Carbon emission and Asset prices

New evidence from machine learning

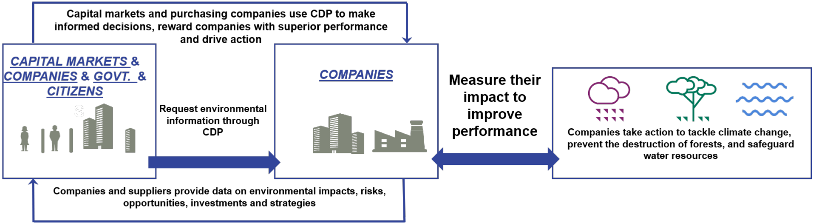
Feng Li, Xingjian Zheng

Shanghai Advanced Institute of Finance, SJTU

June, 2023

Motivation

- Carbon emission data is one of the most important data for every Business/Govt./ entity
  - In the financial industry, in 2022, 680+ financial institutions with US$130 Trillion AUM request information on carbon emissions.
  - Large corporates like LOréal and Diageo use carbon emission data for supplier evaluations and performance reviews.ă
  - Nation-wide/State-level authorities need carbon emission data to regulate firms, as in the Paris Agreement.



- As for economic empiricists, there exist contradictory views on the relationship between carbon emissions and asset prices.
  - The risk story: **higher emissions, higher returns** (Bolton and Kacperczyk, 2020, 2021a, 2021b, 2022a, 2022b)
  - Contrasting findings: **higher emissions, lower returns** (Garvey et al., 2018; In et al., 2017; Matsumura et al., 2014)
  - Inconclusive results: **Insignificant emission-return relationship** (Aswani et al., 2022; Monasterolo and De Angelis, 2020)
- The problem is with a lack of emission DATA! (sampling bias for the high emission firms/fewer disclosures prior to 2016/biased estimation methods)

Empirical challenge

- **Challenge(i):** few firms voluntarily disclose carbon emissions (Busch et al., 2020).
  - In 2021, a record-breaking 13189 firms disclose carbon emissions globally. However, that only accounts for a small fraction of all the firms.
  - In the US, among the S&P MidCap 400, only 28 % of companies disclose their GHG emissions, and more than 4000 thousand companies do not make the disclosure.
  - In the UK, 68% of the firms do not disclose carbon emissions.
- **Challenge(ii):** emission data might be generated by biased estimation algorithms.
  - Carbon emissions estimated by data vendors are largely linear interpolations based on industry-fixed effects and are too clustered with stock returns (Aswani et al., 2022).

This paper

- Estimates a large panel of scope 1 carbon emissions by US firms with XGBoost from 2002 to 2021.
  - This data set has broad coverage of 4111 firms per year as compared to 1675 firms provided by data vendors.
  - We design empirical tests and perform various ML validation tests to examine the robustness & accuracy.
  - We provide various robust analyses from empirical tests to machine learning tests.
- With this data set, we document the following simple facts:
  - Prior to 2016 (the Paris Agreement), there is a **high carbon, high return** relationship
  - Post-2016, the positive relationship drastically reversed to become negative, i.e., **high carbon, low return**.
  - This reversal is driven by institutional investor flows that purchase stocks of low-emission firms.

- We replicate Bolton and Kacperczyk (2021a) by running monthly pooled-OLS regressions and plotting the cumulative return premia estimated with (1) XGBoost estimated data, and (2) data provided by the data vendor.

$$RET_{i,t} = \alpha + \beta GHG_{i,t} + \gamma' X_{i,t-1} + \delta_t + \mu_j + \varepsilon_{i,t}, \tag{1}$$

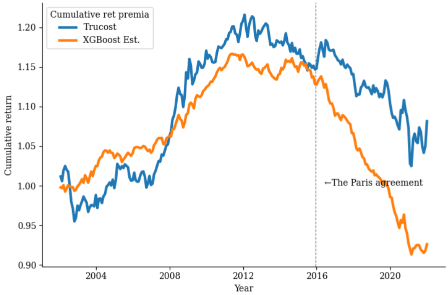


Figure 1: Carbon cumulative return premia

Related literature

- Carbon emission and stock returns
  - Positive emission-return relationship: A series of works by Bolton and Kacperczyk
  - Contrasting evidence: Aswani et al. (2022); Duan et al. (2022); Choi et al. (2022); Cheema-Fox et al. (2021); which may be driven
  - Evidence on investors' raising awareness of ESG investing: P'astor et al. (2021, 2022); Pedersen et al. (2021), and van der Beck (2021)
- Boosting trees in Economics and Finance
  - XGBoost models or other basic boosting trees are used in loan approval (Tantri, 2021, Rossi and Utkus, 2020), patent classification (Zheng, 2022), and return predictions (Teng et al., 2020)
- Economic links and industrial competitions
  - Most notably on business similarity (Hoberg and Philips, 2010, 2016, 2018), or competitions (Li et al. 2013, Bernard et al., 2020, Eisdorfer et al., 2020).
  - Other links like technological linkage (Lee et al., 2019), customers (Cohen and Frazzini, 2008), common analysts (Ali and Hirshleifer, 2020), etc.

1. Background and Intro
2. Emission estimation
3. Empirical results
4. Conclusion
5. Supplementary figures and tables

Methodology

- **Firms that share similar business structures produce carbon emissions on a similar scale.**
- We can use the emission of disclosure firms to predict the emission of similar non-disclosure firms by training a supervised model.

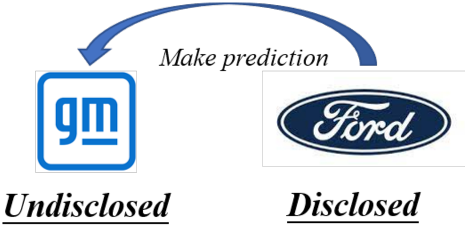


Figure 2: Estimation example

- We input (i) the emission data, (ii) similarity score pairs, (iii) firm fundamentals, and (iv) firms fixed identifiers into the XGBoost algorithm. We train the algorithm on a cross-sectional basis.
  - Similarity score is from the Hoberg and Philips data library; emission data is obtained from Trucost; firm fundamentals and fixed identifiers are from Compustat CCM.
  - Firm fundamentals include firm sales, total assets, non-current assets, and employee numbers.
  - XGBoost model is as follows:

$$\widehat{GHG}_f = \hat{f}\left(GHG_d, \text{score}_{<f,d>}, \cdots\right) = \arg\min L\left(f(X) + \widehat{GHG}_f\right) + R(f(\cdot)) \quad (2)$$

- Sample period: 2002-2021, with a training period from 2002-2018 and a test period from 2019 to 2021.
- We obtain (i) scope 1 emission data from Trucost and link firms by their GVKEY from WRDS, (ii) firm fundamentals from Compustat, and (iii) time series data from CRSP.

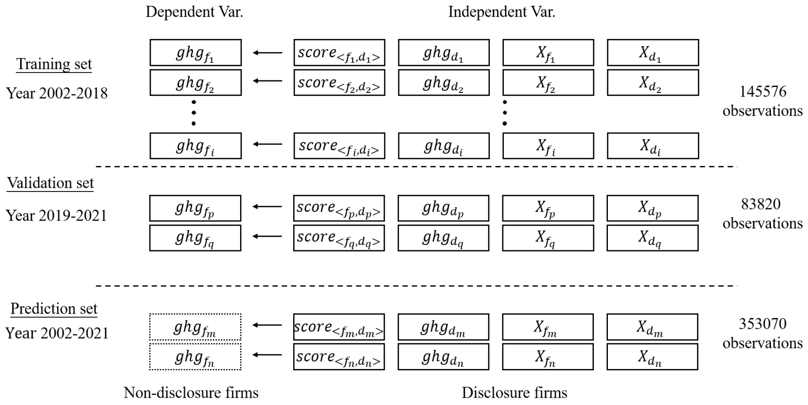


Figure 3: Estimation methodology

Table 1: Number of disclosed firms in the dataset

	Number of disclosed firms in the dataset				
Year	Trucost	Thomson Reuters	CDP	Aswani et al. (2022)	XGBoost Estimated
2002	629	4	4		2952
2003	851	10	1		3703
2004	1026	20	12		4073
2005	1260	50	37	700	4406
2006	1275	171	54	706	4440
2007	1237	309	77	693	4367
2008	1251	367	105	690	4328
2009	1265	500	329	709	4162
2010	1258	550	564	704	4011
2011	1252	588	801	715	3938
2012	1252	597	900	727	3909
2013	1350	572	998	800	3946
2014	1372	585	1217	829	4049
2015	1377	659	1135	859	4117
2016	3265	722	1472	2369	4281
2017	3286	797	1469	2509	4228
2018	3363	895	1501	2645	4242
2019	3393	1066	1418	1992	4279
2020	3154	1110	1436		4329
2021	385	398	356		4453
Average	1675	499	694	1176	4111

