Supplementary Materials for

Robust Estimation of the False Discovery Rate

Stan Pounds and Cheng Cheng

This supplemental contains proofs regarding theoretical properties of the proposed method (Section S1), provides figures for the example applications (Section S2), uses some simulation results to illustrate theoretical properties (Section S3) and plots summarizing the results of many simuations (Section S4).

S1 Proofs of Theoretical Results

S1.1 Notation and Definitions

We use the notation from the manuscript and also introduce some additional notation. Throughout sections S1.2 through S1.4, we let $i = 1, \ldots, m$ index the hypothesis tests and P_i represent the random variable corresponding to the p-value for test i. Let $F_i(\alpha) = \Pr(P_i \leq \alpha)$ and $f_i(\alpha)$ be the corresponding probability distribution function or probability mass function. Additionally, we let $\mathcal{H}_{\mathcal{A}}$ represent the set of all i such that the alternative hypothesis is true and let $\mathcal{H}_0$ represent the set of all i such that the null hypothesis is true. Thus, we say $i \in \mathcal{H}_0$ if the null hypothesis of test i is true. Additionally, let $\bar{P}$ be the average of $P_1, \ldots, P_m$. For each i, define $A_i = 2\min(P_i, 1 - P_i)$ and $G_i(\alpha) = \Pr(A_i \leq \alpha)$. Finally, let $\mathrm{I}(\cdot)$ be the indicator function, so that $\mathrm{I}(\cdot) = 1$ if the enclosed statement is true, and $\mathrm{I}(\cdot) = 0$ if the enclosed statement is false.

Now, in terms of these definitions, we have that

$$\pi = \frac{1}{m} \sum_{i=1}^{m} \mathrm{I}(i \in \mathcal{H}_0) \tag{1}$$

and

$$\upsilon(\alpha) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{I}(i \in \mathcal{H}_0) F_i(\alpha). \tag{2}$$

S1.2 Results for Two-Sided Tests

Throughout this section, we assume that all p-values are based on two-sided tests. We now prove a trivial but important result.

Theorem 1 *If* $\mathrm{E}(P_i) \geq \frac{1}{2}$ *for all* $i \in \mathcal{H}_0$*, then* $\mathrm{E}(\bar{P}) \geq \frac{\pi}{2}$.

Proof:

$$\begin{aligned}
\mathrm{E}(\bar{P}) &= \frac{1}{m}\sum_{i=1}^{m} \mathrm{E}(P_i) \\
&= \frac{1}{m}\left(\sum_{i\in\mathcal{H}_0} \mathrm{E}(P_i) + \sum_{i\in\mathcal{H}_A} \mathrm{E}(P_i)\right) \\
&\geq \frac{1}{m}\sum_{i\in\mathcal{H}_0} \mathrm{E}(P_i) \\
&\geq \frac{1}{m}\sum_{i\in\mathcal{H}_0} \frac{1}{2} \\
&\geq \frac{\pi}{2}. \quad \blacksquare
\end{aligned} \tag{3}$$

Theorem 1 is now used to prove the following theorem regarding the conservativeness of $\widehat{\upsilon}(\alpha)$ as defined for two-sided tests.

Theorem 2 *If* $\Pr(P_i \leq \alpha) \geq \alpha$ *for all* α *and all* $i \in \mathcal{H}_0$ *and* $\Pr(\bar{P} > \frac{1}{2}) = 0$*, then* $\mathrm{E}\left(\widehat{\upsilon}(\alpha)\right) \geq \upsilon(\alpha)$.

Proof: First, note that $\Pr(P_i \leq \alpha) \geq \alpha$ implies that $\mathrm{E}(P_i) \geq \frac{1}{2}$. Therefore,

$$\begin{aligned}
\mathrm{E}\left(\widehat{\upsilon}(\alpha)\right) &= \mathrm{E}(\widehat{\pi}\alpha) = \alpha\mathrm{E}(\widehat{\pi}) \\
&= \alpha\left\{\mathrm{E}\left(2\bar{P}|\bar{P} \leq \frac{1}{2}\right)\Pr\left(\bar{P} \leq \frac{1}{2}\right) + \mathrm{E}\left(1|\bar{P} > \frac{1}{2}\right)\Pr\left(\bar{P} > \frac{1}{2}\right)\right\} \\
&\geq 2\alpha\mathrm{E}(\bar{P}) \\
&\geq \pi\alpha.
\end{aligned} \tag{4}$$

Additionally, observe that

$$\begin{aligned} \upsilon(\alpha) &= \mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m}\mathrm{I}(P_i \leq \alpha)\mathrm{I}(i \in \mathcal{H}_0)\right) \\ &= \frac{1}{m}\sum_{i\in\mathcal{H}_0}\mathrm{E}\left(\mathrm{I}(P_i \leq \alpha)\right) \\ &= \frac{1}{m}\sum_{i\in\mathcal{H}_0}\Pr(P_i \leq \alpha) \\ &\leq \frac{1}{m}\sum_{i\in\mathcal{H}_0}\alpha = \pi\alpha. \end{aligned} \tag{5}$$

Combining (4) and (5), we have $\mathrm{E}\left(\widehat{\upsilon}(\alpha)\right) \geq \upsilon(\alpha)$. ■

Together, Theorems 1 and 2 show that our proposed estimator $\widehat{\upsilon}(\alpha)$ (for two-sided tests) is conservative so long as $\Pr(\bar{P} > \frac{1}{2}) = 0$, which should be the case if π is sufficiently less than 1 and there is reasonable power for each $i \in \mathcal{H}_A$. Otherwise, it is clear from (3) that $\mathrm{E}(\widehat{\pi}) \approx 1$, and thus the proposed $\widehat{\upsilon}(\alpha)$ will not dramatically underestimate $\upsilon(\alpha)$. The following theorem proves a useful relationship between $\widehat{\upsilon}(\alpha)$ and the power of each test $i \in \mathcal{H}_A$.

Theorem 3 *Assume that* $\mathrm{E}(P_i) = \frac{1}{2}$ *for all* $i \in \mathcal{H}_0$ *and* $\Pr\left(\bar{P} > \frac{1}{2}\right) = 0$. *Additionally, assume that there exists some* i *such that* $i \in \mathcal{H}_A$. *Then* $\mathrm{E}(\widehat{\pi}) \to \pi$ *as* $\Pr(P_i \leq \epsilon) \to 1$ *for any fixed* $\epsilon > 0$ *and each* $i \in \mathcal{H}_A$. *That is, the bias of* $\widehat{\pi}$ *approaches 0 as the powers of the tests of false null hypotheses approach 1 for any fixed level.*

Proof: First note that the definition of $\widehat{\pi}$ from the manuscript implies that

$$\begin{aligned} \mathrm{E}(\widehat{\pi}) &= \mathrm{E}\left(2\bar{P}|\bar{P} \leq \frac{1}{2}\right)\Pr\left(\bar{P} \leq \frac{1}{2}\right) + \mathrm{E}\left(1|\bar{P} > \frac{1}{2}\right)\Pr\left(\bar{P} > \frac{1}{2}\right) \\ &= \frac{2}{m}\sum_{i=1}^{m}\mathrm{E}(P_i) \\ &= \frac{2}{m}\left(\sum_{i\in\mathcal{H}_0}\mathrm{E}(P_i) + \sum_{i\in\mathcal{H}_A}\mathrm{E}(P_i)\right) \\ &= \pi + \sum_{i\in\mathcal{H}_A}\mathrm{E}(P_i). \end{aligned} \tag{6}$$

If $\Pr(P_i \leq \alpha) \to 0$ for each $i \in \mathcal{H}_A$, then for each $i \in \mathcal{H}_A$ it follows that $\mathrm{E}(P_i) \to 0$. Thus, the second term in (6) approaches 0. ■

Therefore, under the conditions of Theorem 3, the bias of our proposed estimator $\widehat{v}(\alpha)$ approaches 0 as the statistical power of each test $i \in \mathcal{H}_0$ approaches 1 for an arbitrarily small level.

S1.3 Results for One-Sided Tests with Continuous p-values

Throughout this section, we assume that each test i is one-sided and yields a continuously distributed p-value. First, we note that the definition of A_i implies that

$$G_i(x) = F_i\left(\frac{x}{2}\right) + 1 - F_i\left(1 - \frac{x}{2}\right). \tag{7}$$

Additionally, using integration by parts, we note that

$$\mathrm{E}(P_i) = \int_0^1 x f_i(x)\mathrm{d}x = 1 - \int_0^1 F_i(x)\mathrm{d}x \tag{8}$$

for all i. Clearly, an analogous relationship exists between $\mathrm{E}(A_i)$ and $G_i(x)$ for all i. Therefore, (7) implies that

$$\mathrm{E}(A_i) = 2\int_0^{\frac{1}{2}} \left[F_i\left(x + \frac{1}{2}\right) - F_i(x)\right] \mathrm{d}x. \tag{9}$$

These observations allow us to prove the following result.

Theorem 4 *For each $i \in \mathcal{H}_0$, assume that $F_i(x) \leq 2F_i\left(\frac{1}{2}\right)x$ for $0 < x \leq \frac{1}{2}$ and $F_i(x) \geq 2F_i\left(\frac{1}{2}\right)x$ for $\frac{1}{2} \leq x \leq 1$. Then $v(\alpha) \leq 2\mathrm{E}(\bar{A})\alpha$ for each fixed $\alpha \leq \frac{1}{2}$.*

Proof: Using (9) and the assumptions, we have that

$$\begin{aligned}
\mathrm{E}(A_i) &= 2\int_0^{\frac{1}{2}} \left[F_i\left(\frac{x}{2}\right) + 1 - F_i\left(1 - \frac{x}{2}\right)\right] \mathrm{d}x \\
&\geq 2\int_0^{\frac{1}{2}} \left[2F_i\left(\frac{1}{2}\right)\left(x + \frac{1}{2}\right) - 2F_i\left(\frac{1}{2}\right)x\right] \mathrm{d}x \\
&= 4F_i\left(\frac{1}{2}\right)\int_0^{\frac{1}{2}} \frac{1}{2}\mathrm{d}x \\
&= F_i\left(\frac{1}{2}\right)
\end{aligned} \tag{10}$$

for each $i \in \mathcal{H}_0$. Therefore, for $\alpha \leq \frac{1}{2}$, we have

$$\begin{aligned} 2\mathrm{E}(\bar{A})\alpha &= \frac{1}{m}\sum_{i\in\mathcal{H}_0}(2\mathrm{E}(A_i)\alpha) + \frac{2}{m}\sum_{i\in\mathcal{H}_A}(\mathrm{E}(A_i)\alpha) \\ &\geq \frac{1}{m}\sum_{i\in\mathcal{H}_0} 2F_i\left(\frac{1}{2}\right)\alpha \\ &\geq \frac{1}{m}\sum_{i\in\mathcal{H}_0} F_i(\alpha) \\ &= \upsilon(\alpha). \quad \blacksquare \end{aligned} \tag{11}$$

Therefore, using Theorem 4 and arguments analogous to those of section S1.1, we can prove bias and power properties analogous to those shown in section S1.2.

S1.4 Results for One-Sided Tests with Discrete p-values

Throughout this section, we assume that all tests are one-sided and yield discretely distributed p-values. Let $\boldsymbol{u} = \{u_1, \ldots, u_d\}$ represent the set of d unique possible p-values. Let λ be the largest member of $\boldsymbol{u}$ that is less than or equal to $\frac{1}{2}$. Now, we prove the following result.

Theorem 5 *For each $i \in \mathcal{H}_0$, assume that $F_i(\alpha) \leq \frac{F_i(\lambda)}{\lambda}\alpha$ for all $\alpha \leq \frac{1}{2}$ and that $F_i(\alpha) \geq \frac{F_i(\lambda)}{\lambda}\left(\alpha - \frac{1}{2}\right)$ for all $\alpha \geq 1 - \lambda$. Then $8\mathrm{E}(\bar{A})\alpha \geq \upsilon(\alpha)$ for all $\alpha \leq \frac{1}{2}$.*

Proof: By (9) we have that

$$\begin{aligned}
\mathrm{E}(A_i) &= 2\int_0^{\frac{1}{2}} \left[F_i\left(x+\frac{1}{2}\right) - F_i(x)\right] \mathrm{d}x \\
&= 2\left\{\int_0^{\lambda} \left[F_i\left(x+\frac{1}{2}\right) - F_i(x)\right] \mathrm{d}x + \int_{\lambda}^{\frac{1}{2}} \left[F_i\left(x+\frac{1}{2}\right) - F_i(x)\right] \mathrm{d}x\right\} \\
&\geq 2\left\{\int_0^{\lambda} \left[F_i(\lambda) - \frac{F_i(\lambda)}{\lambda}x\right] \mathrm{d}x + \int_{\lambda}^{\frac{1}{2}} \left[F_i(x+\frac{1}{2}) - F_i(x)\right] \mathrm{d}x\right\} \\
&\geq 2\left\{\int_0^{\lambda} \left[F_i\left(\frac{1}{2}\right) - \frac{F_i(\lambda)}{\lambda}x\right] \mathrm{d}x + \int_{\lambda}^{\frac{1}{2}} \left[\frac{F_i(\lambda)}{\lambda}x - F_i(\lambda)\right] \mathrm{d}x\right\}. \qquad (12) \\
&= 2\left\{\int_0^{\lambda} F_i(\lambda)\mathrm{d}x + \int_{\lambda}^{\frac{1}{2}} \frac{F_i(\lambda)}{\lambda}x\mathrm{d}x - \int_0^{\lambda} \frac{F_i(\lambda)}{\lambda}x\mathrm{d}x - \int_{\lambda}^{\frac{1}{2}} F_i(\lambda)\mathrm{d}x\right\} \\
&= 2\left\{\lambda F_i(\lambda) + \frac{F_i(\lambda)}{\lambda}\left(\frac{1}{4} - \frac{\lambda^2}{2}\right) - \frac{F_i(\lambda)}{\lambda}\frac{\lambda^2}{2} - \left(\frac{1}{2} - \lambda\right)F_i(\lambda)\right\} \\
&= F_i(\lambda)\left[\lambda + \frac{\left(\frac{1}{2}-\lambda\right)^2}{\lambda}\right].
\end{aligned}$$

Therefore,

$$\frac{\mathrm{E}(A_i)}{\lambda + \frac{\left(\frac{1}{2}-\lambda\right)^2}{\lambda}} \geq F_i(\lambda) \qquad (13)$$

and thus

$$\left(\frac{\mathrm{E}(A_i)}{\lambda + \frac{\left(\frac{1}{2}-\lambda\right)^2}{\lambda}}\right)\frac{\alpha}{\lambda} \geq F_i(\lambda)\frac{\alpha}{\lambda} \geq F_i(\alpha) \qquad (14)$$

for all $\alpha \leq \frac{1}{2}$. The denominator of the left-most term of (14) achieves a minimum value of $\frac{1}{8}$ at $\lambda = \frac{1}{4}$. Therefore, for each $i \in \mathcal{H}_0$, $F_i(\alpha) \leq 8\mathrm{E}(A_i)$ for all $\alpha \leq \frac{1}{2}$. ■

Using Theorem 5 and arguments analogous to those of section S1.2, we can prove bias and power properties analogous to those of section S1.2. Additionally, we note that (14) suggests power can be improved without compromising conservativeness, if λ is known. Therefore, in practice, one can let $\widehat{\lambda}$ be the largest observed p-value that is less than or equal to $\frac{1}{2}$. Then by substituting $\widehat{\lambda}$ into (14), one can obtain the constant

$$k(\widehat{\lambda}) = \frac{1}{\widehat{\lambda} + \frac{\left(\frac{1}{2}-\widehat{\lambda}\right)^2}{\widehat{\lambda}}} \qquad (15)$$

so that $\mathrm{E}(k(\widehat{\lambda}))\alpha \geq \upsilon(\alpha)$. By definition, the largest p-value less than or equal to $\frac{1}{2}$ must be less than or equal to λ. Because (15) has a unique minimum at $\widehat{\lambda} = \frac{1}{4}$, one can use (15) to improve power so long as $\widehat{\lambda} \geq \frac{1}{4}$.

S2 Figures for Examples

In this section, we display figures for the two example applications: the randomization test of Gadburdy et al. (2003) and the probe set filtering of Pounds and Cheng (2005a).

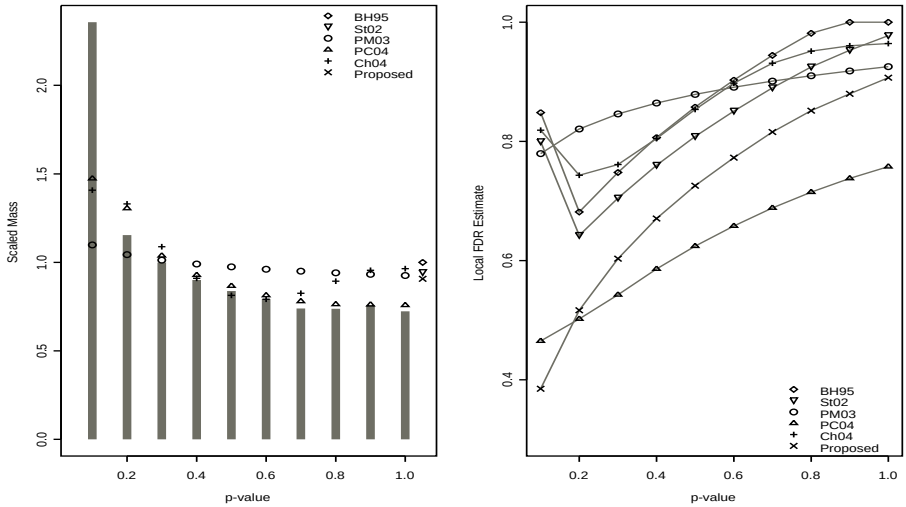

Figure S.1: Results for the p-values reported by Gadbury et al. (2003). The left panel shows the p-value pdf estimates obtained by PM03, PC04, and Ch04 against a properly scaled histogram of the p-values. The values of $\widehat{\pi}$ computed or used by BH95, St02, and the proposed method are shown by points to the right of the histogram. The right panel shows the values of the lFDR estimates t as a function of the p-value cut-off.

