Carbon Uncertainty and Asset Prices*

Cathy Yi-Hsuan Chen† George M. Constantinides‡

Nilesh Hegde§ Abraham Lioui¶

February 15, 2026

*We thank our colleagues and seminar participants for feedback.

†Adam Smith Business School, University of Glasgow, UK (*e-mail: CathyYi-Hsuan.Chen@glasgow.ac.uk*)

‡University of Chicago Booth School of Business, U.S.A. (*e-mail: George.Constantinides@chicagobooth.edu*)

§University of Chicago Booth School of Business, U.S.A. (*e-mail: Nilesh.Hegde@chicagobooth.edu*)

¶EDHEC Business School, France (*e-mail: Abraham.Lioui@edhec.edu*)

Carbon Uncertainty and Asset Prices

Abstract

We find that measurement uncertainty in firms' carbon intensity commands a positive risk premium, distinct from and alongside a robust premium associated with carbon-intensity levels. Because most emissions are estimated rather than disclosed, our main contribution is to construct an observation-level measure of carbon-data precision at scale: for each firm-month, we quantify the uncertainty of the underlying carbon signal. We conceptually separate *estimation risk* (statistical error from inferring emissions under sparse firm information) from *model risk*, proxied by cross-vendor disagreement. Using a replicable Multiple Imputation by Chained Equations (MICE) framework combined with Random Forests, we jointly estimate firm-month emissions, carbon intensity, and uncertainty for the full CRSP universe, overcoming the selection bias inherent in vendor-restricted samples. Training on self-disclosing firms, we show that vendor-estimated data are largely redundant: adding vendor observations provides little incremental information beyond disclosures, consistent with estimates being anchored to the same disclosure backbone. By contrast, imputations within vendor-covered universes can embed selection bias and materially alter asset-pricing inferences. Overall, investors price not only carbon exposure but also carbon opacity.

Keywords: carbon emissions uncertainty; carbon emissions premium; carbon emissions uncertainty premium; Trucost imputation of carbon emissions; random forest imputation of carbon emissions and uncertainty.

JEL Classification: Q3, Q4, Q43, G12.

1 Introduction

The extant literature on the effect of carbon emissions and carbon intensity (emissions normalized by a measure of the firm scale) on the factor-adjusted stock return premium (alpha) is inconclusive and the results are often contradictory.[1] A growing body of work studies whether firms that are more exposed to transition risk - often proxied by greenhouse gas (GHG) emissions or intensity - face a different cost of capital or earn different expected returns. This question matters because emissions-based signals increasingly guide portfolio tilts, risk management, and the evaluation of corporate transition plans. Yet the apparent simplicity of "use emissions as a proxy for carbon risk" hides a basic measurement fact: for a large fraction of firm-dates, emissions are not observed, they are constructed.

Most empirical studies rely on emissions databases produced by data vendors. These datasets are valuable, but they embed two features that are rarely treated as first-order objects in asset pricing tests. First, disclosure remains limited, so vendor panels contain a substantial share of observations that are not firm-reported but estimated using proprietary methodologies. Second, even when vendors provide broad coverage, users typically receive a point estimate with little guidance about the observation-specific reliability of that number. As a result, the empirical carbon-risk signal bundles together two conceptually distinct components: the firm's underlying transition exposure and the uncertainty surrounding the measurement of this exposure. When a firm is labeled a "low emitter," this can mean either (i) it truly operates with low emissions or (ii) its emissions are weakly identified and the reported value is simply an imprecise reconstruction. Treating these two cases as equivalent is problematic for both inference and economic interpretation.

A key premise of our analysis is that carbon data are not only heterogeneous across firms and providers, but also intrinsically noisy, so that investors face uncertainty about firms' true transition exposure. This matters for pricing because carbon intensity captures *transition risk* (the expected cash-flow and discount-rate effects of climate policy, technological change, and demand shifts) whereas imprecision in that measure generates a distinct *information risk*. Even if two firms have the same reported intensity, the one whose footprint is measured with lower precision is harder

[1] See recent contributions by Aswani et al. (2024), Bolton and Kacperczyk (2021, 2023), and Zhang (2025).

to value and to hedge: investors cannot confidently map disclosures and vendor estimates into regulatory liability, abatement costs, or stranded-asset risk. In asset-pricing terms, such opacity acts like a wedge between fundamentals and perceptions, raising the dispersion of beliefs and exposing investors to model uncertainty. Under ambiguity aversion, this informational friction is not innocuous: when carbon intensity is both high and uncertain, investors cannot rule out severe downside realizations, so they behave as if facing a worse distribution and demand additional compensation. Hence, beyond the level of carbon intensity, carbon measurement uncertainty should be priced on its own and should also amplify the carbon premium through a state-dependent interaction between the intensity signal and its noise.

The central object of this paper is a within-method notion of *carbon uncertainty*: the firm - time-specific precision with which emissions can be inferred *given the available information*. This concept is distinct from *disagreement* across data providers documented in the ESG literature (e.g., rating disagreement highlighted by Berg et al. (2022), and subsequently used by Avramov et al. (2022) as an empirical proxy for "uncertainty"). Disagreement is informative about the lack of consensus *across* measurement systems, but it is an equilibrium outcome of combining different models, scopes, mappings, and objectives. It therefore entangles methodological heterogeneity with statistical imprecision and is only observable when multiple vendors cover the same firm-date. Our goal is different: we construct an observation-level uncertainty measure that is available for the full equity universe under a transparent and fixed construction protocol, even when cross-vendor comparisons are infeasible.

We develop a replicable imputation framework that produces both an emissions estimate and a firm-month measure of carbon uncertainty. The starting point is a simple constraint: if one wants to study carbon-related risks in the cross-section of equities, the relevant universe is not a vendor-selected subset but the full set of listed firms. We therefore impute missing emissions over a broad CRSP universe using a high-dimensional information set of firm characteristics. Because characteristics are themselves incomplete and jointly informative, we use an iterative multiple-imputation approach (MICE) to handle missing covariates and to preserve cross-variable dependence. Within each iteration, we use Random Forests to flexibly learn the conditional mapping from character-

istics to emissions, allowing for nonlinearities and interactions that are difficult to capture with parametric models. This procedure yields a point prediction for each missing firm-month emissions observation. Crucially, it also yields an uncertainty measure: we quantify the dispersion of the imputed emissions conditional on observables, leveraging the distribution of predictions implied by the algorithm (and, more broadly, by the multiple-imputation machinery). Intuitively, uncertainty is low when a firm has many close comparables and the conditional mapping is stable, and it is high when the information set is sparse, noisy, or locally inconsistent.

Our contribution is *not* to claim that our imputed level is "truer" than any given vendor estimate. Any imputation is a model-based reconstruction, and different choices may lead to different point estimates. The object we add is the *precision* attached to the reconstruction: a firm-date-specific assessment of how informative the available information is about emissions. This distinction is essential for empirical work. In asset pricing designs that treat emissions as an error-free regressor or sorting signal, heterogeneous measurement noise induces attenuation and can distort portfolio assignments. By measuring uncertainty explicitly, we can separate cases where emissions provide a sharp signal about transition exposure from cases where the signal is intrinsically weak, and we can propagate carbon measurement risk into standard empirical exercises rather than ignoring it.

Using our full-universe panel, we document that carbon uncertainty is substantial, strongly cross-sectionally heterogeneous, and time-varying, and we examine its asset-pricing implications. Carbon information quality matters for how emissions-based signals should be interpreted: emissions (or intensity) measured with high uncertainty are less informative about underlying transition exposure, and treating them symmetrically with precise observations mechanically injects noise into both regressions and characteristic-sorted portfolios. Moreover, carbon uncertainty captures an economically meaningful dimension of information frictions: firms whose emissions are difficult to infer are precisely those for which investors may face higher costs to assess transition exposure and compliance trajectories. Our evidence is consistent with the view that carbon-related pricing tests should account not only for the level of the emissions proxy but also for its reliability.

This framing also clarifies why the asset-pricing implications of *uncertainty* need not mirror

those of *disagreement*. Because vendor disagreement bundles differences in scopes, boundaries, mapping rules, estimation targets, and proprietary modeling choices, it can be high even when each vendor's estimate is internally precise (but built to measure a different object), and it can be low even when underlying emissions are weakly identified (for instance when vendors converge toward similar benchmarking conventions). Our uncertainty measure is designed to isolate a different economic primitive: the precision with which emissions can be inferred *within* a transparent and fixed protocol, conditional on observables. In our framework, uncertainty therefore maps more directly into carbon *information risk* - the difficulty investors face in assessing a firm's transition footprint and its future compliance trajectory. Consistent with this mechanism, we find that carbon uncertainty is positively priced in the cross-section and that its interaction with carbon intensity is also positive: the carbon-related premium is strongest when the emissions signal is simultaneously high and measured with high uncertainty. By contrast, disagreement across vendors is partly driven by methodological heterogeneity rather than by a common notion of conditional precision; it may therefore be a noisier proxy for the information risk that investors actually bear. This perspective offers a natural interpretation of why prior work that operationalizes "uncertainty" as cross-vendor disagreement documents economically meaningful interaction patterns, yet does not find the disagreement proxy itself to be robustly priced.[2]

These results have direct implications for investors, regulators, and data vendors. For investors, incorporating uncertainty refines emissions-based portfolio construction by distinguishing strong from weak signals and by making transparent where carbon tilts rely primarily on noisy reconstructions rather than on well-identified emissions information. For regulators, the findings suggest that comparability is not achieved by mandating a single number alone: the decision-usefulness of emissions disclosures depends on reliability, and reporting standards could require uncertainty bounds, assurance metrics, or scenario-based ranges alongside point estimates. For data vendors, the results highlight the value of reporting observation-level quality indicators rather than delivering only a single imputed series, thereby enabling users to assess (and price) the data-construction

[2]Importantly, this comparison should not be read as a contradiction: disagreement and uncertainty answer different questions. Disagreement is about the lack of consensus *across* measurement systems; uncertainty is about the precision *within* a given measurement system.

risk embedded in climate metrics.

Finally, a natural reaction to the prevalence of imputed observations is to restrict attention to disclosed emissions only, as in approaches that split samples into "disclosed" versus "estimated" (Aswani et al. (2024)). While useful as a descriptive exercise, this is not a solution to the economic problem we study. A disclosed-only sample is small and selected, and it can exclude precisely the firms for which carbon information frictions are largest and potentially most priced. More fundamentally, a binary disclosure flag does not quantify reliability: among non-disclosers, some emissions can be inferred with relatively high precision from a rich set of observables, whereas others remain weakly identified. Our framework therefore complements disclosure-based splits by delivering a continuous, firm-date measure of carbon uncertainty that can be used on the full universe and that directly targets the economic object of interest: the precision with which investors can infer transition exposure.

The rest of the paper is organized as follows. Section 2 describes the emissions sources, the characteristic set, the construction of the full-universe panel and presents the imputation procedure and the definition of carbon uncertainty. Section 3 reports the empirical properties of uncertainty and its implications for emissions-based asset-pricing tests are reported in Section 4. Section 5 provides robustness analyses and alternative constructions. Section 6 concludes with implications for climate disclosure and the design of climate data products.

2 Data

We begin by describing the carbon emissions data in Subsection 2.1 and the return and characteristic data in Subsection 2.2. Subsection 2.3 then summarizes the main sample properties. In Subsection 2.4 we detail our imputation methodology.

2.1 Carbon emissions data

Following Aswani et al. (2024), Bolton and Kacperczyk (2021, 2023), Sautner et al. (2023), and Zhang (2025), among many others, we use the *Trucost* database to measure corporate greenhouse

gas emissions for U.S. listed firms. The sample covers the period from January 2010 to December 2024 and includes ordinary common shares (CRSP share codes 10 and 11) traded on the NYSE, AMEX, or NASDAQ for which return data are available in CRSP.

Trucost provides a comprehensive set of firm-level emissions that has become a standard input in the empirical literature on climate finance. For each firm-year, the vendor reports Scope 1, Scope 2, and Scope 3 emissions following the Greenhouse Gas Protocol.[3] Scope 1 emissions capture direct emissions from a firm's own operations (e.g., combustion of fuels on site), Scope 2 reflect indirect emissions from purchased electricity, heat, or steam, and Scope 3 include other upstream and downstream emissions along the value chain (for example, those associated with suppliers, product use, or transportation). In addition, *Trucost* reports corresponding intensity measures, defined as tonnes of CO_2-equivalent per million U.S. dollars of revenue.

The vendor's raw coverage begins with fiscal year-ends in 2005 and extends through 2024. Coverage was substantially expanded around 2016 to include smaller firms, which leads to a pronounced increase in the number of available observations over time. The earliest effective observations appear only in May 2009, even for firms with fiscal year-ends in 2005 or earlier. As argued by Zhang (2025), this backfilling implies that some of the data would not have been contemporaneously available to investors at the time of the underlying fiscal reports. To avoid mixing realized and backfilled information, we start our main sample in January 2010 and align emissions with the vendor's effective (publication) date rather than the firm's fiscal year-end.

