SplitComp: Stress-Testing Compute Governance Under Strategic Evasion and Jurisdictional Fragmentation

Abutalha Alam[*]
University College Dublin
research.abutalhaalam@gmail.com

Najmul Hasan[*]
University of North Carolina at Pembroke
nh0033@bravemail.uncp.edu

Shivam Dubey
Algoverse
dubeyjiworks@gmail.com

Joel Christoph[‡]
Harvard Kennedy School, Harvard University
jchristoph@hks.harvard.edu

Jonas Kgomo[‡]
Equiano Institute
jonas@equiano.institute [*]

Abstract

Compute is measurable, but measurability is not enforcement. We extend the Becker deterrence model with a third action (splitting a training run across jurisdictions to reduce effective detection) and a parameter σ that captures cross-border information sharing. A 50,000-point Latin hypercube over $[0, 1]^6$ under uniform priors yields compliance on 11.8% of the space, evasion on 83.1%, and splitting on 5.1%. A one-at-a-time sensitivity analysis shows that the private benefit of non-compliance swings compliance by roughly 20 percentage points; the three enforcement parameters (audit probability, detection accuracy, penalty severity) each swing it by about 14 points and enter only through their product. Under strong enforcement ($a = d = F = 0.6$), weak coordination below $\sigma \approx 0.54$ redirects strategic actors into splitting rather than compliance, a fragmentation margin that a two-action Becker model cannot represent. We treat the 98.8% compliance observed in the EU ETS by Calel et al. [7] as a qualitative benchmark rather than a forecast. Code: https://github.com/najmulhasan-code/splitcomp.

1 Introduction

Policymakers treat compute as the most promising lever for frontier AI regulation because cloud providers meter it, data centers host it, and chip supply chains leave a paper trail. Model capabilities and downstream harms leave no comparable trail. That asymmetry explains why the EU AI Act, the OECD, and the UK all anchor their recent proposals on compute thresholds.

Observability is not enforcement. A lab facing a compute threshold has at least three strategies beyond compliance: it can under-report within one jurisdiction, relocate to a jurisdiction with weaker oversight, or split a training run across data centers in several jurisdictions so that no auditor sees the whole run. None of these strategies are exotic. They are familiar from tax enforcement and environmental regulation. The EU AI Act authorises fines up to €35 million or 7% of global turnover

[*]Equal contribution [‡]Advisors Algoverse

Preprint.

and requires reporting above specified compute thresholds [11], but the infrastructure for auditing, detection, and cross-border coordination is still being built. That gap is where a strategic actor has room to manoeuvre.

We study that gap with a minimal model. A representative lab chooses among complying, evading in one jurisdiction, and splitting across several. The regulator is summarized by four parameters: audit probability a, detection accuracy d given an audit, penalty severity F, and cross-border coordination σ. The model is one-shot, static, and risk-neutral. It will not reproduce real compliance rates. That is the point. Any gap between this model and a real regime is a measure of the work done by mechanisms the model omits (repeated play, reputation, administrative routine, risk aversion). Used this way, the gap is informative rather than a weakness.

2 Related Work

Rational-choice deterrence. The classical benchmark is Becker [5]: a rational actor breaks a rule when the private benefit exceeds the audit probability times the detection probability times the sanction. Ehrlich [10] carries this logic to U.S. crime data and estimates deterrence elasticities, and Polinsky and Shavell [24] survey how the theory was extended over the following three decades to handle errors, marginal deterrence, settlements, and self-reporting. The Becker benchmark is known to underfit empirical compliance. Harrington [15] documents this in environmental regulation and traces the residual to state-dependent inspection and dynamic scrutiny; the resulting "Harrington paradox" has been the reference point for enforcement theory since. Nagin [22] reviews the subsequent evidence and concludes that the certainty of apprehension dominates the severity of punishment as a deterrent, which is the reason we keep a and F as separate parameters.

