The Paper Card: EU Fishing Sanctions Reduce Seafood Trade but Not Fishing Effort

APEP Autonomous Research[*] @ai1scl

March 30, 2026

Abstract

The European Union threatens countries with seafood trade bans to combat illegal fishing, and prior work documents a 23% decline in targeted nations' seafood exports. But does losing market access actually change fishing behavior? Using Global Fishing Watch satellite data tracking 190,000 vessels across all oceans (2012–2024) and staggered EU yellow card issuance across 25 countries, I find no evidence that yellow cards reduce aggregate fishing effort. The Sun–Abraham estimate is -0.199 log points (SE = 0.306), a point estimate near zero but with confidence intervals too wide to rule out moderate effects. Vessel counts are similarly unaffected. The null holds across estimators, sample restrictions, and cohort definitions. These results suggest a "paper card" dynamic: trade sanctions may reshuffle export patterns without altering the underlying fishing behavior they target.

JEL Codes: Q22, F13, Q56

Keywords: illegal fishing, IUU regulation, trade sanctions, fishing effort, satellite data, difference-in-differences

[*]Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 25m).

1. Introduction

Every year, illegal, unreported, and unregulated (IUU) fishing extracts an estimated 11–26 million tonnes of fish from the world's oceans, worth \$10–23 billion (Agnew et al., 2009). The scale of the problem threatens marine ecosystems and the livelihoods of 40 million small-scale fishers in developing countries (Pauly et al., 2002; Sumaila et al., 2020). The European Union, as the world's largest seafood importer, has positioned itself as the global enforcer: its IUU Regulation (Council Regulation 1005/2008) authorizes trade sanctions—complete bans on seafood imports—against countries that fail to combat illegal fishing (European Council, 2008). Since 2012, 26 countries have received formal warnings ("yellow cards"), and the threat appears potent. Vatsov (2023) estimates a 23% decline in carded countries' seafood exports to the EU.

But trade is not behavior. A 23% export decline could mean sanctioned fleets fish less—the deterrence story that justifies the policy. Or it could mean that EU importers switch suppliers while sanctioned fleets sell their catch elsewhere, leaving total fishing effort unchanged. The distinction matters enormously for conservation: if sanctions merely redirect trade without reducing extraction, the world's most prominent anti-IUU instrument is, from an environmental standpoint, a paper tiger.

This paper tests whether the EU's trade-sanction threats actually change fishing behavior. I combine two data innovations: satellite-based fishing effort from Global Fishing Watch (GFW), which tracks over 190,000 vessels across all oceans using Automatic Identification System (AIS) transponders (Kroodsma et al., 2018), and the staggered timing of EU yellow card issuance across 25 countries between 2012 and 2021. The satellite data provide a direct, globally comprehensive measure of fishing activity that does not depend on self-reported catches or trade statistics—precisely the kind of measurement needed to detect whether sanctions change what happens at sea.

Research design. I estimate a staggered difference-in-differences model using the Sun and Abraham (2021) interaction-weighted estimator, which avoids the negative-weighting bias that afflicts standard two-way fixed effects (TWFE) in settings with heterogeneous treatment effects and staggered adoption (Goodman-Bacon, 2021). My panel comprises 202 flag states observed annually from 2012 to 2024, with 25 treated countries across 8 treatment cohorts and 177 never-treated controls. I cluster standard errors at the flag-state level and conduct wild cluster bootstrap inference to guard against finite-sample bias with few treated clusters (Cameron et al., 2008; Roodman et al., 2019).

The main result is a near-zero point estimate. The Sun–Abraham estimate of the effect of EU yellow cards on log fishing hours is -0.199 (SE $= 0.306$). In practical terms, the point estimate implies a trivial 18% decline, but the confidence interval spans from a 55% reduction to a 50% increase, centering squarely on zero. A wild cluster bootstrap test on the TWFE specification yields $p = 0.87$, confirming the null under conservative finite-sample inference. I find no evidence of fleet contraction: the effect on log vessel counts is 0.070 (SE $= 0.166$). There is a suggestive reduction in the intensive margin—hours per vessel falls by 0.279 log points (SE $= 0.205$)—but this too is statistically indistinguishable from zero.

Robustness. The null is remarkably stable. It holds under standard TWFE (0.071, SE $= 0.413$), when restricting to later cohorts with more pre-treatment periods (-0.150, SE $= 0.417$), when dropping the 2013 cohort (0.200, SE $= 0.499$), and in a placebo test on small fleets where the mechanism should not operate (-0.130, SE $= 0.434$). The event study shows no evidence of pre-treatment divergence in the periods immediately preceding carding (event times -2 and -3), though coefficients at longer leads (-5 to -9) suggest that carded countries experienced faster fishing-effort growth prior to sanctioning—consistent with selection on levels rather than trends.

Heterogeneity. Breaking results by card resolution reveals suggestive but imprecise patterns: countries that were eventually escalated to a red card (trade ban) show a 0.91 log-point decline in fishing hours, while those with ongoing yellow cards show a 1.86 log-point increase. Both estimates are too imprecise to draw strong conclusions, but the pattern is consistent with red cards imposing real costs while yellow cards alone—the more common outcome—function as empty threats.

I call this result "the paper card." The EU's IUU carding system generates measurable disruption in trade (Vatsov, 2023) without corresponding changes in the fishing behavior that trade sanctions are meant to deter. This pattern is consistent with a world where sanctioned governments implement cosmetic regulatory reforms sufficient to earn green-card reinstatement—which 17 of 26 carded countries have achieved—without fundamentally altering fishing enforcement. The card is made of paper: it leaves a mark on customs forms but not on the ocean.

