Online Appendix to Simulating Collusion:

Challenging Conventional Estimation Methods

Nicole Bellert[1,2] and Andrea Günster[3]

[1]Institute for Wealth & Asset Management, Zurich University of Applied Sciences, Gertrudstrasse 8, Winterthur, 8401, Zurich, Switzerland.
[2]Department of Informatics, University of Zurich, Binzmühlestrasse 14, Zurich, 8050, Zurich, Switzerland.
[3]Institute of Business Information Technology, Zurich University of Applied Sciences, Theaterstrasse 17, Winterthur, 8400, Zurich, Switzerland.

Contributing authors: bell@zhaw.ch; gues@zhaw.ch;

This online appendix contains result tables for the individual models we simulated. Linear regression, hazard rate (HR) estimation, Lasso cross-validation (CV) Regression (Tibshirani (1996)), and regressions corrected for Heckman Sample Selection (Heckman (1979)) are applied on data simulated based on Model I (Stigler (1964), II (Stigler (1964) and Harrington and Wei (2017), and III (Stigler (1964) and Bos et al (2018).

Appendix A Lasso Results

Table A1 Lasso CV Regression and HR for Cartel Duration on Model I - ICC on Stigler - Detection independent of Collusion

	Ln(Duration+1)			Cartel Death		
	LasSample	LasUndetec	LasCartels	HRLasSample	HRLasUnd	HRLasCartels
Firms n_f	−2.42	9.04	5.53	2.08	26.42	−4.29
	(0.31)	(0.25)	(0.19)	(0.25)	(1.58)	(0.16)
n_f^2	0.80	−2.41	−1.45	−0.69	−3.86	1.16
	(0.08)	(0.06)	(0.05)	(0.07)	(0.28)	(0.04)
n_f^3	−0.08	0.18	0.10	0.07	0.19	−0.08
	(0.01)	(0.005)	(0.004)	(0.01)	(0.02)	(0.003)
Detection Prob. σ	4.21	0.14	−1.65	−2.47	2.98	4.18
	(4.70)	(2.49)	(2.28)	(4.07)	(2.09)	(1.76)
σ^2	−23.18	−30.49	−28.72	14.38	23.30	20.83
	(21.34)	(11.66)	(10.67)	(18.51)	(9.23)	(8.18)
σ^3	31.71	43.58	41.73	−19.47	−34.86	−31.98
	(30.41)	(17.19)	(15.63)	(26.37)	(13.61)	(11.97)
$n_f\sigma$	−0.04	1.07	1.32	0.06	−1.25	−1.37
	(0.12)	(0.08)	(0.06)	(0.11)	(0.13)	(0.05)
Start	−0.002	−0.002	−0.002	0.002	0.001	0.001
	(0.0001)	(0)	(0)	(0.0001)	(0)	(0)
Times Caught	0.10	0.56	0.48	−0.08	−1	−0.33
	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
Constant	7.70	−3.10	0.25			
	(0.49)	(0.36)	(0.28)			
Observations	9'526	38'526	48'052	9'526	38'526	48'052
R^2	0.21	0.56	0.6			
Adjusted R^2	0.21	0.56	0.6			
Log Likelihood				−62'542.89	−113'921.10	−186'376.20

Note: This table shows the estimation results of Lasso CV feature selected linear cross-sectional regressions to explain cartel duration (ln(duration+1)) and the estimation results of a Lasso CV feature selected Weibull Hazard Model to explain cartel death, both at the industry level, for data simulated for Model I. Columns 2 - 4 estimate linear regression coefficients, while columns 5 - 7 estimate HR coefficients, both on the sample of detected cartels, the group of undetected cartels and the population of all cartels, respectively. The estimated coefficients show standard errors in the sample, but do not test for the real population. The estimated HR coefficients show the change of risk for cartel breakdown if the covariate increases by 1 unit, keeping all others fixed.

Table A2 Lasso CV Regression and HR for Cartel Duration on Model II - ICC on Stigler - Detection depends on number of Firms