A key feature of *Trucost* is that emissions can be either disclosed by the firm or estimated by the vendor. The vendor distinguishes 28 different sourcing methods that indicate whether reported emissions are taken directly from firm disclosures (such as Annual Reports, 10-K filings, standalone environmental reports, CDP questionnaires, or direct communication), derived mechanically from disclosed quantities (for example, fuel use), or imputed based on proprietary models and industry- or revenue-based benchmarks. Many firms are consistently associated with a single sourcing method throughout the sample, while others switch methods over time. Following Aswani et al. (2024), we classify a firm as *disclosing* in a given year if the vendor does not

[3] See https://ghgprotocol.org for details on the standard.

explicitly flag its emissions as estimated. In particular, we treat as non-disclosing any firm-year observations that rely on methods such as "Estimate based on partial data disclosure," "Estimate scaled according to company-specific data," "Estimate used instead of disclosure - data do not cover global operations," or "Estimated data." Disclosing firm-year observations are those for which emissions are directly disclosed or mechanically derived from disclosed data (for example, fuel use or activity data reported in financial or sustainability reports).

For each month in the sample period, we construct firm-level variables for (log) emissions and intensities at the Scope 1, Scope 1+2, and Scope 1+2+3 levels, separately for disclosing and non-disclosing emissions. The monthly panel is obtained by carrying forward the most recent effective observation within a twelve-month window. All emissions measures are expressed in metric tonnes of CO_2-equivalent, and, unless stated otherwise, we work with natural logarithms of emissions and intensities to reduce skewness and mitigate the influence of extreme observations.

2.2 Stock returns and characteristics

We obtain monthly stock characteristics from the open-source dataset of Chen and Zimmermann (2022).[4] The database provides a broad panel of firm characteristics that have been used in the asset-pricing literature. We start from the full set of 209 monthly characteristics and merge them with CRSP monthly return and market capitalization data for all ordinary common shares with valid price and share information over January 2010 to December 2024. We follow the CRSP convention and define returns as simple returns including distributions. Excess returns are obtained by subtracting the one-month U.S. Treasury bill rate from Kenneth French data library. Market capitalization is defined as price times shares outstanding at the end of each month. We construct an additional characteristic, *Size*, defined as the natural logarithm of market value of equity, which yields a total of 210 candidate characteristics. Because many characteristics are sparsely populated, we impose a minimum coverage threshold. Specifically, for each characteristic we compute the fraction of non-missing firm-month observations over the full sample. We retain only those characteristics with at least 25% non-missing data.

[4] https://www.openassetpricing.com/

2.3 Summary statistics

Table 1 reports descriptive statistics for the *Trucost* and CRSP universes.

[Table 1 about here.]

The summary statistics in Table 1 highlight several defining features of our carbon and asset-pricing data. The *Trucost* universe is large, but disclosure is far from universal: only about one third of firms ever report emissions, and even among disclosers coverage varies widely over time. The bulk of the panel is therefore composed of vendor-estimated emissions, a fact that reinforces the need to explicitly model estimation uncertainty rather than treat all observations as equally reliable. The distribution of firm-month observations per characteristic is also highly uneven, reflecting substantial heterogeneity in data availability across firms and variables.

When we merge *Trucost* with the broader CRSP universe, the number of firms increases dramatically, but disclosure becomes even more sparse. Roughly 14% of CRSP firms ever disclose emissions. This difference is economically meaningful: once we move beyond the relatively transparent *Trucost* subsample, missing data ceases to be an inconvenience and becomes a core empirical feature. The proportion of usable firm characteristics drops accordingly, from 172 in the *Trucost* universe to 136 in the full CRSP universe, echoing the missing-data patterns documented by Chen and McCoy (2024), Freyberger et al. (2025), and Bryzgalova et al. (2025a).

Panel C shows that carbon variables are highly skewed, even in logarithms, and exhibit wide dispersion across firms. Estimated emissions are systematically lower than disclosed emissions, consistent with vendor models extrapolating from firms that tend to be larger and more carbon intensive. The spread between the 5th and 95th percentiles remains sizeable across all scopes, underscoring the presence of persistent cross-sectional heterogeneity. Together, these patterns motivate our empirical design: they justify the log transformation, highlight the economic relevance of estimation uncertainty, and reinforce the need for imputation methods capable of handling the extreme sparsity and noise inherent in carbon data at the firm level.

2.4 Predicting missing values using MICE and random forest

Empirical asset pricing studies that rely on vendor-provided environmental data face several fundamental challenges related to data construction rather than economic modeling. These challenges motivate the need for a more transparent and disciplined imputation framework. First, vendor datasets typically exhibit limited and uneven coverage, often restricting the usable universe from several thousand firms (CRSP) to only a small fraction.[5] This limited coverage implies that empirical analyses are conducted on a selected subset of firms, potentially introducing selection bias driven by data availability rather than economic forces.[6] Second, vendor imputation procedures are inherently opaque. While this opacity is understandable given their proprietary nature, researchers have little visibility into how missing values are constructed, which firm characteristics are used, or how pervasive missing inputs are handled internally. Given the fragmented nature of corporate disclosures, this lack of transparency creates substantial uncertainty about the informational content of imputed observations. Third, vendor datasets typically provide only point estimates, treated as observed data in downstream analyses despite being predictions constructed under incomplete information. As a result, uncertainty arising from data construction is ignored, even though it is distinct from economic risk and can materially affect asset pricing inference when signals rely on sparsely reported characteristics.

To sharpen the nature of this information friction, it is useful to distinguish between *model risk* and *estimation risk*, two sources of opacity that are often conflated in carbon data. Model risk reflects structural disagreement about the measurement recipe itself (boundaries, scopes, weighting schemes, or even which firm inputs should enter the mapping from observables to emissions) so that different vendors can produce conflicting numbers because they are effectively measuring different objects. Estimation risk, by contrast, persists even within a fixed and fully transparent methodology: it arises from the difficulty of calibrating a predictive relationship with finite and imperfect data. When few firms disclose within a sector, when the relevant conditioning set is high-dimensional, or when business models are atypical, the statistical link between observable

[5] For example, in the Trucost data we use to identify disclosed firms, coverage was around one thousand firms until the sample expansion in 2016, after which it increased to several thousand firms, still fewer than half of the firms in CRSP.

[6] The selection bias in carbon data is further addressed in Chen et al. (2025).

traits and emissions becomes fragile; the resulting estimate need not be "wrong" in a structural sense, but it is intrinsically uncertain because the available information is too sparse to pin down a narrow range of plausible values. In short, model risk is disagreement about *what* is being measured, whereas estimation risk is uncertainty about *how precisely* a given firm - time value can be inferred, even conditional on the agreed measurement protocol.

These limitations point to a broader challenge that any alternative imputation strategy must address. Firm-level characteristics are simultaneously incomplete, jointly informative, and linked through complex, nonlinear relationships. No characteristic is observed on a complete conditioning set, so imputations must rely on inputs that are themselves partially missing or previously imputed. A viable approach must therefore allow information to circulate across characteristics, progressively refining imputations rather than freezing early approximations, while remaining flexible enough to capture heterogeneous dependence structures. We address this challenge using a Multiple Imputation by Chained Equations (MICE) framework combined with Random Forests. The iterative structure of MICE allows each characteristic to be updated conditional on the most recent values of all others, reducing sensitivity to arbitrary initializations and yielding internally consistent imputations. Within each iteration, Random Forests provide flexible, non-parametric conditional models that accommodate nonlinearities, interactions, and mixed data types without imposing a global functional form.[7] This combination simultaneously speaks to both frictions: Random Forests relax restrictive functional-form assumptions that amplify model risk when the true dependence structure is heterogeneous and nonlinear, while a MICE approach mitigates the estimation risk via the iterative generation of multiple datasets maps the underlying sparsity of the data into a distribution of plausible outcomes. Both efforts ensure that the resulting uncertainty measure accurately reflects the noise inherent in the estimation process.

While our approach builds on the general MICE framework combined with Random Forests (RF), it fundamentally departs from standard MICE-RF implementations (Doove et al., 2014) in several economically important ways. Rather than treating imputation as a purely mechanical

[7]Elaborations on these properties of Random Forests and their applications to asset pricing can be found in Gu et al. (2020), DeMiguel et al. (2023), and Bryzgalova et al. (2025b). Note that DeMiguel et al. (2023) also use MICE to impute characteristics, but not the outcome variable of interest (mutual fund alpha).

preprocessing step, we tailor the procedure to the structure of financial panel data and to the requirements of asset pricing analysis. First, imputation is conducted month by month, using only information available at each point in time. This design eliminates look-ahead bias and ensures that constructed characteristics can be used in real-time portfolio formation and risk premia estimation, which is essential for measuring the carbon premium from cross-sectional pricing at each date. Second, we explicitly separate the imputation of firm characteristics from that of the outcome variable of interest. Characteristics are imputed iteratively until they stabilize and are then held fixed when imputing the target variable, preventing outcome-driven feedback from contaminating characteristic construction and preserving the economic interpretation of firm-level signals. Third, while the procedure operates on a panel, it deliberately avoids imposing strong time-series structure on the imputation. Environmental and fundamental data are typically reported at low frequency (annual or quarterly), are highly persistent, and often subject to revisions; in this setting, exploiting long time-series windows can create an illusion of additional information while amplifying imputation noise and bias. By focusing on cross-sectional dependence within each month, the procedure prioritizes economically relevant variation while maintaining temporal discipline. Finally, continuous and categorical characteristics are treated within a unified MICE framework, using regression and classification forests as appropriate, avoiding ad hoc practices (e.g., dropping categorical variables or forcing continuous approximations) that can distort joint dependence structures and propagate errors through the panel.

Random Forests naturally generate a distribution of predictions across trees rather than a single point estimate. We exploit this feature to construct an observation-specific measure of imputation uncertainty for each missing value, based on the dispersion of tree-level predictions. Each tree corresponds to a different local partition of the characteristic space and therefore relies on a different comparison set of firms. When the characteristics are collectively informative, tree-level predictions concentrate and dispersion is low. Conversely, when information is sparse, noisy, or conflicting, trees rely on diverging information sets, leading to higher dispersion. As such, dispersion across trees ideally captures "disagreement on the carbon neighborhood".

The appendix provides a detailed description of the algorithmic implementation, including

convergence criteria, calibration choices, and validation procedures. We also present a stylized toy example that illustrates the mechanics of the two-step imputation strategy and highlights the role of iteration, mixed data types, and observation-level uncertainty in a transparent setting.

3 Imputed Emissions, Uncertainty, and Cross-Predictability

This section presents the key empirical facts that emerge from our imputed carbon emissions dataset. The results combine distributional properties (subsection 3.2), industry-level heterogeneity (subsection 3.3) and cross-predictability (subsection 3.4).

Our imputation procedure focuses on carbon emissions, from which both the imputed means and the associated uncertainties naturally emerge as outputs of the statistical model. The uncertainty measures are not imputed separately but arise directly from the estimation process, which yields a posterior mean and posterior standard deviation for each firm-month observation. Because the literature often reports aggregate measures such as total emissions, the sum of Scope 1+2 for example, as in Sautner et al. (2023), an important question is how to quantify uncertainty for such aggregates. Inferring the uncertainty of total emissions from the separate standard deviations of individual scopes would require strong assumptions about the correlation structure of their estimates. Instead, we construct two aggregate emission variables, Scope 1+2 and Scope 1+2+3, and apply the same imputation procedure to these aggregates as we do to the individual scopes. Consequently, we report uncertainty measures for each scope separately (Scopes 1, 2, and 3) as well as for the two aggregated measures, Scope 1+2 and Scope 1+2+3.

For carbon intensity, we depart from the approach used in the *Trucost* dataset, which defines intensity as emissions scaled by revenues. Directly imputing intensity in that form would mechanically conflate the imputation of emissions with that of revenues, potentially distorting the resulting measure. To ensure full internal coherence across data sources, we therefore compute intensity *ex post* by scaling imputed emissions by total assets, a scaling variable that is theoretically as legitimate as sales or revenues for normalizing emissions. We explicitly avoid using firm size as a denominator, like Ilhan et al. (2020), since size later enters the empirical specifications as a control variable, and using it here could potentially introduce spurious correlations. As a result,

our carbon intensity measures are internally consistent with the imputation of emissions, rely on a stable and well-covered denominator, and are log-transformed for subsequent analysis.[8]

3.1 The Origins of Carbon Uncertainty

While our empirical measure of uncertainty is a standard deviation of predictions across the ensemble of trees, it conceptually aggregates two distinct sources of informational friction: input sparsity and neighbor dissonance.