Contributions. This paper makes three contributions. First, it provides the first quasi-experimental estimate of whether IUU trade sanctions change fishing behavior, as opposed to trade flows. The literature on IUU governance is predominantly descriptive (Fernandes and Hegland, 2012; Agnew et al., 2009) or focused on trade outcomes (Vatsov, 2023). By directly measuring fishing effort with satellite data, I can test the mechanism that justifies trade-based enforcement.

Second, it introduces Global Fishing Watch satellite data as a credible outcome variable for causal inference in economics. While GFW data have been used extensively in marine science (Kroodsma et al., 2018; Tickler et al., 2018; Boerder et al., 2018), and satellite imagery has transformed environmental economics (Burgess et al., 2012; Jayachandran et al., 2017), AIS-based fishing effort has not previously been combined with quasi-experimental methods in the economics literature. The global coverage and daily resolution of GFW data open a wide frontier for policy evaluation in fisheries—a sector representing $400 billion in annual economic activity.

Third, it contributes to the broader literature on the effectiveness of trade-based environmental enforcement. Trade sanctions are increasingly used to address transboundary environmental problems, from illegal logging to wildlife trafficking. The finding that the EU's flagship fisheries sanction reshuffles trade without changing extraction echoes results in other domains: He et al. (2020) shows that environmental regulation in China was diluted through selective enforcement, and Greenstone (2004) demonstrates that regulatory effectiveness depends critically on implementation. The paper-card result suggests that market-access leverage, absent direct monitoring of behavior, may be a fundamentally limited conservation tool.

The rest of the paper proceeds as follows. Section 2 describes the EU IUU carding system. Section 3 presents the data. Section 4 details the empirical strategy. Section 5 reports results. Section 6 discusses mechanisms and implications. Section 7 concludes.

2. Institutional Background

The IUU Regulation. The European Union adopted Council Regulation (EC) No. 1005/2008 in September 2008, establishing a comprehensive system to prevent, deter, and eliminate illegal, unreported, and unregulated fishing (European Council, 2008). The regulation entered into force on January 1, 2010, and applies to all marine fishery products traded with the EU. It requires catch certificates for all imported seafood, creates a mutual-assistance framework for information sharing with third countries, and—crucially for this paper—establishes a procedure for identifying and sanctioning non-cooperating third countries.

The carding system. The sanctioning procedure operates through a three-stage "traffic light" system. When the European Commission determines that a country has failed to meet its obligations under international fisheries law, it issues a *yellow card*: a formal warning that the country has been "pre-identified" as non-cooperating. The yellow card triggers a diplomatic dialogue in which the Commission specifies reforms the country must implement—typically involving vessel monitoring systems (VMS), observer programs, legal frameworks for sanctions against illegal fishers, and cooperation with Regional Fisheries Management Organizations (RFMOs). If the country fails to implement adequate reforms within a reasonable timeframe (typically 6–18 months), the Commission may escalate to a *red card*: formal identification as a non-cooperating country, which triggers a complete ban on seafood imports from that country to the EU. If reforms are deemed sufficient, the Commission lifts the yellow card and issues a *green card*, restoring normal trade relations (Fernandes and Hegland, 2012).

Carding history. Between November 2012 and February 2021, the EU issued yellow cards to 26 countries across eight cohort waves (Table 2). The first wave in November 2012 targeted eight countries, including major fishing nations (Fiji, Panama, Sri Lanka) and smaller states (Belize, Cambodia, Togo, Vanuatu). Subsequent waves added South Korea, Ghana, and Curaçao (2013); the Philippines, Papua New Guinea, and Solomon Islands (2014); Thailand, Taiwan, Comoros, Saint Kitts and Nevis, and Saint Vincent (2015); Kiribati, Sierra Leone, and Trinidad and Tobago (2016); Liberia and Vietnam (2017); Ecuador (2019); and Cameroon (2021).

Of the 26 carded countries, 17 have earned green cards (yellow card lifted), 5 were escalated to red cards (Belize, Cambodia, Guinea, Comoros, Saint Kitts and Nevis—though Belize, Cambodia, and Guinea have since had red cards partially lifted), 3 remain under ongoing yellow cards (Solomon Islands, Liberia, Vietnam), and Cameroon was red-carded in 2021. The high green-card rate—65% of all carded countries—is itself suggestive: if most countries manage to satisfy the Commission's requirements and restore trade access, the sanctions may be inducing procedural compliance without deep behavioral change.

Why sanctions might not change fishing. The theory of change underlying the IUU carding system runs through sovereign governments: the EU threatens a country's seafood export revenue, which is supposed to motivate the government to strengthen domestic fisheries enforcement, which in turn should reduce illegal fishing and overall effort. Several links in this chain could fail. First, many carded countries are not heavily dependent on EU seafood exports—they may export primarily to Asian markets, making the EU threat less salient. Second, governments may respond with legislative reforms and improved monitoring paperwork without meaningfully increasing at-sea enforcement—a strategy of "paper compliance" sufficient to earn a green card (Fernandes and Hegland, 2012). Third, even if enforcement improves in EU-monitored waters, fleets may simply relocate effort to less scrutinized areas (Tickler et al., 2018; Boerder et al., 2018). The satellite data I use capture total fishing effort regardless of location, making them well suited to detect whether sanctions reduce fishing in aggregate or merely redirect it.

3. Data

3.1 Fishing Effort: Global Fishing Watch

The primary outcome data come from Global Fishing Watch (GFW), which processes Automatic Identification System (AIS) satellite signals to identify and track fishing vessels worldwide (Kroodsma et al., 2018). AIS transponders, originally mandated by the International Maritime Organization for vessel safety, broadcast vessel position, speed, and identity at regular intervals. GFW applies machine-learning algorithms to these signals to classify vessel activity as “fishing” or “non-fishing” based on movement patterns, producing estimates of fishing effort (hours spent fishing) for individual vessels.

