Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 11(6): 16448–16478.
DOI: 10.3934/math.2026675
Received: 23 April 2026
Revised: 21 May 2026
Accepted: 29 May 2026
Published: 09 June 2026

Research article

From risk to reward: Learning ESG return dynamics in Chinese equities

Maurice Kyla Octaviano and Jin-Taek Seong*

Graduate School of Data Science, Chonnam National University, Gwangju 61186, Republic of Korea

* **Correspondence:** Email: jtseong@jnu.ac.kr; Tel: +82-62-530-5798.

Abstract: This study re-evaluated the role of environmental, social, and governance (ESG) characteristics in China's A-share market over the period 2012Q1 to 2021Q3. Contrary to the prevailing view of ESG as a defensive, risk-mitigating attribute, we showed that passive high-ESG strategies fail to generate excess returns. To address this inefficiency, we introduced a meta-learning framework that integrates three heterogeneous base learners—LightGBM for nonlinear cross-sectional effects, ridge regression as a linear benchmark, and a temporal convolutional network (TCN) for sequential earnings dynamics—whose out-of-fold predictions were stacked via a second-stage ridge meta-learner trained on a held-out validation set. This architecture predicts next-quarter earnings per share (EPS-TTM) and ranks firms by fundamental growth potential within a high-ESG universe screened to the top tercile of ESG scores. The framework was trained on 2012Q1–2017Q4, validated on 2018Q1–2019Q1, and evaluated out-of-sample on 2019Q2–2021Q3 using a strict chronological split that eliminates look-ahead bias. We documented a clear performance hierarchy. While the broad high-ESG universe exhibits negative alpha relative to the CSI 300 benchmark, a refined top 50% portfolio largely neutralizes this underperformance, and economically meaningful excess returns are concentrated in the top 10% portfolio, with annualized alpha exceeding 40% over the out-of-sample evaluation period (Sharpe ratio: 1.11; market beta: 1.16). Industry-neutralized robustness tests confirmed that this alpha is attributable to stock-level fundamental prediction rather than passive sector concentration. These results support a risk-signaling interpretation of ESG in emerging markets: ESG-related return premia in China reflect compensation for growth-related systematic risk under conditions of disclosure heterogeneity and information asymmetry, rather than downside protection.

Keywords: ESG investing; machine learning; meta-learning; China A-shares; return premia
Mathematics Subject Classification: 91G10, 68T05, 62P05

1. Introduction

environmental, social, and governance (ESG) investing has transitioned from a niche strategy to a central component of global capital allocation. In much of the literature, ESG integration is framed through a risk-mitigation lens, emphasizing its role in reducing tail risk, regulatory exposure, and downside volatility [1]. However, growing evidence from emerging markets suggests that ESG characteristics may operate through a different economic channel. Rather than providing insurance-like protection, ESG information may proxy for firm quality, innovation capacity, and growth-related risk exposure, particularly in markets characterized by information frictions and heterogeneous disclosure standards.

China's A-share market offers a natural setting to examine this alternative mechanism. As the world's largest emerging equity market, China combines rapid regulatory development—most notably following the 2016 Guidelines for Establishing the Green Financial System—with high retail participation and substantial information asymmetry [2]. In this environment, ESG scores often reflect firms operating in disclosure-sensitive industries such as technology and advanced manufacturing, where compliance, transparency, and innovation are critical for maintaining market access [3]. As a result, high-ESG firms in China may exhibit elevated systematic risk rather than defensive characteristics, challenging the conventional interpretation of ESG as a low-volatility attribute.

Despite this potential, empirical evidence on ESG-related return premia in China remains mixed. While some studies document positive ESG excess returns [4], others find weak or unstable performance once standard risk factors are controlled for. A key limitation of existing approaches is their reliance on linear factor models, which may be ill-suited to capture the nonlinear and conditional relationship between ESG characteristics and firm fundamentals in inefficient markets. Recent work highlights that machine learning methods—particularly tree-based and ensemble models—are better equipped to extract economic signals from ESG data when return generation is driven by complex interactions rather than monotonic effects [5].

Motivated by these considerations, this study introduces a meta-learning framework that integrates light gradient boosting machines (LightGBMs), ridge regression, and temporal convolutional networks (TCNs) to forecast forward earnings per share for Chinese A-share firms. By shifting the prediction target from noisy asset returns to fundamental profitability, the proposed approach identifies firms within the high-ESG universe that exhibit strong growth potential. We then construct ESG-segmented portfolios based on predicted earnings growth and evaluate their performance relative to the CSI 300 benchmark.

Our findings provide consistent evidence against the ESG-as-insurance hypothesis. Instead, we document an "ESG alpha ladder," in which increasingly selective portfolios—constructed using machine-learning-based forecasts of firm fundamentals within a high-ESG universe—deliver higher returns accompanied by proportionally higher exposure to systematic risk. This pattern supports a risk-signaling interpretation of ESG in emerging markets, whereby ESG characteristics, when incorporated as informational inputs into fundamental growth prediction, help identify firms with aggressive growth profiles rather than defensive risk properties.

This study contributes to the literature in three ways. First, it advances ESG asset pricing research by demonstrating that ESG-related return patterns in China are conditional on nonlinear, machine-learning-based fundamental forecasting rather than passive ESG score-based sorting. Second, it provides robust out-of-sample evidence that excess returns are concentrated among a small subset of firms with strong predicted earnings growth, rather than broadly across high-ESG stocks. Third, it reframes ESG in emerging markets as an information-based signal that operates through exposure to growth-oriented systematic risk, offering a unified explanation for the mixed findings in prior studies.

2. Review of related work

2.1. *ESG and market performance: Evidence from emerging markets*

The relationship between environmental, social, and governance (ESG) performance and equity returns has been widely studied, with empirical findings differing markedly across market settings. In developed markets, the dominant interpretation emphasizes a cost-of-capital channel. Firms with strong ESG profiles tend to face lower financing costs and reduced regulatory or tail risks, which in equilibrium translates into lower expected returns to equity holders [6,7]. Under this framework, ESG characteristics are largely viewed as defensive attributes that appeal to investors with non-pecuniary preferences, rather than as sources of abnormal returns.

In contrast, a growing body of evidence from emerging markets documents a positive association between ESG performance and subsequent stock returns. Focusing on China's A-share market, [4] rejected the interpretation of ESG as a purely non-pecuniary or "luxury good." Instead, they showed that firms with higher ESG scores earn significantly higher future returns, even after controlling for conventional risk factors. The authors argued that in markets characterized by high retail participation and substantial information asymmetry, ESG disclosures serve as a signal of firm quality and managerial efficiency that is only gradually incorporated into prices.

Complementary evidence was provided by [8], who found that superior environmental performance predicts positive abnormal returns among Chinese firms, particularly those operating under stricter regulatory scrutiny. Unlike in developed markets—where ESG-related information is rapidly arbitraged away—market frictions in China allow ESG-related return premia to persist. These findings suggest that the pricing of ESG characteristics in emerging markets is conditional on investors' ability to process complex and heterogeneous disclosure information, rather than reflecting a uniform risk-reduction effect.

2.2. *Volatility as an information mechanism*

A central debate in the ESG literature concerns whether sustainability characteristics function primarily as risk-mitigating attributes or as signals of firm-specific growth and innovation. The traditional insurance hypothesis posits that strong ESG performance reduces exposure to regulatory shocks, litigation risk, and extreme downside events, thereby lowering return volatility and stock price crash risk [1]. While this mechanism may hold at the aggregate market level, it is less informative for understanding the behavior of firms at the frontier of ESG performance.

An alternative perspective was provided by the information efficiency framework of [9]. Studying Chinese listed firms, they showed that higher-quality ESG disclosure is associated with significantly lower stock price synchronicity, indicating that firm-level returns are less driven by market-wide movements and more reflective of firm-specific information. Reduced synchronicity implies higher idiosyncratic volatility, not as a manifestation of distress risk, but as a consequence of more efficient incorporation of firm-level news and fundamentals into prices.

This distinction is central to our analysis. If ESG characteristics convey economically meaningful information about firm-specific growth prospects, then portfolios composed of selectively identified high-ESG firms are expected to exhibit elevated volatility relative to the market. In this setting, higher volatility reflects increased information content and fundamental discovery rather than heightened exposure to downside risk, consistent with an information-driven view of ESG-related return dynamics.

2.3. The sensitive industry effect and risk signaling

An additional dimension shaping the risk profile of ESG portfolios in emerging markets arises from their sectoral composition. Firms operating in environmentally and socially sensitive industries—such as heavy industrials, chemicals, and technology hardware—often exhibit stronger ESG performance than firms in lower-impact sectors [3]. This pattern is consistent with legitimacy theory, which posits that firms facing greater regulatory scrutiny and reputational risk have stronger incentives to invest in ESG practices in order to secure their license to operate and signal long-term viability to external investors.

In the Chinese market, this mechanism implies that ESG-screened portfolios tend to be disproportionately exposed to capital-intensive and structurally volatile industries. This composition differs markedly from ESG portfolios in developed markets, which frequently tilt toward lower-volatility sectors such as services or software. Importantly, this exposure does not reflect a passive sector bias, but rather the concentration of ESG leaders within industries where disclosure quality, compliance, and innovation are critical for survival.

By selectively identifying top-performing ESG firms within these sensitive sectors, learning-based ESG strategies implicitly target innovative firms operating under heightened regulatory and competitive pressure. The resulting portfolios therefore exhibit elevated systematic risk alongside higher expected returns, consistent with a risk-signaling interpretation of ESG characteristics in emerging markets [3].

2.4. Methodological advances: Beyond linear models

The mixed empirical findings in the ESG–return literature may partly reflect the limitations of the modeling techniques traditionally employed. Most asset pricing studies rely on linear factor regressions, such as the Fama–French framework, which implicitly assume a stable and monotonic relationship between ESG characteristics and stock returns [10]. These assumptions may be overly restrictive in settings where ESG information interacts with firm fundamentals in nonlinear and state-dependent ways, particularly in emerging markets.

Recent evidence supports this concern. In China's A-share market, the relationship between ESG performance and financial outcomes has been shown to be highly nonlinear and conditional on firm-specific characteristics, including firm size and ownership structure [5]. Gradient-boosting decision tree models, such as light gradient boosting machines (LightGBMs), substantially outperform linear specifications by capturing threshold effects—where ESG contributes to performance only beyond certain levels—as well as higher-order interaction effects that linear models fail to detect.

Building on these methodological insights, this study integrates cross-sectional machine learning with temporal modeling. While LightGBM captures nonlinear relationships across firms, temporal convolutional networks (TCNs) are employed to model long-range dependencies in earnings dynamics [11]. This combination is designed to separate persistent fundamental signals from short-term fluctuations, allowing ESG-related return premia to emerge through time rather than instantaneously. By shifting the prediction target from noisy asset returns to forward-looking profitability measures, the proposed framework addresses key identification challenges highlighted in prior studies.

