Sanctions, Satellites, Suspect Ships *

Avi Dutt, Abhiroop Mukherjee, George Panayotov, Debjit Roy, Xudong Wen

September 13, 2024

Abstract

Current sanctions policies require third parties, like banks and insurance companies, to identify and avoid doing business with violators. We first demonstrate – using the context of detecting sanctions-violating oil tankers – that such identification can be prohibitively difficult, even when using AI-based models on large and detailed datasets. Second, we find that making detection easier can change the economics of sanctions enforcement. For example, when Reuters released a list of suspect tankers in August 2023, it had a substantial impact on those involved. Such tankers saw their earnings drop by 16.4% compared to other similar vessels, transferring $620 million annually away from them. Additionally, they began avoiding Russian, Iranian, or Venezuelan waters, became substantially harder to sell in the secondary market, and experienced a 53.7% drop in their likelihood of being chartered by U.S. companies. Finally, we present a simple dynamic model of oil shipping to rationalize these findings, and shed further light on consequences for rogue actors on whom we do not have any data.

Keywords: Sanctions enforcement, Oil transport, Machine learning, Satellite data

*Dutt: Trademo; Mukherjee, Panayotov and Wen: Hong Kong University of Science and Technology; Roy: Indian Institute of Management Ahmedabad (IIMA). Emails: x17avi@iima.ac.in; amukherjee@ust.hk; panayotov@ust.hk; debjit@iima.ac.in; xwenam@connect.ust.hk. We thank Vishesh Vatsal and Dfy-graviti for their help in analyzing satellite imagery, and Niranjana Unnithan and Benas Zurauskas for excellent research assistance. We also thank Sumit Agarwal, Tania Babina, Utpal Bhattacharya, Shan Ge, John Griffin, Minkang Gu, Harrison Hong, Yan Ji, Navin Kumar, Erica Li, Vikram Mathur, Jagmeet Makkar, S. Lakshmi Naaraayanan, Jun Pan, Elias Papaioannou, Krishna Prasad, Robert Ready, Vivek Reshamwala, Ashok Srinivasan, Yongxiang Wang, Constantine Yannelis, Jianfeng Yu, Xiaoyun Yu, Jian Zhang and seminar participants at CKGSB, FMA Asia, Hong Kong University Summer Conference, HKUST, SAIF, The Education University of Hong Kong, and Tsinghua University for various helpful comments and suggestions. We thank the UN Global Platform for access to its AIS database (UN Statistics Division, "Global AIS Data Feed.", 2020).

1 Introduction

Most sanctioning regimes today require third-party economic agents to identify and avoid counterparties who may have violated economic sanctions. Is this identification burden too heavy, and is it restraining sanctions enforcement? In this paper, we first examine how difficult it is to identify violators for a relatively sophisticated third party, like a bank, an insurance company, or a large logistics operator. Then we study whether the economics of sanctions could change if these agents were exogenously provided information identifying suspected violators.

Despite their long history and current prevalence, most experts agree that economic sanctions need to be strengthened through better enforcement (see, e.g., Fisman et al. (2024); Hufbauer et al. (1990); Jones (2015); Pape (1997, 1998)). One way in which such a strengthening has been attempted is by extending penalties beyond direct sanctions-violators and applying them also to other actors, who – knowingly or not – facilitate violations (OFAC, 2024; OFSI, 2022). For instance, vessel owners cannot deny accountability for violations by their charterers or managers. Additionally, sanctions laws are sometimes updated to apply retroactively or their statute of limitations can be extended.

However, these changes in sanctions design expose third parties to substantial financial risks. Regulators can now hold these parties accountable for clients' or counterparties' violations, even ones that occurred years earlier and went undetected during due diligence. For example, when a new owner took over Eagle Shipping in 2014, they found that the former owner had violated Myanmar sanctions in 2011-2012. The new owner had to settle with US authorities, despite not being involved themselves in those violations (US Treasury, 2020). As a result, identifying possible sanctions-violators among their counterparties has become a critical issue for a wide range of economic entities – including banks, logistics firms, insurers, ship owners/managers, brokers, traders, and even government organizations like flag registries and port operators (Deloitte, 2015; KPMG, 2023) – attempting to avoid getting stranded with bad assets or penalties.

Sanctioning authorities help with identification, but this help is limited: While they provide lists of sanctioned entities for third parties to avoid, these lists only contain those already sanctioned. However, the number of entities possibly violating sanctions is often an order of magnitude higher than those currently sanctioned by any regime. For example, a recent US Congressional research report (CRS, 2024) estimates that 1600 tankers have carried sanctioned oil, yet the US sanctions lists have in total only 174 sanctioned oil tankers. This makes such lists incomplete for compliance purposes, because they place the burden on third parties to identify counterparties at risk of sanctions, even if not currently listed.

Sanctioning authorities place this burden of identification on third parties – rather than providing information themselves – because they believe that the latter can have an advantage in terms of noticing unusual business practices. For instance, if a company is asked to send high-tech equipment to a location that's actually a pizza shop, it could suspect foul play (see Huneke (2023) for further examples). Given the wide range of such possible detection methods available to businesses, most sanctioning authorities only provide broad guidance – i.e., define loosely what these parties are expected to do to achieve compliance in terms of identifying risky counterparties. Of course, such an approach also allows for flexibility in enforcement, helps accommodate changing political and economic circumstances, and can deprive violators of the ability to game precisely-specified rules or regulations (Lang, 2023; Moiseienko, 2023; Posner and Sykes, 2012). However, loosely defining compliance rules and not providing more help with identification imposes a cost on businesses.

Many third-party economic agents are finding such identification, and therefore sanctions compliance, highly onerous. For example, Daniel Tadros, chief operating officer of the American Club (one of the largest shipping insurance companies in the US), expressed his industry's concerns succinctly to the New York Times: "It's impossible for us to know on a daily basis exactly what every ship is doing, where it's going, what it's carrying, who its owners are. I would like to think that governments have a lot more capability, manpower, resources to follow that."

These concerns are corroborated by our broader analysis of major US financial institutions' annual (10-K) reports. Figure 1, based on machine-reading these reports using an LLM, shows that over 75% of the 50 largest US financial institutions reported concerns about sanctions in 2023. Moreover, nearly all of them (98%) pointed out that identifying sanctions violators is a major challenge for their business.

In this paper, we quantify the costs and benefits of identifying suspected sanctions violators, focusing on the international oil shipping industry. This focus is motivated by two key reasons: First, oil exports represent a significant portion of revenue for most recently sanctioned regimes like Russia, Iran, Venezuela, and Syria. As a result, the success of sanctions is intimately connected to the ability of these countries to market and transport their oil (Babina et al., 2023; Brown, 2020; Farzanegan, 2011). Second, oil is crucial to the global economy, making up over one-third of global energy, influencing transport, trade, growth, and inflation worldwide (Brancaccio et al., 2020, 2023; Hamilton, 1983; Ready, 2018).

To determine if the challenge of identifying non-compliant oil tankers indeed impairs sanctions compliance, we study two key questions: (1) Is it indeed difficult for a third party using public data to detect sanctions violators? (2) If someone else identified suspected violators for such third parties, would the latter indeed start avoiding suspects? and (3)

Would there be any real consequences from the disclosure on the suspect ships and rogue oil exporters? Note that such real effects are key to the process of enforcing sanctions.

In order to address the first question, we take the perspective of a third party using publicly available data to predict which tankers are in violation, and then check the accuracy of these predictions. Doing so, however, requires knowing the "ground truth" – the list of tankers that have actually breached sanctions. We make progress here by using a unique dataset, provided by an anonymous port agent from a Middle Eastern country. This dataset catalogs all oil tankers observed violating sanctions in Iran during January 2021, including 33 tankers that were neither Iranian-flagged nor part of the Iran-Russia Caspian Sea trade. While limited in length, this is the only dataset of its kind, to our knowledge, providing the ground truth needed to evaluate predictive models targeting sanctions evasion in shipping – a key missing component that has hitherto limited research (e.g., Wolsing et al. (2022)).

Using this dataset, we construct machine learning models to identify potential sanctions-violating tankers, based on readily accessible high-frequency shipping signals and satellite data. We focus on factors that industry experts suggest as important for identifying sanctions evasion, such as tankers disappearing from tracking systems or reporting unusual travel paths suggestive of falsifying their actual routes (see Figure 3). Our analysis shows that even a sophisticated third party with the ability to conduct such an exercise would correctly identify a violator only 35.8% of the time, at best (with 95% confidence). This detection rate is not only low, it is likely an overestimate given the geographically limited area of the Persian Gulf (where our ground truth comes from).

Given such difficulty in identifying suspect tankers, we next address our second question: whether revealing their identities could change market participants' behavior towards them. If these suspect tankers were able to conduct business because their counterparties were unaware of their illicit activities, then disclosing this information could hurt the suspects. However, if the counterparties already knew about the suspects' activities but still chose to work with them, then such disclosure is unlikely to have any impact. Here, we go beyond Iran and look at all sanctioned jurisdictions.

To assess the balance of evidence on these alternatives, an ideal experiment would abruptly release a list of suspect tankers to the public. One could then study how this affects the listed vessels compared to "counterfactual" tankers – those not on the suspect list, but otherwise similar to listed vessels. We get close to this ideal by exploiting the disclosure of a list of suspect tankers on Reuters' Refinitiv Eikon platform by Windward.ai on August 16, 2023. While companies like Windward had been selling similar proprietary lists to specific clients even earlier, these data were prohibitively expensive for all but a handful of very large players before this event. For example, in their analysis of the sanctions advisories

for the maritime industry, IHS Markit and the Association of Certified Sanctions Specialists (IHS-ACSS, 2022) reported that "For smaller banks, the platforms and specific shipping tracking software and datasets, via third-party vendors, are considered to be very expensive in terms of their cost outlay. Every transaction per dollar for managing ship screening checks would probably not make sense in terms of cost from the bank's side." Following the disclosure, the Windward module on suspect tankers became available on Reuters for 356 US dollars per month, reflecting a significant cost reduction for accessing sanctions-risk information.

We start by assessing the accuracy of the Windward's sanctions-risk predictions, using our "ground truth" list of actual violators (which was provided to us three months prior to Windward's disclosure). We find that 27 out of 33 of our Iran-list tankers were on Windward's high- or moderate-risk list released in August 2023.[1] In comparison, our machine learning models could at best detect 13 of these tankers at a 95% confidence level (with the caveat that our numbers are not directly comparable, as Windward might have added tankers to their list based on activity in months other than January 2021). Even beyond our own exercise, another publicly available list of violators put up by the prominent advocacy organization United Against Nuclear Iran (UANI) contained only 10 of these tankers as of August 2023. This suggests that Windward's list is likely more accurate than alternatives.

We then examine whether Windward's disclosure makes it easier for third parties to identify suspect oil tankers. In particular, we examine the extent to which Windward's disclosure helps such parties to reduce their chance of unknowingly dealing with tankers at risk of facing legal actions for sanctions violations. We demonstrate that when augmenting a shipping signals-based machine learning model with Windward's list, the false positive rate in predicting which tankers will get sanctioned over the following three quarters is reduced from 51% to 20%. This reduction agrees with our hypothesis that shipping sector players who are responsibly aiming to identify and avoid possible sanction violators would take advantage of Windward's data and adapt accordingly their activities – next we show evidence for such adaptation.

Addressing our third key question, we assess the real effects on suspect tankers from making Windward's list public. Here we use Propensity Score Matching, in a Difference-in-differences setting (PSM-DiD). In this strategy, we compare tankers disclosed as high/moderate risk by Windward with counterfactual tankers, before and after the disclosure. Counterfactual tankers are other tankers that had a similar ex-ante propensity of being classified as high/moderate risk, but were actually disclosed to be low-risk. We also make use of Abadie (2005)'s semiparametric DiD; this estimator is derived under a weaker identification assump-

[1] This overlap also helps cross-validate the authenticity of the Iranian list provided to us.

tion than the PSM estimator, and its validity is not contingent on the correct estimation of matching weights.

In particular, we examine the response of four key outcome variables to this disclosure. First, one might expect that the risk of a tanker being closely investigated and impounded if found guilty of earlier violations will increase if it is on Windward's high/moderate-risk list. As a result customers worried about potential disruption to their shipment plans might be hesitant to hire such tankers following the disclosure, impacting their fixtures (i.e., contractual rental rates). Our evidence shows that the fixture rate for tankers disclosed as high-risk indeed dropped significantly, leading to a 16.4% drop in their earnings. Figure 12 shows no pre-trends, and that most of the drop happens immediately after disclosure. This decline is robust across specifications, and reflects the market's response to the increased risk of suspect tankers having violated sanctions, not violations themselves. We find no significant changes for tankers classified as moderate-risk.

Second, we examine whether newly disclosed suspect tankers avoid – or increasingly specialize in – routes passing close to sanctioned oil-producing countries - Russia, Iran, and Venezuela (before 2024). We find that tankers disclosed as high-risk are 25% less likely to navigate within 12 nautical miles of Russian, Iranian, or Venezuelan territories by the end of the sample, compared to similar tankers. This suggests high-risk tankers, aware of increased surveillance, steer clear of activities that could raise further suspicion of sanctions breaches. The effect on moderate-risk tankers is again insignificant.

Next, we examine changes in tanker ownership. Various sanctions regimes prohibit ownership changes for violating tankers. Therefore, when a tanker is added to the suspect list, demand in the used-tankers market is expected to decrease. Examining proprietary data from Drewry, we find that owners of high-risk tankers struggle to sell their ships once flagged as risky, with the chance of a sale during a year dropping 18 percentage points compared to owners of low-risk tankers. Given a typical 45.4% sale rate, this represents a 40% decrease in the likelihood of selling suspect tankers. Notably, this decrease in sales likelihood also applies to tankers disclosed as moderate risk.

Fourth, we find that chartering companies in the US or allied countries are 53.7% less likely to engage tankers classified as high-risk by Windward after the disclosure, compared to similar low-risk tankers. This could be due to the risk of particularly high scrutiny in their compliance programs, or the need to explain continued engagement with at-risk ships to domestic regulators. Charterers from U.S.-allied countries also tend to avoid these high-risk tankers, although this trend is weaker and not statistically significant.

In the final part of the paper we rationalize our empirical findings using a simple model of the oil shipping market. This model allows us to calculate the impact of Windward's

disclosure on the shipping costs paid by rogue oil exporters – something that we do not observe in the data. The calibrated model suggests that such actors indeed have to pay substantially higher prices to transport their oil. The model also implies that the disclosure results in lower oil exports by rogue exporters, who in addition need to pay higher prices for the tankers they employ.

Overall, our evidence shows that publicly identifying tankers suspected of sanction-violations reduces the profitability of facilitating evasion – by $620 million annually for suspect tankers in freight rates alone. Yet Windward's market value was only $47 million on the day of disclosure, highlighting a gap between violator harm from disclosure and market willingness to pay for it. This evidence ultimately suggests that a proactive information-provision approach could be more effective for sanctions enforcement than relying on private entities to identify violators on their own.

The rest of the paper is organized as follows: Section 2 discusses related literature, Section 3 outlines our motivation and provides a brief contextual background, Section 4 presents data, methods and results from models detecting sanctions-violators, Section 5 examines the effect of Windward's sanctions-risk disclosure on Reuters, Section 7 briefly discusses potential policy implications of our findings, and Section 8 concludes.

2 Related Literature

Our paper is related to the literature studying the difficulty in detecting evasion of economic embargoes and sanctions. Earlier studies on the topic include Hsieh and Moretti (2006), who use the price gap between Iraqi oil and its close substitutes to show that paper trails, even including the CIA's documentation, can underestimate the extent to which Iraq used bribes to circumvent the oil embargo. Dellavigna and La Ferrara (2010) examine how the illegal arms trade undercuts the effectiveness of the embargoes to a far greater degree than would be apparent from case-by-case evidence in investigative reports by groups such as Amnesty International. They propose a method to detect potential arms embargo violations based on reactions by individual stocks. More recently, Fisman et al. (2024) document a different type of sanctions circumvention – that of entrepot trading – using the context of Ukraine's economic blockade of the anthracite-rich breakaway Donbas region. Beyond the detection of sanctioned activity, our study relates to the literature on forensic economics, aimed at detecting illegal activities of various forms and fashion (see, e.g., Fisman and Wang (2017); Fisman and Wei (2004, 2009), Dimmock and Gerken (2012); Duggan and Levitt (2002); Griffin and Kruger (2023); Griffin and Maturana (2016); Griffin and Shams (2018); Jacob and Levitt (2003); Marion and Muehlegger (2008)).

We relate more generally to the literature on the economics of sanctions (e.g., Early and Preble (2019); Eaton and Engers (1992, 1999); Kaplow (1990); Lacy and Niou (2004); Siddiquee and van Bergeijk (2012)). We also relate specifically to the macroeconomic, political, and firm-level effects of trade restrictions with Iran (Dizaji and van Bergeijk, 2013; Haidar, 2017). More recent research discusses sanctions on Russia (e.g., Itskhoki and Mukhin (2022); Mamonov and Pestova (2023)), including a few studies on how firms, banks and other entities affected by embargoes or sanctions respond (see., e.g., Crozet et al. (2021); Efing et al. (2023); Huynh et al. (2023); Lastauskas et al. (2023)). Different from this literature, our paper examines the impact of providing identifying information on suspected sanctions violators, rather than the impact of the imposition of sanctions themselves.

Finally, the disclosure of risk scores aspect of our paper is related to work on sudden disclosure of information through credit registries, as studied by Liberti et al. (2016). More generally, the diffusion of information in financial markets relates to a large literature, e.g., to Hong and Stein (1999); Hong et al. (2007); Hong and Yogo (2012); Peress (2014).

3 Motivation and Background

Sanctions are a key component of economic statecraft, i.e., the pursuit of political goals using economic means (e.g., Baldwin (1985); Clayton et al. (2023); Morgan et al. (2023); Mulder (2022)). Most sanctions regimes, e.g., that of the U.S., can generally be divided into two categories: "primary" sanctions and "secondary" sanctions. Primary sanctions aim to change the behavior of foreign state or non-state actors by targeting economic activities that involve U.S. persons or entities, while secondary sanctions target activities of non-U.S. entities. Secondary sanctions authorize OFAC or the State Department to prevent non-U.S. entities from dealing with sanctioned activities by threatening to restrict their dealings with the U.S. financial system and firms. These sanctions are enforced on a strict liability basis, meaning that sanctioning authorities do not need to prove fault or intent in order to enter an enforcement action and issue a penalty (FCA, 2024; OFAC, 2024). Moreover, sanctions-related penalties can be severe; for example, violation of secondary sanctions can cut a target's access to the U.S. economy and financial system. To avoid such outcomes, global financial institutions have in several cases agreed to pay over a billion dollars in penalties to settle charges (e.g, $8.97 billion by BNP Paribas in 2014, $1.3 billion by Société Générale in 2018, $1.1 billion by Standard Chartered and $1.3 billion by UniCredit in 2019).

Aggravating matters, sanctions can be enforced retrospectively, as outlined earlier – this is true both for sanctions imposed by the US, as well as by the EU.[2] Combined with the

[2]On April 24, 2024, President Biden extended the statute of limitations for violations under U.S. sanctions

strict liability basis of sanctions enforcement, this compels businesses to scrutinise potential counterparties beyond checking that they are not on any sanctions lists (e.g., OFAC's Specially Designated Nationals (SDN) list). This is because such lists are inevitably incomplete and being continuously updated and extended – a firm can thus find itself in an ongoing relation with a party involved in violating sanctions, with all potential consequences.

Therefore, sanctions compliance efforts must include investigations of counterparties' past behavior that may reveal "red flags" and require additional due diligence that could preempt dealing with suspects, even if they are not currently sanctioned. Attesting to such avoidance, Stan Barstad, the CEO of the largest tanker-only company Frontline, commented that charterers are on alert for hiring tankers with suspicious activity in their history that cannot be explained (Lloyd's List, 28 Aug. 2023).

As a result, identifying such red-flag behavior is a first-order concern for many businesses today, leading to costs that extend beyond direct compliance expenditures. For example, identification difficulties not only require economic entities such as banks, insurance and logistics companies, etc. to spend abundant resources on due diligence, but also may lead them to avoid dealing with many "at-risk" counterparties altogether, even if these counterparties were actually legitimate. Such information frictions in the identification problem can lead to non-negligible economic losses (Gordon, 2016). For example, to avoid unwittingly facilitating money transfers for sanctioned entities, some American financial institutions have already cut U.S. dollar clearing services to Western businesses that have some presence in the Russian market, even if they are there legally (S&P Global (2022)). Similarly, wary of violating rules, banks might abandon entire sectors, even if they're not sanctioned, to avoid any potential regulatory run-ins (Bloomberg, 16 Jan. 2024). The financial media reports that such compliance costs for banks to build industrial-scale operations around sanctions compliance could represent up to 15% of a bank's total annual spend in some cases (Lloyd's List, 3 May 2022).

In fact, even the U.S. Treasury has flagged this issue – its 2021 Sanctions review highlighted that "small businesses may lack the resources to bear the costs of sanctions compliance" and "uncalibrated sanctions could unnecessarily lead them to turn down business opportunities in order to avoid these costs". Similar views have been expressed by private actors too – for example, American Club, as cited above, has suggested that governments have a lot more capability, manpower, resources to follow various activities in the shipping sector.

Beyond the profitability of banks and businesses, uncertainty regarding potential violators and the blacklisting of all related businesses can even affect supply of major commodities

laws, from five to ten years.

central to the world economy. For example, in September 2019, the U.S. sanctioned Cosco Dalian, a subsidiary of Chinese shipping giant Cosco, for transporting Iranian oil. Charterers immediately refrained from booking *any* Cosco tanker, *not just* those of the sanctioned Dalian subsidiary. That effectively removed more than 50 tankers from the global charter market overnight, causing spot rates to surge along with concerns about its effect on oil supplies worldwide (Reuters, September 28, 2019).

As mentioned in the introduction, in this paper we focus on the case of oil shipping to quantify the burden on third parties of identifying potential violators. Oil is one of the most important commodities affected by sanctions today. With sanctions in place already against Venezuela and Syria, the U.S. reimposed all sanctions against Iran in November 2018, threatening to sanction countries continuing to buy oil from Iran after a six-month waiver period expiring in April 2019 (OFAC, 2019b). Further, as part of the sanctions imposed on the Russian Federation, in September 2022, G7 finance ministers agreed to cap the price of Russian oil and petroleum products in an effort intended to reduce Russia's ability to finance its war on Ukraine. Any oil sold above that cap is subject to sanctions enforcement.

As the major transportation tool for crude oil – and therefore a key cog in worldwide macroeconomic activity – oil tankers naturally grab the spotlight for sanctions regulation. Given the large volume of oil under sanctions, industry insiders estimate the percentage of "shadow fleet" – i.e., tankers that might violate sanctions but are difficult to detect – is about 10% of the overall market.[3]

This sizeable proportion of suspect tankers increases the probability of unknowingly dealing with a sanctions violator. As we show in this paper, actors in the shipping sector who want to ensure a low probability of such deals will nowadays need to forego a significant amount of their business, even if they employ a fairly sophisticated set of analytical tools. The loss of business in this context is even higher when violators are difficult to detect. Furthermore, if major banks and insurance companies avoid business with a large number of potential violators, that may also lead to higher alternative insurance and risk prices for such tankers, and thus increases in the price for transporting oil – and therefore of oil itself. In the following section, we quantitatively examine the difficulty in identifying sanctions-violating tankers and discuss the resultant economic impact, as implied by our estimates.

[3] "A mysterious fleet is helping Russia ship oil around the world. And it's growing", CNN, Mar 2023; "Dark fleet of tankers now comprises 10% of seaborne oil transport", Lloyd's List, Mar 2023; "'Shadow Fleet' of tankers hauling Russian oil swells to 600 ships, Trafigura says", Bloomberg, Feb 2023

4 How difficult is it to detect sanctions-evading tankers?

To quantify the difficulty in detecting sanctions evaders, we take the view of a responsible entity making its best effort to comply with sanctions. We assume that such an entity employs sophisticated analytical tools, including machine learning methods and satellite image processing, in addition to analyzing radio-signal-based ship-tracking data from the Automatic Identification System (AIS), which was originally developed for collision avoidance. We first assess the detection accuracy only based on AIS data, and then add satellite data to the model to evaluate how far it improves accuracy.

4.1 Detection using AIS signal data

4.1.1 Data and predictor variables

The main dataset we use here is AIS, which is a key resource for monitoring tankers and regulators have emphasized its relevance for detecting sanctions violators.[4] AIS transmits at short intervals (1) fixed information including the vessel's International Maritime Organisation (IMO) number, Maritime Mobile Service Identity (MMSI) number, and vessel type and size, (2) dynamic information such as vessel position and speed, and (3) voyage information such as draught, destination and expected time of arrival. Dynamic information, e.g., position and speed, is updated automatically from the ship sensors, while voyage information is entered manually by the vessel operator. Vessels above 300 gross tonnage are required to carry AIS equipment and share AIS data with other ships and concerned parties to aid navigation (IMO (2000)).

Besides being used for navigation, AIS is a well-recognized tool for sanctions compliance (e.g., U.N. Doc. S/2019/171, 5 March 2019, Kilpatrick (2022)). Even flag state registries have de-flagged ships based on suspicions raised by AIS data (see Lloyd's List, Feb. 2020, Oct. 2020). According to regulators' guidance (e.g., OFAC (2023b)), shipping industry actors are advised to pay specific attention to suspicious gaps in AIS transmission ("Dark activity"), deviations in routes without legitimate reasons, engaging in ship-to-ship transfers, and patterns of shipping trajectories that appear anomalous (e.g., impossibly sharp turns at high speed, see Windward: "Going dark is so 2019", August 2020). The latter is suggestive of geolocation "Spoofing" – wherein AIS data is manipulated to show a ship in one location while it is physically present in another, as if using a VPN. Such spoofing was reported by the United Nations in 2019 (U.N. Doc. S/2019/171, 5 March 2019, pages 9-11) and has spread quickly since (New York Times, 3 Sep. 2022, Economist, 16 Apr. 2022).[5] Figure 3

[4] This dataset has also been used, for example, by Brancaccio et al. (2020).

[5] One method to implement spoofing involves AI-based technology that enables ships to emit machine-

illustrates episodes of spoofing and dark activity for two tankers on our dataset of Iranian sanctions violators. It shows the tankers' trajectories, as per their AIS signals. In one case, the tanker's uninterrupted AIS signals placed it in the northern part of the Persian Gulf at a time when it was observed at an Iranian port. In the second case, the tanker stopped transmitting AIS signals for four days, during which it was observed at an Iranian port.

Following prior studies, and consistent with guidance by regulators and other industry experts (e.g., Wolsing et al. (2022)), we construct six categories of sanctions evasion predictors. Then we test how successful they are in detecting the oil tankers that visited an Iranian port in January 2021, violating sanctions. This dataset of tankers was collected by our source through aggregation of a large number of text message from contacts at ports, brokers, and charterers. Here we list these predictors. The flexibility of machine learning allows us to incorporate various versions of these variables; detailed descriptions are in the Internet Appendix C.