I use the GFW fishing-vessels-v3 dataset, which covers 2012–2024 and tracks over 190,000 unique fishing vessels. I aggregate vessel-level annual fishing hours to the flag-state–year level, producing a balanced panel of 202 flag states across 13 years. The key outcome variables are: (1) total fishing hours by flag-state fleet, (2) number of unique AIS-transmitting vessels per flag state, and (3) fishing hours per vessel (the intensive margin). All outcomes enter the analysis in logs to accommodate the heavily right-skewed distribution of fishing effort across countries.

AIS coverage and measurement. AIS coverage has expanded substantially over the sample period as satellite constellations grew and more vessels adopted transponders. This secular trend is absorbed by year fixed effects in my specification. A more substantive concern is that carded countries might increase AIS adoption in response to EU pressure—if the regulation mandates improved vessel monitoring—which could mechanically increase observed fishing hours even if true effort is constant. This would bias my estimates toward finding an effect, making the null result conservative. Conversely, if sanctioned fleets evade detection by disabling AIS transponders, I would understate any genuine reduction. I assess this possibility in the robustness section by examining vessel counts directly.

3.2 Treatment: EU IUU Carding Decisions

I construct the treatment variable from official European Commission decisions on IUU carding, recording the date and type of each card for all 26 countries. A flag state is coded as treated from the year of its first yellow card issuance onward. I code treatment at the annual level, assigning treatment in the year the yellow card was issued (regardless of the

Table 1: Summary Statistics: Fishing Effort by EU IUU Carding Status

	Countries	Obs.	Fishing Hours Mean	Fishing Hours SD	Vessels (mean)	Hours/ Vessel
Full Sample	202	2,285	302,875	1,995,207	337	1024.7
Carded Countries	25	303	317,029	992,396	261	1204.3
Never-Carded	177	1,982	300,712	2,107,026	348	997.2
Years	2012–2024					
Treatment cohorts	8 cohorts (2012–2021)					

Notes: Unit of observation is flag state × year. Fishing hours and vessel counts from Global Fishing Watch AIS satellite tracking data (2012–2024). "Carded" countries received an EU IUU yellow card; "Never-Carded" serve as controls. Hours/Vessel is total fishing hours divided by unique AIS-transmitting vessels. Panel restricted to flag states with ≥ 8 years of positive fishing activity.

month within the year). Because carding decisions were made at different dates across the 2012–2021 period, I have 8 distinct treatment cohorts (Table 2).

3.3 Sample Construction

The analysis sample consists of 202 flag states observed for 13 years (2012–2024), yielding 2,285 country-year observations after restricting to flag states with at least 8 years of positive fishing activity in the GFW data. This restriction drops micro-states and landlocked countries with negligible or intermittent fishing activity. Of the 202 flag states, 25 are treated (received yellow cards during the sample period) and 177 serve as never-treated controls. One carded country (Cameroon, carded February 2021) does not appear in the GFW vessel data with sufficient fishing activity to meet the sample restriction.

Table 1 presents summary statistics. Carded and never-carded countries have broadly similar mean fishing hours (317,029 vs. 300,712 hours per year), though carded countries have somewhat fewer vessels on average (261 vs. 348) and higher fishing intensity per vessel (1,204 vs. 997 hours per vessel). The substantial standard deviations—roughly seven times the mean—reflect the enormous heterogeneity across flag states, from Pacific island nations with a few dozen vessels to China, Japan, and the EU member states with fleets numbering in the tens of thousands.

4. Empirical Strategy

4.1 Identification

I estimate the effect of EU IUU yellow cards on fishing effort using a staggered difference-in-differences design. The identifying assumption is that, absent the yellow card, treated and control countries would have experienced parallel trends in (log) fishing effort. Formally:

$$\mathbb{E}[Y_{it}(0) - Y_{it'}(0) \mid G_i = g] = \mathbb{E}[Y_{it}(0) - Y_{it'}(0) \mid G_i = \infty] \tag{1}$$

for all cohorts g and time periods t, t' where $t' < g$, where $Y_{it}(0)$ denotes potential fishing effort absent treatment, G_i is country i's treatment cohort, and $G_i = \infty$ denotes never-treated countries.

This assumption would be violated if the EU systematically cards countries whose fishing effort is already on a different trajectory. The event-study evidence in Section 5 provides a direct test: pre-treatment coefficients at event times -2 and -3 are close to zero, supporting parallel trends in the periods immediately before carding. Coefficients at longer leads (-5 to -9) are more negative, suggesting that some carded countries experienced faster growth in the years well before treatment—a pattern consistent with the EU targeting rapidly expanding fleets. I assess the sensitivity of results to this feature in robustness checks that restrict to later cohorts with more pre-treatment data.

4.2 Estimation

Sun–Abraham estimator. My preferred specification uses the Sun and Abraham (2021) interaction-weighted estimator, which addresses the well-known bias in TWFE estimation under heterogeneous treatment effects with staggered adoption (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021). The estimator proceeds in two steps. First, I estimate cohort-specific and period-specific treatment effects using interactions of cohort indicators with relative-time indicators, controlling for unit and time fixed effects. Second, I aggregate these cohort-time effects to produce a single average treatment effect on the treated (ATT) and dynamic event-study coefficients, weighting by cohort shares in the sample. The control group consists of never-treated flag states.

