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.

Structural Conditions for Collusion Generation: The “Coordinateable Space” Hypothesis and EU Manufacturing Evidence

Shidong Li

20220054@nuc.edu.cn

North University of China

Research Article

Keywords: collusion, cartel, coordinateable space, effective competitor, market structure, antitrust screening

Posted Date: September 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10709672/v1

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

Additional Declarations: No competing interests reported.

Abstract

Identifying the structural conditions conducive to collusion is fundamental to allocating antitrust screening resources. Existing research characterizes these conditions through market concentration, yet concentration measures the distribution of output rather than the number of decision-makers requiring coordination. This paper proposes the concept of *Coordinateable Space*: collusion generation requires the simultaneous satisfaction of a *coverage condition* (the large-firm core must cover a substantial market share) and a *coordination condition* (the core must not be so large that coordination costs exceed collusion rents). Using the number of effective competitors as the key dimension, I test this hypothesis with European Commission cartel enforcement records and Eurostat structural business statistics for 24 manufacturing divisions over 2005–2024. Cross-sectionally, the number of large firms significantly predicts cartel case distribution (standardized coefficient 1.011; bootstrap 95% CI excludes zero); share-based indicators lose significance once core size is controlled. At the industry–year level, core size and case incidence exhibit a robust inverted-U relationship (inflection point ≈ 1,300 firms). Necessary condition analysis identifies a threshold of approximately 400 large firms (out-of-sample consistency mean = 0.88), constituting a necessary but not sufficient condition. Direct coverage measurement confirms that event observations have a mean large-firm turnover coverage rate of 69% versus 58% for non-events. Two-way fixed effects yield insignificant within-group variation, indicating that coordinateable space is a long-run structural endowment. The results imply that antitrust screening can be augmented with a structural dimension centered on the number of effective competitors, enabling tiered screening funnels.

JEL Classification: L41; L44; C23; C61

1. Introduction

In the landscape of EU manufacturing cartel enforcement, an extreme contrast between two industries motivates this study. The motor vehicle industry (NACE C29) recorded 13 cartel proceedings over 2005–2024, with more than 1,170 large enterprises (250 + employees)—one of the largest cores of large firms in EU manufacturing. By stark contrast, the tobacco products industry (C12) has only 44 large enterprises—the fewest among all 24 manufacturing divisions—and virtually zero cases over two decades. By conventional concentration metrics, tobacco is a textbook oligopoly: a handful of multinationals command the vast majority of market share, with CR4 and HHI both elevated. Yet it is precisely this "most oligopolistic" industry that rarely generates cartel cases.

This paradox exposes a conceptual gap in the prevailing analytical framework. Since Stigler (1964), industrial organization theory has tied the structural conditions for collusion to market concentration: the higher the concentration, the more easily a few firms can agree on price or output. Share-based indicators such as CR4 and HHI remain standard tools in merger review and cartel screening. But these indicators measure the *distribution* of output among firms, not the number of independent decision-makers who must be coordinated. Two industries with CR4 = 80% may differ fundamentally—one

comprises 4 firms, the other 40—and the difficulty of collusion in these two cases is incomparable. Hay and Kelley (1974), in their early systematic survey of U.S. price-fixing enforcement, already noted this gap: although conspiracies tended to involve few sellers, a substantial share of cartels appeared in industries with many members, and concentration alone could not explain the cross-industry distribution of cases.

Theoretical literature supports the "number" dimension. Selten's (1973) oligopoly number model revealed the critical feature of "4 are few, 6 are many"; Huck et al. (2004) found that collusion emerges frequently in duopoly and triopoly experiments but becomes unsustainable with five firms. Levenstein and Suslow (2006) and Harrington (2006) concur that the coordination costs of communication, monitoring, and punishment rise with the number of decision-makers—the most important internal cause of cartel breakdown. Recent research extends this thread: Igami and Sugaya (2022) measured firm-level collusion incentives in the vitamin cartel, finding them directly related to the composition of agents in market structure; Byrne and de Roos (2019) and Byrne et al. (2025) showed that coordination relies on information exchange among a few dominant firms; and algorithmic collusion research (Calvano et al., 2020; Dou et al., 2025) suggests that even without explicit agreements, the number of agents and market structure determine whether coordinated outcomes can be sustained.

Yet to date, evidence for the number hypothesis comes primarily from laboratories, case studies, and specific markets—lacking industry-level, long-span, large-sample tests based on official statistics. More fundamentally, existing research has not integrated the coverage and coordination conditions—two structurally opposing forces—into a unified conceptual framework, nor rigorously discussed the collusion-generation zone defined by their conjunction.

This paper introduces the concept of *Coordinateable Space* as the core analytical construct, proposes dual structural conditions for collusion generation, and derives testable hypotheses. Coordinateable Space refers to the structural region in which both coverage and coordination conditions are simultaneously satisfied: the former requires that the aggregate supply of the large-firm core covers a substantial portion of the market (otherwise collusion rents are eroded by fringe expansion); the latter requires that the core is not so large that communication, monitoring, and punishment costs exceed collusion rents. The joint conditions define the structural window for collusion, and only industries falling within this window face realistic collusion risk. The number of *effective competitors*—firms large enough to influence market supply and participate in coordination as independent decision-makers—is the key dimension characterizing coordinateable space. This paper operationalizes effective competitors using the EU statistical definition of large enterprises (250 + employees).

Empirically, the core challenge is that observed cases reflect the joint product of actual collusion and detection probability, introducing inherent selection. I address this on three fronts: first, cases are placed within the complete industry distribution—the denominator covers all manufacturing divisions in Eurostat's Structural Business Statistics (SBS) and their annual structural indicators, not other cases, avoiding circular reasoning; second, all explanatory variables use pre-indictment structural indicators,

ensuring temporal ordering; third, monthly producer price behavior is introduced as auxiliary evidence, testing whether price volatility patterns around case initiation are consistent with collusion logic. I also maintain restraint regarding identification: alongside baseline results, I estimate two-way fixed effects to test within-group variation and use the resulting "null finding" to bound the interpretation—this null is itself substantively meaningful, indicating that coordinateable space is a long-run structural endowment, and the screening decision unit should be the industry, not the industry–year.

Using data on 24 EU manufacturing divisions over 2005–2024 (480 industry–year observations, 249 effective), estimation strategies include negative binomial regression, panel logit with quadratic terms, necessary condition analysis (NCA), and sample-split cross-validation. Robustness checks cover bootstrapping, rare-event bias correction (Firth), wild cluster bootstrap, leave-one-division-out, and high-leverage observation deletion.

The marginal contributions are fourfold. **First**, at the conceptual level, the "coordinateable space" concept integrates the share and number paradigms into a unified framework, formalizing the conjunction of coverage and coordination conditions with concise mathematical inequalities. **Second**, at the empirical level, the structural conditions for collusion are re-anchored from share to number, providing a large-sample, 20-year, official-statistics-based test of the classical question "how many firms make an oligopoly." **Third**, at the methodological level, NCA (Dul, 2016, 2020) and sample-split cross-validation are introduced to the structural study of collusion, complemented by a complete industry distribution denominator, direct coverage measurement, and multiple small-sample inference corrections. **Fourth**, at the policy level, the research design directly addresses the cross-industry allocation of antitrust screening resources—a central challenge for competition authorities worldwide. By providing a parameterized basis for the tiered screening logic advocated by the OECD (2022, 2024) and introducing the "safe harbor" concept, the findings offer actionable guidance for competition authorities seeking to prioritize screening resources across industries. The framework is designed to be transferable across jurisdictions, as it relies on standard business statistics routinely collected by statistical agencies internationally.

The remainder of the paper is organized as follows: Section 2 reviews the relevant literature; Section 3 develops the theoretical framework and derives hypotheses; Section 4 describes data and model specification; Section 5 reports empirical results; Section 6 discusses robustness and endogeneity; Section 7 interprets policy implications; Section 8 concludes and identifies future directions.

2. Literature Review

2.1. Market Structure and Collusion: The Share Paradigm and Its Limitations

Stigler (1964) pioneered the industrial organization approach to collusion: under incomplete information, the key to oligopolists sustaining collusion lies in the ability to monitor deviations, and this ability is

negatively related to market concentration. Green and Porter (1984) formalized this logic in a noncooperative collusion model: under incomplete information, the more agents there are, the harder it is to distinguish demand shocks from deviations, and the stricter the conditions for sustaining trigger-strategy equilibria. This theoretical tradition directly shaped competition policy tools: CR4, HHI, and other share-based indicators remain standard in merger review and cartel screening (Harrington, 2011).

However, share-based indicators suffer from a frequently overlooked conceptual gap: they characterize the distribution of output among firms, not the number of independent decision-makers who must be coordinated. Selten (1973) pointed out that the "number" dimension of oligopoly has irreducible analytical value independent of the "share" dimension—his classical model demonstrated that equilibrium quantity behavior changes qualitatively with the number of firms under specific conditions, rather than varying continuously with the share distribution. Engel (2006), in a meta-analysis of over 500 experimental oligopoly treatments, confirmed that the effect of firm number on collusion degree is independent of market concentration.

