Modelling volatility in International Renewable Energy Certificate markets: a machine-learning approach

Federico Toscano[1]

[1]School of Business and Economics, Maastricht University, The Netherlands

Abstract.

Price volatility in renewable energy certificate markets undermines investment incentives, distorts price signals, and weakens the capacity of these instruments to support cost-effective decarbonisation. Although the International Renewable Energy Certificate (I-REC) framework constitutes the world's largest certificate system by geographical coverage, the modelling and forecasting of volatility in I-REC markets has received virtually no attention in the literature. This paper presents a comparative assessment of econometric and machine-learning approaches to modelling and forecasting I-REC price volatility. Four GARCH-class specifications— GARCH, EGARCH, TGARCH, and FIGARCH—augmented with exogenous energy market drivers are benchmarked against an Extreme Gradient Boosting model incorporating Heterogeneous Autoregressive features, regime indicators, and the same set of exogenous predictors. The models are applied to twenty-six I-REC price series spanning four countries, three generation technologies, and two vintage categories, and are evaluated on both in-sample fit and out-of-sample forecasting accuracy. The results indicate that XGBoost consistently outperforms all GARCH benchmarks, achieving superior predictive accuracy across virtually all out-of-sample cases, while no single GARCH specification performs consistently across all market segments. A SHAP-based interpretability analysis reveals that I-REC volatility dynamics are predominantly endogenous, with limited systematic influence from broader energy commodity or electricity spot markets. The findings carry implications for market participants seeking improved risk management tools and for policymakers designing regulatory frameworks for nascent certificate markets.

Key words: renewable energy certificates, I-REC, volatility, GARCH, XGBoost

1. Introduction

Climate policy frameworks increasingly emphasise the decarbonisation of electricity systems as a cornerstone of net-zero commitments (Kriegler et al., 2014). Achieving material reductions in CO_2 emissions requires substantive engagement from the commercial sector (Nurdiawati and Urban, 2021). Among the policy mechanisms deployed to address this challenge, Renewable Energy Certificate (REC) markets have emerged as an important instrument for reducing industrial carbon footprints. Given that on-site renewable generation is frequently constrained by locational factors, and that corporations are actively seeking greater exposure to emissions-free electricity (O'Shaughnessy et al., 2021; Lee et al., 2022), RECs provide a viable pathway through which organisations can demonstrate a commitment to procuring decarbonised electricity while circumventing the geographical constraints inherent in renewable energy project siting.

Within this landscape, the International Renewable Energy Certificate (I-REC) constitutes a particularly important example of a certificate-based market. Whereas many advanced economies have implemented regional, national, or subnational certificate schemes, a large number of emerging economies have adopted the standards established by I-REC. These standards are recognised by major reporting frameworks—including the Greenhouse Gas Protocol, the Carbon Disclosure Project, and RE100—as providing a credible and auditable instrument for tracking renewable electricity claims (Nandagopal and Devaraja, 2018; Jia et al., 2023). At present, more than sixty countries have implemented national certificate markets under the I-REC framework, making it the world's largest certificate system in terms of geographical coverage and penetration of renewable energy claims.

Despite the potentially significant role of I-REC markets, prices and transaction volumes remain a fraction of those observed in more mature certificate schemes, such as European Guarantees of Origin, thereby limiting their capacity to fulfil the promised role of improving the economic outlook of renewable energy projects (Dong et al., 2025). Among the factors constraining the effective functioning of I-REC markets, price volatility is widely identified as one of the most critical, as it undermines the attractiveness of certificate-based support mechanisms. Early work by Fagiani et al. (2013) demonstrates that the relative advantage of certificates over non-market-based mechanisms—such as feed-in tariffs—depends crucially on investors' tolerance for risk and fluctuations in unrealised revenues. Subsequent analyses associate high-volatility regimes with lower investment probabilities in renewable energy projects and with reduced cost-effectiveness of certificate markets overall (Fagiani and Hakvoort, 2014). Related findings further link risk reduction to a lower weighted average cost of capital and an improved levelised cost of energy for renewable energy projects (Đukan and Kitzing, 2023), underscoring the central role of volatility in shaping both investor returns and policymakers' ability to mobilise cost-effective renewable investment.

While the relevance of volatility in shaping the performance of certificate markets is widely acknowledged in the literature, the problem of treating volatility has received comparatively limited attention. In particular, there has been little methodological progress in modelling and forecasting certificate price volatility beyond the approach proposed by Fagiani and Hakvoort (2014), who employ a Generalised Autoregressive Conditional Heteroskedasticity (GARCH) model to characterise the volatility dynamics of Guarantees of Origin in Northern European countries. This stands in contrast to the substantial progress made by the quantitative finance literature over the past decade, which has not only extended the use of more complex and sophisticated GARCH-class models to energy commodities, but has also introduced machine-learning algorithms that have been proven to systematically outperform traditional econometric models in the presence of pronounced and persistent volatility.

Within this framework, this paper proposes a comparative assessment of quantitative methods for modelling and forecasting I-REC price volatility, with the aim of improving the understanding and predictive accuracy of volatility dynamics in these markets. Building on the existing literature, the paper first expands the set of GARCH specifications considered, incorporating asymmetric and non-linear behaviours that are better suited to capture the heterogeneous statistical properties of I-REC price series. Second, the article proposes the use of an Extreme Gradient Boosting (XGBoost) model to comparatively evaluate the performance of GARCH models and to test its ability to deliver superior descriptive and forecasting performance. The proposed models are then applied to a comprehensive dataset of I-REC price series spanning multiple markets, technologies, and vintages, and their performance is assessed in terms of in-sample fit and out-of-sample forecasting accuracy, thereby providing a robust framework for comparing alternative quantitative approaches to modelling I-REC price volatility.

This paper makes several contributions to the literature on certificate markets and energy commodities. First, it documents and analyses volatility patterns in I-REC prices across a range of countries and technologies, offering a comprehensive empirical characterisation of volatility dynamics in these markets. Second, it extends the application of GARCH-type models to I-REC markets by considering EGARCH, TGARCH, and FIGARCH specifications, thereby establishing a broader benchmark for evaluating volatility modelling and forecasting performance. Third, it applies a well-established machine-learning algorithm to a novel empirical setting and systematically compares its performance with that of traditional statistical models, thereby extending the existing literature to an understudied market. Finally, it provides a comprehensive comparison of the performance of different quantitative approaches in modelling and forecasting volatility in I-REC markets, offering insights into the relative strengths and weaknesses of each method and their suitability for different market conditions and data characteristics. Beyond its methodological contribution, the analysis is intended to be informative for market participants and policymakers alike. In particular, it sheds light on the sources and persistence of volatility in I-REC markets, with implications for risk management and trading strategies, as well as for the design of regulatory frameworks aimed at enhancing market stability and effectiveness.

2. Literature Review

2.1. Volatility and renewable energy certificate markets

Volatility is widely recognised as a significant impediment to the development and maturation of green financial markets. Iqbal et al. (2022) document pronounced volatility spikes and spillover effects in sustainable investment indices during episodes of heightened policy uncertainty and market-wide shocks, thereby amplifying risk exposure for socially responsible investors. A closely related dynamic is identified by Adekoya et al. (2021), who show that European carbon markets tend to act as net receivers of volatility in periods of elevated policy uncertainty, with adverse implications for long-term investment planning. Similar patterns are observed in green bond markets, where returns are negatively affected by exogenous shocks stemming from geopolitical risk and policy uncertainty, increasing their vulnerability, particularly over short horizons (Tang et al., 2023). As already highlighted, certificate markets are subject to comparable volatility-related challenges.

The most immediate and intuitive channel through which volatility undermines the functioning of renewable energy certificate schemes operates via its impact on expected future revenues. Building on the analysis of Fagiani and Hakvoort (2014), Ciarreta et al. (2020) argue that under high-volatility regimes investors demand higher risk-adjusted returns from renewable energy projects, returns that are frequently misaligned with prevailing market conditions. This mismatch increases the probability of project delays or cancellations, particularly for capital-intensive technologies, such as onshore and offshore wind. Moreover, the same authors emphasise that heightened uncertainty regarding policy commitment—reflected in volatile certificate prices—can erode investors' long-term confidence in these markets. Collectively, these factors weaken the additionality effects that certificate revenues are intended to deliver.

Beyond its impact on investment incentives, volatility may also give rise to market distortions. Extending evidence from European markets, Husain et al. (2024) observe that in the United States persistent volatility can lead to a decoupling of certificate prices from underlying market fundamentals. Since REC prices are expected to reflect the marginal value of additional renewable energy generation, sustained price spikes or troughs may distort the economic signals faced by investors. As a result, investment decisions may be biased towards over- or under-deployment of specific technologies or locations that are not cost-optimal from a long-term perspective.

Finally, given that certificates often function as implicit long-term hedging instruments against carbon and fuel price risk, elevated volatility increases hedging costs and reduces the effectiveness of such risk-mitigation strategies (Ciarreta et al., 2017). This mechanism is reinforced by empirical evidence provided by Irfan (2021), who documents a cointegration relationship between certificate and electricity prices in India. His findings suggest that instability in either market can be transmitted to the other, thereby weakening the ability of certificates to provide reliable hedging benefits and, in turn, diverting resources away from the efficient operation and expansion of renewable capacity.

Several factors contribute to the persistent presence of volatility in certificate markets. Policy uncertainty is consistently identified as a primary driver, a finding that applies broadly to renewable energy-related markets (Liang et al., 2022; Alharbey and Ben-Salha, 2024) and to certificate markets in particular (Fagiani and Hakvoort, 2014). Geopolitical risk represents a further source of volatility, as documented more recently by Zhang et al. (2023). In addition, imbalances between certificate supply and demand play a non-negligible role. Periods of excess supply are often associated with increased uncertainty surrounding national decarbonisation targets, which in turn undermines revenue expectations and adversely affects the financial viability of renewable energy projects (Wang et al., 2024).

