SuperCENT Simulation

Junhui Cai, Dan Yang, Wu Zhu, Haipeng Shen, Linda Zhao

This file is produced by `SuperCENT_simulation.Rmd` which contains both codes to reproduce the simulation results in "Network Regression and Supervised Centrality Estimation" and the descriptions and instructions of the code chunks. To reproduce this report, `Knit` this file in RStudio. One can set `echo = T` for the each chunk or globally `knitr::opts_chunk$set(echo = T)` to show the code in the report.

1 Simulation

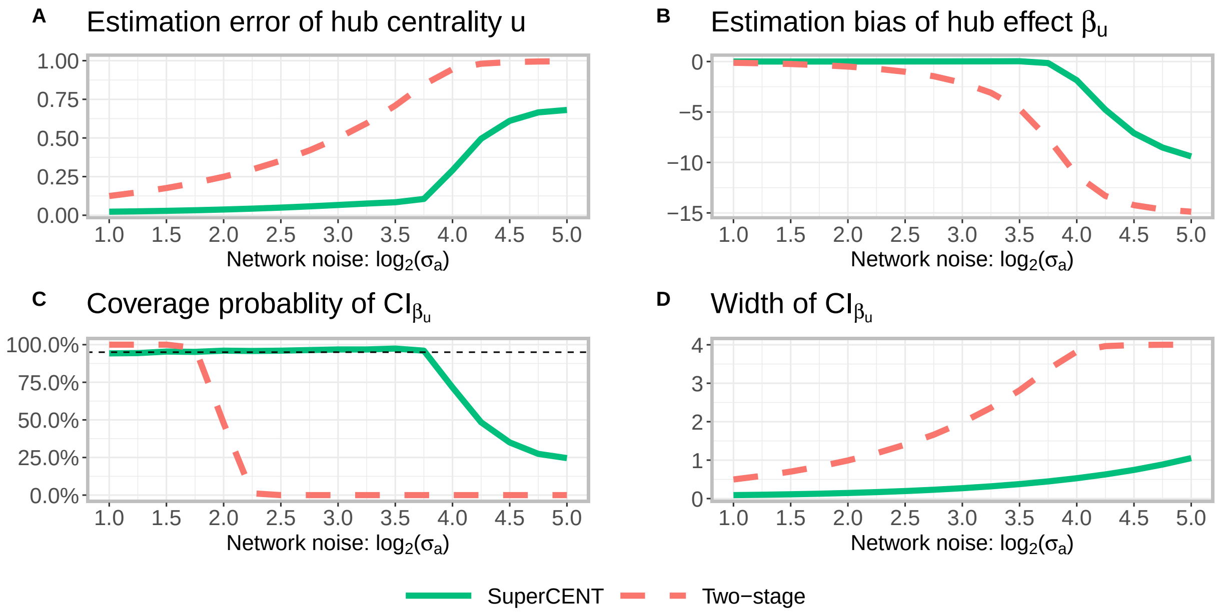
The simulations were run on the Sun Grid Engine (SGE) on Red Hat Enterprise Linux. The main run file is `confint.R` with utility functions in `utils_sim.R`. The setup of parameters $n, d, \beta_u, \beta_v, \beta_x$ therein is set using the `optparse` package and is specified in `option_list` with descriptions for each options.

To run different settings on SGE, specify the parameters grid in `run_confint.sh` and run `qsub run_confint.sh` in the `code` folder. It will write a summary log in `../output/<job-name>_<job-id>.log` and submit multiple jobs each with different setting with the results saved in `../hpcc_output/confint_<job-id>`. To combine the simulation results, run `Rscript reduce_cv.R <job-id>` where `job-id` is the job-id of the SGE. If one does not use SGE, run `Rscript confint.R --help` to print out a brief summary of the options and specify the settings accordingly. For example, `Rscript confint.R --betau 4` to set $\beta_u = 2^4$.

2 Toy experiment

2.1 Figure 1

The following chunk produces the toy experiment in the introduction.



3 Consistent regime of two-stage

This section is for the plots of the consistent regime of the two-stage procedure.

3.1 Calculation of theoretical results

We first calculate the theoretical rate of $\hat{u}$, $\hat{v}$ and $\hat{A}$ as well as $\hat{\beta}_u$ and $\hat{\beta}_v$.

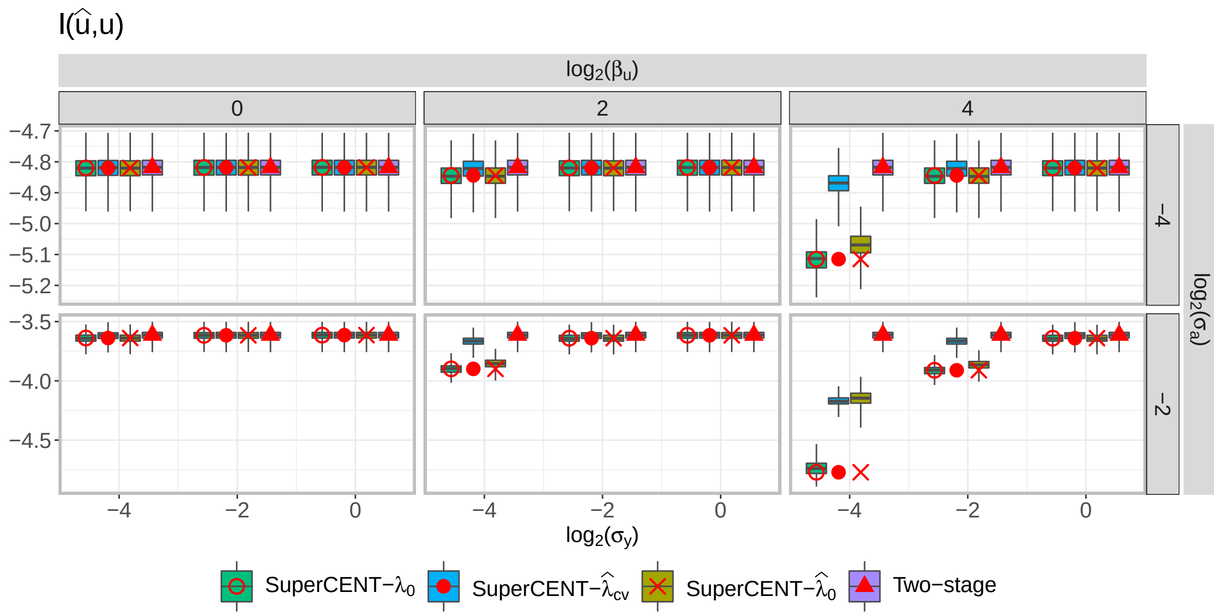
3.2 Estimation property

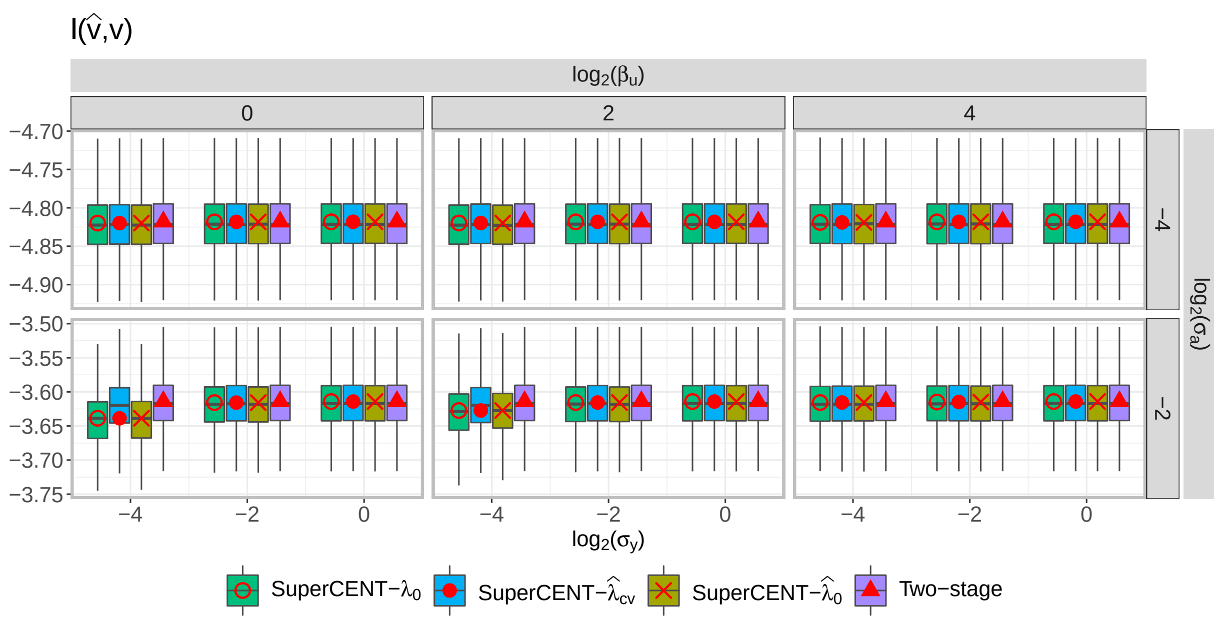
For the estimation accuracy, we compare the following procedures:

1. **Two-stage**: the two-stage procedure as in Algorithm1;
2. **SuperCENT-λ_0**: SuperCENT Algorithm 2 with the optimal $\lambda_0 = n\sigma_y^2/\sigma_a^2$, where the true σ_y, σ_a are used;
3. **SuperCENT-$\hat{\lambda}_0$**: SuperCENT with estimated tuning parameter $\hat{\lambda}_0 = n(\hat{\sigma}_y^{ts})^2/(\hat{\sigma}_a^{ts})^2$, where $(\hat{\sigma}_y^{ts})^2 = \frac{1}{n-p-2}\|\hat{y}^{ts} - y\|_2^2$ and $(\hat{\sigma}_a^{ts})^2 = \frac{1}{n^2}\|\hat{A}^{ts} - A_0\|_F^2$ are estimated from the two-stage procedure;
4. **SuperCENT-$\hat{\lambda}_{cv}$**: SuperCENT with tuning parameter $\hat{\lambda}_{cv}$ chosen by 10-fold cross-validation.