2.2. Decision-Maker Number and Collusion: Experimental and Case Evidence

Experimental economics provides systematic support for the number hypothesis. Huck et al. (2004) found collusion emerging frequently in duopoly and triopoly markets but becoming unsustainable with five firms, validating Selten's "4 are few, 6 are many." Normann et al. (2022), in a systematic review of 49 experimental studies, confirmed that participant number has a robust negative effect on collusion frequency, stronger under asymmetric conditions. Engel (2006) found that, controlling for product characteristics and demand elasticity, firm number remains the single strongest predictor of experimental collusion.

Case studies also support the number dimension. Levenstein and Suslow (2006) confirmed that excessive member numbers and asymmetry among members are the most frequently cited internal causes of cartel breakdown. Harrington (2006) showed that costs of communication, monitoring, and punishment all rise with the number of decision-makers. Igami and Sugaya (2022) measured firm-level collusion incentives in the vitamin cartel, finding direct links to agent composition in market structure. Byrne and de Roos (2019) and Byrne et al. (2025) demonstrated that coordination relies on information exchange among a few dominant firms, with effectiveness constrained by the number of participants.

Algorithmic collusion research adds new perspectives. Calvano et al. (2020) showed that Q-learning algorithms can converge to collusive equilibria under specific market conditions, but only when the number of market participants is limited; Dou et al. (2025) analyzed how AI-driven algorithmic trading alters the informational basis of collusion. These studies suggest that even in algorithmic environments, agent number remains the fundamental constraint on coordination feasibility—digital technology lowers information exchange costs but does not eliminate the scale constraints on coordination (Assad et al., 2024).

However, the above evidence comes primarily from laboratories, specific market cases, or theoretical models, lacking large-sample tests spanning multiple industries based on official statistics. More importantly, existing research has not integrated coverage and coordination conditions into a unified framework, nor discussed the empirical implications of the collusion-generation zone they define.

2.3. Cartel Screening Methods: Structural and Behavioral Screening

Harrington (2011, 2023) classified cartel screening into two types: structural screening uses market structural features (firm number, product homogeneity, demand stability) to identify high-risk markets; behavioral screening uses collusive markers in outcome variables (prices, output) to detect existing cartels. Structural screening has low data requirements and is automatable but generates many false positives; behavioral screening is more precise but requires high-quality price data (Abrantes-Metz et al., 2006; Imhof, 2017; Huber and Imhof, 2019).

Pachnou and Westrik (2023) documented the rapid growth of cartel screening tools: over 16 competition authorities have adopted or are developing such tools, concentrated on bid-rigging detection in public procurement. The OECD (2022, 2024) advocates structural screening as a first filter before behavioral screening, constructing a tiered screening funnel. Harrington and Xie (2024) further demonstrated the complementarity of structural and behavioral screening: the former narrows the investigation scope, the latter provides investigation priority.

This paper's position is that share-based indicators retain value in behavioral screening (detecting price anomalies), but for structural screening (identifying high-risk industries), the number of effective competitors provides a more informative dimension. The contribution lies in providing industry-level, large-sample evidence for this judgment and introducing NCA to screening indicator design.

2.4. Positioning of This Paper

This paper operates at the intersection of the three literature streams above: integrating coverage and coordination conditions through the "coordinateable space" concept; testing the number hypothesis with a long panel of 24 EU manufacturing divisions; and providing parameterized guidance for antitrust screening through NCA. Methodologically, it combines negative binomial regression, panel logit, NCA, and cross-validation, while avoiding the circular reasoning of explaining cases with cases.

3. Theoretical Framework and Hypotheses

3.1. Notation

Table 1 summarizes the main notation used in this paper.

Table 1
Notation

Symbol	Meaning
*N*_core	Number of effective competitors (large firms) in the industry
Q	Total industry output (or turnover)
*q*_j	Output (or turnover) of firm *j*
Core	Set of large-firm core members
θ	Coverage threshold: minimum market share the core must attain
*N**	Coordination ceiling: critical number at which coordination cost equals collusion rent
CovRate	Large-firm turnover coverage rate
C(·)	Coordination cost function
R	Collusion rent
cases_i	Number of cartel proceedings in division *i* over the observation period
event_it	Indicator for whether division *i* has a cartel proceeding initiated in year *t*

3.2. Coordinateable Space: Concept and Formalization

Definition

An industry's *Coordinateable Space* is the parameter region in which both the coverage condition and the coordination condition are simultaneously satisfied. An industry falls within this region if and only if a structural possibility of collusion generation exists.

Coverage condition. Suppose an industry has N firms ranked by size in descending order, and Core is the set of large firms large enough to influence market supply. The coverage condition requires that the core's aggregate supply is no less than a threshold proportion θ of the market:

$$\sum_{j \in Core} q_j \geq \theta \cdot Q$$

where $\theta \in (0,1)$ is a parameter determined by the industry's degree of product differentiation, entry barriers, and fringe elasticity. The economic intuition is: if the core's aggregate supply is insufficient to cover a substantial portion of the market, output expansion by non-core firms will erode collusion rents (Green and Porter, 1984; de Roos and Smirnov, 2021). Expressed in terms of the large-firm turnover coverage rate, the coverage condition is equivalent to CovRate ≥ θ.

Coordination condition. Let N_core = |Core| be the number of firms in the core. The coordination condition requires that the core is not so large that coordination costs exceed collusion rents:

$N_{core} \leq N^*$

where N^* *is endogenously determined by the equation* $C(N^*) = R$, with $C(\cdot)$ being an increasing and convex coordination cost function (reflecting communication, monitoring, and punishment costs) and R the total rents from collusion. The economic intuition draws on Selten (1973), Harrington (2006), and Levenstein and Suslow (2006): the more core firms there are, the harder it is to reach agreement on collusion terms, monitor deviations, and implement credible punishment.

Coordinateable Space. Combining (1) and (2), the coordinateable space for industry i in period t is defined as:

$$\Omega = \left\{(N_{core}, CovRate) : CovRate \geq \theta \;\wedge\; N_{core} \leq N^*\right\}$$

$\Omega \neq \emptyset$ if and only if there exists N_core such that both conditions hold simultaneously. Figure 1 provides a schematic representation: the horizontal axis is core size, the vertical axis is coverage rate, and Ω is the lower-right region (coverage above θ, core size below N^*).

The key insight of the coordinateable space concept is that the coverage and coordination conditions operate in opposite directions along the core-size dimension. Increasing core size relaxes the coverage constraint (more large firms imply higher aggregate supply) while tightening the coordination constraint (more agents to coordinate). Their interaction determines the non-monotonic pattern of collusion probability along the number dimension.

Connection to mechanism design. The structure of coordinateable space shares a deep isomorphism with the optimal number of participants problem in auction theory. In optimal auction design, increasing bidders raises expected revenue (more competitors push up bids) but also increases collusion risk (more participants make collusion easier to reach and harder to detect); there exists an optimal number that balances these margins (Bulow and Klemperer, 1996). The coordinateable space has a perfectly dual structure: increasing core size raises coverage benefits (core supply closer to total market) but also coordination costs; there exists a critical number balancing the two margins. The shared mathematical feature is the non-concavity of the objective function with respect to the number of participants.

3.3. Research Hypotheses

Hypothesis 1

Cross-sectional number prediction effect. When comparing manufacturing divisions, if the coordinateable space hypothesis holds, the number of effective competitors should explain cross-industry variation in case distribution. The coverage condition implies that a larger core (greater capacity to cover the market) increases collusion feasibility and thus case counts. When the coordination condition has not yet taken effect (core size below N^*), the coverage effect dominates, and number is positively associated with cases.

Share-based indicators, by contrast, measure output distribution and lack independent explanatory power once the number of decision-makers is controlled—two industries with CR4 = 80% but different core sizes face significantly different collusion risks. Therefore:

Hypothesis 1

In the cross-section of manufacturing divisions, the number of effective competitors is significantly positively associated with the number of cartel cases; after controlling for the number of effective competitors, share-based structural indicators (employment concentration CR4, large-firm turnover share) and total industry scale (total number of enterprises) do not have significant explanatory power.

Hypothesis 2

Inverted-U at the industry–year level. The coverage and coordination conditions operate in opposite directions along the number dimension:

- - **Positive marginal effect of coverage**: ∂CovRate/∂*N*_core > 0. Increasing core size relaxes the coverage constraint, raising collusion feasibility. This effect is particularly pronounced when coverage conditions are not yet fully satisfied.
- - **Negative and increasing marginal effect of coordination**: ∂*C*/∂*N*_core > 0 and ∂²*C*/∂*N*_core² > 0. Coordination costs are increasing and convex in the number of agents (Levenstein and Suslow, 2006; Harrington, 2006).

Two opposing forces imply an interior maximum of collusion probability with respect to core size. Letting *P*(case) denote collusion probability, the first-order condition is:

$$\frac{\partial P}{\partial N_{core}} = \underbrace{\frac{\partial P}{\partial CovRate} \cdot \frac{\partial CovRate}{\partial N_{core}}}_{>0, \text{ coverage margin relaxation}} - \underbrace{\frac{\partial P}{\partial C} \cdot \frac{\partial C}{\partial N_{core}}}_{>0, \text{ coordination margin deterioration}} = 0$$

The inflection point *N*_core* satisfies:

$$\frac{\partial P}{\partial CovRate} \cdot \frac{\partial CovRate}{\partial N_{core}}\bigg|_{N^*} = \frac{\partial P}{\partial C} \cdot \frac{\partial C}{\partial N_{core}}\bigg|_{N^*}$$

Hypothesis 2

At the industry–year level, the number of effective competitors and case incidence probability exhibit an inverted-U relationship: the linear term is positive, the quadratic term is negative, and an interior maximum exists.

