Graduate Institute of International and Development Studies
International Economics Department
Working Paper Series

Working Paper No. HEIDWP21-2026

An Autopsy of the Voluntary Carbon Market

Ugo Panizza
Geneva Graduate Institute & CEPR

Francesco Tripoli
Harvard Business School

Beatrice Weder di Mauro
Geneva Graduate Institute & CEPR

Chemin Eugène-Rigot 2
P.O. Box 136
CH - 1211 Geneva 21
Switzerland

©The Authors. All rights reserved. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate. No part of this paper may be reproduced without the permission of the authors.

An Autopsy of the Voluntary Carbon Market

Ugo Panizza[†], Francesco Tripoli[‡], and Beatrice Weder di Mauro[†]

[†]Geneva Graduate Institute & CEPR
[‡]Harvard Business School

July 2026 [*]

Abstract

The Voluntary Carbon Market (VCM) rests on the premise that one carbon credit compensates for one tonne of emissions. While a growing body of evidence questions the integrity of many credits, much less is known about whether the ability to offset affects firms' incentives to reduce their own emissions. Using a buyer-linked dataset covering the near-universe of offset retirements, we exploit the 2023 carbon-market scandals as a quasi-natural experiment. Firms that stop offsetting after the shock reduce their operational (Scope 1) emissions substantially more than firms that continue to purchase offsets. This finding is consistent with moral licensing: access to offsets weakens firms' incentives to undertake internal abatement. We also document persistent oversupply, opaque intermediation, and sharp price and volume responses to the scandals. Our results identify a demand-side problem in the VCM that is distinct from concerns about credit quality: even when credits represent genuine emissions reductions, the option to offset may slow decarbonization within purchasing firms.

Keywords: Voluntary Carbon Market; Carbon credits; Integrity controversy; Rebound effects; Moral hazard; GHG Emissions.

JEL: G14, L11, L51, Q54

[*]ugo.panizza@graduateinstitute.ch; ftripoli@hbs.edu; beatrice.wederdimauro@graduateinstitute.ch.
We would like to thank Franklin Allen, Patrick Bolton, Wei Cai, Kimblerley Clausing, Simon Dietz (discussant), David Hemous and Marcin Kacperczyk as well as participants to the TSE–Banque de France Research Seminar, the 1st FISF–CEPR Economics and Finance Workshop on Financing the Global Energy Transition, the 7th World Congress of Environmental and Resource Economists (WCERE 2026), the 2026 Annual Meetign of the Royal Economic Society and the 2nd CEPR Climate Change and the Environment Symposium for helpful comments. The usual caveats apply.

1 Introduction

In the academic tradition, an *anatomy* offers a systematic exploration of a phenomenon, separating it into its constituent parts in order to understand how the whole works. We began this project with precisely that ambition: to provide an anatomy of the voluntary carbon market (VCM). Yet the evidence led us somewhere else. Rather than a functioning body whose organs could be studied in isolation, we found a market showing the unmistakable symptoms of systemic failure. This paper therefore offers not an anatomy, but an *autopsy* of the VCM. We diagnose the market's decline and trace it to four mutually reinforcing forces: chronic oversupply of low quality offsets, controversies over credit integrity, opaque intermediation, and, most importantly, a demand-side moral hazard mechanism through which access to offsets can weaken firms' incentives to cut their own emissions.

The voluntary carbon market emerged in the late 1990s and early 2000s as a complement to regulated carbon markets created under the Kyoto Protocol (Kollmuss et al., 2008; Newell et al., 2013). Its appeal rested on a simple and powerful promise: one carbon credit compensates one tonne of emissions. This promise allowed the market to become a central component of corporate net-zero strategies, especially for sectors in which emissions are costly or technologically difficult to abate, such as aviation, steel, and cement (Tzachor et al., 2023; Black et al., 2024). Yet the same simplicity that made offsets attractive also made them fragile. A growing body of evidence challenges the equivalence at the heart of the market, documenting systematic over-crediting, weak safeguards against non-permanence, and methodological choices that inflate credited reductions (Haya et al., 2020; Badgley et al., 2022; West et al., 2020; Gill-Wiehl et al., 2024; Hanna et al., 2016; Song et al., 2025; Calel et al., 2025). If credits do not represent additional and durable tonnes of avoided or removed emissions, then the accounting identity on which the VCM rests collapses.

For many years, these concerns remained largely confined to specialists, NGOs, and parts of the carbon-market community. They entered the public domain forcefully in

January 2023, when a joint investigation by *The Guardian–Die Zeit–SourceMaterial* alleged severe weaknesses in rainforest offsets.[1] The investigation amplified academic critiques of leading REDD+ (Reducing Emissions from Deforestation and Forest Degradation) methodologies (West et al., 2020, 2023; Guizar-Coutiño et al., 2022) and was followed by intensified NGO campaigns and greenwashing litigation against major buyers, including the challenge to Delta's claim to be the "world's first carbon neutral airline."[2] As the reputational value of offsetting deteriorated, buyers faced a new trade-off: continuing to use offsets could preserve a familiar climate-accounting strategy, but it could also expose firms to accusations of greenwashing. Prices fell, retirements plateaued, and the market's credibility shock became visible in both quantities and valuations.

This collapse in confidence provides the setting for our central question. We take the autopsy metaphor seriously and ask not only why the VCM failed to deliver what it promised, but also what its failure reveals about the behavior of firms that had relied on it. Practitioners and NGOs have long warned that offsets may create moral hazard, or "moral licensing," by allowing firms to claim climate progress while postponing costly internal abatement (Broekhoff et al., 2019; Carton et al., 2020). In principle, well-designed climate guidance limits this risk by reserving offsets for unavoidable residual emissions after deep value-chain decarbonization (SBTi, 2021). In practice, however, firms face strong incentives to buy relatively inexpensive credits rather than undertake more costly changes to production processes, energy use, or capital equipment. The question is therefore not only whether offsets are valid on the supply side, but whether the option to offset changes behavior on the demand side.

Despite the importance of this question, systematic evidence has been missing. We do not know whether firms that rely on offsets subsequently emit more than otherwise similar firms that do not. This gap is mainly due to an inherent identification problem arising when facing this research question: firms with high emissions may be

[1]See "Revealed: more than 90% of rainforest carbon offsets by biggest certifier are worthless, analysis shows".

[2]See *Berrin v. Delta Air Lines Inc.*, Columbia Law School.

more likely to buy more offsets precisely because they emit more, making it difficult to distinguish whether offsets cause weaker abatement or simply reflect pre-existing emissions intensity. Establishing the moral-hazard channel therefore requires variation that changes firms' reliance on offsets for reasons other than their underlying emissions.

We provide such evidence by combining a buyer-linked dataset of voluntary offset transactions with firm-level environmental accounts and financial data. Our empirical strategy exploits the 2023 integrity shock as a quasi-natural experiment. Indeed, the investigation sharply reduced the credibility of major carbon-crediting methodologies and the reputational value of offsetting. For instance, Allen et al. (2026) show that this integrity shock affected how stock prices react to forestry carbon offsetting. In the wake of the shock, some firms stopped buying offsets, while others continued unchanged.[3] We use this divergence to compare the subsequent emissions trajectories of firms that exited the VCM with those that continued to offset.

The results point to a large response. A simple comparison shows that firms that stop purchasing offsets after the integrity scandals reduce their operational greenhouse-gas emissions by about 20% relative to firms that continue to buy offsets. To address endogeneity, we instrument exit using each firm's pre-scandal exposure to the credit segments most affected by the integrity shock. The instrumental-variables estimates are even larger: firms that stop buying offsets reduce emissions by nearly 60% relative to firms that continue offsetting. These effects are concentrated in operational Scope 1 emissions, with no significant differential adjustment in Scope 2 or Scope 3 emissions. This pattern is consistent with a moral-licensing mechanism: when firms lose access to a credible offsetting strategy, they appear to substitute toward internal abatement.

Literature: Our analysis contributes to several strands of literature. First, we speak to work documenting structural weaknesses in offset programmes and crediting methodologies (Gill-Wiehl et al., 2024; West et al., 2020; Cabiyo et al., 2025; OECD, 2024; Stolz

[3] The low-cost airline EasyJet is an example of a firm that decided to exit the offset market: it announced that it would no longer rely on offset purchases and would instead reallocate resources toward fleet renewal, operational efficiencies, and sustainable aviation fuels. See "EasyJet to stop offsetting CO_2 emissions from December".

and Benedict, 2025; NewClimate Institute and Carbon Market Watch, 2024; Günther et al., 2020) and to research on how financial markets react to offset integrity (Allen et al., 2026). This literature shows that many credits may fail to deliver the mitigation they claim. Our first contribution is to extend the diagnosis from the supply-side of the market to the demand one, which has largely remained under-researched. Indeed, we show that even if current efforts to improve credit quality were successful, the mere access to offsetting may still weaken firms' incentives for direct decarbonization.

Second, we contribute to the literature on corporate net-zero strategies and climate governance (Foerster, 2023; Tilsted et al., 2023; Boiral et al., 2024; Burger et al., 2022; Stolz and Benedict, 2025). A related paper is Kim et al. (2026). Their central message is that carbon offset use is concentrated among firms for which offsets are a comparatively cheap way to manage perceived transition performance: low-emission industries, firms with high marginal abatement costs, and firms with weaker climate governance rely more heavily on offsets, often of lower quality, after an ESG-rating shock. In other words, their paper explains *who buys offsets* and shows that offset demand is shaped by the relative cost of internal abatement, ESG incentives, and governance constraints. Building on this work, we instead ask what happens when the credibility of offsetting itself is damaged and firms can no longer rely on offsets as easily. By exploiting a shock that made offsetting less credible and induced some firms to exit the market, we identify a different economic mechanism: while Kim et al. (2026) show that offsets are disproportionately used where internal abatement is less attractive or climate governance is weaker, we show that continued access to offsetting itself weakens real decarbonization incentives.

Third, we bring market-based evidence to debates on rebound effects and moral licensing in environmental policy (Greening et al., 2000; Sorrell et al., 2009; Gillingham et al., 2016; Tiefenbeck et al., 2013; Burger et al., 2022). The literature has long recognized that policies or technologies intended to reduce environmental harm can be partially offset by behavioral responses. We document an analogous mechanism in corporate climate strategy: the availability of offsets can relax the pressure to reduce emissions at the source.

Layout: The remainder of this paper is structured as follows. Section 2 provides a primer on the VCM for readers unfamiliar with its origins, structure, and key participants. Section 3 documents the failure symptoms shown by the voluntary offset market, particularly its persistent oversupply and the role of intermediaries in shaping prices and mark-ups. Section 4 describes the 2023 integrity scandal and shows that, consistent with a negative demand shock, it led to lower prices and lower carbon-offset retirements. Section 5 then exploits this shock to estimate the effect of exiting the VCM on firm-level emissions and interprets the resulting gaps as evidence of moral licensing in the use of offsets. Section 6 concludes by discussing implications for the future role of voluntary offsets in global climate governance.

2 A Primer on the Voluntary Carbon Market

The voluntary carbon market (VCM) emerged in the late 1990s and early 2000s as a complement to compliance carbon markets created under the Kyoto Protocol (Kollmuss et al., 2008; Newell et al., 2013). At its core is the carbon credit: a transferable claim that one metric tonne of carbon dioxide equivalent (CO_2e) has been reduced or removed relative to a counterfactual baseline. Credits are generated by project developers, whose main activity is the design and implementation of projects intended to prevent, reduce, or remove greenhouse-gas emissions, and whose primary source of revenue is the sale of the credits they generate. Purchasers can retire these credits against their own emissions, thus compensating residual emissions by financing mitigation activities undertaken by another actor.

Historically, the voluntary carbon market is closely linked to the Clean Development Mechanism (CDM). This project-based carbon offset system, established under the Kyoto Protocol, allowed emission-reduction projects in emerging and developing economies to generate Certified Emission Reductions (CERs), which advanced economies could use to meet their climate commitments. The CDM introduced many of the institutional features that later migrated into the VCM and, during the late 2000s, generated a large stock of CERs. Demand for these credits collapsed after the end of the first Kyoto commitment period and the tightening of CER eligibility in major compliance markets such as the EU Emissions Trading System. As a result, many CDM methodologies, project developers, and legacy credits transitioned into the voluntary market. This legacy has shaped both the scale and structure of today's VCM, while also carrying over longstanding concerns about integrity.

Unlike compliance systems, such as the EU Emissions Trading System, China's national ETS, or California's cap-and-trade program, which impose legally binding emission caps and trade a fixed supply of allowances, the VCM is not driven by regulatory mandates. Instead, it allows firms and individual consumers to voluntarily offset or "compensate" emissions by purchasing credits. Lacking a centralized authority that defines uniform rules for project eligibility, credit issuance, or credit use, the VCM relies on third-party registries that develop methodologies, oversee project validation and verification, issue credits, and record their retirement. Each registry sets its own rules for eligibility, baselines, monitoring, and crediting, and maintains publicly accessible books of retired credits. Major registries include the American Carbon Registry, Gold Standard, Climate Action Reserve, and Verra.

In addition to registries, a large ecosystem of intermediaries has emerged around the VCM. Brokers, exchanges, and consultants facilitate transactions and market access, while specialized rating agencies, such as Calyx, Sylvera, and BeZero, assess credit quality. These assessments typically examine whether projects generate emissions reductions or removals that would not have occurred in the absence of credit revenues (additionality), whether emissions reductions in one location are accompanied by increases in emissions elsewhere (leakage), the durability of emissions reductions or removals over time (permanence), and the soundness of the methodologies used to quantify and verify outcomes. The coexistence of multiple registries, methodologies, and rating systems contributes to heterogeneity in credit quality and to opacity in market pricing.

Alongside the registries and intermediaries that govern supply, a separate set of institutions shapes how firms disclose their emissions and their use of offsets. The most prominent is CDP, formerly the Carbon Disclosure Project, a not-for-profit organization that runs the largest voluntary corporate environmental disclosure system. Each year, thousands of firms report their Scope 1, Scope 2, and Scope 3 emissions, their climate targets, and whether they retired carbon credits to compensate emissions. Investors, customers, and regulators increasingly treat these disclosures as a benchmark for corporate climate performance, which gives firms a strong incentive to present a favorable net-emissions position. CDP therefore matters for the VCM in two ways. First, it provides much of the visibility into which firms buy offsets and how they fold those offsets into public climate claims. Second, by anchoring evaluation to a net-emissions metric, it sharpens the trade-off at the center of this paper: a firm can improve its reported position either by cutting emissions at the source or by retiring credits.

