Estimates of pi0

David Gerard

2016-05-24

Abstract

I plot estimates of π_0 versus π_0 for various methods under various alternative densities. RUVASH and SUCCOTASH perform the best.

Method

I used π_0 from a grid from 0.01 to 0.99, generated count data from the muscle GTEX data with $p = 1000$, where p is the number of genes. Signal was added via the Poisson thinning procedure with log2-fold-changes drawn from the same alternative distributions as in the ASH paper (see table below). The only difference is that the mixing variances were also divided by $n - 2$, where n is the sample size. This is a little different from my previous simulation settings and is meant to conform to the settings that Mengyin has been looking at.

Scenario	Alternative Distribution
Spiky	$0.4N(0, 0.25^2) + 0.2N(0, 0.5^2) + 0.2N(0, 1^2), 0.2N(0, 2^2)$
Near Normal	$2/3N(0, 1^2) + 1/3N(0, 2^2)$
Flattop	$(1/7)N(-1.5, .5^2) + N(-1, .5^2) + N(-.5, .5^2) + N(0, .5^2) + N(0.5, .5^2) + N(1.0, .5^2) + N(1.5, .5^2)$
Skew	$(1/4)N(-2, 2^2) + (1/4)N(-1, 1.5^2) + (1/3)N(0, 1^2) + (1/6)N(1, 1^2)$
Big-normal	$N(0, 4^2)$
Bimodal	$0.5N(-2, 1^2) + 0.5N(2, 1^2)$

I ran 10 methods, keeping their estimates of π_0:

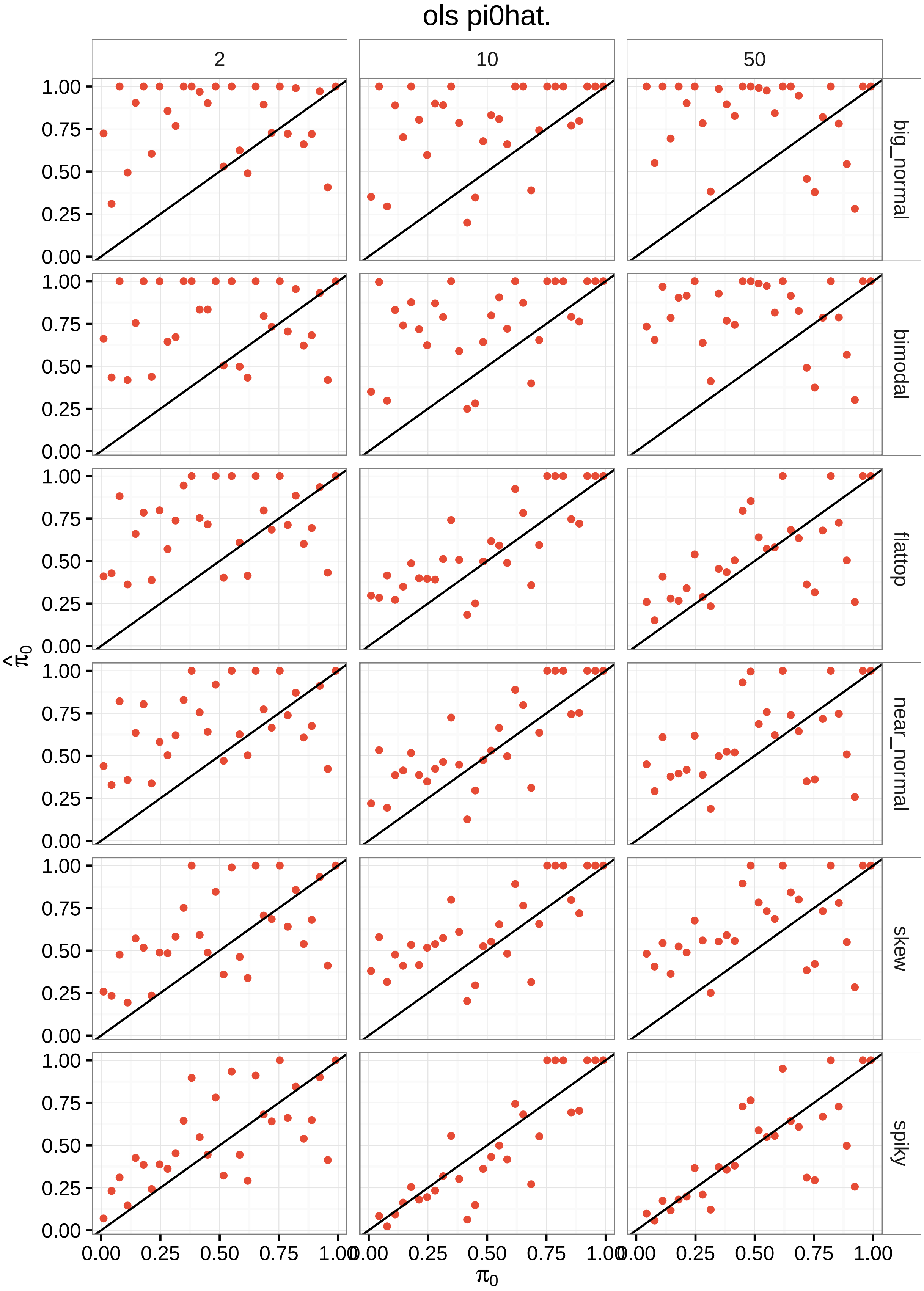
- OLS + qvalue.
- OLS + ASH
- Two-step SUCCOTASH using normal mixtures and heteroscedastic PCA as the factor-analysis method.
- The robust regression version of CATE using PCA as the factor analysis method + qvalue.
- SVA + qvalue.
- Scaled version of RUVASH.
- Negative control version of CATE using PCA as the factor analysis method + qvalue.
- RUV2 + qvalue.
- RUV4 + qvalue.
- Ridge version of LEAPP + qvalue.

