Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Number and Coverage as Structural Margins for Identifying Cartel Markets in EU Manufacturing

Shidong Li

20220054@nuc.edu.cn

North University of China

Article

Keywords: cartels, collusion, market structure, screening, necessary-condition analysis, competition enforcement

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11333771/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Number and Coverage as Structural Margins for Identifying Cartel Markets in EU Manufacturing

Abstract

Competition authorities must choose ex ante where to look for cartels. Market concentration, the usual guide, combines two features acting through different mechanisms: the number of strategically significant firms needed to coordinate, and the share of the market those firms cover. This paper separates the two and asks whether they support a transparent enforcement screen and where that screen stops being informative. The outcome is an established infringement decision rather than the opening of proceedings: of 62 European Commission manufacturing cases, 45 ended in a decision and 23 are analyzed, organized as 630 country-by-industry units across 27 countries and 24 industries. Most cartels span the whole European Economic Area and appear in every member state, so a country holdout leaks the same cartel into training, and the valid design holds out entire industries. Under it the score attains 0.68 ROC and 0.53 precision-recall against a base rate of 0.35; reviewing the top fifth of ranked markets captures a third of infringements at 1.66 times the random rate. It ranks industries but not markets within them, high coverage is not decisive, and no threshold offers a safe harbor. The result is a one-way screen that sharpens targeting within clear limits.

Keywords: cartels, collusion, market structure, screening, necessary-condition analysis, competition enforcement

JEL: L13, L41, K21, D43

1. Introduction

Which markets are prone to collusion is among the oldest questions in industrial organization. The textbook answer is *concentration*: with fewer firms, monitoring is easier, deviation is more visible, and the gains from coordination are larger. This answer has disciplined an enormous body of empirical work and underpins the structural presumptions used in merger control (Davies and Olczak, 2008). It is also something of a puzzle that the empirical returns from this single statistic have been uneven: concentration levels frequently fail to separate collusive from competitive markets, and estimated concentration effects vary in size and sign across studies. Yet concentration is a single summary statistic for what are, on closer inspection, two different structural facts about a market. One is the **number of strategically significant decision-makers** who must reach and maintain an understanding. The other is the **share of the market that those firms collectively cover**, which determines whether a coordinating core can discipline prices against the reaction of

firms outside the agreement. A market can be concentrated because it contains a handful of firms that cover almost the entire market, or because it contains a moderately sized group of firms that cover only part of it. These configurations have different implications for collusion, but concentration alone does not separate them.

This paper develops the two-margin framework and then asks a narrower and, we think, more useful question than "does structure cause collusion": *can the two margins support a predictive screen that a competition authority could deploy, and what are the precise limits of such a screen?* An authority allocating scarce investigative resources does not need a causal estimate of the effect of market structure; it needs an ex ante ranking, built from variables observable before an investigation, that remains informative on markets not used to construct it. Whether that ranking exists is an empirical question with a definite answer, and a large part of this paper is devoted to evaluating it without the optimistic biases that inflate apparent screening performance.

Let a market's firms be ordered by strategic importance and let the *core* be the set of firms significant enough that a workable agreement must include them. Two quantities describe the core: its size n (the number of independent strategic decision-makers) and its coverage κ (the fraction of the market the core accounts for). A market is *coordinateable* when the core is simultaneously few enough that reaching and policing an agreement is feasible and large enough that supply from outside the core cannot undermine a coordinated price. This formulation preserves the classical emphasis on the difficulty of coordination as numbers rise but adds the coverage requirement that the one-dimensional concentration account suppresses, and it predicts that neither feature alone is sufficient.

The framework has to be developed alongside a careful treatment of what is observed. Whether a market's firms collude is latent. The researcher observes enforcement: a market contributes a "case" only if firms collude *and* an authority detects and acts on it. We write this as a three-layer data-generating process comprising structure, behavior, and observation. The observed outcome is the product of the behavioral and observational layers. The distinction matters because an observed zero is uninformative about whether collusion occurred. It is precisely the conflation of "no detected case" with "no collusion" that licenses an inference from a count threshold to a policy safe harbor, and we show that such an inference is not supported. Throughout we separate the *mechanistic* claim (structure is associated with the feasibility of collusion) from the *prognostic* claim (structure helps rank markets for enforcement attention), and we present evidence for the latter without asserting causal identification of the former.

Three methodological choices follow, each made to align measurement with the mechanism and to keep the evaluation honest.

Established infringements rather than openings. Of 62 Commission manufacturing cases opened during our 2011–2024 window, only 45 (73%) resulted in an infringement decision. Seven were closed administratively, two

proceeded only to a statement of objections, one penalized only an obstruction, and seven have no public substantive conclusion. Using openings would misclassify more than a quarter of cases as detected cartels. We therefore define the outcome as an established infringement and account for the remaining cases separately.

The country-by-industry level rather than the EU-wide industry. The theoretical mechanism requires that the firms counted be actual or potential rivals in the same relevant market. An EU-wide count of large firms in a two-digit industry does not identify such a group: 500 large firms across the Union typically constitute several dozen distinct national coordination problems rather than one 500-firm problem. We therefore construct the analysis at the country-by-two-digit-industry level, giving 630 nominal units across 27 member states and 24 manufacturing industries, and adjust for import exposure. This is not the product market of a competition case, but it aligns the empirical unit with the mechanism far more closely than the EU-wide two-digit level, at which consistent structural data exist.

Event-respecting validation rather than a single count threshold or a country holdout. Two distinct evaluation hazards must be avoided. First, reading a threshold from the same events used to judge it overstates how sharp that threshold is; such a threshold should instead be assessed out of sample. Second, and more fundamentally, the country-by-industry data are replicated: an EEA-wide cartel appears as a positive unit in every member state. Holding out a single country while leaving the same cartel positive in the training countries lets the model encounter the very event it is asked to predict, and an industry fixed effect then identifies the event almost directly. The simple pooled design that removes a held-out cartel from both the training and the test data is to hold out an entire industry, a procedure we call leave-one-division-out, and we use this leakage-free protocol as our primary evaluation.

Our findings are four.