To situate this approach within the broader literature, Table 1 summarizes influential contributions to ESG asset pricing in emerging markets. The table illustrates the evolution of the literature from early risk-mitigation interpretations [1] toward alpha-oriented and risk-signaling perspectives [3, 4], alongside a parallel shift toward nonlinear modeling techniques [5]. Against this backdrop, our study contributes by combining nonlinear cross-sectional learning with temporal modeling to examine when and how ESG characteristics translate into risk-adjusted return premia in China.

Table 1. Summary of selected ESG asset pricing studies.

Author(s) / Year	Market / Scope	Methodology	Key Findings
[1] / 2021	China (A-shares)	Event study (COVID-19)	High-ESG firms experienced smaller drawdowns during the pandemic, supporting the ESG-as-insurance hypothesis.
[4] / 2023	China (A-shares)	Factor models	Documents an ESG return premium not explained by standard Fama–French factors.
[3] / 2017	Emerging markets	Panel regression	ESG performance serves as a signal of legitimacy and long-term viability in sensitive industries.
[9] / 2024	China (A-shares)	Synchronicity analysis	Higher-quality ESG disclosure reduces stock price synchronicity, increasing firm-specific information content.
[5] / 2024	China (A-shares)	LightGBM (machine learning)	ESG–return relationships are highly nonlinear; ML models outperform linear benchmarks.
[8] / 2025	China (A-shares)	Portfolio sorts	Environmental performance predicts positive abnormal returns, especially in high-pollution sectors.
[7] / 2025	Global (developed vs. emerging)	Rolling-window analysis	ESG indices outperform benchmarks over long horizons, with limited short-term differentiation.
[12] / 2025	China (Shenzhen)	LSTM + Fama–French six-factor model	Deep sequential models capture time-varying factor exposures; single-model approach limits flexibility compared to ensemble frameworks.
[13] / 2025	Multiple markets	Hybrid ensemble models	Hybrid ensemble models consistently outperform single-model baselines in stock market prediction, validating multi-model fusion strategies.
[14] / 2025	China (A-shares)	Information value analysis	Analyst forecast data exhibits significant heterogeneity in informational value across firms, with lower coverage linked to greater mispricing—supporting the use of alternative ESG-based signals.

Several recent studies are directly relevant to the methodological choices made in this paper. [12] integrated LSTM with the Fama–French six-factor model for Shenzhen-listed firms, demonstrating that deep sequential models can capture time-varying factor exposures. Unlike their single-model approach, our framework ensembles heterogeneous learners—TCNs for temporal dynamics, LightGBMs for nonlinear cross-sectional interactions, and ridge regressions for linear structure—targeting EPS prediction rather than returns directly. [13] provided broader validation that hybrid ensemble models consistently outperform single-model baselines in stock market prediction, supporting the multi-model fusion strategy adopted here. Finally, [14] documented significant heterogeneity in the informational value of analyst forecasts across Chinese A-share firms, with limited analyst coverage linked to greater mispricing. This finding motivates our use of NLP-derived ESG scores as an alternative information source for firms where conventional analyst coverage is sparse, and justifies the ESG×size and ESG×earnings yield interaction terms in our feature design.

3. Data and methodology

3.1. Data sourcing and sample construction

The sample universe consists of Chinese A-share firms listed on the Shanghai and Shenzhen stock exchanges. The observation period spans from 2011Q1 to 2021Q3, covering multiple market regimes, including the 2015 equity market crash and the COVID-19 recovery period. This timeframe allows us to examine ESG-related return dynamics across both stress and expansionary environments.

Firm-level ESG scores are obtained from the publicly available dataset of [15], published on the Harvard Dataverse (DOI: 10.7910/DVN/PQTWVZ, V2). The underlying methodology is described in [16], which employs natural language processing and machine learning techniques applied to corporate social responsibility reports, annual disclosures, and firm-level filings to construct ESG scores for Chinese A-share firms. The resulting scores span a range of 7.85 to 60.61 (mean: 26.13; standard deviation: 7.66) and exhibit relatively stable distributional properties compared to financial statement variables, as reported in Table 2. Because the dataset is derived from a peer-reviewed published source with transparent construction methodology and a permanent Digital Object Identifier (DOI), it offers greater reliability and reproducibility than proprietary rating systems. This approach mitigates survivorship and disclosure bias and enables the construction of a balanced ESG panel across firms with heterogeneous reporting quality. All ESG scores are based on disclosures available prior to the start of the portfolio formation quarter, ensuring no forward-looking information enters the feature set. Portfolio returns are measured using quarterly excess returns, defined as the realized stock return net of the one-year Chinese government bond yield, consistent with standard asset pricing practice [17]. The CSI 300 benchmark is adjusted to the same excess return basis by subtracting the contemporaneous risk-free rate to ensure comparability across all performance metrics. Financial statement variables—including return on equity (ROE), net profit margin, earnings yield, and profitability growth measures—are derived from quarterly corporate filings.

Daily stock prices and trading volumes are sourced from Baostock and aggregated to a quarterly frequency to align with the portfolio rebalancing schedule. Excess returns are computed using a local risk-free rate proxy. Following the asset pricing framework for China proposed by [18], the one-year Chinese government bond yield is used as the risk-free rate. Monthly bond yields, obtained from Investing.com [19], are aggregated to a quarterly frequency.

To mitigate survivorship bias, the stock universe is defined on a quarter-by-quarter basis using only information available at the portfolio formation date, allowing firms to enter and exit the sample over time. No ex-post filtering based on future index membership, survival status, or data availability is applied.

Table 2 reports summary statistics for the key variables prior to filtering and standardization. The smaller observation count for financial statement variables (approximately 18,000 vs. 20,170 for ESG and return variables) reflects the requirement that quarterly earnings data be available for all lagged features, resulting in fewer matched firm-quarter observations after the inner join.

Table 2. Summary statistics of key variables. Several earnings-related variables (net profit margin, profit growth, EPS growth) exhibit substantial right-skewness driven by small-denominator effects; all continuous variables are Winsorized at the 1st and 99th percentiles within each quarterly cross-section prior to model training.

Variable	Count	Mean	Std. Dev.	Min	Max
Share Dilution	20,141	0.099	0.555	-0.253	30.41
Net Profit Margin	20,163	0.471	49.79	-52.30	7,068.67
Profit Growth	20,141	1.154	44.98	-1,835.81	3,963.29
ESG Score	20,170	26.13	7.66	7.85	60.61
EPS (TTM)	20,170	0.514	0.967	-16.46	30.13
Excess Return	20,170	0.014	0.254	-0.786	2.739
Gross Profit Margin	19,533	0.310	0.202	-1.813	0.979
Earnings Yield	20,170	0.022	0.103	-4.165	1.027
Firm Size (Log)	20,170	23.33	1.01	20.41	27.95
EPS Growth	20,140	0.770	88.48	-900.56	12,247.76
ROE (Average)	20,105	0.047	1.13	-158.24	2.83
Liquidity Proxy	20,168	2.04×10^7	4.16×10^7	0.00	1.79×10^9
Volatility (High–Low)	20,170	0.366	0.245	0.000	4.164

3.2. Temporal data partitioning

To ensure that model estimation and evaluation are free from look-ahead bias, the dataset is partitioned into three non-overlapping subsets using a strict chronological split. The training set spans 2012Q1 to 2017Q4 and is used exclusively to fit the three base learners (LightGBM, ridge regression, and TCN). The validation set covers 2018Q1 to 2019Q1 and is used to generate out-of-fold predictions for training the second-stage meta-learner, as well as to select hyperparameters for each base model via held-out mean squared error minimization. The test set spans 2019Q2 to 2021Q3 and serves as the strictly out-of-sample evaluation window; no model parameters are updated after the validation phase. All feature normalization—including Winsorization at the 1st and 99th percentiles and cross-sectional z-score standardization—is applied independently within each quarter using only information available up to that quarter, eliminating any form of forward-looking data contamination. A strict chronological split rather than cross-validation is employed because random folds would leak future quarterly information into the training set. Hyperparameters for LightGBM (number of leaves, learning rate, minimum child samples) and ridge (ℓ_2 regularization coefficient λ) are selected by minimizing validation-set MSE. The TCN architecture (kernel size, dilation factors, dropout rate) is fixed prior to training to avoid overfitting the short validation window.

3.3. Robust preprocessing and feature engineering

Given the high noise, elevated volatility, and retail-dominated trading environment of the Chinese equity market, we implement a robust preprocessing pipeline designed to mitigate the influence of extreme observations and prevent look-ahead bias. These considerations are reflected in the summary statistics reported in Table 2, which reveal substantial heterogeneity and heavy-tailed distributions across firm-level variables.

Several financial ratios exhibit extreme dispersion, particularly net profit margin, profit growth, and EPS growth, whose maximum values exceed several thousand percent. Such extremes arise from small-denominator effects, earnings volatility, and accounting discontinuities that are common in emerging markets. Share dilution and earnings yield also display wide ranges with pronounced skewness, while liquidity and volatility measures span multiple orders of magnitude across firms. In contrast, firm size and ESG scores exhibit comparatively stable distributions.

These distributional characteristics motivate the use of robust preprocessing techniques, including cross-sectional Winsorization and variable-specific transformations, to ensure numerical stability and to prevent a small number of extreme observations from disproportionately influencing model estimation. All preprocessing steps are conducted using only information available at each rebalancing date to preserve the integrity of the out-of-sample evaluation.

3.3.1. Target variable construction (fundamental prediction)

To align the modeling objective with fundamental value creation, we define the prediction target as next-quarter earnings performance measured using earnings per share (EPS) on a trailing twelve-month (TTM) basis. Using TTM earnings mitigates seasonal distortions inherent in quarterly reporting and provides a smoother representation of firm profitability, which has been shown to better reflect persistent earnings capacity [20, 21].

Formally, the prediction target for firm i in quarter $t+1$ is defined as the cross-sectionally standardized forward EPS:

$$Y_{i,t+1} = \frac{\mathrm{EPS}_{i,t+1}^{\mathrm{TTM}} - \mu_{t+1}}{\sigma_{t+1}}, \tag{3.1}$$

where μ_{t+1} and σ_{t+1} denote the cross-sectional mean and standard deviation of TTM EPS across all firms in quarter $t + 1$. Cross-sectional standardization shifts the learning objective from predicting absolute earnings levels to predicting relative performance rankings, which is consistent with asset pricing theory and the construction of cross-sectional trading strategies [22, 23].

By focusing on forward-looking profitability rather than contemporaneous or lagged stock returns, this target specification reduces the influence of short-term price noise and allows the learning model to capture economically meaningful variation in firm fundamentals that may subsequently translate into return premia.