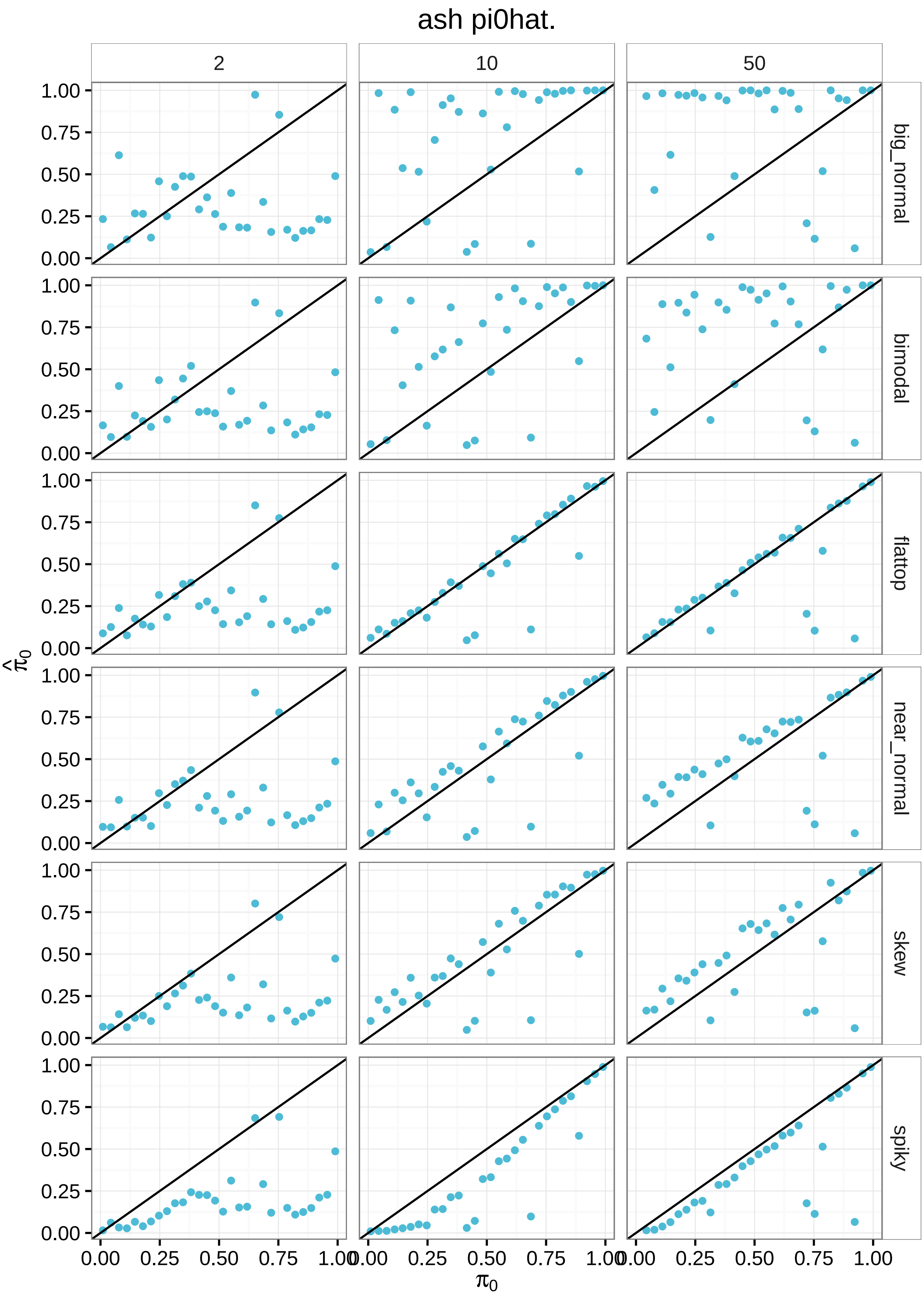
Plots of π_0 vs $\hat{\pi}_0$ for each method in each combination of n by alternative type are plotted below.