	Ln(Duration+1)			Cartel Death		
	LasSample	LasUndetec	LasCartels	HRLasSample	HRLasUnd	HRLasCartels
Firms n_f	−3.07	−3.07	8.33	2.87	25.08	−7.35
	(0.51)	(0.51)	(0.23)	(0.47)	(1.56)	(0.24)
n_f^2	0.98	0.98	−2.17	−0.93	−3.56	2.00
	(0.14)	(0.14)	(0.06)	(0.13)	(0.28)	(0.06)
n_f^3	−0.10	−0.10	0.15	0.10	0.17	−0.15
	(0.01)	(0.01)	(0.004)	(0.01)	(0.02)	(0.004)
Detection Prob. σ	0.86	0.86	1.64	−1.04	4.15	1.63
	(1.17)	(1.17)	(2.46)	(1.08)	(0.79)	(1.91)
σ^2			−26.78			22.87
			(11.55)			(8.83)
σ^3	−6.91	−6.91	39.27	3.26	−0.45	−34.13
	(4.80)	(4.80)	(16.98)	(4.39)	(1.34)	(12.99)
$n_f\sigma$	0.07	0.07	0.70	0.19	−0.67	−1.08
	(0.20)	(0.20)	(0.07)	(0.19)	(0.13)	(0.08)
Start	−0.002	−0.002	−0.003	0.002	0.001	0.001
	(0.0001)	(0.0001)	(0)	(0.0001)	(0)	(0)
Times Caught	0.23	0.23	0.54	−0.17	−1.26	−0.31
	(0.05)	(0.05)	(0.02)	(0.05)	(0.04)	(0.01)
Constant	8.52	8.52	−2.72			
	(0.63)	(0.63)	(0.33)			
Observations	3'097	3'097	41'665	3'097	38'568	41'665
R^2	0.27	0.27	0.62			
Adjusted R^2	0.27	0.27	0.62			
Log Likelihood				−21'044.94	−116'132.20	−142'491.60

Note: This table shows the estimation results of Lasso CV feature selected linear cross-sectional regressions to explain cartel duration (ln(duration+1)) and the estimation results of a Lasso CV feature selected Weibull Hazard Model to explain cartel death, both at the industry level, for data simulated for Model II. Columns 2 - 4 estimate linear regression coefficients, while columns 5 - 7 estimate HR coefficients, both on the sample of detected cartels, the group of undetected cartels and the population of all cartels, respectively. The estimated coefficients show standard errors in the sample, but do not test for the real population. The estimated HR coefficients show the change of risk for cartel breakdown if the covariate increases by 1 unit, keeping all others fixed.

Table A3 Cartel Duration with CV Lasso on Model III - ICC on Harrington et al.

	Ln(Duration+1)			Cartel Death		
	LasSample	LasUndetec	LasCartels	HRLasSample	HRLasUnd	HRLasCartels
Firms n_f	1.04		4.56	−0.78		−2.89
	(0.02)		(0.06)	(0.02)		(0.04)
n_f^2		−0.46	−1.41		1.02	1.00
		(0.004)	(0.02)		(0.01)	(0.01)
n_f^3	−0.03	0.04	0.11	0.02	−0.11	−0.08
	(0.0004)	(0.0005)	(0.001)	(0.0004)	(0.001)	(0.001)
γ		−0.04			0.01	
		(0.68)			(0.50)	
γ^2		−0.04			0.04	
		(0.42)			(0.31)	
γ^3	−0.01		−0.05	0.01		0.03
	(0.02)		(0.01)	(0.02)		(0.01)
Leniency (% Fine) θ	−0.15	0.41	0.18	0.09	−0.89	−0.18
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
$n_f\theta$	0.07	−0.05	0.01	−0.05	0.17	0.01
	(0.01)	(0.01)	(0.004)	(0.01)	(0.01)	(0.003)
Detection Prob. σ	4.89	−10.95	−2.56	−3.27	33.85	8.38
	(1.24)	(0.72)	(0.65)	(1.08)	(0.57)	(0.48)
σ^2	−3.61	−25.64	−28.43	1.48	34.49	17.68
	(5.62)	(3.34)	(3.02)	(4.91)	(2.45)	(2.18)
σ^3	−0.02	47.26	44.05	2.24	−86.43	−36.94
	(8.07)	(4.96)	(4.44)	(7.05)	(3.65)	(3.20)
$n_f\sigma$	−2.03	1.68	0.11	1.51	−6.61	−1.13
	(0.05)	(0.03)	(0.03)	(0.04)	(0.05)	(0.02)
Structured	−0.14	−0.31	−0.32	0.12	0.18	0.21
	(0.01)	(0.004)	(0.004)	(0.01)	(0.003)	(0.003)
Start	−0.002	−0.002	−0.002	0.002	0.001	0.001
	(0)	(0)	(0)	(0)	(0)	(0)
Times Caught	0.01	0.53	0.45	0.001	−0.84	−0.28
	(0.004)	(0.003)	(0.002)	(0.004)	(0.01)	(0.002)
Constant	3.85	10.08	2.73			
	(0.10)	(0.28)	(0.09)			
Observations	128'055	545'857	673'912	128'055	545'857	673'912
R^2	0.22	0.47	0.55			
Adjusted R^2	0.22	0.47	0.55			
Log Likelihood				−833'467	−1'662'869	−2'605'274