While the determinants of volatility in certificate markets are relatively well understood and extensively documented, the problem of forecasting volatility has received comparatively limited attention. Yet, accurate volatility modelling is crucial for effectively managing the chronic risks inherent in environmental markets. Improvements in volatility forecasting can enhance policymakers' ability to plan and to assess the influence of different market drivers (Uddin et al., 2025), while also supporting more effective risk management through improved hedging strategies, as demonstrated in recent empirical contributions (Raza et al., 2024; Wang et al., 2025). In light of the substantial methodological advances in the quantitative literature over the past decade, and given the near absence of empirical evidence on volatility dynamics in I-REC markets, the potential contribution of this research becomes particularly salient.

2.2. Volatility modelling in environmental markets

While volatility modelling in certificate markets remains a relatively underdeveloped strand of the literature, the broader field of environmental finance has witnessed substantial methodological progress, particularly in the application of both traditional econometric techniques and more recent machine-learning approaches.

GARCH-class models have been extensively employed in the analysis of carbon markets, with a large body of evidence emerging from studies of the European Union Emissions Trading System (EU ETS). A notable contribution is provided by Liu et al. (2021a), who adopt a GARCH framework augmented with mixed data sampling to capture the effects of policy uncertainty on the volatility of EUA futures, reporting strong out-of-sample forecasting performance. A related perspective is offered by Niu and Liu (2024), who examine the forecasting ability of the Glosten–Jagannathan–Runkle GARCH model for EU allowance prices. Other studies adopt a broader systemic view. Zhang et al. (2022) and Alkathery and Chaudhuri (2021) employ alternative GARCH specifications to investigate volatility dynamics and cross-market interactions among energy, carbon, and renewable energy markets in China and in Gulf countries, respectively. Collectively, this literature represents the backbone evidence of the relevance of GARCH-class models for capturing time-varying volatility and spillover effects in environmental and energy-related markets.

At the same time, the literature on energy and environmental market forecasting has expanded significantly in recent years, with methodological developments increasingly extending beyond conventional statistical models. A growing number of studies have adopted machine-learning algorithms to model and forecast price dynamics in energy commodity and environmental markets, often with encouraging results (e.g. see Chevallier, 2011; Zhu et al., 2016, 2022). In recent times, various studies in this line of research have focussed directly on the performance of extreme gradient boosting algorithms in modelling price dynamics in energy and environmental markets, showing that similar tree-based ensemble methods are particularly effective in capturing complex, nonlinear relationships and regime-dependent behaviour that traditional econometric models often fail to accommodate, thus leading to notable improvements in forecasting accuracy (Xu et al., 2023; Zhu et al., 2025). A particularly relevant example of this progress is presented by Ding et al. (2022), who compare the performance of XGBoost with that of a GARCH benchmark in forecasting volatility in oil and electricity futures markets. Their findings indicate that the machine-learning approach outperforms the econometric model, especially in capturing pronounced and extreme volatility patterns that characterise energy commodity markets.

Despite the valuable insights offered by this body of work, several important gaps remain. First, the application of GARCH-type models to certificate markets is limited in both scope and depth. To date, Fagiani and Hakvoort (2014) remains the only study providing a detailed analysis of volatility dynamics in certificate markets, focusing on a relatively narrow set of European contexts. This leaves substantial gaps in our understanding of volatility behaviour across other certificate markets, countries, technologies, and vintages, and prevents a systematic evaluation of more advanced GARCH specifications in capturing heterogeneous volatility dynamics. Second, the use of machine-learning models for volatility forecasting in environmental and certificate markets is still at an early stage, with only a small number of studies exploring their potential advantages. This creates a clear opportunity to extend machine-learning approaches that have proven effective in other energy commodity markets to the context of I-REC certificate markets, and to assess their performance within a rigorous and comprehensive forecasting framework that spans multiple markets, technologies, and vintages and benchmarks them against a broad set of GARCH models.

Specifically, the paper addresses these gaps through a structured empirical strategy. First, it extends the application of GARCH-class models to I-REC markets by introducing EGARCH, TGARCH, and FIGARCH specifications alongside the standard GARCH framework, incorporating a set of relevant exogenous variables and applying them to a broad sample of more than twenty certificate markets. The use of more flexible volatility specifications, combined with exogenous drivers, is intended to better capture asymmetries, persistence, and regime-dependent dynamics that characterise I-REC price volatility. Second, the study develops and calibrates an XGBoost model grounded in the energy economics and quantitative finance literature, integrating the exogenous variables employed in the GARCH models with a parsimonious set of additional engineered features designed to enhance the algorithm's ability to respond to abrupt changes and nonlinear volatility patterns. Given the relatively limited data sample, the model is designed to limit the risk of overfitting. Finally, the two classes of models are systematically compared in terms of both in-sample fit and out-of-sample forecasting accuracy, using different evaluation metrics and assessing the consistency and robustness of their performance across markets, technologies, and vintages.

3. Methodology

3.1. Volatility measures and features

From a theoretical perspective, volatility represents the instantaneous variability of asset returns in continuous time. In standard asset-pricing frameworks, returns are typically modelled as diffusion processes driven by Brownian motion, in which volatility determines the scale of random price fluctuations and thus the degree of uncertainty faced by investors (Samuelson, 1965). Within this setting, volatility over a given time horizon is summarised by integrated variance, which measures the cumulative variability of returns along the price path. However, integrated variance is not directly observable in practice, as it would require continuous monitoring of the underlying variance process. This inherent unobservability necessitates the use of discrete-time proxies constructed from observed return data when volatility is analysed empirically.

In empirical applications, volatility must therefore be operationalised using discrete-time measures constructed from observed returns. Squared returns constitute the most widely used proxy, as they provide a direct measure of the magnitude of price fluctuations over the observation interval and are consistent with the definition of variance in continuous-time return processes. This approach was first formalised by Engle (1982) in his seminal contribution on autoregressive conditional heteroskedasticity and has since been extensively applied to daily financial time series. In this context, squared returns serve as an observable estimate of the variance accumulated over a trading day and form the basis of a large class of conditional volatility models (Barndorff-Nielsen and Shephard, 2002).

Nevertheless, early contributions by Andersen and Bollerslev (1998) already emphasised that, while squared returns are an unbiased estimator of volatility, they may constitute a highly unstable measure. This limitation is further corroborated by Andersen et al. (2001), who demonstrate that a single squared daily return is an extremely noisy indicator of the latent volatility process and offers only limited informational content for reliable statistical inference. An alternative class of realised volatility measures relies on the aggregation of returns over rolling windows, typically computed as the standard deviation or variance of returns over a fixed horizon. By construction, this approach reduces measurement noise through temporal aggregation and yields smoother volatility estimates. Within this framework, Mukherjee and Goswami (2017) show that rolling-window measures are better suited to capturing medium- to long-term volatility trends that may be obscured by short-horizon proxies such as squared returns. In addition, Thomakos and Wang (2003) demonstrate that applying a square-root transformation to realised variance substantially mitigates skewness and excess kurtosis, producing volatility measures with more tractable distributional properties for subsequent econometric modelling. For this reason, this measure is adopted to operationalise realised volatility, adopting a short-term rolling window of five days to balance the trade-off between noise reduction and responsiveness to abrupt changes in volatility patterns.

$$RV_t = \sqrt{\frac{1}{4}\sum_{j=0}^{4}\left(r_{t-j} - \bar{r}_t^{(5)}\right)^2}, \qquad \bar{r}_t^{(5)} = \frac{1}{5}\sum_{j=0}^{4} r_{t-j}. \tag{1}$$

This definition matches the sample standard deviation used in `pandas` implementation (i.e., `rolling(5).std()` with `ddof=1`), hence the $1/(5-1)$ denominator.

A substantial empirical literature further documents relevant empirical volatility properties, which are useful to ground the design of additional features for the XGBoost algorithm discussed later on in the section. Volatility's stylised facts include persistence and clustering of large and small volatility episodes (Ding et al., 1993), asymmetric responses to negative and positive return innovations (Dzieliński et al., 2018), fat-tailed behaviour often associated with jump components (Cheng and Hung, 2011), mean reversion over longer horizons (Bali and Demirtas, 2008), and co-movement across assets during periods of common shocks (Baker and Wurgler, 2012). Together, these empirical regularities motivate the use of flexible volatility models and provide the theoretical justification for the additional volatility-related feature engineering introduced later in the calibration of the XGBoost model.

3.2. I-REC and energy markets

Despite the broad geographical coverage of the I-REC framework, only a limited subset of national markets has reached a level of maturity and liquidity sufficient to support the regular publication of daily price data. In addition, several of these markets have only begun reporting prices relatively recently, resulting in short and fragmented time series. To ensure the reliability of the empirical analysis, a minimum data-availability threshold is therefore imposed, retaining only those series with adequate historical depth. This screening step serves to mitigate calibration issues and reduce the risk of overfitting, particularly in the context of the XGBoost model, which can be sensitive to small-sample environments. Accordingly, the threshold is set at five hundred daily observations, in line with established applications of machine-learning techniques in energy and commodity markets (Wang et al., 2020; Shin and Woo, 2022; Kumar et al., 2023).

Following this selection procedure, only I-REC markets in Mexico, Turkey, Brazil and partially India satisfy the data requirements. Nevertheless, the empirical scope extends beyond these four markets in terms of the number of price series analysed. For each country, certificates associated with hydroelectric, solar, and wind generation are considered. Finally, as in many other similar markets, certificates have an expiry period, usually of twelve months, after which they can no longer be redeemed. Most trading platforms report hence separate prices for current-year and previous-year vintages, thereby aggregating in these two macro-categories all certificates with the same technology and country of origin. This yields a total of twenty-six distinct price series and corresponding volatility processes. Such cross-market, cross-technology, and cross-vintage variation provides a rich empirical setting in which to evaluate the robustness and generalisability of the proposed modelling approaches across heterogeneous geographical and technological contexts.