Credits are generated by a wide range of project types that differ markedly in technologies, costs, and environmental risks. A useful organizing distinction is between emissions avoidance or reduction projects, which claim to reduce emissions relative to a baseline, and carbon removal projects, which aim to physically remove CO_2 from the atmosphere and store it over long horizons. Historically, the VCM has been dominated by avoidance-based credits, with removal credits representing a relatively small share of issuance and retirements.

Forestry and land-use projects constitute one of the largest categories by volume. These include afforestation and reforestation, improved forest management (IFM), and avoided deforestation projects, often grouped under REDD+. Such projects generate credits by increasing or preserving carbon stocks in biomass and soils relative to a counterfactual scenario. While they can produce large volumes of low-cost credits and offer biodiversity or development co-benefits, they are also associated with significant concerns regarding additionality, permanence, and leakage. Most forestry credits are best characterized as avoided-loss credits rather than removals, as they rely on continued protection of existing carbon stocks.

Renewable energy projects were among the earliest and most prevalent sources of voluntary offsets. These projects generate credits by displacing fossil-fuel electricity generation, but their importance in the VCM has declined as renewable technologies have become cost-competitive and increasingly supported by public policy. Energy efficiency projects, which reduce emissions by lowering energy consumption through more efficient equipment or processes, share similar characteristics and challenges, relying heavily on modeled baselines and usage assumptions.

Household and community-level projects, such as improved cookstove devices, have historically generated a substantial share of credits, particularly in developing countries. These projects are typically classified as avoidance-based and are often justified by health and development co-benefits, but they face scrutiny over baseline inflation, monitoring difficulties, and the persistence of behavioral change.

By contrast, carbon removal projects remain a small segment of the VCM. These include nature-based removals such as biochar and some soil carbon projects, as well as technological approaches like direct air capture with storage (DACCS) and bioenergy with carbon capture and storage (BECCS). Removal credits are typically more expensive and scarcer but are often viewed as conceptually closer to the "one tonne out" promise of offsetting, particularly when storage is durable and verifiable.

Grounded in these objectives, the VCM expanded rapidly, reaching a peak of roughly $2 billion in market value and approximately 500 $MtCO_2$ in transacted volume (Swinfield and Scott, 2025). Yet integrity concerns have intensified, and high-profile investigations and litigation have increased scrutiny of widely used methodologies and of the credibility of offset-based claims. As confidence eroded, both prices and transaction volumes fell sharply, setting the stage for the market integrity shock analyzed in Section 4.

3 Market Landscape: Supply, Demand, and Prices

Our analysis builds on a comprehensive dataset from AlliedOffsets, which aggregates information from all major public registries and project developers and links credit retirements to identifiable buyers. The database covers 36,444 offsetting and removal projects worldwide and contains detailed project-level information, including registry and standard, sectoral classification, host country and geographic coordinates, crediting period, credit type, methodology, and accreditation program. For each issuance, transfer, and retirement event, the dataset records the project identifier, the volume of credits, the transaction date, buyer information, and – when available – an estimated unit price derived from broker quotes, trading platforms, legacy exchanges, and crypto exchanges.

The dataset allows us to reconstruct both aggregate market dynamics and the composition of offset portfolios at the level of individual buyers. Throughout the paper, we focus on the period from 2006, when retirement volumes become non-negligible and registry coverage relatively complete, to 2025. Transactions are classified as *issuances*, *retirements*, or *other* events, including registry cancellations and offtakes. We construct project-level and buyer-level aggregates by summing the corresponding volumes. On the buyer side, AlliedOffsets links retirements to corporate entities using registry account names, buyer metadata, and external corporate identifiers, yielding a transaction-level dataset that tracks the credit retirements of individual firms over time.

For the firm-level analysis of Section 5, we first aggregate retirements at the buyer-month level. Buyer names are standardized by lowercasing, removing punctuation, harmonizing common legal suffixes such as "Inc", "Corp", "Ltd", and "PLC", and applying a set of audited name aliases. We then compute annual retirement totals and classify firms' post-shock offsetting behavior relative to their 2022 pre-shock activity. We complement the registry-based treatment classification with CDP offset-use disclosures from 2022 and 2023. These disclosures ask whether the company retired any carbon credits in the corresponding year. Firms that report retiring credits in 2022 and again in 2023 are classified as continuing offset users, while firms that report retiring credits in 2022 but not in 2023 are classified as exiters[4]. Because registry account names and corporate names are often imperfectly aligned, the matching procedure combines several layers of information: we first use exact identifiers where available, prioritizing ISINs and then tickers, and then use exact cleaned names and finally a NLP-constructed and manually audited name matching.

We measure firm-level greenhouse-gas (GHG) emissions primarily from CDP, the largest corporate environmental disclosure system, which covers our full 2017–2024 panel. For Scope 1 and Scope 2, we keep company-disclosed values and exclude CDP estimates.[5] For Scope 3, firms typically report Scope 3 separately across upstream, downstream, travel, procurement, use-of-sold-products, and other categories. Thus, we sum all disclosed category-level values a firm reports in a given accounting year, again excluding vendor estimates. We rely on CDP alone for 2017–2023 and use Refinitiv only to fill missing 2024 observations.[6]

Finally, we obtain financial controls from Compustat Capital IQ. All accounting variables are converted to U.S. dollars using average yearly exchange rates. The baseline regressions control for firm size (using log revenue and log total assets), profitability (using return on assets), and leverage.[7]

---

[4] When the registry-based treatment classification and the CDP offset-use disclosure conflict, we give priority to the CDP disclosure to addresses a limitation of registry data: credits are sometimes retired by intermediaries on a firm's behalf, so a firm may not appear under its own name in the registry. However, this issue is only marginal as we account for this by scraping the retirement reasons, which often specifies the firm on whose behalf offsets were canceled.

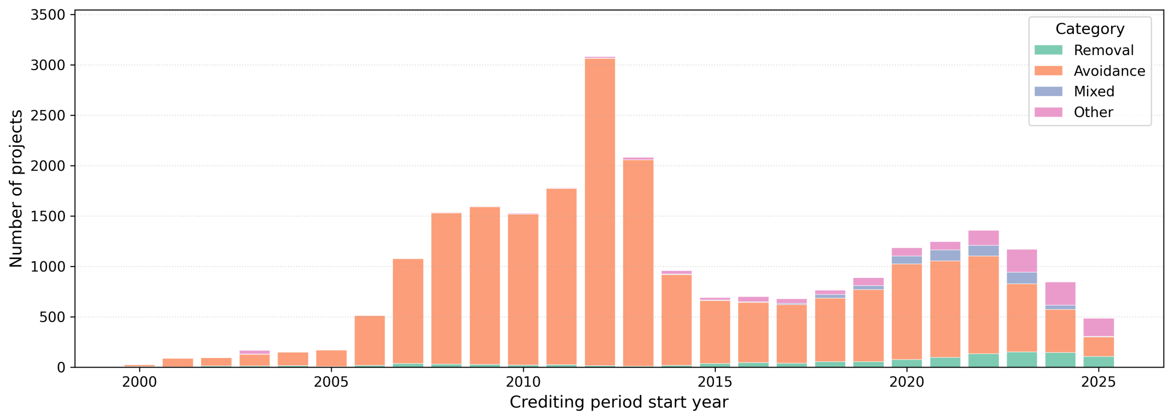
[5] For Scope 2 we use market-based emissions where available and location-based emissions otherwise.

[6] CDP offers the most internally consistent reporting structure over the full sample period, but its 2024 coverage is incomplete for some firms in the matched panel. We therefore use Refinitiv in 2024 only, and only when the corresponding CDP value is missing, which extends the final post-shock year without mixing data-generation processes during the pre-treatment period. We match Refinitiv to the same harmonized firm universe using ISIN, ticker, and cleaned exact names.

[7] Return on assets is $ROA_{it} = \frac{OIBDP_{it}}{AT_{it}}$ and levearge is defined as $Leverage_{it} = \frac{DLTT_{it}+DLC_{it}}{AT_{it}}$. Where AT is total assets, $OIBDP$ is operating income before depreciation, $DLTT$ is long-term debt, and DLC is debt in current liabilities. We winsorize at the 2nd and 98th percentiles.

Figure 1: VCM Projects by Crediting-Period Start Year.

The stacked bars show the number of projects that originate in a given year. Orange bars denote avoidance projects, green bars removal projects, light blue bars denote mixed projects and pink bars other types of projects.



Supply

Our data show that the pipeline of VCM projects grew in two distinct waves. The first wave began in the mid-2000s and ran into the early 2010s, when project origination expanded rapidly under the institutional momentum of the Clean Development Mechanism and the early growth of voluntary standards. Origination then slowed, before a second and smaller wave emerged after 2018, as corporate net-zero pledges revived demand for credits (Figure 1).

Across both waves, one feature stands out: avoidance projects dominate the pipeline, while removals account for only a marginal share of supply. This composition matters for what offsetting can achieve. A removal credit corresponds to a physical withdrawal of CO_2 from the atmosphere, for example through afforestation or soil carbon sequestration, so retiring one such credit for each tonne emitted could in principle restore the atmospheric stock to its counterfactual level. An avoidance credit is different. A project earns it by emitting less than a baseline, for example when a renewable plant displaces fossil generation or an improved cookstove replaces wood burning. Retiring an avoidance credit prevents an additional tonne elsewhere, but it does not undo the tonne the buyer emitted. In the terms of a global carbon balance sheet, avoidance moves the bottom line from two tonnes to one, not from one tonne to zero. Even under optimistic assumptions about additionality and permanence, the VCM has primarily financed the avoidance of incremental emissions rather than the removal of carbon already in the atmosphere. The market has concentrated in the segment whose climate equivalence is hardest to establish, rather than in the removals that come closest to the "one tonne out" promise of offsetting.

Supply is also uneven in geographic space. Project activity is concentrated in China, India, Central America and the Eastern United States (see Figure 2 that maps registered projects by sector). Forestry and land-use projects cluster in tropical forest regions, notably the Amazon basin, Central America, the Congo basin, and Southeast Asia, while renewable-energy projects concentrate in fast-growing emerging economies such as India, China, Brazil and South Africa. Europe and North America host a mix of renewable, forestry and industrial projects, but on a smaller scale relative to their economic size, due to both alternative regulatory instruments and the historical role of the Clean Development Mechanism.

Not all registered projects translate into retired credits, and the gap between the two has widened over time. Between 2006 and 2011, developers issued roughly 1 $GtCO_2e$ of credits, yet buyers retired almost none. By 2025, cumulative issuances had climbed to approximately 5 $GtCO_2e$, leaving a stock of nearly 3.5 $GtCO_2e$ of issued but unretired credits (Figure 3). The market has therefore accumulated a large overhang of unretired supply.

This oversupply need not be a problem in itself. If every project met a uniformly high standard of integrity, abundant supply would simply let environmentally motivated firms and consumers buy genuine mitigation at low cost. The difficulty arises because offset quality varies widely and buyers cannot easily observe it. Additionality, permanence, and leakage are hard to verify, and the developer who designs a project knows far more about its true climate value than the buyer who retires its credits. This information asymmetry places the VCM squarely in the territory of Akerlof's market for lemons (Akerlof, 1970). When buyers cannot tell good offsets from bad, they refuse

Figure 2: Global Distribution of VCM Projects by Sector.

Each point in the map denotes a registered project with differnt colors corresponding to different sectors. The map shows that supply is concentrated in a limited set of host countries, especially in emerging economies and tropical forest regions.

