Trade-off between predictive performance and FDR control for high-dimensional Gaussian model selection
(Supplementary material)

Perrine Lacroix

Contents

This supplementary material is an appendix of the article entitled "Trade-off between predictive performance and FDR control for high-dimensional Gaussian model selection" available in `https://doi.org/10.48550/arxiv.2302.01831`. This material contains two sections. Section 1 contains plots for the bounds $B(K, \hat{\beta}_{\hat{m}(\tilde{K})}, \hat{\sigma}^2)$ for each of the four scenarios described in Table 3 of Subsection 7.1. It is a complementary work to Subsection 7.2 of the article. Section 2 contains plots evaluating the robustness of the model collections and tables of results of the random model collection constructions and applications of variable selection procedures for scenarios (ii), (iii) and (iv) described in Table 3 of Subsection 7.1. It is a complementary work to Subsections 4.3 and 4.4 of the article.

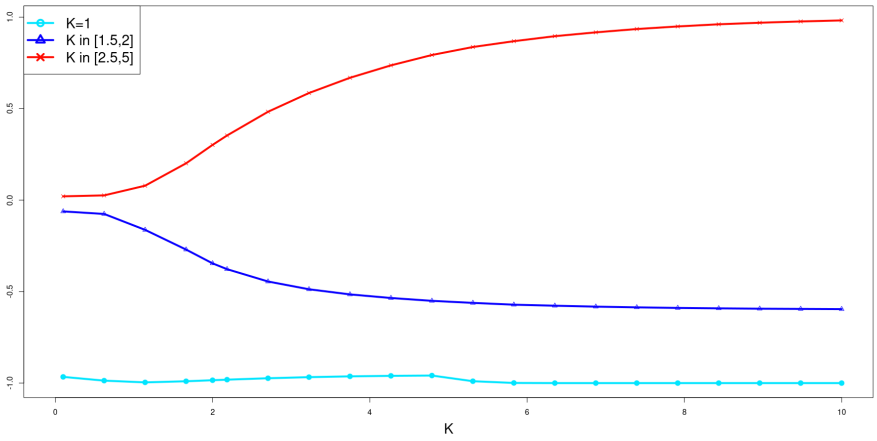
In Section 1, graphs for scenarios (i) to (iv) described in Table 3 are provided. Relative changes and relative standard deviations for the $B(K, \hat{\beta}_{\hat{m}(\tilde{K})}, \hat{\sigma}^2)$ bounds when $\tilde{K} \in \{1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, \log(n)\}$ are plotted in Figures S-1- S-2 for scenario (i), in Figures S-4- S-5 for scenario (ii), in Figures S-7- S-8 for scenario (iii) and in Figures S-10- S-11 for scenarios (iv). The empirical difference in predictions and the empirical FDR$\big(\hat{m}(K)\big)$ functions, the estimated difference in predictions (4.2) and the $B(K, \hat{\beta}_{\hat{m}(4)}, \hat{\sigma}^2)$ functions for a grid of values of $K > 0$ are plotted on Figure S-3 for scenario (i), Figure S-6 for scenario (ii), Figure S-9 for scenario (iii) and Figure S-12 for scenario (iv).
In Section 2, graphs for scenarios (ii) to (iv) described in Table 3 are provided. The empirical FDR, the theoretical FDR, the $B(K, \hat{\beta}_{\hat{m}(4)}, \hat{\sigma}^2)$, the diff-PR$\big(\hat{m}(K)\big)$ and the $\widehat{\text{diff-PR}}\big(\hat{m}(K)\big)$ functions, calculated on the three perturbed model collections described in Subsection 4.3.1, are plotted in Figures S-13- S-15. Table S-1 contains the proportion of active variables in models of size 5, 10, 15 an 20 for random collections built with Bolasso, SLOPE, random forest and the knockoff method. Lastly, Tables S-2- S-4 contain the dimension, PR and FDR of the selected models obtained by LinSelect, the 50-fold CV, the knockoff method and our algorithm, respectively applied on the nested model collection (2.1), the random collection built with Bolasso and the random collection built with the knockoff method.
All the R scripts are available at `https://github.com/PerrineLacroix/Trade_off_FDR_PR`.

When we focus on the scenario (i), the higher the D_{m^*} value is, the smaller the empirical FDR is but the larger the empirical PR for large K is. Moreover, the relative change functions decreases when D_m^* increases, as well as the relative standard deviation ones which remain smaller than 0.5. This can be explained since the higher D_{m^*}, the smaller the number of non active variables, so the smaller the number of the selected non active variables and the smaller FDR value. In the opposite trend, the empirical PR increases with D_{m^*} since penalization tends to select too few variables, especially even K moves away from 2.
As expected, concerning the scenario (ii), when coefficients are smaller than the amplitude of the noise (the second configuration), values of the relative change for the $B(K, \hat{\beta}_{\hat{m}(\tilde{K})}, \hat{\sigma}^2)$ bounds explode (until 10^5) and the relative standard deviation values increase until exceed 1. The best results are obtained for the first β^* configuration of the scenario (ii), but results still remain reasonable with the third one. The relative standard deviation functions increase after $\tilde{K} \geq 4$ whereas in all other scenarios, functions always decrease when $\tilde{K}$ increases. When permutations of the ten first variables are processed from the nested model collection (2.1), $B(K, \hat{\beta}_{\hat{m}(4)}, \hat{\sigma}^2)$ values begin to diverge from those of $B(K, \beta^*, \sigma^2)$ for the second and the third

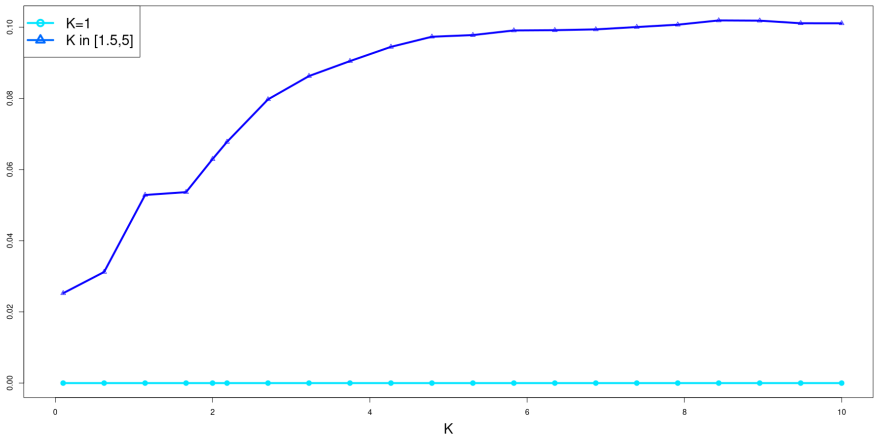
configurations where the coefficients of β^* are close to each other. Unlike the other scenarios, $B(K, \hat{\beta}_{\hat{m}(4)}, \hat{\sigma}^2)$ and $B(K, \beta^*, \sigma^2)$ values are both larger than the empirical FDR ones for the second configuration.
For configuration (iii) and when permutations of the first twelve and fifteen variables are processed, FDR values are the highest all along the collections compared to all other scenarios (similar to scenario (iv) and $\sigma^2 = 4$) and so, distinction between active and non active variables is more difficult. Proportions of active variables in models of size 5, 10, 15 and 20 fall to 0.6 with the third configuration and 0.8 with the second configuration for which the discrimination between active and non active variables is naturally less obvious.
As for the scenario (iii), we observe unsurprisingly that the higher the value of n, the smaller the relative change, the smaller the relative standard deviation, and the tighter the confidence interval of PR (< 0.04 for $n = 30$). However, we note that the computational time to estimate the bounds was significantly higher for $n = 300$.
Lastly, concerning the scenario (iv), as expected, the higher the noise amplitude, the larger the confidence interval for the PR (< 0.45 for $\sigma^2 = 4$), the higher the relative change (which equals 0 when $\sigma^2 = 0.1$ and around 2 when $\sigma^2 = 4$) and the higher the relative standard deviation. However, values remain reasonable even for $\sigma^2 = 4$ excepted for the empirical PR values which are always larger than 5. For $\sigma^2 = 4$ and when permutations of the first twelve and fifteen variables are processed, FDR values are the highest all along the collections compared to all other scenarios (similar to scenario (ii) configuration (iii)) and so, distinction between active and non active variables is more difficult. Unlike the other scenarios, the Bolasso provides the highest values of proportion of active variables in models of size 5, 10, 15 and 20 when $\sigma^2 = 0.1$. Proportions fall to 0.8 when $\sigma^2 = 4$.