First, the structural score gives an enforcement authority a useful concentration of investigative effort. Leave-one-division-out ROC AUC is 0.68 (bootstrap 95% CI 0.54–0.79) under the 250-employee core and 0.69 under the 50-employee core, with precision-recall AUC of about 0.53 against a base rate of 0.35. The operationally decisive quantity is the enrichment at the top of the ranking: reviewing the top fifth of units by out-of-fold risk captures one third of established infringements (lift 1.66 over random targeting), and the top tenth already has a selected precision of 0.58 against 0.35 at random, with lift intervals that stay above one. The paired increment over the global prior is about +0.18 ROC with an interval that excludes zero. This is a prioritization instrument rather than an accurate classifier, and it is at the prioritization margin that its value is clearest.

Second, the number of core firms carries no discrimination increment of its own. Removing the count lowers ROC AUC only from 0.68 to 0.67 (Δ 0.007, 95% CI −0.01 to 0.03); even the upper confidence limit is only about 0.03, so any gain that cannot be ruled out is small. At the EU-industry level an inverted-U is statistically

significant with a turning point near 1,400 firms, but it predicts poorly out of sample (AUC 0.61–0.64) and the count adds no power in a likelihood-ratio test ($p = 0.27$). The classical emphasis on the mere number of independent decision-makers is therefore not borne out as a standalone signal once coverage, size, and import exposure are present.

Third, the screen operates only between industries. Every industry that contains a case is positive in essentially every country, so the country-industry structure cannot separate markets within an industry and within-industry discrimination is not estimable. Combined with the coverage counterexamples, in which the highest-coverage industries have zero events and some low-coverage industries have events, this fixes the screen's scope: it ranks industries for attention, nothing more.

Fourth, there is no safe harbor. The lowest decile of every structural condition still contains established infringements at a rate of 12.7–21.8%, and an observed zero reflects non-detection as well as non-collusion. A lower core count is associated with a lower probability of an established infringement, not with zero probability.

The paper makes three contributions.

The first is conceptual. Decomposing concentration into the number and coverage margins explains why the one-dimensional statistic has had uneven empirical success: it bundles two mechanisms that can move independently and need not point the same way, which a single index cannot disentangle.

The second is a substantive challenge to a structural presumption. The count of strategically significant firms, the feature merger screening gives most weight to, shows no detectable discrimination increment once coverage, size, and import exposure are controlled; what can be ruled in is bounded above by about 0.03 AUC. The discriminating information in market structure instead runs through coverage together with size and import exposure.

The third is methodological and practical. We provide an event-respecting validation protocol for replicated enforcement data, quantify the optimism of the designs usually used in their place (an inflation of about 0.05 to 0.06 ROC), and turn the score into a transparent one-way prioritization instrument whose enrichment and limits are both measured.

The bounded conclusion is the contribution. A transparent structural model, evaluated without event leakage, produces a meaningful ranking of where established infringements are subsequently found and so can help prioritize preliminary inquiry; it cannot identify where collusion is absent, cannot clear markets, and cannot stand as evidence of guilt. We regard the careful demarcation of this boundary, which rests on an event-respecting validation protocol, as important for enforcement as the ranking itself.

The remainder proceeds as follows. Section 2 develops the framework and the three-layer observation model and states the propositions. Section 3 describes the

data and measurement. Section 4 sets out the empirical strategy, including the validation-leakage argument in detail. Section 5 reports results. Section 6 presents the necessary-condition robustness analysis. Section 7 examines detection selection and limitations. Section 8 derives policy implications, and Section 9 concludes.

2. Theory: number, coverage, and the coordinateable region

2.1 Two structural margins

Begin with a relevant market *m*, defined in product and geographic space. Order the firms by strategic importance and define the **core** K_m as the set of firms sufficiently important that a stable agreement must include them:

$$K_m = \{\, i : s_i \geq s^*(m) \,\},$$

where s_i is firm i's share of the market and the cutoff s^* is defined *relative to the market*, not by an absolute size threshold. Two quantities describe the core:

$$n_m = |K_m|, \qquad \kappa_m = \sum_{i \in K_m} s_i.$$

n_m is the number of independent strategic decision-makers in the core, counted on a group basis so that firms under common control count once. κ_m is the **coverage** of the core, that is, the share of market activity accounted for by the firms whose participation an agreement requires. The denominator for κ_m is the relevant market itself, not a broader administrative aggregate.

The two margins correspond to distinct forces. The **coordination margin** concerns reaching and maintaining an agreement: as the number of independent decision-makers rises, agreement is harder to negotiate, deviations harder to attribute, and the shared understanding needed to interpret ambiguous conduct harder to sustain (Stigler, 1964; Friedman, 1971; Tirole, 1988). Case-based evidence confirms that coordination problems recur across market settings (Genesove and Mullin, 2001; Porter, 1983) and that bidding cartels exhibit the same internal organization as price-fixing ones (Asker, 2010). The **coverage margin** concerns whether an agreed price is enforceable against firms outside it: if the core accounts for too little of the market, fringe producers and entrants can expand output and defeat a coordinated price regardless of how easily the core coordinates. This second force is largely absent from a one-dimensional concentration account.

2.2 The coordinateable region

Define the **coordinateable space** as the joint region

$$CS_m = \{\, (n_m, \kappa_m) : n_m \leq \bar{n} \text{ and } \kappa_m \geq \bar{\kappa} \,\}.$$

A market inside the region has a core both few enough to coordinate and large enough to discipline the market. Figure 1 traces the construction.

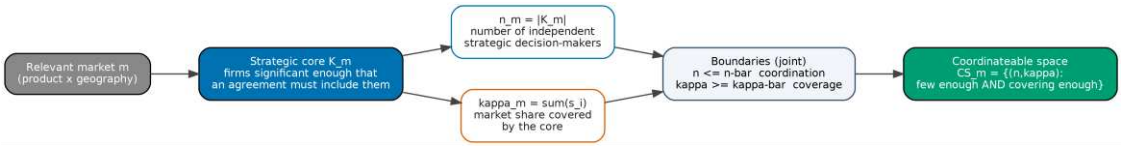


Figure 1. From the relevant market to the coordinateable space.

Figure 1. The strategic core of a relevant market is described by the number of independent decision-makers n_m and their collective coverage κ_m, which jointly define the coordinateable region.

The boundaries are joint rather than independent. As the number of core firms rises, the coverage required for stability may rise with it, because a larger group leaves more sources of fringe supply to police. The region is properly described by a frontier in the (n, κ) plane, not by two unrelated thresholds. Figure 2 depicts this frontier.