Note: This table shows the estimation results of Lasso CV feature selected linear cross-sectional regressions to explain cartel duration (ln(duration+1)) and the estimation results of a Lasso CV feature selected Weibull Hazard Model to explain cartel death, both at the industry level, for data simulated for Models IIIa and IIIb combined. Columns 2 - 4 estimate linear regression coefficients, while columns 5 - 7 estimate HR coefficients, both on the sample of detected cartels, the group of undetected cartels and the population of all cartels, respectively. The estimated coefficients show standard errors in the sample, but do not test for the real population. The estimated HR coefficients show the change of risk for cartel breakdown if the covariate increases by 1 unit, keeping all others fixed.

Appendix B Heckman Sample Selection Correction Results

Table B4 Linear Regression and HR with and without Heckman Correction for Sample Selection on Model I - ICC on Stigler - Detection independent of Collusion

	Ln(Duration+1)				Cartel Death			
	mlrSample	mlrUndetect	mlrCartels	mlrHeck	HRSample	HRUndetect	HRCartels	HRHeck
Firms n_f	−0.22	−1.13	−0.89	0.24	0.15	1.08	0.69	−0.23
	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.03)
Detection Prob. σ	−1.23	−0.84	−1.18	−5.09	1.06	0.65	1.15	4.96
	(0.16)	(0.09)	(0.08)	(0.31)	(0.13)	(0.07)	(0.06)	(0.36)
Start	−0.002	−0.002	−0.002	−0.001	0.002	0.001	0.001	0.0004
	(0.0001)	(0)	(0)	(0.0001)	(0.0001)	(0)	(0)	(0.0001)
Times Caught	0.16	0.59	0.52	−0.22	−0.13	−1.12	−0.35	0.48
	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
IMR				−2.16				2
				(0.12)				(0.14)
Constant	6.32	9.51	8.24	7.26				
	(0.06)	(0.05)	(0.04)	(0.05)				
Observations	9'526	38'526	48'052	7'845	9'526	38'526	48'052	7'845
R^2	0.16	0.53	0.57	0.24				
Adjusted R^2	0.16	0.53	0.57	0.24				
Log Likelihood					−62'789.33	−115'376.30	−187'926	−50'911.10

Note: This table shows the estimation results of linear cross-sectional regressions to explain cartel duration (ln(duration+1)) and the estimation results of a Weibull Hazard Model to explain cartel death, both at the industry level, for data simulated for Model I. Columns 2 - 5 estimate linear regression coefficients, while columns 6 - 9 estimate HR coefficients, both on the sample of detected cartels, the group of undetected cartels, the population of all cartels, and the sample corrected for Heckman Sample Selection, respectively. The estimated coefficients show standard errors in the sample, but do not test for the real population. The estimated HR coefficients show the change of risk for cartel breakdown if the covariate increases by 1 unit, keeping all others fixed.

Table B5 Linear Regression and HR with and without Heckman Correction for Sample Selection on Model II

	Ln(Duration+1)				Cartel Death			
	mlrSample	mlrUndetect	mlrCartels	mlrHeck	HRSample	HRUndetect	HRCartels	HRHeck
Firms n_f	−0.23	−1.23	−1.13	1.18	0.15	1.18	0.99	−0.99
	(0.02)	(0.01)	(0.01)	(0.17)	(0.02)	(0.01)	(0.01)	(0.17)
Detection Prob. σ	−0.18	−0.34	−0.39	−17.40	0.13	0.28	0.37	14.44
	(0.26)	(0.09)	(0.09)	(2.07)	(0.23)	(0.07)	(0.06)	(2.11)
Start	−0.003	−0.003	−0.002	−0.002	0.003	0.002	0.001	0.001
	(0.0001)	(0)	(0)	(0.0001)	(0.0001)	(0)	(0)	(0.0002)
Times Caught	0.35	0.90	0.63	−0.15	−0.28	−1.36	−0.33	0.33
	(0.05)	(0.02)	(0.02)	(0.03)	(0.05)	(0.04)	(0.01)	(0.04)
IMR				−6.2				5.16
				(0.72)				(0.73)
Constant	6.16	10.10	9.51	12.84				
	(0.11)	(0.04)	(0.04)	(0.72)				
Observations	3'097	38'568	41'665	2'981	3'097	38'568	41'665	2'981
R^2	0.21	0.58	0.59	0.22				
Adjusted R^2	0.20	0.58	0.59	0.22				
Log Likelihood					−21'146.67	−117'698.60	−143'358.40	−20'254