The most direct driver of uncertainty is the incompleteness of the conditioning set. When a firm has a high fraction of missing characteristics, the Random Forest algorithm is forced to rely on a disjointed subset of observed characteristics across different trees. Consider the contrast between a firm with complete reporting versus one with gaps. If Firm A reports both Total Assets and Energy Consumption, almost every tree in the forest will prioritize Energy Consumption as the primary splitting variable because of its high predictive power, leading the ensemble to converge on a tight, precise estimate. Conversely, consider Firm B, a non-discloser missing key inputs like Energy Consumption or CO2 Equivalents. The algorithm must hunt for second-best proxies among the remaining variables. Consequently, Tree 1 might split on Total Assets (predicting high emissions), while Tree 2 splits on Employee Count (predicting moderate emissions), and Tree 3 splits on Intangible Assets (predicting low emissions). This divergence causes the resulting distribution of tree predictions for Firm B to become multimodal and highly dispersed. In this context, the uncertainty metric correctly identifies that the firm's emissions are under-identified by the available data.

Even when data is complete, uncertainty arises if a firm resides in a heterogeneous region of the predictor space, i.e. effectively, a "bad neighborhood" where the mapping from financial characteristics to emissions is non-unique. This friction is most acute for conglomerates and diversified holding companies, where consolidated accounting creates a phenomenon we term "Feature

[8]To maintain consistency with the imputation procedure, we retain the AM characteristic from Chen and Zimmermann (2022), defined as total assets divided by market value of equity. We construct our intensity measure by dividing imputed emissions by firm size (market capitalization), and then taking the ratio of this quantity to the AM characteristic reported by Chen and Zimmermann (2022). To avoid scale distortions and improve comparability across firms, we apply a logarithmic transformation to the resulting measure.

Blending". Financial reporting aggregates diverse business lines into a single set of characteristics (e.g., Total Assets, Leverage). This averaging process often dilutes the specific signal of carbon-intensive subsidiaries. Consider Firm C, a conglomerate that operates both a capital-intensive manufacturing division (high emissions) and a large financial services arm (low emissions). Its consolidated financial profile exhibits "hybrid" traits: high leverage (like a bank) but also significant tangible assets (like a manufacturer). Because Random Forests operate as a nearest-neighbor device, different trees will identify different "peers" for Firm C depending on which variables they prioritize at each split. Tree i, splitting on Leverage, may group Firm C into a leaf dominated by Commercial Banks, predicting negligible emissions. Tree j, splitting on Fixed Assets, may group Firm C into a leaf dominated by Heavy Industry, predicting substantial emissions. This disagreement is not a failure of the model but a precise diagnosis of structural ambiguity: the firm's consolidated financials do not map uniquely to a single industrial archetype. The high uncertainty metric correctly signals that the firm's carbon profile cannot be resolved without more granular, segment-level data.

3.2 Summary Statistics of Imputed Emissions

Table 2 reports summary statistics for imputed data in the whole CRSP universe.

[Table 2 about here.]

Average Scope 1 and Scope 3 emissions are of the same order of magnitude and dominate firms' carbon footprints: on average, they are nearly four times larger than Scope 2 emissions. The cross-sectional dispersion reinforces this hierarchy. Scope 1 exhibits the widest spread in the distribution, consistent with the large heterogeneity in firms' direct operational activities, whereas Scope 3, despite capturing broad value-chain effects, shows markedly less variability across firms.

A striking feature of Panel A is the systematic gap between imputed and disclosed emissions. Across all scopes, imputed emissions are consistently higher in levels. Yet, and this is the key empirical pattern, the cross-sectional standard deviation of imputed emissions is *lower* than that of disclosed emissions. This combination of *higher means* and *tighter pooled dispersion* reflects the

nature of imputation on levels: the forest-based procedure corrects downward biases in disclosed emissions while shrinking extreme or idiosyncratic observations toward neighborhoods determined by firm fundamentals. In short, imputation enlarges the carbon footprint in a structurally coherent manner while simultaneously regularizing the cross-section.

The exact same qualitative pattern emerges for carbon intensity. Imputed intensities exceed their disclosed counterparts by an even wider margin, highlighting the extent to which missing or partially reported emissions can distort raw intensity measures. Although the reduction in cross-sectional standard deviation is less pronounced than for emissions, the upward shift in intensity levels is more marked, indicating that intensity is particularly sensitive to missing information and benefits significantly from model-based reconstruction. Together, these patterns show that imputation does not merely "fill in'' missing values; it reshapes the carbon data landscape by correcting systematic underreporting and imposing economically meaningful structure across firms.

Panel B reveals a second set of patterns that are equally striking. First, the cross-sectional dispersion of uncertainty is remarkably stable across all variables: the pooled standard deviations of emissions uncertainty and intensity uncertainty are similar in magnitude, and they are also comparable to the dispersion of the underlying levels. This indicates that uncertainty is not driven by a handful of extreme firms but is instead a pervasive cross-sectional feature. What distinguishes uncertainty from the levels themselves is its *scale*: for Scope 1, the average level of uncertainty is roughly twice as large as the average emission level, highlighting the substantial informational frictions that remain even after imputation. This gap is economically meaningful and entirely consistent with the structure of the data environment: disclosed emissions come from a relatively small subset of firms, and the imputation procedure must extend this information to a much larger universe. The fact that average uncertainty exceeds average emissions for Scope 1 is therefore not a failure of the method but rather an accurate quantification of informational scarcity.

More interestingly, this gap between uncertainty and levels *shrinks monotonically* as we move across scopes. For Scope 1, the ratio of uncertainty to emissions is close to 2; for Scope 3, the ratio falls below 1. This pattern mirrors the hierarchy already observed in the levels. Scope 3 emissions,

though conceptually broad, are less heterogeneous across firms and are more easily predicted from firm fundamentals, leading to lower relative uncertainty. A second pattern emerges when we move from individual scopes to aggregated ones. Mechanically, emissions increase from Scope 1 to Scope 1+2 and Scope 1+2+3. Yet the average uncertainty does *not* increase proportionally. Instead, the uncertainty-to-level ratio steadily decreases and converges toward 1. The reason is that Scope 2 carries substantially less idiosyncratic noise: its own uncertainty is lower, so aggregating Scope 2 with Scope 1 reduces the relative weight of the high-uncertainty component. Adding Scope 3 reinforces this effect, pushing the uncertainty ratio even closer to parity.

In sum, Panel B shows two robust regularities: (i) uncertainty is large in absolute terms, especially for Scope 1, and (ii) the relative importance of uncertainty declines steadily as we move to broader scopes, reflecting the stabilizing effect of aggregating emissions components with progressively lower idiosyncratic variance. Far from undermining the value of the imputed data, this structure reinforces a key insight: uncertainty is largest exactly where the underlying information is scarcest, and it diminishes as the imputation leverages more stable components of the carbon footprint.

A natural robustness check is to restrict the analysis to the Trucost universe and compare our imputed values directly with the emissions that Trucost itself estimates for non-disclosers.[9] This comparison yields two insights. First, when we focus solely on firms covered by Trucost and contrast our imputed values with their *disclosed* emissions, we recover exactly the same pattern documented in the full CRSP universe: imputed emissions are systematically *higher* in levels and exhibit *lower* cross-sectional dispersion than disclosed emissions. This mirroring of results across the two universes confirms that the upward shift and regularization we observe are intrinsic to the imputation mechanism rather than artifacts of the broader CRSP sample composition.

Second, and perhaps more revealing, the behavior of Trucost's own *estimated* emissions differs fundamentally from ours. For the very same Trucost firms, vendor-estimated emissions lie *below* disclosed emissions on average and display *greater* cross-sectional dispersion, the exact opposite of the pattern produced by our approach. Whereas our method reconstructs emissions by borrowing

[9] See Table OA.3.

strength from firms with similar operational and financial structure, leading to higher, smoother, and more coherent levels, Trucost's estimation procedure appears to apply conservative, revenue-based or intensity-factor heuristics that depress the level and amplify cross-sectional noise. The opposing signs of both the level gap and the dispersion gap provide compelling evidence that the forest-based imputation captures structural relationships in the data that vendor estimation does not.

Figure 1 highlights a key stylized fact: in the raw series, the posterior uncertainty is almost perfectly comoving with the level of emissions (and intensity), with time-series correlations typically in the 0.92 - 0.99 range across scopes. This near-collinearity indicates that the uncertainty measure largely inherits the same economic and informational forces that drive the level - as firms become cleaner, smaller, or disclose more regularly, both the magnitude of emissions and the precision with which they can be inferred improve in tandem. Accordingly, raw uncertainty is ill-suited for joint inclusion with levels in predictive regressions. Normalizing uncertainty by the corresponding level breaks this mechanical link: correlations drop substantially (to around 0.40 on average, and near 0.10 in parts of the sample), so scaled uncertainty captures *relative imprecision* rather than tracking the level. The resulting measure provides an additional, economically meaningful dimension of the carbon information set, separating informational precision from emissions magnitude and enabling cleaner downstream econometric tests.

[Figure 1 about here.]

3.3 Industry Analysis

This subsection develops a sectoral perspective using the Fama–French 48 industry classification. We present two complementary rankings. Table 3 reports the ten industries with the highest average Scope 1 emissions (Panel A) and the ten industries with the highest uncertainty-to-level ratios (Panel B). Table 4 provides the analogous rankings for carbon intensity. Because emissions and intensity reflect economic and technological features while uncertainty reflects the information environment, the two tables, each with its dual ranking, provide a clean decomposition of the economic versus informational drivers of carbon data.

[Table 3 about here.]

[Table 4 about here.]

Although some industries may appear sparsely populated, this sparsity is a structural feature of the U.S. industrial landscape rather than an artifact of our procedure. The historical Fama–French 48 portfolios show the same pattern:[10] industries such as Gold, Coal, Steel, or Mines are structurally small, whereas others, such as Drugs, contain several hundred firms. The difference in firm counts between our sample and the original FF48 portfolios arises because the latter intersect CRSP with Compustat and apply additional data filters, whereas our imputation framework allows us to retain a broader cross-section by reconstructing missing firm fundamentals.

The sectoral rankings that emerge from our imputations are both intuitive and revealing. In Panel A of Table 3, the highest-emitting firms belong to industries with well-known carbon-intensive technologies: *Utilities* (9.23 $MtCO_2e$ per firm), *Oil* (4.38), *Transportation* (4.12), *Gold* (4.11), and *Mines* (4.09). These magnitudes are firm-level averages rather than sector totals, but they nonetheless mirror the macro evidence from the EPA, which identifies electric power generation, fossil fuel extraction, and heavy manufacturing as the largest contributors to U.S. greenhouse gas emissions. Smaller sectors in terms of firm count, such as *Coal* (11 firms) and *Soda* (3 firms), also display high emissions for the typical firm, consistent with the concentrated and energy-intensive nature of their production processes.

Panel B shows that uncertainty behaves very differently. The industries with the highest uncertainty ratios are *Insurance* (1.93), *Guns* (1.89), *Banks* (1.87), *Beer* (1.84), and *Aero* (1.83). Here, a ratio near 1.9 means that the posterior standard deviation is almost twice the mean firm-level emissions, reflecting sparse disclosure and weak carbon information within these sectors. The ranking by uncertainty is therefore almost orthogonal to the ranking by emissions: sectors that emit heavily are not necessarily those where emissions are informatively disclosed or easily inferred.

[10] See Table OA.4, which reports the average number of firms per FF48 industry based on Kenneth French's industry portfolios. Several industries contain fewer than 30 firms on average, and many others have fewer than 100, reflecting the intrinsic imbalance of the U.S. industrial structure.

A similar structure emerges for carbon intensity. In Table 4 Panel A, the highest average intensities occur in *Soda* (9.65), *Drugs* (8.90), *Mines* (8.86), *Agriculture* (8.75), and *Ships* (8.63). Again, these are firm-level averages and highlight that high emitters need not be the most intensity-intensive, a distinction driven largely by variation in firm size (total assets) across industries. Uncertainty in intensity (Panel B) closely mirrors the uncertainty ranking for emissions, reinforcing that uncertainty is driven primarily by the information environment rather than by the level of emissions or intensity.

Taken together, these patterns show that firm-level carbon metrics exhibit strong and economically meaningful sectoral structure, while their associated uncertainty captures a distinct informational dimension. This separation between magnitude and uncertainty is essential for the asset-pricing tests that follow.[11]

3.4 Cross-predictability

We conclude our empirical description by examining the informational relationship between carbon intensity, its associated uncertainty, and the standard firm characteristics used throughout the asset-pricing literature. Table 5 reports the panel regressions for Scope 1 intensity and Scope 1 intensity uncertainty; the remaining scopes for intensity, and all scopes for emissions are provided in the Appendix.[12]

[Table 5 about here.]

Two findings stand out. First, firm characteristics explain the *level* of carbon intensity extremely well. Depending on the specification, the R^2 ranges from 0.72 to 0.77, indicating that intensity is tightly linked to economic fundamentals such as profitability, book-to-market, size,

[11] We report in the Online Appendix the time-series patterns at the sector level for carbon emissions (Figure OA.3), carbon intensity (Figure OA.4), carbon emissions uncertainty (Figure OA.5) as well as carbon intensity uncertainty (Figure OA.6).