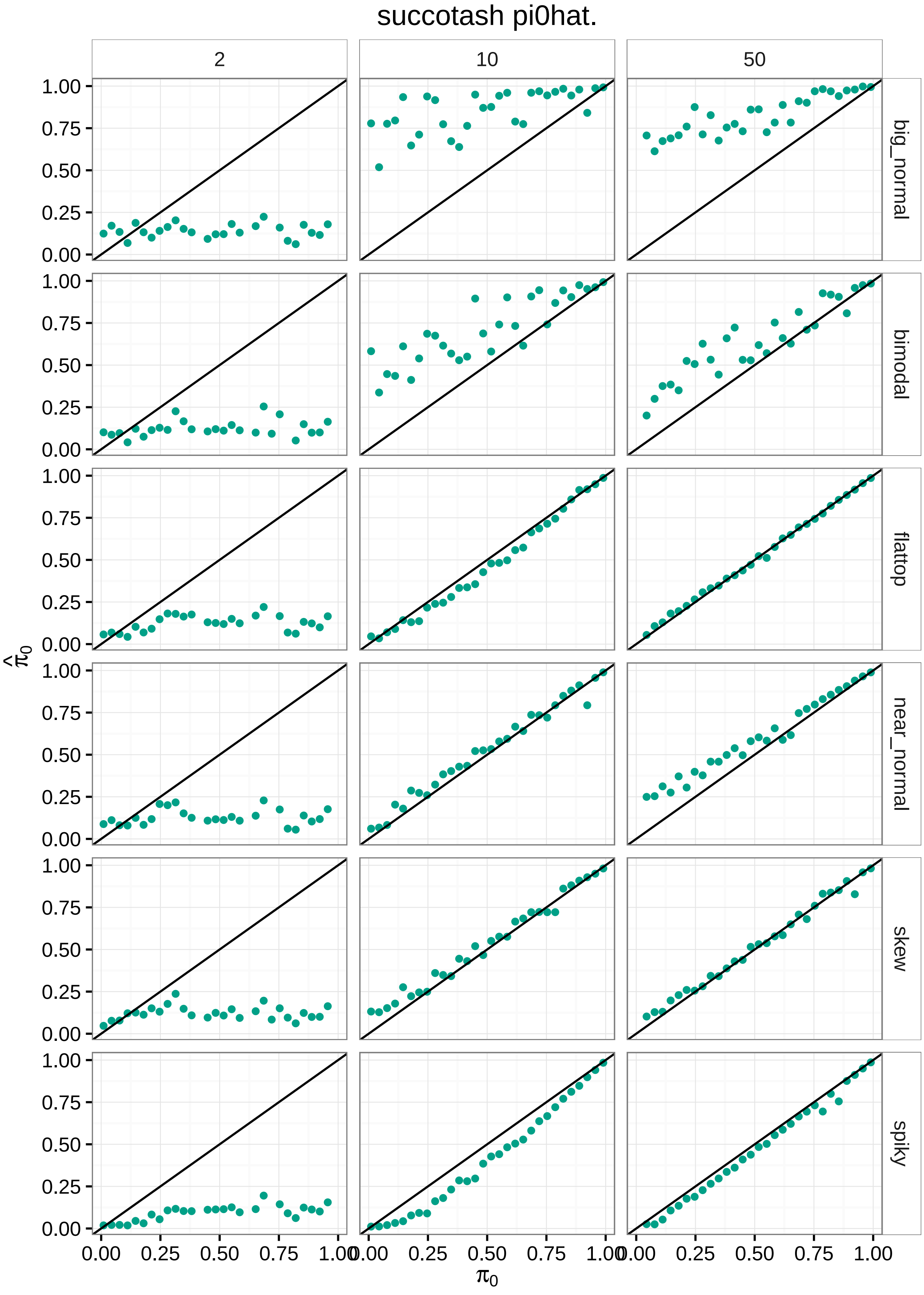
Results

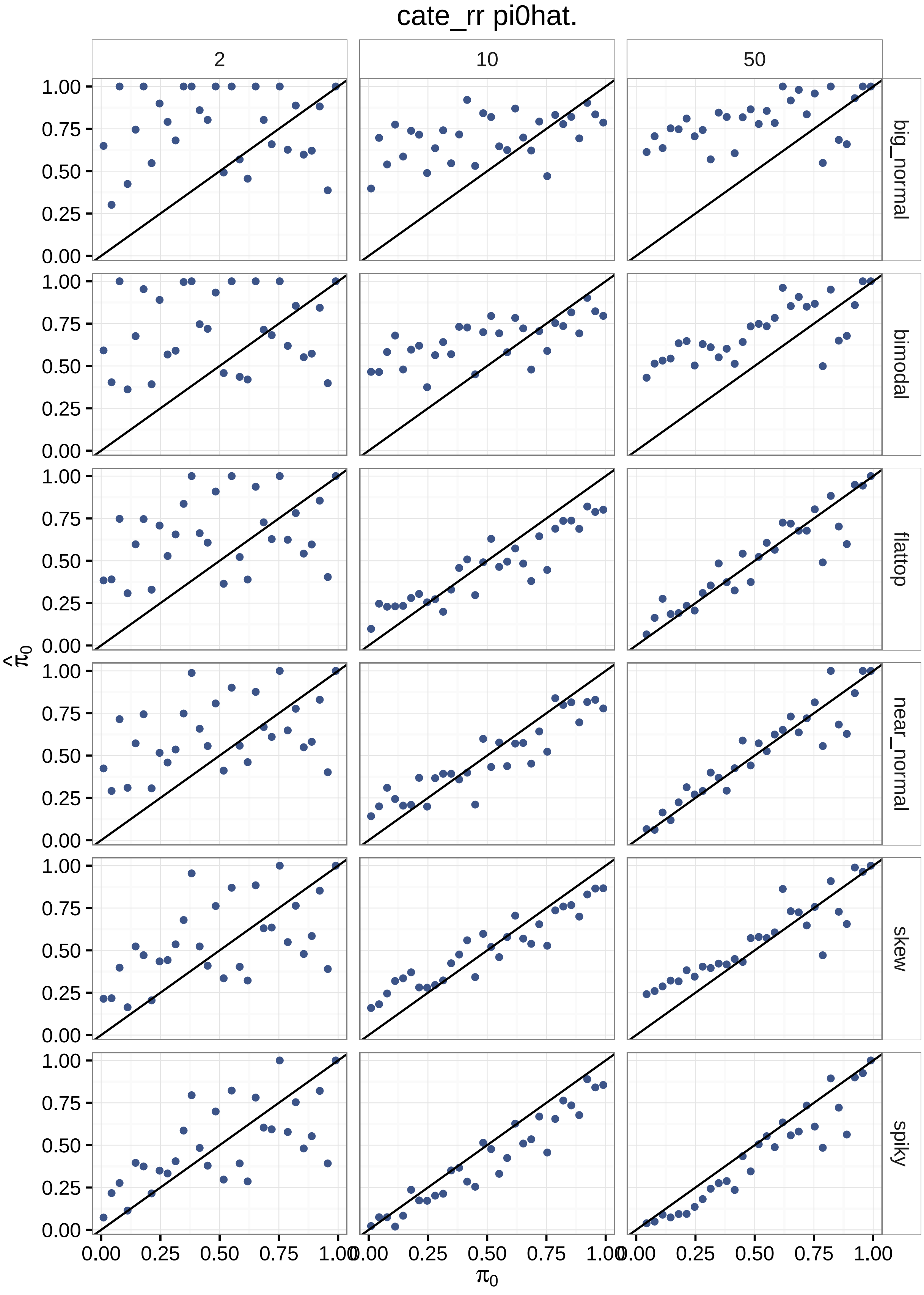
SUCCOTASH and RUVASH seem to perform the best, but all of the methods seem to be getting better at estimating π_0 as n increases.