Note: This table shows the estimation results of linear cross-sectional regressions to explain cartel duration (ln(duration+1)) and the estimation results of a Weibull Hazard Model to explain cartel death, both at the industry level, for data simulated for Model II. Columns 2 - 5 estimate linear regression coefficients, while columns 6 - 9 estimate HR coefficients, both on the sample of detected cartels, the group of undetected cartels, the population of all cartels, and the sample corrected for Heckman Sample Selection, respectively. The estimated coefficients show standard errors in the sample, but do not test for the real population. The estimated HR coefficients show the change of risk for cartel breakdown if the covariate increases by 1 unit, keeping all others fixed.

Table B6 Linear Regression and HR with and without Heckman Correction for Sample Selection on Model III

	Ln(Duration+1)				Cartel Death			
	mlrSample	mlrUndetect	mlrCartels	mlrHeck	HRSample	HRUndetect	HRCartels	HRHeck
Firms n_f	−0.30	−1.35	−1.18	0.17	0.22	1.06	0.85	−0.14
	(0.004)	(0.003)	(0.002)	(0.02)	(0.004)	(0.003)	(0.002)	(0.02)
γ	−0.002	−0.11	−0.09	0.04	0.01	0.10	0.07	−0.01
	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)
Leniency (% Fine) θ	0.05	0.22	0.19	0.002	−0.03	−0.18	−0.14	−0.0003
	(0.01)	(0.005)	(0.004)	(0.01)	(0.01)	(0.004)	(0.003)	(0.01)
Detection Prob. σ	−2.34	−7.82	−6.63	−3.91	1.89	6.24	5.09	3.80
	(0.04)	(0.03)	(0.02)	(0.07)	(0.04)	(0.02)	(0.02)	(0.08)
Structured	−0.13	−0.32	−0.31	−0.04	0.11	0.17	0.21	0.01
	(0.01)	(0.004)	(0.004)	(0.01)	(0.01)	(0.003)	(0.003)	(0.01)
Start	−0.002	−0.002	−0.002	−0.001	0.002	0.001	0.001	0.001
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Times Caught	0.05	0.49	0.42	−0.19	−0.03	−0.88	−0.26	0.42
	(0.004)	(0.003)	(0.002)	(0.003)	(0.004)	(0.01)	(0.002)	(0.003)
IMR				−1.63				1.53
				(0.04)				(0.05)
Constant	6.89	10.85	9.93	7.04				
	(0.04)	(0.03)	(0.02)	(0.03)				
Observations	128'055	545'857	673'912	100'925	128'055	545'857	673'912	100'925
R^2	0.19	0.47	0.54	0.25				
Adjusted R^2	0.19	0.47	0.54	0.25				
Log Likelihood					−835'063.30	−1'676'470.00	−2'613'520	−652'741.70

Note: This table shows the estimation results of linear cross-sectional regressions to explain cartel duration (ln(duration+1)) and the estimation results of a Weibull Hazard Model to explain cartel death, both at the industry level, for data simulated for Model III. Columns 2 - 5 estimate linear regression coefficients, while columns 6 - 9 estimate HR coefficients, both on the sample of detected cartels, the group of undetected cartels, the population of all cartels, and the sample corrected for Heckman Sample Selection, respectively. The estimated coefficients show standard errors in the sample, but do not test for the real population. The estimated HR coefficients show the change of risk for cartel breakdown if the covariate increases by 1 unit, keeping all others fixed.

Appendix C In and Out of Sample Performance

Table C7 In Sample and Out of Sample R^2 and MSE (Model I)

	MLR	MLR IMR	LASSO
Detected Cases (Sample)			
R^2	0.16	0.24	0.21
MSE	1.42	0.79	1.33
Population			
R^2	-0.02	0.35	0.24
MSE	5.06	1.98	3.76
Undetected Cases (Population - Sample)			
R^2	-0.36	0.34	0.00
MSE	5.96	3.95	4.36

This table shows the R^2 and MSE for the linear regression, the linear regression corrected for sample selection bias (Heckman (1979)) and the LASSO CV linear regression for Model I. We calculate it as in-sample R^2 and MSE as in regular regression output. There, we do not see large differences between the models. When we calculate out-of-sample for undetected cartels, or the population of all cartels, correcting for selection bias improves the results compared to linear regression.

