Costs and Benefits of Including or Omitting Interaction Terms: A Monte Carlo Simulation

Małgorzata Mikucka, Francesco Sarracino & Joshua K. Dubrow

Statistical Office of Luxembourg, MZES University of Mannheim, University of Leuvain-la-Neuve, & LCSR - HSE

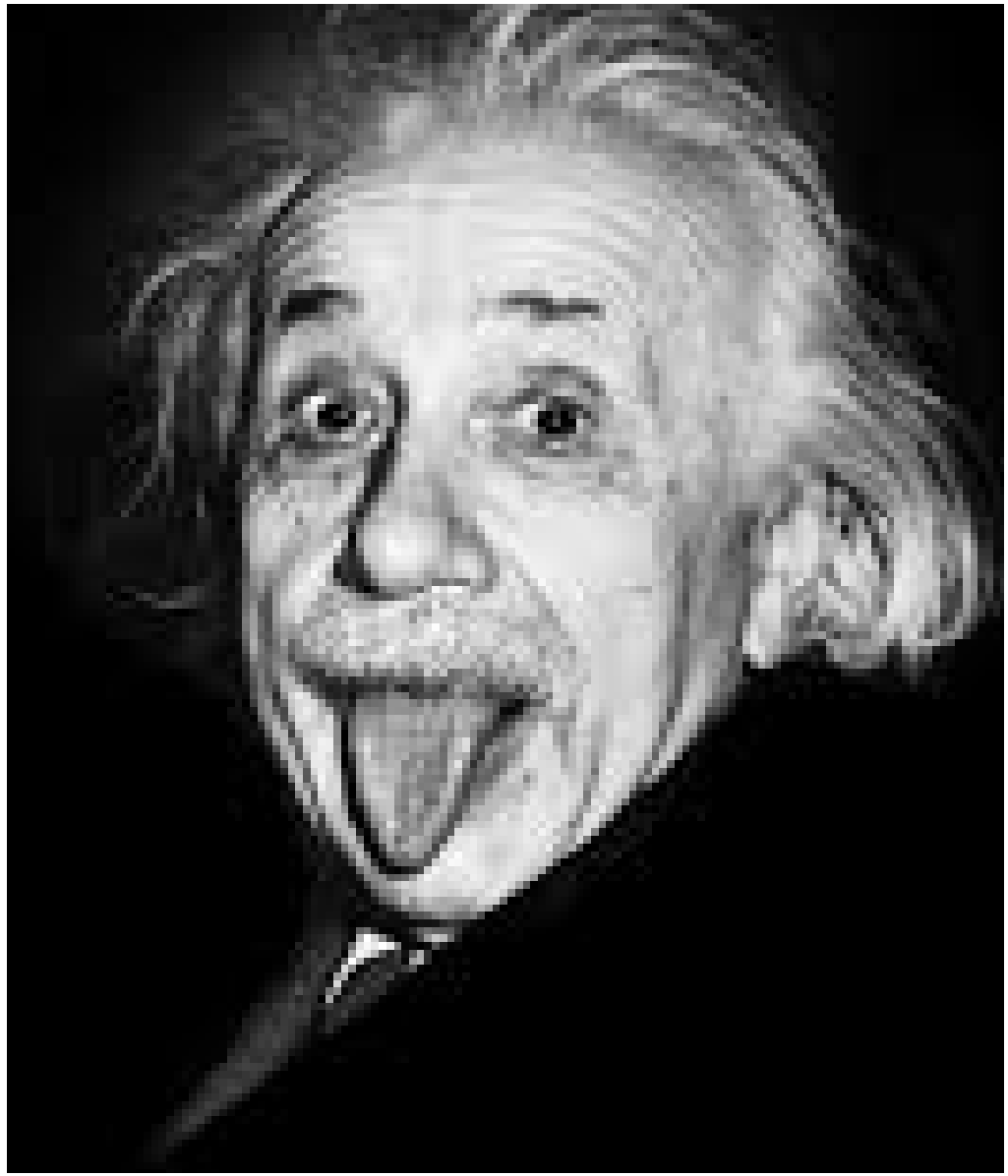
Tuesday 11th April, 2017

Take home messages

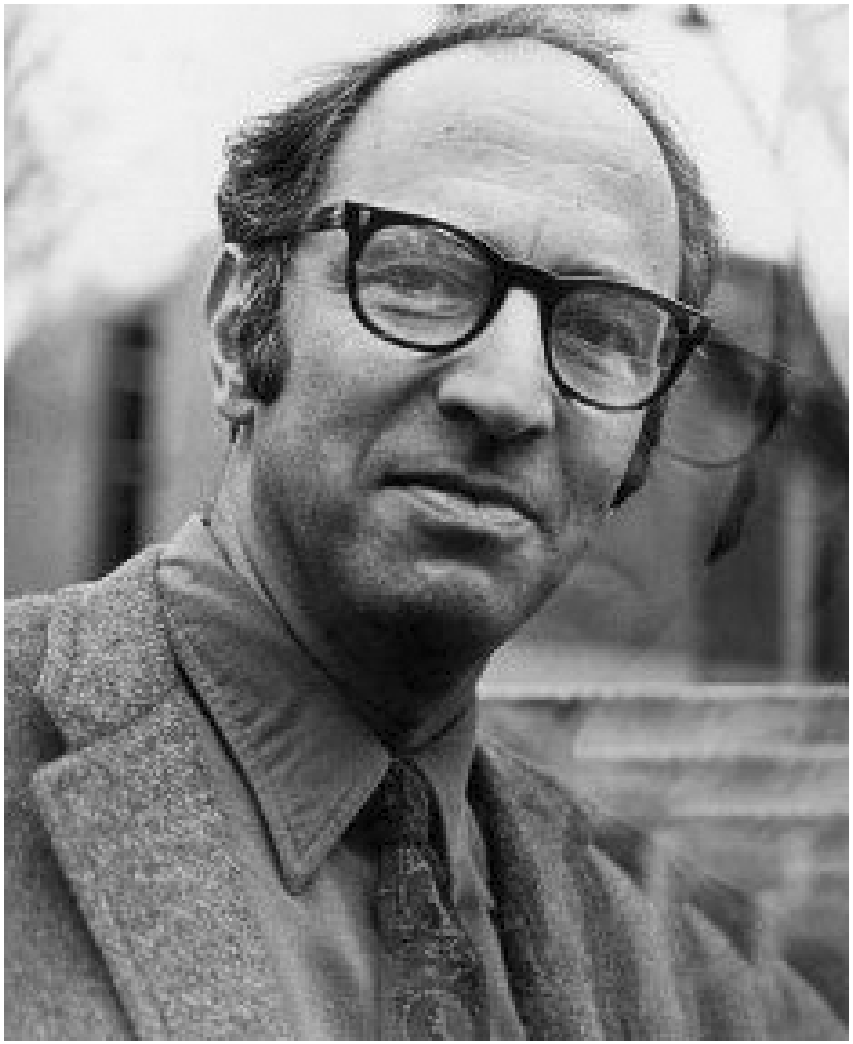
To interact or not to interact?

If you ask Einstein...

It is the theory which decides what we can observe.

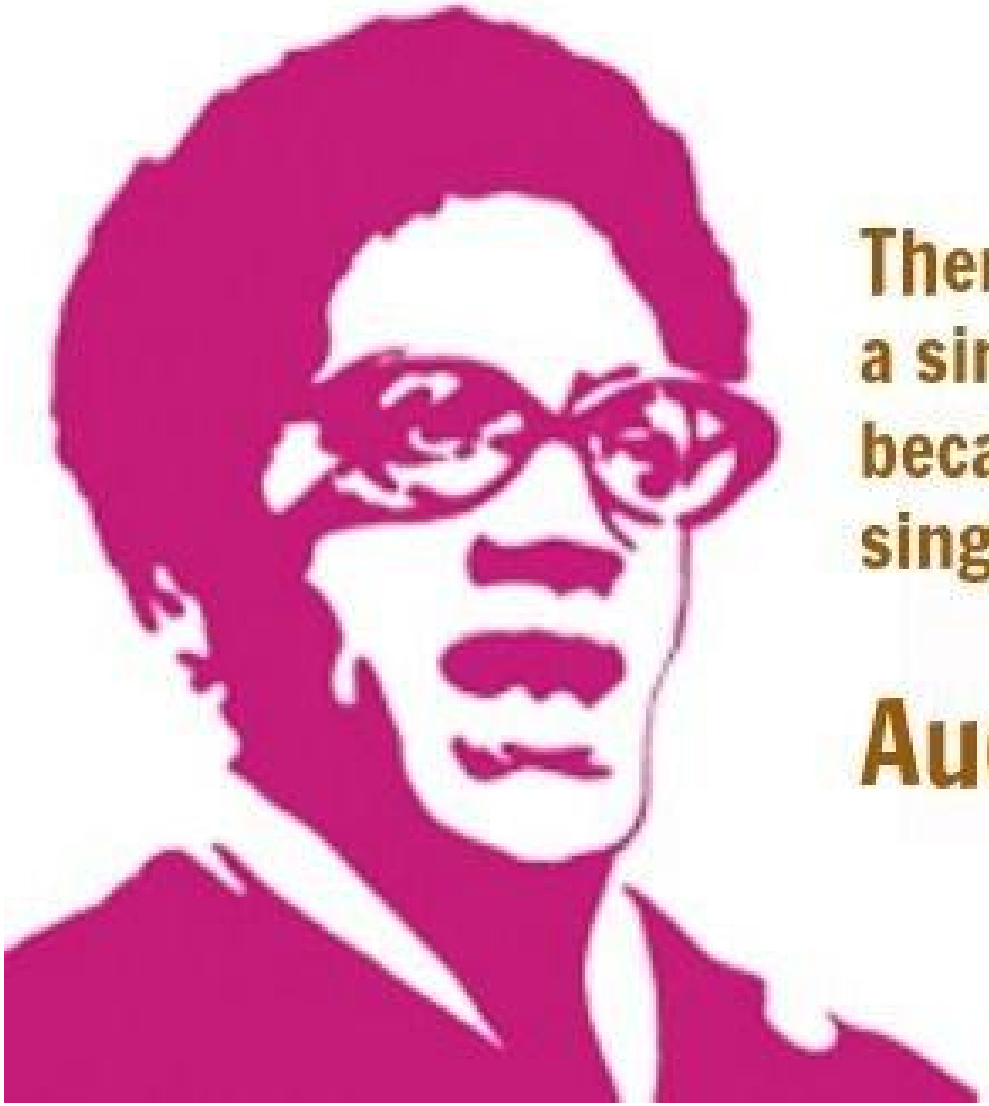


and if you ask Kuhn...



The answers you get depend upon the questions you ask.

