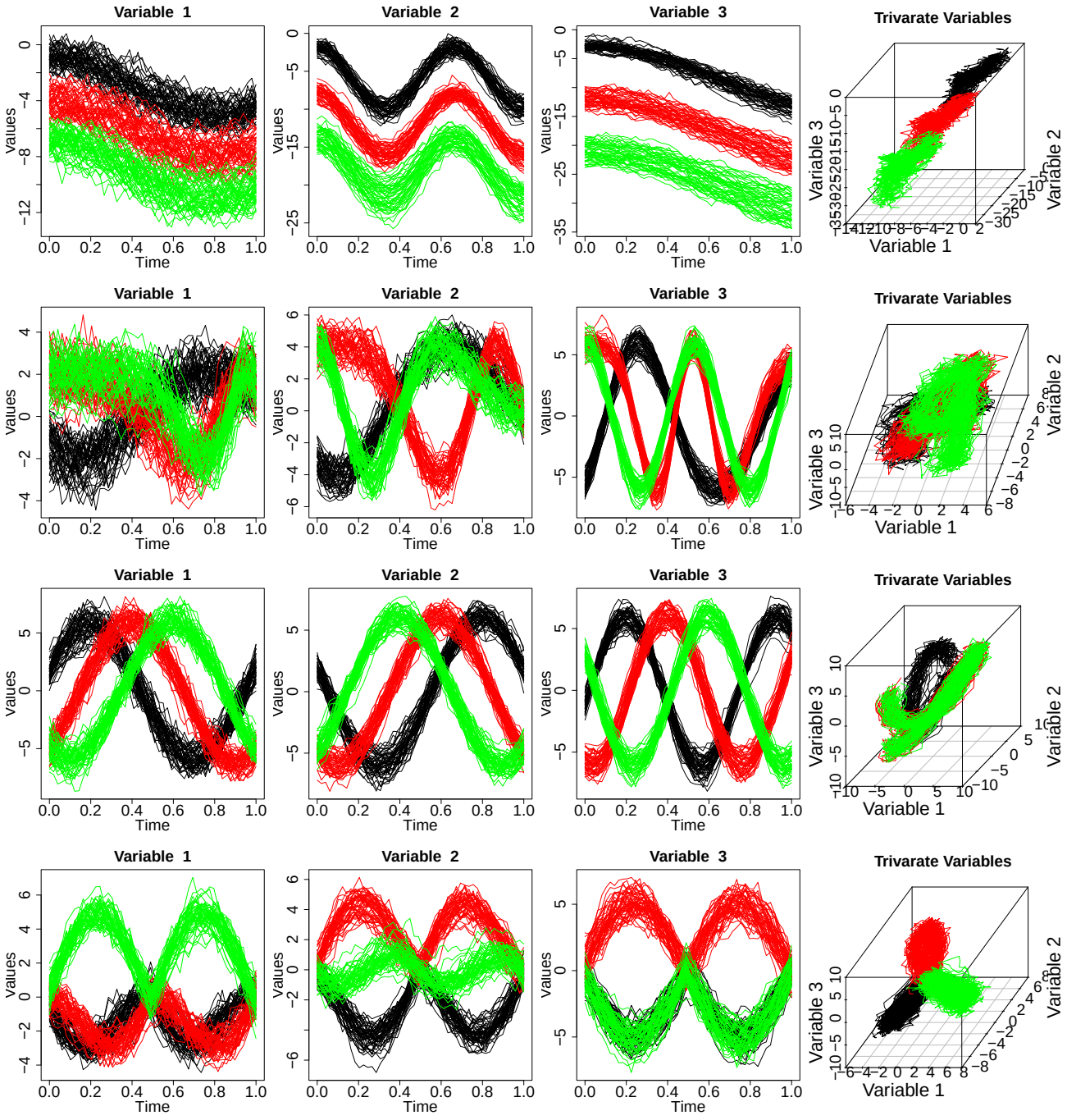
Supplementary Material: Robust Two-Layer Partition Clustering of Sparse Multivariate Functional Data

Zhuo Qu[1], Wenlin Dai[2] and Marc G. Genton[1]

March 15, 2022

This supplementary material provides the results of Scenarios 1-6 in the Simulation Studies. First, we present one simulation of all the above scenarios without the outlier and sparseness corruption. Next, we display the performance of clustering for Scenarios 1-6. Then, we show the performance of outlier detection for Scenarios 1-6.

[1]Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.
E-mail: zhuo.qu@kaust.edu.sa, marc.genton@kaust.edu.sa
This research was supported by the King Abdullah University of Science and Technology (KAUST).
[2]Institute of Statistics and Big Data, Renmin University of China, Beijing 100872, China.
E-mail: wenlin.dai@ruc.edu.cn


Variable 1
Variable 2
Variable 3
Trivarate Variables
Values
Time
Variable 1
Variable 2
Variable 3