[12] The dependent and independent variables are standardized so that the results are directly informative to those in our return regressions, where all regressors, including emissions or intensity and their associated uncertainty, are likewise standardized. Controls correspond to Beta, BM, roaq, AssetGrowth, Mom12m, BookLeverage, VolSD, EarningsSurprise and GrSaleToGrInv. Size and Volume have been constructed using CRSP data on market capitalization and trading volume. Size is the log of market capitalization and Volume is the ratio of the traded volume in shares over total shares outstanding.

growth, and momentum. This is consistent with the evidence in the literature, which shows that emissions scale strongly with economic activity and firm size. In our setting, intensity remains highly predictable even though it is defined as emissions scaled by total assets rather than market capitalization; market capitalization still enters with a strong negative coefficient, as expected, because larger firms tend to operate with proportionally larger asset bases.

Second, and in sharp contrast, the *uncertainty* of carbon intensity is only weakly related to firm characteristics. The R^2 falls by a factor of nearly seven when moving from intensity levels to intensity uncertainty (from roughly 0.75 to barely 0.11). This shows that imputation uncertainty captures a dimension of information that is almost orthogonal to firms' observable attributes. The variable is therefore not a noisy proxy for size, profitability, or any other known cross-sectional determinant; rather, it reflects features of the disclosure environment and the informativeness of firms' carbon data, dimensions that are inherently absent from the standard characteristic set.

Columns (5)–(8) reveal a striking contrast between the cross-sectional and within-firm relationship of carbon intensity and uncertainty. In the pooled specifications without fixed effects (Columns 5 and 7), we find no statistically significant association between carbon intensity and uncertainty. This lack of cross-sectional predictability indicates that knowing a firm is "brown" (high intensity) or "green" (low intensity) does not, in isolation, predict the precision of its carbon estimate. Likewise, knowing that a firm is subject to high or low carbon information risk does not systematically translate into a specific level of carbon intensity. However, the inclusion of firm fixed effects (Columns 6 and 8) dramatically alters this picture, uncovering a strong, negative, and bidirectional relationship within the firm. Once time-invariant structural differences are absorbed, the results reveal a fundamental tension between the magnitude of the carbon signal and the scarcity of carbon information. This finding highlights a critical 'signal-to-noise' trade-off in asset pricing: as the intensity of the carbon signal increases within a firm, the relative noise of measurement uncertainty diminishes, thereby sharpening the market's ability to price transition risk.

This null result in Columns 5 and 7 likely stems from structural heterogeneity, where opposing forces effectively cancel out in the pooled sample. For instance, a high-intensity Utility firm may

exhibit very low uncertainty (because burning coal is easy to measure), while a high-intensity conglomerate with complex supply chains may have high uncertainty. In the pooled sample, these opposing forces cancel out, leaving the relationship statistically indistinguishable from zero.

However, the inclusion of firm fixed effects in columns (6) and (8) dramatically alters this picture, uncovering a strong, negative, and bidirectional relationship within the firm. When referring to intensity to uncertainty (column 8), the effect is most powerful in this direction. The coefficient of -9.76 implies that when a specific firm becomes more carbon intensive relative to its own average, its measurement uncertainty declines sharply. Economically, this supports a "signal-to-noise" interpretation: as a firm scales up its carbon-heavy activities (e.g., expanding manufacturing or transportation), the physical "signal" of emissions becomes stronger and easier to detect relative to the statistical noise of estimation. Conversely, as a firm decarbonizes (lowering intensity), its residual footprint may become dominated by harder-to-measure activities (e.g., R&D, office heating), paradoxically increasing the relative uncertainty of the estimate.

For uncertainty to intensity (column 6), we observe a symmetric negative pattern in the reverse direction -5.65. This suggests that periods of deteriorating information quality (high uncertainty) are contemporaneously associated with lower reported intensity. This is consistent with the idea that opacity often masks the true extent of "brown" activities; as the information environment clears, the full extent of carbon intensity is revealed, leading to higher observed values (Chen et al. (2025)).

By contrasting the results with and without fixed effects, we demonstrate that carbon uncertainty is not merely a static characteristic of 'dirty' or 'clean' industries, a source of heterogeneity typically absorbed by industry fixed effects in the existing literature. Instead, it functions as a dynamic state variable that evolves alongside the firm's operational lifecycle. The data strongly rejects the hypothesis that lower-intensity firms are inherently 'transparent.' Rather, the within-firm evidence suggests that carbon uncertainty acts as a mask on carbon intensity: as the precision of measurement improves, the firm's true carbon intensity becomes manifest.

The patterns documented for Scope 1 intensity and uncertainty replicate almost identically for all other scopes (Scope 2, Scope 3, Scope 1+2, Scope 1+2+3), confirming the robustness of our

cross–predictability results.[13] One might worry, however, that the high R^2 obtained for *intensity* implies that the imputation model is merely reproducing trivial linear relations already embedded in firm characteristics.

This concern disappears once we return to the *numerator*. When the same regressions are run on *emissions levels* (Tables OA.9 - OA.13), the explanatory power of characteristics collapses: the R^2 drops by a factor of five to ten, and most coefficients become economically small. In other words, characteristics explain intensity largely because intensity mechanically scales emissions by total assets, a variable closely tied to size, profitability, and leverage. But these same characteristics do a poor job predicting *emissions themselves*, and they explain even less of imputation uncertainty. The implication thus is that the imputation is far from trivial. Linear relationships among characteristics cannot reconstruct the high–dimensional, nonlinear structure of emissions or their uncertainty. The forest-based procedure adds information that the standard panel framework simply cannot capture, making the uncertainty measure genuinely new rather than redundant with the controls.

4 Should investors care about carbon uncertainty?

A central motivation for constructing a firm-level measure of carbon *uncertainty* is to assess whether markets price the informational frictions inherent in the carbon data environment. If firms differ not only in their carbon footprint but also in how *precisely* that footprint can be inferred from public information, then investors face heterogeneous information risk. Whether such uncertainty commands a risk premium, or is instead orthogonal to expected returns, is ultimately an empirical question.

The uncertainty estimated by our imputation procedure is particularly well suited for this purpose. Unlike the disclosed versus estimated classification used in existing work, our uncertainty measure captures fine-grained, firm-specific dispersion in the posterior distribution of imputed carbon outcomes. It therefore reflects the quality of the carbon information set rather than simple disclosure status. Moreover, by scaling uncertainty by the corresponding carbon level, we extract

[13] See Tables OA.5 - OA.8.

the component of uncertainty that is *independent* of the magnitude of emissions or intensity, isolating informational frictions from economic size effects.

This section investigates the asset-pricing implications of carbon uncertainty. We begin with a set of panel regressions relating stock returns to Scope 1 emissions, their scaled uncertainty, and their interaction, thereby evaluating whether uncertainty amplifies, attenuates, or reverses any return predictability associated with emissions themselves. We present the fully saturated specifications including firm and industry-time fixed effects and a rich set of standard return predictors to test whether any explanatory power of carbon uncertainty survives after controlling for firm fundamentals and time-varying macroeconomic forces. We also provide complementary evidence based on portfolio sorting. The latter is indispensable to disentangle two plausible explanations for the state dependence of the carbon premium on the level of uncertainty.

4.1 Panel evidence

We now present the panel regression results linking stock returns to carbon intensity and its scaled uncertainty. Table 6 reports the findings for Scope 1 which is the core component of firms' direct operational emissions and the most widely studied measure in the climate finance literature. To capture the multidimensional nature of carbon risk, Table 6 incorporates both carbon intensity and carbon uncertainty, along with their interaction term, into our return prediction model. This specification allows us to test whether the market prices the level of carbon exposure differently from the precision of that information. By including firm and industry-time fixed effects, the model isolates within-firm variation, ensuring that our results are driven by changes in a firm's specific risk profile rather than static industrial differences.

[Table 6 about here.]

Columns (1) through (3) provide strong evidence that carbon intensity and carbon uncertainty capture distinct sources of risk. When included jointly (Column 3), both variables retain their statistical significance and economic magnitude. This stability indicates that while the two variables are correlated within-firm (as shown in Table 5), they are not redundant. Intensity proxies

for transition risk exposure, whereas uncertainty proxies for information risk, the ambiguity surrounding that exposure. The fact that both remain significant suggests that investors price the "noise" (uncertainty) separately from the "signal" (intensity). We stress that the results from Table 5 and 6 do not contradict each other. Although high-intensity firms may structurally have lower measurement uncertainty, the market still demands a premium for whatever residual uncertainty remains. By including both, the model separates the market's reaction to "fundamental" shifts in emissions (the physical transition risk) from "informational" frictions (the precision of the estimate). The magnitudes are economically substantial. A one–standard–deviation increase in Scope 1 intensity is associated with an increase in expected returns of roughly 1.30% per month, while a comparable increase in scaled uncertainty raises expected returns by about 0.21%.

We include the interaction between carbon intensity and measurement uncertainty to allow the carbon premium to be state dependent. While our earlier results show that uncertainty is priced per se (i.e., it carries an independent information-risk premium) it may also shape how markets price transition exposure. The sign of this interaction is theoretically ambiguous. On the one hand, under ambiguity aversion, a firm that is both highly carbon intensive and measured imprecisely is precisely the case in which investors cannot bound the worst-case transition scenario.[14] When intensity is high but uncertain, investors have greater difficulty mapping reported footprints into regulatory liability, abatement costs, or stranded-asset risk, and they may therefore demand additional compensation; in this mechanism, uncertainty amplifies the carbon premium and the interaction should be positive. On the other hand, as emphasized in the uncertainty-based explanations in Avramov et al. (2022), poor information can blunt the market's ability to discriminate across firms: when emissions data are noisy, the perceived differences in carbon exposure become less reliable, which can effectively provide "cover" to highly intensive firms and attenuate the pricing of intensity. In that case, the carbon premium is weaker when uncertainty is high and the interaction should be negative. The interaction term is thus designed to capture whether measurement uncertainty acts as an amplifier of transition risk or, alternatively, as an

[14] Rocciolo et al. (2024) study the link between general ambiguity and climate-risk relying on perfectly observed vendor-provided carbon emissions. In contrast, our uncertainty measure is directly tied to carbon emissions themselves, thereby capturing carbon-specific uncertainty rather than general ambiguity.

informational veil that dampens the market's response to carbon intensity.

Without uncertainty (Column 4), the interaction term is irrelevant. The coefficient is positive and significant once uncertainty is included (Column 5), supporting the hypothesis of ambiguity aversion. This suggests that uncertainty acts as an "amplifier": the risk premium demanded for holding carbon-intensive assets is significantly higher when the precise level of that intensity is difficult to verify. Economically, this implies that investors are particularly averse to the "worst-case" scenario, i.e. assets that are both dirty and opaque, demanding a cost of capital that exceeds the sum of the individual risk premiums.

The robustness of the return–uncertainty relation is best assessed by extending the analysis beyond Scope 1 to Scope 2, Scope 3, and the aggregate measures (Scope 1+2 and Scope 1+2+3).[15] A striking pattern emerges. When taken *individually*, both carbon intensity and its scaled uncertainty remain strongly and positively priced across all scopes. For every definition of carbon intensity, from direct operational emissions (Scope 1) to purchased energy (Scope 2) and value-chain emissions (Scope 3), the standalone coefficient on intensity is positive and statistically significant, and the standalone coefficient on uncertainty is also significant with the expected sign. The same holds for the the combined measure Scope 1+2 and Scope 1+2+3 even when including both variables. In this sense, the evidence that carbon uncertainty carries a return premium is remarkably robust.

Once we introduce the interaction between intensity and uncertainty, the picture becomes more nuanced. For Scope 1+2 measure, the interaction term is not significant but both main effects remain statistically significant and economically meaningful, indicating that the uncertainty premium is not subsumed by the level of emissions. In contrast, for Scope 1+2+3, the main effect of uncertainty becomes less stable: conditional on intensity (and the interaction), its coefficient loses statistical significance, while the intensity premium remains positive and robust while the interaction term is significant and negative. A natural interpretation is that, as we move toward Scope 3, the informational environment deteriorates and measurement error becomes more severe and less structured. This makes it harder to separately identify a linear "uncertainty premium"

[15] See Tables OA.14 - OA.17.

once the model allows the price of carbon risk to vary with uncertainty. Mechanically, the interaction term also absorbs part of the variation previously attributed to the uncertainty main effect and can induce sign changes when intensity, uncertainty, and their product are highly correlated or when uncertainty mainly matters in high-intensity states.