Training results

Table 2: Emission comparison between different data sets by year

	Panel A: Trucost data				Panel B: Xgboost estimated			
	Distinct firms	Mean	Std	Median	Distinct firms	Mean	Std	Median
2002	629	12.05	2.54	11.90	2952	9.01	5.37	11.04
2003	851	11.60	2.66	11.35	3703	8.96	5.09	10.80
2004	1026	11.58	2.66	11.34	4073	9.09	4.92	10.79
2005	1260	11.37	2.68	11.19	4406	9.03	4.88	10.67
2006	1275	11.40	2.67	11.22	4440	9.14	4.80	10.71
2007	1237	11.40	2.64	11.21	4367	9.24	4.78	10.82
2008	1251	11.43	2.62	11.25	4328	9.27	4.75	10.83
2009	1265	11.30	2.62	11.07	4162	9.22	4.72	10.73
2010	1258	11.35	2.61	11.16	4011	9.32	4.68	10.79
2011	1252	11.34	2.62	11.08	3938	9.35	4.67	10.80
2012	1252	11.31	2.64	11.10	3909	9.28	4.71	10.73
2013	1350	11.25	2.65	11.07	3946	9.36	4.60	10.73
2014	1372	11.23	2.69	11.01	4049	9.30	4.66	10.71
2015	1377	11.21	2.65	11.02	4117	9.36	4.51	10.68
2016	3265	9.49	2.95	9.44	4281	9.36	3.49	9.76
2017	3286	9.48	2.98	9.41	4228	9.41	3.42	9.78
2018	3363	9.46	3.01	9.42	4242	9.33	3.43	9.69
2019	3393	9.39	3.02	9.35	4279	9.29	3.39	9.62
2020	3154	9.09	3.01	8.95	4329	9.03	3.39	9.38
2021	385	7.61	2.37	7.57	4453	9.08	4.08	9.61

- Estimation performance and data overview.
  - Machine learning performance
  - Variable importance contribution
  - Summary stats (below) and Correlation matrix

	N	Mean	Std	Min	25%	50%	75%	Max
Firm-yearl level observations								
LOGGHG	82213	9.22	4.44	0.00	8.03	10.43	12.02	15.87
GHGINTEN	80469	7.44	20.65	0.00	0.04	0.42	3.78	110.94
LOGSIZE	80955	13.36	2.05	8.53	11.90	13.41	14.83	17.19
LEVERAGE	82166	0.58	0.27	0.08	0.37	0.58	0.80	1.11
INVEST2A	81043	0.04	0.05	0.00	0.00	0.02	0.05	0.23
ROE	82015	0.00	0.42	-1.19	-0.05	0.08	0.15	1.21
HHI	82136	0.09	0.07	0.02	0.05	0.07	0.12	0.35
LOGPPE	78955	4.57	2.59	0.02	2.56	4.49	6.49	9.54
B2M	77172	1.05	1.95	0.06	0.31	0.57	0.95	11.90
SALESGR	76818	0.10	0.31	-0.54	-0.04	0.06	0.18	1.34
EPSGR	77864	-0.03	2.15	-8.50	-0.37	0.08	0.56	5.86
Firm-year-month level observations								
RETX	890602	1.02	16.47	-97.22	-5.79	0.43	6.63	1988.36
MOM	890522	1.12	4.72	-44.98	-1.01	0.90	2.88	169.02
VOLAT	890531	12.61	10.69	0.27	6.69	10.00	15.31	583.47
BETA	890602	1.23	1.08	-21.13	0.59	1.09	1.70	44.39

- Overall, the estimated data set is comparable to Trucosts original emissions in magnitude after 2016, and it has high validity and complements a very large fraction of the data set prior to 2016.

Data validation

- We design several tests to validate the robustness of our data set.
- We design empirical tests to examine the validity.
  - State level regulation shocks
  - Transition matrix
  - Comparison of the determinants of carbon emission
  - Carbon emission and inclusion into ESG-related fund
- We also use machine learning tests to examine the robustness.
  - Cross-validation tests with hyper-parameters
  - Different partitioning results
  - Model comparison: XGBoost versus linear models
  - Business similarity and emission similarity

Main results

- We explore the cross-sectional properties of stock returns with firms carbon emissions.
- We first follow the pooled OLS regression model used in Bolton and Kacperczyk (2021a) as follows,

$$RET_{i,t} = \alpha + \beta GHG_{i,t} + \gamma' X_{i,t-1} + \delta_t + \mu_j + \varepsilon_{i,t}, \tag{3}$$

- The independent variables include three different emission measures: (i) LOGGHG, (ii) GHGGR, and (iii) GHGINTEN. The most important variable is the logarithmic value of carbon emissions LOGHG.
- We partition the sample period before and after 2016 (the Paris Agreement), and examine the emission-return relationship with different samples, i.e., the Trucost sample and the XGBoost sample.

Table 3: Carbon emission and asset prices: replication of BK(2021a)

	Trucost original sample					
Sample period	Panel A: 2002-2016 (before the Paris agreement)					
	(1)	(2)	(3)	(4)	(5)	(6)
LOGGHG	0.0330*	0.0686***				
	(1.89)	(3.86)				
GHGGR			0.5783***	0.5721***		
			(4.84)	(5.17)		
GHGINTEN					-0.0032	-0.0039
					(-0.44)	(-0.52)
LOGSIZE	-0.0837	-0.1026	0.0189	0.0431	-0.0894	-0.0779
	(-0.75)	(-0.93)	(0.18)	(0.37)	(-0.81)	(-0.70)
B2M	-0.0480	-0.0562	-0.021	-0.0033	-0.0502	-0.0454
	(-1.31)	(-1.38)	(-0.63)	(-0.09)	(-1.37)	(-1.11)
LEVERAGE	0.0939	-0.1837	-0.0561	0.0689	0.0641	-0.1126
	(0.37)	(-0.54)	(-0.22)	(0.21)	(0.24)	(-0.33)
MOM	-0.0510	-0.0596	-0.0648	-0.0763	-0.0503	-0.0597
	(-0.53)	(-0.61)	(-0.62)	(-0.73)	(-0.53)	(-0.61)
INVEST2A	-3.3218**	-1.6153	-3.2394*	-1.8206	-3.2162**	-1.9375
	(-2.25)	(-1.20)	(-1.84)	(-1.19)	(-2.15)	(-1.41)
ROE	0.8786***	0.7999***	0.6462***	0.6013***	0.8824***	0.8093***
	(3.17)	(3.56)	(2.96)	(3.06)	(3.16)	(3.59)
HHI	-0.1756	0.0888	0.7088	0.6149	-0.1974	0.1374
	(-0.13)	(0.06)	(0.80)	(0.40)	(-0.14)	(0.10)
LOGPPE	-0.0078	0.003	-0.0019	-0.0077	0.0307	0.0379
	(-0.15)	(0.06)	(-0.04)	(-0.18)	(0.53)	(0.69)
BETA	-0.3914*	-0.4019**	-0.3483	-0.3438	-0.3953**	-0.3992**
	(-1.96)	(-2.11)	(-1.52)	(-1.57)	(-1.98)	(-2.10)
VOLAT	0.1948***	0.2091***	0.1977***	0.2122***	0.1945***	0.2081***
	(3.26)	(3.46)	(2.62)	(2.78)	(3.26)	(3.44)
SALESGR	-0.2561	-0.2186	-0.2900	-0.3059	-0.2583	-0.217
	(-0.63)	(-0.58)	(-0.63)	(-0.68)	(-0.63)	(-0.58)
EPSGR	0.0219	0.0236	0.0111	0.0145	0.0221	0.0238
	(0.84)	(0.85)	(0.46)	(0.59)	(0.85)	(0.87)
Const	T	T	T	T	T	T
Year-Mon FE	T	T	T	T	T	T
Ind FE		T		T		T
R2	0.22	0.22	0.23	0.23	0.22	0.22
N	215808	215808	185490	185490	215760	215760