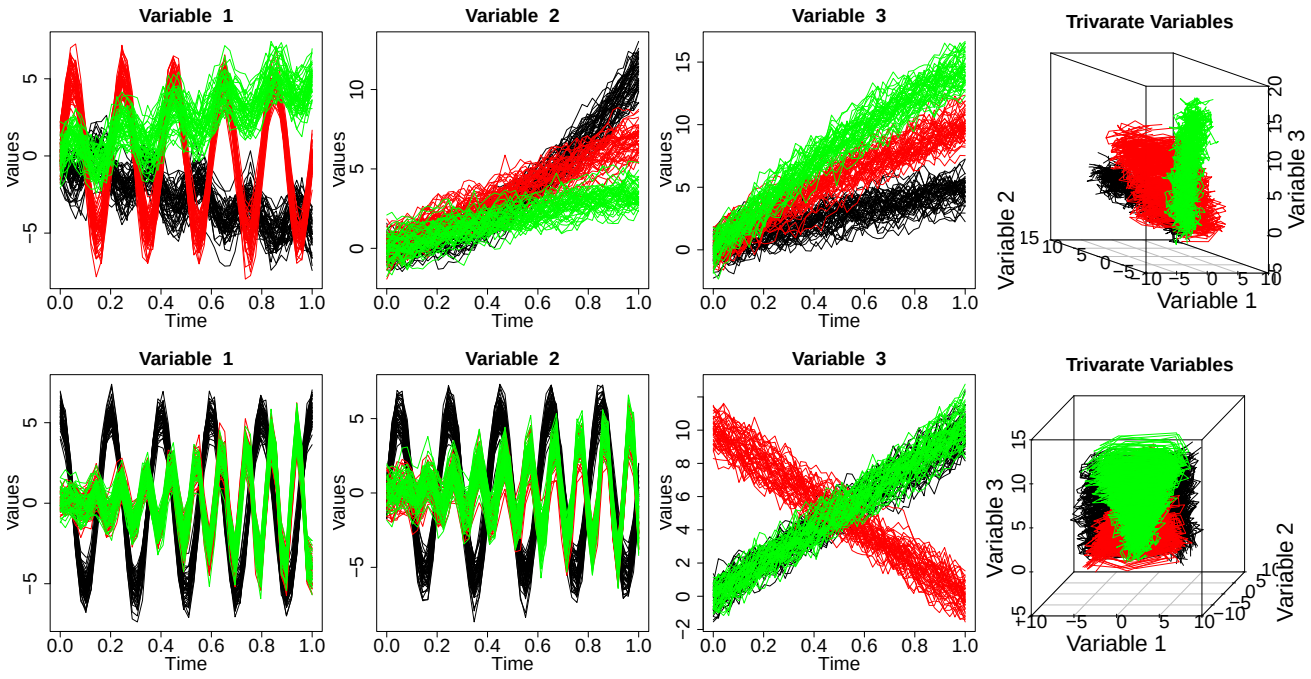


Figure S1: Rows from top to down represent samples from Scenario 1-6, respectively. Three clusters are represented in black, red, and green.

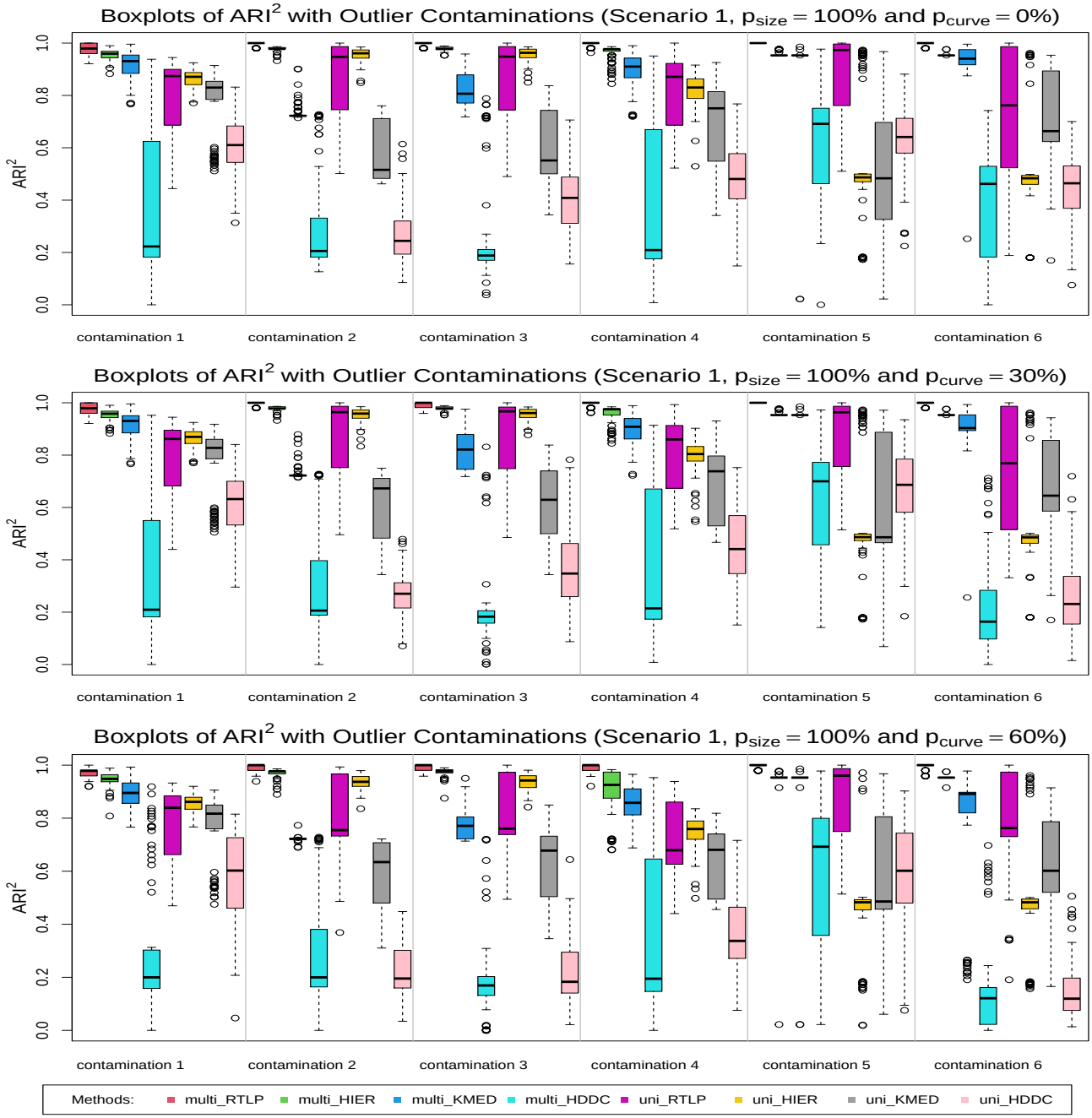


Figure S2: Panels from the top to bottom display the boxplots of ARI^2 in Scenario 1 under $p_{curve} = 0, 30\%$ and 60%. Eight methods are compared in all settings with six contaminations. Here, $K = 3$ and 100 simulation replicates. The methods from left to right are the multivariate and average marginal univariate versions of RTLP, agglomerative hierarchical, K-medoids, and funHDDC methods.