1 Ordered variable selection

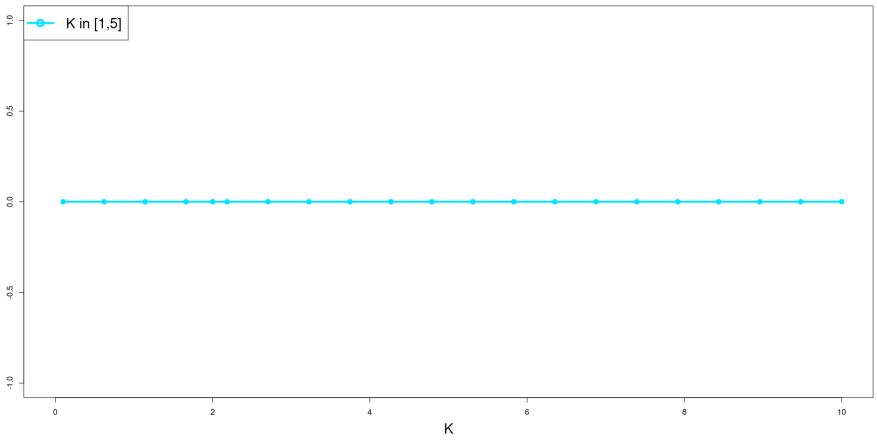
1.1 Scenario (i)



$|\beta^*| = 1$

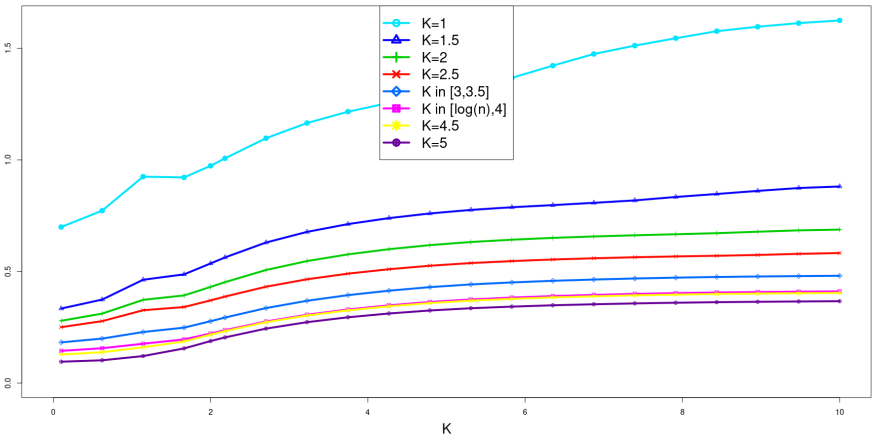


$|\beta^*| = 10$

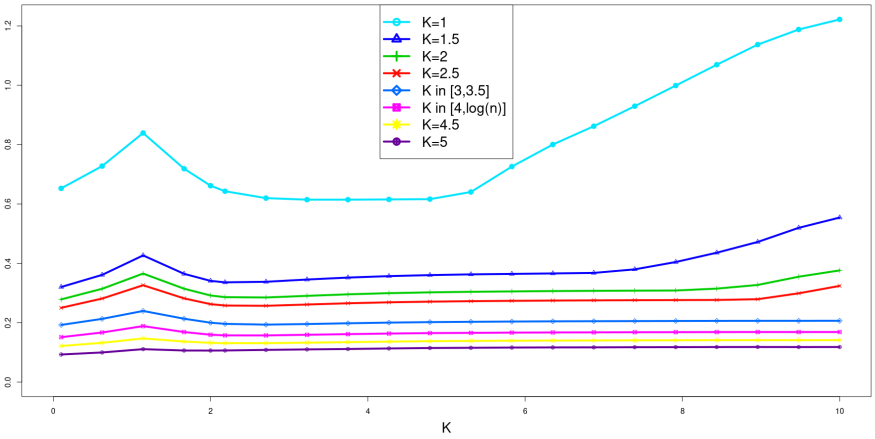


$|\beta^*| = 20$

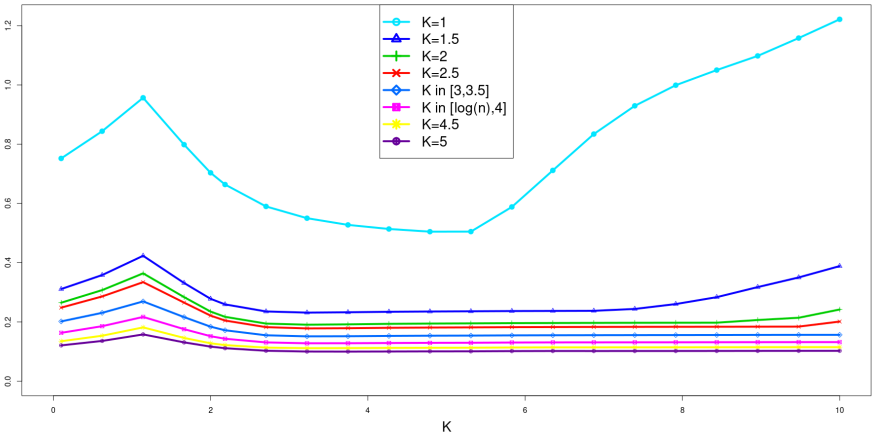
Figure S-1: Curves of the relative change values between the functions $B(K, \beta^*, \sigma^2)$ and the functions $B(K, \hat{\beta}_{\hat{m}(\tilde{K})}, \hat{\sigma}^2)$ with respectively $\hat{\beta}_{\hat{m}(1)}$, $\hat{\beta}_{\hat{m}(1.5)}$, $\hat{\beta}_{\hat{m}(2)}$, $\hat{\beta}_{\hat{m}(2.5)}$, $\hat{\beta}_{\hat{m}(3)}$, $\hat{\beta}_{\hat{m}(3.5)}$, $\hat{\beta}_{\hat{m}(4)}$, $\hat{\beta}_{\hat{m}(4.5)}$, $\hat{\beta}_{\hat{m}(5)}$ and $\hat{\beta}_{\hat{m}(\log(n))}$ where estimators are calculating from a single data set. Top: for $|\beta^*| = 1$. Middle: for $|\beta^*| = 10$. Bottom: for $|\beta^*| = 20$.



$|\beta^*| = 1$

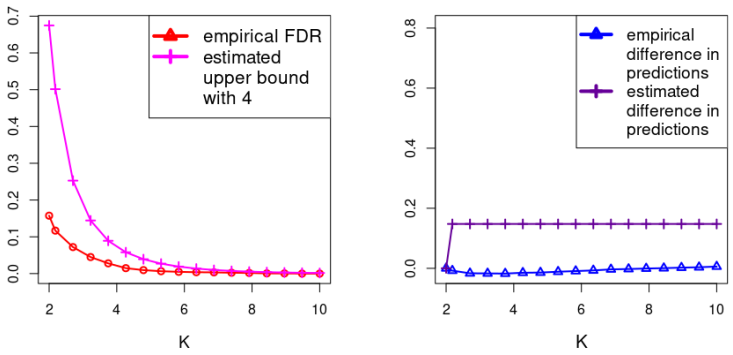


$|\beta^*| = 10$



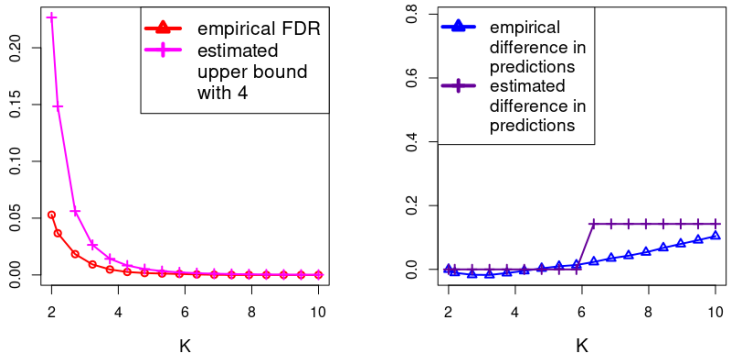
$|\beta^*| = 20$

Figure S-2: Curves of the relative standard deviation (standard deviation normalized by the mean) of the functions $B(K, \hat{\beta}_{\hat{m}(\tilde{K})}, \hat{\sigma}^2)$ obtained from 100 data sets. With each one, $\hat{\beta}_{\hat{m}(1)}, \hat{\beta}_{\hat{m}(1.5)}, \hat{\beta}_{\hat{m}(2)}, \hat{\beta}_{\hat{m}(2.5)}, \hat{\beta}_{\hat{m}(3)}, \hat{\beta}_{\hat{m}(3.5)}, \hat{\beta}_{\hat{m}(4)}, \hat{\beta}_{\hat{m}(4.5)}, \hat{\beta}_{\hat{m}(5)}$ and $\hat{\beta}_{\hat{m}(\log(n))}$ are calculated given $B(K, \hat{\beta}_{\hat{m}(\tilde{K})}, \hat{\sigma}^2)$, variance of the 100 $B(K, \hat{\beta}_{\hat{m}(\tilde{K})}, \hat{\sigma}^2)$ functions and then the relative standard deviation with respect to K. Top: for $|\beta^*| = 1$. Middle: for $|\beta^*| = 10$. Bottom: for $|\beta^*| = 20$.