3.2.1 u and v

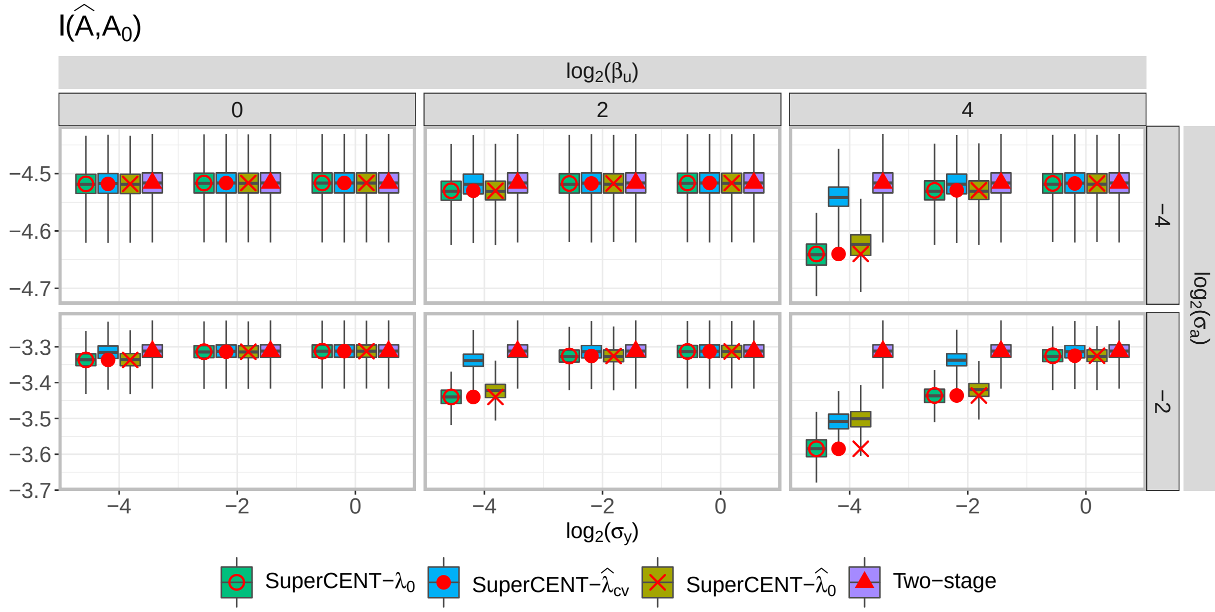
The following chunks produce the plots of $l(\widehat{u}, u)$ and $l(\widehat{v}, v)$.





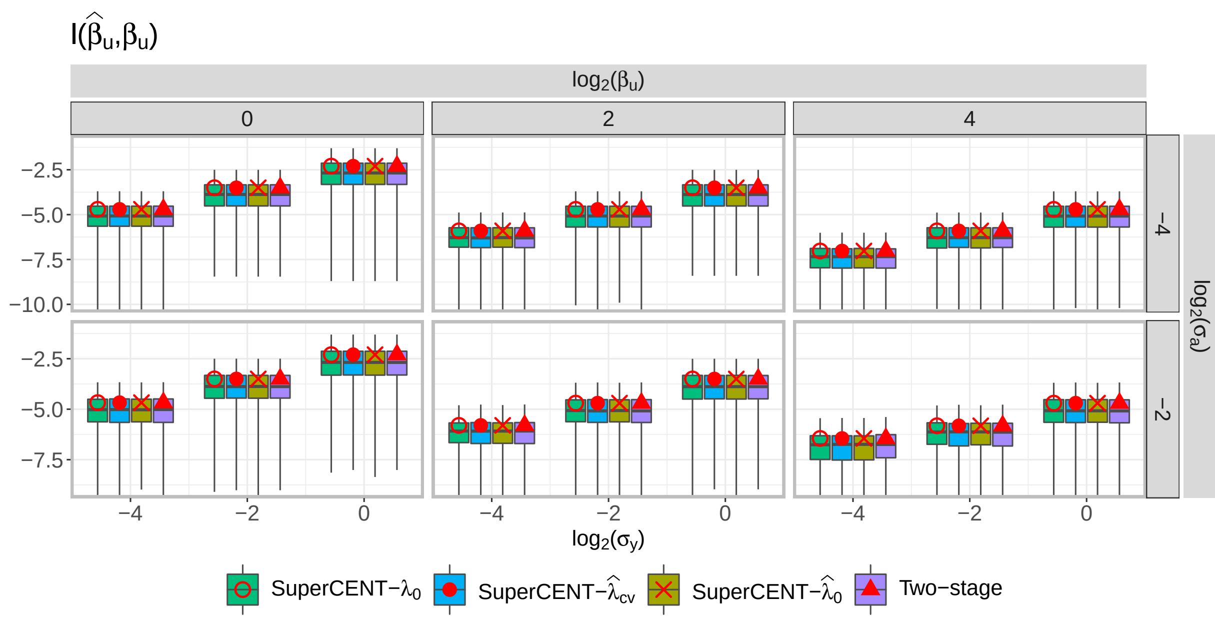
3.2.2 A

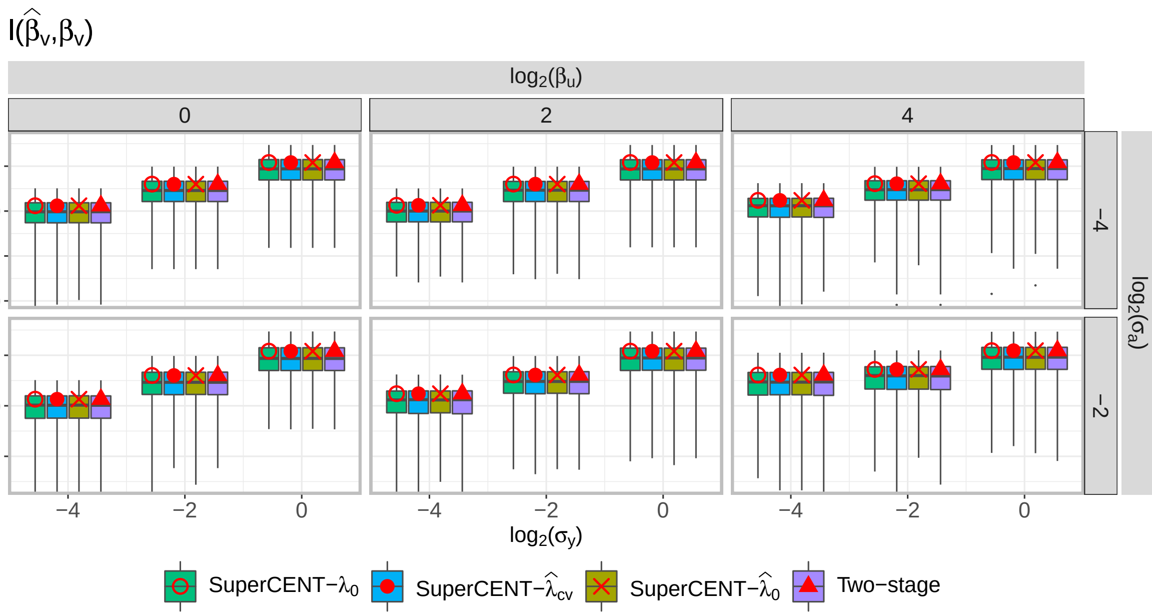
The following chunk produces the plot of $l(\widehat{A}, A_0)$.