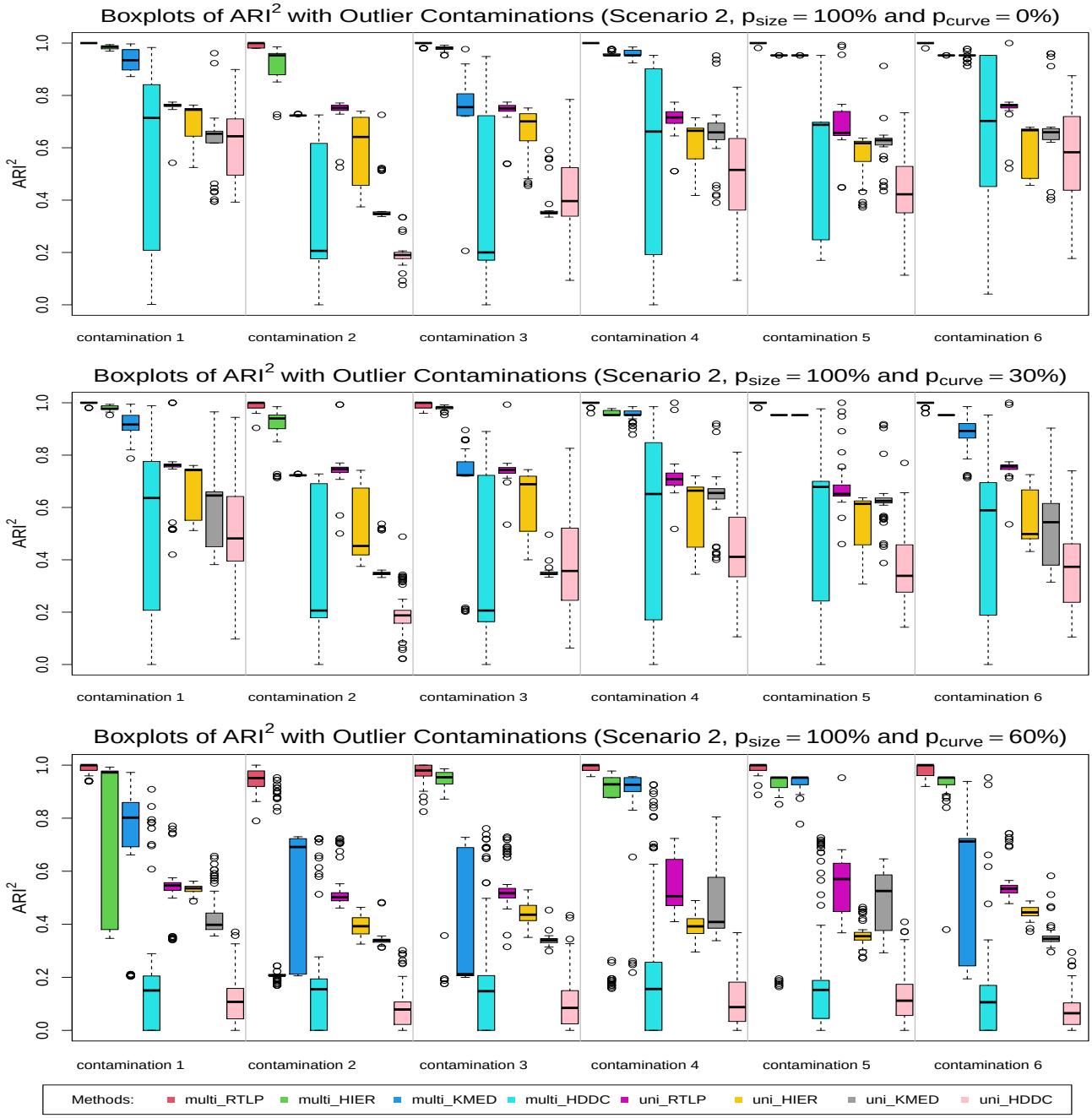


Figure S3: Panels from the top to bottom display the boxplots of ARI^2 in Scenario 2 under $p_{curve} = 0, 30\%$ and 60%. Eight methods are compared in all settings with six contaminations. Here, $K = 3$ and 100 simulation replicates. The methods from left to right are the multivariate and average marginal univariate versions of RTLP, agglomerative hierarchical, K-medoids, and funHDDC methods.