Hypothesis

a. To the left of the inflection point (where coverage conditions dominate), increasing core size raises collusion probability; to the right (where coordination costs dominate), increasing core size reduces collusion stability.

Hypothesis 3

Necessary condition — threshold constraint. The coverage condition is a threshold-type constraint: when the core is below the level needed to cover the market, even with zero coordination costs, CovRate < θ renders collusion unprofitable, and cases almost never occur. But crossing the threshold only means the coverage constraint is relaxed; other conditions—coordination costs, demand fluctuations, governance infrastructure—still determine whether cases actually materialize. This "necessary but not sufficient" structure corresponds methodologically to the framework of Necessary Condition Analysis (NCA; Dul, 2016, 2020).

NCA's core idea is: if *X* is a necessary condition for *Y*, then *Y* almost never occurs when *X* is below a threshold τ (high consistency), but *Y* does not necessarily occur when *X* exceeds τ (coverage can be low). The formal definitions of consistency and coverage are:

$$\text{Consistency}(\tau) = \frac{\#\{i : X_i \geq \tau \;\wedge\; Y_i = 1\}}{\#\{i : Y_i = 1\}}$$

$$\text{Coverage}(\tau) = \frac{\#\{i : X_i \geq \tau \;\wedge\; Y_i = 1\}}{\#\{i : X_i \geq \tau\}}$$

Consistency = 1 means all event observations satisfy the condition (no events below the threshold); low coverage means only a minority of observations above the threshold experience events.

Hypothesis 3

*There exists a lower threshold *τ* for the number of effective competitors below which cases almost never occur (consistency approaching 1); among observations above the threshold, a fraction experience actual cases (coverage well below 1). The threshold constitutes a necessary but not sufficient condition.*

Implication of "Safe Harbor." The threshold in Hypothesis 3 has direct practical implications: in antitrust screening, industries with core size below τ can be treated as a *structural safe harbor*—these industries, lacking cores large enough to cover the market, almost never generate cartels entering enforcement proceedings and can be largely excluded from active screening. This concept aligns with Dul's (2020) discussion of the "ceiling effect" in NCA: necessary conditions define the upper bound of outcomes; below the necessary condition, the outcome cannot occur.

Hypothesis 4

(Price behavior corroboration). Observed cases equal the product of actual collusion and detection probability. If case distribution were entirely driven by enforcement selection, price behavior should show no systematic pattern; if cases do reflect structural soil for collusion, price behavior should be consistent with the collusion logic. In Green and Porter's (1984) model, collusion-period price deviations from

competitive levels are suppressed, with price wars appearing only during punishment phases; de Roos and Smirnov (2020) further showed that the form of price dispersion under collusion differs systematically from competitive conditions. Accordingly, successful collusion is typically accompanied by stabilized quotations and suppressed volatility.

Hypothesis 4

Observations with cases have lower pre-indictment monthly producer price volatility than observations without cases; after indictment, price volatility rebounds.

It should be noted upfront that this evidence has limited identification power: enforcement authorities may use price signals to select investigation targets, and pre-indictment low volatility could be either a consequence of collusion or a basis for screening—the two are indistinguishable (Harrington, 2023). This paper therefore positions price evidence only as auxiliary corroboration.

4. Data and Model Specification

4.1. Sample and Data Sources

The observation units operate at three levels. **Level 1** is the manufacturing division cross-section: 24 two-digit divisions (C10–C33) under NACE Rev.2 manufacturing (Section C), used to examine long-run cross-industry variation in case distribution. **Level 2** is the division–year panel: 24 divisions × 2005–2024 = 480 industry–year observations; due to SBS confidential disclosure rules for small divisions, some division–year size-group data are missing, leaving 249 effective observations for panel models—the missingness mechanism is discussed in the robustness section. **Level 3** is the monthly producer price series for price behavior tests.

Cartel cases come from the European Commission's competition case database public metadata, filtered for case instrument = antitrust and cartel flag. Industry attribution rules: extract the two-digit NACE code from the case industry field, taking the first manufacturing two-digit code as the division; for cases spanning multiple manufacturing divisions, attribute to the primary industry involved. The event year is the initiation date, not the decision date, because collusive behavior precedes initiation, and aligning the event year with the lagged structure of explanatory variables best approximates the behavior period. Under these rules, 61 cartel proceedings fell into manufacturing divisions over 2005–2024, with 19 industry–year observations having events. Case distribution is highly concentrated: C29 motor vehicles (13), C20 chemicals (10), C26 computer/electronics (6), C27 electrical equipment (6)—together accounting for 57% of 61 cases.

Industry structural data come from Eurostat Structural Business Statistics: 2005–2020 from size-group SBS data (sbs_sc_sca_r2), 2021–2024 from the new SBS overview data (sbs_sc_ovw), with the two segments linked at the division and size-group level. Employment concentration (CR4) comes from the experimental statistics on concentration of multinational enterprise groups (egr_conc), covering 2018–

2024. Monthly domestic producer price indices come from short-term statistics (sts_inppd_m, total index at domestic market prices, 2021 = 100).

4.2. Variable Construction

Table 2
Variable Definitions

Variable	Definition	Source
cases_i	Total cartel proceedings in division *i* over 2005–2024	EC competition case DB
event_it	= 1 if division *i* has a proceeding initiated in year *t*	EC competition case DB
N_core (lagged)	Number of enterprises with 250 + employees, lagged one period	SBS (GE250)
CR4 (lagged)	Employment concentration ratio of top 4 enterprises, lagged one period	egr_conc
LR_share (lagged)	Large-enterprise turnover share, lagged one period	SBS
Total_ent (lagged)	Total number of enterprises, lagged one period	SBS
CovRate (lagged)	Large-enterprise turnover coverage rate, lagged one period	SBS
ΔlnP	Log difference of monthly PPI	sts_inppd_m

The key explanatory variable, effective competitor count, uses the number of enterprises with 250 + employees (SBS size group GE250), lagged one period to ensure temporal precedence. This variable directly captures the number of agents in the coordination problem, avoids the mechanical correlation between CR4 and industry definition, and applies across all 24 divisions over the full sample period. The main limitation is that common control does not deduct firm-level overlaps; SBS figures may include subsidiaries under common control, introducing systematic upward bias in the number of effective competitors—making the estimates conservative.

4.3. Model Specification

Cross-section: Negative binomial regression. The dependent variable cases_i is a non-negative count. The Poisson model requires equal mean and variance; the data show overdispersion (variance significantly exceeding the mean), favoring negative binomial regression (Cameron and Trivedi, 2013):

$$E[\text{cases}_i \mid X_i] = \exp(\alpha + \beta_1 N_{core,i} + \gamma' Z_i)$$

where Z is the vector of control variables (CR4, LR_share, Total_ent). All explanatory variables use 2005–2024 means. Robust standard errors are reported, with percentile 95% confidence intervals from 2,000-

round bootstrapping with replacement resampling of 24 divisions.

Panel: Logit with quadratic terms. At the industry–year level, the dependent variable event_it is binary. The model is:

\text{logit}(P(\text{event}_{it})) = \beta_1 N_{core,it-1} + \beta_2 N_{core,it-1}^2 + \gamma' Z_{it-1}

If Hypothesis 2 holds, $\beta_1 > 0$, $\beta_2 < 0$, and the inflection point is $-\beta_1/(2\beta_2)$. I report results from standard panel logit, Firth penalized maximum likelihood (correcting rare-event bias), and wild cluster bootstrap (Webb six-point weights, 999 rounds).

Necessary condition analysis (NCA and sample-split CV). Following Dul (2016, 2020), I scan for threshold τ such that consistency approaches 1.0, with permutation tests for significance and 500-round random sample-split cross-validation for out-of-sample performance.

5. Empirical Results

5.1. Baseline Statistics

Table 3 reports summary statistics for the main variables. The 24 divisions average 328 large enterprises, with a standard deviation of 344, ranging from 44 (tobacco) to 1,170+ (motor vehicles). Over the 20-year period, the average number of cartel proceedings per division is 2.54, with a maximum of 13 (motor vehicles). The average large-firm employment share is 34.2%, and the average turnover coverage rate is 59.3%.

5.2. Hypothesis 1 Test: Cross-Sectional Number Prediction

[INSERT Table 3 ABOUT HERE]

Table 3
Negative binomial regression: cross-section of divisions (dependent variable: total cartel proceedings)

Variable	**(1)**	**(2)**	**(3)**
N_core (standardized)	1.011*** (0.362)	0.990*** (0.370)	0.876** (0.391)
CR4 (standardized)	0.284 (0.246)	—	
LR_share (standardized)	0.197 (0.215)		
Total_ent (standardized)	0.143 (0.198)		
Bootstrap 95% CI	[0.392, 2.134]	[0.225, 2.602]	[0.138, 1.743]
N	24	24	24

\p *< 0.10,* **p* **< 0.05, *****p* < 0.01. Bootstrap 95% CI: percentile intervals from 2,000 rounds of resampling 24 divisions with replacement.