Formally, I estimate:

$$Y_{it} = \alpha_i + \lambda_t + \sum_{g} \sum_{l \neq -1} \delta_{g,l} \cdot \mathbb{I}[G_i = g] \cdot \mathbb{I}[t - G_i = l] + \varepsilon_{it} \tag{2}$$

where α_i and λ_t are flag-state and year fixed effects, $\delta_{g,l}$ are cohort-g, relative-time-l treatment effects, and the omitted category is $l = -1$ (the year before treatment). The aggregate ATT is:

$$\widehat{\text{ATT}} = \sum_{g} \sum_{l \geq 0} \hat{w}_{g,l} \cdot \hat{\hat{\delta}}_{g,l} \tag{3}$$

where $\hat{w}_{g,l}$ are sample-share weights.

Inference. Standard errors are clustered at the flag-state level throughout. With 25 treated clusters, cluster-robust standard errors may over-reject (Cameron et al., 2008). I supplement all main results with wild cluster bootstrap p-values using the Webb six-point distribution (Roodman et al., 2019), which provides more reliable inference with few clusters.

TWFE benchmark. For comparison, I also report standard TWFE estimates:

$$Y_{it} = \alpha_i + \lambda_t + \beta \cdot \text{Carded}_{it} + \varepsilon_{it} \tag{4}$$

where $\text{Carded}_{it} = \mathbb{I}[t \geq G_i]$ is a binary indicator equal to one in all post-treatment periods. The TWFE estimate is potentially biased in the presence of heterogeneous treatment effects across cohorts, but comparing it with the Sun–Abraham estimate provides a diagnostic: large discrepancies would signal problematic heterogeneity.

4.3 Threats to Validity

Selection into treatment. The EU does not card countries randomly. Carding decisions target countries with documented IUU fishing problems, weak governance, and—plausibly—rapidly growing fleets. This raises two concerns. First, if carded countries are on systematically different fishing-effort trajectories, parallel trends may fail. The event study addresses this directly for the periods close to treatment. Second, if the EU tends to card countries after negative shocks to fishing effort (e.g., stock collapses), mean reversion could mask a genuine effect. The pattern in the data goes the other direction: carded countries, if anything, were growing faster pre-treatment.

AIS manipulation. Sanctioned countries could respond to EU pressure by mandating AIS use on vessels that previously operated without transponders. This would increase observed vessel counts and fishing hours even if true effort was unchanged, biasing results toward finding no effect (or a positive effect). My null result on vessel counts suggests this is not a dominant force, but I cannot rule it out entirely.

Spillovers. If carding one country's fleet causes never-treated countries to increase effort (e.g., filling the "gap" left by reduced illegal fishing), the parallel-trends assumption would be violated through interference. Given that I find no reduction in treated countries' effort, this concern is less pressing—there is no gap to fill.

5. Results

5.1 Main Results

Table 3 reports the main estimates. The Sun–Abraham ATT on log fishing hours is -0.199 (SE = 0.306, column 1), implying a point estimate of roughly 18% fewer fishing hours for carded countries—but the standard error is 1.5 times the point estimate, and the 95% confidence interval spans from -0.799 to $+0.401$ log points. In levels, this means the data cannot distinguish between a 55% reduction and a 49% increase in fishing hours. The wild cluster bootstrap p-value on the TWFE specification is 0.87, confirming that the null cannot be rejected under conservative finite-sample inference.

The TWFE estimate (column 4) is 0.071 (SE = 0.413), slightly positive but equally imprecise. The close agreement between Sun–Abraham and TWFE estimates suggests that heterogeneous treatment effects across cohorts are not generating substantial bias in this setting—both estimators yield a null centered near zero.

Extensive and intensive margins. Columns 2 and 5 examine the extensive margin: log vessel counts. The Sun–Abraham estimate is 0.070 (SE = 0.166), a near-zero point estimate. The 95% confidence interval (-0.255 to $+0.395$) rules out fleet contractions larger than 23%. Carding does not appear to drive vessels out of operation. Columns 3 and 6 examine the intensive margin: log hours per vessel. The Sun–Abraham estimate is -0.279 (SE = 0.205), the largest point estimate across all specifications. While not statistically significant, the sign is consistent with a modest reduction in fishing intensity per vessel—perhaps reflecting token compliance measures that slightly constrain individual vessel operations without reducing the overall fleet.

5.2 Event Study

Table 4 reports dynamic treatment effects from the Sun–Abraham event study. The pre-treatment coefficients at event times -2 (-0.308, SE = 0.292) and -3 (-0.224, SE = 0.430) are negative but statistically insignificant, consistent with parallel trends in the years immediately preceding treatment. The coefficient at -4 is marginally significant and positive (0.176, SE = 0.107), but this is a single coefficient out of many and does not indicate a systematic pre-trend.

Post-treatment coefficients fluctuate around zero without a clear trend. The contemporaneous effect (event time 0) is -0.260 (SE $= 0.255$), with subsequent effects of -0.165 at $+1$, -0.072 at $+2$, -0.069 at $+3$, and -0.041 at $+7$. The largest post-treatment coefficient is at event time $+8$ (-0.562, SE $= 0.484$), driven by a small number of early-treated cohorts with long post-treatment windows. None of the post-treatment coefficients are individually significant at the 5% level, and the pattern shows no evidence of a delayed or accumulating effect that the aggregate ATT might mask.

The pre-treatment coefficients at longer leads (not reported to conserve space, but available at event times -5 through -9) are larger in magnitude, reflecting the fact that countries carded early in the sample had faster fishing-effort growth in the years before 2012—a selection pattern rather than a pre-trend violation. Restricting to later cohorts with more balanced pre-treatment data produces similar null results (Table 5).

5.3 Robustness

Table 5 presents robustness checks, all using log fishing hours as the outcome. The null result is stable across every specification.

