overestimate π_0; (ii) for moderate π_0 values such as $\pi_0 = 0.6$, the proposed $\hat{\pi}_0^U$ proves to be an unbiased estimator or slightly overestimates π_0, whereas the other four estimators keep the pattern as that for $\pi_0 = 0.3$; and (iii) for large π_0 values such as $\pi_0 = 0.9$, all five estimators tend to have a small bias, whereas $\hat{\pi}_0^B$ and $\hat{\pi}_0^P$ perform worst due to the large variability in the estimation. In addition, $\hat{\pi}_0^U$ and $\hat{\pi}_0^C$ perform very similarly for $\pi = 0.9$ no matter what values of n and ρ are used. Finally, it is noteworthy that we have also conducted simulation studies for larger n values and the comparison results remain similar. For more details, please refer to Appendix F.

6 Applications to Microarray data

In this section we apply the proposed estimator to several microarray data sets for estimating π_0. The first data set is from the experiment described by Kuo, Duncavage, Mathew, den Besten, Pei, Naeve, Yamamoto, Cheng, Sherr & Roussel (2003). The objective of the experiment was to identify the targets of the Arf gene on the Arf-Mdm2-p53 tumor suppressor pathway. In this study, the cDNA microarrays were printed from a murine clone library available at St. Jude Children's Research Hospital. Samples from reference and Arf-induced cell lines were taken at 0, 2, 4 and 8 hours. At each time point, three independent replicates of cDNA microarray were generated. There were 5776 probe spots on each array. Only 2936 spots that passed a quality control of image analysis were used for differential expression analysis. The p-values used in the study were generated by Pounds & Cheng (2004) where p-values were computed by permutation tests. See Panel A in Figure 4 for the histogram of the p-values. The second data set is the Estrogen data and is described in the "Estrogen 2x2 Factorial Design" vignette by Scholtens, Miron, Merchant, Miller, Miron, Iglehart & Gentleman (2004). The objective of the study was to investigate the effect of estrogen on the genes in ER+ breast cancer cells over time. The p-values of testing null hypothesis of no differential expression in the presence and absence of estrogen were used in our study. See Panel B in Figure 4 for the histogram of the p-values. The third data set is the cancer cell line experiment described by Cui, Hwang, Qiu, Blades & Churchill (2005). The data set is from a cDNA microarray experiment and the objective is to identify differentially expressed genes in two human colon cancer cell lines, CACO2 and HCT116, and three human ovarian cancer cell lines, ES2, MDAH2774 and OV1063. In total, there were 9600 genes tested on each array. The p-values of testing differential expression among these cell lines were then generated by fitting an analysis of variance model to each gene to account for the multiple sources of variation including array, dye and sample effects. See Panel C in Figure 4 for the

histogram of the p-values.

Table 1 reports the estimated values of π_0 for the three data sets using the bootstrap estimator $\hat{\pi}_0^B$, the average estimate estimator $\hat{\pi}_0^A$, the convex estimator $\hat{\pi}_0^C$ and the proposed estimator $\hat{\pi}_0^U$, respectively. Note that the parametric estimator $\hat{\pi}_0^P$ in Qu et al. (2012) is not reported because the two-sided t-statistics are not available for these data sets. Among the four estimators, we observe that $\hat{\pi}_0^U$ is smaller than the other three estimators in most cases, especially for $\hat{\pi}_0^A$. This is consistent with the conclusion in Theorem 1. For the first data set, $\hat{\pi}_0^U$ is the smallest and is followed by $\hat{\pi}_0^B$ and $\hat{\pi}_0^C$, whereas $\hat{\pi}_0^A$ is far above them. For the second data set, there is a high degree of agreement among the estimators except for $\hat{\pi}_0^B$ which is much larger. For the third data set, $\hat{\pi}_0^U$ is similar to $\hat{\pi}_0^B$ and $\hat{\pi}_0^C$ and is less conservative compared to $\hat{\pi}_0^A$.

7 Conclusion

The proportion of true null hypotheses, π_0, is an important quantity in multiple testing and has attracted a lot of attention in the recent literature. It is known that most existing methods for estimating π_0 are either too conservative or suffering from an unacceptably large estimation variance. In this paper, we propose a new method for estimating π_0 that reduces the bias and variance of the estimation simultaneously. To achieve this, we first utilize the probability density functions of false-null p-values and then propose a novel algorithm to estimate the quantity of π_0. The statistical behavior of the proposed estimator is also investigated. Through extensive simulation studies and real data analysis, we demonstrated that the proposed estimator may substantially decrease the bias and variance compared to most existing competitors, and therefore, improve the existing literature significantly. Finally, we note that the paper has focused on the estimation of π_0 only. Some related questions, such as the behavior of false discovery rate using the proposed estimator, may warrant further studies.

Acknowledgements

Yebin Cheng's research was supported in part by National Natural Science Foundation of China grant No.11271241) and Shanghai Leading Academic Discipline Project No.863. Dexiang Gao's research was supported in part by NIH grant R01 CA 157850-02 and 51P30 CA46934. Tiejun Tong's research was supported in part by Hong Kong Research grant HKBU202711 and Hong Kong Baptist University FRG grants FRG2/11-12/110 and FRG1/13-14/018. The authors thank the editor, the associate editor, a referee and Bryan McNair for their constructive comments that led to a substantial improvement of the paper.

References

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

Cui, X., Hwang, J. T. G., Qiu, J., Blades, N. J. and Churchill, G. A. (2005). Improved statistical tests for differential gene expression by shrinking variance components estimates, *Biostatistics* **6**: 59–75.

Dalmasso, C., Broet, P. and Moreau, T. (2005). A simple procedure for estimating the false discovery rate, *Bioinformatics* **21**: 660–668.

Finner, H. and Gontscharuk, V. (2009). Controlling the familywise error rate with plug-in estimator for the proportion of true null hypotheses, *Journal of the Royal Statistical Society, Series B* **71**: 1031–1048.

Genovese, C. and Wasserman, L. (2002). Operating characteristics and extensions of the false discovery rate procedure, *Journal of the Royal Statistical Society, Series B* **64**: 499–517.

Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control, *Annals of Statistics* **32**: 1035–1061.

Hochberg, Y. and Benjamini, Y. (1990). More powerful procedures for multiple significance testing, *Statistics in Medicine* **9**: 811–818.

Hung, H. M., O'nell, R. T., Bauer, P. and Köhne, K. (1997). The behavior of the p-value when the alternative hypothesis is true, *Biometrics* **53**: 11–22.

Jiang, H. and Doerge, R. W. (2008). Estimating the proportion of true null hypotheses for multiple comparisons, *Cancer Informatics* **6**: 25–32.

Kuo, M., Duncavage, E. J., Mathew, R., den Besten, W., Pei, D., Naeve, D., Yamamoto, T., Cheng, C., Sherr, C. J. and Roussel, M. F. (2003). Arf induces p53-dependent and -independent antiproliferative genes, *Cancer Research* **63**: 1046–1053.

Lai, Y. (2007). A moment-based method for estimating the proportion of true null hypotheses and its application to microarray gene expression data, *Biostatistics* **8**: 744–755.

Langaas, M., Lindqvist, B. H. and Ferkingstad, E. (2005). Estimating the proportion of true null hypotheses, with application to DNA microarray data, *Journal of the Royal Statistical Society, Series B* **67**: 555–572.

McLachlan, G. J., Bean, R. W. and Jones, J. B. T. (2006). A simple implementation of a normal mixture approach to differential gene expression in multiclass microarrays, *Bioinformatics* **22**: 1608–1615.

Nettleton, D., Hwang, J. T. G., Caldo, R. A. and Wise, R. P. (2006). Estimating the number of true null hypotheses from a histogram of p-values, *Journal of Agricultural, Biological, and Environmental Statistics* **11**: 337–356.

Nguyen, D. V. (2004). On estimating the proportion of true null hypotheses for false discovery rate controlling procedures in exploratory DNA microarray studies, *Computational Statistics & Data Analysis* **47**: 611–637.

Pawitan, Y., Murthy, K. R. K., Michiels, S. and Ploner, A. (2005). Bias in the estimation of false discovery rate in microarry studies, *Bioinformatics* **21**: 3865–3872.

Pounds, S. and Cheng, C. (2004). Improving false discovery rate estimation, *Bioinformatics* **20**: 1737–1745.

Qu, L., Nettleton, D. and Dekkers, J. C. (2012). Improved estimation of the noncentrality parameter distribution from a large number of t-statistics, with applications to false discovery rate estimation in microarray data analysis, *Biometrics* **68**: 1178–1187.

Ruppert, D., Nettleton, D. and Hwang, J. T. G. (2007). Exploring the information in p-values for the analysis and planning of multiple-test experiments, *Biometrics* **63**: 483–495.

Scholtens, D., Miron, A., Merchant, F. M., Miller, A., Miron, P. L., Iglehart, J. D. and Gentleman, R. (2004). Analyzing factorial designed microarray experiments, *Journal of Multivariate Analysis* **90**: 19–43.

Schweder, T. and Spjøtvoll, E. (1982). Plots of p-values to evaluate many tests simultaneously, *Biometrika* **69**: 493–502.

Storey, J. D. (2002). A direct approach to false discovery rates, *Journal of the Royal Statistical Society, Series B* **64**: 479–498.

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

Storey, J. D. and Tibshirani, R. (2003). SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarrays. In Parmigiani, G., Garrett, E. S., Irizarry, R. A. and Zeger, S. L. (eds), *The Analysis of Gene Expression Data: Methods and Software* . New York: Springer.

