

Readme file- Reproducible package

Readme File for Replication of the Paper:

“Local Visibility vs. Global Integrity: Evidence from Corporate Carbon Offsetting”

This readme file was created to guide the reproducibility package for the analysis of the paper.

Authors: Alvaro Pedraza, Federica Zeni, and Tomas Williams

Overview

This paper studies where firms retire carbon offsets and how project quality varies with firms’ geographic operational footprint. Using global firm-level emissions data matched to voluntary carbon registry retirements and project-level quality ratings, we document three central findings. First, firms disproportionately retire offsets in countries where they maintain real economic operations. This “local visibility” pattern is economically large and persistent over time. Second, offsets retired in operational countries are, on average, of lower quality as measured by independent project ratings. The negative quality gradient is strongest among firms with low emission intensity, consistent with a goodwill or signaling channel. Third, firms with higher emission intensity display a weaker quality gradient, suggesting that when abatement needs are more material, integrity concerns partially discipline offset choice. Overall, the results indicate a systematic tension between local signaling incentives and global environmental integrity, with implications for the credibility and allocation of voluntary carbon markets.

Data Sources and Access

The analysis in this paper relies on six distinct raw data sources. Three are proprietary and cannot be redistributed, while two of the three that are publicly available require substantial processing.

1. ISS STOXX (Proprietary)

Content: Firm-level greenhouse gas emissions, sales, and climate-target disclosures.

Access: ISS STOXX data are available only through institutional subscription agreements. Licensing restrictions prohibit redistribution or inclusion of the raw data in this replication package. Licenses for the following ISS products were obtained for this project: “Climate Core - Impact (DataDesk Data Set)”. The raw data file for 2012-2023 were shared by ISS STOXX via a SFTP connection on March 25th, 2025.

The original data were acquired from ISS STOXX. The dataset “iss_with_fs_id_unique.dta” was generated by processing the raw ISS data in the reproducibility package:

<https://reproducibility.worldbank.org/catalog/441>. The code and data for that package is currently under temporary embargo. The raw ISS data are proprietary and not included in this replication package.

Vendor link: <https://www.issgovernance.com/sustainability/climate-solutions/climate-analytics/>

ISS STOXX. *Climate Core – Impact (DataDesk Data Set): Firm-Level Greenhouse Gas Emissions, Sales, and Climate Target Disclosures*. Proprietary database, ISS STOXX, accessed via institutional subscription. Data for years 2012–2023. Delivered via secure SFTP on March 25, 2025.

2. FactSet GeoRevenue (Proprietary)

Content: Global firm-level geographic revenue breakdowns used to construct firms’ country-level operational footprint and economic exposure across countries.

Access: FactSet GeoRevenue requires an institutional subscription to FactSet. Redistribution is not permitted, so the raw geographic revenue data are not included in this replication package. Licenses for the following FactSet products are required for this project: GeoRevenue, Fundamentals, and Data Management Solutions. With the appropriate license, data can be downloaded via FactSet’s proprietary “FactSet Loader,” which pulls subscribed datasets directly from FactSet’s servers. The results presented in this project are computed based on a data pull from August 26, 2025.

Vendor link: <https://www.factset.com/data>

Researchers seeking to replicate the results must obtain access to ISS STOXX and FactSet GeoRevenue independently through their own institutions. All code in the replication package is structured so that it can be executed seamlessly once these datasets have been obtained and placed in the appropriate directories.

FactSet Research Systems Inc. FactSet GeoRevenue and Firm Fundamentals. Proprietary database, accessed via institutional subscription. Data retrieved using FactSet Loader on August 26, 2025. <https://www.factset.com/data>

3. Carbon Offset Issuances and Retirements (Public, Processed)

Raw source: Berkeley Carbon Trading Project – Offset Issuances & Retirements Database <https://gspp.berkeley.edu/berkeley-carbon-trading-project/offsets-database>

Downloaded on **January 7, 2025**.

The carbon offset issuance and retirement data used in this paper are based on the Berkeley Carbon Trading Project registry files. We use **Version archive-v2024-12-year-end**, which includes data through **December 31, 2024**, available by clicking the “download the registry files here” link on the project website.

Raw Registry Files and Sheets Used

The raw download consists of a single Excel workbook containing multiple sheets. We use four sheets that report **carbon offset retirements** from the major voluntary registries:

- **“ACR Retirements”** (American Carbon Registry)
- **“CAR Retirements”** (Climate Action Reserve)
- **“Gold Retirements”** (Gold Standard)
- **“Verra VCUs”** (Verra / VCS)

Each sheet corresponds to a different registry and uses registry-specific variable names and formats.