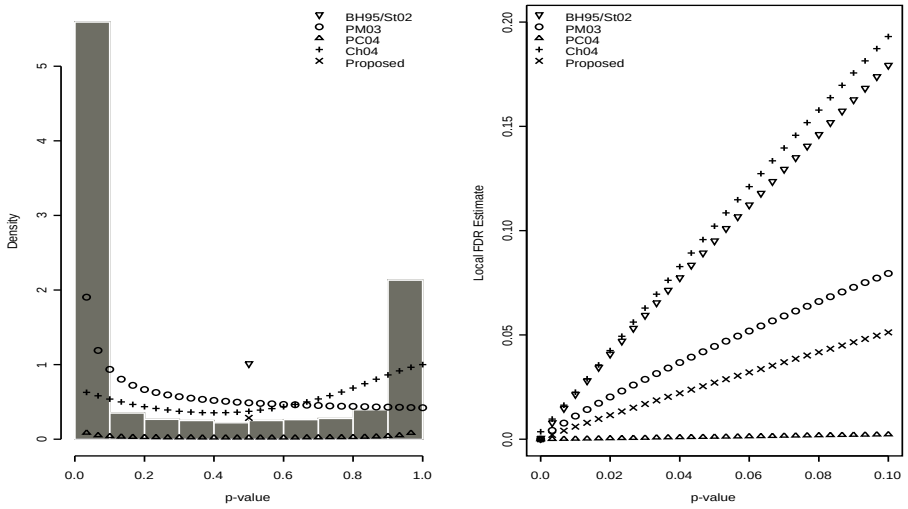

Figure S.2: Results for the pooled p-values used for probe-set filtering by Pounds and Cheng (2005). The left panel shows the p-value pdf estimates obtained by PM03 and PC04 against a histogram of the p-values. The left panel also shows the values of $\widehat{\pi}$ computed or used by BH95, St02, and the proposed method are shown as points above p=0.50. The right panel shows $t(\alpha)$ from each method. In this example, $q = t$ for all p and all methods.

S3 Simulation of p-values for One-sided Tests

In this section, we examine the distribution of the p-values one-sided Wilcoxon rank-sum tests ($H_0 : \delta = 0$ vs. $H_A : \delta > 0$) computed from resampling-based simulation from the data set with $\pi = 0.5$ and $\delta = 1.0$. We consider sample sizes of $n = 3, 5, 10$. Figure S.3 shows the results. We now elaborate on how these plots show the difficulties that FDR methods can encounter when applied to p-values from one-sided tests and how our proposed method overcomes those difficulties.

Figure S.3 clearly shows that the p-values behave as described in section 2.4 of the manuscript. The the distribution of p-values from tests with true nulls roughly follows the line $y = x$, which would correspond to uniformity. The p-values from tests with true "tested alternatives" (i.e. $\delta > 0$) are stochastically less than uniform (distribution is above the line $y = x$). The p-values from tests with true "untested alternatives" (i.e. $\delta < 0$) are stochastically greater than uniform (distribution is below the line $y = x$).

The graphs also show the problems that one-sided testing can introduce in FDR estimation and/or control. For $\delta < 0$, we observe that it is often the case that $\mathbf{F}(\alpha)$ increases at a rate faster than α for some $\alpha > \frac{1}{2}$. This causes $\frac{\alpha}{\mathbf{F}(\alpha)}$ to *decrease* for some large α. In some cases, such as in Pounds and Cheng (2005a), it is possible that $\frac{\alpha}{\mathbf{F}(\alpha)}$ can be smaller for $\alpha \approx 1$ than for $\alpha \approx 0$. Such a case clearly introduces downward bias into equation 3 of the manuscript.

In all panels of Figure S.3 with $\delta = 0$ or $\delta < 0$, the simulation estimate of the average p-value CDF falls below the diagonal line for all $\alpha < \frac{1}{2}$. This suggests that the ratio $\frac{\widehat{\pi}\alpha}{\widehat{\mathbf{F}}(\alpha)}$ is conservative for all $\alpha < \frac{1}{2}$. The proposed method counts all p-values greater than or equal to $\frac{1}{2}$ as if they come from tests with a true null hypothesis; therefore, the proposed method does not lose this conservativeness for $\alpha \geq \frac{1}{2}$.

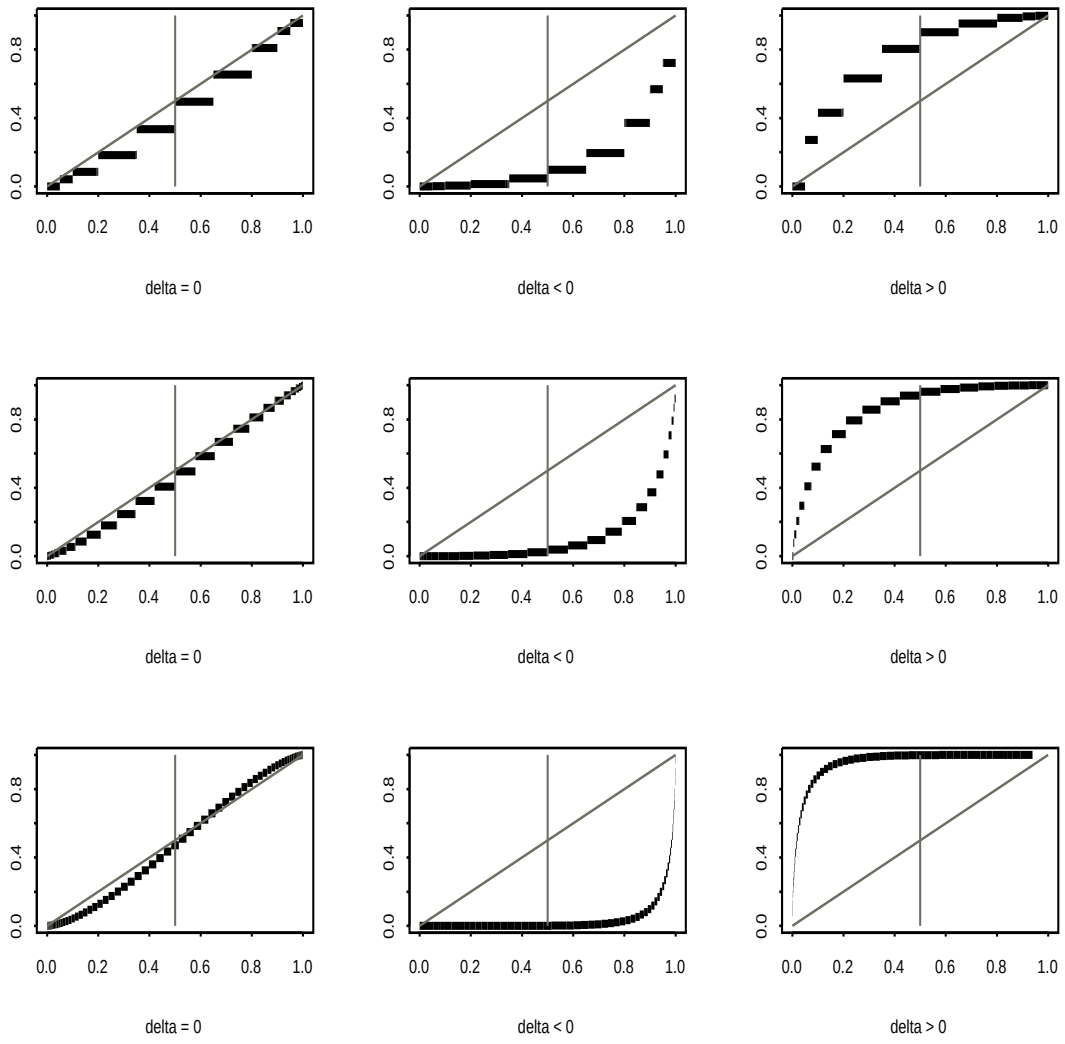

Figure S.3: Null Distribution of P-values from Resampling-Based Simulation of the Data Set with $\pi = 0.5$ and $\delta = 1$. In each panel above, we plot the p-value $\widehat{\mathbf{F}}(\alpha)$ distribution (averaged across reps) against α for the subsets of tests. The panels on the first, middle, and bottom row depict results for sample sizes $n = 3$, $n = 5$, and $n = 10$, respectively. The left, middle, and right columns of panels correspond to p-values from tests with $\delta = 0$, $\delta < 0$, and $\delta > 0$, respectively. In this example, all p-values were computed for a test of the form $H_0 : \delta = 0$ vs. $H_0 : \delta > 0$.

S4 Simulation Results

In this section, we display plots for all simulation results. As mentioned in the manuscript, we performed resampling-based simulations for all combinations of three sample sizes ($n = 3, 5, 10$), number of sides of the tests (one-sided or two-sided) for each of five data sets. Therefore, a total of 30 simulations were performed. We show the results for each of those simulations below. Each figure includes 6 panels, each plotting the simulation results for a method presented in the manuscript: Benjamini and Hochberg (1995), BH95; Storey (2002), St02; Pounds and Morris (2003), PM03; Pounds and Cheng (2004), PC04; and Cheng et al. (2004), Ch04. In each plot, the thick solid line represents the simulation estimate of the FDR incurred by using the p-value threshold; the thin solid line represents the average value of q at p for the indicated method; and the dashed lines represent the value of q at p for each replication. In all simulations, the rank-sum test was used to compute p-values. Therefore, the captions of the figures only indicate the sidedness of the tests (one-sided or two-sided), the per-group sample size ($n = 3, 5$, or 10), the proportion of probe sets that are not differentially expressed ($\pi = 0.5, 0.9$, or 1.0) in the data set, and the effect size of differentially expressed probe sets relative to the standard deviation ($\delta = 0.5$ or 1.0) in data sets with $\pi \neq 1.0$.

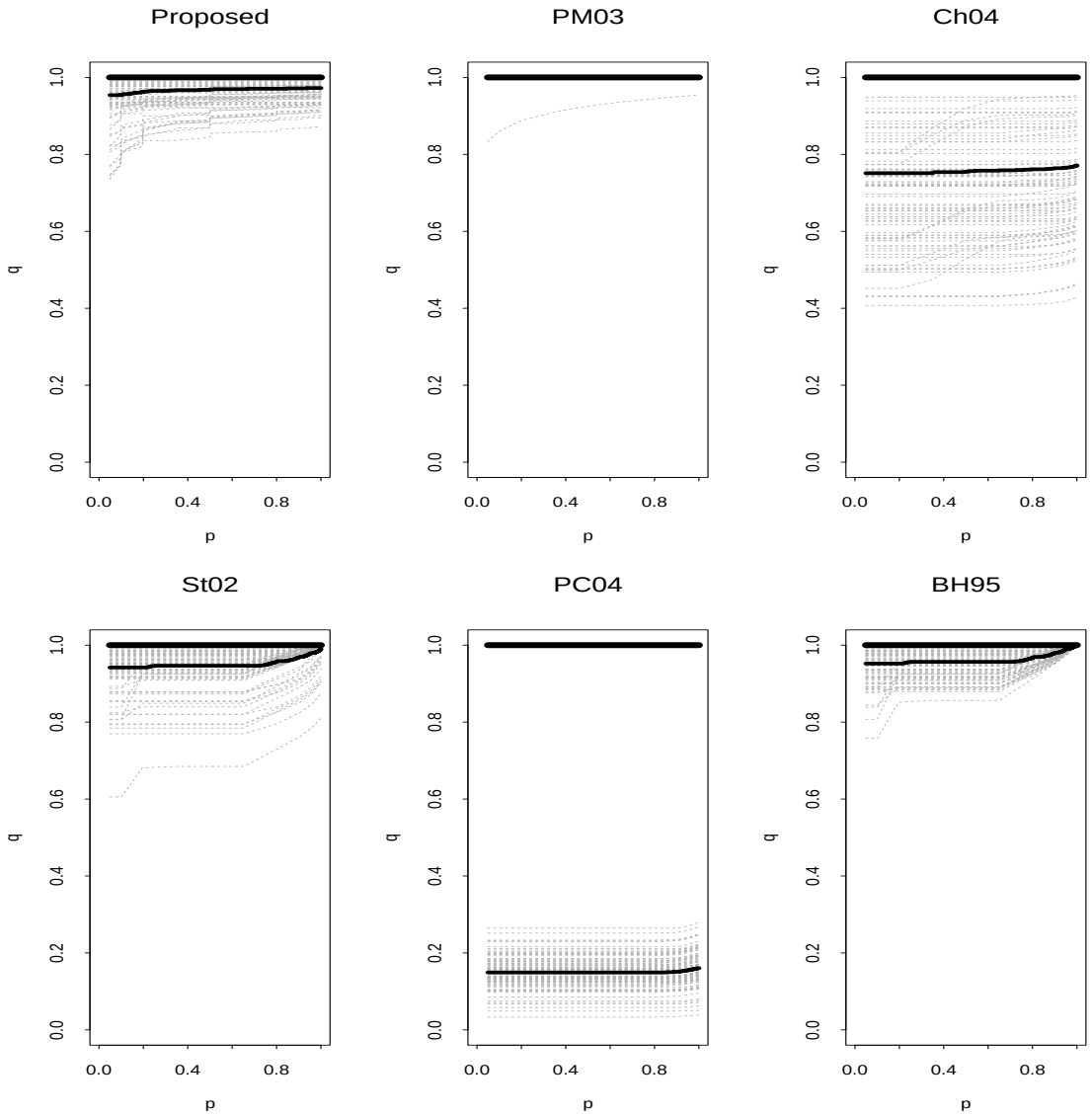

Figure S.4: Simulation Results with One-sided Tests, $n = 3$, and $\pi = 1.0$.

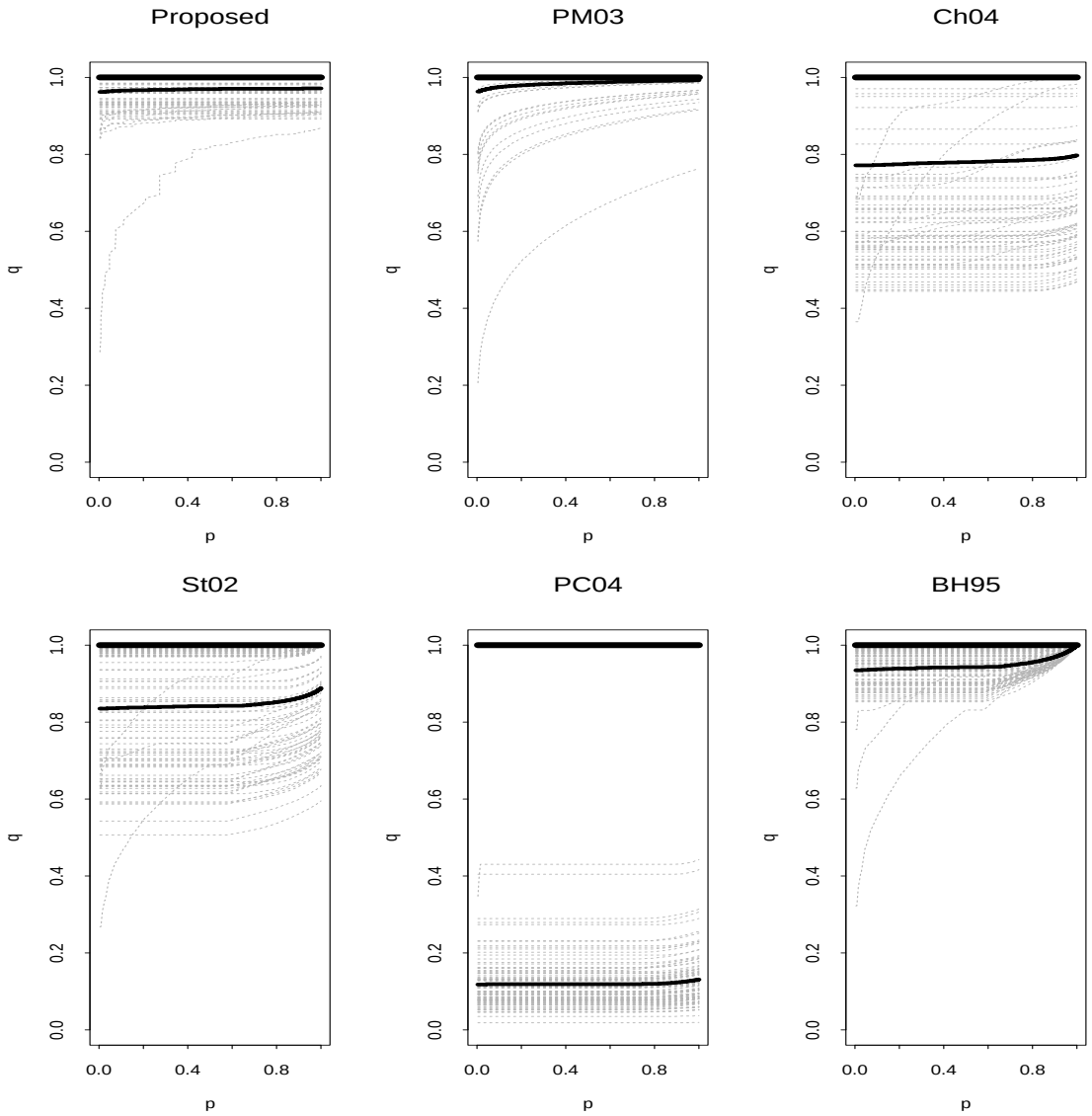

Figure S.5: Simulation Results with One-sided Tests, $n = 5$, and $\pi = 1.0$.

