

Online Appendix to Simulating Collusion: Challenging Conventional Estimation Methods

Nicole Bellert^{1,2} and Andrea Günster³

¹Institute for Wealth & Asset Management, Zurich University of Applied Sciences, Gertrudstrasse 8, Winterthur, 8401, Zurich, Switzerland.

²Department of Informatics, University of Zurich, Binzmühlestrasse 14, Zurich, 8050, Zurich, Switzerland.

³Institute 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 simulate. Hazard rate (HR) estimation, Linear regression, 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

	Cartel Death		Ln(Duration+1)	
	HRLasSample	HRLasCartels	LasSample	LasUndetec
Firms $\ln(n_f)$	0.12*** (0.02)	0.35*** (0.004)	-0.25*** (0.02)	-0.40*** (0.01)
n_f^2	-0.01*** (0.001)	-0.01*** (0.0002)	0.01*** (0.002)	0.02*** (0.0002)
n_f^3	0.0001*** (0)	0.0001*** (0)	-0.0002*** (0)	-0.0002*** (0)
Detection Probability σ	1.23** (0.49)	0.60*** (0.16)	-1.78*** (0.59)	-0.75*** (0.24)
σ^2				
σ^3	-0.79 (2.32)	1.27 (0.80)	1.45 (2.83)	-0.96 (1.26)
$n_f \sigma$	-0.03 (0.04)	-0.05*** (0.01)	0.10*** (0.05)	0.07*** (0.01)
Start	0.002*** (0)	0.0001*** (0)	-0.0002*** (0)	-0.0004*** (0)
Constant			6.53*** (0.11)	4.30*** (0.05)
Observations	11'366	97'045	11'366	97'045
R^2				108'411
Adjusted R^2			0.20	0.09
Log Likelihood	-73'101.01	-324'924.30	0.20	0.09
			-405'501.10	0.17

Note: This table shows the estimation results of Lasso CV feature selected linear cross-sectional regressions to explain cartel duration ($\ln(\text{duration}+1)$) 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

	HRLasSample	Cartel Death HRLasUnd	HRLasCartels	LasSample	Ln(Duration+1) LasUndetec	LasCartels
Firms $\ln(n_f)$		0.27*** (0.004)	0.27*** (0.004)		-0.38*** (0.01)	-0.38*** (0.01)
n_f^2		-0.01*** (0.0002)	-0.01*** (0.0002)		0.02*** (0.0003)	0.02*** (0.0003)
n_f^3		0.0001*** (0)	0.0001*** (0)		-0.0002*** (0)	-0.0002*** (0)
Detection Probability σ		-0.10 (0.16)	-0.12 (0.16)		-0.06 (0.26)	-0.02 (0.26)
σ^3		0.93 (0.80)	0.94 (0.80)		-0.54 (1.33)	-0.58 (1.33)
$n_f \sigma$		-0.002 (0.01)	-0.002 (0.01)		0.02 (0.02)	0.02 (0.02)
Start	0.003*** (0.0004)	0.001*** (0)	0.001*** (0)	-0.003*** (0.0004)	-0.001*** (0)	-0.001*** (0)
Constant				5.82*** (0.12)	4.62*** (0.05)	4.62*** (0.05)
Observations	120	97'226	97'346	120	97'226	97'346
R ²					0.15	0.15
Adjusted R ²				0.35	0.15	0.15
Log Likelihood	-804.79	-331'179.20	-332'075.30			

Note: This table shows the estimation results of Lasso CV feature selected linear cross-sectional regressions to explain cartel duration ($\ln(\text{duration}+1)$) 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.