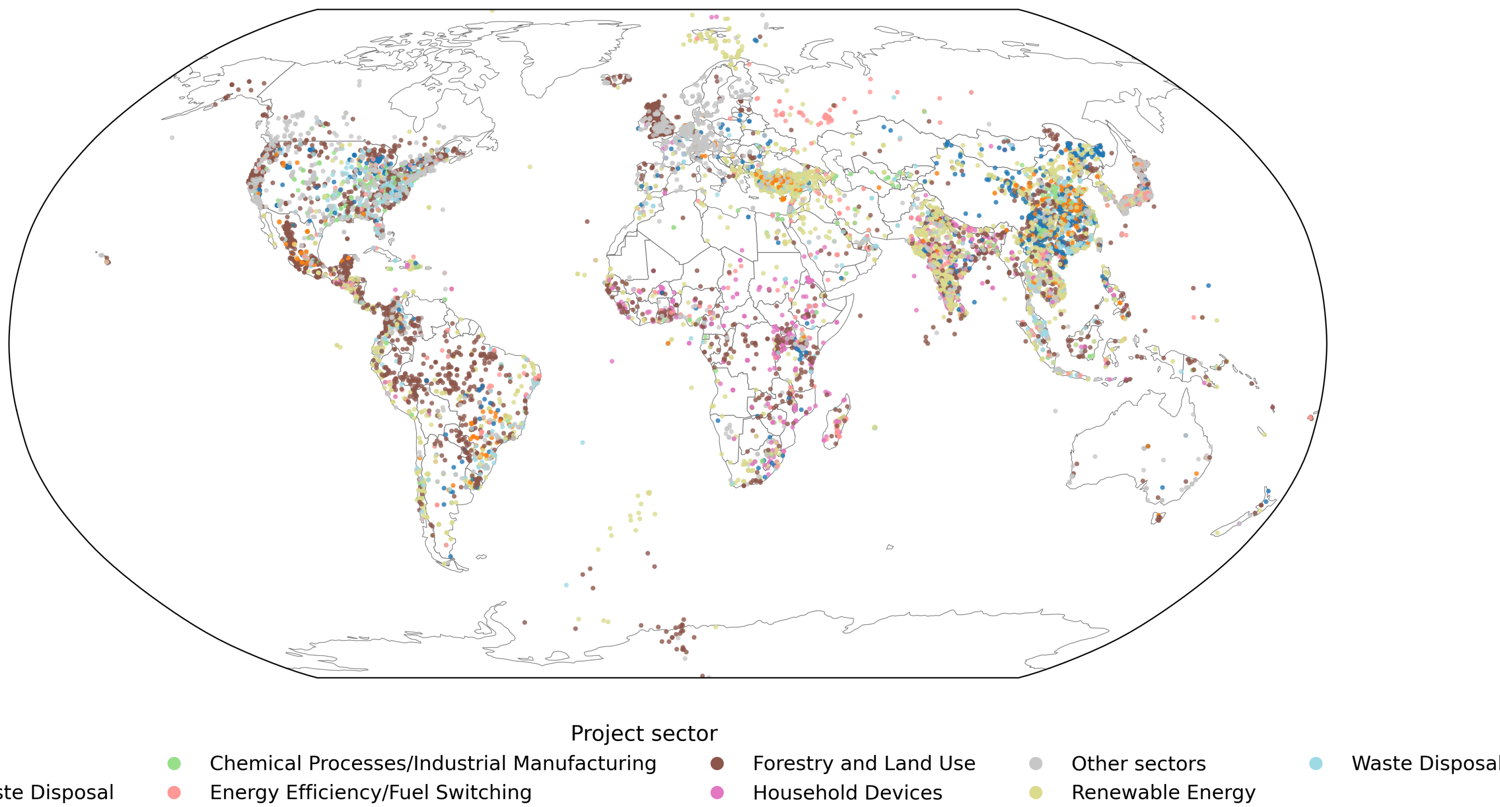
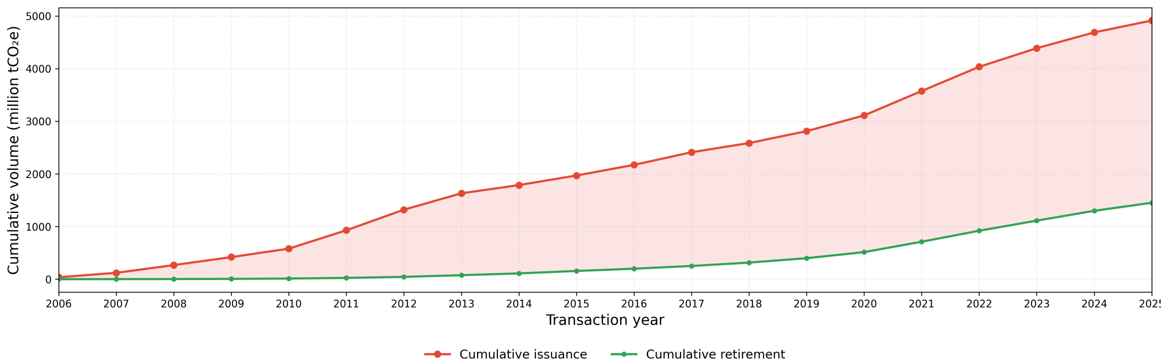


Figure 3: Cumulative Issuance and Retirements in the VCM

The upper line (in red) plots cumulative issuance and the lower line (in green) plots cumulative retirements. The difference between the two lines is the stock of outstanding, unsold credits which amounts to more than 3 $GtCO_2e$.



to pay a premium for quality they cannot confirm, so high-integrity projects, which cost more to develop, struggle to command prices that cover their costs and gradually exit the market.

The demand side adds a further problem. In the textbook lemons market, buyers want high quality, and the information asymmetry frustrates them. Here, some buyers do not care about quality. A firm that uses offsets mainly to support a climate claim or a marketing narrative has little reason to pay for genuine impact, and every reason to select the projects that best fit its budget. For these buyers, low integrity is not a bug but a feature. The result is a market in which some buyers cannot identify quality and others would rather not pay for it.

These two forces converge through a third. If every carbon credit is treated as one tonne, a low-integrity credit and a high-integrity credit carry the same nominal claim at sharply different prices. When buyers either cannot tell the two apart or do not care to, they buy the cheaper one, which sets off a dynamic akin to Gresham's law: cheap, low-quality credits drive out the costly, high-quality ones. A buyer can claim a large volume of offsetting at low cost simply by purchasing the cheapest credits available, which weakens demand for credible but expensive projects.

On the positive side, the market does signal quality to some degree. The pricing regressions discussed below show that removal credits command a premium, and that prices vary systematically with registry, sector, geography, and certification. Observable attributes, however, capture only part of what determines a credit's true integrity. Additionality, permanence, and leakage depend on hidden characteristics that prices reflect imperfectly, so even a market that differentiates on visible features can misprice the quality that matters most. The integrity scandals we analyze in the next section reveal exactly this gap.

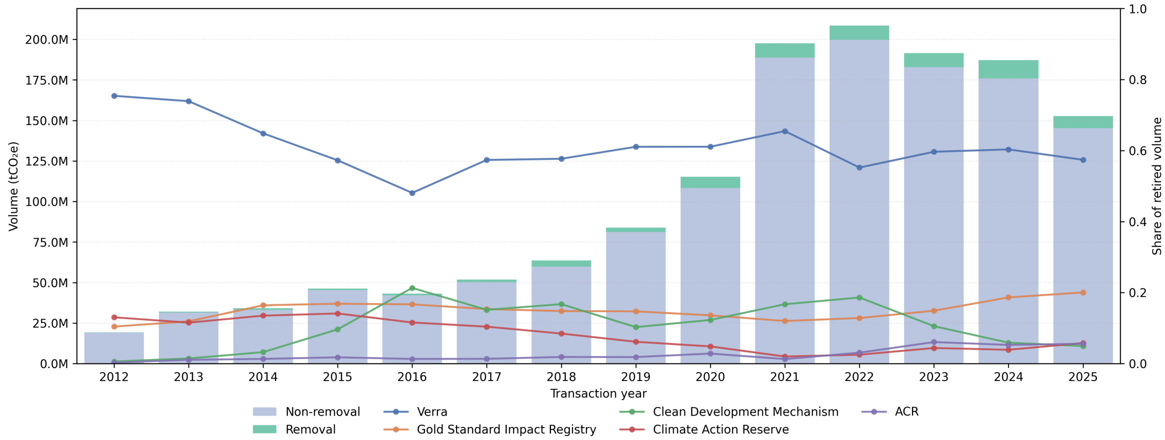
Demand

We now set the behavioral argument aside and document who buys credits, what they buy, and which projects ultimately supply retired volumes. The main buyers are firms in carbon-intensive and consumer-facing sectors, such as airlines, energy utilities, materials and heavy manufacturing, alongside financial institutions and large technology firms with notable climate commitments. Even among these buyers, removal credits account for only a small fraction of portfolios. Most volumes come from renewable-energy, forest-conservation, waste-management, and household-energy projects, mirroring the supply composition in Figure 1. On the project side, a small number of large, often flagship projects account for a disproportionate share of total retirements. Many are renewable-energy or industrial installations registered under the CDM or Verra, with long crediting periods and large installed capacities. Others are large-scale forestry and land-use projects that attracted substantial corporate demand in the late 2010s and early 2020s, when biodiversity co-benefits rose to prominence.

The time profile of demand shows that retirements grew rapidly between 2019 and 2021, then plateaued after the 2022–2023 integrity scandals. The registry composition shifted in parallel. Early retirements drew heavily on CDM credits, reflecting the carry-over of Kyoto-era Certified Emission Reductions into voluntary portfolios. After the first Kyoto commitment period ended, CDM retirements declined and Verra's Verified Carbon Standard expanded rapidly to become the main registry. Gold Standard, the American Carbon Registry, and domestic schemes such as the Australian Carbon Credit

Figure 4: Retirements and Registry Market Shares

The bars show annual retired volumes split into removal and non-removal credits. The lines show the share of each major registry.



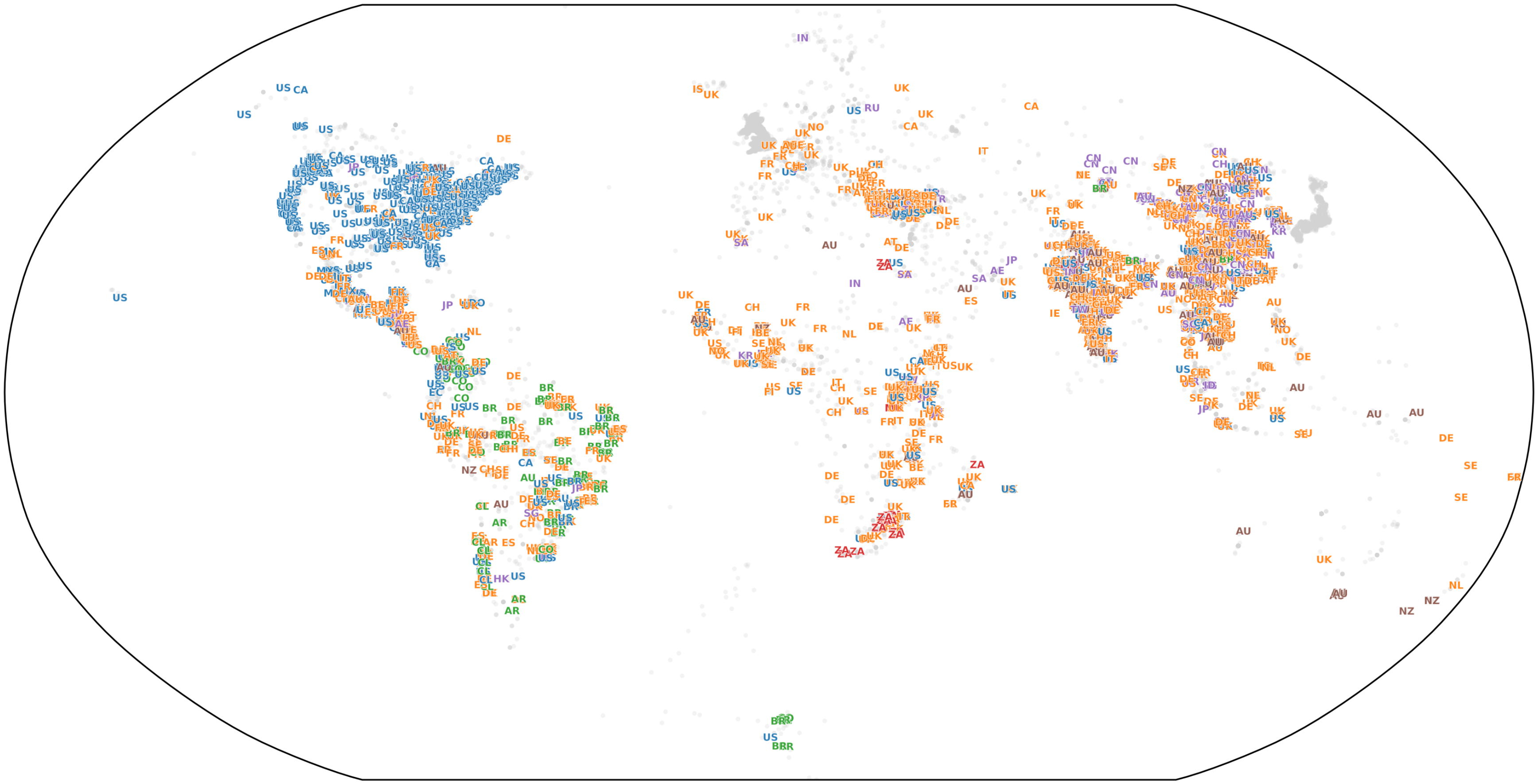
Units play smaller but non-negligible roles (see Figure 4, which plots annual retirement volumes and registry market shares).

Demand also carries a strong geographic signature, one that differs markedly from the geography of supply. Buyers cluster in Europe, North America, East Asia, and Oceania, and they tend to purchase from projects either in their own region or in a limited set of preferred host regions. North and South American firms source credits mainly from projects within their respective continents, and African and Oceanian buyers display a similar regional preference, though less pronounced. European buyers stand apart, with a highly diversified sourcing pattern that spans projects on all major continents. East Asian buyers occupy a middle ground, combining strong regional sourcing from China and Southeast Asia with selected purchases elsewhere (see Figure 5, which annotates each project location with the modal ISO country code of its buyers' headquarters).

Taken together, the supply- and demand-side facts portray a market burdened by a considerable oversupply of credits, with cross-regional flows shaped by registry choices and buyer geography.

Figure 5: Geographical Distribution of Retired Credits by Buyers' Headquarters Country.

Each label marks a project and indicates the modal headquarter country code among buyers retiring credits from that project.



