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 (ZHAW), 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 (ZHAW), Theaterstrasse 17, Winterthur, 8400, Zurich, Switzerland.

*Corresponding author(s). E-mail(s): bell@zhaw.ch; gues@zhaw.ch;

This online appendix contains additional result tables for all estimated models.

Appendix A Linear Models Results

Table A1 Hazard Rate on Cartel Duration for Model I, II and III

	Cartel Death			
	HRSample	HRUndetect	HRCartels	HRHeck
Number of Firms n_f	0.12***	0.60***	0.37***	−0.01***
	(0.002)	(0.002)	(0.001)	(0.002)
Fines γ (% of Profit)	−0.10***	−0.05**	−0.05***	−0.11***
	(0.02)	(0.02)	(0.01)	(0.02)
Leniency (% of Fine) θ	−0.05***	−1.26***	−0.57***	0.02***
	(0.004)	(0.01)	(0.003)	(0.004)
Detection Probability σ	3.34***	4.32***	4.02***	3.02***
	(0.02)	(0.02)	(0.01)	(0.02)
Structured	0.36***	0.35***	0.42***	0.34***
	(0.003)	(0.004)	(0.002)	(0.003)
Model II	−0.48***	−0.34***	−0.37***	−0.58***
	(0.01)	(0.01)	(0.01)	(0.01)
Model III	0.22***	0.87***	0.51***	0.09***
	(0.02)	(0.02)	(0.01)	(0.02)
Start	0.001***	0.001***	0.001***	0.001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Times Caught	0.01***	−0.34***	−0.05***	0.04***
	(0.001)	(0.003)	(0.001)	(0.001)
Repeat Offender	−0.03***	−0.26***	−0.38***	0.29***
	(0.004)	(0.01)	(0.003)	(0.01)
IMR				0.56***
				(0.01)
Observations	475'456	465'402	940'858	475'456
Log Likelihood	−2'806'284.00	−1'445'284.00	−4'442'193.00	−2'804'065.00

Note: This table shows the estimation results of Weibull Hazard Model to explain cartel death at the industry level, for data simulated for all Models I, II, and III. The estimated coefficients show the change of risk for cartel breakdown, if the covariate increases by 1 unit, keeping all others fixed. Columns 2 - 5 use HR estimation 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. All estimators are significant.
Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively.

Table A2 Linear Regression on Cartel Duration for Model I, II and III

	Ln(Duration+1)			
	mlrSample	mlrUndetect	mlrCartels	mlrHeck
Number of Firms n_f	−0.16***	−0.60***	−0.50***	−0.0003
	(0.002)	(0.002)	(0.001)	(0.003)
Fines γ (% of Profit)	0.06***	0.03	0.03	0.07***
	(0.02)	(0.03)	(0.02)	(0.02)
Leniency (% of Fine) θ	0.09***	1.58***	0.84***	0.003
	(0.004)	(0.01)	(0.004)	(0.005)
Detection Probability σ	−3.66***	−4.31***	−4.67***	−3.24***
	(0.02)	(0.03)	(0.02)	(0.02)
Structured	−0.40***	−0.52***	−0.56***	−0.36***
	(0.004)	(0.005)	(0.003)	(0.004)
Model II	0.53***	0.28***	0.23***	0.67***
	(0.01)	(0.01)	(0.01)	(0.01)
Model III	−0.24***	−0.82***	−0.63***	−0.07***
	(0.02)	(0.02)	(0.02)	(0.02)
Start	−0.001***	−0.001***	−0.002***	−0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Times Caught	−0.002**	−0.01***	0.05***	−0.05***
	(0.001)	(0.002)	(0.001)	(0.001)
Repeat Offender	0.04***	1.26***	0.82***	−0.37***
	(0.01)	(0.01)	(0.005)	(0.01)
IMR				−0.71***
				(0.01)
Constant	6.38***	6.98***	7.08***	6.06***
	(0.01)	(0.02)	(0.01)	(0.01)
Observations	475'456	465'402	940'858	475'456
R^2	0.14	0.43	0.42	0.15
Adjusted R^2	0.14	0.43	0.42	0.15

Note: This table shows the estimation results of linear cross-sectional regressions to explain cartel duration (ln(duration+1)) at the industry level, for data simulated for all Models I, II, and III. Columns 2 - 5 estimate linear regression coefficients 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. Except of fines (γ), all coefficients are significant.
Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively.