Figure 1 provides an overview of the realised volatility—as defined in subsection 3.1—observed across the various markets, technologies, and vintages. From the chart, it is evident that volatility displays substantial heterogeneity in terms of level, persistence, and clustering behaviour. This empirical diversity underscores the importance of employing flexible modelling frameworks capable of capturing a wide range of volatility dynamics across different segments of the I-REC market.

Given the dependence of certificate markets on the dynamics affecting the underlying elec-

Figure 1: Annualised realised volatility of selected I-REC markets.

tricity system, the inclusion of exogenous variables is essential in both the GARCH and XG-Boost modelling frameworks. Since certificate prices are influenced by many of the same structural factors that drive electricity prices (Hulshof et al., 2019), the electricity price literature provides a natural reference point for the selection of explanatory variables. In line with established evidence on the relevance of energy price indices for power markets (e.g. Woo et al. 2016; Peura and Bunn 2021), regional benchmark prices are selected according to each market's geographical and institutional context.

For the Latin American markets, West Texas Intermediate (WTI) crude oil and Henry Hub natural gas month-ahead forward prices are employed, reflecting both geographical proximity and strong financial linkages with the United States. In the case of Turkey, UK Brent crude oil, Dutch TTF natural gas month-ahead forward prices, and European Union Allowances year-ahead futures are adopted as representative European benchmarks. These indices are widely recognised as leading reference indices in respective regions and also serve as proxies for broader geopolitical and policy-related risks, providing an additional rationale for their inclusion as exogenous drivers of certificate price volatility (Liu et al., 2021b).

In addition to regional energy indices, national electricity spot prices are incorporated into the modelling framework of the I-REC markets–except for India, for which such data is not readily available. The models also include the daily ratio of electricity generated by a given technology to total power consumption, serving as a proxy for the relevance of that technology in meeting national demand.

Table 1: Exogenous Variables by Market

Market	Exogenous Variables
Brazil	Henry Hub M+1; WTI M+1; Electricity spot price; Hydropower share; Solar share; Wind share
Mexico	Henry Hub M+1; WTI M+1; Electricity spot price; Hydropower share; Solar share; Wind share
Turkey	TTF M+1; Brent M+1; EU Allowances Y+1; Electricity spot price; Hydropower share; Solar share; Wind share
India	JKM M+1; WTI M+1; Hydropower share

Prior to model estimation, both endogenous and exogenous series are transformed to facilitate numerical stability and comparability across specifications. In particular, the target volatility series are rescaled by a constant factor to mitigate convergence issues—especially relevant for GARCH-type models—without altering their informational content. All in-sample and out-of-sample predictions are subsequently rescaled back to the original units to ensure meaningful performance evaluation and direct comparability with the XGBoost forecasts.

Finally, exogenous predictors are standardised to account for differences in scale and dispersion across variables measured in heterogeneous units. This is particularly important to avoid the artificial dominance of variables with larger numerical magnitudes. For each feature j with raw input $x_{t,j}$ in the training sample, the empirical mean μ_j and standard deviation σ_j, computed from the training sample, are then applied to both training and out-of-sample data to prevent data leakage, and inputs are transformed according to

$$z_{t,j} = \frac{x_{t,j} - \mu_j}{\sigma_j}. \tag{2}$$

This standardisation ensures that model estimation is conducted on a variance-consistent basis and prevents predictors with larger numerical magnitudes from exerting disproportionate influence on the estimation and learning process.

3.3. GARCH Framework and Model Specifications

The analysis by Fagiani and Hakvoort (2014) remains one of the few contributions to address volatility in certificate markets from an econometric standpoint. Their work employs a GARCH model—a natural starting point given the extensive and well-documented success of GARCH-type specifications in characterising volatility dynamics across a broad range of asset classes within the financial economics literature. GARCH thus provides a well-established and theoretically grounded benchmark against which the performance of more sophisticated modelling approaches can be meaningfully evaluated.

In the standard GARCH(p, q) formulation of Bollerslev (1986), the return on an asset at time t is expressed as:

$$r_t = \mu + \varepsilon_t, \tag{3}$$

where μ denotes the unconditional mean return and ε_t is a zero-mean error term. The error term is further decomposed into a volatility component and a standardised innovation:

$$\varepsilon_t = \sigma_t e_t, \tag{4}$$

where e_t is an independent and identically distributed process with zero mean and unit variance, and σ_t is the conditional standard deviation. The conditional variance evolves according to:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{5}$$

where $\omega > 0$, $\alpha_i \geq 0$, and $\beta_j \geq 0$ are required for the conditional variance to remain strictly positive. A central advantage of this specification is the relaxation of the homoskedasticity assumption that constrains many linear time-series models. GARCH processes are particularly well suited to capturing volatility clustering—that is, the tendency for large (small) price movements to be followed by large (small) movements—a regularity frequently observed in financial and commodity markets, including those relevant to certificate trading.

Although the standard GARCH model has long served as the principal workhorse of conditional volatility modelling, several well-documented limitations have motivated the development of more flexible extensions. These are especially pertinent in the context of I-REC markets, where the extreme and heterogeneous volatility dynamics observed in the data may not be adequately captured by a symmetric specification. One of the most influential generalisations is the Exponential GARCH (EGARCH) model introduced by Nelson (1991). In its common decomposed form, the logarithm of the conditional variance is specified as:

$$\ln\left(\sigma_t^2\right) = \omega + \sum_{i=1}^{p} \alpha_i \left|z_{t-i}\right| + \sum_{i=1}^{o} \gamma_i \, z_{t-i} + \sum_{j=1}^{q} \beta_j \ln\left(\sigma_{t-j}^2\right), \tag{6}$$

where

$$z_t = \varepsilon_t / \sigma_t \tag{7}$$

denotes the standardised residual. This formulation introduces several noteworthy advantages over the standard GARCH specification. First, by modelling the logarithm of the conditional variance, it eliminates the need for non-negativity constraints on the parameters, as the exponential transformation ensures that σ_t^2 remains strictly positive regardless of the sign of the estimated coefficients. Second, the magnitude effects captured by α_i and the sign-dependent asymmetric effects governed by γ_i are cleanly separated, allowing the model to accommodate the well-documented tendency for negative shocks to exert a larger impact on volatility than positive shocks of comparable magnitude.

A further extension is the Threshold GARCH (TGARCH) model introduced by Zakoian (1994), which specifies volatility dynamics in terms of the conditional standard deviation rather than the variance. A TGARCH(p, o, q) model takes the form:

$$\sigma_t = \omega + \sum_{i=1}^{p} \alpha_i \left|\varepsilon_{t-i}\right| + \sum_{i=1}^{o} \gamma_i \left|\varepsilon_{t-i}\right| \mathbb{1}_{\{\varepsilon_{t-i}<0\}} + \sum_{j=1}^{q} \beta_j \sigma_{t-j}, \tag{8}$$

where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function, which takes the value one when the condition is satisfied and zero otherwise. This specification retains the capacity to model asymmetric responses to shocks, as in the EGARCH framework, while operating directly on the conditional standard deviation. In this formulation, the total impact of a negative shock of magnitude $|\varepsilon_{t-i}|$

on future volatility is given by $(\alpha_i + \gamma_i)|\varepsilon_{t-i}|$, whereas the impact of a positive shock of equal magnitude is captured by $\alpha_i|\varepsilon_{t-i}|$ alone. A positive estimate of γ_i thus implies that negative innovations have a disproportionately larger effect on volatility.

Finally, Baillie et al. (1996) introduce the Fractionally Integrated GARCH (FIGARCH) model, which is designed to accommodate long-memory behaviour in the conditional variance process. Unlike standard GARCH and its asymmetric extensions, in which the impact of past shocks decays at an exponential rate, the FIGARCH specification allows for a hyperbolic rate of decay, thereby capturing the slow, persistent mean reversion of volatility that is characteristic of many financial and commodity time series. The model is expressed as:

$$\sigma_t^2 = \omega + \beta(L)\sigma_t^2 + \left[1 - \beta(L) - \phi(L)(1 - L)^d\right] \varepsilon_t^2, \tag{9}$$

where L denotes the lag operator, $\beta(L)$ and $\phi(L)$ are lag polynomials, and $(1 - L)^d$ is the fractional differencing operator with memory parameter $0 < d < 1$. When $d = 0$, the model reduces to a standard short-memory GARCH, while with $d = 1$, it displays unit persistence. Intermediate values of d allow the model to capture the gradual, hyperbolic decay of autocorrelations in squared returns. FIGARCH is therefore particularly well suited for I-REC market segments exhibiting highly persistent volatility regimes that standard GARCH specifications fail to represent adequately.

Given the heterogeneous and often extreme volatility patterns observed across I-REC price series spanning different markets, technologies, and vintages, no single GARCH-type model can be expected to perform uniformly well across all segments of the data. To substantially extend the scope of the contribution of Fagiani and Hakvoort (2014)—both in geographical breadth and in methodological depth—the present study therefore proposes to employ all four of the aforementioned GARCH specifications as performance benchmarks against which more advanced modelling approaches are evaluated.

In addition to the volatility dynamics specified by the four model architectures, the GARCH framework is augmented with exogenous regressors to account for the influence of external market conditions on certificate price volatility. This is achieved through the GARCH-X extension, which incorporates external variables directly into the variance equation. For illustration, a GARCH-X(p, q) specification can be written as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + \sum_{k=1}^{K} \delta_k x_{t,k}, \tag{10}$$

where $x_{t,k}$ denotes the k-th exogenous variable observed at time t and δ_k is its associated coefficient. The same principle of exogenous augmentation is applied analogously to the EGARCH, TGARCH, and FIGARCH specifications. This extension enables the models to capture the influence of broader energy market conditions and technology-specific factors on the conditional volatility of I-REC prices, which is of particular relevance given the documented sensitivity of certificate markets to external drivers.