Table C8 In Sample and Out of Sample R^2 and MSE (Model II)

	MLR	MLR IMR	LASSO
Detected Cases (Sample)			
R^2	0.21	0.22	0.27
MSE	1.27	1.00	1.17
Population			
R^2	0.07	0.47	0.30
MSE	4.96	2.29	3.73
Undetected Cases (Population - Sample)			
R^2	-0.05	0.48	0.22
MSE	5.25	2.84	3.94

This table shows the R^2 and MSE for the linear regression, the linear regression corrected for sample selection bias (Heckman (1979)) and the LASSO CV linear regression for Model II. We calculate it as in-sample R^2 and MSE as in regular regression output. There, we do not see large differences between the models. When we calculate out-of-sample for undetected cartels, or the population of all cartels, correcting for selection bias improves the results compared to linear regression.

Table C9 In Sample and Out of Sample R^2 and MSE (Model III)

	MLR	MLR IMR	LASSO
Detected Cases (Sample)			
R^2	0.19	0.25	0.22
MSE	1.36	0.79	1.31
Population			
R^2	-0.21	0.33	0.01
MSE	5.82	2.50	4.77
Undetected Cases (Population - Sample)			
R^2	-0.66	0.34	-0.35
MSE	6.87	4.50	5.58

This table shows the R^2 and MSE for the linear regression, the linear regression corrected for sample selection bias (Heckman (1979)) and the LASSO CV linear regression for Model III. We calculate it as in-sample R^2 and MSE as in regular regression output. There, we do not see large differences between the models. When we calculate out-of-sample for undetected cartels, or the population of all cartels, correcting for selection bias improves the results compared to linear regression.

Appendix D Quantifying the Bias

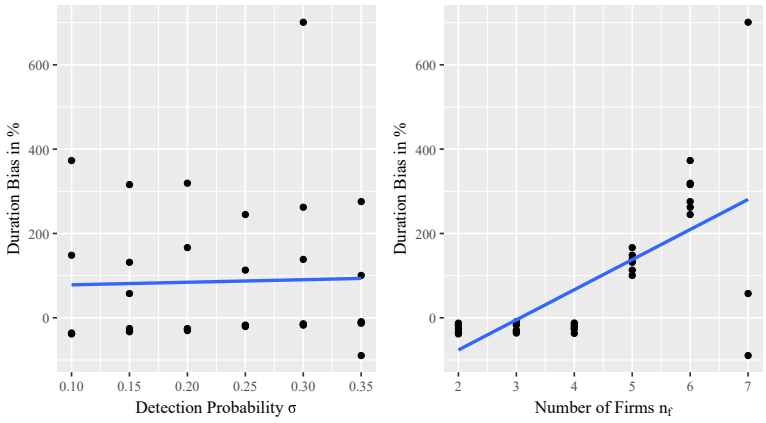

Fig. D1 This figure shows the bias between ADT and ACD for each group of equal variables in Model I.

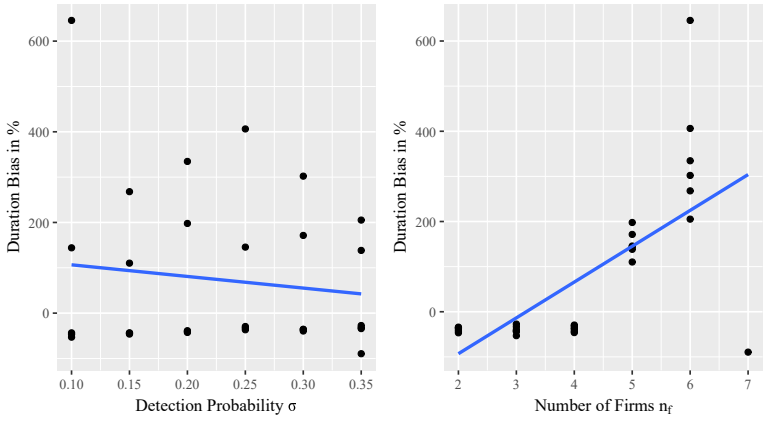

Fig. D2 This figure shows the bias between ADT and ACD for each group of equal variables in Model II.

References

Bos I, Davies SW, Harrington JE, et al (2018) Does Enforcement Deter Cartels? A Tale of Two Tails. International Journal of Industrial Organization 59:372–405

Harrington JE, Wei Y (2017) What Can the Duration of Discovered Cartels Tell Us About the Duration of All Cartels? The Economic Journal 127(604):1977–2005

Heckman JJ (1979) Sample Selection Bias as a Specification Error. Econometrica: Journal of the econometric society pp 153–161

Stigler GJ (1964) A Theory of Oligopoly. Journal of Political Economy 72(1):44–61

Tibshirani R (1996) Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B (Methodological) 58(1):267–288