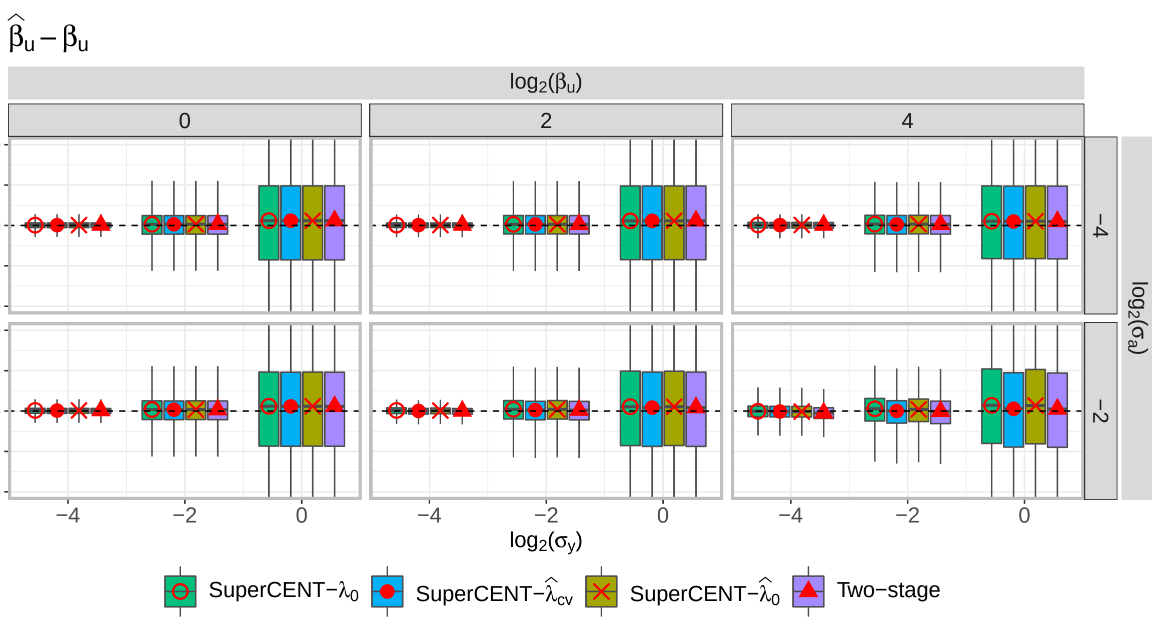
3.2.3 β_u and β_v

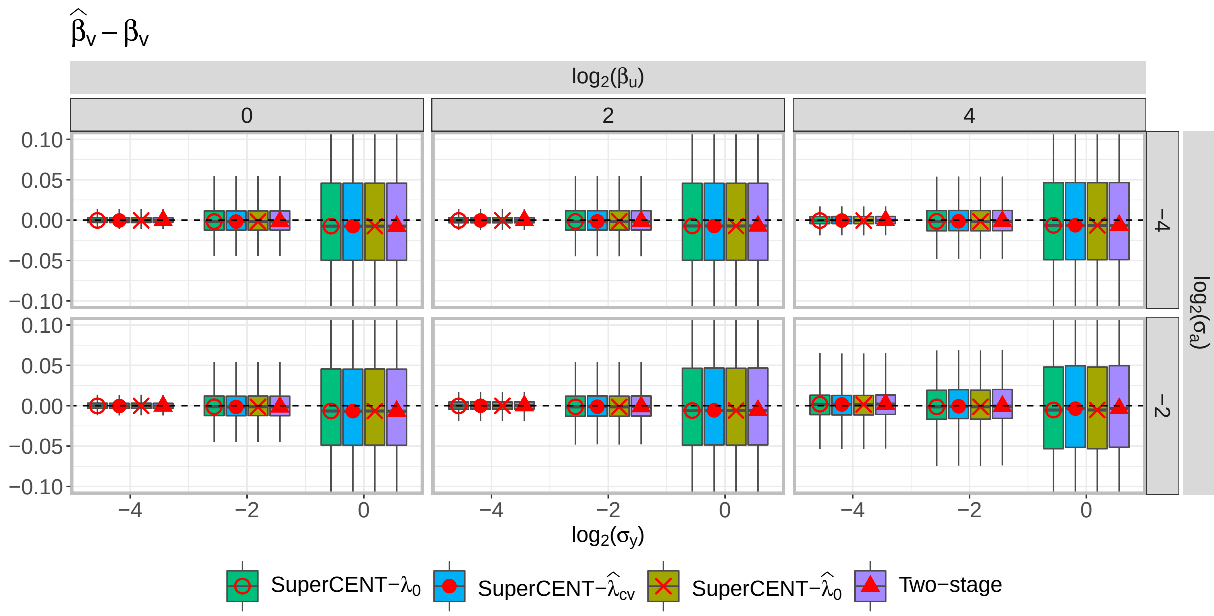
The following chunks produce the plots of $l(\widehat{\beta}_u, \beta_u)$ and $l(\widehat{\beta}_v, \beta_v)$.





The following chunks produce the plots of $\widehat{\beta}_u - \beta_u$ and $\widehat{\beta}_v - \beta_v$.





3.3 Inference property

For the inference property, let $z_{1-\alpha/2}$ denote the $(1-\alpha/2)$-quantile of the standard normal distribution and we consider the following procedures to construct the confidence intervals (CIs) for the regression coefficient, $\{CI_{\beta_u}$ and CI_{β_v}:}

1. **Two-stage-adhoc**: $\hat{\beta}^{ts} \pm z_{1-\alpha/2}\hat{\sigma}^{OLS}(\hat{\beta}^{ts})$, where $\hat{\beta}^{ts}$ is the two-stage estimate of β and $\hat{\sigma}^{OLS}(\hat{\beta}^{ts})$ is the standard error from OLS, assuming $\hat{u}^{ts}, \hat{v}^{ts}$ are fixed predictors;

2. **Two-stage-oracle**: $\hat{\beta}^{ts} \pm z_{1-\alpha/2}\sigma(\hat{\beta}^{ts})$, where $\sigma(\hat{\beta}^{ts})$ is the standard error of $\hat{\beta}^{ts}$, whose mathematical expressions are given in Corollary 2 and the true parameters are plugged into those expressions;

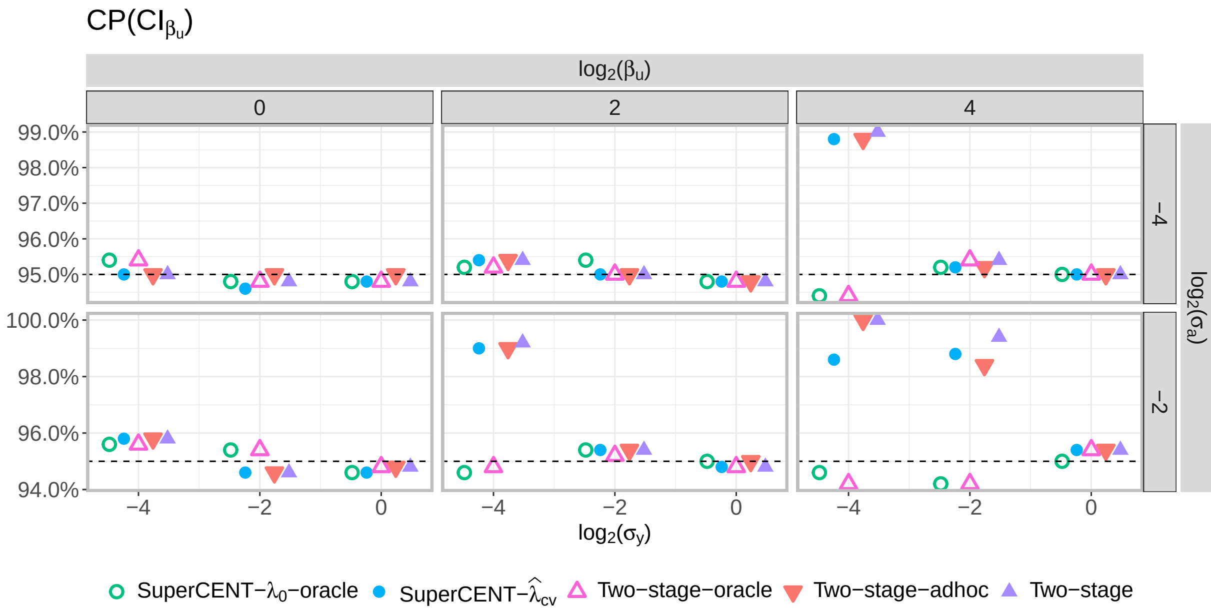
3. **Two-stage**: $\hat{\beta}^{ts} \pm z_{1-\alpha/2}\hat{\sigma}(\hat{\beta}^{ts})$, where $\hat{\sigma}(\hat{\beta}^{ts})$ is the standard error of $\hat{\beta}^{ts}$ by plugging all the two-stage estimators into Corollary 2.

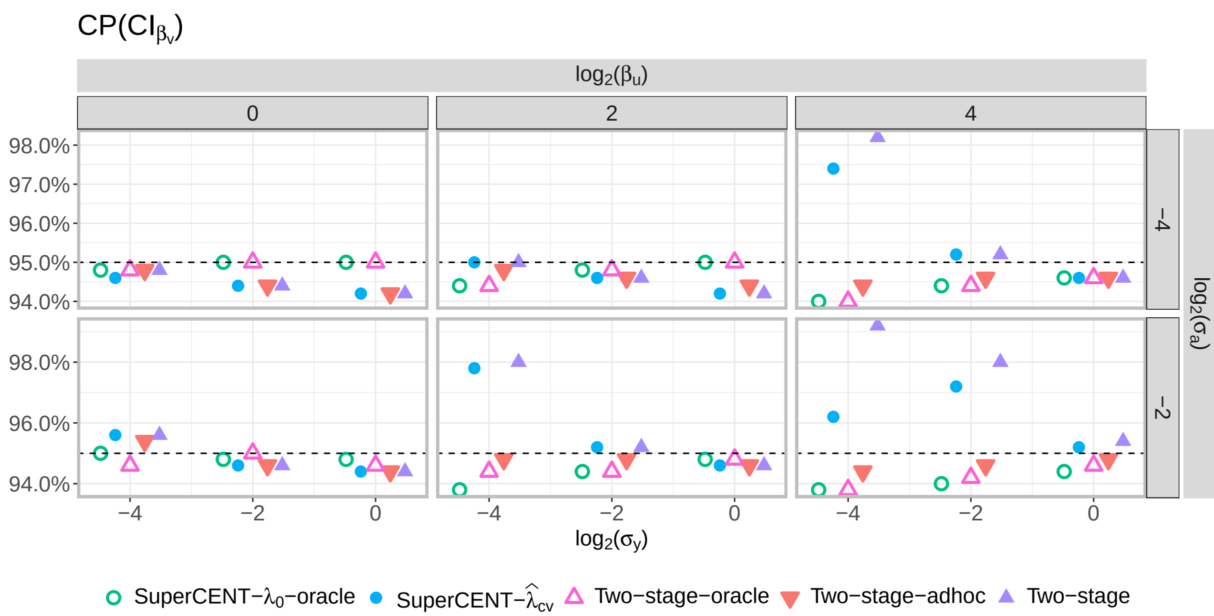
4. **SuperCENT-λ_0-oracle**; $\hat{\beta}^{\lambda_0} \pm z_{1-\alpha/2}\sigma(\hat{\beta}^{\lambda_0})$, where $\hat{\beta}^{\lambda_0}$ is the estimate of β by SuperCENT-λ_0 and $\sigma(\hat{\beta}^{\lambda_0})$ follows Corollary 5, with the true parameters plugged in;

5. **SuperCENT-$\hat{\lambda}_{cv}$**: $\hat{\beta}^{\hat{\lambda}_{cv}} \pm z_{1-\alpha/2}\hat{\sigma}(\hat{\beta}^{\hat{\lambda}_{cv}})$, where $\hat{\beta}^{\hat{\lambda}_{cv}}$ is the estimate of β by SuperCENT-$\hat{\lambda}_{cv}$ and $\hat{\sigma}(\hat{\beta}^{\hat{\lambda}_{cv}})$ is obtained by plugging the SuperCENT-$\hat{\lambda}_{cv}$ estimates into Corollary 5.