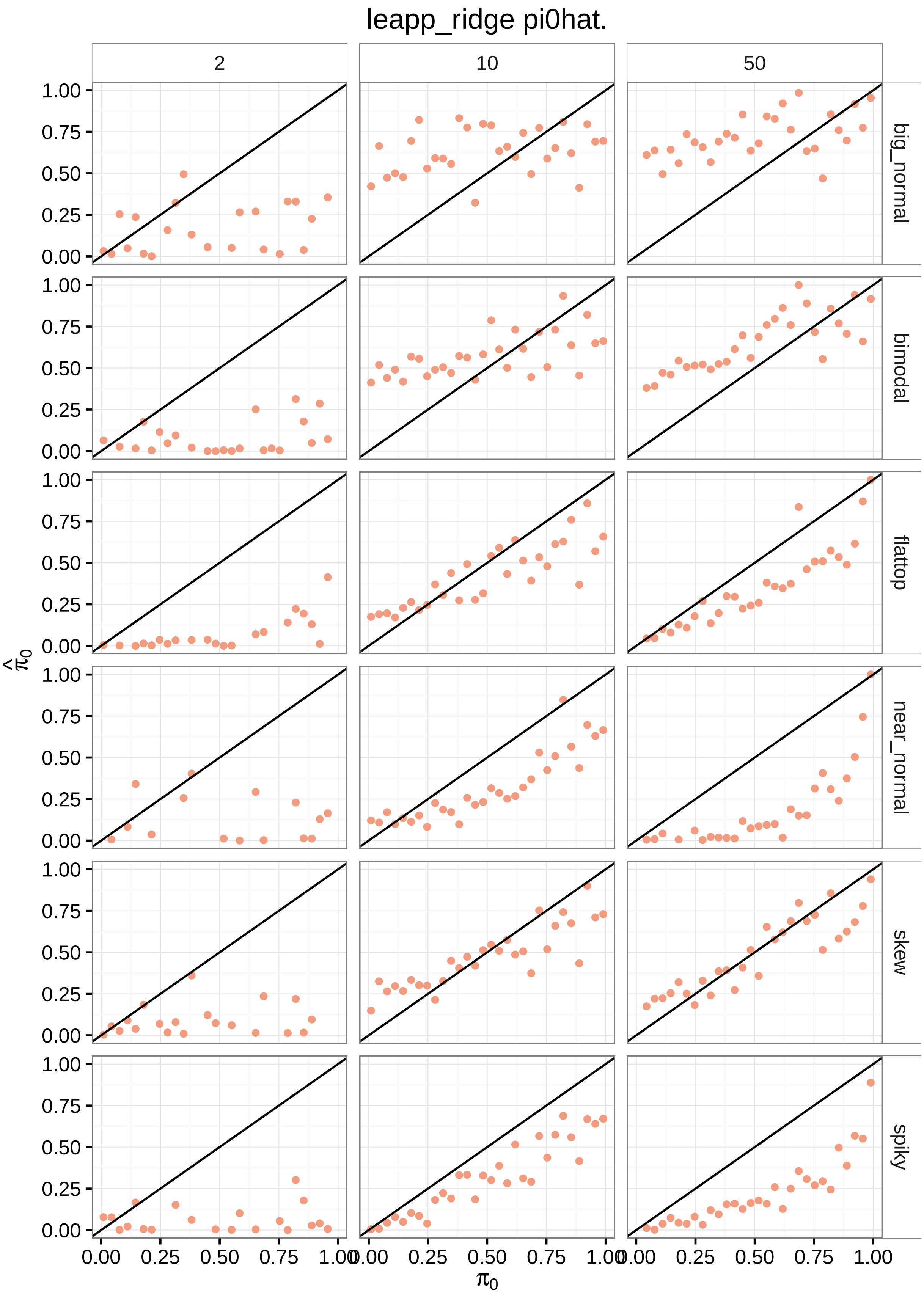
In this case where the mixing variances decrease with n, SUCCOTASH and RUVASH actually perform the worst in the spiky scenario rather than the bimodal scenario. But they perform better than the other methods in the spiky scenario.