Inspection games and imperfect detection. Inspection games formalize a two-player interaction between an actor who may cheat and an inspector who may catch. The earliest applications were in arms-control verification [21]. Avenhaus et al. [4] trace the field from those origins through auditing, accounting, and environmental surveillance. More recently, Boone and Dahan [6] analyze inspection when detection is imperfect at the component level and show that optimal resource allocation responds to heterogeneity in how easy different components are to inspect. We use their result as the reason to expect the three-way symmetry between a, d, and F in our sensitivity analysis to break once detection is modelled as compute-scale dependent rather than as a free parameter.

Empirical enforcement in market-based regulation. Gray and Shimshack [13] review environmental monitoring and enforcement and report that audits produce substantial specific and general deterrence and reduce emissions beyond the facility that was inspected. Martin et al. [20] survey the first decade of EU ETS evaluations across emissions, economic performance, and innovation. Colmer et al. [9] use firm-level data and find that the EU ETS cut CO_2 emissions by 14–16% with no detectable outsourcing to unregulated firms; this result constrains our reading of the split action, which should appear only in specific parameter regions rather than as an automatic response to partial-coverage regulation. Calel et al. [7] show 98.8% compliance across 31 EU ETS regulators despite weak formal enforcement, and find that variation in audit probability and penalty severity across countries explains only about one-tenth of the variation in compliance; this is the reference rate we compare to. Fowlie [12] estimates that exempting out-of-state producers from California's electricity regulation achieved roughly a third of the emission reductions of complete coverage at more than twice the cost per ton. Aichele and Felbermayr [1] find that Kyoto commitments raised committed countries' embodied carbon imports from non-committed countries by about 8%. Both studies are empirical analogues of the cross-jurisdiction leakage that our split action captures.

Tax enforcement and cross-border evasion. Deterrence of tax evasion has an even longer empirical track record than environmental enforcement. Andreoni et al. [3] and Slemrod [26] both conclude that audits, third-party reporting, and information-sharing regimes move compliance more than sanction severity alone. Kleven et al. [18] report the clearest numbers: in a randomized Danish audit experiment, evasion was 0.3% for income subject to third-party reporting and 37% for self-reported income. We read that asymmetry as evidence that our d is largely a function of the reporting environment rather than a free lever. Zucman [27] estimates that about 8% of global household financial wealth sits in tax havens, three-quarters of it unrecorded. Johannesen and Zucman [17] show that the G20 bank-secrecy crackdown pushed deposits into less-compliant havens rather than back

home; Alstadsæter et al. [2] estimate that the top 0.01% of Scandinavian households evade roughly a quarter of their taxes through offshore vehicles. Jurisdictional fragmentation is a documented response to partial-coverage enforcement in a related domain.

AI governance and institutional design. Kolt et al. [19] propose responsible-reporting requirements for frontier developers and argue that regulatory visibility depends on reliable information flowing from developers to policymakers; this is the institutional counterpart of the d parameter in our model. Rauh et al. [25] review more than two hundred safety evaluations applied to generative AI and find systematic gaps across modality, risk type, and deployment context; these gaps tell us where the detection infrastructure our model abstracts is thinnest. Chan et al. [8] characterize the growing agency of algorithmic systems, which makes the case that a static enforcement baseline like ours is incomplete but still useful as a reference. Hadfield and Clark [14] propose regulatory markets as an alternative institutional form for AI oversight; private regulators would still operate under the audit, detection, and coordination constraints our model quantifies. None of this existing work examines the enforcement side explicitly for jurisdictional fragmentation, which is the gap we address.

3 The model

Three actions. The lab's action set is {*comply*, *evade*, *split*}. Complying means following the rule. Evading means violating it within one jurisdiction; a successful audit costs the full expected penalty. Splitting means spreading one training run across data centers in several jurisdictions, each segment below the local reporting threshold, so no single auditor sees the whole run. Splitting is more costly than evasion (network latency, synchronisation overhead, subsidiary legal structures), but any one auditor now sees only a fragment, which lowers expected detection.