Carbon emissions estimated by XGBoost algorithm												
Sample period	Panel B: 2002-2016 (before the Paris agreement)						Panel C: 2016-2021 (after the Paris agreement)					
	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
LOGGHG	**0.0142***	**0.0109***					**-0.0888*****	**-0.0536****				
	(1.84)	**(1.91)**					**(-3.32)**	**(-2.12)**				
GHGGR			0.0082	0.0087					-0.0333	-0.0086		
			(0.67)	(0.74)					(-0.77)	(-0.18)		
GHGINTEN					-0.0089***	-0.0091***					-0.0368***	-0.0339***
					(-3.69)	(-4.01)					(-4.69)	(-5.32)
LOGSIZE	0.0781	0.0597	0.1334	0.1224	0.0609	0.0473	0.4077**	0.3759**	0.4388**	0.3837**	0.3288*	0.2828*
	(0.80)	(0.64)	(1.39)	(1.23)	(0.62)	(0.50)	(2.05)	(2.09)	(2.28)	(2.26)	(1.77)	(1.76)
B2M	0.0238	0.0324	0.0197	0.0273	0.0198	0.0303	-0.0413	-0.0752	-0.0382	-0.0751	-0.0597	-0.1054
	(0.60)	(0.72)	(0.49)	(0.58)	(0.52)	(0.70)	(-0.52)	(-0.95)	(-0.47)	(-0.97)	(-0.74)	(-1.32)
LEVERAGE	0.0980	0.1593	0.0057	-0.0318	0.0234	0.0908	0.3255	-0.9475**	0.3891	-0.8988	0.1917	-1.1771***
	(0.24)	(0.45)	(0.01)	(-0.10)	(0.06)	(0.25)	(0.58)	(-2.28)	(0.72)	(-1.90)	(0.36)	(-2.97)
MOM	-0.2240**	-0.2310**	-0.2049	-0.2123	-0.2243**	-0.2317**	-0.5270***	-0.5744***	-0.5205***	-0.569***	-0.5226***	-0.5675***
	(-2.09)	(-2.13)	(-1.56)	(-1.61)	(-2.08)	(-2.13)	(-2.90)	(-3.09)	(-2.96)	(-3.17)	(-2.73)	(-2.93)
INVEST2A	-5.0657***	-4.5781***	-5.092***	-4.2569***	-4.7319**	-4.5971***	-6.9409***	-5.2963*	-5.9519***	-3.9687	-7.3364***	-5.6558*
	(-2.82)	(-4.44)	(-2.66)	(-4.01)	(-2.59)	(-4.31)	(-2.97)	(-1.78)	(-2.80)	(-1.42)	(-3.33)	(-1.93)
ROE	1.7037***	1.6475***	1.5736***	1.5156***	1.6171***	1.5994***	3.0414***	2.4474***	2.9773***	2.3953**	2.8340***	2.3229***
	(7.53)	(7.75)	(6.08)	(6.40)	(7.08)	(7.37)	(7.64)	(13.56)	(7.19)	(11.80)	(6.92)	(12.53)
HHI	-0.6862	-2.143	-0.5437	-1.7339	-0.625	-2.1255	-1.9027	-8.399	-1.5674	-6.3390	-2.2382	-7.4728
	(-1.45)	(-0.98)	(-1.18)	(-0.69)	(-1.33)	(-0.99)	(-0.77)	(-0.49)	(-0.60)	(-0.40)	(-0.93)	(-0.44)
LOGPPE	0.1223**	0.1458***	0.1134	0.1331***	0.1294*	0.1376***	0.0393	0.1489	-0.0549	0.1043	-0.0298	0.1265
	(2.01)	(2.83)	(1.70)	(2.62)	(1.97)	(2.65)	(0.35)	(1.38)	(-0.42)	(0.89)	(-0.25)	(1.20)
BETA	-0.6399***	-0.6203***	-0.6216**	-0.6081**	-0.6489***	-0.6214***	-0.3427	-0.2134	-0.3324	-0.2079	-0.3378	-0.2075
	(-2.77)	(-2.78)	(-2.57)	(-2.59)	(-2.83)	(-2.78)	(-1.21)	(-0.71)	(-1.14)	(-0.66)	(-1.21)	(-0.70)
VOLAT	0.3083***	0.3136***	0.3063***	0.3138***	0.3101***	0.3153***	0.397***	0.4241***	0.3947***	0.4215***	0.3937***	0.4200***
	(4.50)	(4.68)	(3.61)	(3.75)	(4.49)	(4.67)	(5.21)	(5.52)	(5.33)	(5.64)	(4.98)	(5.31)
SALESGR	-0.5018***	-0.4659***	-0.4297***	-0.3763**	-0.5234***	-0.5105***	-0.7476**	-0.6291*	-0.8482***	-0.7571**	-0.7835**	-0.677*
	(-3.36)	(-3.05)	(-2.75)	(-2.36)	(-3.55)	(-3.45)	(-1.99)	(-1.91)	(-2.10)	(-2.13)	(-2.07)	(-2.03)
EPSGR	0.0946***	0.0974***	0.0475**	0.0511**	0.0994***	0.1003***	0.0298	0.0345	0.027	0.0322	0.0363	0.0369
	(3.75)	(3.80)	(2.20)	(2.35)	(3.85)	(3.82)	(0.69)	(0.87)	(0.64)	(0.85)	(0.87)	(0.98)
Const	T	T	T	T	T	T	T	T	T	T	T	T
Year-Mon FE	T	T	T	T	T	T	T	T	T	T	T	T
Ind FE		T		T		T		T		T		T
R2	0.16	0.17	0.17	0.17	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17
N	533001	533001	406387	406387	531136	531136	231149	231149	217119	217119	229777	229777

Cumulative carbon premia

- We examine the cumulative return premia estimated from monthly cross-sectional returns from equation 3, and the independent var. of interest is LOGGHG.
- Similar to regression results, there seems to be a structural mutation in carbon pricing.
- Adding industry fixed effects makes a huge difference for the Trucost data, but not that significant for XGBoost estimated data.

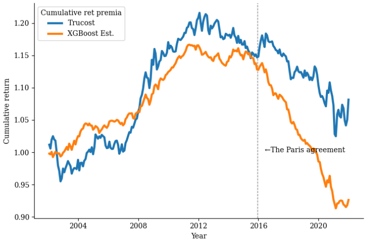


Figure 4: Cumulative premia

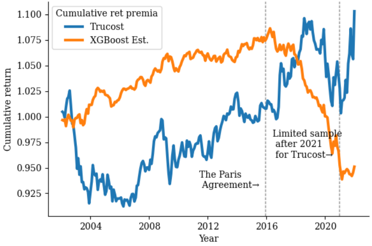


Figure 5: Cumulative premia with industry FE

Portfolio sorting results

- We sort firms into 5 quintiles based on LOGGHG and report cumulative returns for hi-lo portfolio returns.
- XGBoost sample is able to generate a positive emission-return relationship prior to 2016 because of its large sample base.

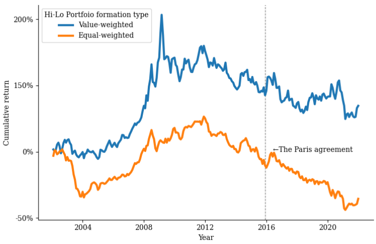


Figure 6: XGBoost data sample

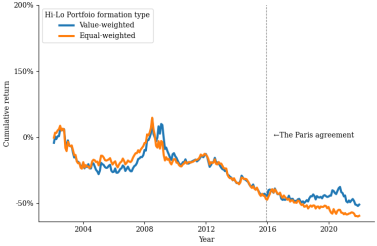


Figure 7: Trucost data sample

Flow-based mechanism

- We examine whether the change in risk premium is driven by institutional investor flow.
- Following van der Beck (2022), we estimate investor flow with equation 4

$$FLOW_{i,t} = A_{i,t} - A_{i,t-1} \times (1 + RET_{i,t}) \tag{4}$$

- we regress investor flows on an interaction term that combines firms' carbon emissions and a time dummy that denotes the date after the Paris Agreement.

Table 4: Mechanism: Carbon emission and investor flow

Dep Var.	LOGFLOW					
	(1)	(2)	(3)	(4)	(5)	(6)
AFTER*GHG	**-0.0276*****	**-0.0279*****	0.0034	-0.0085	0.0001	0.0001
	(-3.07)	**(-2.81)**	(0.24)	(-0.57)	(1.00)	(1.17)
LOGGHG	0.0275***	0.0075				
	(6.69)	(1.12)				
GHGGR			0.0161*	0.0118		
			(1.90)	(1.24)		
GHGINTEN					-0.0004***	-0.0004***
					(-6.50)	(-7.33)
AFTER	0.0234	0.0047	-0.3147**	-0.3037**	-0.2647*	-0.2792**
	(0.15)	(0.03)	(-2.18)	(-2.09)	(-1.88)	(-1.98)
LOGSIZE	1.1195***	1.0904***	1.0451***	1.0040***	1.1040***	1.0825***
	(37.16)	(31.05)	(30.52)	(24.38)	(36.61)	(31.20)
B2M	0.0998***	0.1062***	0.0809***	0.0765***	0.0913***	0.1038***
	(6.76)	(6.87)	(6.37)	(5.11)	(6.68)	(6.92)
LEVERAGE	-0.3403***	0.1860*	-0.0276	0.2112**	-0.6408***	0.1139
	(-3.33)	(1.72)	(-0.30)	(2.28)	(-5.16)	(1.02)
INVEST2A	1.6867***	1.7450***	1.1572**	1.5664***	2.0301***	1.8226***
	(3.19)	(3.47)	(2.14)	(2.86)	(3.79)	(3.67)
ROE	-0.4788***	-0.3187***	-0.3158***	-0.1821***	-0.6121***	-0.3877***
	(-9.51)	(-7.33)	(-6.47)	(-4.57)	(-10.55)	(-8.14)
HHI	0.1120	0.3872	-0.3754	-0.3180	0.2036	0.1700
	(0.31)	(0.64)	(-0.84)	(-0.55)	(0.56)	(0.30)
LOGPPE	-0.0398**	-0.0102	-0.0432**	0.0094	-0.0174	-0.0191
	(-2.20)	(-0.40)	(-2.14)	(0.31)	(-0.91)	(-0.72)
SALESGR	0.2897***	0.2663***	0.2776***	0.2605***	0.3123***	0.2662***
	(3.82)	(3.75)	(3.04)	(3.07)	(3.95)	(3.69)
EPSGR	-0.0366***	-0.0379***	-0.0396***	-0.0405***	-0.0310**	-0.0349***
	(-2.59)	(-2.80)	(-3.10)	(-3.27)	(-2.35)	(-2.67)
Const	T	T	T	T	T	T
Ind FE		T		T		T
R2	0.67	0.65	0.64	0.63	0.67	0.65
N	30007	30007	23584	23584	29935	29935

- We sort stocks of firms of different emission levels into 5 quintiles, and examine the flow-return relationship for each quintile portfolio.
- The flow-induced carbon premium is (more) pronounced among the low-carbon group.

Table 5: Flow-induced stock returns

Dep Var.	RET									
Portfolio type	Lo		2		3		4		Hi	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
LOGFLOW	**1.7186*** (2.73)**	**1.6637*** (2.72)**	**0.6593* (1.82)**	**0.6448* (1.89)**	0.5020 (1.25)	0.4718 (1.04)	1.1486 (1.55)	1.1532 (1.52)	2.0876 (1.22)	2.1667 (1.27)
Const	T	T	T	T	T	T	T	T	T	T
Controls	T	T	T	T	T	T	T	T	T	T
Year-Mon FE	T	T	T	T	T	T	T	T	T	T
Ind FE		T		T		T		T		T
R2	0.18	0.19	0.22	0.23	0.30	0.31	0.21	0.22	0.22	0.23
N	14472	14472	14469	14469	14467	14467	14469	14469	14471	14471

Carbon pricing before and after the Paris Agreement

- We examine the structural mutation of carbon premia around 2016.
- We keep the high emission and low emission quintile firms, and interact their scope 1 emission with a time dummy to examine the carbon pricing.

$$RET_{i,t} = \alpha + HI_{i,t} + After_t + HI_{i,t} \times After_t + \gamma' X_{i,t-1}, \quad (5)$$

- We also report regression results excluding high-emission firms, and the effect is slightly more significant.