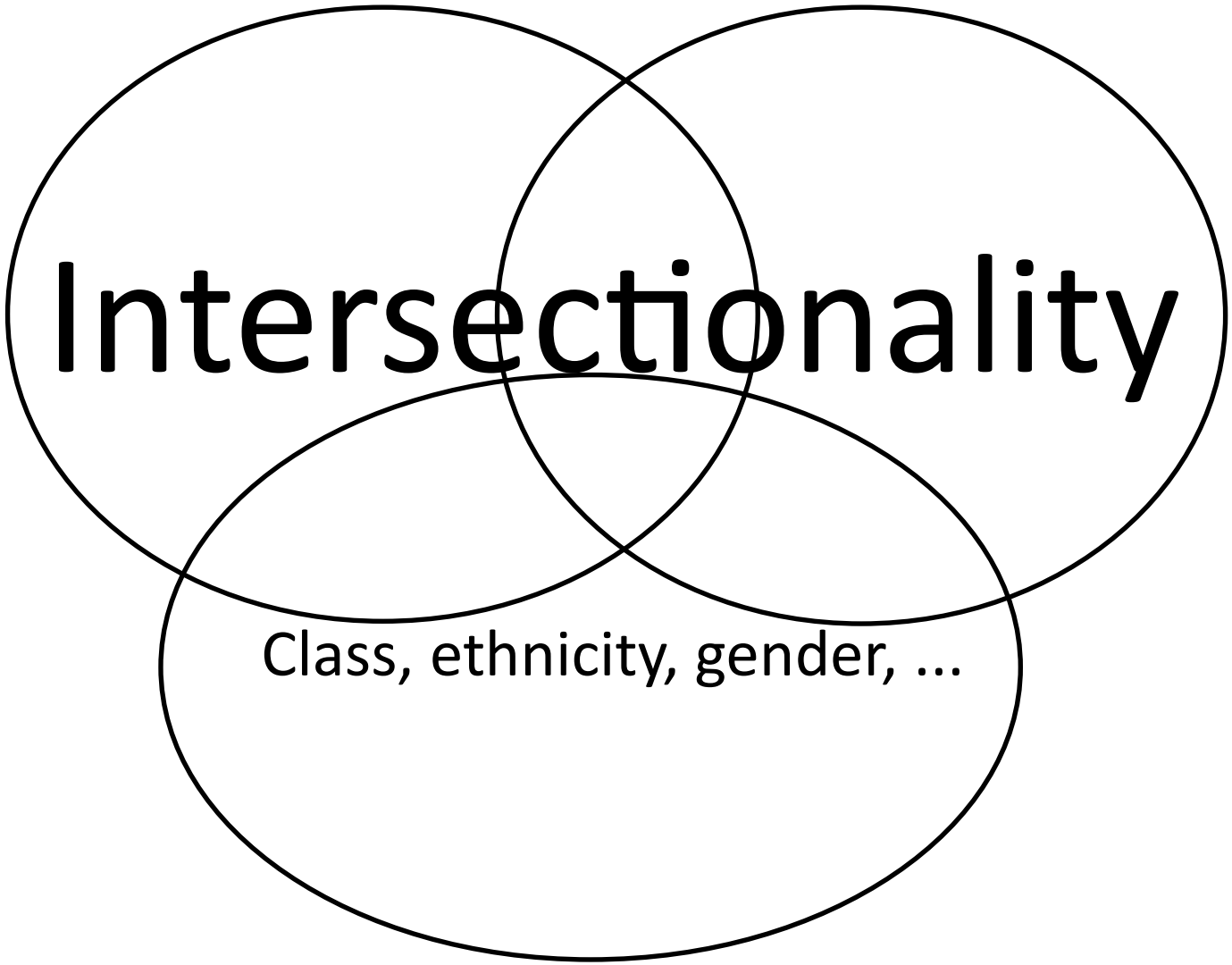
Intersectionality



There is no such thing as a single-issue struggle because we do not live single-issue lives.

Audre Lorde

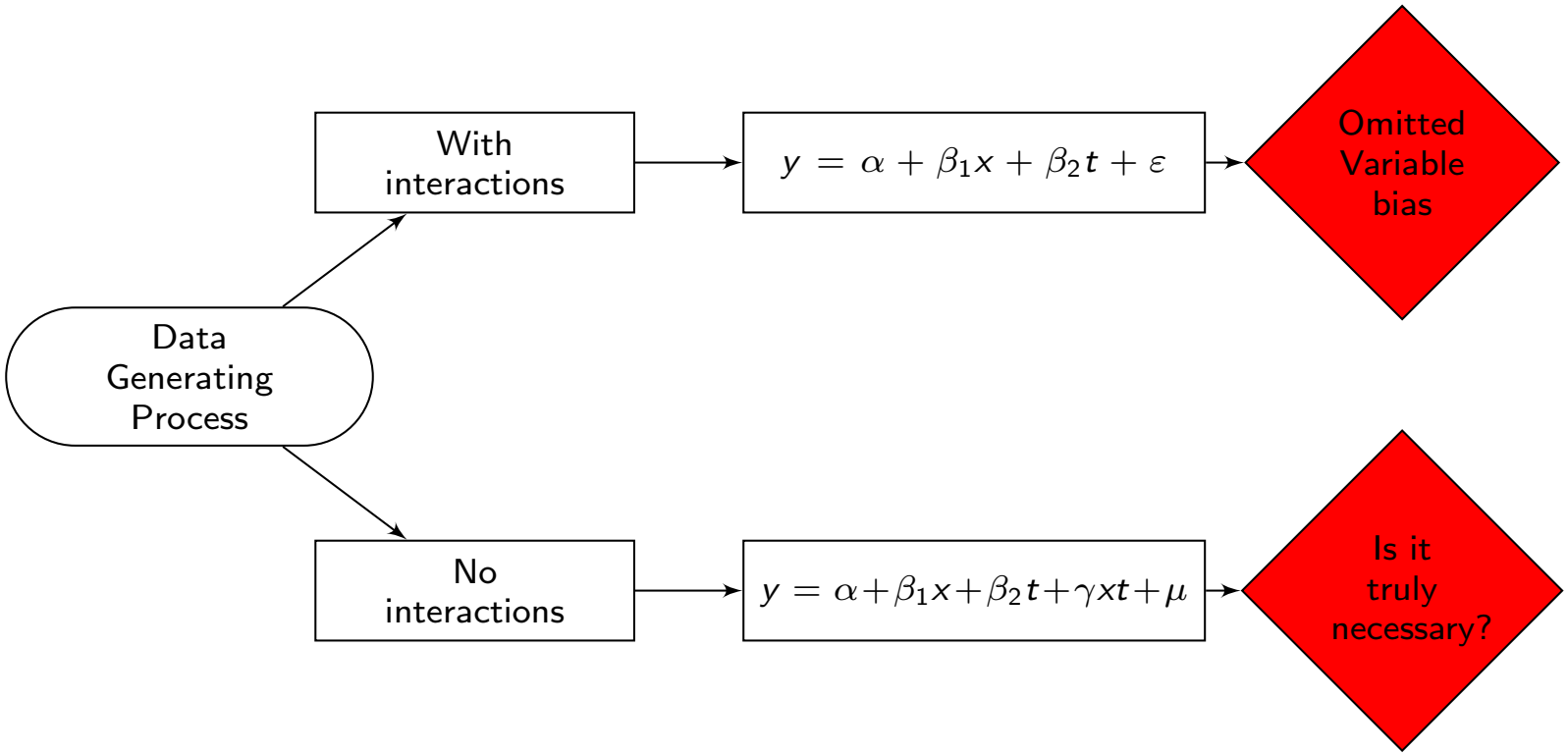
Intersections and interactions



The discovery of regressions with interactions

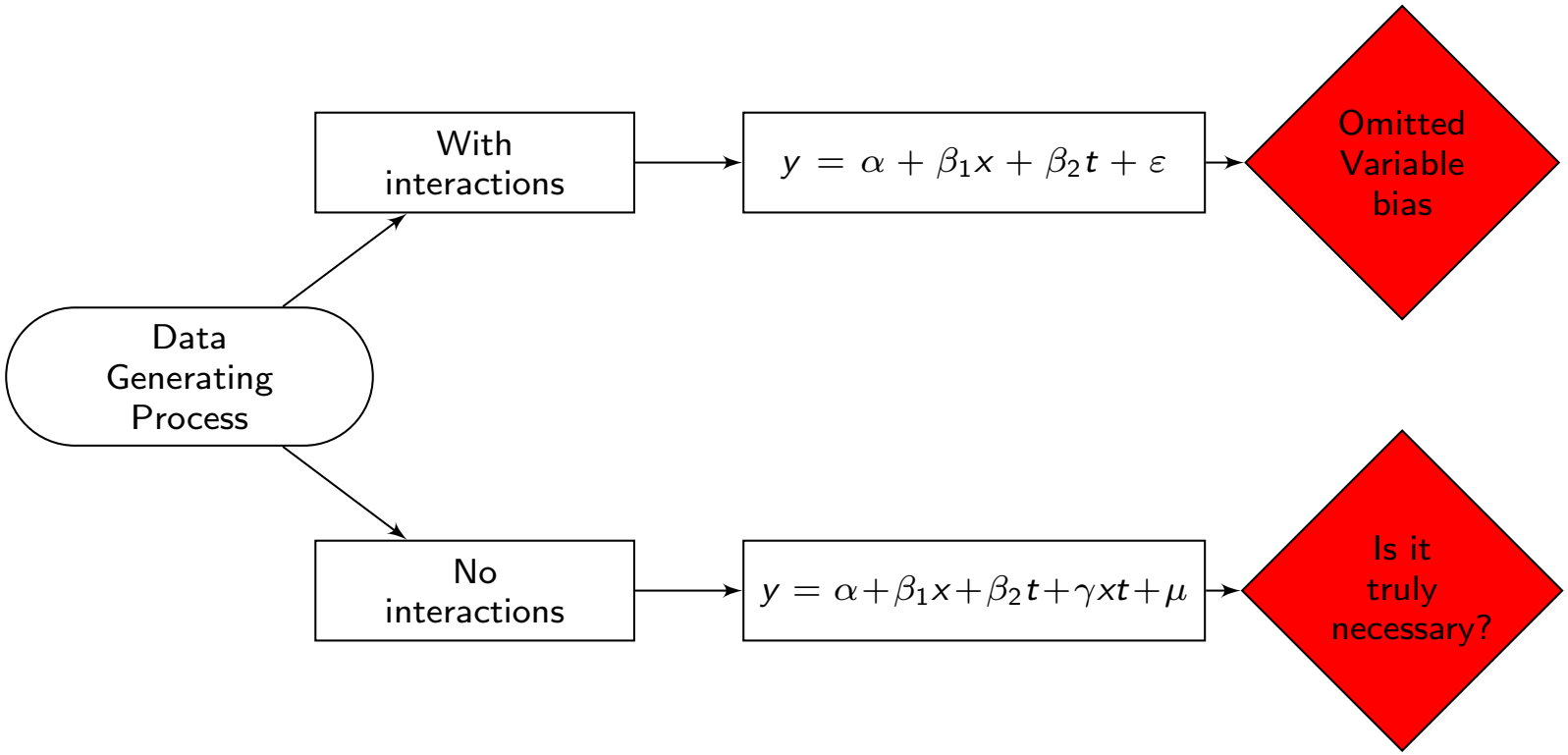
Research questions

- Are interaction models an improvement over additive model?