Table A3 Sample Selection Bias Linear Regression - Model I

Coefficients	β^s	β^l	α_{IMR}	$bias_{IMR}$
Number of Firms n_f	-0.12	-0.02	0.15	-0.10
Start	0	0	0	0
Detection Probability σ	-2.93	-3.11	-0.27	0.18
Times Caught	0.04	-0.03	-0.10	0.07
Repeat Offender	0.21	-0.18	-0.57	0.39
IMR		-0.68		
Constant	5.79	5.87	0.11	-0.07

This table shows, for the Linear Regression of Model I, the sample selection bias that we correct with the IMR following (Heckman (1979)). β^s is the estimated coefficients in the short model without IMR. β^l is the estimated coefficients in the corrected long model including IMR. α_{IMR} is the coefficient in the auxiliary regression between each variable and IMR. The last column shows the sample selection bias: $bias_{IMR} = \beta^l(IMR) * \alpha_{IMR}$.

Table A4 Sample Selection Bias Linear Regression - Model II

Coefficients	β^s	β^l	α_{IMR}	$bias_{IMR}$
Number of Firms n_f	-0.16	-0.09	0.14	-0.07
Start	0	0	0	0
Detection Probability σ	-1.43	-1.67	-0.46	0.25
Times Caught	0.07	-0.17	-0.44	0.24
Repeat Offender	0.27	0.18	-0.16	0.09
IMR		-0.54		
Constant	6.26	6.56	0.56	-0.30

This table shows, for the Linear Regression of Model II, the sample selection bias that we correct with the IMR following (Heckman (1979)). β^s is the estimated coefficients in the short model without IMR. β^l is the estimated coefficients in the corrected long model including IMR. α_{IMR} is the coefficient in the auxiliary regression between each variable and IMR. The last column shows the sample selection bias: $bias_{IMR} = \beta^l(IMR) * \alpha_{IMR}$.

Table A5 Sample Selection Bias Linear Regression - Model III

Coefficients	β^s	β^l	α_{IMR}	$bias_{IMR}$
Number of Firms n_f	-0.18	-0.01	0.25	-0.17
Fines γ (% of Profit)	0.06	0.07	0.01	-0.01
Leniency (% of Fine) θ	0.10	0.01	-0.13	0.09
Start	0	0	0	0
Structured	-0.40	-0.36	0.06	-0.04
Detection Probability σ	-3.85	-3.31	0.79	-0.54
Times Caught	-0.01	-0.05	-0.06	0.04
Repeat Offender	0.01	-0.37	-0.56	0.38
IMR		-0.68		
Constant	6.25	6.02	-0.33	0.23

This table shows, for the Linear Regression of Model II, the sample selection bias that we correct with the IMR following (Heckman (1979)). β^s is the estimated coefficients in the short model without IMR. β^l is the estimated coefficients in the corrected long model including IMR. α_{IMR} is the coefficient in the auxiliary regression between each variable and IMR. The last column shows the sample selection bias: $bias_{IMR} = \beta^l(IMR) * \alpha_{IMR}$.

Table A6 Linear Regression and HR for Cartel Duration on Model I - ICC on Stigler - Detection independent of Collusion

	Ln(Duration+1)				Cartel Death			
	mlrSample	mlrUndetect	mlrCartels	mlrHeck	HRSample	HRUndetect	HRCartels	HRHeck
N Firms n_f	-0.12^{***}	-0.57^{***}	-0.39^{***}	-0.02^{***}	0.09^{***}	0.79^{***}	0.30^{***}	0.01^{*}
	(0.005)	(0.01)	(0.004)	(0.01)	(0.004)	(0.01)	(0.003)	(0.01)
Start	-0.001^{***}	-0.001^{***}	-0.002^{***}	-0.0002^{***}	0.001^{***}	0.001^{***}	0.001^{***}	0.001^{***}
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0001)
Detection Prob. σ	-2.93^{***}	-1.60^{***}	-2.61^{***}	-3.11^{***}	2.78^{***}	1.42^{***}	2.73^{***}	2.94^{***}
	(0.08)	(0.08)	(0.06)	(0.08)	(0.07)	(0.06)	(0.05)	(0.07)
Times Caught	0.04^{***}	-0.01	0.13^{***}	-0.03^{***}	-0.03^{***}	-0.56^{***}	-0.13^{***}	0.02^{***}
	(0.004)	(0.01)	(0.003)	(0.01)	(0.004)	(0.02)	(0.003)	(0.005)
Repeat Offender	0.21^{***}	2.06^{***}	1.20^{***}	-0.18^{***}	-0.22^{***}	-0.19^{***}	-0.59^{***}	0.08^{***}
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.05)	(0.01)	(0.02)
IMR				-0.68^{***}				0.53^{***}
				(0.03)				(0.03)
Constant	5.79^{***}	6.13^{***}	5.81^{***}	5.87^{***}				
	(0.03)	(0.05)	(0.03)	(0.03)				
Observations	36'615	41'207	7'822	36'615	36'615	41'207	7'822	36'615
R^2	0.10	0.47	0.50	0.11				
Adjusted R^2	0.10	0.47	0.50	0.11				
Log Likelihood					$-218'572.00$	$-114'465.60$	$-350'666.30$	$-218'383.30$

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.
Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively.

