# 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.


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

This online appendix contains summary statistics and 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   Summary Statistics

**Table A1** Summary Statistics Population (Detected and Undetected Cartels) of Model I

|  | Mean | Median | SD | Min | Max | Skew | N |
|---|---|---|---|---|---|---|---|
| Number of Firms $n_f$ | 4.74 | 5 | 1.52 | 2 | 8 | -0.59 | 77'822 |
| Detection Probability $\sigma$ | 0.24 | 0.25 | 0.08 | 0.10 | 0.35 | -0.22 | 77'822 |
| Start | 442.35 | 427 | 308.34 | 1 | 1000 | 0.12 | 77'822 |
| End | 545.66 | 539 | 306.41 | 1 | 1000 | 0 | 77'822 |
| Duration | 104.31 | 35 | 159.45 | 1 | 1000 | 2.52 | 77'822 |
| Ln(Duration+1) | 3.30 | 3.58 | 1.91 | 0.69 | 6.91 | -0.08 | 77'822 |
| Detected | 0.47 | 0 | 0.50 | 0 | 1 | 0.12 | 77'822 |
| Times Caught | 2.12 | 1 | 2.54 | 0 | 18 | 1.30 | 77'822 |
| Repeat Offender | 0.47 | 0 | 0.50 | 0 | 1 | 0.14 | 77'822 |

For Model I, this table presents the following summary statistics of the 77'822 simulated detected and undetected cartels within a total period of 1'000 time units: the exogenous industry and enforcement characteristics, the start and end date of the cartel, the duration, if it got detected, detected repeatedly, and how often it got detected.

**Table A2** Summary Statistics Sample (Detected Cartels) of Model I

|  | Mean | Median | SD | Min | Max | Skew | N |
|---|---|---|---|---|---|---|---|
| Number of Firms $n_f$ | 3.79 | 4 | 1.33 | 2 | 8 | 0.15 | 36'615 |
| Detection Probability $\sigma$ | 0.26 | 0.25 | 0.08 | 0.10 | 0.35 | -0.46 | 36'615 |
| Start | 332.28 | 289 | 289.08 | 1 | 999 | 0.43 | 36'615 |
| End | 486.22 | 479 | 291.45 | 1 | 1000 | 0.06 | 36'615 |
| Duration | 154.94 | 103 | 158.53 | 1 | 1000 | 1.80 | 36'615 |
| Ln(Duration+1) | 4.47 | 4.64 | 1.24 | 0.69 | 6.91 | -0.70 | 36'615 |
| Detected | 1 | 1 | 0 | 1 | 1 |  | 36'615 |
| Times Caught | 3.45 | 3 | 2.31 | 1 | 18 | 1.13 | 36'615 |
| Repeat Offender | 0.77 | 1 | 0.42 | 0 | 1 | -1.29 | 36'615 |

For Model I, this table presents the following summary statistics of the 36'615 simulated detected cartels within a total period of 1'000 time units: the exogenous industry and enforcement characteristics, the start and end date of the cartel, the duration, if it got detected repeatedly, and how often it got detected.

**Table A3** Summary Statistics Population (Detected and Undetected Cartels) of Model II

|  | Mean | Median | SD | Min | Max | Skew | N |
|---|---|---|---|---|---|---|---|
| Number of Firms $n_f$ | 5.19 | 6 | 1.38 | 2 | 8 | -1.16 | 53'178 |
| Detection Probability $\sigma$ | 0.23 | 0.25 | 0.09 | 0.10 | 0.35 | -0.10 | 53'178 |
| Start | 422.15 | 401 | 314.10 | 1 | 1000 | 0.18 | 53'178 |
| End | 574.27 | 571 | 309.25 | 1 | 1000 | -0.05 | 53'178 |
| Duration | 153.12 | 11 | 264.18 | 1 | 1000 | 2 | 53'178 |
| Ln(Duration+1) | 3.06 | 2.48 | 2.23 | 0.69 | 6.91 | 0.35 | 53'178 |
| Detected | 0.22 | 0 | 0.41 | 0 | 1 | 1.35 | 53'178 |
| Times Caught | 0.62 | 0 | 1.05 | 0 | 8 | 2.01 | 53'178 |
| Repeat Offender | 0.17 | 0 | 0.37 | 0 | 1 | 1.77 | 53'178 |

For Model II, this table presents the following summary statistics of the 53'178 simulated detected and undetected cartels within a total period of 1'000 time units: the exogenous industry and enforcement characteristics, the start and end date of the cartel, the duration, if it got detected, detected repeatedly, and how often it got detected.