Tong, T., Feng, Z., Hilton, J. S. and Zhao, H. (2013). Estimating the proportion of true null hypotheses using the pattern of observed p-values, *Journal of Applied Statistics* **40**: 1949–1964.

Wang, H., Tuominen, K. L. and Tsai, C. (2011). SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures, *Bioinformatics* **27**: 225–231.

Wu, B., Guan, Z. and Zhao, H. (2006). Parametric and nonparametric FDR estimation revisited, *Biometrics* **62**: 735–744.

Table 1: Estimation of π_0 for the three data sets using the bootstrap estimator $\hat{\pi}_0^B$, the average estimate estimator $\hat{\pi}_0^A$, the convex estimator $\hat{\pi}_0^C$ and the proposed estimator $\hat{\pi}_0^U$, respectively.

	Data Set 1	Data Set 2	Data Set 3
$\hat{\pi}_0^B$	0.447	0.944	0.486
$\hat{\pi}_0^A$	0.658	0.884	0.583
$\hat{\pi}_0^C$	0.463	0.875	0.501
$\hat{\pi}_0^U$	0.431	0.877	0.498

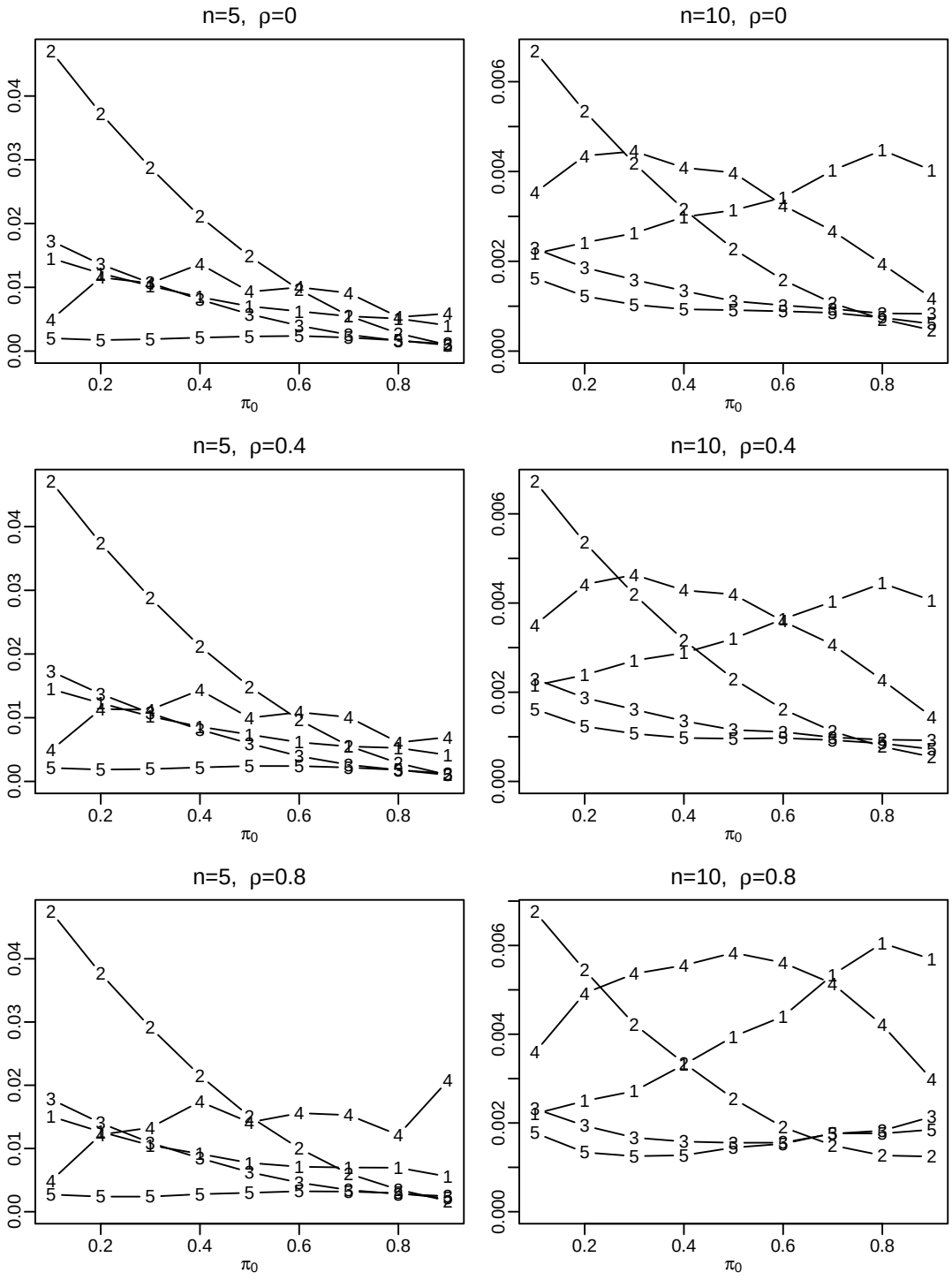

Figure 1: Plots of MSEs as functions of π_0 for various n and ρ values, where "1" represents the bootstrap estimator $\hat{\pi}_0^B$, "2" represents the average estimate estimator $\hat{\pi}_0^A$, "3" represents the convex estimator $\hat{\pi}_0^C$, "4" represents the parametric estimator $\hat{\pi}_0^P$, and "5" represents the proposed new estimator $\hat{\pi}_0^U$.

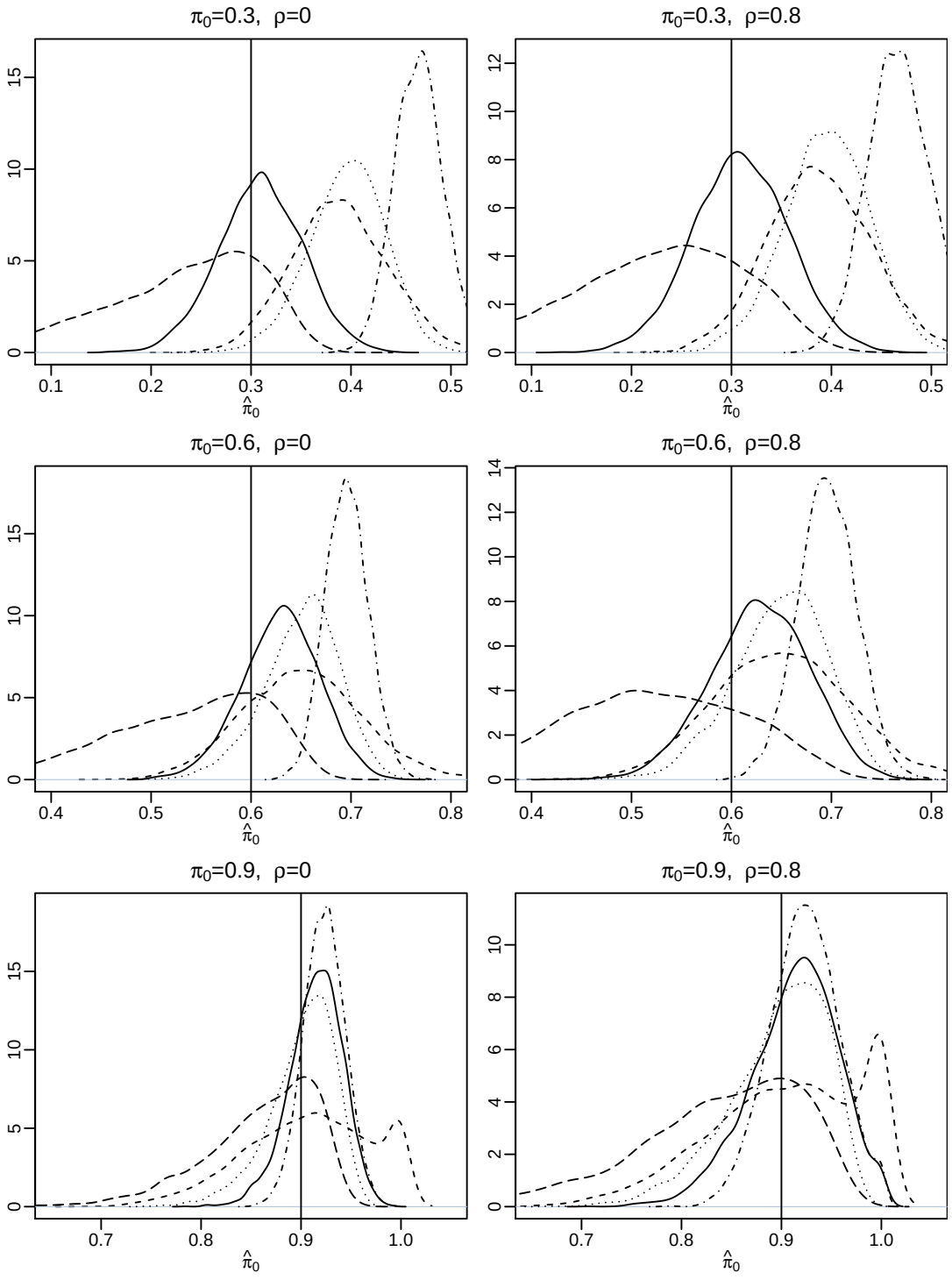

Figure 2: Density estimates of $\hat{\pi}_0$ for $n = 5$, where the short dashed line represents the bootstrap estimator $\hat{\pi}_0^B$, the dash-dotted line represents the average estimate estimator $\hat{\pi}_0^A$, the dotted line represents the convex estimator $\hat{\pi}_0^C$, the long dashed line represents the parametric estimator $\hat{\pi}_0^P$, and the solid line represents the proposed new estimator $\hat{\pi}_0^U$.