- Identity Change: equals one if a tanker has changed its MMSI number or flag in the preceding six months. Windward (2022) has listed identity change as one of "the tried and true techniques of deceptive shipping practices".

- Risky Flag: equals one if a tanker is registered in a country whose flags are often linked to previous violations. According to Lloyd's List (2021), the top 10 such flags belong to Panama, Cameroon, Vietnam, Djibouti, Cook Islands, Tanzania, Togo, Palau, Russia, and Belize.

- Ship-to-ship Transfer (STS): we construct two types of predictors. They equal one if a tanker (i) is moored, or (ii) changed its draft at a location that is x km away from the nearest port (the draft is a measure of how much cargo the tanker is carrying). We set x to be 10 or 20 (i.e., we use two versions of each type). We include these predictors as OFAC has listed STS as one of seven deceptive shipping practices to be vigilant against (OFAC (2020)).

- Irregular Trajectory: equals one if a tanker's total changes of course within a period are abnormally high. We include this predictor following OFAC suggestions (OFAC (2020)). Figure 2 illustrates an irregular trajectory from a tanker in the Persian Gulf in January 2021. Our proprietary dataset reveals that this particular tanker did indeed visit an Iranian port during that week (hence was clearly violating sanctions) and had manipulated its AIS signal so that it appeared to be at sea during that period.

generated signals and appear to AIS as if they following an arbitrary path of their choice (Windward, June 2021).

- DBSCAN Outlier: equals one if a tanker sent signals from anomalous locations, which are obtained using a data-driven clustering method – DBSCAN. The clustering includes tankers' speed and course and we construct two versions of the predictor. Using DBSCAN to detect maritime anomalies follows industry practice (e.g., Wolsing et al. (2022)).

- Dark Activity: equals one if the time gap between two AIS signals is longer than T days, where we set $T = 1, 2, 3$ (i.e., three versions of the predictor). The bottom plot of Figure 3 provides an example of dark activity. We include dark activity predictors following OFAC (2020) and Windward (2022).

In total, we obtain 14 AIS-based predictors in the above six categories. All predictors are constructed at the tanker-day level, except for the Dark Activity ones, which are at the dark event level. The summary statistics of the predictors are shown in Table 1.

4.1.2 Methods

We test the ability of the above predictors to detect sanctions violators when (1) they emit falsified AIS signals (i.e., "spoofing") or (2) turn off their AIS signals (i.e., "going dark"). First, we construct a sample at the tanker-day level to detect tanker spoofing. The dependent variable equals one when a tanker is observed at an Iranian port, as per our proprietary dataset, while its AIS signals show it at a different location; and zero on other days. Second, we construct a sample at the dark event level to detect tankers going dark. A "dark event" corresponds to a tanker not sending an AIS signal for more than 24 hours, whereby a tanker may have multiple such events. When detecting dark activity, the dependent variable equals one if a tanker has been observed at an Iranian port during its dark event, and zero otherwise. Since obviously no AIS-based predictor can be observed during a dark event (when no AIS signals are emitted), we use instead the maximum value of each zero-one predictor on the days immediately before and after each dark event – the intuition here is that a tanker that stops or resumes emitting signals with anomalous behaviors (e.g., at an unusual location) may indicate to visit a nearby sanctioned port.

In our tests, we employ 13 AIS-based predictors for detecting dark (we remove one indicator in Dark Activity with $T = 1$ because it is always one given the sample is conditional on at least 24-hour dark) and 11 AIS-based predictors for detecting spoofing (we remove the three indicators in Dark Activity because they are always zero in this sample). To combine these predictors in an optimal (and potentially non-linear) way, we employ two popular machine-learning tools: decision trees and neural networks. Details are in the Internet Appendix D.

Both decision trees and neural networks need to specify several hyper-parameters, such as the depth of a tree and the number of layers of a neural network, before the training process. We use 10-fold cross-validation to determine the optimal values of these parameters. In this procedure, we randomly divide the sample into 10 equal sub-samples, and train models with different combinations of hyper-parameter values using nine of these sub-samples. The sub-sample left out in each case is used for validating the model predictions. We repeat this procedure ten times, and select the combination of hyper-parameters that yield the best validation performance, based on McFadden's pseudo-R^2.

To evaluate detection performance, we consider two measures: the pseudo-R^2 measure above, corresponding to the best cross-validated model, as well as a more intuitive measure, the implied false positive rate. In our context, the false positive rate indicates the proportion of business from sanctions-compliant tankers that a shipping sector firm must forgo to maintain the probability of engaging with a violator below a pre-specified level.

The map between pseudo-R^2 and the false-positive rate is as follows: Suppose that the unconditional probability of a tanker in the entire market to be a sanctions violator is $\bar{p}$. Suppose also that the false positive rate for a model is α, i.e., the model mis-classifies a compliant tanker as a violator with probability α. Assume further that the model correctly classifies a violator with probability β. Then, $1-\beta$, i.e., the probability of failing to detect a violator, can be viewed as a measure of risk. We assume that a firm will try to keep such a risk below a certain target level and thus will fix beta (e.g., $1-\beta \leq 5\%$). Because pseudo-R^2 is a function of $\bar{p}$, α, and β (see Internet Appendix B), there is a one-to-one mapping between pseudo-R^2 and α, after we fix β and $\bar{p}$. Therefore, given a pre-specified risk level $1-\beta$ and an estimated proportion of violators $\bar{p}$, the statistical performance measure pseudo-R^2 can be translated to a false positive rate α.

4.1.3 Results

Results from this exercise are presented in Table 2. Panel A of this table shows that even the best AIS-based detection models yield Pseudo-R^2 around 10%, for the average of detecting spoofing and dark activity. The implied false positive rates for a firm that would want to ensure that it avoids 95% of true violators are also staggeringly high at about 50%. This implies that to maintain the risk of unknowingly doing business with a violator at 5% (or, conversely, to detect a sanctions violating tanker with 95% probability), a firm has to avoid more than half of its potential counterparties.

In panel C, we examine the incremental contribution to predictive power, as well as to the false positive rate, for each of our underlying predictor variables. Given the low overall Pseudo-R^2-s above, it is not surprising that incremental Pseudo-R^2-s are all low; among

them the Ship-to-ship Transfer and Dark Activity variables seem somewhat important. At the same time, we see that for these two variables the incremental false positive rates are both over 20%. This implies that relative to a model without, say, our Ship-to-ship transfer variables, one with remaining variables would incrementally lead to a firm falsely detecting 20% of their compliant clients as potential violators.

We conclude that despite combining many of the predictors mentioned in industry and regulatory sources – and doing so in a relatively sophisticated way – detecting sanctions evaders based on AIS data is ultimately highly inaccurate, which in turn can have significant economic impact on sanctions-complying firms.

4.2 Detection using AIS and satellite data

Given the limitations of using AIS data alone, as discussed above, in this section we enhance our sanctions detection models by adding satellite data. We expect these data to be particularly useful in the detection of spoofing. In particular, our AIS-based predictors only reflect irregularities that can be indicative of spoofing, while a satellite can observe directly the area where AIS signals place a tanker at a given time and thus help determine whether they are deceptive. In turn, satellite data could also be useful for detecting dark activity, especially if they contain multiple images of the locations where a tanker's signals stopped and then resumed and their vicinity. Given the scope of our satellite data, we use them to detect dark activity when these locations are close to each other. We note that regulators have also recognized the usefulness of commercial satellite imagery for "identifying vessels, monitoring vessel behavior, and pinpointing locations of vessels that are inconsistent with information transmitted via AIS" (OFAC (2023b)).

4.2.1 Data and methods

We use data from the Sentinel-1 mission, which has two satellites that can take an image of the same location on Earth at six-day intervals. It generates images by synthetic aperture radar (SAR) and makes freely available high-resolution, all-weather Earth observations.[6] We use all satellite images that contain some parts of the Persian Gulf and were taken in January 2021. Figure 4 shows three of these images, together with enlargements that make visible individual ships at sea. The processing of satellite images was done by Dfy Graviti – a company specializing in aerospace and maritime AI,[7] so that we can distinguish land vs.

[6]This dataset is widely regarded as one of the most popular satellite sources of imagery currently available and, with its 5-day revisit rates, the closest one can get to free, real-time satellite observations (e.g., https://skywatch.com/free-sources-of-satellite-data/).

[7]See https://www.dfy-graviti.com/

sea area, identify ships at sea (e.g., the yellow objects in Figure 4), and obtain their exact locations, i.e., latitude and longitude.

Using such images, we construct satellite-based predictors and combine them with the AIS-based predictors described in Section 4.1 to improve detection power for both spoofing and going dark. Ideally, we would detect spoofing if a satellite image taken at the exact time of a tanker's AIS signal shows no ship at this time at the location transmitted by this signal (or, conversely, shows the tanker at a location different from the transmitted one); we would also detect dark activity if we could track a tanker's trajectory during a dark period (e.g., we may find violators headed towards an Iranian port). However, this ideal situation would require frequent, if not continuous, satellite observations of multiple locations in the Persian Gulf, whereas in fact such observations are quite sparse – the satellites generating our data revisit each area only at six-days intervals.[8]

To deal with the challenge from sparse satellite images, we proceed as follows: to detect spoofing, we focus on tankers that (1) have emitted AIS signals within a few hours of the time an image is taken, and (2) have been placed, as per those signals, within the area seen in this image. Then we use the available trajectories of other tankers in the same area to define a search area in an image where the tanker of interest is most likely to be found. Figure 5 gives an example of constructing the search area via interpolation and detailed descriptions are in the Internet Appendix C. If, after processing the image, we do not find any tanker in this search area, we conclude that the tanker was likely involved in spoofing. To detect dark activity, we focus on tankers that go dark and then resume emitting AIS signals at about the same place and define search areas around these places – if a satellite image of such a search area that was taken within the dark period does not show any tanker there, we conclude that the respective tanker was likely involved in sanctionable activity during its dark period. Overall, our approach mimics the reasoning underlying hypothesis testing – under the null hypothesis the tanker is compliant and is likely to be seen in a certain area; not seeing it in this area leads to rejection of the null, given some confidence level.

We construct the following satellite-based predictors and add them to the previous list of AIS-based predictors.

- Satellite Detection: we construct these predictors separately for detecting spoofing and

[8]Note that even if other sources of satellite data could provide more frequent observations – for example, one may have access to data from a larger satellite fleet – the use of satellite images poses specific challenges and requires specific approaches to address them. For instance, it is not feasible, even on high-resolution images, to try to identify directly ships at port, where multiple vessels of similar size and shape are densely concentrated and also frequently change positions, as they are waiting to be be loaded or offloaded. Furthermore, sanctions violators have been reported to intentionally modify visual characteristics of their vessels, like repainting in a different color or changing the vessel name or IMO number on the ship's hull (IHS-ACSS, 2022).

dark activity. For spoofing detection, the predictor equals one for day t if a tanker is not seen in its corresponding search area during the period $[t-r, t+r]$. We set $r = 0, 1, 2$ (i.e., three versions), recognizing that spoofing on a given day t may also imply spoofing on adjacent days (i.e., r can be viewed as a smoothing parameter). For dark activity detection, the predictor equals one if a tanker is not seen in its corresponding search area during its dark period, when this area can be reasonably estimated. Because the dark period is observed, we don't construct multiple versions with different periods as in spoofing detection.

In total, we have 14 predictors for spoofing detection and 12 predictors for dark activity detection.

4.2.2 Results

Panel B of Table 2 reports the Pseudo-R^2 of ML models and the implied false positive rates under $\beta = 95\%$ using both satellite-based and AIS-based predictors. Adding satellite images to models improves the detection performance compared to using only AIS-based predictors. For example, when detecting spoofing, the Pseudo-R^2 increases from 6.7% to 11.8% for trees after including satellite images, and from 7.2% to 13.8% for neural networks. The Pseudo-R^2 for the detection of dark activity also increases from 11.5-13.3% to around 18.7%. The implied false positive rates also decline, but not very significantly. For example, to maintain a 95% probability of avoiding sanctions violators, a firm in the shipping sector has to give up 35.8-38% of business from compliant tankers.

Note that we perform this identification exercise in a relatively advanced way – involving multiple machine-learning algorithms, spatial interpolations, and satellite imagery analysis – while such analyses may well be beyond the scope of many actors in the shipping sector. For example, a shipping registry from, say, Panama or Liberia, that provides flags to tankers, or a small trucking company transporting oil from these tankers, perhaps could not reasonably be expected to conduct such analyses on their own. Further, because our ground truth dataset comes only from Iran, we restrict our analysis to the Persian Gulf region – given the geographically limited area of the Gulf, this should make our predictive exercise easier than if, for example, one were to do the same for a country as large as Russia. Therefore, our estimates likely provide an upper bound for how successful such a detection exercise can be.

Overall, our analysis shows that the difficulty in detecting sanctions evasion could lead to large economic losses for third-party businesses.

4.3 Prediction accuracy and detection difficulty

Our results suggest that detecting sanctions evasion is possible to only a limited extent when utilizing readily accessible datasets, even with the application of powerful analytical tools. At this point, one may wonder whether this evidence is sufficient to assert that detection is difficult; after all, the poor performance of our models could just reflect limitations in either our data or our modelling capabilities. In this section, we delve deeper into this possibility, and first show that certain features of the data that are often seen as 'markers' of sanctions-evading activity are associated not only with violating tankers, but are almost equally prevalent for compliant ones too. For example, compliant tankers can also have gaps in AIS transmission and show seemingly anomalous patterns of movements, just like the violators. Moreover, even going beyond our data and our models, other detection algorithms – for example, that used by UANI – are also limited in terms of their accuracy, as we verify by benchmarking their list of suspect vessels with our ground truth dataset.

For an illustration of the similarity between salient features of compliant and non-compliant tankers, consider the irregularity of a tanker's trajectory. In Figure 7, the left panel reproduces Figure 2 – showing an irregular trajectory from a sanctions-violating tanker in January 2021. However, even such a stark pattern is not a reliable indicator on its own. When comparing it with the right panel of Figure 7, we observe a trajectory with similar apparent irregularity, over the same time period and at a comparable area in the Persian Gulf – however, this latter trajectory belongs to a tanker which was not violating any sanctions. AIS data show multiple such "irregular" paths for non-violating tankers. Highlighting the same issue, we plot in Figure 8 the distributions of dark time for violators and compliant tankers, conditional on a tanker going dark (i.e., a gap of at least half a day between consecutive AIS signals). As observed, these distributions overlap to a large extent, and hence compliant tankers resemble violators in terms of dark activities too. This similarity may be due to mechanical failures or data issues in AIS.

The main issue is that tankers may move somewhat irregularly or show periods of dark activity, etc., for various legitimate reasons (e.g., weather conditions, mechanical failures, ship traffic congestion, data issues in AIS itself, etc., see Weng et al. (2022)). Coupled with the fact that our goal here is to detect low probability events (sanctions-violating activity is a small proportion of overall tanker movements), this results in a low signal to noise ratio in the AIS data – even for signals suggested by industry insiders.

Moreover, and contrary to conventional wisdom, satellite data available for public use is sparse. Each satellite sees just a sliver of the earth at a given time, and then revisits the same sliver only after a significant time gap. As AIS data is also often intermittent, this sparsity of satellite images causes problems for detection when a tanker fails to emit an AIS signal

around the time when the satellite sees its location. This is a key reason why our satellite-and-AIS based models do not improve pseudo-R^2 by a large amount, despite the fact that the satellite images are analyzed by experts in maritime situational awareness (Dfy-graviti, the company that helped us analyze satellite images is an AI-deep tech company, that counts the defense and space sectors in India among its clients).

Overall, these issues are likely to make detection difficult for any model – not just ours – that has to rely on readily accessible data. Governments and large corporations (e.g., Starlink) with access to better and/or many satellites, of course, do not face such issues, but their data is not publicly accessible.

To give a deeper perspective of our models and their low accuracy, we plot the general relation between predictive accuracy (pseudo-R^2), sanctions-risk level ($1-\beta$), and the implied false positive rate α (i.e., the amount of business a firm needs to give up to keep sanctions-risk at that level and given that accuracy). This exercise helps visualize how much better accuracy needs to get – relative to that achieved with our models – in order for businesses to be able to avoid suspects while still retaining most of their compliant clients.

In the left panel of Figure 9, we set $\bar{p}$ (i.e., the unconditional probability of a tanker in the entire market to be a sanctions violator) at 2.23%. This is the ratio of the number of sanctions violators in our proprietary dataset to the total number of tankers observed in the Persian Gulf in January 2021 as per AIS data (33/1481). The two curves in this panel correspond to two different risk levels $1-\beta$ (5% and 1%) that a firm might be willing to tolerate – e.g., a pseudo-R^2 of 15% implies a false positive rate of 38% for a risk level of 5%. In other words, a firm needs to give up one third of its business so that it can avoid sanctions violators with a 95% probability. (Note that most of the pseudo-R^2 estimates in Tables 2 are around 15%, and so this example is highly relevant. Importantly, the figure illustrates that even if a shipping industry firm could achieve double the accuracy of our model – perhaps due to having better data or superior algorithms – such a firm would still need to give up about 20% of their current clients in order to maintain sanctions-risk exposure at 5%.

In the right panel, we fix the risk level at 5% and vary $\bar{p}$ (2.23% as in the data, 5%, 10%, or 20%). We do this exercise because the "shadow fleet" of suspect tankers has grown in recent years, and is now estimated to be about 10% of the total, as mentioned in Section 3. Intuitively, if businesses know ex-ante that there are more violators in the population, they need to be more careful with screening clients. However, for a given model accuracy (i.e., fixed pseudo-R^2) and risk level $1-\beta$, this implies that they need to forgo more of their business. For example, with a pseudo-R^2 of 15% and 10% sanctions violators in the population (green line) rather than 2.23% as in our baseline estimate (red line), a firm must give up 48% of their clients rather than 38% in the baseline. Therefore, the false positive

rate has likely worsened significantly in more recent times as the shadow fleet has grown.

In sum, it is indeed difficult to detect sanctions evasion based on public data and standard predictors, as the low accuracy of our models illustrates. This is largely due to similarities along various dimensions between compliant tankers and sanctions violators. We also show that even a significant improvement in prediction accuracy that could come from better data or models will still present actors in the shipping sector with the hard choice between either losing an unacceptably large proportion of their clients or incurring unacceptable levels of sanctions-related risk.

Finally, the difficulty in detecting evaders extends well beyond our data or methods. In order to see this, one can check how many of the 33 tankers on our ground truth Iran-sanctions-violator dataset appeared on alternative publicly available lists aimed at identifying such violators. One such example is UANI's "The Ghost Armada" list of suspect sanctions-violating tankers. UANI is a bi-partisan, non-profit organization in the US whose mission is to prevent Iran from acquiring nuclear weapons capabilities. They started publishing their list in November 2020, and have been periodically updating it. In Figure 6, we compare this UANI list to our ground truth dataset from January 2021, by plotting how many of the 33 violating tankers were included by UANI and when. The figure shows that even as of April 2024, UANI had detected only 13 of these tankers.

5 Public disclosure of suspect tankers

If the difficulty in detecting sanctions evaders indeed poses a key hurdle to the effectiveness of sanctions, then one should observe two crucial features in the data: (1) it is indeed difficult to detect evaders through honest and reasonable effort; and (2) if this detection problem can somehow be solved – e.g., by providing a list of suspects detected with reasonable accuracy – then players who previously engaged with these suspects due to a lack of knowledge should adjust their behavior. We have already demonstrated (1) in the previous section; in this section, we shed light on (2).

Note that in this part of our study (except for Table 3 on comparing accuracy with our previous models) our analysis is no longer restricted to the Persian Gulf – the dataset that we employ here also contains tankers related to sanctioned jurisdictions outside Iran.

5.1 The disclosure event

Given the central role of information in the sanctions context, proprietary information on suspected evaders is now offered by several AI companies, which aim to detect maritime

sanctions evasion using marine, terrestrial, and satellite data, as well as machine learning algorithms. They include Lloyd's List, Planet Labs, Pole Star, Spire, TankerTrackers, and perhaps the most well-known among them, Windward.ai. However, before the August 2023 event that we consider, the cost of obtaining such data was prohibitive for most industry participants except the largest of players, as discussed previously.

The information shock that we exploit in this paper relates specifically to one of these providers – Windward (LON:WNWD) – which focuses on tracking vessels in real time, identifying bad actors and various anomalies. Observers have pointed out that Windward's product is effectively military-style signals intelligence (SIGINT) for commercial use (RAND, 2017).

Windward uses over 100 million data points daily to map events across the world's oceans. These data include not only freely available sources, but also proprietary satellite imagery from Planet.com, weather data from Meteomatics, etc. Most importantly, Windward augments these datasets with information provided by its clients – which include the United Nations, the European Border and Coast Guard Agency (Frontex), and US agencies such as the Drug Enforcement Administration and the Office of Naval Intelligence (See Reuters, 17 Mar. 2016, Wired, 18 Mar. 2020).

This data is processed by Windward's 165 employees across offices in the UK, Israel, the US, Ukraine, and India, who include former naval commanders, maritime data scientists, maritime traders, and shipping practitioners. Many industry observers (e.g, Robinson, 2020) speculate that the vast amount of data collected over 10 years of Windward's operation, especially in the hands of these carefully curated group of industry experts, is the main reason why there are no direct competitors to Windward.

Starting in August 2023, Windward made available on the Refinitiv Eikon information platform (governed by Reuters) its list of tankers that have likely violated sanctions. This list covers the entire global fleet of tankers (above a certain size) and classifies the tankers' riskiness into "Low", "Moderate", "High", and "Sanctioned". It also shows the reasons for the risk classification provided, which include ID or location manipulation, flag hopping, dark activity, STS activity, etc. (A screenshot is provided in the Internet Appendix F.) The subscription fee for Refinitiv Eikon is 280 GBP per month, making it easily affordable.[9] While before this release such risk information would have been prohibitively expensive for all but a few major players, now port officials, logistics companies, foreign ship-flag registries, etc., as well as most banks and insurance companies can easily access the list on Reuters'

[9]Windward continues to sell tailored products to commercial and government clients; if anything its overall sales increased from winning new contracts of this type following the disclosure of their sanctions risk data. See, for example, https://www.proactiveinvestors.co.uk/companies/news/1029531/windward-shares-swell-on-ai-maritime-contract-wins-1029531.html.

well-regarded data platform.

Moreover, such a release of information impacts higher-order beliefs, as the riskiness of each tanker becomes common knowledge (Aumann (1976)). That is, the businesses that knew the list of suspects before August, because they had paid for that information, would now realize that everyone else is also aware of the same list. Those of them who thought earlier that it might be okay to do business with a suspect – perhaps assuming that few others knew the contents of the list – would now have to rethink. As a result, the disclosure is likely to not only change the behavior of those who previously lacked access to the suspect list, but also affect the behavior of those who had it due to this common knowledge aspect.

The disclosure of the Windward's list of suspect tankers on Reuters could have real effects on how businesses behave only if: (1) a substantial number of businesses knew about this disclosure and paid attention to it; and (2) the list itself brought new and accurate information on suspected violators.

First, we look for clues in public data of whether the market indeed paid attention to the disclosure of Windward's list on Reuters – i.e., whether there were many businesses and organizations who are likely to have looked for Windward's list. We make progress here by showing that attention to Windward – as measured by page views of the Windward.ai homepage – increased dramatically right after the list was put up on Reuters. Figure 10 presents this evidence.

Now, in order to truly bring new information about suspected evaders to the market, Windward's list must be significantly more accurate than alternative detection methods. To assess this, we benchmark its accuracy against our methods from the previous section, using our proprietary Iranian port data on actual sanctions evaders as the ground truth.

Table 3 shows first that out of the 33 tankers that have violated Iranian sanctions as per our dataset, 27 have been detected by Windward, i.e., they are labeled high- or moderate-risk (here we drop the Iranian tankers from the sample and ignore the those labeled as "Sanctioned", i.e., those tankers that are known with certainty to be associated with violations). The table then shows the performance of the decision trees and neural networks with optimal structures, as in the Internet Appendix Table A-1. To calculate the number of detected tankers for a given confidence interval, we first collapse the sample from tanker-day level (that we used for spoofing detection) or event level (that we used for detection of tankers "going dark") to the tanker level. We do that by taking the maximum value of the predicted probability of a tanker to be a sanctions violator, and thus likely biasing upwards our models' performance. We report results for several confidence levels.

It is clear from the table that our models fall significantly behind Windward's detection success rate. Even at the 90% confidence level, our models can detect at best only half the

violators, compared to 27 out of 33 for Windward. Further, as discussed before, another prominent public data source on sanctions-violating tankers, UANI's "The Ghost Armada" list, featured only 10 out of the 33 actual violators at the time of the Windward disclosure. These comparisons demonstrate the relative accuracy of Windward's risk evaluations.

5.2 The effect of Windward's disclosure of suspect tankers

We first examine whether the disclosure of a public list of at-risk ships indeed makes it easier for third parties to identify whom to avoid. Next, we examine whether there are any real effects of the disclosure on suspect ships – key to the process of enforcing sanctions.

5.2.1 Effect on the difficulty of identifying at-risk tankers for third parties

We start by examining whether, and by how much, Windward's disclosure can help a third party identify tankers at risk of legal action for sanctions violations. To do so, we again take the point of view of such a third party trying to predict, out-of-sample, the tankers that will actually be sanctioned in the future. Recall that it is a key concern for businesses to be able to detect – and subsequently avoid – such tankers, given the retroactive nature of sanctions as well as their strict legal liability.

We study tankers that were actually sanctioned following Windward's disclosure, i.e., those sanctioned between August 2023 and April 2024. Using this list of tankers, we contrast the predictive ability of public data vs. Windward's list. We start with a sample of 13,997 tankers from Reuters (the sample filtering process is shown in the Internet Appendix Table A-2). For Windward, we take the 3,500 tankers classified as high-/moderate-risk at the end of July 2023 (i.e., right before Windward's disclosure) as those that they predict to be at risk of sanctions.[10]

For comparison, we estimate the probability that a tanker might be sanctioned in the near future based on publicly available data. To do so, we follow a standard training-testing procedure. We construct a training sample using data up to January 2023, where the dependent variable takes a value of one for tankers that were already sanctioned by that date. Predictors based on public data include AIS-based predictors (as described in Section 4.1), satellite-based predictors (as described in Section 4.2)[11], and a zero-one predictor derived from the UANI list (as described in Section 4.3). We train ML models (decision trees and neural networks) as in Section 4.1, and then test them out-of-sample, on ships sanctioned

[10] In this exercise, we focus on all tankers. Results are similar if we use a filtered subsample, e.g., only oil tankers.