**Table A4** Summary Statistics Sample (Detected Cartels) of Model II

|  | Mean | Median | SD | Min | Max | Skew | N |
|---|---|---|---|---|---|---|---|
| Number of Firms $n_f$ | 3.82 | 4 | 1.33 | 2 | 7 | 0.12 | 11'733 |
| Detection Probability $\sigma$ | 0.26 | 0.25 | 0.08 | 0.10 | 0.35 | -0.43 | 11'733 |
| Start | 199.50 | 24 | 263.67 | 1 | 999 | 1.14 | 11'733 |
| End | 487.99 | 480 | 291.23 | 1 | 1000 | 0.06 | 11'733 |
| Duration | 289.49 | 219 | 246.85 | 1 | 1000 | 0.93 | 11'733 |
| Ln(Duration+1) | 5.14 | 5.39 | 1.24 | 0.69 | 6.91 | -1.04 | 11'733 |
| Detected | 1 | 1 | 0 | 1 | 1 |  | 11'733 |
| Times Caught | 1.77 | 1 | 1.02 | 1 | 8 | 1.55 | 11'733 |
| Repeat Offender | 0.48 | 0 | 0.50 | 0 | 1 | 0.10 | 11'733 |

For Model II, this table presents the following summary statistics of the 11'733 simulated detected cartels within a total period of 1'000 time units: the exogenous industry and enforcement characteristics, the start and end date of the cartel, the duration, if it got detected repeatedly, and how often it got detected.

**Table A5** Summary Statistics Population (Detected and Undetected Cartels) of Model III

|  | Mean | Median | SD | Min | Max | Skew | N |
|---|---|---|---|---|---|---|---|
| Number of Firms $n_f$ | 3.62 | 4 | 1.28 | 2 | 7 | 0.16 | 809'858 |
| Fines $\gamma$ (% of Profit) | 0.80 | 0.80 | 0.08 | 0.70 | 0.90 | 0.02 | 809'858 |
| Leniency (% of Fine) $\theta$ | 0.40 | 0.50 | 0.41 | 0 | 1 | 0.37 | 809'858 |
| Detection Probability $\sigma$ | 0.22 | 0.20 | 0.08 | 0.10 | 0.35 | 0.08 | 809'858 |
| Structured | 0.48 | 0 | 0.50 | 0 | 1 | 0.07 | 809'858 |
| Start | 416.30 | 397 | 317.51 | 1 | 1000 | 0.19 | 809'858 |
| End | 524.04 | 516 | 315.19 | 1 | 1000 | 0.02 | 809'858 |
| Duration | 108.75 | 43 | 162.41 | 1 | 1000 | 2.60 | 809'858 |
| Ln(Duration+1) | 3.50 | 3.78 | 1.79 | 0.69 | 6.91 | -0.20 | 809'858 |
| Detected | 0.53 | 1 | 0.50 | 0 | 1 | -0.11 | 809'858 |
| Times Caught | 2.38 | 1 | 2.82 | 0 | 25 | 1.64 | 809'858 |
| Repeat Offender | 0.49 | 0 | 0.50 | 0 | 1 | 0.02 | 809'858 |

For Model III, this table presents the following summary statistics of the 809'858 simulated detected and undetected cartels within a total period of 1'000 time units: the exogenous industry and enforcement characteristics, the start and end date of the cartel, the duration, if it got detected, detected repeatedly, and how often it got detected.