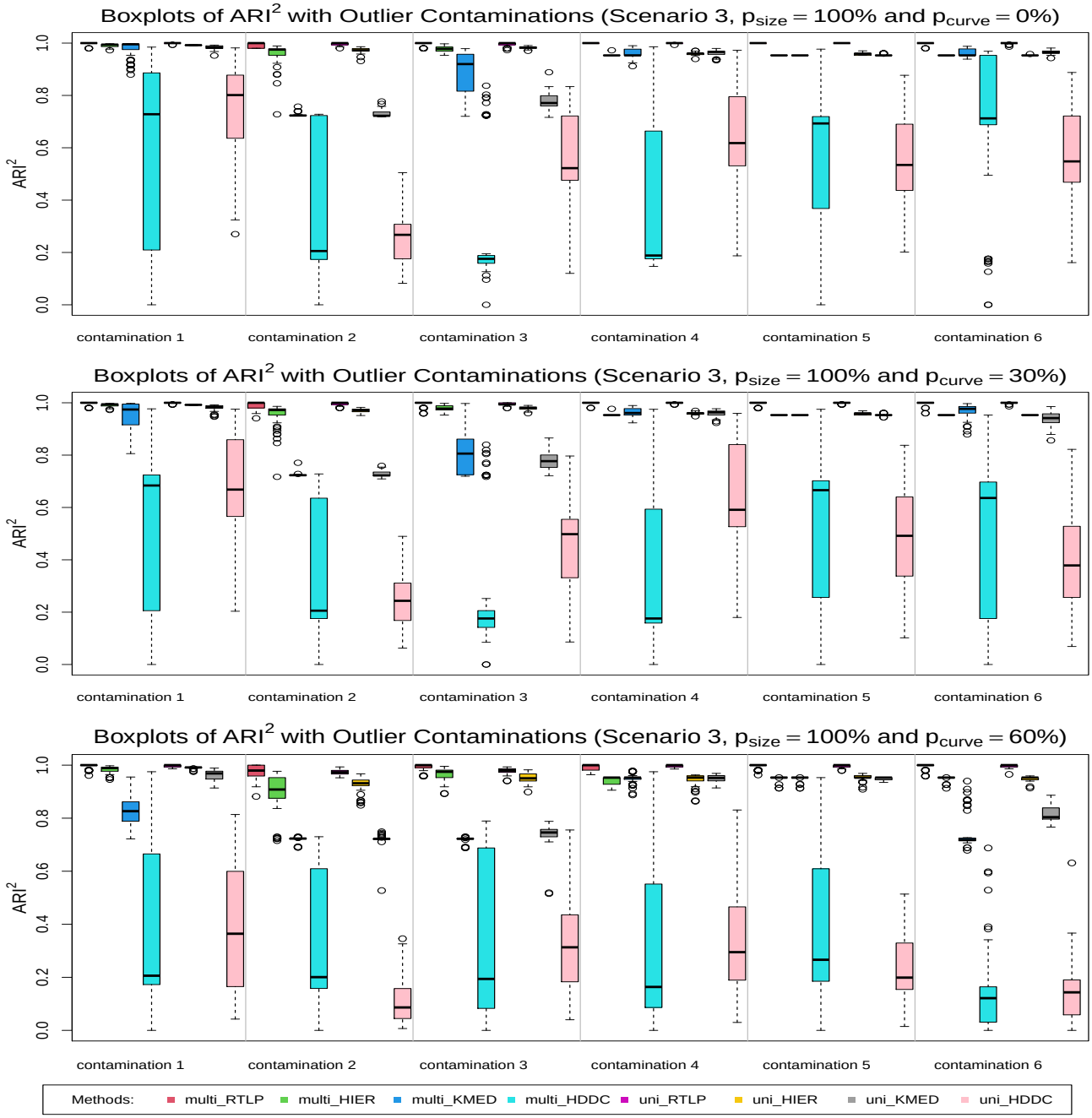


Figure S4: Panels from the top to bottom display the boxplots of ARI^2 in Scenario 3 under $p_{curve} = 0, 30\%$ and 60%. Eight methods are compared in all settings with six contaminations. Here, $K = 3$ and 100 simulation replicates. The methods from left to right are the multivariate and average marginal univariate versions of RTLP, agglomerative hierarchical, K-medoids, and funHDDC methods.

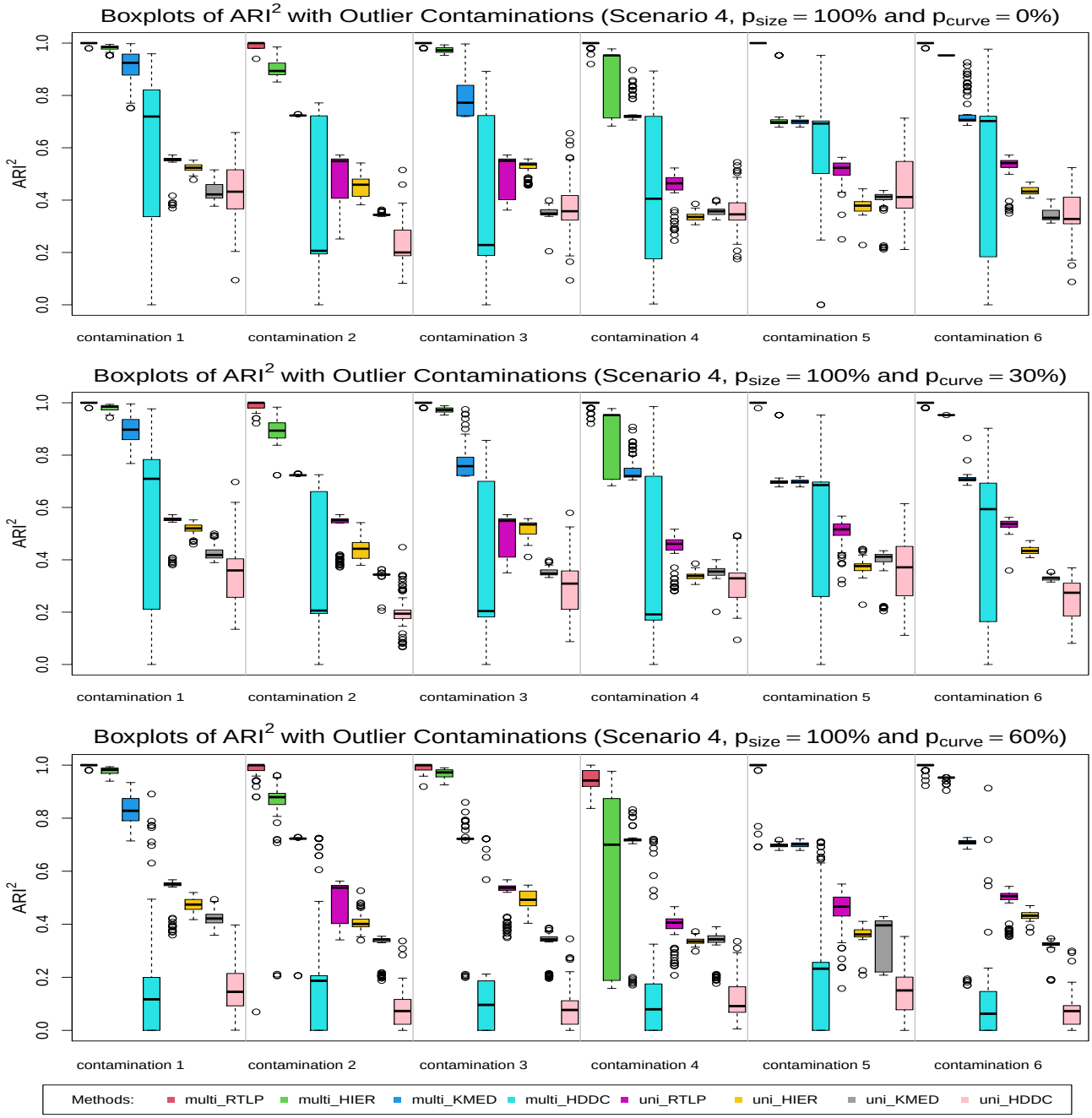


Figure S5: Panels from the top to bottom display the boxplots of ARI^2 in Scenario 4 under $p_{curve} = 0, 30\%$ and 60%. Eight methods are compared in all settings with six contaminations. Here, $K = 3$ and 100 simulation replicates. The methods from left to right are the multivariate and average marginal univariate versions of RTLP, agglomerative hierarchical, K-medoids, and funHDDC methods.