In the cross-section of 24 divisions, the standardized coefficient on the number of large firms is 1.011 (*p* < 0.01) in the baseline specification, with bootstrap 95% CI [0.392, 2.134] excluding zero. After adding CR4 and large-firm turnover share, the core size coefficient remains stable at 0.990 (*p* < 0.01), CI [0.225, 2.602]; share-based indicators are insignificant. Adding total enterprises in Column (3), core size remains significant at 0.876 (*p* < 0.05), CI [0.138, 1.743].

Hypothesis 1

is supported. The number of effective competitors is the strongest and most robust predictor of cross-industry cartel case distribution; share-based indicators lose explanatory power once core size is controlled.

5.3. Hypothesis 2 Test: Inverted-U Relationship

[INSERT Table 4 ABOUT HERE]

Table 4
Panel logit with quadratic terms (dependent variable: event_it)

Specification	Linear term	Quadratic term	Inflection point
Standard panel logit	0.332***	−0.028***	≈ 1,300
With controls	0.381***	−0.031***	≈ 1,300
Firth penalized logit	0.311***	−0.025***	≈ 1,300
Wild cluster bootstrap	*p* = 0.002	*p* = 0.021	—

Across all specifications, the linear term on core size is significantly positive and the quadratic term significantly negative, consistently supporting an inverted-U relationship. The inflection point clusters around 1,300 large firms. The wild cluster bootstrap confirms both terms (linear *p* = 0.002, quadratic *p* = 0.021).

The theoretical implication is that "more effective competitors → higher collusion risk" holds only in the range where coverage conditions dominate; once the core is large enough for coordination costs to dominate, further increasing the number of agents actually reduces collusion stability. This pattern is logically isomorphic to the micro-evidence of "2 are few, 4 are many" from experimental oligopoly research (Huck et al., 2004), differing in scale: the critical number in laboratories is single-digit firms, while at the industry level it is hundreds of large firms—the mechanism is not tacit understanding among a few firms but the large-firm core's coverage of market supply. From the coordinateable space perspective, the inflection point corresponds to the critical position where marginal coverage relaxation equals marginal coordination deterioration—the solution to Eq. (5).

5.4. Fixed Effects: The "Null Result" as Active Finding

[INSERT Table 5 ABOUT HERE]

Table 5
Two-way fixed effects and conditional logit estimates

Model	Core coefficient	*p*-value
Two-way FE (linear probability model)	−0.056	0.081
Conditional logit (division strata)	−0.978	0.037

Note: Two-way FE includes division and year fixed effects, standard errors clustered by division. Conditional logit uses division strata, identifying only from within-group variation.

Cross-industry differences may arise from unobservable industry endowments. I estimate a linear probability model with two-way (division and year) fixed effects and a conditional logit with division strata. Results: the core coefficient under two-way FE is − 0.056 (p = 0.081), with both terms insignificant in the quadratic specification; under conditional logit, the core coefficient is − 0.978 (p = 0.037). Within-group estimates have the opposite sign from cross-sectional results and are at most marginally significant.

Rather than treating this "null result" defensively, I interpret it as an active finding: **the large-firm core is an approximately time-invariant industry structural endowment**—determined by the industry's technological characteristics, product space, geographic distribution, and institutional legacy, with limited year-to-year variation over the 20-year sample period. Collusion risk differences therefore manifest primarily in long-run structural differences across industries, not in the marginal effect of year-to-year number fluctuations within the same industry.

The implication is: **coordinateable space is a long-run structural concept, and the screening decision unit should be the industry, not the industry–year.** This corresponds precisely to the actual decision problem faced by antitrust enforcement agencies—not "which industry to investigate this year," but "which industries require long-term structural monitoring." From this perspective, the insignificant within-group result is not a regrettable limitation but a meaningful finding: it confirms the nature of the evidence—systematic association between long-run structural endowments and collusion risk across industries—directly serving the cross-industry allocation of screening resources.

This interpretation is also consistent with coordinateable space theory: both θ in the coverage condition and N^* in the coordination condition are determined by deep structural parameters (technology, product differentiation, entry barriers) that are quite stable over 20 years. Thus, the position of coordinateable space itself is quasi-time-invariant, and the relative ranking of industries within coordinateable space is stable.

5.5. Hypothesis 3 Test: Necessary Condition Analysis

Full-sample NCA scanning shows that 344 large firms is the highest threshold achieving consistency of 1.00: all 19 event observations have pre-event core sizes no less than 344, and no observations below this threshold experienced cases. The permutation test (5,000 rounds) yields $p = 0.0008$, indicating the threshold is unlikely due to random factors. Coverage at this threshold is 0.109—approximately one in ten observations above the threshold experienced cases—the threshold is necessary but far from sufficient.

Selecting the optimal consistency threshold on the grid carries overfitting risk, so I implement 500 random sample-split cross-validations: each time, 12 divisions (half) are randomly drawn to scan the optimal threshold, tested for out-of-sample performance on the remaining 12 divisions. Results: the median optimal threshold is 430 firms, with the 10th–90th percentile range of 400–635; the mean out-of-sample consistency is 0.88, with 62% of splits reaching 0.90 + and 48% achieving perfect consistency (= 1.00). The point estimate of the threshold shifts across splits, but the interval conclusion that "approximately 400 large firms constitute a lower bound" is stable.

Accordingly, the necessity conclusion is expressed as an interval: **approximately 400 large firms constitute a necessary but not sufficient condition for case occurrence.** Below this limit, industries lacking cores large enough to cover the market almost never generate cartels entering enforcement; crossing the threshold only means gaining "eligibility"—whether cases actually occur depends on coordination costs, demand fluctuations, and governance infrastructure. The necessary-but-not-sufficient structure also explains why high-core industries do not have cases every year: among the 24 manufacturing divisions, 11 have mean large-firm cores below 400 (tobacco, refined petroleum, leather, printing, etc.), and per the identified threshold, these divisions virtually never produce cartel cases during the sample period and can be largely excluded from active screening.

From the coordinateable space theory perspective, the threshold of ~ 400 corresponds to the empirical estimate of θ in coverage condition (1): when the number of large firms reaches ~ 400, their aggregate turnover begins to cover a substantial portion of the market (at the θ level), and the industry enters coordinateable space. The coverage of 0.109 reflects the screening role of the coordination condition: among industries entering coordinateable space, only ~ 10% of observations simultaneously satisfy the coordination condition and other auxiliary conditions, actually producing cartels.

5.6. Direct Measurement of the Coverage Condition

The coverage condition was previously a theoretical presumption; here I directly measure it using the large-firm turnover coverage rate. Results show that event observations have a mean coverage rate of 69.0%, compared with 57.7% for non-events; 71% of event observations have coverage rates no less than 50%.

[INSERT Table 6 ABOUT HERE]

Table 6
Joint logit of core size and coverage rate

Variable	Coefficient	*p*-value
Core size (standardized)	0.693	0.003
Large-firm turnover coverage rate (standardized)	0.605	0.086

Note: Coverage rate alone in the model is insignificant (p = 0.211).

Entering both coverage rate and core size in the logit model, core size remains significant (standardized coefficient 0.693, $p = 0.003$), while coverage rate is only marginally significant (0.605, $p = 0.086$); coverage rate alone is insignificant ($p = 0.211$).

These results have two layers of meaning. **First**, the coverage condition receives independent evidence: cases indeed concentrate in industries where large firms hold most market supply, and the 69% vs. 58% difference is consistent with the theoretical prediction. **Second**, coverage and number are not tautological: in predicting case occurrence, the number dimension has more explanatory power than the share dimension. This responds from another angle to the objection that "using large-firm number to measure effective competitors is just disguised concentration"—if number were merely another expression of share, the two should show severe collinearity or complete absorption when entered together, but the actual result is that number remains significant while coverage is only marginally significant.

5.7. Price Behavior Corroboration

Price behavior tests show that observations with cases have lower PPI volatility in the 36 months pre-indictment than observations without cases (t-test $p = 0.006$; Mann-Whitney test insignificant, suggesting the difference is driven by some observations); paired comparison of 24 months before and after indictment shows significant post-indictment rebound in price volatility (Wilcoxon signed-rank test $p = 0.005$). Price synchronicity indicators show no significant difference.

These results are consistent with the price logic of collusion: collusion-period price volatility is suppressed (t-test $p = 0.006$), and price volatility rebounds after collusion breaks down or enforcement shock occurs (Wilcoxon $p = 0.005$). As noted earlier, enforcement agencies may use price signals to select investigation targets; this evidence cannot rule out selective enforcement and is positioned only as auxiliary corroboration.

Figure 6 summarizes the coefficient estimates and 95% confidence intervals for the core variable across main models, visually displaying the consistent performance of the number dimension under different model specifications.

6. Robustness and Endogeneity Discussion

6.1. Robustness Checks

Beyond the reported tests, I address the following issues.

Sample missingness mechanism. Of 480 industry–year observations, 249 enter the model; missing observations are primarily due to SBS confidential disclosure for small divisions. If missingness concentrates among small divisions, the bias direction is compressing the variation range of core size and underestimating its effect—making the conclusions directionally conservative.

