



Climate Risk in ESG Portfolios: Evidence from Indonesian Equity Markets

Annisa Syalsabila*, Nur Ikhwana, Lalu Ramzy Rahmanda

*Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Makassar,
Makassar 90222, Indonesia*

Abstract

We examine climate variables in a pipeline for dynamic Value-at-Risk (VaR) of the SRI-KEHATI ESG index and the IHSG (Jakarta Composite Index). Using 1,499 daily observations (2014–2025), we combine the Least Absolute Shrinkage and Selection Operator (LASSO) over 50 candidate predictors, a quantile regression neural network (QRNN) for VaR at three quantile levels, and SHAP for attribution. LASSO removes all climate variables from both indices. For IHSG, the partial correlation of clean-energy returns falls from 0.121 to -0.0005 once domestic equity factors enter, consistent with domestic absorption, though constituent overlap cannot exclude a purely compositional explanation. A diagnostic quantile regression shows climate contributions rising 2.82-fold from VaR 90% to VaR 99%, but the pattern is flat under the QRNN itself (ratio 1.03). In the main backtest, QRNN has lower pinball loss than Historical Simulation and GARCH(1,1)- t in all 12 Diebold–Mariano comparisons ($p < 0.001$), although it uses same-day predictors the benchmarks lack and the advantage was not reproduced in the ablation. In a common-sample ablation ($n = 249$ of 300 nominal test days), LASSO pre-selection is associated with 4 of 6 Kupiec passes versus 0 of 6 without it, but IHSG still fails at two of three quantile levels. Whether SRI-KEHATI screening isolates climate risk was not tested; mediation covers IHSG only.

Keywords: climate risk; dynamic VaR; ESG; LASSO; quantile regression neural network

Copyright © 2026 by Authors, Published by CAUCHY Group. This is an open access article under the CC BY-SA License (<https://creativecommons.org/licenses/by-sa/4.0>)

1. Introduction

Between 2013 and 2025, Indonesia experienced several climate-related disasters with substantial economic cost: the 2019 forest and land fires cost about US\$5.2 billion, or 0.5% of GDP, according to the World Bank [1], a loss that BNPB, citing World Bank data, reported as Rp 75 trillion [2], the 1 January 2020 Jabodetabek floods had killed 67 people as of 6 January 2020 [3], with business losses—chiefly from closed retail and commercial activity—estimated at nearly Rp 1 trillion [4], and the November 2025 Sumatera floods caused losses estimated by CELIOS at Rp 68.67 trillion [5]. The Indonesia Stock Exchange (IDX) listed more than 900 companies by 2025. The SRI-KEHATI index, managed by the Indonesian Biodiversity Foundation (*Yayasan Keanekaragaman Hayati Indonesia*), is Indonesia’s primary Environmental, Social, and Governance (ESG) investment proxy. The index selects 25 issuers from the IDX through a multi-stage screening process [6]: exclusion of controversial sectors, a financial and

*Corresponding author. E-mail: annisa.syalsabila@unm.ac.id

liquidity filter (market capitalization, total assets, net income, free float, and average trading value; an earlier secondary description reports thresholds of Rp 1 trillion for market capitalization and total assets, a free-float ratio above 10%, and a positive price-earnings ratio [7]), followed by ESG scoring across environmental, social, and governance aspects (for example, energy use, greenhouse-gas emissions, labor practices, board competence, and business ethics). The index is reviewed semi-annually (May and November), providing a dynamically maintained ESG portfolio with a continuous history from 2009.

SRI-KEHATI is the longest-running of IDX's ESG indices. The other three IDX-endorsed ESG indices (IDX ESG Leaders, December 2020; ESG Sector Leaders IDX KEHATI and ESG Quality 45 IDX KEHATI, both December 2021) did not exist for most of our January 2013 study period. SRI-KEHATI is therefore the only one for which our full sample can be constructed; we do not claim it is the most representative ESG index in the market today.

Its 25-constituent structure (versus 900+ in the IHSG, the Jakarta Composite Index) implies higher idiosyncratic concentration, which we control for through the 22-day realized volatility unpenalized term in the LASSO specification. ESG screening should theoretically reduce climate risk exposure. The empirical question has not been answered: does ESG screening protect against climate risk in the distributional tail, and if so, through what mechanism?

Climate risk enters financial markets through two channels. Physical climate risk operates through asset damage, supply chain disruption, and productivity loss triggered by extreme weather events; these shocks are acute, event-driven, and visible at daily frequency [8]. Transition climate risk operates through policy shifts, technology displacement, and stranded-asset repricing; these shocks are slower-moving, typically detected at monthly or quarterly resolution [9, 10].

In emerging markets, neither channel is separable from the broader market. Domestic equity prices, driven by institutional herding and shallow liquidity, incorporate information across all channels simultaneously before individual risk factors can be identified. This simultaneity points to a possible *mediation problem*: if a climate shock is transmitted through domestic equity prices before it can be measured as a distinct climate signal, variable selection may remove climate variables not because climate risk is economically irrelevant, but because the domestic equity channel absorbs its marginal contribution. Bolton et al. [8] characterize this as a Green Swan property of climate risk: tail risks that are highly probable over long horizons but difficult to price under conventional distributional assumptions because they are continuously embedded in market dynamics.

We examine both sides of this problem, although the partial-correlation and SHAP diagnostics are available for IHSG only: whether ESG screening isolates portfolio returns from the mediation effect (addressed for SRI-KEHATI only through LASSO variable selection), and how much tail risk remains attributable to climate factors once domestic channels are removed. We address both questions through a three-stage pipeline: (1) the Least Absolute Shrinkage and Selection Operator (LASSO) for dimensionality reduction from 50 candidate predictors, (2) a Quantile Regression Neural Network (QRNN) for dynamic VaR estimation, and (3) SHapley Additive exPlanations (SHAP) for variable contribution attribution. We analyze 1,499 daily observations (September 2014–December 2025) for SRI-KEHATI and IHSG as the broad-market comparator.

The results run in three directions. LASSO eliminates all climate variables, and the evidence is consistent with domestic equity factors absorbing their predictive signal: the clean-energy return correlation with IHSG collapses from 0.121 to -0.0005 in partial correlation. Climate contributions are tail-concentrated under a diagnostic quantile regression using all eight climate candidates (rising 2.82-fold from VaR 90% to VaR 99%), though comparatively flat and non-monotonic under the QRNN framework itself, which forces in only three climate variables (ratio 1.03). QRNN outperforms conventional models in all 12 Diebold–Mariano comparisons in the main backtest, though not in the ablation run (Subsection 3.5). The mediation evidence is consistent with Green Swan theory [8], but it rests on IHSG only and was not tested for SRI-KEHATI; we are not aware of prior work that examines it at daily resolution for Indonesia.

This study makes four contributions. *First*, it provides quantitative evidence, for IHSG, that is consistent with domestic equity factors absorbing climate signals at daily resolution, extending the monthly-frequency findings of Faccini, Matin, and Skiadopoulos [10] to a temporal resolution where physical climate shocks are most observable. *Second*, it applies the LASSO-QRNN-SHAP pipeline as a framework for simultaneous variable selection, dynamic VaR estimation, and variable attribution in high-dimensional daily financial data, with backtesting validation across six index-quantile combinations. *Third*, it reports that climate contribution is tail-concentrated under a diagnostic quantile regression ($2.82\times$ higher at VaR 99% than VaR 90%, using all eight climate variables) but comparatively flat under the QRNN itself (ratio 1.03, using three forced climate variables). This gives a qualitative, not threshold-specific, basis for the choice of VaR confidence level in climate stress testing. *Fourth*, it makes explicit that the framework is end-of-day conditional risk measurement rather than predictive forecasting, which limits how the VaR comparisons should be read.

1.1. ESG Investing and Tail Risk

The literature on ESG portfolio performance is extensive but inconclusive at the mean-return level [11]. At the tail risk level, evidence is more limited. Engle et al. [12] show that ESG factors hedge against negative climate news in U.S. equity markets, though hedge ratios are time-varying and small in magnitude. Lashkaripour [13] documents that tail risk in high-ESG stocks rises 14% more than in low-ESG stocks during the COVID-19 market crash. For Indonesian markets, Yasya et al. [14] compare SRI-KEHATI with LQ45 on risk and return (2017–2021) and find no significant difference, but do not examine climate risk exposure at the distributional tail; we extend this line to tail risk.

1.2. Climate Risk in Financial Markets

Bolton et al. [8] characterize climate risk as Green Swan events: tail risks highly probable in the long run but difficult to price under conventional distributional assumptions. Barnett, Brock, and Hansen [9] show that climate uncertainty commands a risk premium concentrated in the tail of the asset return distribution. Faccini, Matin, and Skiadopoulos [10] find U.S. equity markets respond significantly to climate news shocks at monthly frequency. Caldara and Iacoviello [15] construct the geopolitical risk (GPR) index; its daily series, published on the authors' website [16], enters our predictor set as a geopolitical-risk proxy rather than a direct climate measure. We are not aware of a published study that prices climate risk in an ASEAN ESG portfolio at daily frequency.

Prior literature on climate risk in equity markets predominantly operates at monthly frequency [10, 12]. Monthly aggregation may systematically miss acute physical climate shocks (floods, wildfires, extreme heat events) whose market impact is concentrated within a small number of trading days following the event. Distinguishing episodic physical climate shocks from the persistent repricing associated with transition risk therefore requires daily-resolution data as a matter of theoretical design rather than data availability. Indonesia's exposure to frequent high-impact weather events makes this distinction empirically meaningful: if climate risk is episodic at daily scale, monthly models understate peak tail exposure by averaging shocks across the calendar month in which they occur.