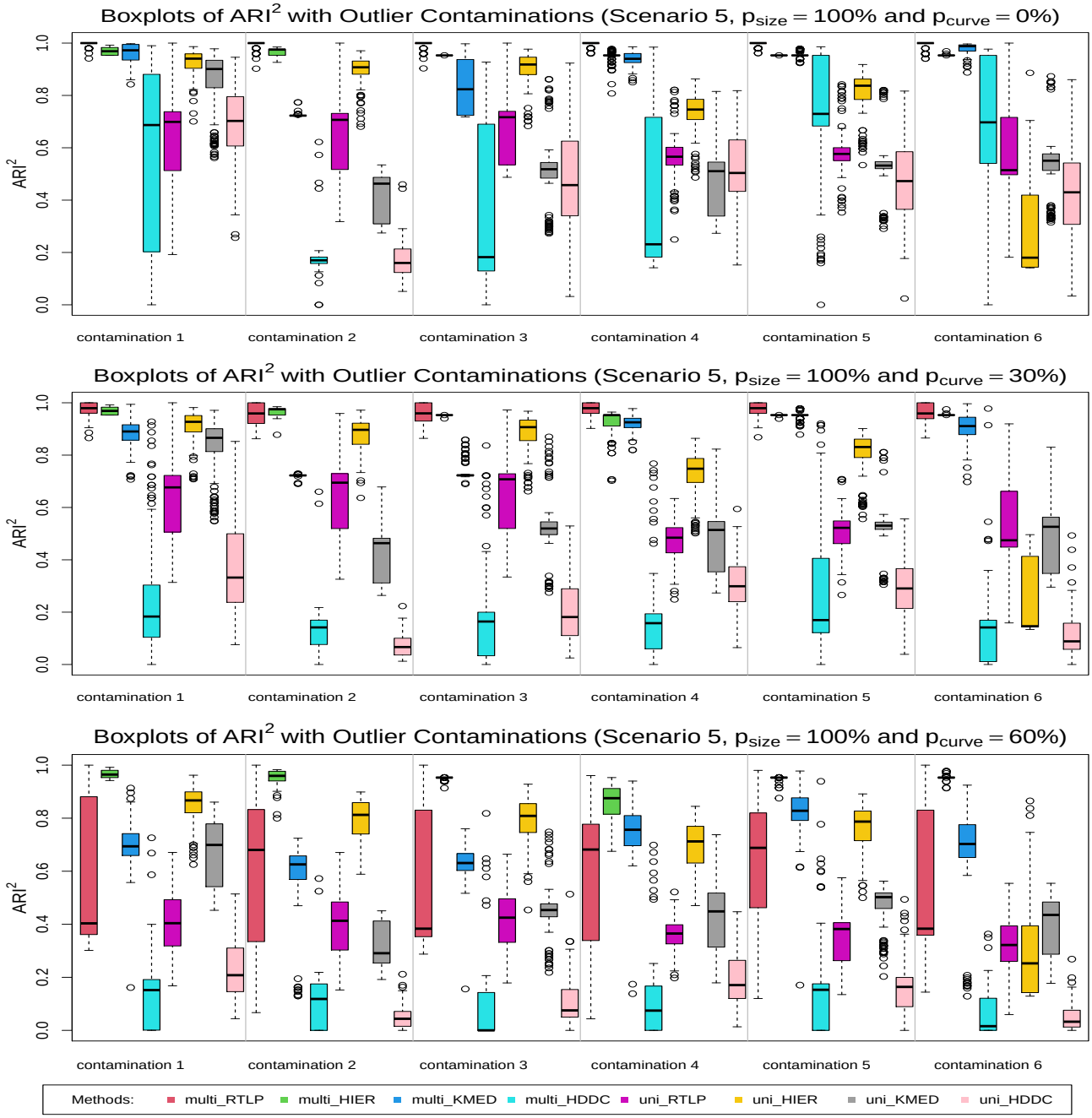


Figure S6: Panels from the top to bottom display the boxplots of ARI^2 in Scenario 5 under $p_{curve} = 0, 30\%$ and 60%. Eight methods are compared in all settings with six contaminations. Here, $K = 3$ and 100 simulation replicates. The methods from left to right are the multivariate and average marginal univariate versions of RTLP, agglomerative hierarchical, K-medoids, and funHDDC methods.