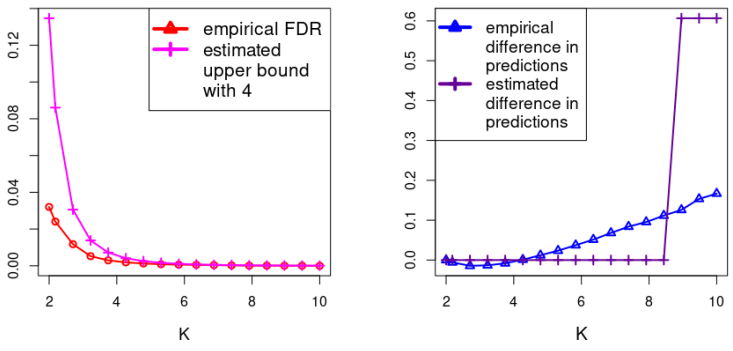
FDR diff-PR

$|\beta^*| = 1$



FDR diff-PR

$|\beta^*| = 10$

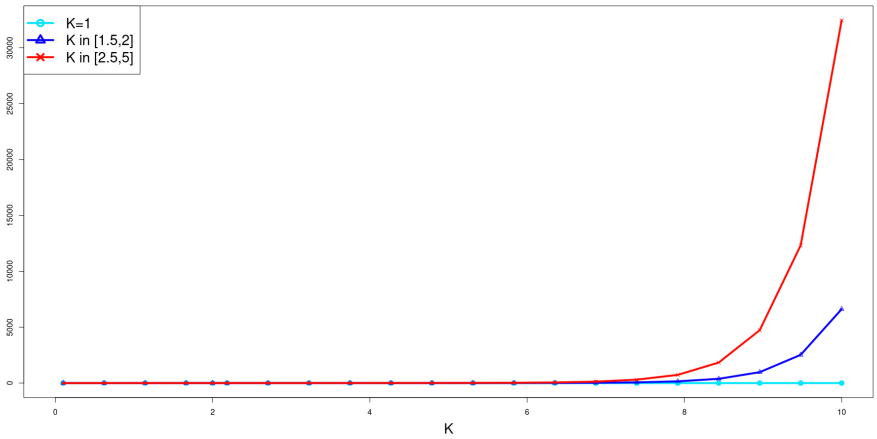


FDR diff-PR

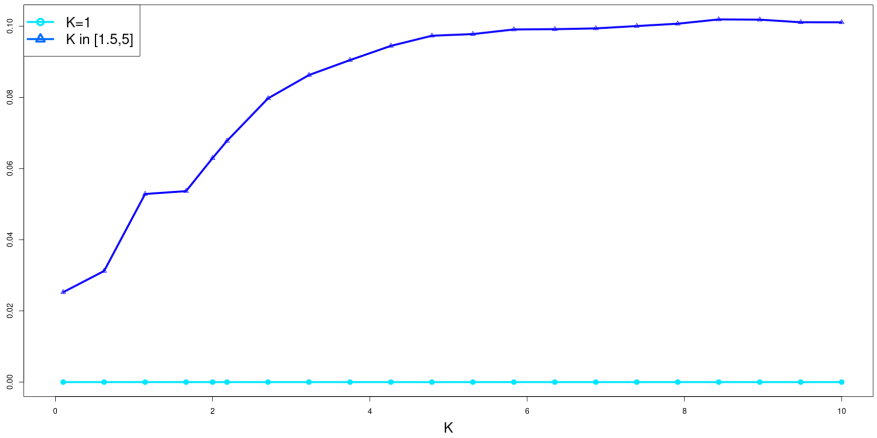
$|\beta^*| = 20$

Figure S-3: Curves of the empirical functions $\mathrm{FDR}\big(\hat{m}(K)\big)$ (red) and $\text{diff-PR}\big(\hat{m}(K)\big)$ (blue), of the $B(K, \hat{\beta}_{\hat{m}(4)}, \hat{\sigma}^2)$ functions (pink) and of $\widehat{\text{diff-PR}}\big(\hat{m}(K)\big)$ (violet) for $K \geq 2$ for the *toy data set*. Top: for $D_m = 1$. Middle: for $D_m = 10$. Bottom: for $D_m = 20$.

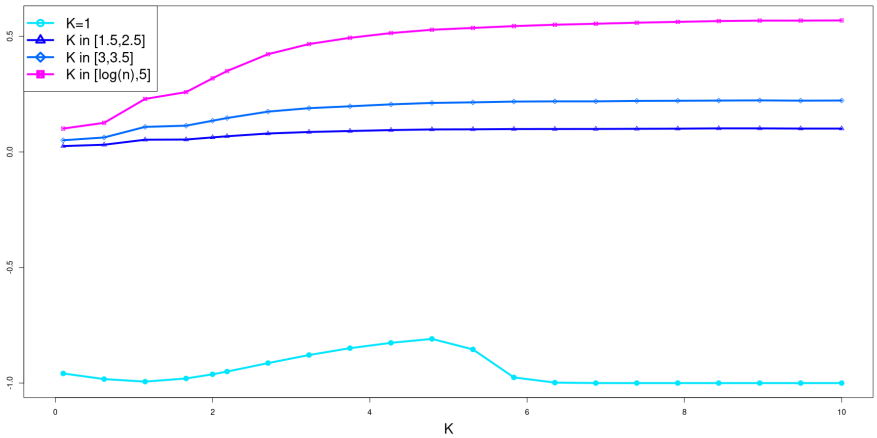
1.2 Scenario (ii)



$\beta^*_{10} = \frac{2}{10}$

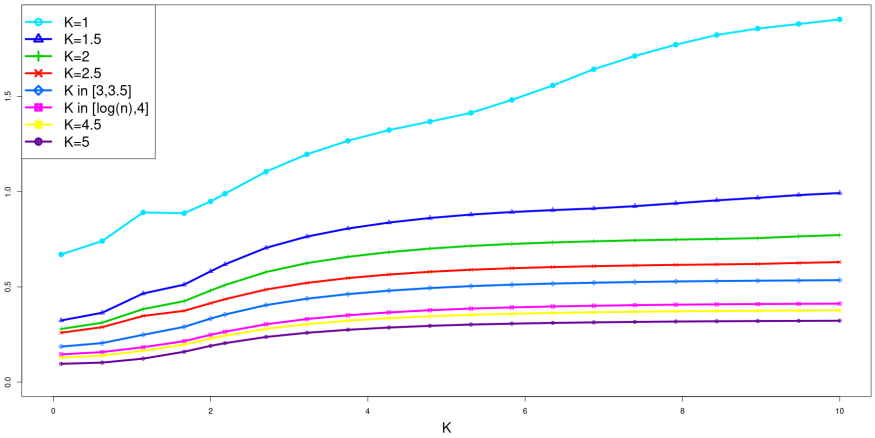


$\beta^*_{10} = 2$ and distant coefficients.

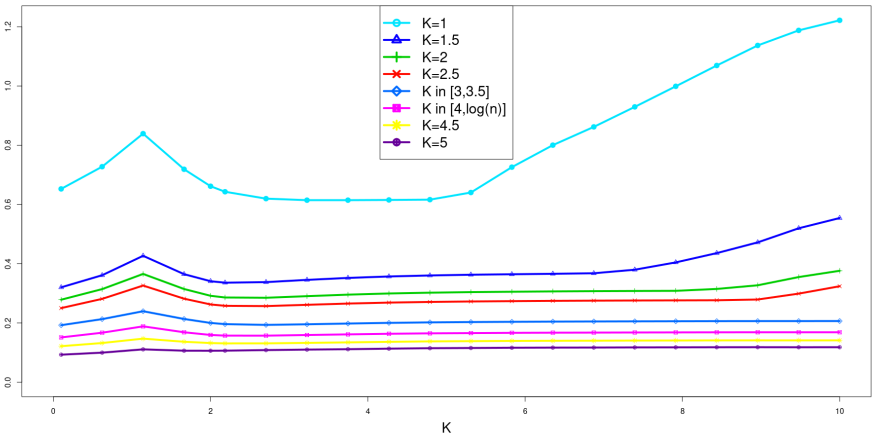


$\beta^*_{10} = 2$ and close coefficients.

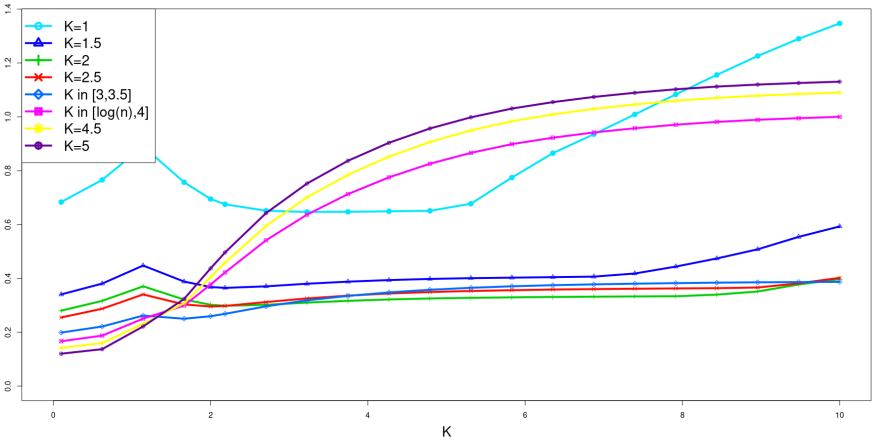
Figure S-4: Curves of the relative change values between the functions $B(K, \beta^*, \sigma^2)$ and the functions $B(K, \hat{\beta}_{\hat{m}(\tilde{K})}, \hat{\sigma}^2)$ with respectively $\hat{\beta}_{\hat{m}(1)}$, $\hat{\beta}_{\hat{m}(1.5)}$, $\hat{\beta}_{\hat{m}(2)}$, $\hat{\beta}_{\hat{m}(2.5)}$, $\hat{\beta}_{\hat{m}(3)}$, $\hat{\beta}_{\hat{m}(3.5)}$, $\hat{\beta}_{\hat{m}(4)}$, $\hat{\beta}_{\hat{m}(4.5)}$, $\hat{\beta}_{\hat{m}(5)}$ and $\hat{\beta}_{\hat{m}(\log(n))}$ where estimators are calculating from a single data set. Top: for $\beta^*_{10} = \frac{2}{10}$. Middle: for $\beta^*_{10} = 2$ and distant coefficients. Bottom: for $\beta^*_{10} = 2$ and close coefficients.