Table 6: Emission-return relationship after the Paris Agreement

Panel A: XGBoost whole sample						
	LOGGHG		GHGGR		GHGINTEN	
	(1)	(2)	(3)	(4)	(5)	(6)
AFTER*HIGHG	**-0.8583***	**-0.9313****	0.4629	0.4887	-1.1188***	-1.065**
	(-1.93)	**(-2.05)**	(0.89)	(0.96)	(-3.00)	(-2.51)
AFTER	1.0684***	1.1492***	4.1305***	4.251***	0.5041	0.5048
	(3.38)	(3.25)	(9.21)	(8.46)	(1.52)	(1.29)
HIGHG	0.3554*	0.4209***	0.2523**	0.2596**	-0.0238	-0.0064
	(1.75)	(3.44)	(2.36)	(2.30)	(-0.18)	(-0.06)
Const	T	T	T	T	T	T
Controls	T	T	T	T	T	T
Year FE	T	T	T	T	T	T
Ind FE		T		T		T
R2	0.16	0.16	0.16	0.16	0.14	0.15
N	308838	308838	241463	241463	295656	295656
Panel B: XGBoost exclude high emission industries						
	LOGGHG		GHGGR		GHGINTEN	
	(7)	(8)	(9)	(10)	(11)	(12)
AFTER*HIGHG	**-0.8922****	**-0.9637****	0.4665	0.4857	-1.0997***	-1.0445**
	(-2.00)	**(-2.08)**	(0.90)	(0.96)	(-2.98)	(-2.43)
AFTER	1.1873***	1.237***	4.123***	4.2275***	0.4106	0.3976
	(3.65)	(3.44)	(8.87)	(8.21)	(1.23)	(1.01)
HIGHG	0.3643*	0.4503***	0.2457**	0.2604**	-0.0163	0.0156
	(1.81)	(3.69)	(2.36)	(2.37)	(-0.12)	(0.14)
Const	T	T	T	T	T	T
Controls	T	T	T	T	T	T
Year FE	T	T	T	T	T	T
Ind FE		T		T		T
R2	0.16	0.16	0.16	0.16	0.14	0.15
N	300021	300021	233743	233743	287194	287194

Reversed carbon premia with Trucost data sample

Table 7: Carbon emission and realized stock returns (Trucost post 2016)

Sample period	2016-2021 (After the Paris Agreement with Trucost sample)					
	(1)	(2)	(3)	(4)	(5)	(6)
LOGGHG	**-0.0898***** **(-3.66)**	**-0.0085** **(-0.16)**				
GHGGR			0.9119* (1.87)	1.0895** (2.09)		
GHGINTEN					-0.0423 (-0.86)	-0.0332 (-0.84)
LOGSIZE	0.7525*** (3.29)	0.6214** (2.48)	0.7759*** (3.26)	0.6725** (2.77)	0.7164*** (3.44)	0.6085*** (2.73)
B2M	0.0373 (1.34)	0.0056 (0.18)	0.0495* (1.84)	0.0260 (0.88)	0.0305 (1.07)	0.0014 (0.05)
LEVERAGE	-0.2922 (-0.59)	-1.1791*** (-3.02)	-0.0177 (-0.05)	-1.0527*** (-3.42)	-0.1547 (-0.30)	-1.1884*** (-3.03)
MOM	-0.3415 (-1.54)	-0.403 (-1.60)	-0.3415 (-1.53)	-0.4052 (-1.60)	-0.3409 (-1.52)	-0.4025 (-1.59)
INVEST2A	-4.0486* (-1.76)	-2.1562 (-1.02)	-5.2882* (-2.01)	-2.9505 (-1.44)	-4.2593* (-1.97)	-2.1882** (-1.10)
ROE	1.9664*** (5.84)	1.7233*** (5.06)	2.0597*** (5.89)	1.8355*** (6.19)	1.9064*** (5.75)	1.7241*** (5.23)
HHI	0.2043 (0.09)	-5.9436 (-1.33)	-0.2678 (-0.12)	-8.5278 (-1.91)	-0.0536 (-0.02)	-5.8526 (-1.28)
LOGPPE	-0.1871* (-1.89)	-0.0311 (-0.40)	-0.2564** (-2.67)	-0.0378 (-0.42)	-0.2349*** (-3.35)	-0.0253 (-0.29)
BETA	-0.5881* (-1.76)	-0.4872 (-1.54)	-0.5841* (-1.76)	-0.4947 (-1.58)	-0.5983* (-1.75)	-0.4839 (-1.54)
VOLAT	0.4642*** (3.41)	0.5014*** (3.51)	0.4479*** (3.19)	0.4873*** (3.31)	0.4625*** (3.37)	0.5011*** (3.51)
SALESGR	-1.0494** (-2.28)	-0.8052 (-1.52)	-1.1018** (-2.71)	-0.8971* (-2.02)	-1.0314** (-2.21)	-0.8129 (-1.55)
EPSGR	0.0694 (1.59)	0.0771 (1.42)	0.0552 (1.24)	0.0626 (1.16)	0.072 (1.55)	0.0770 (1.43)
Const	T	T	T	T	T	T
Year-Mon FE	T	T	T	T	T	T
Ind FE		T		T		T
R2	0.22	0.23	0.23	0.24	0.23	0.23
N	116602	116602	114381	114381	116578	116578

Carbon premia: 2002-2021 full sample period

Table 8: Carbon emission and realized stock returns

Sample period	Panel A: 2002-2021 (Trucost sample)						Panel B: 2002-2021(XGBoost Sample)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
LOGGHG	**0.0098**	**0.0536****					**-0.0006**	**0.0053**				
	(0.48)	**(2.34)**					**(-0.05)**	**(0.68)**				
GHGGR			0.7061***	0.7412***					-0.0067	-0.0027		
			(3.72)	(3.81)					(-0.31)	(-0.13)		
GHGINTEN					-0.0081	-0.0127					-0.0136***	-0.0128***
					(-0.76)	(-1.00)					(-4.07)	(-4.54)
LOGSIZE	0.2110	0.1799	0.3317**	0.3280**	0.2034	0.1959	0.1692*	0.1514	0.2351**	0.2184**	0.1380	0.1245
	(1.31)	(1.15)	(2.06)	(2.03)	(1.30)	(1.30)	(1.72)	(1.61)	(2.34)	(2.30)	(1.45)	(1.40)
B2M	0.0051	-0.0071	0.0356	0.0401	0.0026	0.0004	0.0027	0.0014	0.0022	-0.0021	-0.0040	-0.0047
	(0.15)	(-0.20)	(1.13)	(1.29)	(0.08)	(0.01)	(0.07)	(0.03)	(0.05)	(-0.05)	(-0.10)	(-0.12)
LEVERAGE	0.0248	-0.3707	-0.0686	-0.2306	0.0132	-0.3141	0.2643	-0.0369	0.2143	-0.172	0.1872	-0.1109
	(0.11)	(-1.40)	(-0.30)	(-0.86)	(0.06)	(-1.20)	(0.80)	(-0.12)	(0.70)	(-0.62)	(0.56)	(-0.37)
MOM	-0.1588	-0.1777	-0.1819	-0.2036	-0.1584	-0.1775	-0.32***	-0.3288***	-0.3264***	-0.3372***	-0.3179***	-0.3261***
	(-1.47)	(-1.56)	(-1.56)	(-1.65)	(-1.47)	(-1.56)	(-3.21)	(-3.27)	(-2.89)	(-2.96)	(-3.14)	(-3.20)
INVEST2A	-4.1423**	-1.1284	-4.2961***	-1.3298	-4.0824**	-1.4080	-5.4709***	-3.9786***	-5.1894***	-3.3814***	-5.1657***	-3.9874***
	(-2.66)	(-0.94)	(-2.47)	(-1.13)	(-2.65)	(-1.17)	(-3.32)	(-3.13)	(-3.00)	(-2.50)	(-3.09)	(-3.08)
ROE	1.3541***	1.2599***	1.3136***	1.2813***	1.3487***	1.2730***	2.1456***	1.9609***	2.1062***	1.8976***	2.0225***	1.8924***
	(4.88)	(5.24)	(4.58)	(4.81)	(4.95)	(5.30)	(9.06)	(10.97)	(8.23)	(10.31)	(8.71)	(10.61)
HHI	0.0808	-1.3239	0.5122	-1.4447	0.0584	-1.2548	-1.1072	-2.529	-0.9551	-2.5869	-1.0636	-2.4719
	(0.07)	(-0.77)	(0.54)	(-0.85)	(0.05)	(-0.73)	(-1.30)	(-1.21)	(-0.97)	(-1.10)	(-1.26)	(-1.21)
LOGPPE	-0.0698	-0.0277	-0.0822	-0.0303	-0.0500	0.0026	0.0888	0.1306***	0.0572	0.1090**	0.0825	0.1204**
	(-1.30)	(-0.60)	(-1.41)	(-0.62)	(-0.88)	(0.06)	(1.65)	(2.65)	(0.91)	(2.04)	(1.40)	(2.42)
BETA	-0.5782**	-0.5715**	-0.5272**	-0.5147**	-0.5815**	-0.5690	-0.5642***	-0.5503***	-0.532**	-0.5207***	-0.5670***	-0.5455***
	(-2.66)	(-2.73)	(-2.30)	(-2.35)	(-2.65)	(-2.74)	(-2.80)	(-2.82)	(-2.63)	(-2.64)	(-2.83)	(-2.80)
VOLAT	0.3311***	0.3514***	0.3333***	0.3532***	0.3310***	0.3510***	0.3348***	0.3437***	0.3379***	0.3489***	0.3335***	0.3420***
	(3.49)	(3.64)	(3.37)	(3.51)	(3.49)	(3.64)	(6.70)	(6.96)	(6.10)	(6.33)	(6.61)	(6.87)
SALESGR	-0.5847*	-0.4992*	-0.6767**	-0.6188**	-0.5850**	-0.4968*	-0.5819***	-0.4867***	-0.5943***	-0.4856***	-0.6134***	-0.5333
	(-2.04)	(-1.82)	(-2.23)	(-2.14)	(-2.05)	(-1.81)	(-3.68)	(-3.04)	(-3.35)	(-2.77)	(-3.89)	(-3.38)
EPSGR	0.0510*	0.0525*	0.0375	0.0379	0.0513*	0.0523*	0.0777***	0.0842***	0.0431**	0.0500**	0.0828***	0.0869***
	(1.75)	(1.71)	(1.33)	(1.30)	(1.75)	(1.72)	(3.47)	(3.76)	(2.15)	(2.56)	(3.65)	(3.84)
Const	T	T	T	T	T	T	T	T	T	T	T	T
Year-Mon FE	T	T	T	T	T	T	T	T	T	T	T	T
Ind FE		T		T		T		T		T		T
R2	0.22	0.22	0.23	0.23	0.22	0.22	0.16	0.16	0.17	0.17	0.16	0.16
N	332410	332410	299871	299871	332338	332338	764150	764150	623506	623506	760913	760913