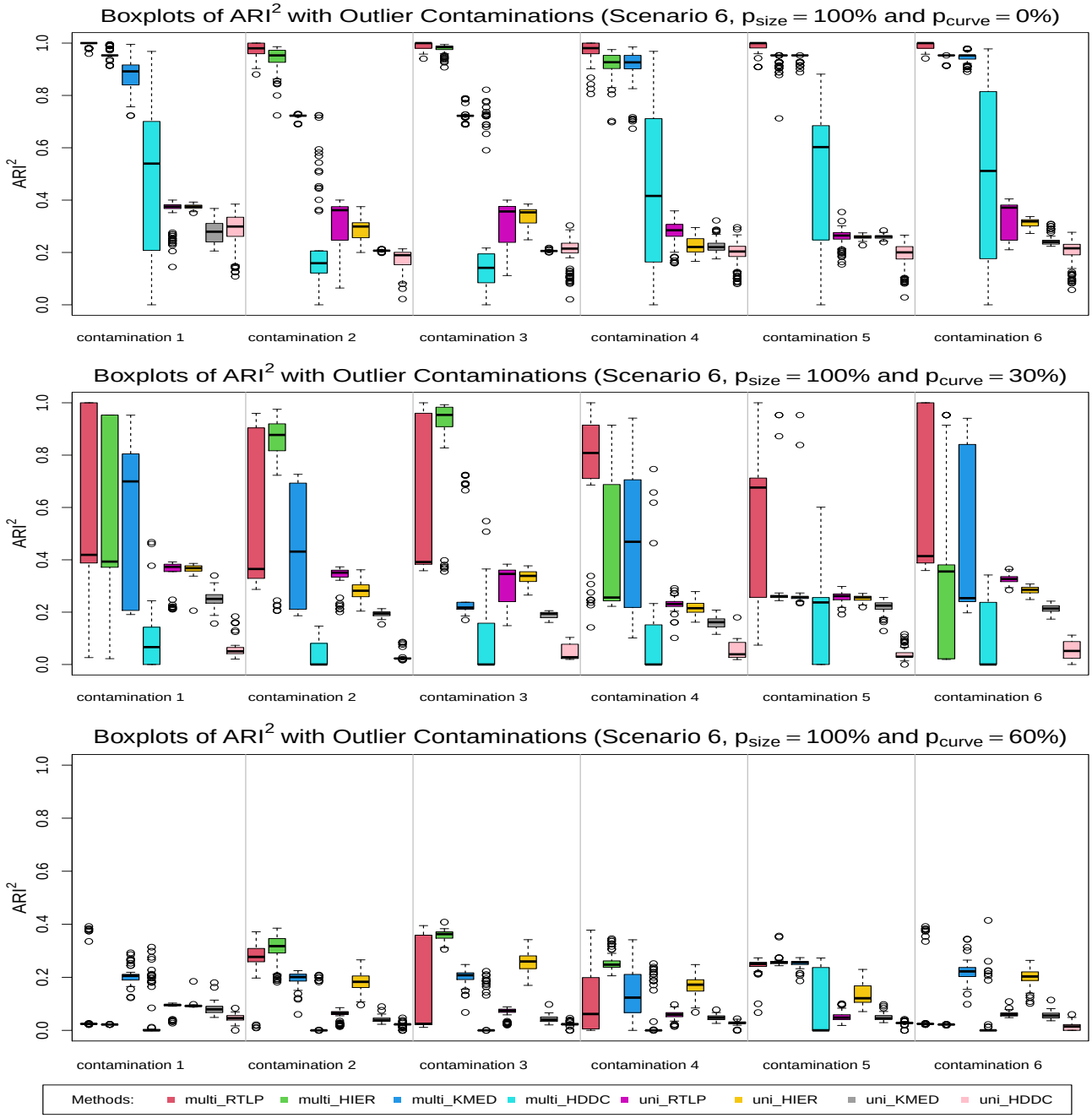


Figure S7: Panels from the top to bottom display the boxplots of ARI^2 in Scenario 6 under $p_{curve} = 0, 30\%$ and 60%. Eight methods are compared in all settings with six contaminations. Here, $K = 3$ and 100 simulation replicates. The methods from left to right are the multivariate and average marginal univariate versions of RTLP, agglomerative hierarchical, K-medoids, and funHDDC methods.

Table S1: Correct outlier detection percentage p_c (%) and false outlier detection percentage p_f (%) of multi_RTLP and uni_RTLP in different outlier contaminations for Scenarios 1-6. We name multi_RTLP (uni_RTLP) multivariate (univariate) in the RTLP version and write the higher p_c and the lower p_f in bold in each setting given the contamination and p_{curve}. The proportion of outliers is 10%, and $p_{size} = 100\%$ in all settings. Simulations with 100 replicates.

(a) Scenario 1

p_{curve}	RTLP version	Contamination 1		Contamination 2		Contamination 3	
		p_c	p_f	p_c	p_f	p_c	p_f
0%	multivariate	**92.3** (6.3)	**0.0** (0.0)	**98.8** (2.6)	**0.0** (0.0)	**98.6** (2.7)	**0.0** (0.0)
	univariate	66.9 (8.0)	0.5 (2.0)	97.6 (2.3)	0.1 (0.2)	97.7 (2.8)	0.3 (1.3)
30%	multivariate	**92.4** (6.4)	**0.0** (0.0)	**98.7** (2.6)	**0.0** (0.0)	**98.3** (2.9)	**0.0** (0.1)
	univariate	67.3 (8.2)	0.4 (1.7)	97.9 (2.3)	2.3 (1.0)	97.3 (2.6)	0.2 (1.1)
60%	multivariate	**91.9** (6.4)	**0.1** (0.3)	**97.5** (3.2)	**0.1** (0.3)	**97.9** (3.1)	**0.1** (0.2)
	univariate	66.4 (7.8)	0.3 (1.3)	94.0 (3.8)	0.4 (1.4)	95.5 (3.2)	0.2 (0.3)
p_{curve}	RTLP version	Contamination 4		Contamination 5		Contamination 6	
		p_c	p_f	p_c	p_f	p_c	p_f
0%	multivariate	**98.4** (3.0)	**0.0** (0.0)	**100.0** (0.0)	**0.0** (0.0)	**99.3** (2.1)	**0.0** (0.0)
	univariate	77.0 (9.1)	1.0 (2.9)	98.7 (5.1)	0.1 (0.2)	99.1 (1.7)	0.6 (1.0)
30%	multivariate	**98.5** (2.9)	**0.0** (0.1)	**100.0** (0.0)	**0.0** (0.0)	**99.2** (0.0)	**0.0** (0.0)
	univariate	75.2 (8.4)	0.8 (2.5)	99.4 (2.1)	0.1 (0.2)	99.0 (1.8)	0.6 (0.9)
60%	multivariate	**97.4** (4.9)	**0.1** (0.3)	**100.0** (0.0)	**0.1** (0.3)	**99.4** (2.0)	**0.0** (0.1)
	univariate	57.3 (14.3)	0.5 (2.0)	99.7 (1.2)	0.1 (0.3)	99.1 (1.4)	0.4 (0.8)