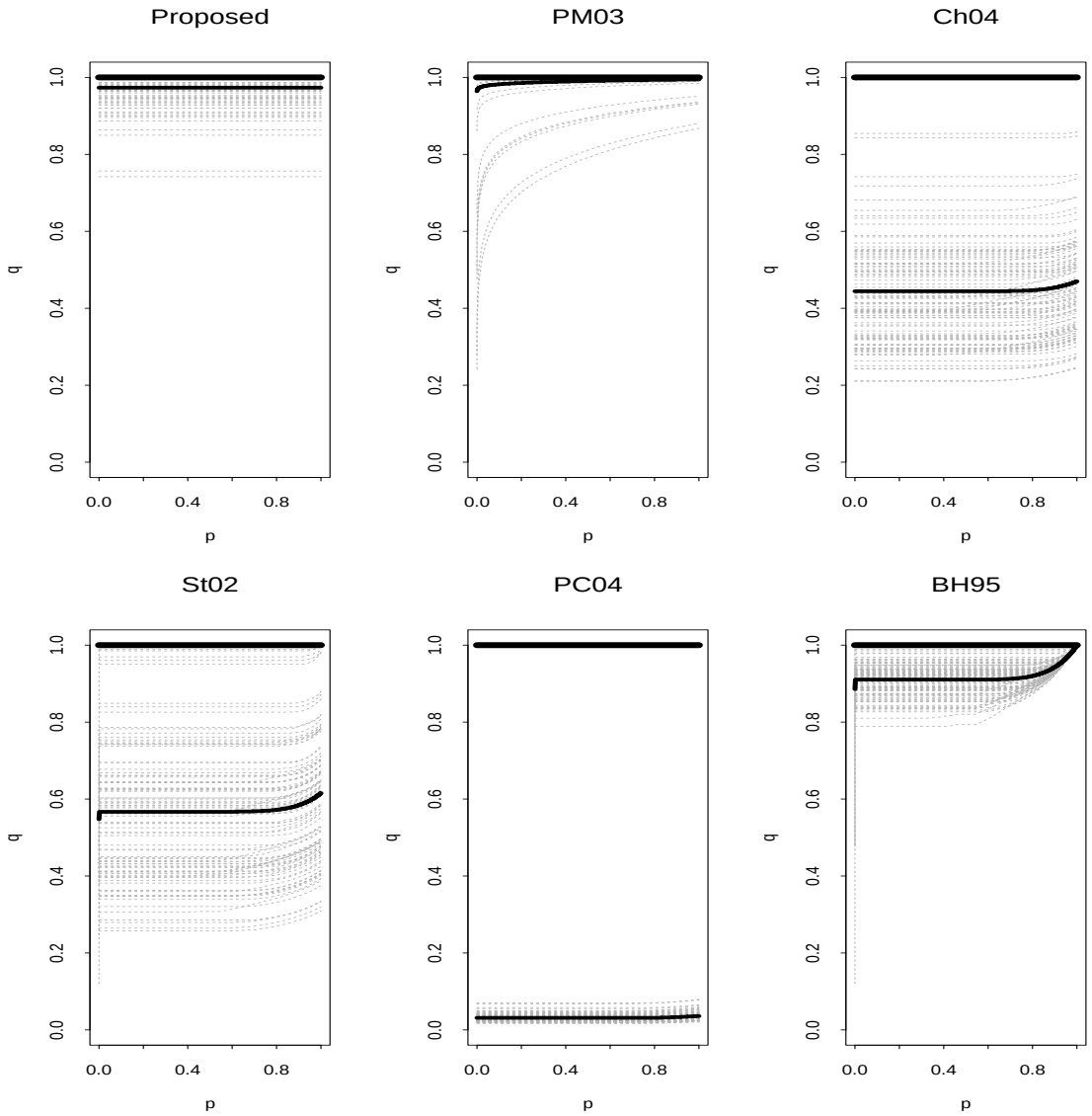

Figure S.6: Simulation Results with One-sided Tests, $n = 10$, and $\pi = 1.0$.

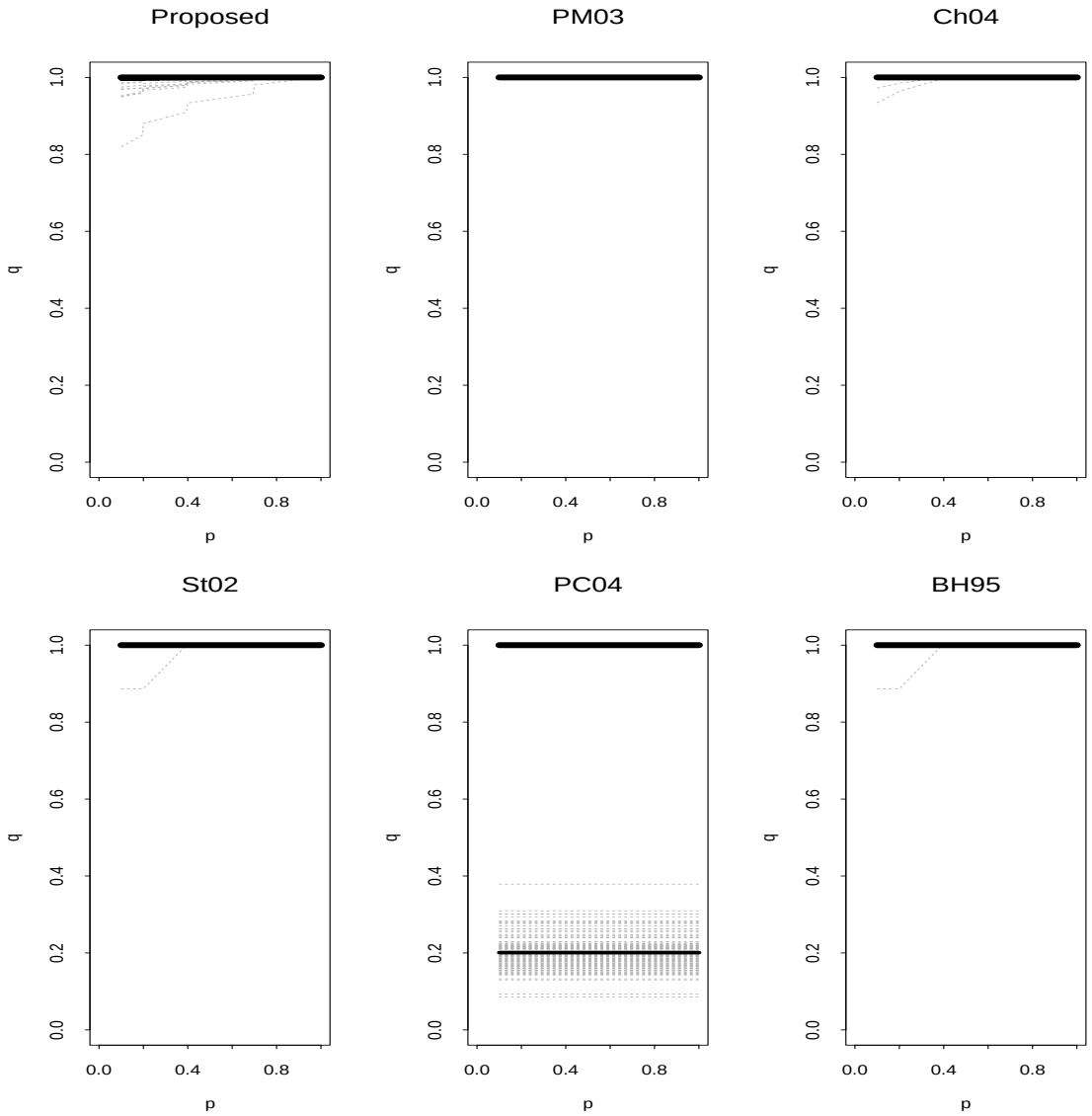

Figure S.7: Simulation Results with Two-sided Tests, $n = 3$, and $\pi = 1.0$

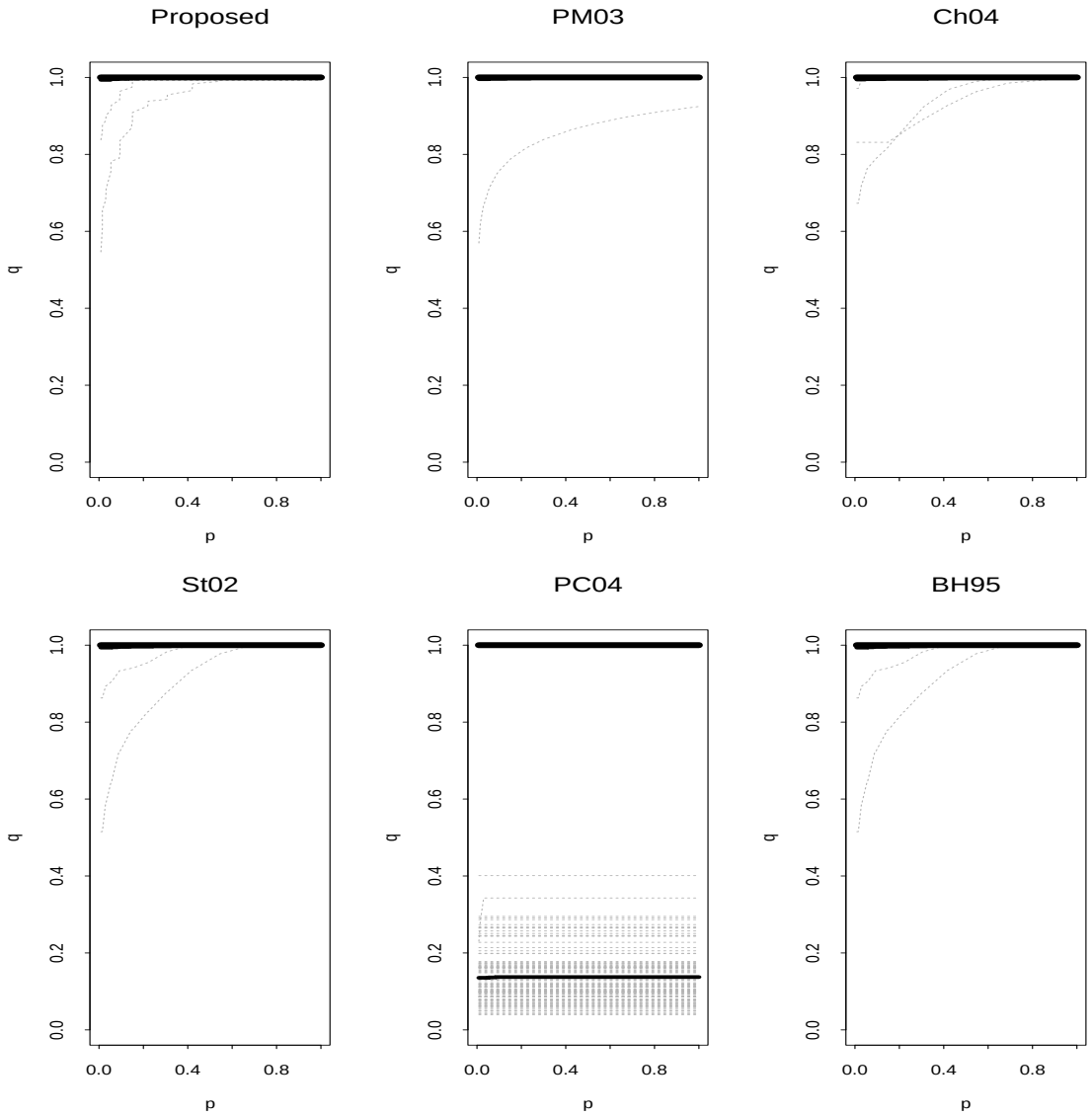

Figure S.8: Simulation Results with Two-sided Tests, $n = 5$, and $\pi = 1.0$.

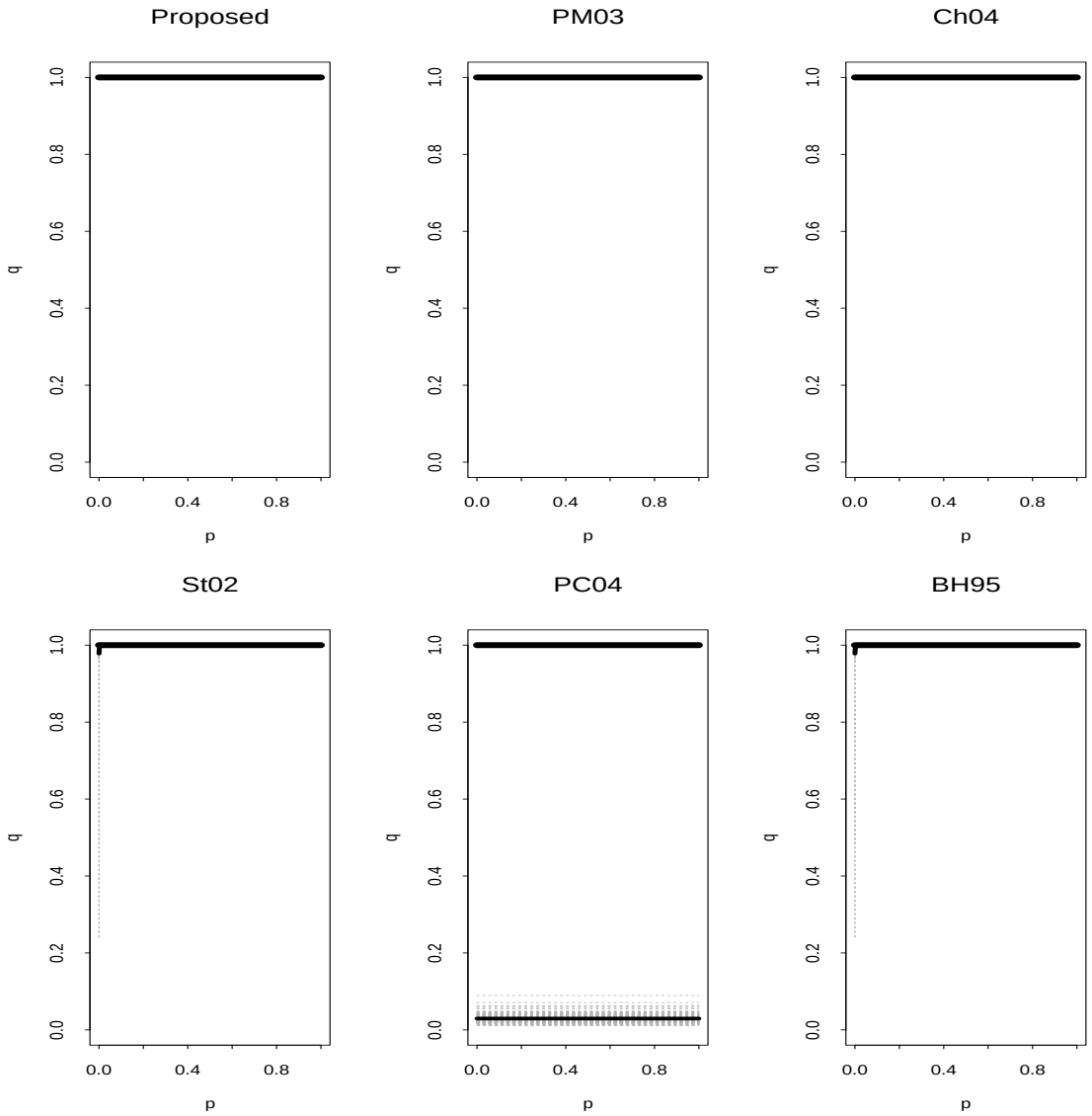

Figure S.9: Simulation Results with Two-sided Tests, $n = 10$, and $\pi = 1.0$.

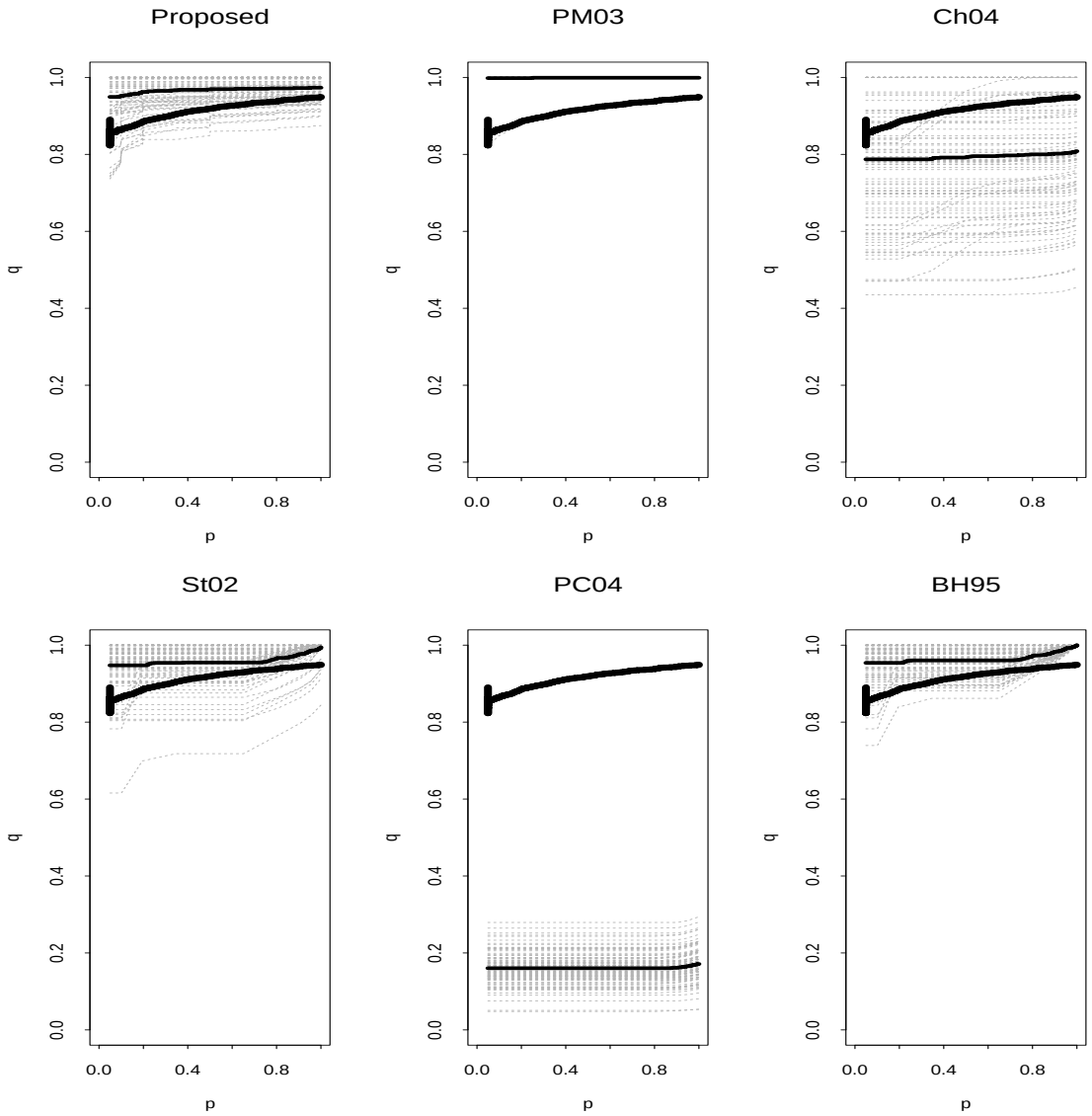

Figure S.10: Simulation Results with One-sided Tests, $n = 3$, $\pi = 0.9$ and $\delta = 0.5$.