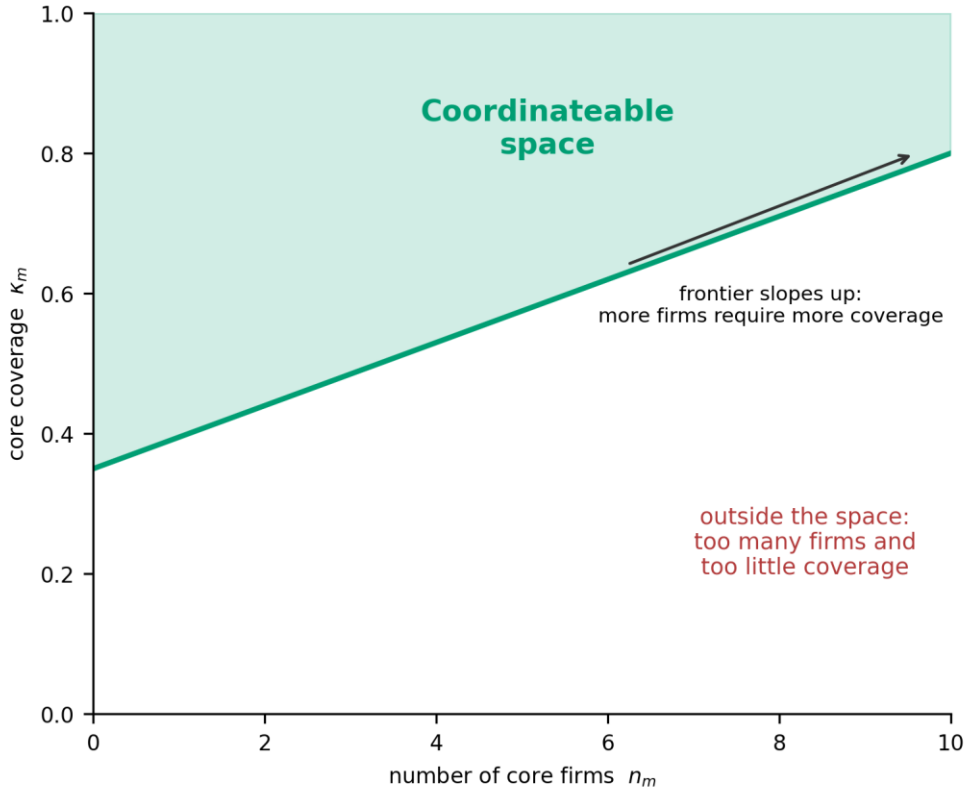


Figure 2. The coordinateable space in the (n, kappa) plane.

Figure 2. The coordinateable space lies above an upward-sloping joint frontier. Markets with too many core firms and too little coverage simultaneously lie outside the feasible region.

Consequently neither condition is sufficient. A core that covers nearly all the market but faces opaque demand, powerful buyers, or a history of failed coordination may not collude; a small group that could coordinate but covers too little cannot impose a coordinated price. The framework predicts necessary-but-not-sufficient structure and predicts specifically that the highest-coverage markets will include both detected cartels and zero-event markets. This is a falsifiable pattern rather than an assumption.

2.3 What the framework does not determine

Falling inside the coordinateable region is a statement about feasibility, not behavior. Within it, whether firms coordinate depends on demand growth and volatility, price transparency, cost and capacity symmetry, buyer power, product differentiation, interaction history, and the expected sanction (Levenstein and Suslow, 2006; Harrington and Chang, 2009). Empirical cartel studies document that stability varies widely with these conditions (Dick, 1996; Ellison, 1994) and that cartels form endogenously among heterogeneous firms (Bos and Harrington, 2010). We do not observe these consistently and therefore do not treat structure as determining an outcome. Structure is a gate on feasibility, not a cause we claim to identify.

2.4 The three-layer data-generating process

- **L1, structural.** The market may or may not fall inside the coordinateable space.
- **L2, behavioral.** Firms may or may not collude: $A_m \in \{0,1\}$, with $\Pr(A_m = 1 \mid CS_m, X_m)$.
- **L3, observational.** An authority may or may not establish an infringement: $D_m \in \{0,1\}$, with $\Pr(D_m = 1 \mid A_m = 1, Z_m)$, where Z_m captures jurisdictional thresholds, leniency, budgets, and priorities.

The researcher observes only

$$Y_m = A_m \times D_m.$$

An observed zero does not identify $A_m = 0$; a market with no detected infringement may contain an undetected cartel. Confounding the L3 zero with an L2 zero is exactly what would convert "no detected case below a threshold" into "no collusion below a threshold." We do not make that conversion. Separating L2 and L3 would require a partial-observability model (Poirier, 1980) or an instrument affecting detection but not collusion; we find no variable that cleanly satisfies the exclusion requirement and do not fit such a model on data too sparse to support it. We position the exercise honestly at L3: we model where established infringements are subsequently detected.

2.5 Propositions

- **P1 (necessary but not sufficient).** Cases are rare outside the coordinateable region, but falling inside it does not imply a case.

- **P2 (number margin, weak form).** Coordination difficulty rises with the number of core firms; an inverted-U is not asserted unless the descending branch is supported at the correct level of analysis.
- **P3 (coverage counterexamples are informative).** High coverage is not sufficient; high-κ zero-event industries are explained by the binding non-structural conditions, not suppressed.
- **P4 (distinct from concentration).** The core constructs are not merely concentration under another name, tested by whether they retain information after conventional controls.
- **P5 (screening utility).** A structural ranking improves leakage-free out-of-sample prediction of established infringements. This proposition is explicitly prognostic rather than causal.

---

3. Data and measurement

3.1 Sources and window

Structural variables come from Eurostat Structural Business Statistics (SBS), retrieved at member-state level through the legacy framework (`sbs_sc_sca_r2`) and the post-2020 framework (`sbs_sc_ovw`). Trade exposure comes from trade by enterprise characteristics (`ext_tec01`). Enforcement outcomes come from Commission cartel cases mapped to the member states and industries each case concerns. The raw counts of large enterprises are consistently available only from **2011**, which sets the window as **2011–2024**. The 2020–2021 SBS regulatory change altered some definitions; we absorb the break with year effects and do not splice the frameworks as identical. Greece is coded **EL**, not GR, in Eurostat; using GR silently returns an empty response.