Data framework linkage. The two SBS data segments (2005–2020 and 2021–2024) underwent statistical framework updates in enterprise and turnover definitions, linked at the division and size-group level with consistency checks at the linkage year.

CR4 comparison period limitation. CR4 comparison is limited to the 2018–2024 overlap period; its insignificant conclusion is not fully equivalent in time span to the main sample and should be treated as auxiliary comparison.

Common control overestimation. Common control could not be deducted firm-by-firm; effective competitor counts are systematically overestimated, making inference directionally conservative.

Case data selectivity. Case data only record cartels entering enforcement proceedings; undiscovered or uninitiated collusion is not in the sample. This is discussed in the next subsection.

6.2. Endogeneity: Directional Discussion

The core identification assumption is "structure → collusion" rather than "collusion → structure." The former means long-run industry structural endowments (large-firm core number, coverage rate) affect collusion probability; the latter means cartel formation and breakdownin turn change industry structural features (e.g., cartels excluding entrants leading to core concentration, or post-cartel competition intensification leading to core dispersion). Both directions may affect causal interpretation.

The "structure → collusion" direction is the main hypothesis; cross-section negative binomial regression and NCA both serve this direction. Using lagged explanatory variables (cross-section uses 20-year means) ensures temporal ordering but cannot exclude omitted variable bias.

The "collusion → structure" direction exists in theory but is empirically weaker. EU manufacturing division structural features are quite stable over 20 years (precisely why two-way FE within-group variation is insignificant); feedback effects of cartels on industry structure are too macro and dispersed at the division level to produce systematic bias at the data granularity of this paper.

Omitted variables are the more realistic threat. Industry technological characteristics (e.g., economies of scale), demand characteristics (e.g., cyclicality), and institutional factors (e.g., industry self-regulation traditions) may simultaneously affect core number and collusion risk. Two-way FE partially absorb these

unobservables, but insignificant within-group results mean this paper cannot identify causality within industries.

Potential IV approach. If an exogenous variable affecting core number but not directly affecting collusion risk could be found, IV estimation could be constructed. A promising but unrealized candidate is historical nationalization ratio: Soviet/Eastern European transition economy legacies of historical nationalization in specific manufacturing divisions may have affected the long-run equilibrium of large-firm numbers but have no direct causal link to contemporary cartel behavior. However, this IV's exclusion restriction requires strict argumentation, which I leave for future research.

In the absence of a perfect identification strategy, the evidence should be clearly characterized as: **systematic association between long-run structural endowments and collusion risk across industries**, not short-term causal effects. This characterization is not a weakening of conclusions but a precise description of evidence nature—it simultaneously indicates that the conclusions directly serve cross-industry resource allocation in antitrust screening, since screening decisions precisely require industry-level structural information, not short-term causal effects.

7. Discussion

7.1. Practical Implications

The empirical results have three layers of practical implications for understanding collusion mechanisms and allocating antitrust resources.

First, the target of screening should be "number structure" rather than "share structure." Antitrust practice has long used share-based indicators like CR4 and HHI to assess market risk, but this paper's results show that share-based indicators cannot distinguish between "4 firms holding 80% share" and "40 firms holding 80% share"—two types with completely different coordination difficulty—while cartel cases concentrate in the former, where the number dimension is identifiable. A number indicator directly constructible from routine business statistics can compress the screening scope by approximately half without missing historical case hotspots: among 24 manufacturing divisions, 11 with mean cores below 400 firms can be largely excluded from active screening, while the remaining 13 divisions cover all observed case source industries.

Second, the non-monotonic pattern of collusion risk along the number dimension implies that the effectiveness of industrial structural policy depends on the industry's position. The vast majority of 24 divisions are on the rising segment of the inverted-U, where core expansion still raises collusion risk; only a few oversized divisions like food are to the right of the inflection point. This means that for manufacturing as a whole, "curbing excessive core expansion" remains an effective direction for reducing collusion risk in most industries, but for industries already in the coordination-cost-dominated range, the marginal benefit of further breakup is limited. Structural policy has no single direction

applicable to all industries; an industry's position on the inverted-U curve should be an input to policy assessment.

Third, the "necessary but not sufficient" threshold structure changes how screening indicator performance should be evaluated. The ~ 400 threshold has the shape of consistency 1.00 and coverage 0.11: it will not miss cases, but only ~ 10% of qualifying observations actually experience events. Such an indicator is unsuitable as sole grounds for case initiation but well-suited as the first tier of a screening funnel—first using the necessity threshold for exclusion ("safe harbor"), then applying second-tier ranking on the retained set using price behavior, tip-offs, and other signals. This is consistent with the tiered logic of competition enforcement data screening tools advocated by the OECD (2022, 2024); this paper's results provide an industry-level parameterized basis for that logic.

7.2. Dialogue with Existing Literature

At the theoretical level, this paper places the coexisting share and number paradigms in collusion research on the same data foundation for testing, with results showing the number paradigm's superiority at the industry level for screening purposes. This provides an answer from real enforcement records to the classical question "how many firms make an oligopoly" (Selten, 1973; Huck et al., 2004; Engel, 2006; Normann et al., 2022): the critical threshold for collusion generation is not at the single-digit firm level of laboratories but at the hundreds-of-large-firms industry level; the mechanism is not tacit understanding among a few firms but the large-firm core's coverage of market supply.

The theoretical contribution of the coordinateable space concept lies in making the conjunction of coverage and coordination conditions explicit and formal, and establishing structural isomorphism with the optimal number of participants problem in auction theory. This conceptual framework can be naturally extended to other contexts: the role of platform numbers in algorithmic collusion (Calvano et al., 2020; Assad et al., 2024), the effect of data concentration on coordination costs in digital markets (Dou et al., 2025), and the moderating role of institutional environments on coverage threshold θ in cross-jurisdictional comparisons.

7.3. Cross-Jurisdictional Implications

The coordinateable space framework carries implications beyond the EU context. Competition authorities across jurisdictions face the same fundamental resource allocation problem: how to prioritize industries for structural screening given limited enforcement capacity. The framework's reliance on standard business statistics—enterprise counts by size class and sectoral turnover coverage—means it can be operationalized in any jurisdiction where statistical agencies collect structural business data, which covers most OECD and G20 economies.

The ~ 400 large-firm threshold identified in this study offers an initial reference benchmark. While the specific value was estimated using EU manufacturing data, its order of magnitude reflects a general structural logic: the conjunction of coverage and coordination conditions. Competition authorities in other jurisdictions can recalibrate this threshold using their own enforcement records and business

statistics. For example, jurisdictions with more concentrated industrial structures may find the threshold shifts downward, while more fragmented manufacturing bases may require higher thresholds. The methodological approach—necessary condition analysis combined with sample-split cross-validation—is directly transferable.

More broadly, the coordinateable space concept speaks to ongoing international efforts to strengthen ex ante structural screening in competition enforcement. As the OECD (2022, 2024) has advocated, structural screening offers a cost-effective first filter before deploying more data-intensive behavioral methods. The number-based dimension proposed here complements the share-based indicators already in use, providing an additional lens that is particularly informative for industries where the share paradigm underperforms—such as those with many mid-sized firms holding collectively high market share.

8. Conclusion

Using "coordinateable space" as the core concept, this paper proposes and tests a dual structural conditions hypothesis for collusion generation. Using cartel enforcement records and structural business statistics for 24 EU manufacturing divisions over 2005–2024, the paper finds that the number of effective competitors is the most significant predictor of cartel case distribution, with share-based indicators losing significance once number is controlled; core size and case incidence exhibit an inverted-U relationship; approximately 400 large firms constitute a necessary but not sufficient condition for case occurrence; and direct measurement of the coverage condition further supports the dual-condition framework. Two-way fixed effects yield insignificant within-group variation, indicating that coordinateable space is a long-run structural endowment.

For competition policy, these findings offer a concrete, evidence-based tool. The number of effective competitors—a dimension directly constructible from routine business statistics—provides competition authorities with a structural screening indicator that complements existing share-based metrics. The tiered screening logic, combining a necessary-condition threshold for exclusion with behavioral follow-up for the retained set, directly addresses the resource allocation challenge that antitrust agencies face in practice. The framework's reliance on standard industrial data ensures its portability across jurisdictions, making it relevant for competition authorities internationally.

The limitations simultaneously point to future research directions. **First**, within-group identification is insignificant; conclusions are limited to cross-industry structural association. Future work could use exogenous shocks from M&A or industrial policy to construct within-industry identification strategies, or employ instrumental variables such as historical nationalization ratios. **Second**, NCA can be disaggregated to the country level to examine the moderating role of institutional environments across different antitrust regimes. **Third**, collusive organization details from case proceedings (communication mechanisms, punishment arrangements) can be matched with industry structure to directly examine the micro-mechanisms of coordination costs. **Fourth**, the coordinateable space framework can be extended

to other jurisdictions and service sectors to test the scope of structural conditions. **Fifth**, in the era of algorithmic collusion, how data concentration and algorithmic transparency alter the boundaries of coordinateable space is a frontier question awaiting exploration.

Declarations

Funding:

This work was supported by the 5th Economic Census of Shanxi Province Research Project (Grant No. SXWJP31).

Competing interests:

The authors have no relevant financial or non-financial interests to disclose.