Turning to emissions levels (in logs), the evidence is particularly striking. When emissions are entered on their own, their level plays only a limited role and is often not statistically significant across scopes: in sharp contrast, emission measurement uncertainty is consistently positively priced for Scope 1, Scope 2, Scope 3, and the aggregated measures (Scope 1+2 and Scope 1+2+3). Importantly, this positive uncertainty premium survives when emissions and uncertainty are included jointly, reinforcing the central message that, for emissions, it is not the level that is priced but the imprecision with which that level is measured. Once we allow for state dependence by introducing the interaction between emissions and uncertainty, coefficients become more volatile, as in the intensity specifications, yet a coherent pattern emerges for Scope 1, Scope 1+2, and Scope 1+3. In these cases, adding the interaction term reveals that logged emissions turn statistically significant and negative, and the interaction itself is also significant and negative (with uncertainty remaining significant). This configuration suggests that the effect of emissions on expected returns is inherently state dependent. Put differently, ignoring the interaction amounts to imposing a constant price of emissions risk, which can wash out significance when the market primarily reacts to emissions conditional on the degree of measurement uncertainty.

By scaling emissions by total assets, intensity removes the level effect and turns carbon exposure into an efficiency measure. This scaling makes intensity directly comparable across firms of different size and embeds information about technology, production processes, supply-chain structure, and transition vulnerability. Because these characteristics are only partially captured by the controls, the uncertainty associated with intensity retains a systematic component: it reflects uncertainty about a firm's *carbon efficiency*, not merely its scale. This uncertainty is economically meaningful for investors and therefore priced. In contrast, uncertainty about the *amount* of emissions a firm produces, after size effects are absorbed, is largely idiosyncratic and does not command a premium.

Taken together, the interaction results are difficult to interpret. The coefficient on $Intensity \times Uncertainty$ is unstable, and its inclusion affects the level and uncertainty estimates in scope-specific ways (Scopes 1, 2, 3, and the aggregated measures), yielding patterns that are not fully consistent across specifications. A plausible unifying explanation is econometric rather than economic: the interaction term is mechanically highly correlated with its components (especially with intensity) so adding it induces severe multicollinearity. Figure 2 reports these correlations. In that setting, the regression has limited power to separately identify the main effects and the interaction, making signs and significance sensitive to small changes in the information set or scope definition. As a result, while the panel evidence cleanly establishes that intensity and measurement uncertainty are priced, it is less informative about the state dependence of the carbon premium and, in particular, about whether uncertainty amplifies transition risk (ambiguity-aversion channel) or instead dampens its pricing by obscuring cross-firm differences (the information-attenuation channel). To sidestep this identification problem, the next section turns to portfolio sorts. By conditioning on intensity and uncertainty nonparametrically, sorts mitigate the multicollinearity inherent in interaction regressions and provide a more transparent assessment of how the carbon premium varies across informational states.

[Figure 2 about here.]

4.2 Portfolio Sorting

While panel regressions allow us to control flexibly for multiple characteristics, they rest on a conditioning structure that differs from the canonical portfolio-sorting approach widely used in asset pricing. In panel specifications, characteristics are controlled for directly through regressors. In contrast, portfolio-sorting controls for characteristics through portfolios, which implicitly define factors and expose the pricing structure in a more non-parametric and model-free way. Because these two approaches rely on different identification philosophies, and are known in the literature to sometimes yield diverging implications, it is natural to complement the panel evidence with a double-sorting exercise.

We therefore perform a double-sorting of all CRSP stocks into five-by-five portfolios based on

intensity (first-stage buckets) and carbon uncertainty (second-stage buckets). We focus the main discussion on Scope 1 intensity, with results for other Scopes reported in the appendix.[16]

[Table 7 about here.]

A striking pattern emerges: the high-minus-low (H - L) portfolio in carbon intensity is positive and statistically significant across all uncertainty quintiles, confirming the strong pricing of intensity. However, the magnitude of the carbon premium declines almost monotonically as uncertainty increases. For instance, in the low-uncertainty quintile, the H - L intensity spread reaches 3.17% per month, whereas in the high-uncertainty quintile it falls to only 1.71% per month. When uncertainty is high, investors' ability to differentiate between "clean" and "dirty" firms is reduced, which mechanically attenuates the carbon premium. In the case of Scope 1, this pattern mirrors the opposite of the interaction term estimated in the panel regressions which was positive. The portfolio sorting reinforces the interpretation that uncertainty dampens the informational content of carbon emissions. This finding is robust across alternative individual or aggregate scopes.

Turning to the uncertainty dimension, the H–L uncertainty portfolios reveal no robust pricing of uncertainty *per se*. Across the five intensity buckets, most H–L uncertainty spreads are statistically insignificant, and one even carries a marginally negative sign. A natural economic interpretation is that, unlike carbon intensity (which is largely systematic and well captured by observable firm characteristics) measurement uncertainty is more granular and idiosyncratic, reflecting disclosure practices, data availability, and vendor imputation noise rather than a common underlying transition-risk exposure. If the dominant component of uncertainty is firm-specific (and weakly related to characteristics), it is largely diversifiable within the portfolios, so it need not command a sizeable standalone premium in sorts. Moreover, the uncertainty signal itself is measured with error, so sorting on it creates non-trivial misclassification that mechanically attenuates H–L spreads. In this sense, the muted portfolio evidence does not necessarily contradict the regression-based results for some scopes; rather, it suggests that any uncertainty premium may be conditional (e.g., concentrated in particular informational states or captured by interactions)

[16] See Table OA.23.

and difficult to detect with coarse, within-intensity sorts when uncertainty is primarily a noisy, non-systematic attribute.

I is worth noting that the raw high-minus-low (H - L) intensity spreads in the full CRSP universe are relatively large, both in terms of raw returns and factor-adjusted alphas. A natural explanation is that our baseline specification includes the entire cross-section of listed firms, including the very small stocks that are known to generate mechanically high average returns and contribute substantial noise to portfolio-level statistics. The summary statistics reported in Table 8 confirm this interpretation.

[Table 8 about here.]

Across the five uncertainty quintiles, the high-intensity portfolios are systematically dominated by small-cap firms. This composition effect inflates, to some extent, the absolute magnitude of the H - L intensity spreads, although it does not affect the key comparative pattern of interest, namely, the sharp contrast between the H - L spread in the low-uncertainty quintile and the much smaller spread observed in the high-uncertainty quintile.

To ensure that our conclusions are not driven by micro-cap distortions, we replicate the entire double-sorting exercise after excluding the bottom 30% of firms by market capitalization. The results, reported in the Table OA.24, show that the main message is unchanged: High uncertainty reduces the price of carbon transition risk. The magnitude is substantially lower even though it is still economically significant without small caps (0.32% (2.51%-2.19%) instead of 1.38% (3.37% - 1.99%) monthly). In the restricted universe, the anomalous negative spread disappears and the H - L uncertainty portfolios are uniformly insignificant.

Finally, as documented in the panel regressions, we do not find any meaningful pattern in the emissions-only sorts, consistent with the interpretation that carbon exposure is priced only through the intensity channel and only when investor uncertainty is low enough for emissions information to be informative.[17]

Overall, the portfolio-sorting results provide a complementary perspective to the panel regressions. The carbon premium associated with emission intensity is economically strong, but it is

[17] See Tables OA.25 and OA.26.

also state dependent. In particular, the evidence suggests that measurement uncertainty acts as the conditioning state: when carbon information is noisier, investors' ability to discriminate across firms weakens and the pricing of carbon intensity becomes less sharp. This pattern is consistent with the view that fuzzy carbon information does not simply add a separate premium, but instead distorts how transition exposure is mapped into expected returns, shaping the strength of the carbon premium across informational environments.

5 Further Analysis

5.1 Expanding the Set of Firm Characteristics: Trucost-restricted imputation

A central feature of our framework is that both carbon emissions and their associated uncertainty are inferred from a high-dimensional set of firm characteristics. In the baseline specification, we retain only characteristics observed for at least 25% of CRSP firms, which reduces the usable set from 210 raw variables to 136. While this restriction limits the extent of imputation performed on characteristics themselves, it raises the question of whether our results are sensitive to the richness of the predictor space.

To address this concern, we re-estimate the full procedure within the Trucost universe only. Because data coverage is lower in Trucost relative to CRSP, this restriction allows us to expand the characteristic set to 172 variables. Importantly, this exercise differs from merely evaluating baseline imputations on Trucost firms: both estimation and imputation are now conducted exclusively within Trucost. This design increases the effective informational content of the predictors and provides a clean robustness check with respect to the dimensionality of the characteristic space.[18]

Expanding the characteristic space leads to a clear level effect. Across all scopes and aggregation schemes, imputed emissions and intensities are uniformly higher. This upward shift is consistent with a richer information set allowing the imputation model to better correct for

[18]Figure OA.7 reports the time-series behavior of imputed emissions, carbon intensity, and uncertainty under Trucost-restricted estimation, while Figure OA.8 shows the corresponding series obtained from CRSP-wide imputation. Summary statistics are reported in Table OA.27, directly comparable to Panel B of Table OA.3.

downward vendor bias and to exploit finer cross-sectional structure. In contrast, the time-series dynamics are virtually unchanged: in both specifications, emissions and intensity decline markedly after 2016, indicating that the temporal evolution reflects underlying economic forces rather than artifacts of the imputation design.

Uncertainty exhibits a distinct but coherent pattern. While its time-series behavior remains stable across specifications, its level declines monotonically as the imputation procedure incorporates additional information. This reduction is consistent with a richer predictor set explaining more systematic heterogeneity and leaving less unexplained dispersion. Moreover, when uncertainty is measured using the ratio-based metric, its correlation with emissions and intensity shifts from slightly negative in the baseline specification to slightly positive in the richer Trucost setting. Although small in magnitude, this change suggests that the richer model disentangles true variation from mechanical compression induced by limited predictors.

Further insight emerges from examining cross-predictability between intensity and uncertainty. Table OA.28 reports the Trucost-restricted results, while Table OA.29 provides the corresponding results for Trucost when learning for the whole CRSP universe is based on disclosed firms only. In the richer specification (former one), firm characteristics retain strong explanatory power for intensity but explain uncertainty only weakly. Once characteristics are controlled for, intensity does not predict uncertainty, and uncertainty does not predict intensity. This contrasts sharply with the baseline CRSP-wide imputation, where both variables exhibited similar predictability and strong cross-dependence.

This reversal is economically informative. With a richer predictor set, the imputation absorbs the systematic component of uncertainty, leaving a more idiosyncratic residual that is no longer mechanically linked to intensity. Intensity reflects persistent technological and operational features well captured by firm characteristics, whereas uncertainty captures higher-order informational frictions that do not align closely with observable traits.

The asset-pricing implications are reported in Table OA.30, with the CRSP-wide benchmark in Table OA.31. In the Trucost-restricted specification, intensity is positively priced, uncertainty carries a negative price, and their interaction is negative and statistically significant. This pattern

indicates that uncertainty moderates rather than scales the intensity premium. By contrast, in the CRSP-wide trained setting uncertainty as well as the interaction are positively priced, consistent with the findings obtained in the whole CRSP universe.

A key implication of our subsample exercises is that the *imputation universe* is not an innocuous implementation detail, it is an economically first-order choice echoing the findings of Chen et al. (2025) about severe selection bias in carbon datasets. When we estimate the imputation model on the full CRSP universe and then restrict the return tests to the Trucost-covered subsample, the Trucost firms behave essentially like the full sample in terms of pricing. In other words, once emissions and uncertainty are inferred using the broad cross-section, the pricing patterns within the Trucost subset align with those obtained in CRSP. By contrast, when we re-estimate the imputation model *only* within the Trucost universe, the resulting pricing evidence changes markedly and no longer mirrors the full-sample conclusions. This divergence indicates that Trucost coverage is not a random slice of the market: it is a selected universe, tilted toward larger, more visible firms with systematically different disclosure incentives, reporting infrastructures, and ESG data availability. Estimating the imputation model on such a selected sample can therefore embed the selection mechanism into the inferred emissions signal itself compressing cross-sectional variation, altering the signal-to-noise structure, and ultimately changing the asset-pricing conclusions. This is precisely the selection-bias concern emphasized by Chen et al. (2025): when the observation process is endogenous (firms are covered, disclose, or are measurable for non-random reasons that correlate with economic fundamentals), imputing within the covered set can produce systematically biased proxies for the latent environmental exposure, and risk-premium estimates can inherit that bias. Our results provide direct empirical support for this warning: broad-universe imputation mitigates selection effects and yields pricing patterns that are stable across samples, whereas imputing within the vendor-covered universe can mechanically generate different "premia" that reflect coverage and measurement regimes rather than underlying transition risk.

5.2 Training on Both Disclosed and Vendor-Estimated Emissions: Sensitivity to Imputed Inputs

An additional validation exercise concerns the choice of the training sample used to construct the imputed carbon signal. In our baseline approach, we train the imputation model on firms that self-disclose emissions and then extrapolate to the broader CRSP cross-section. We now expand the training set by pooling self-disclosers with vendor-estimated observations and re-estimate the entire pipeline.

Summary statistics and time-series evidence are reported in Table OA.34 and Figure OA.9. Incorporating vendor-estimated firms mechanically lowers the level of imputed emissions, intensities, and uncertainty across all scopes. This shift reflects the fact that Trucost estimates are, on average, lower than those produced by our procedure when trained on disclosed data only. Importantly, the time-series dynamics remain unchanged: emissions and intensity decline over time, and uncertainty exhibits no secular trend beyond the level adjustment.