3.2 The country-by-industry units

The most disaggregated level at which consistent structural data exist across the Union is the member state by two-digit manufacturing industry (NACE C10–C33). We construct one unit per country-industry pair, requiring at least eight years of structural observation, yielding **630 nominal units across 27 countries and 24 industries**. We adjust for imports because a national industry is not a closed market. The unit remains coarser than the product market a competition authority would define, but it aligns the empirical unit with the mechanism far more closely than the EU-wide level.

3.3 Structural variables

- **Core-firm count, n.** Because the cutoff is defined relative to the market but public statistics provide size bands, we report two operational definitions: enterprises with **250+** employees (the statistical definition of a large enterprise, primary caliber) and with **50+** employees. Neither is taken to be the

economic definition of an effective competitor, which is why both are reported and the conclusion is framed in terms of core-firm structure.

- **Coverage, κ.** The share of the country-industry market accounted for by the core, on turnover and employment, with the denominator at the same geographic level as the firms counted.
- **Market size.** Log turnover of the unit.
- **Import penetration.** $imports/(turnover + imports - exports)$, with an imports-over-turnover check; genuine re-exporting units above one are left as recorded.
- **Concentration.** CR4 from the European consolidated register where available.

Continuous predictors are standardized within the estimation sample. Where confidentiality suppresses the large-firm band (notably Ireland, Malta, Greece), we fall back to a broader band and exclude units that cannot be constructed rather than impute.

3.4 The outcome: established infringements

Cases are mapped first to the EU industry-year and then expanded to the member states each case covers; the outcome is defined from disposition, not opening. Over 2011–2024, 62 cases were opened but only **45 (73%) ended in an infringement decision** (7 closed administratively, 2 reaching only a statement of objections, 1 penalizing only obstruction, 7 with no public substantive conclusion). Of the 45, **40 (89%) are EEA-wide**, a structural consequence of the Commission's jurisdictional threshold that nearly removes cross-country variation in raw counts; only 3 cases are multi-country-but-not-EEA-wide or single-country. Of the 45, **23 fall inside the 2011–2024 structural window** (the other 22 predate it), corresponding to 17 event industry-years. The 630-unit cross section contains 210 nominal positive units (33.3%) under the full geographic mapping; after units missing predictors are removed the model-ready sample is **549 units with 193 positives (35.2%)**, and the 23 independent events are the genuine unit of replication. We report both the nominal and model-ready figures and do not silently drop units.

3.5 Descriptive patterns

Core-firm counts vary overwhelmingly between rather than within units, with the between-unit standard deviation roughly eight times the within-unit figure, so the structural variation is cross-sectional. Coverage spans nearly the full range, including values above 95%, and import penetration is higher and more variable in small open economies.

4. Empirical strategy

4.1 Two levels

EU-industry benchmark. We aggregate to the EU division-year and model the count of established infringements with a negative-binomial model and the event indicator with a logit, each as a function of the standardized core count and its square, coverage, market size, import penetration, and year effects. With 24 industries, significance is descriptive; we rely on leave-one-division-out and forward-in-time evaluation and a likelihood-ratio test of the count's incremental power.

Country-industry analysis. We model the infringement indicator at the 549-unit model-ready level as a function of the standardized core count and its square, coverage, market size, and import penetration. A logit with full industry dummies encounters separation and returns extreme coefficients and near-zero standard errors; we do not interpret those, relying instead on a robust specification and, centrally, leakage-free validation.

4.2 Why a country-level holdout is not a valid test

The country-industry outcome is a replication of events, not 630 independent draws. An EEA-wide cartel produces a positive unit in every member state. Under a country-level holdout, when country *c* is withheld, the same cartel remains a positive unit in the other training countries. The model therefore does not have to infer anything about a new market: it has already seen the held-out event, labeled, elsewhere. With an industry fixed effect included, the all-positive event industries are identified almost directly. Performance measured under such a split would largely reflect (i) the replication of the same approximately 23 cartels and (ii) industry identity, rather than generalization to unseen markets. Formally, the held-out units and training units are not independent with respect to the outcome once event identity is fixed. This is the grouped-data form of cross-validation leakage: when records that share a dependence structure (here the same cartel replicated across countries) are split across folds, apparent performance measures memorization of that shared structure rather than generalization, the same mechanism documented for spatial and hierarchical data by Roberts et al. (2017) and, more broadly, in machine-learning-based science by Kapoor and Narayanan (2023). The effective sample size is therefore the number of independent events, about 23, not the 630 replicated units. This bias is measurable rather than hypothetical. Two deliberately naïve designs, a pooled random split and leave-one-country-out, return ROC AUC of 0.739 and 0.730 against the leakage-free value of 0.681, an inflation of about 0.05 to 0.06 ROC (and a comparable gap in precision-recall). Leave-one-country-out does not remove the leakage because the same infringement is still positive in the training countries. A valid evaluation must respect this event structure.

4.3 Leakage-free validation: leave-one-division-out

To guarantee that the held-out cartel appears nowhere in training, we hold out an **entire industry**. Every unit in division *d* is the test set, and training uses only the other divisions. No unit of an event in *d* can then occur in training. Predictions are produced for all held-out units (positive and negative), so ROC and precision-recall are well defined. Standardization is fitted inside each fold on training units only. Because folds hold out whole industries, the cross-country replicated labels cannot leak. We complement the pooled metrics with two baselines. One is the global prior under the same holdout, which uses no industry information at all. The other, computed in sample only and explicitly labeled as non-leakage-free, is the per-industry positive frequency, which bounds the optimistic value of memorizing industry identity.

4.4 The core-count increment and its uncertainty

Within the same folds we compare the full model to one omitting the core count. Because every event industry is all-positive, a within-industry AUC is not estimable (each such industry lacks negatives); we therefore quantify the count's value as the division-cluster-bootstrap difference in pooled ROC and precision-recall AUC, resampling the 24 industries. We report means and 95% intervals rather than a single point.

4.5 Robustness across definitions and models

Section 6 ranks the structural conditions by lower-tail empty region and univariate discrimination, across gradient-boosted, XGBoost, and random-forest models, using interventional TreeSHAP with an explicit background, permutation tests, and false-discovery-rate control; paired ceilings examine whether coverage or imports constrain the count channel. These are parallel univariate and pairwise ceilings, not a multivariate joint-necessity operator. Inference is clustered at the level variation occurs, and no threshold or model is selected on the observations used to evaluate it.

