Essays in Carbon Markets

by

Riccardo Cosenza

Faculty of Economics

Università della Svizzera italiana



A thesis submitted for the degree of

Doctor of Philosophy in Finance

September 2026

Board of Examiners:

Prof. Eric Nowak, *Università della Svizzera italiana*, Supervisor

Prof. Francois Degeorge, *Università della Svizzera italiana*, Jury Chair

Prof. Franklin Allen, *Imperial College London*, External Member

Acknowledgements

Looking back, this has been a long and winding road. It is not a journey I ever imagined undertaking, and there were moments, somewhere in the middle of it, when I honestly wondered why I had started at all. I am happy and proud to have reached the end of it, and to have completed my project the way I wanted to. What I am most certain of, however, is that I did not walk this path alone.

My deepest gratitude goes to my supervisor, Eric Nowak. His guidance, patience, and sharp critiques shaped every stage of this research. Thank you for the time you dedicated to me, and for the many discussions that improved this work and, with it, my skills as a researcher. I am also sincerely grateful to Franklin Allen, a member of my thesis committee. His constructive feedback opened up fresh perspectives and helped me strengthen the foundations of this dissertation.

Beyond the committee, thank you to those who stood beside me, and especially to Papa Orgen and Patrick Behr for the great collaboration.

Thank you to my parents, who gave me the principles and the strength to reach this milestone, and to my family for their patience across years of distance and distraction.

Finally, thank you to the colleagues and friends who shared the many stages of this journey with me. It has been a genuinely rewarding one.

Contents

Introduction 4

Chapter 1

Are all COPs busted? The role of climate policy events for carbon pricing in compliance and voluntary markets 7

1.1 Introduction . . . 7

1.2 Institutional background . . . 10

1.2.1 Key institutional facts about COPs . . . 10

1.3 Methodology . . . 12

1.3.1 The constant mean return model . . . 13

1.3.2 Co-integration tests . . . 14

1.4 Results . . . 16

1.4.1 Event study results for the Paris Conference 2015 and the Paris Agreement 2016 16

1.4.2 Event study results for the Glasgow conference (COP 26) . . . 19

1.4.3 Event study results for the UAE Conference (COP 28) . . . 20

1.4.4 Co-integration results for the Paris Conference 2015 and the Glasgow Conference 2021 . . . 21

1.4.5 Robustness tests . . . 24

1.5 Discussion and implications for global carbon policy . . . 26

1.6 Conclusion . . . 27

Appendix A . . . 28

Chapter 2

Do investors care about the rainforest? Evidence from voluntary carbon offsets around the world 42

2.1 Introduction . . . 42

2.2 Institutional background . . . 46

2.2.1 . . . 46

2.2.2 Biodiversity Impact from Forest Carbon Credits . . . 47

2.3 Data and Research Design . . . 50

2.3.1 Data and Events . . . 50

2.3.2 The Guardian Article Shock in January 2023 and the Impact on the VCM . . . 55

2.3.3 Hypothesis Development . . . 58

2.3.4 Empirical Methodology . . . 58

2.4 Results . . . 58
2.4.1 Overall Analysis of Abnormal Returns . . . 59
2.4.2 Abnormal Returns Before and After the Guardian Article Shock . . . 61
2.4.3 Cross-Sectional Results . . . 66
2.4.4 Economic Relevance . . . 67
2.4.5 Robustness Tests . . . 68
2.5 Conclusion . . . 70
Appendix B . . . 72

Chapter 3
Synthetic Control Illusions in Climate Science: Methodological errors in the assessment of forest carbon projects 100
3.1 Introduction . . . 100
3.2 Misuse of Global Forest Change Data . . . 101
3.3 Invalid Implementation of Synthetic Control Method . . . 103
3.3.1 Synthetic Controls Poorly Constrained . . . 104
3.3.2 Methodological concerns with the synthetic control approach . . . 106
3.3.3 Temporal Dynamics and Time-Varying Treatment Effects . . . 106
3.3.4 Spatial Spillovers, Leakage, and the Stable Unit Treatment Value Assumption . . . 107
3.3.5 Unreliable Validation Tests Used . . . 108
3.3.6 Problems with the generalized synthetic control analysis . . . 109
3.4 Sensitivity of Results . . . 111
3.5 Calculation Errors . . . 111
3.6 Sample size . . . 112
3.7 Conclusion . . . 112
Appendix C . . . 114

Introduction

Carbon pricing exists in two forms whose economic properties differ far more than the shared label suggests. Compliance markets, of which the EU Emissions Trading System is the largest, rest on a legislated cap: participants must surrender allowances against verified emissions, and price emerges from politically determined scarcity meeting regulated demand. Voluntary carbon markets have no cap and no legal obligation, trade largely over-the-counter, and carry heterogeneous credit quality with no central authority standing behind the unit. This thesis argues that the deepest difference between them is not liquidity, maturity, or legal obligation, although all three matter. It is the epistemic status of the commodity. An EU allowance is what the regulation says it is: its existence and scarcity are administrative facts, not estimates. A voluntary carbon credit is a certified claim that emissions which would otherwise have occurred did not occur, so its defining property, additionality, is a statement about a world that did not happen. It cannot be observed, only estimated.

The consequences run through the whole thesis. The voluntary carbon market is a market in counterfactuals; the value of the traded unit depends on the credibility of an estimation procedure; certification standards and remote sensing methodologies are not peripheral infrastructure but constitutive of the commodity; and disputes about estimation methods are financially material events rather than academic curiosities. The three essays examine this structure at three levels: the international policy process that sets the rules under which carbon units acquire meaning and eligibility, the corporate offsetting decisions that investors observe and price, and the measurement machinery that determines whether a project delivered what it claimed.

The value of a voluntary carbon credit depends on two things outside the transaction itself, which I call institutional credibility and quality credibility. Institutional credibility concerns whether the framework that defines and recognises the unit is stable and enforceable, and international climate negotiations are the venue in which that question is answered or left open. Quality credibility concerns whether an individual unit represents what it claims. Here the asymmetry between segments is sharpest: compliance markets have no additionality problem, since an allowance is defined into existence by statute, while every voluntary credit carries estimation risk. Because quality is unobservable, the market substitutes labels, and confidence in labels rests on the belief that certifying institutions apply defensible methods. The measurement layer is therefore not downstream of the market but an input to it. If the methods are unstable, prices will respond to methodological artifacts as though they were information about fundamentals.

Chapter 1 asks whether carbon markets price the outcomes of international climate negotiations, and whether the two segments respond differently. Studying four events around the Paris, Glasgow, and

Dubai conferences, using event studies over windows spanning each full negotiation and cointegration tests on the relationship between compliance and voluntary proxies, it finds that participants price COP outcomes significantly and in both directions, positively when outcomes are judged credible and negatively when they lack ambition or implementation certainty. Voluntary proxies build expectations more aggressively during negotiations and correct more sharply afterwards. The cointegrating relationship present before Paris vanishes afterwards, while the relationships around Glasgow remain stationary. Read through the framework, the pattern is not that Glasgow was simply a better conference: Glasgow resolved the specific ambiguity governing recognition of voluntary units through the Article 6 Rulebook, while Paris delivered systemic ambition without operational clarity.

Chapter 2 moves from the market to the firm, asking whether investors distinguish between offsets on the basis of environmental quality. The event is the retirement of a credit, which separates genuine environmental action from mere purchase and is publicly recorded almost immediately, yielding a dataset of 781 listed firms and over three thousand forestry retirements totalling 116 million tons of CO_2. Identification rests on certification under the Climate, Community and Biodiversity Standard as the strongest available marker of biodiversity impact, and on a January 2023 newspaper investigation claiming that most rainforest offsets were worthless, to which every tier of market infrastructure responded within hours. Abnormal returns to retirements are indistinguishable from zero until the end of 2022 and significantly negative afterwards, with the effect concentrated in uncertified credits while certified retirements remain muted. The chapter leaves a puzzle it states rather than resolves: firms undertake costly activity whose measured effect on shareholder value is at best neutral.

Chapter 3 turns to the evidentiary basis of that shock. The investigation drew its central claim from a study applying synthetic control estimators to 24 forest carbon projects. The chapter identifies three compounding problems: pixel counts from a global forest change product are used as direct area estimates against the producers' explicit warning, which systematically understates avoided deforestation; donor pools drawn from randomly generated circular polygons violate the identifying assumptions of the method, omit critical predictors of deforestation risk, and produce estimates that reverse sign when a single filtering step is removed; and two calculation errors, once corrected, substantially raise the reported proportion of credits delivering real carbon benefits. The chapter does not show that these projects work. It shows that the analysis cannot determine whether they do, and that every identified problem biases the estimate against finding effectiveness.

The chapters operate at successively finer levels, and taking all three shows that credibility is produced at the bottom and consumed at the top. Policy events move prices because they change expectations about how project-level claims will be recognised, investors discriminate between credits because they

are reasoning about project-level integrity, and project-level integrity is an estimation problem. The link between Chapters 2 and 3 is literal rather than thematic: the shock identifying Chapter 2 rests on the study Chapter 3 re-examines. The sequence runs from a counterfactual estimate produced with a method whose assumptions are not satisfied, through publication and media amplification, to measurable repricing of corporate offsetting in global equity markets, and together the links show that the price of environmental credibility depends on the reliability of applied econometrics.

Chapters 1 and 2 then report the same finding in different places. Resolving institutional ambiguity about the recognition of voluntary units preserves the pricing relationship between segments; a demanding certification label preserves valuation under an integrity shock. In both cases a credible institutional resolution of quality ambiguity stabilises prices and its absence transmits the shock, which converts two empirical findings into a single proposition about how credibility works in this market. All three chapters estimate counterfactuals, and the discipline applied to the thesis's own estimates is the same discipline Chapter 3 argues was absent in the work it examines.

The thesis makes four contributions. To carbon finance, it provides the first systematic evidence on how climate policy events affect price formation across compliance and voluntary markets, and shows the relationship between segments to be state-dependent on whether policy resolves the ambiguity specific to each. To sustainable and biodiversity finance, it provides the first evidence on investor reactions to voluntary credit retirements and the first isolation of the value attached to biodiversity co-benefits, together with a linked registry-to-equity dataset. To causal inference in environmental applications, it demonstrates how donor pool misspecification and measurement error in remotely sensed outcomes generate unstable counterfactuals. To the understanding of carbon markets as institutions, it identifies the credence-good structure of voluntary credits as the source of the market's characteristic instabilities.

The joint policy argument is stronger than the sum of the separate ones. Prices cannot be stable while the quality of the underlying unit remains contested, quality cannot be settled while the methods for estimating it remain fragile, and no amount of international ambition compensates for either. The principal limitation is common to the design: the event-based approach identifies short-horizon repricing and cannot establish whether revisions persist or translate into changes in real investment, and Chapter 3 identifies conditions under which existing estimates cannot be sustained without producing an alternative estimate of project effectiveness, which remains the most consequential open question.

The remainder proceeds as follows. Chapter 1 examines the pricing of international climate policy events across compliance and voluntary markets. Chapter 2 examines investor reactions to voluntary offset retirements and the role of biodiversity certification. Chapter 3 examines the methodological foundations of synthetic control estimates of forest carbon project effectiveness.

Are all COPs busted? The role of climate policy events for carbon pricing in compliance and voluntary markets

1.1 Introduction

Climate change is one of the most defining problems of our time. Each year, representatives of governments, stakeholders, and other interested parties gather over two weeks to discuss advances, failures, and new measures to stop or slow down the pace of climate change and global warming. These events, referred to as climate conferences (or Conference of the Parties, COP), occur in changing locations around the globe. This study centres on four of these COP- related events - the Paris Climate Conference, the Paris Agreement, the Glasgow Conference, and the Dubai (UAE) Conference [1]. Around the world, carbon policy is often administered nationally and lacks external cohesion (Baranzini et al., 2017; Creti and Ftiti, 2024). COPs act as global policy-signalling events that influence expectations about regulatory ambition and implementation timelines, serving as a melting pot of political and scientific engagement on climate change issues, including carbon pricing. Recent empirical evidence suggests that carbon pricing is an effective tool for emissions reduction, as shown in Sweden (Martinsson et al., 2024), even as market participants hedge climate risk from policy change rather than from physical climate impacts themselves (Faccini and Skiadopoulos, 2023). Antoniuk and Leirvik (2024) show that unexpected climate-related political events, such as the Paris Agreement, generate abnormal returns in climate-sensitive sectors, with clean energy stocks responding positively and fossil-fuel-intensive industries negatively, depending on the policy direction signalled. Furthermore, Jiang and Luo (2018) find that Chinese equity markets reacted positively to environmental policy announcements between 2009 and 2011, particularly in high pollution industries, likely due to expectations of delayed enforcement.

There is a growing body of carbon asset pricing studies [2], e.g., Bolton and Kacperczyk (2021) and Huij et al. (2022), which focus on cross-sectional pricing of climate risk among firms, dynamic portfolio strategies that incorporate climate news signals (Engle et al., 2020), and firm-level responses to environmental regulation (Hsu et al., 2020). Müller et al. (2023) empirically show that portfolios that integrate compliance carbon allowances and voluntary carbon offsets outperform those that do not, on a risk-adjusted basis, though this result has temporal dimensions. Choi et al. (2020) show that carbon-intensive firms underperform during periods of abnormally warm weather, reinforcing the sensitivity of asset prices to climate signals. Barnett (2020) and Bolton et al. (2022) further emphasize that pricing in financial markets reflects expected climate policy interventions and broader transition risks. In the voluntary carbon market (VCM) context, Veld-Merkoulova and Viteva (2015) study the economic and

[1] We discuss the significance of these events and their selection in the Institutional Background section.

[2] We only review the empirical literature. Giglio et al. (2021) provides an overview of theoretical work.

financial effects of carbon trading and regulation on the stock market prices of five individual companies and the joint effects of regulations, the prices of oil and gas, and the volatility of carbon offsets. Kumar and Firoz (2019) find that the issuance of certified emissions reductions (CERs) did not yield any significantly abnormal returns on the share prices of Indian firms in 193 CER announcements in the period 2005–2017. These studies deepen our understanding of firm-level climate risk but leave open the question of how aggregate markets price high-level, policy-driven climate signals.

This paper makes the first systematic empirical attempt at understanding how major climate policy events, proxied by COPs, affect price formation and the co-movement between compliance and voluntary carbon markets. This dual analysis, distinguishing between regulated and voluntary pricing reactions, offers a novel lens on how international climate negotiations shape tangible market outcomes, and fills a key gap in the carbon finance literature. This article addresses that gap by analysing short-term price reactions and long-term structural relationships between compliance and voluntary carbon markets around major COP events.

In a first step, using standard event study methodologies (Campbell et al., 1998); (MacKinlay, 1997), we estimate abnormal carbon price reactions before and after each event. The event studies we conduct allow us to examine how pre-event expectations shift post-event, measured in cumulative abnormal returns (CARs). We make the key assumption that the contemporaneous pricing of new information during the event window offers the clearest insight into market participants' assessment of policy outcomes. To account for market anticipation, we simulate pseudo-event studies using random pre-event dates and test the significance of the carbon policy premium associated with each COP. In a second step, we also assess whether these events altered the long-run pricing relationship between compliance and voluntary carbon markets using co-integration techniques in the spirit of Engle and Granger (1987).

Our results provide several insights into how carbon markets react to climate policy events. These COP events influenced carbon pricing significantly. Market participants priced in the outcomes of all four events. Thus, COPs play a crucial role in signalling policy directions to global carbon markets. Market reactions are positive when pre-event expectations are met and negative when COP outcomes fall short of them. For instance, while the Paris Agreement was positively priced ahead of its signing, carbon prices dropped post-event due to concerns about whether it was ambitious enough (Lesnikowski et al., 2017). The key message, then, is that while COPs can act as carbon market catalysts, at the same time carbon markets are scrutinizing COPs for their effectiveness.

To understand the market-level implications of these reactions, we next examine how COPs shape the long-run relationship between compliance and voluntary markets. COPs are ideal candidates for investigating climate policy events because they have the potential to impact the entire carbon value chain. We

analyse whether these events changed long-run relationships between compliance and voluntary carbon markets post-event. For this, we run co-integration tests according to Engle and Granger (1987) to assess the structural relationships between voluntary and compliance carbon markets, respectively, before and after each policy event. Our results show (i) a relative price inversion between voluntary and compliance carbon market price returns (VCM proxies increase more rapidly than compliance carbon market proxies) two months before the events; and that (ii) co- integration relationships between compliance and voluntary carbon markets before the Paris Conference vanish post-event. We interpret this second finding as a structural break or a decoupling of voluntary and compliance carbon market prices in the aftermath of the Paris Conference. Underlying efficient carbon price trajectories is the need for strong pricing signals and a commitment to maintain them long enough to enforce the required changes. Price decoupling in the case of the Paris Conference suggests that this policy event catered more to the compliance carbon market. The reaction of VCMs to these policy events suggests that policy prescriptions targeting VCMs may yield stronger pricing signals that could boost mitigation efforts. In other words, while VCMs offer non-regulatory participation in carbon markets, COP outcomes can help set rules of engagement in a way that enhances transparency, strengthens accountability, and creates clearer incentives for market participation.

Conceptually, market responses to COPs can be structured along three axes: (i) whether commitments are symbolic or substantive; (ii) whether they target broad systems or specific sectors (e.g., forestry, energy); and (iii) whether they enhance or dilute institutional clarity-especially around mechanisms like Article 6 of the Paris Agreement from 2016. Building on this, we explain the asymmetric responses of compliance and voluntary carbon markets through four structural contrasts. These differences arise from behavioural, institutional, and regulatory factors that shape how each market interprets and integrates new information. First, regulatory anchoring is central to compliance markets, which are governed by binding frameworks such as cap-and-trade systems (e.g., the existing regime in California) and emissions trading schemes (like the EU-ETS). These markets respond predictably, in terms of direction, to policy signals that affect cap levels, allowance allocations, or regulatory scope (Borenstein et al., 2002). VCMs operate on a non-compulsory participation basis and remain fragmented, with heterogeneous credit quality, project-specific pricing, limited secondary trading, and integrity concerns [3]. As a result, their pricing dynamics are partly dependent on how policy outcomes are perceived to affect the credibility, scalability, and future standardization of carbon offset instruments (Marketplace, 2022). Second, the strategic objectives and horizons of market participants differ. Compliance market participants focus on minimizing short- to medium-term costs within established legal mandates. These participants are sen-

[3]The Taskforce for Scaling Voluntary Carbon Markets, for example, is an industry-led consortium of Finance leaders aiming to bring transparency in pricing and quality standards to VCMs, formed after the Paris Conference.

sitive to changes in rules that affect cost and compliance strategies. In contrast, VCM actors engage to demonstrate climate leadership, meet firm net-zero targets, and manage reputational risk. Third, policy signal processing varies between the two systems. Compliance markets have defined agent obligations and timelines, leading to predictable market reactions to regulatory adjustments. Voluntary markets, by contrast, are more exposed to policy ambiguity. Unresolved issues, such as the implementation details of Article 6.4 of the Paris Agreement, can introduce considerable volatility, as these markets lack centralized oversight or backstop mechanisms. Without consistent standards, VCMs are more prone to speculative pricing. These contrasts are reflected in our results. VCMs exhibit pronounced pre-event optimism and post-event corrections when expectations are not met. Differences in market maturity and liquidity amplify these effects. Compliance markets benefit from liquid and standardized instruments and a robust institutional infrastructure, which facilitate efficient price discovery and information absorption. VCMs remain fragmented, with varying verification standards and limited secondary trading. This structural immaturity contributes to greater sensitivity to perceived shifts in policy direction, often resulting in outsized market reactions to symbolic or ambiguous policy outcomes.

1.2 Institutional background

1.2.1 Key institutional facts about COPs

Our analysis focuses on four key COP events: the Paris Conference (COP 21), the Paris Agreement (2016), COP 26 in Glasgow, and COP 28 in the UAE [4]. COP 21 was significant insofar as it represented the first time 196 countries agreed to a common framework aimed at limiting global warming to well below 2°C, and preferably to 1.5°C, relative to pre-industrial levels. The Paris Climate Conference (COP 21) is regar ded as a pivotal moment, which has led to a significant shift in resource mobilization and policy frameworks for addressing climate change. Year-long party negotiations from the Paris Conference culminated in the signing of the Paris Agreement in 2016. A key provision in the Paris Agreement is the country-level obligation to submit five-year ambition plans or nationally determined contributions (NDCs). The Paris conference and Agreement marked a structural shift from the top-down targets of the Kyoto Protocol to a bottom-up architecture based on voluntary ambition and transparency. The Paris Agreement also implemented the Global Stocktake mechanism, established to assess collective progress every five years. A particular sticking point in the Paris Agreement was negotiations surrounding Article 6.4 of the Paris Rulebook. Article 6.4 intended a framework for voluntary cooperation between countries

[4] 4A key consideration for selecting these events is the availability of transparent, exchange-traded pricing data, particularly for voluntary carbon markets. Reliable, liquid price signals are essential for measuring market responses and shaping both our event selection and empirical design. While the history of carbon pricing includes many policy events, we focus on those that are institutionally significant and empirically tractable.

through the Sustainable Development Mechanism (SDM). Subsequent attempts to operationalise Article 6.4 made limited progress until COP26, where the broader Article 6 Rulebook was finalised [5]. The outcomes of COP 21 and the Paris Agreement were systemic in ambition but fell short on institutional clarity, particularly around Article 6, and introduced uncertainty regarding the future interaction between emerging Paris mechanisms and the Clean Development Mechanism (CDM).

COP 26 re-centred the climate agenda on mitigation, adaptation, and the integration of the best available science into national policy planning. A key outcome of COP 26 was the finalisation of the Article 6 Rulebook, which clarified the rules for international carbon trading, including accounting guidelines to prevent double-counting of emissions reductions. This resolution reduced institutional uncertainty for carbon markets and enhanced the credibility of cross-border offset mechanisms. The Glasgow Financial Alliance for Net Zero (GFANZ), a coalition representing over $130 trillion in assets under management, committed to aligning financial flows with net-zero targets. Other headline announcements included a pledge to phase down unabated coal use, the removal of inefficient fossil fuel subsidies, and an agreement among more than 100 countries to end illegal deforestation by 2030. COP 26 therefore combined substantive policy commitments with institutional reinforcement, particularly through the clarification of Article 6, while targeting both economy-wide and sector-specific transition pathways.

COP 28 represents the current state of global climate engagement and marked the most comprehensive assessment of global climate progress since the Paris Agreement. The summit delivered the first Global Stocktake, which concluded that current national pledges fall short of the emission reductions required to meet the 1.5°C target. This formal recognition of the ambition gap adds urgency to policy reforms and strengthens expectations for stricter future regulation. Another outcome was the agreement to operationalise a Loss and Damage Fund to support vulnerable countries facing climate-related losses beyond adaptation. While the fund's financing is significantly lower relative to disclosed global adaptation needs (Markandya et al., 2024), its institutionalisation represents a shift toward climate justice narratives in international finance. COP 28 produced a final decision text that called for a global transition away from fossil fuels. The text also committed countries to tripling renewable energy capacity and doubling energy efficiency by 2030. As with COP 21, COP 28 combined broad system- level ambitions with limited operational detail in several key areas. Nevertheless, the fossil fuel transition reference and institutional advances such as the Loss and Damage Fund provided directional policy signals that may influence long-run expectations, particularly within voluntary carbon markets.

[5] At COP 29 in Baku, a draft set of methodologies and procedures for Article 6.4 was developed, but no consensus was reached on implementation, including the full launch of the international registry and issuance of A6.4ERs (UN-supervised carbon credits). Operational clarity on this subject is now largely achieved as the market transitions from negotiations to practical implementation. The first UN-backed credits under the new mechanism, that is the UN- supervised Paris Agreement Crediting Mechanism, already issued.

1.3 Methodology

The selection of these four climate policy events is motivated by both institutional significance and data availability. Exchange-traded VCM pricing data have only become widely available in the last few years. Organised market access and transparency in price discovery were fundamental requirements in our policy event choices and experiment design. As proxies for compliance markets, we use European Allowances (EUAs) from the EU Emissions Trading System (EU-ETS), which is the largest existing compliance carbon market in the world. We use Certified Emissions Reductions (CER), Global Emissions Offsets (GEO), and Nature-Based Global Emissions Offsets (NGEO) as proxies for voluntary carbon markets. Carbon credits issued from projects certified by the Clean Development Mechanism are referred to as Certified Emission Reductions (CER). CERs were excluded from the EU-ETS from May 2013 onwards. Hence, we consider CER as a proxy for VCM prices in the analyses of the Paris Climate Conference/Agreement. GEO, launched by the CBL (now Xpansiv) in March 2021, provides delivery of physical carbon credits from projects certified by the Verified Carbon Standard (VCS), American Carbon Registry (ACR), and Climate Action Reserve (CAR). GEO-based certificates qualify for the CORSIA program (provided they meet eligibility criteria, including project type, vintage strictions, and Article 6 authorization requirements), while NGEO allows corporations to offset net emissions through nature-based projects registered under VCS and the Climate Community and Biodiversity (CCB) Standard. The emergence of CBL-linked instruments such as GEO and NGEO is particularly important in voluntary carbon markets, where heterogeneous credit quality, fragmented trading, and limited transparency can impede efficient price discovery. These instruments provide practitioners with standardized benchmarks for managing carbon exposures and risks. We use GEO and NGEO as proxies for the voluntary carbon market for the event study on COP 26. EUA and CER price data were sourced from Refinitiv Datastream, while GEO and NGEO prices were obtained from Bloomberg.

Our empirical approach uses standard event study methodology (Campbell et al., 1998); (MacKinlay, 1997) to analyse the price effects of climate policy events on compliance and voluntary carbon markets. Our reference model is the constant mean return model. The event study is premised on the efficient market hypothesis. If carbon prices rapidly incorporate new information, climate policy announcements should be reflected in abnormal returns during the event window. Event windows, defined in Table 1, cover 11 days before and after the two-week Paris, Glasgow, and UAE conferences. We set these windows because day -11 corresponds to the conference's opening plenary and day 0 to the final plenary, so the interval captures every disclosure made during the COPs. These are two-week negotiations where meaningful information (leaks, pledges, draft decisions) is released continuously, not just on day 0. So, the pre-event window is designed to cover the entire duration of the conference, capturing all potential

price-sensitive news. For robustness, we tighten the event window to [-7, +7] and [-5, +5] and apply an anticipation test (reported in Section 4.5).

Table 1
Event study and co-integration windows

The table shows the training windows, the pre-event windows, and the post-event windows for the Paris Conference in 2015, the Paris Agreement in 2016, the Glasgow Conference in 2021, and the UAE conference in 2023. For policy event co-integration tests, the pre-event and post-event windows are defined for the Paris Conference and the Glasgow Conference.

	Event Study			
Period	**Paris Conference**	**Paris Agreement**	**Glasgow Conference**	**UAE Conference**
Training windows	2014-11-20 to 2015-11-26	2015-10-21 to 2016-10-26	2021-02-26 to 2021-10-27	2022-11-20 to 2023-11-27
Pre-event windows	2015-11-27 to 2015-12-11	2016-10-27 to 2016-11-04	2021-10-28 to 2021-11-11	2023-11-28 to 2023-12-12
Post-event windows	2015-12-12 to 2015-12-28	2016-11-05 to 2016-11-14	2021-11-12 to 2021-11-26	2023-12-13 to 2023-12-28
	Co-integration Test			
Period	**Paris Conference**	**Glasgow Conference**		
Pre-event window	2015-05-30 to 2015-11-30	GEO: 2021-04-30 to 2021-10-31 NGEO: 2021-07-30 to 2021-10-31		
Post-event window	2015-12-12 to 2016-06-12	2021-11-13 to 2022-05-13		

1.3.1 The constant mean return model

Let daily observed returns be denoted by r_t obtained by first differencing the price time series:

$$r_t = ln(\frac{P_{t+1}}{P_t}) \tag{1}$$

Using training or estimation windows of 265 days before each event, we estimate mean training window returns as:

$$r_{avg} = \frac{1}{T}\sum_{t}^{T} r_t \tag{2}$$

where T is the T*th* trading day in the training window [6]. Next, define an event window [-t, t] for each for the Paris Climate Conference, the Paris Climate Agreement, the Glasgow Climate Conference and the UAE Conference. Then the abnormal return (AR) for the t*th* day may be written as:

$$AR_t = r_t - r_{avg} \tag{3}$$

We next compute cumulative abnormal returns (CAR) as:

$$CAR = \sum_{\forall t} AR_t$$

[6] Table 2presents the training windows and descriptive statistics for the respective events

Although relatively simple compared to more sophisticated event study models, the constant mean return model remains a widely accepted approach, particularly when estimating abnormal returns at the market level. Brown and Warner (1985) show that the constant mean return model often yields results similar to more complex models. This insensitivity to model choice stems from the fact that the variance of abnormal returns is rarely reduced substantially by using market- based models Campbell et al. (1998); MacKinlay (1997); Binder (1998), particularly when asset betas are unstable and event windows are short or moderate, typically defined as less than one year (Kothari and Warner, 2007). In the context of financial assets with thin trading or weak correlation to broad indices, beta may be noisy and the relevant market index hard to define [7].

In the absence of a large cross-section of 'firms', inferential statistics from our single-commodity event study will suffer from substantial inferential errors using conventional test statistics [8]. Therefore, we employ the Sample Quantile (SQ) test following the inferential frame work proposed by Conley and Taber (2011). For each of EUA, CER, GEO, and NGEO, we sort $\hat{AR}$ in ascending order for all non-event days j:

$$\hat{AR}_{(1)} \leq \hat{AR}_{(2)} \leq \cdots < \hat{AR}_{(j)}$$

Define a ceiling operator $\lceil\rceil$ such that $\lceil x \rceil$ returns the integer q such that for $x < q < x+1, q(\alpha,n) = \lceil \alpha \times n \rceil$ is the $q(\alpha,n)$ order statistic of AR_j. We denote this value as $\hat{r}_{i,\alpha}$ which is the sample α quantile of the realized AR_k values. Finally, for pre and post-event windows, we reject the null hypothesis of no event effect if $|CAR| > |\hat{AR}_\alpha|$ for 90%, 95%, and 99% confidence levels.

1.3.2 Co-integration tests

Co-integration tests assess spurious correlations in non-stationary time series data. Here, we use co-integration to formally test for structural breaks or the decoupling of compliance from voluntary carbon prices following the policy events studied. Specifically, we test whether carbon prices are co-integrated before and after the events, using 6- month windows for the analysis, covering a total of 12 months. Optimal lag lengths are estimated using the Akaike Information Criterion (AIC). In the pre-event training window, we specify a generalized VAR process following Johansen (1988) as follows:

$$\mathrm{X}_t = \mu + \Pi_1 \mathrm{X}_{t-1} + \cdots + \Pi_k \mathrm{X}_{t-k} + \varepsilon_t \tag{4}$$

[7] Nonetheless, we also tested a market model as a robustness check and conduct anticipation tests. We constructed a global carbon market index equally weighted by the ICE Carbon Index and GEO and NGEO, used this as the market benchmark, and tested for the significance of abnormal returns to returns on individual market proxies. Results are qualitatively similar. See Appendix for further details.

[8]

Rewrite the model with first differences using the lag operator Δ:

$$\Delta X_t = \nu + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{k-1} \Delta X_{t-k+1} + \Pi X_{t-k} + \varepsilon_t \tag{5}$$

where $\Gamma_i = -(I - \Pi_1 - \cdots - \Pi_i)$ with $i = 1, \cdots, k-1$

and $\Pi = -(I - \Pi_1 - \cdots - \Pi_k)$

In our case:

X are 2x1 vectors in each case, i.e., EUA and CER, EUA and GEO, and EUA and NGEO.

If:

I. rank $(\Pi) = p$ then X_t is stationary and the rank is full,

II. rank $(\Pi) = 0$ then Π is a zero matrix, and there is no relationship between the variables

III. $0 <$ rank$(\Pi) = r < p$ then Π is of reduced rank and there exists two p x r matrices α and β such that $\Pi = \alpha\beta'$.

In III, the co-integrating vector or matrix β has a stationary process creating property if multiplied with the vector X_t, which means that the two variables are co-integrated. Next, we test the rank of Π using Juselius et al. (1990). We test the null hypothesis:

$$H_0 : \exists\ r \text{ co-integration relationships}$$

$$H_1 : \exists\ r+1 \text{ co-integration relationships}$$

This is the maximum eigenvalue test, which is χ^2-distributed, with the test characterized as $-T \cdot ln(1 - \lambda_{r+i})$ where λ are the eigenvalues.

We also conduct a trace test Juselius et al. (1990) for robustness, characterized by: $-T \sum_{i=h+1}^{p} ln(1 - \lambda_i)$. The last step is to estimate the co-integrating vector β. What we define is the co-integrating space using an arbitrary normalization vector. We apply the Phillips (1991) triangular representation to conveniently simulate a co-integrating vector $\beta = (1 - \hat{\beta})$ for the bivariate co-integrated system $X = (X_1, X_2)$. We subsequently test for the stationarity of the co-integration vector using the Phillips-Perron test by multiplying β with X_t. In the post-event window, the same vector β is used to build the co-integration vector and to test if stationarity is preserved. This would otherwise be proof of a structural shift in the pricing relationship in X.

1.4 Results

We present descriptive statistics for all price time series used in the empirical analyses in Table 2, including the mean, standard deviation, median, and 5 and 95 percentiles of the respective training windows. We apply a 265-day training window for the Paris Conference (2015), Paris Agreement (2016), and UAE Conference (2023). Mean EUA returns (0.78%) are marginally higher than CER returns (0.77%) ahead of the Paris Conference, even though CER exhibits greater price volatility in the training window (26.01%). Ahead of the Paris Agreement, mean returns were negative for EUA and CER, while GCMI reports non- negative returns over the period (0.00%). In the Glasgow Conference training window, we observe higher mean returns in GEO and NGEO compared to EUA and GCMI, while mean returns were flat or negative ahead of the UAE Conference.

Table 2
Summary statistics for training windows for each policy event

The table presents summary statistics for data used in the empirical analyses. Panel A provides summary statistics for the Paris Conference, Panel B for the Paris Agreement and Panel C for the Glasgow Conference. N represents the number of observations, Mean the arithmetic mean, SD is the standard deviation, and 5% and 95% are the 5- and 95-percentiles respectively of the underlying data.

	Descriptive statistics					
	N	Mean	SD	5%	Median	95%
Panel A: Paris Conference						
EUA	265	0.78%	1.57%	-2.32%	0.00%	2.54%
CER	265	0.77%	26.01%	-17.67%	0.00%	6.97%
GCMI	265	0.02%	1.13%	-1.66%	0.01%	1.79%
Panel B: Paris Agreement						
EUA	265	-0.13%	3.02%	-5.07%	-0.21%	4.64%
CER	265	-0.19%	2.02%	-3.37%	0.00%	2.72%
GCMI	265	0.00%	1.66%	-2.82%	-0.14%	2.42%
Panel C: Glasgow Conference						
EUA	170	0.27%	2.37%	-3.43%	0.46%	3.81%
GEO	170	0.71%	3.26%	-3.23%	0.00%	5.80%
NGEO	62	0.67%	2.95%	-2.64%	0.80%	4.84%
GCMI	170	0.05%	2.96%	-2.66%	0.00%	5.04%
Panel D: UAE Conference						
EUA	265	0.00%	2.03%	-3.06%	0.00%	3.26%
GEO	265	-0.65%	7.24%	-11.83%	-0.74%	11.02%
NGEO	255	-0.94%	6.51%	-10.91%	-0.60%	10.01%
GCMI	253	0.00%	1.59%	-2.68%	-0.03%	2.59%

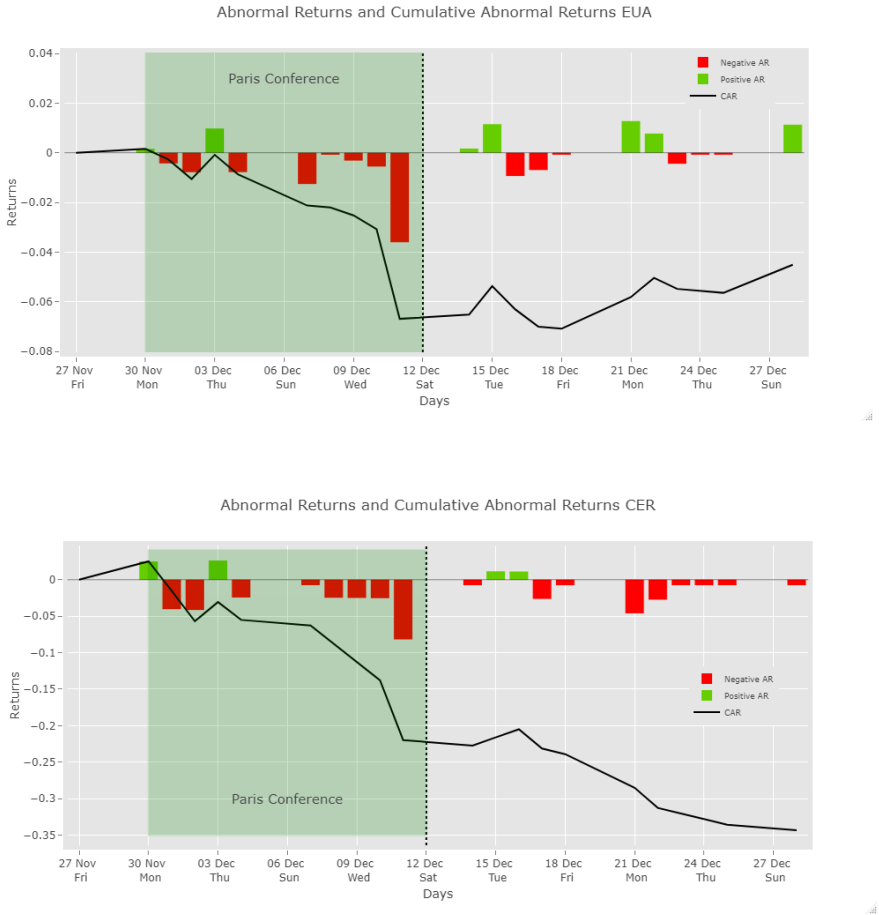
1.4.1 Event study results for the Paris Conference 2015 and the Paris Agreement 2016

We present results from the constant mean return model and the market model for robustness. Fig. 1 illustrates the ARs and CAR for EUA and CER during the Paris Conference. In the pre-conference window, ARs for EUA are mostly negative, except for November 30 and December 3, 2015, leading

to a downward CAR trend, reflecting negative market expectations. Post-conference, AR signs become more balanced, leading to a modest upward CAR trend. For CER, predominantly negative pre- and post-conference ARs result in a persistent downward CAR trend, indicating that conference outcomes failed to reverse the negative expectations. Table 3, Panel A, shows significant pre-conference CARs for EUA (-6.69%) and CER (-21.98%), reflecting negative market expectations.

Figure 1
Paris Conference constant mean return model: EUA/CER Abnormal Return (AR) and Cumulative Abnormal Return (CAR)

The figure shows the abnormal and cumulative abnormal returns of EUA and CER during and after the Paris conference. We calculate abnormal returns as the difference between an observed daily return during the event window and the average training window return estimated using 265 days before the Conference. CAR is the running cumulative sum of estimated AR. The last day of the Paris Conference is the policy event date, marked by the vertical dashed line. The event window is defined as 11 days before (shaded green) and after the Conference. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



The Paris Conference was significant in expanding the normative framework on adaptation, reaffirming the sub-2 C target, and introducing 5-year renewable NDCs. EUA saw a net 8.7% post-event reversal (2.01%), while CER's CAR remained negative (-11.59%) but showed a 47.3% improvement over the pre-event window, indicating some positive reaction to the conference. The uncertainty surrounding

CDM and Article 6 likely contributed to the CER response. Panel B shows similar results in the market model, though the post-event EUA CAR is not significant. Both models show smaller post-event market reactions compared to pre-event.

Table 3
Cumulative Abnormal Returns (CAR) before and after the Paris Conference

The table presents significance test results for estimated CAR for European Emission Allowances (EUA) and Certified Emissions Reductions (CER) using the Sample Quantile (SQ) test Conley and Taber (2011); Gelbach et al. (2013). The significance tests are conducted for the constant mean return model and use a market model with the Global Carbon Market Index (GCMI) as the market benchmark return for robustness. The event window comprises 11 days before (Pre) and after (Post) the Paris Conference. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, respectively.

	Cumulative Abnormal Returns (-11; +11)	
Panel A: Constant Mean Return Model		
	(1) EUA	(2) CER
Pre	-6.69%***	-21.98%**
Post	2.01%*	-11.59%*
Panel B: Market Model		
	(1) EUA	(2) CER
Pre	-4.77%***	-17.18%**
Post	0.91%	-14.35%*

The Paris Climate Agreement was released a year later, on November 4, 2016. The ARs are positive for EUA in the pre-event period, but negative ARs post-agreement, reflecting pre-event optimism that faded afterwards. A similar pattern emerges for CER, with small positive ARs before the event, followed by a strong negative reaction post-event. Table 4 presents event study results for the Paris Agreement. Pre-event CARs for EUA (10.81%) and CER (3.86%) are significantly positive. Post-event, EUA (-13.48%) and CER (-1.89%) report sharply reversed CARs, even though the CER figure does not show conventional significance. The market model confirms a significant post-event EUA CAR (6.35%), though it contrasts with the constant mean return model's negative findings. CER showed weak significance pre-event (2.83%) but no significant response post-event. Compliance markets responded more strongly to the Paris Agreement than voluntary carbon markets. The magnitude of the EUA response may in part be attributed to the structure of compliance markets, which are more responsive to new regulatory and policy initiatives by design. In contrast, uncertainty surrounding CDM, the proposed Sustainable Development Mechanism, and filled adaptation finance pledges likely contributed to CER's post-event response.

Table 4
Cumulative Abnormal Returns (CAR) before and after the Paris Agreement

The table presents significance test results for estimated CAR for European Emission Allowances (EUA) and Certified Emissions Reductions (CER) using the Sample Quantile (SQ) test Conley and Taber (2011); Gelbach et al. (2013). The significance tests are conducted for the constant mean return model and use a market model with the Global Carbon Market Index (GCMI) as the market benchmark return for robustness. The event window comprises 6 days before (Pre) and after (Post) the Paris Agreement. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, respectively.