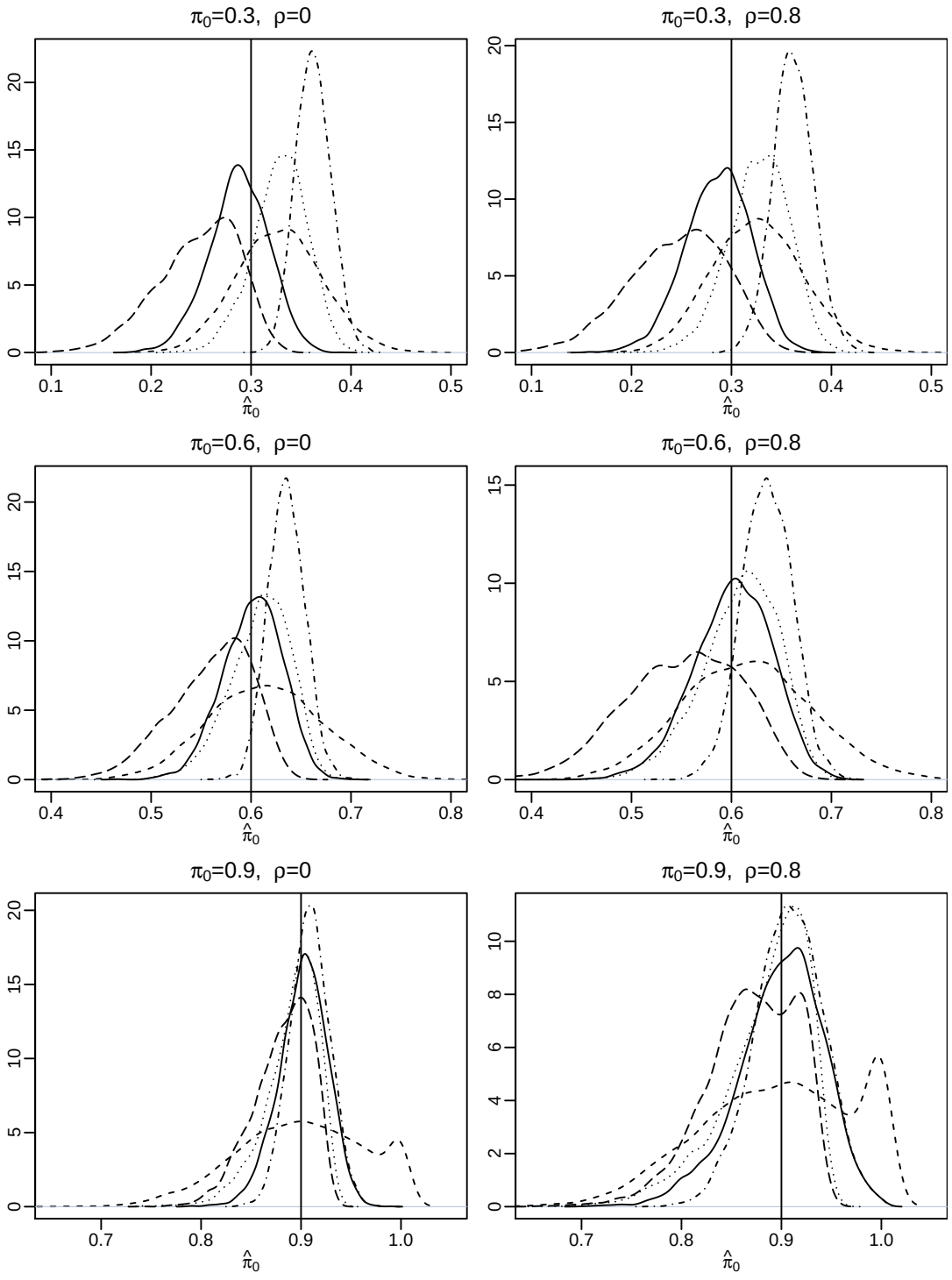

Figure 3: Density estimates of $\hat{\pi}_0$ for $n = 10$, where the short dashed line represents the bootstrap estimator $\hat{\pi}_0^B$, the dash-dotted line represents the average estimate estimator $\hat{\pi}_0^A$, the dotted line represents the convex estimator $\hat{\pi}_0^C$, the long dashed line represents the parametric estimator $\hat{\pi}_0^P$, and the solid line represents the proposed new estimator $\hat{\pi}_0^U$.

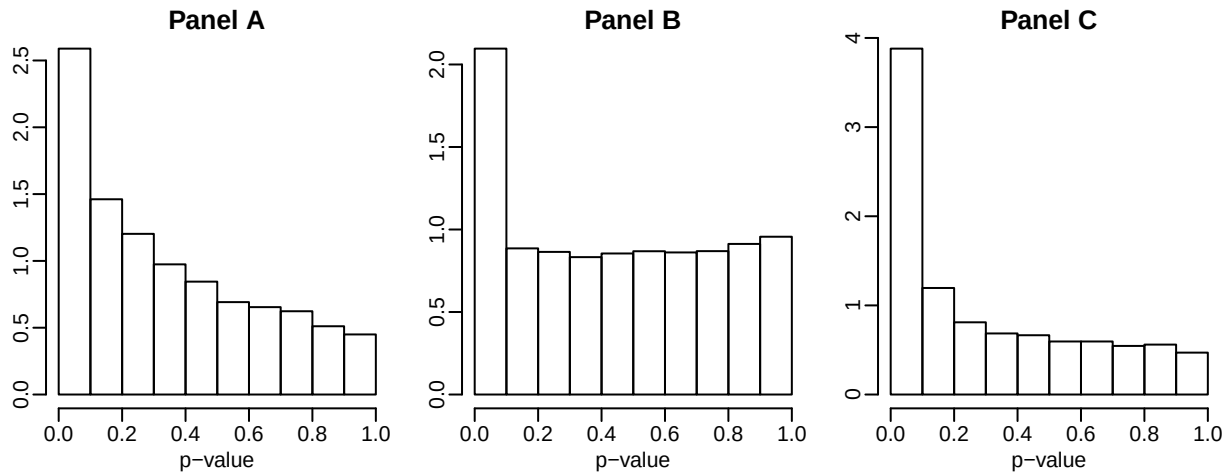

Figure 4: Histograms of p-values for the three data sets, where Panels A, B and C correspond to the p-values for the first, second and third data set, respectively.

Supplementary Materials for "Bias and Variance Reduction in Estimating the Proportion of True Null Hypotheses"

Yebin Cheng[1], Dexiang Gao[2] and Tiejun Tong[3]

[1]School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, P.R.China

[2]Department of Biostatistics and Informatics, University of Colorado, Denver, Colorado, USA

[3]Department of Mathematics, Hong Kong Baptist University, Hong Kong

Email: tongt@hkbu.edu.hk

April 30, 2014

Appendix A: Derivation of (14)

Note that under H_1, the test statistic T follows a normal distribution with mean $\sqrt{n}\delta$ and variance 1, where $\delta = \mu/\sigma$. We have

$$\begin{aligned} P(P \le p) &= P(|T| \ge z_{p/2}) \\ &= P(T \ge z_{p/2}) + P(T \le -z_{p/2}) \\ &= P(T - \sqrt{n}\delta \ge z_{p/2} - \sqrt{n}\delta) + P(T - \sqrt{n}\delta \le -z_{p/2} - \sqrt{n}\delta) \\ &= 1 - \Phi(z_{p/2} - \sqrt{n}\delta) + \Phi(-z_{p/2} - \sqrt{n}\delta). \end{aligned}$$

This leads to the following probability density function of p,

$$\begin{aligned} f_\delta(p) &= \frac{d}{dp} P(P \le p) \\ &= \frac{\phi(z_{p/2} - \sqrt{n}\delta)}{2\phi(z_{p/2})} + \frac{\phi(z_{p/2} + \sqrt{n}\delta)}{2\phi(z_{p/2})}. \end{aligned}$$

Appendix B: Derivation of (17)

Note that under H_1, the test statistic T follows a non-central t distribution with ν degrees of freedom and non-centrality parameter $\sqrt{n}\delta$, where $\delta = \mu/\sigma$. For the two-sided test (13), We have

$$\begin{aligned} P(P \le p) &= P(|T| \ge t_\nu(p/2)) \\ &= P(T \ge t_\nu(p/2)) + P(T \le -t_\nu(p/2)) \\ &= 1 - F_{\nu,\sqrt{n}\delta}(t_\nu(p/2)) + F_{\nu,\sqrt{n}\delta}(-t_\nu(p/2)). \end{aligned}$$

This leads to (17) by noting that

$$f_\delta(p) = \frac{d}{dp}P(P \le p) = \frac{f_{\nu,\sqrt{n}\delta}(t_\nu(p/2))}{2f_\nu(t_\nu(p/2))} + \frac{f_{\nu,\sqrt{n}\delta}(-t_\nu(p/2))}{2f_\nu(t_\nu(p/2))}.$$

Appendix C: Proof of Theorem 1

To show that $\hat{\pi}_0^U$ is less conservative than $\hat{\pi}_0^A$, it suffices to show that for each $\lambda_j \in \Lambda$,

$$\min\left\{1, \max\left\{0, \frac{W(\lambda_j) - m\widehat{Q}(\lambda_j)}{m(1-\lambda_j) - m\widehat{Q}(\lambda_j)}\right\}\right\} \le \min\left\{1, \frac{W(\lambda_j)}{m(1-\lambda_j)}\right\}.$$