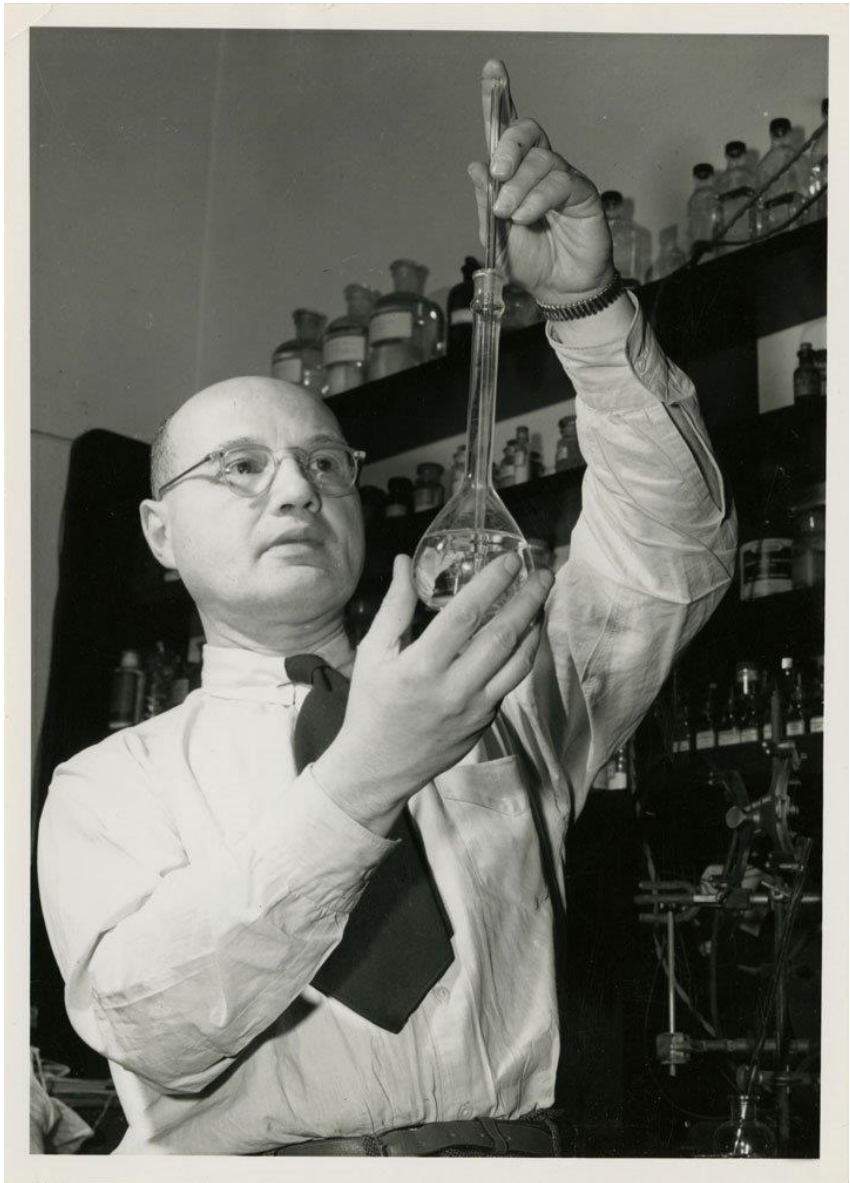
Research questions

- Are interaction models an improvement over additive model?



- Under which conditions wrongly including interaction effects alters the precision of the model?

Do you know this man?



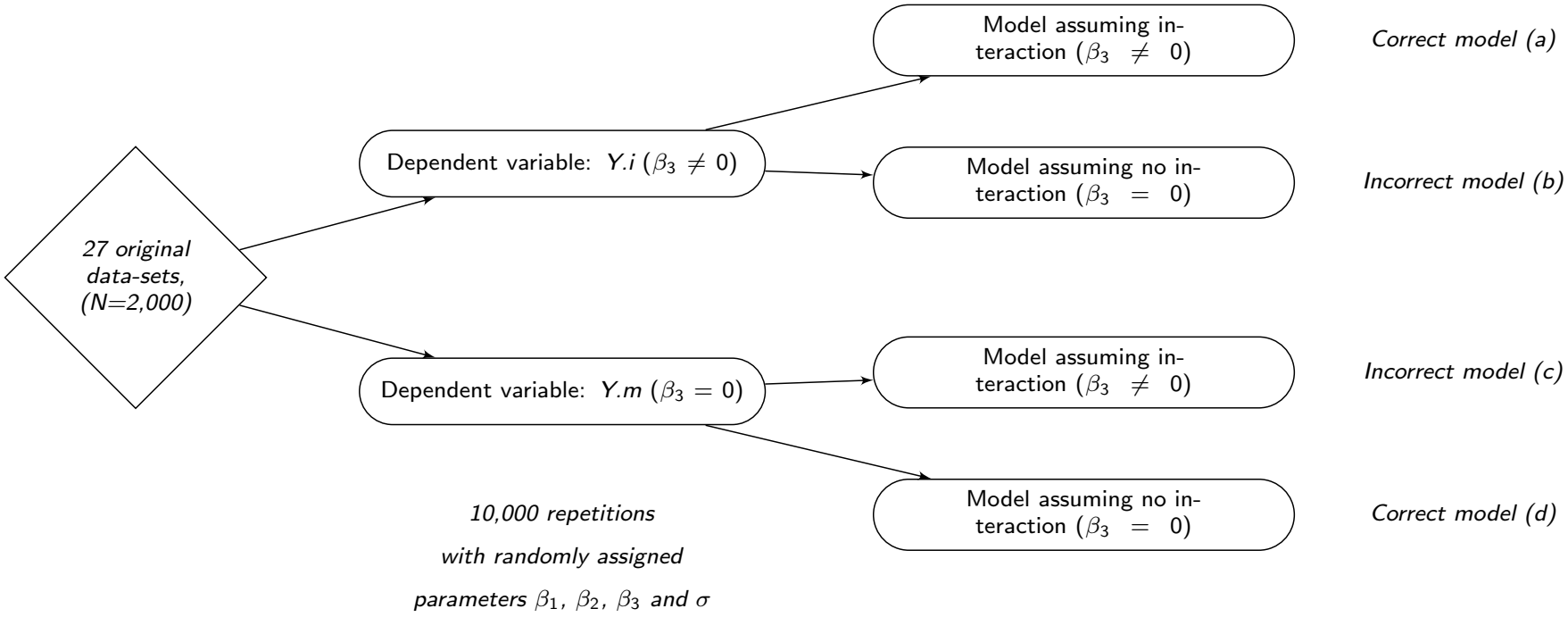
Theory guides, experiment decides.

The ingredients

- Two independent variables:
  - two dummies (DD);
  - one dummy, one continuous (DC);
  - two continuous (CC).
- Varying correlation between variables;
  $(-0.90 \leq Cor(x; t) \leq 0.90)$
- Two dependent variables:
  - $Y.m = \alpha + \beta_1 \cdot x + \beta_2 \cdot t + \epsilon$
  - $Y.i = \alpha + \beta_1 \cdot x + \beta_2 \cdot t + \gamma \cdot x \cdot t + \epsilon$
- Varying ϵ;
  $\epsilon \sim N(0, \sigma); \sigma \sim U(1, 100)$

Overall, we build 27 datasets of 2000 observations each.

Monte Carlo simulation

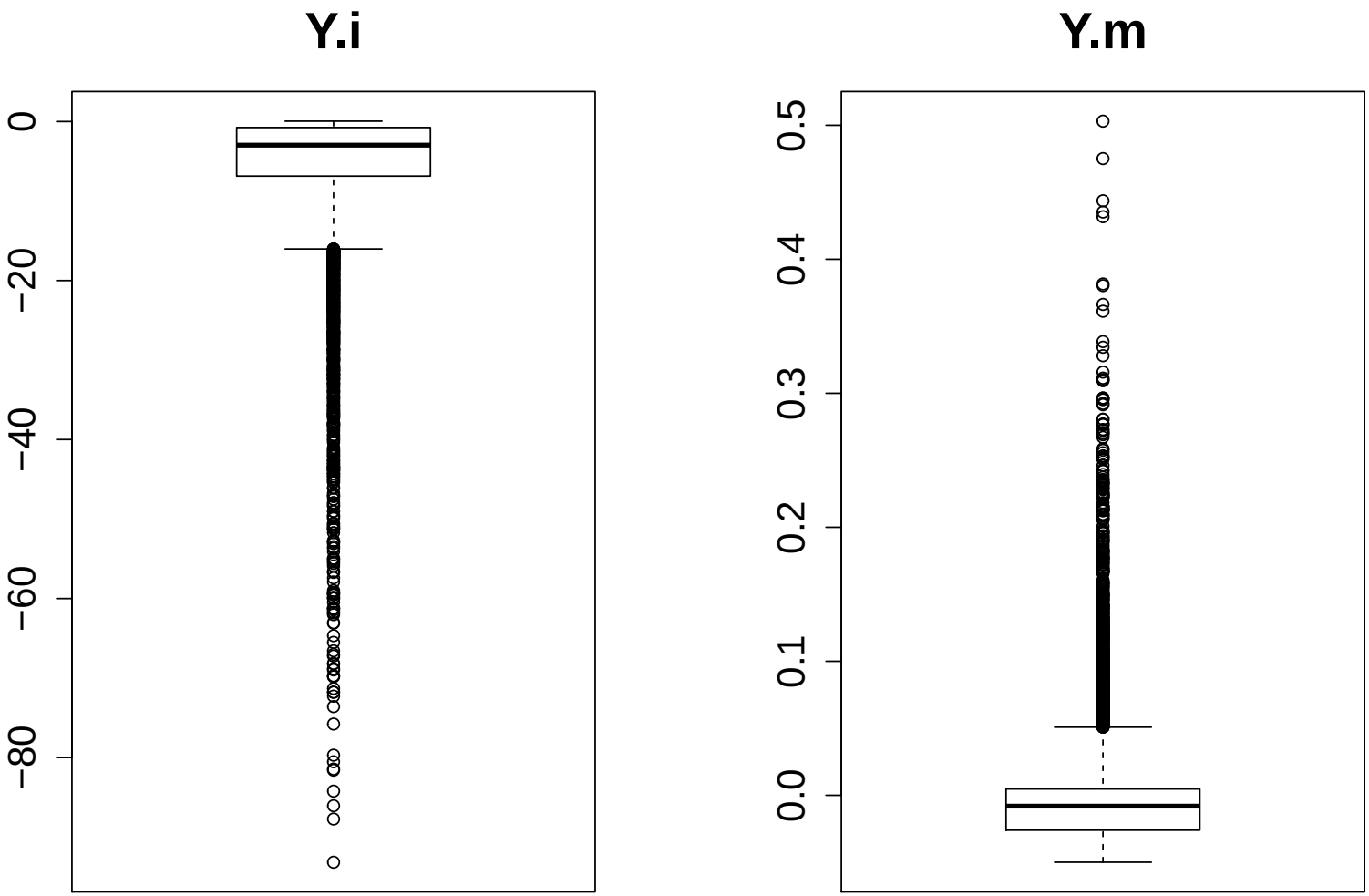


- Estimations: $27 \cdot 10,000 \cdot 4 = 1,080,000$;
- Measurement:
  - $d_1 = adj.R^2_{incorrect} - adj.R^2_{correct}$;
  - $d_2 = ((\hat{\beta}_i - \beta_i^{true})/\beta_i^{true}) * 100$;
  - $d_3 = ((S.E._{incorrect} - S.E._{correct})/S.E._{correct}) * 100$.