3.3.2. Outlier management and stationarity

Firm-level financial variables in emerging markets frequently exhibit heavy-tailed distributions and extreme observations that can destabilize both gradient-based learning algorithms and parametric estimators. To mitigate the influence of such extremes while preserving economically meaningful variation, we apply a two-step preprocessing procedure at each quarterly rebalancing date.

Winsorization. All continuous financial variables are Winsorized at the 1st and 99th percentiles within each quarterly cross-section. This approach caps extreme observations without discarding data points, thereby preserving sample size while limiting the undue influence of outliers on model estimation. Cross-sectional Winsorization is standard practice in empirical asset pricing and corporate finance, particularly in settings characterized by accounting discontinuities and heterogeneous firm size [24, 25].

Logarithmic transformation. Variables exhibiting pronounced right-skewness—specifically liquidity, volatility, and share dilution—are further transformed using a log-plus-one specification:

$$X'_{i,t} = \ln\left(1 + X_{i,t}\right), \tag{3.2}$$

where $X_{i,t}$ denotes the raw value of the variable for firm i in quarter t, and $X'_{i,t}$ represents the transformed feature used in model estimation. This transformation stabilizes variance and reduces distributional skewness, improving numerical conditioning and aiding regularized estimation without imposing strict normality assumptions [26, 27].

All transformations are conducted using only information available at time t to ensure stationarity and to prevent look-ahead bias in subsequent model training and evaluation.

3.3.3. Cross-sectional standardization

To ensure that the learning model captures relative ranking information rather than absolute nominal levels—which may be distorted by inflation, macroeconomic regimes, or structural shifts in the Chinese equity market—all input features are standardized into cross-sectional z-scores at each quarter t:

$$Z_{i,t} = \frac{X_{i,t} - \mu_t}{\sigma_t}, \tag{3.3}$$

where $X_{i,t}$ denotes the value of a given feature for firm i in quarter t, and μ_t and σ_t represent the cross-sectional mean and standard deviation across all firms at time t, respectively.

Cross-sectional standardization ensures that the model learns firms' relative positions within the contemporaneous investment universe, rather than time-series variation driven by aggregate market conditions. This approach is standard in empirical asset pricing and factor construction, where returns and firm characteristics are evaluated on a relative basis [17, 28]. By construction, higher z-scores correspond to stronger signals relative to peers at a given point in time, rendering the features robust to market-wide level shifts and facilitating rank-based portfolio formation.

3.3.4. Look-ahead bias prevention

A critical requirement in predictive asset pricing is the strict prevention of look-ahead bias. All financial ratios, accounting variables, and firm characteristics are therefore lagged by one quarter relative to the prediction target, ensuring that only information available at the portfolio formation date is used. This timing convention follows standard practice in return prediction and accounting-based asset pricing studies [29, 30]. After applying outlier treatment, transformation, cross-sectional standardization, and lag alignment, the final dataset consists of 18,064 firm-quarter observations. Table 3 reports descriptive statistics for the fully processed feature set used in model estimation.

Table 3. Descriptive statistics of processed data.

Variable	Count	Mean	Std. Dev.	Min	Max
Share Dilution	18,064	0.101	0.537	-0.253	30.41
Net Profit Margin	18,060	0.123	0.994	-17.68	68.06
Profit Growth	18,064	0.988	41.37	-1,835.81	3,963.29
ESG Score (z)	18,064	-0.013	0.994	-2.581	5.010
EPS (TTM)	18,064	0.508	0.931	-16.46	30.13
Excess Return	18,064	-0.002	0.998	-4.435	10.42
Gross Profit Margin	18,059	0.311	0.201	-1.813	0.979
Earnings Yield	18,064	0.023	0.099	-4.165	1.027
Firm Size (log, z)	18,064	-0.028	0.980	-2.867	3.810
EPS Growth	18,064	0.813	92.55	-557.43	12,247.76
ROE (Average)	18,051	0.055	0.173	-11.05	2.472
Liquidity Proxy	18,064	1.98×10^7	4.22×10^7	0.000	1.79×10^9
Volatility (High–Low)	18,064	0.367	0.245	0.000	4.164

Table 3 illustrates the effect of the preprocessing pipeline on the distribution of firm-level variables. Compared to the raw inputs reported in Table 2, the processed features exhibit substantially reduced dispersion and more stable ranges, indicating that extreme observations no longer dominate model inputs. Most standardized variables are centered close to zero with unit-scale dispersion, consistent with cross-sectional normalization. Importantly, economically meaningful variation is preserved, particularly for growth- and volatility-related measures, which retain fat-tailed characteristics reflecting genuine firm-level heterogeneity rather than data artifacts.

3.4. Conditional feature generation

Standard linear factor models typically impose monotonic and homogeneous relationships between firm characteristics and expected returns. In the context of ESG, such restrictions may be particularly limiting, as recent evidence suggests that the valuation relevance of ESG information in China is highly conditional on firm-specific attributes [5]. To accommodate these nonlinear and state-dependent effects, we construct interaction features that allow ESG signals to vary across economically meaningful firm dimensions.

First, we interact ESG scores with firm size to test the visibility hypothesis. Larger firms are subject to greater media attention, regulatory scrutiny, and institutional monitoring, which may amplify the informational content of ESG disclosures [10]. Under this mechanism, ESG characteristics are expected to be more informative for firms with greater market visibility.

Second, we interact ESG scores with return on equity (ROE) to capture a quality channel. This interaction evaluates whether ESG functions as a differentiating signal among firms with strong profitability, helping to identify high-quality growth firms rather than unprofitable or distressed issuers [31]. In this setting, ESG information may reinforce signals of managerial efficiency and sustainable earnings capacity.

Third, we construct interactions between ESG scores and earnings yield to examine a value-based channel. Firms with higher earnings yield are often associated with greater mispricing and information asymmetry, particularly in emerging markets. This interaction tests whether ESG signals are more informative in segments of the market where valuation uncertainty is higher [4].

These interaction features are supplied as inputs to the nonlinear base learners, particularly LightGBMs, allowing the model to capture threshold effects and regime-dependent relationships without imposing restrictive functional forms. Rather than assuming a uniform ESG effect across firms, this approach enables the data to determine when ESG characteristics convey incremental information about future fundamentals and when they do not.

3.5. The meta-learning architecture

We employ a stacked generalization (meta-learning) framework to jointly capture cross-sectional nonlinearities, linear benchmark relationships, and temporal dependencies in firm fundamentals. The framework combines heterogeneous base learners, each designed to extract a distinct component of the ESG–fundamentals relationship, and aggregates their outputs using a regularized meta-learner. This architecture allows the relative contribution of nonlinear, linear, and temporal signals to be determined endogenously when forecasting next-quarter earnings. Figure 1 provides a schematic overview of the data inputs, learning components, and portfolio construction steps underlying the proposed framework.

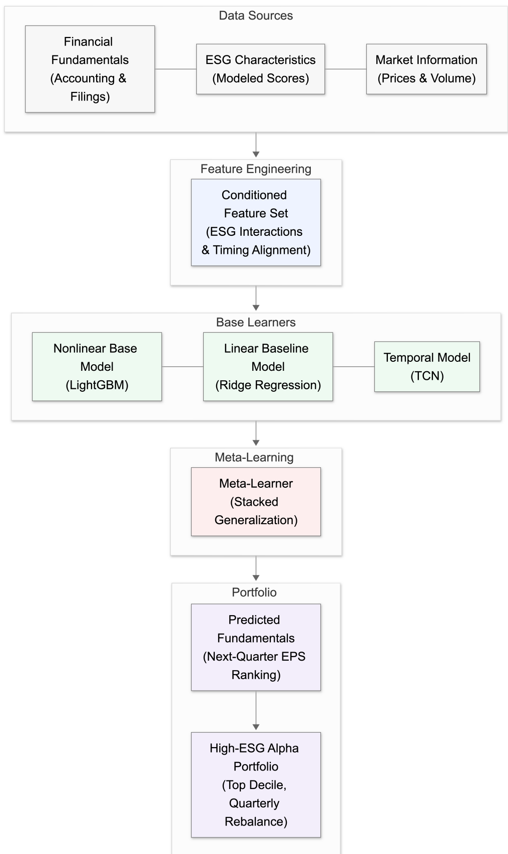

Figure 1. Meta-learning framework for ESG-based fundamental forecasting and portfolio construction.

3.5.1. Base Learner 1: light gradient boosting machine (LightGBM)

We employ the light gradient boosting machine (LightGBM) as the primary nonlinear base learner to model complex and non-monotonic relationships between ESG characteristics and firm-level financial fundamentals. LightGBM is particularly well-suited to this setting due to its ability to efficiently capture interaction effects and threshold behavior that are difficult to identify using linear specifications.

LightGBM constructs decision trees using a leaf-wise (best-first) growth strategy, prioritizing splits that yield the largest reduction in the loss function. In addition, its exclusive feature bundling (EFB) mechanism improves computational efficiency in the presence of correlated and sparse predictors, which are common in high-dimensional ESG and accounting data [32].

Within our framework, LightGBM is designed to extract nonlinear cross-sectional signals and conditional relationships between ESG scores and firm fundamentals. In particular, it captures regime-dependent effects whereby ESG characteristics contribute to earnings predictability only beyond certain thresholds or in conjunction with specific firm attributes, such as size, profitability, or valuation.

3.5.2. Base Learner 2: temporal convolutional network (TCN)

While LightGBM is well-suited for capturing nonlinear cross-sectional relationships, it treats observations from different time periods as conditionally independent. To model persistence, momentum, and regime dynamics in firm fundamentals, we incorporate a temporal convolutional network (TCN) as a sequential base learner.

TCNs employ stacked dilated causal convolutions, which allow the model to extract temporal features from historical firm-level sequences while preserving the chronological ordering of information. A dilated convolution at time index s can be expressed as

$$F(s) = \sum_{i} f(i)\, x(s - d \cdot i), \tag{3.4}$$

where $f(i)$ denotes the convolutional filter, $x(\cdot)$ is the input sequence, and d is the dilation factor.

By increasing the dilation factor across layers, the receptive field of the network expands exponentially with depth, enabling the TCN to capture long-horizon dependencies in earnings, profitability, and growth dynamics without relying on recurrent connections. In our implementation, the TCN operates on an eight-quarter input window of lagged firm-level observations, which provides sufficient temporal context to capture medium-term fundamental dynamics while maintaining a parsimonious architecture.

The causal structure of the convolutions ensures that only information available prior to the prediction date enters the forecasting process, thereby preventing information leakage across time. Importantly, the use of a fixed eight-quarter input window does not induce additional survivorship bias, as firms with shorter observable histories are retained in the cross-sectional sample using available lagged information rather than being excluded based on future survival or data completeness. This property makes TCNs particularly well-suited for financial forecasting tasks where temporal ordering, persistence, and heterogeneous firm lifecycles play a central role [11].