Noting that $W(\lambda_j)/[m(1-\lambda_j)] \ge 0$, to show the above inequality it suffices to show that

$$\min\left\{1, \frac{W(\lambda_j) - m\widehat{Q}(\lambda_j)}{m(1-\lambda_j) - m\widehat{Q}(\lambda_j)}\right\} \le \min\left\{1, \frac{W(\lambda_j)}{m(1-\lambda_j)}\right\}. \tag{1}$$

By the definition of $t_\nu(\cdot)$, it is easy to see that $t_\nu(\lambda_j/2) > 0$ for any $0 < \lambda_j < 1$. Thus by (22), we have $\widehat{Q}_{\delta_i}(\lambda_j) \ge 0$ for any i and $\lambda_j \in \Lambda$ (This can also be seen from the definition of $Q_\delta(\lambda)$ as the upper tail probability can not be negative). In addition, noting that the probability density function of false-null p-value is a monotonically decreasing function in $[0,1]$, we have $\widehat{Q}_{\delta_i}(\lambda_j) < 1 - \lambda_j$. Now since $0 \le \widehat{Q}_{\delta_i}(\lambda_j) < 1 - \lambda_j$ for each i, by (23) we have $0 \le \widehat{Q}(\lambda_j) < 1 - \lambda_j$. Further, we have $m(1-\lambda_j) - m\widehat{Q}(\lambda_j) > 0$ which indicates that the denominator of the estimator $\hat{\pi}_0^U(\lambda_j)$ is always positive.

Finally, to validate (1), we consider the following two cases. (a) When $W(\lambda_j) \ge m(1-\lambda_j)$, it is easy to see that both the left and right hand sides of (1) is 1 so that the inequality holds. (b) When $W(\lambda_j) < m(1-\lambda_j)$, to show (1) it suffices to show that

$$\frac{W(\lambda_j) - m\widehat{Q}(\lambda_j)}{m(1-\lambda_j) - m\widehat{Q}(\lambda_j)} \le \frac{W(\lambda_j)}{m(1-\lambda_j)},$$

which holds under the condition that $m(1-\lambda_j) - m\widehat{Q}(\lambda_j) > 0$.

Appendix D

To investigate how sensitive the method is to the choice of boundaries a, b and τ, we have conducted a new simulation study that considers the following sets of Λ:

(1) $\Lambda_1 = \{0.05, 0.1, \ldots, 0.5\}$ with $a = 0.05$, $b = 0.5$ and $\tau = 9$

(2) $\Lambda_2 = \{0.1, 0.15, \ldots, 0.5\}$ with $a = 0.1$, $b = 0.5$ and $\tau = 8$

(3) $\Lambda_3 = \{0.2, 0.25, \ldots, 0.5\}$ with $a = 0.2$, $b = 0.5$ and $\tau = 6$

(4) $\Lambda_4 = \{0.2, 0.25, \ldots, 0.6\}$ with $a = 0.2$, $b = 0.6$ and $\tau = 8$

(5) $\Lambda_5 = \{0.2, 0.25, \ldots, 0.7\}$ with $a = 0.2$, $b = 0.7$ and $\tau = 10$

(6) $\Lambda_6 = \{0.2, 0.21, 0.22, \ldots, 0.5\}$ with $a = 0.2$, $b = 0.5$ and $\tau = 30$

where Λ_3 is the lambda set used in the proposed algorithm. Note also that Λ_6 is a more dense version of Λ_3 with the same values of a and b, which is used to investigate whether τ is also sensitive to the performance of the method.

With $N = 1000$ simulations, we report the MSEs of the estimators with the six sets of Λ in Figure 4, and the density estimates of the distribution of the estimators in Figure 5 for $\pi_0 = 0.3$, 0.6 and 0.9 and $\rho = 0$ and 0.8. From the simulated results, we observe that (1) the method is not sensitive to the boundaries a and b, as long as b is not too large (preferably not larger than 0.5); (2) the choice of τ has little effect on the performance of the method; and (3) the set Λ used in the manuscript, i.e. Λ_3, is not the best one in terms of the MSE. Nevertheless, we decide to keep it in the paper to avoid over-rating the proposed method.

In addition, for illustration we have also plotted the new estimator $\hat{\pi}_0^U$ with the sets Λ_3, Λ_1, Λ_2, Λ_4 and Λ_5 together with the existing methods in Figure 6. We note that no matter which Λ set is used, the performance of the new estimator performs comparably to most existing methods. It is also noteworthy that we have conducted simulations for $n = 5$ and also for larger n values. The comparison results remain the same.

Appendix E

To investigate how sensitive the method is to the choice of the initial estimator of π_0, we have conducted a new simulation study that considers the following initial estimates of π_0:

(1) $\hat{\pi}_0^{I1} = \hat{\pi}_0^B - 0.05$

(2) $\hat{\pi}_0^{I2} = \hat{\pi}_0^B - 0.025$

(3) $\hat{\pi}_0^{I3} = \hat{\pi}_0^B$

(4) $\hat{\pi}_0^{I4} = \hat{\pi}_0^B + 0.025$

(5) $\hat{\pi}_0^{I5} = \hat{\pi}_0^B + 0.05$

(6) $\hat{\pi}_0^{I6} = \pi_0$

where $\hat{\pi}_0^B$ is the bootstrap estimator in Storey et al. (2004). Note that $\hat{\pi}_0^{I3}$ is the initial estimate of π_0 adopted in the proposed algorithm. The initial value $\hat{\pi}_0^{I6}$ is added as a reference for comparison, where the true π_0 value is used as the initial estimate of π_0. We also note that the range of the selected initial values is $\hat{\pi}_0^{I5} - \hat{\pi}_0^{I1} = 0.1$. That covers a wide range of initial estimates with most existing estimators falling in the interval.

With $N = 1000$ simulations, we report the MSEs of the estimators with the six initial estimates in Figure 1, and the density estimates of the distribution of the estimators in Figure 2 for $\pi_0 = 0.3$, 0.6 and 0.9 and $\rho = 0$ and 0.8. From the simulated results, we observe that (1) the choice of the initial estimator of π_0 is not very sensitive to the performance of the method, as long as the initial estimate is not too small to avoid the over bias-correction (preferably the initial estimate is a conservative estimator); (2) the initial estimator used in the manuscript, i.e. $\hat{\pi}_0^{I3} = \hat{\pi}_0^B$, is not the best one in terms of the MSE. Nevertheless, we decided to keep it in the paper to avoid over-rating the proposed method.

In addition, for illustration we have also plotted the new estimator $\hat{\pi}_0^U$ with the initial estimates $\hat{\pi}_0^{I3}$, $\hat{\pi}_0^{I2}$, $\hat{\pi}_0^{I4}$ and $\hat{\pi}_0^{I5}$ together with the existing methods in Figure 3. We note that no matter which initial estimate of π_0 is used, the performance of the new estimator performs at least comparably to the convex estimator $\hat{\pi}_0^C$ and is better than the others. It is also noteworthy that we have conducted simulations for $n = 5$ and also for larger n values. The comparison results remain the same.

Appendix F

In this appendix, simulation studies are conducted for larger sample sizes. Specifically, we have reported here the result for $n = 30$ with two different settings of effect sizes:

(S1) simulate μ_i i.i.d. from $U[0.5, 1.5]$ for each false null hypothesis,

(S2) simulate μ_i i.i.d. from $U[0.2, 1]$ for each false null hypothesis,

where (S1) is the simulation setting used in the manuscript. For a larger sample size such as $n = 30$, the effect sizes in (S1) may be too large so that the true null p-values and the false-null p-values are too separated. (S2) is used to have a more reasonable setting on the effect sizes.

With $N = 1000$ simulations, we report the MSEs of the estimators in Figures 7 and 8 for the settings (S1) and (S2), respectively. We also report the density estimates of the distribution of the estimators in Figures 9 and 10 for the settings (S1) and (S2), respectively. From the simulated results, we note that the proposed method also perform well under the larger sample size setting.