3.3.1 Coverage of CI_{β_u} and CI_{β_v}

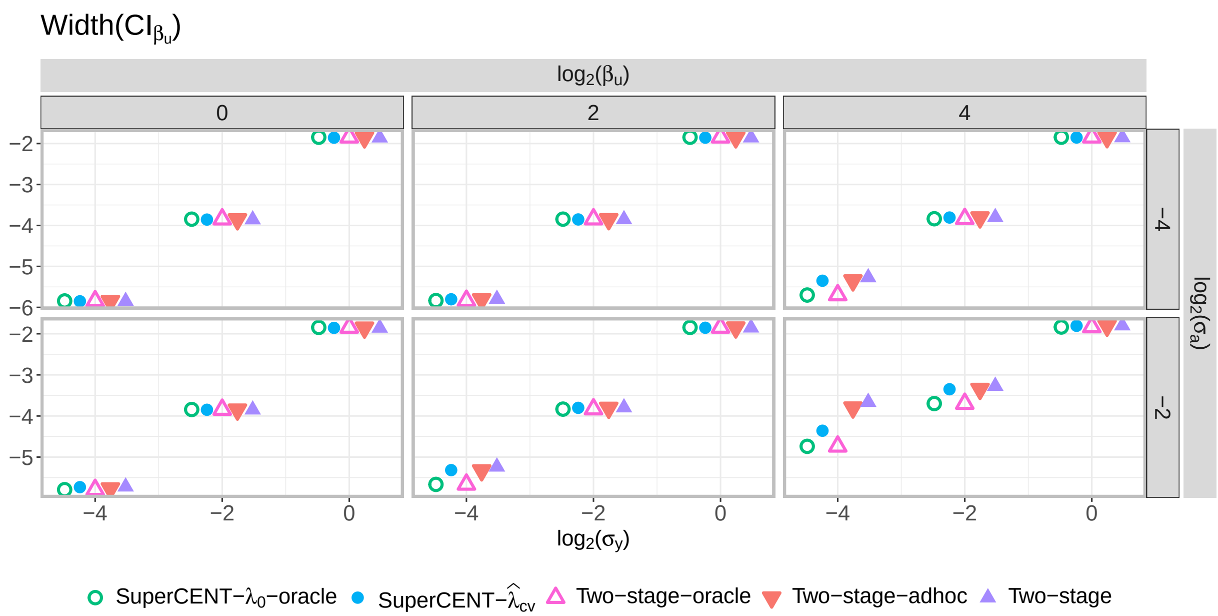
The following chunks produce the plots of the empirical coverage of the 95% confidence interval for β_u and β_v respectively, i.e., $CP(CI_{\beta_u})$ and $CP(CI_{\beta_v})$.

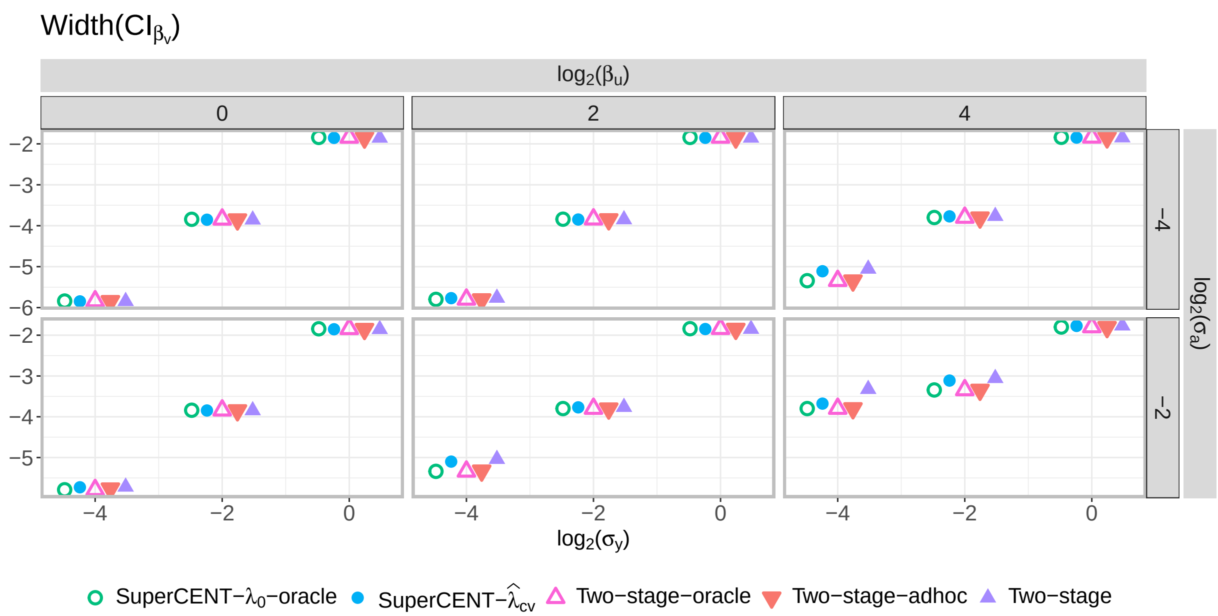




3.3.2 Width of CI_{β_u} and CI_{β_v}

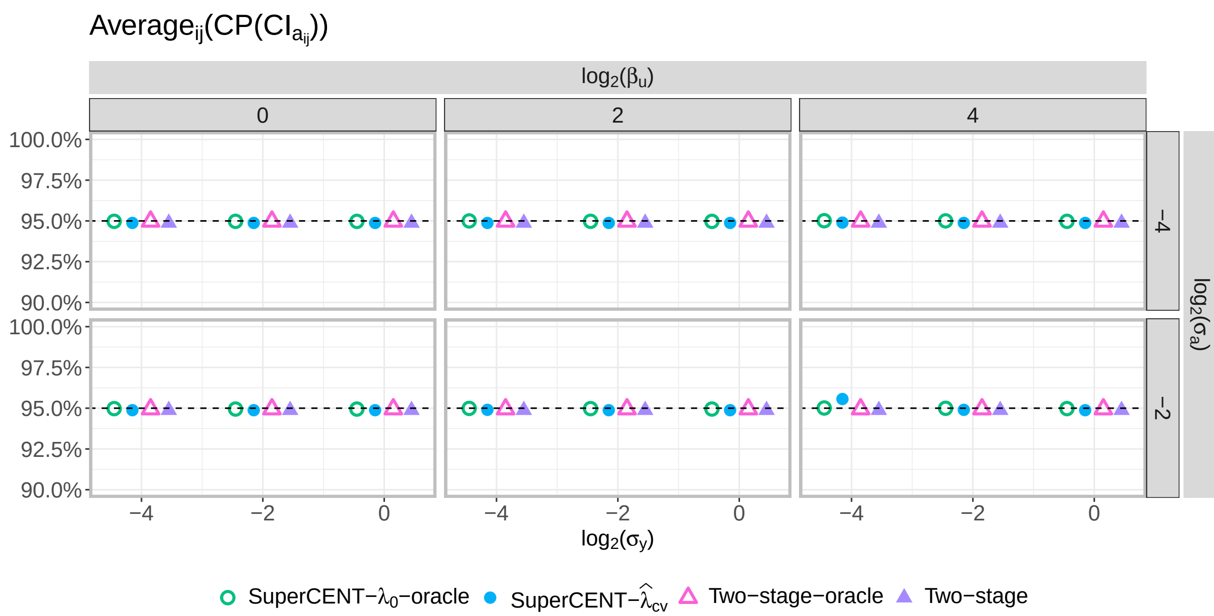
The following chunks produce the plots of the average of width of the 95% confidence interval for β_u and β_v respectively, i.e., $Width(CI_{\beta_u})$ and $Width(CI_{\beta_v})$.





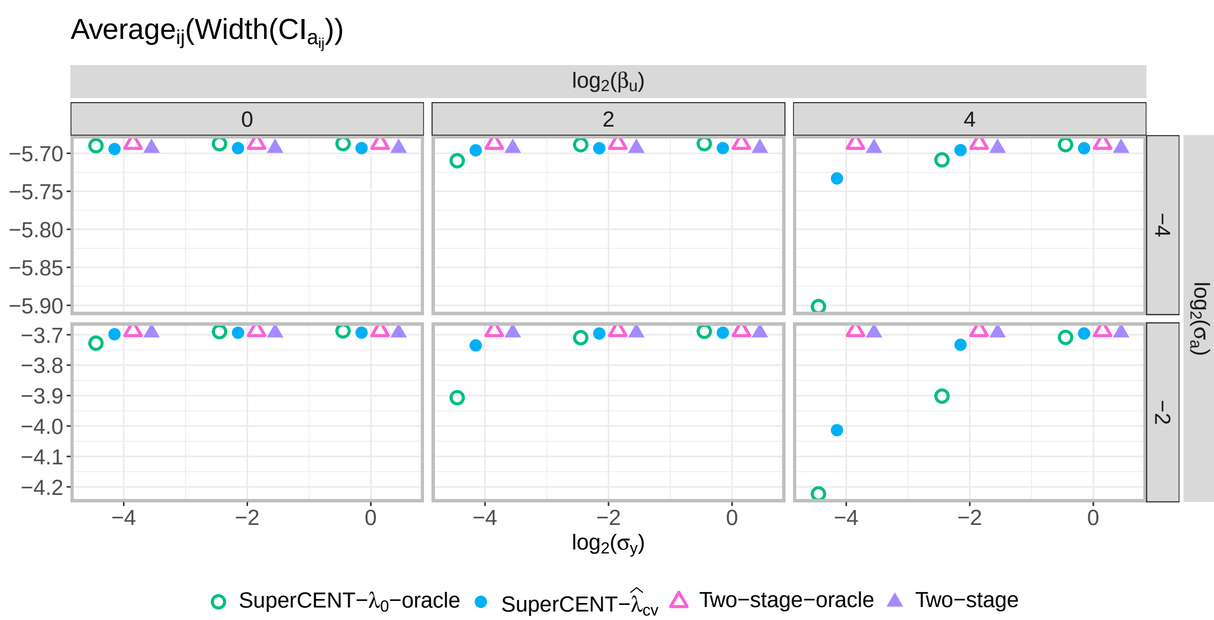
3.3.3 Coverage of $CI_{a_{ij}}$

The following chunk produces the plots of the average empirical coverage of the 95% confidence interval for each entry of A_0, i.e., $CP(CI_{a_{ij}})$.



3.3.4 Width of $CI_{a_{ij}}$

The following chunk produces the plots of the average width of the 95% confidence interval for each entry of A_0, i.e., $Width(CI_{a_{ij}})$.



4 Inconsistent regime of two-stage

This section is for the plots of the inconsistent regime of the two-stage procedure.