Supplementary evidence from XGBoost

- We provide other tests related to the carbon risks in the US equity market.
- Test (i): carbon premia and other common risk factors
- Test (ii):alternative risk stories
  - After 2016, higher emission firms are more profitable/less financially constrained/have lower product failure risk.
  - Firms need to lay off their brown business and divert to cleaner business that they are unfamiliar with and often more costly.
- Test (iii): Univariate sorting results

.

Supplementary evidence from linear models

- We use linear models with the same covariates and firm-fixed identifiers (GVKEYs) to predict GHG.
- Following a similar methodology, we fit a linear model with in-sample data and predict the carbon emissions of non-disclosure firms. We scale the emission of disclosure firms with similarity scores and then estimate the regressions.
- We perform similar pooled-OLS regressions to examine the emission-return relationship before and after 2016.
- Linear models produce noisy fits for data samples before the Paris Agreement, with an average OOS-R2 around 0.18. Moreover, the positive emission-return relationship is insignificant before 2016.
- Linear models produce more significantly negative results post-2016.

Table 9: Carbon premia: sample estimated with linear models

	Carbon emissions estimated by linear models											
Sample period	Panel A: 2002-2016 (before the Paris agreement)						Panel B: 2016-2021 (after the Paris agreement)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
LOGGHG	**0.0064** (0.40)	**0.0011** (0.15)					**-0.1541***** (-5.96)	**-0.1513***** (-3.45)				
GHGGR			0.0610*** (3.72)	0.0653*** (4.05)					-0.0311 (-0.47)	0.0203 (0.30)		
GHGINTEN					-0.0009*** (-5.21)	-0.0011*** (-5.65)					-0.0026*** (-5.26)	-0.0024*** (-4.97)
LOGSIZE	0.0823 (0.85)	0.0677 (0.75)	0.1559 (1.35)	0.1737 (1.47)	0.0620 (0.64)	0.0441 (0.49)	0.3576* (1.81)	0.3660** (2.10)	0.5110** (2.30)	0.4489** (2.28)	0.3223* (1.71)	0.2711* (1.71)
B2M	0.0278 (0.68)	0.037 (0.85)	0.0238 (0.57)	0.0457 (0.96)	0.0190 (0.51)	0.0304 (0.73)	-0.0732 (-0.89)	-0.0739 (-0.94)	-0.0223 (-0.25)	-0.0546 (-0.66)	-0.0757 (-0.91)	-0.1128 (-1.42)
LEVERAGE	0.1677 (0.52)	0.1876 (0.51)	0.1384 (0.38)	0.0549 (0.15)	-0.1329 (-0.31)	-0.0283 (-0.08)	-0.1936 (-0.35)	-0.9025** (-2.14)	-0.1074 (-0.18)	-1.0497** (-2.11)	0.0410 (0.08)	-1.1917** (-3.0)
MOM	-0.2323** (-2.20)	-0.2392** (-2.25)	-0.2443* (-1.87)	-0.2533* (-1.94)	-0.2361** (-2.23)	-0.2440** (-2.28)	-0.5335*** (-2.95)	-0.5801*** (-3.14)	-0.5277*** (-2.94)	-0.5754*** (-3.11)	-0.5305*** (-2.78)	-0.5736*** (-2.98)
INVEST2A	-5.0673*** (-2.81)	-4.6147*** (-4.38)	-5.4579*** (-3.07)	-4.4997*** (-3.93)	-4.5832** (-2.44)	-4.6578*** (-4.30)	-6.8024*** (-2.91)	-5.9003** (-1.98)	-5.5817** (-2.20)	-4.3452 (-1.39)	-7.8397*** (-3.34)	-5.6403* (-1.96)
ROE	1.7729*** (6.66)	1.6987*** (7.27)	1.7127*** (6.39)	1.6068*** (6.56)	1.4971*** (6.30)	1.5224*** (6.75)	2.8767*** (7.70)	2.406*** (13.71)	2.8686*** (7.07)	2.3213*** (11.30)	2.551*** (6.71)	2.1861*** (11.79)
HHI	-0.7581 (-1.58)	-2.3831 (-1.06)	0.1170 (0.16)	3.8975 (2.28)	-0.6862 (-1.40)	-2.3919 (-1.07)	-1.0988 (-0.44)	-8.6217 (-0.51)	-0.8969 (-0.38)	-7.2007 (-0.42)	-1.8518 (-0.75)	-7.7877 (-0.47)
LOGPPE	0.1264** (2.16)	0.1443*** (2.84)	0.1312* (1.67)	0.1423** (2.27)	0.1176* (1.86)	0.1150** (2.35)	0.1457 (1.17)	0.2047* (1.93)	-0.0447 (-0.28)	0.1043 (0.83)	-0.0389 (-0.34)	0.1228 (1.21)
BETA	-0.6271*** (-2.71)	-0.6042*** (-2.69)	-0.6081*** (-2.67)	-0.5880*** (-2.64)	-0.6409*** (-2.78)	-0.6067*** (-2.71)	-0.3380 (-1.18)	-0.2141 (-0.71)	-0.2998 (-1.05)	-0.1766 (-0.58)	-0.3447 (-1.25)	-0.2137 (-0.72)
VOLAT	0.3144*** (4.77)	0.3199*** (4.94)	0.3287*** (3.95)	0.3382*** (4.08)	0.3209*** (4.80)	0.3255*** (4.99)	0.4026*** (5.34)	0.4269*** (5.55)	0.3975*** (5.47)	0.4230*** (5.68)	0.4002*** (5.14)	0.4247*** (5.41)
SALESGR	-0.5257*** (-3.59)	-0.4917*** (-3.20)	-0.5401*** (-3.56)	-0.4830*** (-3.32)	-0.5610*** (-4.07)	-0.5827*** (-4.16)	-0.7503** (-2.00)	-0.6354* (-1.91)	-0.8721** (-2.30)	-0.7284** (-2.21)	-0.7615* (-1.94)	-0.7159** (-2.07)
EPSGR	0.0976*** (3.72)	0.1010*** (3.78)	0.0336 (1.25)	0.038 (1.50)	0.1115*** (4.02)	0.1112*** (3.98)	0.0365 (0.86)	0.0386 (0.99)	0.0229 (0.53)	0.0295 (0.74)	0.0493 (1.30)	0.0444 (1.27)
Const	T	T	T	T	T	T	T	T	T	T	T	T
Year-Mon FE	T	T	T	T	T	T	T	T	T	T	T	T
Ind FE		T		T		T		T		T		T
R^2	0.16	0.16	0.17	0.17	0.16	0.16	0.16	0.16	0.16	0.17	0.16	0.17
N	542066	542066	397310	397310	540011	540011	231285	231285	208027	208027	229913	229913

Conclusion

- This paper estimates a large panel of scope 1 emission data for listed US firms from 2002 to 2021.
  - Prior to 2016, there is a positive emission-return relationship, and it is more pronounced with XGBoost estimated data.
  - Post-2016, stocks of firms with lower emissions earned higher returns, reflecting a shift in investors' ESG-related preferences.
- XGBoost estimated data outperforms (i) the original data set provided by Trucost and (ii) the data set estimated by linear models.
- We cannot rule out other risk-driven stories: Low emission firms are less profitable/more financially constrained/have higher valuation risk.

Thanks!