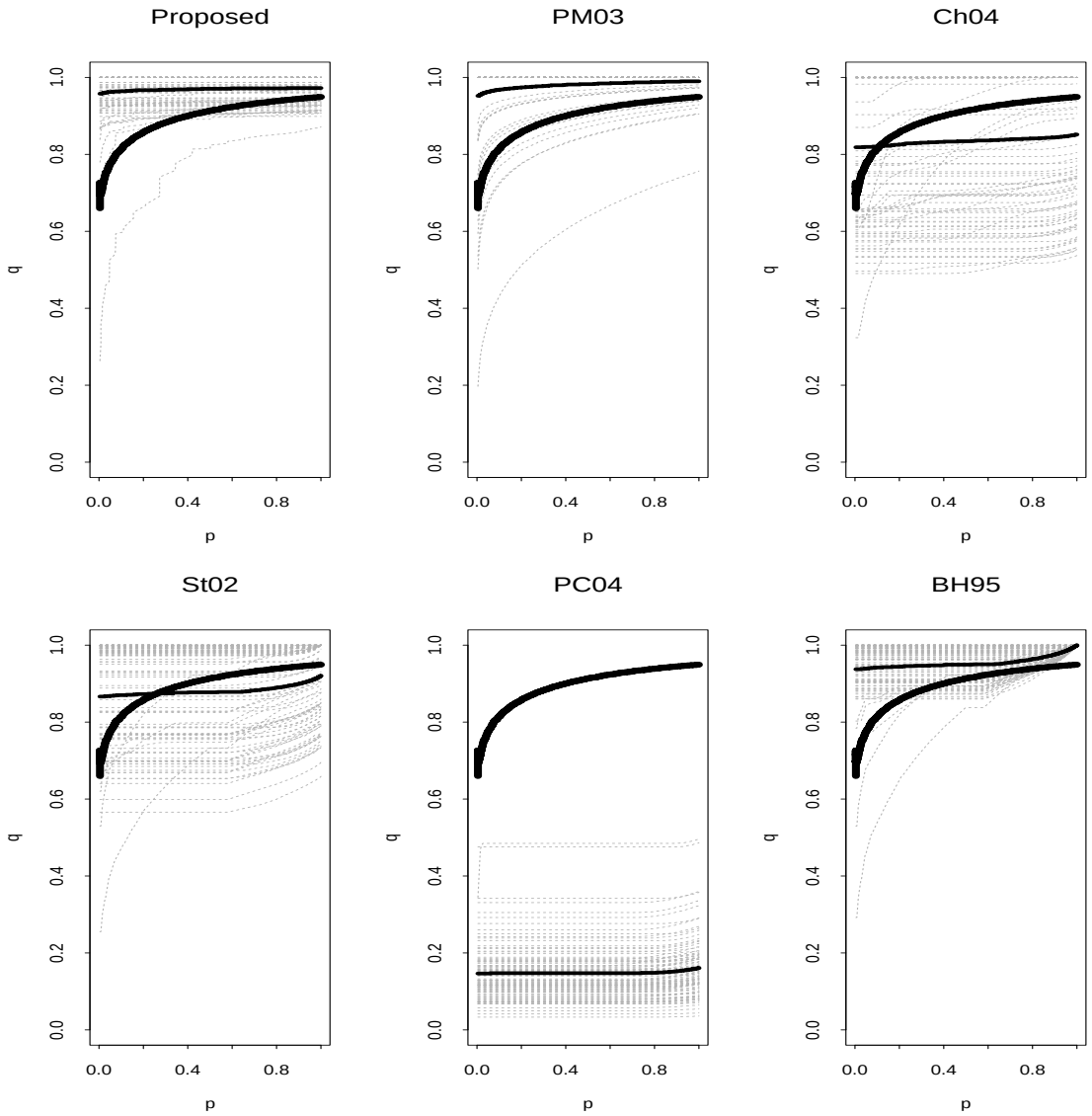

Figure S.11: Simulation Results with One-sided Tests, $n = 5$, $\pi = 0.9$, and $\delta = 0.5$.

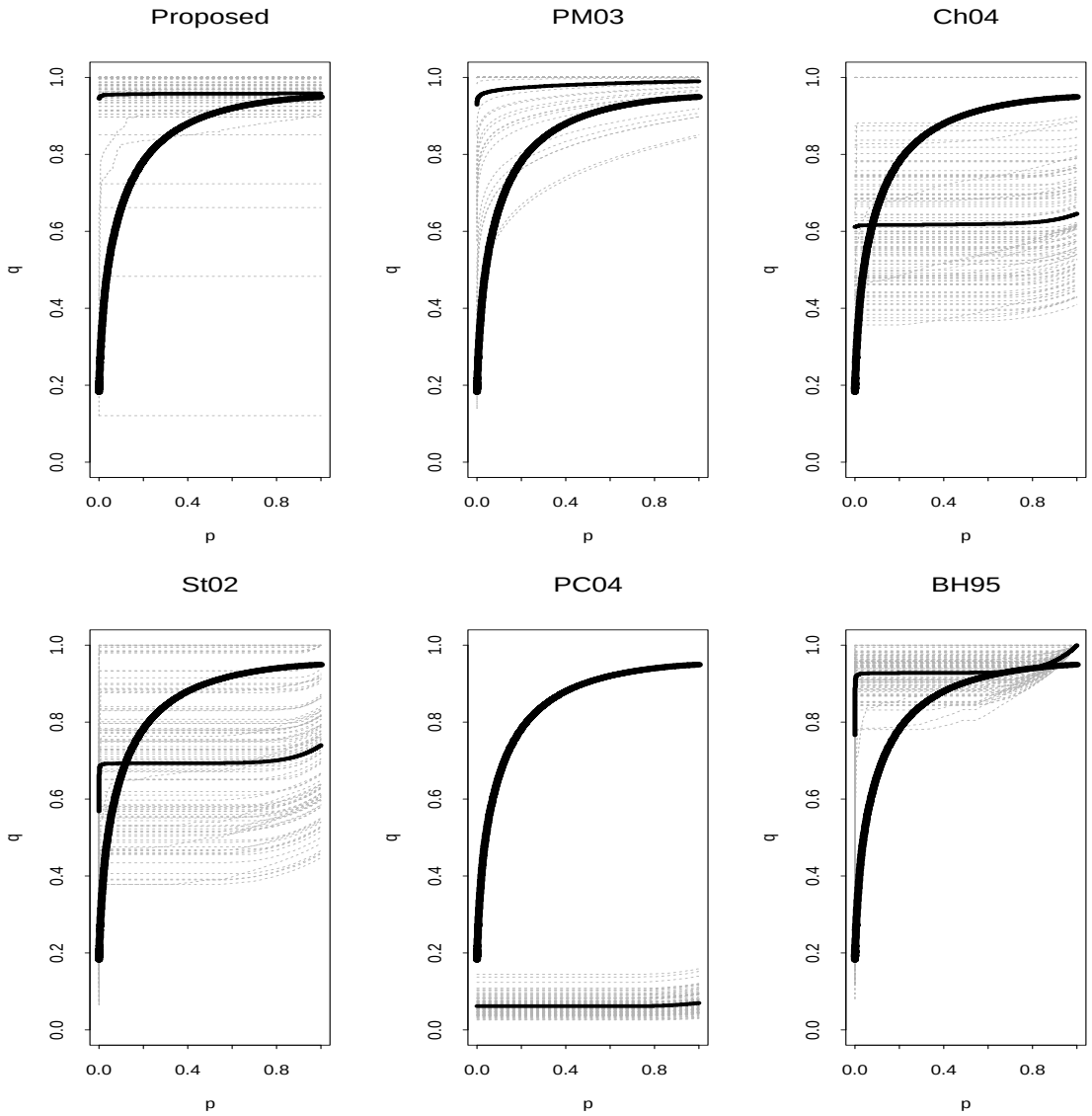

Figure S.12: Simulation Results with One-sided Tests, $n = 10$, $\pi = 0.9$, and $\delta = 0.5$.

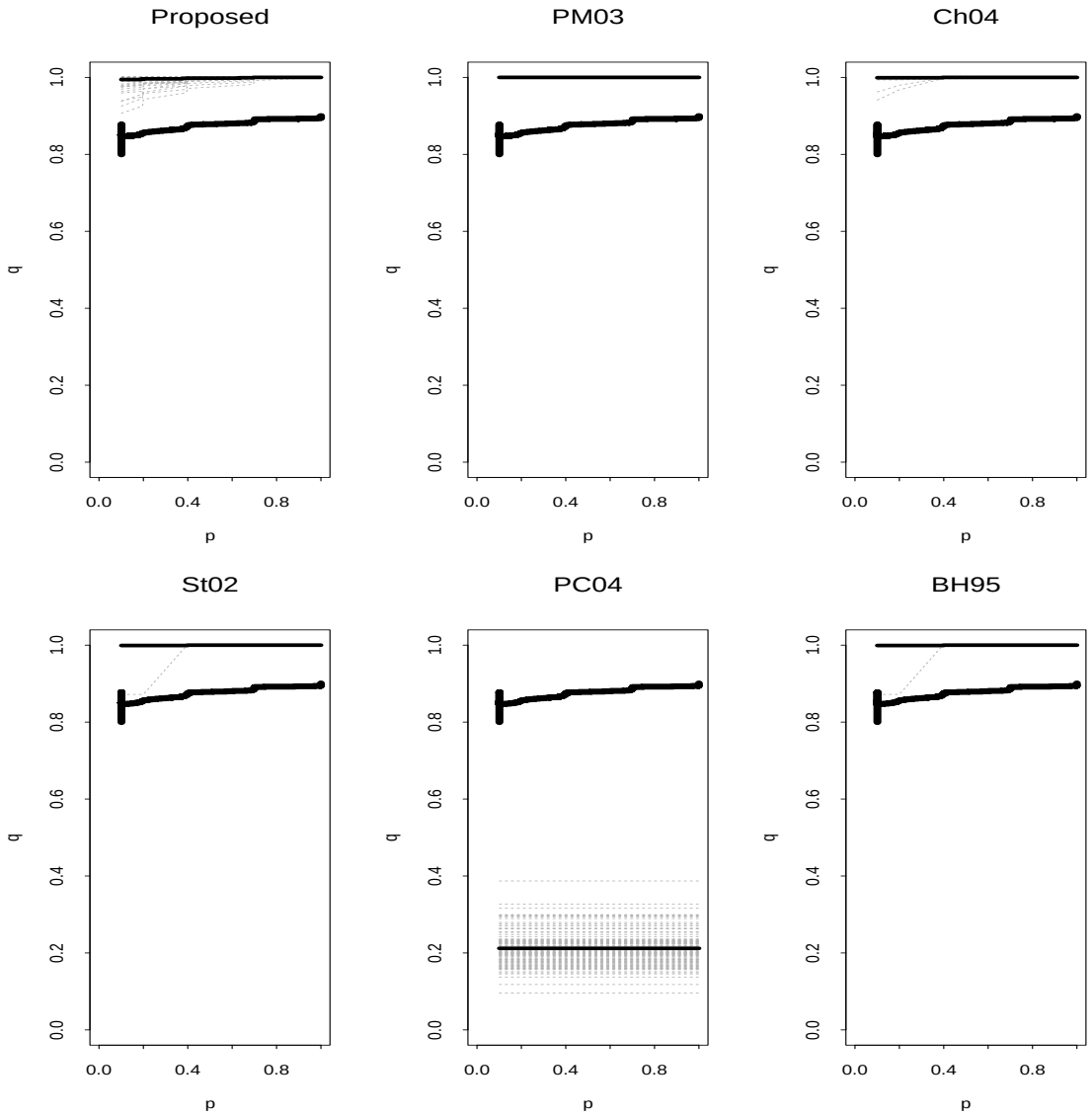

Figure S.13: Simulation Results with Two-sided Tests, $n = 3$, $\pi = 0.9$, and $\delta = 0.5$.

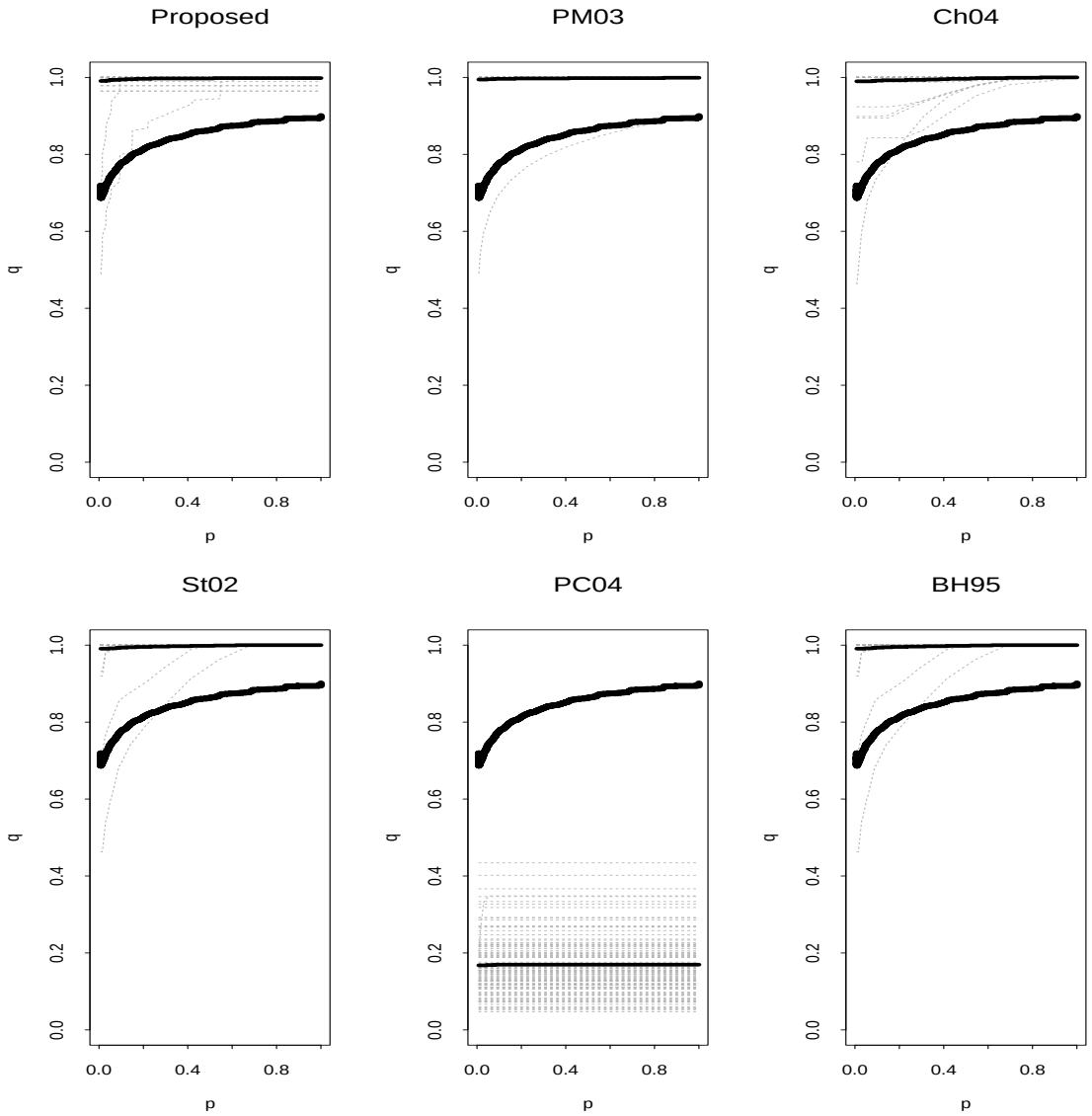

Figure S.14: Simulation Results with Two-sided Tests, $n = 5$, $\pi = 0.9$, and $\delta = 0.5$.