**Table A6** Summary Statistics Sample (Detected Cartels) of Model III

|  | Mean | Median | SD | Min | Max | Skew | N |
|---|---|---|---|---|---|---|---|
| Number of Firms $n_f$ | 2.96 | 3 | 0.98 | 2 | 7 | 0.70 | 427'108 |
| Fines $\gamma$ (% of Profit) | 0.80 | 0.80 | 0.08 | 0.70 | 0.90 | 0.03 | 427'108 |
| Leniency (% of Fine) $\theta$ | 0.52 | 0.50 | 0.41 | 0 | 1 | -0.06 | 427'108 |
| Detection Probability $\sigma$ | 0.23 | 0.25 | 0.08 | 0.10 | 0.35 | -0.07 | 427'108 |
| Structured | 0.49 | 0 | 0.50 | 0 | 1 | 0.04 | 427'108 |
| Start | 331.62 | 286 | 295.59 | 1 | 1000 | 0.43 | 427'108 |
| End | 476.10 | 464 | 291.09 | 1 | 1000 | 0.10 | 427'108 |
| Duration | 145.48 | 91 | 158.39 | 1 | 1000 | 1.97 | 427'108 |
| Ln(Duration+1) | 4.36 | 4.52 | 1.27 | 0.69 | 6.91 | -0.60 | 427'108 |
| Detected | 1 | 1 | 0 | 1 | 1 |  | 427'108 |
| Times Caught | 3.52 | 3 | 2.71 | 1 | 25 | 1.59 | 427'108 |
| Repeat Offender | 0.74 | 1 | 0.44 | 0 | 1 | -1.10 | 427'108 |

For Model III, this table presents the following summary statistics of the 427'108 simulated detected cartels within a total period of 1'000 time units: the exogenous industry and enforcement characteristics, the start and end date of the cartel, the duration, if it got detected repeatedly, and how often it got detected.

# Appendix B   Linear Models Results

**Table B7**  Sample Selection Bias Linear Regression - Model I

| Coefficients | $\beta^s$ | $\beta^l$ | $\alpha_{IMR}$ | $bias_{IMR}$ |
|---|---|---|---|---|
| Number of Firms $n_f$ | -0.12 | -0.02 | 0.15 | -0.10 |
| Start | 0 | 0 | 0 | 0 |
| Detection Probability $\sigma$ | -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)). $\beta^s$ is the estimated coefficients in the short model without IMR. $\beta^l$ is the estimated coefficients in the corrected long model including IMR. $\alpha_{IMR}$ 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 B8**  Sample Selection Bias Linear Regression - Model II

| Coefficients | $\beta^s$ | $\beta^l$ | $\alpha_{IMR}$ | $bias_{IMR}$ |
|---|---|---|---|---|
| Number of Firms $n_f$ | -0.16 | -0.09 | 0.14 | -0.07 |
| Start | 0 | 0 | 0 | 0 |
| Detection Probability $\sigma$ | -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)). $\beta^s$ is the estimated coefficients in the short model without IMR. $\beta^l$ is the estimated coefficients in the corrected long model including IMR. $\alpha_{IMR}$ 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 B9** Sample Selection Bias Linear Regression - Model III

| Coefficients | $\beta^s$ | $\beta^l$ | $\alpha_{IMR}$ | $bias_{IMR}$ |
|---|---|---|---|---|
| Number of Firms $n_f$ | -0.18 | -0.01 | 0.25 | -0.17 |
| Fines $\gamma$ (% of Profit) | 0.06 | 0.07 | 0.01 | -0.01 |
| Leniency (% of Fine) $\theta$ | 0.10 | 0.01 | -0.13 | 0.09 |
| Start | 0 | 0 | 0 | 0 |
| Structured | -0.40 | -0.36 | 0.06 | -0.04 |
| Detection Probability $\sigma$ | -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)). $\beta^s$ is the estimated coefficients in the short model without IMR. $\beta^l$ is the estimated coefficients in the corrected long model including IMR. $\alpha_{IMR}$ 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 B10** Sample Selection Bias Linear Regression - Models I, II and III