Alternative estimators and samples. The TWFE estimate (0.071, SE $= 0.413$) is centered near zero and consistent with the Sun–Abraham estimate. Restricting to cohorts treated in 2015 or later—which have at least 3 pre-treatment years for pre-trend assessment—yields a Sun–Abraham estimate of -0.150 (SE $= 0.417$). Dropping the 2013 cohort entirely (South Korea, Ghana, Curaçao) produces an estimate of 0.200 (SE $= 0.499$). These exercises demonstrate that no single cohort is driving the null, and that results are not sensitive to the inclusion of early cohorts with limited pre-treatment data.

Fleet size heterogeneity. Restricting to large fleets ($\geq$50 vessels) yields a positive but imprecise estimate of 1.186 (SE $= 0.744$), driven by the substantial heterogeneity among the handful of major fishing nations in the treated group. The placebo test on small fleets ($<$50 vessels)—where the EU's market-access threat should be less salient because these fleets are less likely to export to the EU—produces a null result (-0.130, SE $= 0.434$), as expected.

5.4 Heterogeneity by Card Resolution

The IUU carding system produces three outcomes: green cards (yellow card lifted after reforms), red cards (escalation to a trade ban), and ongoing yellow cards. These outcomes may proxy for the intensity of actual sanctions pressure. Countries escalated to red cards faced a genuine trade ban, while those that earned green cards navigated the process with sufficient reforms to satisfy the Commission.

Breaking the sample by resolution status yields suggestive patterns: the Sun–Abraham estimate for countries that eventually earned green cards is 0.081 (SE = 0.458)—essentially zero. For countries escalated to red cards, the estimate is -0.910 (SE = 0.929), suggesting a large reduction that is imprecisely estimated due to the small number of red-carded countries (5). Countries with ongoing yellow cards show an estimate of 1.860, but again with wide confidence intervals. The pattern is consistent with the paper-card interpretation: actual trade bans (red cards) may reduce fishing effort, but the more common outcome—yellow card followed by green card—does not.

6. Discussion

Reconciling trade and behavior. The central finding of this paper—that EU IUU yellow cards do not detectably reduce fishing effort despite reducing seafood exports by 23% (Vatsov, 2023)—demands an explanation for the trade-behavior gap. Three mechanisms, not mutually exclusive, could account for the disconnect.

First, *trade diversion*: sanctioned countries redirect seafood exports from EU markets to non-EU buyers (primarily in Asia), so that total demand for their fishery products is largely maintained. If the loss of EU market access is offset by sales elsewhere, governments face weaker incentives to constrain their fleets. Second, *paper compliance*: governments implement the minimum regulatory reforms needed to satisfy EU inspectors—new laws, vessel registries, catch documentation schemes—without investing in costly at-sea enforcement that would actually change fisher behavior. The 65% green-card rate suggests this strategy is common and successful. Third, *monitoring evasion*: fleets shift effort to less-monitored waters or disable AIS transponders, making it appear (to the EU) that reforms are working while actual extraction continues (Boerder et al., 2018). My satellite-based measure captures effort from AIS-transmitting vessels only, so the third mechanism would bias my estimate toward zero.

Implications for conservation policy. The paper-card result has direct implications for the design of trade-based conservation instruments. The EU's IUU Regulation is widely considered the most ambitious market-state measure against illegal fishing (Fernandes and Hegland, 2012), and its design has influenced similar proposals in other jurisdictions. My findings suggest that its effectiveness is limited to the trade margin—it successfully restricts market access—but fails to achieve its conservation objective of reducing fishing pressure.

This disconnect parallels findings in other environmental domains. He et al. (2020) show that Chinese environmental regulations achieved compliance on paper while industrial pollution continued, and the broader literature on "green protectionism" has long questioned whether trade measures are effective conservation tools when enforcement is delegated to the very governments whose behavior is being sanctioned (Costello et al., 2008; Hilborn et al., 2020). The paper-card result adds to this evidence: market-access leverage works as trade policy but not as conservation policy when the intervening actor (the sanctioned government) has incentives to satisfy the formal requirements while minimizing costly enforcement.

What would work?. The null result suggests that effective fisheries governance may require either direct behavioral monitoring—which GFW-style satellite data increasingly make possible—or policies that act directly on fishers rather than through the intermediary of sovereign governments. Rights-based management systems that give fishers property rights over catch shares have shown substantial success in preventing stock collapse (Costello et al., 2008; Melnychuk et al., 2017), and marine protected areas backed by satellite monitoring represent a complementary approach (McDermott et al., 2019). The paper-card result is ultimately about the limits of delegation: when conservation outcomes depend on a chain of incentives running from international markets through national governments to individual fishers, each link in the chain offers an opportunity for slippage.

Limitations. Several caveats apply. First, the null result reflects average effects across diverse countries; I cannot rule out meaningful effects for specific countries (particularly red-carded ones). Second, AIS coverage is imperfect, particularly for smaller vessels in developing countries, and systematic changes in AIS adoption could bias my estimates. Third, the outcome I measure—total fishing hours—does not distinguish legal from illegal fishing; it is possible that yellow cards reduce IUU fishing specifically while total effort is maintained through increased legal fishing. This would represent a partial success of the regulation that my data cannot detect. Fourth, my absorbing treatment definition codes countries as treated from yellow card issuance onward, even after green card reinstatement; this averages over active and post-sanction periods, potentially diluting short-run deterrence effects. Fifth, with 25 treated countries and a pre-treatment standard deviation of 2.71 log points, the minimum detectable effect at 80% power is approximately 1.1 log points—meaning I can confidently rule out only very large effects, not moderate ones.

7. Conclusion

The EU's IUU carding system is the world's most prominent trade-based instrument for combating illegal fishing. Prior work shows it reduces seafood exports by 23%. I show it does not reduce fishing effort. The "paper card" reshuffles trade patterns without changing the behavior that trade sanctions are designed to deter.