Across all GARCH-type specifications, the integers p, o, and q govern the order of the conditional volatility dynamics. Specifically, p denotes the number of lagged squared innovation terms (or, equivalently, lagged absolute innovations in the TGARCH case and lagged standardised residuals in the EGARCH case) entering the model; q represents the number of lagged conditional variance (or standard deviation) terms; and o captures the order of asymmetric or threshold effects associated with the sign of past shocks. For TGARCH, o is forcibly set to 1, while for all other models it is determined empirically.

Because GARCH order identification is inherently difficult in practice—standard information criteria can favour overly complex specifications or fail to recover the true data-generating process at higher lag orders(Naik et al., 2020)—an exhaustive grid search is conducted over a prespecified set of candidate (p, o, q) combinations. In particular, the search space is defined as $p, q \in \{0, \ldots, 6\}$ and $o \in \{0, 1\}$, and the preferred specification for each model class and market segment is selected on the basis of comparative model fit across this candidate pool. This approach ensures that the chosen lag structure is determined empirically rather than imposed *a priori*, thereby reducing the risk of model misspecification.

3.4. Extreme Gradient Boost

XGBoost, originally introduced by Chen and Guestrin (2016), is an optimised implementation of gradient-boosted regression trees. The algorithm constructs an additive ensemble model of the form:

$$\widehat{y}_t = \sum_{m=1}^{M} f_m(\mathbf{x}_t), \qquad f_m \in \mathcal{F}, \tag{11}$$

where each f_m denotes an individual regression tree, $\mathbf{x}_t$ is the vector of input features at time t, and $\mathcal{F}$ represents the functional space of regression trees.

Trees are added sequentially. At boosting round m, the regularised objective function to be minimised with respect to the new tree f_m consists of a training loss term and a regularisation component:

$$\mathcal{L}^{(m)} = \sum_{i=1}^{n} \ell\Big(y_i,\ \widehat{y}_i^{(m-1)} + f_m(\mathbf{x}_i)\Big) + \Omega(f_m) + \text{const}, \tag{12}$$

where $\ell(\cdot, \cdot)$ is a differentiable loss function, $\widehat{y}_i^{(m-1)} = \sum_{k=1}^{m-1} f_k(\mathbf{x}_i)$ is the prediction from the first $m-1$ trees, $\Omega(f_m)$ penalises model complexity, and the constant collects terms that do not depend on f_m.

A distinguishing feature of XGBoost relative to earlier gradient-boosting algorithms is the use of a second-order Taylor expansion of the loss function to approximate the objective at each boosting round. Dropping constant terms, this yields:

$$\widetilde{\mathcal{L}}^{(m)} \approx \sum_{i=1}^{n} \left[g_i\, f_m(\mathbf{x}_i) + \frac{1}{2}\, h_i\, f_m^2(\mathbf{x}_i) \right] + \Omega(f_m)\,, \tag{13}$$

where $g_i = \partial\, \ell(y_i,\, \widehat{y}_i^{(m-1)}) / \partial\, \widehat{y}_i^{(m-1)}$ and $h_i = \partial^2 \ell(y_i,\, \widehat{y}_i^{(m-1)}) / \partial\, (\widehat{y}_i^{(m-1)})^2$ denote the first- and second-order derivatives of the loss function with respect to the current prediction, respectively. This quadratic approximation enables efficient, closed-form computation of the optimal leaf weights and split gains at each node.

XGBoost possesses several attributes that make it well suited to the problem at hand. Its ensemble-tree architecture enables the approximation of complex nonlinear functional relationships, including threshold effects and higher-order interactions among lagged returns and exogenous predictors. Moreover, its explicit regularisation framework mitigates overfitting and supports strong out-of-sample generalisation even in moderately sized datasets. These properties make XGBoost a compelling alternative to GARCH-class models, particularly in settings where the structural form of volatility dynamics is too intricate to be adequately captured by parametric conditional-variance recursions alone.

In addition to the transformations of target and exogenous variables applied to the GARCH models, an extended set of volatility-related features informed by the empirical finance literature mentioned in subsection 3.1 is engineered to enhance the predictive capacity of the XGBoost model.

The feature set is designed to combine persistence, multi-scale structure, regime information, and exogenous drivers. First, persistence and long-memory effects are captured through aggregates of past volatility over heterogeneous horizons, following the Heterogeneous Autoregressive (HAR) framework, which provides a parsimonious yet effective representation of volatility dependence across short, medium, and long horizons (Andersen et al., 2003; Corsi, 2009). In this context, realised volatility is denoted by RV_t, and the HAR components are specified as:

$$HAR_t^{(d)} = RV_{t-1}, \tag{14}$$

$$HAR_t^{(w)} = \frac{1}{5}\sum_{j=1}^{5} RV_{t-j}, \tag{15}$$

$$HAR_t^{(m)} = \frac{1}{22}\sum_{j=1}^{22} RV_{t-j}, \tag{16}$$

where d, w, and m denote daily, weekly (5 trading days), and monthly (22 trading days) horizons, respectively.The adoption of HAR components is further motivated by evidence that gradient-boosted ensembles with multi-horizon HAR-style features can outperform both linear HAR specifications and neural network alternatives by flexibly accommodating nonlinear effects and complex feature interactions (Hansen et al., 2012; Teller, 2022).

To complement these level-based persistence effects, finite-difference measures are introduced to capture short-run dynamics beyond what the HAR components alone provide (Corsi, 2009). Within the XGBoost framework, these enter as additional covariates that allow tree splits to condition on recent directional changes in realised volatility and on reversals in those changes. Using only information available up to $t-1$, a weekly change measure and a short-run acceleration measure are defined as:

$$\Delta RV_t^{(5)} = RV_{t-1} - RV_{t-6}, \tag{17}$$

$$\mathrm{Acc}(RV)_t = (RV_{t-1} - RV_{t-2}) - (RV_{t-2} - RV_{t-3}) = RV_{t-1} - 2\,RV_{t-2} + RV_{t-3}. \tag{18}$$

Additionally, a high-volatility regime indicator is constructed to serve as an analogue to threshold specifications, enabling distinct dynamics across regimes and potentially improving forecasts by accommodating discrete shifts in both the level and persistence of conditional variance (Hamilton and Susmel, 1994). A binary indicator $\mathcal{R}_t$ is defined to flag a regime change when the most recent realised volatility exceeds a robust trailing baseline:

$$\mathcal{R}_t = \mathbf{1}\{RV_{t-1} > \text{median}(RV_{t-60}, \ldots, RV_{t-1})\}. \tag{19}$$

Interactions between $\mathcal{R}_t$, the HAR components, and exogenous drivers can thus be captured through nonlinear tree splits, yielding regime-dependent responses without the imposition of an explicit parametric switching structure.

Finally, calendar covariates and energy exogenous drivers are included to match the information provided to GARCH-class models, thus extending the feature set beyond endogenous volatility dynamics. This provides a rich system that allows for a combined modelling of typical inherent volatility structures as well as interpretable external factors.

3.5. Calibration and Performance Evaluation

The calibration of the two model classes is performed separately for each market, technology, and vintage, yielding a total of twenty-six model comparisons. For each series, the same training and test split is applied to ensure comparability across models. The training sample comprises the first 70% of observations, while the remaining 30% are reserved for out-of-sample evaluation. This division allows for a robust assessment of predictive performance while mitigating overfitting risks.

As discussed in subsection 3.3, the optimal specification of each GARCH model is selected through an exhaustive search over a predefined grid of candidate orders p, o, q, with the best model chosen on the basis of in-sample fit. The search space is defined with p and q ranging from 0 to 6 and o taking values of 0 or 1. This grid is further extended through the use of alternative distributional assumptions for the standardised innovations—namely normal, Student's t, and generalised error distribution (GED)—to allow greater flexibility in the modelling of heavy-tailed behaviour. The total number of combinations explored per GARCH class and product can reach two hundred, depending on the specific restrictions applicable to each model. The performance of each candidate specification is evaluated against the observed training data using the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), both of which are widely employed in the quantitative literature for parameter estimation and forecast evaluation. Figure 2 summarises the specifications of the best-performing GARCH models across markets, technologies, and vintages, as selected by in-sample MAE and RMSE.

The results lead to several interesting observations. First, it is apparent that there is considerable heterogeneity in the best p and q orders across products. This provides empirical support to the observation made by Aras (2021), who warn against assuming the superiority of simpler GARCH model structures, and reinforces the relevance of the exhaustive grid-search approach adopted here. Second, there is consistency in selected parameters and distributions across MAE and RMSE, with the only major exception being the complete absence of normal distributions among MAE-based best-performing models. This leads to a third observation, namely the predominance of GED in fitting the empirical distribution of many of the markets under consideration. The predominance of GED could be interpreted in light of the inherent dynamics of I-REC markets, where infrequent but concentrated trading and policy-driven shocks generate asymmetric tail behaviour that symmetric distributions cannot adequately represent.

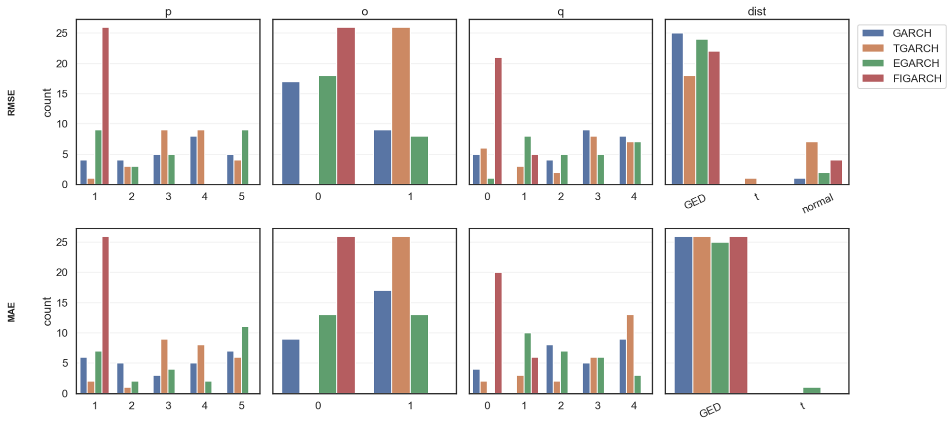

Figure 2: p, o, q orders and distributional assumptions of the best-performing GARCH specifications by MAE and RMSE.