Identification of Retirement Beneficiaries

Information on the **retirement beneficiary** (i.e., the entity on whose behalf the credits are retired) is not standardized across registries and appears in different text fields in each sheet. The relevant raw columns used to identify beneficiaries are:

Registry	Beneficiary Field(s)
ACR	<i>Retired on Behalf of, Retirement Reason, Purpose of Retirement</i>
CAR	<i>Retirement Reason, Retirement Reason Details</i>
Gold Standard	<i>Note</i>
Verra	<i>Retirement Beneficiary, Retirement Reason, Retirement Details</i>

These fields typically contain free-text descriptions that may include company names, subsidiaries, project developers, intermediaries, or other descriptive notes. To standardize beneficiary information across registries, we apply a large-language-model (LLM)–based text-parsing procedure to these fields. While the underlying parsing code is not included in the replication files, the procedure systematically processes the raw text to extract, when available, the following structured identifiers:

- Company name (as mentioned in the registry text)
- Official standardized company name
- sp_entity_id (S&P entity identifier)
- factset_entity_id (FactSet entity identifier)

This procedure allows us to consistently identify publicly listed firms across registries despite heterogeneous reporting formats and unstructured text.

Core Registry Variables Used

In addition to beneficiary information, several registry-specific fields are harmonized and mapped into standardized variables used in the final dataset. The table below lists the raw column names used in each registry and their conceptual role:

Registry	Beneficiary Field (s)	serialnumber	quantity	retirementdate	ProjectID	Vintage
ACR	Retired on Behalf of, Retirement Reason, Purpose of Retirement	Credit Serial Numbers	Quantity of Credits	Status Effective (GMT)	Project ID	Vintage
CAR	Retirement Reason, Retirement Reason Details	Offset Credit Serial Numbers	Quantity of Offset Credits	Status Effective	Project ID	Vintage
GOLD STANDARD	Note	Serial Number	Quantity	Retirement Date	GSID	Vintage
VERRA	Retirement Beneficiary, Retirement Reason, Retirement Details	Serial Number	Quantity Issued	Retirement/Cancellation Date	ID	Vintage Starts

These variables are renamed and harmonized to create consistent fields for credit quantities, retirement timing, project identifiers, and vintage across registries.

Linking to Project-Level Characteristics

Using the registry-specific **Project ID**, retirement records are linked to project-level data from the same Berkeley Carbon Trading Project download. Specifically, we use **version v2024-12-year-end**, sheet “**PROJECTS**”, which provides additional characteristics for each project.

This adds the following project-level variables to the retirement data:

- projectname
- projecttype
- country
- isocode
- region
- subregion
- type_id
- classification_name

This step completes the construction of the processed carbon offset retirement dataset used in the analysis. The resulting file is ***“Retirements_ & ListedFirms.dta”***

Berkeley Carbon Trading Project. *Offset Issuances and Retirements Database*. Version: **archive-v2024-12-year-end**. Goldman School of Public Policy, University of California, Berkeley. Public dataset. Downloaded January 7, 2025. <https://gspp.berkeley.edu/berkeley-carbon-trading-project/offsets-database>

4. Carbon Project Ratings (Public, Manually Collected)

Raw source: MSCI Carbon Project Ratings

<https://www.msci.com/data-and-analytics/carbon-markets/carbon-project-ratings>

MSCI does not provide a bulk download or API access for carbon offset project ratings. As a result, project-level ratings must be retrieved individually through MSCI’s “View our Ratings” web interface (sign-in required).

For each ProjectID appearing in *Retirements_ & ListedFirms.dta*, the project identifier was manually entered into the “View our Ratings” search interface. When a rating was available, the following information was recorded:

- The MSCI letter rating (rating_msci)
- The corresponding numerical rating (numrating_MSCI)

These ratings were manually compiled and merged back to the retirement-level data using the project identifier. Projects without an MSCI rating are retained in the dataset with missing values for the rating variables.

MSCI Inc. MSCI Carbon Project Ratings. Restricted-access web-based database. Accessed via the “View our Ratings” interface (registration required). Retrieved January 2025. <https://www.msci.com/data-and-analytics/carbon-markets/carbon-project-ratings>

Raw source: BeZero Carbon Ratings

<https://bezerocarbon.com/ratings/listings>

BeZero does not provide a bulk download or API access for its carbon offset project ratings. As a result, project-level ratings must be retrieved individually through BeZero's ratings listings interface (log-in required).

For each ProjectID appearing in Retirements_&_ListedFirms.dta, the project was located in the BeZero listings interface. When a rating was available, the following information was recorded:

- The BeZero letter rating (rating)
- The corresponding numerical rating (rating_num)

These ratings were manually compiled and merged back to the retirement-level data using the project identifier. Projects without a BeZero rating are retained in the dataset with missing values for the rating variables.