[11] Besides the Persian Gulf, we extend our satellite-based predictors in this test to cover the Black Sea and the Baltic Sea to predict Russian sanction violators.

between January-July 2023. We choose the best model based on this out-of-sample performance. Finally, using that model, we estimate, standing at the end of July 2023, the probability that a tanker will be sanctioned in the future. The public-data-based set of tankers predicted to be at risk of sanctions is based on these estimates.

One simple way of contrasting Windward's predictive ability with our best model using public data is to sort all tankers based on the estimated probabilities of sanctions based on our model, and take the top 3,500 tankers from this list. Recall that there are 3,500 tankers on Windward's risk list as of that same time point. We find that out of the 69 tankers that were actually sanctioned after July 2023, our best model using public data predicts 33, while Windward predicts 65. This comparison confirms that Windward's disclosure significantly reduces the difficulty of detecting tankers that are subsequently sanctioned.

For further comparison of out-of-sample predictability, we evaluate the performance for weighted averages of Windward's and public-data-based predicted outcomes that use various weights. This exercise is aimed at understanding whether and how much incremental predictive power Windward adds above public-data-based machine learning, and vice-versa.

Figure 11 plots the out-of-sample performance measures (i.e., pseudo-R^2 or implied false positive rate) against the corresponding weights. It shows that the prediction based on purely public data yields an out-of-sample pseudo-R^2 of 7.5% and an implied false positive rate of 51%.[12] In contrast, the prediction based only on Windward's list yields a significantly higher pseudo-R^2 of 17.2% and a lower implied false positive rate of 26%.

Furthermore, a combination of $0.58 \times$ public data $+ 0.42 \times$ Windward's list, achieves an optimal pseudo-R^2 of 21% and implied false positive rate of 20%. This suggests that machine learning on public data (of the type we described above) still has incremental value, adding to the predictability from the Windward's list alone.

5.2.2 Real Effects on Oil Tankers

Is the difficulty in identifying sanctions evaders indeed an important reason for the sometimes limited impact of sanctions (Jones (2015); Pape (1998)) despite their broad reach (especially for secondary sanctions)? If so, then once a reputable source like Reuters identifies suspects, businesses should adjust their behavior and avoid engaging with suspect counterparties. Conversely, if identification itself was not a primary issue – perhaps businesses did not avoid

[12]The implied false positive rates here are not directly comparable to those in Table 2 for several reasons: (1) the dependent variables differ, as we predict future sanctioned tankers here, whereas Table 2 focuses on detecting tankers that went to Iranian ports, i.e., violated Iranian sanctions; (2) the scopes are different, as this analysis encompasses all sanction regimes, while Table 2 is limited to Iranian sanctions; and (3) the evaluation methods differ, with this analysis using out-of-sample measures, whereas Table 2 relies on cross-validation.

suspect counterparties because they thought that the legal burden of prosecuting them is too high – then someone else solving the identification problem for them would likely have no effect. In the following sections, we investigate this issue by examining how Windward's disclosure impacts the shipping sector from different aspects: tankers' freight rates (fixtures), routes, ownership changes and identities of the charterers.

Suppose we want to examine the effect of disclosing the list of suspect tankers on their fixtures (freight rates). We can do so by estimating the change in the fixture of a tanker classified as high/moderate risk in Windward's list after disclosure and comparing it to the change in the fixture of a counterfactual tanker - i.e., one that is otherwise identical to the treated tanker but is not itself labeled as high/moderate risk.

The main challenge for such a strategy lies in the construction of the counterfactual. The key issue is that the untreated tankers, i.e., those classified as low-risk by Windward, may not provide accurate counterfactuals for high/moderate-risk tankers. This is because the pre-treatment characteristics of the low-risk tankers could be so different from those of high/moderate-risk tankers that their fixtures would have trended differently from those of the high/moderate-risk tankers even in without Windward's disclosure. In that case, the conventional difference-in-differences (DiD) estimator would be biased, because its key identification assumption – parallel trends for the treatment and control – no longer holds.

We use two methods to address this issue. The first is propensity score matching (PSM), introduced by Rosenbaum and Rubin (1983) to construct counterfactuals relying on the strong ignorability assumption ("unconfoundedness"). PSM is an intuitive method and makes it easy to visualize the data using standard DiD plots. Second, we follow Abadie (2005) and employ a semiparametric DiD estimator, which is derived under a weaker identification assumption than that of PSM. In particular, the PSM estimator requires conditional unconfoundedness – the assumption that conditional on observable confounders (which determine propensity scores), potential outcomes are independent of treatment status. Abadie's estimator, on the other hand, only requires conditional mean independence – that is, conditional on covariates, the expected value of the outcome variable does not depend on treatment status.[13] Moreover, PSM requires *estimating* matching weight as a function of propensity scores, and choosing the functional form. In contrast, Abadie (2005)'s method is theoretically *derived* and therefore offers lower degrees of researcher freedom regarding specification choice. Further details on our specific implementation of these two estimation strategies are in the Internet Appendix E.

[13]See, e.g., Heckman et al. (1997), about the discussion on strong ignorability assumption and Abadie (2005) for the mean independence assumption.

A. Freight rate (Fixture) changes

We first estimate the treatment effect on tankers' freight rates, using fixtures data from Refinitiv Eikon. More precisely, a fixture is a contract that concludes the negotiations between a shipowner and charterer, generally through shipbrokers, which result in the charter of a ship. A key component of this contract is the freight rate (or rental price for using the tanker) to be paid to the owner, which itself is also referred to as a fixture. Since the information shock occurred in August 2023, we use the preceding six months (February to July 2023) as the pre-period, and August 2023 to March 2024 as the post-period.

We hypothesize that after the information shock, the freight rates for a tanker reported as high/moderate-risk will decrease compared to that for a counterfactual tanker without such a risk indication (i.e., low-risk). Without Windward's disclosure, it would have been more difficult for charterers to identify potential sanctions violators; given this information, however, charterers view high/moderate-risk tankers as less desirable and this leads to lower fixture rates.

In shipping contracts, spot freight rates are often quoted relative to "Worldscale" (WS), i.e., a baseline rate calculated for a standardized vessel on a round-trip voyage between two specified ports (or between ports in two specified areas). Such baseline rates are defined as "WS100" and an actual quote can be given relative to the baseline – for example, "WS75" or "WS200" would denote 75% of the baseline rate or double this rate for a charter between the baseline ports. WS rates are revised annually based on updated bunker prices, port costs, and exchange rates.[14] The key advantage of the Worldscale is that it makes freight rates across different contracts more comparable. About two-thirds of the contracts in our sample are in WS format, and we focus our analysis on WS rates.

To test our hypothesis, we construct the control group based on propensity scores. We calculate propensity scores by regressing the high/moderate-risk indicator on a set of predictors. Here we use the same methods and set of variables that we used to detect sanctions evaders in Section 4, except that: (1) we construct variables using only pre-period information, i.e., Feb 2023 - Jul 2023, and extend the scope to all tankers rather than only tankers that were in the Persian Gulf as in Section 4; (2) since our satellite measures only cover the Persian Gulf in Jan 2021, we don't include them when calculating propensity scores; (3) since here the pre-period includes multiple months, instead of using zero-one indicators, we use counting variables after $\log(1+x)$ transformation as predictors, for example, log of one plus the number of dark events during pre-period as one predictor. The summary statistics of the sample and balance tests are shown in Table 4.

After estimating the propensity score, we match each fixture contract of a treated tanker

[14] See https://www.worldscale.co.uk/ for further details.

to the contracts of untreated tankers by requiring that (1) these contracts are in the same month; and (2) the matched tankers are of the same type (i.e., Medium, Panamax, Aframax, Suezmax, or VLCC). The matching weights are determined by the difference between two propensity scores with a Gaussian kernel. We report results for bandwidth parameters 0.01, 0.03, and 0.05. We first estimate the time series of treatment effects on fixture rates for high-risk tankers only, i.e., we drop moderate-risk tankers from the sample, conduct matching, and estimate the following regression:

$$fixture_{c,i,k,t} = \left[\sum_{l=-6}^{7} \beta_l^H \times high_risk_i \times \mathbb{I}_{\{t=l\}}\right] + \alpha_i + \gamma_{k,t} + \epsilon_{c,i,k,t}, \tag{1}$$

where $fixture_{c,i,k,t}$ is the WS rate for contract c, tanker i (with tanker type k) in month t; $high_risk_i$ equals 1 for tankers labeled as high-risk in Windward's list, and 0 otherwise; α_i is tanker fixed effect, and $\gamma_{k,t}$ is time×tanker-type fixed effect. We are interested in the average treatment effect on the treated (ATT) for high-risk tankers, given by the coefficient series $\{\beta_l^H\}_{l=-6}^{7}$.

Figure 12 shows that the PSM-DiD estimated coefficients in the pre-period $\{\beta_l^H\}_{l=-6}^{-1}$ are insignificant, indicating a lack of pre-trends. After Windward's disclosure, there is an immediate drop of around 20 WS units in fixture rates in August for high-risk tankers compared to the matched control group (recall that fixtures are expressed in percentage points, with a WS of 100 as a benchmark). This drop persists until the end of our sample.

To understand what such a drop in fixtures means for tanker earnings, we can perform the following back-of-the-envelope calculation. Given the average fixture for high-risk tankers in July 2023 is 126.8 WS units as in Table 4, the 20-unit drop in August implies revenue from chartering decreases by 15.8% (20/126.8). From Drewry (a leading provider of research and consulting services to the global shipping industry), the average time charter-equivalent rate of a VLCC tanker in August 2023 was $40,600 per day, and the average operating expense was $10,050, yielding average earnings of $30,550.[15] After disclosure, the revenue would go down by $6,415 ($40,600×15.8%). Suppose operating costs do not change, this would lead to a 21% (e.g., 6,415/30,550) drop in earnings.

While post-period coefficients estimated in Figure 12 lie below the pre-period ones, month-by-month estimation is noisy. To mitigate this noise, we aggregate the indicators in the post-period months into a single dummy, $\mathbb{I}_{\{t\geq 0\}}$. We separately estimate ATT for high-risk and moderate-risk tankers. In a matched sample with high- and low-risk tankers,

[15] Aframax and Suezmax tanker rates were similar to the VLCC ones at the time.

we run the following regression:

$$fixture_{c,i,k,t} = \beta^H \times high_risk_i \times \mathbb{I}_{\{t \geq 0\}} + \alpha_i + \gamma_{k,t} + \epsilon_{c,i,k,t}, \tag{2}$$

where β^H is the ATT for high-risk tankers. We conduct a similar estimation for moderate-risk tankers by replacing $high_risk_i$ with $moderate_risk_i$. Panel A in Table 5 shows the regression results from the above PSM-DiD.

Results from our second estimation strategy, i.e., Abadie (2005)'s semiparametric DiD, are presented in panel B of Table 5. Since Abadie (2005)'s estimator is derived without directly accounting for macro trends through time fixed effects, we do so in two ways. To account for month fixed effects, we manually subtract the cross-sectional mean of the fixtures in each month. To also account for time×tanker-type fixed effects (and thus adjust for tanker-type-level monthly changes in freight rates), we manually subtract the cross-sectional mean of the fixtures, by tanker type, in each month. We report the three versions of the estimated ATT, i.e., using raw fixtures, fixtures demeaned by time, and fixtures demeaned by time and tanker type, respectively. We also separately estimate ATT for high- and moderate-risk tankers as in Eq.(2). Note that the number of observations in panel B is smaller than in panel A. This is because Abadie (2005)'s estimation is conducted on a tanker-level sample instead of a tanker-month level for PSM-DiD. Also, to calculate the change in fixtures as the dependent variable, i.e., post-period average fixture minus pre-period average fixture, tankers without both pre- and post-fixtures will be dropped in panel B.

The results show that, after the information shock, the fixtures of high-risk tankers decreased by an average of 16.45 WS units across specifications, with a range of 13.75-20.71 units. This translates to a drop in earnings of around \$5,300 per day (\$40,600×16.45/126.8), which is a 17.3% decrease relative to the counterfactual earnings of \$30,550. We do not find any significant change in the fixtures for tankers classified as being at moderate risk by Windward.

Another useful measure to think about here is the aggregate value of the earnings drop for a sanctions-violating tanker. At these average charter rates, and with the 16.45 units disclosure-induced fixture drop, each tanker loses about \$5,300 per day. Assuming 200 employed days per year (conservatively), this amounts to a 1.06 million loss per year. Given that Windward's list disclosed 618 large tankers (Panamax, Aframax, Suezmax, and VLCC) to be high-risk (i.e., as in Table 4), this is a total haircut of about \$655 million in earnings for sanctions violators. This calculation does not include "Medium" vessels (as we do not have equivalent time charter rate numbers for these), and the actual aggregate earnings drop is therefore likely to be larger.

Finally, note that these results are not only robust to methods and bandwidth choice across specifications, but also remain similar if we use propensity scores from a logistic regression (rather than the baseline machine learning models), drop August 2023 (to allow about two weeks for the disclosure effect to come to force), drop Russia (to show that our effects are not coming solely from Russia-related changes in maritime activity around this time), and even if we bootstrap standard errors. These results are presented in Table A.2 in the Internet Appendix.

B. Tanker route changes

Here we examine the effect of the information shock on tanker routes. We hypothesize that after the information shock high/moderate-risk tankers are less likely to pass nearby sanctioned countries, including Iran, Venezuela (only before 2024, because the U.S. lifted sanctions on Venezuela in 2024), and Russia. This hypothesis is based on the conjecture that suspect tankers know now that their movements will be under close scrutiny, and therefore they might be especially careful in avoiding any kind of behavior that might catalyze further suspicion.

We use AIS data to trace tankers' trajectories. We define the relevant boundaries for sanctioned countries using territorial seas, which extend to 12 nautical miles (nm) from the coast of a country. These boundaries are downloaded from the website of Marine Regions managed by the Flanders Marine Institute. To test the hypothesis, the dependent variable is a zero-one indicator of AIS signals being emitted within the 12 nm boundaries of Iran, Venezuela (only before 2024), or Russia each month. Similar to the test on fixtures, we first use PSM-DiD to examine the times series of treatment effects for high-risk tankers only via the following regression:

$$near_sanctioned_countries_{i,k,t} = \left[\sum_{l=-6}^{7} \beta_l^H \times high_risk_i \times \mathbb{I}_{\{t=l\}}\right] + \alpha_i + \gamma_{k,t} + \epsilon_{i,k,t}, \quad (3)$$

where α_i is tanker fixed effect, and $\gamma_{k,t}$ is time×tanker-type fixed effect. Figure 13 plots the estimated coefficients $\{\beta_l^H\}_{l=-6}^{-1}$, which can be interpreted as the changes in the probability of passing nearby sanctioned countries. The coefficients in the pre-period are all insignificant at the 95% confidence level, which suggests a lack of pre-trends. In the post-period, there is a gradual drop that reaches around 10 percentage points at the end of our sample. This empirical pattern is consistent with our hypothesis that high-risk tankers avoid getting close to the sanctioned countries after Windward's disclosure.

In Table 6, we present the estimation of ATT on the route change by combining post-

period months into a single dummy and estimating separately for high-risk and moderate-risk tankers. The regression specification for estimating ATT on high-risk tankers after removing the moderate-risk tankers from the sample is:

$$near_sanctioned_countries_{i,k,t} = \beta^H \times high_risk_i \times \mathbb{I}_{\{t \geq 0\}} + \alpha_i + \gamma_{k,t} + \epsilon_{i,k,t}. \tag{4}$$

The specification for the analysis of moderate-risk tankers is similar. As tanker AIS signals are much more frequent than tanker fixtures, we have more observations for this route test than the previous fixture test. The results using PSM-DiD in panel A show that the high-risk tankers are around six percentage points less likely to pass within 12 nm of the sanctioned countries, while the effect on moderate-risk tankers is insignificant. For the magnitude of the effect, note that the average probability of showing signals in those areas is 40 percentage points for high-risk tankers in July, which corresponds to a 17% (6.8/40) decrease for the average effect in the post-period. According to Figure 13, this magnitude is larger towards the end of our sample, e.g., 25% (10/40). In panel B, the estimation using Abadie (2005) semiparametric DiD gives similar results. Overall, the results show that suspect tankers change their navigation routes after being disclosed by Windward.

C. Tanker ownership changes

In this subsection, we examine how conditions in the market for used tankers change after the information shock. We might expect high/moderate-risk tankers to become more difficult to sell in the used tankers market, since various sanctioning regimes explicitly prohibit transactions involving violating entities. For example, Council Regulation (EU) 833/2014 prohibits EU nationals, residents, or entities from directly or indirectly selling or otherwise transferring ownership of tankers falling under Harmonized System (HS) Code 8901 20 for the transport of crude oil or petroleum products listed in Annex XXV of this regulation, see Skadden, 2024.

We use a proprietary set of tanker ownership data (from Drewry), which shows the tankers' owner name and country in six snapshots – as of December 2022, March, July, September, and December 2023, and March 2024. We hypothesize that after Windward's list is released on Reuters, high/moderate-risk tankers become less liquid as shipowners are reluctant to buy them. Note that the time spans between consecutive snapshots are different; so we make them comparable by annualizing the ownership change variable within each snapshot. After making such adjustments, we use both PSM-DiD and Abadie (2005) semiparametric DiD to estimate the treatment effect, as before. The regression to examine

the time series of ATT is:

$$owner_change_{i,k,p} = \left[\sum_{l=-2}^{2} \beta_l^H \times high_risk_i \times \mathbb{I}_{\{p=l\}}\right] + \alpha_i + \gamma_{k,p} + \epsilon_{i,k,p}, \tag{5}$$

and the regression to estimate the average treatment effect in the post-period is:

$$owner_change_{i,k,p} = \beta^H \times high_risk_i \times \mathbb{I}_{\{p\geq 0\}} + \alpha_i + \gamma_{k,p} + \epsilon_{i,k,p}, \tag{6}$$

where *owner_change* equals one if the owner names are different across two snapshots and zero otherwise, and p denotes the period between two adjacent snapshots of the data. The specification for the analysis of moderate-risk tankers is similar. The results are shown in Figure 14 and in Table 7. Panel A of that table shows results from PSM-DiD estimation which indicate that after the information shock, high-risk (moderate-risk) tankers are 17.5-18.4 (16-16.8) percentage points less likely to change owners. Relative to a pre-disclosure annualized turnover of 45.4 percentage points for high-risk tankers, this is a 40% (18/45.4) drop in liquidity. In panel B, the results from Abadie's DiD suggest a similarly significant drop in the probability of tanker owner change. Overall, this suggests that the risk disclosure from Windward makes high/moderate-risk tankers less desirable, reducing their turnover.

D. Charterer identity

In this subsection, we examine whether the U.S. or U.S.-allied companies stop chartering tankers classified as high/moderate-risk by Windward. The charterers' names are provided in the fixture data and we manually search these companies' headquarters via LinkedIn or their websites to determine their countries. We define the U.S.-allied charterers as companies from the UK, European Union, Australia, New Zealand, or Japan. The dependent variable here is at the charterer-month level – it is the number of tankers of a given risk type that each charterer employs in a month. The regression specification is

$$\begin{aligned} n_tankers_{h,t} = \left[\beta^U \times US_charterer_h + \beta^A \times US_allied_charterer_h\right] \times \mathbb{I}_{\{t\geq 0\}} \\ + \alpha_h + \gamma_t + \epsilon_{h,t}, \end{aligned} \tag{7}$$

where h denotes charterer. Consistent with our previous results, we again find a significant drop in the usage of high-risk tankers by U.S. charterers. While the effects from U.S.-allied charterers are not statistically significant here.

To put the economic magnitudes of these results in perspective, the number of tankers, later disclosed as high-risk, employed by the average U.S. charterer in July 2023 is 0.652.

The estimated 0.35 drop therefore implies a 53.7% (0.35/0.652) reduction in usage. In the Internet Appendix F, we list the U.S. charterers to examine their usage of high-risk tankers individually. However, we don't find clear patterns for the drop in high-risk tanker usage (e.g., private company vs. public company).

E. Russian oil trade

In this subsection, we examine the effect of Windward's disclosure on the Russian oil trade. We wish to verify whether risky tankers classified by Windward move less Russian oil as they might be careful in avoiding further suspicion. Refinitiv Eikon provides oil trade flow data at the tanker level (i.e., volumes from a load port to a discharge port). We select the trade flows that start from Russia and end in other countries with departure times between February 2023 and March 2024, similar to our previous tests. This yields a sample of 392 high-risk, 375 moderate-risk, and 47 low-risk tankers. We pool together low- and moderate-risk tankers to construct a control group for the high-risk tankers. As previously, we use PSM-DiD to examine the times series of treatment effects for high-risk tankers via the following regression:

$$Russian_oil_{i,k,t} = \left[\sum_{l=-6}^{7} \beta_l^H \times high_risk_i \times \mathbb{I}_{\{t=l\}}\right] + \alpha_i + \gamma_{k,t} + \epsilon_{i,k,t}, \tag{8}$$

and the regression to estimate the average treatment effect in the post-period is:

$$Russian_oil_{i,k,t} = \beta^H \times high_risk_i \times \mathbb{I}_{\{t\geq 0\}} + \alpha_i + \gamma_{k,t} + \epsilon_{i,k,t}. \tag{9}$$

The dependent variable $Russian_oil_{i,k,t}$ in the above equations is the total volume of crude oil and fuel oil carried by tanker i's (with tanker type k) in month t. To account for both the intensive and extensive margin, we set the value of the dependent variable to zero if there is no trade flow for a given tanker-month.

The results are shown in Table 9 and in Figure 15. Table 9 indicates that the average treatment effect on the total Russian oil trade from high-risk tankers is a reduction of approximately 2 million tonnes (5 kilotonnes $\times$ 392) per month relative to the constructed counterfactuals, which accounts for 7.5% (2/26.4) of the total volume of Russian oil exports in July 2023 (i.e, the month before Windward's disclosure). Figure 15 illustrates the treatment effects in each month, which shows a lack of pre-trends and a persistent effect until the end of our sample. The results indicate the real effect of Windward's disclosure on the Russian oil trade.

5.2.3 Selection into dataset

One question related to our treatment effects estimation could be whether there are potential selection issues: the information shock may affect the high/moderate-risk tankers' incentive to report their fixture contracts to the data vendor, from which we source our numbers. For example, if after the information becomes public, some high/moderate-risk tankers are reluctant to disclose their contracts, there could be a concern that our estimation is biased by the selection issues. We test this hypothesis by counting the number of fixture contracts per month for each tanker. We run the following regression for estimating the effect on high-risk tankers

$$n_fixture_records_{i,k,t} = \beta^H \times high_risk_i \times \mathbb{I}_{\{t \geq 0\}} + \alpha_i + \gamma_{k,t} + \epsilon_{i,k,t}, \tag{10}$$

where $n_fixture_records_{i,k,t}$ is the number of fixture records in our dataset for tanker i (with a tanker type k) in month t, and $n_fixture_records_{i,k,t}$ will be zero if there are no fixture records. The specification for the analysis of moderate-risk tankers is similar. To enter into the sample, we require a tanker to report at least one fixture record during our sample period, i.e., we don't include the tanker with all zero values of $n_fixture_records$. If there are selection issues, we will observe that the number of fixture contracts for high/moderate-risk tankers significantly decreases if they want to avoid disclosing, after Windward publishes their list. The results in Table 10 show that all coefficients are both statistically and economically insignificant for either of our two methods. Thus, it suggests that the owners of high/moderate-risk tankers do not seem to change their reporting of fixture contracts to the dataset, so such selection issues do not seem to be important.

Note also that our ownership change data does not come from reports to Reuters, and is not affected by any such potential selection issues.

6 A Model of the Oil Shipping Market with Sanctions

In this section, we present a simple model of oil shipping with two goals. First, we use the model to shed light on a possible mechanism that reproduces in equilibrium the empirical patterns we have documented, with specific attention to the differences in the likely impact of the disclosure event on actual sanctions violators and sanctions-complying tankers. While we have risk labels for different tankers, for example from Windward, we do not observe who actually violated sanctions in the data (outside of our Iranian dataset). Second, we also use the model to understand what was the impact of this disclosure on costs (and therefore profits) of sanctioned oil producers (or charterers working for them). Again, this is something

that we cannot observe in the data.

6.1 The model

6.1.1 Environment

A. Players

We consider the oil shipping sector consisting of two main groups of players: oil tankers and oil exporters. We distinguish between oil tankers that have previously violated sanctions by carrying sanctioned oil (denoted 'Bad' or B) and those that have never violated sanctions before (denoted 'Good' or G). We also distinguish between exporters of sanctioned and non-sanctioned oil and denote them, respectively, 'Rogue' and 'Clean'.

Importantly, being a 'Bad' tanker does *not* imply that it has actually been sanctioned and put on a designated list - it only implies that this tanker has already breached sanctions in the past and can be added to such a list at any time. Furthermore, a 'Good' tanker will turn irreversibly into 'Bad' if it engages with a 'Rogue' exporter and carries their sanctioned oil. In turn, a 'Clean' exporter may itself fall under secondary sanctions if it is employing a 'Bad' tanker at a time when the tanker is put on a sanctioned list. Note that even if a sanctions clause is included in the contract with the ship owner, such an exporter can still face disruptions in its operations and additional costs, including those related to litigation (e.g., Reed Smith, 2024).

B. Information structure

The model assumes the following information structure: The 'Rogue' exporters know which tankers are 'Bad', because they have employed previously these tankers. The 'Clean' exporters, on the other hand, only know that such 'Bad' tankers exists. 'Clean' exporters are also supposed to do their due diligence to avoid engaging with 'Bad' tankers and thus violating sanctions. To do that, the 'Clean' exporters have access to some detection technology which generates a binary signal and assigns label Low-risk or High-risk for any tanker. The tankers know of course if they are 'Bad'. They also know their risk label, or can deduce it from the prices (freight rates) that 'Clean' exporters quote to them. The 'Rogue' exporters do not know these labels and respective price quotes of the 'Clean' exporters. It is common knowledge that a proportion λ of all tankers are 'Bad'.

C. Tanker classification and sanction probabilities

More formally, there are λ 'Bad' (B) tankers and $1-\lambda$ 'Good' (G) tankers. There are also λ High-risk (H) tankers and $1-\lambda$ Low-risk (L) tankers. The correlation between tanker true

type (B or G) and risk label (H or L) is $\rho \in (0,1)$, which is a measure of the precision of the detection technology employed by the 'Clean' exporters. Using this correlation we define the following conditional probabilities

$$\theta_B \equiv \Pr[H|B] = \lambda + (1-\lambda)\rho, \quad \theta_G \equiv \Pr[H|G] = \lambda(1-\rho),$$

from which it also follows that

$$\Pr[B|H] = \theta_B, \quad \Pr[B|L] = \theta_G.$$

The unconditional probability of a tanker being sanctioned due to past sanction violations is w_0, and the probabilities conditional on risk label are w_L and w_H. The unconditional probability for a G tanker to be sanctioned for carrying sanctioned oil in the current-period is δ_0 (δ_L and δ_H for L and H tankers). A B tanker that currently carries sanctioned oil faces a cumulative probability $w_0 + \delta_0$ of being sanctioned (or $w_L + \delta_L$ and $w_H + \delta_H$ for L and H tankers). By definition, we have

$$\lambda w_H + (1-\lambda) w_L = w_0, \quad \lambda \delta_H + (1-\lambda)\delta_L = \delta_0.$$

Because H tankers could be under more scrutiny relative to L tankers, we assume that the conditional sanction probabilities are proportional to $\Pr[B|H]$ and $\Pr[B|L]$:

$$w_H / w_L = \delta_H / \delta_L = \theta_B / \theta_G,$$

which gives $w_H = \frac{\theta_B}{\lambda} w_0$, $w_L = \frac{\theta_G}{\lambda} w_0$, $\delta_H = \frac{\theta_B}{\lambda}\delta_0$, and $\delta_L = \frac{\theta_G}{\lambda}\delta_0$.