Cross-predictability results are reported in Table OA.35. As in the previous robustness check, characteristics strongly explain intensity but only weakly explain uncertainty. Moreover, the explanatory power of intensity for uncertainty declines sharply: the R^2 of regressions linking the two variables falls from roughly 0.30 in the baseline to about 0.03 when vendor-estimated observations are included. This attenuation indicates that uncertainty becomes increasingly idiosyncratic as the informational environment broadens.

The asset-pricing conclusions shown in Table OA.36 are essentially unchanged. At the firm level, the intensity premium remains positive and strongly significant, and in specifications that include intensity, uncertainty, and their interaction, the qualitative patterns for uncertainty replicate those obtained under disclosure-only training.

Economically, this stability is informative. It suggests that, for the purpose of learning the mapping from firm characteristics into emissions and carbon intensity, the incremental information contained in vendor estimates is limited once self-disclosures are already in the training set. A natural interpretation is that vendor-reported emissions are not an independent signal, but largely a transformed version of the same underlying disclosure infrastructure: vendors typically anchor

their estimation models on disclosed data and propagate it through industry mappings, accounting rules, and auxiliary disclosures. As a result, adding vendor-estimated observations mostly expands the sample mechanically without introducing genuinely new information about the structural relation between firm fundamentals and carbon outcomes. In an asset-pricing setting, this is encouraging for two reasons. First, it indicates that a relatively small set of disclosers can be sufficient to discipline the imputation model and recover robust pricing patterns out of sample. Second, it underscores a broader point about data provenance: when multiple carbon datasets draw on the same disclosure backbone, cross-vendor variation may reflect methodological choices rather than independent measurement, and the apparent "breadth" gained from vendor coverage can be redundant for inference about priced carbon risk.

5.3 Imputing Carbon Intensity Directly

In the baseline analysis, we construct carbon intensity as imputed emissions scaled by an imputed measure of firm scale (total assets). A natural alternative is to impute intensity directly, i.e., to treat the ratio itself as the target variable. We do not adopt this approach as our main specification because the ratio structure introduces an additional layer of non-linearity and measurement error: when intensity is defined as emissions divided by a scale variable (e.g., revenues in Trucost), noise in either component is magnified through division, and the resulting error becomes heteroskedastic and mechanically related to firm size. From an economic standpoint, this matters because the denominator is not a passive normalizer. Scale variables such as revenues embed cyclical demand conditions, accounting conventions, and strategic reporting, so imputing the ratio directly can confound true transition exposure (the numerator) with transient or noisy variation in scale (the denominator), thereby distorting the link between fundamentals and the priced carbon signal.

Table OA.38 and Figure OA.10 show that the directly imputed intensity retains broadly similar distributional features, but the informational content of standard firm characteristics deteriorates materially. Consistent with this, the cross-predictability evidence in Table OA.39 indicates that linear characteristics explain both intensity and uncertainty substantially less well than under the emissions-based construction. This loss of predictability is economically intuitive: by collapsing

two objects (emissions and scale) into a single ratio, direct imputation forces the model to learn a composite outcome whose mapping from fundamentals is intrinsically more nonlinear and more sensitive to denominator noise.

The asset-pricing implications are correspondingly less stable. Relative to the baseline, directly imputing intensity delivers markedly different return patterns in Table OA.40: while measurement uncertainty remains positively priced, the standalone pricing of intensity becomes weak and can fail to survive richer specifications, and the interaction term becomes sensitive to scope definitions and controls. This "blow-up" is consistent with errors-in-variables logic in a ratio setting: if the directly imputed intensity is a noisier proxy for true transition exposure, its estimated premium is mechanically attenuated, and any state-dependence captured by the interaction becomes difficult to interpret because the interaction inherits the same denominator-driven noise. For this reason, we view the direct-imputation strategy as an informative robustness check rather than a reliable baseline. Imputing emissions and scale separately, and forming intensity ex post, better preserves the underlying economics (emissions as the exposure, scale as the normalizer) and yields a cleaner, more interpretable decomposition between the level of transition risk and the information frictions surrounding its measurement.

6 Conclusion

This paper argues that progress in understanding the asset-pricing implications of carbon emissions requires treating measurement uncertainty as a first-order object rather than a nuisance. We jointly estimate firm-month emissions level and uncertainty by combining a Multiple Imputation by Chained Equations (MICE) framework with Random Forests for the full CRSP universe. By separating transition exposure from the reliability with which it is measured, this paper provides a step toward more coherent tests of carbon risk in asset pricing and a clearer foundation for evaluating climate-related financial signals. Much of the existing literature implicitly assumes that emissions data (whether firm-reported or vendor-estimated) are sufficiently precise to be used as sorting variables or regressors without adjustment. We show that this assumption is difficult to justify in practice. For a large share of firm-month observations, emissions are constructed under

sparse information, and the resulting noise is highly heterogeneous across firms and over time. Ignoring this heterogeneity conflates true transition exposure with the precision of its measurement, helping to explain why empirical findings on carbon premia have been mixed and often contradictory.

Our framework delivers two objects: an emissions estimate that spans the full equity universe and, critically, an observation-level measure of carbon uncertainty that quantifies how well emissions can be inferred given available information. This within-method uncertainty differs conceptually from cross-vendor disagreement and is available even when only a single construction protocol is used. Empirically, incorporating this uncertainty sharpens inference and clarifies economic interpretation. It allows researchers to distinguish between firms that are genuinely low (or high) emitters and firms whose classification is driven by weak identification, reducing attenuation bias and mis-sorting in portfolio tests. More broadly, it highlights that carbon risk is not only about exposure levels but also about the information environment surrounding those exposures.

Our findings have direct implications for investors, regulators, and data vendors. For investors, incorporating uncertainty refines emissions-based portfolio construction by distinguishing strong from weak signals. For regulators, the results suggest that disclosure regimes should emphasize reliability, not just point estimates, by encouraging uncertainty bounds or assurance metrics. For data vendors, they highlight the value of reporting observation-level quality indicators alongside emissions estimates.

References

Aswani, J., Raghunandan, A., and Rajgopal, S. (2024). Are carbon emissions associated with stock returns? *Review of Finance*, 28(1):75–106.

Avramov, D., Cheng, S., Lioui, A., and Tarelli, A. (2022). Sustainable investing with esg rating uncertainty. *Journal of financial economics*, 145(2):642–664.

Berg, F., Kölbel, J. F., and Rigobon, R. (2022). Aggregate confusion: The divergence of esg ratings. *Review of Finance*, 26(6):1315–1344.

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

Bolton, P. and Kacperczyk, M. (2023). Global pricing of carbon-transition risk. *The Journal of Finance*, 78(6):3677–3754.

Bryzgalova, S., Lerner, S., Lettau, M., and Pelger, M. (2025a). Missing financial data. *The Review of Financial Studies*, 38(3):803–882.

Bryzgalova, S., Pelger, M., and Zhu, J. (2025b). Forest through the trees: Building cross-sections of stock returns. *The Journal of Finance*, 80(5):2447–2506.

Chen, A. Y. and McCoy, J. (2024). Missing values handling for machine learning portfolios. *Journal of Financial Economics*, 155:103815.

Chen, A. Y. and Zimmermann, T. (2022). Open source cross-sectional asset pricing. *Critical Finance Review*, 27(2):207–264.

Chen, C. Y.-H., Lioui, A., and Scaillet, O. (2025). Green silence: Double machine learning carbon emissions under sample selection bias. *Available at SSRN: 5368583*.

DeMiguel, V., Gil-Bazo, J., Nogales, F. J., and Santos, A. A. (2023). Machine learning and fund characteristics help to select mutual funds with positive alpha. *Journal of Financial Economics*, 150(3):103737.

Doove, L. L., Van Buuren, S., and Dusseldorp, E. (2014). Recursive partitioning for missing data imputation in the presence of interaction effects. *Computational Statistics and Data Analysis*, 72:92–104.

Freyberger, J., Höppner, B., Neuhierl, A., and Weber, M. (2025). Missing data in asset pricing panels. *The Review of Financial Studies*, 38(3):760–802.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Ilhan, E., Sautner, Z., and Vilkov, G. (2020). Carbon tail risk. *The Review of Financial Studies*, 34(3):1540–1571.

Rocciolo, F., Billio, M., Guidolin, M., and Izhakian, Y. (2024). Pricing climate ambiguity. *Available at SSRN: 5049349*.

Sautner, Z., Van Lent, L., Vilkov, G., and Zhang, R. (2023). Firm-level climate change exposure. *The Journal of Finance*, 78(3):1449–1498.

Zhang, S. (2025). Carbon returns across the globe. *The Journal of Finance*, 80(1):615–645.

Table 1: Summary statistics of the *Trucost* and *CRSP* universe

This table presents descriptive statistics for firms covered by *Trucost*. Panel A summarizes the *Trucost* sample composition, including the number of firms, the number of disclosing firms, and the number of firm-month observations for disclosed and estimated emissions and intensities. It also reports the total number of firm characteristics and the average number of firm-month observations per characteristic. Panel B reports the same information for the *CRSP* universe. Panel C reports pooled monthly moments (mean, median, standard deviation, and selected percentiles) for excess returns, firm size, and carbon-related variables (Scopes 1–3 emissions and intensities, both disclosed and estimated). All variables are measured at a monthly frequency, and the sample period spans January 2010 to December 2024.

Panel A: *Trucost* **sample composition and coverage**

Number Firms	3765
Number Disclosing Firms	1423
Number of Firm/month Observations for emissions	211603
Number of Firm/month Observations for disclosed emissions	55639
Number of Firm/month Observations for estimated emissions	155964
Number of Characteristics	172
Mean Number of Firm/month Observations per Characteristic	379042
Minimum Number of Firm/month Observations per Characteristic	171009
Maximum Number of Firm/month Observations per Characteristic	472294

Panel B: *CRSP* **sample composition and coverage**

Number Firms	9845
Number of Disclosing Firms	1423
Number of Firm/month Observations for emissions	211603
Number of Firm/month Observations for disclosed emissions	55639
Number of Firm/month Observations for estimated emissions	155964
Number of Characteristics	136
Mean Number of Firm/month Observations per Characteristic	671923
Minimum Number of Firm/month Observations per Characteristic	445090
Maximum Number of Firm/month Observations per Characteristic	816033

Panel C: Carbon variable-level summary statistics in Trucost Universe

	Mean	**Std**	**5%**	**25%**	**50%**	**75%**	**95%**
Excess Returns (%)	1.09	14.81	-19.96	-5.87	0.61	7.05	22.79
Market Capitalization	14.11	1.96	11.05	12.70	14.05	15.39	17.50
Scope 1 Emissions (Disclosed+Estimated)	9.29	3.10	4.25	7.15	9.38	11.27	14.67
Scope 1 Disclosed Emissions	11.55	2.88	7.11	9.65	11.34	13.42	16.52
Scope 1 Estimated Emissions	8.49	2.76	4.03	6.52	8.62	10.44	12.79
Scope 1 Intensity (Disclosed+Estimated)	2.51	1.83	0.04	1.30	2.53	3.24	6.11
Scope 1 Disclosed Intensity	2.99	2.37	-0.22	1.16	2.60	4.54	7.60
Scope 1 Estimated Intensity	2.34	1.55	0.06	1.34	2.52	3.11	5.22
Scope 2 Emissions (Disclosed+Estimated)	9.49	2.68	4.65	7.74	9.73	11.32	13.63
Scope 2 Disclosed Emissions	11.61	2.05	8.16	10.35	11.65	13.05	14.67
Scope 2 Estimated Emissions	8.73	2.46	4.36	7.11	9.03	10.51	12.37
Scope 2 Intensity (Disclosed+Estimated)	2.63	1.21	0.44	1.89	2.70	3.44	4.50
Scope 2 Disclosed Intensity	2.93	1.37	0.60	2.05	2.94	3.81	5.20
Scope 2 Estimated Intensity	2.52	1.14	0.42	1.86	2.61	3.32	4.14
Scope 3 Emissions (Disclosed+Estimated)	11.35	2.56	7.13	9.69	11.58	13.15	15.18
Scope 3 Disclosed Emissions	13.54	1.68	10.80	12.42	13.57	14.66	16.14
Scope 3 Estimated Emissions	10.57	2.36	6.78	9.07	10.75	12.27	13.99
Scope 3 Intensity (Disclosed+Estimated)	4.41	0.96	3.13	3.57	4.35	5.15	6.00
Scope 3 Disclosed Intensity	4.77	0.95	3.24	3.99	4.81	5.53	6.25
Scope 3 Estimated Intensity	4.28	0.94	3.13	3.43	4.17	4.98	5.91

Table 2: Summary Statistics of Imputed Emissions and Uncertainty

This table presents descriptive statistics for the fully imputed *CRSP* sample. Panel A reports pooled monthly statistics for the levels of carbon emissions and intensities (Scopes 1–3,1+2,1+2+3, disclosed and estimated). Panel B reports pooled monthly statistics for the uncertainty of imputed of emissions and intensities (Scopes 1–3,1+2,1+2+3, disclosed and estimated). All variables are measured at a monthly frequency and are log-transformed.