Prices

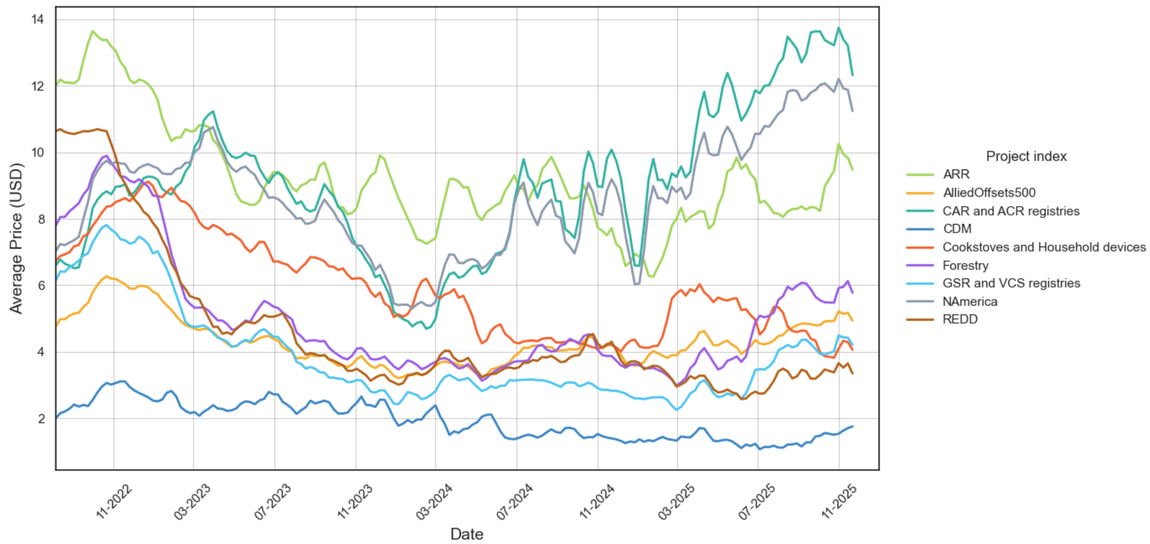
We now turn from quantities to prices and ask how the market values different types of credits. To track price dynamics across project types, we look at price indexes defined as volume-weighted averages of the 20 projects with most retirements in each segment of the market. The overall *AlliedOffsets 500* index, defined as the volume-weighted average of the top 500 projects, serves as a market-wide benchmark, while the remaining series track specific groups such as North American projects, afforestation-reforestation (ARR), and forestry credits.[8]

Three patterns emerge. First, the market splits into two broad price tiers. Credits on CAR and ACR, North American projects, and nature-based removal (ARR) consistently trade above the overall benchmark, while legacy CDM projects and many land-use avoidance credits sit in a lower price band. Second, CDM prices stay persistently depressed and decline slowly over the sample, consistent with their legacy status and limited role in contemporary offsetting. Third, across most non-CDM segments, prices display a pronounced common downturn in early 2023. The fall is particularly sharp for forestry and Gold Standard- or Verra-certified credits, whose prices drop abruptly, immediately after the joint *Guardian–Die Zeit–SourceMaterial* investigation into rainforest offsets. This common movement anticipates the reaction to the integrity shock we analyze in the next section: the price indexes already suggest that the market absorbed the scandals as a negative revaluation. The reaction was not uniform, however: some segments moved little at the time of the scandal, while CAR and ACR registries and North American projects even rose initially before declining only a few months later and then recovering again (Figure 6).

[8]AlliedOffsets does not directly observe prices for all transactions, and confidentiality agreements prevent it from reporting even the prices it does observe directly. It therefore reports estimated credit prices from a machine-learning model that "nowcasts" the price of credits by combining hundreds of weekly samples of bid, ask, and transaction prices from brokers, exchanges, and developers with project-level characteristics. AlliedOffsets reports that this model explains about 97% of the variation in observed sample prices. The share of directly observed prices has risen over time, and AlliedOffsets reports that it directly observes prices for at least 50% of retirements since 2022. We therefore restrict our analysis of price patterns to the 2022–2025 period.

Figure 6: Price Indexes by Project Type

Each line shows the average weekly prices of project indexes constructed from the largest projects by retirements in a given category.



To move beyond the aggregate indexes, we examine how prices vary across project characteristics. Specifically, we regress the log of the credit price on project and issuer attributes by ordinary least squares, including month fixed effects throughout and clustering standard errors at the project level. The estimates confirm the visual patterns from Figure 6. First, registry of issuance shapes prices substantially: the steepest discount in the table is for CDM, which is consistent with its legacy status. Relative to ACR, credits issued under the Clean Development Mechanism trade between 72% and 78% lower in columns 1 to 3 of Table 1 (these values are obtained by applying the transformation $\exp(\beta) - 1$ to coefficients that range between -1.235 and -1.516). Verra credits trade between 33% and 57% below ACR across specifications, while BioCarbon Standard credits trade between 42% and 60% below ACR in the specifications in which they are estimated. Gold Standard credits sell at a significant discount without geographic controls (column 1). However, the latter become indistinguishable from ACR once we condition on project location (columns 2 to 4), suggesting that part of the apparent registry gap reflects geographical differences.The dummies for Bio Carbon standards, CDM, and Woodland Carbon Code are collinear with vintage (these standards concentrate among legacy projects) and scale. We therefore cannot estimate them separately once we control for these variables in column 4 of Table 1.

Second, project type and sector are particularly relevant in explaining the variation in prices: removal credits command a premium that ranges between 77% and 92% over avoidance credits (columns 1 to 4 of Table 1), and mixed avoidance-removal credits a premium that ranges between 38% and 48%. The market, therefore, differentiates and pays substantially more for credits that come closer to the "one tonne out" ideal. Across sectors, the dispersion is wide. Relative to household-device credits, biochar offsets trade roughly thirty-eight times higher and utilization credits about twelve times higher (In the raw data, the median price of offsets linked to household device projects is approximately \$7, that for renewable energy is about \$2, and the median prices for biochar and utilization projects are \$130 and \$160) in the complete specification, while waste-disposal credits are not statistically different from the baseline. Renewable energy is the cheapest segment, trading about 31% below the baseline, in line with the aggregate indexes in Figure 6.

Third, geography also matters: projects in advanced economies carry a large premium, while projects in Asia trade at a discount. Relative to Asian projects, which are the cheapest, European and North American projects price about twice as much, and African and South American projects between 37% and 48% higher in the specifications with geographic controls (columns 2 to 4 of Table 1).

Fourth, in terms of compliance eligibility, CORSIA or Article 6 approval correspond to a premia of about 7% and 9% in the complete specification, respectively, but only CORSIA elegibility is statistically significant (columns 3 and 4 of Table 1). A letter of authorization adds about 34% to the price (column 4). Size and age also matter: micro- and small-scale projects price about 13% and 14% above large projects, and each additional year of vintage raises the price of about 4%, consistent with buyer preferences for credits closer in time to retirement.

Table 1: Determinants of Carbon Credit Prices

This table reports a set of OLS regressions where the dependent variable is the log of the credit price and the explanatory variable are project and issuer characteristics. The omitted categories are ACR for registry, avoidance or reduction for project type, household devices for sector, Asia for geography, and large for scale. Some factor variable are omitted for space constraints. Vintage is scaled by 2005.

	(1)	(2)	(3)	(4)
Registry (ref: ACR)				
BioCarbon Standard	-0.906***	-0.565***	-0.545***	
	(0.100)	(0.107)	(0.107)	
Clean Development Mechanism	-1.516***	-1.235***	-1.276***	
	(0.084)	(0.080)	(0.093)	
Gold Standard	-0.381***	0.020	0.019	-0.005
	(0.071)	(0.072)	(0.072)	(0.068)
Verra	-0.847***	-0.485***	-0.469***	-0.400***
	(0.067)	(0.070)	(0.070)	(0.067)
Woodland Carbon Code	0.206**	0.238**	0.240**	
	(0.089)	(0.101)	(0.100)	
Project type (ref: avoidance/reduction)				
Mixed Avoidance/Removal	0.386***	0.389***	0.379***	0.320***
	(0.062)	(0.057)	(0.057)	(0.061)
Removal	0.626***	0.573***	0.577***	0.652***
	(0.079)	(0.080)	(0.079)	(0.078)
Project sector (ref: household devices)				
Biochar	3.387***	3.834***	3.821***	3.663***
	(0.092)	(0.100)	(0.099)	(0.099)
Renewable energy	-0.733***	-0.421***	-0.433***	-0.372***
	(0.022)	(0.043)	(0.043)	(0.040)
Utilization	2.973***	2.763***	2.752***	2.535***
	(0.088)	(0.101)	(0.100)	(0.097)
Waste disposal	-0.184**	-0.105*	-0.109**	-0.020
	(0.072)	(0.055)	(0.054)	(0.049)
Geography (ref: Asia)				
Africa		0.389***	0.390***	0.323***
		(0.043)	(0.044)	(0.041)
Europe		0.708***	0.717***	0.696***
		(0.040)	(0.038)	(0.048)
North America		0.684***	0.683***	0.639***
		(0.048)	(0.048)	(0.047)
South America		0.348***	0.343***	0.316***
		(0.045)	(0.045)	(0.047)
Governance / quality				
Article 6 registered (True)			0.079	0.087
			(0.079)	(0.075)
CORSIA eligible (True)			0.094***	0.063**
			(0.027)	(0.027)
Scale (ref: large)				
Micro scale				0.124***
				(0.025)
Small scale				0.128***
				(0.021)
Letter of Authorization (True)				0.290***
				(0.098)
Vintage				0.042***
				(0.003)
Intercept	2.392***	1.634***	1.626***	0.971***
	(0.071)	(0.083)	(0.083)	(0.087)
Observations	23005	23000	23000	20422
R^2	0.719	0.752	0.753	0.732
Month FE	Yes	Yes	Yes	Yes

Standard errors in parenthesis are clustered at the project level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Finally, to assess how much of the price variation reflects project features rather than variation in the intermediates involved in it, we estimate an expanded pricing regression that adds developer and verifier to the project and characteristics in column 4 of Table 1, and apply a Shorrocks–Shapley decomposition to this specification. Thus, we assign each group of covariates its average marginal contribution to the model's fit across all possible orderings of independent variable, which yields a model-agnostic allocation of the regression R^2 that does not depend on the order in which we add each group of regressors (Shapley, 1953).[9]

We find that, on average, intermediary identity matters more than the project itself (Table 2). The three intermediary groups, developer, registry, and verifier, jointly account for about 53% of the explained variance, against 27% for project characteristics such as type, sector, methodology, country, vintage, and size. The developer alone explains 31% of the variance, more than all project attributes combined, while the verifier and the registry contribute 14% and 8%. Month fixed effects absorb the remaining 19%.

These findings indicate that branding, reputation, and market power among intermediaries shape carbon-credit prices well beyond project fundamentals. In a market where buyers struggle to observe quality directly, the identity of the developer or verifier serves as a price-relevant signal in its own right, consistent with the lemons logic discussed above. Using observed rather than estimated transaction prices, Berg et al. (2025) similarly find that credit prices are highly dispersed and that buyer and intermediary identity explains much of this variation, well beyond project fundamentals.

[9]Formally, let $\mathcal{G}$ denote the set of groups of regressors, and let $v(S)$ be the value function that maps any subset $S \subseteq \mathcal{G}$ to the regression fit when only variables in S are included (in our case, $v(S)$ is the R^2 from the corresponding specification). The Shapley value for group $g \in \mathcal{G}$ is

$$\phi_g \;=\; \sum_{S \subseteq \mathcal{G} \setminus \{g\}} \frac{|S|!\,\big(|\mathcal{G}| - |S| - 1\big)!}{|\mathcal{G}|!} \Big[\, v(S \cup \{g\}) - v(S) \Big],$$

which averages the marginal contribution of g to the fit over all possible subsets S that could precede it. By construction, the Shapley values satisfy $\sum_{g \in \mathcal{G}} \phi_g = v(\mathcal{G})$, so that the total explained variance is exactly decomposed into additive contributions from each group.

Table 2: Shapley Decomposition of Credit Price Variation.

The table shows the result of a Shorrocks–Shapley decomposition of the expanded pricing regression. Each entry reports the Shapley value (contribution to R^2) and the corresponding percentage of the explained variance. Project characteristics group includes project type, sector, CORSIA / Article 6 approval, CCP likelihood, methodology, country, vintage, size.

Group	Shapley value (R^2)	Share of explained variance (%)
Developer	0.2492	30.61
Registry	0.0689	8.46
Verifier	0.1119	13.74
Project Characteristics	0.2222	27.29
Total	0.8142	100.00

4 The 2023 Integrity Shock

The previous section showed that several project prices plummeted in early 2023. We now focus on the event behind that break: a joint investigation by *The Guardian*, *Die Zeit*, and *SourceMaterial* was published on January 18th, 2023, challenging the integrity of rainforest offsets issued under leading REDD+ methodologies. In the rest of this section, we show how this episode came as an integrity shock to the voluntary carbon market by tracing its effects on prices and retirement volumes.

We estimate a regression discontinuity design (RDD) in time that exploits the sharp timing of the January 2023 piece (Imbens and Lemieux, 2008; Cattaneo et al., 2020). Thus, we estimate the following model:

$$y_t \;=\; \alpha \;+\; \tau\,\mathbf{1}\{t \geq 0\} \;+\; f_-(t)\,\mathbf{1}\{t < 0\} \;+\; f_+(t)\,\mathbf{1}\{t \geq 0\} \;+\; \varepsilon_t, \tag{1}$$

where y_t is the outcome variable (either the average price of offsets or the total volume of retirements in month t), t indexes months and equals zero in January 2023, so that $t < 0$ denotes pre-shock months and $t > 0$ post-shock months, $\mathbf{1}\{\cdot\}$ is an indicator function, $f_-(\cdot)$ and $f_+(\cdot)$ are smooth functions capturing the pre- and post-shock time trends, and τ is the parameter of interest, measuring the discrete jump in the conditional expectation of y_t at the discontinuity.

We estimate Equation (1) by ordinary least squares over the full sample window, letting $f_-(\cdot)$ and $f_+(\cdot)$ be low-order polynomials in event time: a quadratic on each side for the price series and a linear trend on each side for retirement volumes. Because we fit global polynomials over the entire window, the design involves no bandwidth or kernel choice, and the procedure is fully deterministic. We report Newey-West standard errors with six lags, which are robust to heteroskedasticity and serial correlation in the time series. Retirements are expressed as an annualized rate (the quarterly total scaled to an annual figure) so that its level is comparable to standard annual market statistics.

We find that before the scandal offset prices followed a clear upward trend, which the shock halted and partly reversed (top panel of Figure 7). During the pre-shock window, average prices were low and broadly flat in 2018–19 (close to \$3–4 per credit). They then climbed steeply over 2021 and 2022, with the fitted trend approaching \$10 and individual months exceeding \$11 in the run-up to the scandal. In January 2023 the series drops discretely: the estimated jump is about \$2.1 per credit ($t = -2.0$), and the post-shock fitted path starts well below the extrapolated pre-shock trend, bottoming out around \$7 in 2024 before a partial recovery. The discontinuity implies a sharp and economically large fall in average prices relative to the counterfactual.

Retirement volumes tell a similar story. Before the scandal, retirements rose almost linearly through the 2019–2022 boom, from an annualized rate of roughly 60 million credits at the start of the window to about 250 million just before the shock. The scandal halted this expansion as retirements plateaued at the cutoff, with a drop of about 60 million credits a year ($\hat{\tau} = -0.58 \times 10^8$, $t = -3.7$).

Taken together, these patterns are consistent with a textbook negative demand shock associated with a revision of beliefs about credit integrity and heightened reputational risk for buyers. As confidence in credit integrity fell, buyers became less willing to pay for offsets and retired fewer of them, which pushed prices down and halted the growth in volumes.