The resulting project-level ratings file is provided in the replication package as **\Replication Package\Data\Raw\rating_Bezero_msci.dta**

BeZero Carbon. BeZero Carbon Ratings. Web-based database (registration required). Retrieved 2025. <https://bezerocarbon.com/ratings/listings>

Notes on ratings coverage and project identifiers

MSCI and BeZero Carbon Project Ratings are periodically reviewed and updated as new information becomes available or methodologies evolve. As a result, project ratings reported here reflect the status at the time of data collection (January 2025). Researchers accessing MSCI or BeZero at a later date may observe updated ratings for some projects.

In addition, MSCI uses registry-specific prefixes that differ from those commonly used in carbon registries. In particular, projects registered under Verra's Verified Carbon Standard (VCS) are coded by MSCI using the prefix "VER" rather than "VCS." For example, a project identified as VCS1301 in registry data appears in MSCI as VER1301. All project identifiers were harmonized accordingly prior to merging MSCI ratings with the retirement-level dataset.

5. Sylvera (Proprietary)

Content: Project-level carbon credit quality ratings and historical price data used to measure offset integrity and transaction pricing. The subscription files used in this project include:

- **Full_ratings.csv** – Project-level ratings, rating scores, and associated project metadata.
- **Price_history.csv** – Historical transaction price data at the project level.

These data are used to construct the measures of offset quality and to analyze price–quality relationships across projects and countries.

Access: Sylvera data require an institutional subscription and are not publicly available. Redistribution is not permitted, so the raw files are not included in this replication package. The results presented in this project are based on a subscription data extract obtained in 2025. Researchers seeking to replicate the analysis must obtain independent access to Sylvera and place the subscription files in the designated /Data/Proprietary/Sylvera/ directory prior to running the replication scripts.

Vendor link: <https://www.sylvera.com>

Sylvera Ltd. Carbon Credit Ratings and Price History Database. Proprietary database, accessed via institutional subscription. Data retrieved 2025.

6. Worldwide Governance Indicators (Public)

Content: Country-level institutional quality indicators from the World Bank’s Worldwide Governance Indicators (WGI). This project uses the governance estimate variable (WGI composite governance index), which captures institutional quality across six dimensions, including rule of law, regulatory quality, government effectiveness, control of corruption, political stability, and voice and accountability.

The raw file used in this project is:

- **wgidataset_with_sourcedata-2025.xlsx** (sheet: “va”)

From this file, the variable:

- Governanceestimateapprox2

is extracted and renamed as:

- governance

The dataset is merged at the country-year level.

Access: The Worldwide Governance Indicators are publicly available from the World Bank. The data can be downloaded directly from:

<https://www.worldbank.org/en/publication/worldwide-governance-indicators>

World Bank. Worldwide Governance Indicators (WGI). Public dataset. Accessed December 2025.

Replication Package Contents

All code used in the analysis—including data cleaning, merging, construction of treatment variables, and execution of the empirical models—is provided in full. The code will run as-is once the proprietary datasets have been obtained and placed in the specified folders.

Summary of Data Inputs

Data Input	Description	Path in Folder or File name
ISS STOXX	Firm-level emissions, sales, and climate-target variables	Data/Raw/Proprietary/ISS/iss_with_fs_id_unique.dta
FactSet GeoRevenue	Firm revenue footprint	Data/Raw/Proprietary/Factset/factsetgeo.dta
Carbon Offset Issuances and Retirements	Issuances and retirements with firm identifiers and parsed metadata	Data/Raw/Proprietary/Retirements_&_ListedFirms.dta
Carbon Project Ratings	Project-level letter and numeric quality ratings from MSCI	Data/Raw/rating_Bezero_msci.dta
Sylvera Ratings and Prices	Project-level carbon credit quality ratings and historical price data from Sylvera	Data/Raw/Proprietary/Sylvera/
Worldwide Governance Indicators	Country-level institutional quality indicators from the World Bank's Worldwide Governance Indicators	Data/Raw/wgidataset_with_sourcedata-2025.xlsx

Instructions for Replicators

Code

The analysis in this paper is executed through the master script Code/main.do, which organizes all steps required to construct the analytic dataset and reproduce the results presented in the paper. All directory paths are defined at the top of the script. Users must update the global `$_{project}` path before running the replication.

The master script defines the following directory structure:

- `$_{code}` → `$_{project}/Code`
- `$_{DATA}` → `$_{project}/Data`
- `$_{TEMP}` → `$_{project}/Data/Temp`
- `$_{TABLES}` → `$_{project}/Outputs/Tables`
- `$_{FIGURES}` → `$_{project}/Outputs/Figures`

All intermediate datasets are written to the Data/Temp/ folder. Regression outputs and formatted tables are exported to Outputs/Tables/, and all figures are saved to Outputs/Figures/.