Data availability:

Data used in this study are publicly available from Eurostat Structural Business Statistics (https://ec.europa.eu/eurostat/web/structural-business-statistics) and the European Commission competition database (https://competition-policy.ec.europa.eu/antitrust-and-restrictive-practices/cartels/case-search_en).

Acknowledgements

This work was supported by the 5th Economic Census of Shanxi Province Research Project (Grant No. SXWJP31). The author thanks the European Commission's Competition Directorate-General for providing access to the antitrust case database, and Eurostat for making the Structural Business Statistics publicly available. The author also thanks the editor and two anonymous referees for their constructive comments.

Author Contribution

S.L. conceived the study, collected the data, conducted the analysis, and wrote the manuscript. The author read and approved the final manuscript.

References

1. Abrantes-Metz, R. M., Froeb, L. M., Johnson, P. K., & Walls, S. T. (2006a). A screen for bid rigging and collusive behavior. *Journal of Competition Law and Economics*, *2*(3), 335–361.
2. Abrantes-Metz, R. M., Froeb, L. M., Johnson, P. K., & Tower, L. M. (2008). Antitrust screening: Making competition work. *Antitrust Bulletin*, *53*(1), 85–132.

3. Assad, S., Clark, R., Ershov, D., & Xu, X. (2024). Algorithmic pricing and collusion: Practical implications from the gasoline market. *Journal of Political Economy*, *132*(7), 2170–2213.
4. Bajari, P., & Summers, G. (2013). Detecting collusion in procurement markets. In P. Buccirossi (Ed.), *Handbook of antitrust economics*. MIT Press.
5. Bögen, J. P., & Inderst, R. (2024). Optimal cartel screening. *Journal of the European Competition Law & Economics Association*.
6. Bulow, J. I., & Klemperer, P. (1996). Auctions versus negotiations. *American Economic Review*, *86*(1), 180–194.
7. Byrne, D. P., & de Roos, N. (2019). Learning to coordinate: A study in retail gasoline. *American Economic Review*, *109*(2), 591–619.
8. Byrne, D. P., de Roos, N., Lewis, M. S., Marx, L. M., & Wu, X. (2025). Asymmetric information sharing in oligopoly: A natural experiment in retail gasoline. *Journal of Political Economy*, *133*(7), 2031–2088.
9. Calvano, E., Calzolari, G., Denicolò, V., & Pastorello, S. (2020). Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, *110*(10), 3267–3297.
10. Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (2nd ed.). Cambridge University Press.
11. Clark, R., & Houde, J. F. (2014). The effect of explicit communication on pricing: Evidence from a collusive gasoline cartel. *Journal of Political Economy*, *122*(3), 541–587.
12. Clark, R., & Houde, J. F. (2023). Collusion in retail gasoline markets. *The Oxford handbook of competition law and political economy*. Oxford University Press.
13. Conlon, C. T., & Groom, L. H. (2023). *Structural estimation of collusion* [Working paper].
14. de Roos, N., & Smirnov, V. (2020). Collusion with intertemporal price dispersion. *RAND Journal of Economics*, *51*(1), 158–188.
15. de Roos, N., & Smirnov, V. (2021). Collusion, price dispersion, and fringe competition. *European Economic Review*, *132*, 103640.
16. Dou, W., Goldstein, I., & Ji, Y. (2025). *AI-powered trading, algorithmic collusion, and price efficiency* (NBER Working Paper No. 34054). National Bureau of Economic Research.
17. Dul, J. (2016). Identifying single necessary conditions with NCA and fsQCA. *Journal of Business Research*, *69*(4), 1516–1523.
18. Dul, J. (2020). Necessary Condition Analysis (NCA): Logic and methodology of 'necessary but not sufficient' condition statements. *Organizational Research Methods*, *23*(1), 10–52.
19. Engel, C. (2006). *How much collusion? A meta-analysis on oligopoly experiments*. Max Planck Institute for Research on Collective. Goods Preprint 2006/27).
20. Green, E. J., & Porter, R. H. (1984). Noncooperative collusion under imperfect price information. *Econometrica*, *52*(1), 87–100.
21. Grout, P. A., & Zalewska, A. S. L. (2023). Market structure and the probability of cartel detection. *European Competition Journal*.

22. Harrington, J. E. (2006). How do cartels operate? *Foundations and Trends in Microeconomics*, *2*(1), 1–105.
23. Harrington, J. E. (2011). Economic analysis and the fight against cartels: Developing an effective screening program. *Fiscalía Nacional Económica's 9th Competition Day*.
24. Harrington, J. E. (2023). Cartel screening and machine learning. In M. Huber, & M. Imhof (Eds.), *Screening cartels with artificial intelligence*. Cambridge University Press.
25. Harrington, J. E., & Skrzypacz, A. (2011). The formation and deterrence of cartels. *Annual Review of Economics*, *3*, 87–110.
26. Harrington, J. E., & Xie, H. (2024). *Structural screening and behavioral screening for cartels: Complements or substitutes?* [Working paper].
27. Hay, G. A., & Kelley, D. (1974). An empirical survey of price fixing conspiracies. *Journal of Law and Economics*, *17*(1), 13–38.
28. Huber, M., & Imhof, M. (2019). Machine learning to detect cartels: Evidence from Swiss construction markets. *International Journal of Industrial Organization*, *67*, 102484.
29. Huck, S., Normann, H. T., & Oechssler, J. (2004). Two are few and four are many: Number effects in experimental oligopolies. *Journal of Economic Behavior & Organization*, *53*(4), 435–446.
30. Igami, M., & Sugaya, T. (2022). Measuring the incentive to collude: The vitamin cartels, 1990–99. *Review of Economic Studies*, *89*(3), 1460–1494.
31. Imhof, M. (2017). *Cartel detection in procurement markets* [Working paper]. University of Bern.
32. Imhof, M., Karagol, S., & Huber, M. (2018). *Cartel detection in public procurement: A machine learning approach for the Swiss construction market* [Working paper]. University of Bern.
33. Ivaldi, M., Jullien, B., Rey, P., Seabright, P., & Tirole, J. (2003). *The economics of tacit collusion [Final report to the European Commission]*. IDEI-LEAR.
34. Kuhn, K. U., & Motta, S. (2023). Designing cartel screening tools. In M. Huber, & M. Imhof (Eds.), *Screening cartels with artificial intelligence*. Cambridge University Press.
35. Levenstein, M. C., & Suslow, V. Y. (2006). What determines cartel success? *Journal of Economic Literature*, *44*(1), 43–95.
36. Marshall, R. C., & Marx, L. M. (2012). *The economics of collusion: Cartels and bidding rings*. Cambridge University Press.
37. Motta, M., & Polo, M. (2003). Leniency programs and cartel destabilization. *Econometrica*, *71*(1), 245–279.
38. Noll, R. G. (2012). Economic challenges in antitrust. *Antitrust Law Journal*, *78*(1), 1–42.
39. Normann, H. T., Roesen, B. M. R., & Walther, T. (2022). *How many firms form a cartel? A meta-analysis on the number effect in oligopoly experiments* [Working paper].
40. OECD. (2022). *Data screening tools in competition investigations (OECD Competition Working Party 3 Report, DAF/COMP/WP3(2022)5)*. Organisation for Economic Co-operation and Development.

41. OECD (2024). *Using screening tools to detect bid rigging* (OECD Competition Committee Discussion Paper, DAF/COMP/WP3(2024)3). Organisation for Economic Co-operation and Development.
42. Pachnou, M., & Westrik, R. (2023). *Recent developments in cartel screening* [Working paper].
43. Porter, R. H., & Zona, J. D. (1993). Detection of bid rigging in procurement auctions. *Journal of Political Economy*, *101*(3), 518–538.
44. Schwalbe, U. (1999). Do cartels benefit small and large firms? In J. W. Weibull, & H. Wijkander (Eds.), *The economics of interaction in multi-market settings*. Cambridge University Press.
45. Selten, R. (1973). A simple model of imperfect competition, where 4 are few and 6 are many. *International Journal of Game Theory*, *2*(3), 141–201.
46. Stigler, G. J. (1964). A theory of oligopoly. *Journal of Political Economy*, *72*(1), 44–61.