It is important to note that the scandal was driven by external investigations into credit methodologies and registry practices, not by changes in the emissions of indi-

Figure 7: Regression Discontinuity around the January 2023 Shock.

Dots show observed outcomes over time; solid lines plot local–polynomial fits on either side of the cutoff. The top panel shows the effect of the shock on prices and the bottom panel that on retirements.

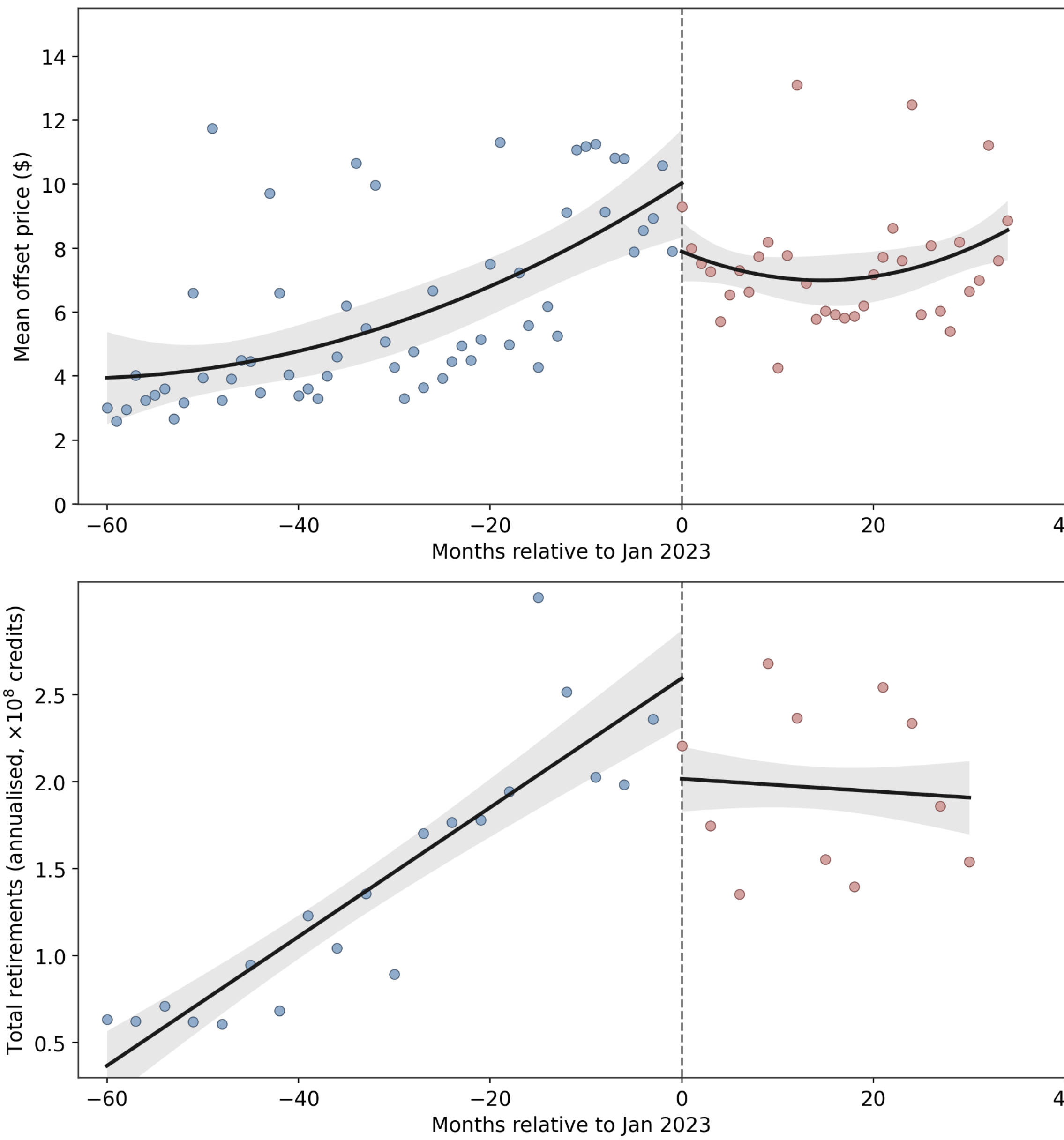


Table 3: Heterogeneous Response to the January 2023 Integrity Shock

The table reports the discontinuity τ at January 2023 from the RDD of Equation (1), estimated separately for each segment with the same specification as Figure 7: monthly average log price with quadratic trends on each side of the cutoff, and log retirement volume with linear trends on each side. Newey–West (HAC, 6 lags) t-statistics are reported in parentheses.

	Price jump (τ)		Volume jump (τ)	
	τ	(t)	τ	(t)
Panel A: Whole market and forestry exposure				
Whole market	-0.346^{**}	(-2.12)	-0.468^{**}	(-2.09)
Forestry & land use	-0.456^{***}	(-3.59)	-0.302	(-1.01)
All non-forestry	-0.282	(-1.21)	-0.558^{**}	(-2.44)
Panel B: By issuing registry				
Verra	-0.482^{**}	(-2.29)	-0.477	(-1.64)
Gold Standard	-0.584^{***}	(-3.31)	-0.338^{*}	(-1.68)
ACR	-0.603^{***}	(-4.01)	$+0.201$	(0.55)
CAR	-0.241^{**}	(-2.02)	$+0.719$	(1.40)
CDM	-0.036	(-0.37)	-0.918^{***}	(-2.87)
Other	-0.270	(-1.40)	$+0.344$	(0.79)
Panel C: By project sector				
Forestry & land use (REDD+)	-0.456^{***}	(-3.59)	-0.302	(-1.01)
Renewable energy	-0.554^{**}	(-2.08)	-0.689^{***}	(-2.77)
Household devices	-0.362^{**}	(-2.54)	-0.003	(-0.01)
Waste & methane	-0.271	(-1.47)	-0.654^{**}	(-2.02)
Industrial	-0.628^{**}	(-2.17)	$+0.911$	(1.31)
Energy efficiency	-0.782^{**}	(-2.16)	$+0.381$	(0.30)
Removal tech	-0.084	(-0.37)	-1.668^{**}	(-1.96)
Agriculture	-0.653^{***}	(-2.62)	-0.034	(-0.03)
Other	-0.278	(-1.21)	$+0.343$	(0.90)

Significance: * 10%, ** 5%, *** 1%.

vidual companies. We can therefore use it as a quasi-natural experiment in which the shock led some firms to stop buying carbon credits altogether. Furthermore, we note that the aggregate discontinuity masks substantial heterogeneity across segments of the market, and this heterogeneity will be central to our identification strategy. Table 3 re-estimates the price and volume discontinuities separately by issuing registry and project sector, using the same specification as in Figure 7. The results show how the repricing is concentrated exactly where the investigation struck: forestry and land-use

(REDD+) credits fall by about 37%, and credits issued under the two large voluntary registries under scrutiny, Verra and Gold Standard, fall by 38% and 44% respectively. Instead, CDM credits, which were not the subject of the integrity concerns, show no discernible price change. The segments most tightly linked to the scandal are thus the ones the market re-evaluated most sharply, while unrelated segments – although subject to some spillovers – were less affected. This differential exposure is precisely the variation we will exploit for identification in Section 5.2.

5 Offsetting and Firm-Level Decarbonization

In this section we provide evidence for the moral-licensing argument, using the 2023 integrity scandal as an exogenous shock to the perceived value of offsetting. We first lay out a simple framework that links a firm's reliance on offsets to its incentive to cut emissions at the source and delivers testable predictions about how firms should respond to the scandal. We then take these predictions to the data.

Consider a firm that wants to keep its reported emissions below a threshold, which can come from a public target or a stakeholder expectation. In a world of high integrity, where one offset equals one tonne and everyone believes it, the firm can meet the target in two ways. It can invest in abatement and cut emissions at the source, or it can buy and retire offsets. Because abatement and offsets are substitutes, each firm picks its own mix of the two. The mix depends on relative costs and on the firm's type. Firms with cheap abatement options lean toward abatement. Firms also differ in how much they care about their true carbon footprint. At one extreme sit firms that care about real emissions, because of regulation, transition risk, financing conditions, or genuine managerial preferences. At the other extreme sit greenwashers, who value the public image of climate action far more than the underlying tonnes. These firms lean toward offsets, and toward cheap ones in particular.

Now assume that an investigation reveals that some offsets carry far less integrity than buyers believed. The shock does not touch every credit equally: it exposes specific segments and leaves others intact. Firms that care about real emissions and held high-integrity offsets see no change in their problem. Their credits remain sound, so they should barely react, unless the scandal spills over to the credibility of the whole offset market, a case we discuss below. Conversely, firms that care about real emissions but held offsets the scandal exposed, face a different situation. For them, a tonne of offset no longer buys a tonne of compensation, so the effective price of meeting the target through offsets rises. They respond in one of two ways. They switch toward segments the scandal left intact, or they buy fewer offsets and spend more on abatement. Either way, exposure pushes them toward real emission cuts. Greenwashers behave differently. They care little about the underlying tonnes, so a fall in offset integrity matters less to them. They react weakly, if at all.

The shock also reshapes the market for credits. Demand falls for the exposed segments, so their price and volume drop. The effect on the other segments depends on how buyers respond. If they rotate toward sound credits, demand rises for the unaffected segments and pushes their price and volume up. If instead the reputational damage spills across the whole market, buyers grow wary of all offsets, not only the exposed ones. Price and volume then fall everywhere, but they fall more for the affected segments. Under this second scenario, every firm that cares about its footprint trims its offset purchases, and firms that held exposed credits trim the most.

We therefore expect that: (i) firms that stop offsetting after the scandal should cut operational emissions more than firms that continue, with the response concentrated in emissions under the firm's direct control (Scope 1); (ii) exit from the offset market should be more likely among firms that, before the scandal, were most exposed to the discredited projects; and (iii) the emissions response should be stronger where reputational stakes and feasible abatement are higher.

Table 4 profiles the firms in our sample that stopped offsetting after the scandal (exiters) against the ones that continued (stayers). Exiters are large, publicly listed multinationals drawn from across the economy. They are headquartered mainly in Europe (42%) and North America (25%), with the remainder in Asia-Pacific and Latin Amer-

Table 4: Characteristics of exiters and stayers

In the left panel, summary statistics for the firms in our sample are reported split by exiters (firms that retired offsets in 2022 but not after the shock) and stayers (firms that retired offsets both in 2022 and after the shock); headquarters region and sector are defined based on CDP data. The right panel lists, for each group, one firm at different percentiles of the pre-shock Scope 1 emissions.

	Exiters	Stayers
Headquarters region (% of firms):		
North America	25	23
Europe	42	42
Asia-Pacific	14	17
Latin America	8	8
Other / missing	11	10
Sector (% of firms):		
Services & finance	41	44
Manufacturing	14	13
Materials & mining	11	6
Infrastructure & utilities	8	8
Fossil fuels	4	5
Retail	7	6
Transport	3	4
Other	12	15
Firm characteristics (median):		
Revenue ($bn)	6.0	9.6
Total assets ($bn)	12.6	23.4
Scope 1 ($ktCO_2e$)	38.1	28.9
Credits retired 2022	1,257	1,880
Return on assets	0.103	0.085
Leverage	0.289	0.261
Hard-to-abate (%)	11	12
Controversy (%)	17	22

Pctile	Firm	Scope 1
Exiters, by Scope 1 percentile:		
p99	Adani Power	53.6 Mt
p96	ConocoPhillips	17.4 Mt
p92	BHP	9.8 Mt
p88	Avis Budget	5.9 Mt
p85	Verallia	2.3 Mt
p80	China State Constr.	875 kt
p70	GeoPark	251 kt
p60	Brown-Forman	112 kt
p50	Viña Concha y Toro	38 kt
p30	Swiss Life	7.3 kt
p20	Emcor UK	3.3 kt
p10	Arcadyan	1.1 kt
Stayers, by Scope 1 percentile:		
p99	Shell	63.4 Mt
p96	Fortum	34.6 Mt
p92	Origin Energy	16.3 Mt
p88	Deutsche Post	6.9 Mt
p85	Nestlé	3.3 Mt
p80	Fonterra	1.6 Mt
p70	Takeda	307 kt
p60	Flex	85 kt
p50	Commerzbank	29 kt
p30	Wolters Kluwer	5.2 kt
p20	AIA Group	2.4 kt
p10	Castellum	0.5 kt

ica, a geographic spread almost identical to that of stayers. They also span sectors: services and financial firms are the largest group (41%), followed by manufacturing, materials and mining, infrastructure and utilities, transport, energy, and retail. Along observable characteristics exiters look much like stayers, somewhat smaller (median revenue of $6.0 billion versus $9.6 billion), but comparable in profitability, leverage, and Scope 1 emissions, no more concentrated in hard-to-abate sectors and, if anything, slightly less likely to have faced an environmental controversy. This similarity is reassuring for a design that identifies the effect of exit from differential exposure to the scandal.

A natural concern is whether exiters can actually cut their own emissions, or whether dropping offsets simply reflects abandoning a target they have no operational means to meet. What makes the abatement channel plausible is that the largest direct emitters among the exiters (right panel of Table 4) operate in sectors where concrete source-abatement technologies exist. Power generators can shift their generation mix; airlines can substitute toward Sustainable Aviation Fuel (SAF) and more fuel-efficient aircraft; logistics firms can electrify ground fleets; and industrial-gas and shipping firms can raise process and fuel efficiency. The extent to which these companies can actually shift towards those abatement options, though, sometimes depends on the availability of climate solutions such as SAF in the case of airlines, which reflects the hard-to-abate nature of some sectors. In any case, absent a moral hazard channel in which firms had substituted abatement for offsets before the shock, we should see no differential change in emissions once those firms stop retiring credits.

We now take the above hypotheses to the data. We first describe the correlation between exiting the offset market and firm-level emissions and then provide causal estimates.