Panel A: Carbon variables summary statistics: levels.

	Mean	**Std**	**5%**	**25%**	**50%**	**75%**	**95%**
Excess Returns (%)	0.57	16.72	-23.26	-6.76	0.20	6.75	24.69
Market Capitalization	12.72	1.71	10.49	11.66	12.44	13.48	16.12
Scope 1 Emissions (Disclosed+Imputed)	14.14	1.24	12.37	13.40	14.15	14.98	15.98
Scope 1 Disclosed Emissions	11.55	2.88	7.11	9.65	11.34	13.42	16.52
Scope 1 Imputed Emissions	14.23	1.05	12.54	13.45	14.18	14.99	15.97
Scope 1 Intensity (Disclosed+Imputed)	7.88	2.34	3.90	6.65	8.00	9.34	11.41
Scope 1 Disclosed Intensity	2.34	2.96	-2.88	0.62	2.32	4.43	6.92
Scope 1 Imputed Intensity	8.06	2.08	4.56	6.79	8.07	9.39	11.45
Scope 2 Emissions (Disclosed+Imputed)	12.67	0.85	11.28	12.20	12.69	13.18	13.98
Scope 2 Disclosed Emissions	11.61	2.05	8.16	10.35	11.65	13.05	14.67
Scope 2 Imputed Emissions	12.70	0.75	11.37	12.22	12.70	13.18	13.96
Scope 2 Intensity (Disclosed+Imputed)	6.40	1.98	2.92	5.31	6.51	7.65	9.47
Scope 2 Disclosed Intensity	2.40	1.98	-1.22	1.29	2.57	3.66	5.22
Scope 2 Imputed Intensity	6.53	1.84	3.37	5.45	6.57	7.69	9.50
Scope 3 Emissions (Disclosed+Imputed)	14.13	0.78	12.86	13.71	14.10	14.59	15.42
Scope 3 Disclosed Emissions	13.54	1.68	10.80	12.42	13.57	14.66	16.14
Scope 3 Imputed Emissions	14.14	0.72	12.92	13.72	14.11	14.59	15.40
Scope 3 Intensity (Disclosed+Imputed)	7.86	1.94	4.48	6.74	7.94	9.09	10.93
Scope 3 Disclosed Intensity	4.31	1.74	0.70	3.43	4.44	5.35	6.76
Scope 3 Imputed Intensity	7.98	1.84	4.84	6.88	8.01	9.14	10.95
Scope 1+2 Emissions (Disclosed+Imputed)	14.45	1.06	12.82	13.80	14.44	15.17	16.09
Scope 1+2 Disclosed Emissions	12.73	2.28	9.14	11.17	12.61	14.33	16.65
Scope 1+2 Imputed Emissions	14.51	0.94	12.96	13.84	14.46	15.18	16.09
Scope 1+2 Intensity (Disclosed+Imputed)	8.19	2.17	4.40	7.00	8.28	9.56	11.54
Scope 1+2 Disclosed Intensity	3.50	2.32	-0.70	2.17	3.51	5.02	7.10
Scope 1+2 Imputed Intensity	8.34	1.99	4.99	7.14	8.35	9.61	11.58
Scope 1+2+3 Emissions (Disclosed+Imputed)	15.05	0.85	13.69	14.52	15.04	15.64	16.40
Scope 1+2+3 Disclosed Emissions	14.16	1.82	11.20	12.94	14.14	15.47	17.13
Scope 1+2+3 Imputed Emissions	15.08	0.78	13.79	14.54	15.05	15.64	16.38
Scope 1+2+3 Intensity (Disclosed+Imputed)	8.79	2.00	5.32	7.66	8.88	10.06	11.91
Scope 1+2+3 Disclosed Intensity	4.93	1.88	0.97	3.95	5.12	6.04	7.59
Scope 1+2+3 Imputed Intensity	8.91	1.87	5.74	7.80	8.94	10.10	11.94

Panel B: Carbon variable-level summary statistics: uncertainty (standard deviations).

	Mean	**Std**	**5%**	**25%**	**50%**	**75%**	**95%**
Scope 1 Emissions	14.26	2.78	11.84	14.00	14.83	15.48	16.20
Scope 1 Intensity	8.28	2.54	3.72	7.12	8.56	9.88	11.80
Scope 2 Emissions	12.34	2.39	10.58	12.18	12.76	13.26	14.16
Scope 2 Intensity	6.36	2.19	2.36	5.26	6.56	7.75	9.56
Scope 3 Emissions	13.53	2.63	11.50	13.30	13.97	14.58	15.52
Scope 3 Intensity	7.55	2.37	3.53	6.39	7.76	9.04	10.94
Scope 1+2 Emissions	14.36	2.78	12.15	14.12	14.92	15.57	16.22
Scope 1+2 Intensity	8.38	2.51	4.02	7.25	8.65	9.96	11.84
Scope 1+2+3 Emissions	14.68	2.80	12.86	14.46	15.24	15.84	16.46
Scope 1+2+3 Intensity	8.70	2.48	4.65	7.61	8.97	10.23	12.04

Table 3: Average emissions levels and uncertainty by Fama - French 48 industries in CRSP Universe.

Emissions are reported in millions of tonnes of CO_2-equivalent (tCO_2e). Uncertainties are computed as the ratio of uncertainty level over imputed emissions. Figures for the top ten industries by Scope 1 Emissions (Panel A) and Scope 1 Uncertainty (Panel B) are reported.

Panel A: Ranking based on Scope 1 Emissions

Sector	# firms	Scope 1		Scope 1+2		Scope 1+2+3	
		Emissions	Unc.	Emissions	Unc.	Emissions	Unc.
Util	113.00	9.23	1.23	9.67	1.08	11.24	0.88
Oil	178.00	4.38	1.43	4.89	1.24	6.96	1.02
Trans	112.00	4.12	1.51	4.58	1.28	6.29	1.03
Gold	41.00	4.11	1.57	4.64	1.35	6.03	1.09
Mines	23.00	4.09	1.51	4.61	1.30	6.17	1.05
Steel	34.00	3.77	1.58	4.37	1.33	6.33	1.05
Drugs	262.00	3.75	1.62	4.13	1.38	5.65	1.10
Agric	12.00	3.51	1.69	3.98	1.43	5.42	1.13
Coal	11.00	3.16	1.58	3.62	1.34	5.21	1.06
Soda	3.00	3.16	1.64	3.71	1.41	5.87	1.13

Panel B: Ranking based on Scope 1 Uncertainty

Sector	# firms	Scope 1		Scope 1+2		Scope 1+2+3	
		Emissions	Unc.	Emissions	Unc.	Emissions	Unc.
Guns	4.00	1.95	1.88	2.52	1.52	4.26	1.16
Banks	468.00	2.46	1.84	2.93	1.51	4.59	1.13
Beer	2.00	2.04	1.79	2.55	1.49	4.24	1.11
Ships	2.00	1.81	1.77	2.22	1.45	3.61	1.10
Insur	139.00	1.92	1.77	2.35	1.43	4.08	1.09
Fin	102.00	2.17	1.75	2.65	1.42	4.22	1.09
Telcm	132.00	2.51	1.74	3.08	1.43	4.81	1.11
Books	36.00	2.52	1.74	2.99	1.43	4.57	1.10
Hlth	70.00	2.60	1.73	3.08	1.43	4.73	1.11
PerSv	153.00	2.46	1.71	2.95	1.41	4.55	1.09

Table 4: Average Intensity and uncertainty by Fama - French 48 industries in CRSP Universe.

Carbon intensity is defined as imputed emissions divided by total assets and is expressed in natural logarithms. The uncertainty measures in this table are ratios of level uncertainty to the corresponding level value. Figures for the top ten industries by Scope 1 Intensity (Panel A) and Scope 1 Uncertainty (Panel B) are reported.

Panel A: Ranking based on Scope 1 Intensity

Sector	# firms	Scope 1		Scope 1+2		Scope 1+2+3	
		Intensity	Unc.	Intensity	Unc.	Intensity	Unc.
Soda	3.00	9.65	1.64	9.88	1.41	10.50	1.13
Drugs	262.00	8.90	1.62	9.13	1.38	9.64	1.10
Mines	23.00	8.86	1.51	9.08	1.30	9.53	1.05
Agric	12.00	8.75	1.69	9.00	1.43	9.50	1.13
Ships	2.00	8.63	1.77	8.97	1.45	9.63	1.10
ElcEq	153.00	8.33	1.71	8.64	1.41	9.25	1.09
MedEq	144.00	8.24	1.67	8.56	1.39	9.22	1.06
Gold	41.00	8.21	1.57	8.44	1.35	8.89	1.09
Beer	2.00	8.20	1.79	8.51	1.49	9.17	1.11
Txtls	9.00	8.20	1.58	8.57	1.30	9.34	1.02

Panel B: Ranking based on Scope 1 Uncertainty

Sector	# firms	Scope 1		Scope 1+2		Scope 1+2+3	
		Intensity	Unc.	Intensity	Unc.	Intensity	Unc.
Guns	4.00	8.17	1.88	8.50	1.52	9.17	1.16
Banks	468.00	6.91	1.84	7.22	1.51	7.82	1.13
Beer	2.00	8.20	1.79	8.51	1.49	9.17	1.11
Ships	2.00	8.63	1.77	8.97	1.45	9.63	1.10
Insur	139.00	5.53	1.77	5.98	1.43	6.80	1.09
Fin	102.00	6.96	1.75	7.34	1.42	8.00	1.09
Telcm	132.00	7.37	1.74	7.73	1.43	8.31	1.11
Books	36.00	7.96	1.74	8.26	1.43	8.86	1.10
Hlth	70.00	7.86	1.73	8.15	1.43	8.74	1.11
PerSv	153.00	7.97	1.71	8.30	1.41	8.92	1.09

Table 5: Scope 1 Cross-predictability between Carbon Intensity Level and Uncertainty in CRSP Universe.

This table reports panel regression results analyzing the cross-predictability between firm-level carbon intensity levels and uncertainty within the $CRSP$ universe. The dependent variable alternates between Scope 1 intensity level and intensity uncertainty. Columns (1) - (2) regress intensity level on standard firm-level controls (size, book-to-market ratio, profitability, leverage, volatility, growth, and momentum), without and with firm and industry–time fixed effects, respectively. Columns (3) - (4) replicate these specifications using intensity uncertainty as the dependent variable. Columns (5) - (6) regress intensity level on intensity uncertainty and controls, while Columns (7) - (8) regress intensity uncertainty on intensity level, again with and without fixed effects. All dependent and independent variables are standardized within month to remove common time variation. All coefficients have been multiplied by 100 for readability. t-statistics are based on two-way clustered (firm and time) standard errors. Asterisks denote statistical significance at conventional levels: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. The table reports coefficients, inclusion of controls and fixed effects, number of observations, and R^2.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	16.60***		-10.76***		16.57***		-10.64***	
	(1.21)		(1.71)		(1.17)		(1.57)	
Intensity							-0.74	-9.76***
							(2.56)	(1.84)
Uncertainty					-0.32	-5.65***		
					(1.11)	(1.08)		
Volume	3.20***	3.31***	-2.91***	-2.97***	3.19***	3.14***	-2.88***	-2.65***
	(0.39)	(0.42)	(0.34)	(0.32)	(0.39)	(0.41)	(0.33)	(0.31)
Beta	0.90*	1.26**	-1.50***	-2.29***	0.90*	1.13**	-1.50***	-2.17***
	(0.48)	(0.56)	(0.55)	(0.53)	(0.47)	(0.55)	(0.55)	(0.52)
Size	-86.66***	-89.52***	-9.94***	-13.39***	-86.69***	-90.28***	-10.58***	-22.13***
	(0.79)	(0.88)	(1.23)	(0.87)	(0.77)	(0.84)	(1.85)	(1.66)
BM	-19.37***	-22.60***	0.03	0.18	-19.37***	-22.59***	-0.11	-2.02***
	(0.70)	(0.77)	(0.63)	(0.63)	(0.70)	(0.76)	(0.83)	(0.77)
ROA	-6.39***	-7.81***	2.04***	2.19***	-6.38***	-7.69***	1.99***	1.43***
	(1.21)	(0.83)	(0.50)	(0.39)	(1.21)	(0.82)	(0.47)	(0.34)
Asset Growth	2.35***	3.14***	-1.17***	-1.17***	2.35***	3.07***	-1.16***	-0.86***
	(0.31)	(0.58)	(0.21)	(0.33)	(0.31)	(0.57)	(0.20)	(0.32)
Momentum	11.04***	10.22***	3.31***	3.47***	11.05***	10.41***	3.39***	4.47***
	(0.44)	(0.42)	(0.51)	(0.41)	(0.44)	(0.42)	(0.61)	(0.48)
Leverage	-1.50***	-2.46***	0.55*	1.06**	-1.50***	-2.40***	0.54*	0.82
	(0.43)	(0.78)	(0.30)	(0.49)	(0.42)	(0.78)	(0.30)	(0.50)
Volatility	0.82	2.03	-2.69***	-3.88***	0.81	1.81	-2.69***	-3.68***
	(1.05)	(1.54)	(0.65)	(0.95)	(1.06)	(1.53)	(0.64)	(0.91)
Earnings Surprise	0.97***	0.82***	-0.17	-0.53***	0.97***	0.79***	-0.16	-0.45**
	(0.24)	(0.21)	(0.21)	(0.20)	(0.24)	(0.20)	(0.21)	(0.20)
Sales Growth	1.10***	1.39**	0.28	0.34	1.10***	1.41***	0.29	0.47
	(0.39)	(0.54)	(0.25)	(0.30)	(0.39)	(0.54)	(0.25)	(0.29)
Controls	Y	Y	Y	Y	Y	Y	Y	Y
FE	N	Y	N	Y	N	Y	N	Y
Observations	810651	810651	810651	810651	810651	810651	810651	810651
R^2	0.696	0.750	0.017	0.048	0.696	0.752	0.017	0.053