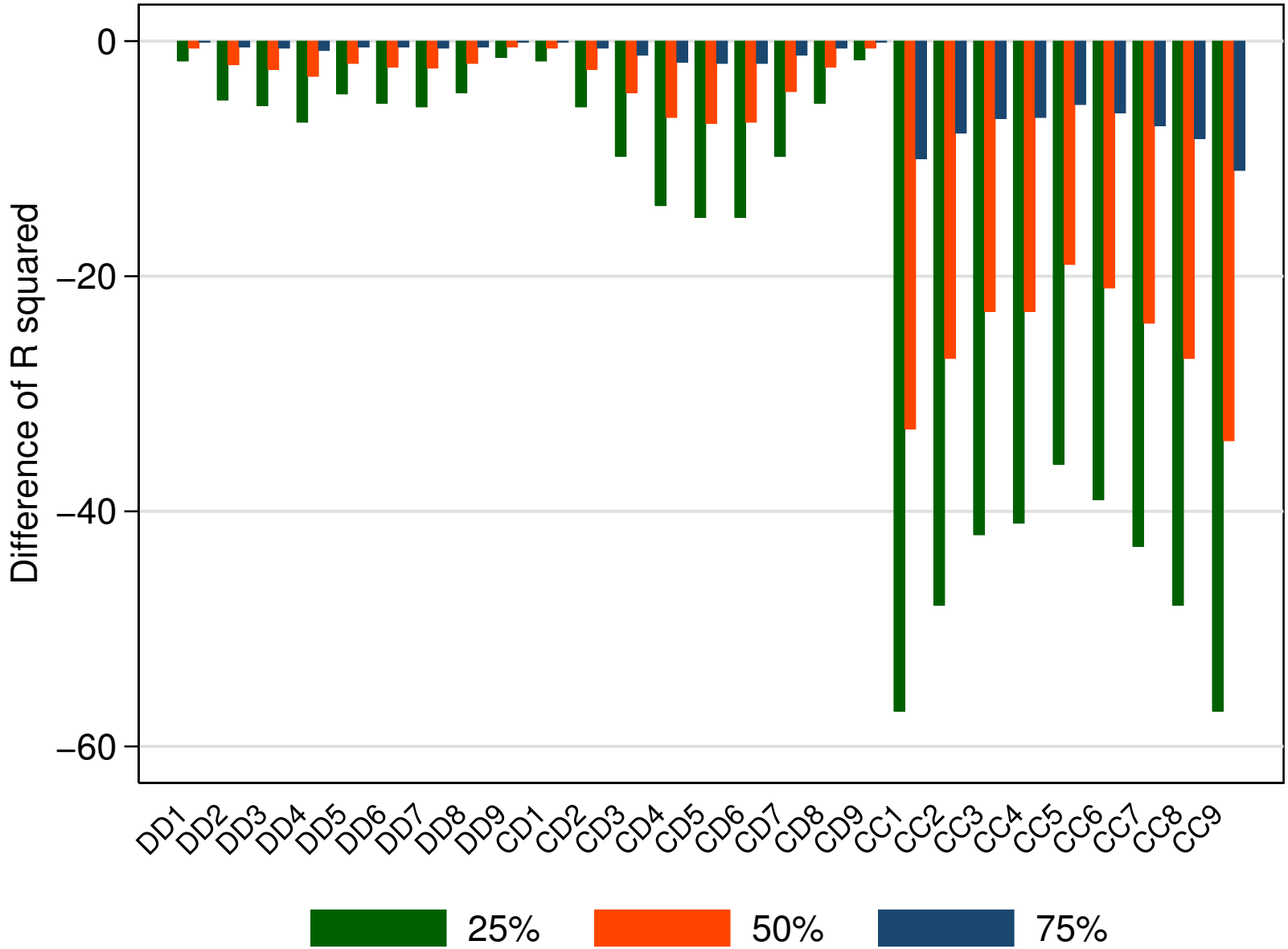
Results

Loss of predictive power: two dummies, equal distribution

Wrongly omitting Vs. Wrongly including



Loss of predictive power: all cases



What explains the loss?

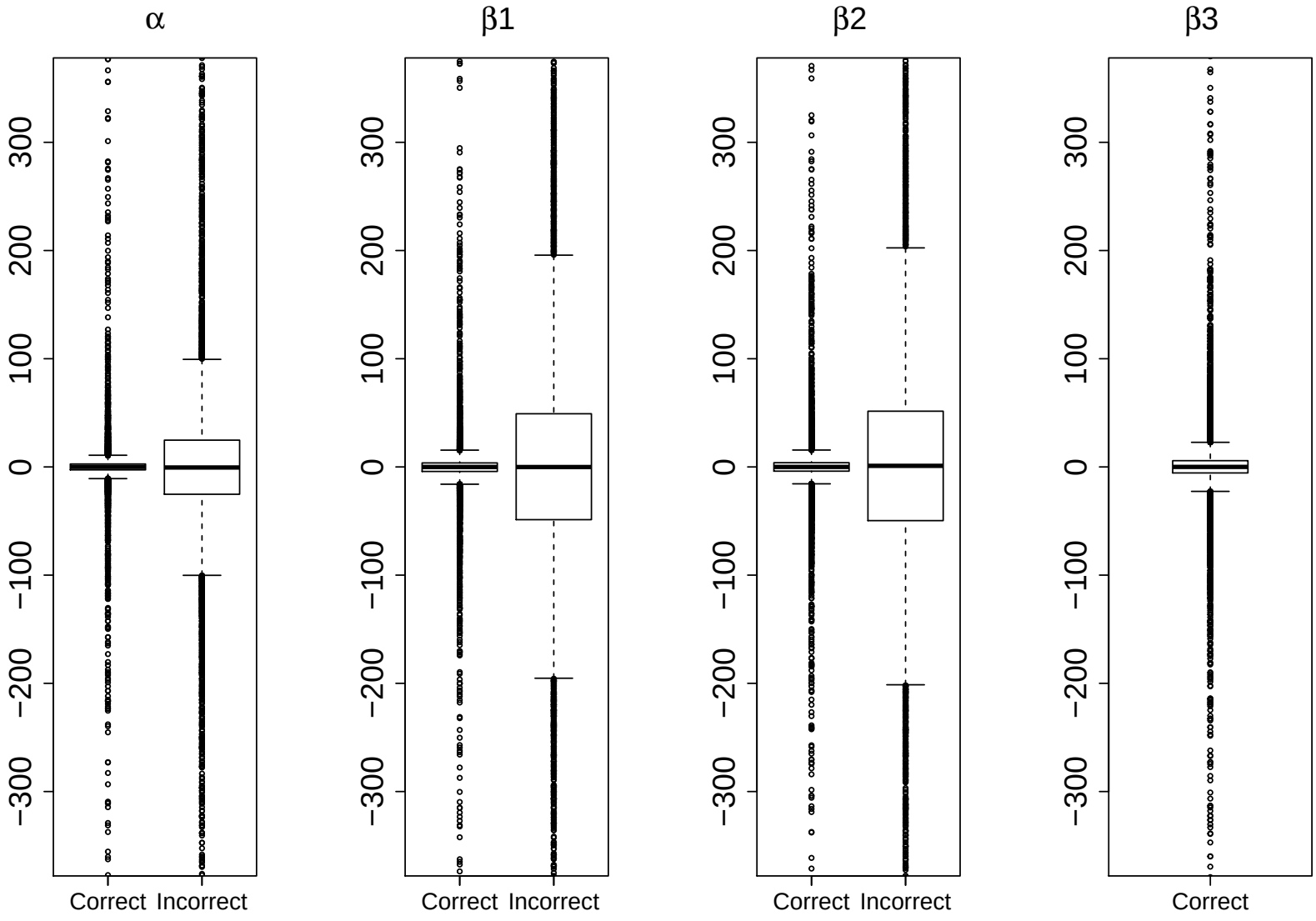
	Dependent variable: Loss of predictive power				
	All data-sets	DD data-sets (1)	DD data-sets (2)	DC data-sets	CC data-sets
DxC data-sets	−3.10***				
	(0.07)				
CxC data-sets	−25.00***				
	(0.07)				
DD1 data		3.80***			
		(0.09)			
DD2 data		1.30***			
		(0.09)			
DD3 data		1.00***			
		(0.09)			
DD5 data		1.60***			
		(0.09)			
DD6 data		1.30***			
		(0.09)			
DD7 data		0.98***			
		(0.09)			
DD8 data		1.60***			
		(0.09)			
DD9 data		4.10***			
		(0.09)			
10% change of rho(X1, X2) (abs)	0.07***		0.30***	1.20***	−1.40***
	(0.01)		(0.01)	(0.01)	(0.03)
10% change of rho(X1, X2)	0.01		0.02***	−0.00	0.00
	(0.01)		(0.00)	(0.01)	(0.01)
10% change of error variance (Sigma2)	1.30***	0.79***	0.79***	1.10***	2.20***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
10% change of b1 (main effect)	−0.00	0.02***	0.02***	0.01	−0.04
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
10% change of b2 (main effect)	−0.02*	0.02***	0.02***	0.02	−0.10***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
10% change of b3 (interaction effect)	−0.02*	−0.04***	−0.04***	−0.01	−0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Constant	−8.70***	−8.30***	−7.70***	−17.00***	−29.00***
	(0.07)	(0.07)	(0.04)	(0.07)	(0.17)
Observations	270,000	90,000	90,000	90,000	90,000
Adjusted R^2	0.33	0.13	0.13	0.14	0.08

Note: *p<0.1; **p<0.05; ***p<0.01; The dependent variable is the difference of adjusted R^2.