	Cartel Death		HRLasCartels		Ln(Duration+1)	
	HRLasSample	HRLasUnd			LasSample	LasUndetec
Firms $\ln(n_f)$	0.06*** (0.004)	0.40*** (0.002)	0.41*** (0.001)		-0.10*** (0.01)	-0.39*** (0.002)
n_f^2		-0.02*** (0.0001)	-0.02*** (0.0001)			0.02*** (0.0001)
n_f^3	-0.0000*** (0)	0.0001*** (0)	0.0002*** (0)		0.0001*** (0)	-0.0002*** (0)
Fines γ (% Profit)		0.02** (0.01)	-0.06 (0.28)			0.25 (0.47)
γ^2			0.06 (0.18)			-0.18 (0.30)
γ^3	0.02 (0.02)				-0.005 (0.02)	
Leniency θ (% Fine)	-0.03*** (0.01)	-0.12*** (0.005)	-0.13*** (0.004)		0.13*** (0.01)	0.19*** (0.01)
$n_f\theta$	0.01* (0.003)	0.01*** (0.001)	0.01*** (0.001)		-0.01*** (0.004)	-0.02*** (0.001)
Detection Probability σ	1.01*** (0.22)	3.36*** (0.05)	2.32*** (0.04)		-1.57*** (0.28)	-1.78*** (0.07)
σ^2	0.81* (0.44)				-0.56 (0.55)	
σ^3		-0.42** (0.21)	0.41** (0.20)			-2.75*** (0.34)
$n_f\sigma$	-0.01 (0.01)	-0.29*** (0.004)	-0.22*** (0.003)		-0.004 (0.02)	0.17*** (0.01)
Structured	0.07*** (0.01)	0.09*** (0.002)	0.08*** (0.002)		-0.09*** (0.01)	-0.10*** (0.003)
Start	0.002*** (0)	0.0002*** (0)	0.0004*** (0)		-0.002*** (0)	-0.001*** (0)
Constant					6.13*** (0.04)	4.88*** (0.19)
Observations	152'747	1'404'446	1'557'193		152'747	1'404'446
R^2					0.19	0.13
Adjusted R^2					0.19	0.13
Log Likelihood	-970'321.40	-4'708'936	-5'782'199			

Note: This table shows the estimation results of Lasso CV feature selected linear cross-sectional regressions to explain cartel duration ($\ln(\text{duration}+1)$) 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

	Cartel Death			Ln(Duration+1)		
	HRSample	HRUndetect	HRCartels	mlrUndetect	mlrCartels	mlrHeck
Firms $\ln(n_f)$	0.28*** (0.02)	0.81*** (0.01)	0.82*** (0.01)	-1.09*** (0.01)	-1.44*** (0.01)	-0.59*** (0.02)
Detection Probability σ	0.92*** (0.12)	0.43*** (0.04)	0.02 (0.04)	-0.39*** (0.06)	0.52*** (0.06)	-0.71*** (0.13)
Start	0.002*** (0)	0.0001*** (0)	0.0003*** (0)	-0.0003*** (0)	-0.001*** (0)	-0.001*** (0.0001)
IMR						382,419.50*** (43,828.61)
Constant				4.34*** (0.03)	5.24*** (0.03)	6.30*** (0.05)
Observations	11'366	97'045	108'411	97'045	108'411	9'507
R ²				0.19	0.17	0.26
Adjusted R ²				0.19	0.17	0.26
Log Likelihood	-73'101.14	-325'270	-405'938.80			

Note: This table shows the estimation results of linear cross-sectional regressions to explain cartel duration ($\ln(\text{duration}+1)$) 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

	Cartel Death			Ln(Duration+1)		
	HRSample	HRUndetect	HRCartels	HRHeck	mlrSample	mlrUndetect mlrCartels mlrHeck
Firms $\ln(n_f)$	-0.13 (0.23)	0.56*** (0.01)	0.56*** (0.01)	-0.11 (0.24)	0.01 (0.25)	-0.98*** (0.01)
Detection Probability σ	0.15 (1.29)	0.04 (0.04)	0.04 (0.04)	0.38 (1.32)	1.39 (1.44)	0.02 (0.07)
Start	0.003*** (0.001)	0.001*** (0)	0.001*** (0)	0.003*** (0.001)	-0.003*** (0.001)	-0.001*** (0)
IMR				2,673,699.00 (8,660,622.00)		14,160,666.00 (9,668,147.00)
Constant					5.46*** (0.47)	4.62*** (0.03)
Observations	120	97'226	97'346	120	120	97'226
R ²					0.35	0.14
Adjusted R ²					0.34	0.14
Log Likelihood	-804.63	-331'810.60	-332'709.30	-805.52		