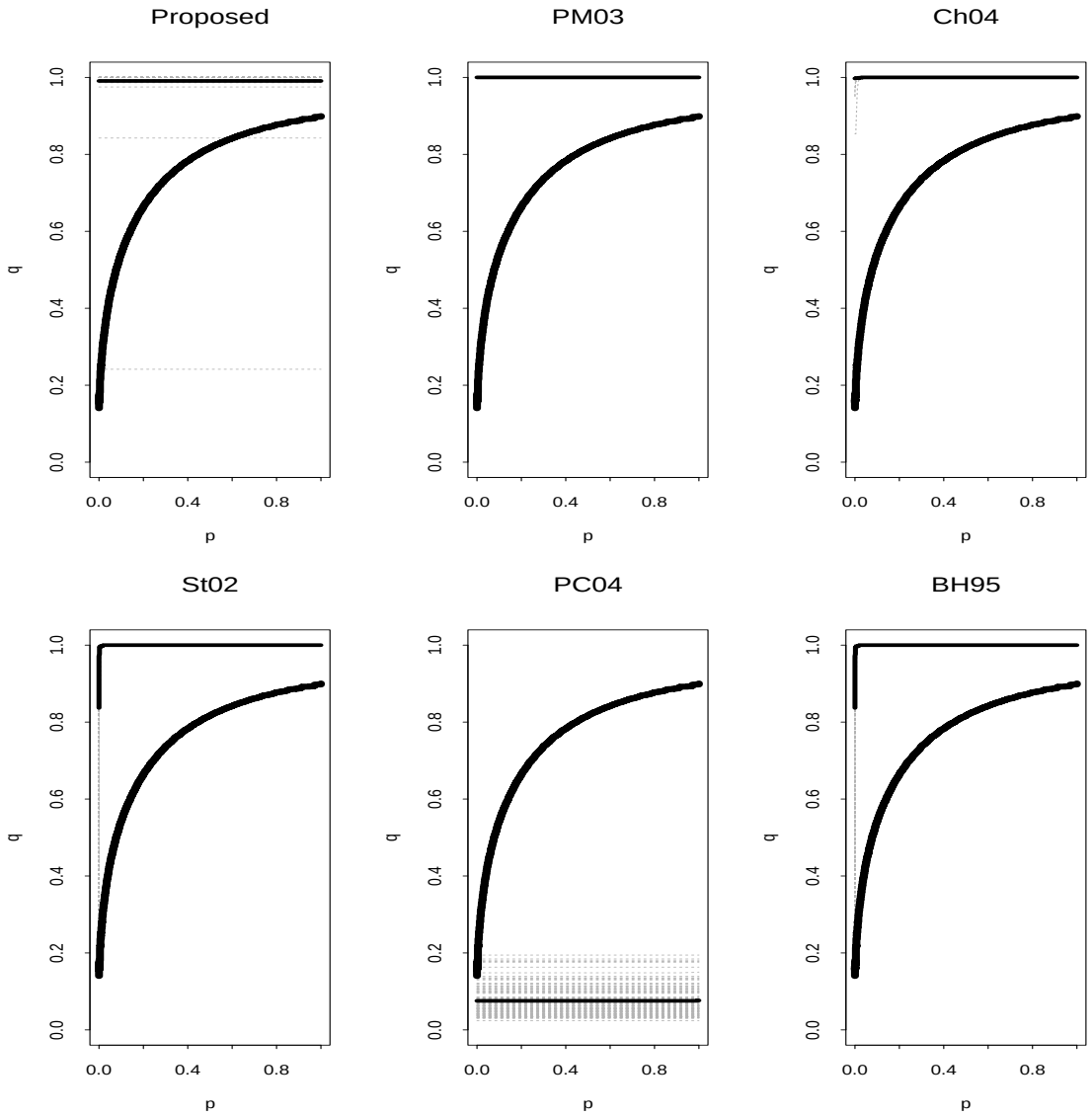

Figure S.15: Simulation Results with Two-sided Tests, $n = 10$, $\pi = 0.9$, and $\delta = 0.5$.

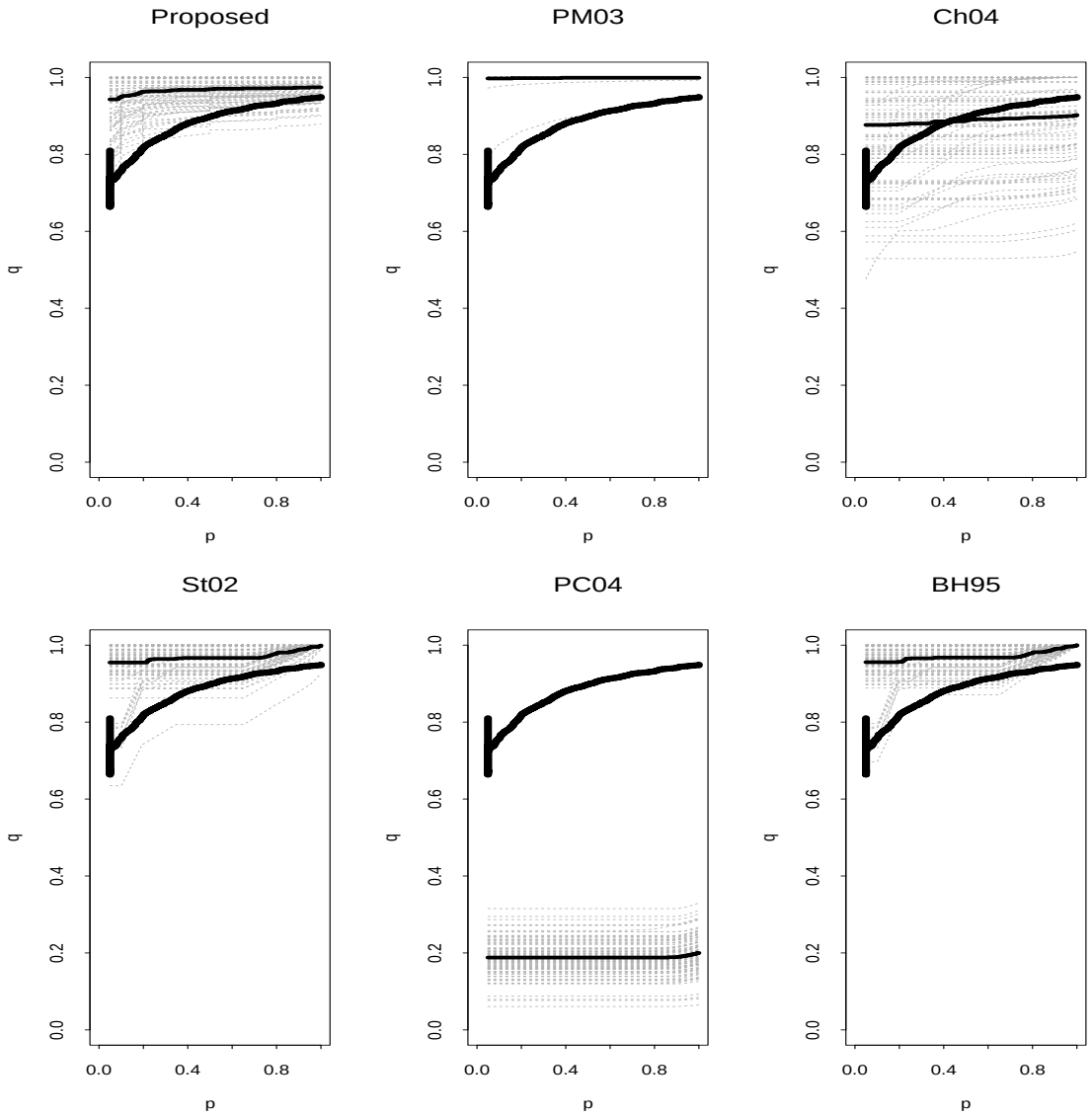

Figure S.16: Simulation Results with One-sided Tests, $n = 3$, $\pi = 0.9$, and $\delta = 1.0$.

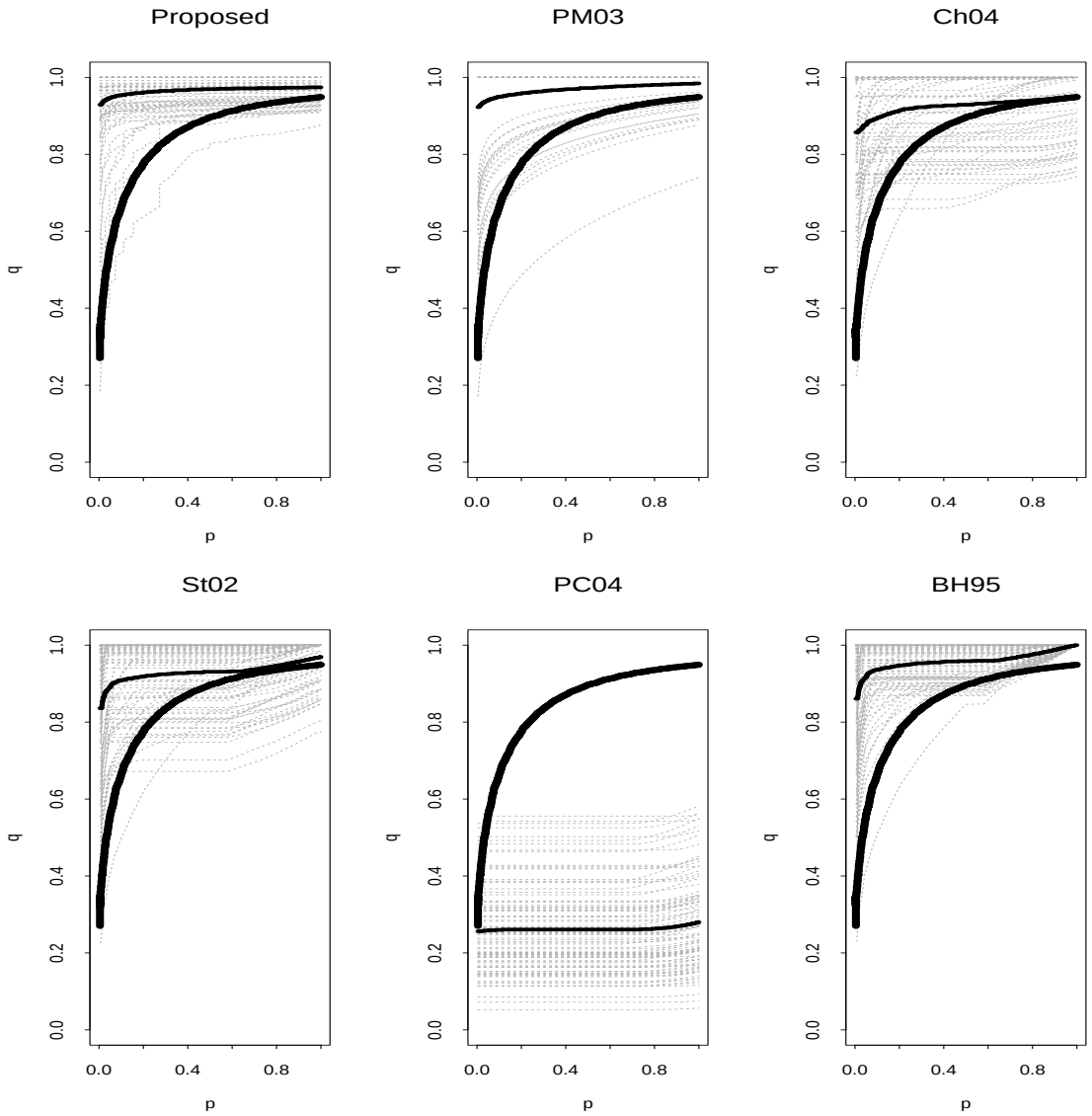

Figure S.17: Simulation Results with One-Sided Tests, $n = 5$, $\pi = 0.9$, and $\delta = 1.0$.

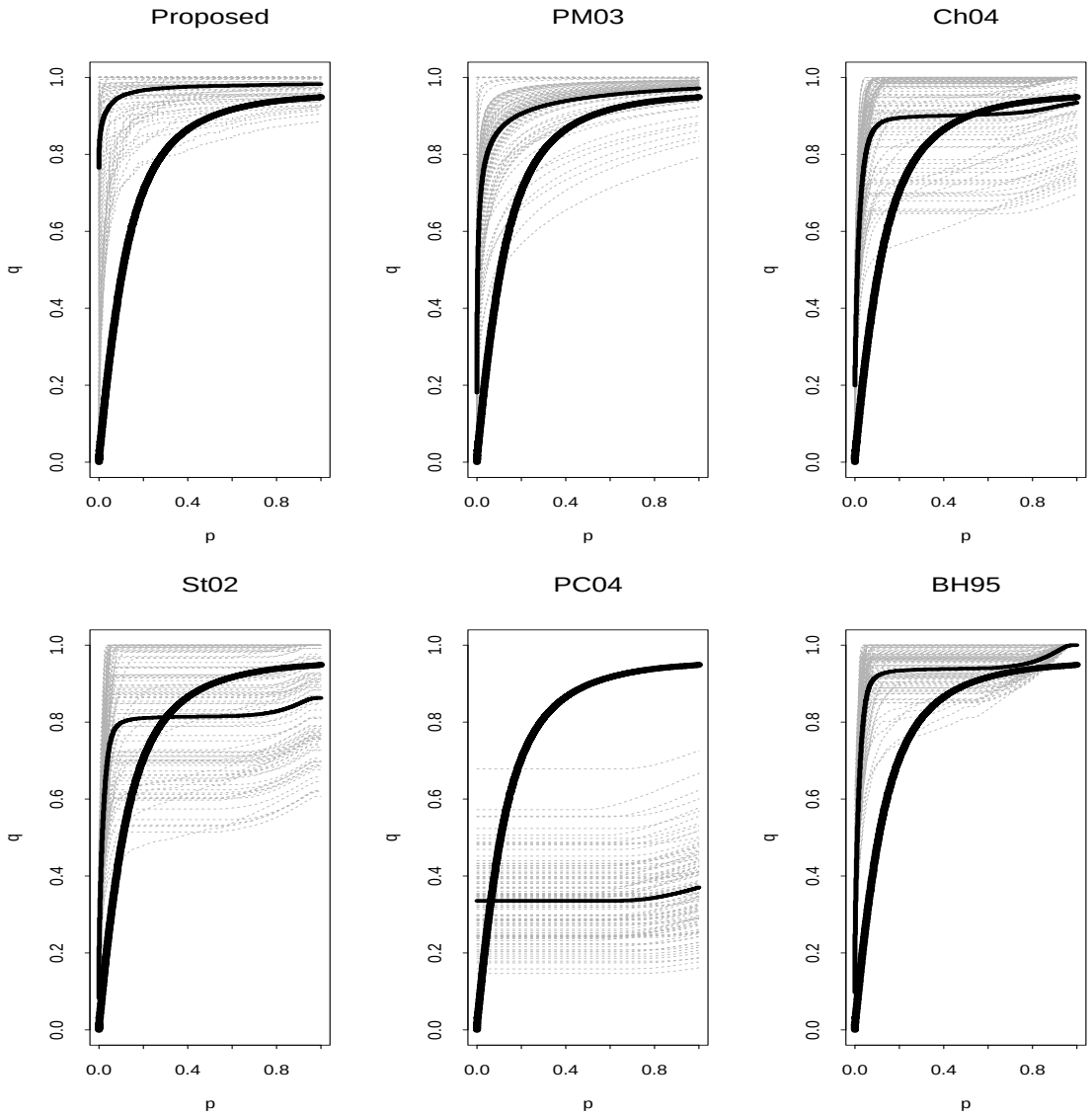

Figure S.18: Simulation Results with One-Sided Tests, $n = 10$, $\pi = 0.9$, and $\delta = 1.0$.

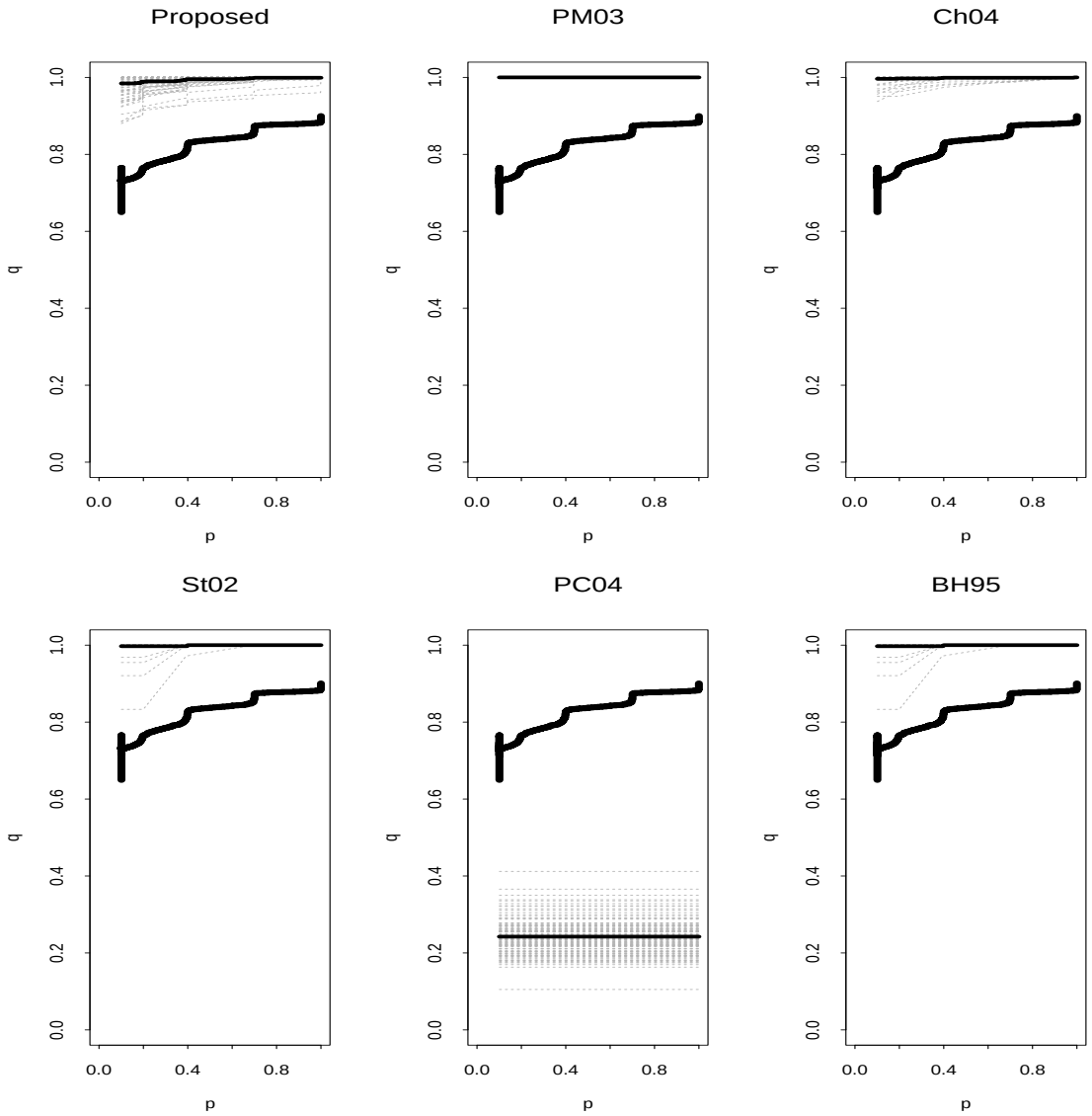

Figure S.19: Simulation Results with Two-Sided Tests, $n = 3$, $\pi = 0.9$, and $\delta = 1.0$.