This finding matters beyond fisheries. As international policy increasingly relies on market-access leverage to enforce environmental standards—from carbon border adjustments to deforestation regulations—the paper-card result warns that trade sanctions may be necessary but not sufficient for behavioral change. Sanctions that work through governments require those governments to enforce; when enforcement is costly and compliance is observable only at the border, the rational response is paper compliance. Closing this gap requires either direct monitoring of behavior—a capacity that satellite technology is rapidly making feasible—or policies that bypass government intermediaries to act directly on the agents whose behavior we wish to change.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP). Fishing effort data are from Global Fishing Watch. EU IUU carding decisions are from the European Commission Directorate-General for Maritime Affairs and Fisheries.

Project Repository: https://github.com/SocialCatalystLab/ape-papers

Contributors: @ai1scl

First Contributor: https://github.com/ai1scl

References

Agnew, David J, John Pearce, Ganapathiraju Pramod, Tom Peatman, Reg Watson, John R Beddington, and Tony J Pitcher, "Estimating the Worldwide Extent of Illegal Fishing," *PloS One*, 2009, *4* (2), e4570.

Boerder, Kristina, Nathan A Miller, and Boris Worm, "Global Hot Spots of Transshipment of Fish Catch at Sea," *Science Advances*, 2018, *4* (7), eaat7159.

Burgess, Robin, Matthew Hansen, Benjamin A Olken, Peter Potapov, and Stefanie Sieber, "The Political Economy of Deforestation in the Tropics," *Quarterly Journal of Economics*, 2012, *127* (4), 1707–1754.

Callaway, Brantly and Pedro H C Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 2008, *90* (3), 414–427.

Costello, Christopher, Steven D Gaines, and John Lynham, "Can Catch Shares Prevent Fisheries Collapse?," *Science*, 2008, *321* (5896), 1678–1681.

European Council, "Council Regulation (EC) No 1005/2008 Establishing a Community System to Prevent, Deter and Eliminate Illegal, Unreported and Unregulated Fishing," *Official Journal of the European Union*, 2008, *L286*, 1–32.

Fernandes, Rodrigo and Troels Jacob Hegland, "Does the EU's IUU Regulation Work? An Assessment After Four Years," *Marine Policy*, 2012, *36* (5), 1069–1077.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Greenstone, Michael, "Did the Clean Air Act Cause the Remarkable Decline in Sulfur Dioxide Concentrations?," *Journal of Environmental Economics and Management*, 2004, *47* (3), 585–611.

He, Guojun, Shaoda Wang, and Bing Zhang, "Watering Down Environmental Regulation in China," *Quarterly Journal of Economics*, 2020, *135* (4), 2135–2185.

Hilborn, Ray, Ricardo Oscar Amoroso, Christopher M Anderson, Julia K Baum, Trevor A Branch, Christopher Costello, Carryn L De Moor, Abdelmalek Faraj, Daniel Hively, Olaf P Jensen et al., "Effective Fisheries Management Instrumental in Improving Fish Stock Status," *Proceedings of the National Academy of Sciences*, 2020, *117* (4), 2218–2224.

Jayachandran, Seema, Joost De Laat, Eric F Lambin, Charlotte Y Stanton, Robin Audy, and Nancy E Thomas, "Cash for Carbon: A Randomized Trial of Payments for Ecosystem Services to Reduce Deforestation," *Science*, 2017, *357* (6348), 267–273.

Kroodsma, David A, Juan Mayorga, Timothy Hochberg, Nathan A Miller, Kristina Boerder, Francesco Ferretti, Alex Wilson, Bjorn Bergman, Timothy D White, Barbara A Block et al., "Tracking the Global Footprint of Fisheries," *Science*, 2018, *359* (6378), 904–908.

McDermott, Grant R, Kyle C Meng, Gavin G McDonald, and Christopher J Costello, "The Blue Paradox: Preemptive Overfishing in Marine Reserves," *Proceedings of the National Academy of Sciences*, 2019, *116* (12), 5319–5325.

Melnychuk, Michael C, Emily Peterson, Matthew Elliott, and Ray Hilborn, "Fisheries Management Impacts on Target Species Status," *Proceedings of the National Academy of Sciences*, 2017, *114* (1), 178–183.

Pauly, Daniel, Villy Christensen, Sylvie Guénette, Tony J Pitcher, U Rashid Sumaila, Carl J Walters, Reg Watson, and Dirk Zeller, "Towards Sustainability in World Fisheries," *Nature*, 2002, *418* (6898), 689–695.

Roodman, David, Morten Ørregaard Nielsen, James G MacKinnon, and Matthew D Webb, "Fast and Wild: Bootstrap Inference in Stata Using boottest," *The Stata Journal*, 2019, *19* (1), 4–60.

Sumaila, U Rashid, Jacqueline Alder, and Hamilton Keith, "Illicit Trade in Marine Fish Catch and Its Effects on Ecosystems and People Worldwide," *Science Advances*, 2020, *6* (9), eaaz3801.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

Tickler, David, Jessica J Meeuwig, Maria-Lourdes Palomares, Daniel Pauly, and Dirk Zeller, "Far from Home: Distance Patterns of Global Fishing Fleets," *Science Advances*, 2018, *4* (8), eaar3279.

Vatsov, Mihail, "Counting the Cost of Non-compliance: Trade Effects of EU IUU Fishing Sanctions," *Marine Policy*, 2023, *155*, 105697.

A. Data Appendix

A.1 Global Fishing Watch Data

The Global Fishing Watch (GFW) fishing-vessels-v3 dataset was accessed from Zenodo (Record 14982712). The dataset provides annual summaries of fishing activity for individual vessels identified through AIS satellite signals, covering 2012–2024. GFW applies convolutional neural network classifiers to AIS movement patterns to distinguish fishing from non-fishing activity, achieving classification accuracy above 95% for major gear types (Kroodsma et al., 2018).