D. Timeline

We assume the following sequence of events each period:

1. Risk labels are assigned to tankers by 'Clean' exporters

2. An aggregate mean-zero shock ϵ to the oil trading profits of all oil exporters is realized. Besides, a tanker-specific shock $\tilde{c}_i$ for each tanker i is realized, which affects tanker costs, especially when dealing with 'Rogue' exporters

3. 'Clean' and 'Rogue' exporters simultaneously choose prices (i.e., freight rates) for the tankers that they can distinguish, given their information. That is, p_L and p_H are given by 'Clean' exporters to L and H tankers, and p_G^R and p_B^R are given by 'Rogue' exporters to G and B tankers, respectively. Hereafter, a superscript R denotes 'Rogue'

exporters.

4. Tankers observe the prices (freight rates) given to them and choose whether to deal with 'Clean' or 'Rogue' exporters

5. Each player generates their payoff and moves to the next period

6.1.2 Exporters' optimization

'Rogue' exporters choose prices to maximize their oil trade profits, which have an average value of r^R and also depend on a random shock ϵ each period - the shock can reflect the pronounced variability in the oil price and also in various components of the cost of extracting oil. We assume ϵ to be i.i.d. uniformly distributed in $[-\sigma, \sigma]$. The supply of tankers that 'Rogue' exporters face depends on tanker type and risk label. These exporters also face a penalty, denoted z^R - in an extreme case, one can think of z^R as the value of the oil cargo that can be impounded or even confiscated. A penalty is incurred when the employed tanker is sanctioned, and hence with the same probabilities.

Denoting by $Q^R_{i,j}$ the supply of tankers of type i and risk label j that engage with 'Rogue' exporters, it follows that these exporters solve

$$\max_{p^R_G, p^R_B} \{Q^R_{GL}(r^R + \epsilon - p^R_G - \delta_L z^R) + Q^R_{GH}(r^R + \epsilon - p^R_G - \delta_H z^R) \\ + Q^R_{BL}[r^R + \epsilon - p^R_B - (\delta_L + w_L) z^R] + Q^R_{BH}[r^R + \epsilon - p^R_B - (\delta_H + w_H) z^R]\}.$$

'Clean' exporters also choose prices (i.e., freight rates) to maximize their oil trade profits, with an average r and depending on the same aggregate shock ϵ as for the 'Rogue' exporters. 'Clean' exporters can incur a sanctions penalty (denoted z) only if employing a B tanker, which happens with the probability w_L or w_H as before. They solve the following optimization problem

$$\max_{p_L, p_H} \{Q_{GL}(r + \epsilon - p_L) + Q_{GH}(r + \epsilon - p_H) \\ + Q_{BL}(r + \epsilon - p_L - w_L z) + Q_{BH}(r + \epsilon - p_H - w_H z)\}.$$

6.1.3 Tankers' optimization

We model tankers as price-takers – they observe the prices (freight rates) quoted to them by 'Clean' or 'Rogue' exporters and then decide which of these exporters to deal with. Let V_G and V_B denote the tanker values before the assignment of risk labels, and $V_{GL}, V_{GH}, V_{BL}, V_{BH}$

denote the values conditional on a risk label. The Bellman equations for tankers are

$$
\begin{aligned}
V_{GL} &= \mathbb{E}\left[\max\begin{Bmatrix} p_L + \beta V_G \\ p_G^R - \tilde{c}_i - \delta_L s + \beta V_B \end{Bmatrix}\right] \\
V_{GH} &= \mathbb{E}\left[\max\begin{Bmatrix} p_H + \beta V_G \\ p_G^R - \tilde{c}_i - \delta_H s + \beta V_B \end{Bmatrix}\right] \\
V_{BL} &= \mathbb{E}\left[\max\begin{Bmatrix} p_L - w_L s + \beta V_B \\ p_B^R - \tilde{c}_i - (w_L + \delta_L) s + \beta V_B \end{Bmatrix}\right] \\
V_{BH} &= \mathbb{E}\left[\max\begin{Bmatrix} p_H - w_H s + \beta V_B \\ p_B^R - \tilde{c}_i - (w_H + \delta_H) s + \beta V_B \end{Bmatrix}\right]
\end{aligned}
$$

where

$$V_G = \theta_G V_{GH} + (1 - \theta_G) V_{GL}, \quad V_B = \theta_B V_{BH} + (1 - \theta_B) V_{BL}$$

When dealing with 'Rogue' exporters, each tanker i incurs in each period an additional random cost $\tilde{c}_i$ that we assume to be i.i.d. uniformly distributed in $[0, \bar{c}]$ – this cost may stem from efforts to circumvent detection or difficulties in financing or other relations with third parties in the shipping sector that seek to avoid falling under secondary sanctions.

The penalty imposed on a tanker when sanctioned is s, which is endogenously determined by

$$s = \beta V_B.$$

We calculate values of the random cost $\tilde{c}_i$ at which tanker i would be indifferent between dealing with a 'Clean' or 'Rogue' exporter. For $\tilde{c}_i$ below (above) such a critical value the tanker prefers to deal with a 'Rogue' ('Clean') exporter. The critical values are given by

$$
\begin{aligned}
c_{GL} &= p_G^R - p_L - \delta_L s + \beta(V_B - V_G) \\
c_{GH} &= p_G^R - p_H - \delta_H s + \beta(V_B - V_G) \\
c_{BL} &= p_B^R - p_L - \delta_L s \\
c_{BH} &= p_B^R - p_H - \delta_H s,
\end{aligned}
$$

Using these critical values, we derive that

$$V_G = \frac{\theta_G \mathbb{E}\left[\frac{(c_{GH})^2}{2\bar{c}} + p_H\right] + (1-\theta_G)\mathbb{E}\left[\frac{(c_{GL})^2}{2\bar{c}} + p_L\right]}{1-\beta}$$
$$V_B = \frac{\theta_B \mathbb{E}\left[\frac{(c_{BH})^2}{2\bar{c}} + p_H\right] + (1-\theta_B)\mathbb{E}\left[\frac{(c_{BL})^2}{2\bar{c}} + p_L\right] - [\theta_B w_H + (1-\theta_B) w_L]s}{1-\beta}$$

6.1.4 Market clearing

The market clearing conditions for the 'Rogue' exporters' market are

$$Q^R_{GL} = (1-\lambda)(1-\theta_G)\left(\frac{c_{GL}}{\bar{c}}\right),$$
$$Q^R_{GH} = (1-\lambda)\theta_G\left(\frac{c_{GH}}{\bar{c}}\right),$$
$$Q^R_{BL} = \lambda(1-\theta_B)\left(\frac{c_{BL}}{\bar{c}}\right),$$
$$Q^R_{BH} = \lambda\theta_B\left(\frac{c_{BH}}{\bar{c}}\right)$$

and for the 'Clean' exporters' market these clearing conditions are

$$Q_{GL} = (1-\lambda)(1-\theta_G)\left(1-\frac{c_{GL}}{\bar{c}}\right),$$
$$Q_{GH} = (1-\lambda)\theta_G\left(1-\frac{c_{GH}}{\bar{c}}\right),$$
$$Q_{BL} = \lambda(1-\theta_B)\left(1-\frac{c_{BL}}{\bar{c}}\right),$$
$$Q_{BH} = \lambda\theta_B\left(1-\frac{c_{BH}}{\bar{c}}\right).$$

Further details on the model solution are in the Appendix.

6.2 Model calibration

A period in the model corresponds to one year in reality. We set the discount factor (β) to 0.9, which corresponds to a 20-year tanker life cycle and 10% scrap value. Based on the estimation in CRS, 2024 that 1600 tankers have carried sanctioned oil, we set the proportion of bad tankers (λ) to 0.114. We calculate the unconditional sanction probability ($w_0 = 0.5\%$) using the number of actually sanctioned tankers divided by the number of total tankers. We assume that 'Clean' and 'Rogue' exporters have the same mean oil trade profit (r and r^R), which we normalize to one. We set the maximum random cost $\tilde{c}_i$ for tankers when dealing with 'Rogue' exporters (i.e., $\bar{c}$) to one. We calibrate the classification precision with

which 'Clean' exporters can detect sanctions violating tankers by converting the Pseudo-R^2 of sanction predictability in Figure 11 to the correlation coefficient (ρ) in the model. The pseudo-R^2 based on public data is 7.5%, which implies that ρ_{pre} is 0.269, while the pseudo-R^2 obtained when including Windward's risk label is 17.2%, implying that ρ_{post} is 0.424.

We calibrate by moment matching the remaining four model parameters: sanction probability for current period violation (δ_0), sanction penalties (z and z^R), and volatility of oil trade profit (σ). We consider four moments: the volatilities of low-risk and high-risk tankers' fixtures, the disclosure effect on high-risk tankers' fixtures (from the DiD estimation in Table 5), and the disclosure effect on high-risk tankers' routes (from the DiD estimation in Table 6). The model analog for the disclosure effect on fixtures is the high-minus-low post-minus-pre percentage change in fixture prices, which is

$$\frac{p_H^{post} - p_H^{pre}}{p_H^{pre}} - \frac{p_L^{post} - p_L^{pre}}{p_L^{pre}}.$$

Similarly, the model analog for the disclosure effect on routes is a high-minus-low post-minus-pre percentage change in the probabilities of dealing with 'Rogue' exporters, which is

$$\frac{b_H^{post} - b_H^{pre}}{b_H^{pre}} - \frac{b_L^{post} - b_L^{pre}}{b_L^{pre}},$$

where the probabilities b_H and b_L are given by

$$b_H = \Pr[B|H] \times \Pr[\tilde{c}_i < c_{BH}] + \Pr[G|H] \times \Pr[\tilde{c}_i < c_{GH}] = \theta_B \frac{c_{BH}}{\bar{c}} + (1 - \theta_B)\frac{c_{GH}}{\bar{c}},$$
$$b_L = \Pr[B|L] \times \Pr[\tilde{c}_i < c_{BL}] + \Pr[G|L] \times \Pr[\tilde{c}_i < c_{GL}] = \theta_G \frac{c_{BL}}{\bar{c}} + (1 - \theta_G)\frac{c_{GL}}{\bar{c}}.$$

We present the parameter values and moment matching results in Table 11, which shows that the model can fit the data quite well. The empirical annualized volatility of low-risk and high-risk tankers' fixtures are 0.77 and 0.81, and the model analogs are 0.75 and 0.82. More importantly, the model can reproduce the disclosure effect on high-risk tankers' fixture changes (a 13% drop in the data vs. 16% in the model) and route changes (a 17% drop in the data vs. 12% in the model).

6.3 Model implications

We solve the model by discretizing the aggregate shock ϵ into grid. The key parameter is the classification precision, ρ, which captures the information shock from Windward's disclosure. For an easier match to the empirical results, we translate ρ to pseudo-R^2 and plot the model outcomes (average across all discrete values of ϵ) as functions of pseudo-R^2 in Figure 16.

The two plots in the first row show that the model reproduces the disclosure effects on high-risk tankers' fixtures and routes, that were shown in Tables 5 and 6. As information precision increases, the discount for H tankers' fixtures relative to those for L tankers increases. This is because the 'H' label now indicates a higher probability of a bad tanker, and a 'Clean' exporter is less willing to hire such tanker, leading to a lower equilibrium p_H. For the route results, as information precision increases, the probability of dealing with 'Rogue' exporters decreases more for H tankers than for L tankers, because H tankers are now under more intense scrutiny and face a higher probability of being sanctioned for dealing with 'Rogue' exporters.

The second row of Figure 16 shows the probabilities and numbers of tankers of different types (GL, GH, BL, and BH) that are dealing with 'Rogue' exporters. Importantly, the probability drop for H tankers comes mainly from the BH tankers. This is because not only BH tankers face higher cost of dealing with 'Rogue' exporters, but also these exporters are less willing to hire them. The number of tankers is the probability multiplied by the pool size for each type. Because the majority of tankers are GL, the drop in the number of tankers dealing with 'Rogue' exporters comes mainly from GL tankers.

The last row of Figure 16 sheds light on additional implications for 'Rogue' exporters, which we can only observe in the model. First, as information precision increases, the total export of 'Rogue' exporters (in number of tankers) decreases. Quantitatively, the information shock accounts for a 23% drop in 'Rogue' exporters' total export, i.e., from 0.199 to 0.153 when pseudo-R^2 increase from the pre-Windward to post-Windward precision. At the same time, the per-tanker spending of 'Rogue' exporters' increases by 10%, i.e, from 0.652 to 0.718. Taken together, these suggest that 'Rogue' exporters indeed are adversely impacted when a more precise information on tanker sanction risks is provided to the market.

7 Policy Implications: A Discussion

Our findings underscore the need for a shift in how sanction violations are approached. Currently, the burden falls disproportionately on third parties, who must navigate the murky waters of international commerce to identify and avoid sanctions violators. This approach is not only inefficient but also fraught with financial risks for these entities.

A more transparent system wherein the identities of suspected violators are disclosed by authoritative sources could streamline the process. Such an approach would not only alleviate the burden on third parties but also enhance the overall effectiveness of sanctions by directly penalizing and isolating violators, thus deterring future breaches.

One way in which sanctioning authorities can promote such transparency is by creating

a list put out on their website. The benefit of such an approach could be higher accuracy, if these authorities (typically governments) have access to better information, e.g., better satellites. Of course, there could be two downsides to such an approach: first, governments might not want to publicly reveal their surveillance capabilities. Second, there could be legal risks for these governments if they incorrectly identify a compliant actor as a violator, and thereby hurt that actor's reputation and business.

An alternative approach could be for sanctioning authorities to pay private companies like Windward.ai a fee to put up these risk scores publicly on the sanctioning authorities' website. Thinking about such a fee – and therefore the value of disclosure – through the lens of economic transfers makes the benefits of such transparency clearer. In our example, the reduction in freight rates alone for violators following Windward's disclosure is about $620 million a year, whereas the total market value of Windward.ai was only $47 million (GBP 37 million, as of August 16, 2023 – the day the list was put up). This suggests that the cost imposed on violators by such disclosure is an order of magnitude higher than what the market is willing to pay for such information – underscoring the need for sanctioning authorities to rethink risk information dissemination.

A third approach could be to create a regulatory framework in which these lists are not shared by the sanctioning authorities, but guidance explicitly adds expectations for economic agents to incorporate sanctions risk scores in compliance programs. In essence, this could be like the use of credit ratings for bonds in financial markets. For example, authorities can require banks to calculate a dollar weighted average sanctions-risk score for their loan portfolios, and mandate such a score to be a factor in capital reserving. This would imply that if a bank lends more to borrowers who have a higher risk of having violated sanctions, then they would have to keep more capital aside.

Of course, one might question the need for any regulation at all – why should the market by itself not end up in an equilibrium where everyone uses Windward-type risk scores? While it is possible that such an equilibrium can arise in time, the process to get there might be too slow. Geo-political expediency of sanctions enforcement suggests that a little help from authorities to nudge the market to do so faster through regulation might help.

Overall, while many approaches might be available to authorities to alleviate the problem third parties face in identifying violators, including those listed above – our evidence makes it clear that something needs to be done urgently.

8 Conclusion

Focusing on the international oil shipping industry, in this paper we examine whether the difficulty of identifying sanctions violators prevents counterparties from avoiding them. We study how oil shipping sector actors – and therefore the economics of sanctions compliance – could change if they are exogenously provided information identifying suspected violators.

Overall, we show that it is hard to identify tankers that might have violated sanctions using readily available data, even for sophisticated players. But when a list of suspect tankers is made widely available by Reuters, the high-risk tankers immediately see a drop of 15.44 WS units (benchmark of WS is 100) in their fixture (rental) rates, corresponding to a 16.4% drop in their earnings, and become much less liquid as assets in the used tanker market.

Our evidence ultimately indicates that if sanctioning bodies or other geopolitical stakeholders were to provide clear and accessible information about the identities of violators, rather than leaving it to market forces to uncover these entities on their own, they could effectively make sanctions violations significantly less profitable.

References

Abadie, A. (2005), 'Semiparametric Difference-in-Differences Estimators', *The Review of Economic Studies* **72**(1), 1–19.

Aumann, R. J. (1976), 'Agreeing to disagree', *The Annals of Statistics* **4**(6), 1236–1239.

Babina, T., Hilgenstock, B., Itskhoki, O., Mironov, M. and Ribakova, E. (2023), 'Assessing the impact of international sanctions on Russian oil exports', *Available at SSRN* .

Baldwin, D. (1985), *Economic Statecraft*, Princeton University Press.

Brancaccio, G., Kalouptsidi, M. and Papageorgiou, T. (2020), 'Geography, transportation, and endogenous trade costs', *Econometrica* **88**(2), 657–691.

Brancaccio, G., Kalouptsidi, M. and Papageorgiou, T. (2023), 'The impact of oil prices on world trade', *Review of International Economics* **31**(2), 444–463.

Brown, P. (2020), *Oil market effects from US economic sanctions: Iran, Russia, Venezuela*, Vol. 3, Congressional Research Service Washington, DC, USA.

Clayton, C., Maggiori, M. and Schreger, J. (2023), A framework for geoeconomics, Working Paper 31852, National Bureau of Economic Research.

Crozet, M., Hinz, J., Stammann, A. and Wanner, J. (2021), 'Worth the pain? Firms' exporting behaviour to countries under sanctions', *European Economic Review* **134**.

Dellavigna, S. and La Ferrara, E. (2010), 'Detecting illegal arms trade', *American Economic Journal: Economic Policy* **2**(4), 26–57.

Deloitte (2015), Managing sanctions compliance is complex. Are you up to the challenge?, Deloitte Flash Survey Series.

Dimmock, S. G. and Gerken, W. C. (2012), 'Predicting fraud by investment managers', *Journal of Financial Economics* **105**(1), 153–173.

Dizaji, S. F. and van Bergeijk, P. A. G. (2013), 'Potential early phase success and ultimate failure of economic sanctions: A VAR approach with an application to Iran', *Journal of Peace Research* **50**, 721–736.

Duggan, M. and Levitt, S. D. (2002), 'Winning isn't everything: Corruption in sumo wrestling', *American Economic Review* **92**(5), 1594–1605.

Early, B. R. and Preble, K. (2019), 'Enforcing Economic Sanctions: Analyzing How OFAC Punishes Violators of U.S. Sanctions', *Working Paper* .

Eaton, J. and Engers, M. (1992), 'Sanctions', *Journal of Political Economy* **100**, 899–928.

Eaton, J. and Engers, M. (1999), 'Sanctions: Some simple analytics', *The American Economic Review* **89**, 409–414.

Efing, M., Goldbach, S. and Nitsch, V. (2023), 'Freeze! Financial sanctions and bank responses', *The Review of Financial Studies* **36**, 4417–4459.

Farzanegan, M. R. (2011), 'Oil revenue shocks and government spending behavior in Iran', *Energy Economics* **33**(6), 1055–1069.

FCA (2024), Handbook of Rules and Guidance, Financial Crime Guide, Box 2.2.4., Financial Conduct Authority, www.handbook.fca.org.uk/handbook/fcg/2/2.html.

Fisman, R. J., Marcolongo, G. and Wu, M. (2024), 'The undoing of economic sanctions: Evidence from the Russia-Ukraine conflict', *Working Paper* .

Fisman, R. and Wang, Y. (2017), 'The Distortionary Effects of Incentives in Government: Evidence from China's "Death Ceiling" Program', *American Economic Journal: Applied Economics* **9**.

Fisman, R. and Wei, S.-J. (2004), 'Tax rates and tax evasion: evidence from "missing imports" in China', *Journal of Political Economy* **112**(2), 471–496.

Fisman, R. and Wei, S.-J. (2009), 'The smuggling of art, and the art of smuggling: Uncovering the illicit trade in cultural property and antiques', *American Economic Journal: Applied Economics* **1**(3), 82–96.

Gordon, J. (2016), 'Extraterritoriality: Issues of Overbreadth and the Chilling Effect in the Cases of Cuba and Iran', *Harvard International Law Journal* **57**, Online.

Griffin, J. and Kruger, S. (2023), 'What is Forensic Finance?', *Available at SSRN 4490028* .

Griffin, J. M. and Maturana, G. (2016), 'Who facilitated misreporting in securitized loans?', *The Review of Financial Studies* **29**(2), 384–419.

Griffin, J. M. and Shams, A. (2018), 'Manipulation in the VIX?', *The Review of Financial Studies* **31**(4), 1377–1417.

Haidar, J. I. (2017), 'Sanctions and export deflection: evidence from Iran', *Economic Policy* **32**, 319–355.

Hamilton, J. (1983), 'Oil and the macroeconomy since World War II', *Journal of Political Economy* **38**, 228–248.

Heckman, J. J., Ichimura, H. and Todd, P. E. (1997), 'Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme', *The Review of Economic Studies* **64**(4), 605–654.

Hong, H. and Stein, J. (1999), 'A unified theory of underreaction, momentum trading, and overreaction in asset markets', *Journal of Finance* **54**, 2143–2184.

Hong, H., Torous, W. and Valkanov, R. (2007), 'Do industries lead stock markets?', *Journal of Financial Economics* **83**, 367–396.

Hong, H. and Yogo, M. (2012), 'What does futures market interest tell us about the macroeconomy and asset prices?', *Journal of Financial Economics* **105**, 473–490.

Hsieh, C.-T. and Moretti, E. (2006), 'Did Iraq cheat the United Nations? Underpricing, bribes, and the oil for food program', *The Quarterly Journal of Economics* **121**, 1211–1248.

Hufbauer, G. C., Schott, J. J. and Elliott, K. A. (1990), *Economic sanctions reconsidered: History and current policy*, Vol. 1, Peterson Institute.

Huneke, M. (2023), The Blind Men and the Elephant, Hughes Hubbard & Reed.

Huynh, L. D. T., Hoang, K. and Ongena, S. (2023), 'The impact of foreign sanctions on firm performance in Russia', *Swiss Finance Institute Research Paper* (23-115).

IMO (2000), Adoption of Amendments to the International Convention for the Safety of Life at Sea, 1974, International Maritime Organisation, Maritime Safety Committee.

Itskhoki, O. and Mukhin, D. (2022), Sanctions and the exchange rate, Technical report, National Bureau of Economic Research.

Jacob, B. A. and Levitt, S. D. (2003), 'Rotten apples: An investigation of the prevalence and predictors of teacher cheating', *The Quarterly Journal of Economics* **118**(3), 843–877.

Jones, L. (2015), *Societies under siege: Exploring how international economic sanctions (do not) work*, Oxford University Press, USA.

Kaplow, L. (1990), 'Optimal Deterrence, Uninformed Individuals, and Acquiring Information about Whether Acts Are Subject to Sanctions', *The Journal of Law, Economics, and Organization* **6**(1), 93–128.

Kilpatrick, R. (2022), 'Monitoring sanctions compliance at sea', *Northwestern Journal of International Law and Business* **42**, 221–251.

KPMG (2023), Complexity is the biggest challenge in sanctions compliance, KPMG forensic sanctions survey results.

Lacy, D. and Niou, E. M. S. (2004), 'A theory of economic sanctions and issue linkage: The roles of preferences, information, and threats', *The Journal of Politics* **66**(1), 25–42.

Lang, A. (2023), 'Alternatives to adjudication in international law: A case study of the ombudsperson to the ISIL and Al-Qaida sanctions regime of the UN Security Council', *American Journal of International Law* **117**(1), 48–91.

Lastauskas, P., Proškutė, A. and Žaldokas, A. (2023), 'How do firms adjust when trade stops?', *Journal of Economic Behavior & Organization* **216**, 287–307.

Liberti, J. M., Seru, A. and Vig, V. (2016), 'Information, Credit, and Organization', Working paper, SSRN.

Mamonov, M. and Pestova, A. (2023), 'The price of war: Macroeconomic and cross-sectional effects of sanctions on Russia'.

Marion, J. and Muehlegger, E. (2008), 'Measuring illegal activity and the effects of regulatory innovation: Tax evasion and the dyeing of untaxed diesel', *Journal of Political Economy* **116**(4), 633–666.

Moiseienko, A. (2023), 'Crime and sanctions: Beyond sanctions as a foreign policy tool', *German Law Journal, Cambridge University Press* pp. 1–30.

Morgan, T. C., Syropoulos, C. and Yotov, Y. V. (2023), 'Economic sanctions: Evolution, consequences, and challenges', *Journal of Economic Perspectives* **37**, 3–30.

Mulder, N. (2022), 'The Sanctions Weapon', *Finance & Development June 2022* .

OFAC (2019b), OFAC Advisory to the Maritime Petroleum Shipping Community (Iran), U.S. Department of the Treasury, Office of Foreign Assets Control.

OFAC (2020), Sanctions Advisory for the Maritime Industry, Energy and Metal Sectors, and Related Communities, U.S. Department of the Treasury, Office of Foreign Assets Control.

OFAC (2023b), Know Your Cargo: Reinforcing Best Practices to Ensure the Safe and Compliant Transport of Goods in Maritime and Other Forms of Transportation, U.S. Department of the Treasury, Office of Foreign Assets Control.

OFAC (2024), Obligations of foreign-based persons to comply with U.S. sanctions and export control laws, U.S. Department of the Treasury, Office of Foreign Assets Control.

OFSI (2022), OFSI enforcement and monetary penalties for breaches of financial sanctions Guidance, HM Treasury, Office of Financial Sanctions Implementation.

Pape, R. (1997), 'Why economic sanctions do not work', *International Security* **22**, 90–136.

Pape, R. (1998), 'Why economic sanctions still do not work', *International Security* **23**, 66–77.

Peress, J. (2014), 'The media and the diffusion of information in financial markets: Evidence from newspaper strikes', *the Journal of Finance* **69**(5), 2007–2043.

Posner, E. A. and Sykes, A. O. (2012), *Economic foundations of international law*, Harvard University Press.

Ready, R. C. (2018), 'Oil consumption, economic growth, and oil futures: The impact of long-run oil supply uncertainty on asset prices', *Journal of Monetary Economics* **94**, 1–26.

Rosenbaum, P. R. and Rubin, D. B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* **70**(1), 41–55.