A two-action Becker setup collapses all non-compliance into evasion and cannot represent the regulator-arbitrage margin. Adding a third action makes cross-border coordination an explicit parameter.

Six parameters and their policy meaning. The model has six parameters, all on $[0, 1]$ (Table 1). Four come from deterrence theory: private benefit B of non-compliance [5], audit probability a, detection accuracy d given an audit [6], and penalty severity F.

Two are specific to the compute-governance setting: c, the extra cost of splitting, and σ, the fraction of detection effectiveness preserved when a run is spread across jurisdictions. At $\sigma = 1$, regulators share everything and splitting buys nothing. At $\sigma = 0$, regulators share nothing and splitting makes the lab invisible to any single enforcer.

Each parameter maps to a real policy lever. B proxies the competitive advantage of an undisclosed frontier model. a reflects inspection capacity and reporting triggers. For scale, IAEA [16] reports more than 3,000 verification activities per year across 1,300+ facilities; no jurisdiction has anything comparable for AI compute. d reflects whether regulators can draw on hardware telemetry and cloud-provider logs or must rely on self-reporting. F reflects sanction credibility. σ reflects cross-border information sharing. The OECD [23] Common Reporting Standard, covering 100+ jurisdictions, is the institutional benchmark for what coordinated monitoring at scale requires.

Table 1: Model parameters and their policy interpretation.

Parameter	Interpretation
B	private benefit of non-compliance (competitive advantage from undisclosed activity)
a	audit probability (inspection staffing, reporting triggers)
d	detection accuracy given audit (telemetry quality, evidence access)
F	penalty severity (credibility and scale of sanctions)
c	additional cost of splitting compute across jurisdictions
σ	cross-border coordination strength (fraction of detection preserved under splitting)

Payoffs and decision thresholds. Following Becker [5], each action's payoff is the private benefit minus expected cost:

$$U(\text{comply}) = 0 \tag{1}$$
$$U(\text{evade}) = B - a \cdot d \cdot F \tag{2}$$
$$U(\text{split}) = B - c - a \cdot d \cdot \sigma \cdot F \tag{3}$$

The lab selects the action with the highest payoff. Three thresholds follow: Compliance beats evasion when $B < adF$; compliance beats splitting when $B - c < ad\sigma F$; and splitting beats evasion when $c < adF(1 - \sigma)$. The third threshold is the contribution of the three-action formulation: it isolates the parameter region in which fragmentation is the preferred form of non-compliance, a region a two-action model cannot represent.

4 Methods

A dense six-dimensional grid is computationally out of reach, so we combine several techniques. We classify each parameter vector by evaluating the three payoffs in closed form and taking the argmax, which avoids numerical error at the decision boundaries. For visualization we generate two-dimensional heatmaps that sweep two parameters while holding the remaining four at illustrative baseline values $B = 0.60$, $a = 0.40$, $d = 0.50$, $F = 0.50$, $c = 0.10$, $\sigma = 0.25$, chosen to describe a regime with weak enforcement and low coordination. At this baseline $adF = 0.10$, so $U(\text{evade}) = 0.50$ and $U(\text{split}) = 0.475$, placing the lab in the evasion region of the action space. To estimate aggregate action shares we draw a 50,000-point Latin hypercube sample over $[0, 1]^6$ and classify each point. The sensitivity analysis fixes each parameter first at 0.2 and then at 0.8, samples the other five with a 10,000-point Latin hypercube per setting, and reports the swing in the compliance share between the two settings.

Figure 1 shows the classifier applied at baseline on four parameter pairs.

5 Results

Finding 1: Evasion dominates the uniform-prior draw. Across the 50,000-point Latin hypercube sample, evasion is the preferred action on 83.1% of the parameter space, compliance on 11.8%, and splitting on 5.1% (Figure 2, Table 2). Splitting is rare because it requires two conditions simultaneously: weak cross-border coordination *and* enforcement strong enough that the detection savings pay for the coordination cost. Most draws fail at least one condition. The 11.8% compliance figure is a property of the model under a non-committal prior, not a forecast of real-world behaviour (Section 6).