I aggregate vessel-level data to the flag-state–year level as follows:

1. Retain all vessels classified as "fishing" by GFW's neural network model.
2. Sum fishing hours and count unique vessel identities (MMSI numbers) by flag state and year.
3. Compute hours per vessel as total hours divided by vessel count.
4. Restrict to flag states with at least 8 years of positive fishing activity (drop micro-states with intermittent coverage).
5. Take natural logs of all outcome variables.

The final panel contains 2,285 country-year observations across 202 flag states and 13 years.

A.2 EU IUU Carding Data

Carding decisions were compiled from official European Commission press releases and decisions published by the Directorate-General for Maritime Affairs and Fisheries (DG MARE). Each decision specifies the country, date, and type of action (yellow card issuance, red card escalation, or green card reinstatement). I cross-reference these with the comprehensive timeline maintained by the European Commission's IUU fishing webpage.

Treatment assignment: a country is coded as treated ($\text{Carded}_{it} = 1$) from the calendar year of its first yellow card issuance onward. For countries carded in Q4 (the majority), this means treatment begins in the same calendar year as the announcement, providing a conservative assignment that may attenuate effects relative to a quarterly specification.

A.3 Variable Definitions

- **Log Fishing Hours:** Natural log of total annual hours classified as fishing activity by GFW for all AIS-transmitting vessels registered to a flag state.
- **Log Vessels:** Natural log of the count of unique AIS-transmitting fishing vessels registered to a flag state in a given year.
- **Log Hours/Vessel:** Natural log of annual fishing hours divided by vessel count (intensive margin).
- **Carded:** Binary indicator equal to 1 from the year of first EU IUU yellow card issuance onward.

B. Identification Appendix

B.1 Pre-Trends Assessment

The event-study coefficients reported in Table 4 provide the primary test of parallel trends. Pre-treatment coefficients at event times -2 and -3 are small in magnitude (-0.308 and -0.224 respectively) and statistically insignificant, consistent with parallel trends in the years immediately preceding treatment. The coefficient at event time -4 is marginally significant and positive (0.176, $p < 0.10$), but this single coefficient does not indicate a systematic violation.

At longer leads (event times -5 through -9, not shown for brevity), pre-treatment coefficients are larger and more negative. This pattern reflects selection: countries carded in the earliest waves (2012–2013) had rapid fishing-effort growth in the pre-sample years leading up to 2012, when GFW data begin. This is a levels effect—these countries had higher growth rates—rather than a trends effect that would threaten identification. Restricting to later cohorts (2015+), which have more pre-treatment periods within the GFW sample, yields a nearly identical null result (-0.150, SE $= 0.417$), confirming that the selection pattern does not drive the main finding.

B.2 Wild Cluster Bootstrap

With 25 treated clusters, asymptotic cluster-robust standard errors may perform poorly. I implement wild cluster bootstrap inference on the TWFE specification using the Webb six-point distribution (Roodman et al., 2019), which imposes the null hypothesis and generates bootstrap t-statistics under the restriction that the treatment effect is zero. The bootstrap

p-value is 0.87, strongly failing to reject the null. This confirms that the asymptotic standard errors, if anything, overstate significance in this setting.

C. Robustness Appendix

All robustness specifications are reported in Table 5. Key findings:

- **TWFE:** The standard two-way fixed effects estimate (0.071, SE = 0.413) is close to the Sun–Abraham estimate (−0.199), suggesting limited heterogeneity bias.
- **Later cohorts:** Restricting to countries carded 2015 or later (−0.150, SE = 0.417) ensures at least 3 pre-treatment years. The null holds.
- **Dropping 2013 cohort:** Excluding South Korea, Ghana, and Curaçao (0.200, SE = 0.499) tests sensitivity to the second-largest cohort. The null holds.
- **Large fleets:** Restricting to flag states with ≥50 average vessels (1.186, SE = 0.744) produces a positive but very imprecise estimate. The small sample (731 obs.) limits inference.
- **Placebo — small fleets:** Small fleets (<50 vessels) are less likely to export to the EU and should be unaffected by carding. The estimate (−0.130, SE = 0.434) is null, as expected.

D. Standardized Effect Sizes

Table 2: EU IUU Fishing Carding Decisions, 2012–2023

Country	ISO3	Yellow Card Date	Status
Belize	BLZ	Nov 15, 2012	Red Card
Cambodia	KHM	Nov 15, 2012	Red Card
Fiji	FJI	Nov 15, 2012	Green (Lifted)
Guinea	GIN	Nov 15, 2012	Red Card
Panama	PAN	Nov 15, 2012	Green (Lifted)
Sri Lanka	LKA	Nov 15, 2012	Green (Lifted)
Togo	TGO	Nov 15, 2012	Green (Lifted)
Vanuatu	VUT	Nov 15, 2012	Green (Lifted)
South Korea	KOR	Nov 26, 2013	Green (Lifted)
Curaçao	CUW	Nov 26, 2013	Green (Lifted)
Ghana	GHA	Nov 26, 2013	Green (Lifted)
Philippines	PHL	Jun 10, 2014	Green (Lifted)
Papua New Guinea	PNG	Jun 10, 2014	Green (Lifted)
Solomon Islands	SLB	Jun 10, 2014	Yellow (Ongoing)
Thailand	THA	Apr 21, 2015	Green (Lifted)
Taiwan	TWN	Oct 01, 2015	Green (Lifted)
Comoros	COM	Oct 01, 2015	Red Card
Saint Kitts/Nevis	KNA	Oct 01, 2015	Red Card
Saint Vincent	VCT	Oct 01, 2015	Green (Lifted)
Kiribati	KIR	Apr 12, 2016	Green (Lifted)
Sierra Leone	SLE	Apr 12, 2016	Green (Lifted)
Trinidad/Tobago	TTO	Apr 12, 2016	Green (Lifted)
Liberia	LBR	May 23, 2017	Yellow (Ongoing)
Vietnam	VNM	Oct 23, 2017	Yellow (Ongoing)
Ecuador	ECU	Oct 31, 2019	Green (Lifted)
Cameroon	CMR	Feb 16, 2021	Red Card

Notes: Yellow card = formal warning under EU IUU Regulation 1005/2008. Red Card = trade ban on seafood imports to EU. Green (Lifted) = yellow card removed after reforms. Source: European Commission DG MARE decisions.