(b) Scenario 2

p_{curve}	RTLP version	Contamination 1		Contamination 2		Contamination 3	
		p_c	p_f	p_c	p_f	p_c	p_f
0%	multivariate	**100.0** (0.0)	**0.0** (0.0)	**98.1** (3.1)	**0.0** (0.0)	**98.5** (2.8)	**0.0** (0.0)
	univariate	**100.0** (0.0)	0.0 (0.2)	96.6 (3.0)	0.0 (0.2)	95.6 (4.7)	0.0 (0.2)
30%	multivariate	99.9 (0.9)	**0.0** (0.1)	**97.0** (4.0)	**0.1** (0.3)	**97.6** (3.2)	**0.1** (0.2)
	univariate	**99.9** (0.7)	0.4 (1.5)	94.2 (4.0)	0.3 (1.5)	93.8 (4.5)	0.1 (0.2)
60%	multivariate	99.5 (0.6)	**0.3** (0.1)	**87.2** (7.6)	0.8 (0.9)	**93.7** (3.1)	**0.5** (0.2)
	univariate	**99.9** (0.8)	0.4 (0.5)	81.1 (7.0)	**0.5** (0.6)	84.4 (6.6)	0.6 (1.3)
p_{curve}	RTLP version	Contamination 4		Contamination 5		Contamination 6	
		p_c	p_f	p_c	p_f	p_c	p_f
0%	multivariate	**100.0** (0.0)	**0.0** (0.0)	**100.0** (0.0)	**0.0** (0.4)	**99.8** (0.0)	**0.0** (0.1)
	univariate	86.0 (9.2)	0.0 (0.2)	74.8 (13.7)	0.1 (0.4)	98.1 (4.2)	**0.0** (0.1)
30%	multivariate	**99.9** (0.9)	**0.1** (0.2)	**100.0** (0.0)	**0.0** (0.1)	**98.9** (2.5)	**0.0** (0.1)
	univariate	84.9 (7.7)	0.2 (0.3)	72.6 (13.1)	0.4 (0.5)	96.7 (4.5)	**0.0** (0.1)
60%	multivariate	**99.9** (0.8)	**0.2** (0.4)	**98.6** (11.6)	**0.3** (0.5)	**97.4** (3.6)	**0.2** (0.4)
	univariate	74.6 (12.5)	0.5 (0.6)	59.4 (16.6)	0.9 (0.7)	91.2 (6.1)	0.4 (0.5)

(c) Scenario 3

p_{curve}	RTLP version	Contamination 1		Contamination 2		Contamination 3	
		p_c	p_f	p_c	p_f	p_c	p_f
0%	multivariate	99.7 (1.5)	**0.0** (0.0)	97.5 (3.3)	**0.0** (0.0)	**99.2** (2.2)	**0.0** (0.0)
	univariate	**100.0** (0.0)	**0.0** (0.0)	**99.0** (1.4)	**0.0** (0.0)	98.7 (1.7)	0.0 (0.1)
30%	multivariate	99.8 (1.2)	0.0 (0.2)	**97.0** (3.5)	0.0 (0.2)	98.6 (2.8)	0.0 (0.2)
	univariate	**100.0** (0.0)	**0.0** (0.1)	94.4 (1.7)	**0.0** (0.1)	**98.8** (1.5)	**0.0** (0.0)
60%	multivariate	**100.0** (0.0)	0.2 (0.4)	**94.4** (4.8)	0.4 (0.5)	**98.8** (2.6)	**0.1** (0.2)
	univariate	**100.0** (0.0)	**0.1** (0.2)	92.9 (3.3)	**0.1** (0.2)	93.5 (3.6)	**0.1** (0.2)
p_{curve}	RTLP version	Contamination 4		Contamination 5		Contamination 6	
		p_c	p_f	p_c	p_f	p_c	p_f
0%	multivariate	**100.0** (0.0)	**0.0** (0.0)	**100.0** (0.0)	**0.0** (0.0)	99.7 (1.5)	**0.0** (0.0)
	univariate	**100.0** (0.0)	**0.0** (0.0)	**100.0** (0.0)	**0.0** (0.0)	**99.9** (0.4)	**0.0** (0.0)
30%	multivariate	**100.0** (0.0)	**0.0** (0.1)	**100.0** (0.0)	0.0 (0.2)	99.7 (1.4)	0.0 (0.1)
	univariate	**100.0** (0.0)	**0.0** (0.1)	**100.0** (0.0)	**0.0** (0.1)	**99.9** (0.4)	**0.0** (0.0)
60%	multivariate	**100.0** (0.0)	0.2 (0.4)	**100.0** (0.0)	0.2 (0.4)	**99.7** (1.4)	**0.0** (0.2)
	univariate	**100.0** (0.0)	**0.1** (0.1)	99.9 (0.7)	**0.1** (0.2)	99.3 (1.3)	0.1 (0.1)