Table 2: Action fractions from 50,000 Latin hypercube samples with uniform priors over $[0, 1]^6$.

Action	Fraction
Comply	11.8%
Evade	83.1%
Split	5.1%

Finding 2: Private benefit dominates the sensitivity ranking. Raising B from 0.2 to 0.8 cuts compliance by roughly 20 percentage points, the largest effect (Figure 3). The three enforcement levers a, d, and F each swing compliance by approximately 14 points (14.0, 13.9, and 14.3 respectively). The near-identical swings are structural: a, d, and F enter only through the product adF, so the model treats them as substitutable. This symmetry is a testable prediction; Boone and Dahan [6] argue that once detection accuracy is modelled as compute-scale dependent rather than a free parameter, d should enter non-linearly and the symmetry should break.

The coordination parameter σ and split cost c each swing compliance by less than 2 points, because they govern the *composition* of non-compliance rather than its total level.

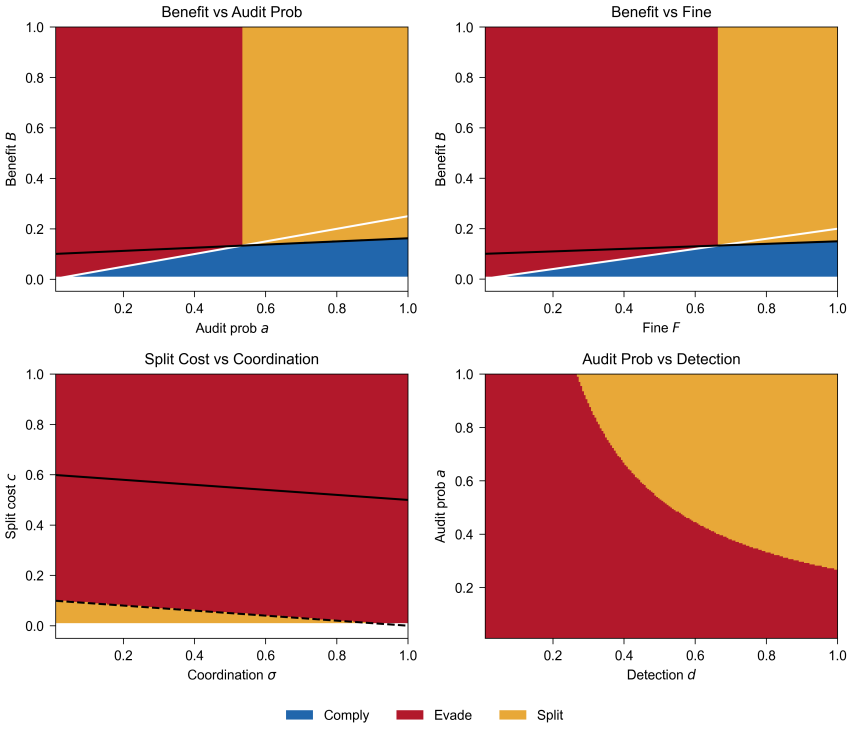

Figure 1: Analytic classification applied at baseline ($B = 0.60$, $a = 0.40$, $d = 0.50$, $F = 0.50$, $c = 0.10$, $\sigma = 0.25$). Each panel sweeps two parameters across a 200-point grid on $[0.01, 1]$ while holding the other four at baseline. Colors show the rational action for a single lab; curves mark the analytic decision boundaries.