4.1 Calculation of theoretical results

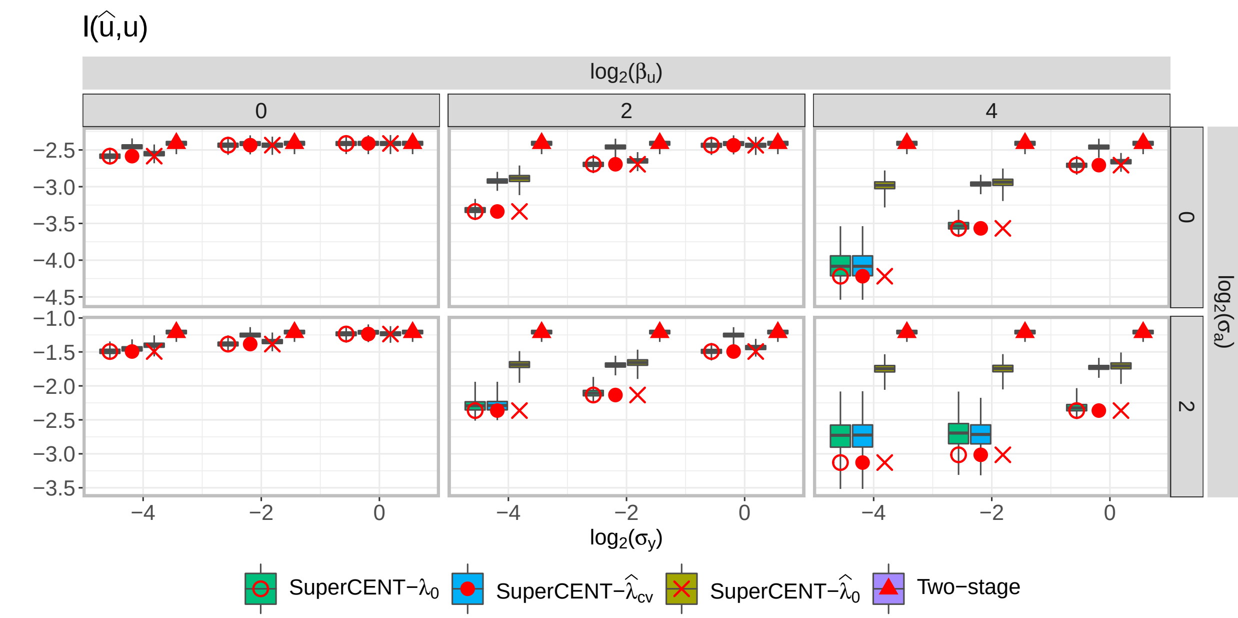
4.2 Estimation property

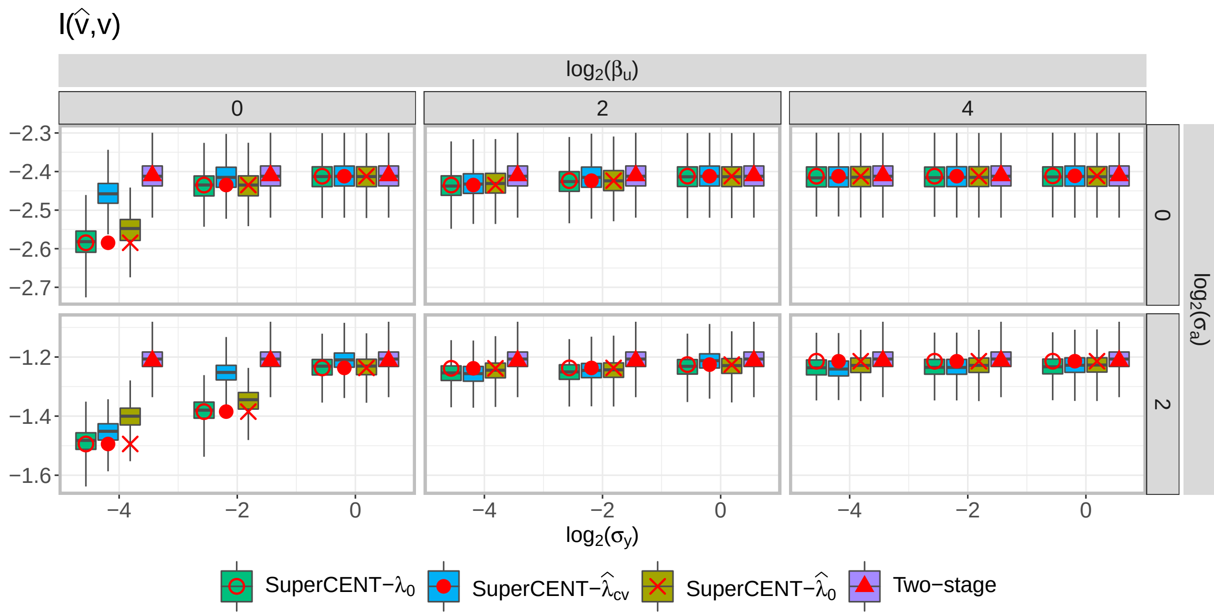
For the estimation accuracy, we compare the following procedures:

1. **Two-stage**: the two-stage procedure as in Algorithm1;
2. **SuperCENT-λ_0**: SuperCENT Algorithm 2 with the optimal $\lambda_0 = n\sigma_y^2/\sigma_a^2$, where the true σ_y, σ_a are used;
3. **SuperCENT-$\hat{\lambda}_0$**: SuperCENT with estimated tuning parameter $\hat{\lambda}_0 = n(\hat{\sigma}_y^{ts})^2/(\hat{\sigma}_a^{ts})^2$, where $(\hat{\sigma}_y^{ts})^2 = \frac{1}{n-p-2}\|\hat{y}^{ts} - y\|_2^2$ and $(\hat{\sigma}_a^{ts})^2 = \frac{1}{n^2}\|\hat{A}^{ts} - A_0\|_F^2$ are estimated from the two-stage procedure;
4. **SuperCENT-$\hat{\lambda}_{cv}$**: SuperCENT with tuning parameter $\hat{\lambda}_{cv}$ chosen by 10-fold cross-validation.

4.2.1 u and v

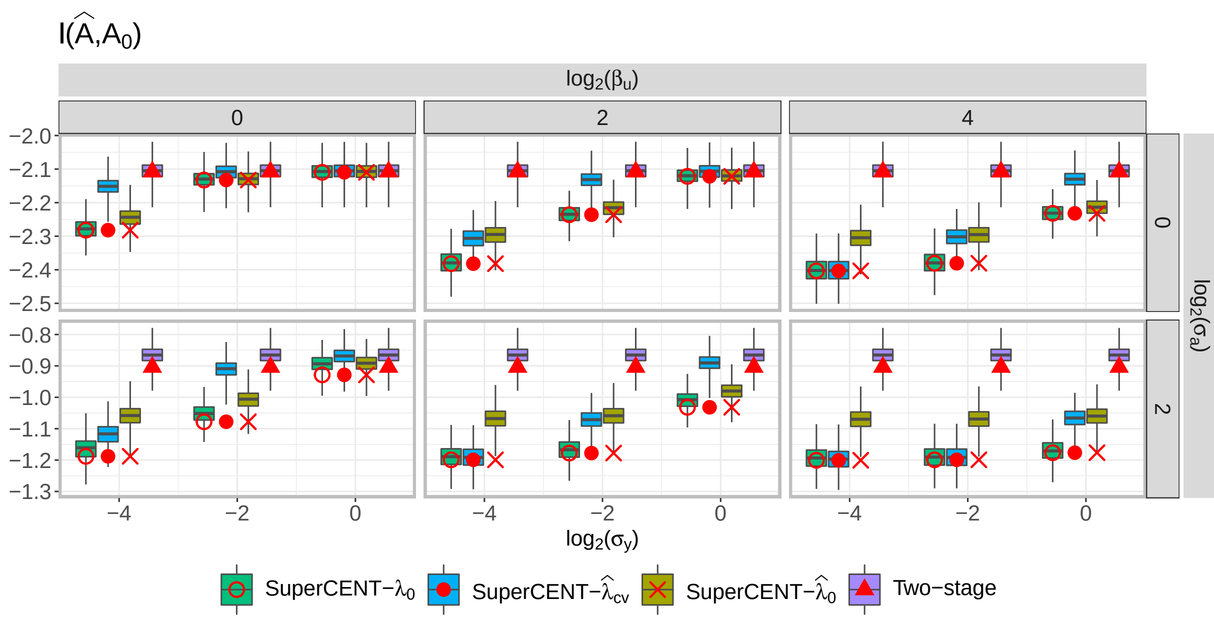
The following chunks produce the plots of $l(\widehat{u}, u)$ and $l(\widehat{v}, v)$.