Table A7 Linear Regression and HR for Cartel Duration on Model II - ICC on Stigler - Detection depends on number of Firms

	Ln(Duration+1)				Cartel Death			
	mlrSample	mlrUndetect	mlrCartels	mlrHeck	HRSample	HRUndetect	HRCartels	HRHeck
N Firms n_f	-0.16^{***}	-0.97^{***}	-0.72^{***}	-0.09^{***}	0.11^{***}	0.96^{***}	0.56^{***}	0.07^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Start	-0.002^{***}	-0.002^{***}	-0.002^{***}	-0.001^{***}	0.002^{***}	0.001^{***}	0.001^{***}	0.002^{***}
	(0.0001)	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0001)
Detection Prob. σ	-1.43^{***}	-0.86^{***}	-1.17^{***}	-1.67^{***}	1.20^{***}	0.67^{***}	1.12^{***}	1.35^{***}
	(0.14)	(0.09)	(0.08)	(0.14)	(0.12)	(0.06)	(0.06)	(0.12)
Times Caught	0.07^{***}	0.75^{***}	0.68^{***}	-0.17^{***}	-0.04^{***}	-1.52^{***}	-0.49^{***}	0.10^{***}
	(0.02)	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)	(0.01)	(0.03)
Repeat Offender	0.27^{***}	-0.54^{***}	-0.31^{***}	0.18^{***}	-0.28^{***}	1.26^{***}	0.34^{***}	-0.21^{***}
	(0.04)	(0.06)	(0.03)	(0.04)	(0.03)	(0.08)	(0.03)	(0.03)
IMR				-0.54^{***}				0.33^{***}
				(0.06)				(0.05)
Constant	6.26^{***}	8.82^{***}	7.60^{***}	6.56^{***}				
	(0.05)	(0.05)	(0.04)	(0.06)				
Observations	11'733	41'445	53'178	11'733	11'733	41'445	53'178	11'733
R^2	0.17	0.51	0.56	0.17				
Adjusted R^2	0.16	0.51	0.56	0.17				
Log Likelihood					$-77'014.56$	$-120'764.40$	$-209'666.50$	$-76'995.30$

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.
Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively.

Table A8 Linear Regression and HR for Cartel Duration on Model III - ICC on Harrington et al.

	Ln(Duration+1)				Cartel Death			
	mlrSample	mlrUndetect	mlrCartels	mlrHeck	HRSample	HRUndetect	HRCartels	HRHeck
N Firms n_f	-0.18^{***}	-0.56^{***}	-0.49^{***}	-0.01^{**}	0.13^{***}	0.58^{***}	0.37^{***}	-0.001
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.003)
Fines γ (% of Profit)	0.06^{***}	0.02	0.02	0.07^{***}	-0.10^{***}	-0.05^{**}	-0.05^{***}	-0.11^{***}
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Leniency (% Fine) θ	0.10^{***}	1.59^{***}	0.87^{***}	0.01	-0.05^{***}	-1.28^{***}	-0.59^{***}	0.02^{***}
	(0.004)	(0.01)	(0.004)	(0.005)	(0.004)	(0.01)	(0.003)	(0.004)
Start	-0.001^{***}	-0.001^{***}	-0.002^{***}	-0.0002^{***}	0.001^{***}	0.001^{***}	0.001^{***}	0.001^{***}
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Structured	-0.40^{***}	-0.52^{***}	-0.56^{***}	-0.36^{***}	0.37^{***}	0.38^{***}	0.44^{***}	0.34^{***}
	(0.004)	(0.005)	(0.003)	(0.004)	(0.003)	(0.004)	(0.002)	(0.003)
Detection Prob. σ	-3.85^{***}	-4.83^{***}	-5.16^{***}	-3.31^{***}	3.51^{***}	5.02^{***}	4.43^{***}	3.10^{***}
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Times Caught	-0.01^{***}	-0.004^{**}	0.04^{***}	-0.05^{***}	0.01^{***}	-0.31^{***}	-0.05^{***}	0.04^{***}
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)
Repeat Offender	0.01^{**}	1.16^{***}	0.76^{***}	-0.37^{***}	-0.01	-0.26^{***}	-0.35^{***}	0.29^{***}
	(0.01)	(0.01)	(0.005)	(0.01)	(0.005)	(0.01)	(0.004)	(0.01)
IMR				-0.68^{***}				0.54^{***}
				(0.01)				(0.01)
Constant	6.25^{***}	6.04^{***}	6.53^{***}	6.02^{***}				
	(0.02)	(0.03)	(0.02)	(0.02)				
Observations	427'108	382'750	809'858	427'108	427'108	382'750	809'858	427'108
R^2	0.14	0.43	0.41	0.14				
Adjusted R^2	0.14	0.43	0.41	0.14				
Log Likelihood					$-2'509'783.00$	$-1'200'409.00$	$-3'867'860.00$	$-2'507'956.00$

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 Models IIIa and IIIb. 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.
Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively.