5. Results

5.1 EU-industry level: a significant but weak inverted-U

At the EU division-year level the inverted-U is present. The negative-binomial linear term is positive and significant ($\beta_1 = 1.62, p = 0.006$), the quadratic negative and significant ($\beta_2 = -0.73, p = 0.020$), with a turning point near **1,433 core firms**; the logit gives the same shape and a linear term alone is positive ($\beta = 0.60, p = 0.021$). But leave-one-division-out AUC is **0.635** and a forward-in-time split **0.608**, and once coverage, imports, and size enter, the count adds no significant power (likelihood-ratio $p = 0.27$). The peak is carried by a few of 24 aggregates and does not establish an independent number construct, even though cross-industry studies

have long found collusion concentrated in particular sectors (Symeonidis, 2003). This is why the subsequent analysis operates at the finer country-by-industry level, where the unit of observation matches the mechanism more closely.

5.2 The screen enforces targeting: measured enrichment without leakage

Leave-one-division-out ROC AUC is **0.681** under the 250-employee core (division-cluster bootstrap 95% CI **0.54–0.79**) and **0.688** under the 50-employee core (CI **0.55–0.80**); precision-recall AUC is **0.527** and **0.522** against a base rate of 0.352. The global-prior baseline is 0.500 ROC and 0.352 precision-recall. The structural model therefore carries real information, with a precision-recall gain over the prior of about 0.17, but it is a modest ranking tool rather than an accurate classifier. The wide intervals reflect that only 24 industries supply the signal.

For an authority the relevant question is not whether the score is an accurate classifier but how much it concentrates cases when resources permit reviewing only a fraction of markets. Ranking the 549 model-ready units by the out-of-fold score and reviewing the top fifth captures 0.332 of established infringement units (lift 1.66; division-cluster 95% interval 1.12 to 2.30), and the top tenth captures 0.166 at a selected precision of 0.582 against a random precision of 0.352 (lift interval 1.19 to 2.61). At the top third the lift is still 1.45. Every lift interval excludes one, so the enrichment is not an artefact of a single point. The paired gain over the global prior, and over the leakage-free sector-only rule (which contains no information for a held-out division and reduces to the prior), is about +0.18 ROC, with intervals of roughly 0.04 to 0.29 that exclude zero. Figure 5 plots the cumulative-gain and lift curves.

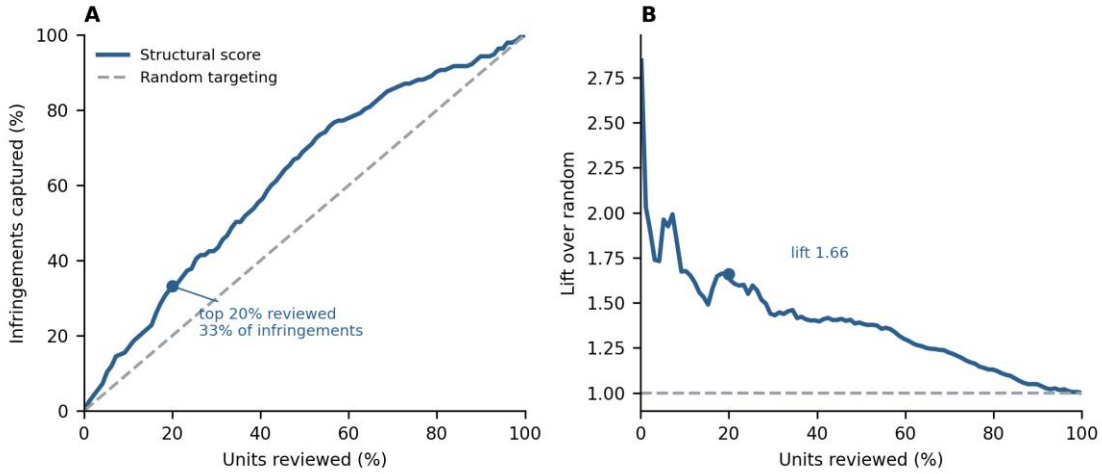


Figure 5. Targeting enrichment of the leakage-free structural score.

Figure 5. (A) Cumulative gain: reviewing the top 20% of units by the out-of-fold risk score captures 33% of established infringement units against 20% under random targeting. (B) Lift over random selection at each review fraction, with lift 1 as the reference.

5.3 The count margin shows no detectable increment

Omitting the count lowers ROC AUC from 0.681 to **0.674** and precision-recall from 0.527 to **0.489**: the paired differences are **Δ ROC 0.007 (CI −0.01 to 0.03)** and **Δ precision-recall 0.032 (CI −0.02 to 0.07)**. The interval contains zero, so the count provides no statistically detectable discrimination gain at this sample size, and even the upper confidence limit is only about 0.03 AUC; any increment that cannot be ruled out is therefore small. We state this as a descriptive bound rather than a formal equivalence result, which would require a pre-specified negligible margin and a two one-sided test. Under P4, core-firm structure as a whole is informative, while the discriminating content runs through coverage together with size and import exposure. The count remains a structural margin with its own economic meaning; it simply does not earn its place through incremental discrimination.

5.4 Discrimination is between industries only

Every industry containing a case, namely C10, C20, C21, C25, C26, C27, C28, and C29, is positive in essentially every country once units are model-ready, while the other sixteen industries are all-negative. There are therefore no within-industry contrasts to predict, and a within-industry AUC cannot be estimated. This is the structural consequence of 89% of cases being EEA-wide: the country dimension expands the sample size without adding independent outcome information. The screen can place an industry higher or lower relative to others; it cannot tell a clean national market in a given industry from a collusive one. Figure 3 documents both the all-positive event industries and the leakage-free performance comparison.