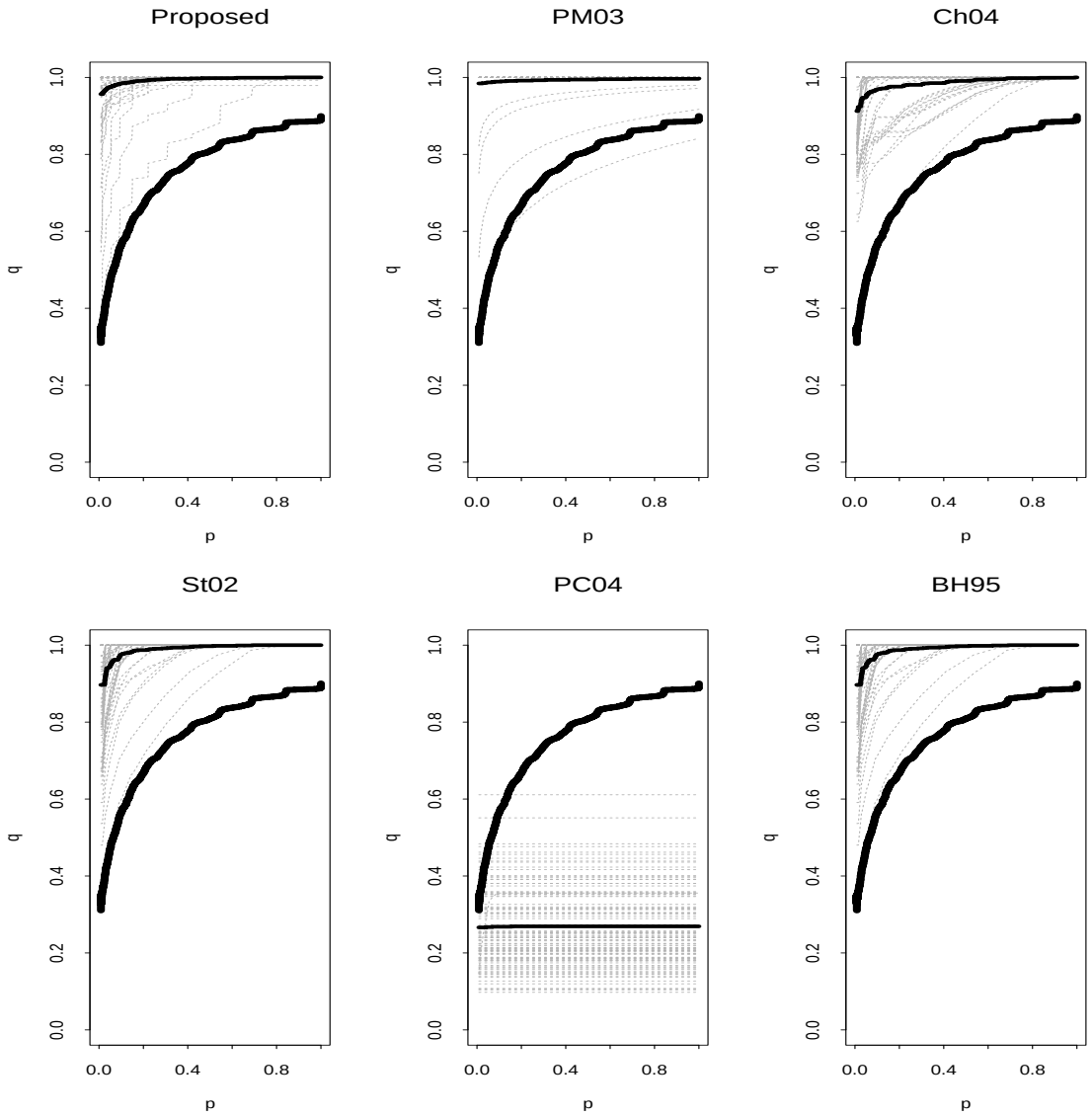

Figure S.20: Simulation Results with Two-Sided Tests, $n = 5$, $\pi = 0.9$, and $\delta = 1.0$.

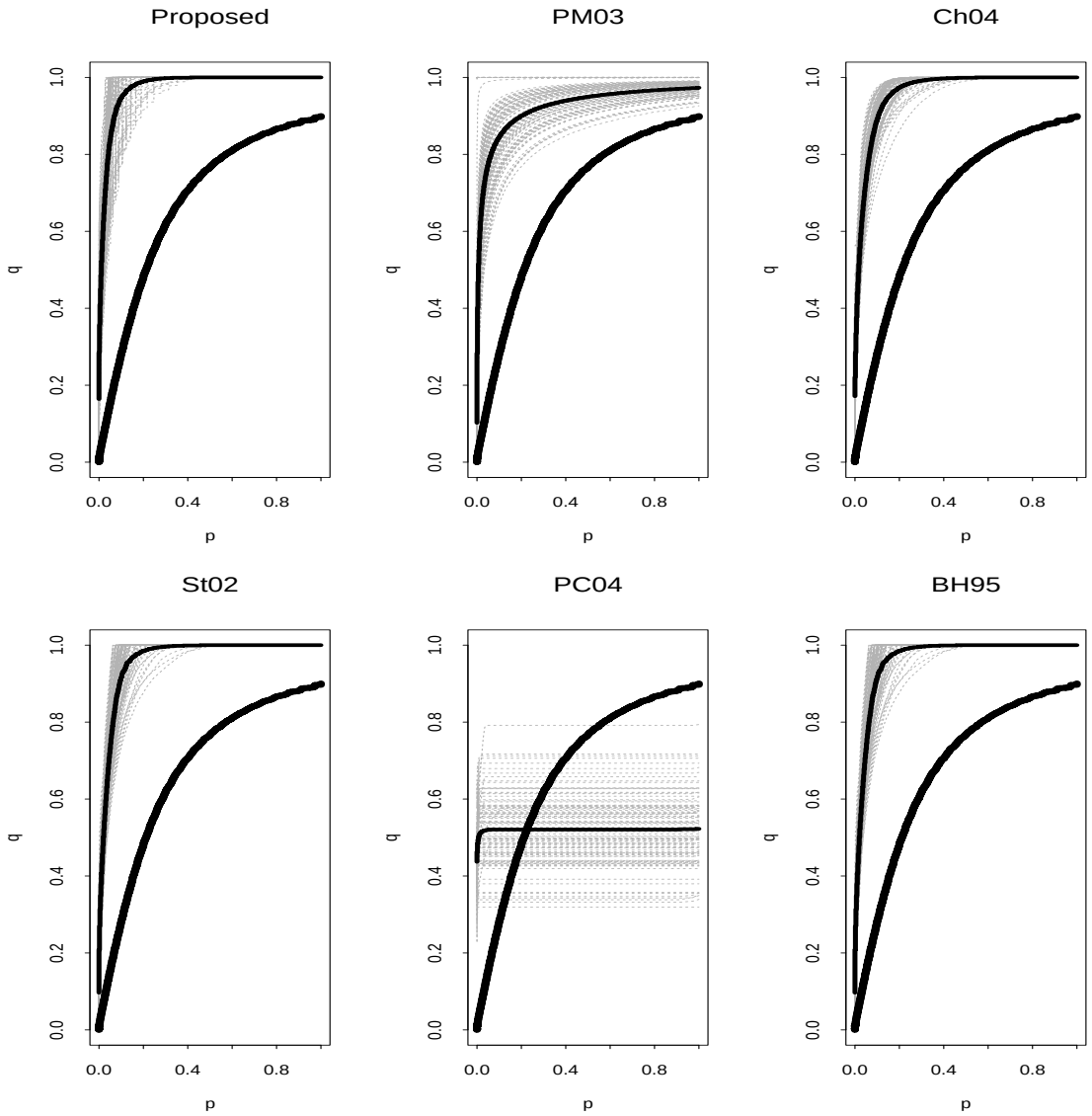

Figure S.21: Simulation Results with Two-Sided Tests, $n = 10$, $\pi = 0.9$, and $\delta = 1.0$.

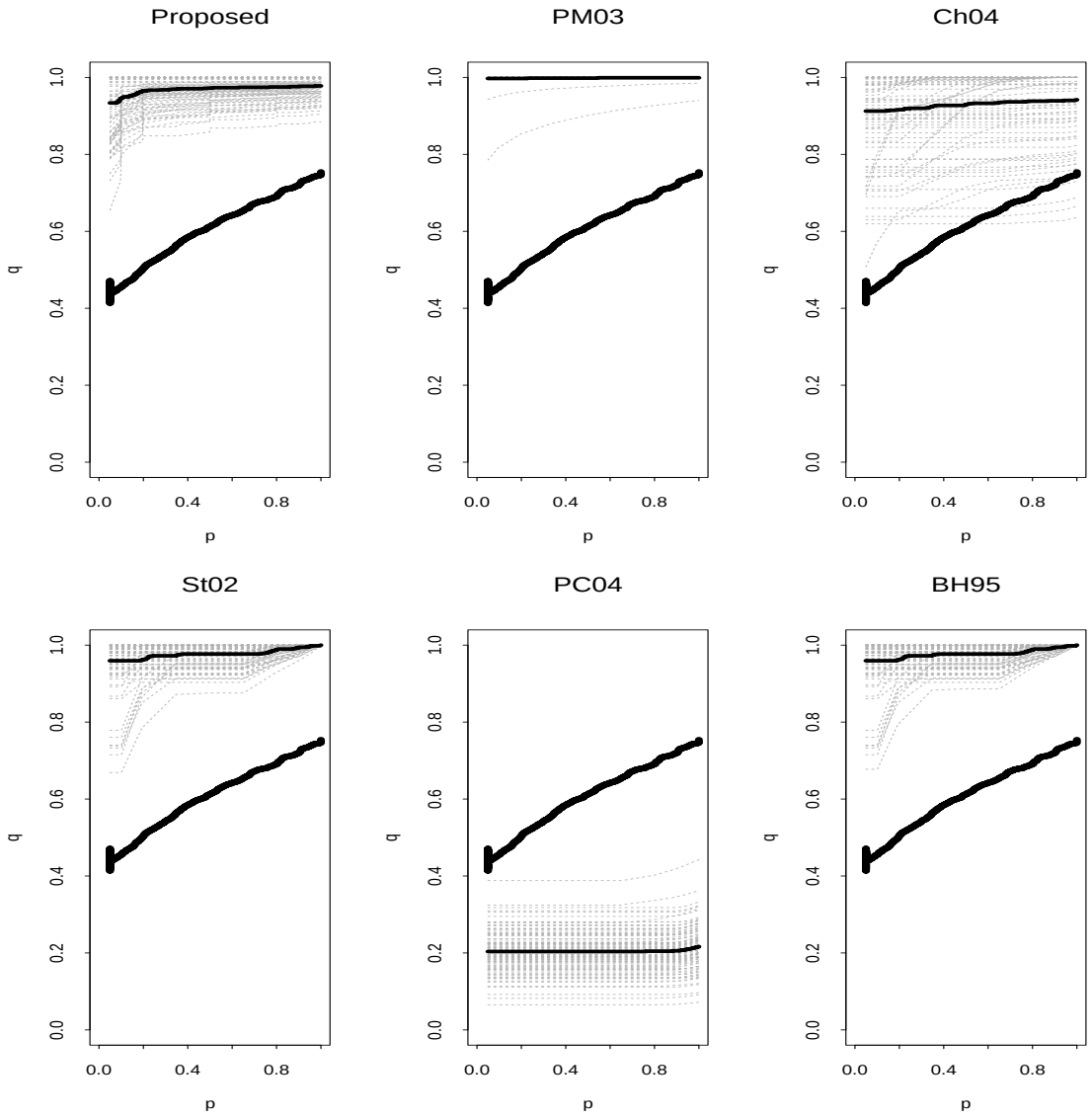

Figure S.22: Simulation Results with One-Sided Tests, $n = 3$, $\pi = 0.5$, and $\delta = 0.5$.

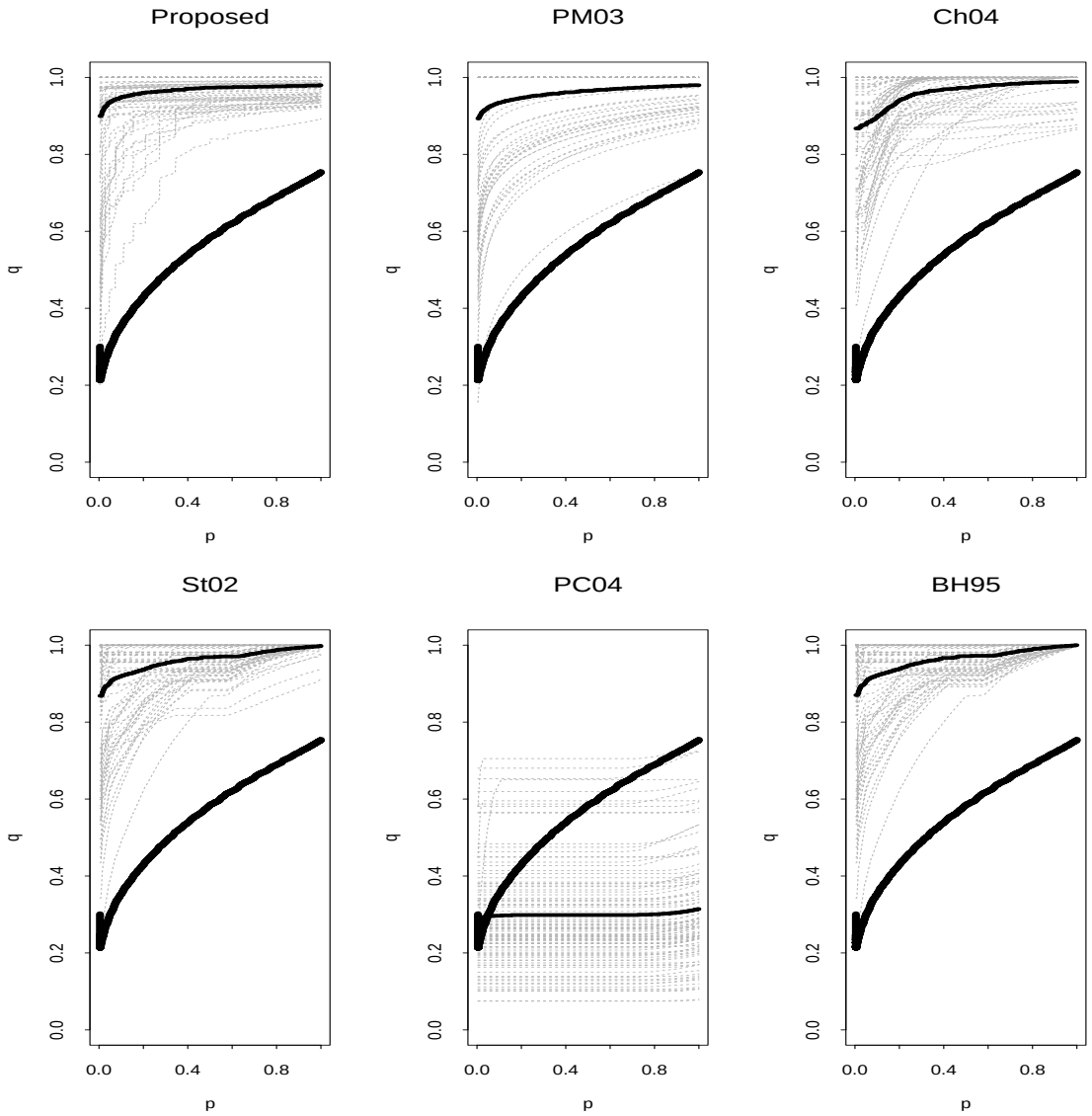

Figure S.23: Simulation Results with One-Sided Tests, $n = 5$, $\pi = 0.5$, and $\delta = 0.5$.

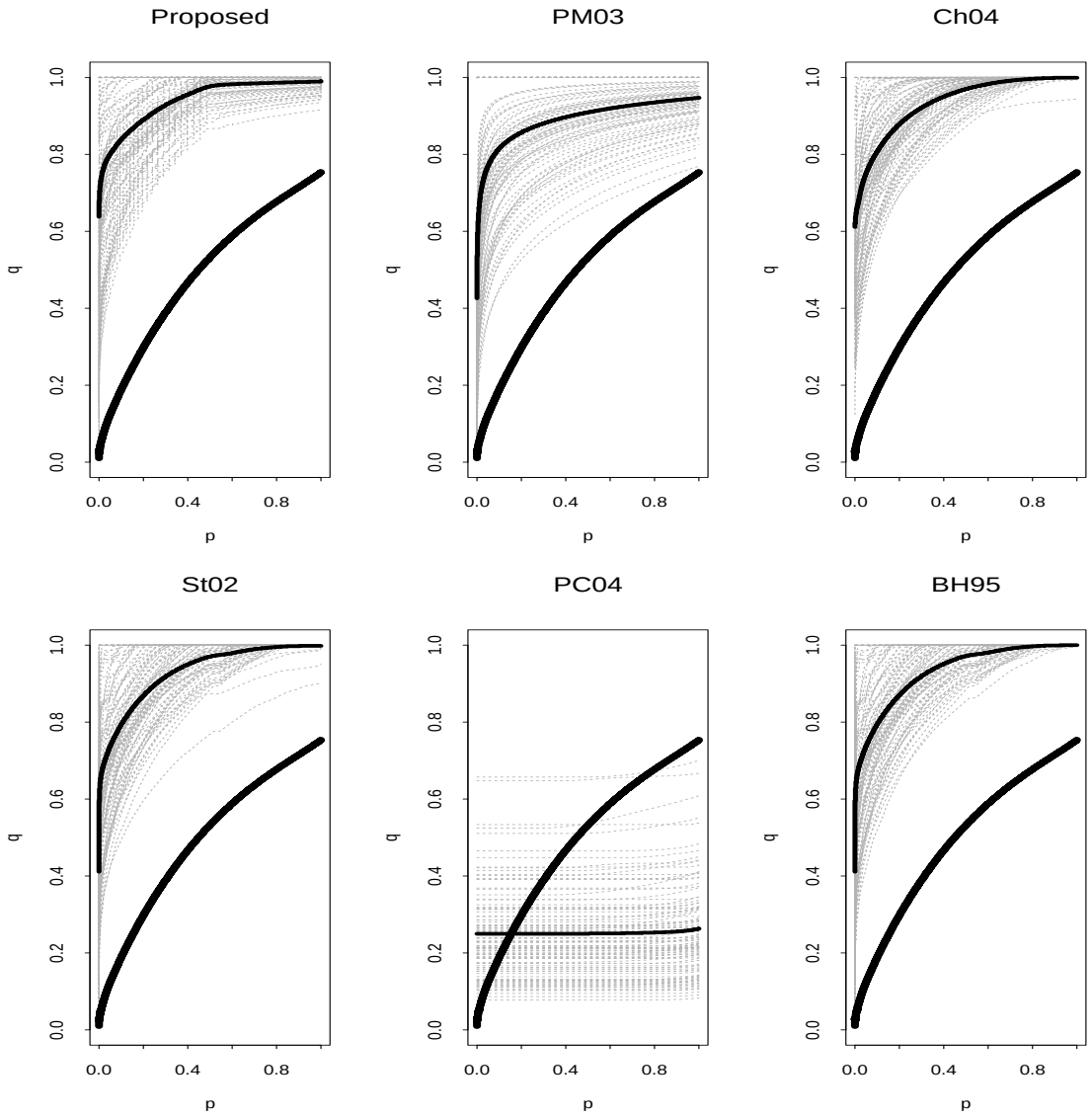

Figure S.24: Simulation Results with One-Sided Tests, $n = 10$, $\pi = 0.5$, and $\delta = 0.5$.