Appendix B Lasso Results

Table B9 Sample Selection Bias Lasso Regression - Models I

Coefficients	β^s	β^l	α_{IMR}	$bias_{IMR}$
Number of Firms n_f	0.26	0.16	-0.22	0.11
n_f^3	-0.01	0	0.01	0
Detection Probability σ	-4.01	-4.25	-0.50	0.24
σ^3	3.93	5.97	4.30	-2.04
$n_f\sigma$	0.12	0.04	-0.17	0.08
Start	0	0	0	0
Times Caught	0.02	-0.02	-0.09	0.04
Repeat Offender	0.18	-0.05	-0.48	0.23
IMR		-0.47		
Constant	5.09	5.54	0.95	-0.45

This table shows, for the Lasso CV Linear Regression of Model I, the sample selection bias that we correct with the IMR following (Heckman (1979)). β^s is the estimated coefficients in the short model without IMR. β^l is the estimated coefficients in the corrected long model including IMR. α_{IMR} is the coefficient in the auxiliary regression between each variable and IMR. The last column shows the sample selection bias: $bias_{IMR} = \beta^l(IMR) * \alpha_{IMR}$.

Appendix C Additional Result Tables

Table B10 Sample Selection Bias Lasso Regression - Models II

Coefficients	β^s	β^l	α_{IMR}	$bias_{IMR}$
Number of Firms n_f	0.33	0.27	-0.20	0.05
n_f^3	-0.01	-0.01	0.01	0
Detection Probability σ	-2.34	-2.64	-1.09	0.30
σ^3	3.19	4.81	5.96	-1.62
$n_f\sigma$	0.11	0.08	-0.13	0.04
Start	0	0	0	0
Times Caught	0.04	-0.08	-0.42	0.11
Repeat Offender	0.21	0.17	-0.15	0.04
IMR		-0.27		
Constant	5.32	5.71	1.44	-0.39

This table shows, for the Lasso CV Linear Regression of Model II, the sample selection bias that we correct with the IMR following (Heckman (1979)). β^s is the estimated coefficients in the short model without IMR. β^l is the estimated coefficients in the corrected long model including IMR. α_{IMR} is the coefficient in the auxiliary regression between each variable and IMR. The last column shows the sample selection bias: $bias_{IMR} = \beta^l(IMR) * \alpha_{IMR}$.

Table B11 Sample Selection Bias Lasso Regression - Models III

Coefficients	β^s	β^l	α_{IMR}	$bias_{IMR}$
Number of Firms n_f	0.40	0.19	-0.46	0.21
n_f^3	-0.01	-0.01	0.01	-0.01
Detection Probability σ	-5.44	-9.09	-8.03	3.65
σ^2	11.03	22.02	24.20	-10.99
σ^3	-6.74	-20.94	-31.27	14.20
$n_f\sigma$	-0.83	-0.36	1.02	-0.46
γ^3	0.03	0.03	0.01	0
Leniency (% of Fine) θ	-0.22	-0.14	0.17	-0.08
$n_f\theta$	0.11	0.06	-0.11	0.05
Structured	-0.39	-0.37	0.04	-0.02
Start	0	0	0	0
Times Caught	-0.01	-0.04	-0.07	0.03
Repeat Offender	-0.04	-0.26	-0.48	0.22
IMR		-0.45		
Constant	5.41	6.18	1.68	-0.76

This table shows, for the Lasso CV Linear Regression of Model III, the sample selection bias that we correct with the IMR following (Heckman (1979)). β^s is the estimated coefficients in the short model without IMR. β^l is the estimated coefficients in the corrected long model including IMR. α_{IMR} is the coefficient in the auxiliary regression between each variable and IMR. The last column shows the sample selection bias: $bias_{IMR} = \beta^l(IMR) * \alpha_{IMR}$.

Table B12 Sample Selection Bias Lasso Regression - Models I, II and III

Coefficients	β^s	β^l	α_{IMR}	$bias_{IMR}$
Number of Firms n_f	0.16	0.02	-0.23	0.14
n_f^3	-0.01	0	0.01	-0.01
Detection Probability σ	-8.44	-11.34	-4.81	2.90
σ^2	14.02	26.38	20.47	-12.36
σ^3	-8.44	-25.83	-28.81	17.40
$n_f\sigma$	-0.07	0.09	0.28	-0.17
Fines γ (% of Profit)	1.77	1.66	-0.17	0.10
γ^3	-1.43	-1.34	0.14	-0.09
Leniency (% of Fine) θ	0.03	-0.05	-0.14	0.08
$n_f\theta$	0.02	0.02	0	0
Structured	-0.37	-0.36	0.02	-0.01
Model II	0.54	0.65	0.18	-0.11
Model III	-0.67	-0.44	0.37	-0.23
Start	0	0	0	0
Times Caught	0	-0.04	-0.07	0.04
Repeat Offender	0.02	-0.30	-0.53	0.32
IMR		-0.60		
Constant	6.25	6.75	0.82	-0.49