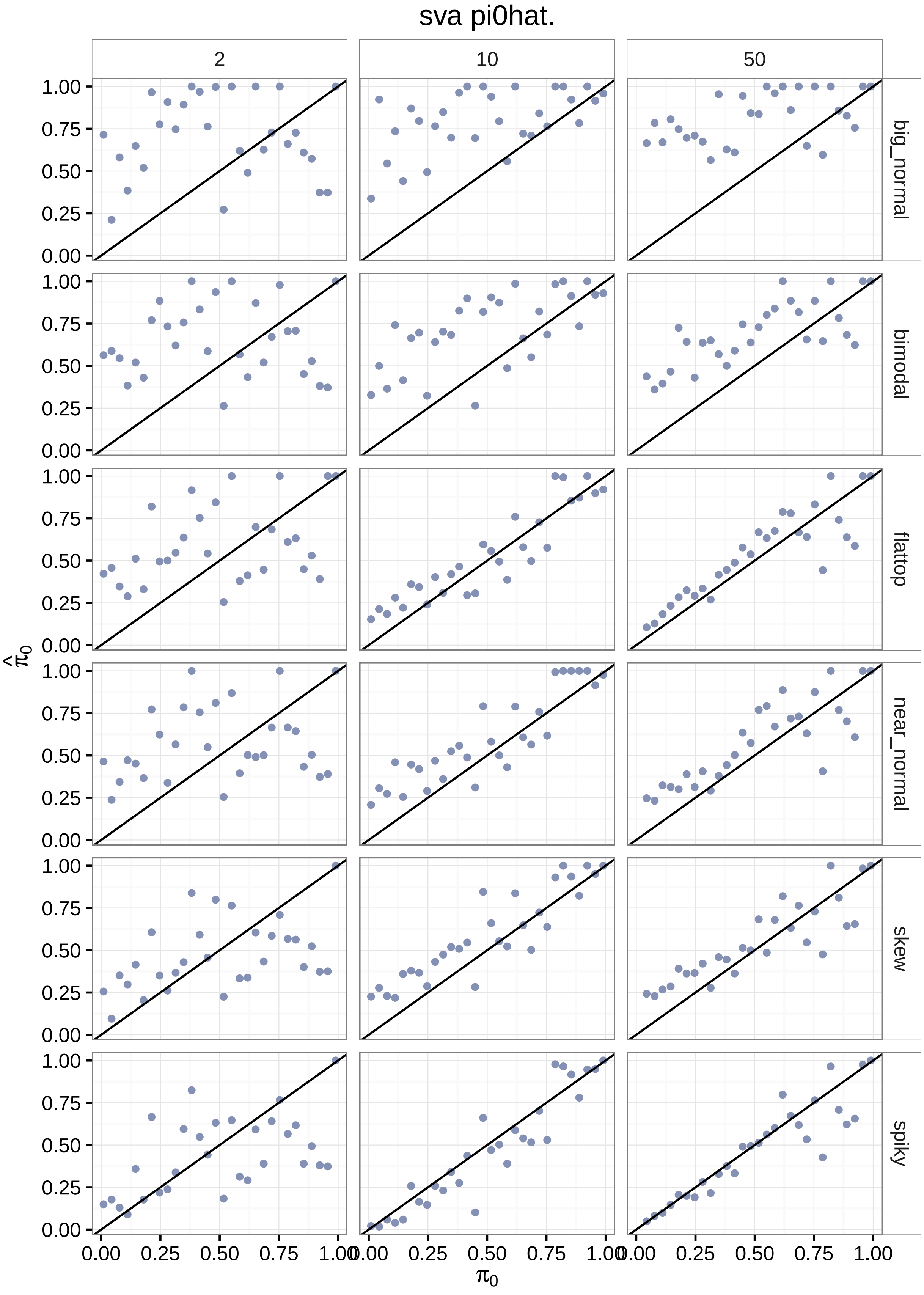

ols pi0hat.
2
10
50
big_normal
bimodal
flattop
near_normal
skew
spiky
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
$\hat{\pi}_0$
π_0