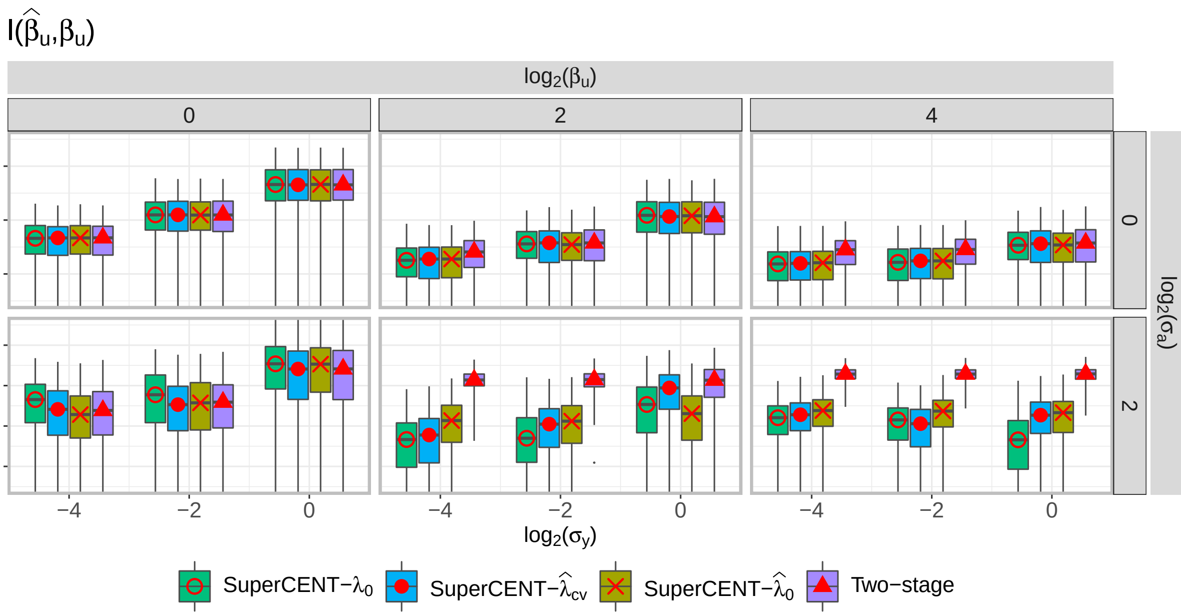
4.2.2 A

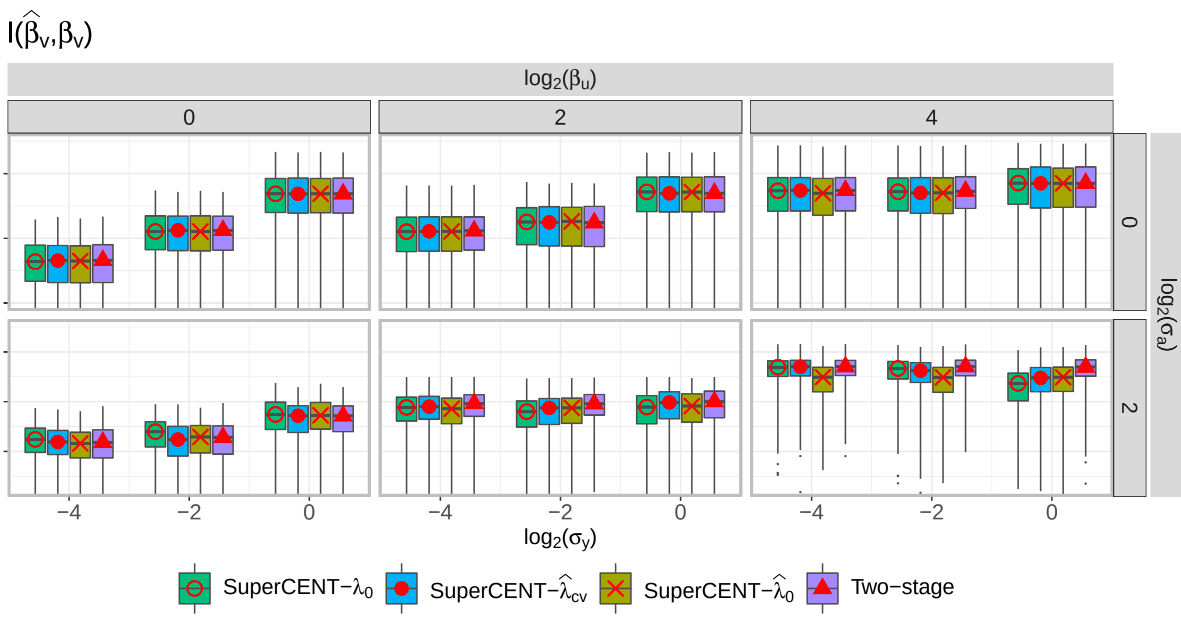
The following chunk produces the plot of $l(\widehat{A}, A_0)$.



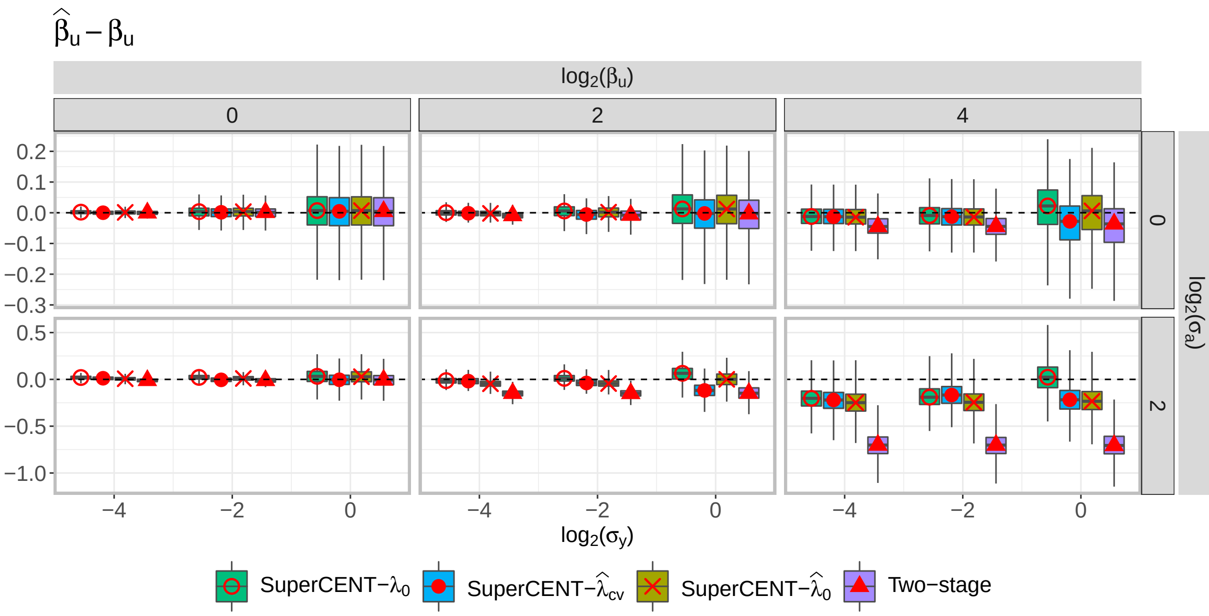
4.2.3 β_u and β_v

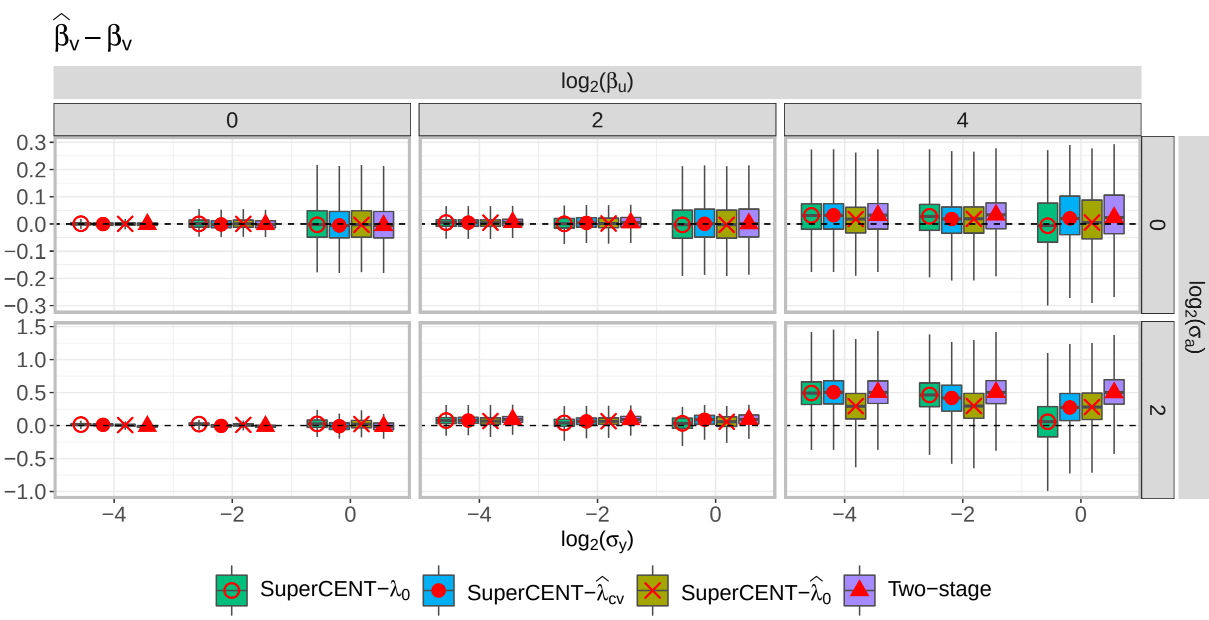
The following chunks produce the plots of $l(\widehat{\beta}_u, \beta_u)$ and $l(\widehat{\beta}_v, \beta_v)$.





The following chunks produce the plots of $\widehat{\beta}_u - \beta_u$ and $\widehat{\beta}_v - \beta_v$.