$\beta^*_{10} = \frac{2}{10}$

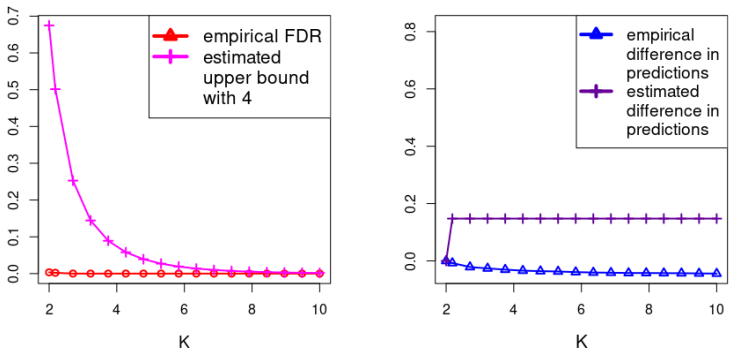


$\beta^*_{10} = 2$ and distant coefficients.



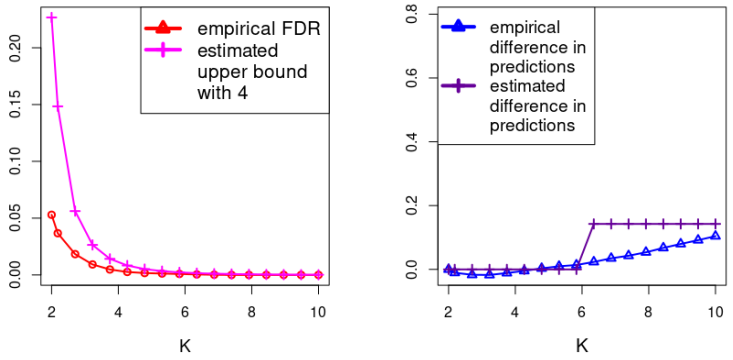
$\beta^*_{10} = 2$ and close coefficients.

Figure S-5: Curves of the relative standard deviation (standard deviation normalized by the mean) of the functions $B(K, \hat{\beta}_{\hat{m}(\tilde{K})}, \hat{\sigma}^2)$ obtained from 100 data sets. With each one, $\hat{\beta}_{\hat{m}(1)}, \hat{\beta}_{\hat{m}(1.5)}, \hat{\beta}_{\hat{m}(2)}, \hat{\beta}_{\hat{m}(2.5)}, \hat{\beta}_{\hat{m}(3)}, \hat{\beta}_{\hat{m}(3.5)}, \hat{\beta}_{\hat{m}(4)}, \hat{\beta}_{\hat{m}(4.5)}, \hat{\beta}_{\hat{m}(5)}$ and $\hat{\beta}_{\hat{m}(\log(n))}$ are calculated given $B(K, \hat{\beta}_{\hat{m}(\tilde{K})}, \hat{\sigma}^2)$, variance of the 100 $B(K, \hat{\beta}_{\hat{m}(\tilde{K})}, \hat{\sigma}^2)$ functions and then the relative standard deviation with respect to K. Top: for $\beta^*_{10} = \frac{2}{10}$. Middle: for $\beta^*_{10} = 2$ and distant coefficients. Bottom: for $\beta^*_{10} = 2$ and close coefficients.



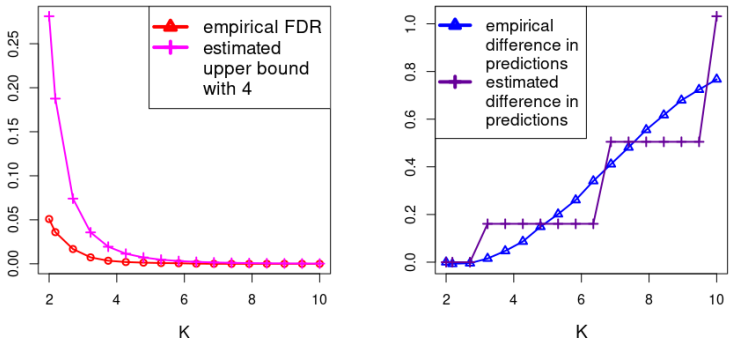
FDR diff-PR

$\beta^*_{10} = \frac{2}{10}$



FDR diff-PR

$\beta^*_{10} = 2$ and distant coefficients

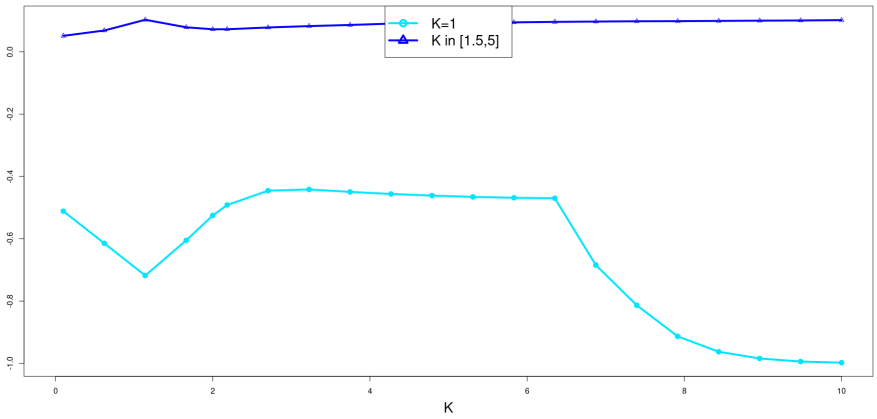


FDR diff-PR

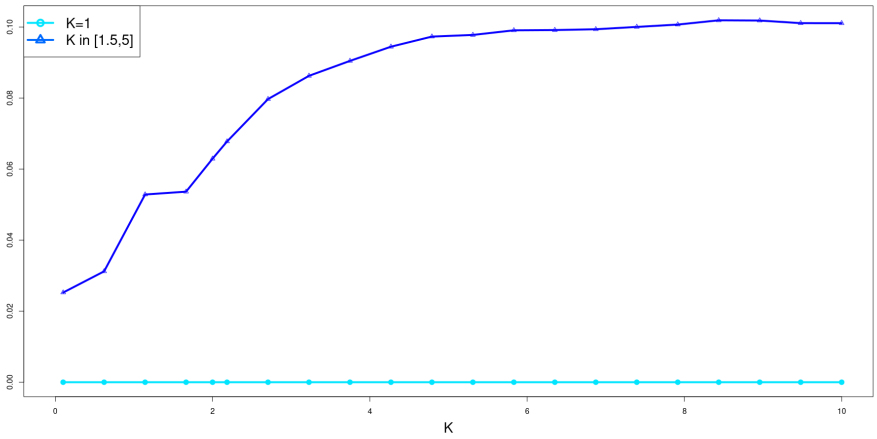
$\beta^*_{10} = 2$ and close coefficients

Figure S-6: Curves of the empirical functions $\mathrm{FDR}\big(\hat{m}(K)\big)$ (red) and $\text{diff-PR}\big(\hat{m}(K)\big)$ (blue), of the $B(K, \hat{\beta}_{\hat{m}(4)}, \hat{\sigma}^2)$ functions (pink) and of $\widehat{\text{diff-PR}}\big(\hat{m}(K)\big)$ (violet) for $K \geq 2$ for the *toy data set*. Top: for $\beta^*_{10} = \frac{2}{10}$. Middle: for $\beta^*_{10} = 2$ and distant coefficients. Bottom: for $\beta^*_{10} = 2$ and close coefficients.

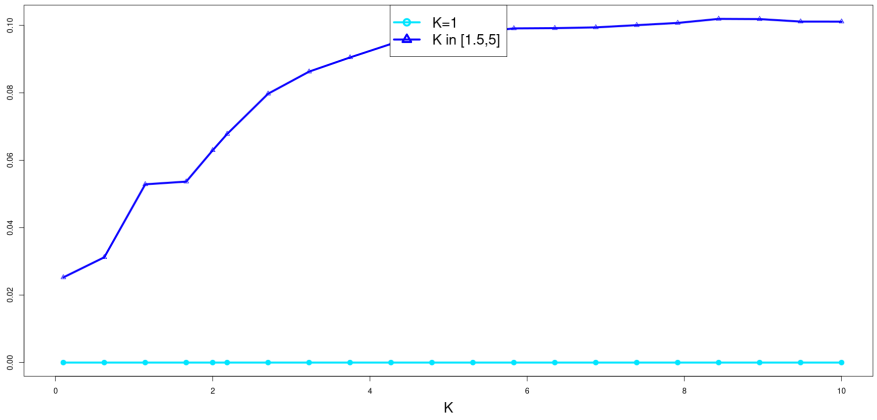
1.3 Scenario (iii)