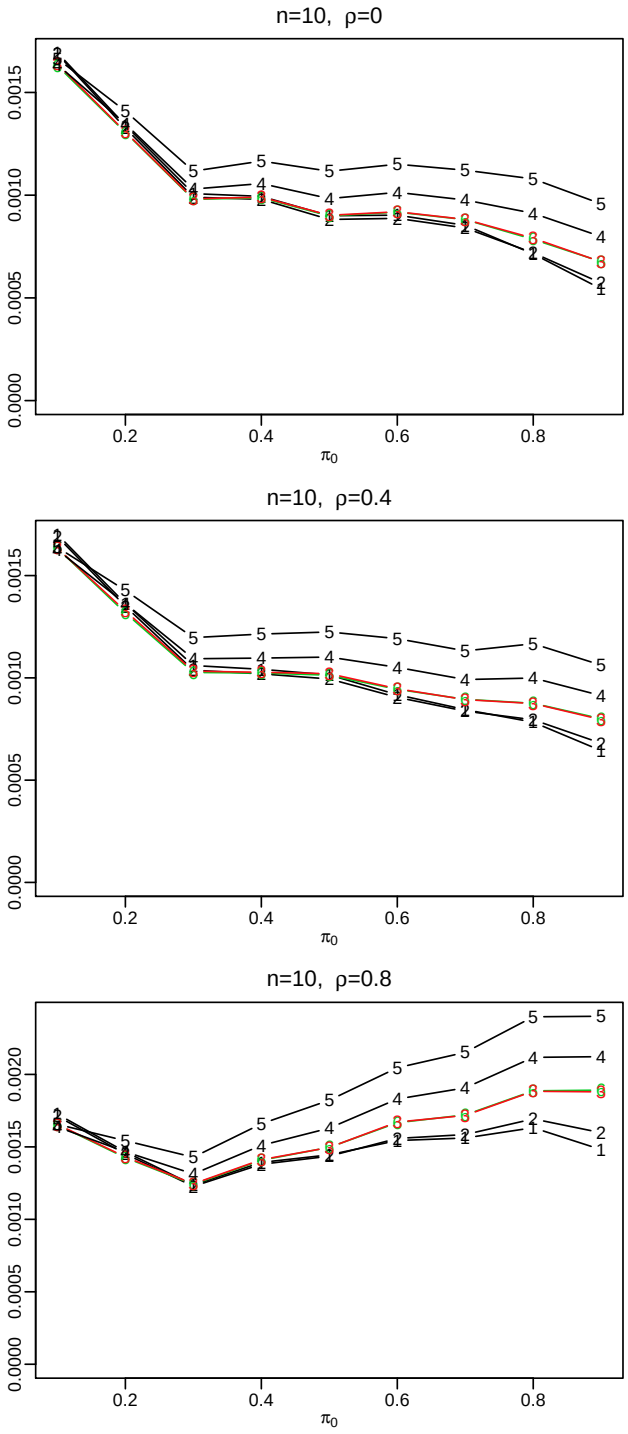

Figure 1: Plots of MSEs as functions of π_0 for $n = 10$ and three ρ values, where "1" represents the estimator with the set Λ_1, "2" represents the estimator with the set Λ_2, "3" represents the estimator with the set Λ_3, "4" represents the estimator with the set Λ_4, "5" represents the estimator with the set Λ_5, and "6" represents the estimator with the set Λ_6.

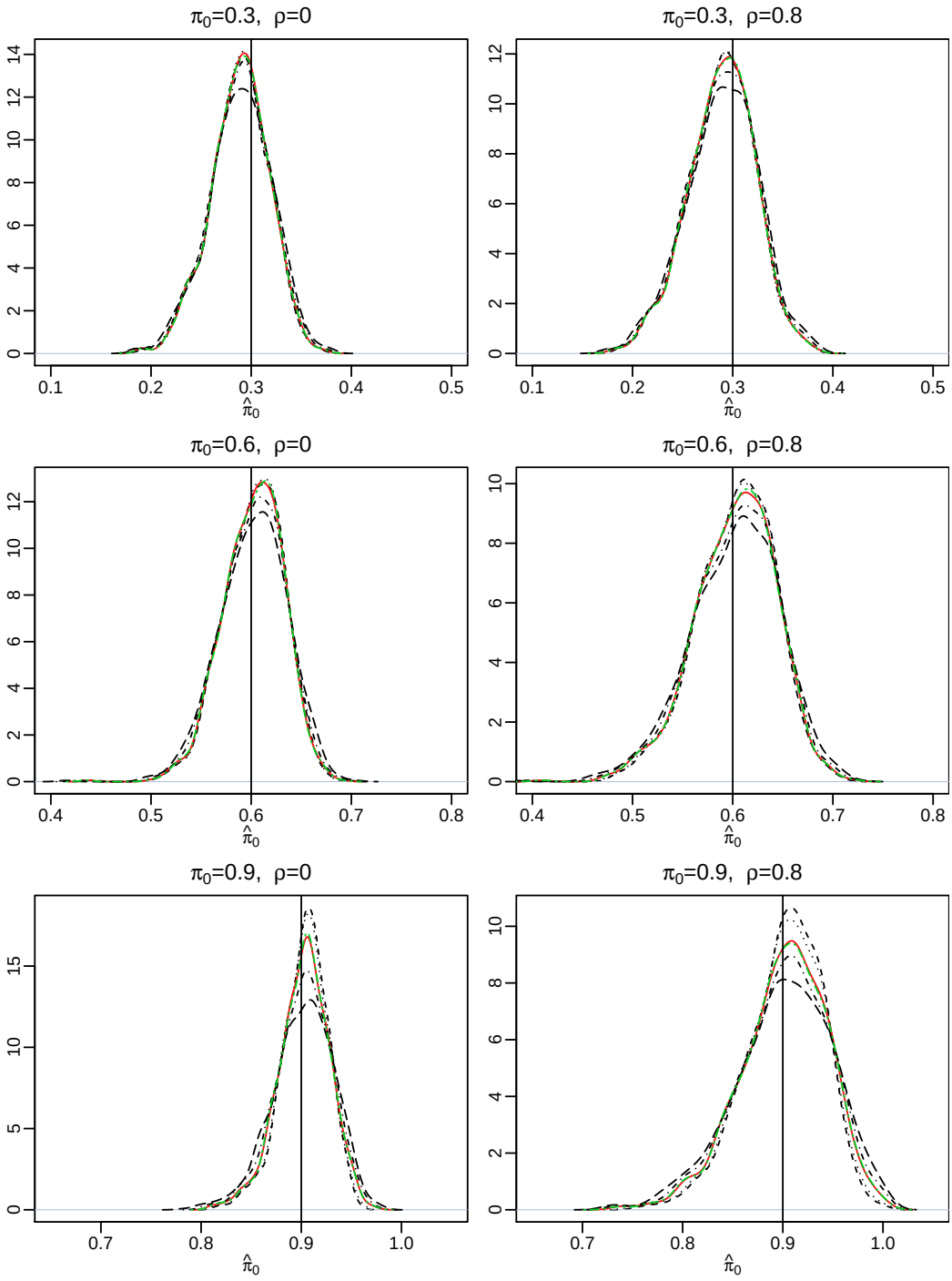

Figure 2: Density estimates of $\hat{\pi}_0$ for $n = 10$, where the short dashed line represents the estimator with the set Λ_1, the dotted line represents the estimator with the set Λ_2, the solid line (the red line) represents the estimator with the set Λ_3, the short dash-dotted line represents the estimator with the set Λ_4, the long dashed line represents the estimator with the set Λ_5, and the long dash-dotted line (the green line) represents the estimator with the set Λ_6.

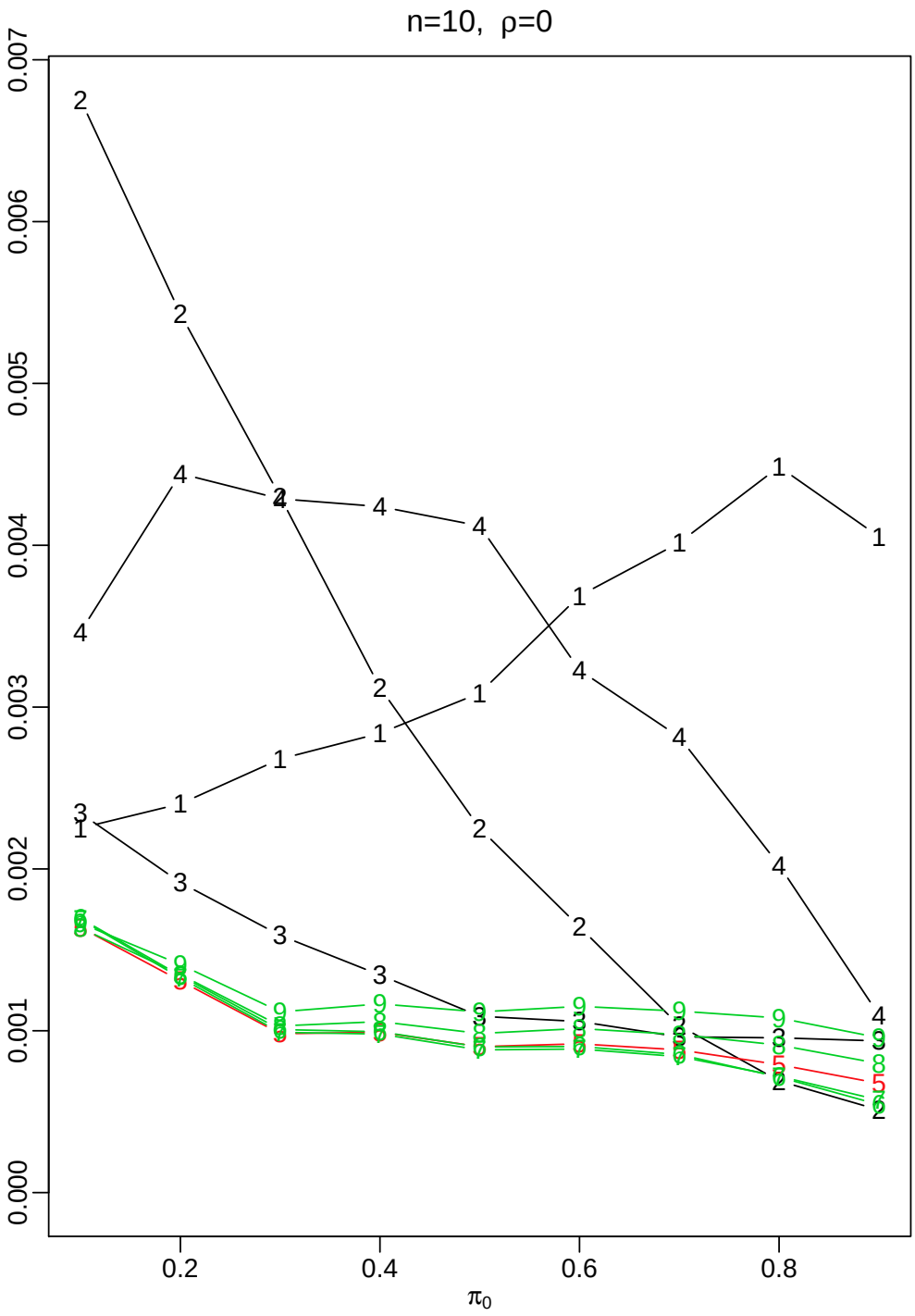

Figure 3: Plot of MSEs as functions of π_0 for $n = 10$ and $\rho = 0$, where "1" represents the bootstrap estimator $\hat{\pi}_0^B$, "2" represents the average estimate estimator $\hat{\pi}_0^A$, "3" represents the convex estimator $\hat{\pi}_0^C$, "4" represents the parametric estimator $\hat{\pi}_0^P$, and "5", "6", "7", "8" and "9" represent the new estimator $\hat{\pi}_0^U$ with the sets Λ_3, Λ_1, Λ_2, Λ_4 and Λ_5, respectively.