For XGBoost, the feature vector combines a parsimonious set of volatility summaries with contemporaneous exogenous drivers, as introduced in subsection 3.4. Specifically, the engineered features include: (i) HAR components based on daily, weekly (5-day), and monthly (22-day) averages of lagged volatility; (ii) local dynamics indicators capturing changes over a weekly horizon and short-run acceleration; and (iii) a binary high-volatility regime indicator, defined by whether lagged volatility exceeds its trailing historical median, serving as a pragmatic analogue to regime-switching dynamics. For all model classes, seasonality is captured via month-of-year dummy variables, and exogenous predictors enter contemporaneously. This construction yields a fully observable information set that retains persistence, regime dependence, and external drivers while keeping dimensionality limited.

Model calibration consists of fitting a gradient-boosted tree ensemble on the training sample for a fixed number of boosting iterations (Chen and Guestrin, 2016). Concretely, the model is trained under a squared-error regression objective using histogram-based tree construction, shallow trees (maximum depth of 3), and a set of regularisation and subsampling controls (minimum child weight of 10, row subsample and column subsample ratios of 0.7, together with additional ℓ_1 / ℓ_2 penalties and a positive minimum split-loss penalty). These specifications are not the product of performance-oriented tuning, but rather they are set *a priori* to enforce strong regularisation on a model class whose high-dimensional and continuous hyperparameter space makes unconstrained tuning—particularly on moderately-sized samples—prone to configurations that overfit the training data. The chosen configuration is therefore deliberately restrictive, designed to maximise out-of-sample generalisation and reduce potential overfitting problems.

Following calibration, a *post-hoc* interpretability analysis is performed using SHAP (SHapley Additive exPlanations) values (Lundberg and Lee, 2017) to examine the contribution of individual predictors and their relative importance in driving the model's forecasts. SHAP decomposes the prediction for each observation into additive contributions from each feature, capturing both the magnitude and direction of their impact on model output. This enables a detailed understanding of how different predictors influence volatility forecasts across markets, technologies, and vintages (Nohara et al., 2021). By comparing SHAP values across models and products, insights can be gained into the specific drivers of volatility in I-REC markets. Given the large number of markets and variables, the analysis focuses on the five most influential features as identified by mean absolute SHAP value.

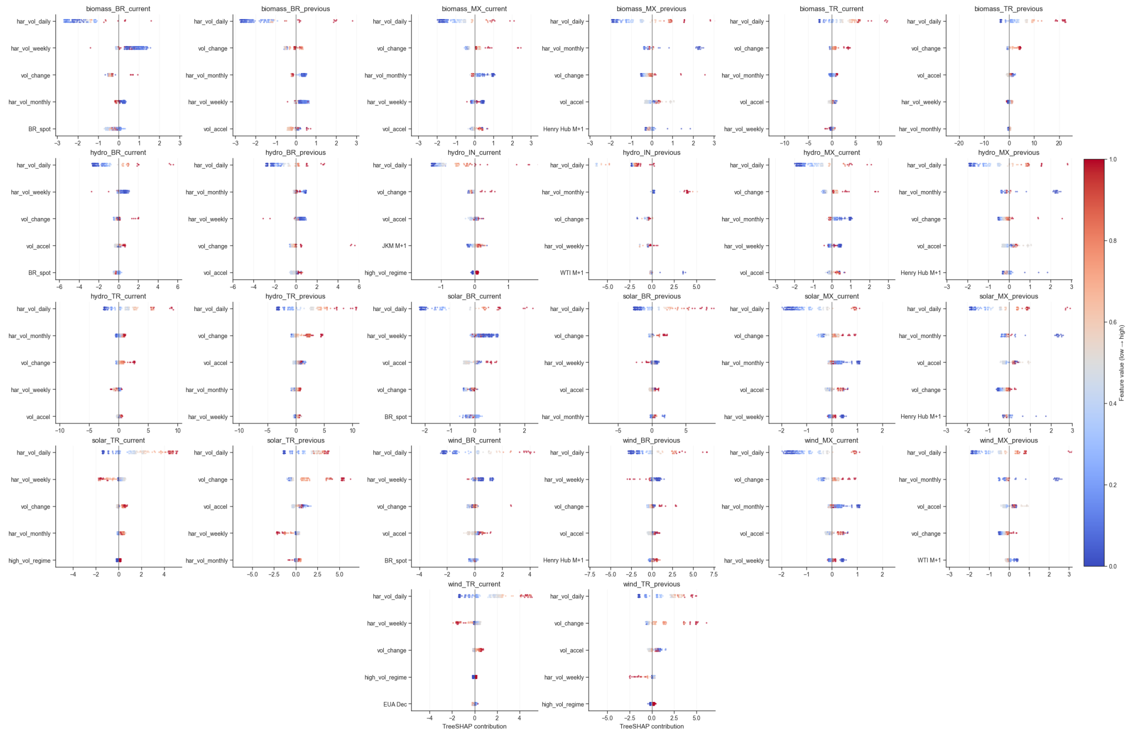

Figure 3: SHAP values for the five most influential features in XGBoost models by market.

The SHAP analysis reports both the magnitude and direction of each feature's contribution to the model's predictions. Features are ranked along the vertical axis in descending order of mean absolute SHAP value, such that those at the top exert the greatest influence on predictive performance. The horizontal axis quantifies the contribution of each observation as well as its sign. The colour of each point encodes the corresponding feature value—red denoting high values and blue low values—thereby enabling a joint reading of directionality and feature magnitude. For instance, a red dot positioned to the right of the origin indicates that an elevated feature value raises predicted volatility, whilst a blue dot to the left indicates that a low feature value depresses it. Conversely, a red dot to the left of the origin—observed for certain features in specific certificate markets—signals that high feature values suppress predicted volatility, a pattern consistent with mean-reverting or regime-dependent dynamics.

A first inspection of the chart reveals that the predominant drivers of I-REC market volatility are endogenous clustering variables rather than exogenous energy-market benchmarks. Across all technologies and geographies considered, the three HAR components consistently occupy the uppermost positions in the feature importance ranking. This finding is strongly consistent with the theoretical underpinnings of the HAR framework, which posits that market participants operating at distinct investment horizons collectively give rise to long-memory-like behaviour in volatility, with each temporal component capturing a different stratum of their activity (Corsi, 2009).

The spread of SHAP values for the daily realised volatility component is substantially wider than that of its weekly and monthly counterparts across most subplots, consistent with the well-established volatility clustering phenomenon. The colour gradients observed for these features—where high values (red) are predominantly associated with large positive SHAP contributions and vice versa—indicate a robust positive relationship between elevated past realised volatility and higher predicted volatility, as expected under a volatility persistence framework.

Exogenous variables appear sporadically across the subplots. Their limited and heterogeneous relevance across certificate markets may be interpreted as an empirical indication that I-REC volatility dynamics are largely self-contained and not systematically driven by either broad energy commodity markets or local electricity spot prices. This is a noteworthy empirical finding, as it suggests that the price discovery process in I-REC markets is still predominantly governed by trading activity specific to each certificate segment, rather than by cross-market transmission channels that are well documented in more established certificate markets.

A further notable feature of the results is the cross-market heterogeneity in feature importance rankings. Whilst the dominance of the HAR structure is a common thread, the relative ordering and SHAP magnitudes of secondary features differ appreciably across technologies and jurisdictions. This heterogeneity confirms that the volatility-generating process in I-REC markets is not homogeneous, and that the same feature may carry markedly different informational content depending on the underlying technology, geographical context, and vintage of the certificate.

4. Results and Discussion

The performance of the GARCH and XGBoost models is evaluated using out-of-sample MAE and RMSE, which provide quantitative measures of predictive accuracy. Results are compared across markets, technologies, and vintages to assess the robustness of each modelling approach in capturing the volatility dynamics of I-REC prices. The full numerical results are reported in Appendix A.

Figure 4 presents comparative in-sample calibration results by product.

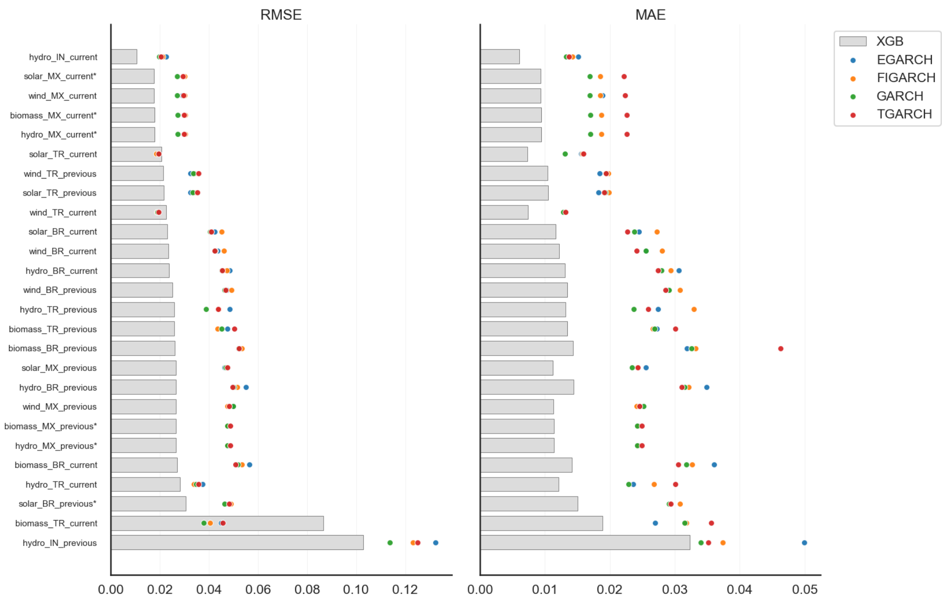

Figure 4: In-sample performance comparison between GARCH and XGBoost models by product.