Siddiquee, S. H. M. and van Bergeijk, P. (2012), Reconsidering Economic Sanctions Reconsidered: Detailed Analysis of the Peterson Institute Sanction Database, Technical report, ISS Working paper, Erasmus University: Rotterdam.

S&P Global (2022), Banks face hidden sanctions risk amid complex correspondent banking system, Market Intelligence.

Weng, J., Li, G. and Zhao, Y. (2022), 'Detection of abnormal ship trajectory based on the complex polygon', *Journal of Navigation* **75**(4), 966–983.

Wolsing, K., Roepert, L., Bauer, J. and Wehrle, K. (2022), 'Anomaly Detection in Maritime AIS Tracks: A Review of Recent Approaches', *Journal of Marine Science and Engineering* **10**, 1–19.

Table 1: Summary statistics of detection samples

This table reports the summary statistics for the detection samples in Section 4. We separately detect spoofing and dark activity. Panel A reports the sample sizes and panel B reports the mean and standard deviation of predictors. The construction of the predictors is described in Section 4.1.1 and 4.2.1. "-" in panel B indicates that the sample does not include the respective predictor.

Panel A: Sample Size

	Spoofing Detection	Dark Detection
No. observations	17,496 (tanker-day)	1,552 (dark periods)
No. unique tankers	1,481	746
No. unique violators	17	20

Panel B: Predictors

	Spoofing Detection		Dark Detection	
	Mean	Std	Mean	Std
Identity Change	0.118	0.322	0.144	0.352
Risky Flag	0.263	0.440	0.288	0.453
Ship-to-ship Transfer				
• anchor 10 km away from ports	0.218	0.413	0.363	0.481
• anchor 20 km away from ports	0.081	0.273	0.137	0.344
• draft change 10 km away from ports	0.036	0.186	0.151	0.358
• draft change 20 km away from ports	0.027	0.161	0.119	0.324
Irregular Trajectory				
• distance cutoff = 1 km	0.050	0.218	0.053	0.225
• distance cutoff = 2 km	0.050	0.218	0.068	0.251
• distance cutoff = 5 km	0.050	0.218	0.079	0.269
DBSCAN Outlier				
• distance parameter = 10 km	0.131	0.338	0.318	0.466
• distance parameter = 20 km	0.049	0.216	0.142	0.349
Dark Activity				
• at least 2-day dark period	-	-	0.406	0.491
• at least 3-day dark period	-	-	0.244	0.429
Satellite Images				
• 1-day window	0.010	0.101	-	-
• 3-day window	0.028	0.165	-	-
• 5-day window	0.042	0.200	-	-
• during dark period	-	-	0.025	0.157

Table 2: Performance of ML detection models

The hyper-parameters of the models are optimally determined by 10-fold cross-validation and reported in Internet Appendix Table A-1. Panel A (Panel B) reports the model performance using only AIS data (both AIS and satellite data). We separately report the detection of spoofing and dark activity, and "total" reports the simple average of these two. The implied false positive rate is calculated from the pseudo-R^2, with $\bar{p} = 2.23\%$ as estimated from our data and a risk level $1 - \beta = 5\%$ (as per Internet Appendix B). Panel C illustrates the importance of individual categories of predictors. Δpseudo-R^2 for a category is the difference between the pseudo-R^2 of a full model, as those reported in panel B, and the same model that sets this particular category to zero (note that all our predictors are zero-one indicators). When calculating Δpseudo-R^2, we treat negative pseudo-R^2 or negative Δpseudo-R^2 as zero. We report the simple averages of the respective Δpseudo-R^2's for the tree and neural network models, spoofing and dark detection. The last column of this panel shows the average increase in the implied false positive rate when setting the respective category to zero.

Panel A: Model Performance Using Only AIS Data			
Detect	Model	Pseudo-R^2	Implied false positive rate
Total	Tree	**9.1%**	**52.9%**
	NN	**10.2%**	**49.8%**
Spoofing	Tree	6.7%	60.5%
	NN	7.2%	58.9%
Dark	Tree	11.5%	46.3%
	NN	13.3%	42.2%
Panel B: Model Performance Using AIS and Satellite Data			
Detect	Model	Pseudo-R^2	Implied false positive rate
Total	Tree	**15.2%**	**38.0%**
	NN	**16.3%**	**35.8%**
Spoofing	Tree	11.8%	45.7%
	NN	13.8%	40.9%
Dark	Tree	18.7%	31.6%
	NN	18.8%	31.4%
Panel C: Predictor Importance			
Category		ΔPseudo-R^2	ΔImplied false positive rate
Satellite Detection		6.2%	14.7%
Identity Change		3.7%	8.0%
Risky Flag		2.8%	6.1%
Irregular Trajectory		0.2%	0.3%
Ship-to-ship Transfer		6.2%	14.8%
DBSCAN Outlier		5.3%	12.2%
Dark Activity		9.4%	24.5%

Table 3: Comparison between Windward's disclosure and ML-model detection

The first two lines in this table show the total number of tankers that violated Iranian sanctions during 2021 January, as per our proprietary dataset, and how many among them are reported on Windward's list as high- or moderate-risk. The bottom panel shows the performance of our decision tree and neural network models. We use either only AIS-based predictors or combine them with predictors based on satellite data. The set of predictors are as described in Section 4. The optimal hyper-parameters of the models, e.g., tree depth or number of layers of a neural network, are determined via 10-fold cross-validation. We show separately the number of sanctions violating tankers that engage in spoofing and going dark that our models detect; we also report the total number of detections for both types of behavior – the totals can be smaller than the sum of the two types of behavior because a tanker can engage in both. Result are shown for different confidence levels ($cl = 99\%,\ 95\%,$ and 90%) – for example, under a 95% confidence level no more than 5% of complying tankers can be mislabelled as violators.

Total Violators:		33			
on Windward's list:		**27**			
			Number of violators detected		
Detect	Model	Predictors	$cl = 99\%$	$cl = 95\%$	$cl = 90\%$
Total	Tree	Satellite + AIS	**5**	**12**	**17**
	NN	Satellite + AIS	**5**	**13**	**15**
	Tree	Only AIS	3	8	14
	NN	Only AIS	0	5	13
Spoofing	Tree	Satellite + AIS	3	5	6
	NN	Satellite + AIS	4	5	6
	Tree	Only AIS	1	2	5
	NN	Only AIS	0	3	4
Dark	Tree	Satellite + AIS	3	9	14
	NN	Satellite + AIS	1	10	12
	Tree	Only AIS	2	6	10
	NN	Only AIS	0	4	11

Table 4: Summary statistics of Windward's tankers sample

This table reports the summary statistics for the sample of tankers we use in Section 5 to examine the effects of the disclosure event. Panel A reports sample sizes and tanker classification. Panel B reports on balance tests on tanker characteristics between treated (i.e., high- and moderate-risk) and control (i.e., low-risk) groups. To construct the control group, we conduct propensity score matching (PSM) for each treated tanker within its tanker type, i.e., we require matched tanker(s) to have the same tanker type. The propensity score is calculated by regressing the high/moderate-risk tanker indicator on pre-period tanker characteristics using ML methods: average outputs from a decision tree and a neural network. The variables used to construct the propensity score are the same as the list in panel B. We calculate matching weights based on a Gaussian kernel with a bandwidth of 0.05. The variable differences and corresponding t-value are reported. For this tanker-level regression, standard errors are clustered at the flag level. Panel C reports sample averages of freight rates (or fixtures).

Panel A: Tanker sample

No. unique tankers	5,363			
	Sanction risk			
Tanker type	Low	Moderate	High	Total
Medium	1,579	423	306	2,308
Panamax	297	88	55	440
Aframax	533	295	280	1,108
Suezmax	309	179	167	655
VLCC	672	64	116	852
Total	3,390	1,049	924	5,363

Panel B: Tanker characteristics balance test

	Treated (High/Mod.-risk)	Control (Low-risk)	Diff.	t-value
Identity Change	0.169	0.165	0.003	[0.12]
Risky Flag	0.276	0.264	0.012	[0.24]
Ship-to-ship Transfer				
• anchor 10 km away from ports	2.652	2.733	-0.082	[-1.61]
• anchor 20 km away from ports	1.424	1.475	-0.051	[-0.67]
• draft change 10 km away from ports	1.620	1.668	-0.048	[-1.63]
• draft change 20 km away from ports	1.297	1.334	-0.037	[-0.98]
Irregular Trajectory				
• distance cutoff = 1 km	1.087	1.064	0.024	[0.40]
• distance cutoff = 2 km	1.426	1.414	0.012	[0.23]
• distance cutoff = 5 km	1.420	1.434	-0.014	[-0.27]
DBSCAN Outlier				
• distance parameter = 10 km	3.744	3.670	0.075	[1.33]
• distance parameter = 20 km km	3.023	2.946	0.077	[1.31]
Dark Activity				
• at least 2-day dark period	0.883	0.872	0.010	[0.27]
• at least 3-day dark period	0.600	0.589	0.011	[0.31]

Panel C: Average fixture

	Average fixture in Jul 2023 (one month before the event)		
Full sample	Low risk	Moderate risk	High risk
165.3	129.8	135.2	126.8

Table 5: Difference-in-differences analysis of fixture changes

This table reports the coefficients (i.e., average treatment effects on the treated, ATT) estimated with PSM-DiD and Abadie (2005)'s semiparametric DiD to examine the change in freight rates (or fixtures) for high/moderate-risk tankers after Windward's disclosure. The sample period is from Feb 2023 to Mar 2024. The post-period starts from Aug 2023. The propensity score is calculated by regressing the high/moderate-risk tanker indicator on pre-period tanker characteristics using ML methods: average outputs from a decision tree and neural networks. The variables used to construct the propensity score are shown in Table 4. In panel A, we match tankers within each tanker type and calculate weights based on propensity scores and a Gaussian kernel with a bandwidth of 0.01, 0.03, or 0.05. Then, we do DiD estimation in the matched sample, controlling for tanker fixed effects and time×tanker-type fixed effects. The standard errors in panel A are double clustered at the tanker and time×tanker-type levels. In panel B, we implement Abadie's method. We first collapse our sample into two cross-sections by calculating each tanker's pre- and post-period average fixtures and then use the Stata command "absdid", which needs to specify the dependent variable (i.e., fixture changes from pre- to post-period), treatment dummy (i.e., high/moderate-risk tanker indicator), and propensity score. Since Abadie's estimator is derived without directly accounting for macro trends through time fixed effects, we manually subtract the cross-sectional mean to account for time fixed effects or the cross-sectional mean by tanker type to account for time×tanker-type fixed effects. We trim the fixtures at the 1st and 99th percentiles each month to avoid the influence of outliers. The standard errors in panel B are as derived in Abadie (2005). *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

Panel A: PSM-DiD			
	Dependent variable: Fixtures		
	Bandwidth = 0.01	Bandwidth = 0.03	Bandwidth = 0.05
ATT for High Risk	-16.063***	-15.150**	-14.605**
	[-2.69]	[-2.39]	[-2.31]
Obs. (tanker-month)	5,378	5,574	5,574
ATT for Moderate Risk	-0.684	-1.197	-1.493
	[-0.17]	[-0.32]	[-0.41]
Obs. (tanker-month)	6,081	6,083	6,083
Time × Tanker Type FE	Yes	Yes	Yes
Tanker FE	Yes	Yes	Yes
Panel B: Abadie (2005) Semiparametric DiD			
	Raw fixtures	Fixtures demeaned by time	Fixtures demeaned by time × tanker type
ATT for High Risk	-18.417***	-20.709***	-13.749***
	[-3.56]	[-4.47]	[-3.16]
Obs. (tanker)	1,036	1,036	1,036
ATT for Moderate Risk	-3.736	-5.087	-0.507
	[-0.88]	[-1.40]	[-0.15]
Obs. (tanker)	1,190	1,190	1,190

Table 6: Difference-in-differences analysis of route changes

This table reports the coefficients (i.e., average treatment effects on the treated, ATT) estimated with PSM-DiD and Abadie (2005)'s semiparametric DiD to examine the change of the routes of high/moderate-risk tankers after Windward's disclosure. The dependent variable is a zero-one indicator of passing during each month within 12 nautical miles (i.e., the boundary of the territorial sea) of Iran, Russia, or Venezuela (only before 2024). The sample period is from Feb 2023 to Mar 2024. The post-period starts from Aug 2023. The propensity score is calculated by regressing the high/moderate-risk tanker indicator on pre-period tanker characteristics using ML methods. The variables used to construct the propensity score are shown in Table 4. In Panel A, we match tankers within each tanker type and calculate weights based on propensity scores and a Gaussian kernel with a bandwidth of 0.01, 0.03, or 0.05. Then, we do DiD estimation in the matched sample, controlling for tanker fixed effects and time×tanker-type fixed effects. The standard errors in panel A are double clustered at the tanker and time×tanker-type levels. In Panel B, we implement Abadie's method in the same way mentioned in Table 5. The standard errors in panel B are as derived in Abadie (2005). *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

Panel A: PSM-DiD			
	Dependent variable: Showing AIS signals near sanctioned countries		
	Bandwidth = 0.01	Bandwidth = 0.03	Bandwidth = 0.05
ATT for High Risk	-0.073***	-0.072***	-0.071***
	[-3.69]	[-3.76]	[-3.79]
Obs. (tanker-month)	39,246	39,246	39,246
ATT for Moderate Risk	-0.016	-0.015	-0.015
	[-1.01]	[-1.02]	[-1.05]
Obs. (tanker-month)	43,820	43,820	43,820
Time × Tanker Type FE	Yes	Yes	Yes
Tanker FE	Yes	Yes	Yes
Panel B: Abadie (2005) Semiparametric DiD			
	Probability	Probability demeaned by time	Probability demeaned by time × tanker type
ATT for High Risk	-0.067***	-0.067***	-0.060***
	[-4.01]	[-4.03]	[-3.60]
Obs. (tanker)	2,543	2,543	2,543
ATT for Moderate Risk	-0.012	-0.012	-0.008
	[-0.97]	[-0.96]	[-0.62]
Obs. (tanker)	3,139	3,139	3,139

Table 7: Difference-in-differences analysis of ownership changes

This table reports the coefficients (i.e., average treatment effects on the treated, ATT) estimated with PSM-DiD and Abadie (2005)'s semiparametric DiD to examine the effect of Windward's disclosure on the probability of tanker owner changes (i.e., liquidity). Our proprietary tanker ownership data contains six snapshots: Dec 2022, Mar, Jul, Sep, Dec 2023, and Mar 2024. The dependent variable is a zero-one indicator of owner changes by comparing the owner names across two snapshots. Note that the time spans between consecutive snapshots are different, so we make them comparable by annualizing the owner change variable. The pre-period owner change includes Dec 2022 - Mar 2023 and Mar - Jul 2023. The post-period owner change includes Jul - Sep 2023, Sep - Dec 2023, and Dec 2023 - Mar 2024. The propensity score is calculated by regressing the high/moderate-risk tanker indicator on pre-period tanker characteristics using ML methods. The variables used to construct the propensity score are shown in Table 4. In panel A, we match tankers within each tanker type and calculate weights based on propensity scores and a Gaussian kernel with a bandwidth of 0.01, 0.03, or 0.05. Then, we do DiD estimation in the matched sample, controlling for tanker fixed effects and time×tanker-type fixed effects. The standard errors in panel A are double clustered at the tanker and time×tanker-type levels. In Panel B, we implement Abadie's method in the same way mentioned in Table 5. The standard errors in panel B are as derived in Abadie (2005). *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

Panel A: PSM-DiD			
	Dependent variable: Owner change		
	Bandwidth = 0.01	Bandwidth = 0.03	Bandwidth = 0.05
ATT for High Risk	-0.183***	-0.169***	-0.164***
	[-3.25]	[-2.91]	[-2.87]
Obs. (tanker-month)	11,402	11,402	11,402
ATT for Moderate Risk	-0.168***	-0.159**	-0.152**
	[-3.30]	[-2.79]	[-2.50]
Obs. (tanker-month)	11,776	11,776	11,776
Time × Tanker Type FE	Yes	Yes	Yes
Tanker FE	Yes	Yes	Yes
Panel B: Abadie (2005) Semiparametric DiD			
	Owner change	Owner change demeaned by time	Owner change demeaned by time × tanker type
ATT for High Risk	-0.185***	-0.184***	-0.170***
	[-3.30]	[-3.29]	[-3.04]
Obs. (tanker)	2,346	2,346	2,346
ATT for Moderate Risk	-0.168***	-0.167***	-0.146***
	[-3.20]	[-3.18]	[-2.77]
Obs. (tanker)	2,421	2,421	2,421

Table 8: Difference-in-differences analysis of changes of charterers

In this table, we examine whether U.S. or U.S.-allied companies tend to stop chartering high/moderate-risk tankers after Windward's disclosure. U.S.-allied companies are those from the UK, European Union, Australia, New Zealand, or Japan. The dependent variable is the number of tankers of a given risk type that each charterer employs in a month. The sample period is from Feb 2023 to Mar 2024. The post-period starts from Aug 2023. We regress the dependent variable on the indicators of U.S. and U.S.-allied charterers interacted with the post-period indicator controlling for charterer fixed effects and time fixed effects. The coefficients of the DiD estimation are reported. The standard errors are double clustered at the charterer and time level. *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

	Dependent Variable: No. tankers for each charterer-month		
	No. high-risk tankers	No. mod.-risk tankers	No. low-risk tankers
U.S. charterer $\times \mathbb{I}_{\{t\geq 0\}}$	-0.350*	-0.130	0.063
	[-1.93]	[-0.67]	[0.09]
U.S.-allied charterer $\times \mathbb{I}_{\{t\geq 0\}}$	0.013	-0.030	0.138
	[0.08]	[-0.11]	[0.31]
Obs. (charterer-month)	1,824	1,824	1,824
Time FE	Yes	Yes	Yes
Charterer FE	Yes	Yes	Yes

Table 9: Difference-in-differences analysis of the volume of carried Russian oil

This table reports the coefficients (i.e., average treatment effects on the treated, ATT) estimated with PSM-DiD and Abadie (2005)'s semiparametric DiD to examine the change in the volume of Russian oil carried by high-risk tankers relative to the matched low- or moderate-risk tankers. We include moderate-risk tankers in the control group because there are very few low-risk tankers in this Russian oil trade sample. The sample period is from Feb 2023 to Mar 2024. The post-period starts from Aug 2023. The propensity score is calculated as in the previous sections. In Panel A, we match tankers within each tanker type and calculate weights based on propensity scores and a Gaussian kernel with a bandwidth of 0.01, 0.03, or 0.05. Then, we do DiD estimation in the matched sample, controlling for tanker fixed effects and time×tanker-type fixed effects. The standard errors in panel A are double clustered at the tanker and time×tanker-type levels. In Panel B, we implement Abadie's method as in Table 5. The standard errors in panel B are as derived in Abadie (2005). *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

Panel A: PSM-DiD			
	Dependent variable: Russian oil carried volume		
	Bandwidth = 0.01	Bandwidth = 0.03	Bandwidth = 0.05
ATT for High Risk	-5.750***	-4.829**	-4.635**
	[-2.96]	[-2.62]	[-2.55]
Obs. (tanker-month)	11,396	11,396	11,396
Time × Tanker Type FE	Yes	Yes	Yes
Tanker FE	Yes	Yes	Yes
Panel B: Abadie (2005) Semiparametric DiD			
	Volume	Volume demeaned by time	Volume demeaned by time × tanker type
ATT for High Risk	-4.892***	-4.892***	-5.055***
	[-2.59]	[-2.59]	[-2.68]
Obs. (tanker)	814	814	814

Table 10: Difference-in-differences analysis of sample selection

This table examines whether there are potential selection issues by estimating the average treatment effects on the treated (ATT) on the number of fixture records for high/moderate-risk tankers after Windward's disclosure using PSM-DiD and Abadie (2005)'s semiparametric DiD. The dependent variable is the number of fixture records for each tanker-month, and it is set to zero if there are no fixture records. We exclude the tankers with all zero values of the dependent variable. The sample period is from Feb 2023 to Mar 2024. The post-period starts from Aug 2023. The propensity score is calculated by regressing the high/moderate-risk tanker indicator on pre-period tanker characteristics using ML methods. The variables used to construct the propensity score are shown in Table 4. In panel A, we match tankers within each tanker type and calculate weights based on propensity scores and a Gaussian kernel with a bandwidth of 0.01, 0.03, or 0.05. Then, we do DiD estimation in the matched sample, controlling for tanker fixed effects and time×tanker-type fixed effects. The standard errors in panel A are double clustered at the tanker and time×tanker-type levels. In Panel B, we implement Abadie's method in the same way mentioned in Table 5. The standard errors in panel B are as derived in Abadie (2005). *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

Panel A: PSM-DiD			
	Dependent variable: No. fixture records in each month		
	Bandwidth = 0.01	Bandwidth = 0.03	Bandwidth = 0.05
ATT for High Risk	0.020	0.013	0.011
	[0.68]	[0.48]	[0.43]
Obs. (tanker-month)	37,338	37,338	37,338
ATT for Moderate Risk	0.009	0.002	0.002
	[0.45]	[0.11]	[0.10]
Obs. (tanker-month)	41,776	41,776	41,776
Time × Tanker Type FE	Yes	Yes	Yes
Tanker FE	Yes	Yes	Yes
Panel B: Abadie (2005) Semiparametric DiD			
	Raw no. records	No. records demeaned by time	No. records demeaned by time × tanker type
ATT for High Risk	0.009	0.009	0.015
	[0.45]	[0.47]	[0.78]
Obs. (tanker)	2,447	2,447	2,447
ATT for Moderate Risk	-0.003	-0.003	0.000
	[-0.18]	[-0.19]	[0.02]
Obs. (tanker)	2,984	2,984	2,984

Table 11: Parameter values and moments matching

This table presents the calibrated values for the model described in Section 6 (in Panel A) and reports the moments matching results (in Panel B).

Panel A: Parameter values

Variable	Symbol	Value	Calibration method
Discount factor	β	0.9	20-year life cycle and 10% scrapped value
Proportion of bad tankers	λ	0.114	based on the estimation in CRS (2024)
Unconditional sanction probability	w_0	0.50%	# sanctioned tankers / # total tankers
Incremental sanction probability	δ_0	0.29%	by moments matching
Sanction penalty ('Clean' exporters)	z	1	by moments matching
Sanction penalty ('Rogue' exporters)	z^R	41.66	by moments matching
Mean oil trade profit	r, r^R	1	normalize to one
Volatility of oil trade profit	σ	0.256	by moments matching
Max operation cost when dealing with 'Rogue' exporters	$\bar{c}$	1	set equal to the mean oil trade profit
Classification precision (pre-period)	ρ_{pre}	0.269	implied by a pseudo-R^2 of 7.5% (Figure 11)
Classification precision (post-period)	ρ_{post}	0.424	implied by a pseudo-R^2 of 17.2% (Figure 11)

Panel B: Moments matching

Moments	Model analog	Data value	Model value
Annualized volatility of low-risk tankers' fixtures before Windward's disclosure (divided by mean)	$\frac{\text{std}(p_L^{pre})}{\text{mean}(p_L^{pre})}$	0.77	0.75
Annualized volatility of high-risk tankers' fixtures before Windward's disclosure (divided by mean)	$\frac{\text{std}(p_H^{pre})}{\text{mean}(p_H^{pre})}$	0.81	0.82
Disclosure effect on high-risk tankers' fixtures (DiD estimation in Table 5)	$\frac{p_H^{post}-p_H^{pre}}{p_H^{pre}} - \frac{p_L^{post}-p_L^{pre}}{p_L^{pre}}$	-0.13	-0.16
Disclosure effect on high-risk tankers' routes (DiD estimation in Table 6)	$\frac{b_H^{post}-b_H^{pre}}{b_H^{pre}} - \frac{b_L^{post}-b_L^{pre}}{b_L^{pre}}$	-0.17	-0.12

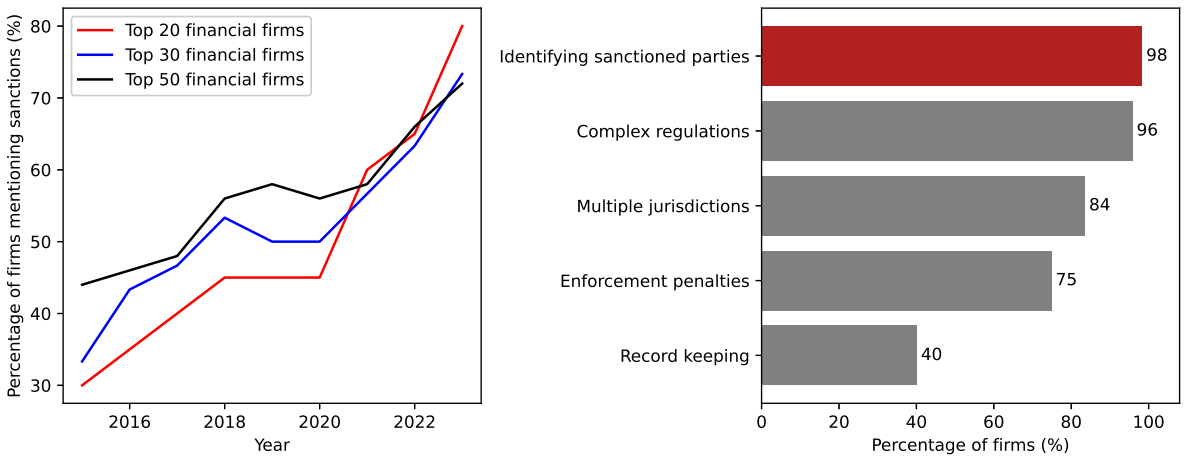


Figure 1: This figure shows the results from a large-language-model-based analysis of 10-K reports of the largest (by total assets) U.S. financial firms between 2015 and 2023. We first ask GPT-4 to identify the reports that mention sanctions enforced by OFAC, or sanctions-related risk, cost, or uncertainty. Then, we ask it to identify the reasons why the firms think sanctions compliance is challenging (further details are in Internet Appendix A). The left plot shows the percentage of firms among the 20, 30, and 50 largest financial firms, respectively, that mention economic sanctions-related risk (or cost or uncertainty) in their 10-K reports over the 2015-2023 period. The right plot shows the top five challenges in sanctions compliance for the 50 largest financial firms.