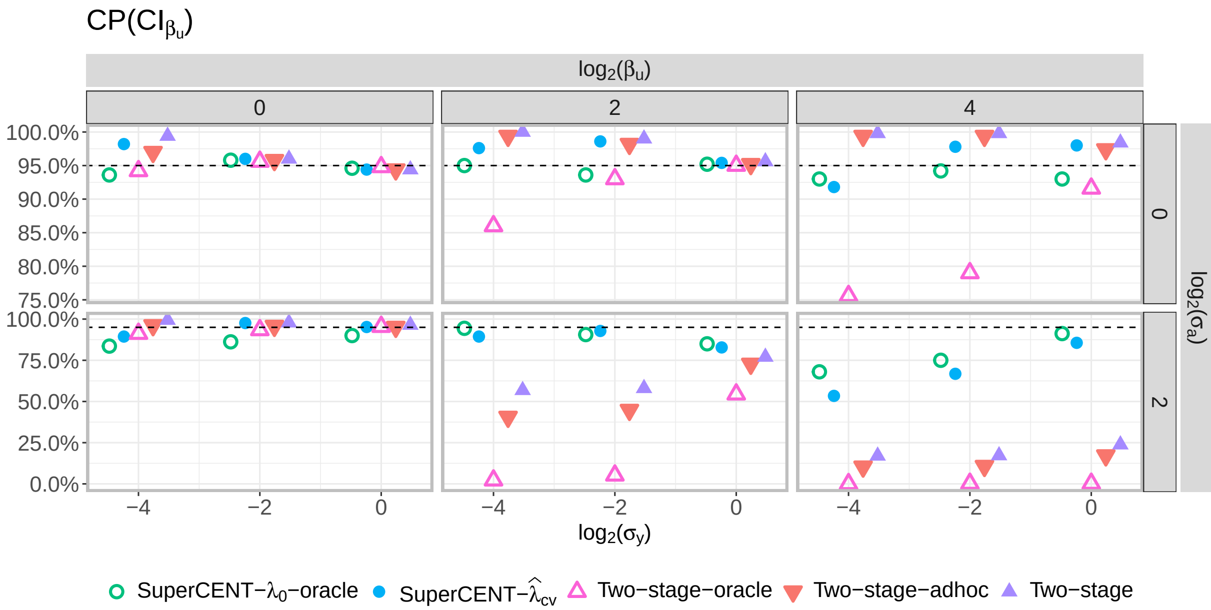


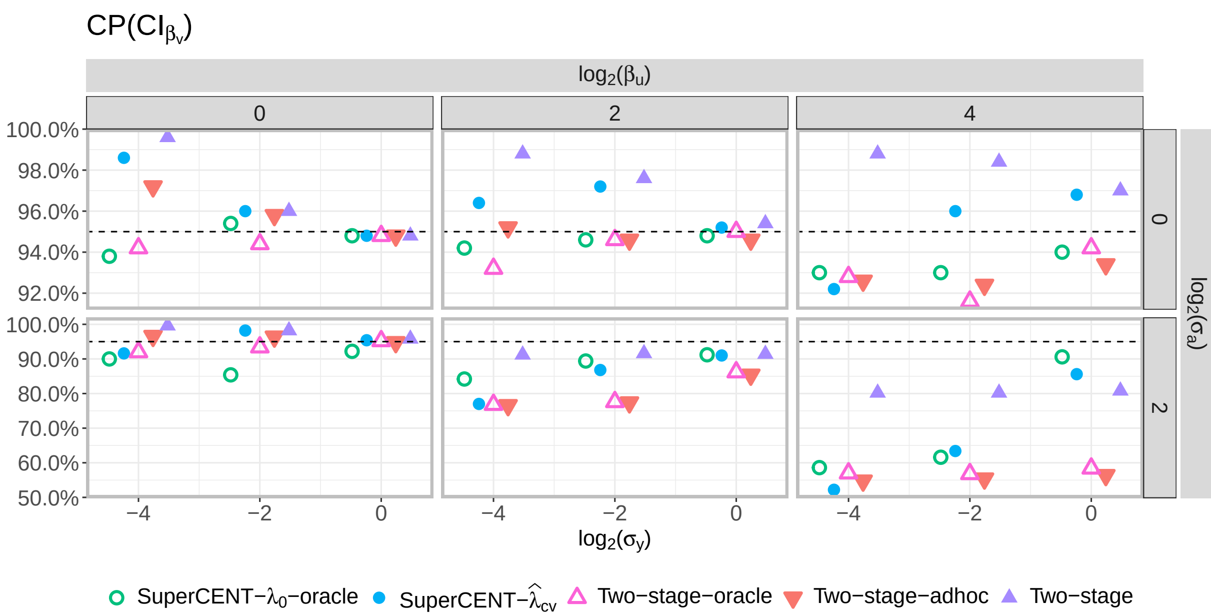
4.3 Inference property

This subsection is for the confidence intervals (CIs) for the regression coefficient, $\{CI_{\beta_u}$ and CI_{β_v}.

4.3.1 Coverage of CI_{β_u} and CI_{β_v}

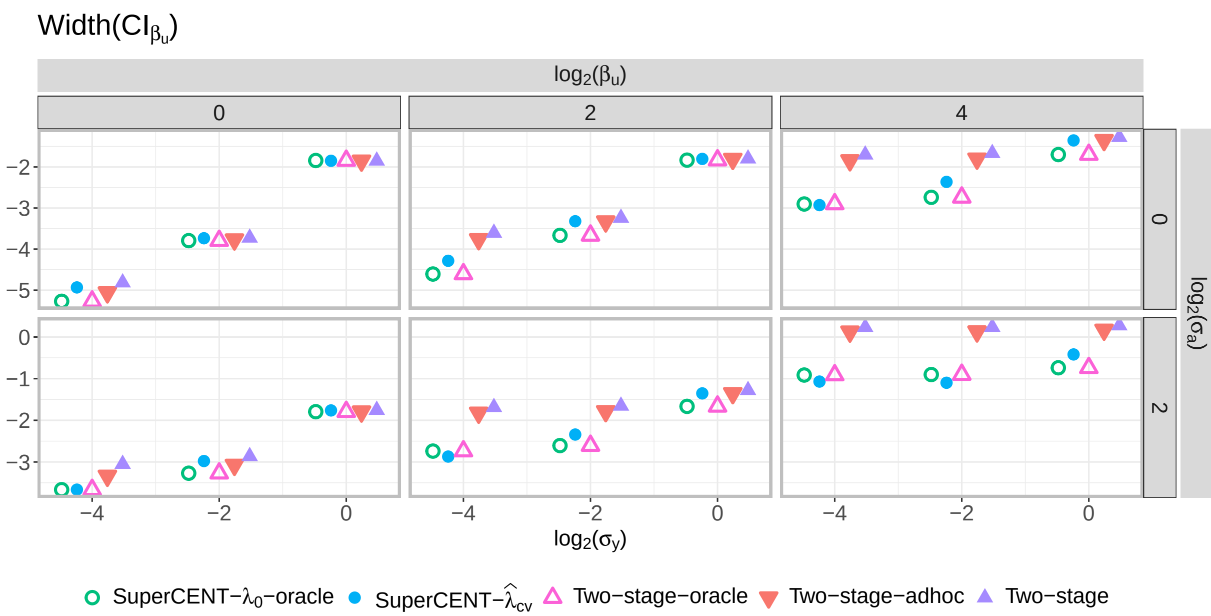
The following chunks produce the plots of the empirical coverage of the 95% confidence interval for β_u and β_v respectively, i.e., $CP(CI_{\beta_u})$ and $CP(CI_{\beta_v})$.

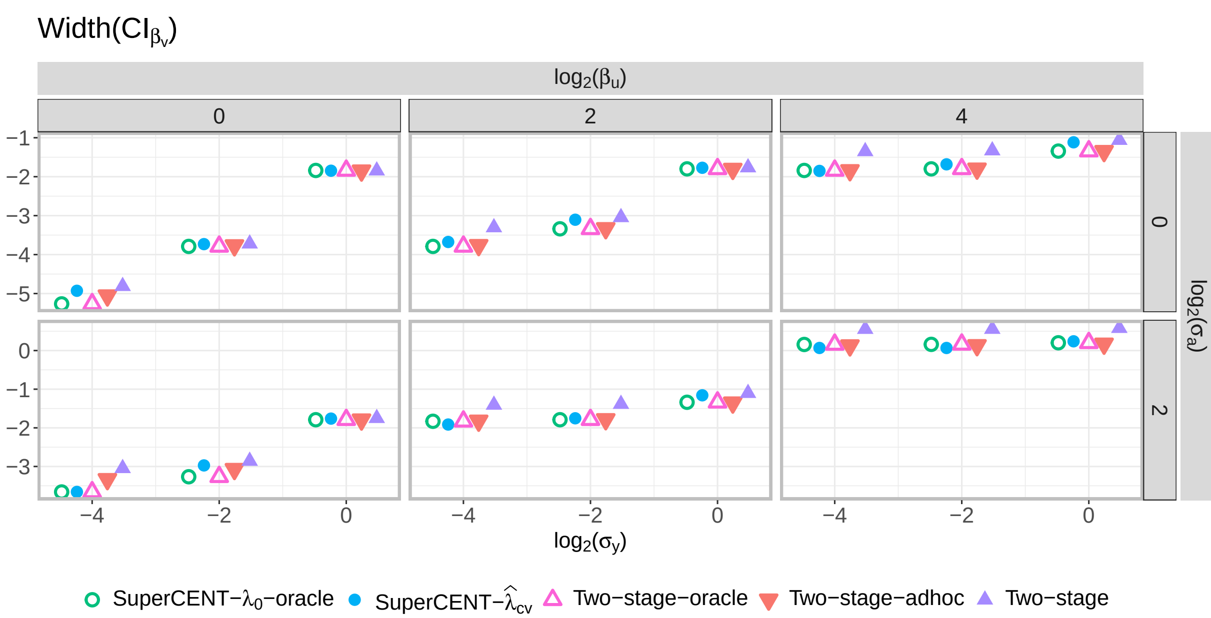




4.3.2 Width of CI_{β_u} and CI_{β_v}

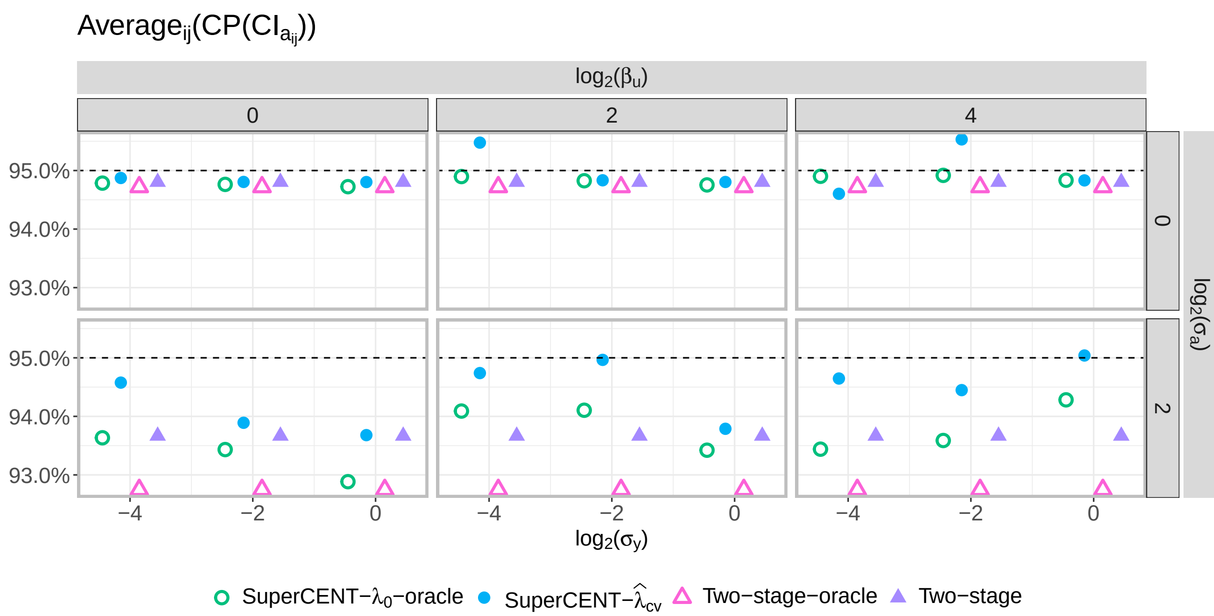
The following chunks produce the plots of the average of width of the 95% confidence interval for β_u and β_v respectively, i.e., $Width(CI_{\beta_u})$ and $Width(CI_{\beta_v})$.