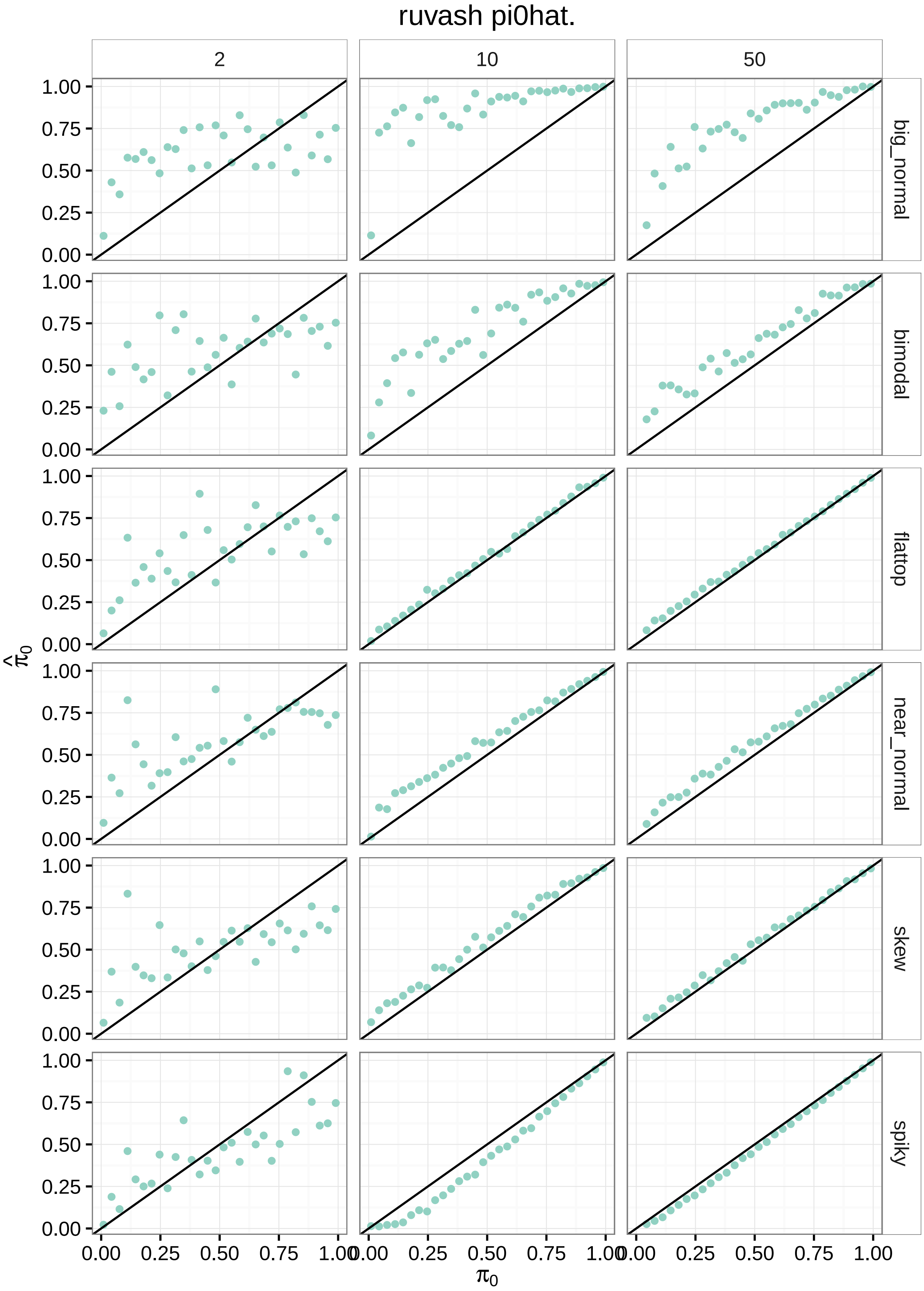

ash pi0hat.
2
10
50
big_normal
bimodal
flattop
near_normal
skew
spiky
0.00
0.25
0.50
0.75
1.00
0.00
0.25
0.50
0.75
1.00
$\hat{\pi}_0$
π_0