| Coefficients | $\beta^s$ | $\beta^l$ | $\alpha_{IMR}$ | $bias_{IMR}$ |
|---|---|---|---|---|
| Number of Firms $n_f$ | -0.16 | 0 | 0.23 | -0.16 |
| Fines $\gamma$ (% of Profit) | 0.06 | 0.07 | 0.01 | -0.01 |
| Leniency (% of Fine) $\theta$ | 0.09 | 0 | -0.13 | 0.09 |
| Detection Probability $\sigma$ | -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)). $\beta^s$ is the estimated coefficients in the short model without IMR. $\beta^l$ is the estimated coefficients in the corrected long model including IMR. $\alpha_{IMR}$ 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** 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. $\sigma$ | −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 | 7822 | 36'615 | 36'615 | 41'207 | 7822 | 36'615 |
| R² | 0.10 | 0.47 | 0.50 | 0.11 | | | | |
| Adjusted R² | 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 B12** 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. $\sigma$ | −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 |
| R² | 0.17 | 0.51 | 0.56 | 0.17 | | | | |
| Adjusted R² | 0.16 | 0.51 | 0.56 | 0.17 | | | | |
| Log Likelihood | −77'014.56 | −120'764.40 | −209'666.50 | | −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.

8

**Table B13** 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 $\gamma$ (% 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) $\theta$ | 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. $\sigma$ | −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. Significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively. Standard errors are in parentheses.

9

**Table B14** Linear Regression and HR for Cartel Duration on Models I, II and III - ICC on Harrington et al.

| | Ln(Duration+1) | | | | Cartel Death | | | |
|---|---|---|---|---|---|---|---|---|
| | mlrSample | mlrUndetect | mlrCartels | mlrHeck | HRSample | HRUndetect | HRCartels | HRHeck |
| N Firms $n_f$ | −0.16*** | −0.60*** | −0.50*** | 0.08*** | 0.12*** | 0.60*** | 0.37*** | −0.07*** |
| | (0.002) | (0.002) | (0.001) | (0.01) | (0.002) | (0.002) | (0.001) | (0.01) |
| Fines $\gamma$ (% of Profit) | 0.06*** | 0.03 | 0.03 | 0.12*** | −0.10*** | −0.05** | −0.05*** | −0.24*** |
| | (0.02) | (0.03) | (0.02) | (0.03) | (0.02) | (0.02) | (0.01) | (0.04) |
| Leniency $\theta$ (% Fine) | 0.09*** | 1.58*** | 0.84*** | 0.15*** | −0.05*** | −1.26*** | −0.57*** | −0.10*** |
| | (0.004) | (0.01) | (0.004) | (0.01) | (0.004) | (0.01) | (0.003) | (0.01) |
| Start | −0.001*** | −0.001*** | −0.002*** | −0.0003*** | 0.001*** | 0.001*** | 0.001*** | −0.0001*** |
| | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| Structured | −0.40*** | −0.52*** | −0.56*** | −0.33*** | 0.36*** | 0.35*** | 0.42*** | 0.40*** |
| | (0.004) | (0.005) | (0.003) | (0.005) | (0.003) | (0.004) | (0.002) | (0.01) |
| Detection Prob. $\sigma$ | −3.66*** | −4.31*** | −4.67*** | −3.66*** | 3.34*** | 4.32*** | 4.02*** | 5.05*** |
| | (0.02) | (0.03) | (0.02) | (0.03) | (0.03) | (0.02) | (0.01) | (0.04) |
| Model II | 0.53*** | 0.28*** | 0.23*** | 0.99*** | −0.48*** | −0.34*** | −0.37*** | −1.37*** |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.01) | (0.01) | (0.01) | (0.02) |
| Model III | −0.24*** | −0.82*** | −0.63*** | −0.09*** | 0.22*** | 0.87*** | 0.51*** | 0.20*** |
| | (0.02) | (0.02) | (0.02) | (0.03) | (0.02) | (0.02) | (0.01) | (0.03) |
| Times Caught | −0.002** | −0.01*** | 0.05*** | −0.09*** | 0.01*** | −0.34*** | −0.05*** | 0.19*** |
| | (0.001) | (0.002) | (0.001) | (0.001) | (0.001) | (0.003) | (0.001) | (0.001) |
| Repeat Offender | 0.04*** | 1.26*** | 0.82*** | −0.15*** | −0.03*** | −0.26*** | −0.38*** | 0.52*** |
| | (0.01) | (0.01) | (0.005) | (0.01) | (0.004) | (0.01) | (0.003) | (0.01) |
| IMR | | | | −1.58*** | | | | 1.73*** |
| | | | | (0.02) | | | | (0.02) |
| Constant | 6.38*** | 6.98*** | 7.08*** | 6.09*** | | | | |
| | (0.01) | (0.02) | (0.01) | (0.02) | | | | |
| Observations | 475'456 | 465'402 | 940'858 | 125'201 | 475'456 | 465'402 | 940'858 | 125'201 |
| R² | 0.14 | 0.43 | 0.42 | 0.35 | | | | |
| Adjusted R² | 0.14 | 0.43 | 0.42 | 0.35 | | | | |
| Log Likelihood | | | | | −2'806'284.00 | −1'445'284.00 | −4'442'193.00 | −742'964.20 |