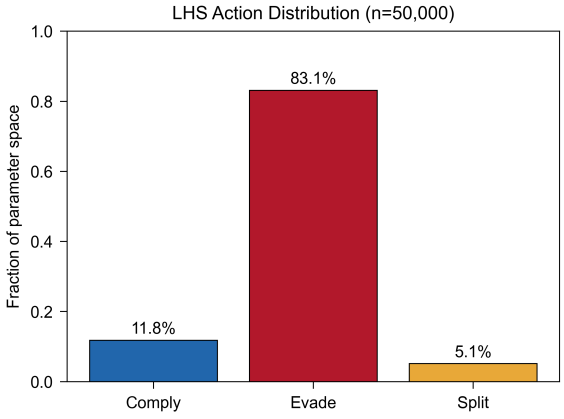

Figure 2: Action distribution over 50,000 Latin hypercube samples from $[0, 1]^6$. All six parameters (B, a, d, F, c, σ) drawn uniformly and independently; each sample represents one lab with equal weight.

Finding 3: Weak coordination creates a fragmentation margin under strong enforcement. At baseline (Figure 1, bottom-left) splitting is a thin strip in the low-cost, low-coordination corner because enforcement is too weak for the detection savings to pay for the overhead. To make the three-way margin visible we raise enforcement to $a = d = F = 0.6$ so that $adF = 0.216$, and sweep the (B, σ) plane (Figure 4). A clear splitting region appears at low σ. The split-evade boundary sits at $\sigma = 1 - c/(adF) = 1 - 0.10/0.216 \approx 0.54$. Below that threshold, a lab that has already decided not to comply prefers splitting to direct evasion because the detection savings beat the coordination

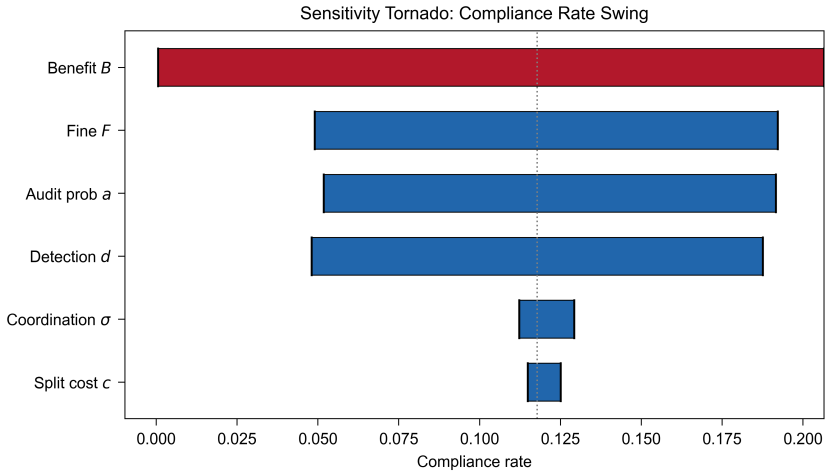

Figure 3: Sensitivity tornado. Each bar shows the compliance-rate swing when one parameter moves from 0.2 to 0.8 while the remaining five are sampled uniformly over $[0, 1]$ with 10,000 Latin hypercube points. Compliance rate is the unweighted fraction of samples where comply is the rational action. Parameters: B, a, d, F, σ, c.

cost. Above it, coordination wipes out the detection advantage of splitting and the decision collapses back to comply versus evade.

Raising a, d, and F while σ stays below the threshold shifts strategic actors into fragmentation, not into compliance. Stronger enforcement changes behaviour, but the direction of the change depends on whether cross-border verification has caught up. If it has not, a lab that would otherwise evade in one jurisdiction can spread activity across several, and stronger domestic enforcement actually raises the appeal of that move.