Note: This table shows the estimation results of linear cross-sectional regressions to explain cartel duration ($\ln(\text{duration}+1)$) 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

	Cartel Death		Ln(Duration+1)	
	HRSample	HRUndetect	mlrSample	mlrUndetect
Firms $\ln(n_f)$	0.29*** (0.01)	0.71*** (0.002)	-0.54*** (0.01)	-1.28*** (0.003)
Fines γ (% Profit)	0.04 (0.03)	0.02 (0.01)	-0.01 (0.04)	-0.04** (0.02)
Leniency θ (% Fine)	-0.01** (0.01)	-0.04*** (0.002)	0.03*** (0.01)	0.08*** (0.003)
Detection Probability σ	1.39*** (0.03)	1.56*** (0.01)	-1.94*** (0.04)	-1.45*** (0.02)
Structured	0.08*** (0.01)	0.09*** (0.002)	-0.10*** (0.01)	-0.10*** (0.003)
Start	0.002*** (0)	0.0002*** (0)	-0.002*** (0)	-0.001*** (0)
IMR				
Constant				
Observations	152'747	1'404'446	152'747	1'557'193
R ²				
Adjusted R ²				
Log Likelihood	-970'106.10	-4'711'575	-782'818.70	

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 Quantifying the Bias

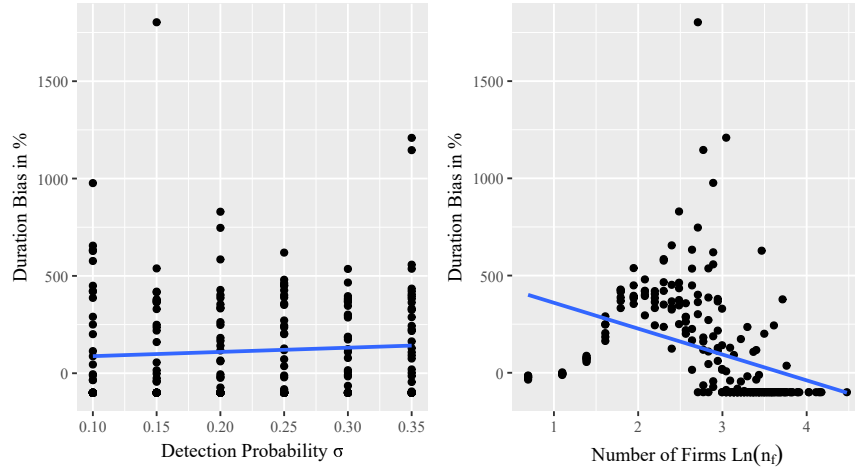


Fig. C1 This figure shows the bias between ADT and ACD for each group of equal variables in Model I.

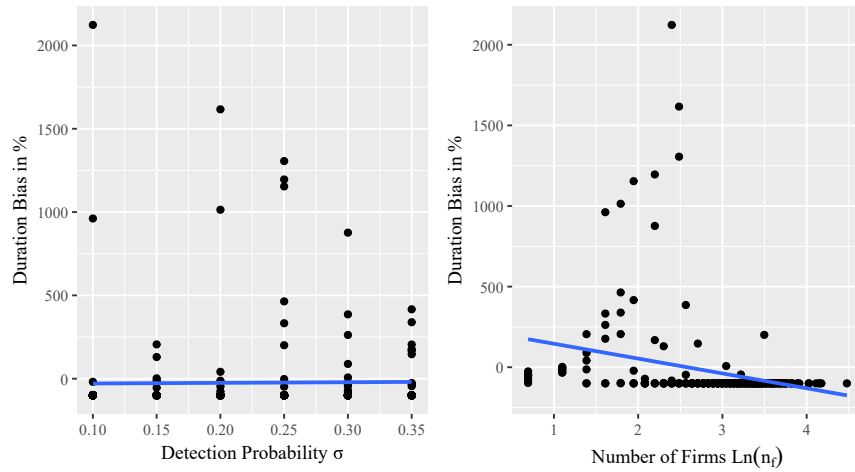


Fig. C2 This figure shows the bias between ADT and ACD for each group of equal variables in Model II.