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 C    Lasso Results

**Table C15** Lasso CV Regression and HR for Cartel Duration on Model I - ICC on Stigler - Detection independent of Collusion

| | LasSample | Ln(Duration+1) LasUndetec | LasCartels | HRLasSample | Cartel Death HRLasUnd | HRLasCartels |
|---|---|---|---|---|---|---|
| Start | −0.001*** | −0.001*** | −0.002*** | 0.001*** | 0.001*** | 0.001*** |
| | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| N Firms $n_f$ | −1.37*** | 4.87*** | 2.50*** | 1.04*** | 12.61*** | −1.49*** |
| | (0.14) | (0.17) | (0.09) | (0.12) | (1.42) | (0.07) |
| $n_f^2$ | 0.43*** | −1.29*** | −0.64*** | −0.33*** | −1.56*** | |
| | (0.04) | (0.04) | (0.02) | (0.03) | (0.25) | |
| $n_f^3$ | −0.04*** | 0.09*** | 0.04*** | 0.03*** | 0.07*** | −0.02*** |
| | (0.003) | (0.003) | (0.002) | (0.002) | (0.01) | (0.001) |
| Detection Prob. $\sigma$ | −5.04** | −17.14*** | −12.64*** | 4.97** | 18.48*** | 12.14*** |
| | (2.38) | (2.25) | (1.71) | (2.05) | (2.12) | (1.38) |
| $\sigma^2$ | 4.74 | 26.19** | 13.83* | −6.61 | −15.51* | |
| | (10.82) | (10.53) | (7.93) | (9.34) | (9.02) | |
| $\sigma^3$ | −2.64 | −34.76** | −17.19 | 5.93 | 22.24* | 18.06* |
| | (15.46) | (15.51) | (11.53) | (13.35) | (13.29) | (9.30) |
| $n_f\sigma$ | 0.12** | 1.69*** | 1.39*** | −0.06 | −2.31*** | −1.12*** |
| | (0.06) | (0.07) | (0.04) | (0.05) | (0.16) | (0.03) |
| Times Caught | 0.02*** | 0.03*** | 0.12*** | −0.02*** | −0.49*** | −0.12*** |
| | (0.004) | (0.01) | (0.003) | (0.004) | (0.02) | (0.003) |
| Repeat Offender | 0.17*** | 1.75*** | 0.99*** | −0.19*** | −0.23*** | −0.51*** |
| | (0.02) | (0.03) | (0.02) | (0.02) | (0.05) | (0.01) |
| Constant | 7.03*** | 1.06*** | 3.17*** | | | |
| | (0.24) | (0.28) | (0.17) | | | |
| | | | | | | |
| Observations | 36'615 | 41'207 | 7822 | 36'615 | 41'207 | 7822 |
| $R^2$ | 0.12 | 0.49 | 0.53 | | | |
| Adjusted $R^2$ | 0.12 | 0.49 | 0.53 | | | |
| Log Likelihood | | | | −218'190.40 | −113'185.60 | −348'310.50 |