The master script sequentially executes the following do-files:

- **0_data_preparation.do**
Constructs the analytic dataset used in all empirical analyses.
This script imports raw firm-level emissions data, operational footprint data, carbon registry retirements, project ratings, and country-level controls. It harmonizes identifiers across data sources and generates the key analysis variables.
Intermediate datasets are written to the Data/Temp/ folder.
- **1_sum_stats.do**
Produces summary statistics and descriptive patterns reported in the paper. This includes construction of tables and all figures illustrating operational presence and offset quality.

- **2_regressions.do**

Runs all baseline specifications, heterogeneity analyses, and robustness checks. The script exports regression tables to the Outputs/Tables/ folder and saves figures to the Outputs/Figures/ folder.

Running main.do from start to finish reproduces all results in the paper.

Computational settings

The script was developed and tested using **Stata version MP 18.0** on a **64-bit Windows system** (MinGW environment) with **32 GB of RAM (32000 MB)**. While the script ran successfully in this environment, we expect it would also run with less RAM.

Under the described settings, the reproducibility package takes around **10** minutes to run from beginning to end.

List of Exhibits

Exhibit name	Output file name	Script	Folder Path
Figure 1. Retirements by Geographical Presence	Share_time.pdf	1_sum_stats.do	/Outputs/Figures
Figure 2. Revenue Exposure and Concentration of Offset Retirements	Share_by_operation.pdf	1_sum_stats.do	/Outputs/Figures
Figure 3. Carbon offset retirements by project quality and firm presence	Rating_operation.pdf (Panel A) Rating_operation_type.pdf (Panel B)	1_sum_stats.do	/Outputs/Figures
Figure 4. Offset Quality and Firm Experience within Project Countries	Rating_new_old.pdf	1_sum_stats.do	/Outputs/Figures
Figure 5. Carbon offset retirements by project quality and firm presence: High vs. Low governance countries	Rating_operation_low_governance.pdf (Panel A) Rating_operation_high_governance.pdf (Panel B)	1_sum_stats.do	/Outputs/Figures
Exhibit name	Output file name	Script	Folder Path
Table 1. Carbon Offset Projects: Summary Statistics	Table_summary_project.xlsx	1_sum_stats.do	/Outputs/Tables

Exhibit name	Output file name	Script	Folder Path
Table 2. Top Countries by Corporate Origin and Carbon Project Location	Top15_Origin_vs_Retire.xlsx	1_sum_stats.do	/Outputs/Tables
Table 3. Listed companies retiring carbon credits: Summary statistics	Summary_stats_selected.xlsx	1_sum_stats.do	/Outputs/Tables
Table 4. Operational Presence, Revenue Footprint, and Offset Quality	Findings.xlsx Sheet: Baseline	2_regressions.do	/Outputs/Tables
Table 5. Operational Presence, Experience, and Offset Quality	Findings.xlsx Sheet: Early_late	2_regressions.do	/Outputs/Tables
Table 6. Retirement Intensity, Operational Presence, and the Quality Gradient	Findings.xlsx Sheet: Robustness	2_regressions.do	/Outputs/Tables
Table 7. Prices, Buyer Composition, and Project Quality	Findings.xlsx Sheet: Price_quality	2_regressions.do	/Outputs/Tables
Appendix Table A1. Operational Presence, Revenue Footprint, and Offset Quality (Sylvera and BeZero ratings)	Findings.xlsx Sheet: SYL_BeZero	2_regressions.do	/Outputs/Tables

Folder structure

C:\Users\WB468173\OneDrive - WBG\Documents\Research\Operational Footprint\Replication Package

- Code
 - 0_data_preparation.do
 - 1_sum_stats.do
 - 2_regression
 - main.do
- Data
 - Data/Raw
 - Rating_Bezero_msci.dta
 - Wgidataset_with_source-2025.xlsx
 - Data/Proprietary
 - Retirements_&_ListedFirms.dta
 - Data/Proprietary/Factset/factsetgeo.dta
 - Data/Proprietary/ISS/iss_with_fs_id_unique.dta
 - Data/Temp
 - factsetgeo_oper.dta

- factsetgeo_orig.dta
 - governance.dta
 - HighI_firm.dta
 - HighR_firm.dta
 - iss_with_fs.dta
 - Retirements_& ListedFirms_Emissions.dta
 - Retirements_& ListedFirms_Full.dta
 - Sylvera_prices.dta
 - Sylvera_ratings.dta
- Readme
 - Readme.docx
- Outputs
 - Outputs/Figures
 - Rating_new_old.pdf
 - Rating_operation.pdf
 - Rating_operation_high_governance.pdf
 - Rating_operation_low_governance.pdf
 - Rating_operation_type.pdf
 - Share_by_operation.pdf
 - Share_time.pdf
 - Outputs /Tables
 - Findings.xlsx
 - summary_stats_selected.xlsx
 - table_summary_projects.xlsx
 - Top15_Origin_vs_Retire.xlsx