1.3. Dynamic VaR Methods and Machine Learning

VaR estimation has shifted from parametric approaches toward semiparametric and nonparametric methods as empirical evidence of fat tails has grown. QRNN, first proposed by Taylor [17], estimates conditional quantiles without distributional assumptions by optimizing the pinball loss function. Liu and Luger [18] propose a quantile-based model of scale dynamics for Value-at-Risk and Expected Shortfall forecasting. For high-dimensional variable selection, LASSO [19] is now widely used. Belloni and Chernozhukov [20] derive convergence rates for ℓ_1 -penalized quantile

regression and its post-selection refit in high-dimensional sparse models. Lundberg and Lee [21] propose SHAP for decomposing individual predictions into variable-level contributions using cooperative game theory. Awijen, Ben Jabeur, and Pillot [22] apply SHAP to decompose ESG factor contributions to stock prices with climate control variables.

2. Methods

This section describes the data and the three-stage pipeline used to estimate dynamic Value-at-Risk (VaR) for the SRI-KEHATI index and the IHSG. The first subsection documents the sample, the 50 candidate predictors, and the data-quality corrections applied before estimation. The following subsections then present, in order, LASSO variable selection, the Quantile Regression Neural Network (QRNN) that produces VaR at three quantile levels, and the SHAP attribution and backtesting procedures used to interpret and evaluate the resulting forecasts. Each stage takes the output of the previous one as its input, so the subsections are best read as a single sequence.

2.1. Data

The panel covers 1,499 trading days from September 2014 to December 2025, effective after listwise deletion from 3,146 raw observations spanning January 2013 to December 2025. Two target variables are the log daily returns of SRI-KEHATI and IHSG. Table 1 summarizes 50 predictor variables across seven categories covering global equity markets, commodities, exchange rates, domestic macro, domestic equity markets, internal index factors, and climate and ESG indicators. Two data quality corrections are documented explicitly. First, the BI Policy Rate was excluded from the predictor set due to frequency mismatch: as a monthly-determined variable, its daily forward-filled values lack within-month variation and are theoretically inappropriate as predictors in a daily return model. Second, CPO futures trade on Bursa Malaysia Derivatives, which observes different public holidays from IDX; this 141-day calendar mismatch generated artificial return spikes (SD 3.46% before correction), resolved by forward-filling CPO closing prices before computing log returns, reducing SD to 1.46%.

Table 1: Variable categories and counts.

Category	N	Representative Variables
Global Markets	10	S&P 500, DJIA, Nasdaq, MSCI EM, Nikkei, Hang Seng, FTSE 100, DAX, VIX, Shanghai
Commodities	10	WTI, Brent, CPO, coal, gold, silver, copper, nickel, tin, aluminum
Exchange Rates	4	USD/IDR, EUR/IDR, JPY/IDR, CNY/IDR
Domestic Macro	5	SBN 10Y & 5Y yields, UST 10Y, CDS Indonesia 5Y, yield spread 10Y–5Y
Domestic Markets	7	LQ45, JK Finance, ISSI, JK Mining, MSCI Indonesia, ESG Global, Bitcoin
Internal Index Factors	6	Lagged return, daily volume, 22-day realized volatility (SRI-KEHATI and IHSG, 3 each)
Climate & ESG	8	Physical Risk Index, Transition Risk Index, GPR Daily [15, 16], ERA5 temperature anomaly, CHIRPS rainfall, carbon EUA return, natural gas return, clean energy equity return

All predictor variables enter the model as end-of-day conditioning information available at time t , consistent with the end-of-day conditional risk measurement framework [23]. After listwise deletion, the effective estimation sample contains 1,499 observations, split 80/20 chronologically: 1,199 training observations through October 2023 and 300 test observations from November 2023 through December 2025. We chose listwise deletion over imputation because the gaps are structural rather than random: CPO futures (the 141-day calendar mismatch above), Bitcoin (pre-2016 availability), and aluminum futures (late-starting coverage). Imputing daily returns

by smoothing would introduce artificial serial correlation into the left tail, precisely where VaR accuracy is evaluated. Excluding these three variables would recover a much larger sample; we leave full backtesting of that specification to future work, since it requires a separate LASSO selection and QRNN training run.

2.2. LASSO Variable Selection

We chose LASSO over alternative dimensionality reduction approaches. Unlike Principal Component Analysis, LASSO produces a sparse solution that retains interpretable variable identity, a prerequisite for SHAP attribution, which requires named variables. Unlike Elastic Net, LASSO's pure ℓ_1 penalty enforces harder sparsity, appropriate when the true signal is expected to be concentrated in a small subset of 50 candidates. And unlike stepwise regression, LASSO's penalization is continuous and stable under cross-validation, which avoids the multiple-testing inflation that sequential hypothesis testing introduces.

LASSO [19] is implemented via the `glmnet` package in R as a dimensionality reduction step prior to QRNN estimation. With 50 candidate predictors and an effective sample of 1,499 observations, fitting QRNN directly risks severe overfitting; LASSO reduces the input dimension to a manageable subset while retaining variables with meaningful conditional association. In an end-of-day conditional framework, selected variables reflect contemporaneous market comovement rather than predictive causation; LASSO functions as dimensionality reduction to prevent QRNN overfitting, not as identification of causal risk determinants. Three variables enter without penalty (lagged return, 22-day realized volatility, daily volume) to retain basic autoregressive structure. The penalty parameter λ is selected by five-fold expanding-window cross-validation, where each fold trains on the expanding historical window and validates on the next block to preserve temporal order. As a robustness check, three LASSO-QR variants using `rqPen` were tested; all produced non-sparse models (λ approaching zero); we therefore retained the LASSO-OLS two-step.

LASSO is applied separately at each of the three quantile levels ($\tau \in \{0.01, 0.05, 0.10\}$), and Table 4 reports the resulting quantile-specific variable sets, which are nested for both indices (Section 3). Because a single QRNN produces simultaneous estimates for all three quantiles from one shared input layer (Subsection 2.3), the model requires one fixed feature set rather than three separate ones. We construct this feature set as the union of the quantile-specific LASSO selections across $\tau = 0.01, 0.05, 0.10$, together with the three unpenalized autoregressive terms: 6 variables in total for SRI-KEHATI (`lq45_ret`, `jk_finance_ret`, `issi_ret`, plus the three unpenalized terms) and 11 for IHSG (adding `cds_id_5y`, `jk_mining_ret`, `nikkei_ret`, `hang_seng_ret`, and `gold_ret` to the same three domestic variables and three unpenalized terms).

Because LASSO selection is nested across quantiles for both indices (Section 3), the union is equivalent to the $\tau = 0.10$ selection, the widest quantile-specific set, and no variable is included in the QRNN input solely on the basis of a narrower quantile's selection. This union-based construction is a design choice with a known limitation, discussed in Subsection 3.5 and the Limitations section, for indices where the quantile-specific sets differ substantially: variables selected only for a wider quantile are present throughout training even when estimating the narrowest quantile, which may introduce noise not present under a τ -specific model. Because the LASSO selection itself is performed once using only the 1,199 training observations described above (through October 2023), and this fixed feature set is then used unchanged throughout the entire 300-day rolling-window test period, the variable set predates the test period in its entirety; no test-period information informs which variables are selected, so fixing the feature set for the QRNN's full estimation horizon introduces no look-ahead bias.

2.3. Quantile Regression Neural Network (QRNN)

The QRNN architecture follows Taylor [17] with modifications for high-dimensional financial data. All QRNN results reported in this paper, including the main model in Table 8, the architecture

comparison in Table 2, and the ablation study in Subsection 3.5, are generated using an identical 250-day rolling training window: at each forecast date, the model is retrained on the preceding 250 trading days only, rather than an expanding window that would accumulate all historical observations. We chose QRNN over GARCH-family models and CAViaR [24]. GARCH models impose parametric assumptions on the conditional variance (typically normality or Student- t) that are inconsistent with the excess kurtosis documented in Table 3 (9.21 for SRI-KEHATI, 7.75 for IHSG). CAViaR, meanwhile, estimates conditional quantiles as autoregressive functions of lagged VaR but cannot incorporate high-dimensional predictor sets without intractable interaction terms; QRNN scales naturally to the predictor sets identified by LASSO.

The hidden-layer size was determined empirically rather than by appeal to a universal parameter-to-observation rule: classical statistical theory treats a parameter count exceeding the sample size as a hallmark of underdetermined, overfit estimation, but this guidance does not transfer directly to regularized neural networks, where dropout and batch normalization can support stable generalization even when nominal parameter counts exceed the number of training observations, consistent with the double-descent phenomenon documented by Belkin et al. [25]. We therefore treat the parameter-to-observation ratio as a heuristic starting point, validated empirically rather than assumed: for SRI-KEHATI, with a 250-day rolling training window and 6 LASSO-selected inputs (Table 4), a Linear(16)–Linear(8) architecture yields approximately 323 trainable parameters (ratio 1.3:1); for IHSG, whose union feature set across quantile levels comprises 11 inputs, the same architecture yields approximately 403 trainable parameters (ratio 1.6:1). We also report an informal comparison against a larger architecture, Linear(64)–Linear(32), with approximately 2,819–2,883 trainable parameters (ratio 11.3–11.5:1).

It was run on a comparable 6–7 variable feature set from an auxiliary LASSO-QR specification developed for a related project, which retains global equity spillover variables (`msci_em_ret`, `djia_ret`, and, for IHSG, `sp500_ret`) in place of the domestic IDX variables selected in Table 4. This architecture failed Kupiec calibration in all 6 of 6 index–quantile configurations (Table 2). Because the feature set also differs from the main analysis, we read this as suggestive only: the failure may reflect the larger parameter count, the weaker predictive content of globally sourced features, or both, and we cannot separate them. It is nonetheless consistent with our use of the smaller Linear(16)–Linear(8) architecture throughout.