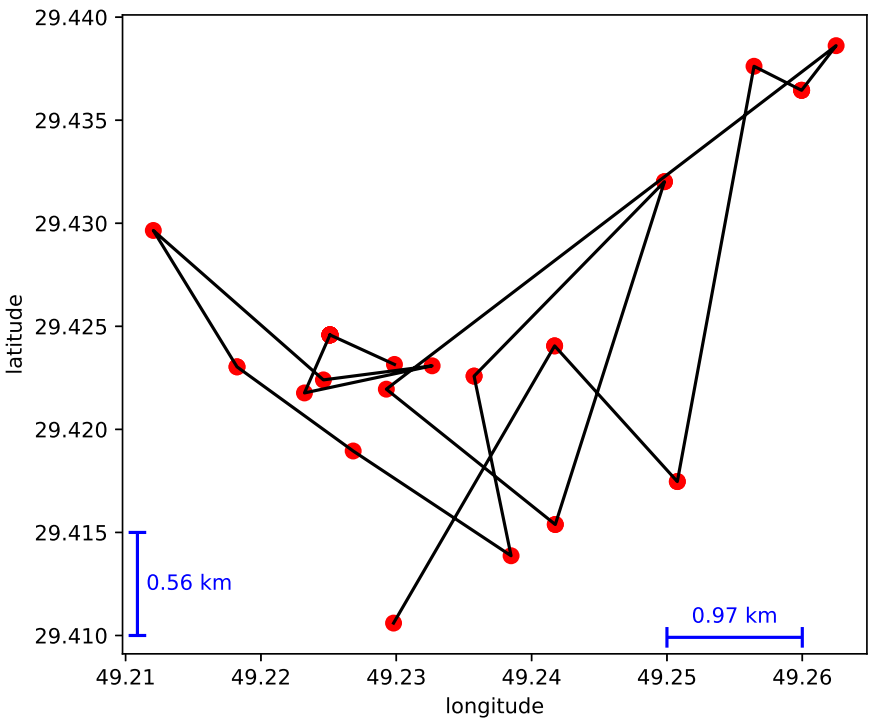


Figure 2: This figure illustrates an irregular trajectory. We define a trajectory as irregular if a tanker's total changes of course within a period are abnormally high. The axes show degrees of latitude and longitude. Red dots show the coordinates of each AIS signal emitted by each tanker.

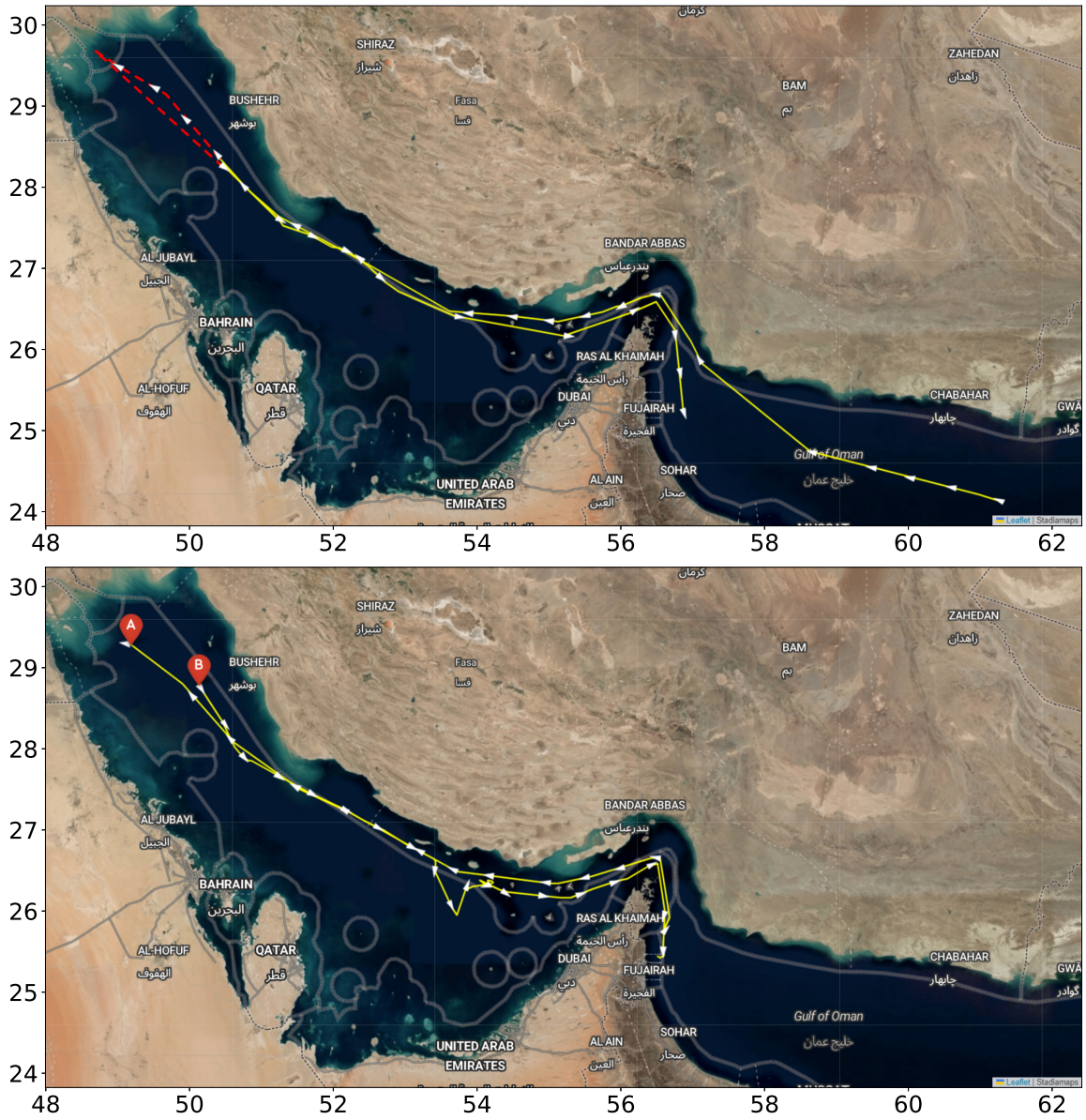


Figure 3: In this figure, solid and dashed lines show the trajectories of two tankers in the Persian Gulf, as given by their AIS signals (AIS is the Automatic Identification System for ship-tracking). White arrows indicate the direction on movement. The top plot shows a case of spoofing, i.e., at a time when the tanker was physically seen at an Iranian port (as per our dataset of sanctions-violating tankers), its AIS signals were showing it in the northern part of the Gulf near Iraq (red dashed lines). Therefore, the tanker was manipulating its signals. The yellow solid lines show this tanker's path before and after its Iranian port visit. The bottom plot shows a case of dark activity, i.e., the tanker stopped emitting signals in a period when it was observed at an Iranian port (again as per our dataset). The last signal before it went dark is indicated by marker "A" and was emitted at 02:16:12 on 2021-01-23. The first signal after emerging from its dark period is indicated by marker "B", not too far from A and near the Iranian coast, and was emitted at 04:54:45 on 2021-01-27, for a four-day dark period.

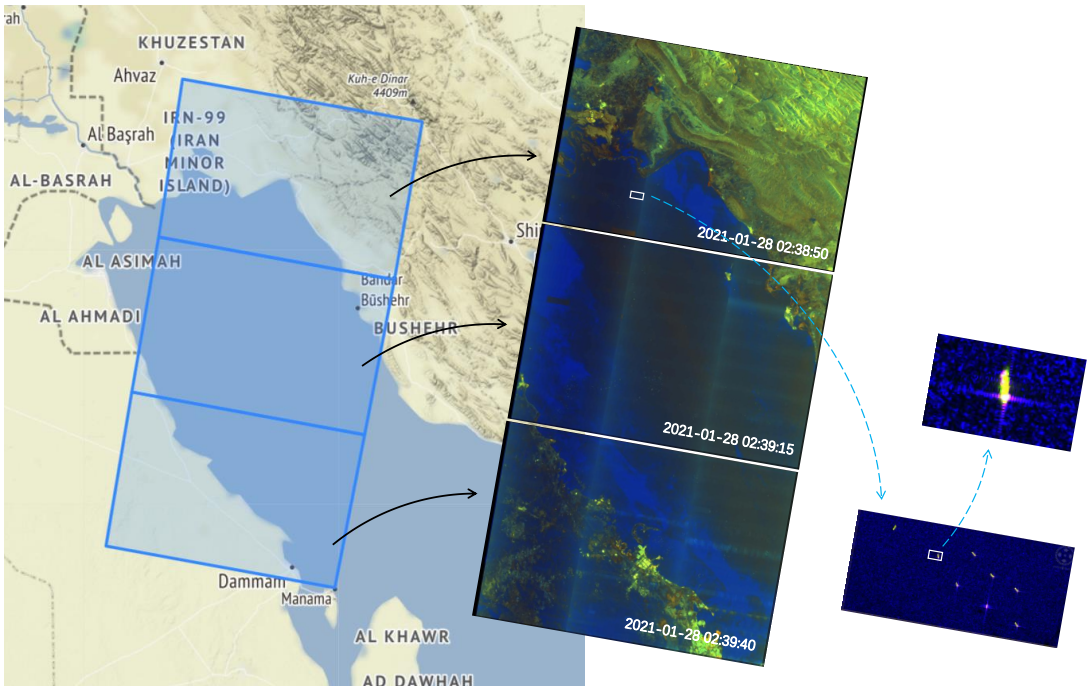


Figure 4: This figure shows examples of satellite images. These three images were taken a little after 2:30 a.m. on the 28th of January 2021 in the Persian Gulf area. Each image covers a rectangular sector, 250 kilometers wide. One insert on the right shows eight ships as little yellow objects at sea, while the other insert shows at the largest enlargement level one of these ships, with individual pixels clearly seen.

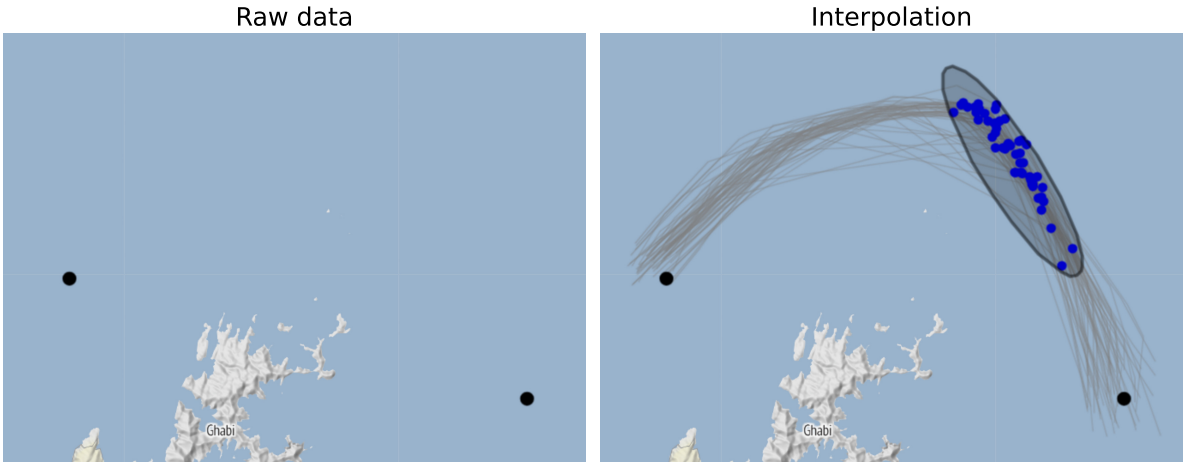


Figure 5: This figure illustrates the construction of the search area via interpolation. The black dots show the locations of a given tanker (call it A) at time t_0 (t_1) of the closest AIS signal before (after) the time t^I when a satellite could see the plotted area (the Strait of Hormuz). Let x_{t_0} and x_{t_1} denote tanker A's locations as per the AIS signals at t_0 and t_1. We require that t_0 and t_1 are within six hours of t^I. Tanker A was moving between these two times, and the distance between x_{t_0} and x_{t_1} is large. We determine the search area using the trajectories of other comparable tankers that have passed near both x_{t_0} and x_{t_1}, in the same direction and over a similar time interval, and hence at similar speed. We require a comparable tanker to have sent a signal within five kilometers from x_{t_0}, and then another one within the same distance from x_{t_1}, whereby the time it took to travel between these two signals is within 30 minutes of the difference $t_1 - t_0$. To increase the number of matches, we search for comparable tankers in the three-month period between December 2020 and February 2021. If multiple pairs of signals from a given tanker satisfy these matching criteria, we chose the pair with starting and ending locations closest to x_{t_0} and x_{t_1}. The grey lines in the figure show the trajectories of comparable tankers. Then we find by interpolation the location of each comparable tanker at its time corresponding to t^I for tanker A. These locations are shown as blue dots in the figure. Finally, we define the search area as the smallest ellipse that contains all of these interpolated locations (i.e., the blue dots), except those which are more than five standard deviations away from the center of the interpolation locations. We use ellipses to reduce the size of the search area.

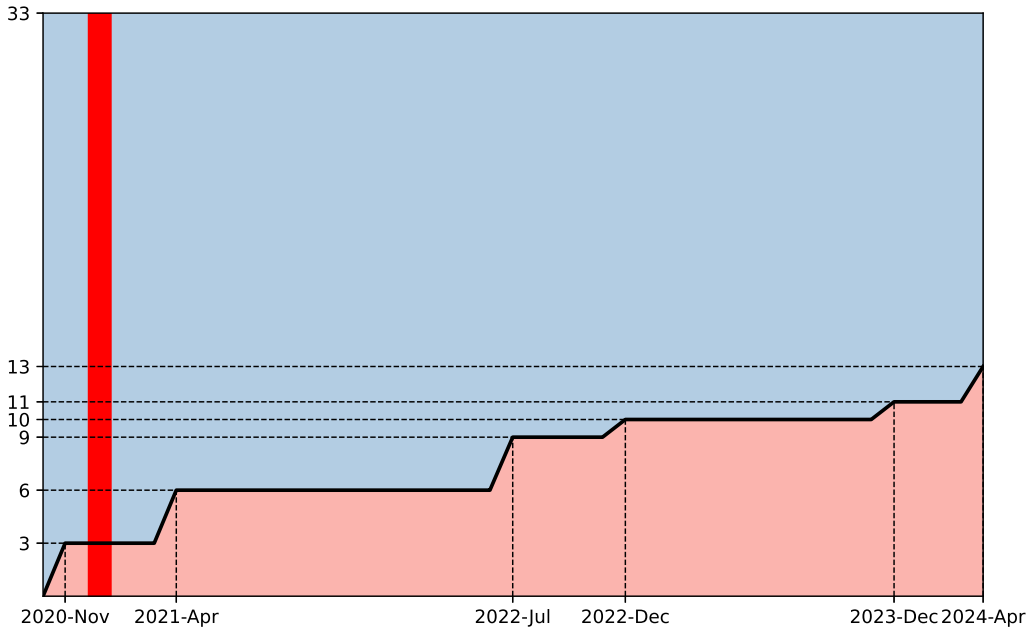


Figure 6: This figure examines the performance of UANI's "Ghost Armada" list of suspect sanctions-violating tankers. The solid black line shows at different dates the number of vessels in UANI's list that were also present in our dataset of 33 sanctions-violating tankers, that were observed physically at Iranian ports during January 2021. The line shifts up on dates when UANI's list is updated (shown on the x-axis). A red bar denotes the observation time of our dataset.

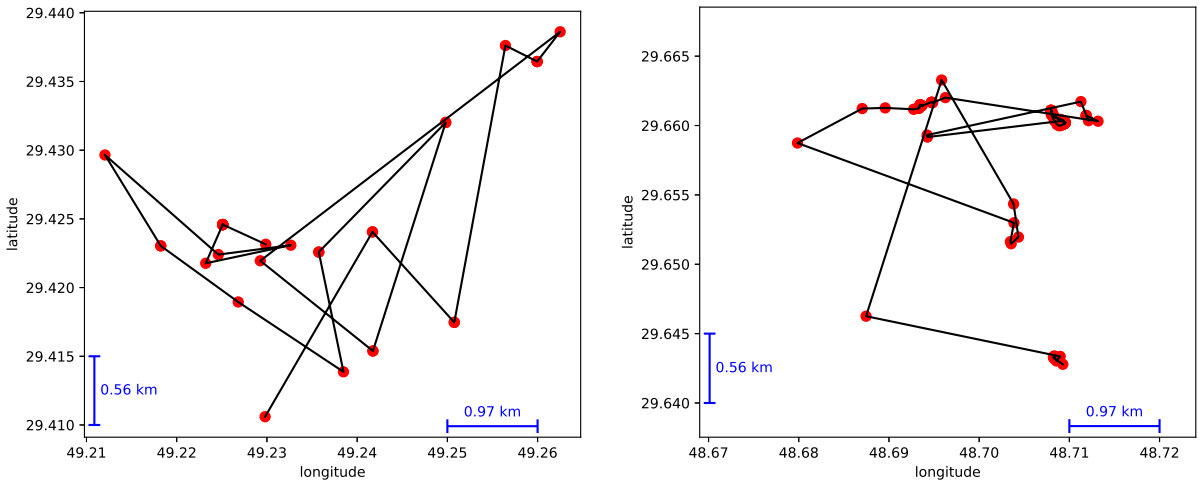


Figure 7: This figure shows irregular trajectories from a sanctions-violating tanker and a sanctions-compliant tanker. The left plot is for a sanctions-violating tanker from our proprietary dataset (it reproduces Figure 2 for direct comparability), and the right plot is for a sanctions-compliant tanker. The axes show degrees of latitude and longitude. Red dots show the coordinates of each AIS signal emitted by each tanker.

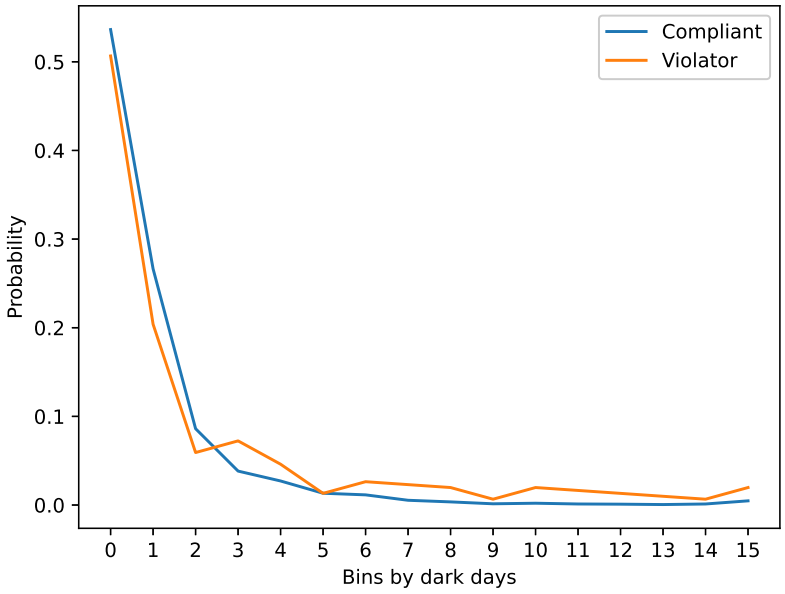


Figure 8: This figure shows the distribution of dark days in Jan 2021 for tankers entering the Persian Gulf conditional on the tanker going dark for at least half a day. Each bin in x-axis represents a range of dark time: bin 0 is from 12 hours to 24 hours; bin x ($x = 1, 2, \cdots, 14$) is from x to $x+1$ days; bin 15 is equal or greater than 15 days. The blue line represents compliant tankers, while the yellow line represents sanctions violators.

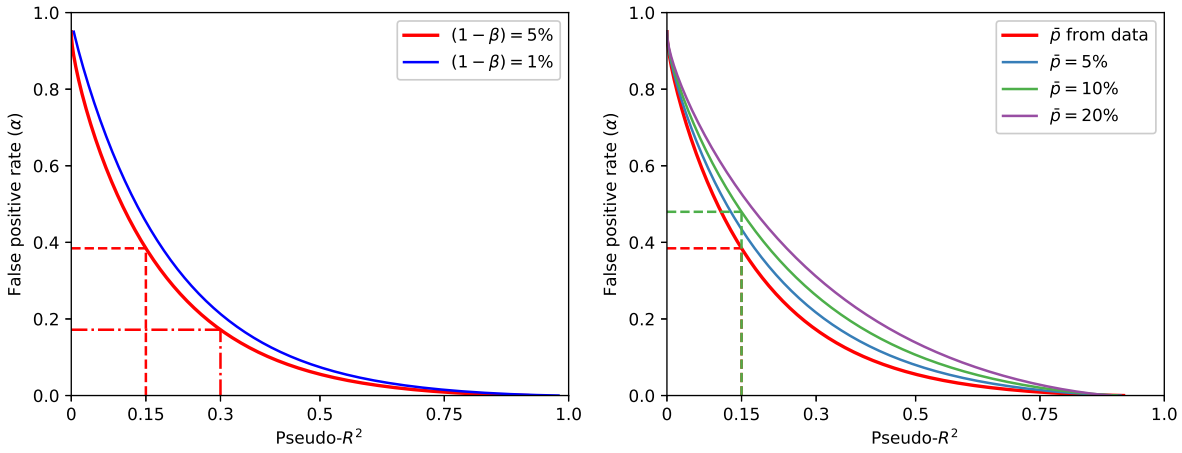


Figure 9: This figure plots the relation between Pseudo-R^2 (x-axis) and false positive rate (denoted by α, y-axis) when fixing the unconditional probability of sanctions-violating tankers (denoted $\bar{p}$) or the true positive rate (denoted β). In the left plot, $\bar{p}$ is set to 2.23% (thus equal to the one calculated from our proprietary data sample: 33/1481), and we set β to be 90%, 95%, or 99%. In the right plot, we fix $\beta = 95\%$ and vary $\bar{p}$.

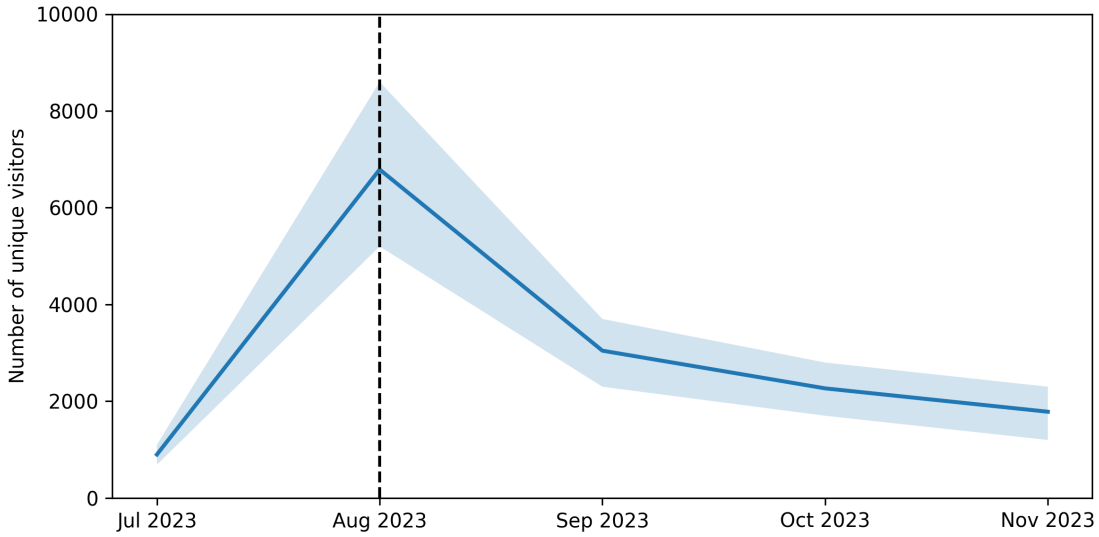


Figure 10: This figure plots the number of unique visitors of the domain https://windward.ai/ (and its sub-domains), from July 2023 to November 2023. The data is extracted from Semrush, a competitive analysis platform operated by Semrush Holdings, Inc. The shadow area represents the estimation deviation (i.e., the error band provided by the platform). The vertical dashed line indicates the event month of Windward's disclosure of high/moderate-risk tankers (August 2023).

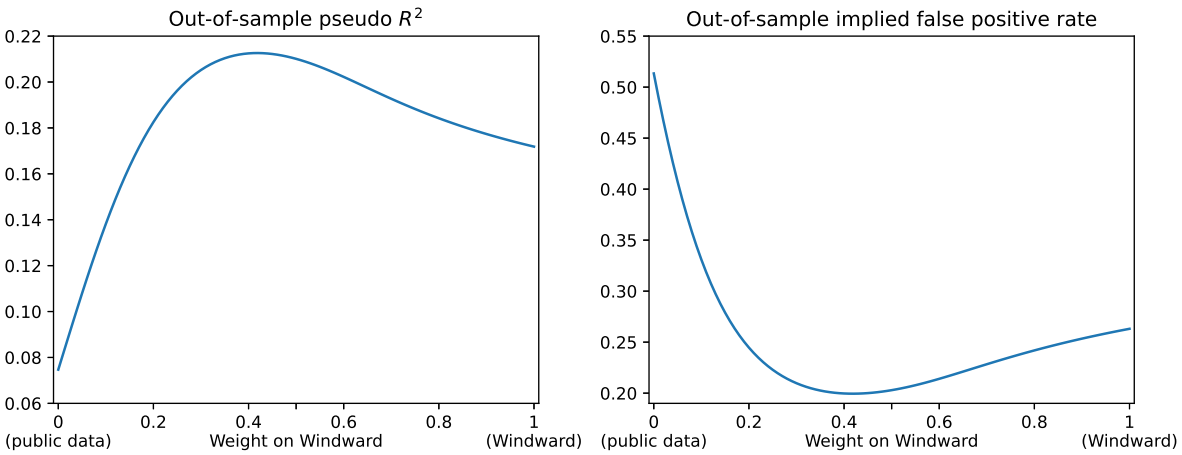


Figure 11: This figure compares the out-of-sample predictability of future sanctions achieved using weighted averages of Windward's and public-data-based predicted outcomes with various weights. We plot the corresponding out-of-sample performance measures against the weight for Windward's list. Windward's predicted outcome is a variable equal to one for high-/moderate-risk tankers and zero otherwise. For the public-data-based predicted outcome, we take the view of a third party predicting future sanctioned tankers at the end of July 2023 (i.e., right before Windward's disclosure). For our training sample the predictors are constructed from data from July 2022 till January 2023 and the dependent variable is a zero-one indicator of tankers sanctioned between July 2022 and July 2023. The public data include AIS-based predictors (as described in Section 4.1), satellite-based predictors (as described in Section 4.2), and a zero-one predictor derived from the UANI list (as described in Section 4.3). We train ML models (decision trees and neural networks) with hyper-parameters optimally determined by 10-fold cross-validation. The public-data-based predicted outcome is the estimated probability of a tanker to be sanctioned after July 2023. The out-of-sample evaluation does not include tankers that were already sanctioned at the prediction time, i.e., the end of July 2023.