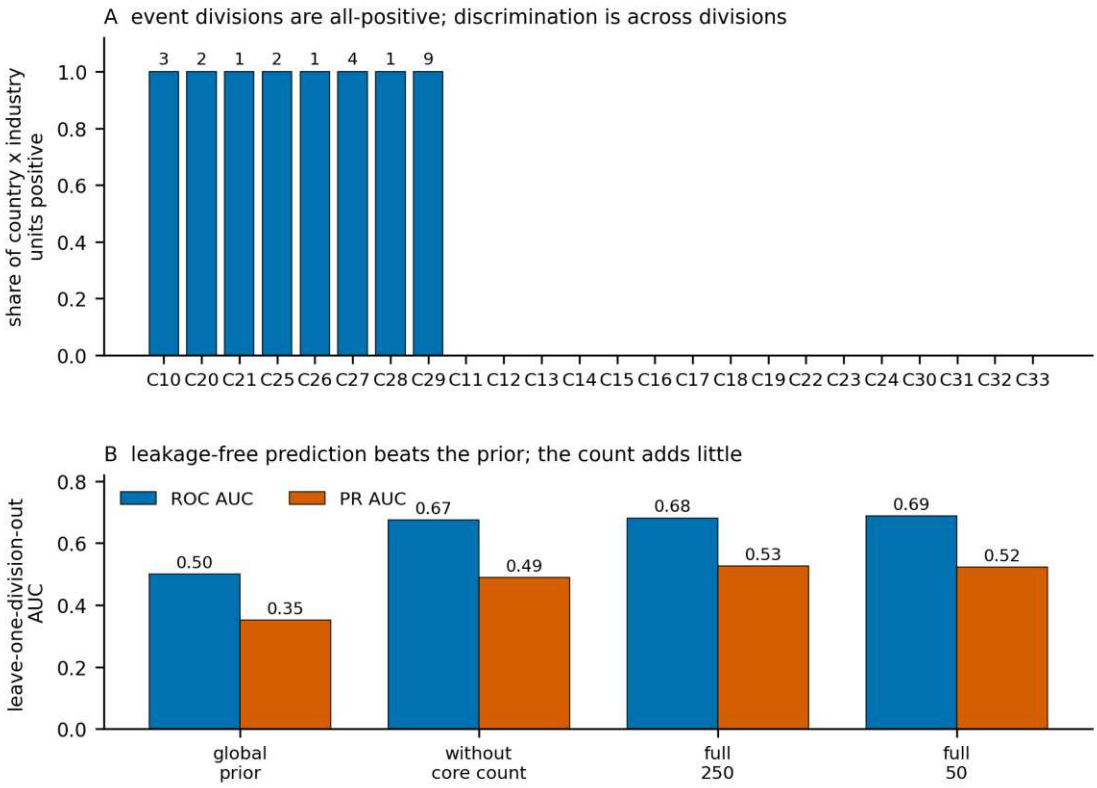


Figure 3. Event-division composition and leakage-free LODO performance.

Figure 3. (A) The eight industries containing established cases are all-positive across countries (labels give event counts); the other sixteen are all-negative. (B) Under leakage-free leave-one-division-out validation the structural model (0.68–0.69 ROC; 0.52–0.53 precision-recall) beats the global prior (0.50; 0.35), while removing the core count changes performance little (0.67; 0.49).

5.5 Coverage is not sufficient

The two industries with the highest coverage, C19 at 96.9% and C30 at 85.8%, had zero detected events, whereas the much lower-coverage industries C14 at 35.3% and C31 at 33.6% did. Coverage alone does not identify collusive markets. High coverage removes the fringe-discipline constraint but leaves the non-structural conditions of Section 2.3 still to be satisfied, which in C19 and C30 they evidently are not. This is P3 in the data.

5.6 Prices: corroboration only

The producer price level is elevated in case industries before proceedings (roughly 92.3 to 95.8 on the indexed basis, $p < 0.001$), consistent with coordinated conduct and with the overcharges documented for established cartels (Connor, 2014), but there is no discontinuity at the proceeding date and inflation and volatility results are specification-dependent. We report this as context, not as an enforcement effect.

5.7 Against the propositions

P1 is supported in the risk sense but the region is not sufficient. P2's strong inverted-U is not robust, so the claim is limited to separable number and coverage effects. P3 is directly confirmed. P4 is supported only for core structure as a whole, not the count's marginal increment. P5 is supported at 0.68–0.69 ROC AUC under leakage-free validation, and is sharpest as targeting enrichment (top 20% review capturing 33% of established infringements). Exploratory forward-in-time evidence points the same way: training on decisions through 2015 and scoring industries that were then silent gives ROC 0.63 over a forward base rate of 0.16, although with only six later independent events the clustered interval is wide and includes 0.5, and window-averaged predictors make this corroboration rather than a strict nowcast. The gain is real but bounded.

6. Robustness: necessary conditions and independent increment

This section asks which structural features exhibit a lower-tail empty region and whether any condition constrains the count channel, following the necessary-condition logic formalized by Dul (2016). Univariate analyses use all units with the relevant variable (about 600 of 630; the multivariate model-ready sample of 549 is used where all predictors enter). We fit gradient-boosted, XGBoost, and random-forest models and compute SHAP with interventional TreeSHAP (explicit background), with pairwise ceilings at the 90th percentile, permutation tests, within-model Benjamini–Hochberg correction, and cross-model aggregation. These are parallel univariate and pairwise ceilings, not a multivariate joint-necessity operator.

6.1 Conditions ranked

Table 1 reports the lower-tail necessary-condition results for the 250-employee core and should appear at this point in the text.

Table 1. Lower-tail necessary-condition analysis (250-employee core).

Condition	Lower-tail threshold	Share in lower tail	Positive rate in lower tail	Overall positive rate	AUC	Permutation p
Core count n_{250}	0.286	9.9%	0.214	0.349	**0.687**	0.002
Market size (log turn)	5.193	9.9%	0.218	0.349	0.668	0.002
Coverage κ_{250}	0.000	9.9%	0.205	0.349	0.661	0.002
Import	0.140	9.9%	0.127	0.349	0.612	0.002

penetration

All four have significant monotone discrimination, with the count strongest at this caliber. But the lower-tail positive rate is **0.127–0.218, not zero**: the lowest decile is lower-risk, not empty. This is the precise reason for a probability statement rather than a strict necessity or safe-harbor claim. Figure 4 illustrates both the coverage counterexamples and the non-empty lower tails.