Table 2: Preliminary architecture comparison (SRK denotes SRI-KEHATI): Linear(16)–Linear(8) (dropout 0.1, the architecture used throughout the main analysis, evaluated on the Table 4 feature set) versus Linear(64)–Linear(32) (dropout 0.2, evaluated but rejected, on a comparable-size auxiliary feature set; see Subsection 2.3 for the feature-set difference), both on the same 250-day rolling window. This comparison varies both architecture size and feature set simultaneously and should be read as suggestive rather than a controlled experiment isolating architecture alone.

Index	τ	Viol. Rate	Kupiec p	Status
SRK	0.01	2.67%	0.016	FAIL
SRK	0.05	8.33%	0.015	FAIL
SRK	0.10	15.00%	0.007	FAIL
IHSG	0.01	4.67%	0.000	FAIL
IHSG	0.05	10.33%	0.000	FAIL
IHSG	0.10	14.67%	0.011	FAIL

Note. Linear(64)–Linear(32) architecture on an auxiliary 6–7 variable LASSO-QR feature set (global equity spillover variables, not the domestic IDX variables in Table 4); same rolling design as the main model. Kupiec summary: 0/6 PASS.

The network structure is: Input $\rightarrow$ Linear(16) $\rightarrow$ BatchNorm $\rightarrow$ ReLU $\rightarrow$ Dropout(0.1) $\rightarrow$ Linear(8) $\rightarrow$ BatchNorm $\rightarrow$ ReLU $\rightarrow$ Dropout(0.1) $\rightarrow$ Linear(3). The output layer produces simultaneous estimates for $\tau \in \{0.01, 0.05, 0.10\}$. The loss function is the aggregate pinball loss:

$$L(y, \hat{q}, \tau) = \frac{1}{N} \sum_{i=1}^N \max[\tau(y_i - \hat{q}_i), (\tau - 1)(y_i - \hat{q}_i)] \quad (1)$$

Rolling-window predictions introduce distributional shift because return scale varies across windows. We z-score normalize the target variable within each 250-day training window and initialize the output layer bias from the empirical quantile of that window (warm-start bias). Without this adjustment, BatchNorm produces divergent predictions across windows due to the mismatch between running statistics and the local distribution.

2.4. SHAP Attribution and Backtesting

SHAP was selected over permutation importance and LIME because it satisfies three axiomatic properties required for valid attribution: efficiency (SHAP values sum to the model output), symmetry (equivalent variables receive equivalent attribution), and dummy (variables with no effect receive zero attribution) [21]. These properties ensure that the climate contribution metric in Eq. (2) is not inflated by correlated variables that would bias simpler importance measures. KernelSHAP [21] is applied with 100 background observations per training window. The climate contribution per observation is defined as:

$$\text{Climate_SHAP}(t) = \frac{\sum_{j \in \mathcal{C}} |\phi_j(t)|}{\sum_{\text{all } j} |\phi_j(t)|} \times 100\% \quad (2)$$

where $\phi_j(t)$ is the SHAP value of variable j at observation t , and $\mathcal{C}$ denotes the set of climate variables. Computing the ratio per observation and then averaging avoids the bias that arises when observations with large total SHAP magnitude dominate the aggregate. Backtesting uses three complementary procedures: the Kupiec [26] test for violation frequency, the Christoffersen [27] test for temporal clustering of violations, and the Dynamic Quantile (DQ) test of Engle and Manganelli [24] for violation predictability. Comparative accuracy uses the Diebold–Mariano test [28] with Newey–West HAC standard errors evaluated on pinball loss. Two benchmarks are Historical Simulation VaR (empirical quantile of the 250-day rolling window) and GARCH(1,1)- t [29], both refitted monthly.

3. Results and Discussion

This section reports the empirical results in the order of the pipeline and discusses each result alongside its interpretation. It begins with the distributional properties of the two return series, which show why a normal benchmark is inadequate for tail risk, and then reports which predictors LASSO retains for each index. The next subsection uses partial correlations to examine why LASSO eliminates the climate variables, and the one after it reports SHAP attribution across quantile levels. The final two subsections present the ablation study of LASSO pre-selection and the backtesting and benchmark comparison, which together evaluate the forecasting performance of the QRNN.

3.1. Distributional Properties

Table 3 reports distributional statistics for both return series. SRI-KEHATI shows excess kurtosis of 9.21, versus 7.75 for IHSG, and both series reject normality overwhelmingly (Jarque–Bera $p < 0.001$). A normal distribution underestimates VaR 99% by 18.5% for SRI-KEHATI and 21.8% for IHSG, meaning that risk models built on normality will systematically miss the losses that matter most. This concentration effect is consistent with the SRI-KEHATI portfolio containing only 25 constituents versus 900+ in IHSG: idiosyncratic shocks not diversified away in the smaller portfolio produce index-level extremes more frequently. Fig. 1 displays the empirical return distributions; the heavy tails relative to the fitted normal curve indicate the inadequacy of the normal benchmark at the VaR 99% level.

Table 3: Distributional statistics for SRI-KEHATI and IHSG daily log returns, January 2013–December 2025.

Statistic	SRI-KEHATI	IHSG
Observations	3,145	3,146
Mean (% per day)	0.016	0.022
Standard deviation (%)	1.292	1.006
Minimum (%)	−8.36	−8.23
Maximum (%)	+14.73	+9.70
Skewness	0.09	−0.39
Excess kurtosis	9.21	7.75
Historical VaR 99% (%)	−3.67	−2.97
Parametric normal VaR 99% (%)	−2.99	−2.32
Normal model underestimate (%)	18.5	21.8

Note. Jarque–Bera rejects normality for both series at $p < 0.001$. The one-observation difference (3,145 vs. 3,146) reflects a single day with no recorded SRI-KEHATI price in the raw data; it predates and is independent of the listwise deletion applied in Section 2.

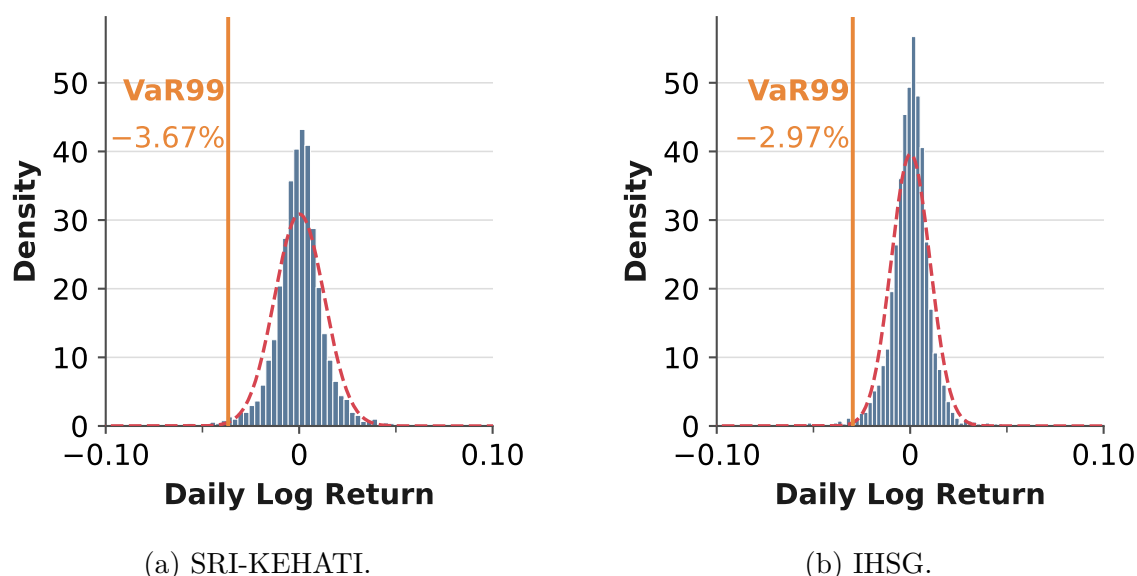


Fig. 1: Return distributions for (a) SRI-KEHATI and (b) IHSG daily log returns, January 2013–December 2025. This is the full raw-price sample ($N = 3,145/3,146$, Table 3) prior to the listwise deletion described in Section 2; the 1,499-observation effective panel used for LASSO-QRNN estimation is a subset of the dates shown here. The dashed red curve represents the fitted normal distribution, and the solid orange vertical line marks the historical 99% VaR (−3.67% for SRI-KEHATI, −2.97% for IHSG; Table 3); the horizontal axis shows returns in decimal units ($-0.05 = -5\%$). Excess kurtosis is 9.21 (SRI-KEHATI) and 7.75 (IHSG), and skewness is 0.09 and −0.39, respectively, confirming fat tails that normal models systematically understate.

3.2. LASSO Feature Selection Results

Table 4 reports LASSO-retained variables across indices and quantile levels. For SRI-KEHATI, three domestic IDX market variables dominate: LQ45, JK Finance sector, and ISSI Islamic equity index; all three trade on the same exchange with identical closing times as SRI-KEHATI, so their selection reflects contemporaneous co-movement rather than predictive causation. For IHSG, a perfectly nested pattern emerges: the variable set at $\tau = 0.01$ is a strict subset of $\tau = 0.05$, which is a subset of $\tau = 0.10$, meaning that global market variables (Nikkei, Hang Seng, gold) enter only at the less extreme quantiles ($\tau = 0.05$ and 0.10). No climate variable enters any specification across both indices. Fig. 2 shows the five-fold cross-validation pinball-loss curve used to select λ for each index, confirming that the reported λ is the interior minimum of a well-behaved curve

rather than an artifact of the search boundary. That no climate variable is selected at any of the three independently cross-validated τ levels in Table 4 (each corresponding to its own λ) is separate, weaker evidence against a boundary artifact than a full coefficient path would provide, since it speaks to three discrete regularization strengths rather than a continuum; we did not retain the full per-variable coefficient path data needed to trace continuous trajectories, and flag this as a direction for a future robustness figure.

Table 4: LASSO-retained variables by index and quantile level.