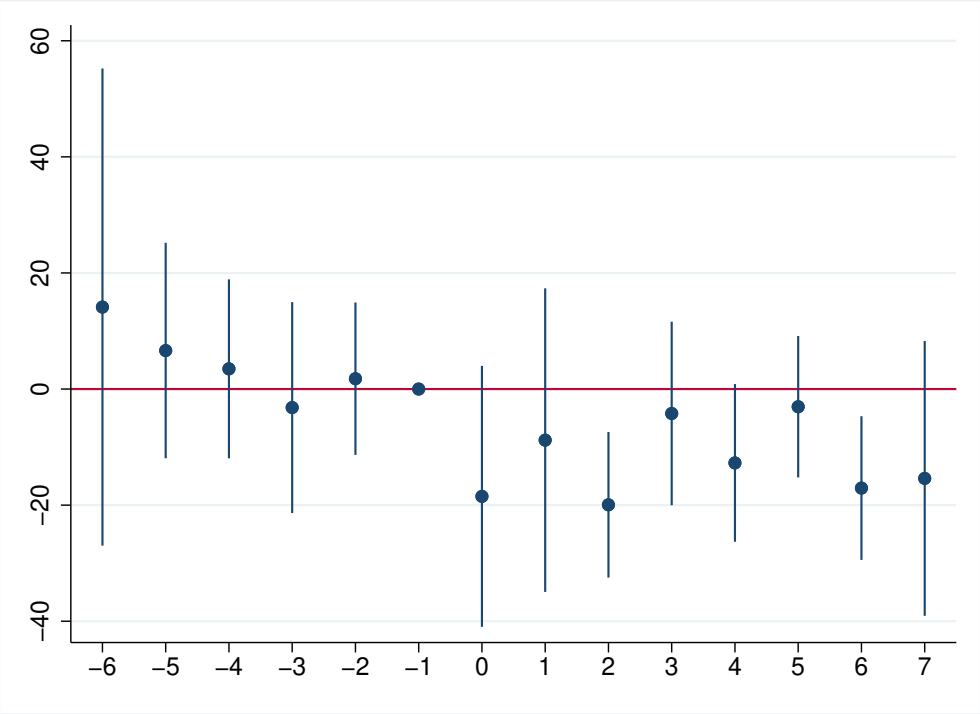


Figure 12: This figure plots the time series of coefficients (i.e., average treatment effects on the treated, ATT) estimated with PSM-DiD to examine the change in freight rates (or fixtures) for high-risk tankers only (i.e., we remove moderate-risk tankers from the sample). The sample period is from Feb 2023 to Mar 2024. Period 0 indicates the month of Windward's disclosure, i.e., Aug 2023. The propensity score is calculated by regressing the high/moderate-risk tanker indicator on pre-period tanker characteristics using ML methods: average outputs from a decision tree and neural networks. The variables used to construct the propensity score are shown in Table 4. We match tankers within each tanker type and calculate weights based on propensity scores and a Gaussian kernel with a bandwidth of 0.01. Then, we regress fixtures on indicators of high-risk tankers interacted by period dummies controlling for tanker fixed effects and time×tanker-type fixed effect. We trim the fixtures at the 1st and 99th percentiles each month to avoid the influence of outliers. Vertical bars represent 95% confidence intervals. The standard errors are double clustered at the tanker and time×tanker-type levels.

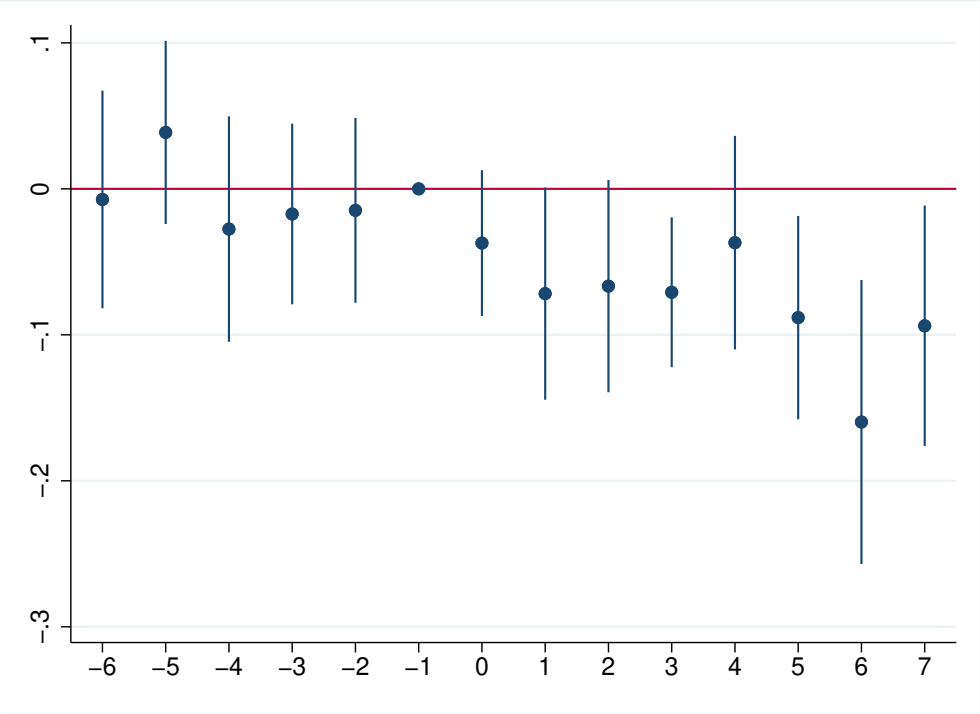


Figure 13: This figure plots the time series of coefficients (i.e., average treatment effects on the treated, ATT) estimated with PSM-DiD to examine the change of the routes of high-risk tankers only (i.e., we remove moderate-risk tankers from the sample). The dependent variable is a zero-one indicator of passing during each month within 12 nautical miles (i.e., the boundary of the territorial sea) of Iran, Russia, or Venezuela (only before 2024). The sample period is from Feb 2023 to Mar 2024. Period 0 indicates the month of Windward's disclosure, i.e., Aug 2023. The propensity score is calculated by regressing the high/moderate-risk tanker indicator on pre-period tanker characteristics using ML methods. The variables used to construct the propensity score are shown in Table 4. We match tankers within each tanker type and calculate weights based on propensity scores and a Gaussian kernel with a bandwidth of 0.01. Then, we regress the dependent variable on indicators of high-risk tankers interacted by period dummies controlling for tanker fixed effects and time×tanker-type fixed effect. Vertical bars represent 95% confidence intervals. The standard errors are double clustered at the tanker and time×tanker-type levels.

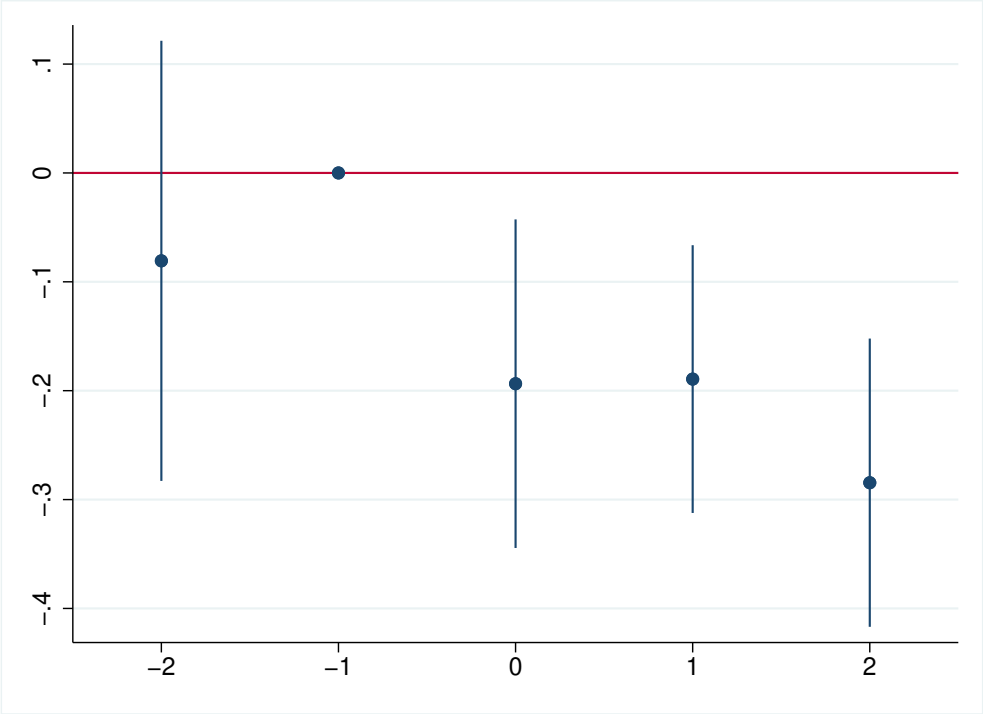


Figure 14: This figure plots the time series of coefficients (i.e., average treatment effects on the treated, ATT) estimated with PSM-DiD to examine the effect of Windward's disclosure on the probability of tanker owner changes for high-risk tankers only (i.e., we remove moderate-risk tankers from the sample). Our proprietary tanker ownership data contains six snapshots: Dec 2022, Mar, Jul, Sep, Dec 2023, and Mar 2024. The dependent variable is a zero-one indicator of owner changes by comparing the owner names across two snapshots. Since the time spans between consecutive snapshots are different, we make them comparable by annualizing the owner change variable. The pre-period includes Dec 2022 - Mar 2023 and Mar - Jul 2023. The post-period includes Jul - Sep 2023, Sep - Dec 2023, and Dec 2023 - Mar 2024. The propensity score is calculated by regressing the high/moderate-risk tanker indicator on pre-period tanker characteristics using ML methods. The variables used to construct the propensity score are shown in Table 4. We match tankers within each tanker type and calculate weights based on propensity scores and a Gaussian kernel with a bandwidth of 0.01. Then, we regress the dependent variable on indicators of high-risk tankers interacted by period dummies controlling for tanker fixed effects and time×tanker-type fixed effect. Vertical bars represent 95% confidence intervals. The standard errors are double clustered at the tanker and time×tanker-type levels.

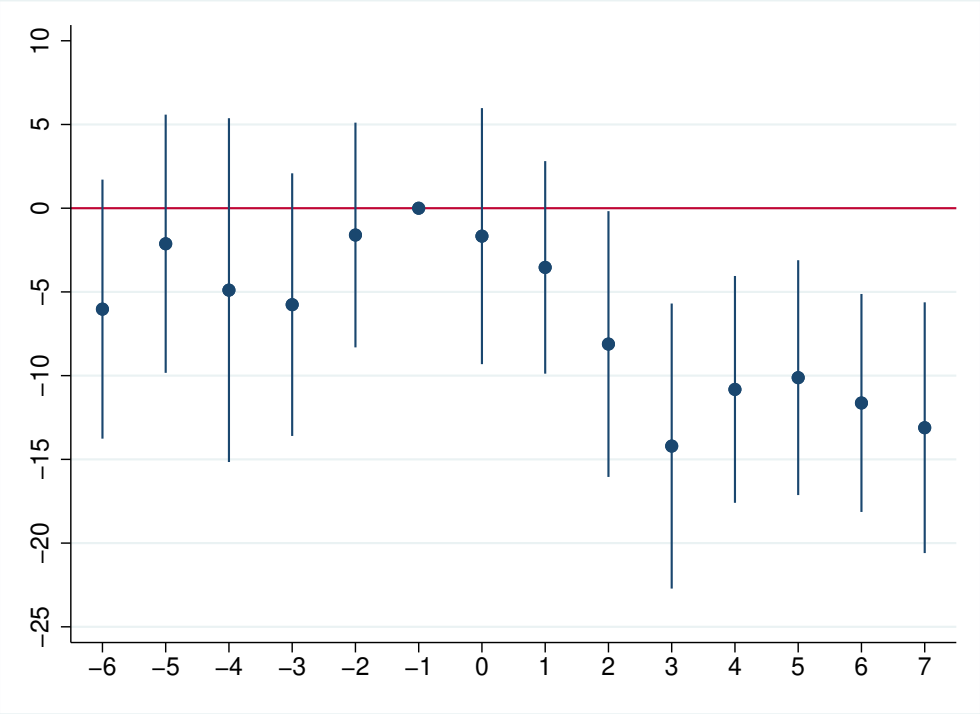


Figure 15: This figure plots the time series of coefficients (i.e., average treatment effects on the treated, ATT) estimated with PSM-DiD to examine the change in the volume of Russian oil carried by high-risk tankers relative to the matched low- or moderate-risk tankers. We include moderate-risk tankers in the control group because there are very few low-risk tankers in this Russian oil trade sample. The sample period is from February 2023 to March 2024. Period 0 indicates the month of Windward's disclosure, i.e., Aug 2023. The propensity score is calculated as in the previous sections. We match tankers within each tanker type and calculate weights based on propensity scores and a Gaussian kernel with a bandwidth of 0.01. We regress the volume of carried Russian oil on indicators of high-risk tankers interacted by period dummies controlling for tanker fixed effects and time×tanker-type fixed effect, as per equation 8 in the text. Vertical bars represent 95% confidence intervals. The standard errors are double clustered at the tanker and time×tanker-type levels.

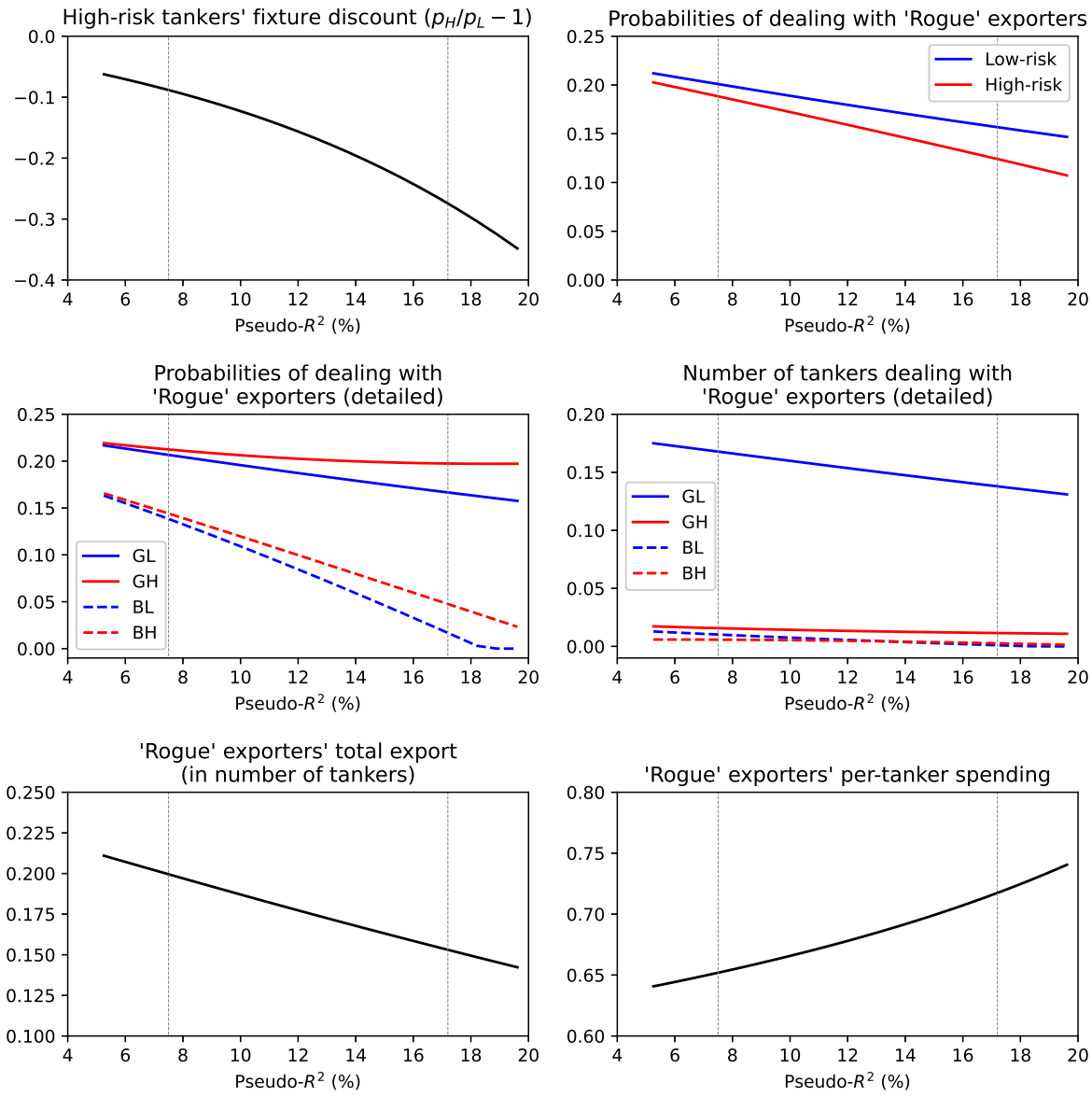


Figure 16: This figure presents the predictions from the calibrated model from Section 6. We solve the model by discretizing the aggregate shock ϵ on a grid. We plot the average model outcomes for the values of ϵ as functions of pseudo-R^2, which is backed out from the classification precision ρ in the model. In each subplot, two vertical dashed lines indicate the pre- and post-Windward precisions, which are 7.5% and 17.2%, respectively. The total number of tankers is normalized to one.

Internet Appendix for "Sanctions, Satellites, Suspect Ships"

Avi Dutt, Abhiroop Mukherjee, George Panayotov, Debjit Roy, Xudong Wen

September 13, 2024

A Large-language-model-based analysis

We first obtain a list of major U.S. financial firms from Compustat by selecting firms with a GICS sector code of 40 and total assets ranked in the top 50. Then, we download these firms' 10-K annual reports from SEC websites with filing dates from 2015 to 2023. To focus on the information related to economic sanction, for each report, we select all paragraphs containing the word "sanction" as materials for textual analysis. Our textual analysis is based on GPT-4. We input the selected paragraph(s) and ask GPT-4 the following question:

> **Question**: Does the following paragraph mention economic sanction, for example, enforced by OFAC, and the economic-sanctions-related risk, cost, uncertainty to their business? Output Yes or No, and give a brief reason.

Here are two examples of responses from GPT-4:

> **Input**: for example, the current tensions between the United States and Russia due to the Russian invasion of Ukraine that began in February 2022 and the resulting actions by the United States and a number of other countries in response (including economic sanctions imposed on Russia and the provision of military supplies to Ukraine) could result in retaliatory cyber attacks by Russian threat actors on our business or on third parties with which we do business that have a material impact on our business.
>
> **Response**: Yes, the paragraph mentions economic sanctions imposed on Russia and the resulting potential risks, costs, and uncertainties to the business due to possible retaliatory cyber attacks.

Input: governmental regulatory authorities may institute administrative or judicial proceedings that may result in censure, fines, the issuance of cease-and-desist orders, trading prohibitions, the suspension or expulsion of a broker-dealer or member, its officers, registered representatives or employees or other similar sanctions.

Response: No, the paragraph does not mention economic sanctions, specifically enforced by OFAC, or the related risk, cost, and uncertainty to their business. It only discusses possible outcomes of administrative or judicial proceedings that can be brought against a broker-dealer or member.

As the examples show, despite both inputs containing the word "sanction", GPT-4 can further distinguish whether it relates to what we want to identify.

Then, to analyze the underlying reasons for sanctions compliance risk, we let GPT-4 generate a list of potential reasons and mark the related reason(s) for each 10-K report. Specifically, we first ask GPT-4 the following question without any input text:

Question: Please list potential reasons why a firm thinks sanctions compliance is challenging.

Based on the replied 10 reasons, we input the paragraphs that are related to economic sanctions previously identified by GPT-4, and further ask the following question for each report:

Question: Read the provided firm annual report and mark the reasons from the following list why this firm thinks sanctions compliance is challenging.

1. Complex regulations
2. Frequent updates
3. Multiple jurisdictions
4. Identifying sanctioned parties
5. Internal communication
6. Staff training
7. Screening technology
8. Risk assessment
9. Record keeping
10. Enforcement penalties

Give your answer in a sequence of reason IDs separated by a comma, for example, if your answer is the first 5 reasons, output 1,2,3,4,5

Based on GPT-4's responses, we calculate the percentage of firms related to each reason and

report the five most important reasons in Figure 1.

B Calculation of implied false positive rate

Let $\bar{p}$ denote a tanker's unconditional probability of being a sanctions violator. Consider the following conditional classification precision:

	Predicted "Compliant"	Predicted "Violator"
True "Compliant"	$1-\alpha$	α
True "Violator"	$1-\beta$	β

A perfect classification means $\alpha = 0$ and $\beta = 1$. To represent pseudo-R^2 as a function of $\bar{p}$, α, and β, we first calculate the conditional probability of being true "Bad" under each predicted label, i.e., we define

$$p_C = \Pr[\text{true "violator"}|\text{predicted "compliant"}],$$

and

$$p_V = \Pr[\text{true "violator"}|\text{predicted "violator"}].$$

It is easy to show that

$$p_C = \frac{\bar{p}(1-\beta)}{\bar{p}(1-\beta)+(1-\bar{p})(1-\alpha)}, \quad p_V = \frac{\bar{p}\beta}{\bar{p}\beta+(1-\bar{p})\alpha}. \tag{A-1}$$

The pseudo-R^2 defined as one minus log-likelihood ratio is given by

$$\text{pseudo-}R^2 = 1 - \frac{E[y\ \log(p)+(1-y)\log(1-p)]}{\bar{p}\ \log(\bar{p})+(1-\bar{p})\log(1-\bar{p})},$$

where $y = 1$ ($y = 0$) represents a tanker is true "violator" (true "compliant"), and p is the predicted probability of true "violator" satisfying

$$p = \begin{cases} p_C, & \text{for predicted "compliant" tankers,} \\ p_V, & \text{for predicted "violator" tankers.} \end{cases}$$

Finally, the pseudo-R^2 is given by

$$1 - \frac{\bar{p}(1-\beta)\log(p_C)+\bar{p}\beta\log(p_V)+(1-\bar{p})(1-\alpha)\log(1-p_C)+(1-\bar{p})\alpha\log(1-p_V)}{\bar{p}\log(\bar{p})+(1-\bar{p})\log(1-\bar{p})}, \tag{A-2}$$

where p_C and p_V are functions of $\bar{p}$, α, and β defined in equation (A-1). Based on equation (A-2), we can map pseudo-R^2 to α when given $\bar{p}$ and β.

C Predictor construction

C.1 AIS-based predictors

The detailed descriptions for the construction of AIS-based predictors are listed below:

- Identity Change: the indicator equals one if a tanker has changed its MMSI number or flag in the preceding six months, as reported in AIS, and zero otherwise.
- Risky Flag: the indicator equals one if a tanker's flag is among the top 10 that have been associated with moving Iranian and Venezuelan oil (Lloyd's List (2021)). These are the flags of Panama, Cameroon, Vietnam, Djibouti, Cook Islands, Tanzania, Togo, Palau, Russia, and Belize.
- Ship-to-ship Transfer (STS): two types of predictors are constructed. The first equals one when the AIS navigation status is recorded as "At Anchor" or "Moored" at locations that are at least x km away from the nearest port, and zero otherwise. We set x to be 10 or 20 (i.e., two versions of this predictor). We obtain port coordinates from the World Port Index (WPI) published by the Maritime Safety Information and supplement this list with estimated ports via a data-driven clustering method, DBSCAN, to reflect ports that are newly-built or have been omitted in the WPI. DBSCAN method is described in Section D.4. The second equals one when the draft (i.e., the vertical distance between the waterline and the bottom of the hull, which indicates how much cargo the tanker carries) changes by more than one meter while the tanker is at least x km away from the nearest port. We again set x to be 10 or 20 (i.e., two versions of this predictor).
- Irregular Trajectory: we define a trajectory as irregular if a tanker's total changes of course within a day are abnormally high. A course is calculated from two adjacent AIS signals, ranging from 0° to 360°, with 0°, 90°, 180°, 270° denoting north, east, south, and west, respectively. A course change is the difference between two adjacent courses, ranging from -180° to 180°. We sum all absolute values of the course changes in a day. To filter out small-scale movements, we calculate the course when the distance between two signals is above a threshold. If the distance doesn't exceed the threshold, we move to the following signals until the requirement is satisfied. We set the distance cutoff

to be 1, 2, or 5 km (i.e., three versions). The indicator we use equals one when the total course changes of a given tanker in a day are above the 95th percentile of the distribution of all such changes within the respective month.

- DBSCAN Outlier: this is a measure aiming to capture anomalous locations conditional on speed and course as identified by DBSCAN - a data-driven clustering method that has been employed to detect marine anomalies in a shipping context (e.g., Pallotta et al. (2013)). We first sort the data points into several groups based on speed and course, and then perform DBSCAN within each group, i.e., we cluster conditional on speed and course. The outliers that we find are also conditional, i.e., they deviate in certain ways from tankers to which they are *similar* in other ways. The indicator in this case equals one if a ship is identified as an outlier on a given day, whereby we set the distance parameter of DBSCAN to be 10 or 20 km (i.e., two versions of the predictor). The DBSCAN method is described in the Internet Appendix D.4.

- Dark Activity: equals one if the time gap between two signals is longer than T days. We set $T = 1, 2, 3$ (i.e., three versions of this predictor).

C.2 Satellite-based predictors

We construct the following satellite-based predictors and provide detailed descriptions for constructing the search area.

- Satellite Detection: we construct the predictors separately for detecting spoofing and dark activity. For spoofing detection, the predictor equals one for day t if a tanker is not seen in its corresponding search area during the period day $[t - r, t + r]$. We set $r = 0, 1, 2$ (i.e., three versions) recognizing that spoofing on a given day t may also imply spoofing on adjacent days (e.g., r can be viewed as a smoothing parameter for a time series). For dark activity detection, the predictor equals one if a tanker is not seen in its corresponding search area during its dark period. Because the dark period is observed, we don't construct multiple versions with different periods as in spoofing detection.

We define search areas on satellite images that should contain given tankers, under the null hypothesis that the tankers are not violating sanctions. Then, if we do not find a tanker in such a search area, we conclude that it was likely involved in sanctionable activity. The construction of a search area mainly involves an estimation of its center and size. For each tanker in our sample and each satellite image, we use two adjacent AIS signals around the satellite image time, i.e., the last before and the first after it, to construct the search areas.

Using both these AIS signals helps to narrow down significantly the search area – this is a crucial concern, given the trade-off between increasing the false positive rate (i.e., a smaller search area yields a lower probability of actually seeing the tanker and thus not rejecting the null) and reducing the false negative rate (i.e., a smaller search area helps to catch an actual violator as it makes it less likely to be mistaken for another ship that happens to be in the same search area). Our construction aims to balance these two aspects.

In the case of spoofing detection, we consider the following three cases: First, a stationary tanker, whereby the two adjacent signals around the satellite image are emitted roughly at the same location. This is the simplest case as the search area can be constructed as a circle centered around that location with a properly chosen radius. To reduce the concern that a compliant tanker following a round-trip route may not show up in the circle, we only flag the tankers if they don't show up and the time between the two signals is within a threshold. Second, a moving tanker with signal(s) close to the satellite image time (i.e., within 15 minutes of it), whereby the distance traveled between the two signals is large. In this case, we focus on the signal closer to the satellite image time and define the search area to be a circle centered around the location of this closer signal and with a properly chosen radius. Third, again a moving tanker, but with signals only within a few hours of the satellite time. In this most challenging case, we match the respective tanker with other tankers whose trajectories have passed through the same locations, and in the same direction, and for roughly the same time, and while emitting signals. When trying to detect dark activity, we focus only on the first of the above three cases, whereby the two adjacent signals are the last before and first after a dark period.