$n = 30$

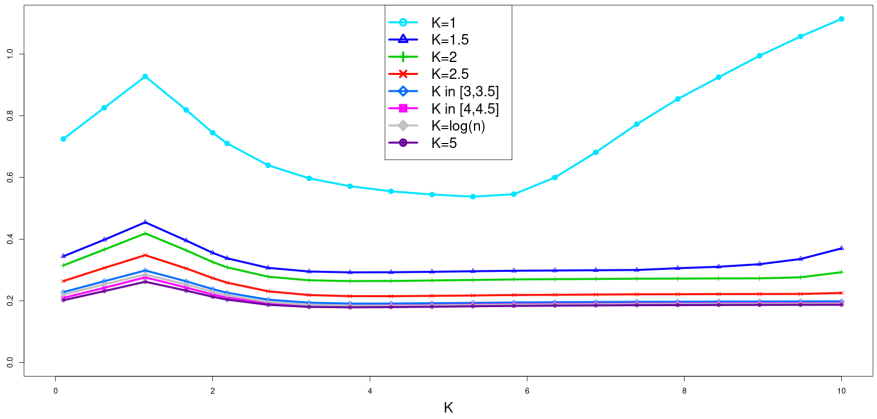


$n = 50$

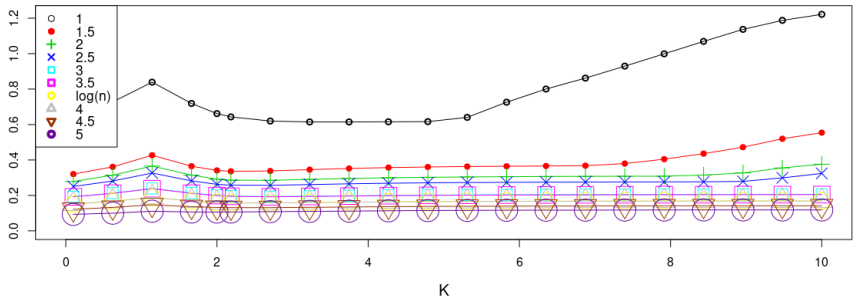


$n = 300$

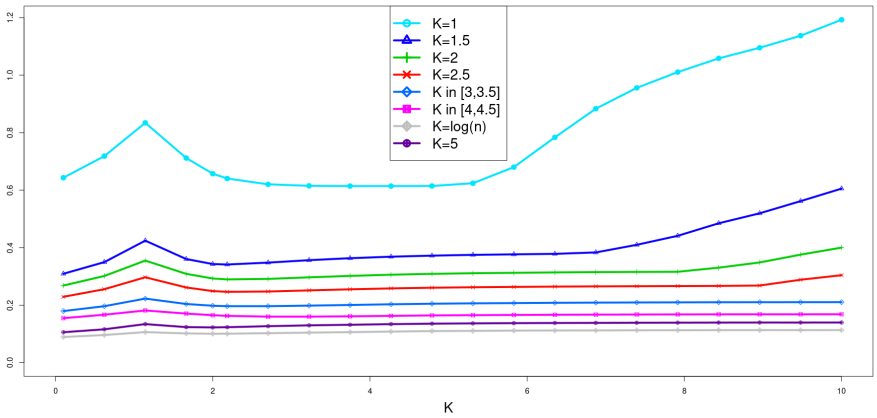
Figure S-7: Curves of the relative change values between the functions $B(K, \beta^*, \sigma^2)$ and the functions $B(K, \hat{\beta}_{\hat{m}(\tilde{K})}, \hat{\sigma}^2)$ with respectively $\hat{\beta}_{\hat{m}(1)}$, $\hat{\beta}_{\hat{m}(1.5)}$, $\hat{\beta}_{\hat{m}(2)}$, $\hat{\beta}_{\hat{m}(2.5)}$, $\hat{\beta}_{\hat{m}(3)}$, $\hat{\beta}_{\hat{m}(3.5)}$, $\hat{\beta}_{\hat{m}(4)}$, $\hat{\beta}_{\hat{m}(4.5)}$, $\hat{\beta}_{\hat{m}(5)}$ and $\hat{\beta}_{\hat{m}(\log(n))}$ where estimators are calculating from a single data set. Top: for $n = 30$. Middle: for $n = 50$. Bottom: for $n = 300$.



$n = 30$

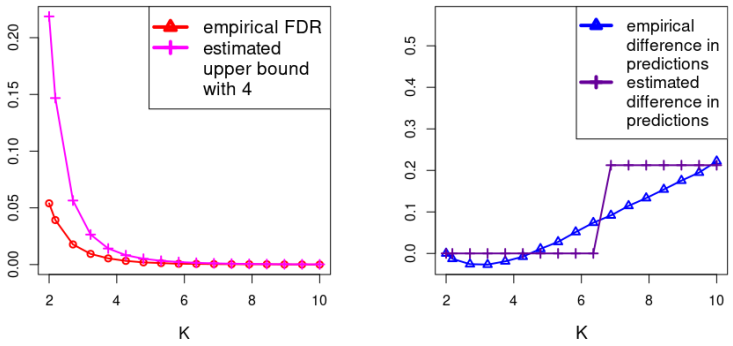


$n = 50$



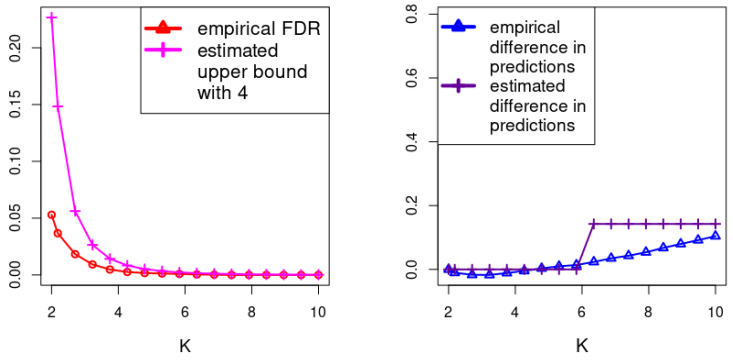
$n = 300$

Figure S-8: Curves of the relative change values between the functions $B(K, \beta^*, \sigma^2)$ and the functions $B(K, \hat{\beta}_{\hat{m}(\tilde{K})}, \hat{\sigma}^2)$ with respectively $\hat{\beta}_{\hat{m}(1)}$, $\hat{\beta}_{\hat{m}(1.5)}$, $\hat{\beta}_{\hat{m}(2)}$, $\hat{\beta}_{\hat{m}(2.5)}$, $\hat{\beta}_{\hat{m}(3)}$, $\hat{\beta}_{\hat{m}(3.5)}$, $\hat{\beta}_{\hat{m}(4)}$, $\hat{\beta}_{\hat{m}(4.5)}$, $\hat{\beta}_{\hat{m}(5)}$ and $\hat{\beta}_{\hat{m}(\log(n))}$ where estimators are calculating from a single data set. Top: for $n = 30$. Middle: for $n = 50$. Bottom: for $n = 300$.



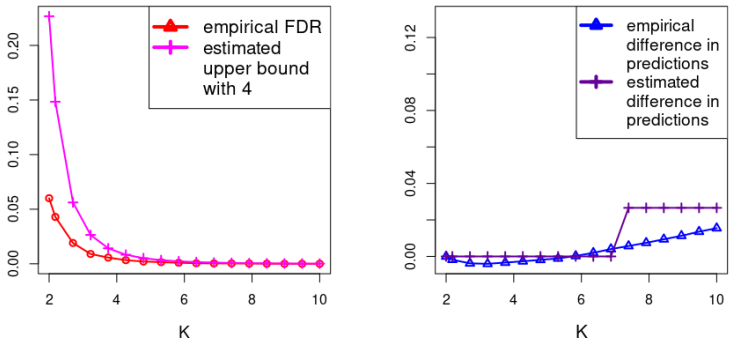
FDR diff-PR

$n = 30$



FDR diff-PR

$n = 50$

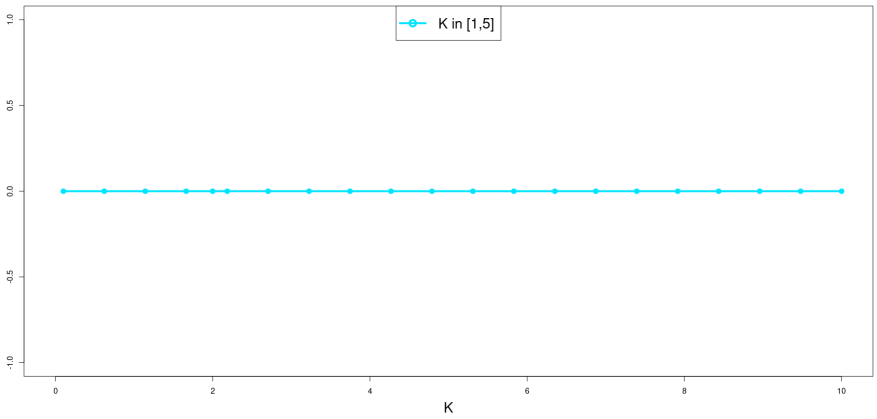


FDR diff-PR

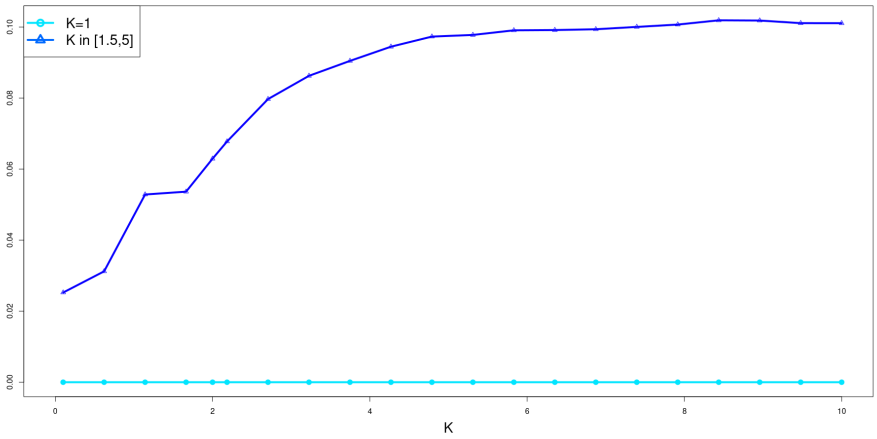
$n = 300$

Figure S-9: Curves of the empirical functions $\text{FDR}\big(\hat{m}(K)\big)$ (red) and $\text{diff-PR}\big(\hat{m}(K)\big)$ (blue), of the $B(K, \hat{\beta}_{\hat{m}(4)}, \hat{\sigma}^2)$ functions (pink) and of $\widehat{\text{diff-PR}}\big(\hat{m}(K)\big)$ (violet) for $K \geq 2$ for the *toy data set*. Top: for $n = 30$. Middle: for $n = 50$. Bottom: for $n = 300$.