Index	τ	Selected Variables (excl. unpenalized terms)
SRI-KEHATI	0.01	lq45_ret, jk_finance_ret
SRI-KEHATI	0.05–0.10	lq45_ret, jk_finance_ret, issi_ret
IHSG	0.01	cds_id_5y (first difference), lq45_ret, jk_finance_ret, issi_ret, jk_mining_ret
IHSG	0.05	Previous set + nikkei_ret
IHSG	0.10	Previous set + hang_seng_ret, gold_ret
Both indices	All τ	No climate variable selected

Note. Unpenalized terms (forced): 1-day lagged return, 22-day realized volatility, daily volume proxy.

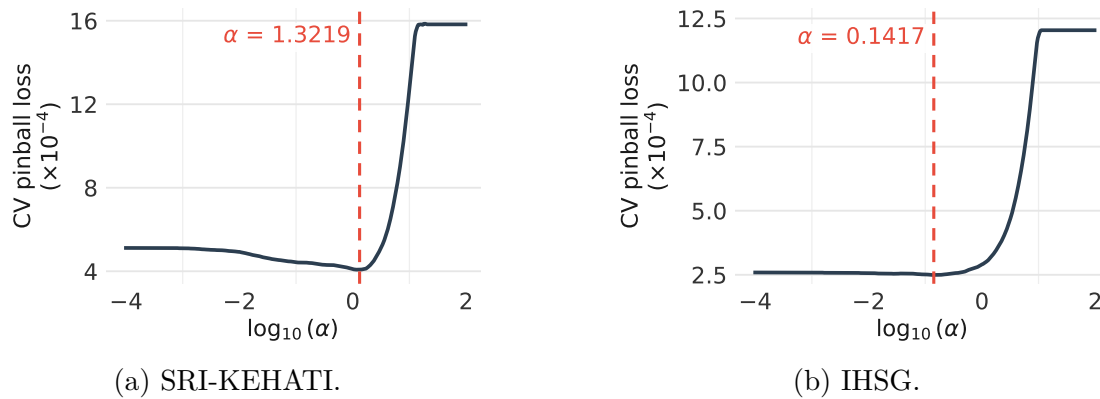


Fig. 2: Five-fold cross-validation pinball loss versus regularization strength $\alpha = \lambda \times n_{\text{train}}$, where λ is the `glmnet` penalty (see Methods), for the LASSO fit at $\tau = 0.05$, for (a) SRI-KEHATI and (b) IHSG. The dashed line marks the cross-validation minimum. This figure shows the loss curve used for λ selection, not per-variable coefficient trajectories; it does not by itself show which variables are climate-related or how their coefficients evolve with λ (see Table 4 for the selected-variable outcome at each τ).

3.3. Why LASSO Eliminates Climate Variables: Partial Correlations

Table 5 shows the pattern behind the elimination of climate variables. The raw correlation of `clean_energy_ret` with IHSG is 0.121, which is distinguishable from zero ($p < 0.001$ at $n = 1,499$, our calculation). After six domestic equity and volatility controls enter the partial correlation model, the coefficient collapses to -0.0005 , a pattern consistent with, though not a formal test of, mediation in the Baron–Kenny sense [30]: the association of clean-energy returns with IHSG disappears once domestic equity factors are included. Clean-energy returns are themselves equity returns, so co-movement with domestic equity factors is expected.

Of the eight climate variables, one (`clean_energy_ret`) loses a statistically significant raw association, three (`carbon_eua_ret`, `chirps_rainfall`, `era5_temp_anomaly`) have no distinguishable raw association to begin with (raw $|r| \leq 0.027$, $p \geq 0.29$, our calculation), three (`pri`, `tri`, `nat_gas`) are weak throughout, and one is an exception: the Geopolitical Risk Daily index [15], whose partial correlation (0.042) exceeds its raw correlation (0.012). This is a suppressor-type pattern, but with $n = 1,499$ a partial correlation of 0.042 corresponds to $p \approx 0.10$ (our calculation), so it is not statistically distinguishable from zero. Two of the six control variables,

LQ45 and JK Finance, are large-capitalization sub-indices whose constituents account for a substantial share of IHSG’s own market capitalization; ISSI and JK Mining are narrower sectoral indices with smaller but non-trivial overlap. Controlling for indices that mechanically share constituents with the dependent variable absorbs variance for compositional as well as economic reasons, and we did not compute standard variance inflation diagnostics for this specification.

The partial-correlation collapse is therefore consistent with two non-exclusive explanations: genuine economic mediation, in which climate information is priced into domestic assets before it can be detected as a separate daily signal, and a partly definitional absorption of variance driven by the compositional overlap between IHSG and its own sub-indices. We cannot distinguish between these explanations with the present specification. Under either explanation, LASSO retains no climate variable once contemporaneous domestic price action is observed; this is a statement about variable selection in this sample, not a formal test of predictive value. We use the Baron–Kenny framing as the standard vocabulary for this pattern. A stricter causal reading would require instrumenting the domestic controls or restricting them to indices with minimal constituent overlap with IHSG, which we leave to future work.

Table 5: Raw and partial correlations of climate variables with IHSG daily returns.

Variable	Raw	Partial	Interpretation
clean_energy_ret	0.121	−0.0005	Association absorbed
carbon_eua_ret	0.027	−0.0003	No raw association
chirps_rainfall	0.024	0.002	No raw association
era5_temp_anomaly	−0.002	−0.008	No raw association
gpr_daily	0.012	0.042	Suppressor effect
pri, tri, nat_gas	<0.03	0.02–0.03	Weak throughout

Note. Control variables: `lq45_ret`, `jk_finance_ret`, `issi_ret`, `jk_mining_ret`, `rv_ihsg_22`, `ihsg_lag1`. LQ45 and JK Finance are large-capitalization sub-indices with substantial constituent overlap with IHSG; see main text for the resulting interpretive caveat.

3.4. SHAP Attribution across Quantile Levels

Table 6 reports aggregate SHAP climate contributions for three model variants. In the main QRNN models, climate SHAP is zero by construction because no climate variable enters the feature set. When climate variables are forced into a diagnostic QRNN (denoted IHSG + Climate, as in the figures), their combined SHAP share is 13.08% at $\tau = 0.01$, computed as the per-observation ratio of climate to total SHAP magnitude, averaged across observations, consistent with Eq. (2) and the Diagnostic LASSO-QR computation below. The forced inclusion is motivated by the hypothesis, not established here, that LASSO removes climate variables because domestic factors absorb their marginal contribution rather than because climate risk is economically irrelevant. Because this diagnostic model retains the same domestic control variables (LQ45, JK Finance, ISSI, JK Mining) flagged in Subsection 3.3 as having substantial constituent overlap with IHSG, the reported climate SHAP shares depend on those controls: to the extent they also absorb variance compositionally, a specification with non-overlapping controls could give a different climate share, and we claim no bound in either direction.

The IHSG + Climate specification forces in only three climate variables (`gpr_daily`, `carbon_eua_ret`, `chirps_rainfall`) rather than all eight climate candidates in Table 1; the Diagnostic LASSO-QR specification discussed below uses all eight. Among these three forced variables, at $\tau = 0.01$ and computed under the same per-observation-then-average procedure used for the totals above, CHIRPS rainfall contributes 49.0% of the IHSG + Climate model’s total climate SHAP (6.41% of total model SHAP), GPR Daily contributes 28.2% (3.68% of total model SHAP), and carbon EUA returns contribute 22.9% (2.99% of total model SHAP, not shown among the ten variables in Fig. 3). Excluding GPR Daily, which is a geopolitical-risk proxy rather than a direct climate measure, the climate share at $\tau = 0.01$ is 9.40% (6.41% +

2.99%). Because only three variables were forced into this specification, this is not a ranking among a larger candidate set, and we do not intend a substantive claim about which of the eight climate variables would dominate under a specification that forced in all eight.

Fig. 3 and Fig. 4 present the SHAP feature importance and category breakdown, respectively, and Fig. 5 presents the time-varying contribution analysis. Under the per-observation-then-average procedure, the physical indicator (CHIRPS rainfall, 6.41% of total model SHAP) and the two remaining variables combined (GPR Daily, a geopolitical-risk proxy, and carbon EUA returns, a transition-policy proxy, $3.68\% + 2.99\% = 6.67\%$) contribute comparably; we do not read this near-even split as evidence that either channel dominates at the daily tail frequency. In Fig. 3, the order of the variables and the percentage labels follow the per-observation-then-average share, whereas bar length shows raw mean $|\text{SHAP}|$; the two need not agree (CHIRPS rainfall, 6.4%, ranks above `rv_ihsg_22`, 6.0%, but has the shorter bar). Carbon EUA returns (2.99%) rank eleventh and are not shown. Two of the ten variables shown are climate variables. Fig. 4 compares three variants, but the first two bars use $\tau = 0.05$ and the third $\tau = 0.01$, so the bars are not quantile-matched; climate segments appear only in the third because no climate variable enters the main specifications (Subsection 3.3).

Table 6: SHAP climate contribution by model and quantile level (%).

Model	$\tau=0.01$	$\tau=0.05$	$\tau=0.10$	Method
QRNN Main (SRI-KEHATI)	0.00	0.00	0.00	Natural
QRNN Main (IHSG)	0.00	0.00	0.00	Natural
QRNN+Climate (IHSG)	13.08	13.32	12.73	Forced (3 climate vars)
Diagnostic LASSO-QR	17.25	8.40	6.12	Diagnostic (8 climate vars)

Note. SHAP computed per observation and averaged for both rows. Diagnostic LASSO-QR uses separate quantile regression on all eight climate variables; QRNN+Climate forces in three. The two rows are not directly comparable on either methodology or variable count.