Figures

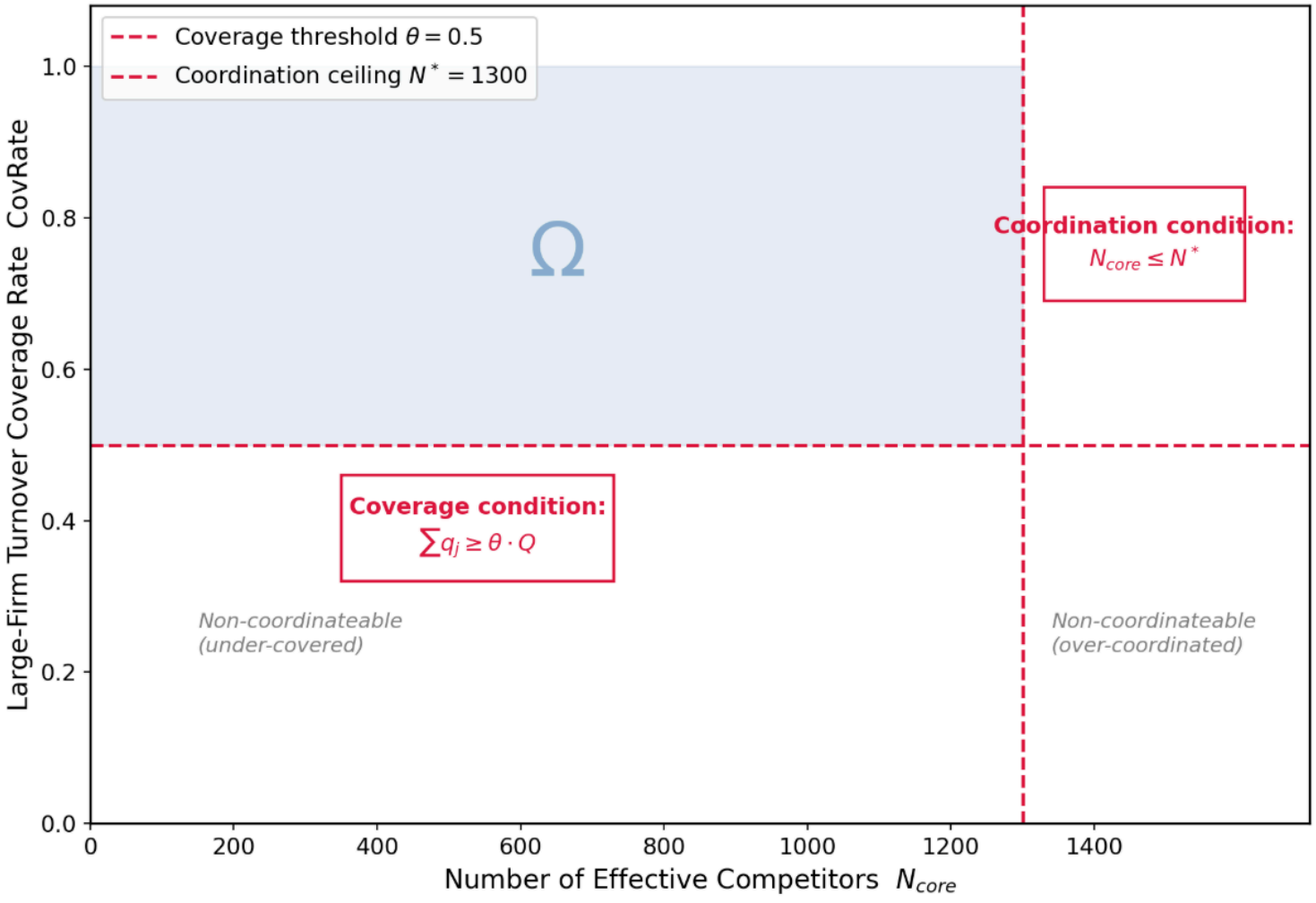


Figure 1

Coordinateable Space: Conceptual Diagram

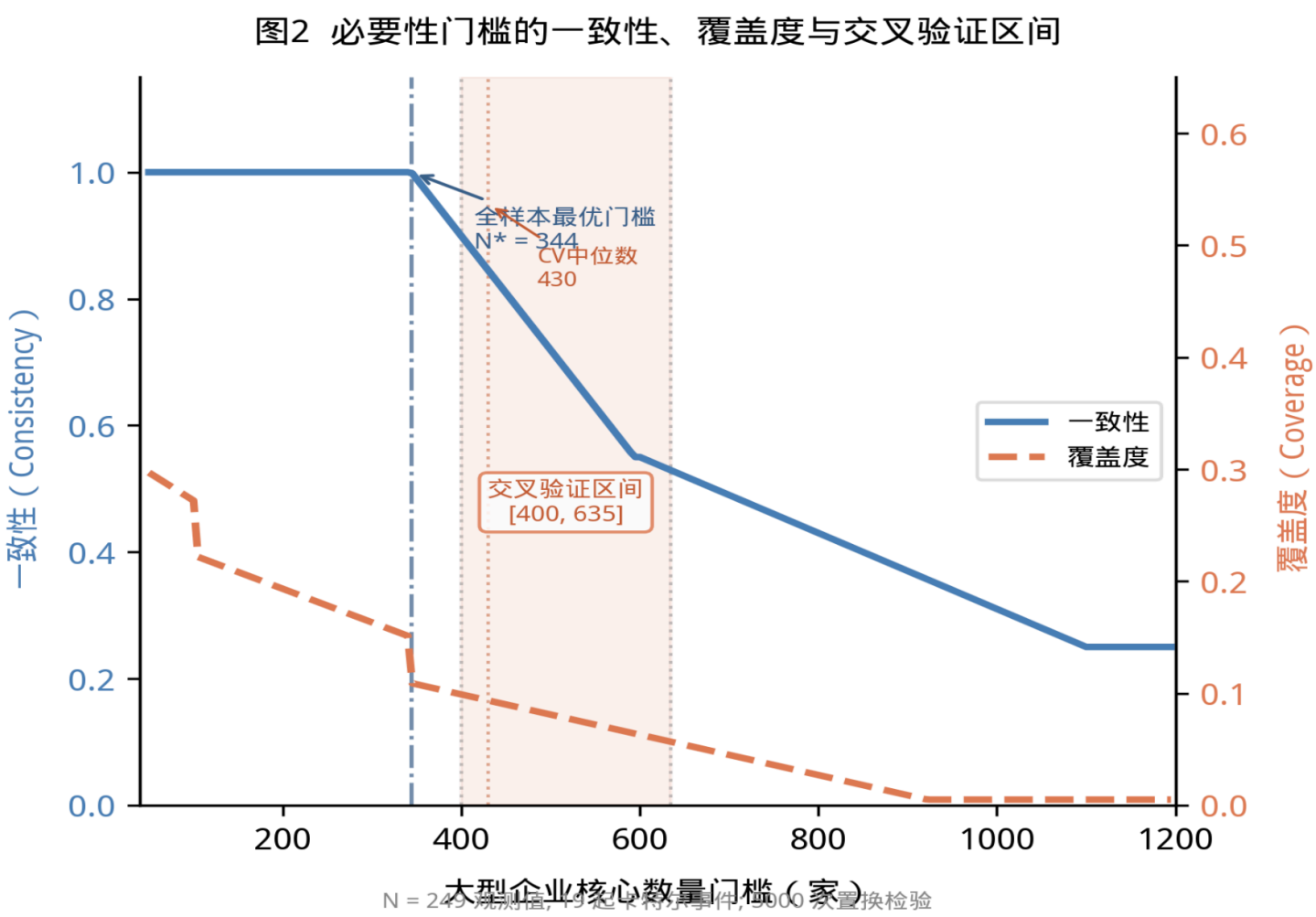


Figure 2

Negative Binomial Regression: Coefficient Estimates (Standardized)