The plot reports all markets analysed, with coloured dots representing the metrics for the best-performing GARCH model within each class and the grey bar denoting the performance of the XGBoost model. Upon initial inspection, it was observed that EGARCH entirely fails to yield meaningful results in a number of cases, predominantly concentrated on Mexican products. Where these values exceed the scale of the plot, they are omitted from the figure given their limited analytical relevance. The corresponding markets are indicated with an asterisk on the vertical axis.

From an overall assessment of the in-sample results, it is apparent that the XGBoost model outperforms the GARCH models in nearly every market. Whilst this does not in itself constitute evidence of superior forecasting accuracy, it provides a strong indication that XGBoost is better able to capture the complex, nonlinear relationships that characterise volatility dynamics in I-REC markets during the calibration process.

XGBoost performance proves considerably more consistent than that of the GARCH models, whose MAE and RMSE increase markedly across different markets, technologies, and vintages. Nevertheless, XGBoost also registers a notable deterioration in performance for current-year vintage Turkish biomass and previous-year vintage Indian hydro certificates. This may be attributable to the particularly extreme volatility patterns observed in these product. Figure 5 displays a comparative violin plot of the volatility distribution for the two aforementioned markets alongside Indian hydro current-year vintage certificates, which is the product yielding the best MAE and RMSE scores for XGBoost models. Turkish biomass current-year and Indian hydro previous-year certificates exhibit visibly fatter tails and a higher frequency of extreme values, which might account for the decline in XGBoost performance in this segment.

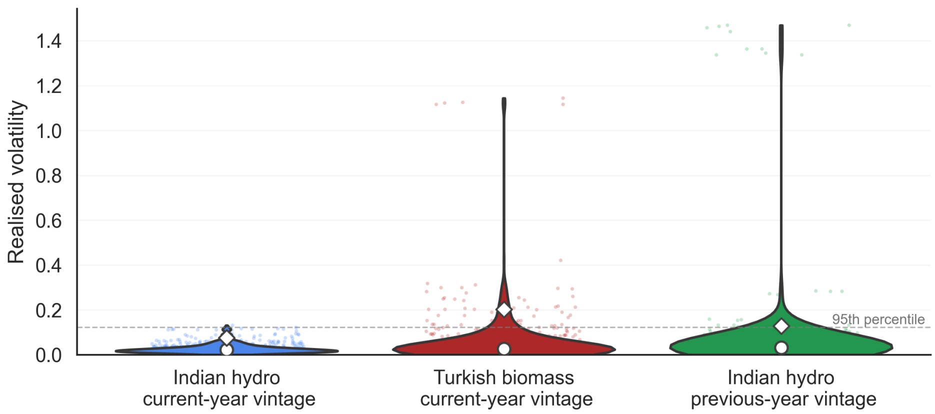

Figure 5: Violin plot for the in-sample volatility distribution of best and worst-performing products.

Closely comparable conclusions can be drawn from the out-of-sample performance comparison, reported in Figure 6. First, the superior performance of the XGBoost model is confirmed also out-of-sample across all markets, technologies, and vintages, with the machine-learning algorithm outperforming the GARCH benchmarks in 96% to 100% of cases. Second, a comparison between in-sample and out-of-sample metrics indicates that the XGBoost model maintains a consistent level of accuracy across both samples, suggesting that the model does not overfit the training data and generalises effectively to unseen observations. Third, out-of-sample performance tends to be more stable than in-sample performance, with no sharp deterioration observed for any individual product. This could mean that the drop in MAE and RMSE performance observed in the current-year Turkish biomass and previous-year Indian hydro might have been limited to a certain past time window, without extensions in the out-of-sample dataset.

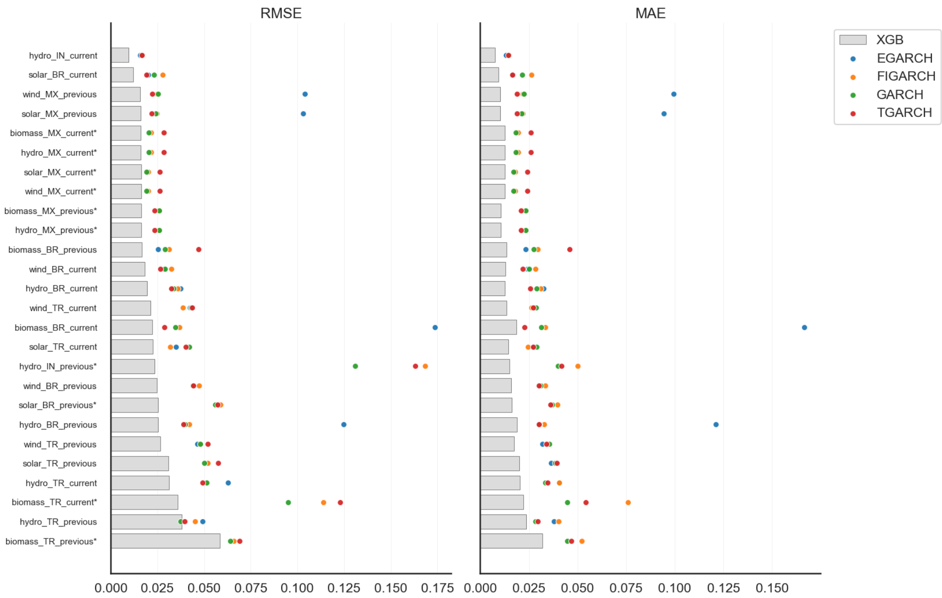

Figure 6: Out-of-sample performance comparison between GARCH and XGBoost models by product.

A further noteworthy observation concerns the absence of a clear pattern in the relative performance of individual GARCH specifications. Whilst the standard GARCH model outperforms all other variants in absolute terms, it does so in only half of the markets considered. Moreover, the performance of TGARCH and FIGARCH varies considerably depending on whether MAE or RMSE is used as the evaluation criterion. The practical consequence is that it is difficult to identify a single GARCH specification that can be reliably and consistently applied across the full range of markets, technologies, and vintages. This conclusion extends to the out-of-sample results, making it challenging for practitioners and policymakers to select *a priori* a GARCH model that is expected to perform well across the different products available within a given national certificate market.

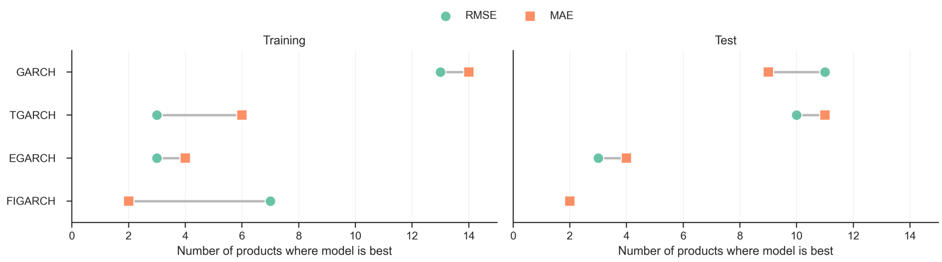

Figure 7: In-sample and out-of-sample comparison of best-model frequency across GARCH.

5. Conclusion and future areas of work

This paper sets out to investigate various modelling options for volatility in I-REC markets, providing a comparison between an extended benchmark of GARCH-class models against the adoption of an Extreme Gradient Boosting architecture. By assembling a systematic dataset of I-REC prices spanning four national markets, four generation technologies, and two vintage classes, the analysis provides an empirical foundation for a market segment that has grown rapidly yet received almost no attention from the forecasting literature.

Three findings deserve emphasis. First, I-REC price series exhibit pronounced heterogeneity in their volatility dynamics across countries, technologies, and vintages. No single GARCH specification—whether symmetric, asymmetric, or power-transformed—dominates uniformly, and the best-performing parametric model varies considerably across the twenty-six series examined. This heterogeneity cautions against importing modelling conventions from more mature certificate markets such as the European Guarantees of Origin system without careful re-evaluation. Second, XGBoost systematically outperforms the best-fitting GARCH benchmark on both MAE and RMSE criteria, proving consistently robust through in and out-of-sample modelling and across virtually all series. Third, SHAP-based decompositions reveal that the drivers of I-REC volatility are market-specific: fossil-fuel reference prices, electricity spot markets, and carbon-related indicators enter with varying importance depending on the national and technological context, reinforcing the view that I-REC markets do not converge towards a single homogeneous price-formation behaviour.

These results carry practical implications for at least three audiences. For corporate buyers managing renewable-energy procurement portfolios, the evidence suggests that hedging strategies calibrated on standard GARCH forecasts may systematically understate or mischaracterise risk, and that ensemble-based alternatives could provide a credible alternative. For I-REC market operators and regulators, the cross-country heterogeneity documented here highlights the extent to which local market structure—grid composition, regulatory frameworks, and the depth of secondary trading—shapes price behaviour, information that is relevant to ongoing efforts to harmonise tracking standards across jurisdictions. For researchers, the paper demonstrates that the I-REC market, despite its relative opacity, generates price signals that are sufficiently structured to support systematic quantitative analysis.

Whilst the paper advances the understanding of volatility in I-REC markets, as well as the application of econometric and machine-learning models to this domain, several avenues for future research remain open. First, subsequent work could explore the inclusion of additional exogenous variables, such as weather conditions, which are known to influence renewable energy generation and could therefore affect certificate prices and their volatility. By introducing direct measures of some of the underlying drivers of production—particularly for solar and wind technologies—additional dynamics could potentially be uncovered, with possible marginal gains in forecasting performance. Second, future research could extend the analysis to other geographies, including markets in South-East Asia, where I-REC markets are developing rapidly and may exhibit distinct volatility patterns owing to differences in market structure and regulatory frameworks. Third, further work could explore the application of alternative machine-learning models. Whilst markets will need to accumulate substantially longer price histories before more data-intensive approaches such as deep learning can be meaningfully deployed, other methods—such as random forests and support vector machines—could be examined to test the robustness of the results obtained with XGBoost and to provide additional insights into the drivers of volatility in I-REC markets.