	Cumulative Abnormal Return (-6; +6)	
Panel A: Constant Mean Return Model		
	(1) EUA	(2) CER
Pre	10.81%***	3.86%**
Post	-13.48%***	-1.89%
Panel B: Market Model		
	(1) EUA	(2) CER
Pre	-2.37%**	2.79%*
Post	6.36%***	-0.25%

1.4.2 Event study results for the Glasgow conference (COP 26)

The third policy event we study is the Glasgow Conference held in November 2021. Table 5shows the event study results from the constant mean return model, with EUA (5.67%), GEO (8.62%), and NGEO (22.18%) building up significant expectations during the conference. Post-event, EUA (11.25%) and NGEO (5.25%) show strong and significant responses, with the market model confirming a larger posi tive CAR for EUA (6.21%) compared to the pre-event window (-2.24%). NGEO's post-event CAR, while not statistically significant, reflected continued optimism [9]. These results indicate optimistic expectations ahead of COP26 from both compliance and voluntary carbon markets. In the short run, compliance markets responded positively to conference outcomes, while GEO participants were generally disappointed. NGEO, meanwhile, maintained a flat CAR in the post-event window.

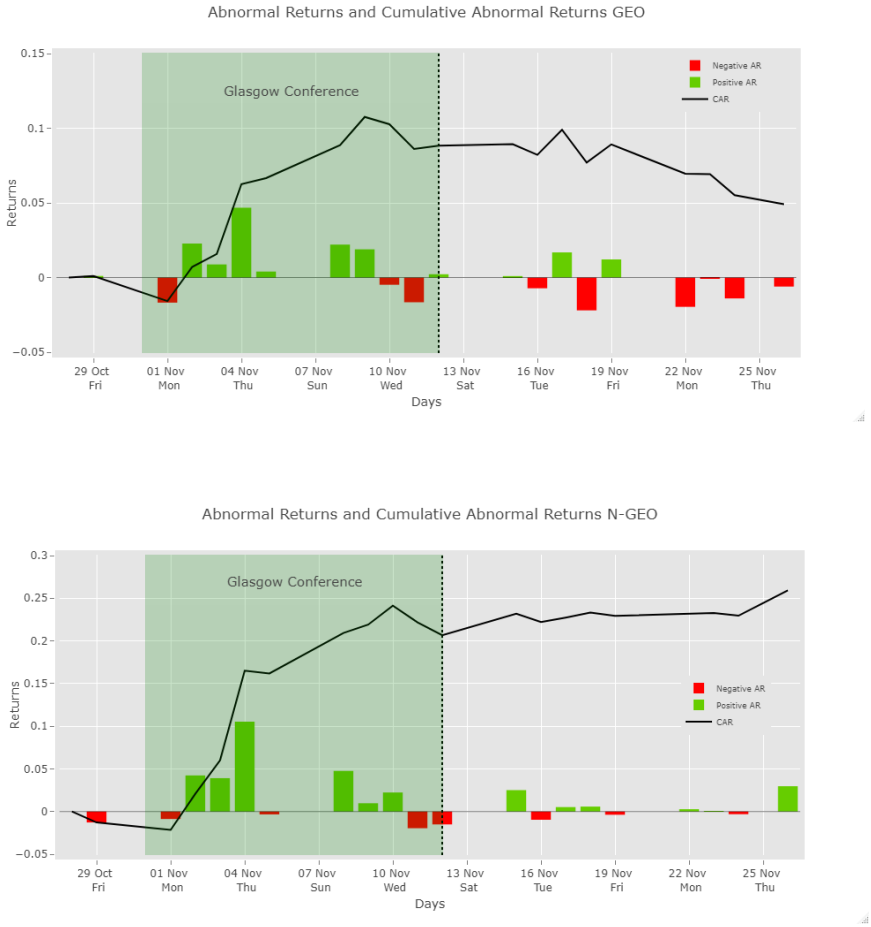
The strong performance of GEO and NGEO during the pre-event window coincided with discussions on finalizing Article 6 of the Paris Rulebook. COP26 also produced key pledges for carbon markets, including phasing down coal use by 2030 and removing fossil fuel subsidies. A coal phase-down pressures utilities to adopt cleaner technologies or face higher compliance costs. Parties also agreed to end illegal deforestation by 2030 and incentivize sustainability-conscious finance. Such pledges, if implemented, may provide enduring and consistent pricing signals to carbon markets. While GEO saw negative CARs post-event, a key focus for voluntary carbon markets at COP26 was adaptation finance, particularly the

[9] These results are validated in Fig. 2 (GEO and NGEO). Fig. 2 shows daily GEO and NGEO ARs and CAR for the Glasgow Conference. Pre-event, both GEO and NGEO report mostly positive AR. Estimated AR values, however, are mostly negative in the post-event window for GEO, resulting in a downward trend for CAR. For NGEO, we find an initially flat CAR post-event with an upward spike at the end of the event window as positive and negative ARs largely cancel out.

100 billion U.S. Dollar annual target now to be replaced by the New Quantitative Goal on Climate Finance. Fund flows to GEO-related sectors, such as Waste Disposal and Household Devices, could help low-income countries develop climate resilience. Meanwhile, NGEO's pre-event increase was driven by the pledge to end illegal deforestation by 2030.

Figure 2
Glasgow Conference constant mean return model: GEO/N-GEO Abnormal Return (AR) and Cumulative Abnormal Returns (CAR)

The figure shows the abnormal and cumulative abnormal returns of GEO and N-GEO during and after the Glasgow Conference. We calculate abnormal returns as the difference between an observed daily return during the event window and the average training window return estimated training windows in Table 1. CAR is the running cumulative sum of estimated AR. The last day of the Glasgow Conference is the policy event date, marked by the vertical dashed line. The event window is defined as 11 days before (shaded green) and after the Conference.



1.4.3 Event study results for the UAE Conference (COP 28)

Table 6 provides the event study results of the COP 28 conference. The results are broadly consistent across both the constant mean return and market models, except for EUA in the pre-event window, where there is a sign reversal. For both models, CAR is statistically significant for the compliance market. For GEO, we find a strong positive reaction (16%) in the pre-event window while the post-event response

Table 5
Cumulative abnormal returns before and after the Glasgow Conference

The table presents significance test results for estimated CAR for European Emission Allowances (EUA), Global Emissions Offsets (GEO), and Nature-based Global Emissions Offsets (NGEO) using the Sample Quantile (SQ) test Conley and Taber (2011); Gelbach et al. (2013). The significance tests are conducted for the constant mean return model and use a market model with the Global Carbon Market Index (GCMI) as the market benchmark return for robustness. The event window comprises 11 days before (Pre) and after (Post) the Glasgow Conference. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, respectively.

	Cumulative Abnormal Return (-11; +11)		
Panel A: Constant Mean Return Model			
	(1) EUA	(2) GEO	(3) NGEO
Pre	5.67%***	8.62%**	22.18%***
Post	11.25%***	-3.92%*	5.25%**
Panel B: Market Model			
	(1) EUA	(2) GEO	(3) NGEO
Pre	-2.24%***	7.68%**	20.87%***
Post	6.21%***	-4.31% **	4.70%**

is insignificant. NGEO results are noteworthy, as pre-event expectations are flat (CAR insignificant) but report a significant collapse in NGEO prices in the post-event period. This result is validated by the market model. It is difficult to place the UAE Conference in the proper context from a compliance carbon market perspective. The UAE Conference signalled a global exit from fossil fuels but lacked concrete, actionable steps on the phasing out of fossil fuels. Parties could not agree on a new Paris-aligned global carbon market mechanism that would replace the CDM (Article 6) yet agreed on the need to triple renewable capacity, double the annual rate of energy efficiency improvements by 2030, and finalise commitments on Loss and Damage. The UAE Conference 'said the right things' but lacked sufficient concrete resolutions to move compliance carbon markets overall. The same cannot be said about the voluntary carbon markets. GEO and NGEO responses are broadly net negative for reasons similar to those of compliance markets and the lack of progress on Article 6. In particular, countries were divided on reporting requirements and concerns over confidential information (Article 6.2), as well as disagreement on how to mobilize voluntary carbon markets in terms of connectivity and eligibility (Article 6.4).

1.4.4 Co-integration results for the Paris Conference 2015 and the Glasgow Conference 2021

Our co-integration results offer insights into how policy events shaped the pricing relationship between compliance and voluntary carbon markets, particularly whether a decoupling occurred after the events. Fig. 3 shows the co-integrating process and the price differential between EUA and CER using a time

Table 6
Cumulative abnormal returns before and after the UAE Conference

The table presents significance test results for estimated CAR for European Emission Allowances (EUA), Global Emissions Offsets (GEO), and Nature-based Global Emissions Offsets (NGEO) using the Sample Quantile (SQ) test Conley and Taber (2011); Gelbach et al. (2013). The significance tests are conducted for the constant mean return model and use a market model with the Global Carbon Market Index (GCMI) as the market benchmark return for robustness. The event window comprises 11 days before (Pre) and after (Post) the UAE Conference. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, respectively.

	Cumulative Abnormal Return (-11; +11)		
Panel A: Constant Mean Return Model			
	(1) EUA	(2) GEO	(3) NGEO
Pre	-6.52%***	16.00%**	5.74%
Post	11.40%***	-1.28%	-99.19%***
Panel B: Market Model			
	(1) EUA	(2) GEO	(3) NGEO
Pre	4.65%***	19.11%**	7.70%*
Post	-2.12%**	-5.20%	-101.59%***

window of 6 months before (top figure) and after the Paris Conference. Visually, the co-integrating process appears stationary before the 2015 Paris Conference. There is a relative price inversion on October 4, 2015, two months before the conference. This implies in our case, increased volatility in CER prices lead to higher price returns than for EUA. We interpret this as an increased price-in of expectations for voluntary carbon markets ahead of the Paris Conference. Fig. 3 also shows the co-integrating process and price differential between EUA and CER post-Paris. We observe the non-stationarity of the co-integration process post-event [10]. The visual results are statistically confirmed in Table 7. Table 7 presents the co-integration test results before and after the Paris and Glasgow conferences using the Johansen Eigenvalue and Johansen Trace tests. The Phillips-Perron (PP) test results for stationarity of the post-event co-integrating process for each of the Johansen tests are included [11]. We test the null hypothesis that the rank (r) of the co-integrating vector is equal to zero. Rejecting the null hypothesis implies at least one co-integration relationship between the prices. Rejecting r$\leq$ 1 implies at least two co-integration relationships between pairs in our case. We argue that the absence of a co-integration relationship post-event reflects a structural shift in the pricing rela tionship between compliance and voluntary carbon markets. The absence of a co-integration relationship represents a breakdown in the long-run relationship between compliance and voluntary markets post a policy event. This may be viewed as empirical evidence of the disruptive nature of the policy event on carbon markets. In the Johansen

[10] Graphical evidence for the Glasgow COP is presented in the Appendix.

[11] The beta generation of the co-integration process derived from the generalized VAR specification is not necessarily unique and hence, depends on the co-integration test.

Eigen-value test, we observe weak stationarity of the co-integrating process generated from EUA and CER pre-Paris in PP tests (p-value of 0.02), which does not exist post-Paris (p-value 0.87). This result is confirmed in the Trace Test. For the Glasgow Conference, we observe the existence of stationary co-integration relationships before and after the conference for EUA-GEO (p-values: 0.01 (pre), 0.09 (post)) and EUA-NGEO (p- values: 0.03 (pre), 0.02 (post)) using the Johansen Eigenvalue test. The trace test confirms these results.

Table 7
Co-integration results before and after the Paris and Glasgow Conferences

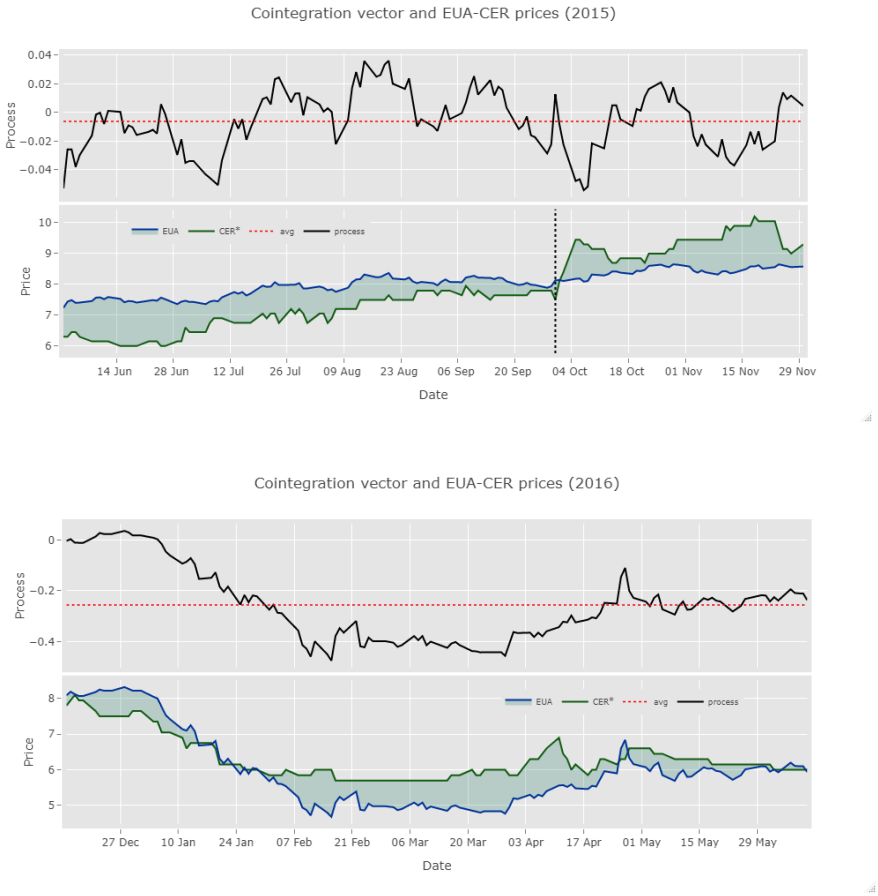
The table presents the Johansen (Eigenvalue and Trace) co-integration test results on the price time series and the Philipps-Perron stationarity test for stationarity on the cointegrated vector pairs EUA - CER, EUA - GEO and EUA - NGEO. Co-integration windows are described in Table 1. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, respectively.

	Johansen Eigenvalues Test		Philipps-Perron Test on the co-integration vector	Johansen Trace Test		Philipps-Perron Test on the co-integration vector
	Pairs Paris Conference					
	H_0	Test	p-value	H_0	Test	p-value
EUA–CER pre	$r \leq 1$ $r = 0$	3.29 13.32*	0.02**	$r \leq 1$ $r = 0$	3.29 16.61*	0.02**
EUA–CER post			0.89			0.89
	Pairs Glasgow Conference					
EUA–GEO pre	$r \leq 1$ $r = 0$	6.51 18.58**	0.01***	$r \leq 1$ $r = 0$	6.51 25.09***	0.01**
EUA–GEO post			0.09*			0.09*
EUA–NGEO Pre	$r \leq 1$ $r = 0$	8.30* 20.36***	0.03**	$r \leq 1$ $r = 0$	8.30* 28.66***	0.03**
EUA–NGEO Post			0.02**			0.02**

The results of the co-integration tests suggest that voluntary carbon markets priced in expectations ahead of the Paris and Glasgow conferences more strongly than compliance markets. With regards to the Paris Conference, the co-integration relationship vanishes post-event. We interpret this as a decoupling of the voluntary and compliance carbon market prices post-Paris as negotiations for the Agreement progressed. While the Paris Conference has been described as a watershed moment in global climate policy, its non-binding nature, unfulfilled expectations on adaptation finance, and the lack of finality on Article 6 have been broadly criticized. Our results demonstrate this in both the event studies and the co-integration tests. The breakdown of pricing relationships between compliance and voluntary carbon markets after the Paris Conference suggests policy outcomes catered more to a section of carbon markets than another.

Figure 3
Co-integration between European Emission Allowances (EUA) and Certified Emissions Reductions (CER) prices before and after the Paris Conference

The figure shows the co-integration vector, EUA, and CER prices six months before and six months after the Paris Conference. The horizontal dashed line represents the average of the co-integration process while the vertical dashed line marks the price relative price inversion between EUA and CER two months before the conference. The green shade is the difference between the prices. CER is scaled by factor 15.



1.4.5 Robustness tests

The study period inevitably overlaps with exogenous shocks that could have distorted carbon prices irrespective of COP-related information. To guard against this risk, we implemented two complementary tests whose combined evidence shows that the COP effects we document are not artefacts of these wider disturbances.

When an event is pre-empted, asset returns around the announcement of the event may have an inconsistent sign: a positive sign around negative news, or vice versa. To address anticipation effects, we run simulations of the constant mean return model using random event windows for the events we study. These event windows are created using a random event date selected one month before the beginning of the conference/agreement. The choice of the event date is random to the extent that the event window does not intersect with the actual policy event window. The design elements of the anticipation tests

are otherwise analogous to the event studies discussed above. The anticipation tests aim to compare investors' expectations about carbon price developments before the respective policy event windows. Our prior then is that the policy events induce a price recalibration in carbon (both compliance and voluntary) price returns, measured in terms of CAR compared to the status quo. If the large cumulative abnormal returns (CARs) observed in the true windows were driven by other shocks, similarly large and systematic CARs would emerge in the simulations. They do not: across the Paris, Glasgow and UAE COPs the simulated CAR distribution is not significantly different from zero in almost every case, whereas the real windows yield large, statistically significant CARs (see Tables Appendix). The anticipation tests therefore rule out both simple macro drift and pre-emptive price adjustment as explanations for our findings. The market model, explained in detail in the appendix, provides a second robustness test. It replaces the constant-mean return benchmark with a regression on the Global Carbon Market Index (GCMI). This specification explicitly nets out any price movement common to the whole carbon complex, such as the broad sell-off during the first COVID lockdowns or the EUA rally during the 2022 gas crisis. Re-estimating every window under this alternative yields CARs of the same sign and, in most instances, the same significance level as the constant-mean results. Thus, even after controlling for contemporaneous shocks that moved all carbon instruments together, the abnormal returns around COPs remain intact.

Taken together, the anticipation tests eliminate bias from price for mation that precedes the event windows, while the market model filters out shocks that affect the carbon markets system- wide. The convergence of results across both methods demonstrates that the price reactions we attribute to COP negotiations are robust to external shocks that occurred within the sample period. The co-integration analysis is also subjected to a structural-break robustness check. The Johansen test presupposes that the long-run parameters remain constant over time, yet one purpose of global climate-policy summits is precisely to shift market expectations abruptly; in that circumstance, the null of no co-integration can be rejected spuriously even when a stable relationship prevails within shorter, distinct regimes (Gregory and Hansen, 1996). To eliminate that risk, we re-estimate the systems allowing for multiple unknown breaks by applying the Bai–Perron sequential breakpoint algorithm as implemented by Zeileis et al. (2003). The breakpoint procedure detects two statistically significant regimes as shown in Fig. A.8 that align closely with the Paris window [12]. Within each regime, the residuals of the cointegrating vector remain stationary, confirming that the linkage is genuine but state-dependent. Because co-integration tests that ignore structural instability can over-reject or misdate integration (Qu and Perron, 2007), demonstrating residual stationarity inside each stable regime strengthens our claim that COP meetings act as discrete structural shocks rather than mere statistical noise, fully consistent with the event-study evidence and the

[12] Glasgow is presented in the Appendix

anticipation and market-model checks already discussed.

1.5 Discussion and implications for global carbon policy

We have shown through the event study and co-integration analyses that the four COP events attracted significant policy premiums in compliance and voluntary carbon markets. Carbon markets are a critical market-based mechanism for driving the green transition by incentivizing lower emissions. However, carbon policy administration remains fragmented across jurisdictions, as countries pursue independent mitigation and adaptation strategies, leading to inconsistent pricing signals. Our results highlight that market participants derive the most value from COPs when they serve as platforms for coordinated global engagement, producing binding, actionable, and ambitious policy outcomes with a clear trickle-down effect to national regulatory works. For example, the Paris Agreement's establishment of NDCs provides a direct and actionable commitment to parties. The EUA's positive response to the Glasgow commitments illustrates how regulatory actions, such as the phasing out of coal, can directly affect policy, e. g., reduced emissions caps in the EU ETS system, which further incentivize cost-efficient production adjustments across the economy.

In contrast, voluntary carbon markets (VCMs) exhibit a more nuanced response, often marked by pre-event optimism that fails to materialize post-event if outcomes lack ambition or clarity. For instance, GEO and NGEO markets rallied on specific commitments at the Glasgow Conference, such as the pledge to end illegal deforestation, while the CER markets declined following the Paris Conference due to ongoing uncertainty surrounding the future of the CDM under Article 6. This reflects the sensitivity of VCMs to targeted policy outcomes that align with specific sectors and goals.

Our results suggest that carbon market participants evaluate policy credibility not only based on the ambition of COP outcomes but also on their enforceability, clarity, and track record of implementation. Investors interpret credibility through a combination of factors: the presence of institutional enforcement mechanisms, the specificity and measurability of commitments, whether historical pledges have been fulfilled (e.g., on adaptation finance), and procedural transparency. These elements map onto three key policy areas: substantive vs. symbolic commitments, sector-specific vs. systemic focus, and institutional clarity vs. ambiguity. For example, the finalisation of the Article 6 Rulebook at COP 26 provided a clear framework for cross-border crediting, which stabilised expectations across both compliance and coluntatary carbon markets. In contrast, the delays in reaching agreements on operationalising Article 6.4 at COP 28, along with vague language on fossil fuel transition, introduced ambiguity and undermined investor confidence.

Voluntary carbon markets, by design, are crucial for innovation in mitigation and resilience, fostering the

development of bankable technologies, financed through project-linked carbon credits. Our analysis suggests that the value of COPs to carbon markets will diminish if policymakers fail to implement pledges or if market signals transmitted through these events are weak or unclear. Looking ahead to COP 30 in Bel'em, Brazil, in 2025, our findings underscore the importance of delivering policy instruments that are not only ambitious but also institutionally grounded and transparently implementable, particularly in the case of voluntary carbon markets, where credibility hinges on standardization, governance, and transparent price discovery.

1.6 Conclusion

Conferences of Parties are pivotal in shaping global carbon policy. This paper analysed the impact of four key climate policy events, COP 21 Paris (2015), the Paris Agreement (2016), COP 26 in Glasgow (2021), and COP 28 in the UAE (2023), on the pricing of carbon in compliance and voluntary markets. Using publicly available data, we applied the constant mean returns model and, for robustness, a market model to estimate market reactions to these events. Our results show that carbon markets respond significantly to COP outcomes. Positive, statistically significant price reactions occur when policy outcomes are perceived as credible, actionable, and broadly consistent with pre-event expecta tions, while negative reactions follow outcomes viewed as lacking ambition, clarity, or implementation certainty. Events with weak commitments risk disrupting pricing relationships, as illustrated by the breakdown in long-run co-movement between compliance and voluntary carbon markets after the Paris Conference. Conversely, when expectations are partially fulfilled, as in the Glasgow Conference, the pricing relationship between these markets remains stable.

These findings carry clear policy implications for future COPs. First, COPs must deliver actionable, enforceable, and sector-specific policy instruments. Ambition alone is insufficient; and credibility depends on implementation. Second, credibility is cumulative. Investors assess new commitments in light of past delivery. Delays in fulfilling pledges weaken trust and diminish the pricing power of COP announcements. Third, policies must be tailored to the structure of each market. Compliance markets are shaped by regulatory enforcement, while voluntary carbon markets depend on policy signals aligned with project- level dynamics and robust verification standards. Our analysis shows that markets are already distinguishing between COPs on their effectiveness. Preserving the relevance of future COPs requires translating international ambition into credible and operational policy frameworks that strengthen price discovery, guide investment, and accelerate decarbonisation.

Appendix A

The market model for climate policy event studies

To obtain a market proxy, we construct an equally-weighted global carbon market index (GCMI) for the respective events as follows:

$$GCMI = \frac{1}{2}(ICE\ Carbon\ Index + CER) \tag{A.1}$$

$$GCMI = \frac{1}{2} ICE\ Carbon\ Index + \frac{1}{2} \begin{cases} GEO & pre\,NGEO \\ GEO(\frac{volume\ GEO}{total\ volume}) + NGEO(\frac{volume\ NGEO}{total\ volume}) & post\,NGEO \end{cases} \tag{A.2}$$

The equally-weighted GCMI assumes that carbon markets respond equally to climate policy events [13]. The ICE Carbon Index measures the per formance of a long-only basket of EUA, UK ETS, ICE California Carbon Allowance, and the ICE Regional Greenhouse Gas Initiative Futures Contracts. The first equation above is our reference market for the Paris Conference/Agreement, while the second is the same for the Glasgow and UAE conferences. The OLS regression for the market model is:

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \varepsilon_{i,t} \tag{A.3}$$

$r_{i,t}$ in the equations represents log returns for compliance and voluntary carbon market price proxies, and $r_{m,t}$ is the benchmark market return for carbon using the GCMI. These OLS regressions are run in the training windows described in Table 1. Next, for each event or non-event date t, abnormal returns are calculated as follows:

$$\hat{AR}_t = r_{i,t} - \hat{\alpha}_i - \hat{\beta} r_{m,t} \tag{A.4}$$

[13] Ignoring relative market sizes, equally weighted fits well with the theoretical framework we build on, that news should equally diffuse through compliance and voluntary carbon markets. GCMI weights both carbon market sectors equally and accounts for GEO and NGEO in voluntary carbon markets on the basis of when they are launched.

Appendix figures

Figure A.1
European Emissions Allowances (EUA) and Certified Emissions Reduction (CER) prices; Global Emissions Offsets (GEO) and Nature-based Global Emissions Offsets (NGEO) prices compared to EUA prices

The top figure shows EUA and CER prices from 2009 to 2016. EUA represents rolling year-ahead futures of EU ETS prices. CER prices are underlined by Clean Development Mechanism - certified projects. The underlying data was obtained from Datastream. The bottom figure shows GEO, NGEO, and EUA prices. GEO represents the Global Emissions Offset Futures underlined by voluntary carbon projects that are non-agricultural, forestry, and other land-use (AFOLU) projects. GEO-based certificates meet the eligibility criteria for the Carbon Offsetting and Reduction Scheme for International Aviation (CORSIA). NGEO offers the possibility to offset carbon emissions using nature-based AFOLU projects. The underlying data was obtained from DataStream and Bloomberg.

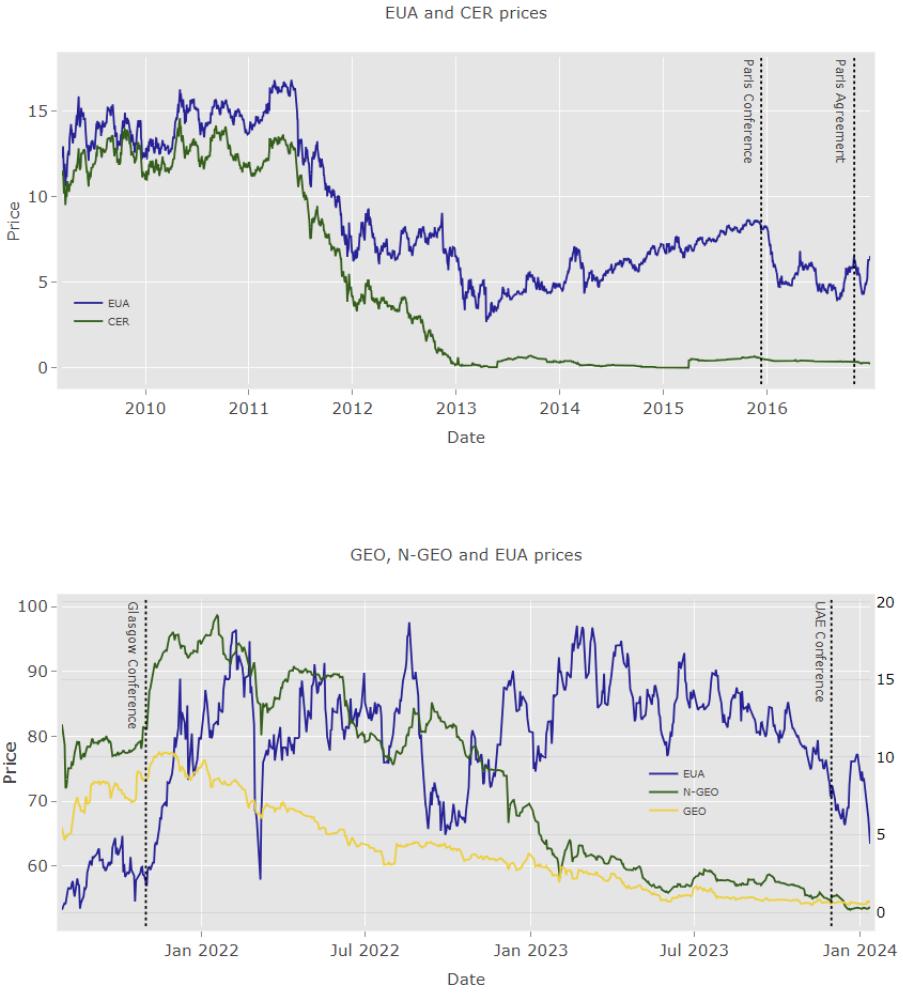


Figure A.2
Paris Agreement constant mean return model: EUA/CER Abnormal Returns (AR) and Cumulative Abnormal Returns (CAR)

The figure shows the abnormal and cumulative abnormal returns of EUA and CER during and after the Paris Agreement. We calculate abnormal returns as the difference between an observed daily return during the event window and the average training window return estimated using 265 days before the Agreement. CAR is the running cumulative sum of estimated AR. The last day of the Paris Agreement is the policy event date, marked by the vertical dashed line. The event window is defined as 6 days before (shaded green) and after the Agreement.

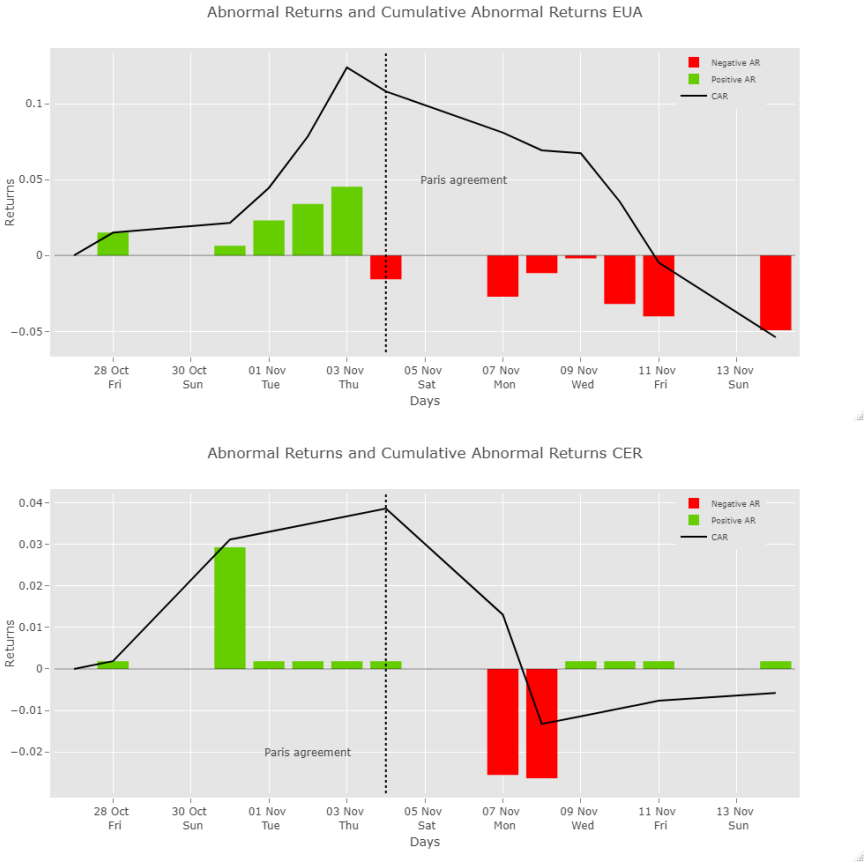


Figure A.3
Glasgow Conference constant mean return model: European Allowances (EUA) Abnormal Returns (AR) and Cumulative Abnormal Returns (CAR)

The figure shows the abnormal and cumulative abnormal returns of EUA during and after the Glasgow Conference. We calculate abnormal returns as the difference between an observed daily return during the event window and the average training window return estimated using 265 days before the Glasgow Conference. CAR is the running cumulative sum of estimated AR. The last day of the Glasgow Conference is the policy event date, marked by the vertical dashed line. The event window is defined as 11 days before (shaded green) and after the Conference.

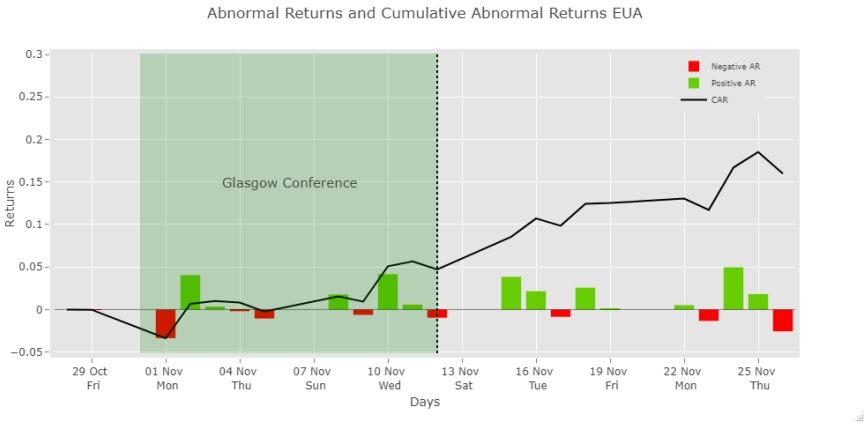


Figure A.4
UAE Conference constant mean return model: EUA Abnormal Return (AR) and Cumulative Abnormal Return (CAR)

The figure shows the abnormal and cumulative abnormal returns of European Allowances (EUA) during and after the UAE conference. We calculate abnormal returns as the difference between an observed daily return during the event window and the average training window return estimated using 265 days before the Conference. CAR is the running cumulative sum of estimated AR. The last day of the UAE Conference is the policy event date, marked by the vertical dashed line. The event window is defined as 11 days before (shaded green) and after the Conference.

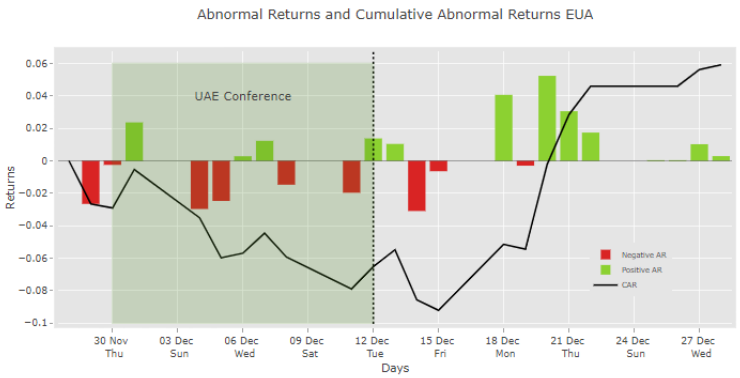


Figure A.5
UAE Conference constant mean return model: GEO/N-GEO Abnormal Return (AR) and Cumulative Abnormal Returns (CAR)

The figure shows the abnormal and cumulative abnormal returns of GEO and N-GEO during and after the UAE Conference. We calculate abnormal returns as the difference between an observed daily return during the event window and the average training window return estimated using training windows in Table 1. CAR is the running cumulative sum of estimated AR. The last day of the UAE Conference is the policy event date, marked by the vertical dashed line. The event window is defined as 11 days before (shaded green) and after the Conference.

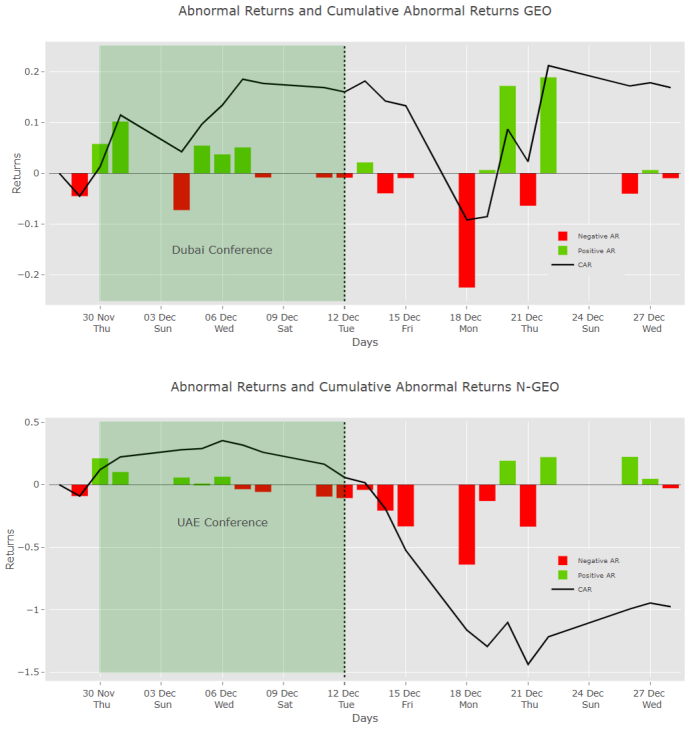


Figure A.6
Co-integration between European Emission Allowances (EUA) and Global Emission Offsets (GEO) prices before and after the Glasgow Conference

The figure shows the co-integration vector, EUA, and GEO prices six months before and six months after the Glasgow Conference. The horizontal dashed line represents the average of the co-integration process while the vertical dashed line marks the price relative price inversion between EUA and GEO two months before the conference. The green shade is the difference between the prices. GEO is scaled by factor 10.

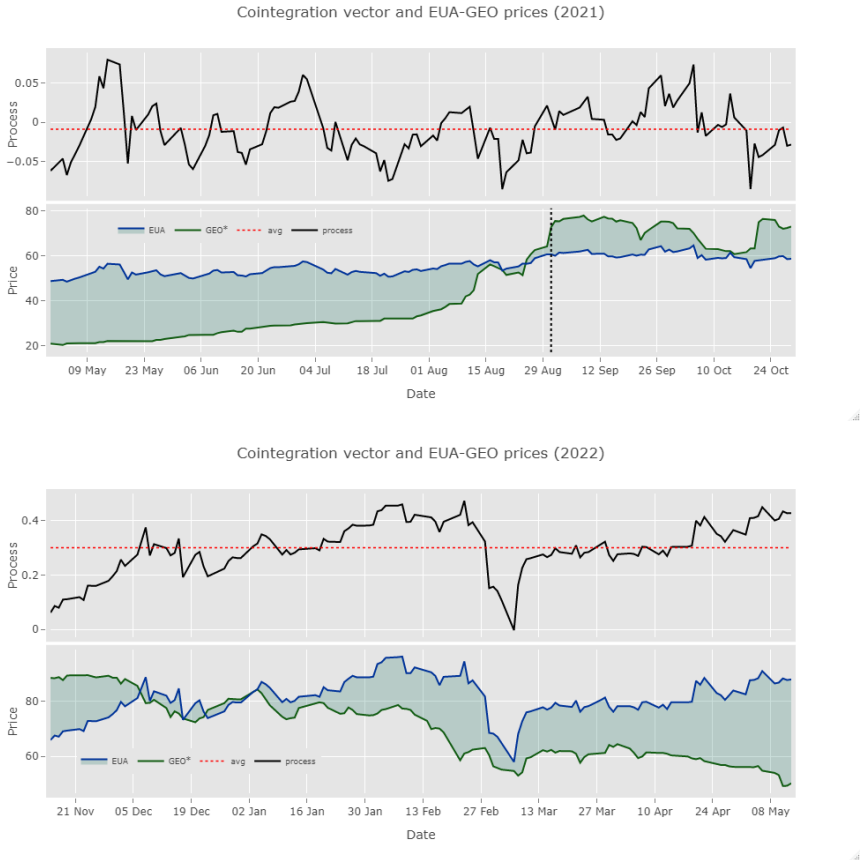


Figure A.7
Co-integration between European Emission Allowances (EUA) and Nature-based Global Emission Offsets (NGEO) prices before and after the Glasgow Conference

The figure shows the co-integration vector, EUA, and NGEO prices three months before and six months after the Glasgow Conference. The horizontal dashed line represents the average of the co-integration process. The green shade is the difference between the prices. NGEO is scaled by factor 10.

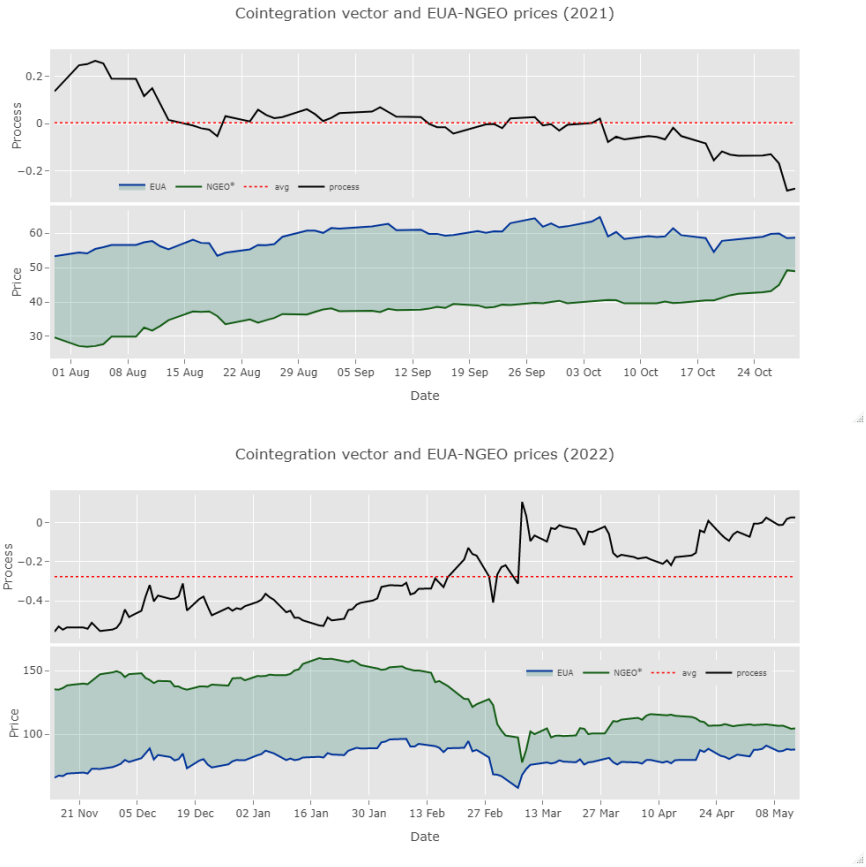


Figure A.8
Structural break between European Emission Allowances (EUA) and Certified Emissions Reductions (CER) prices before and after the Paris Conference

The figure shows the structural break (blue dotted line) between EUA and CER prices six months before and six months after the Paris Conference.