Table 6: Stock Returns and Scope 1 Intensity Level and Uncertainty in CRSP Universe

This table reports panel regression results linking stock returns to firm-level Scope 1 carbon intensity metrics and financial characteristics within the *CRSP* universe. Columns (1)–(5) present specifications with fixed effects and controls, using successively Scope 1 intensity level, intensity uncertainty, and both jointly as explanatory variables as well as their interaction. Firm-level controls include trading volume, market beta, size, book-to-market ratio, profitability (ROA), asset growth, momentum, leverage, volatility, earnings surprise, and sales growth. All independent variables are standardized within month to remove common time variation. Returns, the dependent variable, are in percentage for readability. t-statistics are based on two-way clustered (firm and time) standard errors. Asterisks denote statistical significance at conventional levels: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. All regressions are estimated by OLS after applying the appropriate within transformations for fixed effects. The table reports coefficients, the inclusion of controls and fixed effects, sample size, and R^2.

	(1)	(2)	(3)	(4)	(5)
Intensity	1.28***		1.30***	1.29***	1.38***
	(0.09)		(0.08)	(0.09)	(0.09)
Uncertainty		0.13***	0.21***		0.27***
		(0.05)	(0.05)		(0.06)
Intensity × Uncertainty				0.02	0.09***
				(0.02)	(0.03)
Volume	0.93***	0.98***	0.94***	0.93***	0.94***
	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)
Beta	-0.08	-0.06	-0.07	-0.08	-0.07
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Size	2.17***	1.04***	2.21***	2.17***	2.25***
	(0.12)	(0.05)	(0.11)	(0.12)	(0.11)
BM	0.72***	0.42***	0.72***	0.72***	0.74***
	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)
ROA	0.20***	0.10**	0.20***	0.20***	0.20***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Asset Growth	-0.21***	-0.17***	-0.21***	-0.21***	-0.21***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Momentum	-0.30***	-0.18**	-0.31***	-0.30***	-0.32***
	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)
Leverage	0.04	0.01	0.04	0.04	0.04
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Volatility	-0.46***	-0.43***	-0.45***	-0.46***	-0.46***
	(0.09)	(0.08)	(0.09)	(0.09)	(0.09)
Earnings Surprise	0.08*	0.09**	0.08*	0.08*	0.09*
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Sales Growth	-0.01	0.00	-0.01	-0.01	-0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Controls	Y	Y	Y	Y	Y
FE	Y	Y	Y	Y	Y
Observations	794137	794137	794137	794137	794137
R^2	0.015	0.012	0.016	0.015	0.016

Table 7: Excess Returns and Risk Adjusted Returns for Double Sorted Portfolios in CRSP Universe.

This table reports excess returns and risk-adjusted returns for portfolios formed from a double *sequential* sort on carbon intensity and on *scaled* imputation uncertainty. Each month, stocks are first sorted into quintiles based on carbon intensity (Low, Q2, Q3, Q4, High), which define the *rows* of the table. Within each carbon-intensity quintile, stocks are then sorted into quintiles based on scaled imputation uncertainty, which define the *columns* of the table. For each row - column portfolio, the table reports: (i) the average monthly excess return (in percent), and (ii) the alpha from a seven-factor asset pricing model (in percent). Newey - West t-statistics are reported in parentheses. The H - L *row* reports the excess return and alpha of a long - short portfolio that is long the highest carbon-intensity quintile and short the lowest carbon-intensity quintile, within each uncertainty quintile. The H - L *column* reports the excess return and alpha of a long - short portfolio that is long the highest scaled-uncertainty quintile and short the lowest scaled-uncertainty quintile, within each carbon-intensity quintile.

Returns, Scope 1

	Low	**Q2**	**Q3**	**Q4**	**High**	**H-L**
Low	**1.45**	**1.28**	**1.48**	**1.55**	**1.38**	-0.07
	(5.78)	(3.95)	(5.11)	(5.81)	(4.69)	(-0.58)
Q2	**1.42**	**1.76**	**1.80**	**1.86**	**1.68**	0.26
	(5.15)	(5.15)	(4.93)	(4.71)	(4.66)	(0.95)
Q3	**1.76**	**2.05**	**2.11**	**2.10**	**2.20**	0.44
	(5.48)	(5.09)	(5.29)	(4.70)	(4.93)	(1.08)
Q4	**2.52**	**3.09**	**2.79**	**3.22**	**2.88**	0.36
	(5.18)	(6.61)	(5.07)	(5.54)	(5.08)	(0.89)
High	**4.62**	**3.67**	**3.77**	**3.07**	**3.09**	**-1.53**
	(5.10)	(5.02)	(5.42)	(4.25)	(4.60)	(-2.23)
H-L	**3.17**	**2.14**	**2.14**	**1.52**	**1.71**	
	(3.73)	(3.67)	(3.74)	(2.30)	(3.05)	

Alphas, Scope 1

	Low	**Q2**	**Q3**	**Q4**	**High**	**H-L**
Low	**0.37**	**0.31**	**0.37**	**0.48**	**0.36**	-0.01
	(5.50)	(2.27)	(3.72)	(4.33)	(3.58)	(-0.08)
Q2	**0.43**	**0.65**	**0.70**	**0.64**	**0.62**	0.19
	(3.11)	(5.46)	(6.00)	(5.74)	(4.18)	(0.96)
Q3	**0.83**	**1.08**	**1.22**	**1.13**	**1.11**	0.28
	(3.61)	(5.94)	(6.79)	(6.25)	(6.32)	(0.91)
Q4	**1.55**	**2.13**	**1.79**	**2.39**	**1.94**	0.39
	(4.78)	(7.56)	(7.15)	(6.07)	(6.55)	(0.89)
High	**3.74**	**2.89**	**2.81**	**2.10**	**2.35**	**-1.39**
	(5.72)	(6.44)	(7.36)	(4.33)	(5.11)	(-2.43)
H-L	**3.37**	**2.19**	**2.25**	**1.62**	**1.99**	
	(5.04)	(6.89)	(6.08)	(3.32)	(4.49)	

Table 8: Features of Double Sorted Portfolios in CRSP Universe.

This table reports detailed summary statistics of characteristics for portfolios formed from a double *sequential* sort on carbon intensity and scaled imputation uncertainty. For each characteristic, firms are first sorted into quintiles based on carbon intensity (Low, Q2, Q3, Q4, High), which define the *rows* of the table. Within each carbon-intensity quintile, firms are then sorted into quintiles based on (scaled) imputation uncertainty, which define the *columns* of the table. Each cell reports the value-weighted mean of the corresponding characteristic within the double-sorted portfolio. "# Firms" reports the average number of firms in each of the 25 (5×5) double-sorted portfolios. Size corresponds to log of market capitalization.

 # Firms, Scope 1

	Low	**Q2**	**Q3**	**Q4**	**High**
Low	250	129	155	172	176
Q2	177	176	176	176	176
Q3	176	176	176	176	176
Q4	176	176	176	176	176
High	176	176	176	176	176

Carbon Intensity, Scope 1

	Low	**Q2**	**Q3**	**Q4**	**High**
Low	1.02	2.93	2.89	2.92	2.72
Q2	6.01	6.02	6.03	6.02	6.00
Q3	7.46	7.53	7.54	7.53	7.50
Q4	9.03	9.09	9.05	9.01	8.94
High	11.21	11.12	11.08	11.01	10.83

Carbon Uncertainty, Scope 1

	Low	**Q2**	**Q3**	**Q4**	**High**
Low	0.07	0.78	1.34	1.89	2.86
Q2	0.49	1.32	1.70	2.11	2.93
Q3	0.73	1.38	1.69	2.06	2.82
Q4	0.96	1.35	1.61	1.95	2.61
High	1.00	1.25	1.44	1.70	2.24

Size, Scope 1

	Low	**Q2**	**Q3**	**Q4**	**High**
Low	18.41	15.64	16.46	17.89	17.87
Q2	16.75	15.86	15.62	15.59	15.45
Q3	15.53	14.77	14.60	14.47	14.37
Q4	14.15	13.93	13.77	13.66	13.44
High	12.73	12.72	12.65	12.49	12.22

Figure 1: Imputed Emissions Level and Uncertainty Correlation in the CRSP universe

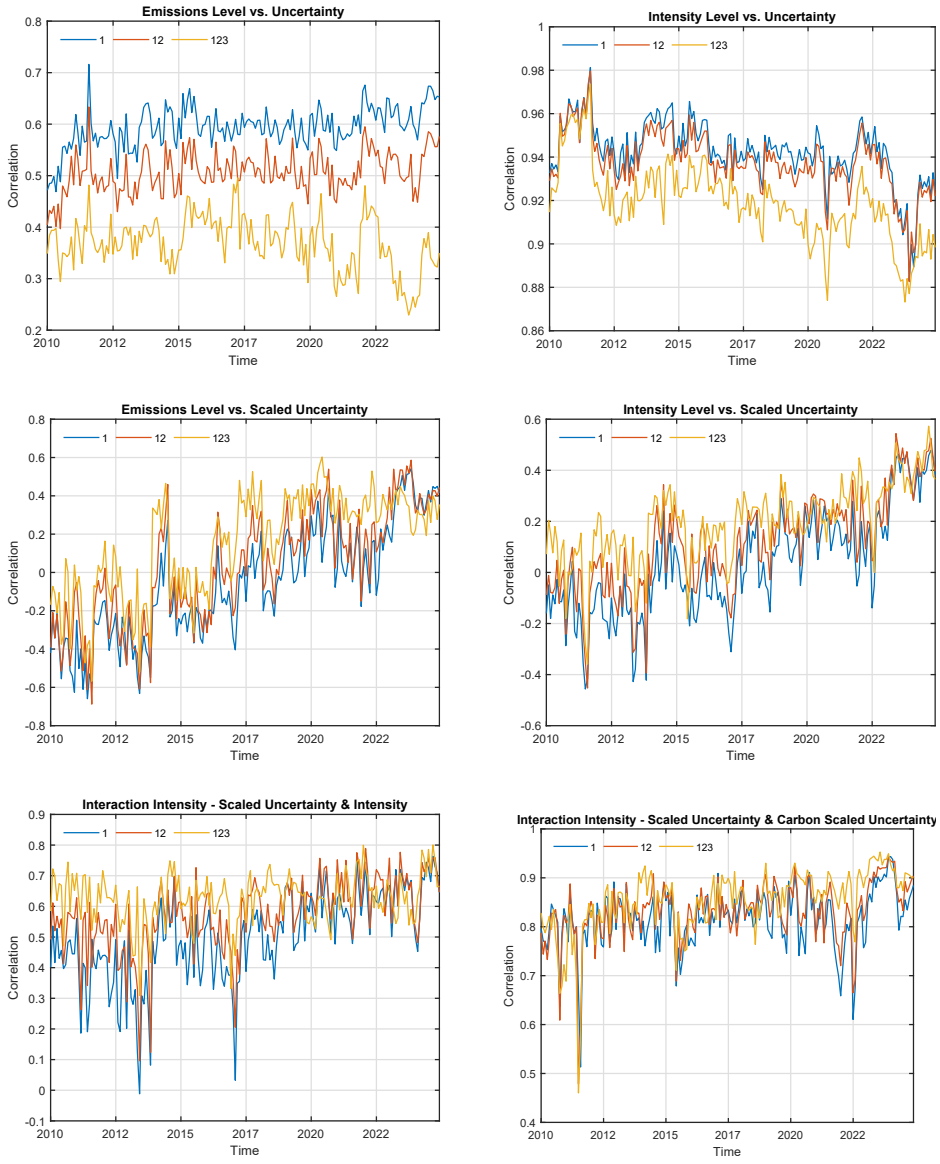


Figure 2: Levels and Uncertainty Interactions Correlation with Levels in the CRSP universe