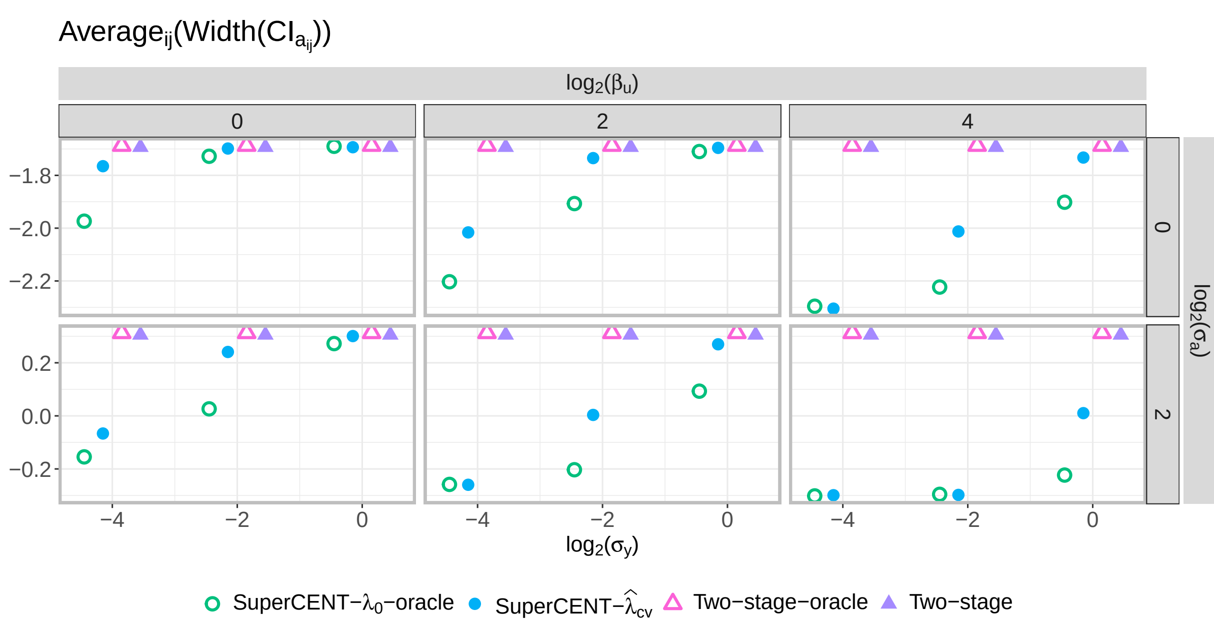
4.3.3 Coverage of $CI_{a_{ij}}$

The following chunk produces the plots of the average empirical coverage of the 95% confidence interval for each entry of A_0, i.e., $CP(CI_{a_{ij}})$.



4.3.4 Width of $CI_{a_{ij}}$

The following chunk produces the plots of the average width of the 95% confidence interval for each entry of A_0, i.e., $Width(CI_{a_{ij}})$.

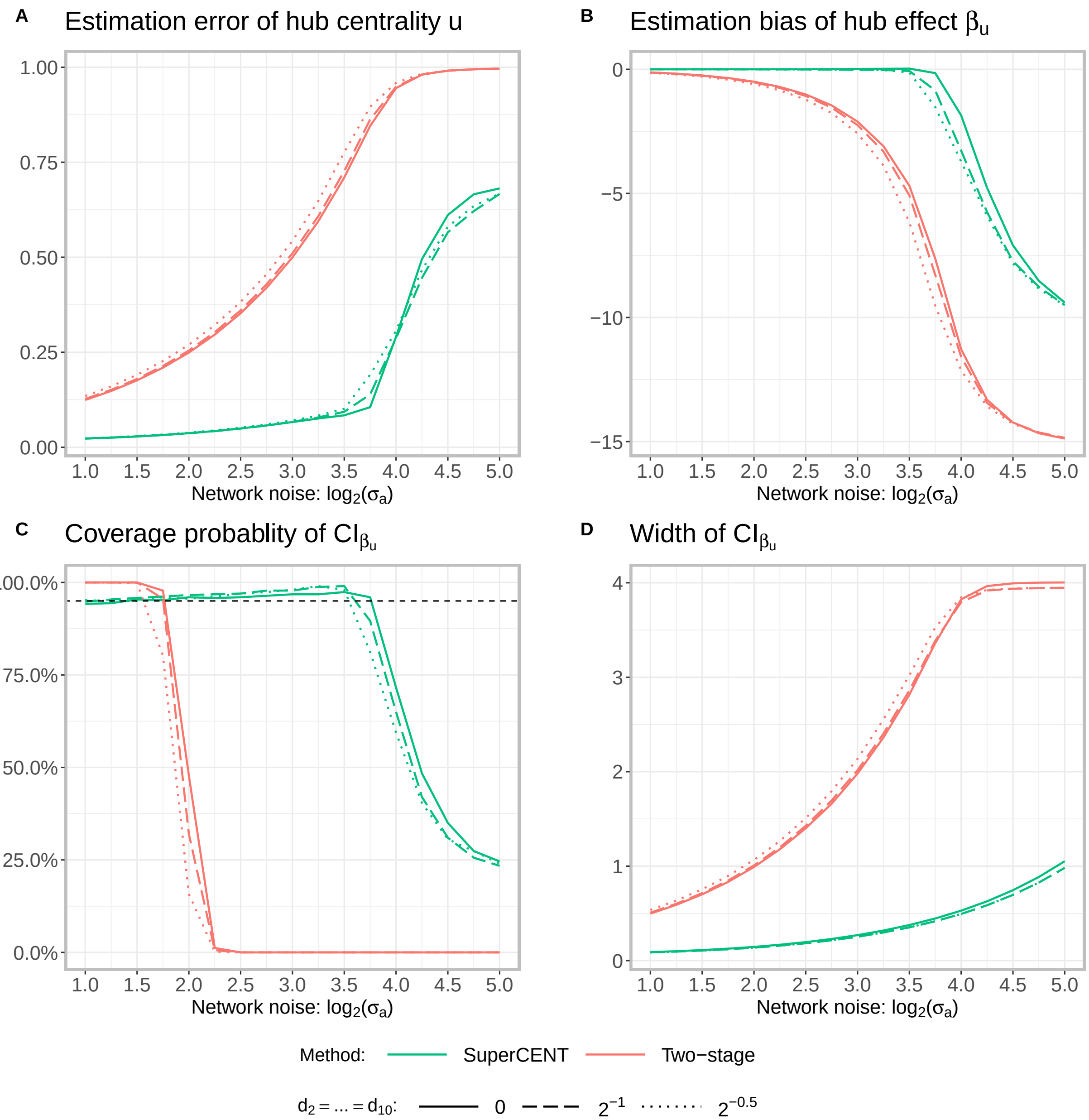


5 Rank-10 network example

The section for the simulation results of the multi-rank unified framework with a rank-10 network.

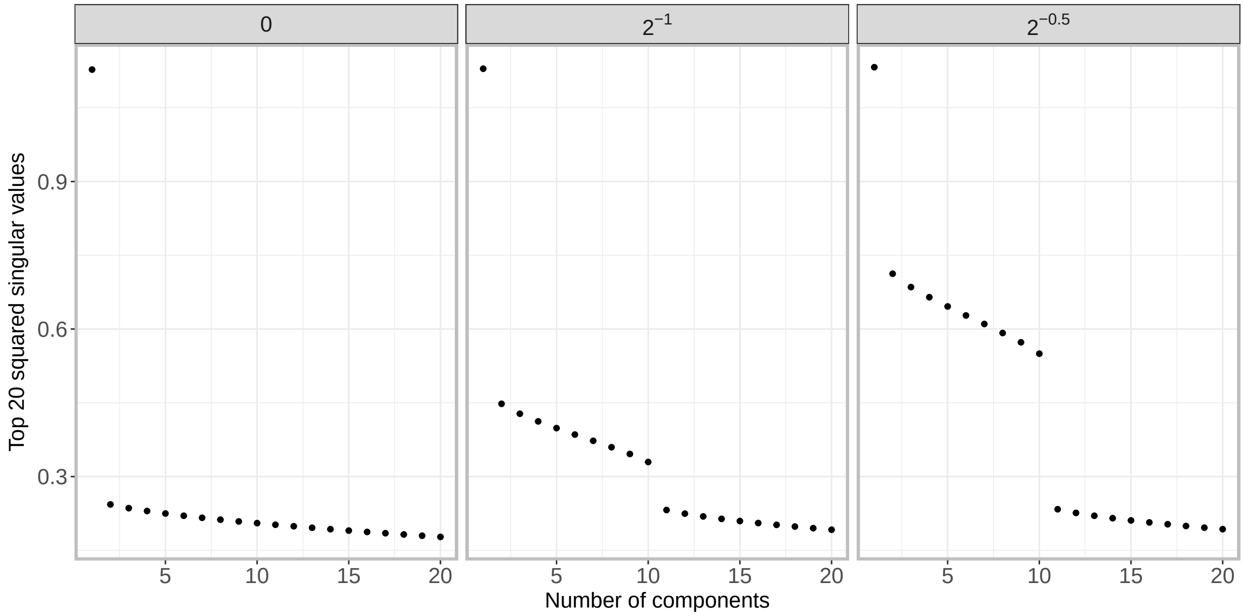
5.1 Four plots

The following chunk produces the plot for the comparison between the two-stage and SuperCENT in terms of the estimations of u and β_u as well as the coverage probability and the width of CI_{β_u} varying the network noise level σ_a and the gap between the leading singular value and the non-leading ones.



5.2 Top 20 singular values

This section shows the top 20 squared singular values of the observed network adjacency matrix A with fixed leading singular value $d_1 = 1$ for the noiseless component A_0, fixed noise level $\sigma_a = 2$ and various non-leading singular values for A_0.



0
2^{-1}
$2^{-0.5}$
Top 20 squared singular values
0.9
0.6
0.3
5
10
15
20
Number of components