Climate contributions under the IHSG + Climate specification are 13.08% at $\tau = 0.01$, 13.32% at $\tau = 0.05$, and 12.73% at $\tau = 0.10$. The pattern is essentially flat: the differences across quantiles (0.6 percentage points at most) are too small to describe as a hump or a trend. The VaR 99%-to-VaR 90% ratio of 1.03 confirms that climate share is flat across quantile levels under this specification. The Diagnostic LASSO-QR specification, a separate quantile regression that includes all eight climate variables rather than three, shows a larger and monotonic tail concentration: climate share rises from 6.12% to 17.25% (ratio 2.82) over the same quantile range. The two estimates are not directly comparable: IHSG + Climate is closer to the proposed pipeline but includes only three of the eight climate variables, while the Diagnostic LASSO-QR figure comes from a linear model without the QRNN's regularization and nonlinearity and includes all eight. We cannot attribute the gap between the ratios (1.03 versus 2.82) to either factor alone.

The time-varying SHAP in Fig. 5 tracks the IHSG + Climate contribution across the 250 valid observations of the test period (8 November 2023–26 August 2025), with an overall mean of 13.1%. We split the test period into three mutually exclusive sub-periods: El Niño 2023 (nominally 1 Jun–31 Dec 2023, mean 7.0%; the test window begins on 8 November 2023, so only 8 Nov–31 Dec 2023 enters, and only 28 of the panel's dates fall in that span—itsself a subset of IDX trading days in this window, for reasons we did not trace), Flood 2024 (1 Jan–30 Jun 2024, mean 7.9%), and a residual period labelled “Residual period” in Fig. 5 (nominally 1 Jul 2024–31 Dec 2025, mean 16.4%, matching the 300-day test window's nominal end of 23 December 2025; but the final 50 dates of the panel, 27 Aug–23 Dec 2025, have zero total SHAP and are excluded—see below—so the climate-share observations actually entering this group span 1 Jul 2024–26 Aug 2025, roughly 14 calendar months). Only the first two are anchored on documented climate events, and the third is a residual span whose usable data end well before its nominal close. The windows do not overlap, as a between-period comparison requires, but the group sizes are unequal and the smallest (El Niño 2023) has the least power.

A Kruskal–Wallis test on the three periods yields $H = 13.05$, $df = 2$, $p = 0.0015$ against the null hypothesis that the climate-share values are drawn from the same distribution. The test is rank-based and omnibus: it indicates that at least one period’s distribution differs, but it is not a test of the period means reported above (it equals a comparison of medians only if the distributions have similar shape and spread), and it does not identify which period differs. That would require a post-hoc procedure (e.g., Dunn’s test with a multiple-comparison correction), which we did not conduct. Descriptively, Fig. 5 shows the highest and most volatile climate share in the residual period rather than discrete spikes coinciding with the named 2023–2024 events; we report this as a visual pattern, not as a conclusion the test supports.

The test is run on the daily (per-observation) climate-share values, not on the 20-day rolling mean shown for visual smoothing in Fig. 5; the group sizes are $n = 28$, 66 and 156 for El Niño 2023, Flood 2024 and the residual period (250 valid observations, i.e., 300 minus the 50 zero-SHAP days at $\tau = 0.01$). The test assumes independent observations, which does not hold here: each daily SHAP value comes from a QRNN retrained on a 250-day window that shifts by one day at a time, so nearby dates share most of their training data. We did not attempt a correction for this dependence. The reported $p = 0.0015$ should be read as suggestive of a distributional difference rather than as an exact tail probability; positive dependence of this kind typically makes such a p -value too small. Climate contribution under IHSG + Climate is not uniformly persistent across the test window; it appears to have risen and become more volatile in the second half of the test period, and we have no explanation for this pattern.

The per-observation climate-share distribution is right-skewed. The mean (13.08% at $\tau = 0.01$) exceeds the median (8.79%) by a wide margin, with a standard deviation of 15.74 percentage points. A minority of low-total-SHAP-magnitude days generate disproportionately large climate-share ratios and pull the mean upward. The mean climate-contribution percentages in Table 6 should therefore be read as a central-tendency summary of a skewed distribution, not as a consistently observed daily share.

Separately, 50–52 of the 300 test-period observations (depending on quantile level) have zero total SHAP magnitude across all features; we excluded these from the per-observation average. At $\tau = 0.01$, these 50 dates are not scattered across the test window: they form a single contiguous block at its end, from 27 August to 23 December 2025 (the panel’s final date), with no zero-SHAP date occurring earlier. This pattern is more consistent with a structural gap—for example, the SHAP computation or an input series being unavailable for that final stretch—than with instability isolated to scattered dates. We do not have a confirmed explanation for the underlying mechanism and flag this as a data-quality limitation rather than resolve it here. On the policy side, OJK Regulation No. 51/POJK.03/2017 [31] concerns sustainable-finance action plans and sustainability reporting and contains no provision on VaR or stress testing; our results support only a qualitative recommendation that stress tests should not rely on VaR 90% alone, since the LASSO-QR diagnostic (though not the IHSG + Climate specification, whose pattern is comparatively flat) shows higher climate contribution at VaR 95% and VaR 99% than at VaR 90%; we do not find QRNN-based evidence for a specific VaR 99% threshold as strongly as the LASSO-QR diagnostic alone would suggest.

3.5. Ablation Study: Contribution of LASSO Pre-Selection

To isolate the contribution of LASSO pre-selection from model architecture effects, we conduct a matched-architecture ablation in which Raw-QRNN (no LASSO) and LASSO-QRNN (the feature set in Table 4) share an identical hidden-layer architecture (Linear(16)–BN–ReLU–Dropout(0.1)–Linear(8)–BN–ReLU–Dropout(0.1)–Linear(3)) and are evaluated on the same 300-day test window used for the main model in Table 8 (8 November 2023 to 23 December 2025), using the same 250-day rolling training window. The ablation employs the extended 2013–2025 panel, which excludes Bitcoin (pre-2016 availability) and aluminum futures (late coverage onset) as candidate inputs, yielding 48 predictor candidates for Raw-QRNN compared with the 50

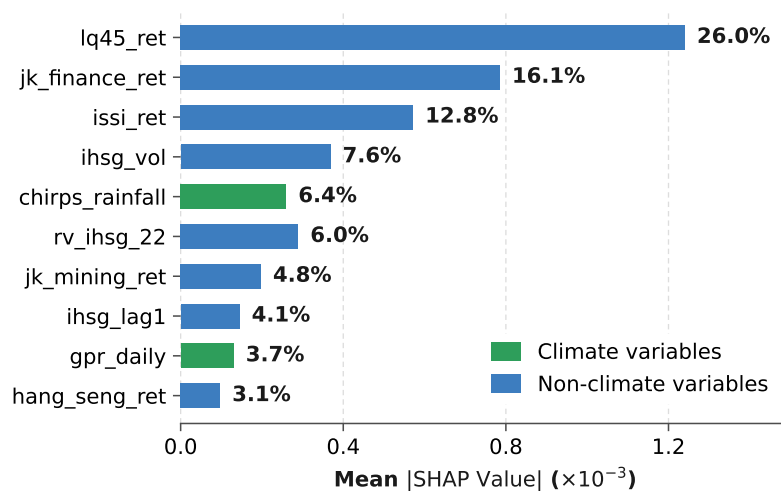
SHAP Importance: IHSG + Climate ($\tau = 0.01$)

Fig. 3: Mean absolute SHAP by variable for IHSG + Climate (IHSG, $\tau = 0.01$); the ten variables shown are ranked by their per-observation-then-average share (labels), while bar length shows raw mean |SHAP|. Climate variables are green. See main text for the ranking criterion and the excluded carbon EUA series.

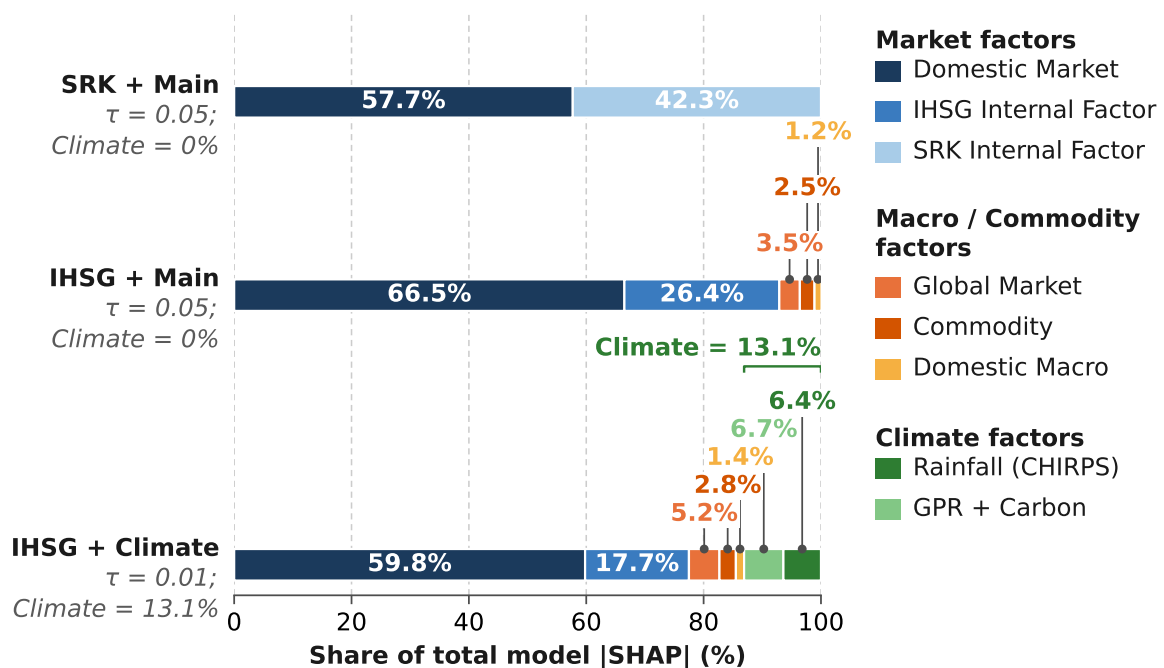
SHAP Contribution by Risk Category

Fig. 4: SHAP contribution by risk category for three model variants: SRK + Main (SRI-KEHATI, $\tau = 0.05$), IHSG + Main (IHSG, $\tau = 0.05$), and IHSG + Climate (IHSG, $\tau = 0.01$). The two green segments, “GPR + Carbon” and “Rainfall (CHIRPS)”, are climate variables and appear only in the forced-inclusion bar; their combined share is bracketed. The quantile level differs between the first two bars and the third.