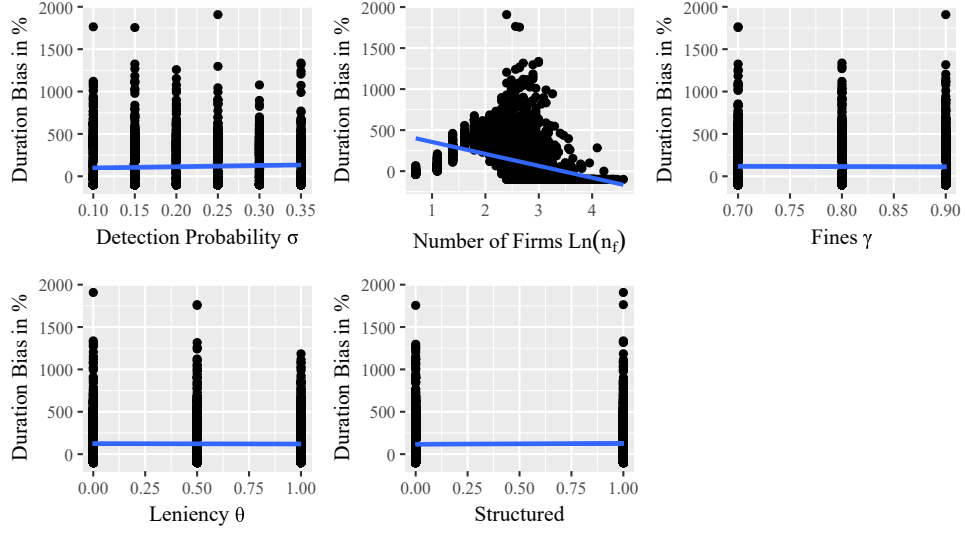


Fig. C3 This figure shows the bias between ADT and ACD for each group of equal variables in Model III.

Appendix D Confidence Intervals

Table D7 Hazard Rate 95% Confidence Intervals (Model I)

	Sample		Undetected		Population	
	Lower	Upper	Lower	Upper	Lower	Upper
Firms $\text{Ln}(n_f)$	0.235	0.330	0.790	0.823	0.809	0.840
Detection Probability σ	0.633	1.204	0.332	0.519	-0.071	0.105
Start	0.002	0.002	0	0	0	0
N	11'366		97'045		108'411	

This table compares the confidence intervals of the estimated HR coefficients for the sample of detected cartels, the group of undetected cartels, and the population of detected and undetected cartels. The confidence intervals of the estimations do not overlap between estimations of the population and sample.

Table D8 Hazard Rate 95% Confidence Intervals (Model II)

	Sample		Undetected		Population	
	Lower	Upper	Lower	Upper	Lower	Upper
Firms $\text{Ln}(n_f)$	-0.674	0.423	0.543	0.579	0.542	0.578
Detection Probability σ	-2.880	3.184	-0.049	0.136	-0.056	0.130
Start	0.002	0.004	0.001	0.001	0.001	0.001
N	120		97'226		97'346	

This table compares the confidence intervals of the estimated HR coefficients for the sample of detected cartels, the group of undetected cartels, and the population of detected and undetected cartels. Except for detection probability (σ), the confidence intervals of the estimations do not overlap between estimations of the population and sample.

Table D9 Hazard Rate 95% Confidence Intervals (Model III)

	Sample		Undetected		Population	
	Lower	Upper	Lower	Upper	Lower	Upper
Firms $\text{Ln}(n_f)$	0.275	0.299	0.705	0.712	0.756	0.763
Fines γ (% Profit)	-0.023	0.099	-0.004	0.038	0.004	0.044
Leniency θ (% Fine)	-0.027	-0.002	-0.046	-0.038	-0.049	-0.041
Detection Probability σ	1.330	1.460	1.543	1.585	1.133	1.173
Structured	0.068	0.088	0.088	0.095	0.080	0.086
Start	0.002	0.002	0	0	0	0
N	152'747		1'404'446		1'557'193	

This table compares the confidence intervals of the estimated HR coefficients for the sample of detected cartels, the group of undetected cartels, and the population of detected and undetected cartels. Except for fine (γ) and *Structured*, the confidence intervals of the estimations do not overlap between estimations of the population and sample.

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