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

12

**Table C16** 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 |
| Start | $-0.002^{***}$ | $-0.002^{***}$ | $-0.002^{***}$ | $0.002^{***}$ | $0.001^{***}$ | $0.001^{***}$ |
| | (0.0001) | (0.0000) | (0.0000) | (0.0001) | (0.0000) | (0.0000) |
| N Firms $n_f$ | $-1.98^{***}$ | $8.53^{***}$ | $5.88^{***}$ | $1.37^{***}$ | $7.67^{***}$ | $-4.30^{***}$ |
| | (0.24) | (0.18) | (0.13) | (0.20) | (1.48) | (0.11) |
| $n_f^2$ | $0.61^{***}$ | $-2.16^{***}$ | $-1.48^{***}$ | $-0.43^{***}$ | $-0.55^{**}$ | |
| | (0.06) | (0.04) | (0.03) | (0.05) | (0.26) | (0.03) |
| $n_f^3$ | $-0.06^{***}$ | $0.15^{***}$ | $0.10^{***}$ | $0.04^{***}$ | $-0.0004$ | $-0.07^{***}$ |
| | (0.01) | (0.003) | (0.002) | (0.004) | (0.01) | (0.002) |
| Detection Prob. $\sigma$ | $-2.36^{***}$ | $-5.81^{***}$ | $-7.30^{***}$ | $4.17$ | $5.79^{***}$ | $8.33^{***}$ |
| | (0.61) | (0.51) | (0.40) | (3.61) | (2.11) | (1.68) |
| $\sigma^2$ | | | | $-11.18$ | $0.25$ | $-5.39$ |
| | | | | (16.44) | (9.01) | (7.83) |
| $\sigma^3$ | $3.12$ | $-2.29$ | $-0.19$ | $12.46$ | $1.97$ | $7.29$ |
| | (2.52) | (1.64) | (1.46) | (23.52) | (13.27) | (11.46) |
| $n_f \sigma$ | $0.13$ | $0.97^{***}$ | $1.19^{***}$ | $0.01$ | $-0.94^{***}$ | $-1.11^{***}$ |
| | (0.10) | (0.08) | (0.05) | (0.09) | (0.16) | (0.05) |
| Times Caught | $0.02$ | $0.62^{***}$ | $0.54^{***}$ | $-0.01$ | $-1.30^{***}$ | $-0.41^{***}$ |
| | (0.02) | (0.02) | (0.01) | (0.02) | (0.03) | (0.01) |
| Repeat Offender | $0.20^{***}$ | $-0.44^{***}$ | $-0.23^{***}$ | $-0.22^{***}$ | $1.07^{***}$ | $0.26^{***}$ |
| | (0.03) | (0.05) | (0.03) | (0.03) | (0.08) | (0.03) |
| Constant | $7.99^{***}$ | $-2.81^{***}$ | $-0.25$ | | | |
| | (0.30) | (0.26) | (0.19) | | | |
| Observations | 11'733 | 41'445 | 53'178 | 11'733 | 41'445 | 53'178 |
| $R^2$ | 0.20 | 0.54 | 0.59 | | | |
| Adjusted $R^2$ | 0.20 | 0.54 | 0.59 | | | |
| Log Likelihood | | | | $-76'836.46$ | $-118'987.30$ | $-207'949.80$ |

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

13

**Table C17** Cartel Duration with CV Lasso on Model III - ICC on Harrington et al.

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

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 - 4 estimate linear regression coefficients, while columns 5 - 7 estimate HR coefficients, 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 C18** Cartel Duration with CV Lasso on Models I, II and III

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

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

# 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