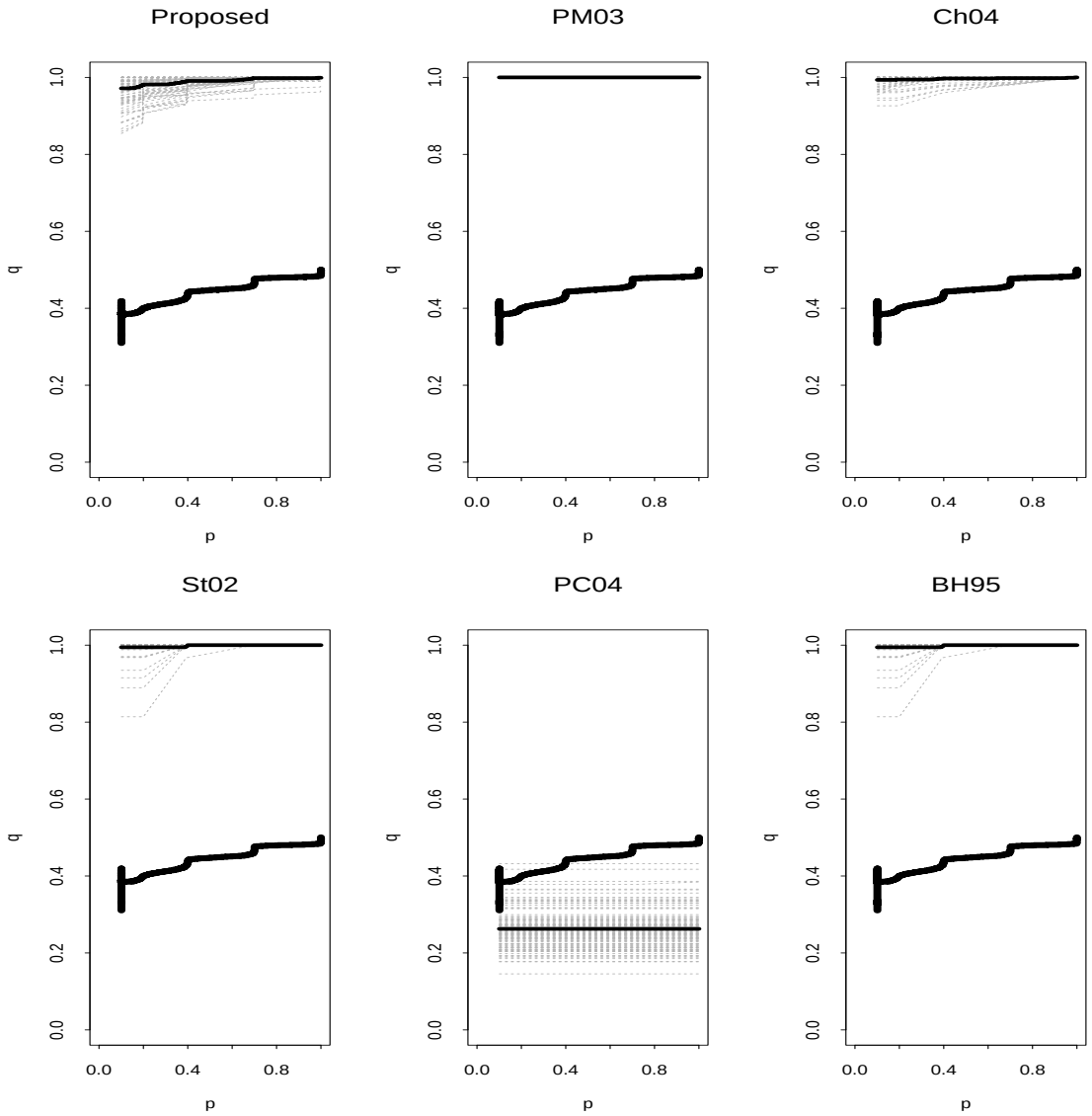

Figure S.25: Simulation Results with Two-Sided Tests, $n = 3$, $\pi = 0.5$, and $\delta = 0.5$.

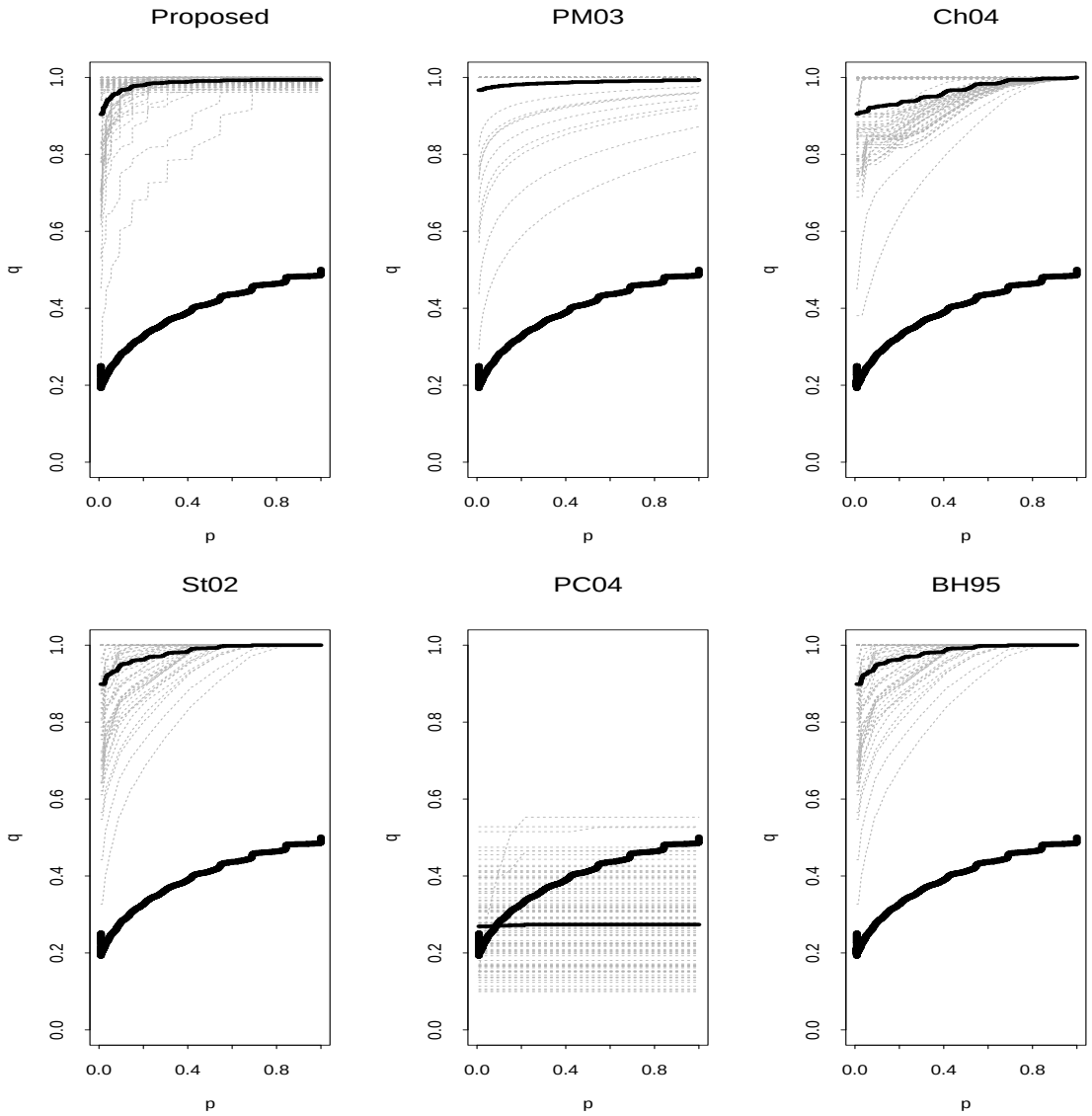

Figure S.26: Simulation Results with Two-Sided Tests, $n = 5$, $\pi = 0.5$, and $\delta = 0.5$.

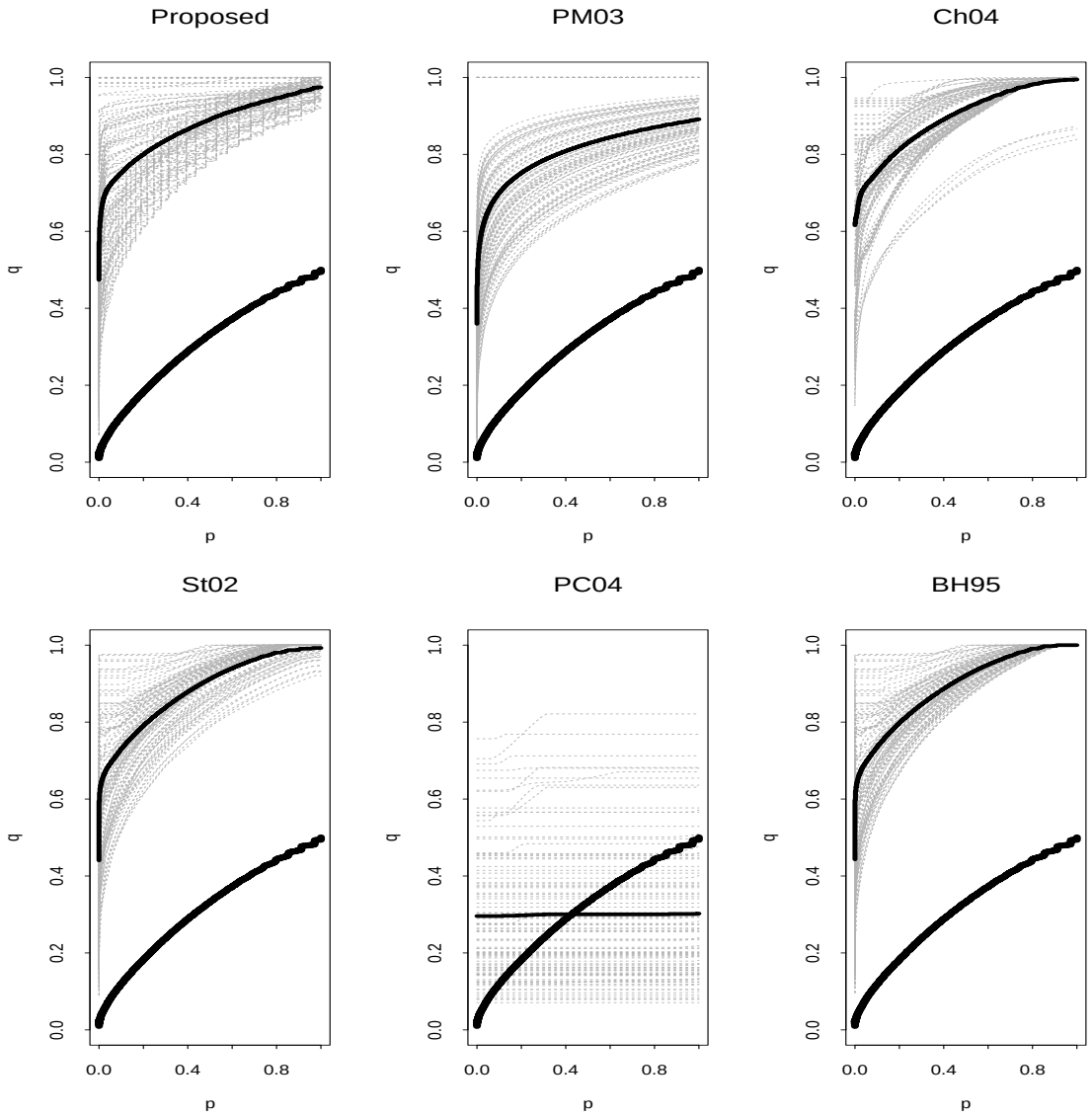

Figure S.27: Simulation Results with Two-Sided Tests, $n = 10$, $\pi = 0.5$, and $\delta = 0.5$.

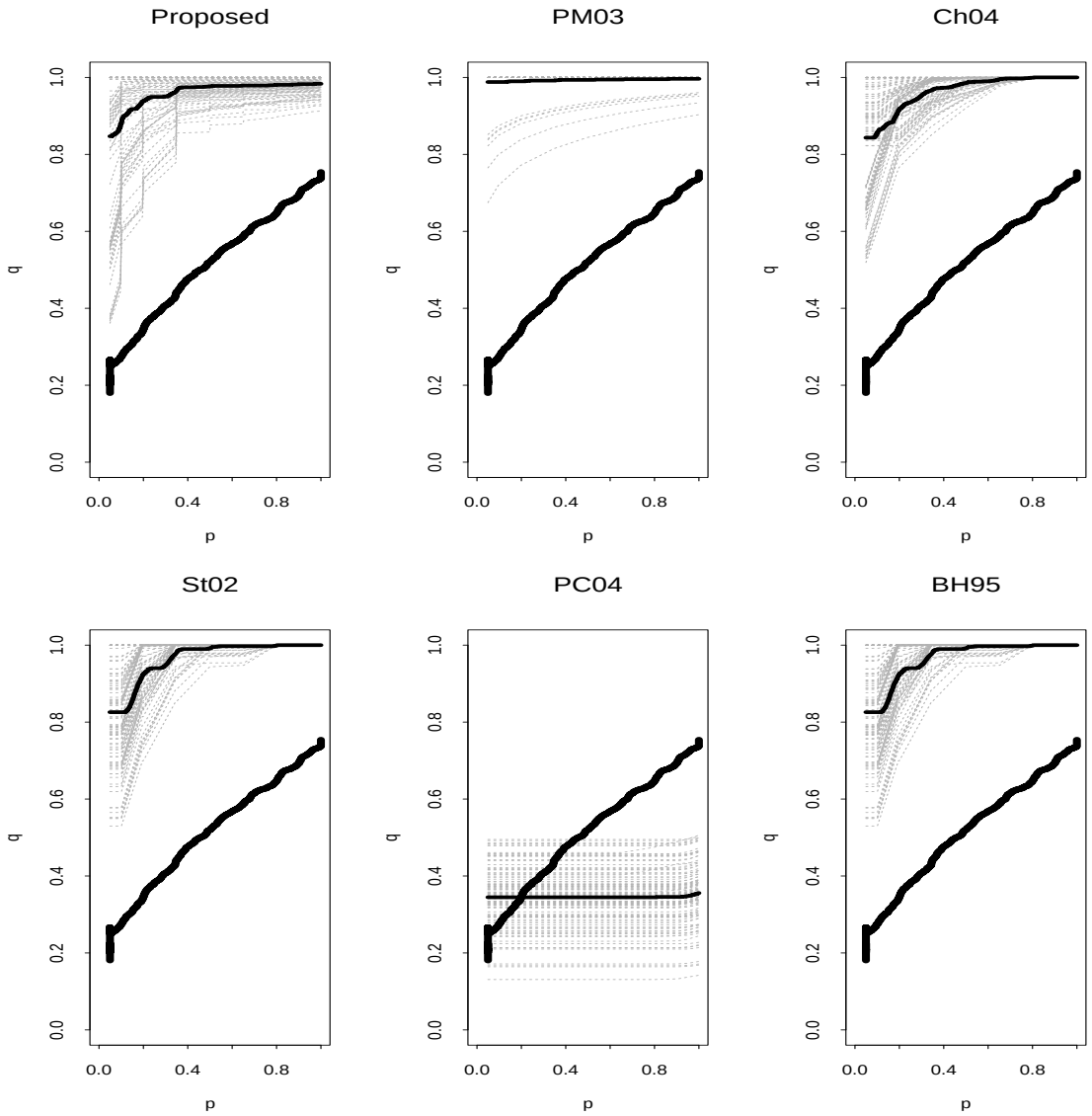

Figure S.28: Simulation Results with One-Sided Tests, $n = 3$, $\pi = 0.5$, and $\delta = 1.0$.