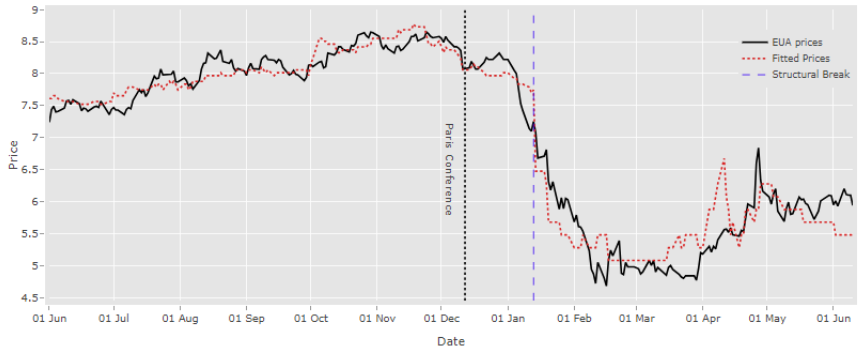


Figure A.9
Structural break between European Emission Allowances (EUA) and Global Emission Offsets (GEO) prices before and after the Glasgow Conference

The figure shows the structural break (blue dotted line) between EUA and CER prices six months before and six months after the Paris Conference.

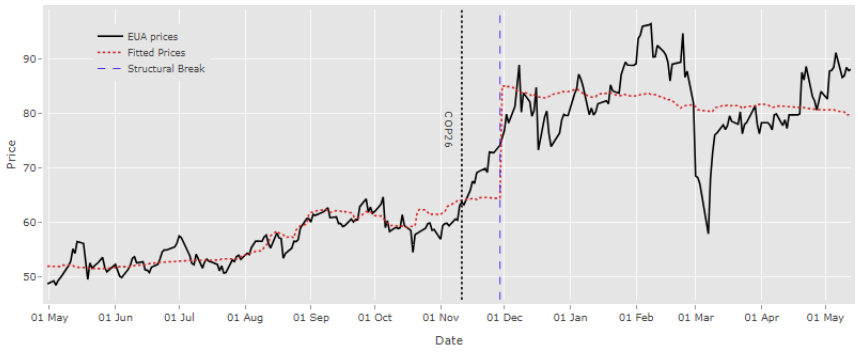
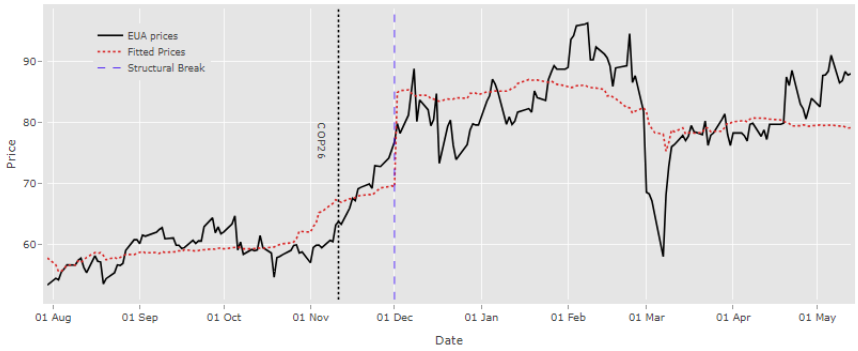


Figure A.10
Structural break between European Emission Allowances (EUA) and Nature-based Global Emission Offsets (NGEO) prices before and after the Glasgow Conference

The figure shows the structural break (blue dotted line) between EUA and CER prices six months before and six months after the Paris Conference.



Appendix tables

Table A.1
Variable definitions

The table summarizes the variables and abbreviations used in the analysis.

Abbreviation	In full	Meaning
AR	Abnormal Return	Difference between observed and expected return over the event window.
CAR	Cumulative Abnormal Return	Sum of abnormal returns over the event window, capturing total market reaction.
COP	Conference of the Parties	Annual UN climate negotiations where global agreements are discussed.
EUA	European Union Allowance	Tradable compliance carbon units under the EU ETS.
CER	Certified Emissions Reduction	Offsets issued under the Kyoto Protocol's Clean Development Mechanism (CDM).
GEO	Global Emissions Offset	Futures contract for non-AFOLU voluntary carbon offsets traded on exchanges.
NGEO	Nature-based Global Emissions Offset	Futures contract for AFOLU-based voluntary carbon offsets.
VCM	Voluntary Carbon Market	Markets for carbon offsets outside regulatory obligations.
EU ETS	European Union Emissions Trading System	The European Union's compliance carbon market, operating under a cap-and-trade scheme.
GFANZ	Glasgow Financial Alliance for Net Zero	Coalition of financial institutions committed to net-zero targets.
NDC	Nationally Determined Contribution	Country-specific emissions reduction commitments under the Paris Agreement.
SDM	Sustainable Development Mechanism	Mechanism under Article 6.4 for trading emission reductions between countries.
CDM	Clean Development Mechanism	Kyoto Protocol mechanism allowing developed countries to fund emission reduction projects in developing countries.
A6.4ERs	Article 6.4 Emission Reductions	Carbon credits issued under the SDM of the Paris Agreement.
ITMO	Internationally Transferrable Mitigation Outcome	Emission reduction unit transferred between countries under Article 6.2.
GCMI	Global Carbon Market Index	Custom benchmark used to proxy carbon market performance in estimation model.

Table A.2
Cumulative Abnormal Returns (CAR) for Paris Conference and Paris Agreement Anticipation Tests

The table presents significance test results for estimated CAR for European Emission Allowances (EUA) and Certified Emissions Reductions (CER) using the sample T-test. The anticipation tests are conducted using a simulation of event studies (constant mean return model) using a set of random dates one month before the beginning of the conferences. The event window comprises 11 days before (Pre) and after (Post) this random event date (akin to the Paris Conference). Similarly, we consider 6 days before and after a random event date to be akin to the Paris Agreement. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, respectively.

Paris Conference		
	Cumulative Abnormal Return (-11; +11)	
	(1) EUA	(2) CER
Pre	-0.19%	-1.18%
Post	0.41%	-4.41% **
Paris Agreement		
	Cumulative Abnormal Return (-6; +6)	
	(1) EUA	(2) CER
Pre	-2.07%	-2.66%
Post	-6.03%	3.74%

Table A.3
Cumulative Abnormal Returns (CAR) for Glasgow and UAE Conference Anticipation Tests

The table presents significance test results for estimated CAR for European Emission Allowances (EUA), Global Emissions Offsets (GEO), and Nature-based Global Emissions Offsets (NGEO)) using the sample T-test. The anticipation tests are conducted using a simulation of event studies (constant mean return model) using a set of random dates one month before the beginning of the conferences. The event window comprises 11 days before (Pre) and after (Post) this random event date (akin to the Glasgow (UAE) Conference). *, **, and *** indicate significance at the 10, 5, and 1 percent levels, respectively.

Glasgow Conference			
	Cumulative Abnormal Return (-11; +11)		
	(1) EUA	(2) GEO	(3) NGEO
Pre	-2.81%	-5.11%	-7.95%***
Post	-4.43%	-10.31%**	-2.04%
UAE Conference			
	Cumulative Abnormal Return (-11; +11)		
Pre	-0.88%	-2.77%	-16.09%***
Post	-1.11%	-2.91%	-14.63%*

Table A.4
Cumulative Abnormal Returns (CAR) before and after the Paris Conference

The table presents significance test results for estimated CAR for European Emission Allowances (EUA) and Certified Emissions Reductions (CER) using the Sample Quantile (SQ) test Conley and Taber (2011); Gelbach et al. (2013). The significance tests are conducted for the constant mean return model and use a market model with the Global Carbon Market Index (GCMI) as the market benchmark return for robustness. The event window comprises 7 and 5 days before (Pre) and after (Post) the Paris Conference. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, respectively.

	Cumulative Abnormal Returns (-7; +7)	
Panel A: Constant Mean Return Model		
	(1) EUA	(2) CER
Pre	-6.60%***	-18.91%*
Post	1.48%	-8.52%*
Panel B: Market Model		
	(1) EUA	(2) CER
Pre	-2.68%***	-9.08%*
Post	0.99%*	-9.75%*
	Cumulative Abnormal Returns (-5; +5)	
Panel A: Constant Mean Return Model		
	(1) EUA	(2) CER
Pre	-4.56%***	-15.69%*
Post	-0.01%	-1.11%
Panel B: Market Model		
	(1) EUA	(2) CER
Pre	-2.13%***	-9.59%*
Post	1.29%**	3.05%

Table A.5
Cumulative Abnormal Returns (CAR) before and after the Paris Agreement

The table presents significance test results for estimated CAR for European Emission Allowances (EUA) and Certified Emissions Reductions (CER) using the Sample Quantile (SQ) test Conley and Taber (2011); Gelbach et al. (2013). The significance tests are conducted for the constant mean return model and use a market model with the Global Carbon Market Index (GCMI) as the market benchmark return for robustness. The event window comprises 3 days before (Pre) and after (Post) the Paris Agreement. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, respectively.

	Cumulative Abnormal Return (-3; +3)	
Panel A: Constant Mean Return Model		
	(1) EUA	(2) CER
Pre	6.35%**	0.56%
Post	-4.55%*	-2.26%
Panel B: Market Model		
	(1) EUA	(2) CER
Pre	-1.04%*	-0.04%
Post	3.97%***	-1.55%

Table A.6
Cumulative abnormal returns before and after the Glasgow Conference

The table presents significance test results for estimated CAR for European Emission Allowances (EUA), Global Emissions Offsets (GEO), and Nature-based Global Emissions Offsets (NGEO) using the Sample Quantile (SQ) test Conley and Taber (2011); Gelbach et al. (2013). The significance tests are conducted for the constant mean return model and use a market model with the Global Carbon Market Index (GCMI) as the market benchmark return for robustness. The event window comprises 7 and 5 days before (Pre) and after (Post) the Glasgow Conference. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, respectively.

	Cumulative Abnormal Return (-7; +7)		
Panel A: Constant Mean Return Model			
	(1) EUA	(2) GEO	(3) NGEO
Pre	4.65%**	7.03%**	16.18%***
Post	8.33%***	-1.88%	2.54%
Panel B: Market Model			
	(1) EUA	(2) GEO	(3) NGEO
Pre	-0.89%*	6.38%**	15.27%***
Post	3.04%***	-2.50%	1.66%
	Cumulative Abnormal Return (-5; +5)		
Panel A: Constant Mean Return Model			
	(1) EUA	(2) GEO	(3) NGEO
Pre	5.89%***	1.96%	6.00%**
Post	7.77%***	-1.13%	2.26%
Panel B: Market Model			
	(1) EUA	(2) GEO	(3) NGEO
Pre	-1.07%*	4.85%**	1.14%
Post	1.79%***	1.67%	-1.82%

Table A.7
Cumulative abnormal returns before and after the UAE Conference

The table presents significance test results for estimated CAR for European Emission Allowances (EUA), Global Emissions Offsets (GEO), and Nature-based Global Emissions Offsets (NGEO) using the Sample Quantile (SQ) test Conley and Taber (2011); Gelbach et al. (2013). The significance tests are conducted for the constant mean return model and use a market model with the Global Carbon Market Index (GCMI) as the market benchmark return for robustness. The event window comprises 7 and 5 days before (Pre) and after (Post) the UAE Conference. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, respectively.

	Cumulative Abnormal Return (-7; +7)		
Panel A: Constant Mean Return Model			
	(1) EUA	(2) GEO	(3) NGEO
Pre	-2.48%*	11.76%**	-22.21%***
Post	8.87%***	-15.85%**	-145.01%***
Panel B: Market Model			
	(1) EUA	(2) GEO	(3) NGEO
Pre	0.86%	12.92%**	-21.37%***
Post	0.79%	-18.52%**	-147.06%***
	Cumulative Abnormal Return (-5; +5)		
Panel A: Constant Mean Return Model			
	(1) EUA	(2) GEO	(3) NGEO
Pre	-0.46%	2.59%	-29.54%***
Post	0.39%	-26.67%***	-131.05%***
Panel B: Market Model			
	(1) EUA	(2) GEO	(3) NGEO
Pre	-0.73%	2.53%	-29.57%***
Post	-2.29%**	-27.55%***	-131.58%***

Do investors care about the rainforest? Evidence from voluntary carbon offsets around the world

2.1 Introduction

According to the OECD, *biodiversity finance* is defined as any "expenditure that contributes, or intends to contribute, to the conservation, sustainable use, and restoration of biodiversity" (OECD, 2020). Giglio et al. (2024) provide a theoretical framework where biodiversity loss can have major negative economic consequences. According to recent research, the mean damage costs of "business as usual" action on climate change and biodiversity loss could reach $5.4 trillion/year by 2070 and $31 trillion/year by 2200, which could translate into a 10% (25%) reduction in the global GDP growth rate by 2050 (2070) (Ferguson et al., 2020). While direct estimates of the economic costs of biodiversity loss remain scarce, related ecosystem degradation, such as deforestation and land degradation, has been estimated to cost the global economy $6.3 to $10.6 trillion per year (United Nations Convention to Combat Desertification (UNCCD), 2022; Dasgupta, 2021). Yet, the loss of biodiversity is not only an economic concern. The disappearance of species, the degradation of ecosystems, and the erosion of genetic diversity represent an irreversible loss to humanity's cultural and natural heritage.

The loss of biodiversity extends well beyond mere economic calculations, encompassing intrinsic, ecological, cultural, and ethical dimensions. Biodiversity embodies inherent value independent of human utility, as argued prominently by Wilson (1986) who emphasizes the moral obligation to preserve species diversity for future generations. Additionally, biodiversity underpins the resilience of ecosystems, providing essential services such as pollination, water purification, and disease regulation, which, when disrupted, lead to irreversible consequences for human well-being (Millennium Ecosystem Assessment, 2005). Ecological literature further stresses the interconnectedness and interdependence of species within ecosystems, with the disappearance of one species potentially triggering cascading negative impacts across ecological networks (Hooper et al., 2012; Dirzo et al., 2014). The loss of species disrupts food webs, reduces ecosystem productivity, and diminishes the ability of ecosystems to adapt to environmental changes, such as climate change. For instance, the decline of pollinators, such as bees and butterflies, threatens global food production. Klein et al. (2007) estimate that 75% of global food crops depend on animal pollination, with an economic value of $153 billion annually. Our study is one of the first to try to shed light on the importance of biodiversity loss, voluntary carbon markets, and natural finance with a specific focus on rainforests around the world.

Measuring actions to combat climate change that allow the functioning of voluntary offset carbon markets is straightforward. The focus is typically on reductions in greenhouse gas emissions or sequestration

of carbon. In contrast, measuring actions to combat biodiversity loss is much more complex. Measures of biodiversity involve multiple attributes: species diversity, population dynamics, and ecosystem functionality, which are all inherently challenging to measure and verify (International Union for Conservation of Nature (IUCN), 2015). Many existing pilot projects to combat biodiversity loss require specialized metrics or lengthy field surveys, creating higher costs and longer timelines for issuing voluntary biodiversity credits (Paulson Institute et al., 2020). Several biodiversity-specific metrics must be tracked, including species abundance, habitat condition, and ecosystem resilience (Pollination, 2024). This complexity translates into higher monitoring and verification costs, limiting the standardization needed to spur robust trading in voluntary biodiversity markets (International Finance Corporation (IFC), 2023). Furthermore, biodiversity improvements often require lengthy time horizons to manifest. Landowners and investors may be discouraged by the protracted timelines before biodiversity credits can be verified and monetized (Bahrami et al., 2024).

Biodiversity finance represents an emerging asset class that aims to generate financial returns through investments in conservation, typically requiring blended finance structures due to high risk and uncertain returns (Flammer et al., 2025). The inherent difficulty in monetizing biodiversity underscores challenges in setting credible market signals for biodiversity-linked carbon offsets.

Another barrier to the market expansion of voluntary biodiversity credits is the absence of internationally recognized standards for biodiversity credit verification (Bennett et al., 2017). Several entities, including the IUCN and other conservation-focused organizations, have developed methodologies for biodiversity offsetting. However, these frameworks are not yet universally accepted. Regulatory requirements, such as "no net loss" or mandatory biodiversity offset policies, vary widely across jurisdictions, limiting their predictability and scale (Taskforce on Nature-related Financial Disclosures (TNFD), 2023). While some companies are exploring "nature-positive" supply chains and are willing to invest in biodiversity credits, demand remains modest (Paulson Institute et al., 2020). Investors' motivations in sustainable finance often oscillate between financial value and intrinsic environmental values, a distinction that affects how biodiversity and carbon offset projects are evaluated and valued in financial markets (Starks, 2023). Recognizing these divergent motivations is crucial for interpreting market reactions to biodiversity-linked offsets. With few buyers and inconsistent project supply, the market struggles to establish clear pricing signals or liquidity (Bennett et al., 2017). Bahrami et al. (2024) highlight the stark discounts faced by landowners subject to restrictive biodiversity protections. While this underscores the need to unlock private finance (that is, natural capital finance), it also reflects the lack of a market for biodiversity credits. Without transparent pricing signals or deep market participation, credit suppliers struggle to forecast revenue, and buyers worry about overpaying or inadvertently engaging in "greenwashing". As a result, by

2024, less than 2 million USD worth of genuine biodiversity credits have been sold globally (Pollination, 2024).

An important alternative to biodiversity credits is carbon credits with biodiversity impact as a "co-benefit". Given the challenges of pure biodiversity credits and the fact that so few have been issued, our analysis compares voluntary carbon credits with and without biodiversity conservation certified by the Climate, Community & Biodiversity (CCB) Standard. This standard represents a multiple-benefit certification framework that evaluates both biodiversity and social co-benefits in land-based carbon offset projects. While the standard does not disaggregate biodiversity impact from community outcomes, given its minimum requirements-based framework rather than being a scoring or indexing system, it is explicitly designed to distinguish these co-benefits from pure carbon or climate mitigation outcomes. According to Verra, Biodiversity and community benefits are bundled together since, without local community support, the biodiversity efforts are unlikely to be fully successful. The CCB Standard requires projects to adopt best practices throughout their lifecycle and to demonstrate distinct, measurable objectives in the domains of biodiversity. Importantly, the standard mandates that projects generate a net positive biodiversity impact over their lifetime. This includes maintaining or enhancing high conservation values within the project zone and prohibiting the use of invasive species or genetically modified organisms. Biodiversity, as defined in the CCB documentation, follows the Convention on Biological Diversity of 1992 and encompasses variability within species, between species, and across ecosystems. CCB-certified carbon credits are widely understood to deliver higher biodiversity co-benefits relative to non-certified offsets. As such, they serve as the most viable available proxy for biodiversity impact in the Voluntary Carbon Market (VCM). Given the absence of robust, liquid markets for stand-alone biodiversity credits and the difficulty of directly measuring biodiversity gains, the use of CCB certification as a biodiversity signal is not only defensible but also arguably the strongest indicator presently accessible to investors. We, therefore, treat CCB labeling as a credible marker of biodiversity impact and use it to differentiate high- vs low-certified biodiversity carbon offsets in our empirical analysis.

Corporate sustainability efforts have, to date, focused mainly on greenhouse gas mitigation using forest carbon credits rather than on biodiversity itself (Natural Capital Coalition, 2016).

Our analysis uses data on forestry projects with and without CCB certification. We construct a novel dataset by integrating multiple high-quality sources to examine the stock market response to voluntary carbon credit retirements. The initial source of carbon offset data is the Berkeley Voluntary Registry Offsets Database, which consolidates information from the four major voluntary carbon registries: Verra, Gold Standard, the Climate Action Reserve, and the American Carbon Registry. We verify and augment data from these registries, which collectively cover the vast majority of carbon offset activity in the VCM.

To identify market reactions, we link retirement announcements to publicly listed firms using the ISIN as a unique identifier and manually retrieve historical stock price data from Bloomberg. This allows us to conduct a high-resolution event study centered on the announcement date of credit retirements, which we define as the moment when a carbon credit is officially retired and claimed against a company's emissions footprint, which immediately is reported in the public carbon registry. Additional firm-level data are gathered to support our cross-sectional analysis, including sectoral classification, geographic location, and financial characteristics.

We compare how investors react to information about voluntary carbon offsetting with and without CCB certification to understand the effects of biodiversity benefits. What should we expect in terms of investors' reactions? Given the novelty of this topic, there is no previous literature that studies the reaction of investors to voluntary carbon credit retirements and their biodiversity impact. A study in a similar spirit was conducted about the market reaction to green bond issuances (Flammer, 2021). This shows that there is a positive response to the issue announcements that is especially strong for first-time issuers and third-party verified bonds. Other related work has shown that the stock market responds positively to companies' eco-friendly behavior (Flammer, 2015; Krüger, 2015). Hsu et al. (2023) study how a pollution premium can impact portfolio performance. Heeb et al. (2022) use experimental evidence to show that investors experience positive emotions when choosing sustainable investments. We seek to analyze whether forestry-related carbon credit "retirements" to offset a specific amount of carbon emissions are perceived by shareholders in a similar or different way depending crucially on whether their biodiversity impact meets the standard set by the CCB certification.

Despite being the most viable mechanism to preserve tropical rainforests and to protect and preserve biodiversity in developing countries, tropical rainforest projects and the principle of carbon offsetting itself are constantly under scrutiny from investigative journalists. A much-noticed article by the English newspaper The Guardian (2023) claims that "over 90% of rainforest offsets by biggest providers are worthless", that "of nearly 100 million carbon credits found, only a fraction of them resulted in real emissions reductions", and "companies are making false claims and convincing customers they can fly guilt-free" with "rubber-stamped projects" of "lower environmental integrity". These claims are based on the results of two studies by West et al. (2020) and West et al. (2023), arguing that the majority of those projects are failing to reduce deforestation rates.

We find that biodiversity impact plays an important role in shaping investor reactions to carbon offset retirements. Companies retiring carbon credits certified under the CCB Standard experience neutral or muted stock market responses. In contrast, retirements of standard forestry-related credits without CCB certification are associated with significantly negative abnormal returns, particularly after the ar-

ticle published by The Guardian in January 2023 questioned the integrity of rainforest offsets. These findings suggest that investors differentiate between offsets based on their perceived environmental credibility, with biodiversity-linked certifications like CCB helping to mitigate (but not eliminate) concerns over greenwashing and enhance investor trust. We believe that CCB-labeled carbon offsets are a way for buyers to contribute to biodiversity impact while at the same time mitigating emissions on their net-zero path.

The remainder of the paper is organized as follows. In Section 2, we present facts about the institutional background, in particular, specific details about the impact global deforestation and degradation of the rainforest have on climate change and biodiversity loss. Section 3 describes the data used for our empirical analyses, presents descriptive statistics, and details our empirical approaches. Section 4 presents the empirical results, including a test for the robustness of our findings. Finally, Section 5 provides a conclusion of our findings and offers some policy implications.

2.2 Institutional background

2.2.1

Carbon Markets and Biodiversity Carbon units are traded in two primary markets: the *compliance* market and the *voluntary* carbon market (VCM). The compliance market operates under regulatory frameworks, where entities are legally obliged to meet emissions reduction targets and can buy and sell *carbon allowances* on Emissions Trading Systems (ETS) to comply with these obligations. On the other hand, the VCM enables organizations and individuals to voluntarily offset their emissions beyond regulatory requirements (or in the absence thereof), demonstrating their commitment to sustainability.

The primary difference between compliance and voluntary carbon markets lies in their design, operation, and the price discovery mechanisms. VCMs do not set an emissions cap but allow participants to offset, i.e., reduce, remove, or avoid carbon emissions through specific projects like, for instance, avoiding deforestation of the tropical rainforest.

Carbon offsets are generated by project developers according to the standards of carbon registries, the most important ones being the Verified Carbon Standard (VCS), Gold Standard (GS), American Carbon Registry (ACR), and Climate Action Reserve (CAR). Specific methodologies under these standards allow corporations to offset carbon emissions through nature-based projects in agriculture, forestry, and land use (AFOLU), added with co-benefit labels like the CCB. One of the key institutional developments of voluntary VCM is the approval of seven carbon crediting methodologies by the Integrity Council for the Voluntary Carbon Market (ICVCM) in 2024 [14]. Experimental evidence indicates that consumers place

[14] https://icvcm.org/.

value on environmental integrity and transparency, suggesting that willingness to support biodiversity-linked offsets may increase significantly when their broader environmental impacts are clearly communicated (Rodemeier, 2025). This is particularly relevant in light of research suggesting that firms may strategically use carbon offsets not only to achieve their carbon-transition goals but also to enhance their ESG score following negative rating shocks (Kim et al., 2024) raising concerns for 'transition-washing' where firms claim emissions reductions without significant direct abatement.

The value of voluntary carbon credits depends on various factors, including project type, location, vintage, and so-called *co-benefits* like biodiversity impact, gender involvement, or local community benefits. This distinct market structure ensures that voluntary offsets are not directly comparable to traditional assets like equities, cash, or bonds [15]. Price discovery in the voluntary market is driven by project characteristics. However, Aldy and Halem (2024) point out that a lack of carbon price convergence and potential information asymmetries may contribute to limited price discovery and lead to market segmentation and fragmentation.

It is important to note that carbon offset prices, including those with biodiversity co-benefits, are not standardized or easily observable due to the nature of the VCM, which is primarily an over-the-counter (OTC) market. As a result, reliable, transparent price data is difficult to obtain. While price differentiation is indeed a key indicator of perceived value, particularly in assessing whether buyers are willing to pay more for biodiversity-linked offsets, the fragmented and opaque structure of the VCM limits price visibility. Nonetheless, aggregate data from Ecosystem Marketplace provides indicative insights. In 2021, credits with co-benefits were priced at \$5.02/tCO_2e, compared to \$3.37/tCO_2e for offsets without co-benefits, on average. In 2022, this price gap widened, with credits with co-benefits averaging \$10.51/tCO_2e versus \$6.46/tCO_2e, a premium of over 60%. In 2023, this trend persisted, though average prices declined across the board: credits with co-benefits traded at \$8.11/tCO_2e, while those without averaged \$5.91/tCO_2e. These consistent price differentials suggest that CCB-labeled offsets too are indeed more valued by buyers, possibly also due to their higher perceived environmental integrity and reputational benefits. However, due to the market's OTC nature and limited public reporting, constructing a comprehensive and reliable time series of offset prices remains a future challenge outside the scope of this paper.

2.2.2 Biodiversity Impact from Forest Carbon Credits

One of the biggest drivers of climate change is the deforestation of tropical forests and land degradation for agriculture all over the world. Forests cover about one-third of the global land area, and half of the

[15] Another form of offset is renewable energy certificates, which are not analyzed in this study.

global forests are in only five countries: Brazil, Canada, China, Russia, and the United States (FAO-UNEP, 2020). Besides providing ecosystem services, forests represent a strong carbon sink (Stephens et al., 2007; Federici et al., 2015). Tropical forests alone cover about 15% of the world's land surface and contain about 25% of the carbon in the terrestrial biosphere (Bonan, 2008). Araujo et al. (2023) show that forest degradation in the Amazon Rainforest generates huge negative externalities. Finance for forest-based climate mitigation amounts to only 2.3 billion USD per year, which is estimated to be 1 percent of the necessary total [16]. Deforestation not only contributes to climate change through the release of carbon dioxide but also affects biodiversity loss, which was recently highlighted by the United Nations Biodiversity Conference (COP15) in Montreal (IPBES, 2019). We directly quote from this:

"The biggest driver of biodiversity loss is how people use the land and sea. This includes the conversion of land covers such as forests, wetlands and other natural habitats for agricultural and urban uses. Since 1990, around 420 million hectares of forest have been lost through conversion to other land uses. Agricultural expansion continues to be the main driver of deforestation, forest degradation and forest biodiversity loss."

Thus, deforestation and degradation (of tropical rainforests) are not only one of the biggest single drivers of climate change but also of biodiversity loss.

Forest carbon emissions reduction potential and biodiversity conservation are positively correlated, suggesting that these objectives are strictly intertwined, and they can be pursued simultaneously by rainforest conservation projects (Palomo et al., 2021). Not acting on deforestation, and, thereby, on climate change and biodiversity loss, might have large economic impacts. Giglio et al. (2024) discuss theoretically how biodiversity loss affects the economy. The mean damage costs of "business as usual" action on climate change could reach $5.4 trillion/year by 2070 and $31 trillion/year by 2200, which translates into a 10% (25%) reduction in global GDP growth rate by 2050 (2070) (Ferguson et al., 2020). Fenichel et al. (2024) provide a framework for measuring biodiversity loss and point out that a process for measuring and valuing changes in biodiversity must play a key role. The cost of recent losses of ecosystem services has been estimated at USD 4 trillion–USD 20 trillion per year (Costanza et al., 2014). Land degradation is estimated to cost USD 6 trillion–USD 11 trillion per year (Economics of Land Degradation Initiative, 2015). Alroy (2017), Giam (2017), and Wolf and Ripple (2022) suggest that biodiversity losses will cause a mass extinction event if tropical deforestation continues unabated, implying that the costs of biodiversity loss might rapidly escalate. Garel et al. (2024) argue that investors have already started to require a corporate *biodiversity risk premium*.

Overall, "restoration of trees remains among the most effective strategies for climate change mitigation"

[16] http://www.forestdeclaration.org/resources/forest-finance-theme-3-assessment.

(Bastin et al., 2019) and biodiversity protection. But despite the global importance of deforestation for climate change and biodiversity loss, there is not enough action to halt it. The VCM, in many opinions, including our own, is so far the only viable solution. Streck (2012) discusses these and related issues. The Forestry and Land Use segment of the VCM had a transaction value of $351.3 million in 2023. Early 2025 data suggest that carbon credit prices have stabilized[17]. Biodiversity credit markets on the other hand are only in early stages even though some people project they will grow in the coming decades [18]. However, actual transactions in biodiversity credits have been quite limited. A 2024 report by Pollination finds that a mere $325,000 to $1.85 million worth of voluntary biodiversity credits had been sold globally, highlighting the nascent state of the market [19]. However, a more recent and comprehensive research by Gradeckas (2025) estimates biodiversity credit sales to be closer to a still tiny $5.9 million globally, combining both confirmed and estimated transactions. This includes approximately $4 million in public business-to-business transactions, with additional volumes in consumer-facing and private deals.

So while demand-side use-cases for biodiversity credits might exist, according to Boston Consulting Group, the short- to medium-term demand will be limited since by definition, corporates cannot make claims against their biodiversity footprint, while they can offset their carbon footprint with carbon credits that carry a biodiversity impact as a co-benefit [20]. Thus, we can conclude that at the moment, forest carbon credits with biodiversity co-benefits are the only viable tool for those who aim to create a positive biodiversity impact, which has to be additional to the carbon benefits. Biodiversity benefits are conditional on endorsement and intertwined with the involvement of the local community in a CCB project; without local community endorsement, it is unlikely the biodiversity benefits will be fully realised. In more detail, a CCB project must embed distinct objectives for people and nature and then prove it delivers a net-positive balance on both fronts. The core criteria make this explicit (VCS Association, 2017): Criterion CM2 demands measurable improvements in the well-being of all community groups together with the protection of any local areas important to them, while criterion B2 requires that biodiversity in the project zone ends up better off than in the without-project scenario. Because CM2 ties community well-being to the maintenance or enhancement of the same attributes that B2 uses to gauge ecological success, a project cannot pass the biodiversity test unless it simultaneously secures positive outcomes for the people who rely on those resources. All of these outcomes must be additional to what would have happened anyway: developers must demonstrate that neither the community nor the biodiversity benefits

[17] https://www.spglobal.com/commodity-insights/en/news-research/latest-news/energy-transition/122324-commodities-2025-nature-based-avoidance-prices-to-remain-steady-despite-expected-increase-in-demand

[18] Projections by the World Economic Forum (WEF) suggest that the market could reach $2 billion by 2030 and between $6 billion and $69 billion by 2050, depending on various development scenarios.
https://initiatives.weforum.org/financing-for-nature/insights

[19] https://pollinationgroup.com/global-perspectives/state-of-voluntary-biodiversity-credit-markets/

[20] https://www.bcg.com/publications/2024/the-reality-of-biodiversity-credits

were legally required or financially viable without the project. Consequently, the CCB label is perceived by the market as an additional biodiversity certification beyond the sole carbon benefits.

2.3 Data and Research Design

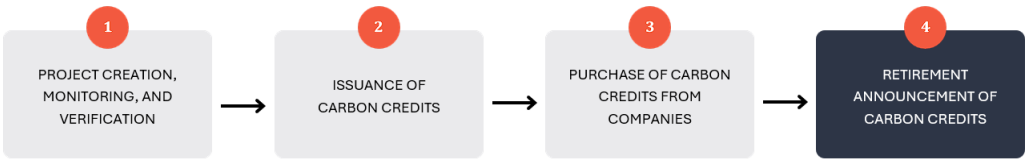
In this section, we describe the data collection process and provide a detailed description of the dataset we employ. We also outline our empirical strategy to address our research questions.

2.3.1 Data and Events

This study uses data collected from multiple sources. The starting point is the Berkeley Voluntary Registry Offsets Database, which includes all carbon offset projects listed globally by the four major voluntary offset project registries: the ACR, the CAR, GS, and VCS. These registries account for the vast majority of voluntary market offsets worldwide and include projects eligible for the California/Quebec linked cap-and-trade programs, as well as the United Nations Clean Development Mechanism projects that have transitioned into one of these voluntary registries (So et al., 2023). These registries offer a clear summary of the lifecycle of voluntary carbon credits, from the creation of the project to the retirement of the offsets. We augment the Berkeley data and validate all the information directly from these registries. From our final dataset, it is possible to understand who the project developer is, the amount of carbon credits issued, the number of carbon credits retired, and the two key variables in our analysis, i.e., the retirement date and the retiring company. The retirement date is the moment when a carbon credit is finally used by its holder to offset one ton of emissions to lower its carbon footprint. Until that date, a carbon credit is an asset that can still be sold to another entity. Figure 1 depicts the typical "lifecycle" of a carbon credit.

Figure 1
The lifecycle of a carbon credit

This illustrates the stages in the lifecycle of a voluntary carbon credit, from project registration and credit issuance to eventual retirement. The process includes project verification, credit issuance by carbon registries, transactions in the voluntary market, and final retirement for emissions offsetting.



Across the major registries, a retirement becomes publicly observable essentially at the moment the retirement transaction is finalized and recorded: in Gold Standard's Impact Registry, completing "Initiate

Retire" generates the retirement certificate instantly and the public retirement entry is viewable right away (Marketplace purchases trigger a follow-up certificate email with links to the public retirement within minutes), indicating minutes-level latency at most [21]. In Verra (VCS), the retirement workflow creates a retirement certificate that is attached to the notification email at submission and the retirement then appears under the Registry's Public Records section; Verra does not state any batching delay, implying publication upon completion of the retirement action [22]. In the American Carbon Registry (ACR), the email sent at the time of retirement serves as confirmation, and the same retirement is visible to anyone on the public Retired Credits report. ACR likewise does not document any lag beyond the transaction time [23]. In Climate Action Reserve (CAR), the public serial-number search and reports show whether credits are active or retired and include the time and date of retirement, which functions as the publicly visible timestamp [24]. None of the registries publish a formal "SLA" for posting speed, but their user guides and help articles collectively indicate that retirements become visible immediately or within minutes of the transaction being finalized on the registry.

In our dataset, we focus only on public companies that have retired forestry-related carbon credits (ARR Afforestation, Reforestation & Revegetation, IFM Improved Forest Management, REDD+ Reducing Emissions from Deforestation and Forest Degradation plus the Role of Conservation, Sustainable Development of Forests, and Enhancement of Forest Carbon Stocks in Developing Countries). Table 1 provides an overview of the sample firms, split into three mutually exclusive groups: firms that retire both CCB-certified and non-CCB credits (Both), only CCB credits, and only non-CCB credits. For each group, the table reports the number of firms and events [25], the scale of offsetting activity (total and average tons of (CO_2 tons) retired, plus medians), and basic firm characteristics: total emissions, total assets, revenues, carbon-intensity, Bloomberg ESG scores, and the Bloomberg ESG-disclosure score. Two patterns stand out. First, although 60% of the 781 listed firms in our universe rely exclusively on non-CCB credits, those firms account for barely 6% of all tons retired; most of the entire volume (92%) comes from companies that use both CCB-certified and non-CCB offsets. Second, firms that use both credit types are on average larger and more carbon-intensive. As metrics of environmental performance and transparency, we compare the average ESG scores and the ESG Disclosure scores to find that they are slightly higher for firms that use both offset types compared to those that retire only CCB offsets, but are the same for the set of firms that only use non-CCB-certified carbon retirements.

The event date in this analysis is the announcement date of the retirement. First, when a company re-

[21] https://goldstandardhelp.freshdesk.com/support/solutions/articles/44002545138-how-to-retire-credits
[22] https://vcmintegrity.org/wp-content/uploads/2023/09/Verra-Registry-User-Guide.pdf
[23] https://acrcarbon.org/wp-content/uploads/2024/08/ACR-Registry-Operating-Procedures_Nov-25-2024.pdf
[24] https://www.climateactionreserve.org/wp-content/uploads/2022/05/Reserve-User-Guide_May2022.pdf
[25] The number of events is different from the analysis because we exclude multiple retirements in a 2-day window.

Table 1
Summary statistics by CCB category

This table presents summary statistics for the 781 companies in the sample, classified according to their use of carbon credits (in metric tons). Firms are grouped into three mutually exclusive categories: those that retire both CCB and non-CCB credits, those that retire only CCB credits, and those that retire only non-CCB credits.
*The number of events is different from the analysis because we exclude multiple retirements in a 2-day window.

	Both	Only CCB	Only non-CCB	Total
Number of companies	173	136	472	781
Percentage of companies	22%	18%	60%	100%
One-time only users	0	93	256	349
Number of retirements	4,473	308	1,480	6,261
Number of retirements (clean*)	1,730	294	1,048	3,072
Average credits retired (CO_2 tons)	23,924	7,206	4,918	10,216
Median credits retired (CO_2 tons)	1,619	240	506	1040
Total credits retired (CO_2 tons)	107,015,989	2,219,721	7,278,759	116,514,469
Total credits retired (%)	92%	2%	6%	100%
Average emissions (CO_2 tons)	6,749,759	1,902,483	3,659,763	4,081,841
Median emissions (CO_2 tons)	315,721	46,859	110,269	119,990
Average total assets (millions)	171,942	68,417	98,588	113,087
Median total assets (millions)	24,582	5,377	12,632	14,686
Average total revenue (millions)	42,014	11,165	23,153	26,157
Median total revenue (millions)	13,908	2,922	6,613	7,157
Average carbon intensity	111.56	134.16	137.64	130.91
Median carbon intensity	12.23	5.93	7.52	7.31
Average ESG score	44.43	42.27	39.94	41.36
Median ESG score	46.00	40.65	40.00	41.20
Average ESG disclosure score	0.59	0.55	0.55	0.56
Median ESG disclosure score	0.60	0.55	0.56	0.57

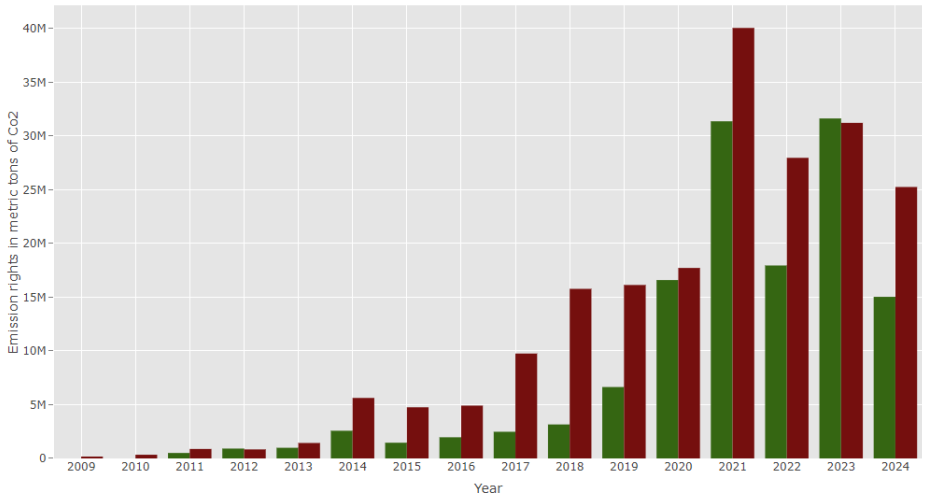
tires carbon credits on the registry, they are permanently removed from the market and used to offset a company's carbon emissions. This is a critical event that distinguishes the mere purchasing of credits from actual environmental action, as the company can claim the offset against its carbon footprint. Credit retirements, particularly in highly scrutinized markets such as the market for REDD+, could have significant effects on stock prices. Retiring high-quality credits in a transparent and timely manner might positively impact a company's stock by signaling authentic commitments to sustainability. By focusing on carbon credits from forestry-related projects only, we can also study how investors perceive the value of high vs. low biodiversity certified co-benefits. The first panel of Figure 2 shows the yearly volume of CCB-related carbon credits and generally forestry-related carbon credits. The second panel of the figure shows the same volume for publicly listed companies.

Table 2 presents absolute numbers for companies by sector and region for the full dataset (Panel A) and for the dataset of companies by sector and region that retired CCB carbon credits in our sample period (Panel B).

To study how investors perceive the value of high vs. low biodiversity impact co-benefits, we investigate firms that retire CCB-labeled carbon credits. The CCB Standard, developed by the largest GHG stan-

Figure 2
Forestry carbon credits universe, overall and public firms

The figure (top) shows the yearly volume of CCB-related credits retired (green bars) and general forestry-related credits (red). The figure (bottom) shows the yearly volume of carbon credits retired by publicly listed firms of CCB-related credits (green bars) and general forestry-related credits (red).



dard setter *Verra*, provides a certification framework for carbon offset projects that demonstrate not only climate benefits but also positive social and biodiversity impacts. These projects include forest conservation, reforestation, and ecosystem restoration efforts that contribute to biodiversity conservation while addressing carbon emissions. The CCB label is distinct in that it emphasizes the broader co-benefits of carbon offset projects beyond simple carbon sequestration. While traditional carbon credits focus primarily on mitigating greenhouse gas emissions, CCB-certified projects integrate environmental and social safeguards, ensuring that local communities and biodiversity also benefit from the investment. This distinction may influence investor perceptions to differ from standard forestry-related carbon credits, as companies that engage in CCB-labeled projects may be viewed as more committed to not just carbon benefits but biodiversity impact and local community goals specifically. We analyze whether firms retiring CCB-certified credits experience different market reactions compared to those retiring standard forestry-related carbon credits. By partitioning our dataset, we examine if raising investor concerns over carbon offset integrity might be mitigated by certified biodiversity impact and whether companies that retire CCB credits receive more positive market responses than those relying on non-CCB-certified offsets.