This table shows, for the Lasso CV Linear Regression of the combined Models I, II, and III, the sample selection bias that we correct with the IMR following (Heckman (1979)). β^s is the estimated coefficients in the short model without IMR. β^l is the estimated coefficients in the corrected long model including IMR. α_{IMR} is the coefficient in the auxiliary regression between each variable and IMR. The last column shows the sample selection bias: $bias_{IMR} = \beta^l(IMR) * \alpha_{IMR}$.

Table B13 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	LasHeck	HRLasSample	HRLasUnd	HRLasCartels	HRLasHeck
Start	−0.001***	−0.001***	−0.002***	−0.0004***	0.001***	0.001***	0.001***	0.001***
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0001)
N Firms n_f	−1.37***	4.87***	2.50***	0.16***	1.04***	12.61***	−1.49***	−0.14***
	(0.14)	(0.17)	(0.09)	(0.02)	(0.12)	(1.42)	(0.07)	(0.02)
n_f^2	0.43***	−1.29***	−0.64***		−0.33***	−1.56***	0.38***	
	(0.04)	(0.04)	(0.02)		(0.03)	(0.25)	(0.02)	
n_f^3	−0.04***	0.09***	0.04***	−0.005***	0.03***	0.07***	−0.02***	0.004***
	(0.003)	(0.003)	(0.002)	(0.0004)	(0.002)	(0.01)	(0.001)	(0.0004)
Detection Prob. σ	−5.04**	−17.14***	−12.64***	−4.25***	4.97**	18.48***	12.14***	3.72***
	(2.38)	(2.25)	(1.71)	(0.36)	(2.05)	(2.12)	(1.38)	(0.31)
σ^2	4.74	26.19**	13.83*		−6.61	−15.51*	−15.50**	
	(10.82)	(10.53)	(7.93)		(9.34)	(9.02)	(6.41)	
σ^3	−2.64	−34.76**	−17.19	5.97***	5.93	22.24*	18.06*	−4.94***
	(15.46)	(15.51)	(11.53)	(1.49)	(13.35)	(13.29)	(9.30)	(1.29)
$n_f\sigma$	0.12**	1.69***	1.39***	0.04	−0.06	−2.31***	−1.12***	0.005
	(0.06)	(0.07)	(0.04)	(0.06)	(0.05)	(0.16)	(0.03)	(0.05)
Times Caught	0.02***	0.03***	0.12***	−0.02***	−0.02***	−0.49***	−0.12***	0.01
	(0.004)	(0.01)	(0.003)	(0.01)	(0.004)	(0.02)	(0.003)	(0.004)
Repeat Offender	0.17***	1.75***	0.99***	−0.05*	−0.19***	−0.23***	−0.51***	−0.03
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.05)	(0.01)	(0.02)
IMR				−0.47***				0.33***
				(0.03)				(0.03)
Constant	7.03***	1.06***	3.17***	5.54***				
	(0.24)	(0.28)	(0.17)	(0.09)				
Observations	36'615	41'207	7'822	36'615	36'615	41'207	7'822	36'615
R^2	0.12	0.49	0.53	0.13				
Adjusted R^2	0.12	0.49	0.53	0.13				
Log Likelihood					−218'190.40	−113'185.60	−348'310.50	−218'178.10

Note: This table shows the estimation results of Lasso CV linear cross-sectional regressions to explain cartel duration (ln(duration+1)) and the estimation results of a Lasso CV 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.
Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively.