5.1 A First Look at the Exit-Emissions Gap

We start by comparing firms that were active offset users before the January 2023 shock but made different offsetting choices afterward. We focus on firms with positive retirements in 2022 and define *exiters* as firms that retired no credits from February 2023 onward, and *stayers* as firms that kept retiring credits after the shock.[10]

We then compare how emissions evolve for exiters and stayers before and after the

[10] To expand the sample, we rely on CDP disclosure on carbon-offset use and classify firms whose status is otherwise unknown as exiters if they report retiring credits in 2022 but not in 2023, and as stayers if they report retiring credits in both years. The result is robust to how we classify exit: using registry retirements only, using CDP offset-use disclosures only, and varying the post-shock window used to define exit all leave the Scope 1 estimate in a tight range (Appendix Table A1).

shock. Let i index firms and t index years. For each outcome y_{it}, defined as the natural logarithm of a firm's GHG emissions, we estimate the following two-way fixed-effects difference-in-differences specification:

$$y_{it} = \alpha_i + \lambda_t + \beta \left(\text{Post}_t \times \text{Exit}_i\right) + X'_{it}\Gamma + \varepsilon_{it}, \tag{2}$$

where α_i are firm fixed effects, λ_t are year fixed effects, Post_t is an indicator equal to one for years $t \geq 2023$, and Exit_i is a dummy equal to one for firms that stop offsetting after the shock. The coefficient of interest is β, which captures the differential post-shock change in emissions for exiters relative to stayers. The vector X_{it} includes log revenue, log assets, profitability, and leverage. We cluster standard errors at the firm level to allow for serial correlation in firms' emissions. We use emissions levels rather than an emissions-intensity ratio because a ratio would impose a unit elasticity of emissions with respect to revenue, an assumption the data reject.

The identifying assumption is that exiters and stayers would have followed parallel emissions trends absent their post-shock divergence in offsetting behavior, once we condition on firm fixed effects, common year shocks, and time-varying controls. This assumption is more plausible here than in a comparison between offset users and non-users, because we focus only on firms that were active VCM participants before the shock. Since the decision to exit remains endogenous, Equation 2 does not estimate a causal treatment effect. It provides preliminary evidence that firms that responded differently in their offsetting decisions also adjusted their emissions. The next subsection addresses causality directly.

We find that firms that exit offsetting cut operational (Scope 1) emissions by about 19% more than comparable firms that continue, a difference that is significant at the 1% level (column 1 of Table 5, where the point estimate of -0.210 corresponds to $\exp(-0.210) - 1 \simeq -0.189$). This gap is hard to reconcile with the view that offsetting is only used as a last resort for unabatable emissions. Conditional on firm and year fixed effects and on financial and size controls, firms that stop offsetting reduce their

Table 5: Difference-in-Differences Estimates of GHG Emissions.

The table reports difference-in-differences estimates of greenhouse-gas (GHG) emissions following the 2023 integrity shock. Columns (1), (3)–(5) compare firms that *stopped* retiring offsets after the shock to firms that continued offsetting. Column (2) compares firms that reduce retirements by at least 70% without fully exiting to firms that continue offsetting. Dependent variables are log absolute emissions. Reported in parentheses are t-statistics (robust, clustered at the firm level).

	Scope 1		Scope 2	Scope 3	Total
	(1)	(2)	(3)	(4)	(5)
Post shock × Exit	-0.210^{***}		-0.178	0.147	-0.000
	(-2.71)		(-1.60)	(0.70)	(-0.00)
Post shock × Reduce		-0.084			
		(-0.86)			
Log(Revenue)	0.320^{***}	0.212^{**}	0.195	0.337	0.343
	(3.44)	(2.22)	(1.61)	(1.30)	(1.60)
Log(Assets)	0.178	0.206	0.414^{***}	0.271	0.124
	(1.52)	(1.48)	(2.63)	(0.94)	(0.55)
ROA	-0.441	-0.346	-1.013^{*}	-1.481	-1.253
	(-1.31)	(-1.07)	(-1.68)	(-1.47)	(-1.56)
Leverage	0.165	-0.015	0.021	-0.208	-0.646
	(0.62)	(-0.05)	(0.05)	(-0.30)	(-1.36)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	2,616	2,059	2,616	2,616	2,616

Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

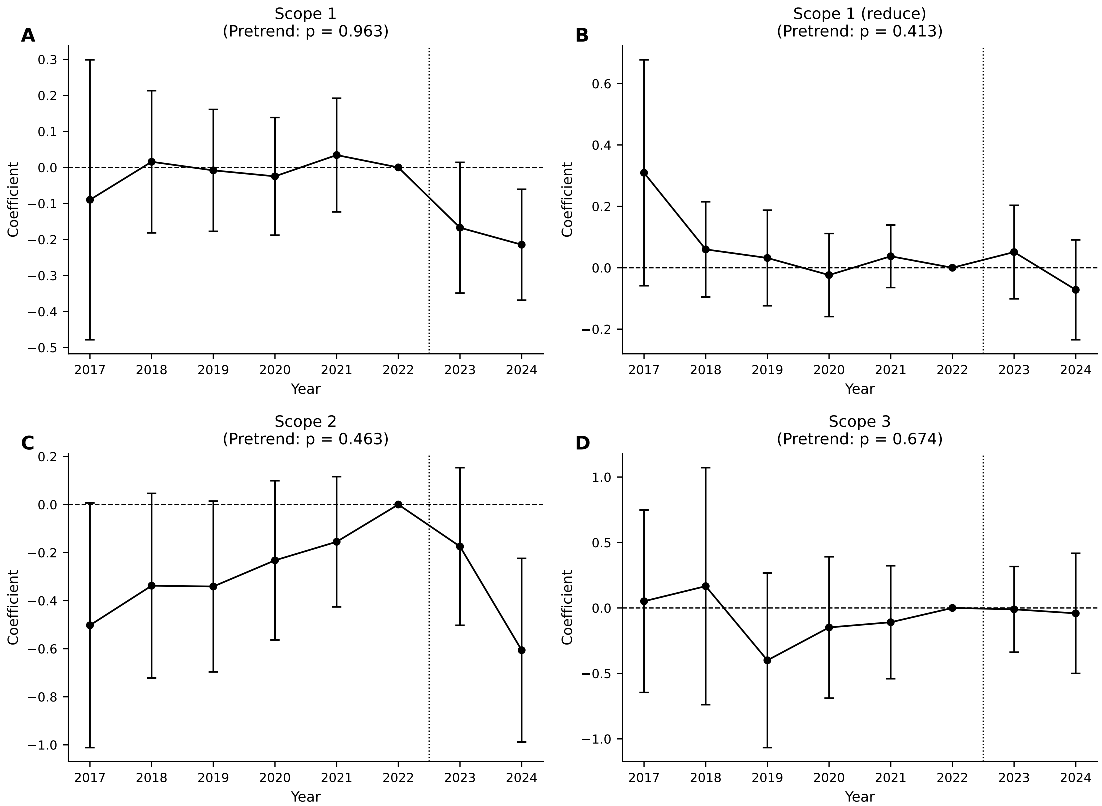
own emissions more than firms that keep relying on offsets.

What matters is stopping altogether, not simply buying fewer credits. We define *reducers* as firms whose annualized post-shock retirements fall by at least 70% relative to their 2022 baseline but that do not fully exit the market. For these firms the coefficient is negative but much smaller and not statistically significant (column 2).

The response concentrates in operational emissions, the category most directly tied to a firm's own production and process choices. For emissions that managers control less directly over short horizons, or that carry greater measurement and boundary

Figure 8: Event-study of the January 2023 Integrity Shock.

The figure plots year-specific treatment coefficients from two-way fixed-effects event-study regressions, omitting 2022 as the baseline. The top left panel report estimates for Scope 1 emission for firms that exit offset, the top right panels report results for Scope 1 emissions for firms that reduce but do not fully exit offsetting, the bottom panels report results for Scope 2, and Scope 3 emissions for firms that exit the offset market.



complexity, we find no differential adjustment: Scope 2 emissions, which are still within firm control but less compared to Scope 1, are reduced but the coefficient is imprecisely estimated; the Scope 3 coefficient is positive but not statistically significant and total-emissions, of which Scope 3 are the largest contributor, are basically unchanged.

An event study confirms the baseline pattern and shows that exiters and stayers shared common trends before the shock (Figure 8, which omits 2022, the last pre-shock year, as the baseline). For Scope 1 emissions, the pre-treatment coefficients are small and indistinguishable from zero, both for firms that exit the offset market (panel A) and for firms that only reduce their purchases (panel B). The joint pre-trend tests fail to reject parallel trends in both cases (p = 0.963 and p = 0.413). After the shock, Scope 1 emissions fall significantly for exiters, while reducers show no comparable break. This reinforces the result in Table 5 that exit, rather than a partial cut in purchases, is the relevant margin. These patterns do not rule out that exit is endogenous, but they show that exiters and stayers were not already on different emissions trajectories before the shock.

The evidence for Scope 2 and Scope 3 emissions is weaker (panels C and D). Confidence intervals are wide both before and after the shock, so we cannot pin down a clear effect, although Scope 2 emissions of exiters drop sharply in 2024.

5.2 The Effect of Discontinuing Offsetting

The baseline difference-in-differences estimates described above compare firms that were all active offset users before the January 2023 shock but made different offsetting choices afterwards. This design has the advantage of avoiding a direct comparison between offset users and non-users, and it exploits a common shock to the credibility and reputational value of voluntary offsetting. It nonetheless treats exit as exogenous, which is not. Whether a firm stops buying and retiring carbon credits after the scandal is ultimately an endogenous choice that may hinge on firm characteristics and preferences.

If *exiters* differ systematically from firms that continue relying on carbon offsets along unobserved dimensions that also affect post-2023 emissions trajectories, the estimates of the previous section suffer from endogeneity bias. For example, exiters may have already planned abatement investments, may differ in governance quality, or may have experienced operational shocks that coincided with the scandal. The difference-in-differences design would then conflate the causal effect of quitting the VCM with selection into exit, biasing the estimated treatment effect in an indeterminate direction.

To address this concern, we recast the analysis as an encouragement design and estimate it with an instrumental-variables strategy. The strategy starts from the observation that the scandal raised the reputational and perceived-effectiveness costs of using offsets, but not equally across firms. The strength of the shock depended on the composition of each firm's pre-scandal offset portfolio. Firms whose 2022 retirements concentrated in segments that the scandal hit hardest, as signaled by a sharp drop in price, faced a stronger encouragement to exit. Indeed, having used these now-discredited offsets to back carbon-neutral claims, they were the most exposed to the reputational scrutiny that followed, as in the suit over Delta's "carbon-neutral airline" claim (cfr. Section 1). We measure this exposure from pre-shock portfolio composition, which provides variation in the intensity of the shock-induced encouragement to quit that is plausibly unrelated to a firm's own post-shock emissions choices. In practice, we instrument exit with a Bartik-style shift-share design (Bartik, 1991; Goldsmith-Pinkham et al., 2020), where the pre-scandal composition of a firm's offset portfolio provides the shares and the segment-level price drop provides the shift.

We proceed as follows. We first define offset-market segments as cells given by the interaction of issuing registry and project sector. We group registries into Verra/VCS, Gold Standard, CDM, ACR, CAR, and Other, and we organize projects sectors as: forest and land use, household devices, renewable energy, waste and methane, industrial, energy efficiency, removal technologies, agriculture, and other. For each registry-sector cell, we estimate the market-adjusted price decline following the scandal. This cell-level price decline is the shift component of the instrument, while each firm's pre-scandal retirement portfolio provides the share component. Because the instrument is built from each firm's pre-shock portfolio of retirements, the IV sample is restricted to firms with matched 2022 registry transactions. Indeed, in case of firms classified only from CDP offset-use disclosures we would be unable to identify the projects whose credits those firms purchased. This leads to a slightly smaller sample than the difference-in-differences specification.

The instrument takes the form

$$Z_i = \sum_k s_{ik}\, g_k,$$

where s_{ik} is the exposure of firm i to registry-sector cell k before the shock, and g_k is the intensity of the integrity shock for cell k (Appendix B gives details on the construction of the instrument). We instrument the endogenous treatment $D_{it} = Post_t \times Exit_i$ with $Post_t \times Z_i$ and estimate the following two-stage least squares (2SLS):

$$D_{it} = \alpha_i + \lambda_t + \pi(Post_t \times Z_i) + X_{it}'\Gamma_1 + u_{it}, \tag{3}$$

$$y_{it} = \alpha_i + \lambda_t + \beta\widehat{D}_{it} + X_{it}'\Gamma_2 + \varepsilon_{it}, \tag{4}$$

where y_{it} is log Scope 1 emissions, α_i and λ_t are firm and year fixed effects, and the control vector X_{it} includes log revenue, log assets, profitability (ROA), and leverage. Because the excluded instrument is a shift-share exposure built from common registry–sector price shocks, we base inference on standard errors that are robust to this structure, following Adão et al. (2019). However, our inference does not hinge on the choice of clustering as the results are consistent with conventional standard errors clustered at the firm level, leaving every significance conclusion unchanged.

The first stage confirms that the instrument is relevant: pre-scandal exposure to adversely repriced segments strongly predicts post-scandal exit (column 1 of Table 6). The coefficient on $Post_t \times Z_i$ is 0.545 ($t = 3.88$), so a one-log-point increase in portfolio-weighted price-drop exposure raises the post-shock exit indicator by 0.545. The cluster-robust first-stage F-statistic is 15.04, above the conventional weak-instrument threshold of 10 (Kleibergen and Paap, 2006; Stock and Yogo, 2005).

The exclusion restriction requires that pre-shock exposure to adversely repriced segments affects post-2023 emissions only through exit from offsetting. Several features of the design support this condition. First, we measure the portfolio shares entirely before the shock, using 2022 retirements, so they cannot mechanically reflect post-scandal operational changes. Second, we demean the price shocks by market-month and estimate them at the registry×broad-sector level, so they capture differential segment repricing. Third, the instrument rests on the composition of a firm's offset portfolio, not on the total level of offset use. Conditional on firm and year fixed effects and time-varying financial controls, it is hard to see why pre-shock exposure to particular segments would affect a firm's subsequent Scope 1 emissions except through the decision to stop relying on offsets. Consistent with this, we show in the Appendix Table B2 that firms with high and low instrument intensity are balanced on pre-shock size, profitability, leverage, emissions, sectoral abatement type, and reputational exposure, while differing only in their exit rate, i.e. the intended first-stage channel. Monotonicity is also natural here: greater exposure to segments whose relative prices fell after the scandal should weakly raise the incentive to exit.