We identified 781 publicly listed companies that collectively announced 3,072 retirements of forestry-related carbon credits, representing 116 million tons of CO2 over the period 2009–2024. Stock price data were gathered manually using the International Securities Identification Number (ISIN) as the primary identifier for each company's security. To obtain historical and current stock prices, we utilized Bloomberg as our data source. We then categorized the dataset as explained before. First, we examined

Table 2
Carbon credit retirements (totals and averages) by sector, region, and period

This table displays the total quantity of carbon credits retired (in metric tons) and the average quantity retired per event, both overall and broken down by "Before Guardian" vs. "After Guardian." Panel A includes all firms in the dataset, while Panel B focuses on companies retiring Climate, Community & Biodiversity (CCB) carbon credits.

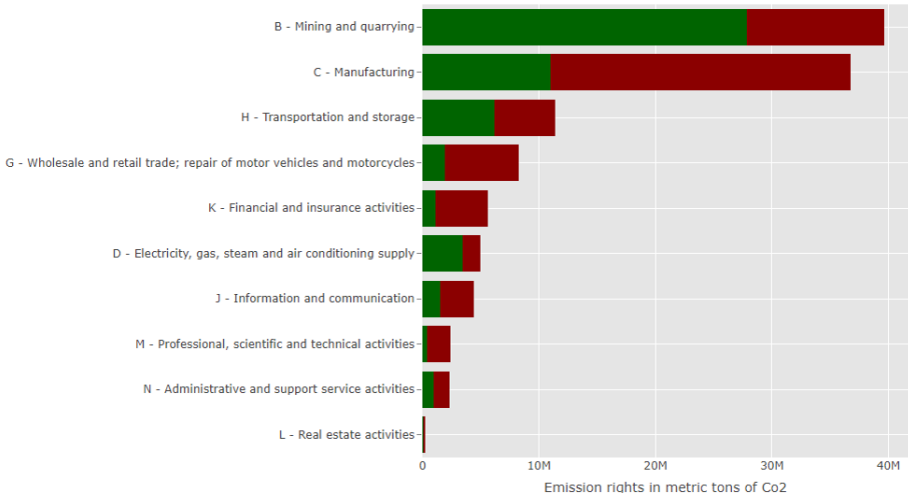
	Overall		Before Guardian		After Guardian	
	Total	Average	Total	Average	Total	Average
PANEL A: Full Dataset						
All	116,514,469	18,609	73,193,635	20,682	43,313,426	15,947
Sectors						
Mining	39,637,650	61,837	16,514,085	55,790	23,123,565	67,024
Manufacturing	36,774,782	17,808	25,952,237	21,699	10,815,387	12,517
Motor vehicles	8,297,468	15,804	7,899,063	19,360	398,405	3,405
Transportation	11,387,417	19,838	10,344,402	24,866	1,043,015	6,601
Financials	5,644,104	6,391	3,725,349	8,169	1,918,755	4,493
Areas						
Asia	12,252,156	20,591	8,030,489	34,172	4,221,665	11,759
North America	29,268,074	14,744	22,780,205	17,755	6,487,619	9,254
Oceania	2,488,342	6,252	1,204,384	6,581	1,283,958	5,971
South and Central America	4,255,851	6,897	3,091,014	7,307	1,164,837	6,004
Western Europe	67,528,759	25,559	37,422,132	26,730	30,099,471	24,312
PANEL B: Companies Retiring CCB Carbon Credits						
All	54,521,266	27,030	33,573,698	30,246	20,947,568	23,095
Sectors						
Mining	27,856,270	76,951	13,412,174	71,341	14,444,096	83,012
Manufacturing	10,994,362	17,534	7,325,039	20,068	3,669,323	14,005
Motor Vehicles	1,889,591	19,683	1,847,963	23,999	41,628	2,190
Transportation	6,161,153	21,027	5,989,553	27,601	171,600	2,257
Financials	1,153,993	4,543	827,120	8,030	326,873	2,164
Areas						
Asia	7,326,107	36,630	4,640,836	48,342	2,685,271	25,819
North America	12,058,700	21,044	9,778,269	25,732	2,280,431	11,815
Oceania	1,457,035	6,070	811,355	9,658	645,680	4,138
South and Central America	582,484	5,200	284,278	3,841	298,206	7,847
Western Europe	32,578,298	37,020	17,550,750	37,421	15,027,548	36,563

companies that retired CCB-related carbon credits, identifying a total of 325 companies and 1,051 retirements, representing 54 million tons of CO2. Figure 3 provides an overview by industrial sector (top part) and region (bottom part). Figure 4 gives an overview of the 30 companies that retired the most carbon offsets between 2009 and 2024.

Table 3 presents the number of events and companies by sector and region for the full dataset (Panel A) and for CCB-related retirements in our sample periods (Panel B).

Figure 3
Sectoral and regional distribution of carbon credit retirements

Breakdown of companies retiring CCB-related credits (green bars) and general forestry-related credits (red) by industrial sector, categorized according to NACE classifications (top) and geographic distribution of corporate carbon credit retirements (bottom).



2.3.2 The Guardian Article Shock in January 2023 and the Impact on the VCM

The publication of an article in The Guardian on 18th January 2023, which purports that over 90% of REDD+ rainforest carbon offsets certified by Verra were "worthless", triggered significant shifts in both corporate behavior and public opinion regarding carbon credits. This collaborative investigation, published jointly with climate activists from SourceMaterial.org and German weekly newspaper Die Zeit, raised doubts about the effectiveness of forest carbon offsets, especially those used by major companies to meet their sustainability and net-zero targets.

Unlike earlier scandals or negative news coverage, VCM constituents, registries, project developers, and industry associations issued statements or announced reviews within days of the publication of the article. Within hours, Verra issued a statement, denouncing the article's conclusions as sensational and incorrect[26], and on 31st January it issued a technical review[27]. In addition, it announced it will retire and replace its entire REDD methodology set by mid-2025, explicitly referencing credibility concerns raised by the Guardian investigation [28]. On the same day as the Guardian publication, the 18th of January, the International Emissions Trading Association (IETA) argued that the article offers an unbalanced view of forest-carbon projects, drawing on research that is not yet peer-reviewed and ignoring experts who support the existing REDD+ methodology[29]. On the 20th of January, South Pole (one of the largest project developers) posted a blog acknowledging the investigation and pledging to review all REDD+ projects

[26] https://verra.org/verra-response-guardian-rainforest-carbon-offsets/
[27] https://verra.org/technical-review-of-west-et-al-2020-and-2023-guizar-coutino-2022-and-coverage-in-britains-guardian/
[28] https://www.theguardian.com/environment/2023/mar/10/biggest-carbon-credit-certifier-replace-rainforest-offsets-scheme-verra-aoe
[29] https://www.ieta.org/the-importance-of-representing-all-voices-when-it-comes-to-the-climate/

Table 3
Firm selection and sample composition by sector and region

This table shows the number of companies and events analyzed in each industry sector and geographic region for each of the three samples considered. The table also disaggregates data into pre- and post-Guardian article periods. Panel A shows the full dataset; Panel B only includes companies retiring CCB carbon credits.

	Overall		Before Guardian		After Guardian	
	Companies	Events	Companies	Events	Companies	Events
PANEL A: Full Dataset						
All	781	3,072	530	1,838	504	1,230
Sectors						
Mining	37	231	22	149	25	82
Manufacturing	244	1,023	170	611	154	409
Motor Vehicles	55	253	31	189	40	64
Transportation	36	263	22	172	24	91
Financials	150	460	97	240	105	220
Areas						
Asia	104	288	57	189	40	64
North America	209	932	136	583	141	348
Oceania	48	209	37	101	30	109
South and Central America	65	348	43	228	49	120
Western Europe	348	1,278	252	796	209	480
PANEL B: Companies Retiring CCB Carbon Credits						
All	325	1,051	209	626	201	425
Sectors						
Mining	21	147	14	107	12	40
Manufacturing	103	306	68	180	61	126
Motor Vehicles	16	53	13	42	7	11
Transportation	18	160	11	105	12	55
Financials	64	148	37	68	44	80
Areas						
Asia	46	129	31	63	24	66
North America	80	251	53	160	54	91
Oceania	32	146	20	59	22	87
South and Central America	18	78	14	52	11	26
Western Europe	145	439	89	287	87	152

and publish the findings transparently [30]. Gold Standard in July 2024, updated its news page about the argument to say it would file an official complaint, accusing the article of relying on cherry-picked secondary evidence and inflammatory language [31]. Although criticisms of individual projects were not new, the speed and synchrony of the ensuing reactions from every tier of the VCM architecture were unprecedented and mark the event as a market-wide shock.

As a result, the voluntary carbon market, previously viewed as an essential tool for mitigating corporate carbon footprints, faced widespread criticism and skepticism. This could have prompted companies to

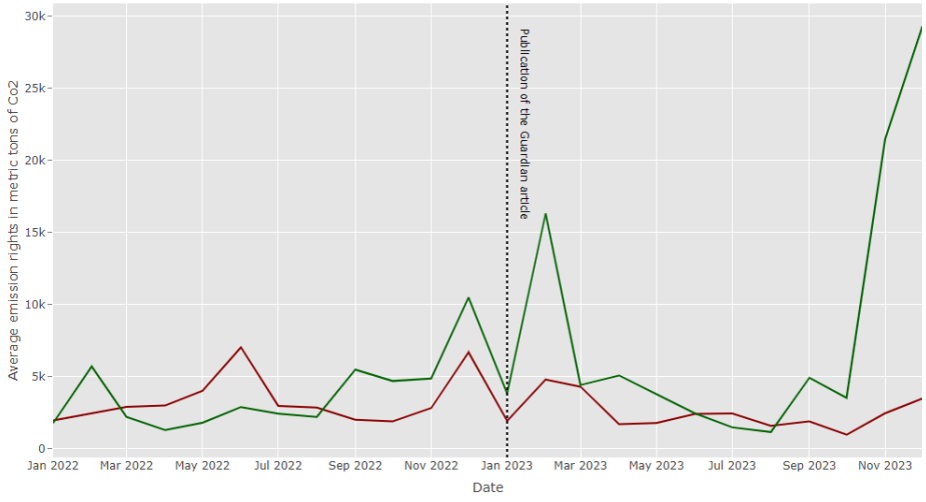
[30] https://www.southpole.com/blog/south-pole-reaction-to-criticisms-of-redd-and-the-vcm
[31] https://www.goldstandard.org/news/response-to-guardian-article-about-the-effectivene

pause or slow down their purchasing of carbon credits ("Greenhushing"). Some companies moved to retire carbon credits associated with rainforest projects under scrutiny, seeking alternative ways to address their environmental impacts. There was also a growing shift away from carbon or climate neutrality and net-zero claims based on offsets to strategies focused on direct emissions reductions and avoidance. This shift could reflect a desire to maintain credibility in the face of mounting scrutiny of carbon offset effectiveness. Public trust in carbon offsets diminished after the revelations, with consumers and environmental groups raising concerns about greenwashing, where companies made false or exaggerated claims about being carbon neutral. Our dataset incorporates an analysis of this important event as a quasi-exogenous shock on the VCM, enabling us to partition our sample into periods before and after the publication of the Guardian article. This will allow us to conduct an analysis on how the markets reacted to the article, providing insights into potential shifts in behavior and market dynamics during these two distinct time periods. Furthermore, it will allow us to draw conclusions about the perception by investors of carbon credits that have high vs. low biodiversity impact as certified by the CCB label, enabling us to analyze the value of biodiversity co-benefits when there is a shock to public perception of the carbon market.

Figure 4 shows the monthly average carbon credit retirements from January 2022 to December 2023. A noticeable spike in CCB-labeled credit retirements follows the Guardian publication.

Figure 4
Carbon credit retirements before and after the Guardian article

This figure shows the monthly average carbon credit retirements from January 2022 to December 2023. The red line represents general forestry carbon credits, while the green line represents higher-integrity, CCB-certified forestry carbon credits. The vertical dotted line marks the publication of a Guardian article in January 2023. A noticeable spike in CCB-labeled credit retirements follows the article, suggesting a shift toward more credible offsetting practices in response to media attention.



2.3.3 Hypothesis Development

We hypothesize that firms retiring forestry-related carbon credits will experience muted or negative abnormal stock returns due to increasing investor scrutiny regarding the credibility of carbon offsets. However, companies retiring CCB-labeled credits may observe more neutral or positive returns, as these credits carry additional biodiversity impact (and social) co-benefits, potentially mitigating concerns over greenwashing. Furthermore, the impact of the Guardian's 2023 article is expected to amplify negative investor reactions.

2.3.4 Empirical Methodology

We employ standard event study methodologies (Campbell et al., 1998; MacKinlay, 1997) to investigate how investors react to purchases of voluntary carbon credits related to rainforest projects around the globe. For the event study, we make use of two standard models to estimate our counterfactual. Abnormal returns are calculated using the *Fama-French Three-Factor Model* [32]. In addition, we test different event windows and compositions over the period of the announcements; in particular, [-1:+1], [-1:+5], [-1:+10] days (and more event windows in the online appendix). The estimation window is always one year before the event window.

Let the daily simple return be r_t, defined as $r_t = P_t/P_{t-1} - 1$, the abnormal returns (ARs) for each model are calculated using the following regressions:

$$\textit{Fama-French Three-Factor Model: } AR_t = r_t - \alpha - \beta_1 r_{m,t} - \beta_2 r_{\text{SMB,t}} - \beta_3 r_{\text{HML,t}}$$

where the coefficients $\alpha, \beta_1, \beta_2, \beta_3$ are obtained evaluating the model over the estimation window. In this case, the returns for the small-minus-big and high-minus-low portfolios were obtained from the online data library Fama and French (2023). The cumulative abnormal returns, defined as CAR for each company, are the sum of the ARs over the event window. The cumulative average abnormal return, defined as CAAR, is the average of a company's CARs. We employed two tests to study the abnormal returns obtained: a non-parametric Wilcoxon signed-rank test and a standard CAAR t-test for robustness. We also apply the non-parametric Corrado (1989) test and the Boehmer et al. (1991) test for event-induced variance, but don't report them as results are the same.

2.4 Results

Our study evaluates the Cumulative Average Abnormal Returns (CAARs) across different event windows by summing the daily abnormal stock returns for each company over the event window and then

[32] For robustness, we employ the *Fama-French Four-Factor Model* and the *Fama-French Three-Factor Model* with regional factors. The results are consistent with the main analysis, and the tables are available in the online appendix.

computing the mean across the sample. First, we test for the impact after the event, defining the event windows from -1 to +1,+3,+5, and +10 days. Finally, we also test for the effects of the days before the event with event windows [-3:+3], [-5:+5], and [-10:+10] (Not all the results are shown in the tables, but all the windows are available in the Online Appendix). The dataset is partitioned into companies from five different industry sectors (the most active ones in the market) according to their NACE classifications (as described in the previous section) and five geographical areas to find effects specific to the industry and geographical area. The sectors are *Mining and Quarrying, Manufacturing, Wholesale and Retail Trade, Motor Vehicles and Motorcycles, Transportation and Storage, and Financial and Insurance Activities*. The geographical areas are *Asia, North America, Oceania, South and Central America, and Western Europe*. Not all the results are shown and discussed in the main text, but only the most significant ones. Special attention is given to firms associated with (CCB) Standards certification, as these categories may exhibit distinct investor responses. The statistical significance of the CAARs is tested with the null hypothesis that they equal zero by applying two different tests: a non-parametric Wilcoxon signed-rank test and a standard CAAR t-test for robustness. For the consistency of our results, abnormal returns are calculated using the Constant Mean Return Model and the Fama-French Three-Factor Model.

2.4.1 Overall Analysis of Abnormal Returns

This section presents the findings for the overall sample and the entire sample period. For the full dataset, we observe that CAARs tend to be negative (but not significantly different from zero) across most event windows. The corresponding exact absolute values, together with their statistical significance, are shown in Table 4.

This result confirms that investors generally react neutrally or even slightly unfavorably to carbon credit retirements, but the overall findings do not indicate any results that are statistically significant from zero at the aggregate level. However, a key difference emerges between firms retiring CCB-certified credits and those retiring non-CCB credits. Firms retiring non-CCB credits experience stronger and statistically significantly negative responses. On the contrary, CCB-certified credit retirements show a neutral market reaction, often with, albeit mostly negative, smaller, and statistically insignificant CAAR values. This suggests that investors differentiate between standard forestry offsets and those with additional biodiversity (and social) co-benefits, even though the differences are only economically, but not statistically significant for most event windows. Figure 5 shows these results in a graph.

The mining sector exhibits the strongest negative reaction, with CAARs significantly declining across most event windows. The overall column shows particularly negative effects, indicating that carbon offset reliance in this high-emission industry is viewed negatively by investors, but the results are not

Table 4
Results for the entire sample

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The counterfactual returns are constructed using the Fama-French model presented in the paper. Panel A presents results for the full dataset, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
-1:+1)	-0.09	0.49	0.24	0.01	0.83	0.95	-0.24*	0.08	0.04
(-1:+5)	-0.23	0.28	0.09	-0.14	0.63	0.42	-0.42	0.12	0.06
(-1:+10)	-0.29	0.22	0.13	0.02	0.84	0.93	-0.78**	0.01	0.02
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	-0.02	0.66	0.89	0.07	0.40	0.67	-0.17	0.30	0.31
(-1:+5)	-0.11	0.75	0.60	0.01	0.83	0.97	-0.38	0.20	0.18
(-1:+10)	-0.06	0.80	0.82	0.16	0.67	0.67	-0.57	0.12	0.14
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.14	0.19	0.14	-0.08	0.44	0.51	-0.25	0.12	0.08
(-1:+5)	-0.28	0.29	0.09	-0.29	0.21	0.14	-0.39	0.38	0.17
(-1:+10)	-0.38	0.21	0.11	-0.12	0.71	0.65	-0.87**	0.04	0.05

statistically significant.

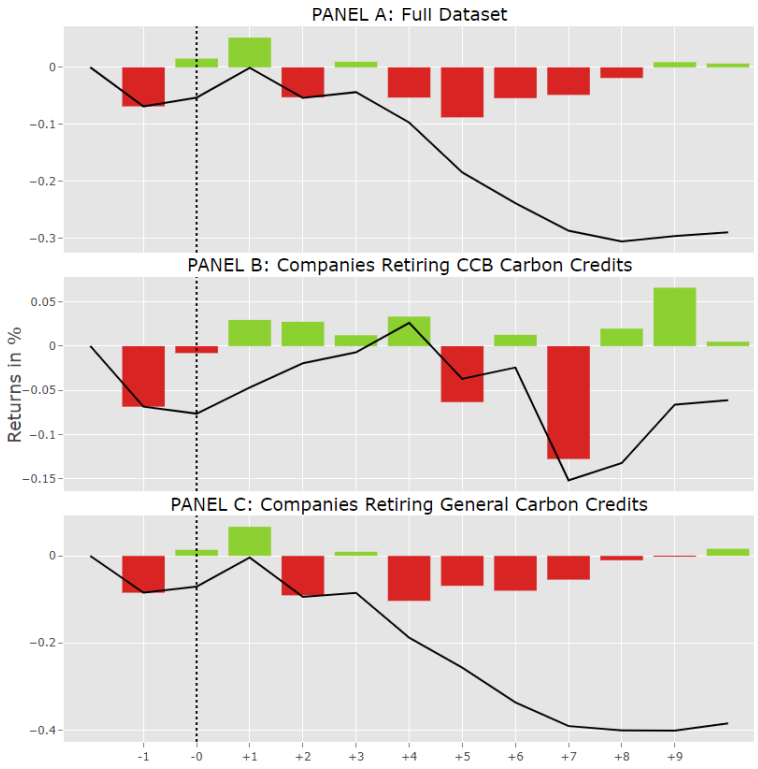
For manufacturing firms, CAARs tend to be negative, with several windows showing statistical significance, reinforcing the notion that investor sentiment has shifted against firms perceived as relying on carbon offsets rather than direct emission reductions. However, CCB-certified credits within this sector tend to buffer some of the negative reactions, as indicated by their less pronounced CAAR declines.

Companies operating in South and Central America (where most of the forestry-carbon related credits are generated) experience some of the largest negative CAARs, with statistical significance across multiple event windows, reflecting heightened investor skepticism in regions more directly linked to forestry-related carbon offsets. This suggests that firms operating in areas associated with rainforest conservation projects face the highest levels of investor scrutiny.

To summarize, the overall market response to forestry-related carbon credit retirements is negative but without strong statistical significance across all event windows. Mining companies face the strongest negative market reaction, while financial firms show more neutral or even slightly positive market responses. South and Central American firms are disproportionately affected, possibly due to their direct association with rainforest-related carbon projects. The difference in CAARs between CCB and non-CCB credits is notable in the results. Overall, firms associated with CCB-certified projects exhibit less negative or neutral CAARs compared to the full dataset, suggesting a less negative investor response. These findings indicate that while market reactions to the retirement of forestry-related carbon credits

Figure 5
Results for the entire sample

This figure shows the abnormal and cumulative abnormal returns for the full dataset (Panel A), companies retiring CCB-labeled carbon credits (Panel B), and companies retiring general carbon credits (Panel C). The sample used in this case is the overall dataset for the overall period. The dotted line represents the event, i.e. the carbon credit retirement.



are weak in general, firms engaged in CCB-certified projects may experience a relatively better or muted investor response, potentially reflecting a more favorable market perception of their sustainability efforts.

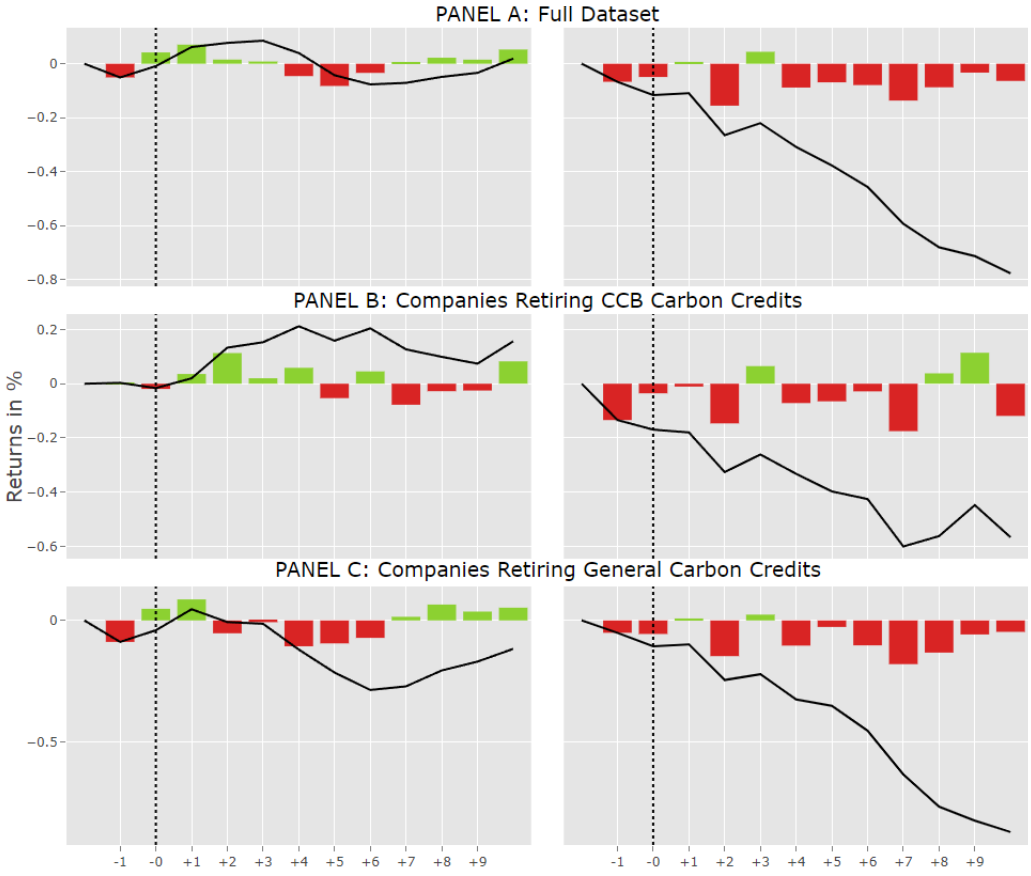
2.4.2 Abnormal Returns Before and After the Guardian Article Shock

This section employs the same event study methodology as before to examine the stock market reaction before and after the publication of the Guardian article on the 18th of January 2023, thereby assessing the potential impact of the controversy on firms engaged in forestry carbon offset projects. By analyzing abnormal returns across multiple event windows, this study evaluates whether investor sentiment shifted significantly in response to allegations regarding the alleged overaccreditation and inefficacy of forest-based carbon offsets. The analysis is conducted at both aggregate and sectoral levels, as well as across different geographical regions, allowing for a comprehensive assessment of market reactions. To ensure consistency and comparability, the same subsample structure as in the overall analysis is applied, facilitating a robust evaluation of how different categories of firms were affected by the controversy. Figure 6 provides a graphical overview of the overall event study results before the article publication (Panel A), as well as the results for companies retiring CCB carbon credits (Panel B) and for companies retiring

general carbon credits (Panel C).

Figure 6
Results before and after the Guardian article

This figure shows the abnormal and cumulative abnormal returns for the full dataset (Panel A), companies retiring CCB-labeled carbon credits (Panel B), and companies retiring general carbon credits (Panel C). In this case, the figure on the left represents the event study before the publication of the Guardian article while the figure on the right represents the event study after. The sample used is the overall dataset. The dotted line represents the event, i.e. the carbon credit retirement.

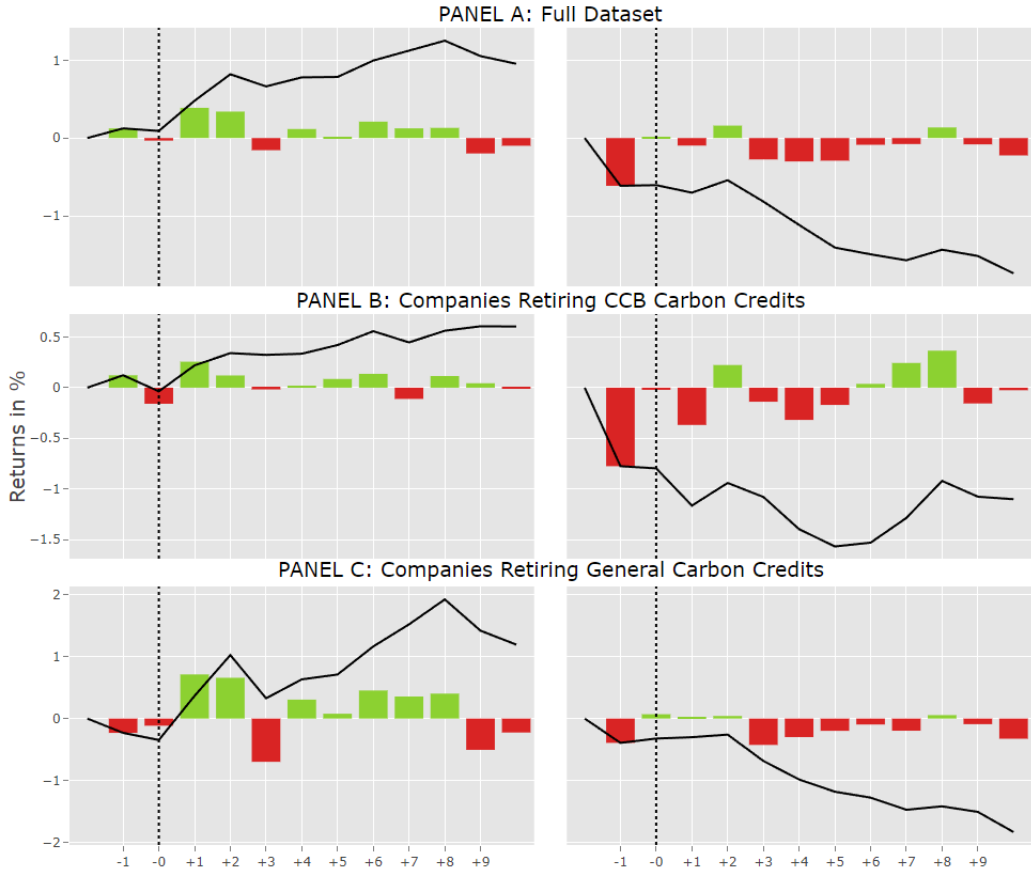


For the full dataset, CAARs tend to be weakly negative across most event windows, with statistical significance in some cases. This confirms that investors generally react unfavorably to forestry carbon credit retirements. However, a clear shift occurs after the Guardian article, with the post-publication period exhibiting significantly stronger negative reactions. A key contrast emerges between firms retiring CCB-labeled credits and those retiring non-CCB credits. In the pre-Guardian period, the reaction to carbon credit retirements is relatively muted for both credit types, with CAARs fluctuating but not consistently significant. However, in the post-Guardian period, non-CCB credit retirements experience a much stronger negative response, indicating a substantial decline in investor confidence. Conversely, CCB-certified credit retirements maintain relatively stable and statistically insignificant CAAR values, suggesting that investors perceive these offsets as more credible and/or more attractive in light of their biodiversity impact. The results for the mining sector (shown in Figure 7) exhibit a strong negative

reaction overall, with CAARs significantly declining in most event windows.

Figure 7
Results before and after the Guardian article, mining sector

This figure shows the abnormal and cumulative abnormal returns for the full dataset (Panel A), companies retiring CCB-labeled carbon credits (Panel B), and companies retiring general carbon credits (Panel C). In this case, the figure on the left represents the event study before the publication of the Guardian article while the figure on the right represents the event study after. The sample used is the mining sector. The dotted line represents the event, i.e. the carbon credit retirement.



In the pre-Guardian period, CAARs are relatively stable, but in the post-Guardian period, they drop sharply in some cases. This suggests that carbon offset reliance in high-emission industries was increasingly viewed negatively by investors. The divergence between CCB and non-CCB credits is also evident, with non-CCB retirements facing the brunt of the market skepticism. The absolute values of the results are depicted in Table 5.

For manufacturing firms, CAARs tend to be negative, with several event windows showing statistical significance. Again, we first present graphical evidence Figure 8) and then the absolute values (Table 6). In the pre-Guardian period, CAAR values are more variable but mainly stable; in the post-Guardian period, the decline becomes more pronounced. However, firms retiring CCB-certified credits experience smaller declines, indicating that investors differentiate between different types of offsets, showing more trust in those carbon credit retirements with additional social and biodiversity co-benefits.

Table 5
Results for the mining sector

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The sample includes only mining and quarrying companies. The counterfactual returns are constructed using the Fama-French model presented in the paper. Panel A presents, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	0.06	0.98	0.81	0.35	0.20	0.25	-0.59**	0.04	0.07
(-1:+5)	0.12	0.55	0.77	0.59	0.65	0.30	-0.89**	0.04	0.11
(-1:+10)	0.10	0.78	0.85	0.95	0.17	0.17	-1.74**	0.01	0.03
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	0.08	0.59	0.79	0.21	0.29	0.56	-0.37	0.27	0.44
(-1:+5)	-0.06	0.86	0.91	0.18	0.69	0.76	-0.87	0.15	0.35
(-1:+10)	0.22	0.69	0.74	0.60	0.31	0.43	-1.10	0.30	0.40
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.31	0.14	0.35	0.25	0.69	0.65	-0.90**	0.01	0.01
(-1:+5)	0.17	0.33	0.78	1.10	0.98	0.31	-0.81	0.11	0.21
(-1:+10)	-0.27	0.20	0.71	1.19	0.57	0.30	-1.83**	0.01	0.04

Table 6
Results for the manufacturing sector

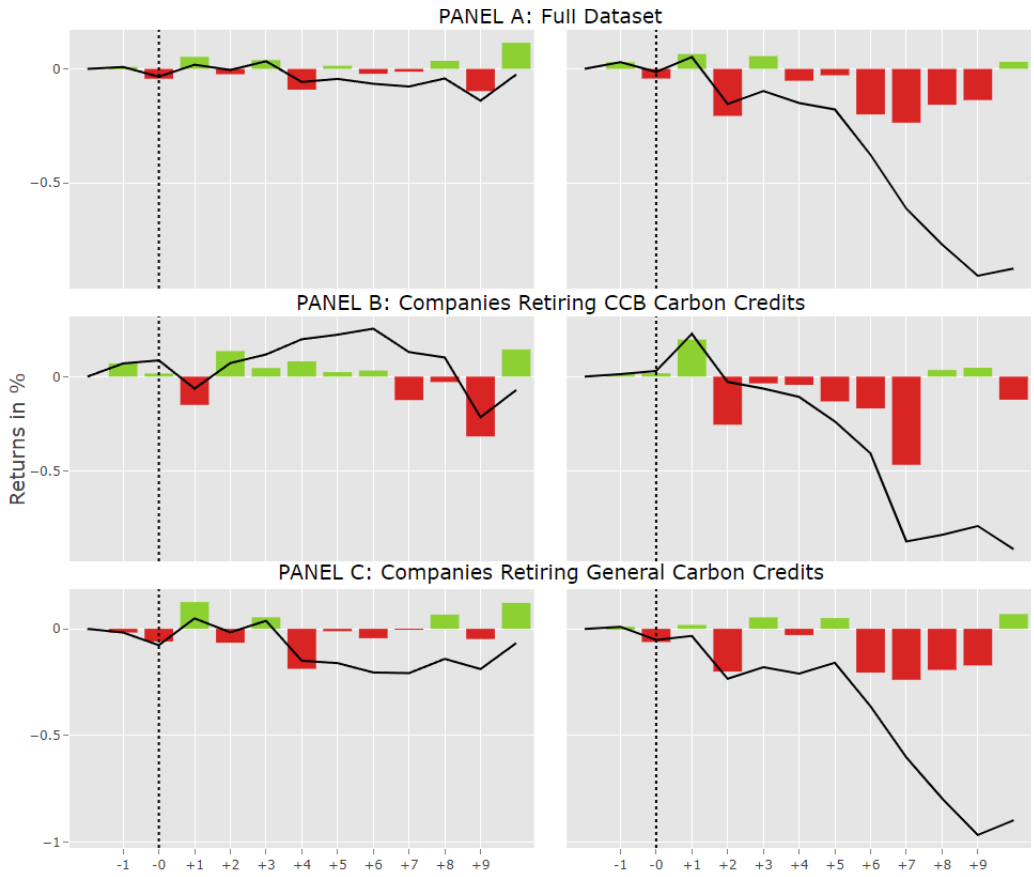
This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The sample includes only manufacturing companies. The counterfactual returns are constructed using the Fama-French model presented in the paper. Panel A presents results for the full dataset, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	-0.04	1.00	0.74	0.02	0.89	0.90	-0.21	0.45	0.17
(-1:+5)	-0.18	0.48	0.27	-0.09	0.92	0.67	-0.58**	0.02	0.03
(-1:+10)	-0.27	0.23	0.24	-0.02	0.95	0.93	-0.88**	0.01	0.01
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	0.09	0.79	0.63	0.20	0.61	0.44	-0.33	0.37	0.19
(-1:+5)	-0.01	0.96	0.96	0.16	0.63	0.69	-0.85*	0.05	0.04
(-1:+10)	-0.17	0.48	0.67	-0.07	0.71	0.90	-0.91	0.17	0.10
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.08	0.78	0.54	-0.07	0.62	0.65	-0.17	0.75	0.38
(-1:+5)	-0.23	0.45	0.23	-0.26	0.61	0.28	-0.47	0.10	0.13
(-1:+10)	-0.29	0.31	0.27	-0.07	0.94	0.84	-0.90**	0.02	0.03

Companies operating in South and Central America experience some of the largest negative CAARs, with statistical significance across multiple event windows, as shown in Table 7.

Figure 8
Results before and after the Guardian article, manufacturing sector

This figure shows the abnormal and cumulative abnormal returns for the full dataset (Panel A), companies retiring CCB-labeled carbon credits (Panel B), and companies retiring general carbon credits (Panel C). In this case, the figure on the left represents the event study before the publication of the Guardian article while the figure on the right represents the event study after. The sample used is the manufacturing sector. The dotted line represents the event, i.e. the carbon credit retirement.



This effect intensifies after the Guardian publication, with CAARs declining in the post-Guardian period. The heightened skepticism in this region could be attributed to its direct association with rainforest-related carbon offset projects, which were at the center of the article's controversy. This suggests that investors are particularly wary of offsets linked to these REDD+ rainforest projects, likely due to concerns about their integrity and effectiveness.

To summarize, the overall market response to forestry-related carbon credit retirements is negative, with statistical significance increasing after the Guardian article, particularly for non-CCB offsets. Firms retiring CCB-certified credits experience more muted market reactions, with close to no statistically significant CAAR declines in most cases. This suggests that investors differentiate between standard forestry carbon offsets and those with additional biodiversity impact and social co-benefits. However, most of the differences in CAARs between CCB-certified and non-CCB credits are only economically,

Table 7
Results for companies from South and Central America

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The sample includes only South and Central American companies. The counterfactual returns are constructed using the Fama-French model presented in the paper. Panel A presents results for the full dataset, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	-0.07	0.85	0.82	0.03	0.94	0.95	-0.23	0.67	0.59
(-1:+5)	-1.02*	0.09	0.07	-0.94	0.22	0.23	-1.20	0.24	0.14
(-1:+10)	-1.57**	0.01	0.03	-0.96	0.30	0.32	-2.65**	0.01	0.01
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	0.75*	0.07	0.20	0.65	0.16	0.34	1.52	0.21	0.18
(-1:+5)	0.79	0.54	0.48	0.80	0.69	0.53	0.90	0.67	0.72
(-1:+10)	1.47	0.30	0.30	1.27	0.44	0.44	2.96	0.40	0.35
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.40	0.17	0.25	-0.37	0.30	0.48	-0.43	0.37	0.33
(-1:+5)	-1.75**	0.01	0.01	-2.04*	0.05	0.04	-1.46	0.16	0.09
(-1:+10)	-2.79***	0.00	0.00	-2.38**	0.04	0.04	-3.32***	0.00	0.00

but not statistically significant at conventional levels [33]. Notably, mining companies face the strongest negative market reaction post-Guardian, with significantly larger declines in CAARs, while financial firms show a decline in previously observed positive responses.

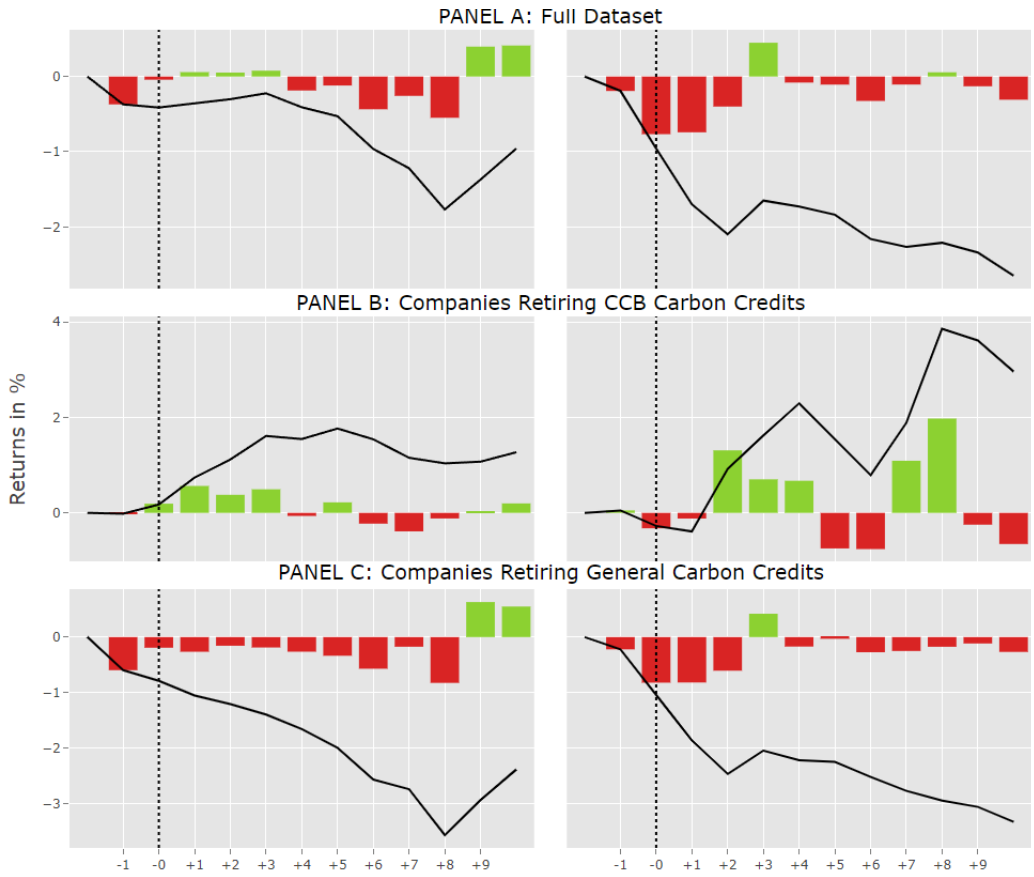
2.4.3 Cross-Sectional Results

In this section, we summarize a regression analysis to examine the determinants of the Cumulative Abnormal Returns associated with firms' announcements of forestry-related carbon credit retirements. Table 8 displays these results. Each column corresponds to a specific event window. The CCB dummy, which equals one if a firm retires carbon credits that carry a CCB label, has a positive and statistically significant coefficient in the two longer windows. This suggests that when companies use offsets with higher biodiversity co-benefits, the stock market may respond less negatively, or even somewhat favorably, consistent with the previous findings. Company size shows minimal explanatory power, as indicated by the small and generally insignificant coefficients. Similarly, the Market-to-book ratio appears to matter only marginally. Overall, these results suggest that CCB-certified credits may mitigate negative market reactions, while larger firm fundamentals such as Size and Market-to-book play a relatively minor role in explaining abnormal returns around these carbon credit retirement announcements.

[33] In our opinion, the lack of statistical significance stems from limited power in a highly heterogeneous sample rather than from a true absence of market discrimination

Figure 9
Results before and after the Guardian article, South and Central America area

This figure shows the abnormal and cumulative abnormal returns of the full dataset (Panel A), companies retiring CCB-labeled carbon credits (Panel B), and companies retiring general carbon credits (Panel C). In this case, the figure on the left represents the event study before the publication of the Guardian article while the figure on the right represents the event study after. The sample used is the South and Central America area. The dotted line represents the event, i.e. the carbon credit retirement.