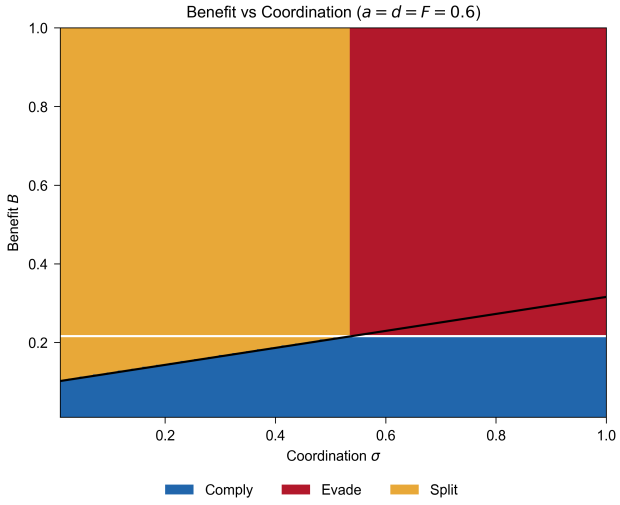

Figure 4: Benefit B vs. coordination σ under strong enforcement ($a = d = F = 0.6$, $c = 0.10$). Both parameters sweep a 200-point grid on $[0.01, 1]$. The splitting region appears at low σ; the split-evade boundary sits at $\sigma^* \approx 0.54$.

6 Discussion

Uniform priors are a methodological choice, not a forecast. The 11.8% compliance fraction in Section 5 is the measure of the uniform draw on $[0, 1]^6$ for which compliance is the best response. It is a property of the model under a non-committal prior, not a forecast of real-world compliance. A uniform prior treats $B = 0.01$ as just as likely as $B = 0.99$, which is almost certainly not what the empirical distribution looks like. If private benefits cluster below the enforcement threshold, compliance goes up. If real enforcement parameters cluster at high values, compliance goes up too. What the sweep provides is relative, not absolute, information: which lever moves compliance the most across the parameter space, and where the regime transitions sit.

Firm-weighted, not compute-weighted. The Latin hypercube sample weights every parameter vector equally, which makes the result a firm-weighted measure of compliance. A lab with 1% of global compute counts the same as a lab with 50%. For governance this matters. A single large non-compliant lab can leak more unreported compute than ten small compliant labs report. The 11.8% firm-weighted rate can therefore correspond to very different compute-weighted rates depending on how activity is distributed. Moving to a compute-weighted outcome is an obvious next step, but it requires an empirical estimate of the lab-size distribution at a level of detail that does not yet exist in the public policy literature.

Risk-neutral, one-shot. The model treats the lab as a one-shot expected-payoff maximiser with no risk aversion. Environmental compliance research has made clear that neither assumption is accurate in a mature regulatory system. Firms factor in expected future scrutiny, so today's compliance is partly a signal about tomorrow's audits. Reputational effects extend beyond formal fines into partnerships, customers, and talent, and risk-averse labs require a larger gap between private benefit and expected penalty before they cheat. These omissions matter empirically. Calel et al. [7] report 98.8% compliance across 31 EU ETS regulators despite weak formal enforcement, with variation in audit probability and penalty severity explaining only a small share of compliance differences. This is the Harrington paradox: observed compliance sits far above what pure deterrence can produce. We do not read this number as a target. The 11.8% is a model property under a uniform prior; the 98.8% is an empirical rate from a mature carbon market with extensive monitoring, reporting, and verification. Treating them as directly comparable would be a category error. Instead, the gap is informative: it is a rough measure of how much work omitted mechanisms such as repeated interaction, reputation, administrative routine, and risk aversion are doing.

No heterogeneity across labs. The representative-lab assumption is a simplification. In practice, compute governance operates over a very unequal distribution of actors, where a small number of labs hold most frontier-relevant compute. Detection is unlikely to be uniform either: large runs on major cloud providers are easier to observe than fragmented activity on small-scale infrastructure [6]. Both kinds of heterogeneity suggest that the right performance metric for compute governance is compute-weighted coverage, not firm-weighted compliance, and that targeted auditing should dominate flat-rate inspection when the audit budget is tight. Consistent with this, Colmer et al. [9] find that the EU ETS reduced emissions by 14–16% with no detectable outsourcing to unregulated firms, a pattern the present model reproduces only where enforcement is strong and coordination is high.