Moving to the second stage, we find that firms induced to exit the offset market cut Scope 1 emissions sharply. The raw IV estimate is negative, significant, and large in magnitude, implying a reduction of more than 70% ($\exp(-1.23) - 1 \approx -0.705$; column 2 of Table 6).

This effect is larger than the difference-in-differences estimate of the previous section, which implied a reduction of about 20%. Three factors plausibly account for this. First, the two estimators answer different questions. The difference-in-differences coefficient averages over all firms that exit, while the IV coefficient is a local average treatment effect for compliers, the firms whose exit the shock itself triggered (Imbens and Angrist, 1994). Compliers are firms whose portfolios concentrated in the discredited segments and who left the market once those credits lost credibility. These are the firms our framework expects to respond most, because they had leaned on cheap, now-exposed credits to meet their targets. The average exiter, by contrast, includes firms that left for unrelated reasons and might have adjusted less. Second, exit is measured with error. We classify it from registry matches and CDP disclosures, and both misclassify some firms. Such measurement error attenuates the difference-in-differences estimate toward zero, and the IV corrects part of this attenuation. Third, a few firms with extreme emissions changes may carry heavy weight in the raw IV. We now turn to this latter issue.

Changes in Scope 1 emissions are fat-tailed. This issue concerns the change in emissions rather than the cross-sectional level, so winsorizing the level of emissions

Table 6: IV Estimates of Scope 1 Emissions.

The table reports two-stage least squares estimates where the endogenous regressor is $D_{it} = Post_t \times Exit_i$ and the excluded instrument is $Post_t \times Z_i$. All regression controls for firm and year fixed effects and log revenue, log assets, profitability (ROA), and leverage (coefficients not reported). Column (1) reports the first-stage regression for the full sample. Column (2) reports the second stage using the raw Scope 1 log-emissions outcome. Columns (3)–(7) use the 10/90 winsorized Scope 1 outcome. The bottom panel reports the relevant first-stage$F$-statistic for each sample. Second-stage t-statistics in parentheses are exposure-robust to the shift-share structure of the instrument, following Adão et al. (2019). The first stage reports the cluster-robust coefficient and the Kleibergen–Paap F-statistic.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Scope 1		Hard-to-abate		Controversy	
	1^{st} stage	Level	Bound	Yes	No	Yes	No
D		-1.236^{**}	-0.894^{***}	0.285	-1.060^{***}	-1.372	-0.853^{***}
		(-2.57)	(-3.34)	(0.27)	(-2.92)	(-0.71)	(-3.32)
Z	0.545^{***}						
	(3.88)						
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs	1,657	1,657	1,657	183	1,474	485	1,172
1^{st} stage F	15.04	15.04	15.04	3.45	12.33	1.63	13.38
R^2	0.297	—	—	—	—	—	—

Significance: *** p<0.01, ** p<0.05, * p<0.10.

does not address the variation that identifies β in the IV specification. Because the endogenous regressor is D_{it}, identification comes from changes in firms' emissions after the shock. A firm with a moderate emissions level can still post an unusually large change, and such changes exert substantial leverage on the IV estimate. To address this problem, we apply a bounded-change transformation to Scope 1 emissions.[11]

[11] Let

$$\Delta y_i = \frac{1}{|\mathcal{T}_1|}\sum_{t\in\mathcal{T}_1} y_{it} - \frac{1}{|\mathcal{T}_0|}\sum_{t\in\mathcal{T}_0} y_{it}, \qquad \mathcal{T}_0 = \{2021, 2022\}, \quad \mathcal{T}_1 = \{2023, 2024\},$$

where y_{it} is log absolute Scope 1 emissions. The pre-period uses the last two full years before the integrity shock, so the capped change corresponds to the emissions adjustment around the event rather than to longer-run firm drift. Let $q_{0.10}$ and $q_{0.90}$ denote the 10th and 90th percentiles of Δy_i in the IV sample, and define $\Delta y_i^B = \min\{\max(\Delta y_i, q_{0.10}), q_{0.90}\}$. We then construct the bounded-change panel outcome as $y_{it}^B = y_{it} + \mathbf{1}[t \geq 2023]\left(\Delta y_i^B - \Delta y_i\right)$. The 10/90 threshold is intentionally conservative since the issue in this sample is not a single outlier observation, but that post-pre Scope 1 changes display broad tails. The pre- and post-shock averages are taken over the years actually observed

This transformation is the panel implementation of winsorizing the firm-level post-pre change. It leaves all pre-shock observations unchanged and preserves within-post-year variation, while keeping each firm's average post-pre Scope 1 change inside the 10th to 90th percentile range of the IV sample. We prefer it to two alternatives. Winsorizing annual emissions levels targets large emitters rather than large identifying changes, and collapsing the panel to one change per firm would discard the year fixed effects, the annual controls, and part of the panel information.

The bounded-change estimate confirms the result and trims the influence of the outliers that inflated the raw coefficient. Exit from the offset market still produces a significant reduction in Scope 1 emissions, but the effect is now smaller: the point estimate of -0.89 implies a reduction of about 60% (column 3 of Table 6; the first stage in column 1 applies to both columns 2 and 3, since the transformation changes only the outcome, not the treatment or the instrument). The gap between this estimate and the raw -1.22 shows that a handful of firms with extreme emissions changes drove much of the difference, and that the conservative figure is the one to read.

The framework layed out at the beginning of this Section predicts that the shift toward internal abatement should be stronger where reputational stakes and feasible abatement are higher. To test this hypothesis, we re-estimate the 2SLS on different sub-samples. We first split firms by whether they experienced an environmental controversy during the sample window, and then by whether they operate in a sector where abatement is relatively easy or relatively hard.

The split by abatement feasibility is informative. Among non-hard-to-abate firms, the first stage stays strong, with an F-statistic of 12.33 on 1,474 firm-year observations, and the second-stage coefficient is -1.060 ($t = -2.92$), which implies a reduction of about 65% (column 5). This fits the idea that firms with feasible short-run operational margins respond most when the credibility of offsets deteriorates. Among hard-to-abate firms, by contrast, the first-stage F-statistic falls to 3.45 (column 4). The effect for the latter group is positive but imprecisely estimated. We read this estimate with

within each window, so $|T_0|$ and $|T_1|$ equal the number of non-missing years for firm i

caution for two reasons. The subsample contains only 183 firm-year observations, so the test has little power, and the instrument is weak, so the second-stage coefficient is uninformative. However, the lack of relevance of the instrument for firms operating in hard-to-abate sectors is not inconsistent with our prior: those firms are less likely to be able to substitute toward internal abatement and, therefore, a higher exposure to the integrity controversy is unlikely to be the reason why these firms quit the market.

The split by environmental controversy raises the same concern. Firms without a controversy display a strong first stage, with an F-statistic of 13.38 on 1,172 observations, and a negative second-stage coefficient of -0.853 ($t = -3.32$), or a reduction of about 57% (column 7). Firms with a controversy, instead, number only 485 and have a very weak first stage, with an F-statistic of 1.63, so the IV is uninformative for this group (column 6). The coefficient estimate suggests that controversy-experiencing firms, who are more reputationally exposed, respond more strongly. However, the estimate is imprecise, the subsample is small and pre-shock portfolio exposure predicts their exit poorly, so we refrain from assigning an interpretation to this coefficient.

Taken together, the IV evidence is sharpest where the framework predicts a response and where we have both enough firms and a strong instrument: among non-hard-to-abate firms and firms without prior controversies. For the smaller hard-to-abate and controversy subsamples, the first stage is too weak to support a causal inference.

6 Conclusions

Our autopsy of the voluntary carbon market combines transaction-level evidence on supply, demand, and prices with firm-level emissions accounts. We study how integrity controversies reshaped market outcomes and corporate decarbonization. Drawing on the near-universe of transactions, we document a market that displayed structural fragilities well before the 2023 scandal: persistent oversupply, fragmented intermediation, and limited transparency in pricing and buyer behavior. We then show that the January 2023 investigation acted as a sharp and plausibly exogenous negative demand shock, lowering prices and halting the growth of retirements. Using this shock as a quasi-natural experiment, we find evidence consistent with a moral-licensing mechanism: firms that *exit* offsetting after the scandal reduce Scope 1 emissions substantially more than firms that *stay*.

Recent scrutiny has focused on the *supply side* of the market: methodologies, baselines, additionality, permanence, and certification. Our results point to a different issue specific to the *demand side*. Even if every credit were of high quality, the option to substitute offsets for abatement can slow real reductions in emissions intensity at the firm level, because it relaxes the constraint that drives operational change. Despite speculation about the likely existence of the documented moral hazard, the policy and academic debate has so far paid little attention to this channel, focusing far more on the supply-side of the market, due to lack of empirical evidence on the substitution mechanism. In this paper, we provide such evidence, which contrasts with firms' narratives that often frame offsetting as a last-resort instrument for residual emissions. Offsetting may be the only option in hard-to-abate sectors, but cheap and easy access to offsets may also weaken firms' incentives to develop new abatement strategies. It would have been interesting to test whether firms purchasing Carbon Dioxide Removal (CDR) credits behaved differently, since these are considered to be of higher quality. However, the number of firms in our sample purchasing those credits is quite limited and mainly concentrated among large firms in the tech industry. Therefore, we were unable to conduct such a test because of statistical power limitations, but we recognize this as a future research opportunity.

Our findings carry important implications for climate policy for several reasons. First, the current structure of the market means that many credits are imperfect substitutes for carbon emissions along the dimensions that matter, namely time and risk. Most traded credits come from *avoidance* activities rather than durable removals, and even when these projects deliver real mitigation, the mapping from "one credit" to "one tonne of permanent compensation" remains contested, both conceptually and empirically. Using offsets in place of feasible abatement therefore risks delaying the technological and organizational changes needed to cut gross emissions intensity for good. Second, a forward-looking view points to scarcity and prioritization. Today's oversupply reflects what we call *nominal* credit abundance. Many units exist on registries and trade at low prices, but this abundance does not imply an equally large supply of high-integrity tonnes that can serve as residual-emissions instruments as the climate transition tightens. If hard-to-abate sectors must decarbonize fast, the binding constraint may become a shortage of credible, durable mitigation units. Using offsets to replace abatement where it is viable then misallocates a resource that will grow scarce.

The trade-off is sharpest for genuinely hard-to-abate activities. Aviation is a good example. Scalable substitutes such as sustainable aviation fuel remain limited relative to demand, so near-term decarbonization will still leave residual emissions, for which firms turn to offsets or other mechanisms. The right response here is not to abolish offsets, but to govern their use: to prioritize them for residual emissions and to hold firms accountable for the claims they make. The challenge is to stop offsetting from becoming a cheap substitute for feasible abatement, while preserving a role for credible instruments in sectors where technological options cannot yet scale.

Our results point to three practical implications. First, regulators and investors should judge corporate climate claims not only on the integrity of the credits a firm buys, but also on whether the firm cuts its own *gross* emissions. Second, governance reforms should compress the wedge between the private cost of offsetting and the social value of decarbonization. One promising route is a more centralized or tightly supervised transaction architecture that prices credits closer to benchmark abatement costs, limits intermediary mark-ups, and channels any surplus toward public purposes. Revenue recycling works in practice. Switzerland's CO_2 levy returns receipts to households through lower health-insurance premiums, one concrete channel for lump-sum recycling. In the EU, ETS auction revenues fund mechanisms such as the Innovation Fund and sit within a broader framework that links carbon-pricing proceeds to climate and energy goals. Third, policy should treat credible offsets as a scarce resource and reserve them for residual emissions in sectors that cannot yet abate, instead of letting them substitute for abatement that is already feasible.

We thus conclude by calling for more research on how to design institutions that preserve a legitimate role for offsets while preventing substitution away from feasible abatement. The current research efforts have largely prioritized reforms of the supply-side of the Voluntary Carbon Market, which remains a priority, but have overlooked the demand-side: the Voluntary Carbon Market 2.0 should reserve offsetting for residual emissions and align private incentives with real, durable, and measurable abatement. In this sense, our autopsy is not only diagnostic. It is also a starting point for a governance architecture in which voluntary climate finance supports genuine decarbonization rather than substituting for it.

References

Adão, R., Kolesár, M., and Morales, E. (2019). Shift-share designs: Theory and inference. *The Quarterly Journal of Economics*, 134(4):1949–2010.

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Allen, F., Behr, P., Cosenza, R., and Nowak, E. (2026). Do investors care about the rainforest? Evidence from voluntary carbon offsets around the world. *Review of Finance*, 30(1):321–349.

Badgley, G., Freeman, J., Hamman, J. J., and Haya, B. (2022). Systematic Over-Crediting of Forest Offsets. *Global Change Biology.*

Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute for Employment Research, Kalamazoo, MI.

Berg, F., Ceccarelli, M., Heeb, F., Ivashchenko, A., Rigobon, R., and Zwinkels, R. C. J. (2025). The market for voluntary carbon offsets. *Available on SSRN.*

Black, A. et al. (2024). State of the Voluntary Carbon Market 2024. Ecosystem Marketplace Report.

Boiral, O., Heras-Saizarbitoria, I., and Brotherton, M. (2024). Achieving Corporate Carbon Neutrality: A Critical Analysis of Disclosures, Offsetting Practices, and Claims. *Journal of Cleaner Production*, 434:139846.

Broekhoff, D., Piggot, G., and Erickson, P. (2019). Securing Climate Benefit: A Guide to Using Carbon Offsets. Stockholm Environment Institute Report.

Burger, A. M., Schuler, J., and Eberling, E. (2022). Guilty Pleasures: Moral Licensing in Climate-Related Behavior. *Global Environmental Change*, 72:102415.

Cabiyo, B. et al. (2025). Carbon Offset Markets Must Learn to Mitigate the Risk of Overcrediting. *PNAS Nexus*. Discusses overcrediting as a structural risk in offset systems.