2.4.4 Economic Relevance

Table 9 provides some insights into the economic magnitude of investor reactions to forestry-related carbon credit retirements, underscoring that even seemingly small abnormal returns can translate into small changes in firm value [34]. These effects are particularly pronounced after the Guardian article, where magnitudes often triple in size. The cross-sectional breakdown reveals that companies retiring general carbon credits face the steepest declines, whereas firms retiring CCB-certified credits experience more muted impacts. This divergence highlights that markets appear to recognize and assign value to the biodiversity and social co-benefits embedded in CCB credits, thereby softening investor skepticism. At the same time, the positive but small pre-Guardian magnitudes suggest that prior to heightened scrutiny, some retirements may have been interpreted as credible commitments to sustainability. However, the shift after the Guardian publication indicates that reputational concerns can quickly reverse market perceptions,

[34] We thank the Editor Marcin Kacperczyk for this important suggestion.

Table 8
Determinants of cumulative abnormal returns for forestry-related carbon credit retirements

This table presents the results of ordinary least squares (OLS) regressions where the dependent variable is the Cumulative Abnormal Return (CAR) for various event windows. The CCB dummy is an indicator variable set to one for retirements involving CCB carbon credits. Standard errors are shown in parentheses; the bottom rows of the table report the number of observations and the adjusted R-squared. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	CARs (-1:+1)	*CARs (-1:+5)*	*CARs (-1:+10)*
Cons	0.05	-1.18	0.55
	(0.44)	(0.73)	(0.98)
CCB dummy	0.12	0.49*	0.70**
	(0.12)	(0.2)	(0.27)
Size	-0.01	0.03	-0.04
	(0.02)	(0.03)	(0.04)
Market-to-book	-0.02	-0.04	0.05
	(0.05)	(0.08)	(0.1)
Obs.	2136		
$Adj.R^2$	≤ 0.01		

amplifying the economic costs for firms relying on offsets perceived as less credible. Considering the median market capitalization of sample firms, these losses translate into material percentage changes in firm value, demonstrating that announcements of carbon credit retirements-particularly those involving general credits-carry tangible financial risks. In short, while the absolute basis-point changes in CAARs might seem significant, their dollar magnitudes are not large enough to be economically meaningful and relevant for investors assessing the credibility and impact of firms' climate strategies.

2.4.5 Robustness Tests

To ensure the reliability of our findings, we conduct a robustness test by removing overlapping events occurring within the same month for the same company. This addresses concerns that closely timed retirements of voluntary carbon credits may introduce distortions in market reactions, either by diluting individual event effects or by creating artificial event clustering that does not reflect true investor sentiment. In event studies, multiple occurrences within a short time frame can lead to event clustering, where investors react not to each specific event but to an aggregate perception of corporate behavior. Suppose carbon credit retirements within the same month are perceived as part of a continuous process rather than isolated incidents. In that case, the stock market response may be muted or spread out, reducing the measured abnormal returns. Filtering out such clustered events allows for more precise identification of how investors react to distinct credit retirements. To isolate independent corporate actions, we systematically removed all events that occurred within the same calendar month for the same company. Finally, we

Table 9
Economic magnitudes

This table reports median cumulative average abnormal returns (CAAR, in basis points) and the corresponding economic magnitude (in USD) across different event windows. Results are shown for the full dataset (Panel A), for companies retiring CCB carbon credits (Panel B), and for companies retiring general carbon credits (Panel C). Columns group results for the overall sample, for the pre-Guardian period, and for the post-Guardian period.

	Overall		Before Guardian		After Guardian	
Windows	CAAR	Econ. magnitude	CAAR	Econ. magnitude	CAAR	Econ. magnitude
PANEL A: Full Dataset						
(-1:+1)	-0.42	-1,662.34	3.67	14,459.63	-5.31	-15,331.49
(-1:+5)	-4.40	-17,364.07	0.00	0.00	-21.99	-63,469.66
(-1:+10)	-9.42	-37,138.54	0.00	0.00	-44.00	-127,018.09
PANEL B: Companies Retiring CCB Carbon Credits						
(-1:+1)	-0.29	-1,115.66	4.58	18,045.32	-7.16	-17,728.76
(-1:+5)	-3.42	-13,172.41	5.66	22,309.49	-22.14	-54,795.45
(-1:+10)	-2.86	-11,025.63	0.21	840.87	-42.76	-105,854.82
PANEL C: Companies Retiring General Carbon Credits						
(-1:+1)	-1.40	-6,247.43	-2.34	-9,192.82	1.23	5,589.41
(-1:+5)	-9.64	-43,010.53	-14.72	-57,747.17	-10.32	-46,912.61
(-1:+10)	-30.79	-137,286.45	-5.96	-23,384.45	-44.56	-202,592.95

collected earnings announcement dates for approximately 600 companies, primarily located in Europe and North America, and excluded retirement events that occurred within the same week as the earnings announcement events. This filtered dataset is then subjected to the same event study methodology as outlined in Section 3.4, employing the Fama-French Three-Factor Model.

In addition, we recompute abnormal returns using region-appropriate factor sets (we use regional factors for Europe, Asia and Oceania, Japan, and the US.) and include Momentum (Carhart). All the tables are available in the Online Appendix.

The robustness tests confirm our main findings while demonstrating that the event clustering initially dampened the observed stock price reaction to carbon credit retirements. By filtering out overlapping events within the same month for the same company, we observe a more pronounced negative impact of retirements, particularly in the post-Guardian period.

CCB-certified carbon credits continue to exhibit a less severe negative stock price reaction compared to standard forestry-related carbon credits, even after controlling for event clustering. This suggests that investors recognize the added value of biodiversity protection and social co-benefits associated with CCB-certified projects, which may mitigate concerns over greenwashing in the carbon-integrity part of the offset. While standard carbon credit retirements face increasing skepticism, especially following the Guardian's critical article, CCB-certified offsets appear to retain stronger credibility, resulting in more neutral investor responses. These findings reinforce the importance of offset quality and biodiversity impact in shaping market reactions and highlight that not all carbon credits are perceived equally by

investors. Companies that retire carbon offsets with verified biodiversity impact and social co-benefits may experience more stable market responses, even as broader concerns over carbon offset integrity intensify. By controlling for event clustering, this robustness test validates our conclusions while demonstrating that investor responses are amplified when sustainability actions are clearly distinguishable and strategically communicated.

2.5 Conclusion

We explore empirically how investors react to firms' biodiversity-focused activities in the controversially debated voluntary carbon market. We develop and analyze a unique dataset, manually assembled from carbon registry data that includes announcements of forestry-carbon offset retirements from 2009 to 2024. Most carbon offsets retired by publicly listed companies are related to rainforest projects around the world, because deforestation and land degradation are arguably the biggest drivers of climate change and biodiversity loss.

Our results suggest that the overall average stock price reaction of carbon offsetting is not significantly different from zero at the time of retirement of forestry carbon credits until the end of 2022. However, following a Guardian article claiming rainforest carbon offsets to be 'worthless' in January 2023, the announcement returns to forestry carbon offset retirements turn significantly negative. This effect is not significant for carbon credits contributing to biodiversity conservation as certified by the CCB Standard, but is significantly negative for other credits. Firms that rely heavily on carbon offsets, particularly those within the mining and manufacturing sectors, face stronger negative market reactions, highlighting concerns about greenwashing. Sectoral and regional variations further suggest that manufacturing and financial firms, as well as companies operating in emerging markets, may be more vulnerable to reputational risks associated with carbon offset retirements. The significant shift in investor sentiment after The Guardian's 2023 article highlights the growing scrutiny of the carbon offset mechanism, attributed to low transparency, concerns about carbon integrity, and regulatory uncertainty.

From a policy perspective, our results underscore the need for enhanced disclosure standards to improve the transparency and credibility of corporate carbon offsetting strategies. Strengthening governance mechanisms in VCMs through independent auditing and stricter certification criteria will be critical in restoring investor confidence, especially to ensure investors receive clear and consistent information on offset quality and biodiversity co-benefits. Further regulatory developments should focus on ensuring market integrity, preventing misleading net-zero claims, and promoting accountability in corporate sustainability disclosures. As investor scrutiny continues to rise, firms that integrate credible, high-quality offsets with certified biodiversity impact will likely benefit from stronger investor trust and reduced rep-

utational risk.

Our results show that companies retiring CCB-certified credits experience less negative investor reactions, implying that carbon offsets with additional biodiversity impact and social co-benefits seem to be viewed more favorably. Thus, investors seem to care about both the biodiversity impact of carbon credits as well as the climate-change mitigation integrity of offsetting activities. Still, our findings raise the puzzling question of why firms voluntarily engage in costly (climate change mitigation and biodiversity impact) activities, which are at best neutral to and potentially harmful to shareholder value.

Appendix B

Economic relevance

Table B1 presents a summary of the economic relevance of carbon credit retirements for the largest offsetting companies in our sample. We value each firm's retired credits using both the maximum, mean, and minimum aggregate forestry project prices provided by another provider, ViridiosAI. These values were then scaled by the companies' total assets and total revenues to evaluate the economic impact of the retirements. The results show that, in nearly all cases, the economic relevance of offsets is fairly limited, typically amounting to less than 1% of assets. Only a few firms exhibit higher relative values, but even these remain modest in scale. However, while the total amount of offsetting is not trivial or negligible, it still supports our interpretation that the financial impact of offset retirements is unlikely to drive stock market reactions on its own, so a signaling effect too is likely driving the results.

First-time versus repeated users of carbon offsets

Across Tables B2 and B3, the evidence paints a consistent picture: once the Guardian article was published, investors began to punish companies that repeatedly retire non-CCB credits. One-off retirements elicit a milder version of this penalty, indicating that market scrutiny intensifies with repeat offset use. Within firms that use credits with and without CCB certification, the initial pre-2023 premium for CCB tranches disappears and even reverses, suggesting that the market's baseline expectation has ratcheted upwards: CCB status now merely shields companies from downside rather than conferring upside. Taken together, the tables reveal a post-2023 flight-to-quality dynamic in which investors do not reward high-biodiversity- impact offsets but decisively penalize those without a CCB label, especially when market-wide suspicion of low carbon integrity of forest offsets forms a discernible pattern.

Covariate Balance

Table B4 presents two complementary cross-sectional analyses. Panel A compares the pure subsample of firms that use only general carbon offsets vs. firms that use only carbon offsets with CCB certification; the Wilcoxon tests reveal that CCB-only firms are smaller and have lower revenues, but environmental traits such as carbon-intensity, ESG score, and ESG disclosure score are not statistically different. Panel B widens the lens to compare firms that have ever retired at least one CCB-certified credit against firms that have never done so; here, all size and emissions variables balance out, and the only residual gap is a modestly higher ESG score for the ever-CCB group (which makes intuitive sense) while ESG disclosure is not statistically different. Because the significant differences in Panel A are confined to scale rather

than greenness, and because they disappear once mixed-type users are admitted (Panel B), we view them as selection effects rather than threats to identification.

Difference-in-means tests

We run difference-in-means tests for the CAARs in all tables. Across the whole sample, the average gap in CAARs between the two carbon offset types (CCB vs non-CCB) is economically small and statistically indistinguishable from zero in (almost) every event window (Table B6). In our opinion, however, the lack of statistical significance stems from limited power in a highly heterogeneous sample rather than from a true absence of market discrimination. On its own, the statistical evidence therefore does not imply that investors reward a premium for CCB offsets. What it does show instead is a penalty/discount for the non-CCB alternative that is statistically different from zero. In line with the consistently negative and statistically significant CAARs we document for non-CCB retirements after January 2023, our interpretation is that the market penalizes credits whose integrity has come under doubt, rather than awarding a premium for the CCB label. Sector and area subsamples seem to confirm this asymmetric pattern. CCB certification seems to function primarily as downside protection when market concerns about offset carbon integrity escalate; it does not create an upside premium, but it shields firms from the penalty investors impose on low-integrity offsets (without biodiversity impact).

Table B.1
Summary of economic relevance

This table reports the economic relevance of the largest offsetting companies' carbon-credit retirements. For each firm, the total tons of credits retired were valued using the maximum and minimum forestry-project prices published by ViridiosAI. Those two valuation figures were then divided by the firm's total assets and revenue, respectively, to express the results as percentages.

Company name	Total Asset		
	Max	Average	Min
AIR FRANCE – KLM SA	0.01344	0.00708	0.00072
APPLE INC.	0.00320	0.00169	0.00017
AUDI AG	0.01188	0.00626	0.00063
BAYER AG	0.00194	0.00102	0.00010
BRITISH AMERICAN TOBACCO PLC	0.00222	0.00117	0.00012
CHURCH & DWIGHT CO., INC.	0.05699	0.02951	0.00204
CONTINENTAL AG	0.02241	0.01180	0.00120
DELTA AIR LINES, INC.	0.00359	0.00189	0.00019
ENI S.P.A.	0.02992	0.01576	0.00160
INTERFACE, INC.	0.09986	0.05260	0.00533
KERING	0.05023	0.02645	0.00268
MICROSOFT CORPORATION	0.00004	0.00002	0.00000
NESTLÉ S.A.	0.01150	0.00605	0.00061
NETFLIX, INC.	0.01021	0.00538	0.00054
ORIGIN ENERGY LIMITED	0.00381	0.00200	0.00020
OSAKA GAS CO. LTD	0.00043	0.00022	0.00002
PETROCHINA COMPANY LIMITED	0.00052	0.00027	0.00003
PUBLICIS GROUPE S A	0.00009	0.00005	0.00000
SHELL PLC	0.01053	0.00554	0.00056
TAKEDA PHARMACEUTICAL COMPANY LIMITED	0.00002	0.00001	0.00000
TOKYO GAS CO., LTD.	0.00052	0.00027	0.00003
VOLKSWAGEN AG	0.00474	0.00250	0.00025
WALT DISNEY COMPANY (THE)	0.00037	0.00019	0.00002
Average	**0.01472**	**0.00773**	**0.00074**

Table B.2
Results for repeated users of carbon credits

This table presents significance test results for the estimated average CAR of repetitive users of carbon credits using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The counterfactual returns are constructed using the Fama-French model presented in the paper. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$, respectively.

	Both CCB & non-CCB		Only CCB		Only non-CCB	
Window	CAAR	Wilcoxon	CAAR	Wilcoxon	CAAR	Wilcoxon
PANEL A: Overall						
(-1:+1)	0.01	0.44	-0.42	0.20	-0.33**	0.02
(-1:+3)	0.03	0.39	-0.40	0.24	-0.59*	0.08
(-1:+5)	-0.09	0.98	-0.24	0.59	-0.44	0.40
(-1:+10)	-0.57	0.40	-0.92	0.47	-0.87	0.30
(-3:+3)	0.06	0.17	-0.25	0.61	-0.44*	0.07
(-5:+5)	-0.05	0.64	-0.42	0.28	-0.52*	0.07
(-10:+10)	-0.04	0.98	-0.61	0.26	-0.78*	0.10
PANEL B: Before Guardian						
(-1:+1)	0.08	0.17	-0.56	0.54	-0.23*	0.05
(-1:+3)	0.05	0.48	-1.43	0.24	-0.30	0.32
(-1:+5)	0.08	0.38	-1.97	0.47	0.38	0.91
(-1:+10)	-0.16	0.51	-4.87	0.34	0.73	0.79
(-3:+3)	0.09	0.24	-0.74	0.73	-0.36*	0.06
(-5:+5)	-0.03	0.62	-1.67	0.40	-0.19	0.28
(-10:+10)	0.15	0.41	-2.95	0.27	-0.09	0.54
PANEL C: After Guardian						
(-1:+1)	-0.14	0.27	-0.15	0.30	-0.45	0.18
(-1:+3)	-0.02	0.83	0.11	0.42	-0.90*	0.09
(-1:+5)	-0.25	0.35	0.23	0.70	-1.34	0.14
(-1:+10)	-1.08**	0.03	0.18	0.51	-2.59**	0.04
(-3:+3)	0.01	0.54	-0.03	0.45	-0.55	0.44
(-5:+5)	-0.11	0.83	-0.02	0.34	-0.91*	0.08
(-10:+10)	-0.38	0.16	0.19	0.64	-1.56*	0.05

Table B.3
Results for one-time only users of carbon credits

This table presents significance test results for the estimated average CAR of One-Time only users of carbon credits using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The counterfactual returns are constructed using the Fama-French model presented in the paper. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$, respectively.

	Both CCB & non-CCB		Only CCB		Only non-CCB	
Window	CAAR	Wilcoxon	CAAR	Wilcoxon	CAAR	Wilcoxon
PANEL A: Overall						
(-1:+1)	–	–	0.14	0.40	0.09	0.94
(-1:+3)	–	–	0.21	0.90	0.14	0.63
(-1:+5)	–	–	-0.96	0.33	-0.71	0.12
(-1:+10)	–	–	-2.49	0.19	-1.73*	0.09
(-3:+3)	–	–	-0.36	0.53	-0.15	0.79
(-5:+5)	–	–	-0.89	0.29	-0.51	0.55
(-10:+10)	–	–	0.02	0.96	-0.53	0.39
PANEL B: Before Guardian						
(-1:+1)	–	–	0.14	0.89	0.18	0.58
(-1:+3)	–	–	-0.38	0.32	1.01	0.65
(-1:+5)	–	–	-2.12	0.11	-0.14	0.33
(-1:+10)	–	–	-4.78	0.11	-0.99	0.46
(-3:+3)	–	–	-0.52	0.36	0.33	0.40
(-5:+5)	–	–	-1.01	0.21	-0.21	0.78
(-10:+10)	–	–	0.28	0.83	0.14	0.97
PANEL C: After Guardian						
(-1:+1)	–	–	0.14	0.25	-0.03	0.58
(-1:+3)	–	–	0.81	0.53	-1.18	0.19
(-1:+5)	–	–	0.20	0.72	-1.59	0.21
(-1:+10)	–	–	-0.20	0.83	-2.87*	0.07
(-3:+3)	–	–	-0.20	1.00	-0.90	0.14
(-5:+5)	–	–	-0.77	0.72	-0.99	0.50
(-10:+10)	–	–	-0.24	0.89	-1.57	0.19

Table B.4
Covariate balance

This table compares key firm characteristics for two groups. In Panel A, we have those who retire only CCB credits and those who retire only non-CCB credits. In Panel B, Ever CCB, firms that retire at least one CCB-certified carbon credit. For each variable, we report the cross-sectional mean with the standard deviation in parentheses. The final column shows the p-values from a two-sample Wilcoxon test.

PANEL A	Only CCB	Only non-CCB	Wilcoxon
Log Assets	3.73 (1.06)	4.10 (0.99)	0.00
Log Revenues	3.32 (0.97)	3.75 (0.88)	0.00
Carbon intensity	131.94 (402.67)	138.42 (485.06)	0.75
ESG score	42.27 (13.84)	39.84 (12.75)	0.37
ESG disclosure score	0.55 (0.12)	0.55 (0.12)	0.34
Obs.	136	472	

PANEL B	Ever CCB	Only non-CCB	Wilcoxon
Log Assets	4.14 (1.04)	4.10 (0.98)	0.49
Log Revenues	3.75 (0.97)	3.75 (0.88)	0.75
Carbon intensity	121.63 (335.42)	138.42 (485.06)	0.31
ESG score	43.45 (12.56)	39.84 (12.75)	0.00
ESG disclosure score	0.57 (0.12)	0.55 (0.12)	0.15
Obs.	309	472	

Table B.5
Difference in means results

This table presents significance test results for the estimated average difference in means CAR of CCB and non-CCB retirements using the non-parametric Wilcoxon Test (used for significance). The counterfactual returns are constructed using the Fama-French model presented in the paper. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Entire sample									
(-1:+1)	0.12	0.26	0.44	0.15	0.25	0.46	0.08	1.00	0.69
(-1:+3)	0.16	0.32	0.45	0.26	0.17	0.32	-0.05	0.66	0.88
(-1:+5)	0.17	0.77	0.52	0.30	0.39	0.38	0.01	0.56	0.98
(-1:+10)	0.32	0.58	0.36	0.28	0.55	0.54	0.30	0.87	0.61
(-3:+3)	0.26	0.26	0.43	0.41	0.41	0.48	0.09	0.87	0.85
(-5:+5)	0.23	0.68	0.61	0.10	0.95	0.93	0.55	0.57	0.42
(-10:+10)	0.64	0.80	0.35	0.48	0.70	0.80	1.04	0.86	0.33
PANEL B: Mining sector									
(-1:+1)	0.39	0.13	0.38	-0.04	0.71	0.95	0.53	0.40	0.37
(-1:+3)	0.01	0.57	0.98	-0.59	0.87	0.46	0.16	0.97	0.86
(-1:+5)	-0.23	0.66	0.78	-0.92	1.00	0.46	-0.06	0.89	0.96
(-1:+10)	0.49	0.21	0.61	-0.59	0.91	0.67	0.73	0.51	0.64
(-3:+3)	0.73*	0.08	0.42	0.25	0.88	0.54	0.66	0.33	0.65
(-5:+5)	1.28*	0.09	0.29	0.38	0.86	0.47	1.24	0.32	0.49
(-10:+10)	2.39**	0.04	0.13	0.79	0.71	0.48	2.19	0.29	0.32
PANEL C: Manufacturing sector									
(-1:+1)	0.17	0.72	0.46	0.27	0.50	0.37	-0.16	0.54	0.61
(-1:+3)	0.17	0.60	0.55	0.25	0.59	0.51	-0.20	0.78	0.65
(-1:+5)	0.22	0.73	0.53	0.42	0.55	0.37	-0.38	0.57	0.45
(-1:+10)	0.12	0.92	0.79	0.00	0.79	1.00	-0.01	0.82	0.98
(-3:+3)	0.42	0.72	0.37	1.15	0.88	0.46	-0.67	0.40	0.29
(-5:+5)	0.55	0.57	0.37	0.95	0.83	0.94	-0.07	0.97	0.94
(-10:+10)	0.45	0.58	0.57	0.37	0.89	0.93	0.54	0.49	0.64
PANEL D: Financial sector									
(-1:+1)	-0.01	0.62	0.98	-0.22	0.72	0.72	0.21	0.79	0.53
(-1:+3)	-0.31	0.57	0.52	-0.58	0.72	0.49	-0.02	0.66	0.96
(-1:+5)	-0.69	0.21	0.20	-1.17	0.19	0.17	-0.17	0.68	0.80
(-1:+10)	-1.35*	0.07	0.06	-2.81**	0.04	0.01	0.14	0.60	0.88
(-3:+3)	-0.90	0.58	0.26	-1.93	0.23	0.19	0.19	0.51	0.80
(-5:+5)	-1.76	0.17	0.06	-3.57*	0.06	0.07	0.09	0.74	0.94
(-10:+10)	-2.08	0.25	0.11	-4.47**	0.04	0.02	0.50	0.79	0.76
PANEL E: South and Central America									
(-1:+1)	1.15**	0.03	0.09	1.02*	0.08	0.23	1.95	0.16	0.11
(-1:+3)	1.87**	0.03	0.07	2.10**	0.05	0.10	1.99	0.30	0.31
(-1:+5)	2.54*	0.09	0.05	2.84	0.11	0.08	2.36	0.36	0.39
(-1:+10)	4.26***	0.01	0.01	3.65*	0.06	0.07	6.28*	0.05	0.08
(-3:+3)	3.03***	0.01	0.08	3.43*	0.07	0.05	0.61	0.97	0.82
(-5:+5)	3.91*	0.06	0.09	4.29	0.25	0.21	0.33	0.78	0.94
(-10:+10)	7.80*	0.09	0.11	9.97	0.37	0.22	2.70	0.70	0.68

Filtered analysis

We conduct a robustness test by removing overlapping events occurring within the same month for the same company. In addition, with this filter, we are also able to exclude all the events in the same week of earnings announcements from companies. This addresses concerns that closely timed retirements of voluntary carbon credits may introduce distortions in market reactions, either by diluting individual event effects or by creating artificial event clustering that does not reflect true investor sentiment. This filtered dataset is then subjected to the same event study methodology, the Fama-French Three-Factor Model.

Fama-French Four-Factor model

For robustness, we employ the *Fama-French Four-Factor Model* (FF4).

$$\textit{Fama-French Three-Factor Model: } AR_t = r_t - \alpha - \beta_1 r_{m,t} - \beta_2 r_{\text{SMB,t}} - \beta_3 r_{\text{HML,t}} - \beta_4 r_{\text{CMA,t}}$$

where the coefficients $\alpha, \beta_1, \beta_2, \beta_3, \beta_4$ are obtained evaluating the model over the estimation window. In this case, the returns for the small minus big, high minus low and the Carhart momentum portfolios were obtained from the online data library.

Fama-French Three-Factor model with regional factors

For robustness, we employ the *Fama-French Three-Factor Model* with regional factors (FF3r). We use regional factors for Europe, Asia and Oceania, Japan, and the US.

$$\textit{Fama-French Three-Factor Model: } AR_t = r_t - \alpha - \beta_1 r_{m,t} - \beta_2 r_{\text{SMB,t}} - \beta_3 r_{\text{HML,t}}$$

where the coefficients $\alpha, \beta_1, \beta_2, \beta_3, \beta_4$ are obtained evaluating the model over the estimation window. In this case, the returns for the small minus big and high minus low portfolios, and the regional factors were obtained from the online data library.

DiD around the Guardian shock

We complement the event study with a difference-in-differences specification that interacts Post-Guardian × CCB, with firm and date fixed effects. Even if the coefficients are consistent with our main results, they are not significant.

Table B.6
Filtered results for the entire sample

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). In this filtered sample, all events that occur within the same calendar month for the same company are removed. The counterfactual returns are constructed using the Fama-French model presented in the paper. Panel A presents results for the full dataset, Panel B for companies retiring CCB credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	-0.26**	0.04	0.02	-0.16	0.42	0.26	-0.40**	0.02	0.03
(-1:+3)	-0.43**	0.02	0.01	-0.32	0.22	0.09	-0.65**	0.02	0.02
(-1:+5)	-0.64***	0.00	0.00	-0.45	0.16	0.08	-1.01***	0.00	0.01
(-1:+10)	-0.72**	0.01	0.02	-0.30	0.50	0.35	-1.47***	0.00	0.01
(-3:+3)	-0.50**	0.01	0.02	-0.32	0.29	0.19	-0.86***	0.00	0.03
(-5:+5)	-0.66**	0.01	0.03	-0.29	0.35	0.37	-1.29***	0.00	0.02
(-10:+10)	-1.48***	0.00	0.02	-1.00	0.27	0.18	-2.37***	0.00	0.02
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	-0.05	0.77	0.77	0.01	0.74	0.97	-0.16	0.29	0.44
(-1:+3)	-0.10	0.94	0.65	-0.06	0.71	0.83	-0.33	0.43	0.28
(-1:+5)	-0.28	0.29	0.33	-0.16	0.93	0.70	-0.71	0.06	0.06
(-1:+10)	-0.22	0.38	0.53	0.02	0.72	0.96	-0.97*	0.03	0.06
(-3:+3)	-0.12	0.58	0.64	-0.04	0.86	0.91	-0.47	0.22	0.22
(-5:+5)	-0.44	0.13	0.17	-0.18	0.95	0.70	-1.10**	0.01	0.01
(-10:+10)	-0.98*	0.05	0.04	-0.88	0.40	0.18	-1.57**	0.02	0.03
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.31**	0.03	0.02	-0.28	0.11	0.12	-0.37	0.14	0.09
(-1:+3)	-0.54**	0.01	0.00	-0.51**	0.02	0.02	-0.61	0.13	0.08
(-1:+5)	-0.76***	0.00	0.00	-0.72**	0.01	0.01	-0.93*	0.06	0.05
(-1:+10)	-0.98***	0.00	0.01	-0.61	0.11	0.10	-1.59***	0.00	0.04
(-3:+3)	-0.66**	0.01	0.02	-0.57	0.10	0.06	-0.88**	0.02	0.08
(-5:+5)	-0.78**	0.01	0.04	-0.40	0.16	0.32	-1.48**	0.01	0.04
(-10:+10)	-1.81**	0.01	0.03	-1.12	0.27	0.26	-2.96***	0.00	0.03

Table B.7
Filtered results for the mining sector

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). In this filtered sample, all events that occur within the same calendar month for the same company are removed. The counterfactual returns are constructed using the Fama-French model presented in the paper. Panel A presents results for the full dataset, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	0.02	0.91	0.95	0.39	0.24	0.43	-0.83	0.14	0.17
(-1:+3)	-0.24	0.91	0.66	0.26	0.39	0.69	-1.41	0.11	0.13
(-1:+5)	0.05	0.69	0.95	0.75	0.61	0.46	-1.63*	0.09	0.12
(-1:+10)	0.52	0.79	0.58	1.92	0.12	0.11	-2.77**	0.03	0.05
(-3:+3)	0.23	0.85	0.73	1.21	0.17	0.14	-2.08*	0.08	0.06
(-5:+5)	-0.11	0.85	0.91	0.66	0.43	0.60	-1.92	0.11	0.10
(-10:+10)	-0.49	0.78	0.70	1.02	0.24	0.53	-4.04**	0.03	0.04
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	0.14	0.44	0.77	0.17	0.34	0.74	-0.06	0.67	0.97
(-1:+3)	-0.08	0.76	0.92	0.04	0.62	0.96	-0.96	0.67	0.74
(-1:+5)	-0.07	0.88	0.94	0.10	0.69	0.92	-1.37	0.67	0.65
(-1:+10)	0.91	0.41	0.46	1.22	0.33	0.34	-1.42	0.67	0.76
(-3:+3)	0.56	0.43	0.53	0.83	0.33	0.37	-1.46	0.83	0.67
(-5:+5)	0.13	0.63	0.92	0.14	0.58	0.92	0.11	1.00	0.97
(-10:+10)	-0.08	0.75	0.96	-0.24	0.80	0.90	1.16	1.00	0.80
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.73*	0.07	0.18	-0.55	0.35	0.56	-0.94*	0.09	0.05
(-1:+3)	-0.68	0.26	0.26	0.15	0.96	0.87	-1.63*	0.06	0.03
(-1:+5)	-0.45	0.19	0.64	0.61	0.70	0.71	-1.67*	0.09	0.06
(-1:+10)	-1.18*	0.07	0.23	0.46	0.94	0.77	-3.06**	0.01	0.00
(-3:+3)	-1.28*	0.07	0.10	-0.31	0.77	0.80	-2.40**	0.02	0.01
(-5:+5)	-1.52*	0.05	0.12	-0.10	0.73	0.95	-3.15**	0.01	0.00
(-10:+10)	-2.44**	0.03	0.08	0.30	0.99	0.88	-5.59***	0.00	0.00

Table B.8
Filtered results for the manufacturing sector

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). In this filtered sample, all events that occur within the same calendar month for the same company are removed. The counterfactual returns are constructed using the Fama-French model presented in the paper. Panel A presents results for the full dataset, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	-0.11	0.40	0.50	-0.04	0.90	0.87	-0.23	0.20	0.29
(-1:+3)	-0.25	0.40	0.24	-0.08	0.85	0.78	-0.50	0.10	0.11
(-1:+5)	-0.48	0.12	0.06	-0.20	0.90	0.56	-0.92**	0.01	0.02
(-1:+10)	-0.70*	0.05	0.06	-0.27	0.79	0.60	-1.34**	0.01	0.01
(-3:+3)	-0.46	0.21	0.09	-0.29	0.96	0.43	-0.75*	0.03	0.08
(-5:+5)	-0.40	0.29	0.25	0.00	0.83	1.00	-1.01*	0.03	0.04
(-10:+10)	-0.72	0.15	0.16	-0.17	1.00	0.82	-1.57**	0.01	0.02
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	-0.06	0.36	0.81	0.10	0.64	0.77	-0.31	0.40	0.35
(-1:+3)	-0.20	0.59	0.53	-0.01	0.89	0.99	-0.50	0.31	0.32
(-1:+5)	-0.51	0.15	0.19	0.08	0.89	0.88	-1.41**	0.01	0.01
(-1:+10)	-0.78	0.11	0.12	-0.08	0.86	0.90	-1.86*	0.02	0.01
(-3:+3)	-0.34	0.23	0.39	-0.02	0.71	0.97	-0.83	0.16	0.18
(-5:+5)	-0.65	0.06	0.16	0.11	0.81	0.85	-1.82**	0.01	0.01
(-10:+10)	-1.82**	0.01	0.02	-1.17	0.20	0.25	-2.82*	0.02	0.03
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.16	0.30	0.41	-0.14	0.54	0.62	-0.21	0.33	0.40
(-1:+3)	-0.34	0.26	0.16	-0.31	0.49	0.34	-0.45	0.20	0.21
(-1:+5)	-0.56	0.11	0.06	-0.56	0.30	0.17	-0.82	0.06	0.06
(-1:+10)	-0.78*	0.05	0.08	-0.47	0.58	0.44	-1.38**	0.01	0.03
(-3:+3)	-0.52	0.32	0.11	-0.55	0.82	0.21	-0.69	0.08	0.15
(-5:+5)	-0.48	0.30	0.25	-0.25	0.93	0.67	-1.18*	0.02	0.03
(-10:+10)	-0.74	0.25	0.23	-0.21	0.98	0.82	-1.88**	0.01	0.02

Table B.9
Filtered results for the South and Central America

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). In this filtered sample, all events that occur within the same calendar month for the same company are removed. The counterfactual returns are constructed using the Fama-French model presented in the paper. Panel A presents results for the full dataset, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	-0.65	0.12	0.06	-0.74	0.17	0.10	-0.49	0.43	0.35
(-1:+3)	-1.55**	0.01	0.00	-1.79*	0.02	0.01	-1.07	0.15	0.15
(-1:+5)	-2.24***	0.00	0.00	-2.48***	0.00	0.01	-1.78	0.12	0.11
(-1:+10)	-2.60***	0.00	0.00	-2.30*	0.04	0.04	-3.20**	0.01	0.01
(-3:+3)	-1.62***	0.00	0.01	-1.58*	0.02	0.06	-1.71	0.06	0.06
(-5:+5)	-2.80***	0.00	0.00	-2.92**	0.01	0.04	-2.57	0.06	0.03
(-10:+10)	-7.17***	0.00	0.07	-9.31*	0.03	0.12	-3.01*	0.02	0.07
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	0.26	0.33	0.67	0.14	0.43	0.84	1.25	0.86	0.47
(-1:+3)	0.06	0.65	0.95	-0.01	0.68	0.99	0.71	0.86	0.81
(-1:+5)	-0.04	0.82	0.97	-0.06	0.75	0.96	0.13	1.00	0.97
(-1:+10)	0.79	0.63	0.62	0.72	0.66	0.68	1.41	0.86	0.76
(-3:+3)	1.27	0.48	0.27	1.11	0.69	0.37	2.58	0.58	0.43
(-5:+5)	-0.75	0.72	0.66	-1.07	0.50	0.56	1.88	0.86	0.69
(-10:+10)	-1.95	0.37	0.39	-3.25	0.16	0.17	8.81	0.20	0.24
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-1.08**	0.01	0.01	-1.41**	0.01	0.01	-0.70	0.28	0.21
(-1:+3)	-2.28***	0.00	0.00	-3.13***	0.00	0.00	-1.29	0.12	0.10
(-1:+5)	-3.24***	0.00	0.00	-4.31***	0.00	0.00	-2.01	0.11	0.09
(-1:+10)	-4.18***	0.00	0.00	-4.62***	0.00	0.00	-3.76***	0.00	0.01
(-3:+3)	-2.99***	0.00	0.00	-3.67***	0.00	0.00	-2.23*	0.02	0.02
(-5:+5)	-3.70***	0.00	0.00	-4.27**	0.01	0.03	-3.10*	0.02	0.01
(-10:+10)	-9.47***	0.00	0.09	-13.74	0.08	0.18	-4.44***	0.00	0.01

Table B.10
Results for the entire sample (FF4)

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The counterfactual returns are constructed using the Fama-French Four-Factor model. Panel A presents results for the full dataset, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	-0.08	0.62	0.27	0.01	0.79	0.93	-0.23	0.11	0.04
(-1:+3)	-0.10	0.83	0.33	-0.04	0.84	0.77	-0.22	0.69	0.20
(-1:+5)	-0.23	0.27	0.09	-0.16	0.53	0.36	-0.40	0.15	0.07
(-1:+10)	-0.32	0.18	0.09	-0.06	0.88	0.79	-0.75**	0.02	0.03
(-3:+3)	-0.14	0.52	0.32	-0.06	0.66	0.72	-0.30	0.33	0.20
(-5:+5)	-0.28	0.19	0.14	-0.01	0.88	0.97	-0.63*	0.06	0.05
(-10:+10)	-0.81**	0.02	0.02	-0.29	0.75	0.49	-1.55***	0.00	0.01
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	0.01	0.49	0.95	0.08	0.38	0.62	-0.13	0.43	0.41
(-1:+3)	0.03	0.37	0.85	0.12	0.37	0.56	-0.21	0.63	0.37
(-1:+5)	-0.11	0.80	0.60	-0.01	0.84	0.97	-0.37	0.21	0.19
(-1:+10)	-0.07	0.86	0.78	0.13	0.67	0.72	-0.57	0.12	0.14
(-3:+3)	-0.02	0.96	0.91	0.11	0.76	0.65	-0.38	0.26	0.20
(-5:+5)	-0.39	0.13	0.11	-0.09	0.82	0.78	-0.84**	0.02	0.02
(-10:+10)	-0.64*	0.09	0.07	-0.49	0.40	0.31	-1.19**	0.02	0.02
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.14	0.23	0.14	-0.08	0.50	0.51	-0.26	0.14	0.07
(-1:+3)	-0.14	0.54	0.28	-0.15	0.24	0.36	-0.19	1.00	0.38
(-1:+5)	-0.28	0.25	0.09	-0.31	0.14	0.11	-0.37	0.43	0.20
(-1:+10)	-0.43	0.14	0.08	-0.22	0.41	0.40	-0.83*	0.07	0.07
(-3:+3)	-0.20	0.37	0.27	-0.21	0.23	0.35	-0.24	0.73	0.42
(-5:+5)	-0.24	0.50	0.32	-0.04	0.75	0.90	-0.64	0.23	0.13
(-10:+10)	-0.84	0.12	0.07	-0.22	0.98	0.68	-1.82***	0.00	0.02

Table B.11
Results for the mining sector (FF4)

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The sample includes only mining and quarrying companies. The counterfactual returns are constructed using the Fama-French Four-Factor model. Panel A presents, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	0.06	0.95	0.80	0.35	0.22	0.24	-0.58**	0.04	0.07
(-1:+3)	0.07	0.90	0.83	0.45	0.29	0.26	-0.76**	0.05	0.12
(-1:+5)	0.11	0.52	0.81	0.56	0.69	0.32	-0.89**	0.04	0.11
(-1:+10)	0.07	0.78	0.90	0.91	0.17	0.19	-1.77***	0.01	0.03
(-3:+3)	0.05	0.53	0.90	0.64	0.40	0.23	-1.22***	0.01	0.05
(-5:+5)	-0.36	0.20	0.51	0.44	0.64	0.54	-2.10***	0.00	0.01
(-10:+10)	-0.79	0.21	0.28	0.52	0.48	0.58	-3.61***	0.00	0.00
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	0.08	0.62	0.79	0.20	0.30	0.57	-0.36	0.26	0.46
(-1:+3)	0.00	0.93	1.00	0.21	0.47	0.65	-0.70	0.17	0.39
(-1:+5)	-0.09	0.80	0.85	0.14	0.76	0.81	-0.89	0.15	0.34
(-1:+10)	0.17	0.67	0.79	0.55	0.30	0.46	-1.15	0.28	0.38
(-3:+3)	-0.12	0.62	0.82	0.32	0.58	0.58	-1.60**	0.03	0.12
(-5:+5)	-0.71	0.25	0.31	-0.16	1.00	0.85	-2.60**	0.02	0.04
(-10:+10)	-0.83	0.41	0.37	-0.34	0.96	0.75	-2.50*	0.09	0.11
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.30	0.14	0.36	0.27	0.65	0.61	-0.90***	0.01	0.01
(-1:+3)	0.03	0.48	0.94	0.89	0.43	0.18	-0.88*	0.06	0.11
(-1:+5)	0.19	0.34	0.76	1.13	0.94	0.29	-0.80*	0.10	0.21
(-1:+10)	-0.29	0.20	0.69	1.18	0.58	0.30	-1.85***	0.01	0.03
(-3:+3)	-0.15	0.18	0.80	0.60	0.95	0.52	-0.94**	0.05	0.19
(-5:+5)	-0.49	0.14	0.52	0.86	0.87	0.47	-1.92**	0.02	0.03
(-10:+10)	-1.39*	0.08	0.17	1.07	0.69	0.48	-4.01***	0.00	0.00

Table B.12
Results for the manufacturing sector (FF4)

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The sample includes only manufacturing companies. The counterfactual returns are constructed using the Fama-French Four-Factor model. Panel A presents results for the full dataset, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	-0.05	0.96	0.62	-0.03	0.93	0.85	-0.19	0.58	0.22
(-1:+3)	-0.03	0.84	0.80	-0.06	0.88	0.71	-0.14	0.78	0.53
(-1:+5)	-0.21	0.40	0.21	-0.16	0.80	0.45	-0.54**	0.04	0.04
(-1:+10)	-0.33	0.15	0.15	-0.19	0.58	0.53	-0.78**	0.02	0.03
(-3:+3)	-0.09	0.97	0.60	-0.19	0.72	0.37	-0.11	0.84	0.69
(-5:+5)	-0.15	0.50	0.50	-0.07	0.92	0.81	-0.54*	0.09	0.10
(-10:+10)	-0.66**	0.04	0.04	-0.35	0.52	0.42	-1.49***	0.00	0.00
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	0.10	0.70	0.60	0.18	0.65	0.51	-0.28	0.51	0.27
(-1:+3)	0.15	0.39	0.53	0.13	0.64	0.69	-0.23	0.78	0.51
(-1:+5)	0.00	0.95	0.99	0.12	0.62	0.76	-0.79*	0.08	0.05
(-1:+10)	-0.21	0.54	0.60	-0.20	0.68	0.71	-0.83	0.19	0.14
(-3:+3)	0.18	0.59	0.55	0.05	0.93	0.90	-0.25	0.99	0.58
(-5:+5)	0.00	0.82	1.00	0.01	0.95	0.99	-0.82	0.16	0.14
(-10:+10)	-0.60	0.36	0.33	-0.78	0.40	0.34	-1.31	0.16	0.15
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.10	0.71	0.41	-0.12	0.48	0.44	-0.15	0.84	0.42
(-1:+3)	-0.07	1.00	0.67	-0.15	0.54	0.43	-0.07	0.99	0.78
(-1:+5)	-0.27	0.33	0.16	-0.35	0.35	0.14	-0.43	0.15	0.17
(-1:+10)	-0.35	0.19	0.19	-0.23	0.55	0.49	-0.79**	0.04	0.06
(-3:+3)	-0.13	0.90	0.51	-0.29	0.61	0.24	-0.05	0.88	0.88
(-5:+5)	-0.23	0.37	0.36	-0.18	0.73	0.59	-0.61*	0.07	0.12
(-10:+10)	-0.73**	0.04	0.05	-0.30	0.63	0.55	-1.78***	0.00	0.00