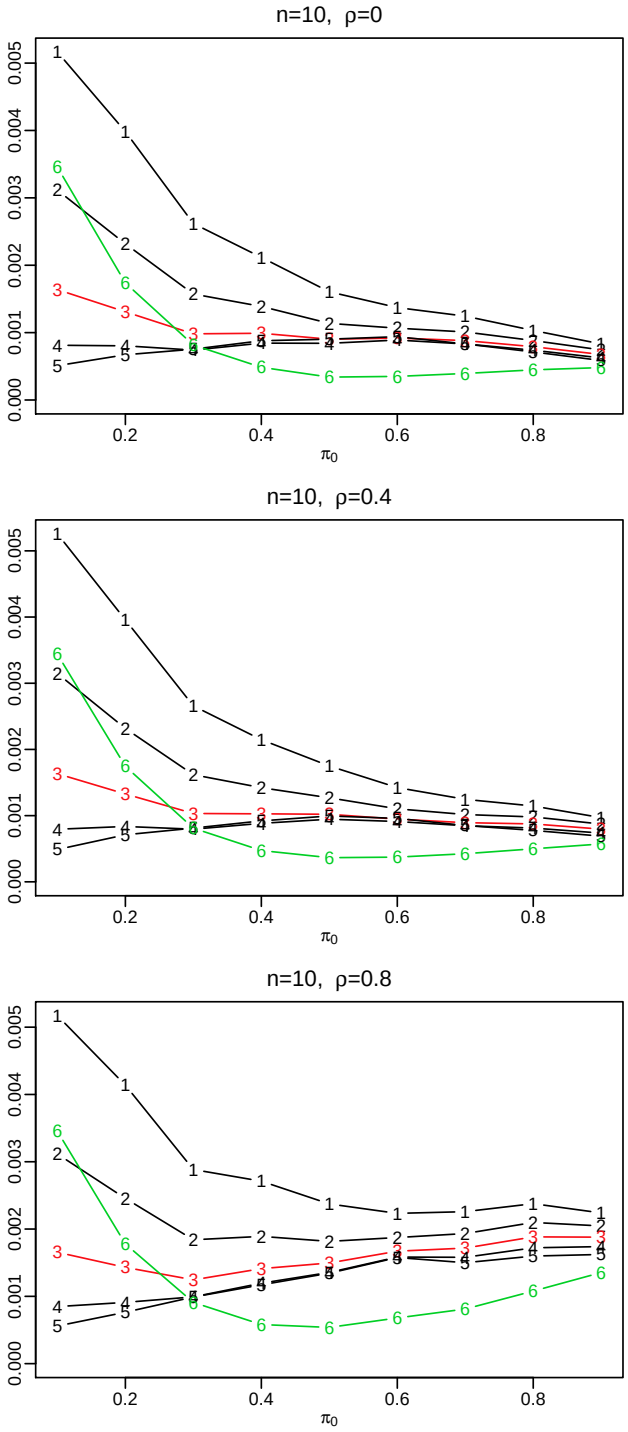

Figure 4: Plots of MSEs as functions of π_0 for $n = 10$ and three ρ values, where "1" represents the estimator with $\hat{\pi}_0^{I1}$, "2" represents the estimator with $\hat{\pi}_0^{I2}$, "3" represents the estimator with $\hat{\pi}_0^{I3}$, "4" represents the estimator with $\hat{\pi}_0^{I4}$, "5" represents the estimator with $\hat{\pi}_0^{I5}$, and "6" represents the estimator with $\hat{\pi}_0^{I6}$.

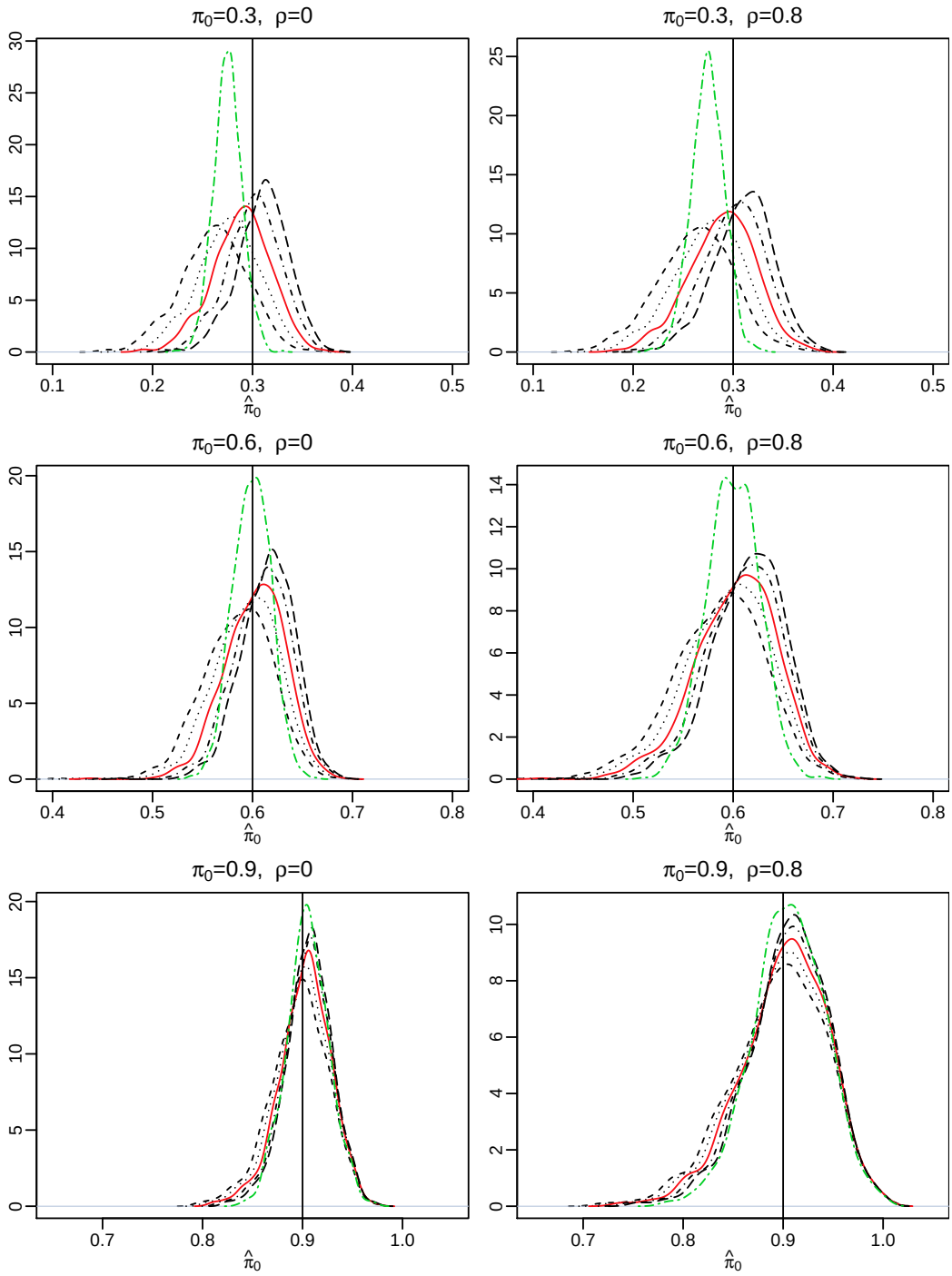

Figure 5: Density estimates of $\hat{\pi}_0$ for $n = 10$, where the short dashed line represents the estimator with $\hat{\pi}_0^{I1}$, the dotted line represents the estimator with $\hat{\pi}_0^{I2}$, the solid line (the red line) represents the estimator with $\hat{\pi}_0^{I3}$, the short dash-dotted line represents the estimator with $\hat{\pi}_0^{I4}$, the long dashed line represents the estimator with $\hat{\pi}_0^{I5}$, and the long dash-dotted line (the green line) represents the estimator with $\hat{\pi}_0^{I6}$.