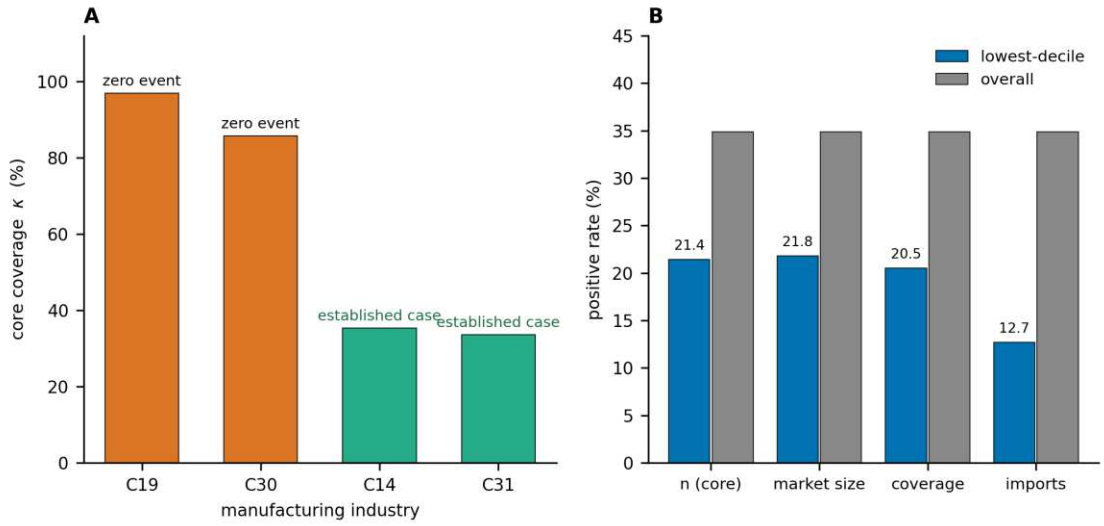


Figure 4. Coverage counterexamples and non-empty lower tails.

Figure 4. (A) Highest-coverage industries have zero events while lower-coverage industries contain cases. (B) The positive rate in the lowest decile of every condition (12.7–21.8%) is below the overall rate (34.9%) but not zero.

6.2 Sensitivity to core definition

Table 2 reports the same ranking under the fifty-employee core definition and should appear at this point in the text.

Table 2. Fifty-employee core.

Condition	AUC	Permutation p
Coverage κ_{50}	**0.689**	0.002
Market size	0.671	0.002
Core count n_{50}	0.650	0.002
Import penetration	0.613	0.002

Lowering the band reverses the top rank: coverage becomes strongest and the count weakens. The conclusion is stated at the level of core-firm structure with both calibers reported; this sensitivity is the empirical echo of the concern that an absolute cutoff is statistical rather than economic, and reporting both is what answers it.

6.3 Paired ceilings: exploratory

With the count as mechanism and coverage or imports as condition, the ceiling effect is $d \approx 0.64$ and 0.63 and the slope is positive in all three models, consistent with the joint-region logic. But no pairing survives false-discovery-rate correction (best $p \approx 0.13$); we report this as consistent direction only, not a significant joint boundary.

6.4 Necessity and predictive importance differ

Mean absolute SHAP ranks variables by total predictive role and orders them differently from the lower-tail ranking: at 250, imports (0.220), size (0.181), count (0.149), coverage (0.127); at 50, coverage (0.238), imports (0.163), count (0.151), size (0.092). Imports do the most predictive work with the weakest lower-tail region. We do not conflate *how much a variable moves predictions* with *whether its absence constrains the outcome*.

6.5 Method boundary

A single held-out country yields only about 23 test units, below the minimum needed to estimate a ceiling, so fold-wise ceilings are undefined and necessary-condition analysis runs only on pooled units. What the country holdout could provide is a fold distribution of an AUC increment, which, given Section 4.2, is itself not a valid estimate. We state the boundary rather than implying cross-validated ceilings. On the smaller restricted geographic mapping coverage is strongest (0.718) ahead of the count (0.668), directional corroboration only.

6.6 What this settles

The count is a caliber-sensitive structural predictor with no demonstrated marginal predictive increment; no condition defines an empty lower region; paired ceilings are directionally consistent but not significant. The method does not address L3 detection selection and does not rescue the safe harbor.

7. Detection, selection, and limitations

7.1 Selection precedes the model

The detection literature has long stressed that observed cartels are a selected sample (Harrington, 2008). We observe $A \times D$ and D is non-random. The Commission's jurisdictional threshold (why 89% of cases are EEA-wide), leniency programs (which cartels are revealed and when, with strategic effects documented by Miller (2009)), and enforcement budgets and priorities all enter the outcome before estimation, and none is independent of the structural predictors. The threshold producing EEA-wide cases is itself correlated with market size and firm numbers. Interventional SHAP conditions on observed covariates but does not correct outcome-level selection. We therefore claim that structure predicts where

established infringements are found, not that it causes cartel formation or that zero-case markets lack cartels.

7.2 Why prevalence is not identifiable

Separating prevalence would require a detection model with an exclusion restriction or a capture–recapture/partial-observability structure with adequate events. The most standardized cross-jurisdiction source, OECD competition statistics (about 69 jurisdictions, 2015–2023), reports sector-level decisions and openings but its sector detail is not openly downloadable, is not standard NACE, and covers a short window. There is no unified European NCA case repository, since the ECN coordinates without maintaining one. Member-state cases must therefore be gathered by hand from annual reports and separated from vertical and RPM matters. A German sample of 18 decisions over 2003–2025 confirms the material exists but illustrates the cost. With 23 window cases and 14 aggregate event cells a partial-observability model is not credibly estimable, and we do not fit one for show.

7.3 Construct limitations

The employee-band cutoff means a capital-intensive 150-employee firm can be a major competitor while a 500-employee firm can be marginal; dual-caliber reporting mitigates but does not close this gap, and firm-level data with common-control identifiers would be needed to define relative-share cores and consolidate groups, which the current count does only approximately. The country-industry unit is coarser than the relevant product market and at times wider or narrower than the geographic market served. The non-structural conditions governing behavior within the region are largely unobserved, which is precisely why prediction reaches 0.68 rather than approaching one.

7.4 What would strengthen the result

Firm-level data for relative-share cores and group consolidation; a harmonized NCA case file supplying within-country variation and detection instruments; and a clear leniency or institutional reform used as a detection-only instrument. These are the path to identifying L2; we do not claim to have taken it.