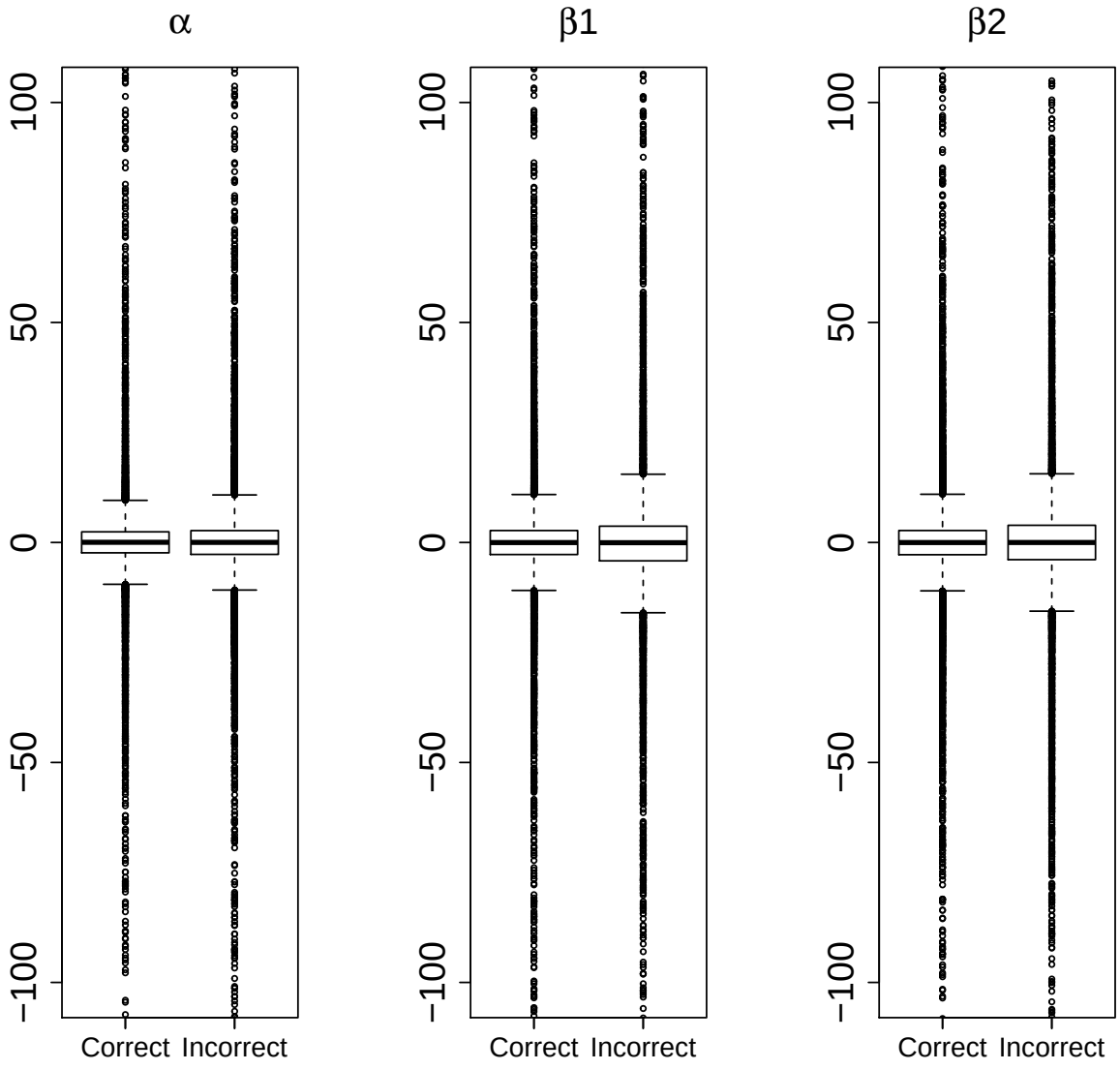
Findings

- Wrongly omitting decreases the predictive power of the model;
- CxC are particularly problematic;
- Wrongly including does not raise problem;
- What explains the loss of predictive power in case of wrongly omitting?
  - $\rho(X_1, X_2)$:
    - decreases the loss for CxC;
    - increases the loss for at least 1 Dummy.
  - σ^2 increases the loss.

Bias of the estimates (wrongly omitting)



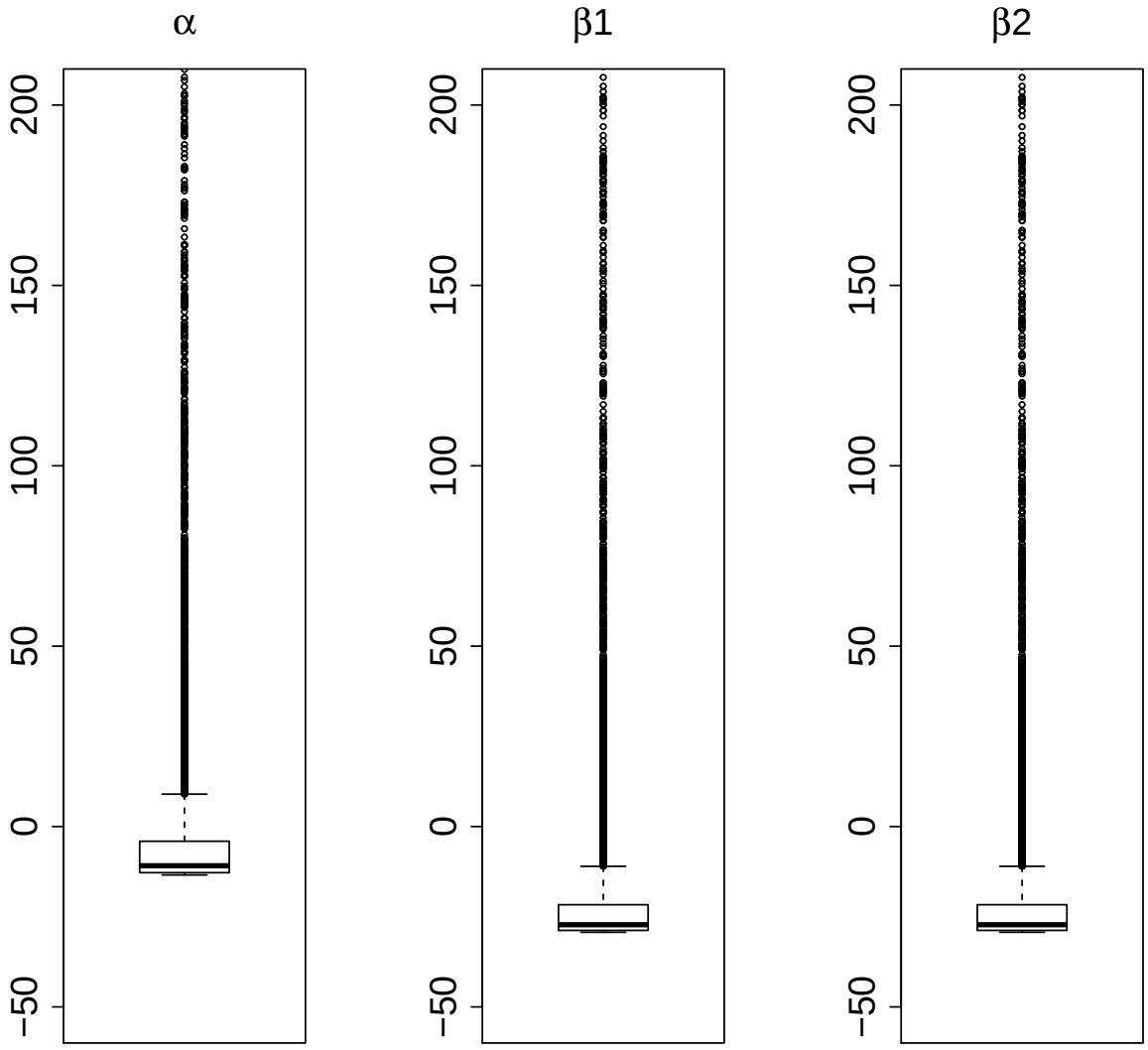
Bias of the estimates (wrongly including)



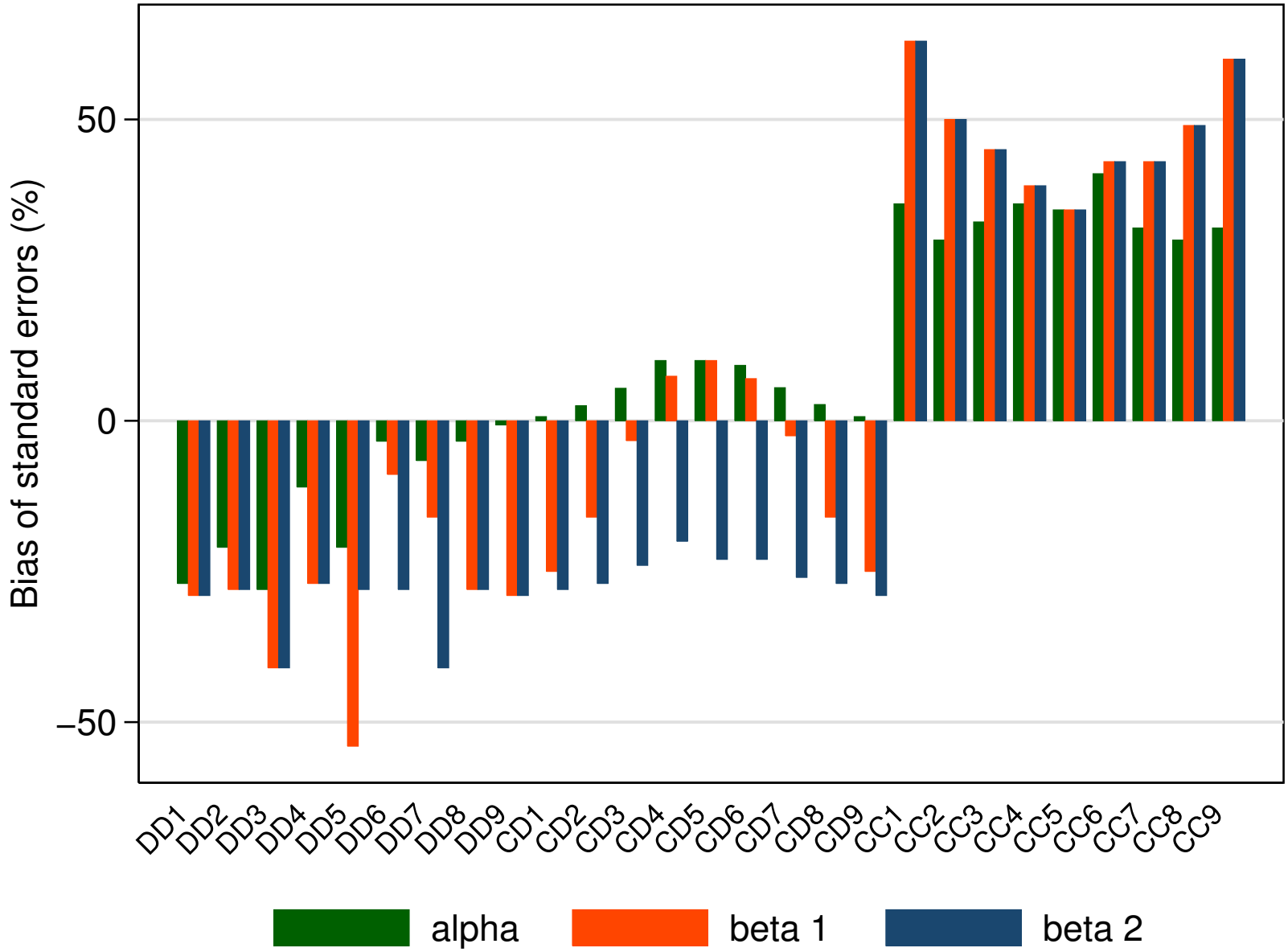
Findings:

- Wrongly omitting leads to biased coefficients:
  - Highest bias for DxD;
  - Lowest for CxC;
- Wrongly including leads to moderate bias;
- What does explain the bias?
  - We don't know: $adj.R^2 = 0$.

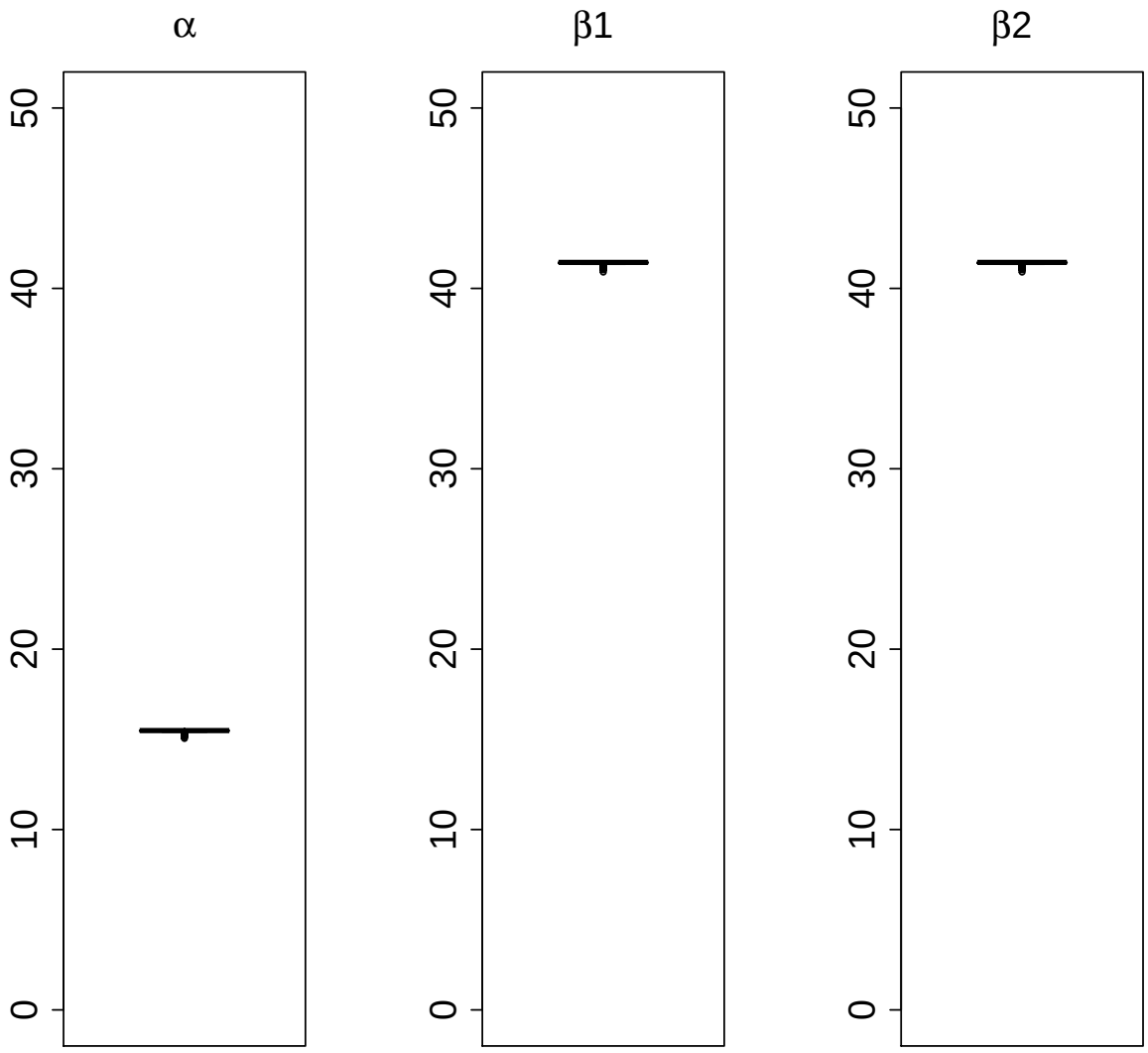
Percentage bias of the standard errors (wrongly omitting)



Percentage bias of the standard errors (wrongly omitting)



Percentage bias of the standard errors (wrongly including)



Conclusion: $Adj.R^2$

Wrongly omitting:

- loss of predictive power:
  - in particular for CxC;
  - $\rho(X_1; X_2)$: reduces the difference of $adj.R^2$;
  - σ^2: increases the difference of $adj.R^2$.

Wrongly including:

- No problem.

Conclusion: bias of coefficients

Wrongly omitting:

- large bias:
  - highest for DxD;
  - lowest for CxC;
  - but we do not know why.

Wrongly including:

- No bias.

Conclusion: bias of standard errors

Wrongly omitting:

- strongly affects the standard errors:
  - type 1 error: risk of false positive (DxD);
  - type 2 error: risk of false negative(CxC).

Wrongly including:

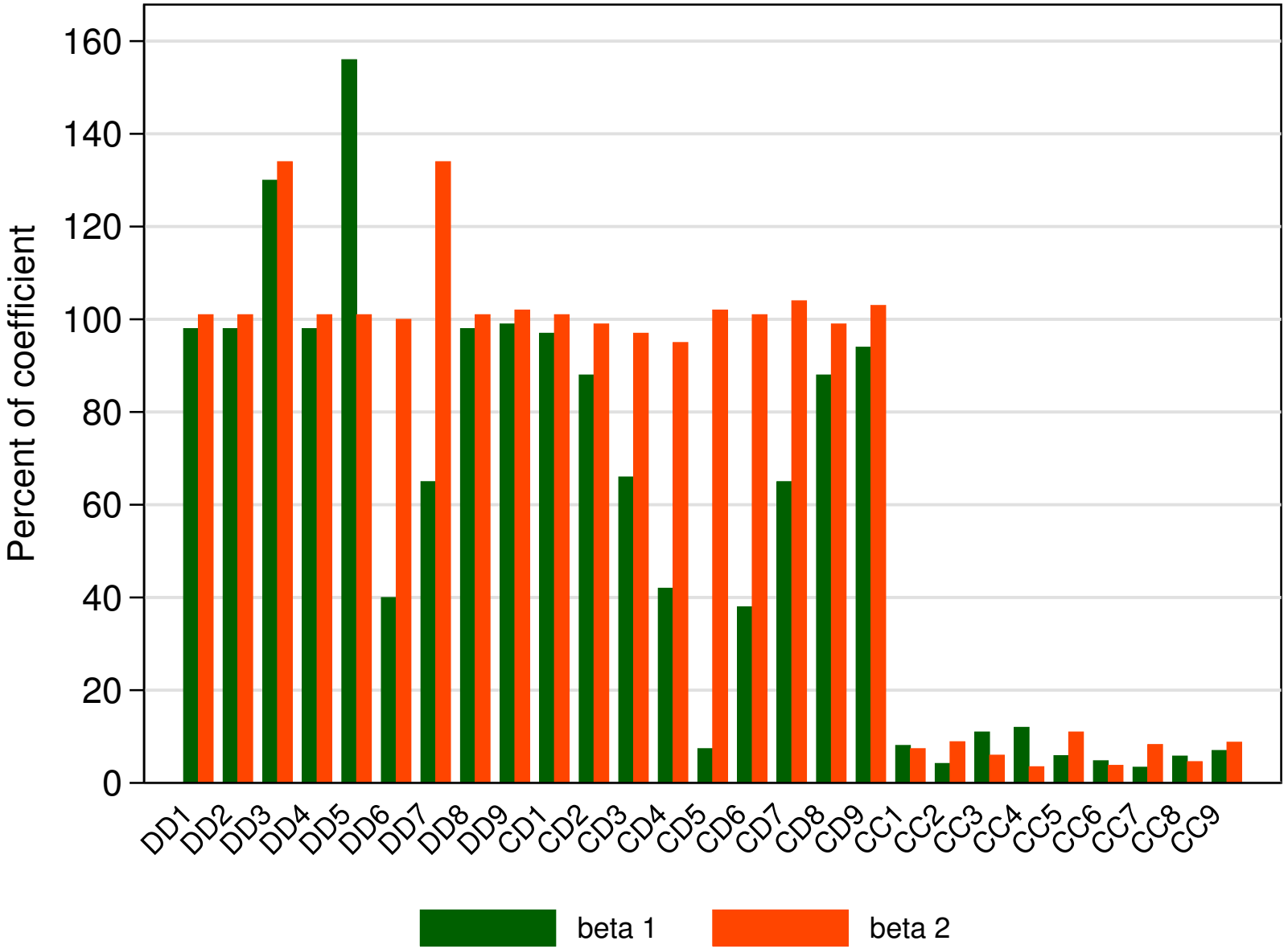
- type 2 error

Conclusion

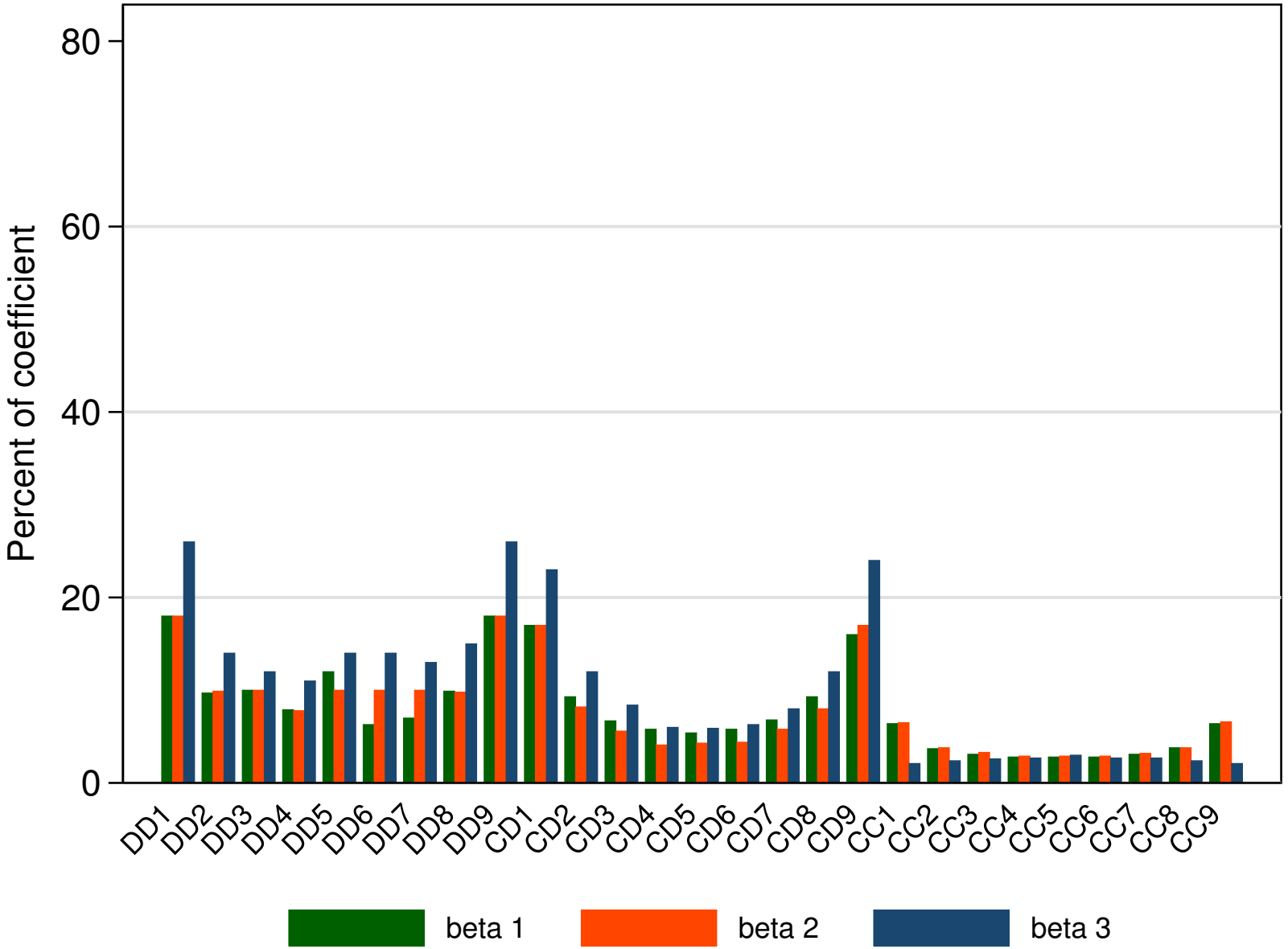
Thanks a lot for your attention!

Francesco.Sarracino@statec.etat.lu

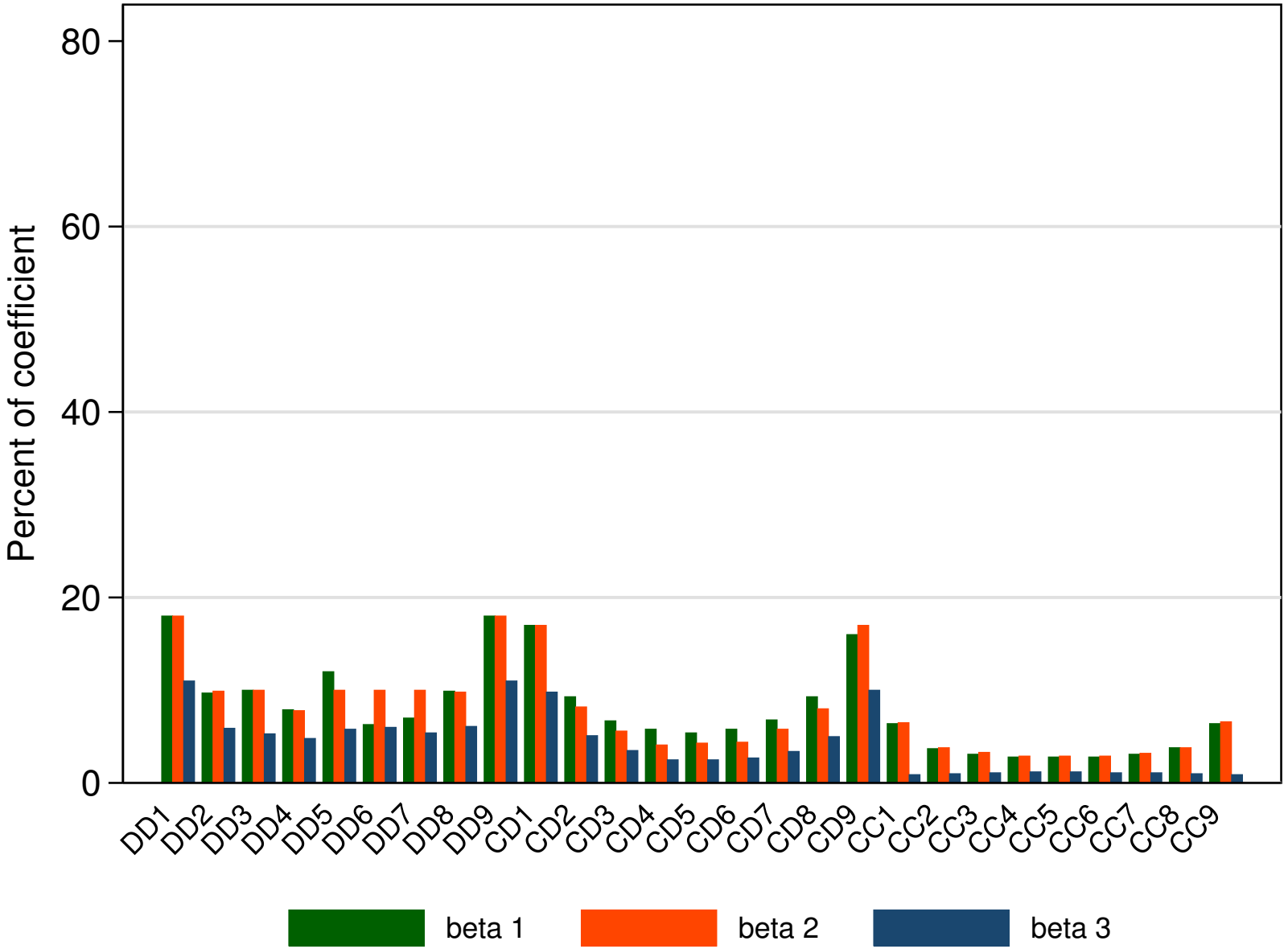
Coefficients bias: wrongly omitting



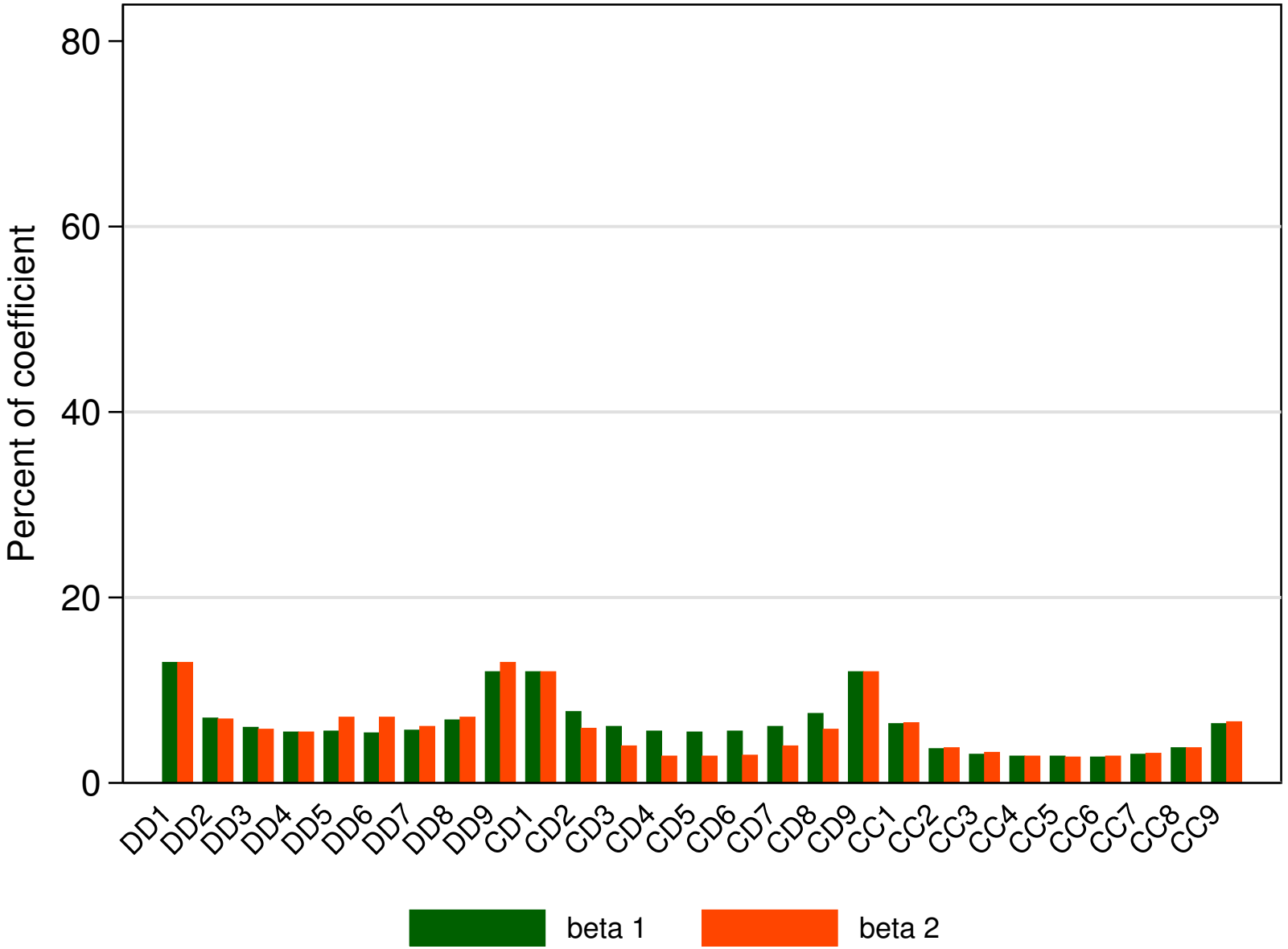
Coefficients bias: correctly including



Coefficients bias: wrongly including



Coefficients bias: correctly omitting



Explaining the percentage bias β_1 (wrongly omitting)