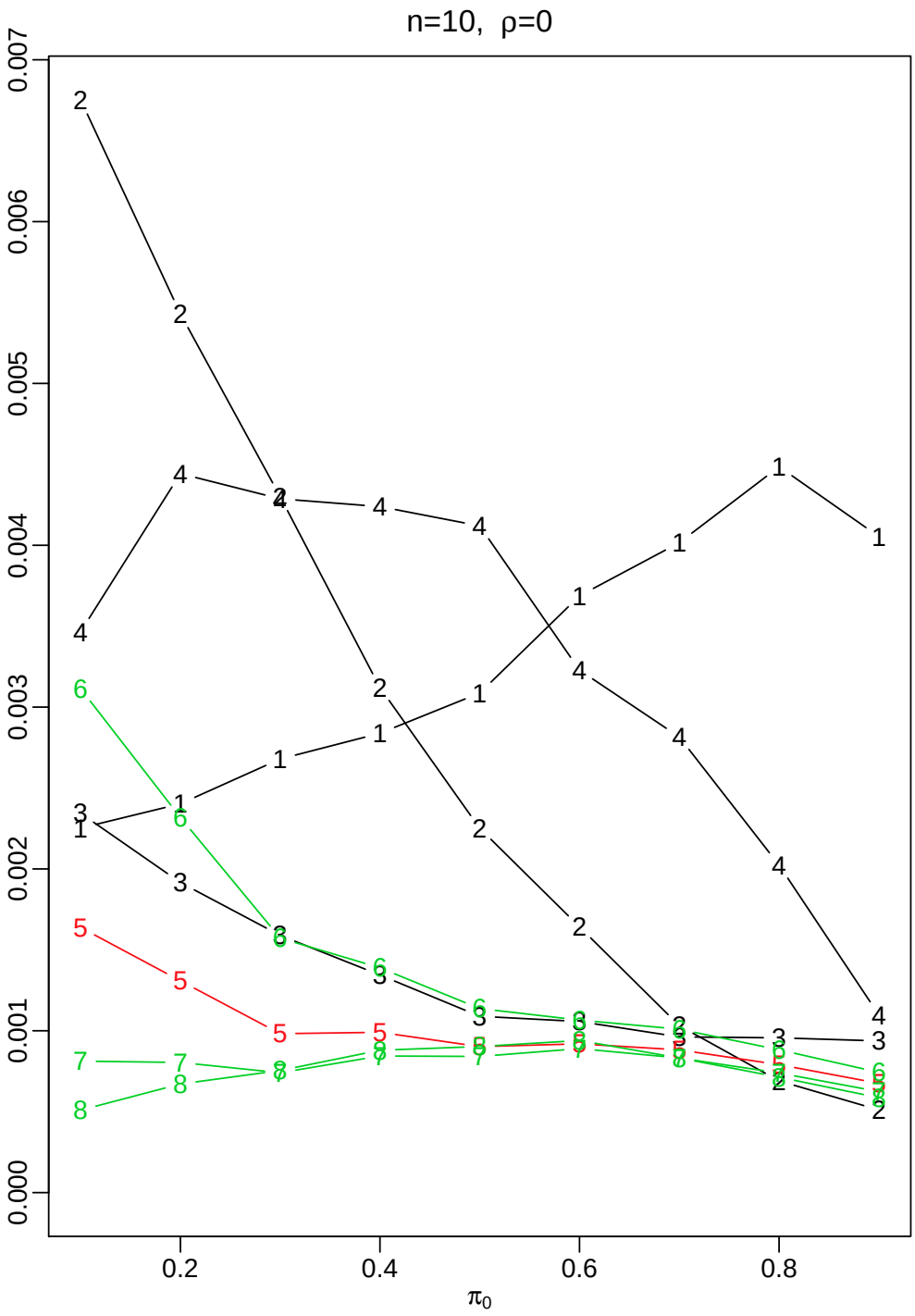

Figure 6: Plot of MSEs as functions of π_0 for $n = 10$ and $\rho = 0$, where "1" represents the bootstrap estimator $\hat{\pi}_0^B$, "2" represents the average estimate estimator $\hat{\pi}_0^A$, "3" represents the convex estimator $\hat{\pi}_0^C$, "4" represents the parametric estimator $\hat{\pi}_0^P$, and "5", "6", "7" and "8" represent the new estimator $\hat{\pi}_0^U$ with the initial estimates $\hat{\pi}_0^{I3}$, $\hat{\pi}_0^{I2}$, $\hat{\pi}_0^{I4}$ and $\hat{\pi}_0^{I5}$, respectively.

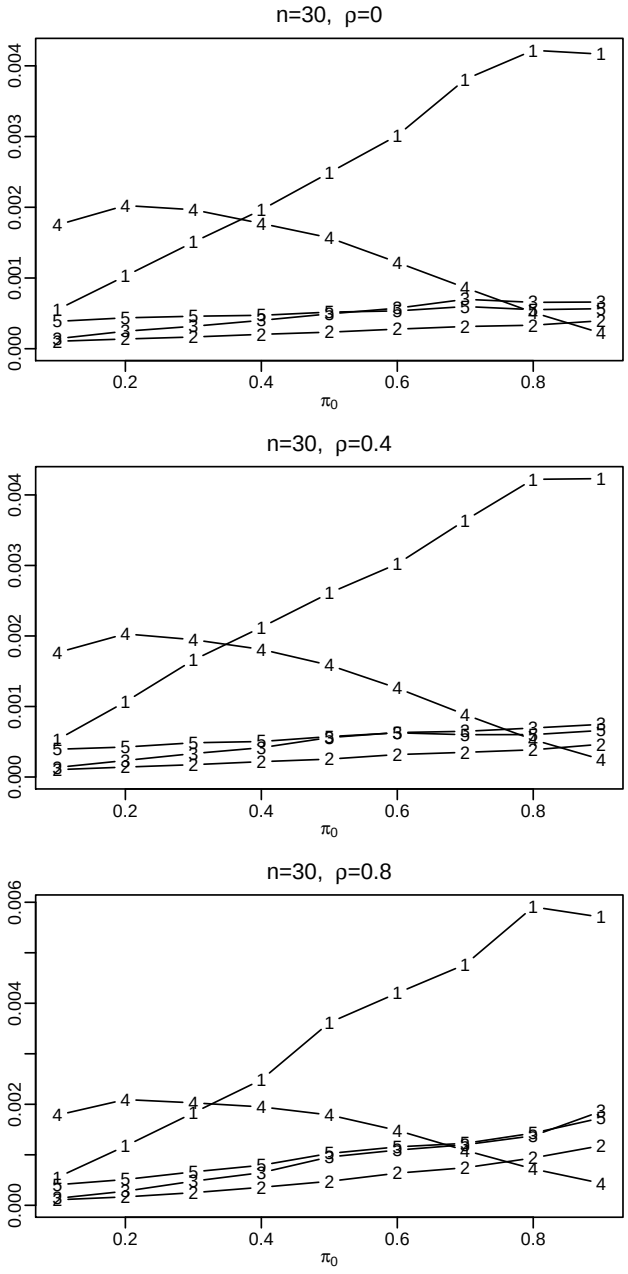

Figure 7: Plots of MSEs as functions of π_0 for $n = 30$ and three ρ values under the setting (S1), where "1" represents the bootstrap estimator $\hat{\pi}_0^B$, "2" represents the average estimate estimator $\hat{\pi}_0^A$, "3" represents the convex estimator $\hat{\pi}_0^C$, "4" represents the parametric estimator $\hat{\pi}_0^P$, and "5" represents the proposed new estimator $\hat{\pi}_0^U$.

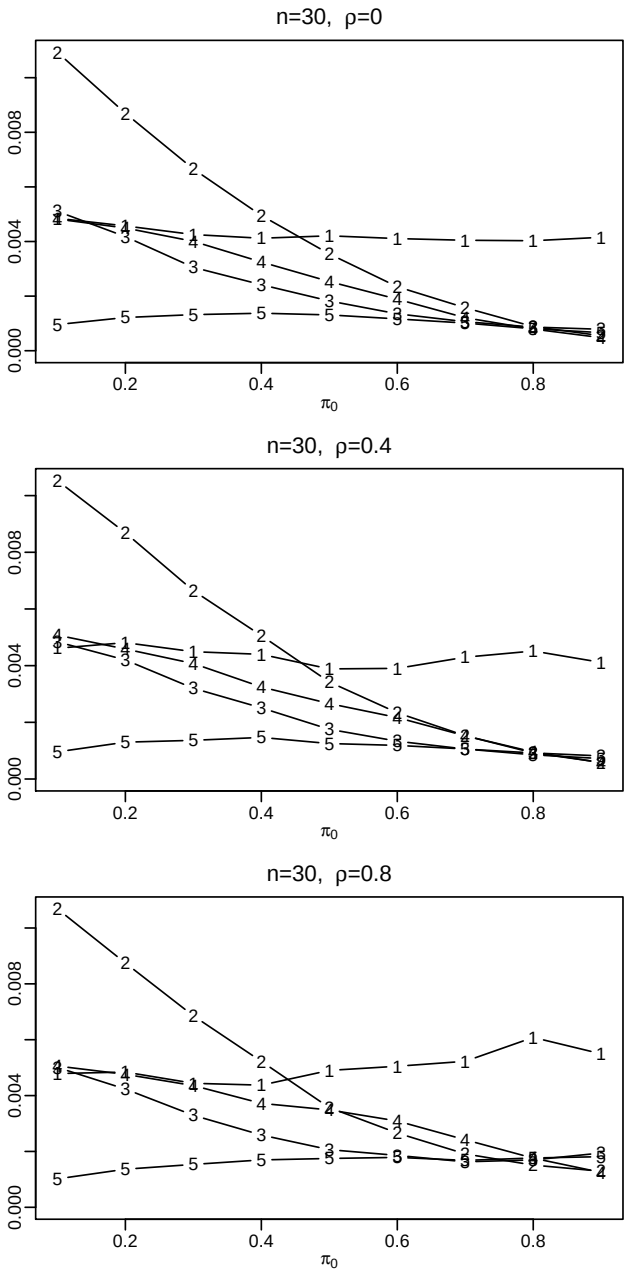

Figure 8: Plots of MSEs as functions of π_0 for $n = 30$ and three ρ values under the setting (S2), where "1" represents the bootstrap estimator $\hat{\pi}_0^B$, "2" represents the average estimate estimator $\hat{\pi}_0^A$, "3" represents the convex estimator $\hat{\pi}_0^C$, "4" represents the parametric estimator $\hat{\pi}_0^P$, and "5" represents the proposed new estimator $\hat{\pi}_0^U$.