Table B.13
Results for companies from South and Central America (FF4)

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The sample includes only South and Central American companies. The counterfactual returns are constructed using the Fama-French Four-Factor model. Panel A presents results for the full dataset, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	0.02	0.93	0.95	0.13	0.77	0.76	-0.17	0.80	0.69
(-1:+3)	-0.47	0.42	0.27	-0.49	0.56	0.42	-0.45	0.59	0.39
(-1:+5)	-0.97*	0.09	0.08	-0.87	0.21	0.24	-1.16	0.28	0.15
(-1:+10)	-1.60***	0.01	0.02	-1.09	0.24	0.26	-2.52***	0.00	0.01
(-3:+3)	-0.58	0.25	0.26	-0.27	0.61	0.71	-1.13	0.20	0.12
(-5:+5)	-1.61**	0.05	0.04	-1.32	0.28	0.24	-2.16*	0.07	0.02
(-10:+10)	-5.42***	0.00	0.05	-6.22	0.15	0.15	-4.26***	0.00	0.00
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	0.74*	0.08	0.21	0.60	0.17	0.37	1.69	0.29	0.16
(-1:+3)	0.68	0.25	0.44	0.62	0.32	0.55	1.19	0.83	0.52
(-1:+5)	0.66	0.70	0.55	0.68	0.86	0.60	0.67	0.83	0.79
(-1:+10)	1.18	0.38	0.41	1.07	0.49	0.52	2.06	0.53	0.50
(-3:+3)	1.73	0.18	0.09	1.72	0.31	0.15	2.05	0.29	0.30
(-5:+5)	-0.44	0.83	0.77	-0.40	0.87	0.82	-0.74	0.83	0.82
(-10:+10)	-1.51	0.30	0.46	-2.38	0.21	0.31	3.60	0.83	0.47
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.27	0.32	0.44	-0.17	0.53	0.75	-0.39	0.48	0.38
(-1:+3)	-0.93*	0.08	0.05	-1.19	0.13	0.11	-0.65	0.42	0.24
(-1:+5)	-1.62**	0.02	0.01	-1.86*	0.07	0.04	-1.38	0.22	0.11
(-1:+10)	-2.72***	0.00	0.00	-2.47**	0.04	0.04	-3.07***	0.00	0.00
(-3:+3)	-1.52**	0.03	0.01	-1.55	0.15	0.09	-1.51*	0.09	0.05
(-5:+5)	-2.08**	0.03	0.02	-1.90	0.23	0.20	-2.33*	0.06	0.01
(-10:+10)	-7.00***	0.01	0.06	-8.67	0.34	0.21	-5.20***	0.00	0.00

Table B.14
Results for the entire sample (FF3r)

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The counterfactual returns are constructed using the Fama-French model with regional factors. Panel A presents results for the full dataset, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	-0.09	0.49	0.24	0.01	0.84	0.95	-0.24*	0.08	0.04
(-1:+3)	-0.10	0.88	0.33	-0.02	0.97	0.85	-0.23	0.62	0.18
(-1:+5)	-0.23	0.28	0.09	-0.14	0.62	0.41	-0.42	0.13	0.06
(-1:+10)	-0.29	0.22	0.13	0.02	0.86	0.94	-0.77***	0.01	0.03
(-3:+3)	-0.14	0.56	0.32	-0.03	0.88	0.88	-0.33	0.28	0.16
(-5:+5)	-0.24	0.30	0.21	0.05	0.85	0.84	-0.62*	0.07	0.05
(-10:+10)	-0.76**	0.05	0.03	-0.20	0.97	0.63	-1.54***	0.00	0.01
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	-0.02	0.65	0.89	0.07	0.40	0.67	-0.16	0.31	0.31
(-1:+3)	0.02	0.40	0.90	0.13	0.33	0.53	-0.24	0.59	0.32
(-1:+5)	-0.11	0.75	0.60	0.01	0.83	0.97	-0.38	0.21	0.18
(-1:+10)	-0.06	0.81	0.82	0.16	0.67	0.67	-0.56	0.12	0.14
(-3:+3)	-0.03	0.98	0.88	0.12	0.75	0.63	-0.39	0.24	0.19
(-5:+5)	-0.36	0.15	0.14	-0.05	0.87	0.87	-0.83**	0.03	0.02
(-10:+10)	-0.59*	0.10	0.10	-0.40	0.48	0.41	-1.17**	0.02	0.02
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.14	0.19	0.14	-0.08	0.43	0.50	-0.25	0.13	0.08
(-1:+3)	-0.14	0.62	0.29	-0.13	0.35	0.42	-0.19	0.96	0.37
(-1:+5)	-0.28	0.29	0.09	-0.29	0.20	0.13	-0.39	0.39	0.18
(-1:+10)	-0.39	0.20	0.11	-0.12	0.69	0.63	-0.86**	0.04	0.06
(-3:+3)	-0.19	0.49	0.30	-0.17	0.40	0.47	-0.28	0.66	0.35
(-5:+5)	-0.21	0.70	0.40	0.02	0.97	0.94	-0.64	0.25	0.13
(-10:+10)	-0.79	0.24	0.09	-0.15	0.72	0.79	-1.82***	0.01	0.02

Table B.15
Results for the mining sector (FF3r)

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The sample includes only mining and quarrying companies. The counterfactual returns are constructed using the Fama-French model with regional factors. Panel A presents, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	0.06	0.98	0.80	0.36	0.20	0.25	-0.58**	0.04	0.07
(-1:+3)	0.08	0.88	0.80	0.46	0.31	0.25	-0.75**	0.05	0.12
(-1:+5)	0.13	0.55	0.77	0.59	0.65	0.30	-0.88**	0.04	0.11
(-1:+10)	0.10	0.79	0.85	0.96	0.17	0.17	-1.74***	0.01	0.03
(-3:+3)	0.07	0.55	0.87	0.65	0.42	0.22	-1.20***	0.01	0.05
(-5:+5)	-0.29	0.25	0.59	0.52	0.58	0.47	-2.05***	0.00	0.01
(-10:+10)	-0.65	0.27	0.38	0.67	0.41	0.48	-3.50***	0.00	0.00
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	0.08	0.59	0.79	0.21	0.29	0.56	-0.37	0.27	0.44
(-1:+3)	0.02	0.96	0.95	0.23	0.48	0.61	-0.69	0.17	0.39
(-1:+5)	-0.05	0.87	0.91	0.18	0.68	0.76	-0.86	0.15	0.35
(-1:+10)	0.22	0.68	0.73	0.61	0.31	0.43	-1.10	0.30	0.40
(-3:+3)	-0.09	0.67	0.87	0.35	0.58	0.55	-1.58**	0.03	0.12
(-5:+5)	-0.60	0.29	0.39	-0.04	0.94	0.96	-2.53**	0.02	0.05
(-10:+10)	-0.65	0.49	0.48	-0.14	0.94	0.90	-2.39*	0.10	0.13
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.31	0.14	0.35	0.25	0.69	0.65	-0.90***	0.01	0.01
(-1:+3)	0.01	0.47	0.99	0.82	0.47	0.21	-0.86*	0.06	0.12
(-1:+5)	0.18	0.33	0.78	1.10	0.97	0.31	-0.80	0.11	0.21
(-1:+10)	-0.27	0.20	0.71	1.19	0.58	0.30	-1.82***	0.01	0.04
(-3:+3)	-0.16	0.18	0.78	0.57	0.88	0.54	-0.94*	0.06	0.19
(-5:+5)	-0.47	0.15	0.53	0.85	0.85	0.48	-1.87**	0.02	0.03
(-10:+10)	-1.31*	0.09	0.19	1.08	0.71	0.48	-3.86***	0.00	0.00

Table B.16
Results for the manufacturing sector (FF3r)

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The sample includes only manufacturing companies. The counterfactual returns are constructed using the Fama-French model with regional factors. Panel A presents results for the full dataset, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	-0.04	0.99	0.73	0.02	0.90	0.91	-0.21	0.45	0.18
(-1:+3)	-0.01	0.71	0.94	0.00	0.79	0.99	-0.17	0.64	0.42
(-1:+5)	-0.19	0.47	0.26	-0.09	0.94	0.66	-0.58**	0.02	0.03
(-1:+10)	-0.27	0.23	0.24	-0.03	0.97	0.91	-0.87***	0.01	0.01
(-3:+3)	-0.06	0.79	0.74	-0.10	0.86	0.63	-0.16	0.73	0.56
(-5:+5)	-0.08	0.66	0.72	0.07	0.71	0.80	-0.58*	0.06	0.08
(-10:+10)	-0.52	0.12	0.11	-0.11	1.00	0.80	-1.50***	0.00	0.00
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	0.09	0.78	0.63	0.20	0.61	0.44	-0.32	0.38	0.19
(-1:+3)	0.14	0.43	0.55	0.18	0.59	0.59	-0.30	0.65	0.39
(-1:+5)	-0.01	0.95	0.96	0.16	0.63	0.69	-0.85**	0.05	0.04
(-1:+10)	-0.17	0.48	0.67	-0.07	0.71	0.90	-0.91	0.17	0.11
(-3:+3)	0.18	0.65	0.53	0.11	0.97	0.79	-0.29	0.96	0.50
(-5:+5)	0.03	0.80	0.94	0.12	0.99	0.80	-0.90	0.14	0.10
(-10:+10)	-0.49	0.43	0.42	-0.54	0.56	0.50	-1.38	0.14	0.13
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.08	0.77	0.53	-0.08	0.60	0.64	-0.16	0.76	0.39
(-1:+3)	-0.03	0.82	0.84	-0.07	0.86	0.70	-0.10	0.89	0.70
(-1:+5)	-0.24	0.44	0.22	-0.26	0.59	0.27	-0.47	0.11	0.14
(-1:+10)	-0.30	0.30	0.27	-0.08	0.97	0.82	-0.89**	0.02	0.04
(-3:+3)	-0.09	0.78	0.66	-0.19	0.89	0.45	-0.10	0.78	0.77
(-5:+5)	-0.16	0.55	0.54	-0.03	0.86	0.92	-0.64**	0.04	0.10
(-10:+10)	-0.58	0.14	0.13	-0.06	0.91	0.91	-1.78***	0.00	0.00

Table B.17
Results for companies from South and Central America (FF3r)

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The sample includes only South and Central American companies. The counterfactual returns are constructed using the Fama-French model with regional factors. Panel A presents results for the full dataset, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	-0.07	0.85	0.82	0.03	0.94	0.95	-0.22	0.67	0.59
(-1:+3)	-0.58	0.27	0.17	-0.58	0.41	0.33	-0.60	0.48	0.28
(-1:+5)	-1.02*	0.09	0.07	-0.94	0.22	0.23	-1.20	0.24	0.14
(-1:+10)	-1.57***	0.01	0.03	-0.96	0.30	0.32	-2.65***	0.01	0.01
(-3:+3)	-0.76	0.12	0.15	-0.36	0.47	0.62	-1.45	0.11	0.05
(-5:+5)	-1.58*	0.06	0.04	-1.17	0.37	0.30	-2.34*	0.06	0.01
(-10:+10)	-5.40***	0.01	0.06	-6.12	0.21	0.18	-4.37***	0.00	0.01
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	0.75*	0.07	0.20	0.65	0.16	0.33	1.52	0.21	0.18
(-1:+3)	0.75	0.21	0.40	0.70	0.31	0.49	1.18	0.40	0.52
(-1:+5)	0.79	0.55	0.48	0.80	0.69	0.53	0.90	0.83	0.72
(-1:+10)	1.47	0.30	0.30	1.28	0.44	0.43	2.96	0.40	0.35
(-3:+3)	1.81	0.17	0.08	1.83	0.28	0.13	1.95	0.29	0.33
(-5:+5)	-0.02	0.96	0.99	0.00	0.96	1.00	-0.08	0.83	0.98
(-10:+10)	-0.74	0.60	0.72	-1.73	0.39	0.45	5.32	0.40	0.28
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.40	0.17	0.25	-0.37	0.30	0.47	-0.43	0.37	0.33
(-1:+3)	-1.12**	0.03	0.02	-1.40*	0.06	0.06	-0.81	0.34	0.17
(-1:+5)	-1.75***	0.01	0.01	-2.04**	0.05	0.04	-1.45	0.16	0.09
(-1:+10)	-2.79***	0.00	0.00	-2.38**	0.04	0.04	-3.32***	0.00	0.00
(-3:+3)	-1.79***	0.01	0.00	-1.76*	0.07	0.05	-1.86**	0.04	0.02
(-5:+5)	-2.21**	0.03	0.02	-1.91	0.25	0.20	-2.61**	0.04	0.01
(-10:+10)	-7.28***	0.00	0.06	-8.92	0.36	0.22	-5.53***	0.00	0.00

Table B.18
Difference-in-Differences

This table reports estimates from a difference-in-differences (DiD) specification examining the impact of the Guardian article on cumulative abnormal returns around carbon credit retirement announcements. The dependent variable is CAR over the specified event window. The key independent variable is the interaction term Post-Guardian × CCB, which equals one for CCB-certified retirements announced after the Guardian article. All regressions include firm fixed effects and date fixed effects. Standard errors are clustered at the firm level. The coefficient on Post-Guardian × CCB captures whether CCB certification mitigated the negative stock market reaction following the Guardian shock.

	(-1:+1)	(-1:+3)	(-1:+5)	(-1:+10)	(-3:+3)	(-5:+5)	(-10:+10)
Post × CCB	0.009	0.010	0.013	0.022	0.013	0.014	0.048
	(0.009)	(0.013)	(0.017)	(0.028)	(0.017)	(0.028)	(0.054)
Adj. R^2	0.124	0.128	0.109	0.098	0.110	0.094	0.038
Observations	2,136	2,136	2,136	2,136	2,136	2,136	2,136

Additional Tables

Table B.19
Results for the entire sample

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The counterfactual returns are constructed using the Fama-French model presented in the paper. Panel A presents results for the full dataset, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	-0.09	0.49	0.24	0.01	0.83	0.95	-0.24*	0.08	0.04
(-1:+3)	-0.10	0.88	0.33	-0.02	0.96	0.86	-0.23	0.61	0.18
(-1:+5)	-0.23	0.28	0.09	-0.14	0.63	0.42	-0.42	0.12	0.06
(-1:+10)	-0.29	0.22	0.13	0.02	0.84	0.93	-0.78**	0.01	0.02
(-3:+3)	-0.14	0.57	0.33	-0.03	0.89	0.88	-0.34	0.27	0.16
(-5:+5)	-0.24	0.30	0.21	0.05	0.84	0.83	-0.63*	0.07	0.05
(-10:+10)	-0.76**	0.05	0.03	-0.20	0.95	0.64	-1.55***	0.00	0.01
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	-0.02	0.66	0.89	0.07	0.40	0.67	-0.17	0.30	0.31
(-1:+3)	0.02	0.41	0.90	0.13	0.32	0.53	-0.24	0.58	0.31
(-1:+5)	-0.11	0.75	0.60	0.01	0.83	0.97	-0.38	0.20	0.18
(-1:+10)	-0.06	0.80	0.82	0.16	0.67	0.67	-0.57	0.12	0.14
(-3:+3)	0.07	0.46	0.80	0.25	0.26	0.50	-0.20	0.86	0.59
(-5:+5)	0.03	0.84	0.94	0.13	0.71	0.79	-0.09	0.92	0.86
(-10:+10)	-0.15	0.73	0.77	0.34	0.61	0.62	-0.79	0.22	0.28
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.14	0.19	0.14	-0.08	0.44	0.51	-0.25	0.12	0.08
(-1:+3)	-0.14	0.63	0.30	-0.13	0.36	0.42	-0.19	0.95	0.37
(-1:+5)	-0.28	0.29	0.09	-0.29	0.21	0.14	-0.39	0.38	0.17
(-1:+10)	-0.38	0.21	0.11	-0.12	0.71	0.65	-0.87**	0.04	0.05
(-3:+3)	-0.19	0.49	0.30	-0.16	0.41	0.48	-0.29	0.65	0.35
(-5:+5)	-0.20	0.71	0.40	0.03	0.95	0.93	-0.64	0.24	0.13
(-10:+10)	-0.79	0.24	0.09	-0.14	0.70	0.80	-1.83**	0.01	0.02

Table B.20
Results for the mining sector

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The sample includes only mining and quarrying companies. The counterfactual returns are constructed using the Fama-French model presented in the paper. Panel A presents, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	0.06	0.98	0.81	0.35	0.20	0.25	-0.59**	0.04	0.07
(-1:+3)	0.08	0.87	0.81	0.46	0.31	0.25	-0.75*	0.05	0.12
(-1:+5)	0.12	0.55	0.77	0.59	0.65	0.30	-0.89**	0.04	0.11
(-1:+10)	0.10	0.78	0.85	0.95	0.17	0.17	-1.74**	0.01	0.03
(-3:+3)	0.06	0.55	0.88	0.65	0.41	0.22	-1.21**	0.01	0.05
(-5:+5)	-0.30	0.25	0.59	0.52	0.58	0.47	-2.05***	0.00	0.01
(-10:+10)	-0.66	0.27	0.37	0.66	0.41	0.49	-3.51***	0.00	0.00
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	0.08	0.59	0.79	0.21	0.29	0.56	-0.37	0.27	0.44
(-1:+3)	0.02	0.96	0.95	0.23	0.48	0.61	-0.70	0.18	0.39
(-1:+5)	-0.06	0.86	0.91	0.18	0.69	0.76	-0.87	0.15	0.35
(-1:+10)	0.22	0.69	0.74	0.60	0.31	0.43	-1.10	0.30	0.40
(-3:+3)	0.57	0.15	0.41	0.82	0.10	0.32	-0.28	0.83	0.83
(-5:+5)	0.81	0.22	0.40	1.23	0.12	0.29	-0.64	0.70	0.69
(-10:+10)	1.07	0.21	0.38	1.87*	0.09	0.21	-1.68	0.46	0.37
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.31	0.14	0.35	0.25	0.69	0.65	-0.90**	0.01	0.01
(-1:+3)	0.01	0.46	0.99	0.82	0.47	0.21	-0.86*	0.07	0.12
(-1:+5)	0.17	0.33	0.78	1.10	0.98	0.31	-0.81	0.11	0.21
(-1:+10)	-0.27	0.20	0.71	1.19	0.57	0.30	-1.83**	0.01	0.04
(-3:+3)	-0.16	0.18	0.78	0.57	0.88	0.54	-0.94*	0.06	0.19
(-5:+5)	-0.47	0.15	0.52	0.85	0.86	0.47	-1.88**	0.02	0.03
(-10:+10)	-1.32*	0.09	0.19	1.08	0.71	0.48	-3.87***	0.00	0.00

Table B.21
Results for the manufacturing sector

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The sample includes only manufacturing companies. The counterfactual returns are constructed using the Fama-French model presented in the paper. Panel A presents results for the full dataset, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	-0.04	1.00	0.74	0.02	0.89	0.90	-0.21	0.45	0.17
(-1:+3)	-0.01	0.70	0.94	0.00	0.77	0.98	-0.17	0.63	0.42
(-1:+5)	-0.18	0.48	0.27	-0.09	0.92	0.67	-0.58**	0.02	0.03
(-1:+10)	-0.27	0.23	0.24	-0.02	0.95	0.93	-0.88**	0.01	0.01
(-3:+3)	-0.05	0.78	0.75	-0.10	0.84	0.64	-0.16	0.73	0.56
(-5:+5)	-0.07	0.67	0.73	0.08	0.69	0.78	-0.58*	0.05	0.08
(-10:+10)	-0.52	0.12	0.11	-0.10	0.98	0.82	-1.51***	0.00	0.00
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	0.09	0.79	0.63	0.20	0.61	0.44	-0.33	0.37	0.19
(-1:+3)	0.14	0.43	0.55	0.18	0.58	0.59	-0.30	0.64	0.39
(-1:+5)	-0.01	0.96	0.96	0.16	0.63	0.69	-0.85*	0.05	0.04
(-1:+10)	-0.17	0.48	0.67	-0.07	0.71	0.90	-0.91	0.17	0.10
(-3:+3)	0.34	0.61	0.43	0.97	0.16	0.13	-0.77	0.24	0.16
(-5:+5)	0.40	0.74	0.48	0.92	0.32	0.26	-0.71	0.24	0.36
(-10:+10)	-0.12	0.83	0.87	0.33	0.73	0.74	-1.25	0.27	0.23
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.08	0.78	0.54	-0.07	0.62	0.65	-0.17	0.75	0.38
(-1:+3)	-0.03	0.81	0.85	-0.07	0.89	0.72	-0.10	0.88	0.69
(-1:+5)	-0.23	0.45	0.23	-0.26	0.61	0.28	-0.47	0.10	0.13
(-1:+10)	-0.29	0.31	0.27	-0.07	0.94	0.84	-0.90**	0.02	0.03
(-3:+3)	-0.08	0.77	0.67	-0.18	0.88	0.46	-0.10	0.78	0.76
(-5:+5)	-0.15	0.56	0.55	-0.03	0.83	0.94	-0.64*	0.04	0.10
(-10:+10)	-0.57	0.15	0.13	-0.04	0.89	0.93	-1.79***	0.00	0.00

Table B.22
Results for companies from South and Central America

This table presents significance test results for estimated average CAR using the non-parametric Wilcoxon Test (used for significance) and the T-test (applied for robustness). The sample includes only South and Central American companies. The counterfactual returns are constructed using the Fama-French model presented in the paper. Panel A presents results for the full dataset, Panel B for companies retiring CCB carbon credits, and Panel C for companies retiring general carbon credits. *, **, and *** indicate significance when $p \leq 0.1$, $p \leq 0.05$, $p \leq 0.01$ respectively.

	Overall			Before Guardian			After Guardian		
Windows	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test	CAAR	Wilcoxon	T-test
PANEL A: Full Dataset									
(-1:+1)	-0.07	0.85	0.82	0.03	0.94	0.95	-0.23	0.67	0.59
(-1:+3)	-0.58	0.27	0.17	-0.58	0.41	0.33	-0.60	0.48	0.28
(-1:+5)	-1.02*	0.09	0.07	-0.94	0.22	0.23	-1.20	0.24	0.14
(-1:+10)	-1.57**	0.01	0.03	-0.96	0.30	0.32	-2.65**	0.01	0.01
(-3:+3)	-0.76	0.12	0.15	-0.36	0.47	0.62	-1.45	0.11	0.05
(-5:+5)	-1.58*	0.06	0.04	-1.17	0.36	0.30	-2.34*	0.06	0.01
(-10:+10)	-5.40**	0.01	0.06	-6.12	0.21	0.18	-4.37***	0.00	0.01
PANEL B: Companies Retiring CCB Carbon Credits									
(-1:+1)	0.75*	0.07	0.20	0.65	0.16	0.34	1.52	0.21	0.18
(-1:+3)	0.75	0.21	0.40	0.70	0.31	0.50	1.18	0.40	0.52
(-1:+5)	0.79	0.54	0.48	0.80	0.69	0.53	0.90	0.67	0.72
(-1:+10)	1.47	0.30	0.30	1.27	0.44	0.44	2.96	0.40	0.35
(-3:+3)	1.24	0.18	0.45	1.67	0.12	0.38	-1.25	0.53	0.64
(-5:+5)	1.70	0.29	0.43	2.38	0.22	0.33	-2.28	0.83	0.62
(-10:+10)	0.52	0.75	0.86	1.06	0.67	0.75	-2.83	0.67	0.66
PANEL C: Companies Retiring General Carbon Credits									
(-1:+1)	-0.40	0.17	0.25	-0.37	0.30	0.48	-0.43	0.37	0.33
(-1:+3)	-1.12**	0.03	0.02	-1.40*	0.06	0.06	-0.81	0.34	0.17
(-1:+5)	-1.75**	0.01	0.01	-2.04*	0.05	0.04	-1.46	0.16	0.09
(-1:+10)	-2.79***	0.00	0.00	-2.38**	0.04	0.04	-3.32***	0.00	0.00
(-3:+3)	-1.79**	0.01	0.00	-1.76*	0.07	0.05	-1.86**	0.04	0.02
(-5:+5)	-2.21**	0.03	0.02	-1.91	0.25	0.21	-2.61**	0.04	0.01
(-10:+10)	-7.28***	0.00	0.06	-8.91	0.37	0.22	-5.53***	0.00	0.00

Table B.23
Economic magnitudes

This table reports median cumulative average abnormal returns (CAAR, in basis points) and the corresponding economic magnitude (in USD) across different event windows. Results are shown for the full dataset (Panel A), for companies retiring CCB carbon credits (Panel B), and for companies retiring general carbon credits (Panel C). Columns group results for the overall sample, for the pre-Guardian period, and for the post-Guardian period.

	Overall		Before Guardian		After Guardian	
Windows	CAAR	Econ. magnitude	CAAR	Econ. magnitude	CAAR	Econ. magnitude
PANEL A: Full Dataset						
(-1:+1)	-0.42	-1,662.34	3.67	14,459.63	-5.31	-15,331.49
(-1:+3)	-0.19	-758.46	1.91	7,522.00	-2.46	-7,107.69
(-1:+5)	-4.40	-17,364.07	0.00	0.00	-21.99	-63,469.66
(-1:+10)	-9.42	-37,138.54	0.00	0.00	-44.00	-127,018.09
(-3:+3)	-1.55	-6,109.37	3.73	14,716.91	-15.02	-43,343.09
(-5:+5)	-2.90	-11,442.25	0.10	398.76	-12.21	-35,238.68
(-10:+10)	-38.28	-150,921.32	-9.25	-36,472.74	-96.59	-278,812.16
PANEL B: Companies Retiring CCB Carbon Credits						
(-1:+1)	-0.29	-1,115.66	4.58	18,045.32	-7.16	-17,728.76
(-1:+3)	0.04	165.88	4.78	18,864.71	-1.20	-2,969.55
(-1:+5)	-3.42	-13,172.41	5.66	22,309.49	-22.14	-54,795.45
(-1:+10)	-2.86	-11,025.63	0.21	840.87	-42.76	-105,854.82
(-3:+3)	0.06	237.12	5.24	20,661.33	-2.46	-6,078.98
(-5:+5)	-2.90	-11,178.74	0.55	2,164.72	-17.43	-43,143.40
(-10:+10)	-38.35	-147,724.41	-13.16	-51,873.56	-99.68	-246,740.59
PANEL C: Companies Retiring General Carbon Credits						
(-1:+1)	-1.40	-6,247.43	-2.34	-9,192.82	1.23	5,589.41
(-1:+3)	-4.56	-20,325.66	-1.72	-6,755.29	-8.67	-39,400.13
(-1:+5)	-9.64	-43,010.53	-14.72	-57,747.17	-10.32	-46,912.61
(-1:+10)	-30.79	-137,286.45	-5.96	-23,384.45	-44.56	-202,592.95
(-3:+3)	-15.29	-68,193.94	-2.42	-9,501.60	-45.18	-205,415.61
(-5:+5)	-2.51	-11,215.41	-1.17	-4,570.07	16.03	72,859.08
(-10:+10)	-29.24	-130,372.71	7.67	30,081.90	-64.02	-291,052.48

Additional Figures

Figure B.1
Top 30 carbon offset buyers between 2009-2024

This figure shows the list of the leading 30 companies based on the cumulative volume of carbon offsets retired over the period 2009–2024. The retirements are divided into: CCB-certified forestry credits (green bars) and general forestry-related credits (red).

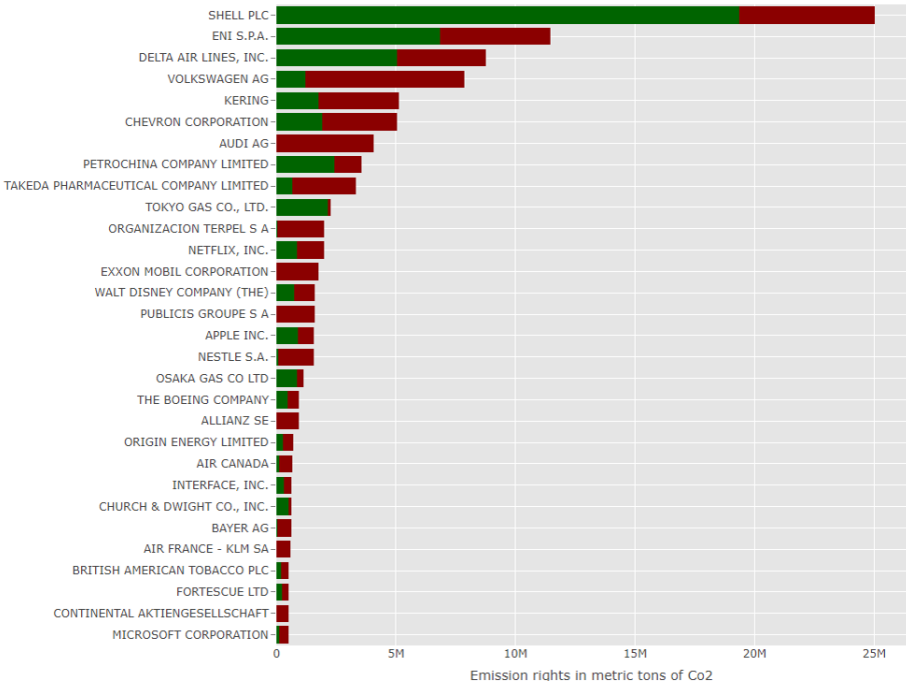


Figure B.2
Results for the entire sample, mining and quarrying sector

This figure shows the abnormal and cumulative abnormal returns for the full dataset (Panel A), companies retiring CCB-labeled carbon credits (Panel B) and finally companies retiring general carbon credits (Panel C). The sample used in this case is the mining and quarrying sector for the overall period. The dotted line represents the event, i.e. the carbon credit retirement.

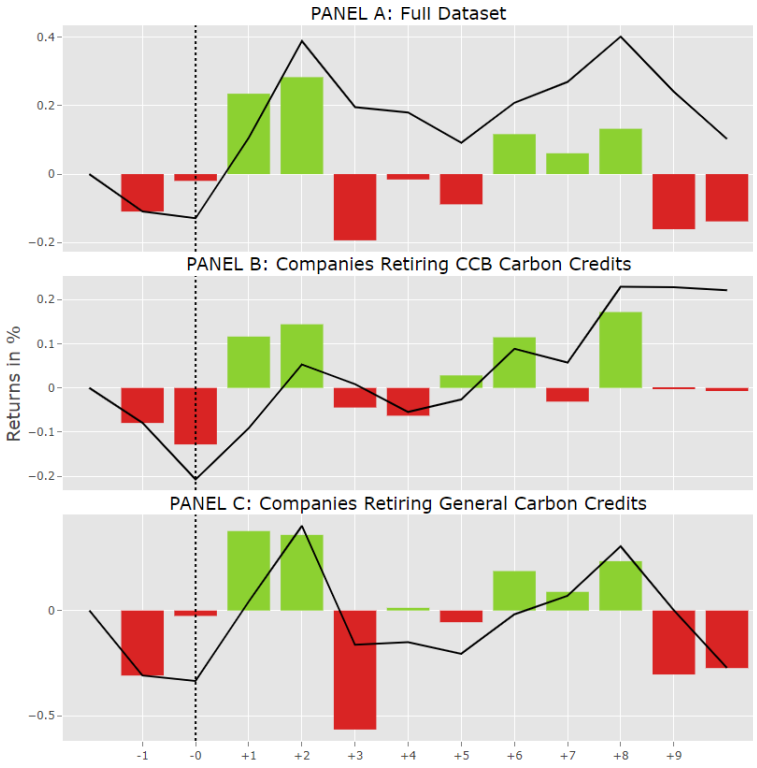


Figure B.3
Results for the entire sample, manufacturing sector

This figure shows the abnormal and cumulative abnormal returns for the full dataset (Panel A), companies retiring CCB-labeled carbon credits (Panel B), and companies retiring general carbon credits (Panel C). The sample used in this case is the manufacturing sector for the overall period. The dotted line represents the event, i.e. the carbon credit retirement.

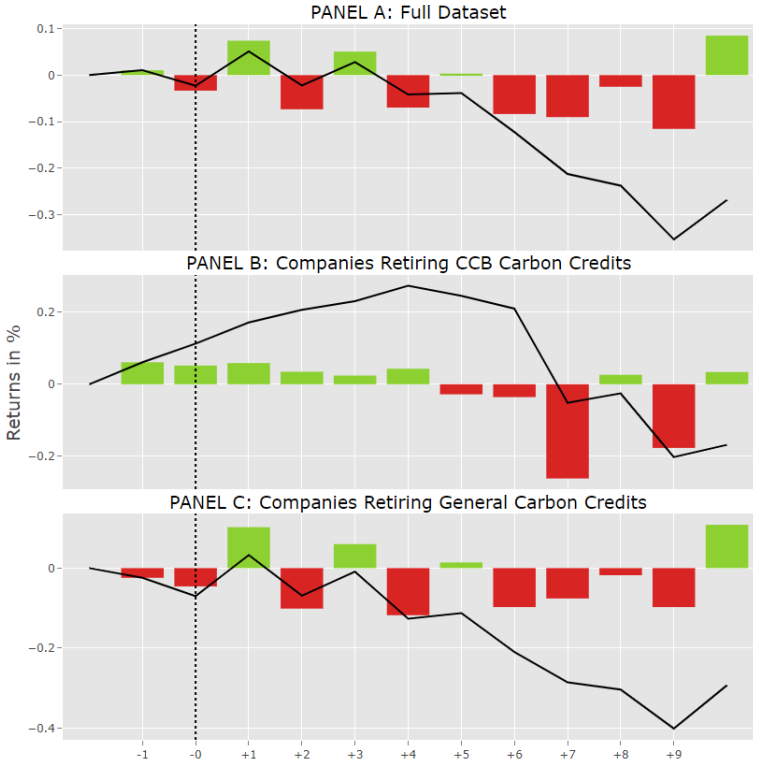
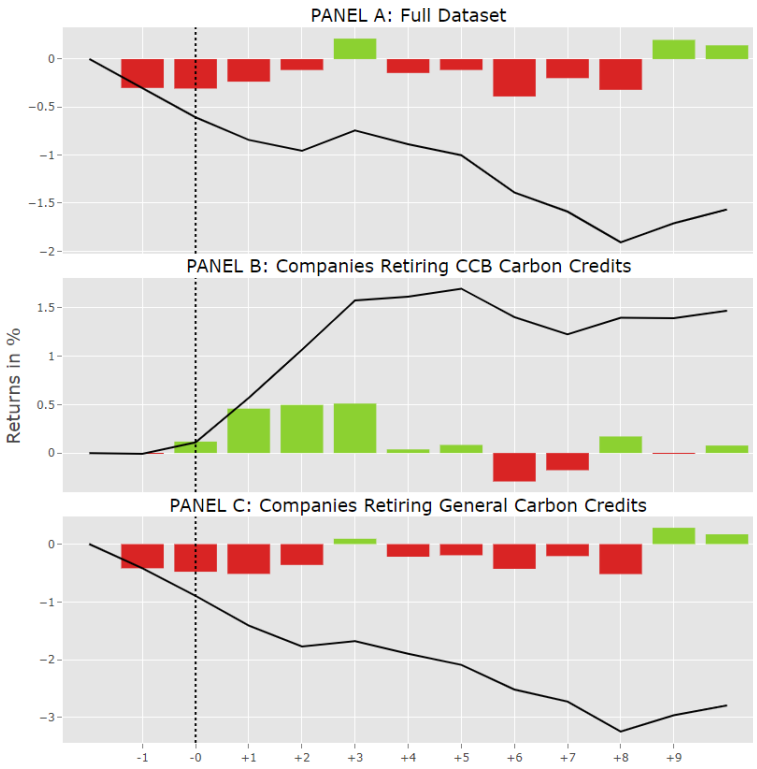


Figure B.4
Results for the entire sample, South and Central America area

This figure shows the abnormal and cumulative abnormal returns for the full dataset (Panel A), companies retiring CCB-labeled carbon credits (Panel B), and companies retiring general carbon credits (Panel C). The sample used in this case is the South and Central American area for the overall period. The dotted line represents the event, i.e. the carbon credit retirement.



Synthetic Control Illusions in Climate Science: Methodological errors in the assessment of forest carbon projects

3.1 Introduction

The synthetic control method has been one of the most influential methodological innovations in causal inference over the past two decades, with first seminal contributions published in the American Economic Review by Abadie and Gardeazabal (2003) on the economic costs of the Basque conflict, and described by Athey and Imbens (2017) as "arguably the most important innovation in the policy evaluation literature in the last 15 years." SCM constructs a weighted combination of untreated units to approximate the counterfactual trajectory of a treated unit in the absence of intervention. Because the method is transparent and intuitive, it has rapidly diffused across economics, political science, epidemiology, and environmental policy.

Recent methodological advances have further expanded SCM applications, including the Augmented Synthetic Control Method (ASCM) (Ben-Michael et al., 2021), generalized synthetic controls, and synthetic difference-in-differences estimators (Arkhangelsky et al., 2021). At the same time, the methodological literature has increasingly emphasized the importance of donor pool selection, covariate balance, interpolation constraints, placebo validation, and sensitivity analysis for ensuring credible inference (Abadie et al., 2010); (Abadie, 2021). Despite these advances, comparatively little attention has been paid to the behavior of SCM in spatially dependent environmental systems characterized by measurement error, heterogeneous institutions, dynamic treatment effects, and spatial spillovers. Tropical deforestation represents an especially difficult setting for SCM because deforestation dynamics are influenced simultaneously by biophysical conditions, transport infrastructure, commodity markets, governance systems, legal institutions, and spatial leakage processes, many of which are either imperfectly observed or omitted entirely from available datasets.

This study contributes to the growing methodological literature on synthetic control methods in three ways. First, it demonstrates how donor pool misspecification and omitted spatial covariates generate unstable counterfactuals in environmental applications. Second, it illustrates how measurement error in remotely sensed outcome variables propagates through SCM estimation and validation procedures. Third, it highlights the importance of explicitly addressing spatial dependence, leakage, and spillover effects in conservation policy evaluation.

West et al. (2023) estimate the effectiveness of 24 REDD projects from six countries, representing around a quarter of projects registered under the Verified Carbon Standard (VCS). The core of their method is to match each REDD project area (PA) to synthetic control (SC) sites: circular proxies of the same size as

the PA that are assumed to undergo the same trend in deforestation as the PA would have, if the REDD project had not been implemented.

The area deforested in the PAs and SCs were estimated using a global scale forest loss dataset (Hansen et al., 2013), and West et al. attempted to quantify the carbon credits successfully delivered according to the difference between deforestation in the PAs and the SCs. They concluded that the majority of projects did not generate any carbon benefits, and of those that did, most issued more carbon credits than were justified. We argue that the implementation of SCM in West et al. violates several core principles established in the canonical synthetic control literature, raising serious concerns about the credibility of their causal estimates. Specifically, the SCM estimator as implemented suffers from three compounding problems:

1. The satellite-derived data (Hansen et al., 2013) they use in all analyses has errors and biases that its creators state make it inappropriate for this type of comparison. These led to effective projects being incorrectly labeled as ineffective; and a potential underestimate of avoided deforestation of 90% or more.

2. Their project-level SC and country-level Generalized Synthetic Control (GSC) methods failed to select appropriate donor areas, with the SC sets failing common tests in 24 of 30 cases. The resulting models are also highly unstable, with fine-tuning of the method causing wildly different results. This is likely due to omitting key variables, such as 'distance to roads'.

3. West et al. made two different calculation errors, that mean the proportion of credits they report that deliver real carbon benefits should be increased by 62%.

West et al. further did not attempt to calculate the uncertainty on their estimates of carbon credits generated, nor did they adopt a formal hypothesis testing framework. These weaknesses have led to their results appearing more precise than they are, and may have led to a widespread misinterpretation that REDD projects are worthless (Greenfield, 2023). This is compounded by extending their conclusions beyond their dataset: they generalize about REDD success on the basis of a non-random sample of 24 projects from a potential dataset of 103 registered Verra REDD projects, itself a subset of 377 known REDD projects (Atmadja et al., 2022).

Details on all these points follow below.

3.2 Misuse of Global Forest Change Data

West et al. based their analysis on the 'Global Forest Change' (GFC) dataset, a product designed for use as a "stratified in targeting forest extent and/or change by a probability sample" (Hansen et al., 2013),

e.g. as used by Tyukavina et al. (2015). The website where the data are accessed explicitly states that "definitive area estimation should not be made using pixel counts from the forest loss layers", and that the data does not meet standards for areal land-cover change estimation set by the Intergovernmental Panel on Climate Change. However, West et al. did not adopt a point-based sample approach, instead using pixel counts directly as area estimates. The use of pixel counts from the GFC dataset as a direct measure of deforestation area introduces systematic measurement error into the outcome variable, which propagates through subsequent SCM estimation. Due to the nature of their comparisons, between REDD PAs and either control sites or REDD PAs before their start date, this choice substantially inflates uncertainty in the estimated treatment effects and reduces the probability that a successful project will be correctly identified as such.

The GFC data should not be used in this way because they contain false and missed detections that combine to reduce the apparent success of all projects. These errors can be large: for example, a study using field plots and Synthetic Aperture Radar data covering drier countries in sub-Saharan Africa, including all of Tanzania and Zambia, and the southern portion of DRC (3 of the 6 countries covered by West et al.) found the GFC data overestimated forest cover by a factor of two, but then underestimated forest loss rates by a factor of five, with these overall rates of error varying greatly by region and starting canopy cover (McNicol et al., 2018). We calculate that in this region this error rate would lead to an 89% understatement of project carbon benefits, meaning a project that stopped all deforestation (100% successful) would appear to have only stopped 11% if the GFC data were used to do the assessment. The African regional analysis in West et al. found an overall effectiveness of 16% (414 ha year-1 or 423 ha year-1 effect size, depending on method, compared to 2700 ha year-1 baseline); the error rate in the GFC data used suggests this could really represent an effectiveness of well over 100%.

We calculate this based on assuming the detection rate (R_d = proportion of true deforestation detected) and false alarm rate (R_f = proportion of stable forest marked as deforestation) were the same in the PA and the SC, then the area of avoided deforestation would be underestimated by a factor $(R_d - R_f)$. To understand this, consider two sites with the same area A, of which D_i is deforested: the predicted area of deforestation in site i will be:

$$\hat{D}_i = R_f(A - D_i) + R_d D_i \tag{1}$$

The true difference in the areas of deforestation ($\Delta D_{1,2}$) predicted between sites 1 and 2 would therefore equal:

$$\Delta D_{1,2} = (D_1 - D_2)(R_d - R_f) \tag{2}$$

For study (McNicol et al., 2018) $R_d = 0.19$ and $R_f = 0.08$ leads to an 89% underestimate of effectiveness.