candidates used for LASSO selection in the main analysis (Section 2). Within the ablation, Raw-QRNN and LASSO-QRNN share an identical architecture and a common set of evaluation dates (established below), differing primarily in whether LASSO pre-selection is applied; this is the comparison Table 7 reports. The ablation uses the extended 2013–2025 panel just described,

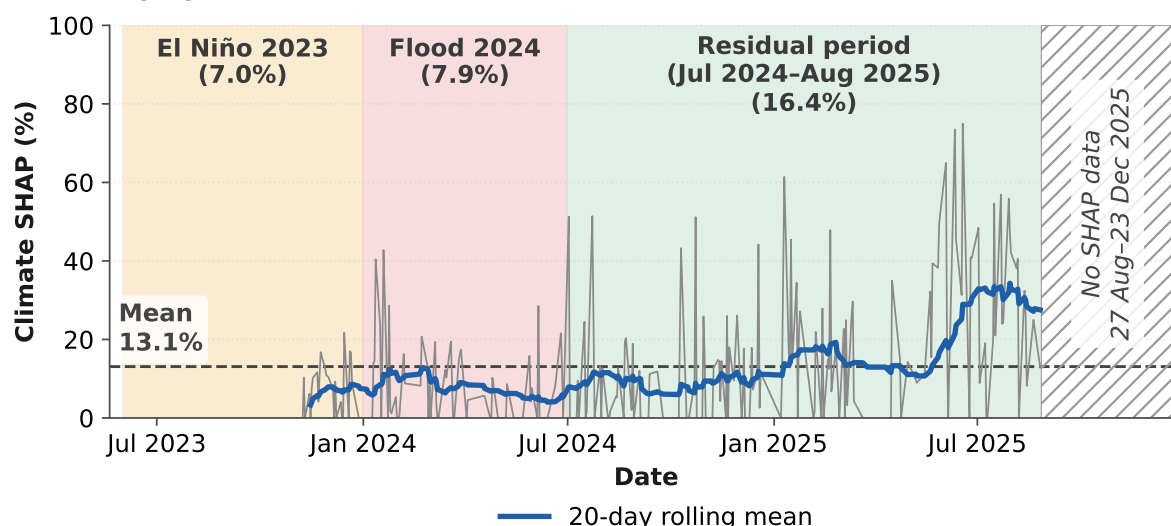
Time-Varying Climate SHAP: IHSG + Climate ($\tau = 0.01$)

Fig. 5: Time-varying climate SHAP for IHSG + Climate (IHSG, $\tau = 0.01$), per-observation-then-average as in Table 6 (grey: daily value; blue: 20-day rolling mean; dashed: overall mean, 13.1%). Shaded regions mark El Niño 2023 (mean 7.0%), Flood 2024 (7.9%) and a residual period, Jul 2024–Aug 2025 (16.4%). Shading shows nominal windows; valid data run from 8 November 2023 to 26 August 2025, as the final 50 dates have zero total SHAP and are excluded. Kruskal–Wallis: $H = 13.05$, $df = 2$, $p = 0.0015$ (omnibus; p -value indicative only, see main text).

whereas the main model uses the panel of Section 2 (September 2014 onward; see Limitations), so the ablation-run Diebold–Mariano statistics discussed below should not be read as directly comparable to Table 8's.

The rolling-window procedure produces a prediction only when both the 250-day training window and the test-date feature row are fully observed. With 48 candidate columns, Raw-QRNN's predictions are missing on 51 of the 300 nominal test dates, most plausibly because at least one globally-sourced series is unavailable on a foreign holiday. The effective sample is $n = 249$. Raw-QRNN's 249 valid dates are a strict subset of LASSO-QRNN's own valid dates (299 for SRI-KEHATI, 256 for IHSG), reflecting LASSO-QRNN's smaller and more reliably available feature set. We restrict Table 7 to this common $n = 249$ sample for both models so that every cell is evaluated on identical dates.

Table 7 reports the results on this common sample. Raw-QRNN fails the Kupiec proportion-of-failures test in all 6 index–quantile configurations, with violation rates between 1.7 and 6.0 times their nominal levels (e.g., 20.88% observed against a 10% nominal level for SRI-KEHATI, about 2.1 times). This failure is consistent with overfitting: with 48 predictors and a 250-day rolling window, the parameter-to-observation ratio for Raw-QRNN is approximately 4:1, roughly three times less favorable than the 1.3:1 ratio of the LASSO-selected 16–8 architecture discussed in Subsection 2.3.

With LASSO pre-selection, LASSO-QRNN passes Kupiec in 4 of 6 configurations, with all three SRI-KEHATI quantile levels passing by a wide margin ($p = 0.747, 0.874, 0.817$ at $\tau = 0.01, 0.05, 0.10$ respectively). IHSG remains harder to calibrate for both models; LASSO-QRNN improves IHSG from 0 of 3 to 1 of 3 Kupiec passes, but at $\tau = 0.05$ the failure mode reverses direction rather than simply persisting: the main model under-violates (2.0% against a 5.0% target, Table 8) while this ablation-run LASSO-QRNN over-violates (10.0%, $n_{\text{viol}} = 25/249$). We do not read this as the same miscalibration appearing in both runs. The two runs differ in at least two respects—the predictor panel (Section 2 panel vs. the extended panel here) and the evaluation sample ($N = 300$ nominal dates vs. this run's $N = 249$ common-sample dates)—and we have not isolated which difference, if either alone, drives the reversal.

We also compared LASSO-QRNN against the Historical Simulation and GARCH(1,1)- t

benchmarks using the Diebold–Mariano test on this same ablation run. Unlike the main model reported in Table 8, this ablation-run LASSO-QRNN is not significantly different from either benchmark in 11 of 12 comparisons; the one exception is IHSG at $\tau = 0.10$, where HS-VaR outperforms significantly ($DM = -2.51$, $p = 0.012$). Training-window design does not explain the attenuated Diebold–Mariano statistics, since it is identical across runs (see Limitations); we have not isolated whether the predictor panel, the evaluation sample, or both together are responsible. The Kupiec calibration result is not unaffected—as noted above, IHSG $\tau = 0.05$ reverses from under- to over-violation between the two runs, which we have likewise not attributed to a specific cause.

Raw-QRNN produces a lower mean pinball loss than LASSO-QRNN (0.000903 versus 0.001096, common $n = 249$ sample). This does not favor the unrestricted model. Raw-QRNN’s lower pinball loss reflects under-dispersed (insufficiently negative) VaR estimates, which minimize point-wise quantile loss while producing the severe Kupiec under-coverage documented above. A model can achieve a lower average loss while providing unreliable risk estimates. Calibration tests such as Kupiec, rather than pinball loss alone, are the appropriate criterion for evaluating VaR validity. Taken together, the ablation indicates that LASSO pre-selection improves calibration relative to an unregularized 48-variable QRNN; it does not show that pre-selection is necessary, because no other regularization scheme was compared. Without it, the 48-variable QRNN fails calibration uniformly across both indices and all quantile levels, which is consistent with overfitting within the 250-day rolling window (an explanation we did not test); with it, calibration is achieved for SRI-KEHATI at all three quantile levels and for IHSG at one of three, at performance statistically indistinguishable from established benchmarks in nearly all configurations.

Table 7: Ablation study: effect of LASSO pre-selection on VaR calibration, evaluated on the identical 300-day nominal test window as the main model (8 Nov 2023–23 Dec 2025). The common effective sample across both models is $n = 249$ for both indices (see main text for why 51 of the 300 nominal dates are dropped). Raw-QRNN and LASSO-QRNN share the Linear(16)–Linear(8) architecture and this common sample.

Index	τ	Raw-QRNN (no LASSO, 48 vars)			LASSO-QRNN (see Table 4)		
		n_{viol}	Viol.%	Kupiec p	n_{viol}	Viol.%	Kupiec p
SRI-KEHATI	0.01	13	5.22%	$<0.001^\times$	2	0.80%	0.747 $^\checkmark$
	0.05	33	13.25%	$<0.001^\times$	13	5.22%	0.874 $^\checkmark$
	0.10	52	20.88%	$<0.001^\times$	26	10.44%	0.817 $^\checkmark$
IHSG	0.01	15	6.02%	$<0.001^\times$	9	3.61%	0.001 $^\times$
	0.05	31	12.45%	$<0.001^\times$	25	10.04%	0.001 $^\times$
	0.10	42	16.87%	$<0.001^\times$	32	12.85%	0.149 $^\checkmark$

Note. $n_{\text{viol}}/249$ recovers the reported violation rate exactly (e.g., $13/249 = 5.22\%$); $n = 249$ (common effective sample, see main text) for every cell in both columns, not the 300-day nominal test window. $^\checkmark =$ Kupiec PASS ($p \geq 0.05$); $^\times =$ FAIL. LASSO-QRNN feature set: union across quantile levels of the variables in Table 4 (6 variables for SRI-KEHATI, 11 for IHSG, each including three unpenalized terms). Raw-QRNN Kupiec summary: 0/6 PASS. LASSO-QRNN Kupiec summary: 4/6 PASS.