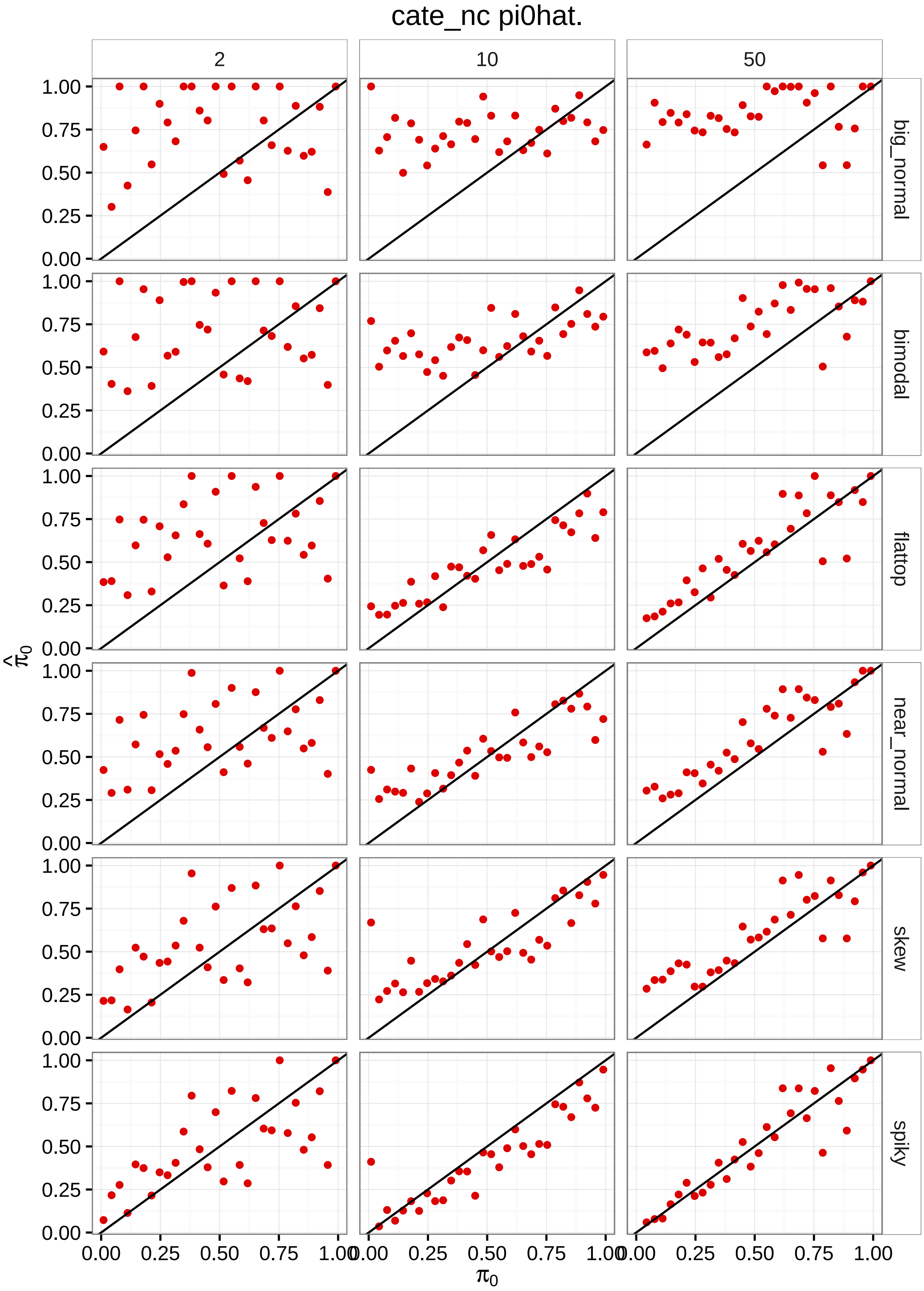

succotash pi0hat.
2
10
50
big_normal
bimodal
flattop
near_normal
skew
spiky
1.00
0.75
0.50
0.25
0.00
$\hat{\pi}_0$
π_0
0.00
0.25
0.50
0.75
1.00

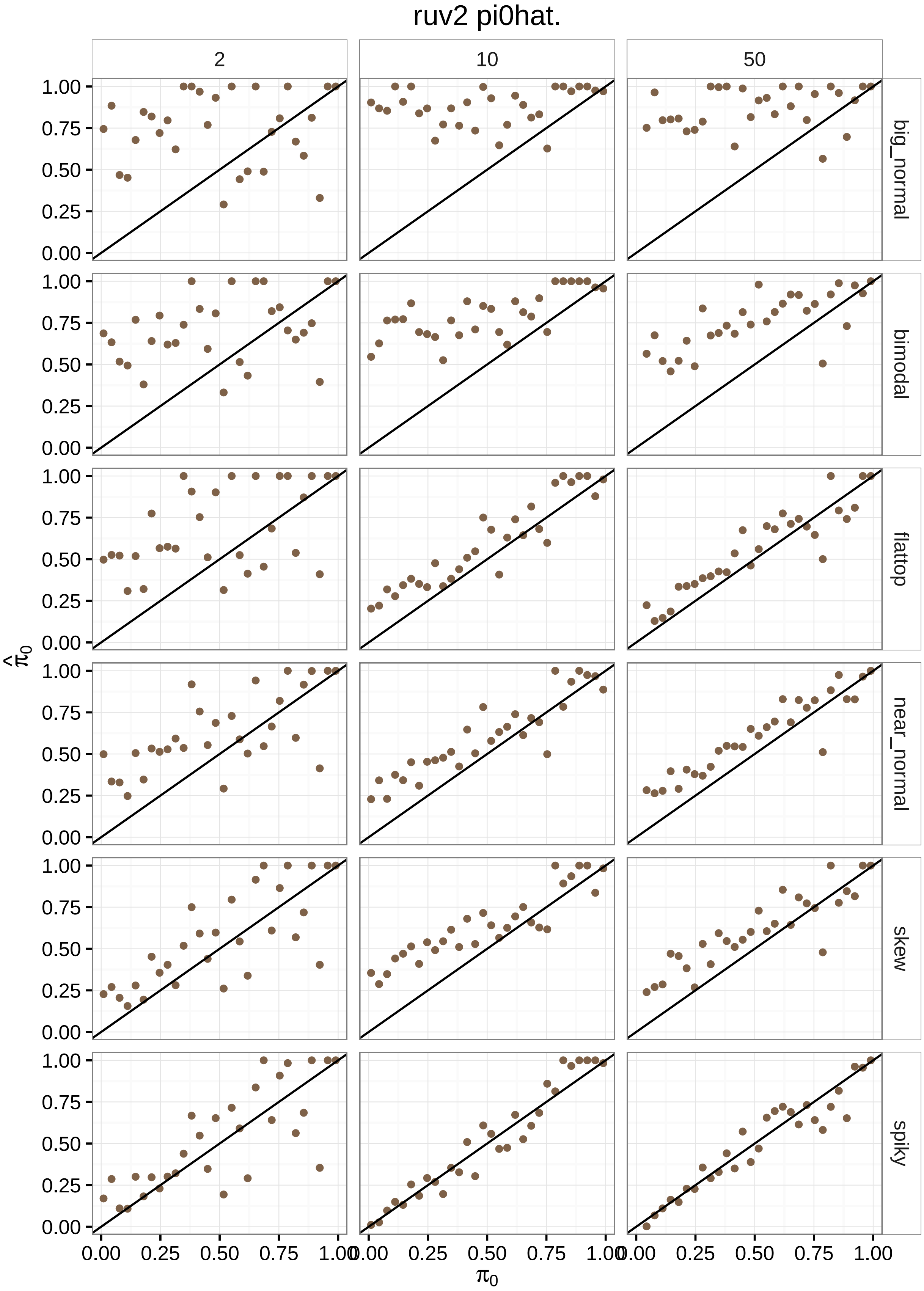

cate_rr pi0hat.
2
10
50
big_normal
bimodal
flattop
near_normal
skew
spiky
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
$\hat{\pi}_0$
π_0