Table B14 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	LasHeck	HRLasSample	HRLasUnd	HRLasCartels	HRLasHeck
Start	−0.002***	−0.002***	−0.002***	−0.001***	0.002***	0.001***	0.001***	0.002***
	(0.0001)	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0001)
N Firms n_f	−1.98***	8.53***	5.88***	0.27***	1.37***	7.67***	−4.30***	−0.26***
	(0.24)	(0.18)	(0.13)	(0.04)	(0.20)	(1.48)	(0.11)	(0.03)
n_f^2	0.61***	−2.16***	−1.48***		−0.43***	−0.55**	1.08***	
	(0.06)	(0.04)	(0.03)		(0.05)	(0.26)	(0.03)	
n_f^3	−0.06***	0.15***	0.10***	−0.01***	0.04***	−0.0004	−0.07***	0.01***
	(0.01)	(0.003)	(0.002)	(0.001)	(0.004)	(0.01)	(0.002)	(0.001)
Detection Prob. σ	−2.36***	−5.81***	−7.30***	−2.64***	4.17	5.79***	8.33***	1.84***
	(0.61)	(0.51)	(0.40)	(0.61)	(3.61)	(2.11)	(1.68)	(0.55)
σ^2					−11.18	0.25	−5.39	
					(16.44)	(9.01)	(7.83)	
σ^3	3.12	−2.29	−0.19	4.81*	12.46	1.97	7.29	−4.34*
	(2.52)	(1.64)	(1.46)	(2.55)	(23.52)	(13.27)	(11.46)	(2.29)
$n_f\sigma$	0.13	0.97***	1.19***	0.08	0.01	−0.94***	−1.11***	0.04
	(0.10)	(0.08)	(0.05)	(0.10)	(0.09)	(0.16)	(0.05)	(0.09)
Times Caught	0.02	0.62***	0.54***	−0.08**	−0.01	−1.30***	−0.41***	0.03
	(0.02)	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)	(0.01)	(0.03)
Repeat Offender	0.20***	−0.44***	−0.23***	0.17***	−0.22***	1.07***	0.26***	−0.21***
	(0.03)	(0.05)	(0.03)	(0.04)	(0.03)	(0.08)	(0.03)	(0.03)
IMR				−0.27***				0.12**
				(0.06)				(0.06)
Constant	7.99***	−2.81***	−0.25	5.71***				
	(0.30)	(0.26)	(0.19)	(0.16)				
Observations	11'733	41'445	53'178	11'733	11'733	41'445	53'178	11'733
R^2	0.20	0.54	0.59	0.20				
Adjusted R^2	0.20	0.54	0.59	0.19				
Log Likelihood					−76'836.46	−118'987.30	−207'949.80	−76'867.61

Note: This table shows the estimation results of Lasso CV linear cross-sectional regressions to explain cartel duration (ln(duration+1)) and the estimation results of a Lasso CV 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.
Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively.

Table B15 Cartel Duration with CV Lasso on Model III - ICC on Harrington et al.

	Ln(Duration+1)				Cartel Death			
	LasSample	LasUndetec	LasCartels	LasHeck	HRLasSample	HRLasUnd	HRLasCartels	HRLasHeck
Start	−0.001***	−0.001***	−0.002***	−0.0004***	0.001***	0.001***	0.001***	0.001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
N Firms n_f	0.41***	−0.40***	0.58***	0.19***	−0.33***	1.96***	−0.73***	−0.14***
	(0.01)	(0.06)	(0.04)	(0.01)	(0.01)	(0.06)	(0.03)	(0.01)
n_f^2		−0.13***	−0.22***			0.08***	0.22***	
		(0.02)	(0.01)			(0.01)	(0.01)	
n_f^3	−0.01***	0.01***	0.01***	−0.01***	0.01***	−0.02***	−0.01***	0.005***
	(0.0002)	(0.001)	(0.001)	(0.0003)	(0.0002)	(0.001)	(0.001)	(0.0002)
Fines γ (% of Profit)	1.40***		0.04**		−0.54**		−0.06***	
	(0.31)		(0.02)		(0.26)		(0.01)	
γ^3	−0.70***	0.02*		0.03***	0.23*	−0.04***		−0.05***
	(0.16)	(0.01)		(0.01)	(0.14)	(0.01)		(0.01)
Leniency (% Fine) θ	−0.22***	1.06***	−0.60***	−0.14***	0.14***	−3.78***	0.50***	0.08***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)
$n_f\theta$	0.11***	0.12***	0.42***	0.06***	−0.07***	0.52***	−0.31***	−0.03***
	(0.004)	(0.005)	(0.003)	(0.005)	(0.004)	(0.01)	(0.002)	(0.004)
Structured	−0.39***	−0.49***	−0.54***	−0.37***	0.37***	0.36***	0.42***	0.35***
	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.002)	(0.003)
Detection Prob. σ	−3.09***	−8.71***	−11.16***	−9.09***	2.83***	27.77***	9.44***	7.12***
	(0.12)	(0.78)	(0.10)	(0.66)	(0.11)	(0.68)	(0.09)	(0.57)
σ^2		−60.58***		22.02***		47.40***		−14.55***
		(3.67)		(3.04)		(3.06)		(2.59)
σ^3	9.32***	121.65***	21.48***	−20.94***	−7.71***	−122.98***	−21.25***	11.95***
	(0.46)	(5.49)	(0.39)	(4.44)	(0.39)	(4.58)	(0.31)	(3.79)
$n_f\sigma$	−0.83***	2.69***	0.56***	−0.36***	0.72***	−5.49***	−0.30***	0.29***
	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.04)	(0.01)	(0.03)
Times Caught	−0.01***	0.05***	0.07***	−0.04***	0.01***	−0.31***	−0.06***	0.03***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)
Repeat Offender	−0.04***	1.00***	0.62***	−0.26***	0.03***	−0.18***	−0.27***	0.20***
	(0.01)	(0.01)	(0.005)	(0.01)	(0.005)	(0.01)	(0.004)	(0.01)
IMR				−0.45***				0.36***
				(0.01)				(0.01)
Constant	4.52***	7.19***	6.05***	6.18***				
	(0.16)	(0.09)	(0.05)	(0.06)				
Observations	427'108	382'750	809'858	427'108	427'108	382'750	809'858	427'108
R^2	0.15	0.45	0.45	0.15				
Adjusted R^2	0.15	0.45	0.45	0.15				
Log Likelihood					−2'507'443.00	−1'183'409.00	−3'841'901.00	−2'506'846.00