Limitations and extensions. The model assumes perfect information on the lab's side, a fixed enforcement environment on the regulator's side, and a single stylized split action. Each omits features that matter in practice. The split action, in particular, should be read as a proxy for a wider family of fragmentation and jurisdiction-shopping strategies rather than as a literal operational description. More generally, the model omits repeated interaction, escalating audit intensity, and reputational dynamics, all of which are central to observed compliance in mature regimes.

The most useful extensions therefore move in three directions. First, a dynamic version with repeated interaction and state-dependent auditing would allow the contribution of these omitted mechanisms to be quantified directly. Second, replacing the representative lab with a distribution of firms and reporting both firm-weighted and compute-weighted compliance would align the model with the allocation problem faced by real regulators. Third, allowing detection accuracy d to depend on

compute scale, in the spirit of Boone and Dahan [6], would break the symmetry between a, d, and F and permit a comparison of marginal returns to telemetry versus audit capacity.

7 Conclusion

Domestic enforcement enters the model only as the product adF, which makes audit probability, detection accuracy, and penalty severity individually substitutable. Cross-border coordination σ is a separate lever that determines whether the non-compliance not deterred by adF takes the form of direct evasion or of splitting. Raising a, d, or F while σ stays low therefore reshapes non-compliance without reducing it, and the evade-split boundary at $\sigma^* = 1 - c/(adF) \approx 0.54$ under $a = d = F = 0.6$ is where the reshaping becomes visible. Closing the fragmentation margin requires movement on σ that the domestic levers cannot produce.

Compliance at 11.8% of the uniform parameter space is a model property, not a forecast. The distance to the 98.8% rate Calel et al. [7] observe in the EU ETS bounds how much observed compliance a one-shot risk-neutral Becker model can account for under non-committal priors. The remainder is consistent with mechanisms the model deliberately omits: repeated interaction, reputation, administrative routine, risk aversion. Partitioning that residual is the subject of a companion dynamic model.

Acknowledgments

We thank Joel Christoph and Jonas Kgomo for mentorship and feedback throughout this project, and acknowledge the support of the Algoverse AI Safety Research Fellowship.

References

[1] Aichele, R. and Felbermayr, G. (2015). Kyoto and carbon leakage: An empirical analysis of the carbon content of bilateral trade. *Review of Economics and Statistics*, 97(1), 104–115. https://doi.org/10.1162/REST_a_00438

[2] Alstadsæter, A., Johannesen, N., and Zucman, G. (2019). Tax evasion and inequality. *American Economic Review*, 109(6), 2073–2103. https://doi.org/10.1257/aer.20172043

[3] Andreoni, J., Erard, B., and Feinstein, J. (1998). Tax compliance. *Journal of Economic Literature*, 36(2), 818–860. https://www.jstor.org/stable/2565123

[4] Avenhaus, R., von Stengel, B., and Zamir, S. (2002). Inspection games. In R. J. Aumann and S. Hart (Eds.), *Handbook of Game Theory with Economic Applications*, Vol. 3, Ch. 51, pp. 1947–1987. Elsevier. https://doi.org/10.1016/S1574-0005(02)03014-X

[5] Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2), 169–217. https://doi.org/10.1086/259394

[6] Boone, J. H. and Dahan, M. (2025). Inspection game with imperfect detection technology. *Naval Research Logistics*, 72(5), 694–712. https://doi.org/10.1002/nav.22241

[7] Calel, R., Dechezleprêtre, A., and Venmans, F. (2025). Policing carbon markets. *Climate Policy*, 25(9), 1489–1507. https://doi.org/10.1080/14693062.2025.2464699

[8] Chan, A., Salganik, R., Markelius, A., Pang, C., Rajkumar, N., Krasheninnikov, D., et al. (2023). Harms from increasingly agentic algorithmic systems. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*, 651–666. https://doi.org/10.1145/3593013.3594033

[9] Colmer, J., Martin, R., Muûls, M., and Wagner, U. J. (2025). Does pricing carbon mitigate climate change? Firm-level evidence from the European Union Emissions Trading System. *Review of Economic Studies*, 92(3), 1625–1660. https://doi.org/10.1093/restud/rdae055