A. Appendix

Table 2: Model Performance Comparison – Brazil

Set	Metric	Model	Biomass *Curr.*	Biomass *Prev.*	Hydro *Curr.*	Hydro *Prev.*	Solar *Curr.*	Solar *Prev.*	Wind *Curr.*	Wind *Prev.*
Training										
	MAE	EGARCH	0.0361	0.0318	0.0307	0.0349	0.0244	–	0.0255	0.0289
		FIGARCH	0.0327	0.0332	0.0293	0.0321	0.0273	0.0308	0.0281	0.0308
		GARCH	0.0318	0.0326	0.0279	0.0315	0.0238	0.0291	0.0256	0.0291
		TGARCH	0.0305	0.0463	0.0275	0.0310	0.0227	0.0293	0.0241	0.0286
		XGB	**0.0141**	**0.0143**	**0.0131**	**0.0144**	**0.0117**	**0.0151**	**0.0122**	**0.0135**
	RMSE	EGARCH	0.0564	0.0522	0.0484	0.0550	0.0423	–	0.0435	0.0472
		FIGARCH	0.0534	0.0534	0.0472	0.0515	0.0452	0.0491	0.0462	0.0491
		GARCH	0.0517	0.0519	0.0451	0.0500	0.0404	0.0463	0.0426	0.0463
		TGARCH	0.0508	0.0523	0.0453	0.0497	0.0409	0.0483	0.0424	0.0467
		XGB	**0.0271**	**0.0261**	**0.0238**	**0.0266**	**0.0231**	**0.0306**	**0.0236**	**0.0251**
Test										
	MAE	EGARCH	0.1666	0.0232	0.0326	0.1211	0.0169	–	0.0243	0.0304
		FIGARCH	0.0335	0.0297	0.0311	0.0327	0.0263	0.0395	0.0283	0.0336
		GARCH	0.0313	0.0275	0.0289	0.0304	0.0216	0.0369	0.0253	0.0310
		TGARCH	0.0229	0.0460	0.0259	0.0301	0.0166	0.0360	0.0220	0.0301
		XGB	**0.0186**	**0.0136**	**0.0128**	**0.0188**	**0.0093**	**0.0162**	**0.0128**	**0.0160**
	RMSE	EGARCH	0.1737	0.0253	0.0373	0.1248	0.0201	–	0.0284	0.0442
		FIGARCH	0.0367	0.0311	0.0359	0.0419	0.0277	0.0589	0.0323	0.0474
		GARCH	0.0347	0.0291	0.0335	0.0399	0.0232	0.0561	0.0289	0.0438
		TGARCH	0.0287	0.0469	0.0323	0.0391	0.0192	0.0574	0.0266	0.0441
		XGB	**0.0223**	**0.0166**	**0.0194**	**0.0254**	**0.0121**	**0.0253**	**0.0182**	**0.0246**

Table 3: Model Performance Comparison – India

Set	Metric	Model	**Hydro** *Curr.*	**Hydro** *Prev.*
Training				
	MAE	EGARCH	0.0152	0.0500
		FIGARCH	0.0143	0.0374
		GARCH	0.0132	0.0340
		TGARCH	0.0137	0.0352
		XGB	**0.0061**	**0.0323**
	RMSE	EGARCH	0.0226	0.1322
		FIGARCH	0.0210	0.1232
		GARCH	0.0198	0.1137
		TGARCH	0.0205	0.1249
		XGB	**0.0105**	**0.1028**
Test				
	MAE	EGARCH	0.0133	–
		FIGARCH	0.0147	0.0500
		GARCH	0.0147	0.0400
		TGARCH	0.0145	0.0417
		XGB	**0.0075**	**0.0150**
	RMSE	EGARCH	0.0157	–
		FIGARCH	0.0169	0.1685
		GARCH	0.0168	0.1308
		TGARCH	0.0165	0.1633
		XGB	**0.0096**	**0.0235**

Table 4: Model Performance Comparison – Mexico

Set	Metric	Model	Biomass *Curr.*	Biomass *Prev.*	Hydro *Curr.*	Hydro *Prev.*	Solar *Curr.*	Solar *Prev.*	Wind *Curr.*	Wind *Prev.*
Training										
		EGARCH	–	–	–	–	–	0.0256	0.0189	0.0246
		FIGARCH	0.0187	0.0244	0.0187	0.0244	0.0185	0.0242	0.0185	0.0241
	MAE	GARCH	0.0170	0.0242	0.0170	0.0242	0.0169	0.0234	0.0170	0.0252
		TGARCH	0.0226	0.0249	0.0226	0.0249	0.0222	0.0243	0.0223	0.0245
		XGB	**0.0095**	**0.0114**	**0.0095**	**0.0114**	**0.0094**	**0.0112**	**0.0094**	**0.0113**
		EGARCH	–	–	–	–	–	0.0464	0.0292	0.0498
		FIGARCH	0.0302	0.0475	0.0302	0.0475	0.0302	0.0474	0.0302	0.0475
	RMSE	GARCH	0.0272	0.0475	0.0272	0.0475	0.0271	0.0469	0.0271	0.0498
		TGARCH	0.0298	0.0487	0.0298	0.0487	0.0294	0.0476	0.0296	0.0483
		XGB	**0.0177**	**0.0267**	**0.0177**	**0.0267**	**0.0175**	**0.0265**	**0.0175**	**0.0267**
Test										
		EGARCH	–	–	–	–	–	0.0942	–	0.0992
		FIGARCH	0.0194	0.0237	0.0194	0.0237	0.0181	0.0218	0.0181	0.0218
	MAE	GARCH	0.0183	0.0235	0.0183	0.0235	0.0171	0.0214	0.0171	0.0226
		TGARCH	0.0259	0.0208	0.0259	0.0208	0.0243	0.0189	0.0241	0.0189
		XGB	**0.0126**	**0.0107**	**0.0126**	**0.0107**	**0.0127**	**0.0104**	**0.0127**	**0.0103**
		EGARCH	–	–	–	–	–	0.1031	–	0.1039
		FIGARCH	0.0217	0.0260	0.0217	0.0260	0.0202	0.0246	0.0202	0.0246
	RMSE	GARCH	0.0203	0.0259	0.0203	0.0259	0.0190	0.0241	0.0190	0.0254
		TGARCH	0.0283	0.0235	0.0283	0.0235	0.0264	0.0220	0.0262	0.0222
		XGB	**0.0159**	**0.0164**	**0.0159**	**0.0164**	**0.0163**	**0.0159**	**0.0163**	**0.0158**

Table 5: Model Performance Comparison – Turkey

Set	Metric	Model	Biomass Curr.	Biomass Prev.	Hydro Curr.	Hydro Prev.	Solar Curr.	Solar Prev.	Wind Curr.	Wind Prev.
Training										
	MAE	EGARCH	0.0270	0.0273	0.0236	0.0274	0.0156	0.0183	0.0132	0.0184
		FIGARCH	0.0317	0.0266	0.0268	0.0329	0.0157	0.0198	0.0131	0.0198
		GARCH	0.0315	0.0269	0.0228	0.0237	0.0130	0.0193	0.0128	0.0194
		TGARCH	0.0356	0.0300	0.0301	0.0259	0.0160	0.0191	0.0132	0.0194
		XGB	**0.0189**	**0.0134**	**0.0121**	**0.0132**	**0.0073**	**0.0105**	**0.0074**	**0.0104**
	RMSE	EGARCH	0.0448	0.0476	0.0374	0.0485	0.0196	0.0324	0.0194	0.0324
		FIGARCH	0.0405	0.0435	0.0338	0.0439	**0.0185**	0.0340	**0.0187**	0.0341
		GARCH	**0.0378**	0.0453	0.0348	0.0388	0.0196	0.0334	0.0191	0.0337
		TGARCH	0.0455	0.0504	0.0358	0.0438	0.0196	0.0353	0.0195	0.0357
		XGB	0.0867	**0.0259**	**0.0282**	**0.0258**	0.0207	**0.0216**	0.0225	**0.0213**
Test										
	MAE	EGARCH	–	–	0.0405	0.0379	0.0276	0.0365	0.0275	0.0321
		FIGARCH	0.0760	0.0521	0.0406	0.0403	0.0245	0.0392	0.0264	0.0346
		GARCH	0.0447	0.0448	0.0334	0.0284	0.0290	0.0385	0.0288	0.0355
		TGARCH	0.0541	0.0468	0.0348	0.0296	0.0271	0.0393	0.0272	0.0340
		XGB	**0.0222**	**0.0320**	**0.0202**	**0.0235**	**0.0144**	**0.0202**	**0.0134**	**0.0175**
	RMSE	EGARCH	–	–	0.0627	0.0492	0.0348	0.0521	0.0423	0.0465
		FIGARCH	0.1139	0.0659	0.0515	0.0452	0.0318	0.0518	0.0388	0.0479
		GARCH	0.0950	0.0641	0.0513	**0.0373**	0.0421	0.0501	0.0430	0.0478
		TGARCH	0.1228	0.0691	0.0490	0.0396	0.0401	0.0577	0.0436	0.0519
		XGB	**0.0358**	**0.0583**	**0.0311**	0.0380	**0.0226**	**0.0310**	**0.0214**	**0.0266**

References

Adekoya, O. B., Oliyide, J. A., and Noman, A. (2021). The volatility connectedness of the eu carbon market with commodity and financial markets in time-and frequency-domain: The role of the us economic policy uncertainty. *Resources Policy*, 74:102252.

Alharbey, M. and Ben-Salha, O. (2024). Does climate policy uncertainty predict renewable energy stocks? a quantile-based (a) symmetric causality analysis. *Energy Strategy Reviews*, 54:101465.