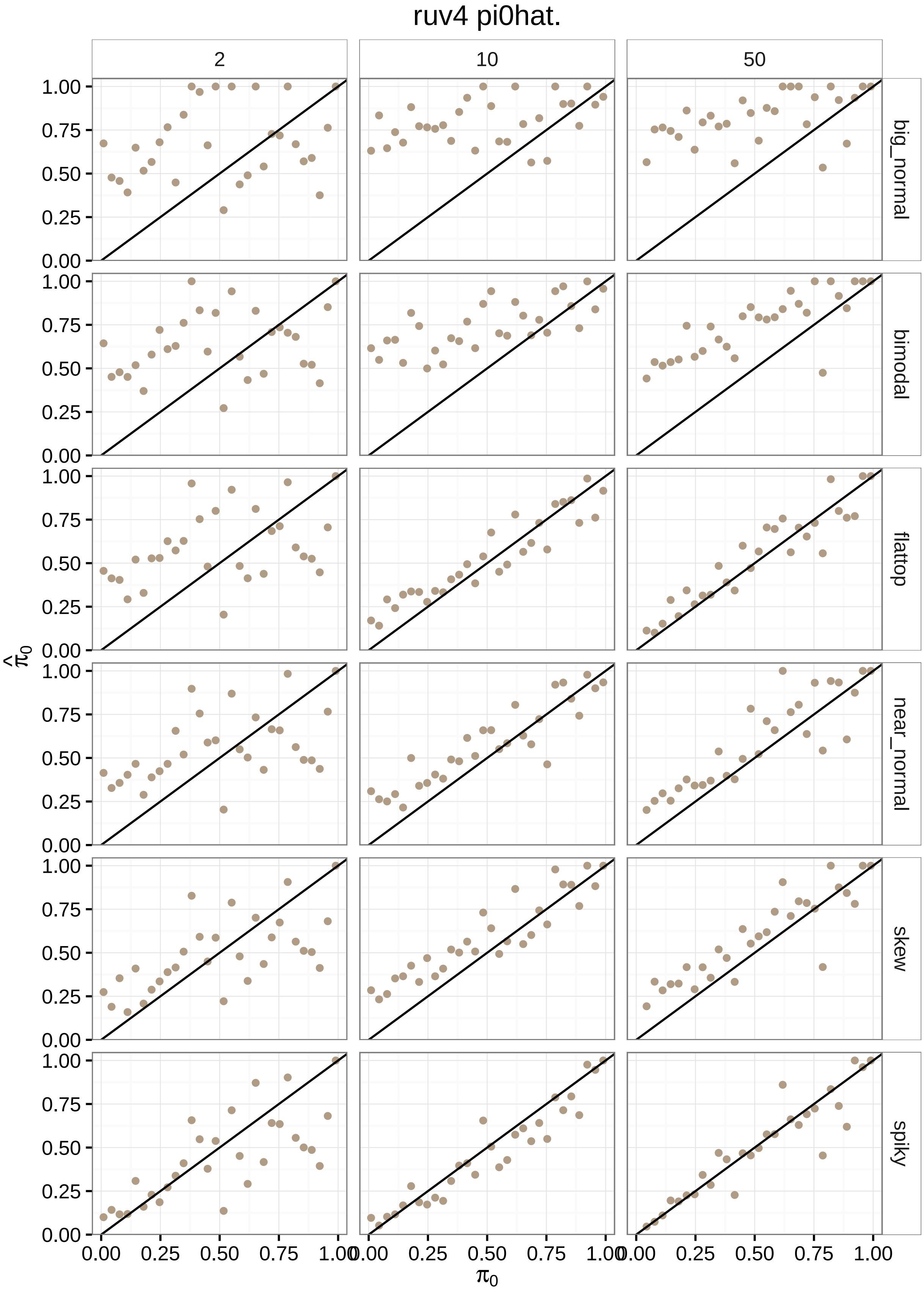

leapp_ridge pi0hat.
2
10
50
big_normal
bimodal
flattop
near_normal
skew
spiky
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
$\hat{\pi}_0$
π_0


sva pi0hat.
2
10
50
big_normal
bimodal
flattop
near_normal
skew
spiky
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
$\hat{\pi}_0$
π_0


ruvash pi0hat.
2
10
50
big_normal
bimodal
flattop
near_normal
skew
spiky
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
$\hat{\pi}_0$
π_0


cate_nc pi0hat.
2
10
50
big_normal
bimodal
flattop
near_normal
skew
spiky
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
$\hat{\pi}_0$
π_0


ruv2 pi0hat.
2
10
50
big_normal
bimodal
flattop
near_normal
skew
spiky
$\hat{\pi}_0$
π_0
0.00
0.25
0.50
0.75
1.00


ruv4 pi0hat.
2
10
50
big_normal
bimodal
flattop
near_normal
skew
spiky
$\hat{\pi}_0$
π_0
0.00
0.25
0.50
0.75
1.00

```
sessionInfo()
```

```
## R version 3.3.0 (2016-05-03)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 14.04.4 LTS
## 
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C              
##  [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8    
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8   
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C                 
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C            
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C       
## 
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base     
## 
## other attached packages:
## [1] ggsci_1.1      reshape2_1.4.1 ggplot2_2.1.0 
## 
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.5      codetools_0.2-14 digest_0.6.9     plyr_1.8.3      
##  [5] grid_3.3.0       gtable_0.2.0     formatR_1.3      magrittr_1.5    
##  [9] evaluate_0.9     scales_0.4.0     stringi_1.0-1    rmarkdown_0.9.6 
## [13] labeling_0.3     tools_3.3.0      stringr_1.0.0    munsell_0.4.3   
## [17] yaml_2.1.13      compiler_3.3.0   colorspace_1.2-6 htmltools_0.3.5 
## [21] knitr_1.12.28
```