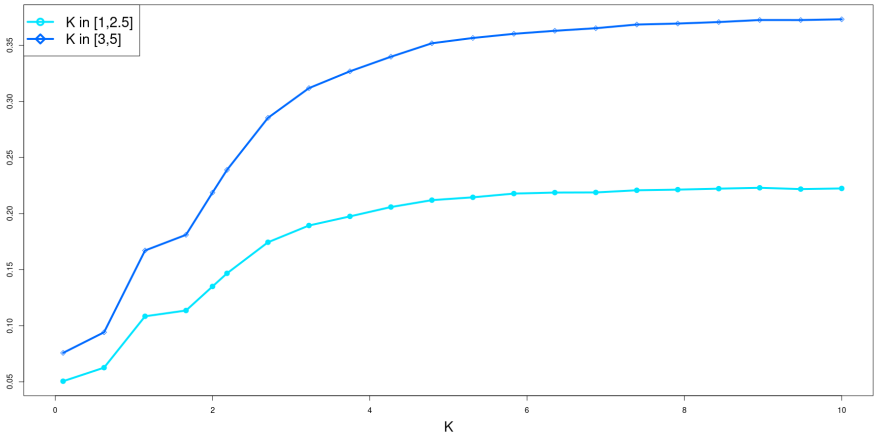
1.4 Scenario (iv)



$\sigma^2 = 0.1$

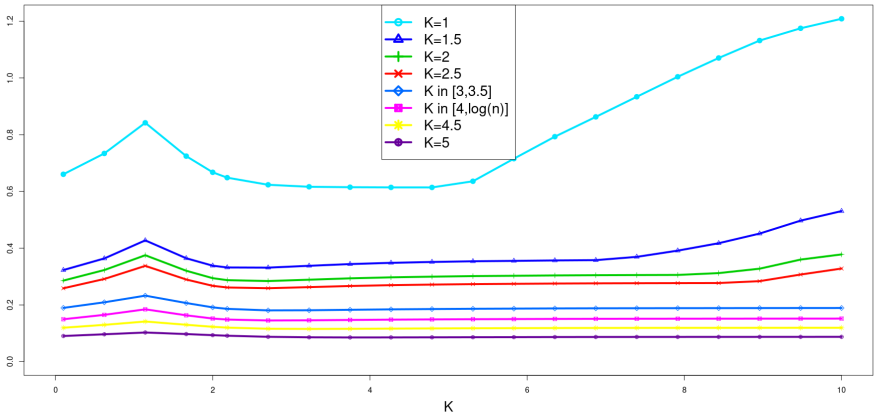


$\sigma^2 = 1$

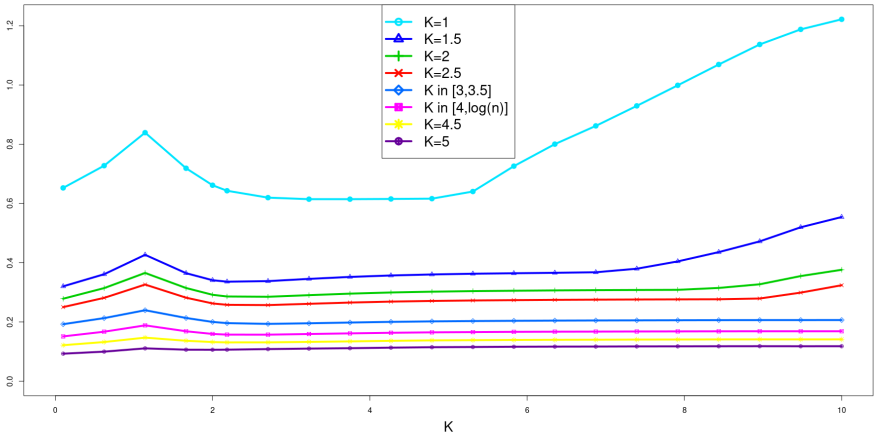


$\sigma^2 = 4$

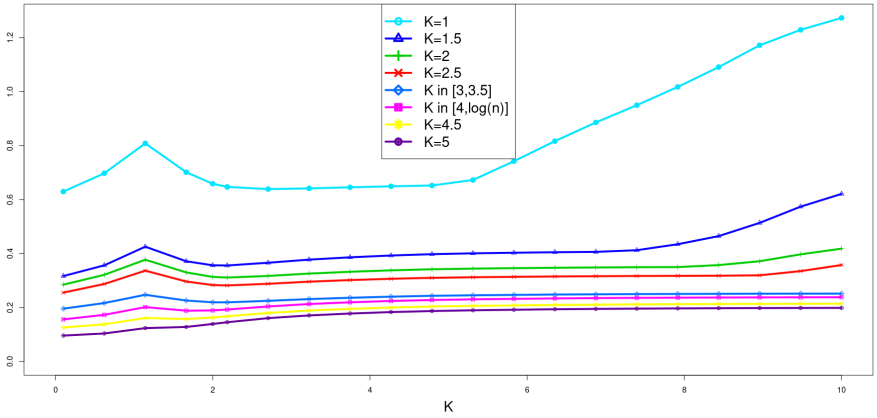
Figure S-10: Curves of the relative change values between the functions $B(K, \beta^*, \sigma^2)$ and the functions $B(K, \hat{\beta}_{\hat{m}(\tilde{K})}, \hat{\sigma}^2)$ with respectively $\hat{\beta}_{\hat{m}(1)}$, $\hat{\beta}_{\hat{m}(1.5)}$, $\hat{\beta}_{\hat{m}(2)}$, $\hat{\beta}_{\hat{m}(2.5)}$, $\hat{\beta}_{\hat{m}(3)}$, $\hat{\beta}_{\hat{m}(3.5)}$, $\hat{\beta}_{\hat{m}(4)}$, $\hat{\beta}_{\hat{m}(4.5)}$, $\hat{\beta}_{\hat{m}(5)}$ and $\hat{\beta}_{\hat{m}(\log(n))}$ where estimators are calculating from a single data set. Top: for $\sigma^2 = 0.1$. Middle: for $\sigma^2 = 1$. Bottom: for $\sigma^2 = 4$.



$\sigma^2 = 0.1$

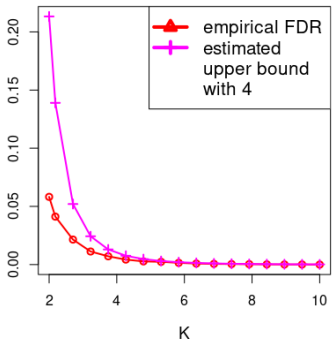


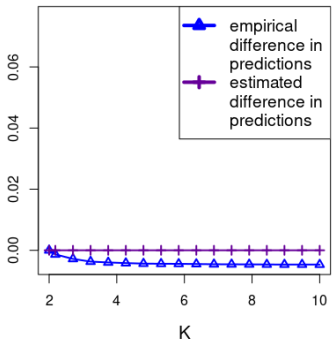
$\sigma^2 = 1$



$\sigma^2 = 4$

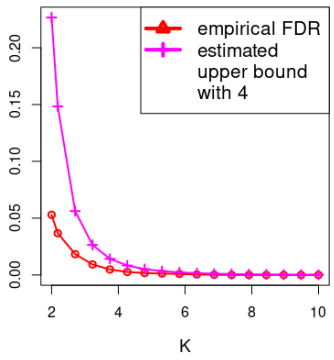
Figure S-11: Curves of the relative change values between the functions $B(K, \beta^*, \sigma^2)$ and the functions $B(K, \hat{\beta}_{\hat{m}(\tilde{K})}, \hat{\sigma}^2)$ with respectively $\hat{\beta}_{\hat{m}(1)}$, $\hat{\beta}_{\hat{m}(1.5)}$, $\hat{\beta}_{\hat{m}(2)}$, $\hat{\beta}_{\hat{m}(2.5)}$, $\hat{\beta}_{\hat{m}(3)}$, $\hat{\beta}_{\hat{m}(3.5)}$, $\hat{\beta}_{\hat{m}(4)}$, $\hat{\beta}_{\hat{m}(4.5)}$, $\hat{\beta}_{\hat{m}(5)}$ and $\hat{\beta}_{\hat{m}(\log(n))}$ where estimators are calculating from a single data set. Top: for $\sigma^2 = 0.1$. Middle: for $\sigma^2 = 1$. Bottom: for $\sigma^2 = 4$.

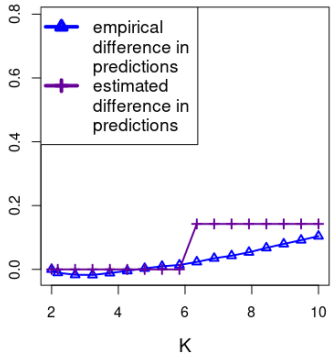




FDR diff-PR

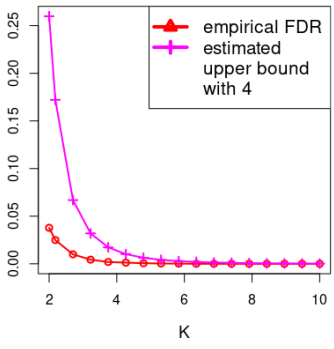
$\sigma^2 = 0.1$

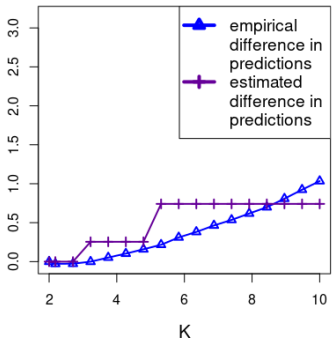




FDR diff-PR

$\sigma^2 = 1$





FDR diff-PR

$\sigma^2 = 4$

Figure S-12: Curves of the empirical functions $\mathrm{FDR}\big(\hat{m}(K)\big)$ (red) and $\text{diff-PR}\big(\hat{m}(K)\big)$ (blue), of the $B(K, \hat{\beta}_{\hat{m}(4)}, \hat{\sigma}^2)$ functions (pink) and of $\widehat{\text{diff-PR}}\big(\hat{m}(K)\big)$ (violet) for $K \geq 2$ for the *toy data set*. Top: for $\sigma^2 = 0.1$. Middle: for $\sigma^2 = 1$. Bottom: for $\sigma^2 = 4$.