More formally, for a given tanker (call it A) and a given satellite image at time t^I, let t_0 (t_1) denote the time of the closest AIS signal just before (after) t^I. Let x_{t_0} and x_{t_1} denote tanker A's locations as per the AIS signals at t_0 and t_1. For spoofing detection, we construct search areas in the following three cases:

Case I: Both t_0 and t_1 are reasonably close to t^I and the distance between x_{t_0} and x_{t_1} is small. This case describes a stationary ship and is thus most relevant, given our proprietary dataset where all but one spoofers are stationary. In our calculations, we use a maximum distance of five kilometers and a maximum interval of six hours before and after t^I. In this case, we define the search area as a circle centered at the mid-point between x_{t_0} and x_{t_1} and diameter equal to the distance between them, but not smaller than some lower bound, to account for potential measurement errors (we use a bound 0.5 km).

Case II: One of t_0 and t_1 is too far from t^I (i.e., beyond six hours in our calculations), but the other one is very close (within 15 minutes of t^I in our calculations). We define the search area simply as a circle centered at the location of the close signal. Since the maximum speed

for tankers is around 15 knots (about 28 km per hour), we define the radius assuming that the ship will move (or has moved) at this maximum speed until (or since) t^I. This radius is bounded below as before. By ignoring the more distant AIS signal we effectively accept that any direction of movement is possible in this case.

Case III: Both t_0 and t_1 are again reasonably close to t^I (within six hours), but the distance between x_{t_0} and x_{t_1} is large. This case describes a moving ship and is the most challenging one. For this case, we determine the search area based on a data-driven interpolation method. The main idea is to interpolate the location of tanker A using the trajectories of other tankers that have passed near both x_{t_0} and x_{t_1}, in the same direction and over a similar time interval.

Figure 6 in the paper provides an example of Case III. Essentially, we match tanker A to some comparable tankers in terms of locations and average speed. We extract by interpolation the location of each comparable tanker at its time corresponding to t^I for tanker A (the blue dots in the figure). We define the search area as the smallest ellipse that contains all of these interpolated locations, except those which are more than five standard deviations away from the center of the interpolation locations. We use ellipses to reduce the size of the search area.

To be included in the matching, we require a tanker to have sent a signal within a five-kilometer distance from x_{t_0}, and then another one within the same distance from x_{t_1}, whereby the time it traveled between these two signals is within 30 minutes of the difference $t_1 - t_0$. To increase the number of matches, we search for comparable tankers in the three-month period between December 2020 to February 2021. If multiple (pairs of) signals from a given tanker satisfy these matching criteria, we chose the one with starting and ending locations closest to x_{t_0} and x_{t_1}. Finally, we validate our method by sequentially leaving out signals as an out-of-sample test and show that our interpolation method yields a successful coverage for 96% of the test cases.

When detecting dark activity, Cases II and III are not applicable, since by design the time interval between the adjacent AIS signals is larger than a few hours – we have defined a dark period to be at least 24 hours long. Therefore, we only consider here Case I, but modify it so that t_0 is the last signal before a dark period, and t_1 is the first signal after a tanker emerges from the dark period.

D Machine learning methods

D.1 Decision trees

Decision trees are a non-parametric machine learning approach for incorporating nonlinearities and interactions. A tree groups together observations based on predictors and "grows" in a sequence of steps. At each step, the sample leftover from the preceding step is split based on one predictor variable, with the goal of separating the sanctions-violating tankers from the other tankers. The tree will try every possible cutoff for each predictor and choose the split to minimize forecast errors before the next step. The splitting stops when further partitioning cannot reduce forecast errors, or when a tree attribute (e.g., tree depth or the minimum number of observations in each group) reaches a pre-specified threshold. We use a decision tree with a loss function to minimize Gini impurity, which is calculated at each node m via $H_m = 2\hat{p}_m(1 - \hat{p}_m)$. $\hat{p}_m$ is the proportion of sanctions-violating tankers at node m. A perfect classification at node m with $\hat{p}_m = 0$ or $\hat{p}_m = 1$ implies $H_m = 0$.

D.2 Neural networks

Neural networks are a parametric machine learning approach consisting of one "input layer" of raw predictors, one or more "hidden layers" that interact and nonlinearly transform the predictors, and an "output layer" that aggregates hidden layers into a final outcome prediction. By employing linear transformation and nonlinear activation in these layers, neural networks can generate flexible and complex relations between raw inputs and final output. We use the activation function of rectified linear unit (ReLU), which is defined as

$$\text{ReLU}(x) = \begin{cases} 0, & \text{if } x < 0, \\ x, & \text{otherwise.} \end{cases} \tag{A-3}$$

We employ a standard network structure with a loss function to minimize mean squared error. We train the model using the Adam optimizer, i.e., an optimization algorithm to determine the learning rate adaptively.

D.3 Cross validation

We use 10-fold cross-validation to determine the optimal values of hyper-parameters. The search space is as follows. For decision trees, the depth is in the set $\{2, 3, 4, 5\}$ and the minimum number of observations in each leaf is in $\{2, 3, 4, 5\}$, hence we have 4×4 combinations for a grid search. For neural networks, the number of hidden layers is in $\{1, 2, 3\}$,

the number of units is in $\{2, 4, 6\}$, and the batch size is 100 or 200, for a total of $3 \times 3 \times 2$ combinations. Table A-1 reports the optimal hyper-parameters for decision trees and neural networks determined by 10-fold cross-validation.

D.4 DBSCAN

We use DBSCAN (Density-Based Spatial Clustering of Applications with Noise) to construct "DBSCAN Outlier" predictors, and cluster ports that are omitted in the WPI for the construction of "Ship-to-ship Transfer" predictors. DBSCAN is a non-parametric clustering algorithm: given a set of points in some space, it groups together points that are closely packed together (points with many nearby neighbors) and marks as outliers for points that lie alone in low-density regions (whose nearest neighbors are too far away). Two key parameters need to be given: (i) a radius parameter to determine how close two points will be viewed as neighbors, and (ii) a minimum number of points that can be counted as a cluster. A smaller radius is helpful to identify the cluster with really nearby members but may split or fail to identify clusters if the actual distance among members is not that small.

E Estimation strategies: Detail

Consider the following setup to fix ideas in our context. Let $t = 0$ ($t = 1$) denote the pre-treatment period (post-treatment period), and X is a vector of tanker observed characteristics. Let $D = 1$ ($D = 0$) indicate that the tanker is treated (untreated), i.e., it is classified as high/moderate risk (low-risk) in the Windward disclosure. Let Y_t be the tanker's observed fixture at period t. Further, suppose $Y_1^T|D = 1$ is the fixture for a treated tanker at $t = 1$, and $Y_1^T|D = 0$ is its counterfactual fixture – i.e., what its fixture would have been at $t = 1$ had it not been treated. Similarly, $Y_1^U|D = 0$ denote the fixture for an untreated tanker at $t = 1$, and $Y_1^U|D = 1$ denote its counterfactual fixture – i.e., what its fixture would have been at $t = 1$ had it actually been treated. The quantity of interest is the average treatment effect on the treated (ATT):

$$ATT = E\left[Y_1^T - Y_1^U|D = 1\right], \tag{A-4}$$

which is the difference between the fixture for a disclosed-as-high/moderate-risk tanker and what its fixture would have been, had it not been disclosed as high/moderate risk. When there is a need to indicate a particular tanker, e.g., tanker i, we add a subscript i to the variables, e.g., X_i and $Y_{i,t}$. In the cases without confusion, we drop the subscript of i for simplicity.

E.1 Propensity score matching (PSM)

PSM addresses the challenge that the conventional DiD identification assumption

$$E\left[Y_1^U - Y_0 | D = 1\right] = E\left[Y_1^U - Y_0 | D = 0\right], \tag{A-5}$$

may not hold when the differences between the characteristics of treated and control tankers, i.e., between $X|D = 1$ and $X|D = 0$, affect the dynamics of outcome variables. For every treated tanker i, PSM tries to find tanker(s) j which have very *similar* characteristics X_j to those of tanker i but were actually classified by Windward as low-risk; then it treats such tanker(s) j as counterfactual. Rosenbaum and Rubin (1983) show that, under unconfoundedness (i.e., $Y_1^T, Y_1^U \perp\!\!\!\perp D \mid X$), it is sufficient to focus solely on tankers with similar propensity $p = \Pr(D = 1|X)$ to have been labeled as high/moderate risk ex-ante. That is, we can define 'similar' tankers relying only on dimensions of X that matter for risk labeling, rather than considering matching on the full vector of characteristics X.

In our notation, PSM replaces $[Y_{i,1}^U - Y_{i,0}|D_i = 1]$, i.e., the counterfactual fixture changes of high/moderate-risk tanker i in the absence of treatment, with $\sum_{j:D_j=0} w(j,i)(Y_{j,1} - Y_{j,0})$, i.e., the weighted average of observed fixture changes of low-risk tanker(s) j that had a similar ex-ante propensity to be labeled as high/moderate risk. $w(j,i)$ is determined by comparing p_i and p_j (Rosenbaum and Rubin, 1983). specifically, we give a higher weight for a low-risk tanker j if p_j, the conditional probability of being labeled as high/moderate risk based on X_j, is closer to that of the target tanker i. Once the weights $w(j,i)$ of matching tankers are obtained, ATT can be estimated as follows:

$$\hat{ATT} = \frac{1}{N_1} \sum_{i:D_i=1} \left[(Y_{i,1} - Y_{i,0}) - \sum_{j:D_j=0} \hat{w}(j,i)(Y_{j,1} - Y_{j,0}) \right], \tag{A-6}$$

where N_1 is the total number of treated tankers and $\hat{w}(j,i)$ is the estimated value of $w(j,i)$. ATT averages the differences between the fixture changes around the disclosure event for treated tankers and for the matched untreated tankers.

In our empirical estimation, we follow Heckman et al. (1997) and use kernel matching to determine the PSM weights. The weight $\hat{w}(j,i)$ equals the absolute value of difference between estimated propensity scores $\hat{p}_i$ and $\hat{p}_j$, transformed via a kernel function $G(x;\theta)$ and normalized:

$$\hat{w}(j,i) = \frac{G(|\hat{p}_j - \hat{p}_i|;\theta)}{\sum_{k:D_k=0} G(|\hat{p}_k - \hat{p}_i|;\theta)}. \tag{A-7}$$

In our application, $G(x;\theta)$ is a Gaussian kernel and we show robustness for the bandwidth parameter θ. $\hat{p}_i$ can be estimated by regressing D on X using logistic regression or other ML methods. To extend the above two-period estimation to multiple periods and add fixed effects, we estimate equation (A-6) by regressing $Y_{i,t}$ on an indicator of post-period high/moderate-risk tankers, i.e., $\mathbb{I}_{\{t=1\}} \times \mathbb{I}_{\{D_i=1\}}$, plus fixed effects. To include the PSM weights in the regression, we follow Imbens (2004) and use weighted least squares (WLS) with a weight of one for each high/moderate-risk tanker, and weights equal to $\sum_{i:D_i=1} \hat{w}(j,i)$ as per equation (A-7) for each low-risk tanker j. For example, suppose that the low-risk tanker C is matched to high/moderate-risk tankers A and B with PSM weights $\hat{w}(C,A)$ and $\hat{w}(C,B)$, respectively. The WLS weight of tanker C is the sum of these two, i.e., $\hat{w}(C,A) + \hat{w}(C,B)$, reflecting the importance of tanker C as being counterfactual for both A and B.

E.2 Abadie's (2005) semiparametric difference-in-differences

Abadie (2005) relies on the following identification assumption:

$$E\left[Y_1^U - Y_0 | X = x, D = 1\right] = E\left[Y_1^U - Y_0 | X = x, D = 0\right]. \tag{A-8}$$

This conditional identification assumption is more suitable to our setting because it is more plausible that conditional on the same tanker size, AIS emission pattern, trajectories, owner structure, etc, the fixture of a disclosed high/moderate-risk tanker in the absence of disclosure would follow parallel dynamics with that of a low-risk tanker having the same value of all these characteristics in X. To proceed, Abadie (2005) shows that ATT defined in equation (A-4) can be written as an integral along all dimensions of X: $E\left[...|D=1\right] = \int E\left[...|X=x, D=1\right] dP(X=x|D=1)$. After substituting into the identification assumption in equation (A-8) and conducting some algebra manipulation, ATT equals to

$$ATT = \frac{1}{\Pr(D=1)} \cdot E\left[\frac{D-p}{1-p} \cdot (Y_1 - Y_0)\right], \tag{A-9}$$

where $p = \Pr(D=1|X)$, as defined before, is the propensity score. To gain some intuition, we can partition the expectation in equation (A-9) based on $D=0$ and $D=1$, which gives

$$ATT = E\left[Y_1 - Y_0 | D=1\right] - E\left[\frac{p/\Pr(D=1)}{(1-p)/(1-\Pr(D=1))} \cdot (Y_1 - Y_0) | D=0\right]. \tag{A-10}$$

In equation (A-10), the first term is the average of changes of observed fixtures for high or moderate-risk tankers, and the second term is a propensity-score-based-weighted average of changes of observed fixtures for low-risk tankers. Under the perfect randomized treatment assignment, $p = \Pr(D = 1|X) = \Pr(D = 1)$ and the propensity-score-based weight is always one, which returns back to the conventional DiD estimator. When p differs from $\Pr(D = 1)$, the method will give higher (lower) weights to the low-risk tankers with higher (lower) p, i.e., those who are closer (not close) to the high/moderate-risk tankers in terms of propensity scores.

F Results on the information shock effects

F.1 Screenshot of Windward's sanction risk measure

Figure A-1 shows a screenshot from Refinitiv Eikon illustrating Windward's sanctions risk measures for one tanker (A STAR).

F.2 Sample filtering

We report the sample filtering process and the sample size after each step, i.e., the number of tankers, in Table A-2.

F.3 Robustness checks

We consider the following robustness checks on our fixture changes estimation. First, we use propensity scores from a logistic regression, rather than the baseline machine learning models. Second, we drop the fixtures in August 2023 to allow about two weeks for the disclosure effect to come into force. Third, we drop the fixtures with Russian ports, i.e., ports of Novorossiysk, St. Petersburg, Primorsk, Tuapse, and Taman in our data, to check that our effects are not coming solely from Russia-related changes in maritime activity around this time. Finally, we use bootstrap standard errors with 500 replications instead of clustered standard errors. The results are shown in Table A-3.

F.4 U.S. charterers' usage of high-risk tankers

Table A-4 lists the U.S. charterers to examine their usage of high-risk tankers individually. However, we don't find clear patterns for the drop in high-risk tanker usage (e.g., private company vs. public company).

G Solving the model

The first-order conditions (FOC) for the 'Rogue' exporters' optimization are

$$\frac{(1-\lambda)(1-\theta_G)}{\bar{c}}(r^R+\epsilon-p_G^R-\delta_L z^R)+\frac{(1-\lambda)\theta_G}{\bar{c}}(r^R+\epsilon-p_G^R-\delta_H z^R)=Q_{GL}^R+Q_{GH}^R$$
$$\frac{\lambda(1-\theta_B)}{\bar{c}}[r^R+\epsilon-p_B^R-(\delta_L+w_L)z^R]+\frac{\lambda\theta_B}{\bar{c}}[r^R+\epsilon-p_B^R-(\delta_H+w_H)z^R]=Q_{BL}^R+Q_{BH}^R$$

Similarly, the first-order conditions for the 'Clean' exporters' optimization are

$$\frac{(1-\lambda)(1-\theta_G)}{\bar{c}}(r+\epsilon-p_L)+\frac{\lambda(1-\theta_B)}{\bar{c}}(r+\epsilon-p_L-w_L z)=Q_{GL}+Q_{BL}$$
$$\frac{(1-\lambda)\theta_G}{\bar{c}}(r+\epsilon-p_H)+\frac{\lambda\theta_B}{\bar{c}}(r+\epsilon-p_H-w_H z)=Q_{GH}+Q_{BH}$$

Given V_B and V_G, we need to solve for p_L, p_H, p_G^R, p_B^R as functions of ϵ. With indifference cost points for tankers as defined in the text, the following equations relate prices and costs

$$p_G^R+(1-\theta_G)c_{GL}+\theta_G c_{GH}=r^R+\epsilon-(1-\theta_G)\delta_L z^R-\theta_G\delta_H z^R$$
$$p_B^R+(1-\theta_B)c_{BL}+\theta_B c_{BH}=r^R+\epsilon-(1-\theta_B)(\delta_L+w_L)z^R-\theta_B(\delta_H+w_H)z^R$$
$$p_L-\frac{(1-\theta_B)}{(1-\lambda)}c_{BL}-\frac{\lambda(1-\theta_B)}{(1-\lambda)}c_{BL}=r+\epsilon-\bar{c}-\frac{\lambda(1-\theta_B)}{(1-\lambda)}w_L z$$
$$p_H-\frac{(1-\lambda)\theta_G}{\lambda}c_{GH}-\theta_B c_{BH}=r+\epsilon-\bar{c}-\theta_B w_H z$$

These equations can be summarized in matrix form as

$$\begin{bmatrix} c_{GL} & c_{GH} & c_{BL} & c_{BH} & p_L & p_H & p_G^R & p_B^R \\ 1 & 1 & 1 & 1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} c_{GL} \\ c_{GH} \\ c_{BL} \\ c_{BH} \\ p_L \\ p_H \\ p_G^R \\ p_B^R \end{bmatrix} =$$

$$
= \begin{bmatrix}
-\delta_L s + \beta(V_B - V_G) \\
-\delta_H s + \beta(V_B - V_G) \\
-\delta_L s \\
-\delta_H s \\
r^R + \epsilon - (1-\theta_G)\delta_L z^R - \theta_G \delta_H z^R \\
r^R + \epsilon - (1-\theta_B)(\delta_L + w_L) z^R - \theta_B(\delta_H + w_H) z^R \\
r + \epsilon - \bar{c} - \frac{\lambda(1-\theta_B)}{(1-\lambda)} w_L z \\
r + \epsilon - \bar{c} - \theta_B w_H z
\end{bmatrix}
$$

Table A-1: Optimal hyper-parameters of ML models

This table reports the optimal hyper-parameters of the ML models (i.e., those in Tables 2 and 3) determined by 10-fold cross-validation (CV).

			Optimal hyper-parameters determined by 10-fold CV		
Model	Detection	Predictors	Depth of tree	Min no. obs. in each leaf	
Tree	Spoofing	Satellite + AIS	4	2	
	Spoofing	Only AIS	3	2	
	Dark	Satellite + AIS	3	5	
	Dark	Only AIS	3	5	
			No. layers	No. units	Batch size
NN	Spoofing	Satellite + AIS	3	6	200
	Spoofing	Only AIS	2	6	100
	Dark	Satellite + AIS	2	6	200
	Dark	Only AIS	3	4	100

Table A-2: Sample filtering

This table reports the sample filtering process and the sample size after each step, i.e., the number of tankers.

Data cleaning steps	**# tankers**
Full tanker list from Reuters	13,997
Keep oil tankers	10,322
Keep large tankers (i.e., Medium, Panamax, Aframax, Suezmax, VLCC)	5,466
Drop tankers that were sanctioned before Windward's disclosure	5,392
Drop tankers without AIS data (i.e., unable to calculate propensity score)	5,363
Reg. sample 1 (merge to fixture data)	3,628
Reg. sample 2 (merge to fixture data and fixture type is WS)	2,852
Reg. sample 3 (merge to ownership data)	3,312

Table A-3: Robustness checks

This table reports on robustness checks for the fixture changes estimation. The sample and construction of propensity score are the same as in Table 6, except for the features indicated in the first row of the table. In panel A, PSM uses a Gaussian kernel with a bandwidth of 0.01. Standard errors are double-clustered at the tanker and time×tanker-type levels (except for the last column). In panel B, we manually subtract the cross-sectional mean by tanker type to account for time×tanker-type fixed effects, since Abadie's estimator is derived without directly accounting for macro trends through time fixed effects. Standard errors are as derived in Abadie (2005) (except for the last column). *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

	Propensity score from logistic reg.	Drop Aug 2023	Drop Russian ports	Bootstrap s.e. (500 replications)
Panel A: PSM-DiD				
		Dependent Variable: Fixtures		
ATT for High Risk	-10.373*	-15.314**	-9.257*	-14.958***
	[-1.72]	[-2.39]	[-1.79]	[-3.42]
Obs. (tanker-month)	7,056	6,493	6,854	7,056
Time × Tanker Type FE	Yes	Yes	Yes	Yes
Tanker FE	Yes	Yes	Yes	Yes
Panel B: Abadie (2005) Semiparametric DiD				
	Dependent Variable: Fixtures demeaned by time × tanker type			
ATT for High Risk	-11.943**	-13.395***	-9.841**	-13.460***
	[-2.56]	[-2.94]	[-2.23]	[-3.14]
Obs. (tanker)	965	913	957	965

Table A-4: List of U.S. charterers

This table reports the average number of chartered high-risk tankers per month for U.S. charterers in the pre-period (i.e., Feb 2023 to Jul 2023) and post-period (i.e., Aug 2023 to Mar 2024). We list the U.S. charterers who had fixture records in both pre- and post-periods. Charterers are sorted in descending order by the number of unique tankers they chartered.

		Average no. high-risk tankers per month		
U.S. charterer	Company type	Pre	Post	Post − Pre
ExxonMobil	Public	4.667	2.625	-2.042
Chevron Corp	Public	6.833	5.000	-1.833
Phillips 66	Public	0.333	1.500	1.167
Valero Energy Corp	Public	0.833	0.375	-0.458
ATMI	Private	1.000	0.000	-1.000
Petco Petroleum	Private	0.200	0.125	-0.075
SeaRiver Maritime	Private	0.333	0.286	-0.048
Marathon Oil	Public	0.000	0.125	0.125
Hess Corporation	Public	0.250	0.857	0.607
Occidental Petroleum	Public	0.000	0.143	0.143
Newton Oil Company	Private	0.333	0.833	0.500
Houston Refining	Private	0.400	0.000	-0.400
Citgo Petroleum	Private	0.200	0.000	-0.200
Koch Industries	Private	0.000	0.000	0.000
PBF Energy	Public	0.000	0.000	0.000
Dow Chemical Company	Public	0.000	0.000	0.000
Cargill	Private	0.000	0.000	0.000
Musket Corporation	Private	0.000	0.000	0.000
Core Petroleum	Private	0.000	0.000	0.000
Sunoco LP	Public	0.000	0.000	0.000
Hartree Partners	Private	0.000	0.000	0.000
ConocoPhillips	Public	0.000	0.000	0.000
Hunt Oil Company	Private	1.000	0.000	-1.000
Novum Energy	Private	0.000	0.000	0.000
PDC Energy	Public	0.000	0.000	0.000
Tartan Oil	Private	0.000	0.000	0.000
Norvic Shipping	Private	0.000	0.000	0.000

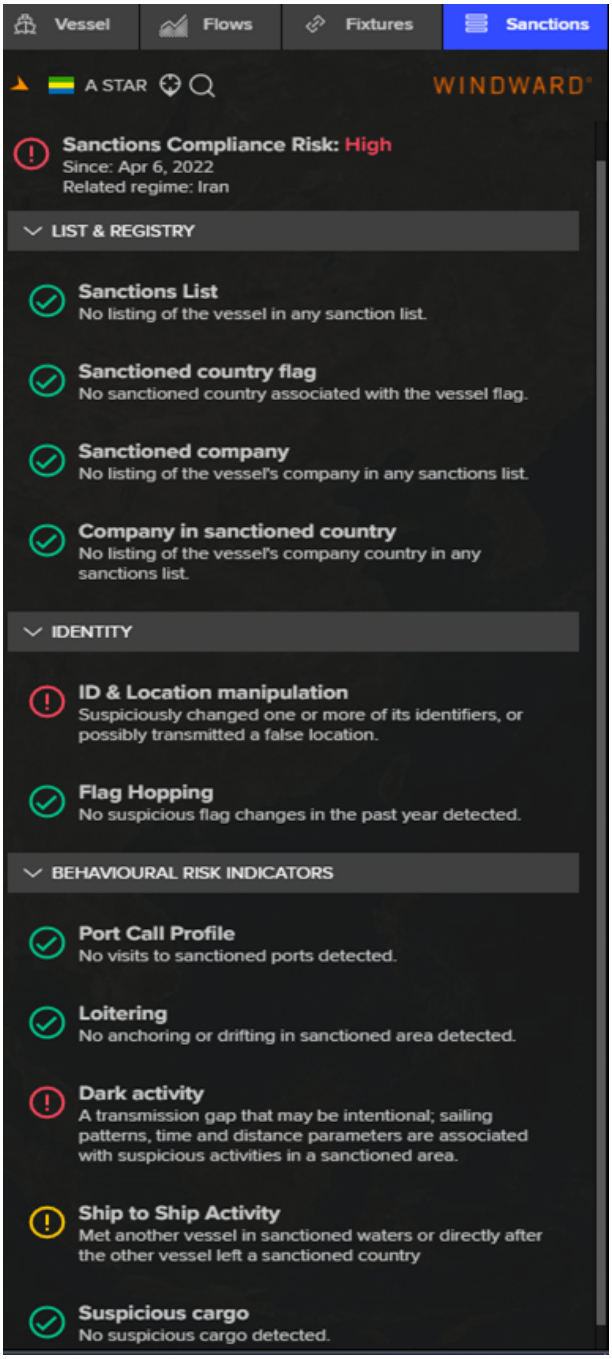


Figure A-1: This figure shows a screenshot from Refinitiv Eikon illustrating Windward's sanctions risk measures for one tanker (A STAR). The tanker is classified as high-risk with respect to the Iranian sanctions, because Windward has found strong evidence that it has changed its ID or/and has been involved in spoofing, and also has had long dark periods (transmission gaps) that can be associated with suspicious activities. Windward has also found some (weaker) evidence that the tanker was involved in ship-to-ship transfers.

References

Abadie, A. (2005), 'Semiparametric Difference-in-Differences Estimators', *The Review of Economic Studies* **72**(1), 1–19.

Heckman, J. J., Ichimura, H. and Todd, P. E. (1997), 'Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme', *The Review of Economic Studies* **64**(4), 605–654.

Imbens, G. W. (2004), 'Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review', *The Review of Economics and Statistics* **86**(1), 4–29.

Pallotta, G., Vespe, M. and Bryan, K. (2013), 'Vessel pattern knowledge discovery from ais data: A framework for anomaly detection and route prediction', *Entropy* **15**(6), 2218–2245.

Rosenbaum, P. R. and Rubin, D. B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* **70**(1), 41–55.