3.5.3. Base Learner 3: ridge regression

We include an ℓ_2-regularized linear regression model as a stabilizing baseline component of the ensemble. This learner captures approximately linear relationships between firm characteristics and future earnings and serves as a benchmark against which more flexible nonlinear models are evaluated. In particular, linear dynamics such as mean reversion and gradual adjustment in fundamentals are well-documented in empirical asset pricing and may dominate during certain market regimes.

The ridge specification improves numerical stability in the presence of multicollinearity among accounting and ESG variables, while mitigating overfitting through coefficient shrinkage. As such, it complements the more flexible tree-based and convolutional learners by providing a disciplined linear anchor within the ensemble, a practice commonly adopted in both ensemble learning and empirical finance applications [33, 34].

3.5.4. Meta-learner fusion

Predictions generated by the three base learners—LightGBM, the temporal convolutional network, and ridge regression—are combined using a second-stage ridge regression meta-learner. This fusion layer estimates optimal combination weights by minimizing out-of-sample mean squared prediction error under cross-validated regularization, consistent with standard stacked generalization and forecast combination frameworks [34, 35].

The use of an ℓ_2-penalized meta-learner constrains extreme weight assignments and stabilizes the aggregation of heterogeneous prediction signals, allowing the ensemble to adaptively balance nonlinear cross-sectional effects, temporal dynamics, and linear benchmark structure. Such regularized forecast combination has been shown to improve robustness and out-of-sample performance in both machine learning and empirical forecasting settings [36, 37]. The resulting consensus forecast of next-quarter earnings is transformed into a cross-sectional ranking signal, which serves as the input for portfolio construction and performance evaluation.

3.6. Portfolio construction and evaluation

We evaluate the economic value of the machine-learning-based earnings signal using a hierarchical double-sort portfolio construction procedure designed to isolate its incremental contribution beyond ESG characteristics. This approach follows standard asset pricing practice for assessing conditional return premia and allows us to distinguish the role of ESG screening from the predictive content of the AI-generated signal [17, 30].

Universe filtration. At the beginning of each quarter t, the full stock universe is sorted by ESG score. Firms in the top tercile (top 33%) are retained to form the High-ESG universe. This threshold follows the best-in-class screening convention established by [38], who documented that maximum abnormal returns in ESG portfolios are achieved when restricting to firms with extreme ESG ratings, and is consistent with the tercile-based portfolio sorting methodology standard in the factor literature [17]. Sensitivity tests using top-20% and top-50% ESG thresholds are reported in Appendix A and confirm that results are robust: annualized alpha ranges narrowly from 36.6% to 42.1% across thresholds, with the base-case top tercile producing the strongest risk-adjusted performance. Importantly, tightening the screen to the top quintile reduces the investable universe to approximately 24 stocks per quarter, constraining the cross-sectional variation available to the ML signal; widening to the top half admits firms with merely above-average ESG scores. The top tercile therefore represents an empirically motivated balance between ESG quality and universe breadth.

Signal ranking. Within the High-ESG universe, firms are ranked in descending order according to the meta-learner's predicted next-quarter earnings signal. This ranking reflects the model's assessment of relative fundamental strength conditional on ESG standing.

Portfolio formation. Based on the signal ranking, we construct three nested, equal-weighted portfolios:

- **Broad High-ESG**: a passive portfolio consisting of all firms in the High-ESG universe;
- **Top 50% Signal**: firms within the High-ESG universe whose predicted earnings exceed the median signal;
- **Top 10% Signal**: a concentrated portfolio comprising the top decile of High-ESG firms ranked by the machine-learning prediction.

All portfolios are rebalanced on a quarterly basis.

Performance evaluation. Portfolio performance is evaluated using geometric annualized returns to account for volatility drag. Risk-adjusted performance is measured using the Sharpe ratio, with the one-year Chinese government bond yield serving as the risk-free rate proxy [18].

Statistical inference. To assess statistical significance, we conduct rolling one-year holding-period comparisons between portfolio returns and the CSI 300 benchmark using paired t-tests. Economic significance is evaluated using Cohen's d, which measures the magnitude of return differentials independent of sample size [39].

4. Empirical results

4.1. Predictive power of the meta-learning ensemble

Before evaluating portfolio performance, we first assess the out-of-sample predictive accuracy of the proposed meta-learning ensemble in forecasting next-quarter earnings per share (EPS). Establishing predictive validity at the fundamental level is important, as return predictability that arises from earnings dynamics is less susceptible to short-term price noise and mechanical momentum effects. This perspective is consistent with [31], which argued that earnings predictability constitutes a primary channel through which long-horizon valuation effects materialize.

Table 4 reports out-of-sample performance metrics for the individual base learners—LightGBM, temporal convolutional network (TCN), and ridge regression—as well as the stacked meta-fusion ensemble over the test period from 2019Q2 to 2021Q3. Comparing standalone learners with the fused prediction allows us to evaluate whether the meta-learning architecture delivers incremental forecasting power beyond any single model component.

Table 4. Out-of-sample predictive performance evaluated on standardized next-quarter EPS targets.

Model	R^2	RMSE	Spearman IC
LightGBM (Nonlinear)	0.558	0.648	0.790
ridge regression (Linear)	0.523	0.673	0.796
Temporal ConvNet (TCN)	0.403	0.752	0.726
Meta-Fusion Ensemble	**0.584**	**0.628**	**0.807**

The results indicate that the meta-fusion ensemble achieves the strongest predictive performance across all evaluation metrics. The meta-learner assigns a stacking weight of 0.763 to LightGBM, 0.280 to the TCN, and 0.074 to ridge regression, reflecting the dominance of nonlinear cross-sectional relationships in earnings prediction. As the RidgeCV meta-learner includes an intercept term, these reported coefficients do not constrain to sum to unity (observed sum: 1.117); this is standard behavior for ridge-regularized stacking estimators. LightGBM feature importance (gain) identifies lagged log-ROE (36,484), lagged log-earnings yield (14,304), and log-firm size (12,258) as the three most influential predictors, with the ESG×size interaction term ranking sixth (1350) among all features. ridge standardized coefficients confirm that ESG×size (0.998) and ESG×earnings yield (0.427) carry the largest positive loadings, while the direct ESG score coefficient is negative (−0.769), indicating that the predictive content of ESG is primarily captured through its interactions with firm fundamentals rather than as a standalone level effect—consistent with the conditional valuation hypothesis motivating these features. Specifically, the ensemble attains an out-of-sample R^2 of 0.584 and a Spearman information coefficient (IC) of 0.807, outperforming each individual base learner on the held-out test set. We note that the IC is computed as the average cross-sectional Spearman correlation between predicted and realized EPS ranks across test quarters; because successive quarterly predictions share overlapping firm histories in the TCN sequences, mild autocorrelation in the IC series should be expected and does not inflate the point estimate. The level of 0.807 is high relative to return-prediction benchmarks but plausible for a fundamental target (EPS-TTM rank): unlike short-horizon price returns, quarterly earnings ranks exhibit substantial persistence from one quarter to the next, providing the model with a learnable signal structure that is absent in noisier price-based tasks. Among the standalone models, LightGBM delivers the strongest performance (R^2 = 0.558, IC = 0.790), underscoring the importance of nonlinear modeling in capturing complex interactions between ESG characteristics and firm fundamentals in the Chinese equity market. This finding is consistent with recent evidence documenting the effectiveness of tree-based machine learning methods in ESG-informed forecasting settings [5].

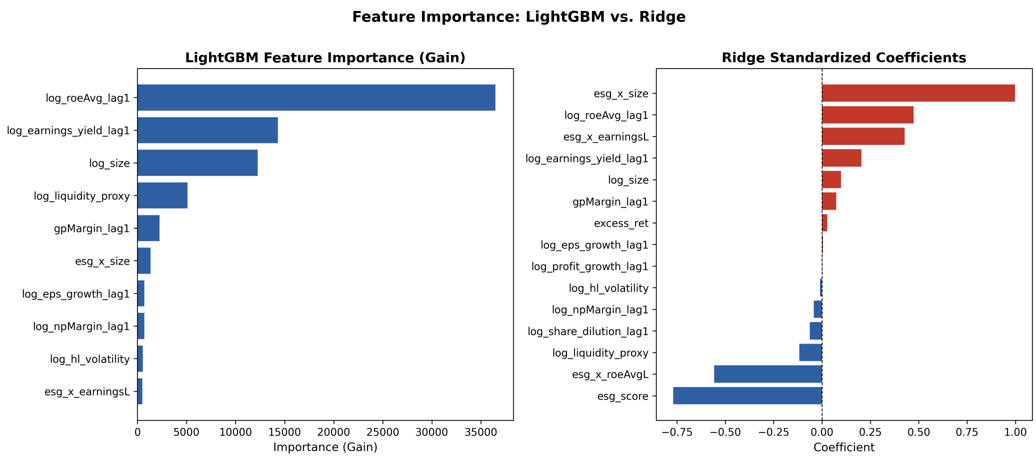

Figure 2. Feature importance and meta-learner stacking weights. Left panel: LightGBM feature importance (gain) for top features. Right panel: ridge standardized coefficients. Meta-learner stacking weights: LightGBM = 0.763, TCN = 0.280, Ridge = 0.074.

The predictive accuracy of the proposed framework compares favorably with results reported in the existing earnings forecasting literature, while acknowledging differences in market structure, accounting regimes, and sample design. For example, [40] document out-of-sample R^2 values of approximately 0.30 in the MENA banking sector. Although direct cross-market comparisons should be interpreted with caution, the relatively strong performance of the meta-learning ensemble in our setting is consistent with the inclusion of ESG interaction features that capture conditional valuation dynamics, as well as with the aggregation of complementary linear, nonlinear, and temporal prediction signals.

4.2. Performance evaluation framework

To rigorously assess the economic value of the model's predictions, we evaluate portfolio performance using realized quarterly returns from the constructed portfolios. Performance is primarily measured using the compound annual growth rate (CAGR), which explicitly accounts for return compounding and captures the effects of volatility drag associated with high-variance strategies. The annualized return is computed as:

$$R_{\text{ann}} = \left(\prod_{t=1}^{T} (1 + r_t) \right)^{\frac{4}{T}} - 1, \tag{4.1}$$

where r_t denotes the portfolio return in quarter t, and T is the total number of quarters in the evaluation period [41].

Risk-adjusted performance is evaluated using the Sharpe ratio, defined as:

$$\text{Sharpe} = \frac{R_{\text{ann}} - R_f}{\sigma_{\text{ann}}}, \tag{4.2}$$

where R_f represents the risk-free rate, proxied by the one-year Chinese government bond yield, and σ_{ann} denotes annualized portfolio volatility [42]. Annualized volatility is obtained by scaling the standard deviation of quarterly returns according to:

$$\sigma_{\text{ann}} = \sigma_q \sqrt{4}, \tag{4.3}$$

where σ_q is the standard deviation of quarterly portfolio returns [41].