8. Screening and policy implications

8.1 A one-way prioritization instrument

Screens based on observable data are an established part of cartel detection, alongside price-variance screens such as that of Abrantes-Metz et al. (2006). An authority can use the structural ranking to allocate preliminary inquiry: an ex ante model, built from number-and-coverage structure, market size, and import exposure and validated on industries entirely absent from training, raises the precision of where established infringements are subsequently found (precision-

recall 0.53 versus a 0.35 base rate). The appropriate use is prioritization, not presumption of liability, and the ranking does not substitute for the evidence an infringement decision requires.

8.2 No safe harbor

Every condition's lowest decile still contains established infringements at 12.7–21.8%, and an observed zero reflects non-detection as well as non-collusion. No count threshold clears a market; a threshold selected on the same data used to assess it is not a stable basis for such clearance. The only licensed policy statement is that lower core counts are associated with a lower probability of an established infringement, not zero. The screen is one-way: useful for raising attention, not for granting clearance. This accords with the standard treatment that structural indicators alone cannot establish the absence of coordination (Motta, 2004).

8.3 Design lesson

Using concentration alone conflates the coordination and coverage margins and flags high-coverage zero-event markets that satisfy only one condition while missing the joint requirement. Structural enforcement analysis should ask both whether the firms that must agree are few enough and whether they cover enough, and in small open economies should account for foreign supply captured only partly by imports. The number/coverage decomposition changes which markets are prioritized.

9. Conclusion

Concentration combines two structural facts that should be kept apart: the number of strategically significant decision-makers and the share of the market they cover. The coordinateable region is where a core is few enough to coordinate and large enough to discipline the market, but structure within it is necessary rather than sufficient because feasibility still depends on transparency, buyer power, symmetry, and history. Taking the observation problem seriously, since collusion is latent and only established infringements are seen, we used infringement decisions rather than proceedings, worked at the country-industry level, and, crucially, evaluated the screen with an event-respecting design that holds whole industries out rather than sharing replicated cartels across country folds.

Three findings emerge. First, evaluated without leakage, the structural score is a useful targeting instrument: at 0.68–0.69 ROC AUC and 0.53 precision-recall it concentrates established infringements clearly at the top of the ranking, where reviewing one fifth of markets captures one third of cases at 1.66 times the random rate, and it adds about 0.18 ROC over the uninformative prior. Exploratory forward-in-time evidence points in the same direction. Second, the count of strategically significant firms does not carry this signal on its own: its discrimination increment is statistically undetectable with an upper confidence limit near 0.03, and the

discriminating content runs through coverage together with size and import exposure. Third, the screen ranks industries rather than markets within them, high coverage is not by itself decisive, and no threshold defines an empty lower region.

The contribution is deliberately bounded. The model helps an authority decide where to look, and the targeting gain at the top of the ranking is concrete; with enforcement-selected data the same score cannot establish where collusion is absent, clear a market, or serve as evidence of guilt. Making that boundary precise is itself useful to enforcement, because it rules out the safe-harbor inference that a count threshold would otherwise invite.

Declarations

Conflict of interest. The author declares no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Code availability. The full replication code (data construction, model estimation, leave-one-division-out validation, and the necessary-condition analysis) is provided with this submission as Supplementary Information (Supplementary Code), so that it is available to editors and reviewers without registration or login and without revealing author identity. The code builds the reported results from the raw public files without manual steps. Upon acceptance, the code will additionally be archived in a public repository with a frozen, citable snapshot.

Use of AI. Language-assistance tools were used to improve the readability of the manuscript. The author reviewed and is fully responsible for all content; no AI tools were used to generate or analyze the underlying data.

References

Abrantes-Metz, R. M., Froeb, L. M., Geweke, J. & Taylor, C. T. A variance screen for collusion. *International Journal of Industrial Organization* 24, 467–486 (2006).

Asker, J. A study of the internal organization of a bidding cartel. *American Economic Review* 100, 724–762 (2010).

Bos, I. & Harrington, J. E., Jr. Endogenous cartel formation with heterogeneous firms. *RAND Journal of Economics* 41, 92–117 (2010).

Connor, J. M. *Price-Fixing Overcharges* rev. 3rd edn (Purdue University, 2014).

Davies, S. & Olczak, M. Tacit collusion, firm asymmetries and numbers: evidence from EC merger cases. *International Journal of Industrial Organization* 26, 244–258 (2008).

Dick, A. R. When are cartels stable contracts? *Journal of Law and Economics* 39, 241–283 (1996).

Dul, J. Necessary condition analysis (NCA): logic and methodology of "necessary but not sufficient" causality. *Organizational Research Methods* 19, 10–52 (2016).

Ellison, G. Theories of cartel stability and the Joint Executive Committee. *RAND Journal of Economics* 25, 37–57 (1994).

Friedman, J. W. A non-cooperative equilibrium for supergames. *Review of Economic Studies* 38, 1–12 (1971).

Genesove, D. & Mullin, W. P. Rules, communication, and collusion: narrative evidence from the Sugar Institute case. *American Economic Review* 91, 379–398 (2001).

Kapoor, S. & Narayanan, A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns* 4, 100804 (2023).

Harrington, J. E., Jr. Detecting cartels. In *Handbook of Antitrust Economics* (ed. Buccirossi, P.) 213–258 (MIT Press, 2008).

Harrington, J. E., Jr. & Chang, M.-H. Modeling the birth and death of cartels with an application to evaluating competition policy. *Journal of the European Economic Association* 7, 1400–1435 (2009).

Levenstein, M. C. & Suslow, V. Y. What determines cartel success? *Journal of Economic Literature* 44, 43–95 (2006).

Miller, N. H. Strategic leniency and cartel enforcement. *American Economic Review* 99, 750–768 (2009).

Motta, M. *Competition Policy: Theory and Practice* (Cambridge University Press, 2004).

Poirier, D. J. Partial observability in bivariate probit models. *Journal of Econometrics* 12, 209–217 (1980).

Porter, R. H. A study of cartel stability: the Joint Executive Committee, 1880–1886. *Bell Journal of Economics* 14, 301–314 (1983).

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F. & Dormann, C. F. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40, 913–929 (2017).

Stigler, G. J. A theory of oligopoly. *Journal of Political Economy* 72, 44–59 (1964).

Symeonidis, G. In which industries is collusion more likely? Evidence from the UK. *Journal of Industrial Economics* 51, 45–74 (2003).

Tirole, J. *The Theory of Industrial Organization* (MIT Press, 1988).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx
- SupplementaryCodeandData.zip