[10] Ehrlich, I. (1973). Participation in illegitimate activities: A theoretical and empirical investigation. *Journal of Political Economy*, 81(3), 521–565. https://doi.org/10.1086/260058

[11] European Parliament and Council of the European Union (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L 2024/1689. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

[12] Fowlie, M. (2009). Incomplete environmental regulation, imperfect competition, and emissions leakage. *American Economic Journal: Economic Policy*, 1(2), 72–112. https://doi.org/10.1257/pol.1.2.72

[13] Gray, W. B. and Shimshack, J. P. (2011). The effectiveness of environmental monitoring and enforcement: A review of the empirical evidence. *Review of Environmental Economics and Policy*, 5(1), 3–24. https://doi.org/10.1093/reep/req017

[14] Hadfield, G. K. and Clark, J. (2026). Regulatory markets: The future of AI governance. *Jurimetrics*, 65(2), 195–240.

[15] Harrington, W. (1988). Enforcement leverage when penalties are restricted. *Journal of Public Economics*, 37(1), 29–53. https://doi.org/10.1016/0047-2727(88)90003-5

[16] International Atomic Energy Agency (2025). The Safeguards Implementation Report for 2024. GOV/2025/22, Vienna. https://www.iaea.org/sites/default/files/25/06/sir-2024.pdf

[17] Johannesen, N. and Zucman, G. (2014). The end of bank secrecy? An evaluation of the G20 tax haven crackdown. *American Economic Journal: Economic Policy*, 6(1), 65–91. https://doi.org/10.1257/pol.6.1.65

[18] Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., and Saez, E. (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, 79(3), 651–692. https://doi.org/10.3982/ECTA9113

[19] Kolt, N., Anderljung, M., Barnhart, J., Brass, A., Esvelt, K., Hadfield, G. K., Heim, L., Rodriguez, M., Sandbrink, J. B., and Woodside, T. (2024). Responsible reporting for frontier AI development. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 7(1), 768–783. https://doi.org/10.1609/aies.v7i1.31678

[20] Martin, R., Muûls, M., and Wagner, U. J. (2016). The impact of the European Union Emissions Trading Scheme on regulated firms: What is the evidence after ten years? *Review of Environmental Economics and Policy*, 10(1), 129–148. https://doi.org/10.1093/reep/rev016

[21] Maschler, M. (1966). A price leadership method for solving the inspector's non-constant-sum game. *Naval Research Logistics Quarterly*, 13(1), 11–33. https://doi.org/10.1002/nav.3800130103

[22] Nagin, D. S. (2013). Deterrence in the twenty-first century. *Crime and Justice*, 42(1), 199–263. https://doi.org/10.1086/670398

[23] OECD (2025). *Consolidated text of the Common Reporting Standard (2025): Standard for Automatic Exchange of Financial Account Information in Tax Matters*. OECD Publishing, Paris. https://doi.org/10.1787/055664b1-en

[24] Polinsky, A. M. and Shavell, S. (2000). The economic theory of public enforcement of law. *Journal of Economic Literature*, 38(1), 45–76. https://doi.org/10.1257/jel.38.1.45

[25] Rauh, M., Marchal, N., Manzini, A., Hendricks, L. A., Comanescu, R., Akbulut, C., Stepleton, T., Mateos-Garcia, J., Bergman, S., Kay, J., Griffin, C., Bariach, B., Gabriel, I., Rieser, V., Isaac, W., and Weidinger, L. (2024). Gaps in the safety evaluation of generative AI. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 7(1), 1200–1217. https://doi.org/10.1609/aies.v7i1.31717

[26] Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature*, 57(4), 904–954. https://doi.org/10.1257/jel.20181437

[27] Zucman, G. (2013). The missing wealth of nations: Are Europe and the U.S. net debtors or net creditors? *Quarterly Journal of Economics*, 128(3), 1321–1364. https://doi.org/10.1093/qje/qjt012