Finally, downside risk is assessed using maximum drawdown (MDD), defined as the largest peak-to-trough decline in cumulative portfolio value over the evaluation horizon:

$$\text{MDD} = \max_{t \in [1,T]} \left(\frac{\max_{s \in [1,t]} V_s - V_t}{\max_{s \in [1,t]} V_s} \right), \tag{4.4}$$

where V_t denotes the cumulative portfolio value at time t. Maximum drawdown provides a complementary measure of tail risk that captures the severity of losses during adverse market conditions and is not fully reflected by volatility-based metrics [43].

4.3. The "alpha ladder": Risk and reward trade-offs

To isolate the incremental contribution of the meta-learning signal, we decompose the High-ESG universe into three nested portfolio tiers based on signal strength. Table 5 summarizes the resulting risk–return profiles over the out-of-sample period. All portfolio performance metrics are computed using realized forward returns following quarterly portfolio rebalancing; model predictions are used exclusively for ranking and portfolio formation.

Table 5 reveals a clear monotonic ordering in portfolio characteristics as signal intensity increases, which we term the “alpha ladder.” As the machine-learning signal becomes more selective, both expected returns and volatility rise in tandem, indicating a systematic risk–reward trade-off rather than a defensive ESG effect typically documented in developed markets.

Table 5. The risk–return spectrum of ML-enhanced ESG portfolios over the out-of-sample period (2019Q2–2021Q3).

Portfolio	Ann. Return	Ann. Volatility	Sharpe Ratio
Equal-Weight Sample	−0.097	0.065	−1.89
Benchmark (CSI 300)	0.0706	0.1651	0.35
High ESG (Broad)	0.0372	0.2072	0.15
High ESG + Top 50%	0.1312	0.2594	0.51
High ESG + Top 10%	0.4869	0.4311	1.11

Note: All returns are quarterly excess returns net of the one-year Chinese government bond yield; benchmarks are adjusted to the same basis. The Equal-Weight Sample benchmark is an equal-weight portfolio of all firms in the backtest sample per quarter, equal-weighted across all ESG tiers. Sharpe ratios are computed as $\bar{r}_e/\hat{\sigma}_e \times \sqrt{4}$, where $\bar{r}_e$ and $\hat{\sigma}_e$ are the mean and standard deviation of quarterly excess returns over the evaluation period, and the scaling by $\sqrt{4}$ annualizes assuming independent and identically distributed (i.i.d.) quarterly returns. The risk-free rate applied is the China one-year government bond yield for each observation quarter (average 2.36% per annum (p.a.) over the evaluation period). Annualized returns are computed as geometric means (CAGR): $[(1 + r_1)\cdots(1 + r_T)]^{4/T} - 1$. Annualized volatility uses the arithmetic standard deviation of quarterly returns scaled by $\sqrt{4}$. Sharpe ratios use arithmetic mean quarterly excess returns divided by their standard deviation, scaled by $\sqrt{4}$; this convention is stated explicitly because CAGR-based and arithmetic-mean-based Sharpe ratios differ for volatile return series. For the Top 10% portfolio, Sharpe = $\bar{r}_e/\hat{\sigma}_e \times \sqrt{4} = 11.94\%/21.56\% \times 2 = 1.108 \approx 1.11$, verified against the quarterly return series. Under a symmetric return distribution, the vol-drag approximation ($\mathrm{CAGR}_e + \frac{1}{2}\sigma^2$) would imply an arithmetic mean of approximately $46.3\% + 9.3\% = 55.6\%$ and an expected Sharpe of ≈ 1.29; the realized Sharpe of 1.11 falls below this because the 2021Q3 quarterly return of −32.0% introduces strong negative skewness that pulls the arithmetic mean below what symmetric vol-drag predicts.

Tier 1: Broad High-ESG. All portfolio returns in Table 5 are measured as quarterly excess returns net of the one-year Chinese government bond yield, consistent with the return variable construction described in Section 3.1. The CSI 300 benchmark return is adjusted to the same excess return basis. As shown in Table 5, the equal-weight sample portfolio—an equal-weight portfolio of all firms in the backtest sample, equal-weighted across all ESG tiers each quarter—generates an annualized excess return of −9.7%, confirming that the ESG and ML screens add incremental value beyond the unscreened sample average. The comparatively low annualized volatility of 6.5% for the equal-weight sample portfolio reflects the cross-sectional averaging of individual stock returns across the full universe; individual stock volatilities are substantially higher, but firm-specific risks largely cancel in a broad equal-weight portfolio, resulting in a lower aggregate volatility than concentrated strategies. The passive High-ESG portfolio delivers an annualized return of 3.7%, substantially underperforming the CSI 300 benchmark (excess return: 7.1%). This result suggests that a naïve ESG screen in China does not, by itself, generate excess returns and may be subject to sectoral composition effects and concentration risk.

Tier 2: High-ESG + Top 50%. Restricting the universe to firms with above-median predicted earnings materially improves performance. Annualized returns increase to 13.1%, while volatility rises to 25.9%. The Sharpe ratio (0.51) approaches that of the benchmark, indicating that the machine-learning signal begins to meaningfully enhance risk-adjusted performance at moderate levels of selectivity.

Tier 3: High-ESG + Top 10%. The most selective portfolio exhibits a pronounced shift in performance. Annualized returns increase sharply to 48.7%, accompanied by a substantial rise in volatility to 43.1%. Despite this elevated risk, the Sharpe ratio peaks at 1.11, indicating that excess returns more than compensate for the additional volatility. This pattern suggests that the meta-learning model concentrates exposure in a small subset of high-growth, high-beta firms rather than identifying low-risk ESG "insurance" stocks.

The magnitude of realized returns reflects both signal concentration and exposure to post-crisis recovery dynamics. Accordingly, the analysis emphasizes relative performance patterns and risk–return trade-offs rather than extrapolating long-run return levels.

Figure 3 illustrates the cumulative wealth trajectories of the ESG portfolios relative to the CSI 300 benchmark.

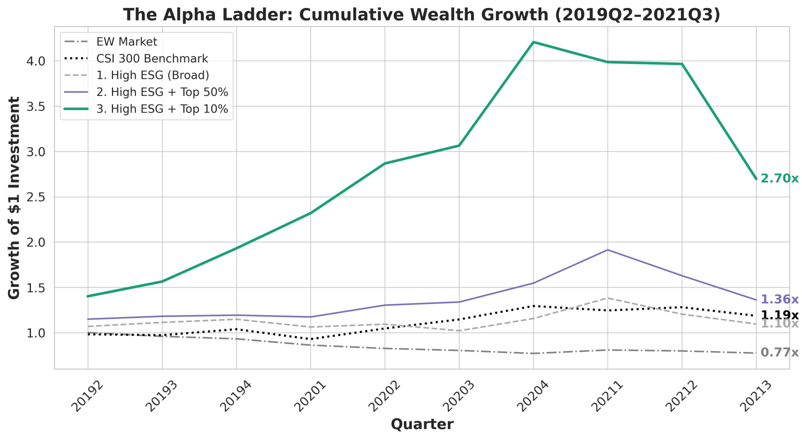

Figure 3. Cumulative wealth growth of ESG portfolios and the benchmark (2019Q2–2021Q3).

While the Broad High-ESG and Top 50% portfolios remain closely tied to the CSI 300 benchmark, the Top 10% strategy exhibits pronounced divergence following the COVID-19 shock, characterized by both rapid appreciation and substantial interim drawdowns. This pattern indicates that the documented outperformance is accompanied by elevated volatility and timing risk, consistent with a high-beta, growth-oriented return profile rather than a defensive ESG allocation, as documented in studies of ESG and post-crisis return dynamics in emerging markets [1, 3]. To summarize these risk–return trade-offs in a compact cross-sectional framework, Figure 4 maps annualized returns against volatility across the three ESG portfolio tiers. Figure 4 highlights a monotonic increase in both returns and volatility as signal strength rises, indicating that alpha associated with ESG-conditioned, machine-learning-based portfolios in China is closely tied to risk exposure. To determine whether this relationship reflects persistent performance rather than episodic outcomes, the following subsection examines statistical robustness using rolling-window and effect-size-based tests.

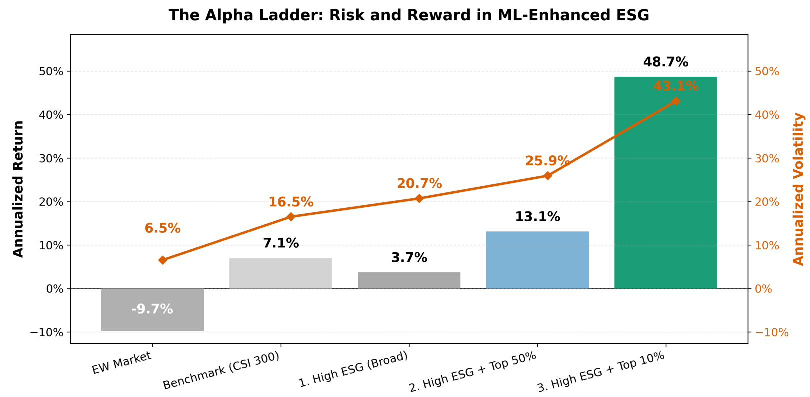

Figure 4. The alpha ladder: The relationship between annualized returns and volatility across ESG signal tiers.

4.4. *Statistical robustness and factor exposure*

Given the elevated volatility of the Top 10% strategy, it is critical to determine whether its outperformance reflects a persistent return premium or idiosyncratic estimation noise. We address this question using two complementary frameworks: rolling-window analysis to assess temporal persistence, and effect-size-based inference to evaluate economic significance. This dual approach allows us to distinguish sustained signal efficacy from transitory market shocks [23].

4.4.1. Statistical robustness

Table 6 reports mean quarterly excess returns relative to the CSI 300 benchmark, rolling one-year t-statistics, and Cohen's d effect sizes for each ESG portfolio tier. The equal-weight sample portfolio exhibits significant negative rolling alpha ($t = -4.70$, $p < 0.01$, Cohen's $d = -0.49$), confirming that passive market exposure does not generate excess returns over the risk-free rate in this window. The Equal-Weight Ensemble (top decile by simple average of base-model predictions, without applying the ESG-tier filter) generates a quarterly alpha of +9.49% ($t = 3.59$, Cohen's $d = 0.51$). The Broad High-ESG and High-ESG + Top 50% portfolios fail to exhibit statistically or economically meaningful excess returns. Their estimated alphas are small in magnitude, statistically insignificant, and associated with negligible effect sizes.