3.6. Backtesting and Benchmark Comparison

Table 8 reports backtesting results for the main QRNN models. QRNN passes the Kupiec and Dynamic Quantile tests for SRI-KEHATI at all quantile levels (Christoffersen independence results are not tabulated here). For IHSG at $\tau = 0.05$, the Kupiec test fails with a violation rate of 2.0% against the expected 5.0%; this under-violation indicates the model is overly conservative for the test period, not under-fitted, as a conservative model errs toward greater capital protection relative to the regulatory benchmark. The Dynamic Quantile Test passes in all six index–quantile configurations ($p = 0.425$ – 0.934), so we find no evidence of predictable violations; with 300 observations and only 1 to 22 violations per configuration, the DQ test has limited power. QRNN

wins all 12 Diebold–Mariano comparisons against both benchmarks (all $p < 0.001$) on pinball loss, with the caveats stated in the Conclusion (same-day predictors unavailable to the benchmarks; not reproduced in the ablation run); this advantage holds across all quantile levels and both indices in this run.

Table 8: Backtesting results and Diebold–Mariano (DM) statistics, main QRNN models (test set $N = 300$, November 2023–December 2025). SRK denotes SRI-KEHATI; HS-VaR denotes Historical Simulation VaR. Violation counts (n_{viol}) are reported alongside violation rates to allow direct verification of the Kupiec statistic without back-calculation.

Index	τ	n_{viol}	Viol. Rate	Kupiec p	DM vs HS-VaR	DM vs GARCH- t
SRK	0.01	1	0.33%	0.178	4.31***	7.24***
SRK	0.05	11	3.67%	0.267	6.43***	6.93***
SRK	0.10	21	7.00%	0.069	7.23***	8.17***
IHSG	0.01	1	0.33%	0.178	3.44***	4.08***
IHSG	0.05	6	2.00%*	0.007	6.06***	6.05***
IHSG	0.10	22	7.33%	0.108	6.73***	7.37***

n_{viol}/N recovers the reported violation rate exactly (e.g., $1/300 = 0.33\%$, $11/300 = 3.67\%$); Kupiec p -values were independently recomputed from n_{viol} and $N = 300$ and match the reported values to three decimal places.

*** $p < 0.001$. *Kupiec fails (under-violation: model over-conservative, not under-fitted).

Positive DM statistic = QRNN achieves lower pinball loss. DQ test: 6/6 PASS ($p = 0.425$ – 0.934).

4. Conclusion

We examined how climate variables behave in a three-stage LASSO-QRNN-SHAP pipeline for the VaR of the SRI-KEHATI index and IHSG, using 1,499 daily observations. Whether SRI-KEHATI screening isolates climate risk was not tested directly. Four results stand out.

First, LASSO eliminates all climate variables from both indices. For IHSG, the `clean_energy_ret`–IHSG correlation falls from 0.121 to -0.0005 in partial correlation once domestic equity factors are controlled. This pattern is consistent with domestic equity factors absorbing the climate signal, though we note in [Subsection 3.3](#) that constituent overlap between the domestic controls and IHSG itself means this economic interpretation cannot be fully disentangled from a partly definitional absorption of variance.

Second, climate contributions are tail-concentrated only under a diagnostic quantile regression using all eight climate variables (rising 2.82-fold from VaR 90% to VaR 99%); the pattern is comparatively flat under the QRNN framework itself, which forces in only three climate variables (ratio 1.03; see [Section 3](#)). The diagnostic regression, not the QRNN, provides a qualitative basis for the recommendation that stress tests should not rely on VaR 90% alone, though we stop short of recommending a specific VaR 99% threshold on QRNN-based evidence alone, given the weak and near-flat QRNN-based pattern. Third, in the main backtest QRNN outperforms both Historical Simulation VaR and GARCH(1,1)- t across all 12 Diebold–Mariano comparisons (all $p < 0.001$) on pinball loss. Two caveats apply: QRNN conditions on same-day predictors that the benchmarks do not use, and in the ablation run the LASSO-QRNN was not significantly different from either benchmark in 11 of 12 comparisons ([Subsection 3.5](#)). The comparison therefore supports the pipeline’s fit, not a like-for-like forecasting advantage.

Fourth, a matched-architecture ablation indicates that calibration is better with LASSO pre-selection: on a common effective sample of 249 evaluation dates drawn from the identical 300-day nominal test window ([Subsection 3.5](#)), a QRNN trained on all 48 candidate predictors without LASSO fails Kupiec calibration uniformly (0 of 6 index–quantile configurations), while with LASSO pre-selection the QRNN passes in 4 of 6, with SRI-KEHATI passing all three quantile levels by a wide margin; IHSG still fails at two of three levels. LASSO pre-selection is therefore associated with better calibration, but does not by itself deliver adequate calibration

for IHSG.

For portfolio managers, we note a practical implication with a caveat on its evidentiary basis. LASSO excludes climate variables from the SRI-KEHATI model on the same basis as for IHSG (Table 4), and SRI-KEHATI's LASSO-selected domestic controls (LQ45, JK Finance, ISSI) overlap with those used in the IHSG mediation analysis. This is consistent with the same domestic-absorption mechanism operating in SRI-KEHATI, but does not independently confirm it: the partial correlation and SHAP forced-inclusion diagnostics that quantify the absorption mechanism (Subsection 3.3, Table 6) were conducted for IHSG only. On this basis, we draw no conclusion on whether SRI-KEHATI screening isolates climate risk relative to a broad-market position; the evidence for SRI-KEHATI is limited to LASSO variable selection. Replicating the IHSG partial-correlation and forced-SHAP analysis directly on SRI-KEHATI should precede any implication for portfolio managers.

Limitations include the 52.4% sample reduction from listwise deletion, the end-of-day conditioning framework that limits interpretation to risk attribution rather than pure forecasting, and the single-country, single-index scope. A related limitation concerns the temporal structure of the panel after listwise deletion. Global predictors (e.g., foreign market indices) are missing when their home exchanges are closed but IDX is open, and several domestic-macro series have intermittent gaps, so the retained observations are not contiguous trading days. The 300-observation test set spans 8 November 2023 to 23 December 2025, about 26 months that would ordinarily contain roughly 515–535 IDX trading days. Each 250-day rolling training window therefore spans more than 250 calendar days, and autoregressive terms (lagged return, realized volatility) skip the missing dates. We accept this as a trade-off against imputing missing predictor values (see Section 2), but did not test sensitivity to the size or placement of the gaps.

A further limitation concerns the comparison between the ablation in Subsection 3.5 and the main model in Table 8. Both use the same 250-day rolling training window, architecture, and hyperparameters (seed, epochs, learning rate, batch size, dropout) on the Table 4 feature set, so training-window design does not explain the attenuated Diebold–Mariano statistics or the reversed Kupiec failure mode at IHSG $\tau = 0.05$ (Subsection 3.5) in the ablation. Two things verifiably differ between the runs, not one: the predictor panel (main model: Section 2 panel, September 2014 onward, $N_{\text{eff}} = 1,499$; ablation: extended panel, January 2013 onward, excluding Bitcoin and aluminum futures) and the evaluation sample ($N = 300$ nominal test dates for the main model versus this ablation's $N = 249$ common-sample dates, Table 7).

We have not isolated the contribution of either difference alone, so attributing the attenuation or the reversed failure mode to the panel specifically is not something we verified. The Kupiec results, which are internally consistent within each run and do not rely on a benchmark, remain the primary evidence of LASSO-QRNN's advantage over Raw-QRNN. Finally, whether the domestic absorption mechanism identified here is specific to Indonesia's shallow, retail-dominated equity market or generalizable across ASEAN emerging markets remains an open question; markets with stronger institutional investor presence (such as Singapore or Thailand) would be expected to show a weaker pattern.

CRedit Authorship Contribution Statement

Annisa Syalsabila: Conceptualization; Methodology; Software; Formal Analysis; Investigation; Visualization; Writing—Original Draft; Writing—Review & Editing; Project Administration; Funding Acquisition; Supervision.

Nur Ikhwana: Data Curation; Formal Analysis; Writing—Review & Editing.

Lalu Ramzy Rahmanda: Data Curation; Formal Analysis; Validation; Writing—Review & Editing.

Declaration of Generative AI and AI-assisted Technologies

The authors used Claude (Anthropic) during manuscript preparation for the following purposes: writing assistance (drafting and editing prose sections). The authors take full responsibility for the integrity of this work.

Declaration of Competing Interest

The authors declare no competing interests.

Ethics Approval and Consent to Participate

Ethics approval was not required for this study. All data used are publicly available from financial data providers and open-access climate databases.

Funding and Acknowledgments

This research was funded by the BIMA Penelitian Dasar Pemula (PDP) Affirmasi scheme, under contract numbers 242/C3/DT.05.00/PL-BARU/2026 and 2816/DST/UN36.12/TU/2026, granted by the Direktorat Riset, Teknologi, dan Pengabdian kepada Masyarakat (DRTPM), Ministry of Higher Education, Science, and Technology of the Republic of Indonesia (Kemdiktisaintek) through Universitas Negeri Makassar.

Data and Code Availability

The panel dataset and R/Python analysis scripts are available from the authors upon reasonable request. Climate data are publicly available from the Copernicus Climate Data Store¹ and UCSB CHIRPS². Financial data are sourced from Yahoo Finance and Investing.com.

References

- [1] World Bank. *Indonesia Economic Quarterly, December 2019: Investing in People*. Box “An economic analysis of the 2019 forest and land fires and haze”, p. 7. Washington, DC: World Bank, Dec. 2019. DOI: [10.1596/33033](https://doi.org/10.1596/33033).
- [2] BNPB. *Kerugian Kebakaran Hutan dan Lahan Sepanjang 2019 Capai Rp 75 Triliun*. Tech. rep. News item, 30 December 2019; the Rp 75 trillion total economic loss is attributed to World Bank data. Jakarta: Badan Nasional Penanggulangan Bencana, Dec. 2019. <https://bnpb.go.id/berita/kerugian-kebakaran-hutan-dan-lahan-sepanjang-2019-capai-rp-75-triliun>.
- [3] BNPB. *UPDATE Banjir dan Longsor: Korban Meninggal Bertambah, Menjadi 67 Orang dan Satu Hilang*. Tech. rep. News item, 6 January 2020; reports 67 fatalities and 1 missing person as of 6 January 2020 for the Jabodetabek floods and landslides. Jakarta: Badan Nasional Penanggulangan Bencana, Jan. 2020. <https://www.bnpb.go.id/berita/update-banjir-korban-meninggal-bertambah-menjadi-67-orang-dan-satu-hilang>.