State regulation and firm carbon emission

- Empirical background: California pioneered sustainable operation by setting emission reduction targets in 2005.
- Until 2022, 23 states followed quickly.
- Emission policies include carbon pricing, emission limits, renewable portfolio standards, and steps to promote cleaner transportation.
- We investigate firms carbon emissions in these Green states before and after the policy shock. We expect to see a significant decrease in these firms.
- Our identification strategy is very similar to a staggered DID as follow:

$$LOGGHG_{i,t} = \alpha + \beta REGU_{i,t} + \gamma' X_{i,t} + \mu_j + \lambda_s + \varepsilon_{i,t} \tag{6}$$

[BACK to validation page]

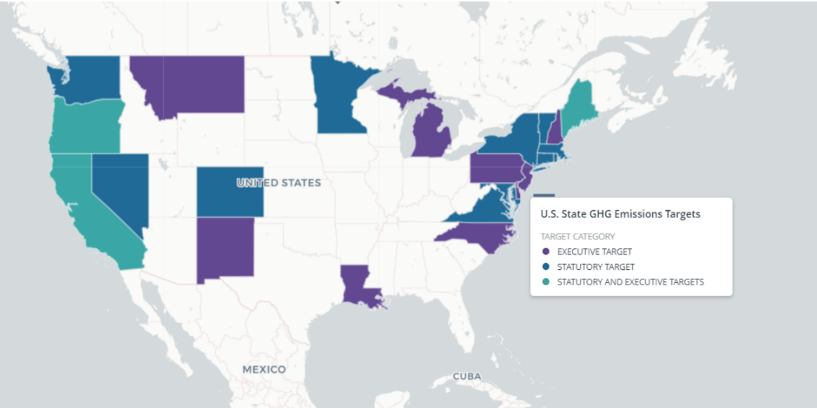


Figure 8: State emission targets

[BACK to validation page]

Table 10: State regulation and firm carbon emission

	LOGGHG			
	(1)	(2)	(3)	(4)
Regulated	**-0.2317*****	**-0.3383*****		
	(-3.66)	**(-5.18)**		
RegulateYears			**-0.0136*****	**-0.0211*****
			(-2.93)	**(-3.81)**
LOGSIZE	0.3612***	0.3626***	0.3332***	0.3351***
	(10.27)	(10.25)	(7.04)	(7.06)
B2M	0.2701***	0.2728***	0.2918***	0.2941***
	(7.16)	(7.25)	(4.80)	(4.82)
LEVERAGE	1.0599***	1.0333***	0.9260***	0.9185***
	(6.28)	(6.15)	(4.18)	(4.15)
INVEST2A	-1.5413**	-1.4396**	-2.9039***	-2.9326***
	(-2.39)	(-2.24)	(-2.74)	(-2.76)
ROE	-0.0576	-0.0856	-0.0594	-0.0773
	(-0.72)	(-1.07)	(-0.52)	(-0.68)
HHI	-1.5277*	-1.5027*	-1.2401	-1.2758
	(-1.73)	(-1.71)	(-1.11)	(-1.14)
PPE	0.3098***	0.3118***	0.3061***	0.3078***
	(8.97)	(8.95)	(6.47)	(6.48)
SALESGR	-0.0128	-0.0044	0.0190	0.0268
	(-0.22)	(-0.08)	(0.25)	(0.36)
EPSGR	0.0043	0.0056	0.0008	0.0014
	(0.54)	(0.71)	(0.07)	(0.13)
Const	T	T	T	T
Ind FE	T	T	T	T
State FE		T		T
R2	0.11	0.11	0.10	0.09
N	61739	61739	61739	61739

[BACK to validation page]

Transition Matrix

Table 11: Transition matrix of firms in each emission quntiles

	Panel A: Transition Prob. after 1 year						Panel B: Transition Prob. after 3 years				
	Q1 L0	Q2 L0	Q3 L0	Q4 L0	Q5 L0		Q1 L0	Q2 L0	Q3 L0	Q4 L0	Q5 L0
Q1 L1	**70.94%**	14.66%	8.78%	5.37%	1.83%	Q1 L3	**59.10%**	19.51%	11.95%	8.34%	2.68%
Q2 L1	13.40%	**65.16%**	17.92%	4.93%	1.09%	Q2 L3	17.40%	**52.63%**	25.78%	8.42%	1.84%
Q3 L1	8.37%	14.20%	**55.93%**	19.13%	2.18%	Q3 L3	11.84%	17.91%	**41.04%**	25.33%	3.39%
Q4 L1	5.33%	4.99%	15.69%	**60.75%**	11.56%	Q4 L3	8.45%	7.77%	17.96%	**46.70%**	15.96%
Q5 L1	1.95%	0.98%	1.67%	9.82%	**83.34%**	Q5 L3	3.21%	2.18%	3.27%	11.21%	**76.13%**
N	13952	14440	14528	14733	15157	N	10391	11027	11125	11676	12509
	Panel C: Transition Prob. after 5 years						Panel D: Transition Prob. after 7 years				
	Q1 L0	Q2 L0	Q3 L0	Q4 L0	Q5 L0		Q1 L0	Q2 L0	Q3 L0	Q4 L0	Q5 L0
Q1 L5	**51.45%**	21.03%	14.18%	9.86%	2.91%	Q1 L7	**46.86%**	21.33%	14.59%	10.15%	3.05%
Q2 L5	19.99%	**46.58%**	28.76%	12.23%	2.29%	Q2 L7	21.68%	**43.36%**	31.41%	14.25%	3.16%
Q3 L5	14.40%	19.48%	**33.71%**	26.46%	4.68%	Q3 L7	15.42%	20.23%	**29.55%**	27.88%	5.29%
Q4 L5	10.15%	9.85%	19.32%	**38.50%**	18.63%	Q4 L7	11.61%	11.20%	19.85%	**34.59%**	18.82%
Q5 L5	4.01%	3.06%	4.04%	12.94%	**71.49%**	Q5 L7	4.43%	3.87%	4.60%	13.14%	**69.69%**
N	7805	8356	8500	9093	10148	N	5849	6248	6463	7055	8135

[BACK to validation page]

[Supplementary auto-correlation test]

Emission persistency with auto-correlation test

We follow Bolton and Kacperczyk (2021a) by examining the auto-correlation of carbon emissions.

Table 12: Emission persistency with auto-correlation test

	$LOGGHG_t$		$GHGGR_t$		$GHGINTEN_t$	
	(1)	(2)	(3)	(4)	(5)	(6)
$LOGGHG_{t-1}$	0.6881***	0.5448***				
	(35.21)	(24.15)				
$GHGGR_{t-1}$			-0.0974***	-0.1088***		
			(-13.87)	(-15.96)		
$GHGINTEN_{t-1}$					0.7407***	0.7024***
					(40.16)	(33.66)
Const	T	T	T	T	T	T
Control		T		T		T
Year FE	T	T	T	T	T	T
R2	0.49	0.40	0.01	0.02	0.59	0.58
N	76113	63463	54499	48103	71256	63283

[BACK to validation page]

Emission determinants comparison

Table 13: Comparison of the determinants of carbon emission

	LOGGHG		GHGGR		GHG_INTEN	
	(1)	(2)	(3)	(4)	(5)	(6)
LOGSIZE	0.3414***	0.3542***	-0.0038*	-0.0052	-0.3366***	-1.9057***
	(14.06)	(11.23)	(-1.86)	(-0.25)	(-5.61)	(-11.28)
B2M	0.1554***	0.177***	-0.0011	-0.0228***	-0.1300***	-0.5718***
	(12.56)	(12.47)	(-0.58)	(-3.37)	(-2.76)	(-7.30)
ROE	0.2726***	-0.0876	-0.0264***	-0.1110***	-0.1391	-5.9785***
	(5.42)	(-1.19)	(-3.52)	(-2.23)	(-1.04)	(-8.28)
LEVERAGE	0.9781***	1.0461***	0.0015	0.0233	0.1629	-6.2313***
	(8.38)	(7.08)	(0.17)	(0.41)	(0.60)	(-7.57)
INVEST2A	-4.1697***	-2.2379***	0.0464	-0.2078	-4.7228***	5.0805
	(-8.31)	(-3.49)	(0.42)	(-0.75)	(-2.36)	(1.08)
HHI	0.4701	-2.3518***	0.3405***	-0.6824*	-0.4627	2.8716
	(0.98)	(-2.38)	(3.73)	(-1.75)	(-0.56)	(0.57)
LOGPPE	0.4958***	0.3356***	0.0009	0.0007	0.3357***	-0.8784***
	(19.90)	(11.14)	(0.49)	(0.09)	(7.54)	(-5.18)
SALESGR	-0.0902*	-0.0275	0.8895***	0.5855***	-0.0828	-1.7184***
	(-1.71)	(-0.61)	(19.84)	(5.41)	(-0.64)	(-3.82)
EPSGR	-0.0044	0.0082	-0.0025**	-0.0001	0.0186	0.3005***
	(-1.07)	(0.98)	(-2.03)	(-0.02)	(1.30)	(5.59)
Const	T	T	T	T	T	T
Year FE	T	T	T	T	T	T
Ind FE	T	T	T	T	T	T
R2	0.56	0.13	0.28	0.01	0.01	0.12
N	29146	67912	26089	54992	29143	67720
Data sample	Trucost	XGB	Trucost	XGB	Trucost	XGB

[BACK to validation page]

Emission and ESG fund inclusion

Table 14: Carbon emission and inclusion into ESG-related fund

	Total inclusion			Average inclusion		
	(1)	(2)	(3)	(4)	(5)	(6)
LOGGHG	-0.0752*	-0.1465***	-0.1336***	-0.0001	-0.0027*	-0.0025*
	(-1.72)	(-3.36)	(-3.33)	(-0.08)	(-1.84)	(-1.77)
IO			0.0700***			0.0102***
			(7.38)			(6.01)
Controls	T	T	T	T	T	T
Ind FE	T	T	T	T	T	T
Year FE		T	T		T	T
R2	0.50	0.52	0.57	0.68	0.71	0.72
N	67912	67912	67912	67912	67912	67912

[BACK to validation page]

XGBoost training results

- Machine learning performance.