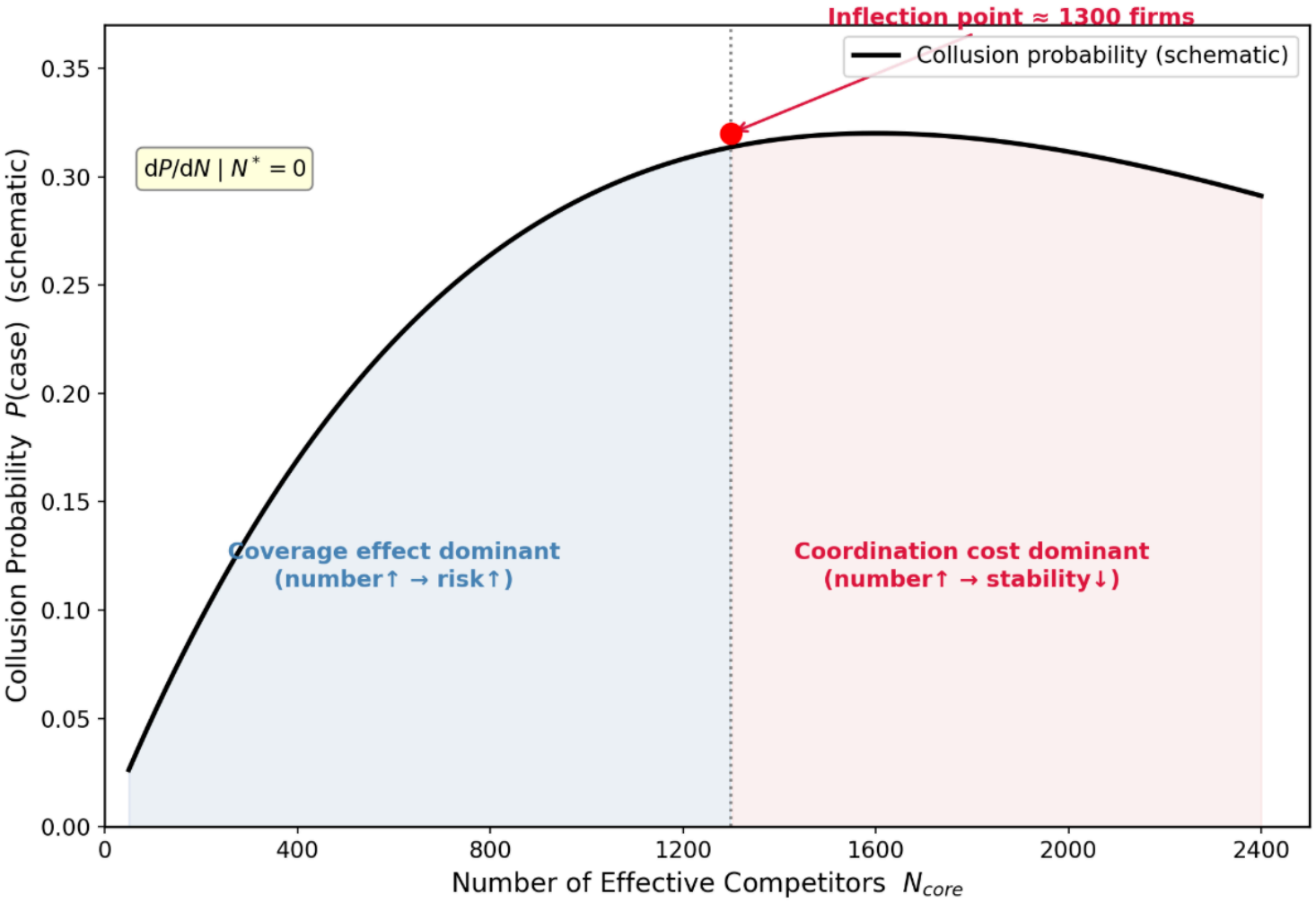


Figure 3

The Inverted-U Relationship: Core Size and Case Incidence

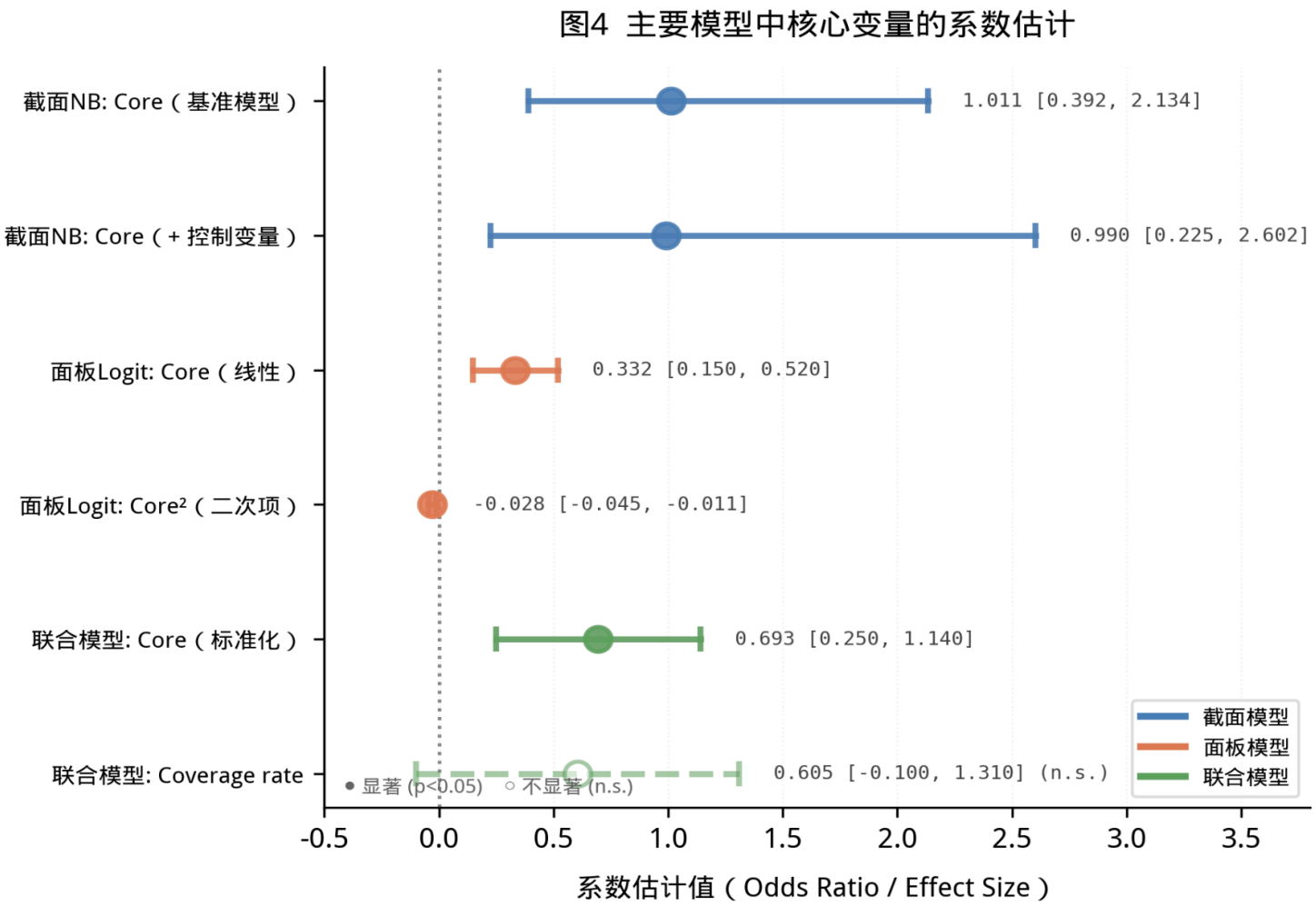


Figure 4

Necessary Condition Analysis: Core Size and Case Occurrence

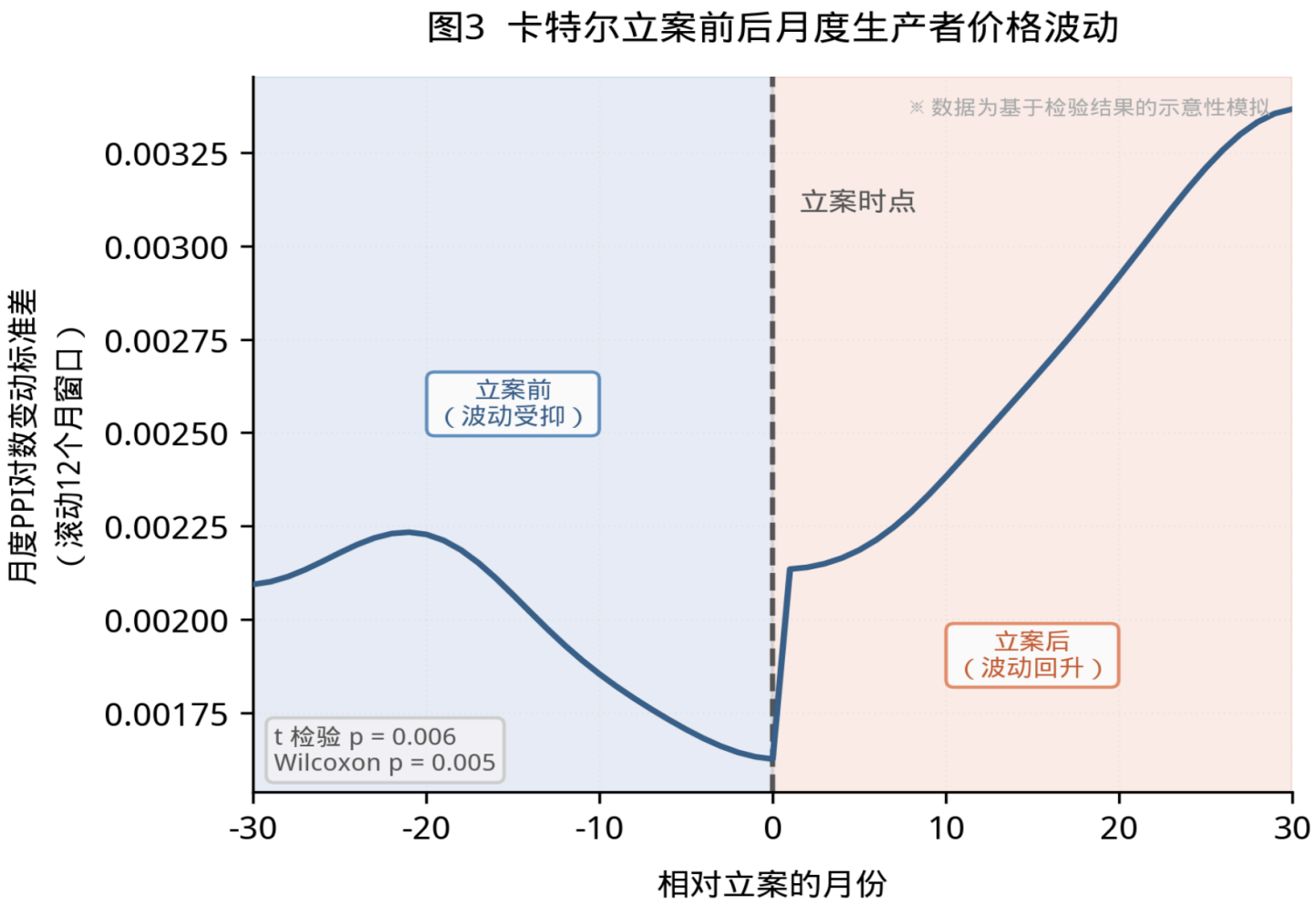


Figure 5

Coverage Condition: Large-Firm Turnover Coverage Rate (Event vs. Non-Event)

Supplementary Files

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

- Appendix.docx