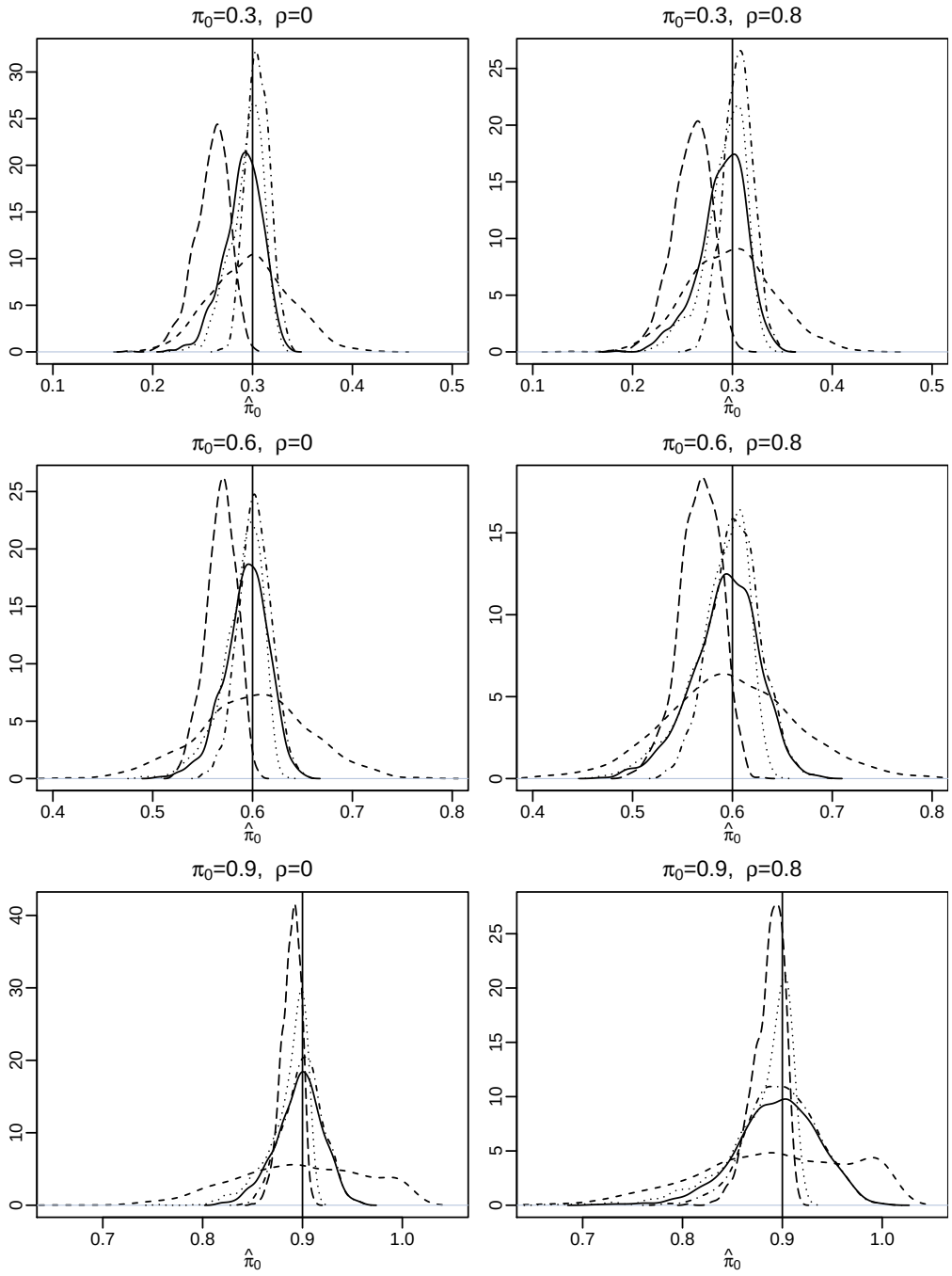

Figure 9: Density estimates of $\hat{\pi}_0$ for $n = 30$ under the setting (S1), where the short dashed line represents the bootstrap estimator $\hat{\pi}_0^B$, the dash-dotted line represents the average estimate estimator $\hat{\pi}_0^A$, the dotted line represents the convex estimator $\hat{\pi}_0^C$, the long dashed line represents the parametric estimator $\hat{\pi}_0^P$, and the solid line represents the proposed new estimator $\hat{\pi}_0^U$.

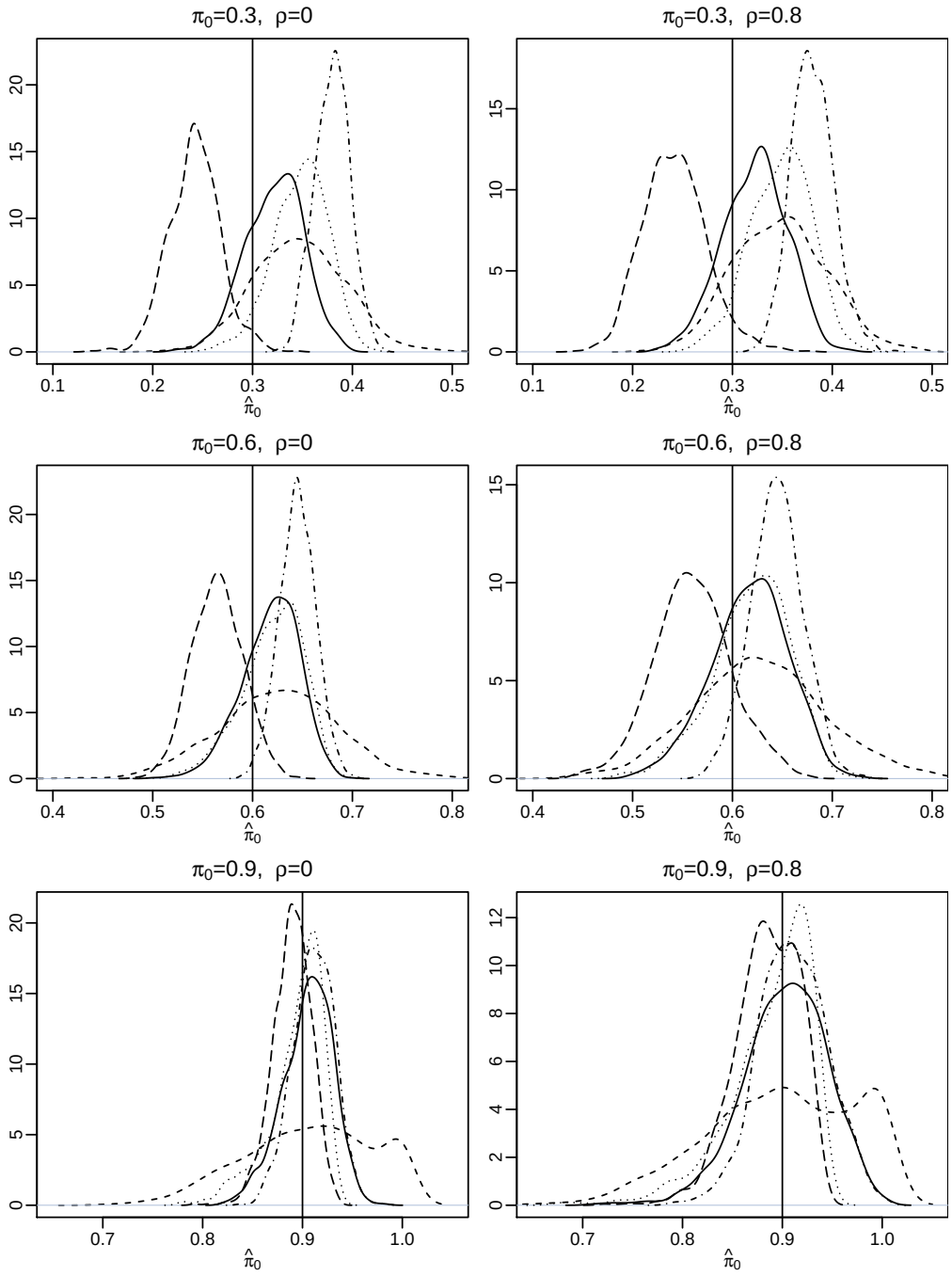

Figure 10: Density estimates of $\hat{\pi}_0$ for $n = 30$ under the setting (S2), where the short dashed line represents the bootstrap estimator $\hat{\pi}_0^B$, the dash-dotted line represents the average estimate estimator $\hat{\pi}_0^A$, the dotted line represents the convex estimator $\hat{\pi}_0^C$, the long dashed line represents the parametric estimator $\hat{\pi}_0^P$, and the solid line represents the proposed new estimator $\hat{\pi}_0^U$.