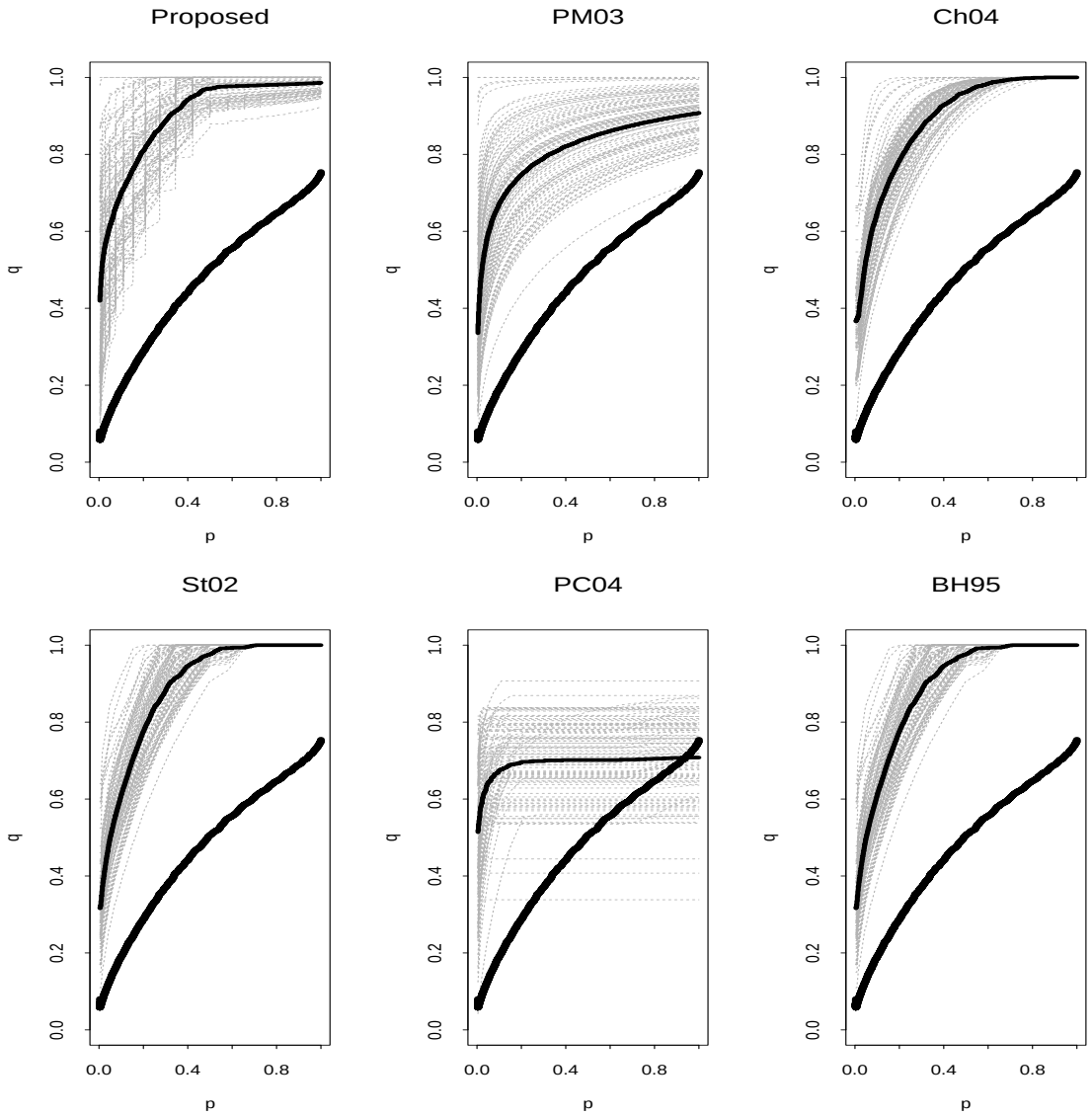

Figure S.29: Simulation Results with One-Sided Tests, $n = 5$, $\pi = 0.5$, and $\delta = 1.0$.

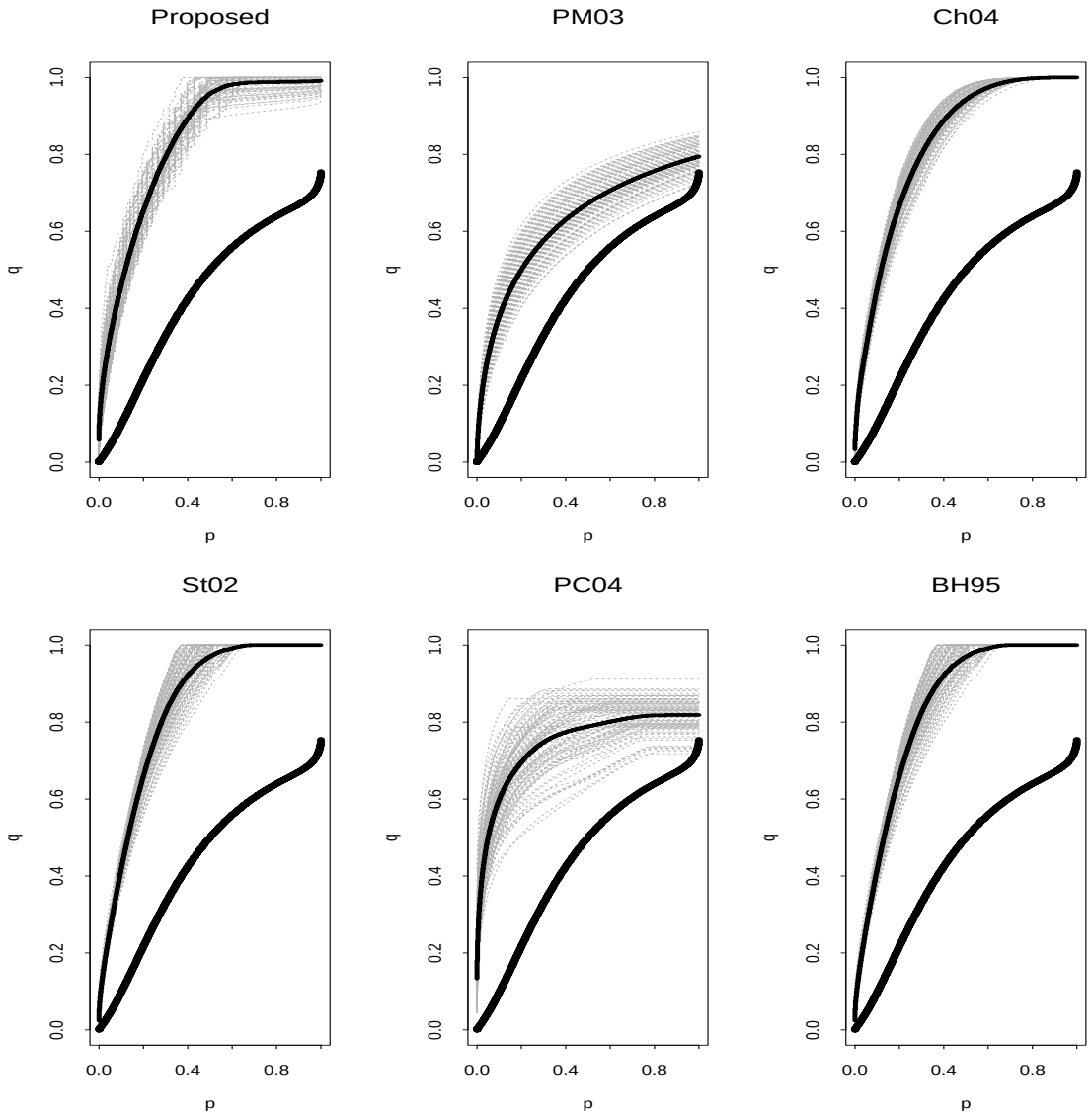

Figure S.30: Simulation Results with One-Sided Tests, $n = 10$, $\pi = 0.5$, and $\delta = 1.0$.

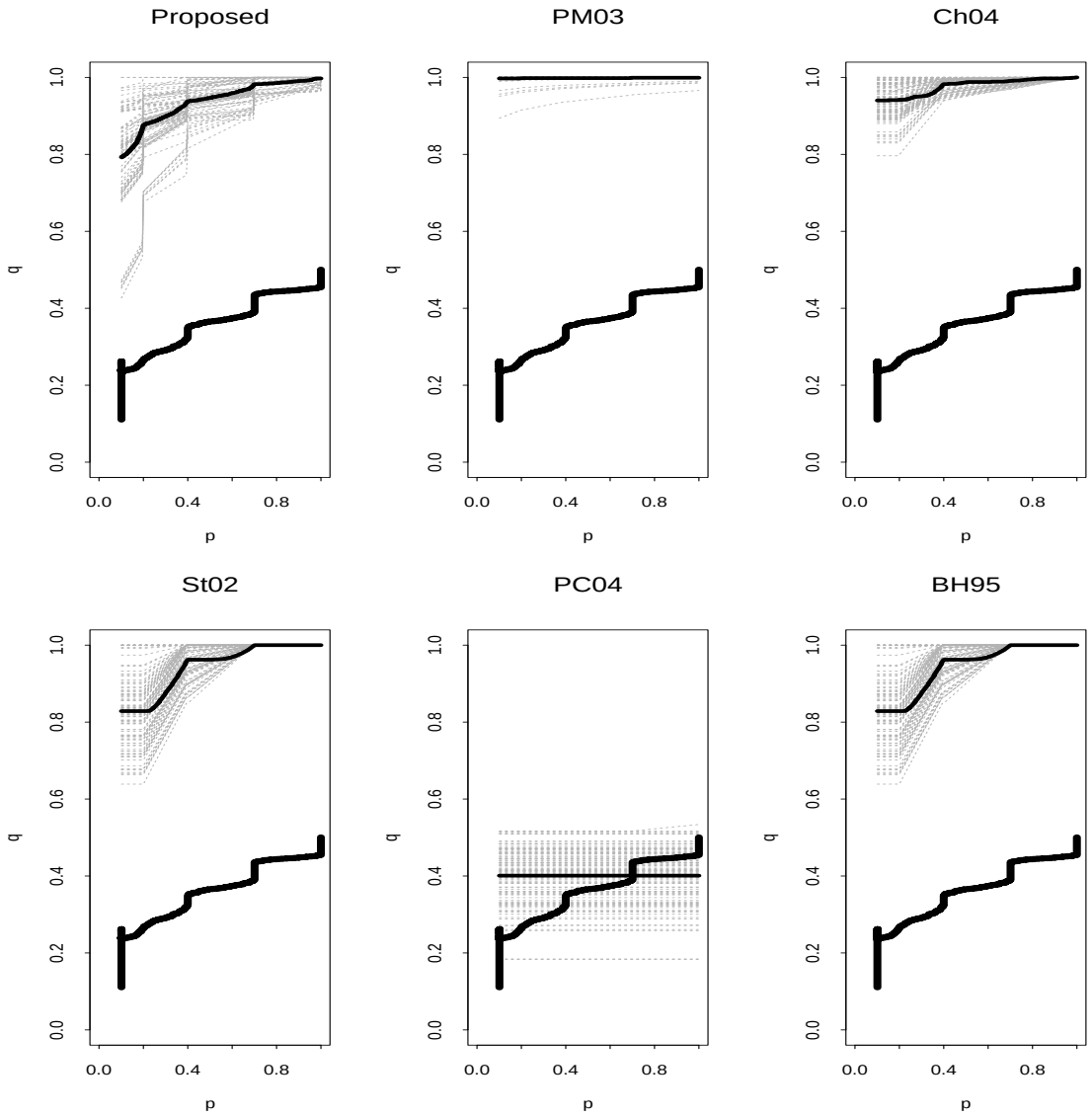

Figure S.31: Simulation Results with Two-Sided Tests, $n = 3$, $\pi = 0.5$, and $\delta = 1.0$.

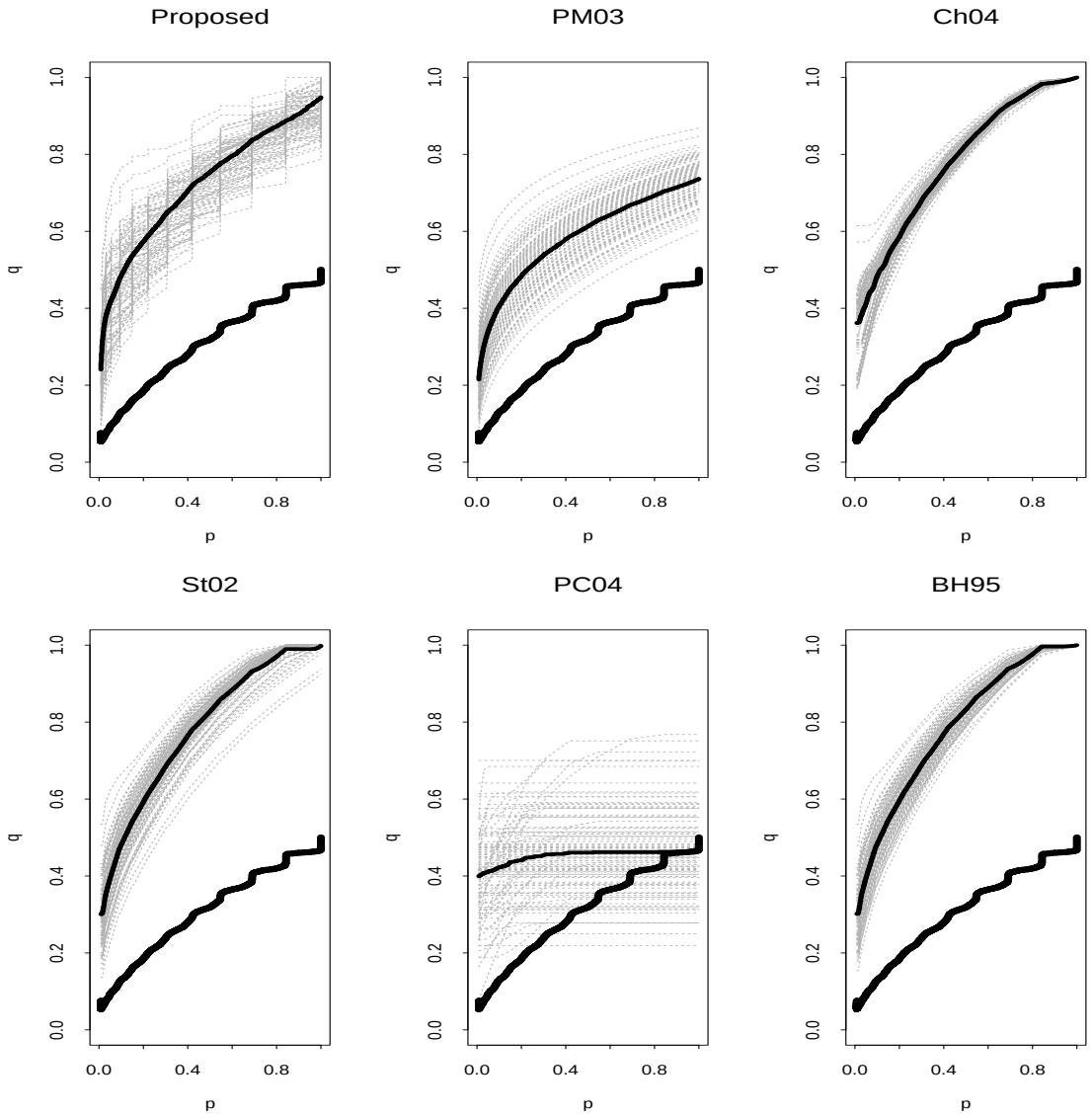

Figure S.32: Simulation Results with Two-Sided Tests, $n = 5$, $\pi = 0.5$, and $\delta = 1.0$.

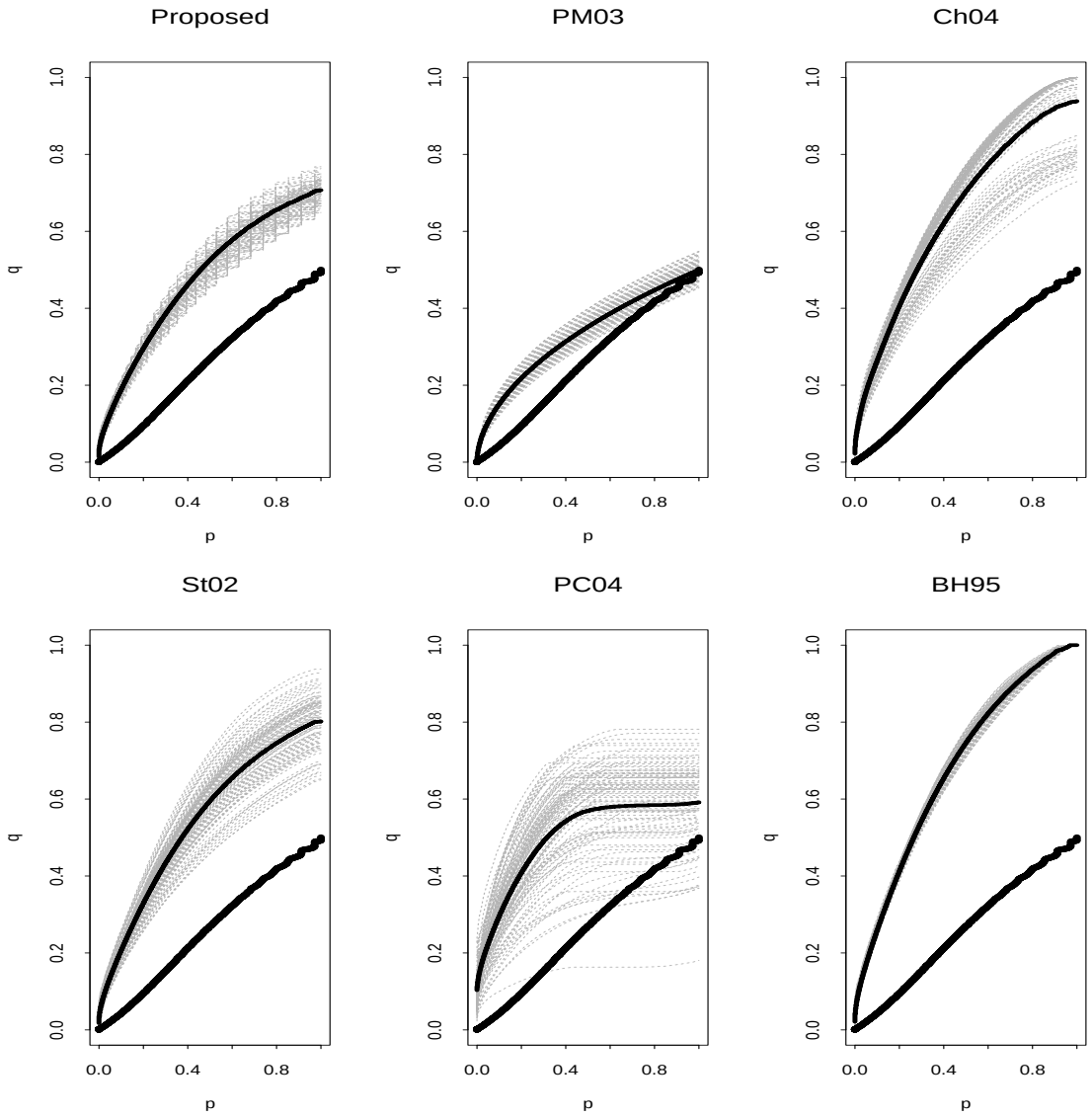

Figure S.33: Simulation Results with Two-Sided Tests, $n = 10$, $\pi = 0.5$, and $\delta = 1.0$.