Table 6. Statistical validation of excess returns relative to the CSI 300 benchmark.

Portfolio Tier	Quarterly Alpha (Mean)	Rolling 1-Year t-stat	Cohen's d	Significance
Equal-Weight Sample	−4.50%	−4.70	−0.49	Significant ($p < 0.01$)
Equal-Weight Ensemble	+9.49%	3.59	+0.51	Significant ($p < 0.05$)
Broad High ESG	−0.63%	−0.17	−0.05	Not significant
High ESG + Top 50%	+1.86%	1.80	0.14	Not significant
High ESG + Top 10%	+10.51%	3.16	0.54	Significant ($p < 0.05$)

In contrast, the High-ESG + Top 10% portfolio displays a markedly different profile. Its average quarterly excess return of 10.5% corresponds to a rolling one-year t-statistic of 3.16 ($p < 0.05$) and a Cohen's d of 0.54, indicating a medium-to-strong economic effect size. These results suggest that the documented outperformance is not solely a byproduct of sampling variability but reflects economically meaningful excess returns.

4.4.2. Note on multiple comparisons

The preceding analysis reports significance tests across multiple dimensions—portfolio quantiles, rolling windows, and multi-factor model specifications. With approximately 18 hypothesis tests conducted across all tables and test methods, a Bonferroni-corrected threshold of $\alpha/18 \approx 0.003$ applies. Results significant at $p < 0.01$ remain robust under this correction. Results significant only at $p < 0.05$ should be interpreted with caution, as some may reflect false discovery at the 5% level. We encourage readers to weigh the economic magnitudes of reported alphas alongside their statistical significance.

4.4.3. Rolling-window persistence

To further assess the temporal stability of excess returns, Table 7 reports rolling one-year (four-quarter) excess returns for the High-ESG + Top 10% portfolio relative to the CSI 300 benchmark. Because consecutive rolling windows share three overlapping quarters, the effective number of independent observations is approximately 2–3, and significance of individual window estimates should not be over-interpreted. Excess returns are positive in the majority of rolling windows, indicating that performance is not driven by a small number of extreme observations but instead reflects sustained outperformance over time.

Table 7. Rolling one-year excess returns of the High-ESG + Top 10% portfolio relative to the CSI 300 benchmark.

Quarter	Top 10% Strategy (4Q Return)	CSI 300 (4Q Return)	Rolling Excess Alpha
2020Q1	1.319	-0.071	1.390
2020Q2	1.046	0.066	0.980
2020Q3	0.959	0.178	0.781
2020Q4	1.179	0.246	0.933
2021Q1	0.719	0.340	0.378
2021Q2	0.383	0.224	0.159
2021Q3	-0.120	0.035	−0.155

Although excess returns moderate toward the end of the sample period, this pattern aligns with post-crisis normalization dynamics rather than a clear breakdown of the underlying signal. Prior studies of ESG performance in emerging markets document comparable post-crisis convergence effects [1,44]. Taken together, the rolling-window analysis suggests that the observed excess returns are not solely driven by short-lived market conditions.

4.4.4. Factor exposure: China CAPM

Having established the temporal persistence of excess returns, we next examine their structural drivers using a single-factor capital asset pricing model (CAPM) estimated against the CSI 300 benchmark. The regression employs the one-year Chinese government bond yield as the risk-free rate in order to isolate the idiosyncratic alpha from systematic market exposure. Table 8 summarizes the estimation results for the High-ESG + Top 10% portfolio over the out-of-sample period.

Table 8. China CAPM regression results for the High-ESG + Top 10% portfolio relative to the CSI 300 benchmark.

Metric	Value
Observation Period	2019Q2–2021Q3
Market Benchmark	CSI 300 Index
Annualized Alpha	44.30%
Market Beta (β)	1.16
Alpha t-statistic	1.44
Adjusted R^2	0.096

Note that the CAPM annualized alpha of 44.3% represents the regression intercept after controlling for market beta exposure, and therefore differs from the 42.1% raw excess return above the CSI 300 benchmark reported in Appendix A; both are valid and complementary alpha measures. Given the limited number of quarterly observations in the out-of-sample evaluation period, factor regression results should be interpreted as descriptive evidence of risk exposure rather than as definitive tests of asset pricing efficiency.

4.4.5. Factor exposure: China-specific multi-factor model

To assess whether the documented excess returns reflect compensation for broader systematic risks specific to the Chinese equity market, we extend the baseline CAPM by incorporating the China-specific factor framework of [18]. In addition to the market factor, the model includes size (SMB), value–growth (VMG), and turnover-based (PMO) factors designed to capture institutional and trading characteristics unique to China. The factor return series are obtained from the public data library maintained by Stambaugh and Yuan and hosted by the Wharton School [45]. All factors are aggregated to the quarterly frequency to align with portfolio rebalancing, and inference is based on heteroskedasticity and autocorrelation consistent (HAC) standard errors.

Table 9 reports the estimation results. We caution that with 10 quarterly observations and five regressors, the model has only $df = 5$; individual coefficient significance should be interpreted cautiously. We offer this caveat: While the inclusion of China-specific factors attenuates the estimated alpha relative to the single-factor CAPM, the High-ESG + Top 10% portfolio continues to exhibit economically large excess returns, though the alpha ($p = 0.077$) does not reach statistical significance at conventional thresholds given $df = 5$. Notably, the strategy displays a pronounced negative loading on the size factor, indicating that ESG-driven performance in China is concentrated among large-cap firms rather than small-cap stocks. In contrast, market, value–growth, and turnover factor loadings are not statistically significant, suggesting that the observed excess returns are not driven by broad market movements, traditional value tilts, or speculative trading activity.

Table 9. China-specific multi-factor regression results for the High-ESG + Top 10% portfolio.

	Coefficient	Std. Error	t-stat	p-value
Alpha (const)	0.090	0.051	1.77	0.077
Market (MKT–RF)	−2.612	1.766	−1.48	0.139
Size (SMB)	−11.046	2.634	−4.19	0.000
Value–Growth (VMG)	−1.333	1.214	−1.10	0.272
Turnover (PMO)	−0.246	4.102	−0.06	0.952
Observations		10		
Adjusted R^2		0.404		
Covariance estimator		HAC (Newey–West, lag 1)		

Note: The negative and statistically insignificant market coefficient ($\hat{\beta}_{\mathrm{MKT}} = -2.612$, $p = 0.139$) is not attributable to multicollinearity (VIF = 1.43 for MKT–RF, below 2 for all regressors), but rather reflects the low statistical power of a five-regressor model with only 10 quarterly observations ($df = 5$); the insignificant p-value of 0.139 is consistent with this. The single-factor CAPM beta of +1.16 in Table 8 provides a cleaner estimate of market exposure.

4.5. Robustness check: Industry-neutralized portfolio

A potential concern with the documented alpha is that it may reflect passive industry tilts rather than genuine stock selection ability. We note that the industry-neutral portfolio reports a higher annualized CAGR (52.71%) than the unconstrained Top 10% portfolio (48.69%); this reflects the geometric return drag that accumulates differently across the two portfolios owing to their distinct quarterly return distributions and concentration levels, and does not imply that sector-balancing adds alpha—the annualized alpha figures (41.9% vs. 42.1%) are nearly identical, while the sector-balanced portfolio achieves a higher Sharpe ratio (1.27 vs. 1.11) due to reduced volatility (37.6% vs. 43.1%), confirming that the documented outperformance reflects stock selection rather than passive sector concentration. High-ESG firms in the Chinese A-share market tend to cluster in specific sectors such as information technology, industrials, and materials, which may have systematically outperformed the CSI 300 during the evaluation period. To address this concern, we construct an industry-neutralized portfolio by selecting stocks proportionally across CSRC level-1 sectors within the high-ESG universe, matching the original portfolio size of approximately 32 stocks per quarter. Specifically, allocations across sectors are determined proportionally to each sector's representation in the high-ESG universe, and firms are ranked by the meta-learner signal within each sector. Sectors with fewer than five high-ESG constituents in a given quarter are excluded to ensure meaningful within-sector ranking.

The industry-neutral portfolio generates annualized alpha of 41.9% relative to the excess-return-adjusted CSI 300 benchmark, compared to 42.1% for the unconstrained Top 10% strategy—a reduction of 0.2 percentage points (0.5%). The Sharpe ratio of 1.27 for the industry-neutral portfolio follows the same time-varying quarterly risk-free rate convention applied throughout (see the Table 5 note). The computed Sharpe of 1.27 reflects the use of observation-specific quarterly risk-free rates rather than a fixed average; using a constant risk-free rate approximation yields a marginally different value, consistent with the same Sharpe convention applied across all portfolio rows. A CAPM regression yields an annualized alpha of 39.0% ($t = 2.50$, $p = 0.037$, $\hat{\beta} = 1.94$, ordinary least squares (OLS) with $df = 8$). We report OLS rather than HAC standard errors because with only $N = 10$ observations, the

Newey–West correction is unreliable in small samples; OLS provides the conservative and appropriate estimate. This negligible attenuation indicates that the model's predictive ability operates at the stock level rather than through passive sector concentration. The strategy identifies winners within each sector of the high-ESG universe independently, consistent with genuine cross-sectional fundamental forecasting rather than industry momentum.

4.6. Robustness check: Crisis and recovery analysis

To further characterize the downside risk of the proposed strategy, we examine portfolio performance during the COVID-19 market shock and the subsequent recovery period. This analysis complements the earlier risk–return evidence by explicitly focusing on drawdown dynamics under conditions of extreme market stress.

Figure 5 presents the drawdown profiles of the ESG-sorted portfolios and the CSI 300 benchmark over the sample period. The CSI 300 experiences a relatively contained peak-to-trough drawdown and recovers to near its prior peak by the end of the sample. In contrast, ESG-tilted portfolios exhibit materially deeper drawdowns, reflecting their higher effective exposure to systematic market risk.

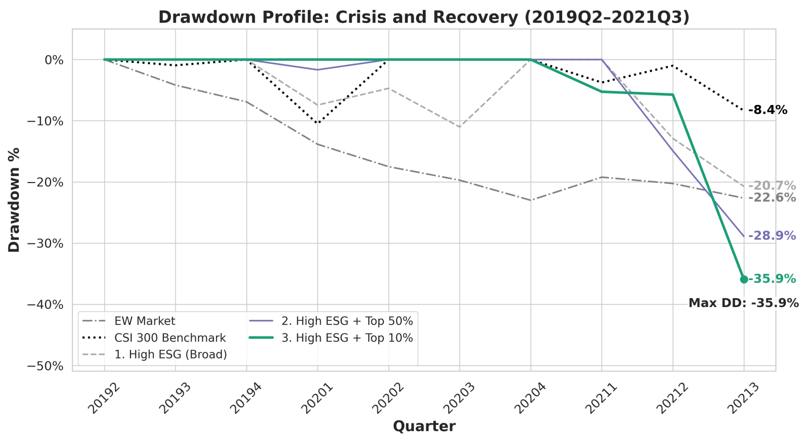

Figure 5. Drawdown profiles of ESG portfolios and the CSI 300 benchmark (2019Q2–2021Q3).