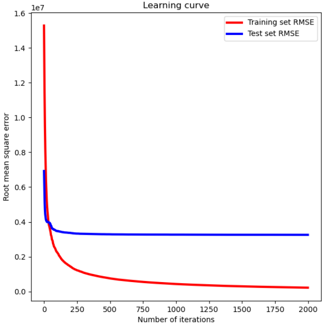


Figure 9: XGBoost learning curve

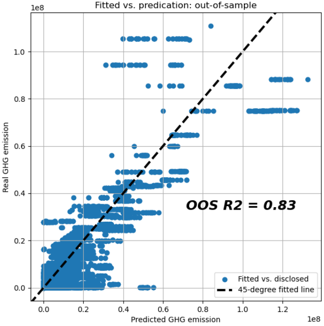


Figure 10: Fitted vs. Disclosed

[BACK TO PREVIOUS PAGE]

- Variable importance contribution plot.

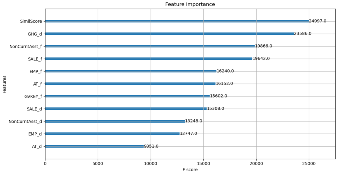


Figure 11: Importance plot

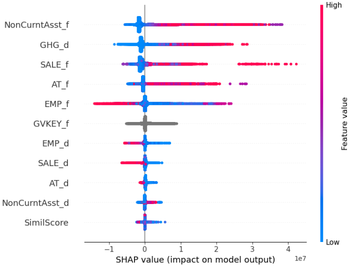


Figure 12: SHAP value plot

[BACK TO PREVIOUS PAGE]

- Cross-validation tests for different hyper-parameters.

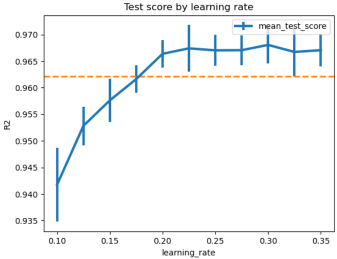


Figure 13: Cross-validation test on learning rate

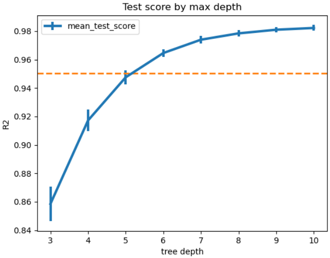


Figure 14: Cross-validation test on tree depth

[BACK to validation page]

- Training with different partitioning period.

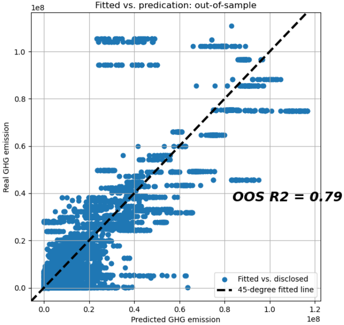


Figure 15: OOS validation (Training set: 2002-2017)

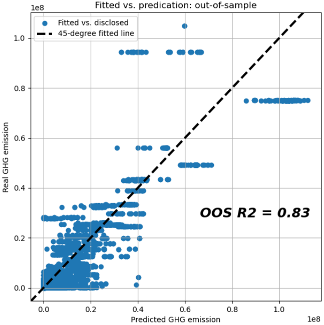


Figure 16: OOS validation (Training set: 2002-2019)

[BACK to validation page]

- Model comparison:XGBoost versus Linearmode1s.

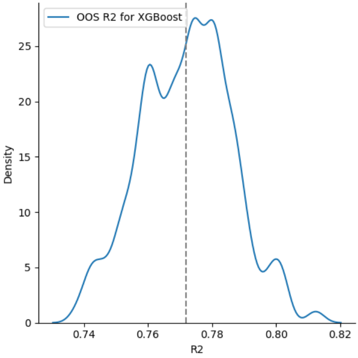


Figure 17: XGBoost model density

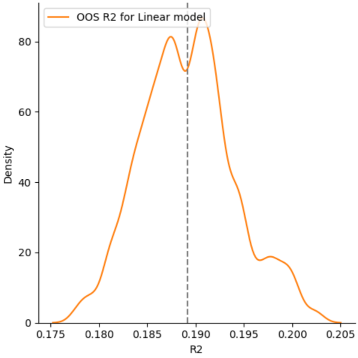


Figure 18: Linear model density

[BACK to validation page]

- Business similarity and emission similarity.

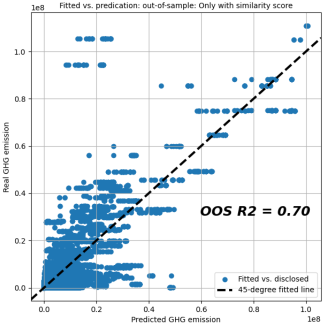


Figure 19: Prediction only with similarity score

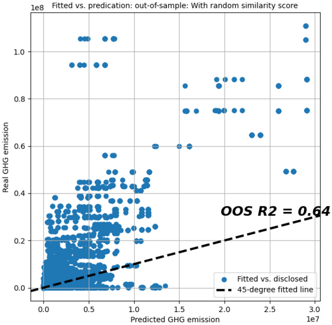


Figure 20: Prediction with random similarity score

[BACK to validation page]

- Correlation matrix.

Table 15: Correlation matrix

	LOGGHG	GHGINTEN	LOGSIZE	LEVERAGE	INVEST2A	ROE	HHI	LOGPPE	B2M	SALESGR	EPSGR
LOGGHG	1.00										
GHGINTEN	0.23	1.00									
LOGSIZE	0.32	-0.30	1.00								
LEVERAGE	-0.01	-0.09	0.07	1.00							
INVEST2A	0.25	0.03	0.08	-0.09	1.00						
ROE	0.08	-0.16	0.27	0.14	0.05	1.00					
HHI	0.00	0.02	-0.03	-0.08	-0.03	-0.04	1.00				
LOGPPE	0.47	-0.25	0.69	0.23	0.39	0.23	-0.05	1.00			
B2M	0.10	-0.03	-0.18	0.08	0.04	0.00	-0.01	0.23	1.00		
SALESGR	0.00	-0.01	0.08	-0.06	0.06	0.05	0.00	-0.04	-0.07	1.00	
EPSGR	0.02	-0.04	0.15	-0.04	-0.02	0.26	-0.02	0.05	-0.06	0.19	1.00

[BACK TO PREVIOUS PAGE]

Carbon premia and common risk factors

- We estimate the following regression, where the premia are estimated from monthly cross-sectional regressions from equation 3.
- We especially examine the post-2016 period in panel B.

$$RISKPRMM_t = \alpha + \beta' FACTOR_{i,t} + \varepsilon_t \tag{7}$$

Table 16: Carbon premia and common risk factors

	Panel A: 2002-2021						Panel B: 2016-2021					
	LOGGHG		GHGGR		GHGINTEN		LOGGHG		GHGGR		GHGINTEN	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Intercept	**-0.0209**	**-0.0135**	0.0214*	0.0158	-0.012***	-0.0132***	**-0.0957*****	**-0.0796*****	0.0274	0.0171	-0.0272***	-0.0288***
	(-1.53)	**(-1.06)**	(1.79)	(1.36)	(-4.01)	(-4.35)	**(-8.02)**	**(-4.85)**	(1.34)	(0.71)	(-5.68)	(-6.23)
RMRF	0.0013		-0.0003		-0.0006		0.0093***		-0.0151***		-0.0001	
	(0.64)		(-0.10)		(-1.17)		(2.95)		(-2.81)		(-0.10)	
SMB	-0.0079		-0.0022		-0.0037***		-0.0218***		0.0032		-0.0081***	
	(-1.54)		(-0.33)		(-3.49)		(-2.56)		(0.27)		(-6.05)	
HML	-0.0065		-0.0146**		-0.0004		-0.0036		-0.0224***		0.0004	
	(-1.42)		(-2.29)		(-0.30)		(-0.53)		(-3.13)		(0.23)	
RMW	0.0055		-0.0064		0.0003		0.0082		0.0131		-0.0023	
	(1.16)		(-1.12)		(0.23)		(1.20)		(1.00)		(-1.12)	
CMA	0.0153**		-0.0146*		-0.0022*		0.0256**		-0.0166*		-0.0034**	
	(2.57)		(-1.69)		(-1.80)		(2.37)		(-1.96)		(-2.10)	
BAB	0.0046		-0.0057		0.0001		0.0052		-0.0119		-0.0003	
	(1.31)		(-1.23)		(0.01)		(0.76)		(-1.10)		(-0.17)	
LIQ	0.001		-0.0032		0.0003		-0.0035		0.0116		-0.0011**	
	(0.39)		(-0.86)		(0.53)		(-0.95)		(1.66)		(-1.93)	
Mom	0.0001		0.0008		-0.0001		0.0046		-0.0042		-0.0020*	
	(0.05)		(0.29)		(-0.27)		(1.10)		(-0.47)		(-1.89)	
R2	0.10	0.00	0.10	0.00	0.14	0.00	0.33	0.00	0.28	0.00	0.33	0.00
N	240	240	228	228	240	240	72	72	72	72	72	72

[BACK to previous page]

Alternative risk stories

- The structural mutation in carbon premia may not be only driven by a shift in investors' ESG preference.
- Higher emission firms may be more profitable/less financially constrained/product failure risks...etc.
  - Reasoning: Firms need to adopt ESG technologies after the Paris Agreement because of regulatory changes.
  - Firms need to lay off their brown business and divert to cleaner business that they are unfamiliar with and often more costly.
- We estimate the following regression,

$$RISK_{i,t} = \alpha + GHG_{i,t} + After_t + GHG_{i,t} \times After_t + \gamma' X_{i,t-1} + \varepsilon_{i,t}, \quad (8)$$

where the dependent variable includes different risks such as profitability/liquidity/solvency/innovation/valuation, and we interact with firms' carbon emissions with a time dummy after the Paris Agreement.