¹<https://cds.climate.copernicus.eu>

²<https://www.chc.ucsb.edu/data/chirps>

- [4] Bank Indonesia. *Kerugian Ekonomi Akibat Banjir Jakarta Awal 2020 Capai Rp1 Triliun*. Tech. rep. Estimate by Bank Indonesia's DKI Jakarta office, compiled from Aprindo, Hipmi, and Kadin data, as reported by Alinea.id (28 February 2020); covers trade, transport, warehousing/logistics, and financial-services losses from the 1 January 2020 Jabodetabek floods. Jakarta: Bank Indonesia, Kantor Perwakilan DKI Jakarta, Feb. 2020. <https://www.alinea.id/bisnis/bi-kerugian-akibat-banjir-jakarta-awal-2020-capai-rp1-triliun-b1ZIr9rWp>.
- [5] Center of Economic and Law Studies (CELIOS). *Dampak Kerugian Ekonomi Bencana Banjir Sumatera*. Tech. rep. Projected economic loss from the November 2025 Sumatra floods and landslides (housing, household income, roads and bridges, agricultural land). Jakarta: Center of Economic and Law Studies, Dec. 2025. <https://celios.co.id/dampak-kerugian-ekonomi-bencana-banjir-sumatera/>.
- [6] Indonesia Stock Exchange (IDX) and Yayasan Keanekaragaman Hayati Indonesia (KEHATI). *Panduan Indeks SRI-KEHATI*. Tech. rep. Official index methodology guide (PDF). Jakarta: PT Bursa Efek Indonesia, Apr. 2022. https://www.idx.co.id/media/11195/panduan-indeks-sri-kehati_apr22.pdf.
- [7] CRMS Indonesia. *Mengenal Indeks Keberlanjutan Perusahaan dari SRI-KEHATI*. Web article, 27 September 2018; secondary source for the selection thresholds. Sept. 2018. <https://crmsindonesia.org/publications/mengenal-indeks-keberlanjutan-perusahaan-dari-sri-kehati/> (visited on 04/01/2026).
- [8] Patrick Bolton, Morgan Despres, Luiz Awazu Pereira da Silva, Frédéric Samama, and Romain Svartzman. *The Green Swan: Central Banking and Financial Stability in the Age of Climate Change*. Basel: Bank for International Settlements, 2020. <https://www.bis.org/publications/green-swan.pdf>.
- [9] Michael Barnett, William Brock, and Lars Peter Hansen. "Pricing Uncertainty Induced by Climate Change". In: *Review of Financial Studies* 33.3 (2020), pp. 1024–1066. DOI: [10.1093/rfs/hhz144](https://doi.org/10.1093/rfs/hhz144).
- [10] Renato Faccini, Rastin Matin, and George Skiadopoulos. "Dissecting Climate Risks: Are They Reflected in Stock Prices?" In: *Journal of Banking & Finance* 155 (2023), p. 106948. DOI: [10.1016/j.jbankfin.2023.106948](https://doi.org/10.1016/j.jbankfin.2023.106948).
- [11] Gunnar Friede, Timo Busch, and Alexander Bassen. "ESG and financial performance: aggregated evidence from more than 2000 empirical studies". In: *Journal of Sustainable Finance & Investment* 5.4 (2015), pp. 210–233. DOI: [10.1080/20430795.2015.1118917](https://doi.org/10.1080/20430795.2015.1118917).
- [12] Robert F. Engle, Stefano Giglio, Bryan Kelly, Heebum Lee, and Johannes Stroebel. "Hedging Climate Change News". In: *Review of Financial Studies* 33.3 (2020), pp. 1184–1216. DOI: [10.1093/rfs/hhz072](https://doi.org/10.1093/rfs/hhz072).
- [13] Mohammadhossein Lashkaripour. "ESG tail risk: The Covid-19 market crash analysis". In: *Finance Research Letters* 53 (2023), p. 103598. DOI: [10.1016/j.frl.2022.103598](https://doi.org/10.1016/j.frl.2022.103598).
- [14] Nada Yasya, Beatrice Alicia Maharani, Nabiila Putri Ardiansyah, Adinda Sabelia, and Taufiq Hidayat. "Risk and Return of Sustainable Awarded Company: Comparison between Stock Index Sri-Kehati and LQ45". In: *Jurnal Keuangan dan Perbankan* 19.1 (2023), pp. 37–50. DOI: [10.35384/jkp.v19i1.340](https://doi.org/10.35384/jkp.v19i1.340).
- [15] Dario Caldara and Matteo Iacoviello. "Measuring Geopolitical Risk". In: *American Economic Review* 112.4 (2022), pp. 1194–1225. DOI: [10.1257/aer.20191823](https://doi.org/10.1257/aer.20191823).
- [16] Dario Caldara and Matteo Iacoviello. *Geopolitical Risk (GPR) Index*. Authors' website. Daily GPR index data, updated weekly. 2026. <https://www.matteoiacoviello.com/gpr.htm>.

- [17] James W. Taylor. “A quantile regression neural network approach to estimating the conditional density of multiperiod returns”. In: *Journal of Forecasting* 19.4 (2000), pp. 299–311. DOI: [10.1002/1099-131X\(200007\)19:4<299::AID-FOR775>3.0.CO;2-V](https://doi.org/10.1002/1099-131X(200007)19:4<299::AID-FOR775>3.0.CO;2-V).
- [18] Xiaochun Liu and Richard Luger. “Quantile-based modeling of scale dynamics in financial returns for Value-at-Risk and Expected Shortfall forecasting”. In: *International Journal of Forecasting* 42.3 (2026), pp. 872–888. DOI: [10.1016/j.ijforecast.2025.12.002](https://doi.org/10.1016/j.ijforecast.2025.12.002).
- [19] Robert Tibshirani. “Regression shrinkage and selection via the lasso”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1 (1996), pp. 267–288. DOI: [10.1111/j.2517-6161.1996.tb02080.x](https://doi.org/10.1111/j.2517-6161.1996.tb02080.x).
- [20] Alexandre Belloni and Victor Chernozhukov. “ ℓ_1 -penalized quantile regression in high-dimensional sparse models”. In: *The Annals of Statistics* 39.1 (2011), pp. 82–130. DOI: [10.1214/10-AOS827](https://doi.org/10.1214/10-AOS827).
- [21] Scott M. Lundberg and Su-In Lee. “A Unified Approach to Interpreting Model Predictions”. In: *Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017, pp. 4765–4774. DOI: <https://dl.acm.org/doi/10.5555/3295222.3295230>.
- [22] Haithem Awijen, Sami Ben Jabeur, and Julien Pillot. “Interpretable machine learning models for ESG stock prices under transition and physical climate risk”. In: *Annals of Operations Research* 357.1 (2026), pp. 191–221. DOI: [10.1007/s10479-024-06231-x](https://doi.org/10.1007/s10479-024-06231-x).
- [23] Tobias Adrian and Markus K. Brunnermeier. “CoVaR”. In: *American Economic Review* 106.7 (2016), pp. 1705–1741. DOI: [10.1257/aer.20120555](https://doi.org/10.1257/aer.20120555).
- [24] Robert F. Engle and Simone Manganelli. “CAViaR: Conditional Autoregressive Value at risk by Regression Quantiles”. In: *Journal of Business & Economic Statistics* 22.4 (2004), pp. 367–381. DOI: [10.1198/073500104000000370](https://doi.org/10.1198/073500104000000370).
- [25] Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal. “Reconciling modern machine-learning practice and the classical bias-variance trade-off”. In: *Proceedings of the National Academy of Sciences* 116.32 (2019), pp. 15849–15854. DOI: [10.1073/pnas.1903070116](https://doi.org/10.1073/pnas.1903070116).
- [26] Paul H. Kupiec. “Techniques for Verifying The Accuracy of Risk Measurement Models”. In: *Journal of Derivatives* 3.2 (1995), pp. 73–84. DOI: [10.3905/jod.1995.407942](https://doi.org/10.3905/jod.1995.407942).
- [27] Peter F. Christoffersen. “Evaluating Interval Forecasts”. In: *International Economic Review* 39.4 (1998), pp. 841–862. DOI: [10.2307/2527341](https://doi.org/10.2307/2527341).
- [28] Francis X. Diebold and Roberto S. Mariano. “Comparing Predictive Accuracy”. In: *Journal of Business & Economic Statistics* 13.3 (1995), pp. 253–263. DOI: [10.1080/07350015.1995.10524599](https://doi.org/10.1080/07350015.1995.10524599).
- [29] Christian Francq and Jean-Michel Zakoian. *GARCH Models: Structure, Statistical Inference and Financial Applications*. 2nd. Chichester: Wiley, 2019. DOI: [10.1002/9781119313472](https://doi.org/10.1002/9781119313472).
- [30] Reuben M. Baron and David A. Kenny. “The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations”. In: *Journal of Personality and Social Psychology* 51.6 (1986), pp. 1173–1182. DOI: [10.1037/0022-3514.51.6.1173](https://doi.org/10.1037/0022-3514.51.6.1173).
- [31] Otoritas Jasa Keuangan. *Peraturan Otoritas Jasa Keuangan Nomor 51/POJK.03/2017 tentang Penerapan Keuangan Berkelanjutan bagi Lembaga Jasa Keuangan, Emiten, dan Perusahaan Publik*. Lembaran Negara Republik Indonesia Tahun 2017 Nomor 169. 2017. <https://peraturan.go.id/files/ojk51-2017bt.pdf>.