But even using the estimates for wet tropical regions by the providers of the GFC data (Hansen et al., 2013), $R_d = 0.83$ and $R_f = 0.002$, leading to underestimation of real avoided deforestation by 17%. West et al. (2023) could have corrected for this best-case scenario by increasing their observed effectiveness in tables and figures by 17%, and recognizing that the impact could be bigger in many areas.

These errors are further compounded because the satellites used and methodology applied in the GFC data have changed significantly during the time period analyzed by West et al. In particular a change in method and the use of Landsat 8 caused a big increase in detections post 2012, with particularly increased sensitivity to degradation and small-scale deforestation (Pendrill et al., 2022). This will have a spatially heterogeneous impact that may only become apparent while projects are in operation, but not during the training period of selecting Synthetic Controls. For example two sites, one with small-scale degradation and the other with no disturbance, might appear identical in GFC from 2000-2012, but diverge from 2013 onwards when the small-scale degradation starts being detected. This step change in sensitivity could significantly bias all the analyses in West et al.

Beyond data quality, which is itself sufficient to cast serious doubt on the results, the causal identification strategy introduces additional and independent sources of bias, as we demonstrate in the following section.

3.3 Invalid Implementation of Synthetic Control Method

The identification assumptions underlying SCM, shared latent factor structures, donor comparability, convex hull feasibility, and pre-treatment parallel trends, require careful validation in any application. We demonstrate that these assumptions are systematically violated in West et al.'s implementation, producing unreliable counterfactual estimates. Beyond these methodological concerns, the broader lessons for applying SCM in spatial environmental systems with noisy remotely sensed outcomes make this analysis an important contribution to the literature on causal inference in conservation policy evaluation.

Leaving aside the quality of the underlying data, though it is itself sufficient to invalidate the results, we conducted an empirical evaluation of the synthetic control method (SCM) employed by West et al. SCMs were designed as a tool to understand the effects of state policy interventions (Abadie and Gardeazabal, 2003), and in a conservation context it is vital that steps are taken to account for domain specific complexity, select appropriate covariates, and test that valid controls have been obtained (Schleicher et al., 2020). Here, we show that the approach taken by West et al. leads to inappropriate matching, ignores key drivers of deforestation, and that in most cases (24 of 30) the SCs set selected for PAs fail to pass standard validation tests.

3.3.1 Synthetic Controls Poorly Constrained

Canonical SCM practice requires donor units to share the latent factor structure of the treated unit, and for treated units to lie within the convex hull of donor unit characteristics (Abadie, 2021). West et al. draw potential donor areas from randomly generated circles spanning entire countries, violating these comparability assumptions. Furthermore, the covariates used to generate synthetic controls lack critical variables: 'distance to roads', identified by Busch and Ferretti-Gallon (2023) as the single most important driver of deforestation risk, is entirely absent. In fact, only two of the top five predictors are included; 'Population' and 'Wealth' are also omitted. The resulting SCM estimator is therefore both biased and unstable: omitting important predictors increases bias because the synthetic control fails to reproduce the true counterfactual trajectory, and increases variance because the optimizer must overfit remaining covariates to compensate for missing information (Ben-Michael et al., 2021); (Abadie, 2021). Drawing donor areas from across entire countries generates donor units that violate comparability assumptions in fundamental ways. For example, donor units for two PAs in Peru and all Colombian PAs are drawn from sites on the opposite side of the Andes (West et al. Fig. S4), where biomes, market access, disease ecology, and agricultural systems are structurally different, all of which directly determine deforestation dynamics.

Elsewhere, West et al. make comparisons across huge distances, not least in the DRC, a country with a similar land area to western Europe, and corresponding huge variance in modes and types of agricultural production, transport networks, access to markets, laws and culture. A simple comparison to satellite data shows that only 1 of the 4 selected donor areas for site DRC 934 is in a closed canopy forest like the PA, with the other three in the savanna woodland ecosystem of southern DRC (Fig. 1). Even the single matched rainforest site is located over 600 km from the PA, at a much greater travel distance (multiple weeks by boat, the dominant means of transport in this region of DRC) from Kinshasa and its 17 million people.

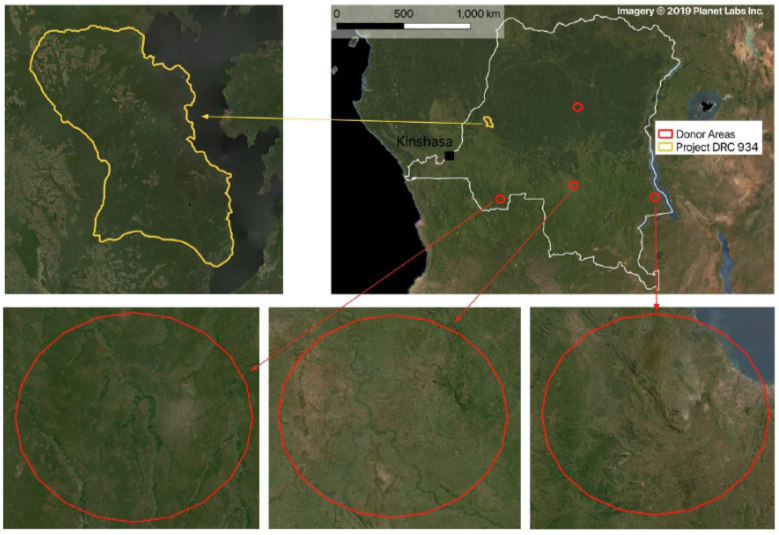
Even more concerning is that the chosen SCs for site DRC 934 all have 0% 'Logging concession cover', whereas the project has 100% 'Logging Concession Cover (West et al. Table A17). This variable was matched for, but is not a strict condition, so SCs were allowed if they fitted other variables better even if they failed entirely to match this condition. However, forests located in or out of a logging concession likely have a totally different risk profile: the REDD project activity here was to prevent a logging concession becoming active.

This poor quality of matching is systematic: every PA in Peru that showed more deforestation than its SC was matched to an SC with less timber concession coverage; every Colombian PA that showed more deforestation than its SC was matched to an SC with a higher level of pre-existing legal protection; and over

80% of PAs were matched to an SC that had lower levels of deforestation in a 10 km buffer zone around it. Given these biases, a lack of significant difference in deforestation rates between PAs and SCs after project implementation is not surprising, and indeed could even itself indicate project success. Theoretical work on synthetic control methods demonstrates that omitting important predictors of the outcome increases both bias and variance in treatment effect estimates (Ben-Michael et al., 2021); (Abadie, 2021). The bias arises because the synthetic control fails to reproduce the counterfactual trajectory of the treated unit, while the variance increases because the weights must overfit to remaining covariates to compensate for missing information. Both effects work against finding true treatment effects.

Figure 1
Project area (PA) DRC 934 and the four donor areas used by West et al. (1) to generate its synthetic control (SC)

The white outline in the overview panel shows the border of the Democratic Republic of the Congo (DRC), while the zooms show the Project Area (in dense rainforest with small patches of savanna), and in the second row three of the four selected Synthetic Controls, where savanna dominates. The satellite basemap is from December 2019, with data provided by Planet and accessed through Norway's International Climate and Forest Initiative (NICFI).



Recent advances in synthetic control methods, particularly the Augmented Synthetic Control Method (ASCM), are designed specifically to reduce sensitivity to covariate choice and improve robustness (Ben-Michael et al., 2021). ASCM incorporates an outcome model that can partially compensate for imperfect covariate balance, providing more reliable estimates even when some predictors are omitted or poorly matched. Our reanalysis using ASCM (Supplementary Materials) shows that while the fundamental problems with West et al.'s donor pool selection remain, the ASCM approach does improve pre-treatment fit for some projects, suggesting that methodological refinements could yield more credible results.

We emphasize that the appropriate response to covariate imbalance is not to proceed with analysis using

poorly matched controls, but rather to refine the donor pool, expand the set of matching variables, or adopt alternative methods. The Verra methodology for REDD+ baseline setting, while imperfect, does prioritize geographic proximity and similar forest characteristics precisely to avoid the kinds of incomparable matches we document in West et al.'s analysis.

3.3.2 Methodological concerns with the synthetic control approach

The use of randomly generated circular polygons as donor units raises fundamental concerns about the identifying assumptions underlying the synthetic control estimator. Random circular polygons possess no institutional identity, no land tenure regime, no governance framework, no community engagement structure, and therefore cannot plausibly share the latent factor structure that determines deforestation trajectories in REDD+ project sites.

The validity of the synthetic control as a counterfactual depends critically on the assumption that donor units share the relevant factor structure with treated units, that is, donors must be drawn from a population where the same unobserved factors govern outcome trajectories.

Random circular polygons cannot plausibly satisfy this requirement. REDD+ project sites are characterized by specific institutional arrangements, land tenure regimes, community engagement structures, and governance frameworks that jointly determine deforestation trajectories. These factors are entirely absent from arbitrary circular areas, which possess no institutional identity whatsoever. Consequently, even if pre-treatment deforestation trends appear similar between REDD+ sites and their synthetic controls, this similarity may be coincidental rather than reflecting shared underlying factor structures.

Furthermore, the synthetic control method requires that treated units lie within the convex hull of donor unit characteristics (Abadie, 2021). This constraint ensures that the synthetic control represents an interpolation rather than an extrapolation from the donor pool. With randomly generated circular polygons, there is no guarantee that REDD+ sites satisfy this constraint for the institutional and governance variables that are omitted from the analysis but likely determine deforestation outcomes. West et al. match on biophysical and deforestation covariates but cannot match on unobserved institutional factors because their donor pools lack such characteristics entirely. This represents a fundamental limitation that cannot be remedied through more sophisticated weighting or additional observed covariates.

3.3.3 Temporal Dynamics and Time-Varying Treatment Effects

An implicit assumption in West et al.'s analysis is that REDD+ project effects are constant over time, that projects either work or don't work, with a fixed treatment effect throughout the implementation period. This assumption is both unrealistic and potentially misleading, as conservation interventions typically

exhibit time-varying effects that evolve as projects mature, funding changes, and local conditions shift. Empirical evidence supports these theoretical expectations. Studies of protected areas show that conservation benefits often increase over time as enforcement becomes more effective and community buy-in strengthens (McNicol et al., 2023). Conversely, some projects show declining effectiveness if initial enthusiasm wanes or if deforestation pressure simply shifts to later time periods (a form of temporal leakage). Averaging effects over multi-year implementation periods, as West et al. do, obscures these dynamics and may misclassify projects that are improving over time as failures, or projects experiencing early success but later problems as successes.

3.3.4 Spatial Spillovers, Leakage, and the Stable Unit Treatment Value Assumption

Synthetic control methods rely on the Stable Unit Treatment Value Assumption (SUTVA), which requires that the treatment applied to one unit does not affect outcomes in other units (Abadie et al., 2010); (Abadie, 2021). This assumption is frequently violated in spatial contexts where units are geographically proximate and connected through markets, migration, or ecological processes. For REDD+ projects, SUTVA violations can arise through two primary mechanisms: leakage (deforestation displaced from project areas to surrounding regions) and positive spillovers (project activities reducing deforestation in nearby non-project areas).

Leakage is a well-documented concern in forest conservation (Busch and Ferretti-Gallon, 2023). If a REDD+ project prevents logging or agriculture in one location, economic incentives may simply shift those activities to nearby unprotected forests, resulting in zero net conservation benefit despite apparent project success. The magnitude of leakage depends on the mobility of deforestation drivers: highly mobile logging operations can easily relocate, while smallholder agriculture tied to specific communities may exhibit less leakage. Verra methodologies require projects to define leakage belts and monitor for displaced deforestation, but these procedures assume leakage occurs in predictable nearby zones rather than through complex regional dynamics.

West et al.'s analysis ignores leakage entirely, treating project areas as if they exist in isolation. This is problematic because if leakage occurs to areas within their donor pools, those control areas become contaminated, they are no longer valid counterfactuals because they have been affected by the treatment. Contaminated controls will show artificially high deforestation, making projects appear more effective than they truly are when accounting for total landscape-level impacts.

Conversely, positive spillovers can occur if REDD+ projects create incentives or capacity that reduces deforestation beyond project boundaries. For example, alternative livelihood programs might benefit neighboring communities, or enhanced law enforcement might deter illegal activities in surrounding ar-

eas. Such spillovers violate SUTVA in the opposite direction, potentially causing project effectiveness to be understated if control areas experience spillover benefits.

The spatial scale of West et al.'s donor pools exacerbates these concerns. Drawing controls from across entire countries means some donor areas may be hundreds or thousands of kilometers from project sites, far enough to avoid spillovers but also far enough to be fundamentally incomparable in terms of deforestation drivers. Closer controls would better satisfy the comparability requirement but increase the risk of SUTVA violations. This tradeoff between comparability and independence is fundamental to spatial causal inference and cannot be resolved without explicit modeling of spatial processes.

Recent methodological developments offer potential solutions. Spatial synthetic control methods explicitly model spillover effects and leakage, allowing for valid inference even when SUTVA is violated. These methods require additional assumptions about the functional form of spatial dependence but can provide unbiased estimates of both direct effects (within project areas) and indirect effects (in surrounding regions). Alternatively, pre-specifying leakage buffers and treating them separately from the donor pool, as Verra methodologies require, provides a transparent approach to leakage quantification, though it assumes the buffer zones are correctly defined.

3.3.5 Unreliable Validation Tests Used

In their Figure S6, West et al. show a "validation" in which they divide the pre-treatment period into two parts (training and validation periods). If the method is reliable, then the SC and the PA should show parallel deforestation trends in the validation period, but this requirement appears to be violated. The authors claim that 26 out of 30 PAs pass the validation by setting a threshold in terms of absolute deforestation by the final year, equal to a deviation of less than 0.5% of the total project area. Replicating this same test, we find that 24 of the PAs pass using normal SCM, and 25 pass using an Augmented synthetic control method (ASCM) proposed by Ben-Michael et al. (2021).

However, this "validation" cannot be considered a rigorous test. The absolute size of error allowed is very large, especially for PAs with low deforestation rates (10 of the 32 PAs had pre-project total deforestation less than 0.5%, meaning a deviation of over 100% is allowed in this test).

The credibility of an SC estimator depends on its ability to track the trajectory of the outcome variable (in this case deforestation d) for an extended period in time t (Abadie, 2021), so a better numerical test is given by:

$$\frac{\max\left(|d_{PA}(t) - d_{SC}(t)|\right)}{d_{PA}(t_{final})} < 0.2 \tag{3}$$

This threshold ensures the SC can track deforestation dynamics with sufficient precision to detect treatment effects of policy-relevant magnitude. We derive this threshold empirically: validation deviations

exceeding 20% are systematically associated with matching failures on key covariates.

In addition, following best practice (Abadie et al., 2010) we applied a second test:

$$\frac{\text{RMSPE(validation)}}{\text{RMSPE(training)}} < 5 \tag{4}$$

The canonical SC literature suggests exclusion thresholds ranging from 2 to 10. We adopt 5 as a conservative standard appropriate for environmental applications with noisy outcome data.

We find that only 7 out of 30 PAs pass the requirement in Equation 3, and that these results are also more consistent with the graphical trends: only PAs that exhibit a parallel trend with the SC pass the test (Fig S2). For example, Project 1900 (Fig 2a) fails our test, despite passing that in West et al. We find that only 14 PAs pass the requirement of Equation 4, and suggest that if the authors truly wish to apply an "abundance of caution" (West et al., 2023), then they should apply both tests, which 24 out of 30 PAs would fail. Our results show the same pattern when using ASCM: 6 PAs pass both validation tests, 4 of which also passed the tests using SCM (see supplementary materials).

To illustrate the sensitivity of validation outcomes to the choice of threshold, we computed the number of projects that would pass the RMSPE ratio test (Equation 4) across a range of threshold values from 1 to 20, using both SCM and ASCM (Fig. S1). This analysis reveals that at the threshold of 2, only 4 projects pass under SCM and only 2 under ASCM. Even at the relatively permissive threshold of 5 that we adopt, fewer than half of the projects pass: 14 under SCM and 13 under ASCM. The two methods yield broadly similar results across the threshold range, with SCM marginally outperforming ASCM at lower thresholds and the pattern reversing at higher thresholds. Notably, even at a threshold of 10, which would represent an extremely lenient validation criterion, approximately one-third of projects still fail under both methods. This threshold sensitivity analysis underscores the fragility of the synthetic control approach as implemented by West et al., and demonstrates that their validation procedure does not provide a robust basis for causal inference about project effectiveness.

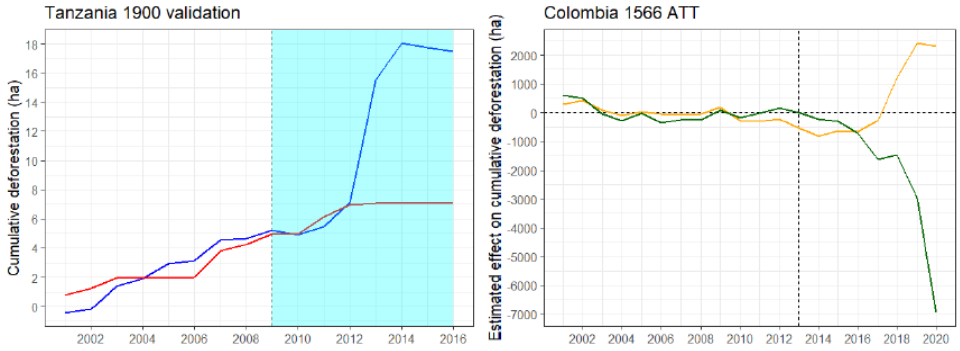
Given this failure rate (and arguably even their stated failure rate of over 10% of PAs), it should have been concluded that their SC model was not reliable and required adjustments to its input datasets, rather than continuing with the analysis. It is likely that improving the number of variables known to be important for deforestation, using a more reliable remote sensing datasets, and putting in hard controls related to biomes or land use types available for SC matches, would have improved these results.

3.3.6 Problems with the generalized synthetic control analysis

The generalized synthetic control (GSC) analysis pools projects across vastly heterogeneous contexts to estimate average treatment effects on the treated (ATT). While aggregation can improve statistical preci-

Figure 2
Examples of validation and average treatment effect (ATT) sensitivity

a) cumulative deforestation in hectares for a REDD project area (PA) in Tanzania (project 1900, red line) and the average of its synthetic controls (SCs) over a training period and subsequent validation period (blue). Project 1900 in Tanzania passes the "validation" test in West et al. despite a clear large divergence over the validation period. This figure replicates West et al. Figure S6, where it is obscured by excessive range in the y-axis. b) difference in cumulative deforestation between PAs and SCs using an initial filter that prioritizes donor areas with similar buffer deforestation (green line) and without the filter (yellow line) for a project in Colombia (1566) that passes all validation steps. The REDD project starts in 2013 (dotted grey line). Datasets for all projects shown in Figures Supplementary Information.



sion, it raises interpretive challenges when treatment effects are plausibly heterogeneous. ATT represents the average effect across all treated units. However, the REDD+ projects in the sample vary enormously in scale, baseline deforestation pressure, institutional context, implementing organizations, and intervention strategies. Averaging across such heterogeneity produces an estimate that may not characterize any individual project or policy context.

More critically, a null or small average effect is consistent with multiple underlying realities: all projects having near-zero effects, some projects succeeding and others failing with effects that cancel on average, or projects having effects that vary systematically with observable characteristics. The policy implications differ dramatically across these scenarios. If some project types or contexts yield substantial deforestation reductions while others do not, the appropriate policy response is to identify and scale successful approaches, not to conclude that REDD+ is ineffective.

(West et al., 2023) acknowledge heterogeneity by presenting individual project results alongside the GSC estimates. However, their primary conclusions emphasize the pooled findings, stating that "most projects have not significantly reduced deforestation." This framing obscures the potential for meaningful heterogeneity. A more informative analysis would systematically examine effect modification by project characteristics, implementation features, and contextual factors, moving beyond the question of whether REDD+ works on average to identify conditions under which it works.

3.4 Sensitivity of Results

Due to the flaws in the SCM approach discussed above, the results presented by West et al. are not robust to small changes in the algorithm used to generate SCs. This robustness testing is a common feature of SCM approaches, with reference 9 for example varying their models in many ways to see how the results change. To demonstrate this, we replicated their method with and without the application of a filter they used that prioritizes donor areas with similar rates of buffer deforestation. If the method were robust, we would expect the filter to have some effect, but for the overall results to be similar without it given that buffer deforestation remains one of the covariates.

In fact, removing the filtering step leads to reversal of outcomes for 7 of the PAs. For example, Colombia 1566 has 7,000 ha less deforestation than its SC when the filter is used, but 2,500 ha more deforestation than its SC when the filter is removed. On average, these numbers (which West et al. carry forward into their calculation of carbon credits), differ by over 200%. Such discrepancies indicate that the SCM estimator lacks the pre-treatment fit and stability required for credible causal inference, and that specification choices, rather than underlying project effectiveness, are driving the reported results.

This instability in results is seen to some degree in West et al's analysis. Their reported effectiveness for the same projects, using the same datasets, varies widely: for example overall effectiveness of all African projects covered varies from 0% (project-level SC analysis) to 16% (regional GSC) to 50% (regional GSC using cardinality matching by year 5, West et al. Table S12).

3.5 Calculation Errors

The authors make two further errors. Firstly, they compare their estimates of avoided deforestation to pre-project ex ante estimates of carbon credits, rather than to the lower ex post issuance figures that reflect actual verified performance. This specification error ignores the fact that carbon credits are actually issued ex post following an independently verified assessment of performance, and the number is normally lower than the original estimate due to reductions for leakage and residual deforestation.

In reality, the projects assessed by West et al., as stated in West et al. Table S4, issued 64,966,812 credits, but in their analysis West et al. compare to the 88,951,394 total in the ex ante baselines of Table S1 (note Table S4 records 0 issuances for project 2278, but we added 3,002,760 issued credits as now stated for the project on the Verra registry). As a result, West et al. understate the proportion of carbon credits that are 'real' according to their own analysis by 37%.

Secondly, when converting from avoided deforestation to carbon offsets (West et al. Table S1), the wrong numbers for "project baseline deforestation" are used. In all cases the deforestation between the year 2000 and the start of the project has been added to the baseline deforestation that occurred after the

project started. This has the effect of inflating the baseline deforestation by a mean of 18.0% (range 0 - 91% with bigger disparities for more recent projects), and means that the average project was given an "expected carbon offset per hectare of avoided deforestation" value of 207 Mg CO2 ha-1, compared to a true value of 268 Mg CO2 ha-1 (Table S1). The exact magnitude of the error this causes varies by the set of projects West et al. consider to be successful in their different analyses, but we estimate it understates the carbon benefits of avoided deforestation by between 18 and 35%.

These two errors are independent, and should be combined to understand the overall impact of these calculation errors. If this is done with West et al.'s individual project based analysis, the 'proportion of credits that deliver real carbon benefits' should be increased by 61.6% (Table S2), purely due to calculation missteps. While this appears to have only a small impact on West et al.'s estimate of the overall effectiveness of the REDD projects in their analysis (an increase in effectiveness from 6.1% to 9.8% is still far from 100%), when combined with uncertainties resulting from the input satellite data and poorly selected comparison sites, all of which likely decreased the observed estimated impact of a truly impactful project, they are significant.

3.6 Sample size

West et al. extend their conclusions from the 24 REDD projects they study to all REDD projects. But there are 103 registered Verra REDD projects, so they cover only 23%. Further these 24 are only 6% of the 377 REDD projects identified in the study by Atmadja et al. (2022), that West et al. cite. Furthermore, their sample is not random: projects were included only from pre-selected countries. Generalizing from a non-random sample of 24 projects to the broader universe of REDD+ interventions is statistically unsupported; valid inference should be limited to projects in the six countries covered.

3.7 Conclusion

West et al. conclude that "only a minority of projects achieved statistically significant reductions in comparison with ex post counterfactuals", but our analysis demonstrates that the SCM implementation in West et al. (2023) systematically violates core assumptions of the synthetic control estimator, including donor pool comparability, covariate completeness, and the Stable Unit Treatment Value Assumption. The resulting estimator lacks the pre-treatment fit and stability required for credible causal inference. Combined with systematic measurement error in the outcome variable and two specification and calculation errors, each of which independently biases results against finding project success, these compounding problems mean that the substantive conclusions about REDD+ effectiveness cannot be reliably supported

by the methods as implemented.

More broadly, this analysis highlights the fragility of SCM-based causal claims in spatial environmental applications. When deforestation dynamics are governed by unobserved institutional factors, spatially autocorrelated processes, and measurement error in remotely sensed outcomes, SCM estimates become highly sensitive to specification choices, a pattern likely to recur in future applications of quasi-experimental methods to conservation policy evaluation. We note that all project areas studied have low rates of deforestation (average 0.21%/yr according to West et al. figures, compared to 0.47%/yr for forests across the six countries covered, despite often being located in higher-risk areas, a pattern consistent with project effectiveness that the methods employed cannot reliably detect.

Future applications of SCM and related quasi-experimental methods in conservation policy evaluation should satisfy four requirements before causal conclusions are drawn. First, donor pool comparability must be validated using both observed covariate diagnostics and pre-treatment parallel trend tests, with hard constraints on geographic proximity and ecological similarity. Second, outcome data must have well-characterized measurement properties appropriate to the inferential task; use of pixel counts from global forest change datasets as direct area estimates does not meet this standard without design-based correction. Third, spatial leakage and spillover effects must be explicitly modeled or accounted for, rather than ignored. Fourth, pre-specified sensitivity analyses across plausible specification choices must be reported before conclusions are drawn. Specifically, we call for future studies on REDD+ effectiveness to use locally tuned forest change data with known accuracies or point-based sampling approaches to quantify deforestation. Pixel-based matching methods (McNicol et al., 2023); (Guizar-Coutiño et al., 2022) may be more appropriate given the volatile nature of deforestation as an outcome variable. Finally, we emphasize that broad conclusions about avoided deforestation projects as a whole require not only methodological rigor, but also a formally representative sample of projects and an explicit hypothesis testing framework.

Appendix C

Validity of Synthetic Control approach

We conducted an empirical evaluation of the synthetic control method (SCM) employed by West et al. (2023). Our approach and the methodology are as close as possible, given the example scripts provided by the authors did not specify the years used to create training and validation periods or the exact cutoffs used for filtering potential donor areas by buffer deforestation. The analysis was conducted with the Synth and augsynth packages available for R software.

We replicate the results in West et al. Figure S6 (our Figure C1) and Table S3 (Table C3) using the dataset made available by West et al. in each REDD+ project using both the SCM (that they adopt) and the ASCM. The simple SCM used in their analysis does not allow negative weights while ASCM relaxes this constraint allowing for negativity and extrapolation. ASCM is particularly useful for improving the covariate balance between treated and control groups and often leads to a better pre-treatment fit (Busch and Ferretti-Gallon, 2023). Weights are chosen to balance pre-treatment outcomes and other covariates as in the SCM.

The first step when applying a SCM is to test the validity of the methodology. The approach consists of dividing the pre-treatment period in two parts (training and validation period) using a random year. This step is shown in fig. C3. We then apply three numerical tests to determine the quality of the fits in the validation period between project areas (PAs) and synthetic controls (SCs). In table C3 we replicate the same test used by West et al. We calculated the difference between project and synthetic control deforestation in the final year of the validation period and then we compared this number with the total project area. This is done for both the ASCM and the SCM.

In table C4 we applied a variant of the validation test proposed by (Abadie and Gardeazabal, 2003). This test is based on the ratio between the root mean squared prediction error (RMSPE) for the validation period and for the training period. The smaller the value, the better is the validity of the SC. Based on experimentation and following guidance in the literature, we set 5 as a limit value for this ratio; if the post RMSPE / Pre RMSPE ratio is bigger than 5 the project fails the validation. The benefit of using this methodology is that it is not biased by the quality of the fit in the training period.

In addition, we proposed a second more reasonable numerical test using the ratio between the maximum of the absolute value differences in the validation period and total deforestation. This approach not only allows us to capture the divergence between project and SCs across the entire validation period, but also eliminates the bias of preference of low deforestation areas. Empirically we choose 20% (of total deforestation in the last year of the validation period) as a threshold to pass the validation test, as this seemed to keep only projects where by eye the match appeared strong in the validation period. The results are

shown in table C5. We continued to use ASCM as this algorithm led to better validation results and therefore should be more robust.

Addressing the second concern around the sensitivity of SC results, we ran the analysis both with and without the filter applied by West et al. This filter selectively includes only those donors with deforestation patterns analogous to the project sites (ideally 10 percent deviation, but increasing this range to ensure an SC with a good match in the pre-implementation period is obtained). We used different percentage deviations (20 or 30 percent) for each country based on the ability of the SC to find enough controls for the analysis.

Figure C2 shows the cumulative deforestation of the PAs relative to the SCs created with and without the filter. We replicate West et al.'s figures CZ7 and S8 in our figures C3 and C4 (without filtering) and figures C5 and C6 (with filtering). Finally, in Table C6 we show the average treatment effects on the treated (ATT) for each project when the filter is applied and when it is not, which we defined as the mean difference between the deforestation in the PS and the SC over the implementation period.

Table C.1
Corrected carbon credits generated, estimated using synthetic control (SC) results

Replication of West et al. Table S5. The carbon offsets generated, according to their SC results, are calculated assuming a fixed value of carbon per hectare of avoided deforestation. This value is underestimated due to the incorrect addition of pre-implementation deforestation to the project baselines. Overall, this increases the number of credits generated by 18%. *The large disparity in Peru 844 is due to the wrong number for baseline deforestation being taken from the project documents.

	Avoided deforestation from SC results (ha)	Project baseline deforestation (ha)		Carbon credits expected (Mg CO_2)	Carbon offsets per hectare (Mg CO_2 ha^{-1})		Carbon offsets from SC results (Mg CO_2)	
Project		West et al.	Correct		West et al.	Correct baseline	West et al.	Correct baseline
Peru 1882	0	1,268	1,016	356,960	282	351	0	0
Peru 2278	0	13,299	12,581	5,891,253	443	468	0	0
Peru 1067	4,727	13,581	11,893	4,817,471	355	405	1,676,768	1,914,755
Peru 958	2,223	2,924	1,525	630,937	216	414	479,676	919,720
Peru 944	3,457	23,685	20,842	5,151,165	217	247	751,850	854,408
Peru 844	0	125,561*	21,983	12,475,134	99	567	0	0
Colombia 1400	0	12,425	12,157	1,657,098	133	136	0	0
Colombia 1566	0	103,908	94,345	31,325,923	301	332	0	0
Colombia 1396	2,992	8,076	7,417	1,489,786	184	201	551,937	600,976
Colombia 1395	1,440	16,835	15,944	2,791,723	166	175	238,793	252,138
Colombia 1392	117	10,722	10,633	1,455,141	136	137	15,879	16,012
Cambodia 904	0	21,252	20,142	1,626,420	77	81	0	0
Cambodia 1650	4,110	30,446	28,304	12,432,277	408	439	1,678,272	1,805,280
DRC 1359	0	11,949	10,030	4,735,361	396	472	0	0
Tanzania 1325	0	10,578	7,122	359,834	34	51	0	0
Tanzania 1900	0	11,407	11,400	348,019	31	31	0	0
Tanzania 1897	0	35,472	25,301	1,406,892	40	56	0	0
Sum							**5,393,174**	**6,363,289**

Table C.2
Corrected proportion of carbon credits linked to reduced deforestation using SC results

Replication of the result in West et al. that 6.1% of carbon offsets are linked to reduced deforestation. Two corrections are made: first, the correct values for the number of offsets generated per hectare are used (Table C.1); second, the comparison is to credits actually issued (values from West et al. Table S4) instead of expected credits (as used in West et al. Table S5). Combined, this increases the proportion of "real" carbon credits according to their SC analysis by 62%. *West et al. Table S4 records 0 issuances for project 2278, but 3,002,760 issued credits are added here as now stated for the project on the Verra registry.

	Carbon offsets from project baseline (Mg CO_2)		Carbon offsets from SC method (Mg CO_2)		Percentage of credits "real"	
Project	Expected	Issued	West et al.	Correct baseline	West et al.	Correct baseline
Peru 1882	356,960	171,673	0	0	0%	0%
Peru 2278	5,891,253	3,002,760*	0	0	0%	0%
Peru 1067	4,817,471	3,678,270	1,676,768	1,914,755	35%	52%
Peru 958	630,937	566,843	479,676	919,720	76%	>100%
Peru 944	5,151,165	5,282,313	751,850	854,408	15%	16%
Peru 844	12,475,134	9,658,069	0	0	0%	0%
Colombia 1400	1,657,098	544,278	0	0	0%	0%
Colombia 1566	31,325,923	22,274,745	0	0	0%	0%
Colombia 1396	1,489,786	567,286	551,937	600,976	37%	>100%
Colombia 1395	2,791,723	1,620,202	238,793	252,138	9%	16%
Colombia 1392	1,455,141	477,432	15,879	16,012	1%	3%
Cambodia 904	1,626,420	48,000	0	0	0%	0%
Cambodia 1650	12,432,277	14,568,314	1,677,170	1,805,280	13%	12%
DRC 1359	4,735,361	1,620,202	0	0	0%	0%
Tanzania 1325	359,834	10,000	0	0	0%	0%
Tanzania 1900	348,019	150,425	0	0	0%	0%
Tanzania 1897	1,406,892	726,000	0	0	0%	0%
Total	**88,951,394**	**64,966,812**	**5,392,073**	**6,363,289**	**6.06%**	**9.79%**

Table C.3
Replication of West et al. Table S3

Synthetic control "validation": the difference between project and synthetic control deforestation in the final year of the validation period. The difference in % is the % of the project area. Projects with a difference between project and synthetic control deforestation $> 0.5\%$ of the project area are assumed to have failed validation and are marked with an asterisk.

			ASCM		SCM	
Country	**Project ID**	**End of validation**	Difference (ha)	Difference (%)	Difference (ha)	Difference (%)
Cambodia	904	2007	−272.5	−0.4	−497.6	*−0.8
Cambodia	1650	2010	−3,122.7	*−1.6	−702.9	−0.4
Colombia	856	2011	20.8	0.2	54.4	0.4
Colombia	1389	2013	−1,994.8	*−2.9	−1,420.7	*−2.1
Colombia	1390	2014	127.7	0.1	348.0	0.3
Colombia	1391	2013	−102.5	−0.2	67.9	0.1
Colombia	1392	2013	−26.9	0.0	−60.5	−0.1
Colombia	1395	2013	257.9	0.3	61.0	0.1
Colombia	1396	2014	−36.9	−0.1	67.1	0.1
Colombia	1400	2013	−34.1	−0.1	195.8	0.3
Colombia	1566	2013	553.8	0.0	−842.7	0.0
DRC	934	2011	−2,052.4	*−0.7	56.3	0.0
DRC	1359	2009	−826.6	−0.4	−2,946.0	*−1.6
Peru	844	2009	−10.0	0.0	0.3	0.0
Peru	944	2009	275.3	0.2	368.6	0.2
Peru	958	2011	212.8	0.1	−7.5	0.0
Peru	985	2009	−1,008.8	−0.1	−26,682.6	*−2.0
Peru	1067	2011	−829.9	−0.1	−1,254.1	−0.2
Peru	1182	2013	−48.2	−0.1	−74.2	−0.1
Peru	2278	2018	−863.8	−0.5	−611.2	−0.3
Peru	1360-1	2010	−1,198.0	*−1.3	−6,706.3	*−7.5
Peru	1360-2	2010	−27.3	−0.1	−3.1	0.0
Peru	1360-3	2010	5.8	0.0	11.5	0.1
Tanzania	1325	2011	351.5	0.5	480.4	0.7*
Tanzania	1897	2017	4,685.5	*2.3	3,396.6	*1.7
Tanzania	1900	2016	−10.4	0.0	−260.6	−0.2
Zambia	1202	2009	−53.4	−0.1	−24.4	−0.1
Zambia	1775-1	2015	−3,147.8	−0.5	−2,446.0	−0.4
Zambia	1775-2	2015	−310.2	−0.2	290.5	0.2
Zambia	1775-3	2015	−252.4	−0.1	−177.2	−0.1

Table C.4
Post-RMSPE / pre-RMSPE ratios

Validation test based on the ratio between the root mean squared prediction error (RMSPE) for the validation period and for the training period. Projects with a post-RMSPE / pre-RMSPE ratio > 5 are assumed to have failed validation and are marked with an asterisk.

		ASCM	SCM
Country	**Project ID**	Post RMSPE / Pre RMSPE	Post RMSPE / Pre RMSPE
Cambodia	904	*11.0	3.8
Cambodia	1650	*7.8	1.4
Colombia	856	*5.5	*9.3
Colombia	1389	*203.4	*7.4
Colombia	1390	5.0	*44.0
Colombia	1391	4.6	*8.9
Colombia	1392	*6.4	*26.5
Colombia	1395	*17.1	*6.2
Colombia	1396	1.0	4.8
Colombia	1400	3.5	*45.7
Colombia	1566	1.5	4.2
DRC	934	*85.0	3.0
DRC	1359	4.3	3.2
Peru	844	*12.3	1.0
Peru	944	3.0	4.4
Peru	958	2.9	0.9
Peru	985	4.7	2.5
Peru	1067	2.2	3.1
Peru	1182	4.3	*6.0
Peru	2278	*13.1	*9.9
Peru	1360-1	*18.3	2.8
Peru	1360-2	4.1	1.5
Peru	1360-3	4.8	*6.0
Tanzania	1325	*9.8	*6.3
Tanzania	1897	*42.1	*25.6
Tanzania	1900	*7.8	*18.0
Zambia	1202	*5.6	3.3
Zambia	1775-1	*10.5	*13.3
Zambia	1775-2	*6.1	*11.3
Zambia	1775-3	*9.7	*27.6

Table C.5
Synthetic control validation

The maximum of the absolute value differences between project and synthetic control deforestation in each year of the validation period. The difference in % is the % of the project deforestation in the final year of the validation period. Projects with a percentage difference $> 20\%$ are assumed to have failed validation and are marked with an asterisk.

			ASCM		SCM	
Country	**Project ID**	**Years**	Difference (ha)	Difference (%)	Difference (ha)	Difference (%)
Cambodia	904	2004–2007	272.5	*26.4	497.6	*48.2
Cambodia	1650	2007–2010	3,122.7	*145.8	702.9	*32.8
Colombia	856	2006–2011	20.8	14.1	54.4	*36.8
Colombia	1389	2008–2013	1,994.8	*215.2	1,420.7	*153.2
Colombia	1390	2008–2014	127.7	*22.6	348.0	*61.5
Colombia	1391	2008–2013	102.5	*53.7	90.4	*47.3
Colombia	1392	2007–2013	26.9	*30.2	60.5	*67.9
Colombia	1395	2008–2013	276.0	*31.0	228.2	*25.6
Colombia	1396	2008–2014	36.9	5.6	69.1	10.5
Colombia	1400	2007–2013	42.9	16.0	195.8	*72.9
Colombia	1566	2007–2013	553.8	5.8	851.5	8.9
DRC	934	2006–2011	2,052.5	*24.2	667.6	7.9
DRC	1359	2006–2009	826.6	*43.1	2,946.0	*153.5
Peru	844	2005–2009	10.0	16.6	5.6	9.3
Peru	944	2005–2009	326.6	11.5	507.9	17.9
Peru	958	2006–2011	212.8	12.2	109.5	6.3
Peru	985	2006–2009	1,723.0	*77.5	26,682.6	*1,199.6
Peru	1067	2007–2011	882.4	*52.3	1,254.1	*74.3
Peru	1182	2007–2013	48.2	19.1	74.2	*29.4
Peru	2278	2011–2018	863.8	*120.2	611.2	*85.1
Peru	1360-1	2007–2010	1,198.0	*328.4	6,706.3	*1,838.3
Peru	1360-2	2007–2010	27.3	*86.4	14.1	*44.6
Peru	1360-3	2006–2010	22.2	*30.1	21.2	*28.8
Tanzania	1325	2006–2011	468.0	13.5	627.5	18.2
Tanzania	1897	2010–2017	4,685.5	*46.1	3,396.6	*33.4
Tanzania	1900	2009–2016	11.0	*155.4	260.6	*3,691.6
Zambia	1202	2006–2009	53.4	*131.4	36.8	*90.6
Zambia	1775-1	2009–2015	3,147.8	*395.8	2,329.0	*292.8
Zambia	1775-2	2009–2015	310.2	*67.9	290.5	*63.6
Zambia	1775-3	2009–2015	252.4	*159.9	177.2	*112.2

Table C.6
Averages of the average treatment effects on the treated (ATT)

Differences between the average ATT for each project when the filter is applied and when it is not. The ATT is calculated as the mean difference between the deforestation in the project area and the synthetic control deforestation over the implementation period.

Country	Project ID	Without filter	With filter	Difference (%)
Cambodia	1650	−9,375.14	−17,464.24	46%
Cambodia	904	−1,385.98	−5,709.93	76%
Colombia	1392	−49.20	−15.37	−220%
Colombia	1566	504.60	−2,035.37	125%
Colombia	856	−3.22	76.51	104%
Colombia	1391	−188.37	−98.99	−90%
Colombia	1396	−464.40	−404.86	−15%
Colombia	1400	−353.82	−364.89	3%
Colombia	1389	−4,506.85	−3,019.10	−49%
Colombia	1395	−582.62	−820.55	29%
Colombia	1390	−107.12	182.65	159%
DRC	1359	837.89	−3,169.86	126%
DRC	934	−1,261.76	379.42	433%
Peru	1360-3	−94.74	−132.38	29%
Peru	844	−697.61	−35.25	−1,879%
Peru	1067	−1,545.24	−6,432.88	76%
Peru	944	−2,541.10	−2,922.78	13%
Peru	985	−9,873.20	−108,971.30	91%
Peru	958	−450.49	−854.08	47%
Peru	1360-2	−122.58	−274.98	55%
Peru	2278	−567.84	−506.51	−12%
Peru	1360-1	−28.20	−73.12	61%
Peru	1182	−99.55	−125.87	21%
Tanzania	1325	1,604.99	47.46	−3,282%
Tanzania	1897	1,794.07	1,985.91	10%
Tanzania	1900	−0.01	0.19	105%
Zambia	1775-1	−4,121.52	−4,337.38	5%
Zambia	1775-2	−72.75	328.68	122%
Zambia	1775-3	−236.41	−316.88	25%
Zambia	1202	−62.27	−1,107.55	94%

Figure C.1
Sensitivity of validation test outcomes to RMSPE ratio threshold

Number of REDD+ projects passing the RMSPE ratio validation test (Equation 4) as a function of the threshold value, comparing standard synthetic control method (SCM, red) and augmented synthetic control method (ASCM, blue). At the commonly used threshold of 5, fewer than half of projects pass validation under either method. The similar performance of both methods across the threshold range indicates that the validation failures reflect fundamental problems with donor pool selection and covariate specification rather than estimation approach.