[BACK to previous page]

Table 17: Carbon emission and asset prices: alternative risk stories

	XGBoost sample					
Dep.Var	Profit margin		Operating CF		R&D	
	(1)	(2)	(3)	(4)	(5)	(6)
LOGGHG	-0.0089***	0.0005	-0.0004*	-0.0002	0.0009***	0.0003**
	(-6.36)	(0.33)	(-1.79)	(-1.09)	(6.21)	(2.33)
LOGGHG*POST	**0.0487*** **	**0.0366*** **	**0.0049*** **	**0.0033*** **	**-0.0022*** **	**-0.0008*** **
	(5.76)	**(5.31)**	**(5.92)**	**(4.27)**	**(-5.44)**	**(-3.70)**
POST	-0.5982***	-0.4361***	-0.0768***	-0.0587***	0.0250***	0.0095***
	(-7.31)	(-6.67)	(-6.66)	(-5.35)	(5.53)	(3.45)
LOGSIZE	-0.0445***	-0.029***	-0.0011	0.0035***	0.0126***	0.0075***
	(-6.62)	(-3.86)	(-1.14)	(3.33)	(16.55)	(13.12)
B2M	-0.0048	-0.0081**	-0.0031***	-0.0011**	0.0012***	0.0015***
	(-1.56)	(-2.54)	(-5.47)	(-2.20)	(2.63)	(5.13)
ROE	1.6349***	1.2696***	0.2988***	0.2585***	-0.1023***	-0.0487***
	(20.94)	(25.82)	(32.03)	(31.94)	(-20.79)	(-14.81)
LEVERAGE	0.5333***	0.3014***	-0.0292***	-0.0151*	-0.0792***	-0.0001
	(6.06)	(4.85)	(-2.84)	(-1.96)	(-23.54)	(-0.03)
INVEST2A	-0.2785*	-0.0476	0.3801***	0.3386***	-0.0209	0.0575***
	(-1.76)	(-0.32)	(10.55)	(10.05)	(-1.41)	(4.88)
HHI	-0.3902***	-0.1018	0.0119	-0.0415*	0.0887***	-0.0058
	(-3.96)	(-0.70)	(0.77)	(-1.79)	(5.12)	(-0.57)
LOGPPE	0.0546***	0.0557***	0.0134***	0.0097***	-0.0129***	-0.0097***
	(8.85)	(7.94)	(14.19)	(8.82)	(-16.94)	(-17.08)
SALESGR	0.0221	0.1681***	0.0264***	0.0395***	0.0271***	0.0143***
	(0.80)	(5.39)	(4.41)	(8.03)	(9.33)	(10.39)
EPSGR	-0.0252***	-0.0131***	-0.0047***	-0.0032***	0.0033***	0.0017***
	(-5.52)	(-4.52)	(-5.83)	(-5.07)	(9.51)	(8.84)
Const	T	T	T	T	T	T
Ind FE		T		T		T
R2	0.22	0.35	0.39	0.60	0.10	0.24
N	67720	67720	67356	67356	67911	67911

[BACK to previous page]

Table 17 Cont'd

	XGBoost sample					
Dep.Var	Solvency		Tobin's Q		Divid payout	
	(7)	(8)	(9)	(10)	(11)	(12)
LOGGHG	-0.7007***	-0.1924	-0.001	-0.0124***	-0.0037***	-0.0012
	(-3.70)	(-1.06)	(-0.37)	(-4.41)	(-2.72)	(-0.87)
LOGGHG*POST	**1.246****	**1.0284****	**-0.0121****	**-0.0089**	**0.0058*****	**0.0056*****
	(2.46)	**(2.20)**	**(-2.13)**	**(-1.51)**	**(2.86)**	**(2.74)**
POST	-21.2084***	-17.5476***	0.2503***	0.1798**	-0.1027***	-0.0942***
	(-3.48)	(-3.10)	(2.87)	(2.08)	(-3.33)	(-3.32)
LOGSIZE	4.7327***	4.2763***	0.4184***	0.4686***	0.0891***	0.0796***
	(5.67)	(4.86)	(19.61)	(21.07)	(14.99)	(12.45)
B2M	0.9898*	0.722	0.0211**	0.0678***	0.0112***	0.0064*
	(1.78)	(1.39)	(2.28)	(8.44)	(3.13)	(1.74)
ROE	55.2045***	45.1487***	-0.3129***	-0.0245	0.4342***	0.3837***
	(18.98)	(15.74)	(-6.40)	(-0.55)	(24.11)	(19.96)
LEVERAGE	-105.5669***	-112.1008***	-0.8742***	0.019	0.2578***	0.0899***
	(-11.91)	(-13.43)	(-10.39)	(0.24)	(9.26)	(3.29)
INVEST2A	81.4961***	100.5752***	5.4758***	5.497***	-0.4401***	-0.5383***
	(4.76)	(5.73)	(15.90)	(17.46)	(-3.12)	(-3.33)
HHI	7.5768	-16.6693	1.8527***	0.8196**	-0.0701	-0.0173
	(0.61)	(-0.79)	(7.85)	(2.59)	(-0.78)	(-0.09)
LOGPPE	-1.8233**	-0.8269	-0.2960***	-0.3662***	-0.0251***	-0.008
	(-2.39)	(-0.98)	(-19.40)	(-21.30)	(-4.56)	(-1.32)
SALESGR	3.3622**	8.3616***	0.3992***	0.3027***	-0.2147***	-0.194***
	(2.04)	(5.13)	(14.18)	(11.51)	(-9.01)	(-8.33)
EPSGR	0.2583	0.6511***	0.0141**	0.0062	0.022***	0.0234***
	(1.12)	(3.11)	(2.47)	(1.35)	(7.39)	(7.67)
Const	T	T	T	T	T	T
Ind FE		T		T		T
R2	0.08	0.16	0.08	0.27	0.17	0.20
N	51379	51379	67873	67873	63766	63766

[BACK to previous page]

Univariate sorting results

.

Table 18: Univariate sorting results

Panel A: Sorting results based on XGBoost estimated data								
Portfolio	VW return	EW return	LOGGHG	GHGGR	GHGINTEN	SALE	AT	LOGSIZE
Lo	0.6336*	1.1785***	3.98	-0.16	1.39	746.49	1980.58	12.55
	(1.78)	(2.81)						
2	0.8837***	1.1544***	9.74	0.30	6.59	1515.48	2453.22	13.11
	(2.88)	(2.85)						
3	0.7291**	1.0347**	10.98	0.51	9.03	2322.34	3121.30	13.35
	(2.31)	(2.51)						
4	0.7652***	1.0837***	12.20	0.62	9.97	5541.45	7575.81	13.88
	(2.92)	(2.80)						
Hi	0.5343**	0.8782**	14.55	0.54	11.21	19320.72	27068.58	14.99
	(2.02)	(2.36)						
Hi-Lo	-0.0993	-0.3003						
	(-0.49)	(-1.41)						
Portfolio	LEVERAGE	B2M	INVEST2A	ROE	HHI	LOGPPE	SALESGR	EPSGR
Lo	0.49	0.73	0.03	-0.08	0.10	3.18	0.12	-0.13
2	0.50	0.88	0.04	-0.04	0.10	4.04	0.11	-0.12
3	0.54	0.81	0.05	0.00	0.10	4.75	0.11	-0.08
4	0.58	1.06	0.06	0.07	0.10	5.92	0.10	0.01
Hi	0.64	1.69	0.06	0.10	0.09	7.87	0.08	0.07

[BACK to previous page]

Table 18 Cont'd

Panel A: Sorting results based on XGBoost estimated data								
Portfolio	VW return	EW return	LOGGHG	GHGGR	GHGINTEN	SALE	AT	LOGSIZE
Lo	0.6336*	1.1785***	3.98	-0.16	1.39	746.49	1980.58	12.55
	(1.78)	(2.81)						
2	0.8837***	1.1544***	9.74	0.30	6.59	1515.48	2453.22	13.11
	(2.88)	(2.85)						
3	0.7291**	1.0347**	10.98	0.51	9.03	2322.34	3121.30	13.35
	(2.31)	(2.51)						
4	0.7652***	1.0837***	12.20	0.62	9.97	5541.45	7575.81	13.88
	(2.92)	(2.80)						
Hi	0.5343**	0.8782**	14.55	0.54	11.21	19320.72	27068.58	14.99
	(2.02)	(2.36)						
Hi-Lo	-0.0993	-0.3003						
	(-0.49)	(-1.41)						
Portfolio	LEVERAGE	B2M	INVEST2A	ROE	HHI	LOGPPE	SALESGR	EPSGR
Lo	0.49	0.73	0.03	-0.08	0.10	3.18	0.12	-0.13
2	0.50	0.88	0.04	-0.04	0.10	4.04	0.11	-0.12
3	0.54	0.81	0.05	0.00	0.10	4.75	0.11	-0.08
4	0.58	1.06	0.06	0.07	0.10	5.92	0.10	0.01
Hi	0.64	1.69	0.06	0.10	0.09	7.87	0.08	0.07

[BACK to previous page]