2 Non-ordering variable selection

2.1 Robustness to variable ordering

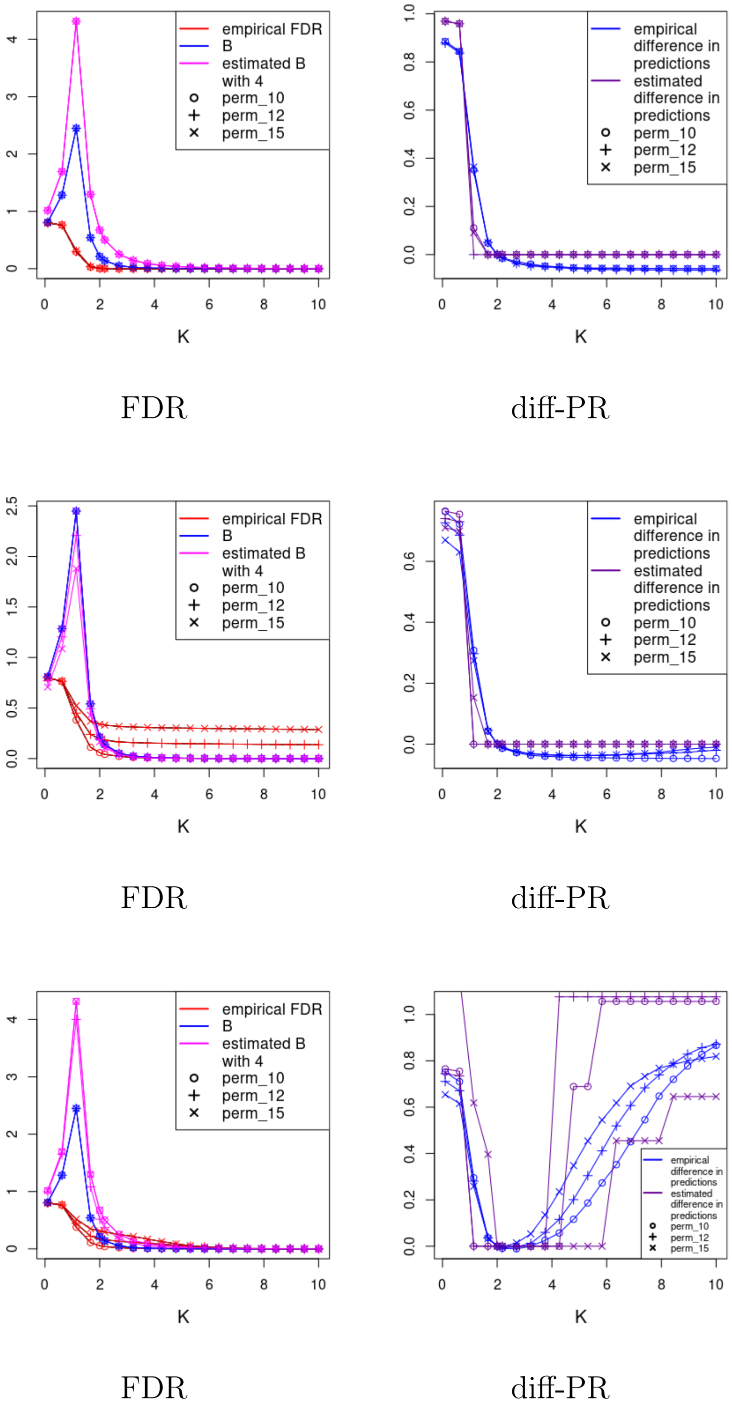
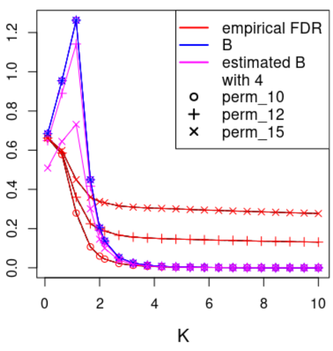
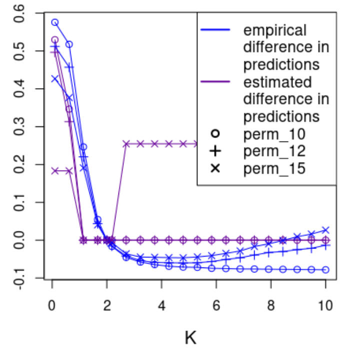


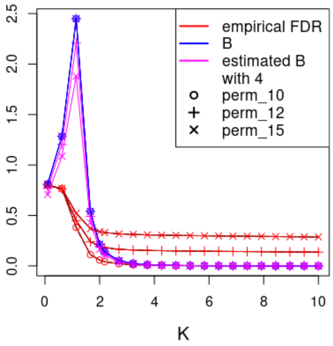
Figure S-13: Curves of the empirical functions $\text{FDR}\big(\hat{m}(K)\big)$ (red) and diff-PR$\big(\hat{m}(K)\big)$ (blue), of the $B(K, \beta^*, \sigma^2)$ functions (blue), the $B(K, \hat{\beta}_{\hat{m}(4)}, \hat{\sigma}^2)$ functions (pink) and $\widehat{\text{diff-PR}}\big(\hat{m}(K)\big)$ (violet) for the *toy data set* and for the three perturbed collections. Top: for $\beta^*_{10} = \frac{2}{10}$. Middle: for $\beta^*_{10} = 2$ and distant coefficients. Bottom: for $\beta^*_{10} = 2$ and close coefficients.