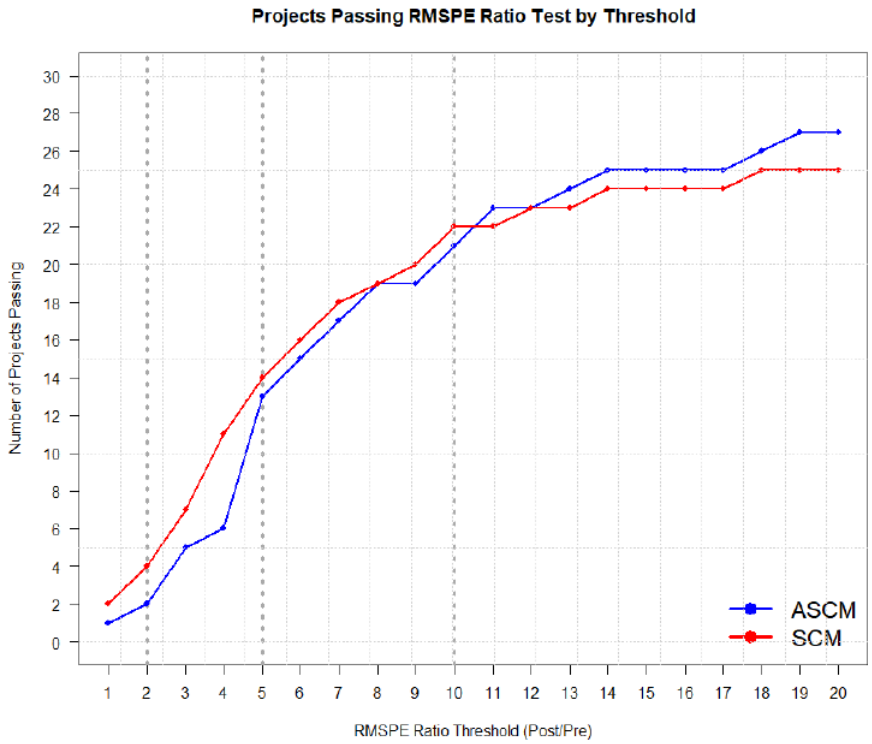


Figure C.2
Replication of West et al. Figure S6

"Validation" of the synthetic control method in each REDD+ project using an augmented synthetic control method (ASCM). Pre-REDD+ deforestation in "to-be" REDD+ project areas (red) versus synthetic controls (blue). Shaded blue areas represent the validation periods.

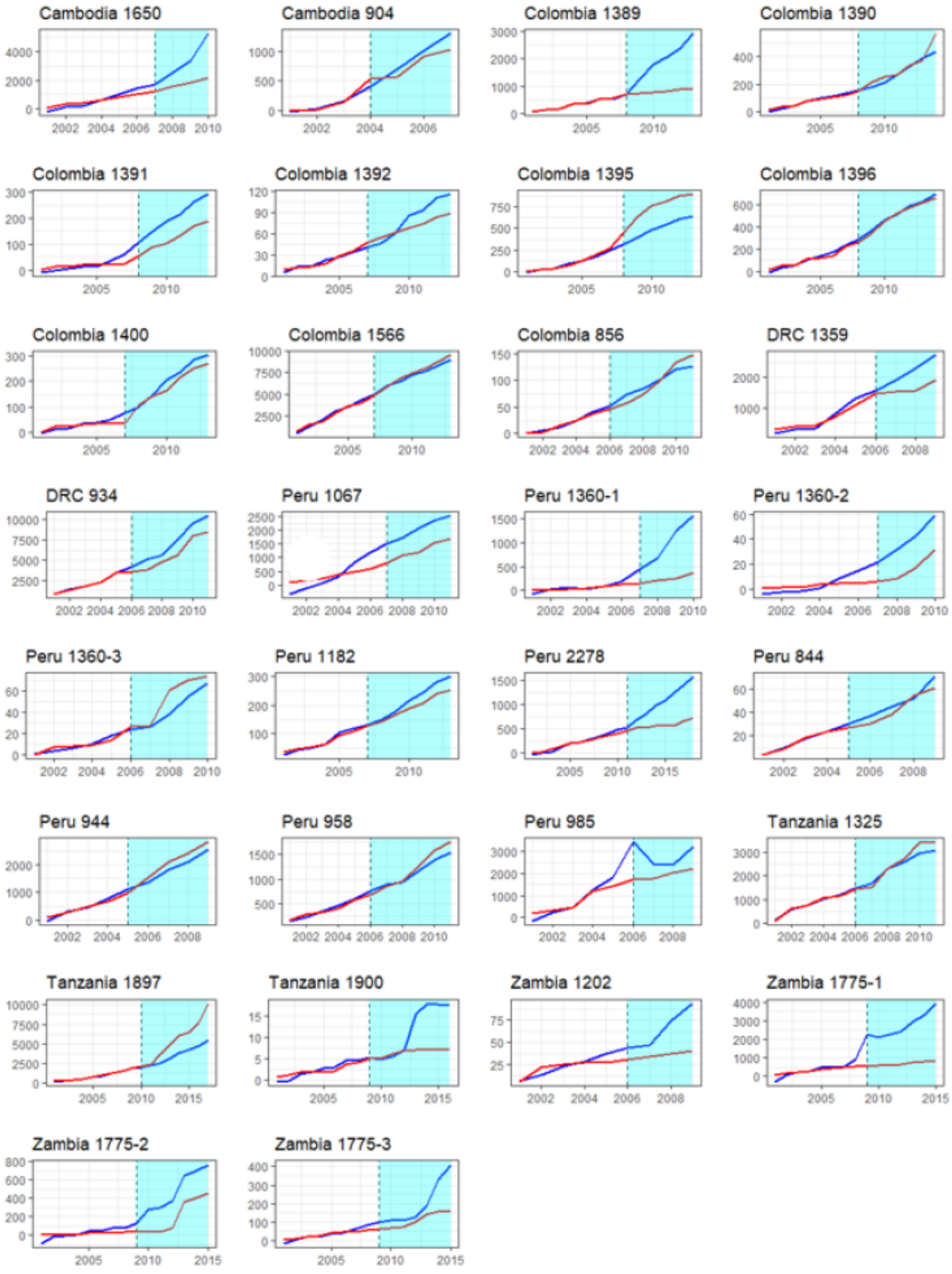


Figure C.3
Sensitivity of average treatment effects on the treated (ATT)

Differences between the average ATT for each project when the filter is applied (green line) and when it is not (orange line). The ATT is calculated as the difference between the deforestation in the project area and the synthetic control deforestation.

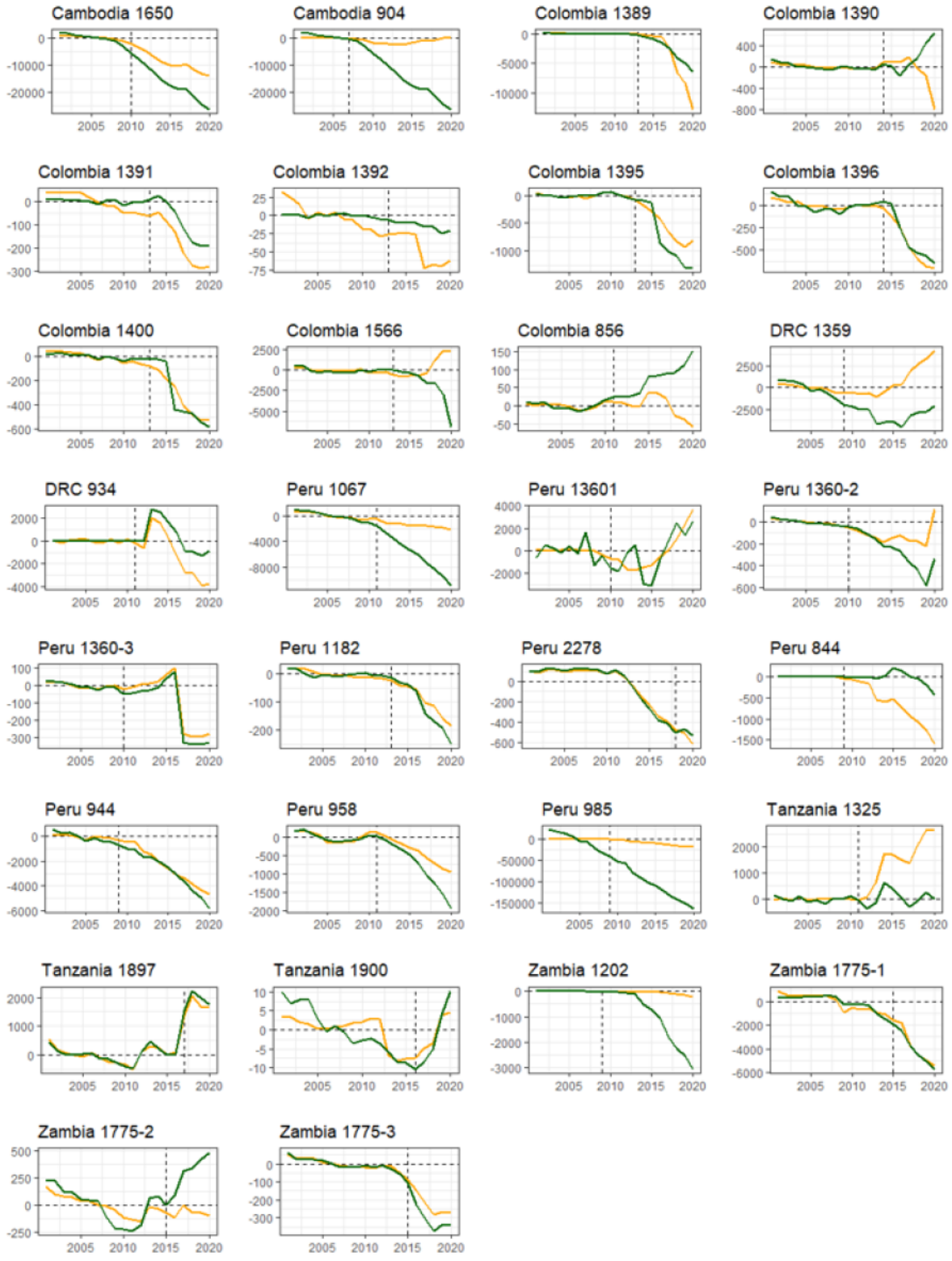


Figure C.4
Replication of West et al. Figure S7 (without filter)

Cumulative post-2000 deforestation in REDD+ project areas (red) versus synthetic controls (blue). Dashed black lines indicate the project implementation year.

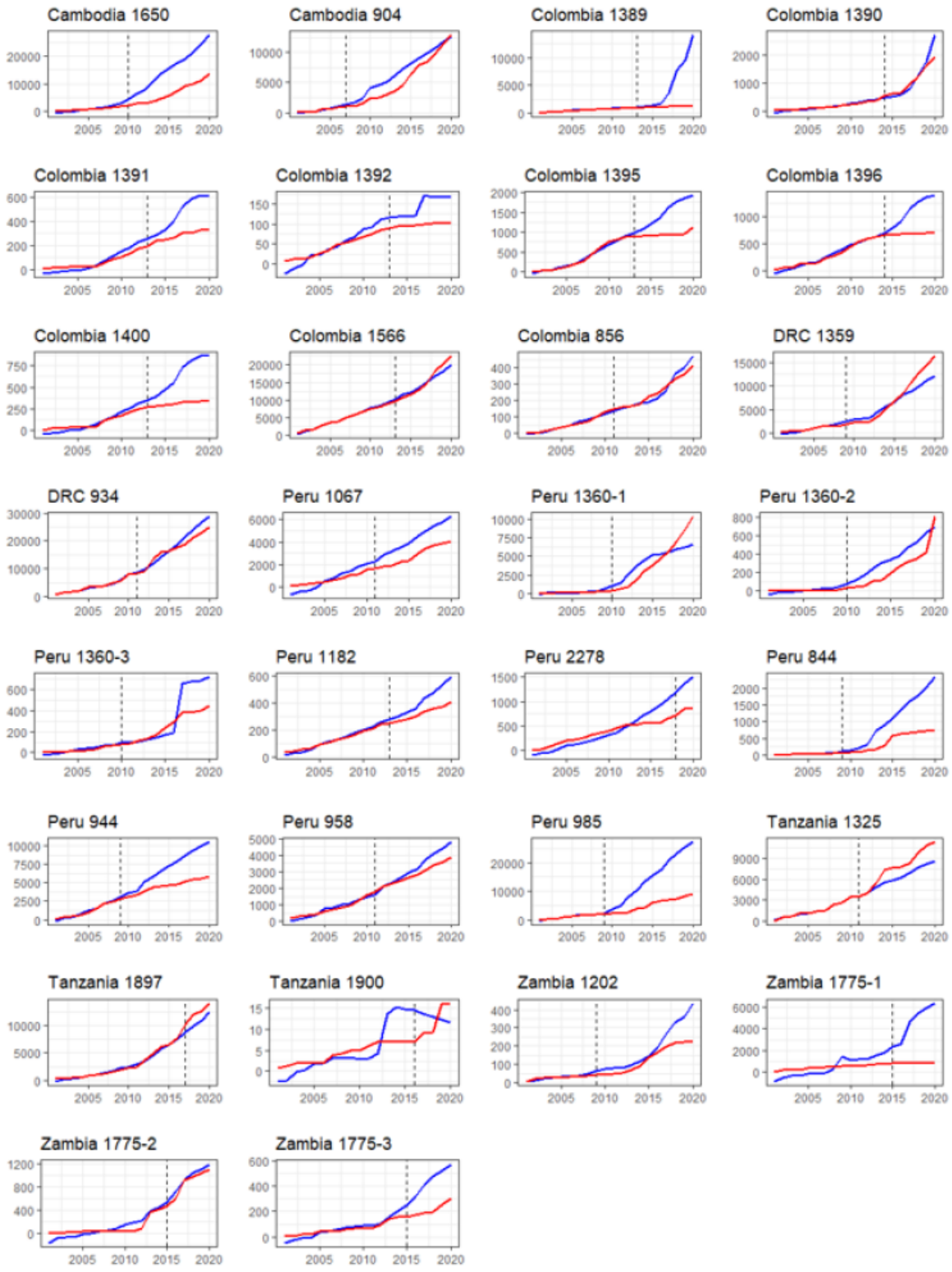


Figure C.5
Replication of West et al. Figure S8 (without filter)

Cumulative deforestation in REDD+ project areas minus deforestation in their respective synthetic controls. Dashed black lines indicate the project implementation year. Grey shaded areas represent 95% confidence intervals, obtained using the jackknife+ method (Barber et al., 2021).

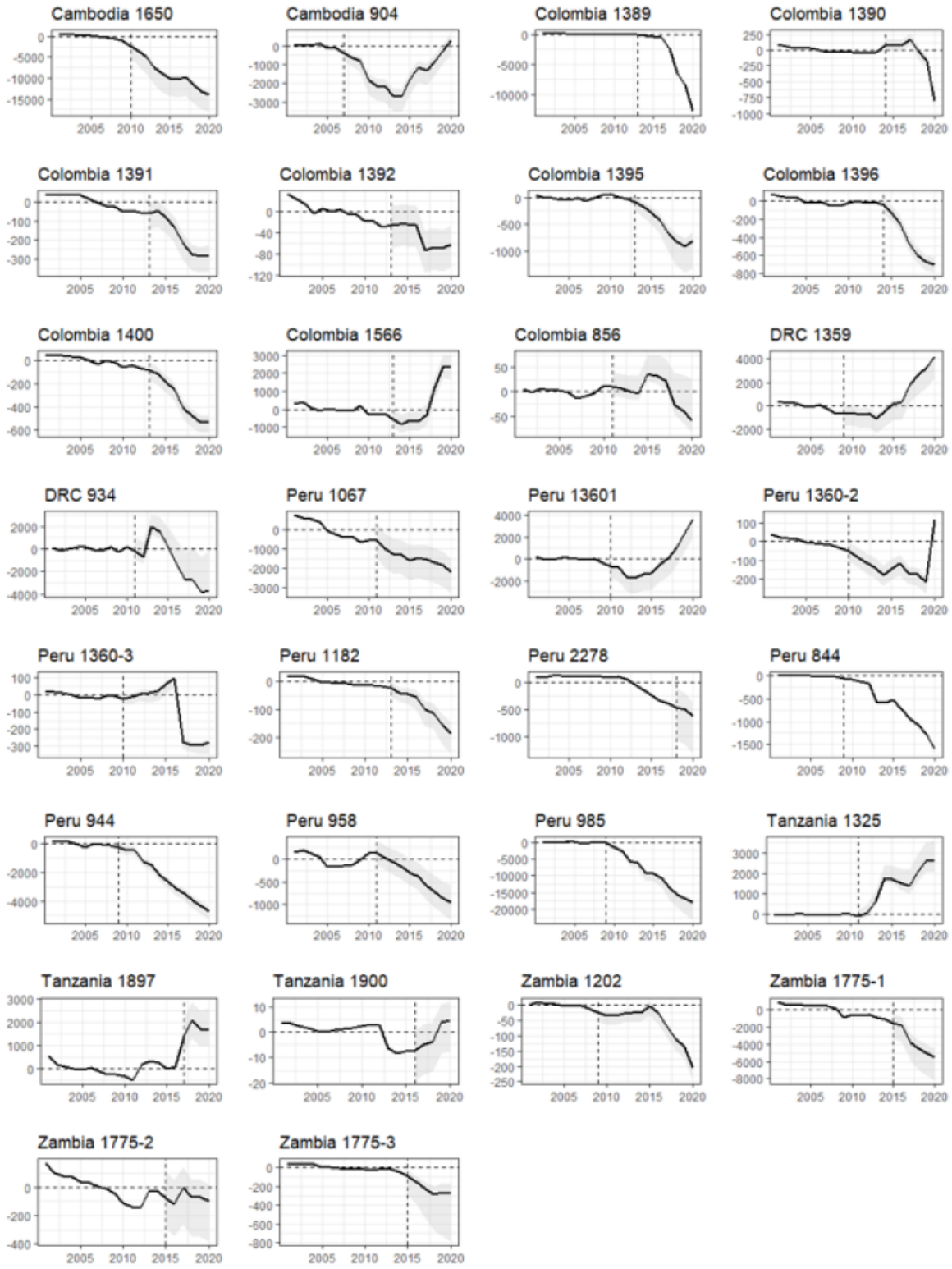


Figure C.6
Replication of West et al. Figure S7 (with filter)

Cumulative post-2000 deforestation in REDD+ project areas (red) versus synthetic controls (blue). Dashed black lines indicate the project implementation year.

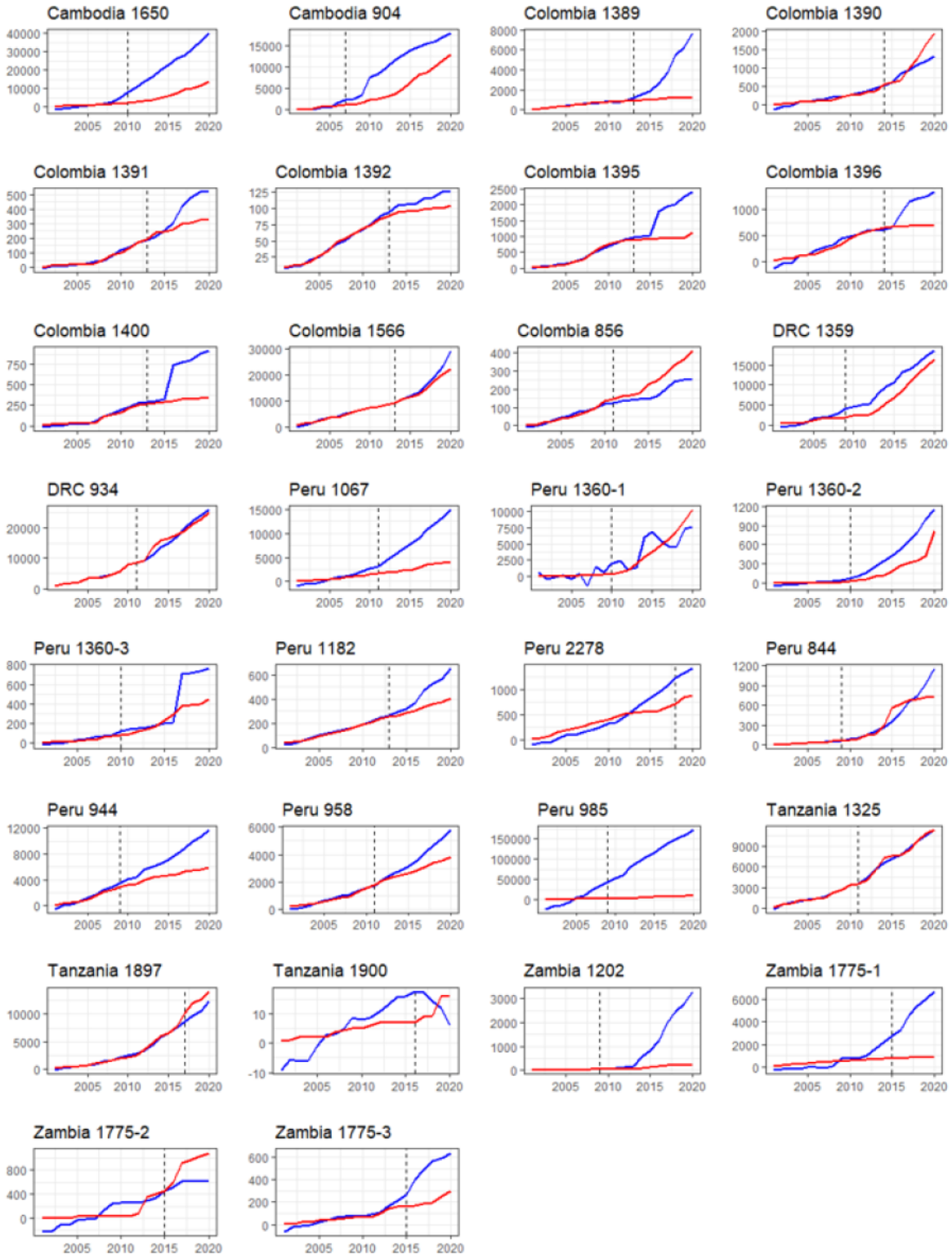
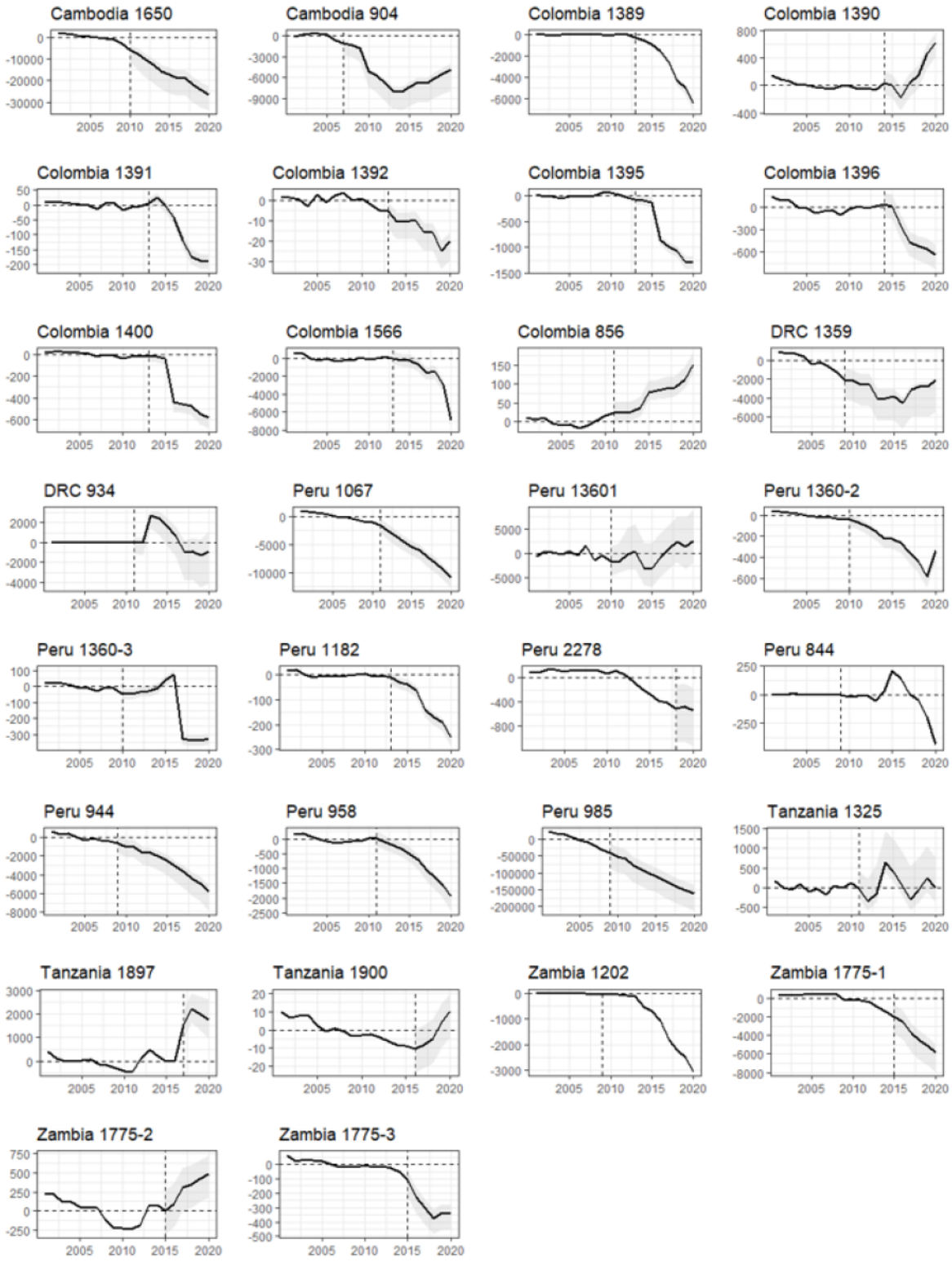


Figure C.7
Replication of West et al. Figure S8 (with filter)

Cumulative deforestation in REDD+ project areas minus deforestation in their respective synthetic controls. Dashed black lines indicate the project implementation year. Grey shaded areas represent 95% confidence intervals, obtained using the jackknife+ method as in Barber et al. (2021).



References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of economic literature*, 59(2):391–425.

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, 105(490):493–505.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American economic review*, 93(1):113–132.

Aldy, J. E. and Halem, Z. M. (2024). The evolving role of greenhouse gas emission offsets in combating climate change. *Review of Environmental Economics and Policy*, 18(2):175–209.

Alroy, J. (2017). Effects of habitat disturbance on tropical forest biodiversity. *Proceedings of the National Academy of Sciences*, 114(23):6056–6061.

Antoniuk, Y. and Leirvik, T. (2024). Climate change events and stock market returns. *Journal of Sustainable Finance & Investment*, 14(1):42–67.

Araujo, R., Assunção, J., Hirota, M., and Scheinkman, J. A. (2023). Estimating the spatial amplification of damage caused by degradation in the Amazon. *Proceedings of the National Academy of Sciences*, 120(46):e2308478120.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2021). Synthetic difference-in-differences. *American economic review*, 111(12):4088–4118.

Athey, S. and Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic perspectives*, 31(2):3–32.

Atmadja, S. S., Duchelle, A. E., De Sy, V., Selviana, V., Komalasari, M., Sills, E. O., and Angelsen, A. (2022). How do redd+ projects contribute to the goals of the paris agreement? *Environmental Research Letters*, 17(4):044038.

Bahrami, G., Gustafson, M., and Steiner, E. (2024). The biodiversity protection discount. Available at SSRN.

Baranzini, A., Van den Bergh, J. C., Carattini, S., Howarth, R. B., Padilla, E., and Roca, J. (2017). Carbon pricing in climate policy: seven reasons, complementary instruments, and political economy considerations. *Wiley Interdisciplinary Reviews: Climate Change*, 8(4):e462.

Barber, R. F., Candes, E. J., Ramdas, A., and Tibshirani, R. J. (2021). Predictive inference with the jackknife+. *The Annals of Statistics*, 49(1):486–507.

Barnett, M. (2020). The estimated impacts of climate policy risk. Technical report, Working paper.

Bastin, J.-F., Finegold, Y., Garcia, C., Mollicone, D., Rezende, M., Routh, D., Zohner, C. M., and Crowther, T. W. (2019). The global tree restoration potential. *Science*, 365(6448):76–79.

Ben-Michael, E., Feller, A., and Rothstein, J. (2021). The augmented synthetic control method. *Journal of the American Statistical Association*, 116(536):1789–1803.

Bennett, G., Gallant, M., and ten Kate, K. (2017). State of biodiversity mitigation 2017: Markets and compensation for global infrastructure development. Technical report, Forest Trends' Ecosystem Marketplace, Washington, DC.

Binder, J. (1998). The event study methodology since 1969. *Review of quantitative Finance and Accounting*, 11(2):111–137.

Boehmer, E., Masumeci, J., and Poulsen, A. B. (1991). Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics*, 30(2):253–272.

Bolton, P. and Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of financial economics*, 142(2):517–549.

Bolton, P., Kacperczyk, M., and Samama, F. (2022). Net-zero carbon portfolio alignment. *Financial Analysts Journal*, 78(2):19–33.

Bonan, G. B. (2008). Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science*, 320(5882):1444–1449.

Borenstein, S., Bushnell, J. B., and Wolak, F. A. (2002). Measuring market inefficiencies in california's restructured wholesale electricity market. *American Economic Review*, 92(5):1376–1405.

Brown, S. J. and Warner, J. B. (1985). Using daily stock returns: The case of event studies. *Journal of financial economics*, 14(1):3–31.

Busch, J. and Ferretti-Gallon, K. (2023). What drives and stops deforestation, reforestation, and forest degradation? an updated meta-analysis. *Review of Environmental Economics and Policy*, 17(2):217–250.

Campbell, J. Y., Lo, A. W., MacKinlay, A. C., and Whitelaw, R. F. (1998). The econometrics of financial markets. *Macroeconomic Dynamics*, 2(4):559–562.

Choi, D., Gao, Z., and Jiang, W. (2020). Attention to global warming. *The Review of Financial Studies*, 33(3):1112–1145.

Conley, T. G. and Taber, C. R. (2011). Inference with "difference in differences" with a small number of policy changes. *The Review of Economics and Statistics*, 93(1):113–125.

Corrado, C. J. (1989). A nonparametric test for abnormal security-price performance in event studies. *Journal of Financial Economics*, 23(2):385–395.

Costanza, R., De Groot, R., Sutton, P., Van der Ploeg, S., Anderson, S. J., Kubiszewski, I., Farber, S., and Turner, R. K. (2014). Changes in the global value of ecosystem services. *Global Environmental Change*, 26:152–158.

Creti, A. and Ftiti, Z. (2024). Energy, just transition, and sustainability: What's new?

Dasgupta, P. (2021). The economics of biodiversity: The Dasgupta review. Technical report, HM Treasury, London.

Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J. B., and Collen, B. (2014). Defaunation in the anthropocene. *Science*, 345(6195):401–406.

Engle, R. F., Giglio, S., Kelly, B., Lee, H., and Stroebel, J. (2020). Hedging climate change news. *The Review of Financial Studies*, 33(3):1184–1216.

Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276.

Faccini, Renato, M. R. and Skiadopoulos, G. (2023). Dissecting climate risks: are they reflected in stock prices? *Journal of Banking Finance*.

Fama, E. F. and French, K. R. (2023). Production of US rm-rf, smb, and hml in the Fama-French data library. *Chicago Booth Research Paper*, (23-22).

FAO-UNEP (2020). The state of the world's forests 2020: Forests, biodiversity and people. Technical report, Food and Agriculture Organization of the United Nations and United Nations Environment Programme.

Federici, S., Tubiello, F. N., Salvatore, M., Jacobs, H., and Schmidhuber, J. (2015). New estimates of CO2 forest emissions and removals: 1990–2015. *Forest Ecology and Management*, 352:89–98.

Fenichel, E. P., Dean, M. F., and Schmitz, O. J. (2024). The path to scientifically sound biodiversity valuation in the context of the global biodiversity framework. *Proceedings of the National Academy of Sciences*, 121(34):e2319077121.

Ferguson, C., Andandarajah, G., Butnar, I., Calzadilla, A., and Dessens, O. (2020). Costing the earth–climate damage costs and GDP. Technical report, UCL Energy Institute for CDP.

Flammer, C. (2015). Does corporate social responsibility lead to superior financial performance? A regression discontinuity approach. *Management Science*, 61(11):2549–2568.

Flammer, C. (2021). Corporate green bonds. *Journal of Financial Economics*, 142(2):499–516.

Flammer, C., Giroux, T., and Heal, G. M. (2025). Biodiversity finance. *Journal of Financial Economics*, 164:103987.

Garel, A., Romec, A., Sautner, Z., and Wagner, A. F. (2024). Do investors care about biodiversity? *Review of Finance*, 28(4):1151–1186.

Gelbach, J. B., Helland, E., and Klick, J. (2013). Valid inference in single-firm, single-event studies. *American Law and Economics Review*, 15(2):495–541.

Giam, X. (2017). Global biodiversity loss from tropical deforestation. *Proceedings of the National Academy of Sciences*, 114(23):5775–5777.

Giglio, S., Kelly, B., and Stroebel, J. (2021). Climate finance. *Annual Review of Financial Economics*, 13:15–36.

Giglio, S., Kuchler, T., Stroebel, J., and Wang, O. (2024). The economics of biodiversity loss. Working paper, National Bureau of Economic Research.

Gradeckas, S. (2025). Deep dive: Biodiversity credit sales. https://sgradeckas.substack.com. Accessed: 2025-04-30.

Greenfield, P. (2023). Revealed: more than 90% of rainforest carbon offsets by biggest certifier are worthless, analysis shows. *The Guardian*, 18(2023):2023.

Gregory, A. W. and Hansen, B. E. (1996). Residual-based tests for cointegration in models with regime shifts. *Journal of econometrics*, 70(1):99–126.

Guizar-Coutiño, A., Jones, J. P., 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.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., et al. (2013). High-resolution global maps of 21st-century forest cover change. *science*, 342(6160):850–853.

Heeb, F., Kölbel, J. F., Paetzold, F., and Zeisberger, S. (2022). Do investors care about impact? *The Review of Financial Studies*, 36(5):1737–1787.

Hooper, D. U., Adair, E. C., Cardinale, B. J., Byrnes, J. E. K., Hungate, B. A., Matulich, K. L., Gonzalez, A., Duffy, J. E., Gamfeldt, L., and O'Connor, M. I. (2012). A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, 486(7401):105–108.

Hsu, P., Li, K., and Tsou, C. (2020). The pollution premium (working paper). *Available at SSRN 3508808*.

Hsu, P.-H., Li, K., and Tsou, C.-Y. (2023). The pollution premium. *The Journal of Finance*, 78(3):1343–1392.

Huij, J., Laurs, D., Stork, P. A., and Zwinkels, R. (2022). Carbon beta: A market-based measure of climate risk. *Available at SSRN*, 3957900.

International Finance Corporation (IFC) (2023). Biodiversity finance reference guide: Building on the green bond principles and green loan principles. Technical report, International Finance Corporation, Washington, DC.

International Union for Conservation of Nature (IUCN) (2015). No net loss and net positive impact approaches for biodiversity. Technical report, IUCN, Gland, Switzerland. ISBN 978-2-8317-1718-0.

IPBES (2019). Global assessment report on biodiversity and ecosystem services. Technical report, Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Bonn, Germany.

Jiang, Y. and Luo, L. (2018). Market reactions to environmental policies: Evidence from china. *Corporate Social Responsibility and Environmental Management*, 25(5):889–903.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, 12(2-3):231–254.

Juselius, K. et al. (1990). Maximum likelihood estimation and inference on cointegration–with applications to the demand for money. *Oxford Bulletin of Economics & Statistics*, 52(2):169.

Kim, S., Li, T., and Wu, Y. (2024). Carbon offsets: Decarbonization or transition-washing? Available at SSRN 4762799.

Klein, A.-M., Vaissière, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C., and Tscharntke, T. (2007). Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society B: Biological Sciences*, 274(1608):303–313.

Kothari, S. P. and Warner, J. B. (2007). Econometrics of event studies. In *Handbook of empirical corporate finance*, pages 3–36. Elsevier.

Krüger, P. (2015). Corporate goodness and shareholder wealth. *Journal of Financial Economics*, 115(2):304–329.

Kumar, P. and Firoz, M. (2019). How does capital market respond to certified emission reductions (cers) announcements in india: An event study methodology. *Managerial Finance*, 45(7):950–965.

Lesnikowski, A., Ford, J., Biesbroek, R., Berrang-Ford, L., Maillet, M., Araos, M., and Austin, S. E. (2017). What does the paris agreement mean for adaptation? *Climate Policy*, 17(7):825–831.

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of economic literature*, 35(1):13–39.

Markandya, A., Orgen, P., and Abdulla, S. (2024). Adaptation responses to date in asia and the pacific. In *1st Asia-Pacific Flagship Report on Climate Change*. Asian Development Bank.

Marketplace, F. T. (2022). The art of integrity: state of voluntary carbon markets, q3 insights briefing. *Washington DC: Forest Trends Association*.

Martinsson, G., Sajtos, L., Strömberg, P., and Thomann, C. (2024). The effect of carbon pricing on firm emissions: Evidence from the swedish co2 tax. *The Review of Financial Studies*, page hhad097.

McNicol, I. M., Keane, A., Burgess, N. D., Bowers, S. J., Mitchard, E. T., and Ryan, C. M. (2023). Protected areas reduce deforestation and degradation and enhance woody growth across african woodlands. *Communications Earth & Environment*, 4(1):392.

McNicol, I. M., Ryan, C. M., and Mitchard, E. T. (2018). Carbon losses from deforestation and widespread degradation offset by extensive growth in african woodlands. *Nature communications*, 9(1):3045.

Millennium Ecosystem Assessment (2005). *Ecosystems and human well-being*, volume 5. Island Press, Washington, DC.

Müller, C., Orgen, P., and Behr, P. (2023). Portfolio allocation and optimization with carbon offsets: is it worth the while? *Journal of Climate Finance*, 5:100019.

Natural Capital Coalition (2016). Natural capital protocol. Technical report, Natural Capital Coalition, London.

OECD (2020). A comprehensive overview of global biodiversity finance. Oecd environment policy paper, Organisation for Economic Co-operation and Development, Paris.

Palomo, I., Locatelli, B., Otero, I., Colloff, M., Crouzat, E., Cuni-Sanchez, A., Gómez-Baggethun, E., González-García, A., Grêt-Regamey, A., Jiménez-Aceituno, A., et al. (2021). Assessing nature-based solutions for transformative change. *One Earth*, 4(5):730–741.

Paulson Institute, The Nature Conservancy, and Cornell Atkinson Center for Sustainability (2020). Financing nature: Closing the global biodiversity financing gap. Technical report, Paulson Institute / The Nature Conservancy / Cornell Atkinson Center for Sustainability, Chicago.

Pendrill, F., Gardner, T. A., Meyfroidt, P., Persson, U. M., Adams, J., Azevedo, T., Bastos Lima, M. G., Baumann, M., Curtis, P. G., De Sy, V., et al. (2022). Disentangling the numbers behind agriculture-driven tropical deforestation. *Science*, 377(6611):eabm9267.

Phillips, P. C. (1991). Optimal inference in cointegrated systems. *Econometrica: Journal of the Econometric Society*, pages 283–306.

Pollination (2024). State of voluntary biodiversity credit markets: Current supply & demand dynamics. Technical report, Pollination, Sydney.

Qu, Z. and Perron, P. (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica*, 75(2):459–502.

Rodemeier, M. (2025). Willingness to pay for carbon mitigation: Field evidence from the market for carbon offsets. *Review of Financial Studies*. Forthcoming.

Schleicher, J., Eklund, J., D. Barnes, M., Geldmann, J., Oldekop, J. A., and Jones, J. P. (2020). Statistical matching for conservation science. *Conservation Biology*, 34(3):538–549.

So, S. I., Haya, B. K., and Elias, M. (2023). Voluntary registry offsets database, v7.1. Technical report, University of California, Berkeley, Berkeley Carbon Trading Project.

Starks, L. T. (2023). Presidential address: Sustainable finance and ESG issues—value versus values. *The Journal of Finance*, 78(4):1837–1872.

Stephens, B. B., Gurney, K. R., Tans, P. P., Sweeney, C., Peters, W., Bruhwiler, L., Ciais, P., Ramonet, M., Bousquet, P., Nakazawa, T., et al. (2007). Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO2. *Science*, 316(5832):1732–1735.

Streck, C. (2012). Financing REDD+: Matching needs and ends. *Current Opinion in Environmental Sustainability*, 4(6):628–637.

Taskforce on Nature-related Financial Disclosures (TNFD) (2023). TNFD beta framework v0.4. Technical report, TNFD. Version 0.4.

Tyukavina, A., Baccini, A., Hansen, M., Potapov, P., Stehman, S., Houghton, R., Krylov, A., Turubanova, S., and Goetz, S. (2015). Aboveground carbon loss in natural and managed tropical forests from 2000 to 2012. *Environmental Research Letters*, 10(7):074002.

United Nations Convention to Combat Desertification (UNCCD) (2022). Global land outlook, second edition: Land restoration for recovery and resilience. Technical report, UNCCD, Bonn, Germany.

VCS Association (2017). Climate, community & biodiversity standards: Third edition. Technical report, Verra, Washington, DC. Standard Version 3.1.

Veld-Merkoulova, Y. and Viteva, S. (2015). Do regulations affect carbon market returns and volatility? In *Carbon Finance: How Carbon and Stock Markets are affected by Energy Prices and Emissions Regulations*, pages 19–78. Springer.

West, T. A., 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.

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.

Wilson, E. O. (1986). *Biophilia*. Harvard University Press.

Wolf, C. and Ripple, W. J. (2022). Putting a face on carbon with threatened forest primates. *Proceedings of the National Academy of Sciences*, 119(28):e2207255119.

Zeileis, A., Kleiber, C., Krämer, W., and Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational statistics & data analysis*, 44(1-2):109–123.