	Dependent variable:				
	Bias of β_1				
	All data-sets	DxD data-sets		DxC data-sets	CxC data-sets
DxC data-sets	−412.00***				
	(132.00)				
CxC data-sets	−385.00***				
	(132.00)				
DD1 data		−97.00			
		(567.00)			
DD2 data		0.47			
		(567.00)			
DD3 data		143.00			
		(567.00)			
DD5 data		321.00			
		(567.00)			
DD6 data		−190.00			
		(567.00)			
DD7 data		−121.00			
		(567.00)			
DD8 data		128.00			
		(567.00)			
DD9 data		97.00			
		(567.00)			
10% change of rho(X1, X2) (abs)	−0.49		−2.40	0.07	1.50
	(17.00)		(38.00)	(29.00)	(2.80)
10% change of rho(X1, X2)	19.00**		7.50	45.00***	0.58
	(9.50)		(26.00)	(15.00)	(1.40)
10% change of error variance (Sigma2)	23.00	75.00*	75.00*	−6.00	0.29
	(18.00)	(45.00)	(45.00)	(29.00)	(2.80)
10% change of b1 (main effect)	0.92	2.90	2.90	−0.18	0.01
	(18.00)	(46.00)	(46.00)	(30.00)	(2.80)
10% change of b2 (main effect)	−50.00***	−159.00***	−159.00***	10.00	−0.44
	(19.00)	(47.00)	(47.00)	(30.00)	(2.90)
10% change of b3 (interaction effect)	−49.00***	−156.00***	−156.00***	9.90	−0.86
	(19.00)	(47.00)	(47.00)	(31.00)	(2.90)
Constant	311.00**	106.00	146.00	−6.90	−6.40
	(129.00)	(429.00)	(251.00)	(198.00)	(19.00)
Observations	270,000	90,000	90,000	90,000	90,000
Adjusted R^2	0.00	0.00	0.00	0.00	-0.00

Note: *p<0.1; **p<0.05; ***p<0.01.

Explaining the percentage bias of β_2 (wrongly omitting)

	Dependent variable: Bias of β_2				
	All data-sets	DD data-sets (1)	DD data-sets (2)	DC data-sets	CC data-sets
DxC data-sets	5.90				
	(17.00)				
CxC data-sets	48.00***				
	(17.00)				
DD1 data		−5.20			
		(65.00)			
DD2 data		−2.60			
		(65.00)			
DD3 data		−15.00			
		(65.00)			
DD5 data		1.90			
		(65.00)			
DD6 data		−1.50			
		(65.00)			
DD7 data		−14.00			
		(65.00)			
DD8 data		−0.09			
		(65.00)			
DD9 data		0.86			
		(65.00)			
10% change of rho(X1, X2) (abs)	−0.03		0.10	−0.59	0.36
	(2.20)		(4.40)	(4.70)	(0.36)
10% change of rho(X1, X2)	0.23		0.29	0.17	0.23
	(1.20)		(2.90)	(2.40)	(0.18)
10% change of error variance (Sigma2)	8.70***	13.00**	13.00**	13.00***	0.27
	(2.40)	(5.20)	(5.20)	(4.70)	(0.36)
10% change of b1 (main effect)	−2.20	−4.00	−4.00	−3.50	0.78**
	(2.40)	(5.30)	(5.30)	(4.80)	(0.37)
10% change of b2 (main effect)	0.22	0.35	0.35	0.31	−0.02
	(2.40)	(5.40)	(5.40)	(4.90)	(0.37)
10% change of b3 (interaction effect)	17.00***	26.00***	26.00***	24.00***	0.87**
	(2.40)	(5.40)	(5.40)	(4.90)	(0.38)
Constant	−80.00***	−91.00*	−95.00***	−85.00***	−4.20*
	(17.00)	(49.00)	(29.00)	(32.00)	(2.40)
Observations	270,000	90,000	90,000	90,000	90,000
Adjusted R^2	0.00	0.00	0.00	0.00	0.00

Note: *p<0.1; **p<0.05; ***p<0.01.

Explaining the percentage bias β_1 (wrongly including)

	Dependent variable:			
	Percentage bias of β_1			
	All data-sets	DxD data-sets	DxC data-sets	CxC data-sets
DxC data-sets	−0.33			
	(26.00)			
CxC data-sets	31.00			
	(26.00)			
DD1 data		−106.00		
		(114.00)		
DD2 data		4.40		
		(114.00)		
DD3 data		−14.00		
		(114.00)		
DD5 data		26.00		
		(114.00)		
DD6 data		77.00		
		(114.00)		
DD7 data		52.00		
		(114.00)		
DD8 data		113.00		
		(114.00)		
DD9 data		176.00		
		(114.00)		
10% change of rho(X1, X2) (abs)	2.90		1.30	6.80	1.20
	(3.40)		(7.70)	(4.90)	(2.60)
10% change of rho(X1, X2)	4.30**		14.00***	−0.67	1.70
	(1.90)		(5.20)	(2.50)	(1.30)
10% change of error variance (Sigma2)	−2.20	−5.00	−5.00	−4.10	2.50
	(3.60)	(9.10)	(9.10)	(4.90)	(2.60)
10% change of b1 (main effect)	−0.06	−0.17	−0.17	−0.11	0.10
	(3.60)	(9.30)	(9.30)	(5.00)	(2.60)
10% change of b2 (main effect)	3.30	8.20	8.20	6.70	−5.20*
	(3.70)	(9.40)	(9.40)	(5.10)	(2.70)
10% change of b3 (interaction effect)	2.90	7.80	7.80	6.50	−5.40**
	(3.70)	(9.40)	(9.40)	(5.10)	(2.70)
Constant	−24.00	−40.00	−9.00	−38.00	−1.00
	(25.00)	(87.00)	(51.00)	(33.00)	(17.00)
Observations	270,000	90,000	90,000	90,000	90,000
Adjusted R^2	0.00	-0.00	0.00	0.00	0.00

Note: *p<0.1; **p<0.05; ***p<0.01.

Explaining the percentage bias β_2 (wrongly including)

	Dependent variable:			
	Percentage bias of β_2			
	All data-sets	DxD data-sets	DxC data-sets	CxC data-sets
DxC data-sets	6.60***			
	(2.30)			
CxC data-sets	3.90*			
	(2.30)			
DD1 data		−11.00		
		(9.90)		
DD2 data		0.10		
		(9.90)		
DD3 data		0.27		
		(9.90)		
DD5 data		0.82		
		(9.90)		
DD6 data		2.70		
		(9.90)		
DD7 data		6.10		
		(9.90)		
DD8 data		2.00		
		(9.90)		
DD9 data		−2.40		
		(9.90)		
10% change of rho(X1, X2) (abs)	−0.38		−0.81	−0.18	0.03
	(0.29)		(0.66)	(0.46)	(0.20)
10% change of rho(X1, X2)	0.26		0.43	0.26	0.11
	(0.16)		(0.44)	(0.24)	(0.10)
10% change of error variance (Sigma2)	−0.21	−0.08	−0.08	−0.71	0.15
	(0.31)	(0.79)	(0.79)	(0.47)	(0.20)
10% change of b1 (main effect)	0.53*	0.13	0.13	0.64	0.82***
	(0.32)	(0.80)	(0.80)	(0.48)	(0.20)
10% change of b2 (main effect)	−0.00	0.01	0.01	0.01	−0.03
	(0.32)	(0.81)	(0.81)	(0.48)	(0.21)
10% change of b3 (interaction effect)	1.10***	2.60***	2.60***	0.14	0.66***
	(0.32)	(0.82)	(0.82)	(0.48)	(0.21)
Constant	−2.30	−4.10	−1.20	5.00	−1.70
	(2.20)	(7.50)	(4.40)	(3.10)	(1.40)
Observations	270,000	90,000	90,000	90,000	90,000
Adjusted R^2	0.00	0.00	0.00	-0.00	0.00

Note: *p<0.1; **p<0.05; ***p<0.01.

Explaining the difference of S.E.

	Dependent variable:			
	Percentage bias of the standard error of the β_1			
	All data-sets	DxD data-sets	DxC data-sets	CxC data-sets
DxC data-sets	44.00***			
	(1.60)			
CxC data-sets	216.00***			
	(1.60)			
DD1 data		−15.00***		
		(0.88)		
DD2 data		−5.50***		
		(0.88)		
DD3 data		−19.00***		
		(0.88)		
DD5 data		−37.00***		
		(0.88)		
DD6 data		17.00***		
		(0.88)		
DD7 data		11.00***		
		(0.88)		
DD8 data		−5.50***		
		(0.88)		
DD9 data		−15.00***		
		(0.88)		
10% change of rho(X1, X2) (abs)	−1.10***		−0.75***	−11.00***	8.30***
	(0.21)		(0.06)	(0.17)	(0.63)
10% change of rho(X1, X2)	0.16		0.30***	−0.04	0.25
	(0.12)		(0.04)	(0.09)	(0.32)
10% change of error variance (Sigma2)	−27.00***	−5.10***	−5.10***	−13.00***	−62.00***
	(0.22)	(0.07)	(0.07)	(0.17)	(0.63)
10% change of b1 (main effect)	−1.00***	−0.28***	−0.28***	−0.59***	−2.10***
	(0.23)	(0.07)	(0.07)	(0.18)	(0.64)
10% change of b2 (main effect)	0.74***	0.18**	0.18**	0.41**	1.60**
	(0.23)	(0.07)	(0.07)	(0.18)	(0.65)
10% change of b3 (interaction effect)	1.60***	0.36***	0.36***	0.86***	3.60***
	(0.23)	(0.07)	(0.07)	(0.18)	(0.65)
Constant	79.00***	9.10***	4.40***	125.00***	366.00***
	(1.60)	(0.66)	(0.40)	(1.20)	(4.20)
Observations	270,000	90,000	90,000	90,000	90,000
Adjusted R^2	0.11	0.11	0.06	0.10	0.10

Note: *p<0.1; **p<0.05; ***p<0.01.