Note: This table shows the estimation results of Lasso CV linear cross-sectional regressions to explain cartel duration (ln(duration+1)) and the estimation results of a Lasso CV Weibull Hazard Model to explain cartel death, both at the industry level, for data simulated for Models IIIa and IIIb combined. 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.
Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively.

Table B16 Cartel Duration with CV Lasso on Model I, II and III

	Ln(Duration+1)				Cartel Death			
	LasSample	LasUndetec	LasCartels	LasHeck	HRLasSample	HRLasUnd	HRLasCartels	HRLasHeck
N Firms n_f	0.26***	0.45***		0.02**	−0.28***	1.15***		−0.01
	(0.05)	(0.05)		(0.01)	(0.04)	(0.05)		(0.01)
n_f^2	−0.03**	−0.30***	−0.11***		0.04***	0.02**	0.07***	
	(0.01)	(0.01)	(0.001)		(0.01)	(0.01)	(0.001)	
n_f^3	−0.01***	0.02***	−0.001***	−0.003***	0.003**	−0.002***	0.003***	0.002***
	(0.001)	(0.001)	(0.0002)	(0.0002)	(0.001)	(0.001)	(0.0001)	(0.0001)
γ^2				1.66***				−0.64*
				(0.38)				(0.33)
γ^3	0.03***	0.02	0.03***	−1.34***	−0.05***	−0.03**	−0.03***	0.48*
	(0.01)	(0.01)	(0.01)	(0.32)	(0.01)	(0.01)	(0.01)	(0.27)
Leniency (% Fine) θ	0.03***	0.66***	−0.55***	−0.05***	−0.05***	−2.69***	0.44***	0.03**
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)
$n_f\theta$	0.02***	0.21***	0.40***	0.02***	−0.001	0.29***	−0.28***	−0.002
	(0.004)	(0.005)	(0.003)	(0.004)	(0.004)	(0.01)	(0.002)	(0.003)
Detection Prob. σ	−5.43***	−18.35***	−13.46***	−11.34***	4.77***	23.45***	11.52***	9.07***
	(0.11)	(0.14)	(0.08)	(0.62)	(0.09)	(0.15)	(0.07)	(0.53)
σ^2				26.38***				−18.72***
				(2.88)				(2.46)
σ^3	11.91***	24.02***	23.64***	−25.83***	−10.07***	−29.63***	−22.95***	16.65***
	(0.42)	(0.50)	(0.35)	(4.21)	(0.36)	(0.43)	(0.27)	(3.60)
$n_f\sigma$	−0.09***	2.19***	1.26***	0.09***	0.10***	−2.97***	−0.93***	−0.07***
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Structured	−0.37***	−0.51***	−0.51***	−0.36***	0.35***	0.35***	0.40***	0.34***
	(0.004)	(0.005)	(0.003)	(0.004)	(0.003)	(0.004)	(0.002)	(0.003)
Model II	0.54***	0.30***	0.36***	0.65***	−0.48***	−0.39***	−0.40***	−0.57***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Model III	−0.30***	−0.90***	−0.90***	−0.44***	0.24***	0.62***	0.73***	0.21***
	(0.01)	(0.01)	(0.01)	(0.08)	(0.01)	(0.01)	(0.01)	(0.07)
Start	−0.001***	−0.001***	−0.002***	−0.0003***	0.001***	0.001***	0.001***	0.001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Times Caught	−0.001	0.05***	0.07***	−0.04***	0.005***	−0.35***	−0.07***	0.04***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)
Repeat Offender	0.02***	1.06***	0.69***	−0.30***	−0.02***	−0.21***	−0.32***	0.24***
	(0.01)	(0.01)	(0.004)	(0.01)	(0.004)	(0.01)	(0.003)	(0.01)
IMR				−0.60***				0.48***
				(0.01)				(0.01)
Constant	5.95***	7.57***	7.83***	6.75***				
	(0.06)	(0.07)	(0.02)	(0.05)				
Observations	475'456	465'402	940'858	475'456	475'456	465'402	940'858	475'456
R^2	0.15	0.45	0.46	0.15				
Adjusted R^2	0.15	0.45	0.46	0.15				
Log Likelihood					−2'804'453.00	−1'432'521.00	−4'409'298.00	−2'803'121.00

Note: This table shows the estimation results of Lasso CV linear cross-sectional regressions to explain cartel duration (ln(duration+1)) and the estimation results of a Lasso CV Weibull Hazard Model to explain cartel death, both at the industry level, for data simulated for Models I, II, IIIa, and IIIb combined. 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.
Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively.