(d) Scenario 4

p_{curve}	RTLP version	Contamination 1		Contamination 2		Contamination 3	
		p_c	p_f	p_c	p_f	p_c	p_f
0%	multivariate	**99.8** (1.1)	**0.0** (0.0)	**97.6** (3.4)	**0.0** (0.1)	**99.2** (2.2)	**0.0** (0.0)
	univariate	98.2 (6.3)	0.9 (2.9)	92.7 (10.2)	4.1 (5.6)	93.0 (9.8)	3.6 (5.1)
30%	multivariate	**99.9** (0.7)	**0.0** (0.1)	**97.5** (4.0)	**0.0** (0.2)	**99.2** (2.2)	**0.0** (0.0)
	univariate	98.3 (5.7)	1.0 (3.0)	95.1 (8.1)	2.7 (4.7)	92.7 (9.9)	3.4 (5.0)
60%	multivariate	**100.0** (0.0)	**0.0** (0.1)	**95.8** (7.2)	**0.6** (3.6)	**98.3** (3.1)	**0.0** (0.2)
	univariate	97.5 (7.2)	1.4 (3.4)	90.9 (8.2)	2.8 (4.7)	92.2 (7.2)	2.1 (4.2)
p_{curve}	RTLP version	Contamination 4		Contamination 5		Contamination 6	
		p_c	p_f	p_c	p_f	p_c	p_f
0%	multivariate	**98.6** (4.0)	**0.0** (0.0)	**100.0** (0.0)	**0.0** (0.0)	**99.7** (1.3)	**0.0** (0.0)
	univariate	62.2 (15.2)	1.8 (4.0)	86.0 (11.1)	1.3 (3.9)	92.2 (8.2)	1.1 (3.3)
30%	multivariate	**97.8** (5.0)	**0.0** (0.0)	**100.0** (0.0)	**0.0** (0.1)	**99.1** (2.3)	**0.0** (0.0)
	univariate	60.4 (13.0)	2.3 (4.1)	84.8 (11.1)	1.1 (3.5)	92.2 (5.3)	0.2 (1.2)
60%	multivariate	**81.5** (15.1)	**0.0** (0.2)	**95.7** (19.4)	**0.0** (0.1)	**99.0** (2.8)	**0.0** (0.2)
	univariate	39.4 (10.9)	1.1 (2.9)	67.5 (19.9)	2.2 (4.5)	80.8 (6.9)	1.8 (3.9)

(e) Scenario 5

p_{curve}	RTLP version	Contamination 1		Contamination 2		Contamination 3	
		p_c	p_f	p_c	p_f	p_c	p_f
0%	multivariate	**100.0** (0.0)	**0.1** (0.3)	**98.6** (2.9)	**0.1** (0.3)	**99.7** (1.5)	**0.0** (0.1)
	univariate	99.9 (0.7)	0.8 (0.8)	98.0 (2.2)	0.5 (0.6)	98.7 (2.1)	0.4 (0.5)
30%	multivariate	**100.0** (0.6)	**0.7** (0.7)	95.1 (5.1)	**0.7** (0.7)	**99.6** (1.6)	**0.2** (0.4)
	univariate	99.2 (2.4)	1.1 (0.9)	**96.1** (2.8)	0.8 (1.1)	97.2 (2.5)	0.7 (0.5)
60%	multivariate	**100.0** (0.0)	**4.0** (3.3)	**87.6** (10.4)	4.9 (7.8)	**98.6** (2.7)	**2.8** (3.0)
	univariate	88.0 (5.7)	4.4 (1.8)	80.1 (6.2)	**3.7** (2.0)	85.9 (4.6)	3.4 (1.6)
p_{curve}	RTLP version	Contamination 4		Contamination 5		Contamination 6	
		p_c	p_f	p_c	p_f	p_c	p_f
0%	multivariate	**99.8** (1.1)	**0.1** (0.3)	**100.0** (0.0)	**0.1** (0.3)	**99.9** (0.9)	**0.0** (0.1)
	univariate	42.2 (9.1)	1.3 (2.4)	49.7 (10.1)	1.1 (1.5)	97.7 (3.9)	1.4 (1.6)
30%	multivariate	**98.5** (5.4)	**0.8** (0.8)	**99.5** (4.3)	**1.0** (1.0)	**99.9** (0.8)	**0.2** (0.3)
	univariate	40.6 (9.8)	6.0 (3.3)	46.9 (5.1)	4.8 (3.2)	82.0 (7.0)	1.8 (1.5)
60%	multivariate	**57.5** (32.8)	7.1 (7.1)	**64.5** (41.5)	**6.8** (7.1)	**100.0** (0.6)	**3.1** (5.3)
	univariate	37.9 (9.2)	**6.8** (1.2)	47.2 (4.8)	7.3 (1.7)	66.4 (3.0)	7.9 (2.4)

(f) Scenario 6

p_{curve} \ RTLP version		Contamination 1		Contamination 2		Contamination 3	
		p_c	p_f	p_c	p_f	p_c	p_f
0%	multivariate	**99.9** (0.7)	**0.1** (0.3)	**94.4** (4.7)	**0.3** (0.5)	**98.1** (3.0)	**0.1** (0.2)
	univariate	96.9 (7.5)	2.7 (4.4)	89.5 (12.5)	7.1 (7.5)	88.4 (13.6)	7.1 (7.4)
30%	multivariate	**100.0** (0.0)	**0.4** (1.4)	73.3 (12.8)	**0.4** (0.7)	**97.1** (4.5)	**0.4** (0.6)
	univariate	99.0 (4.3)	1.3 (2.4)	**86.6** (7.5)	2.2 (4.0)	86.7 (8.7)	3.8 (5.3)
60%	multivariate	**100.0** (0.0)	**1.6** (3.3)	55.7 (10.4)	**1.9** (7.8)	**94.9** (2.7)	**1.5** (3.0)
	univariate	98.5 (5.7)	1.9 (1.8)	**66.2** (6.2)	7.7 (2.0)	76.4 (4.6)	5.5 (1.6)
p_{curve} \ RTLP version		Contamination 4		Contamination 5		Contamination 6	
		p_c	p_f	p_c	p_f	p_c	p_f
0%	multivariate	**92.6** (1.6)	**0.2** (0.4)	**96.8** (0.0)	**0.2** (0.4)	**99.4** (0.0)	**0.0** (0.1)
	univariate	60.5 (12.6)	2.2 (3.5)	42.5 (9.8)	3.2 (4.3)	92.1 (11.2)	4.5 (5.1)
30%	multivariate	**54.6** (31.7)	**0.8** (1.4)	26.7 (44.3)	3.8 (8.3)	**99.7** (1.4)	**0.2** (0.4)
	univariate	37.1 (11.1)	1.7 (2.9)	**34.1** (14.9)	**1.3** (1.0)	80.0 (7.0)	0.7 (0.5)
60%	multivariate	16.6 (22.7)	**14.2** (9.3)	0.0 (0.0)	**3.4** (5.9)	**100.0** (0.0)	**1.9** (2.4)
	univariate	**33.1** (6.5)	14.9 (3.5)	**12.9** (9.4)	15.1 (3.7)	51.1 (4.6)	9.0 (1.9)