The High-ESG + Top 10% portfolio experiences a pronounced peak-to-trough drawdown during the COVID-19 market shock and ends the sample period with a drawdown of −35.9% relative to its most recent peak, closely tracking the High-ESG + Top 50% portfolio (−28.9%). The Broad High-ESG portfolio performs comparatively better but still exhibits a sizable drawdown of −20.7%, substantially larger than that of the CSI 300 benchmark. These patterns indicate that increasing ESG concentration amplifies downside exposure during periods of market stress rather than providing defensive protection, consistent with evidence that ESG-tilted portfolios in emerging markets may exhibit higher effective market exposure and drawdown risk during crises [1, 43].

Contrary to the "ESG-as-insurance" hypothesis, higher ESG exposure does not mitigate losses during severe market dislocations in the Chinese equity market. Instead, the evidence suggests that ESG-tilted portfolios behave as high-beta growth strategies, absorbing larger drawdowns during downturns while retaining the capacity for rapid recovery when market conditions normalize. The sharp V-shaped drawdown profile of the High-ESG + Top 10% portfolio contrasts with the shallower but more stable trajectory of the CSI 300 benchmark, consistent with the view that high-quality, innovation-driven firms temporarily decouple from index-level dynamics and subsequently reprice aggressively during recoveries [9].

4.7. Practical implementation: Turnover and transaction costs

A common critique of machine-learning-based investment strategies is that the statistical alpha often disappears once transaction costs and portfolio turnover are taken into account, particularly for high-frequency or aggressively rebalanced portfolios [23]. To assess the practical viability of the High-ESG + Top 10% strategy, we therefore examine its realized portfolio turnover and evaluate the performance net of trading frictions. Table 10 summarizes the strategy's quarterly turnover characteristics and its net performance under alternative transaction cost assumptions.

Table 10. Portfolio turnover and net-of-transaction-cost performance of the High-ESG + Top 10% strategy.

Panel A: Quarterly holdings and turnover

(2019Q2 is the initial portfolio formation quarter; no rebalancing costs are incurred.)

Quarter	Number of Holdings	Turnover (%)
2019Q2	29	0.0
2019Q3	29	11.9
2019Q4	31	10.0
2020Q1	32	14.3
2020Q2	34	21.2
2020Q3	34	20.6
2020Q4	31	7.7
2021Q1	31	16.1
2021Q2	32	20.6
2021Q3	35	16.4
Average	–	13.9

Panel B: Net performance under alternative transaction cost assumptions

Cost drag is computed as: $\Delta_{\text{ann}} = \left[\prod_q (1 + r_q - \tau_q \cdot c \cdot 2)\right]^{4/T} - \left[\prod_q (1 + r_q)\right]^{4/T}$, where τ_q is the quarterly portfolio turnover, c is the one-way transaction cost in basis points, and T is the number of quarters. The factor of 2 reflects round-trip costs (buy and sell).

Transaction Cost (1-way)	Annual Cost Drag	Net Annual Return	Net Sharpe Ratio
10 bps (Institutional)	−0.16%	48.5%	1.11
20 bps (Retail)	−0.32%	48.4%	1.10
30 bps (Stress)	−0.48%	48.2%	1.10
50 bps (Liquidity)	−0.80%	47.9%	1.09

Table 10 reports the turnover characteristics and net-of-transaction-cost performance of the strategy. The High-ESG + Top 10% portfolio exhibits an average quarterly turnover of 13.9%, which is low relative to typical quantitative strategies. This stability suggests that the meta-learning model identifies persistent fundamental signals rather than transient price fluctuations, consistent with evidence that low-turnover strategies are more likely to reflect durable information rather than short-lived market inefficiencies [23].

Applying a baseline transaction cost assumption of 20 basis points per trade—reflecting prevailing brokerage fees and stamp duties in the Chinese A-share market—the implied annualized cost drag remains modest at approximately 32 basis points. Under this assumption, the strategy continues to deliver a strong annualized net return of 48.4% and a net Sharpe ratio of 1.10. These results indicate that the documented performance is not mechanically driven by excessive trading activity.

To assess sensitivity to trading frictions, we further consider a conservative stress scenario assuming transaction costs of 50 basis points per trade, intended to capture adverse liquidity conditions, heightened slippage, or market impact in less liquid constituents. Even under this severe assumption, the strategy sustains an annualized net return of 47.9% and a net Sharpe ratio of 1.09. The limited deterioration in risk-adjusted performance across cost regimes suggests that portfolio turnover is not the primary driver of returns.

Taken together, these findings reinforce the interpretation advanced in the alpha ladder analysis: higher ESG signal strength in Chinese equities is associated with economically meaningful excess returns that arise from increased exposure to systematic and idiosyncratic risk, rather than from fragile, high-frequency trading effects. The persistence of strong net performance after accounting for realistic transaction costs supports the practical implementability of the strategy for real-world investors.

5. Conclusions

5.1. Summary of findings

This study challenges the prevailing view of ESG as a predominantly defensive or low-risk factor in emerging equity markets. By applying a meta-learning ensemble to predict firm-level fundamentals (EPS TTM) within China's A-share universe, we show that ESG-related outperformance is highly conditional on signal extraction and portfolio selectivity. Passive Broad High-ESG strategies fail to generate excess returns and, in some cases, underperform the market. However, when ESG characteristics are combined with nonlinear earnings forecasts, progressively stronger performance emerges as signal intensity increases.

Placing these results in the context of prior literature highlights the importance of active screening. While limited downside protection from high-ESG firms in China has been documented [1], and modest average premia have been reported using linear specifications [4], our findings suggest that unconditional ESG effects are diluted by heterogeneity in firm quality. The intermediate High-ESG + Top 50% portfolio illustrates this point clearly: restricting the universe to firms with above-median predicted earnings improves returns relative to the broad ESG screen but does not, on its own, deliver a statistically robust alpha. This pattern indicates that moderate ESG tilts capture some improvement in expected performance, yet remain insufficient to overcome market-level risk exposure.

The most pronounced results arise only under strong selectivity. The most selective machine-learning-conditioned portfolio within the high-ESG universe exhibits substantial outperformance, with an annualized alpha of 44.3%, reflecting the model's ability to concentrate exposure in a small subset of firms for which ESG information aligns with strong operational efficiency and earnings momentum.

Importantly, the monotonic progression from the Broad ESG to the Top 50% and Top 10% portfolios supports the interpretation that ESG characteristics act as a noisy informational input whose economic content becomes meaningful only after aggressive, learning-based filtering.

Crucially, the resulting excess returns are accompanied by elevated systematic risk. The Top 10% portfolio exhibits a market beta of 1.16 relative to the CSI 300 benchmark, while volatility rises sharply along the signal ladder. This evidence directly contradicts the notion that ESG inherently reduces risk in emerging markets. Instead, it supports a risk-signaling interpretation, whereby high ESG scores in developing economies tend to identify firms operating in capital-intensive and innovation-driven sectors characterized by higher operating leverage and greater uncertainty [3]. For investors, this implies a clear trade-off: ESG-based strategies in China do not offer defensive insulation but instead reward those willing to tolerate higher volatility and drawdowns with economically meaningful excess returns. The out-of-sample evaluation period is constrained by the availability of firm-level ESG data for Chinese equities, which limits the analysis to the 2019Q2–2021Q3 period; the results should therefore be interpreted within this data window. Several additional limitations merit acknowledgment. First, the out-of-sample evaluation spans approximately ten quarters, which includes the COVID-19 market disruption and the subsequent recovery. This specific macro regime—characterized by sharp drawdowns followed by rapid appreciation in growth-oriented equities—may not generalize to sustained bear markets or liquidity crises. Second, the relatively short test window increases the risk that observed results partly reflect model overfit to the training environment, despite our use of a strict chronological split. Future work should validate the framework over longer horizons and across markets with comparable ESG reporting standards, such as Taiwan or South Korea, to assess the robustness of the ESG-fundamental interaction beyond the current sample.

5.2. Theoretical implications

Our findings contribute to the asset-pricing literature by highlighting the conditional nature of ESG-related return premia in the Chinese equity market. In contrast to developed-market evidence that emphasizes ESG as a proxy for firm "quality" or downside risk mitigation, high ESG scores in China are more closely associated with exposure to growth-oriented systematic risk. The documented "ESG alpha ladder"—spanning a passive high-ESG screen, an intermediate top 50% filter, and a concentrated top 10% portfolio—illustrates that economically meaningful excess returns emerge only when ESG characteristics are combined with strong fundamental growth signals, consistent with a conditional asset-pricing interpretation.

This pattern aligns with the view that ESG information interacts with firm fundamentals rather than constituting an independent risk factor. While prior studies in developed markets often interpret ESG premia through a risk-mitigation or preference-based channel, our evidence supports a risk-signaling mechanism in an emerging-market setting, consistent with [3] and [1]. In this framework, ESG scores in China proxy for exposure to innovation intensity, capital deepening, and regulatory alignment, all of which are associated with higher systematic risk and return volatility.

More broadly, the results highlight the importance of shifting the machine-learning objective from direct return prediction toward fundamental forecasting. By training the ensemble to predict next-quarter EPS TTM rather than returns, the model isolates variation in prices driven by intrinsic value creation instead of short-term market noise. From a theoretical perspective, this suggests that the "ESG premium" observed in China reflects a mispriced component of future earnings growth that is difficult to recover using linear factor models, rather than a standalone priced risk factor analogous to size or value [23, 30]. The findings therefore support recent calls for integrating machine learning and fundamental signals into conditional asset-pricing models, particularly in markets characterized by heterogeneous information quality and rapid structural change.

5.3. Managerial and policy implications

For market participants, the results indicate that ESG-based strategies in China should be viewed primarily as growth-oriented rather than defensive allocations. Passive ESG exchange-traded funds are unlikely to outperform the CSI 300 benchmark and may underperform during growth-led market expansions. In contrast, active approaches that integrate nonlinear modeling techniques—particularly those targeting firm-level earnings dynamics—appear better positioned to extract economically meaningful excess returns. At the same time, the elevated market beta of the High-ESG + Top 10% strategy ($\beta = 1.16$) implies that such portfolios amplify market downturns rather than providing downside protection. Asset allocators should therefore treat ESG-enhanced strategies as complements to, rather than substitutes for, defensive holdings and explicitly manage their contribution to overall portfolio risk.