Calel, R., Colmer, J., Dechezleprêtre, A., and Glachant, M. (2025). Do Carbon Offsets Offset Carbon? *American Economic Journal: Applied Economics*, 17(1):1–40.

Carton, W., Asiyanbi, A., Beck, S., Buck, H. J., and Lund, J. F. (2020). Negative Emissions and the Long History of Carbon Removal. *WIREs Climate Change*, 11(6):e671.

Cattaneo, M. D., Idrobo, N., and Titiunik, R. (2020). *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge Elements: Quantitative and Computational Methods for Social Science. Cambridge University Press.

Foerster, A. (2023). Corporate Net Zero Pledges: A Triumph of Private Climate Governance? *Australian Journal of International Affairs*, 77(5):587–606.

Gill-Wiehl, A., Kammen, D. M., Haya, B., et al. (2024). Carbon Offsets for Clean Cookstoves are Pervasively Over-Credited. *Nature Sustainability*.

Gillingham, K., Rapson, D. S., and Wagner, G. (2016). The Rebound Effect and Energy Efficiency Policy. *Review of Environmental Economics and Policy*, 10(1):68–88.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review*, 110(8):2586–2624.

Greening, L. A., Greene, D. L., and Difiglio, C. (2000). Energy Efficiency and Consumption — The Rebound Effect — A Survey. *Energy Policy*, 28(6–7):389–401.

Guizar-Coutiño, A., Jones, J. P. G., Balmford, A., Carmenta, R., and Coomes, D. A. (2022). A Global Evaluation of the Effectiveness of Voluntary REDD+ Projects at Reducing Deforestation and Degradation in the Moist Tropics. *Conservation Biology*, 36(6):e13970.

Günther, S. A., Staake, T., Schöb, S., and Tiefenbeck, V. (2020). The Behavioral Response to a Corporate Carbon Offset Program: A Field Experiment on Adverse Effects and Mitigation Strategies. *Global Environmental Change*, 64:102123.

Hanna, R., Duflo, E., and Greenstone, M. (2016). Up in Smoke: The Influence of Household Behavior on the Long-Run Impact of Improved Cooking Stoves. *American Economic Journal: Economic Policy*, 8(1):80–114.

Haya, B., Cullenward, D., Strong, A., et al. (2020). Managing Uncertainty in Carbon Offsets: Improved Forest Management and the US Forest Offset Market. *Climate Policy*, 20(9):1112–1131.

Imbens, G. W. and Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475.

Imbens, G. W. and Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142(2):615–635.

Kim, S., Li, T., and Wu, Y. (2026). Carbon offsets: Decarbonization or transition-washing? *Journal of Finance*. Forthcoming.

Kleibergen, F. and Paap, R. (2006). Generalized Reduced Rank Tests Using the Singular Value Decomposition. *Journal of Econometrics*, 133(1):97–126.

Kollmuss, A., Lazarus, M., Lee, C., and Polycarp, C. (2008). A Review of the Voluntary Carbon Market. *Stockholm Environment Institute Working Paper*.

NewClimate Institute and Carbon Market Watch (2024). Corporate Climate Responsibility Monitor 2024.

Newell, R. G., Pizer, W. A., and Raimi, D. (2013). Carbon Markets 15 Years After Kyoto: Lessons Learned, New Challenges. *Journal of Economic Perspectives*, 27(1):123–146.

OECD (2024). The Interplay Between Voluntary and Compliance Carbon Markets. Technical report, Organisation for Economic Co-operation and Development.

Science Based Targets initiative (2021). SBTi Corporate Net Zero Standard.

Shapley, L. S. (1953). A Value for n-Person Games. *Contributions to the Theory of Games*, 2:307–317.

Song, Y., Aldy, J. E., Holbrook, N. M., Gao, X., and Thompson, J. (2025). Selection and Over-Crediting in Forest-Based Carbon Offset Projects. Working paper.

Sorrell, S., Dimitropoulos, J., and Sommerville, M. (2009). Empirical Estimates of the Direct Rebound Effect: A Review. *Energy Policy*, 37(4):1356–1371.

Stock, J. H. and Yogo, M. (2005). Testing for weak instruments in linear IV regression. In Andrews, D. W. K. and Stock, J. H., editors, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pages 80–108. Cambridge University Press, Cambridge.

Stolz, N. and Benedict, P. (2025). The Negligible Role of Carbon Offsetting in Corporate Climate Strategies. *Nature Communications*, 16(7963).

Swinfield, T. and Scott, E. (2025). Scientific Credibility for High-Integrity Voluntary Carbon Markets. Cambridge Centre for Carbon Credits, University of Cambridge. Cambridge Open Engage preprint.

Tiefenbeck, V., Staake, T., Roth, K., and Sachs, O. (2013). For Better or for Worse? Empirical Evidence of Moral Licensing in a Behavioral Energy Conservation Campaign. *Energy Policy*, 57:160–171.

Tilsted, J. P., Palm, E., Bjørn, A., and Lund, J. F. (2023). Corporate Climate Futures in the Making: Why We Need Research on the Politics of Science-Based Targets. *Energy Research Social Science*, 103:103229.

Tzachor, A. et al. (2023). Voluntary Carbon Markets and Net-Zero Corporate Strategies. *Climate Policy*.

West, T. A. P., Börner, J., Sills, E. O., and Kontoleon, A. (2020). Overstated Carbon Emission Reductions from Voluntary REDD+ Projects in the Brazilian Amazon. *Proceedings of the National Academy of Sciences*, 117(39):24188–24194.

West, T. A. P., Wunder, S., Sills, E. O., Börner, J., Rifai, S. W., Neidermeier, A. N., Frey, G. P., and Kontoleon, A. (2023). Action Needed to Make Carbon Offsets from Forest Conservation Work for Climate Change Mitigation. *Science*, 381(6660):873–877.

Appendix

A Robustness tests

Table A1: Robustness of the Exit Classification (Difference-in-Differences)

Each entry reports the Post×Exit coefficient (with the firm-clustered t-statistic in parentheses) from the baseline difference-in-differences (Equation 2, log emissions on Post×Exit with firm and year fixed effects and the full control set). Panel A varies the classification source: "registry only" uses only matched registry retirements; "CDP only" uses only firms with a CDP offset-use disclosure. Panel B varies the post-shock window used to define exit.

	Scope 1	Scope 2	Scope 3	Obs.
Panel A: Classification source				
Baseline (expanded, CDP priority)	-0.210**	-0.178	0.147	2550
	(-2.71)	(-1.60)	(0.70)	
Registry retirements only	-0.218*	-0.083	-0.137	1853
	(-2.43)	(-0.60)	(-0.54)	
CDP disclosure only	-0.182	-0.413*	0.401	1171
	(-1.41)	(-2.40)	(1.31)	
Panel B: Post-shock exit window				
No retirements from Jan. 2023 (no grace)	-0.216**	-0.219	0.146	2550
	(-2.73)	(-1.96)	(0.69)	
No retirements from Apr. 2023 (3-mo. grace)	-0.203**	-0.184	0.131	2550
	(-2.70)	(-1.71)	(0.65)	
No retirements from Jul. 2023 (6-mo. grace)	-0.153*	-0.187	0.216	2550
	(-2.08)	(-1.85)	(1.09)	

Significance: *** $p < 0.001$; ** $p < 0.01$; * $p \leq 0.05$.

Because our treatment combines two sources (matched registry retirements and CDP offset-use disclosures) in order to expand the sample, we show here that the classification does not drive the result. Table A1 re-estimates the baseline Scope 1 difference-in-differences under alternative exit definitions. Restricting the classification to registry records only leaves the estimate essentially unchanged (-0.218, significant at 5%); this matters because the registry-only definition is entirely independent of CDP, so the effect is not an artifact of the CDP priority rule. Using CDP disclosures only yields the same point estimate (-0.182) but loses significance, as expected from the smaller and noisier disclosure-based sample. Varying the post-shock window from no grace period (exit requires zero retirements from January 2023) to a six-month grace, also leaves the estimate stable, attenuating only mildly under the longest window, which might indeed confound slow exiters with stayers. Across all definitions the differential Scope 1 reduction remains economically large and, except in the smallest sample, statistically significant.

B Construction of the Instrument Z_i

We estimate the price shock for each cell as follows. For transaction τ in month $m(\tau)$, let P_τ denote the estimated retirement price and v_τ the number of credits retired. We restrict the price-shock calculation to transactions from January 2022 through June 2023.[12] We first compute the volume-weighted market log price in each month,

$$\bar{\ell}_m^{\text{mkt}} = \frac{\sum_{\tau:m(\tau)=m} v_\tau \ln P_\tau}{\sum_{\tau:m(\tau)=m} v_\tau},$$

and define the market-month-demeaned transaction price as $\widetilde{\ell}_\tau = \ln P_\tau - \bar{\ell}_{m(\tau)}^{\text{mkt}}$. This demeaning step removes common monthly movements in the aggregate VCM price level and isolates differential repricing across project segments.

For each cell k, we compute the volume-weighted mean of $\widetilde{\ell}_\tau$ in the pre-scandal window, January–December 2022, and in the post-scandal window, January–June 2023:

$$\bar{\widetilde{\ell}}_k^{\text{pre}} = \frac{\sum_{\tau\in k,\, \tau\in\text{pre}} v_\tau \widetilde{\ell}_\tau}{\sum_{\tau\in k,\, \tau\in\text{pre}} v_\tau}, \qquad \bar{\widetilde{\ell}}_k^{\text{post}} = \frac{\sum_{\tau\in k,\, \tau\in\text{post}} v_\tau \widetilde{\ell}_\tau}{\sum_{\tau\in k,\, \tau\in\text{post}} v_\tau}.$$

The raw cell-level price chnage is $\widehat{\delta}_k = \bar{\widetilde{\ell}}_k^{\text{pre}} - \bar{\widetilde{\ell}}_k^{\text{post}}$. A positive value means that credits in cell k became cheaper relative to the market after the scandal. To reduce the influence of noisy shocks in thinly traded cells, we shrink the raw shock toward the

[12]To limit noise, we only use data with vendor defined price-confidence scores classified as Medium, High, or Very High

volume-weighted global mean $\bar{\delta}$:

$$\widehat{\delta}_k^* = w_k \widehat{\delta}_k + (1 - w_k)\bar{\delta}, \qquad w_k = \frac{V_k}{V_k + \kappa},$$

where V_k is total transaction volume in cell k across the two windows and $\kappa = 100{,}000$ credits. Finally, because the encouragement to quit should come from adverse revaluation rather than from relative price increases, we define the cell-level shock intensity as $g_k = \max\{\widehat{\delta}_k^*, 0\}$. Thus, cells whose relative prices did not fall after the scandal contribute zero to the instrument.

The shrinkage target $\bar{\delta}$ is the volume-weighted mean of the raw cell shocks over the cells for which a raw shock is estimable,

$$\bar{\delta} \;=\; \frac{\sum_{k \in \mathcal{K}} V_k \, \hat{\delta}_k}{\sum_{k \in \mathcal{K}} V_k},$$

where $\hat{\delta}_k$ is the raw (market-adjusted, pre-minus-post) log-price change in cell k; $V_k = V_k^{\text{pre}} + V_k^{\text{post}}$ is the total volume of credits transacted in cell k across the pre- and post-shock pricing windows, restricted to the priced, medium-or-higher-confidence retirements used to construct the shocks; and $\mathcal{K}$ is the set of cells whose raw shock is well-defined, namely those with at least three priced transactions in the pre-window and at least two in the post-window. Cells outside $\mathcal{K}$ are assigned the target, $\hat{\delta}_k^* = \bar{\delta}$, while cells in $\mathcal{K}$ are shrunk using the above formula. In our sample $\mathcal{K}$ contains 18 cells and $\bar{\delta} = 0.31$.

The firm-level exposure measure is based on each firm's project-level retirement portfolio in 2022. We use 2022 because it is the last full pre-shock year and therefore best captures the portfolio exposed to the January 2023 shock. Let $R_{ik,2022}$ be the number of credits retired by firm i in cell k during 2022, using all matched positive-volume retirement transactions regardless of whether the transaction has an observed price. Define

$$s_{ik} = \frac{R_{ik,2022}}{\sum_{\ell} R_{i\ell,2022}},$$

Table B2: Instrument Validity: Covariate Balance by Exposure Intensity

Each row compares the pre-shock (2017–2022) average of a covariate across two groups split at the median of the firm-level Bartik exposure Z. The difference is High−Low, with a Welch t-statistic in parentheses. Financials and emissions are in logs. "Exiter", "Hard-to-abate", and "Any controversy" are firm-level shares.

Covariate	Low Z	High Z	Difference
ln(Revenue)	23.052	22.984	−0.067 (-0.35)
ln(Assets)	23.996	23.806	−0.190 (-0.82)
ROA	0.105	0.106	0.001 (0.10)
Leverage	0.287	0.277	−0.010 (-0.52)
ln(Scope 1)	11.138	10.986	−0.152 (-0.37)
Exiter (share)	0.129	0.206	0.076* (1.89)
Hard-to-abate (share)	0.118	0.118	0.000 (0.00)
Any controversy (share)	0.265	0.235	−0.029 (-0.62)

Significance: * 10%, ** 5%, *** 1%.

and construct the raw exposure index as

$$Z_i = \sum_k s_{ik} g_k.$$

The panel instrument is then $Z_{it}^{IV} = Post_t \times Z_i$, where $Post_t = \mathbf{1}[t \geq 2023]$.

Finally, we assess whether pre-shock portfolio exposure proxies for firm type rather than isolating the integrity shock. Table B2 splits the IV sample at the median of the Bartik exposure Z and compares the two groups' pre-shock characteristics. High- and low-exposure firms are statistically indistinguishable on size (log revenue and log assets), profitability, leverage, the level of Scope 1 emissions, the share operating in hard-to-abate sectors, and the share with an environmental controversy: none of the differences is significant, and most are close to zero. The one margin on which the groups differ is the exit rate itself, which is exactly showing the instrument's relevance: firms more exposed to the discredited segments are more likely to leave the market. Exposure intensity is therefore unrelated to observable firm type and operates through exit, as the design requires.