Table C17 Sample Selection Bias Linear Regression - Models I, II and III

Coefficients	β^s	β^l	α_{IMR}	$bias_{IMR}$
Number of Firms n_f	-0.16	0	0.23	-0.16
Fines γ (% of Profit)	0.06	0.07	0.01	-0.01
Leniency (% of Fine) θ	0.09	0	-0.13	0.09
Detection Probability σ	-3.66	-3.24	0.61	-0.43
Start	-0.40	-0.36	0	0
Structured	0.53	0.67	0.05	-0.04
Model II	-0.24	-0.07	0.19	-0.14
Model III	0	0	0.24	-0.17
Times Caught	0	-0.05	-0.07	0.05
Repeat Offender	0.04	-0.37	-0.58	0.41
IMR		-0.71		
Constant	6.38	6.06	-0.45	0.32

This table shows, for the Linear Regression of the combined Models I, II, and III, the sample selection bias that we correct with the inverse Mill's ratio (IMR) following (Heckman (1979)). β^s is the estimated coefficients in the short model without IMR. β^l is the estimated coefficients in the corrected long model including IMR. α_{IMR} is the coefficient in the auxiliary regression between each variable and IMR. The last column shows the sample selection bias: $bias_{IMR} = \beta^l(IMR) * \alpha_{IMR}$.

Table C18 Out of Sample R^2 for Undetected Cartels

Model	mlrSamp	mlrHeck	lasSamp	lasHeck
I	-0.65	-0.05	-0.39	-0.07
II	-0.42	-0.05	-0.19	-0.05
III	-0.59	-0.19	-0.36	-0.16
I, II, III	-0.64	-0.2	-0.45	-0.15

Table C19 Out of Sample R^2, for Detected and Undetected Cartels

Model	mlrSamp	mlrHeck	lasSamp	lasHeck
I	0.04	0.33	0.17	0.32
II	-0.03	0.23	0.13	0.23
III	0.01	0.2	0.13	0.22
I, II, III	-0.02	0.2	0.08	0.23

Table C20 Out of Sample MSE for Undetected Cartels

Model	mlrSamp	mlrHeck	lasSamp	lasHeck
I	5.39	3.42	4.54	3.41
II	6.18	4.57	5.19	4.52
III	5.12	3.85	4.37	3.73
I, II, III	5.48	3.99	4.83	3.91

References

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

Table C21 Out of Sample MSE for Detected and Undetected Cartels

Model	mlrSamp	mlrHeck	lasSamp	lasHeck
I	3.5	2.45	3.03	2.47
II	5.1	3.84	4.32	3.85
III	3.15	2.55	2.79	2.49
I, II, III	3.41	2.67	3.08	2.59

Table C22 Mean Duration and Probability of Death Estimation

	Mean Arr. T. τ^{-1}	Births τ	Mean Dur. λ^{-1}	Deaths λ	Alive τ/λ	Obs. N
Model I						
Detected Cases (Sample)	0.027	36.62	154.94	0.006	5'673	36'615
Undetected Cases (Population - Sample)	0.021	48.12	50.81	0.020	2'445	48'119
All Cases (Population)	0.012	84.73	95.81	0.010	8'118	84'734
Model II						
Detected Cases (Sample)	0.085	11.73	289.49	0.003	3'397	11'733
Undetected Cases (Population - Sample)	0.021	48.36	98.15	0.010	4'746	48'357
All Cases (Population)	0.017	60.09	135.51	0.007	8'143	60'090
Model IIIa						
Detected Cases (Sample)	0.005	218.12	160.20	0.006	3'4943	218'123
Undetected Cases (Population - Sample)	0.004	281.70	59.82	0.017	16'852	281'703
All Cases (Population)	0.002	499.83	103.63	0.010	51'795	499'826
Model IIIb						
Detected Cases (Sample)	0.005	208.98	130.12	0.008	27'194	208'985
Undetected Cases (Population - Sample)	0.004	264.29	34.36	0.029	9'081	264'286
All Cases (Population)	0.002	473.27	76.65	0.013	36'275	473'271

$\tau^{-1} = timespan/N$
$\lambda^{-1} = \sum duration/N$