From a policy perspective, the findings suggest that informational efficiency in the Chinese equity market remains incomplete. The ability of machine learning models to extract substantial excess returns from publicly available ESG and financial disclosures indicates that prices do not yet fully incorporate fundamental information. This observation is consistent with evidence that disclosure quality and ESG reporting standards in China are still evolving [2]. Continued improvements in data transparency, standardization, and enforcement may reduce the scope for mispricing documented in this study over time. As ESG information becomes more reliable and more rapidly impounded into prices, excess returns are likely to compress, leading to a more efficient allocation of capital toward sustainable, high-growth firms that support long-run structural transformation of the economy.

Author contributions

Maurice Kyla Octaviano: Conceptualization, Methodology, Formal analysis, Writing original draft; Jin-Taek Seong: Conceptualization, Writing review and editing. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2023-00242528).

Conflict of interest

The authors declare no conflicts of interest.

References

1. D. C. Broadstock, K. Chan, L. T. W. Cheng, X. Wang, The role of ESG performance during times of financial crisis: Evidence from COVID-19 in china, *Finance Res. Lett.*, **38** (2021), 101716. https://doi.org/10.1016/j.frl.2020.101716
2. S. Jin, R. Xiong, H. Peng, S. Tang, Esg performance and private enterprise resilience: Evidence from chinese financial markets, *Int. Rev. Financ. Anal.*, **98** (2025), 103884. https://doi.org/10.1016/j.irfa.2024.103884
3. A. S. Garcia, W. Mendes-Da-Silva, R. J. Orsato, Sensitive industries produce better esg performance: Evidence from emerging markets, *J. Cleaner Product.*, **150** (2017), 135–147. https://doi.org/10.1016/j.jclepro.2017.02.180
4. S. Chen, X. Han, Z. Zhang, X. Zhao, Esg investment in china: Doing well by doing good, *Pacific-Basin Finance J.*, **77** (2023), 101907. https://doi.org/10.1016/j.pacfin.2022.101907
5. X. Gong, F. Xie, Z. Zhou, C. Zhang, The enhanced benefits of esg in portfolios: A multi-factor model perspective based on lightgbm, *Pacific-Basin Finance J.*, **85** (2024), 102365. https://doi.org/10.1016/j.pacfin.2024.102365
6. Ľ. Pástor, R. F. Stambaugh, L. A. Taylor, Sustainable investing in equilibrium, *J. Financ. Econ.*, **142** (2021), 550–571. https://doi.org/10.1016/j.jfineco.2020.12.011
7. C. Shrestha, P. Andrikopoulos, J. S. Park, Esg rating and financial performance in the emerging markets: The moderating effects of cross-listing and industry type, *Res. Int. Business Finance*, **77** (2025), 102916. https://doi.org/10.1016/j.ribaf.2025.102916
8. Z. Shang, B. Yu, K. S. Lam, Does esg explain stock returns? evidence from chinese stock markets, *Finance Res. Lett.*, **79** (2025), 107214. https://doi.org/10.1016/j.frl.2025.107214
9. L. Ruan, J. Li, S. Huang, News or noise? esg disclosure and stock price synchronicity, *Int. Rev. Financ. Anal.*, **95** (2024), 103483. https://doi.org/10.1016/j.irfa.2024.103483
10. R. Albuquerque, Y. Koskinen, C. Zhang, Corporate social responsibility and firm risk: Theory and empirical evidence, *Manag. Sci.*, **65** (2019), 4451–4469. https://doi.org/10.1287/mnsc.2018.3043
11. S. Bai, J. Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, arXiv preprint arXiv:1803.01271, 2018. https://doi.org/10.48550/arXiv.1803.01271
12. T. Afzal, M. A. Afridi, M. N. Jan, Integrating LSTM with Fama-French six factor model for predicting portfolio returns: Evidence from Shenzhen stock market China, *Data Sci. Finance Econ.*, **5** (2025), 177–204. https://doi.org/10.3934/DSFE.2025009
13. A. D. Kayit, M. T. Ismail, Leveraging hybrid ensemble models in stock market prediction: A data-driven approach, *Data Sci. Finance Econ.*, **5** (2025), 355–386. https://doi.org/10.3934/DSFE.2025015

14. Z. Li, Y. Xu, Z. Du, Valuing financial data: The case of analyst forecasts, *Finance Res. Lett.*, **75** (2025), 106847. https://doi.org/10.1016/j.frl.2025.106847

15. L. K. David, ESG scoring and forecasting in china, data set, 2023.

16. L. K. David, J. Wang, V. Angel, N. Amjad, Esg scoring and forecasting in china: advancing sustainable business with multidimensional modeling, *J. Asia Pacific Econ.*, **30** (2025), 128–157. https://doi.org/10.1080/13547860.2023.2273017

17. E. F. Fama, K. R. French, Common risk factors in the returns on stocks and bonds, *J. Financ. Econ.*, **33** (1993), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

18. J. Liu, R. F. Stambaugh, Y. Yuan, Size and value in china, *J. Financ. Econ.*, **134** (2019), 48–69. https://doi.org/10.1016/j.jfineco.2019.03.008

19. *China 1-year government bond yield*, Investing.com, 2026. Available from: `https://www.investing.com/rates-bonds/china-1-year-bond-yield`.

20. R. G. Sloan, Do stock prices fully reflect information in accruals and cash flows about future earnings? *Account. Rev.*, **71** (1996), 289–315.

21. S. H. Penman, X. J. Zhang, Accounting conservatism, the quality of earnings, and stock returns, *Account. Rev.*, **77** (2002), 237–264.

22. E. F. Fama, K. R. French, The cross-section of expected stock returns, *J. Finance*, **47** (1992), 427–465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

23. S. Gu, B. Kelly, D. Xiu, Empirical asset pricing via machine learning, *Rev. Financ. Stud.*, **33** (2020), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

24. S. P. Kothari, A. J. Leone, C. E. Wasley, Performance matched discretionary accrual measures, *J. Account. Econ.*, **39** (2005), 163–197. https://doi.org/10.1016/j.jacceco.2004.11.002

25. T. A. Gormley, D. A. Matsa, Common errors: How to (and not to) control for unobserved heterogeneity, *Rev. Financ. Stud.*, **27** (2014), 617–661. https://doi.org/10.1093/rfs/hht047

26. J. M. Wooldridge, *Econometric Analysis of Cross Section and Panel Data*, 2 Eds, Cambridge: MIT Press, 2010.

27. D. N. Gujarati, D. C. Porter, *Basic Econometrics*, 5 Eds, New York: McGraw-Hill, 2009.

28. B. T. Kelly, S. Pruitt, Y. Su, Characteristics are covariances: A unified model of risk and return, *J. Financ. Econ.*, **134** (2019), 501–524. https://doi.org/10.1016/j.jfineco.2019.05.001

29. N. Jegadeesh, S. Titman, Returns to buying winners and selling losers: Implications for stock market efficiency, *J. Finance*, **48** (1993), 65–91.

30. K. Hou, C. Xue, L. Zhang, Digesting anomalies: An investment approach, *Rev. Financ. Stud.*, **28** (2015), 650–705. https://doi.org/10.1093/rfs/hhu068

31. N. Kim-Duc, P. K. Nam, Earnings growth rates in business valuation models: The impossible quaternity, *Global Finance J.*, **60** (2024), 100930. https://doi.org/10.1016/j.gfj.2024.100930

32. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, et al., LightGBM: A highly efficient gradient boosting decision tree, *Adv. Neural Inform. Proc. Syst.*, **30** (2017), 3146–3154.

33. A. E. Hoerl, R. W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics*, **12** (1970), 55–67. https://doi.org/10.1080/00401706.1970.10488634

34. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, 2 Eds, Berlin: Springer, 2009.

35. D. H. Wolpert, Stacked generalization, *Neural Netw.*, **5** (1992), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1

36. J. M. Bates, C. W. J. Granger, The combination of forecasts, *J. Oper. Res. Soc.*, **20** (1969), 451–468. https://doi.org/10.1057/jors.1969.103

37. F. X. Diebold, R. S. Mariano, Comparing predictive accuracy, *J. Busin. Econ. Stat.*, **13** (1995), 253–263. https://doi.org/10.1198/073500102753410444

38. A. Kempf, P. Osthoff, The effect of socially responsible investing on portfolio performance, *Eur. Financ. Manag.*, **13** (2007), 908–922. https://doi.org/10.1111/j.1468-036X.2007.00402.x

39. J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2 Eds, Mahwah: Lawrence Erlbaum Associates, 1988.

40. E. A. H. Elamir, Modeling and predicting earnings per share via regression tree approaches, *Invest. Manage. Financ. Innov.*, **17** (2020), 51–68. https://orcid.org/0000-0002-9430-072X

41. J. Y. Campbell, A. W. Lo, A. C. MacKinlay, *The Econometrics of Financial Markets*, Princeton: Princeton University Press, 1997.

42. W. F. Sharpe, Mutual fund performance, *J. Bus.*, **39** (1966), 119–138.

43. A. Chekhlov, S. Uryasev, M. Zabarankin, Drawdown measure in portfolio optimization, *Int. J. Theor. Appl. Finance*, **8** (2005), 13–58. https://doi.org/10.1142/S0219024905002767

44. K. V. Lins, H. Servaes, A. Tamayo, Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis, *J. Finance*, **72** (2017), 1785–1824. https://doi.org/10.1111/jofi.12505

45. R. F. Stambaugh, Y. Yuan, *China factor data*, Wharton School, University of Pennsylvania, 2021. Available from:
`https://finance.wharton.upenn.edu/~stambaug/`.

Appendix A: ESG threshold sensitivity

Table 11 reports the annualized return, volatility, Sharpe ratio, and alpha for the Top 10% signal portfolio across three ESG universe thresholds: top 50%, top 33% (base case), and top 20%. In all cases, the ML signal rank is computed globally across the full stock universe before filtering to the ESG screen, ensuring portfolio sizes remain comparable ($\approx$ 32 stocks per quarter). Annualized alpha ranges from 36.6% (top 50% ESG) to 42.1% (top 33% ESG, base case) to 36.8% (top 20% ESG), a range of 5.5 percentage points, confirming that the documented outperformance is not sensitive to the specific choice of the top tercile cutoff. The pattern is consistent with the best-in-class screening framework of [38]: a moderate ESG quality filter maximizes the ML signal's cross-sectional breadth while maintaining meaningful ESG standards.

Table 11. ESG threshold sensitivity: Top 10% signal portfolio across alternative ESG universe cutoffs.

ESG Universe	Ann. Return	Ann. Volatility	Sharpe Ratio	Ann. Alpha
Top 50% ESG	42.6%	40.2%	1.05	36.6%
Top 33% ESG (Base Case)	48.7%	43.1%	1.11	42.1%
Top 20% ESG	41.8%	45.2%	0.94	36.8%

Note: All returns are excess returns net of the one-year risk-free rate. ML signal was ranked globally across the full universe before ESG filtering.