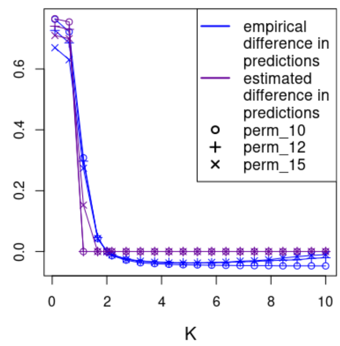
FDR



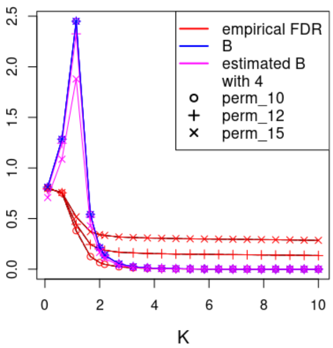
diff-PR



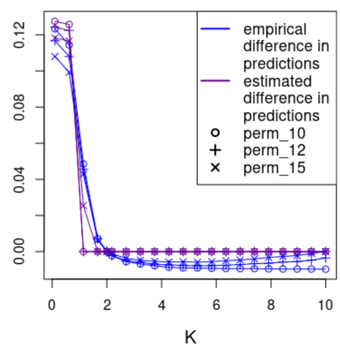
FDR



diff-PR

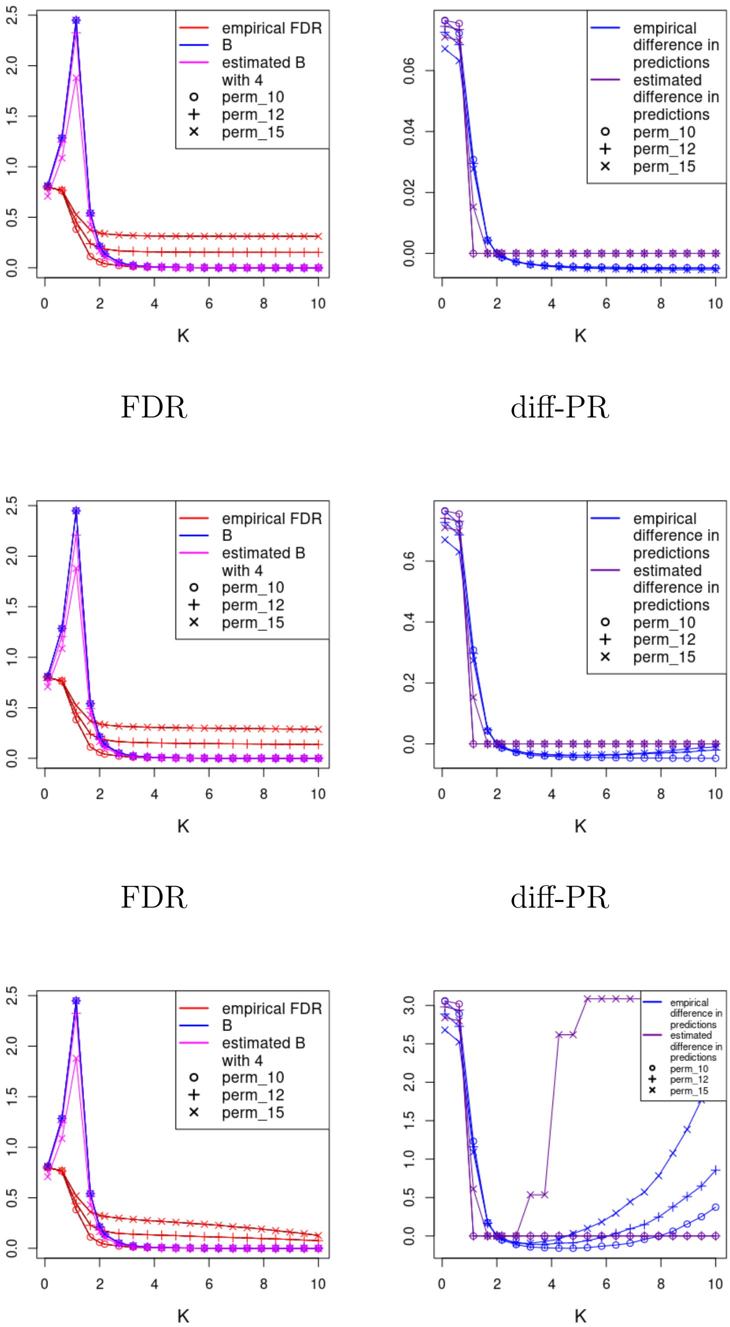


FDR



diff-PR

Figure S-14: Curves of the empirical functions $\text{FDR}\big(\hat{m}(K)\big)$ (red) and diff-$\text{PR}\big(\hat{m}(K)\big)$ (blue), of the $B(K, \beta^*, \sigma^2)$ functions (blue), the $B(K, \hat{\beta}_{\hat{m}(4)}, \hat{\sigma}^2)$ functions (pink) and $\widehat{\text{diff-PR}}\big(\hat{m}(K)\big)$ (violet) for the *toy data set* and for the three perturbed collections. Top: for $n = 30$. Middle: for $n = 50$. Bottom: for $n = 300$.



FDR diff-PR

FDR diff-PR

FDR diff-PR

Figure S-15: Curves of the empirical functions $\mathrm{FDR}\big(\hat{m}(K)\big)$ (red) and diff-PR$\big(\hat{m}(K)\big)$ (blue), of the $B(K, \beta^*, \sigma^2)$ functions (blue), the $B(K, \hat{\beta}_{\hat{m}(4)}, \hat{\sigma}^2)$ functions (pink) and $\widehat{\text{diff-PR}}\big(\hat{m}(K)\big)$ (violet) for the *toy data set* and for the three perturbed collections. Top: for $\sigma^2 = 0.1$. Middle: for $\sigma^2 = 1$. Bottom: for $\sigma^2 = 4$.

2.2 Random variable order

	Bolasso	SLOPE	random forests	the knockoff method
Scenario (ii) config. (ii)				
$D_m = 5$	0.25	0.24	0.24	0.26
$D_m = 10$	0.23	0.23	0.23	0.24
$D_m = 15$	0.34	0.34	0.34	0.34
$D_m = 20$	0.44	0.44	0.44	0.43
Scenario (ii) config. (iii)				
$D_m = 5$	0.63	0.60	0.63	0.74
$D_m = 10$	0.54	0.52	0.53	0.58
$D_m = 15$	0.69	0.68	0.69	0.69
$D_m = 20$	0.79	0.78	0.79	0.75
Scenario (iii), $n = 30$				
$D_m = 5$	0.96	0.95	0.95	*the knockoff method*
$D_m = 10$	1.00	1.00	1.00	*is not adapted*
$D_m = 15$	1.00	1.00	1.00	*to the* $n < p$ *case*
$D_m = 20$	1.00	1.00	1.00	
Scenario (iii), $n = 300$				
$D_m = 5$	0.98	0.91	0.92	1.00
$D_m = 10$	0.83	0.73	0.78	0.92
$D_m = 15$	0.92	0.86	0.91	0.96
$D_m = 20$	0.96	0.92	0.96	0.97
Scenario (iv), $\sigma^2 = 0.1$				
$D_m = 5$	1.00	1.00	1.00	1.00
$D_m = 10$	0.99	0.99	0.94	0.98
$D_m = 15$	1.00	1.00	0.98	0.98
$D_m = 20$	1.00	1.00	0.99	0.98
Scenario (iv), $\sigma^2 = 4$				
$D_m = 5$	0.86	0.85	0.82	0.96
$D_m = 10$	0.66	0.66	0.65	0.70
$D_m = 15$	0.78	0.77	0.77	0.78
$D_m = 20$	0.85	0.84	0.84	0.82

Table S-1: Active variable proportions in models of size 5, 10, 15 and 20 for random collections built with Bolasso, SLOPE, random forest and the knockoff method for scenarios (ii)-(iv) of Table 3. Values are the average over 100 iterations.

2.3 Comparison with other variable selection methods

	$D_{\hat{m}}$	PR($\hat{m}$)	FDR($\hat{m}$)
Scenario (ii) config. (ii)			
LinSelect	0.00	1.06	0.00
50-fold CV	23.15	1.47	0.45
Our algorithm	0.27	1.08	0.00
Scenario (ii) config. (iii)			
LinSelect	0.03	2.21	0.00
50-fold CV	18.65	1.78	0.33
Our algorithm	7.55	1.39	0.00
Scenario (iii), $n = 30$			
LinSelect	2.00	14.41	0.00
50-fold CV	12.00	3.57	0.20
Our algorithm	9.00	1.43	0.01
Scenario (iii), $n = 300$			
LinSelect	2.07	14.41	0.00
50-fold CV	11.81	3.57	0.20
Our algorithm	9.38	1.43	0.01
Scenario (iv), $\sigma^2 = 0.1$			
LinSelect	10.27	0.12	0.02
50-fold CV	28.18	0.35	0.47
Our algorithm	10.07	0.12	0.00
Scenario (iv), $\sigma^2 = 4$			
LinSelect	3.37	10.94	0.00
50-fold CV	25.13	7.74	0.44
Our algorithm	7.85	5.12	0.00

Table S-2: Results of the dimension, PR and FDR of the selected models obtained by LinSelect, the 50-fold CV and our algorithm, applied on the nested model collection (2.1) for the scenarios (ii), (ii) and (iv) described in Table 3. Each value is the average over 100 independent iterations. PR and FDR of each selected model are the empirical quantities. Input parameters of our algorithm are fixed to $\gamma = 0.1$ and $\alpha = 0.05$.

	$D_{\hat{m}}$	PR($\hat{m}$)	FDR($\hat{m}$)
Scenario (ii) config. (ii)			
LinSelect	0.02	1.06	0.00
50-fold CV	23.51	1.71	0.44
Our algorithm	6.80	1.50	0.05
Scenario (ii) config. (iii)			
LinSelect	0.06	2.22	0.00
50-fold CV	24.83	1.98	0.47
Our algorithm	13.86	1.85	0.25
Scenario (iii), $n = 30$			
LinSelect	1.47	15.89	0.00
50-fold CV	14.66	3.66	0.28
Our algorithm	12.65	1.79	0.19
Scenario (iii), $n = 300$			
LinSelect	11.58	1.17	0.13
50-fold CV	21.68	1.34	0.40
Our algorithm	15.20	1.11	0.30
Scenario (iv), $\sigma^2 = 0.1$			
LinSelect	11.24	0.14	0.08
50-fold CV	27.58	0.47	0.46
Our algorithm	13.60	0.15	0.24
Scenario (iv), $\sigma^2 = 4$			
LinSelect	3.44	12.91	0.01
50-fold CV	24.23	8.53	0.44
Our algorithm	13.93	6.73	0.25

Table S-3: Results of the dimension, PR and FDR of the selected models obtained by LinSelect, the 50-fold CV and our algorithm, applied on the random collections built with Bolasso for the scenarios (ii), (ii) and (iv) described in Table 3. Each value is the average over 100 independent iterations. PR and FDR of each selected model are the empirical quantities. Input parameters of our algorithm are fixed to $\gamma = 0.1$ and $\alpha = 0.05$.

	$D_{\hat{m}}$	PR($\hat{m}$)	FDR($\hat{m}$)
Scenario (ii) config. (ii)			
LinSelect	0.04	1.07	0.00
50-fold CV	24.97	1.70	0.45
Knockoff	0.00	1.06	0.00
Our algorithm	4.31	1.43	0.00
Scenario (ii) config. (iii)			
LinSelect	0.07	2.22	0.00
50-fold CV	24.25	1.99	0.43
Knockoff	0.00	2.22	0.00
Our algorithm	9.42	1.84	0.07
Scenario (iii) $n = 30$			
LinSelect			*the knockoff method*
50-fold CV			*is not adapted*
Knockoff			*to the $n < p$ case*
Our algorithm			
Scenario (iii), $n = 300$			
LinSelect	9.84	1.07	0.03
50-fold CV	23.21	1.21	0.44
Knockoff	0.21	2.65	0.01
Our algorithm	13.39	1.10	0.23
Scenario (iv), $\sigma^2 = 0.1$			
LinSelect	10.07	0.31	0.04
50-fold CV	23.64	0.47	0.38
Knockoff	0.00	13.18	0.00
Our algorithm	21.02	0.16	0.32
Scenario (iv), $\sigma^2 = 4$			
LinSelect	4.44	10.17	0.00
50-fold CV	21.12	7.91	0.38
Knockoff	0.00	17.12	0.00
Our algorithm	10.58	6.62	0.10

Table S-4: Results of the dimension, PR and FDR of the selected models obtained by LinSelect, the 50-fold CV, the knockoff method and our algorithm, applied on the random collections built with the knockoff method for the scenarios (ii), (ii) and (iv) described in Table 3. Each value is the average over 100 independent iterations. PR and FDR of each selected model are the empirical quantities. Input parameters of our algorithm are fixed to $\gamma = 0.1$ and $\alpha = 0.05$.