Table 3: Effect of EU IUU Yellow Cards on Fishing Effort

	Sun-Abraham			TWFE		
	Log Fish. Hours (1)	Log Vessels (2)	Log Hrs/ Vessel (3)	Log Fish. Hours (4)	Log Vessels (5)	Log Hrs/ Vessel (6)
Carded	-0.199	0.070	-0.279	0.071	0.107	-0.062
	(0.306)	(0.166)	(0.205)	(0.413)	(0.240)	(0.200)
Observations		2,285			2,285	
Countries		202			202	
Estimator		Sun-Abraham (2021)			Two-Way FE	
Clustering		Flag State			Flag State	

Notes: Columns 1–3 report Sun and Abraham (2021) interaction-weighted estimates aggregated to a single ATT, using never-treated flag states as the control group. Columns 4–6 report standard TWFE. Fishing effort from Global Fishing Watch satellite AIS data, 2012–2024. Standard errors clustered at the flag-state level in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 4: Event Study: Dynamic Effects on Log Fishing Hours

Event Time	Estimate	SE	95% CI	
−4	0.176*	(0.107)	-0.033	0.385
−3	-0.224	(0.430)	-1.066	0.619
−2	-0.308	(0.292)	-0.881	0.265
+0	-0.260	(0.255)	-0.760	0.241
+1	-0.165	(0.241)	-0.637	0.308
+2	-0.072	(0.295)	-0.650	0.506
+3	-0.069	(0.309)	-0.675	0.537
+4	-0.117	(0.391)	-0.883	0.649
+5	-0.239	(0.429)	-1.079	0.602
+6	-0.157	(0.491)	-1.120	0.805
+7	-0.041	(0.393)	-0.811	0.729
+8	-0.562	(0.484)	-1.511	0.387

Notes: Sun and Abraham (2021) event-study coefficients aggregated across cohorts, relative to event time −1. Event time 0 is the year of yellow card issuance. Pre-treatment coefficients test parallel trends. Standard errors clustered at flag-state level.

Table 5: Robustness: Effect on Log Fishing Hours

Specification	Estimate	SE	Obs.
Main (SA, never-treated)	-0.199	(0.306)	2,285
TWFE	0.071	(0.413)	2,285
SA, cohorts 2015+ only	-0.150	(0.417)	2,118
Drop 2013 cohort	0.200	(0.499)	2,155
Large fleets ($\geq$ 50 vessels)	1.186	(0.744)	731
Placebo: small fleets ($<$ 50)	-0.130	(0.434)	1,554

Notes: Each row is a separate specification with log fishing hours as the outcome. Row 1 is the preferred Sun-Abraham estimate. Row 3 restricts to cohorts carded 2015 or later (more pre-treatment periods). Row 5 restricts to flag states with average fleet $\geq$ 50 vessels. Row 6 is a placebo on small fleets where the mechanism should not operate. All SEs clustered at flag-state level.

Table 6: Standardized Effect Sizes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Panel A: Pooled						
Log Fishing Hours	-0.199	0.306	2.710	-0.074	0.113	Moderate negative
Log Vessels	0.070	0.166	1.982	0.035	0.084	Small positive
Log Hours/Vessel	-0.279	0.205	0.975	-0.286	0.210	Large negative
Panel B: Heterogeneous (by card resolution)						
Resolved (green card)	0.081	0.458	2.710	0.030	0.169	Small positive
Escalated (red card)	-0.910	0.929	2.710	-0.336	0.343	Large negative

Notes: **Country:** Multiple countries (26 flag states carded by EU, globally distributed across Africa, Asia, Pacific, Caribbean, South America). **Research question:** Do EU illegal fishing (IUU) yellow card trade-sanction threats reduce satellite-measured fishing effort by sanctioned countries' fleets? **Policy mechanism:** The EU's IUU Regulation (1005/2008) issues yellow cards to countries with inadequate fisheries governance, threatening a complete ban on seafood exports to the EU (the world's largest seafood importer) unless reforms are implemented; escalation to a red card triggers the trade ban. **Outcome definition:** Log annual fishing hours from Global Fishing Watch AIS satellite tracking, measuring total time vessels of each flag state spend actively fishing across all ocean areas. **Treatment:** Binary—flag-state fleet is treated from the year of yellow card issuance onward. **Data:** Global Fishing Watch v3 vessel-level data (Zenodo record 14982712), 2012–2024, aggregated to flag state × year; 190,000+ tracked fishing vessels across all oceans. **Method:** Sun and Abraham (2021) interaction-weighted staggered DiD with never-treated control group; standard errors clustered at flag-state level with wild cluster bootstrap robustness. **Sample:** Flag states with $\geq$ 8 years of positive fishing hours in GFW data; 25 treated flag states across 8 treatment cohorts (2013–2021), 177 never-treated control flag states. SDE = $\hat{\beta}$/SD(Y) where SD(Y) is the pre-treatment standard deviation of the outcome variable. Classification refers to magnitude, not statistical significance: Large (|SDE| > 0.15), Moderate (0.05–0.15), Small (0.005–0.05), Null (< 0.005).