Alkathery, M. A. and Chaudhuri, K. (2021). Co-movement between oil price, co2 emission, renewable energy and energy equities: evidence from gcc countries. *Journal of Environmental Management*, 297:113350.

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review*, pages 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1):43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

Aras, S. (2021). Stacking hybrid garch models for forecasting bitcoin volatility. *Expert Systems with Applications*, 174:114747.

Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30.

Baker, M. and Wurgler, J. (2012). Comovement and predictability relationships between bonds and the cross-section of stocks. *The Review of Asset Pricing Studies*, 2(1):57–87.

Bali, T. G. and Demirtas, K. O. (2008). Testing mean reversion in financial market volatility: Evidence from s&p 500 index futures. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 28(1):1–33.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17(5):457–477.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, San Francisco, California, USA. ACM.

Cheng, W.-H. and Hung, J.-C. (2011). Skewness and leptokurtosis in garch-typed var estimation of petroleum and metal asset returns. *Journal of Empirical Finance*, 18(1):160–173.

Chevallier, J. (2011). Nonparametric modeling of carbon prices. *Energy Economics*, 33(6):1267–1282.

Ciarreta, A., Espinosa, M. P., and Pizarro-Irizar, C. (2017). Optimal regulation of renewable energy: A comparison of feed-in tariffs and tradable green certificates in the spanish electricity system. *Energy economics*, 67:387–399.

Ciarreta, A., Pizarro-Irizar, C., and Zarraga, A. (2020). Renewable energy regulation and structural breaks: An empirical analysis of spanish electricity price volatility. *Energy Economics*, 88:104749.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Ding, S., Cui, T., and Zhang, Y. (2022). Futures volatility forecasting based on big data analytics with incorporating an order imbalance effect. *International Review of Financial Analysis*, 83:102255.

Ding, Z., Granger, C. W., and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance*, 1(1):83–106.

Dong, F., Shi, M., and Kang, K. (2025). Review of existing policies and prospects for green power and green certificates in a dual-carbon context. *Energy for Sustainable Development*, 88:101819.

Dzieliński, M., Rieger, M. O., and Talpsepp, T. (2018). Asymmetric attention and volatility asymmetry. *Journal of Empirical Finance*, 45:59–67.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Fagiani, R., Barquín, J., and Hakvoort, R. (2013). Risk-based assessment of the cost-efficiency and the effectivity of renewable energy support schemes: Certificate markets versus feed-in tariffs. *Energy policy*, 55:648–661.

Fagiani, R. and Hakvoort, R. (2014). The role of regulatory uncertainty in certificate markets: A case study of the swedish/norwegian market. *Energy Policy*, 65:608–618.

Hamilton, J. D. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64(1–2):307–333.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized garch: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

Hulshof, D., Jepma, C., and Mulder, M. (2019). Performance of markets for european renewable energy certificates. *Energy Policy*, 128:697–710.

Husain, S., Sohag, K., and Wu, Y. (2024). The responsiveness of renewable energy production to geopolitical risks, oil market instability and economic policy uncertainty: Evidence from united states. *Journal of Environmental Management*, 350:119647.

Iqbal, N., Naeem, M. A., and Suleman, M. T. (2022). Quantifying the asymmetric spillovers in sustainable investments. *Journal of International Financial Markets, Institutions and Money*, 77:101480.

Irfan, M. (2021). Integration between electricity and renewable energy certificate (rec) markets: factors influencing the solar and non-solar rec in india. *Renewable Energy*, 179:65–74.

Jia, H., Liu, D., Zhuo, G., Feng, Y., and Han, Y. (2023). Review for global practices of renewable energy certificates. In *2023 8th Asia Conference on Power and Electrical Engineering (ACPEE)*, pages 642–646. IEEE.

Kriegler, E., Weyant, J., Blanford, G., Krey, V., Clarke, L., Edmonds, J., and Vuuren, D. (2014). The role of technology for achieving climate policy objectives: Overview of the emf 27 study on global technology and climate policy strategies. *Climatic Change*, 123(3):353–367.

Kumar, V., Kedam, N., Sharma, K., Mehta, D., and Caloiero, T. (2023). Advanced machine learning techniques to improve hydrological prediction: A comparative analysis of streamflow prediction models. *Water*, 15(14).

Lee, J. W., Choi, E. J., Jeong, M. J., Moragriega, R. C., Zaragoza, P. G., and Kim, S. W. (2022). Optimal sourcing strategy for enterprises to achieve 100% renewable energy. *Energy Reports*, 8:14865–14874.

Liang, C., Umar, M., Ma, F., and Huynh, T. L. (2022). Climate policy uncertainty and world renewable energy index volatility forecasting. *Technological Forecasting and Social Change*, 182:121810.

Liu, J., Zhang, Z., Yan, L., and Wen, F. (2021a). Forecasting the volatility of eua futures with economic policy uncertainty using the garch-midas model. *Financial Innovation*, 7:1–19.

Liu, Y., Han, L., and Xu, Y. (2021b). The impact of geopolitical uncertainty on energy volatility. *International Review of Financial Analysis*.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

Mukherjee, I. and Goswami, B. (2017). The volatility of returns from commodity futures: evidence from India. *Financial Innovation*, 3(1):15.

Naik, N., Mohana, B. R., and Jha, R. A. (2020). Garch-model identification based on performance of information criteria. In *Procedia Computer Science*, volume 171, pages 1935–1942. Elsevier.

Nandagopal, P. and Devaraja, T. S. (2018). International rec standard for empowering renewable power generation. *International journal of current research and academic review*, 6:13–16.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.

Niu, H. and Liu, T. (2024). Forecasting the volatility of european union allowance futures with macroeconomic variables using the gjr-garch-midas model. *Empirical Economics*, 67(1):75–96.

Nohara, Y., Matsumoto, K., Soejima, H., and Nakashima, N. (2021). Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Computer methods and programs in biomedicine*, 214.

Nurdiawati, A. and Urban, F. (2021). Towards deep decarbonisation of energy-intensive industries: A review of current status, technologies and policies. *Energies*, 14(9).

O'Shaughnessy, E., Heeter, J., Shah, C., and Koebrich, S. (2021). Corporate acceleration of the renewable energy transition and implications for electric grids. *Renewable and Sustainable Energy Reviews*, 146:111160.

Peura, H. and Bunn, D. (2021). Renewable power and electricity prices: The impact of forward markets. *Management Science*, 67:4772–4788.

Raza, S. A., Khan, K. A., Benkraiem, R., and Guesmi, K. (2024). The importance of climate policy uncertainty in forecasting the green, clean and sustainable financial markets volatility. *International Review of Financial Analysis*, 91:102984.

Samuelson, P. A. (1965). Rational theory of warrant pricing. In *Henry P. McKean Jr. Selecta*, pages 195–232. Springer.

Shin, S. and Woo, H. (2022). Energy consumption forecasting in korea using machine learning algorithms. *Energies*, 15(13).

Tang, Y., Chen, X. H., Sarker, P. K., and Baroudi, S. (2023). Asymmetric effects of geopolitical risks and uncertainties on green bond markets. *Technological Forecasting and Social Change*, 189:122348.

Teller, D. (2022). Short- to long-term realized volatility forecasting using extreme gradient boosting. *Available at SSRN 4267541.*

Thomakos, D. D. and Wang, T. (2003). Realized volatility in the futures markets. *Journal of Empirical Finance*, 10(3):321–353.

Uddin, M. A., Chang, B. H., Aldawsari, S. H., and Li, R. (2025). The interplay between green finance, policy uncertainty and carbon market volatility: A time frequency approach. *Sustainability*, 17(3):1198.

Đukan, M. and Kitzing, L. (2023). A bigger bang for the buck: The impact of risk reduction on renewable energy support payments in europe. *Energy Policy*, 173:113395.

Wang, N., Wang, Q., and Li, Y. (2025). Estimation and forecast of carbon emission market volatility based on model averaging method. *Economic Modelling*, 143:106976.

Wang, Y., Li, J., O'Leary, N., and Shao, J. (2024). Excess demand or excess supply? a comparison of renewable energy certificate markets in the united kingdom and australia. *Utilities Policy*, 86:101705.

Wang, Z., Hong, T., and Piette, M. (2020). Building thermal load prediction through shallow machine learning and deep learning. *Applied Energy*, 263.

Woo, C., Moore, J., Schneiderman, B., Ho, T., Olson, A., Alagappan, L., Chawla, K., Toyama, N., and Zarnikau, J. (2016). Merit-order effects of renewable energy and price divergence in california's day-ahead and real-time electricity markets. *Energy Policy*, 92:299–312.

Xu, K., Xia, Z., Cheng, M., and Tan, X. (2023). Carbon price prediction based on multiple decomposition and xgboost algorithm. *Environmental Science and Pollution Research*, 30(38):89165–89179.

Zakoian, J. M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5):931–955.

Zhang, D., Chen, X. H., Lau, C. K. M., and Cai, Y. (2023). The causal relationship between green finance and geopolitical risk: implications for environmental management. *Journal of Environmental Management*, 327:116949.

Zhang, W., He, X., and Hamori, S. (2022). Volatility spillover and investment strategies among sustainability-related financial indexes: Evidence from the dcc-garch-based dynamic connectedness and dcc-garch t-copula approach. *International Review of Financial Analysis*.

Zhu, B., Shi, X., Chevallier, J., Wang, P., and Wei, Y.-M. (2016). An adaptive multiscale ensemble learning paradigm for nonstationary and nonlinear energy price time series forecasting. *Journal of Forecasting*, 35(7):633–651.

Zhu, B., Wan, C., Wang, P., and Chevallier, J. (2025). Forecasting carbon market volatility with big data. *Annals of Operations Research*, 348(1):317–343.

Zhu, B., Ye, S., Wang, P., Chevallier, J., and Wei, Y.-M. (2022). Forecasting carbon price using a multi-objective least squares support vector machine with mixture kernels. *Journal of Forecasting*, 41(1):100–117.