Can Transition Risk Drive European Stock Market Predictions? An XGBoost Approach

Bogdan Ionut ANGHEL*
Bucharest University of Economic Studies, Bucharest, Romania
**Corresponding author, bogdan.anghel@rei.ase.ro*

Radu LUPU
Bucharest University of Economic Studies, Bucharest, Romania
Romanian Academy, Bucharest, Romania, Institute for Economic Forecasting
radu.lupu@rei.ase.ro

Abstract. *This study investigates the predictive power of transition risk for the European stock market by incorporating a text-based Transition Risk Index into an XGBoost forecasting model for the European Index STOXX600. Utilizing daily data from January 2005 to December 2023, we integrate standard macroeconomic factors—such as exchange rates, gold prices, and interest rates—with an NLP-derived Transition Risk Index that captures shifts in climate policies, technological innovations, and investor sentiment. Through an extensive feature engineering process and rigorous hyperparameter tuning (via cross-validation), we assess the relative contribution of each predictor using both feature importance rankings and SHAP (SHapley Additive exPlanations) analysis. Our findings reveal that, while conventional macro-financial variables remain the dominant drivers of STOXX600 price dynamics, transition risk exerts only a modest influence on short-term market forecasts. This suggests that near-term valuation processes may not yet fully integrate sustainability considerations. However, as climate policies evolve and investor awareness grows, the transition risk's role may become more pronounced over longer horizons. The results underscore both the potential and challenges in quantifying transition risk, offering a robust perspective on incorporating sustainability-driven metrics into real-time market analytics and predictive models. Furthermore, this research highlights the importance of refining transition risk measurement methodologies, as existing models may not fully capture the speed and complexity of regulatory shifts, technological advancements, and investor sentiment changes.*

Keywords: financial markets, transition risk, machine learning, forecast, European market

Introduction

Financial markets can amplify the challenges tied to carbon-intensive assets, especially as efforts to shift toward lower-carbon operations gain momentum. When this transition speeds up, investors often find it difficult to shed carbon-heavy liabilities, sparking liquidity issues that spread across multiple market segments. This vulnerability underscores the importance of understanding transition risk and the need for reliable forecasting tools. Such tools help us see how changing climate policies and market preferences can impact asset valuations and overall financial stability.

Against this backdrop, transition risk—which includes the economic and financial pressures arising from new climate regulations, novel technologies, and shifting market behaviors—has taken center stage in the European Union. Even so, it's not yet clear whether transition risk indicators truly increase the accuracy of forecasts for major stock indices like the STOXX600. To explore this question, our paper incorporates a NLP (Natural Language Processing) Transition Risk Index into an XGBoost (Extreme Gradient Boosting) model, assessing

DOI: 10.2478/picbe-2025-0097

whether these climate-related factors add valuable explanatory insights beyond traditional drivers such as interest rates and global equity trends.

The study draws on daily data from January 2005 through December 2023—a period that saw several significant economic upheavals, including the 2008 global financial crisis, the European sovereign debt crisis, the COVID-19 pandemic and the invasion of Russia into Ukraine. These events tested financial systems worldwide, offering a diverse setting in which to investigate how transition risk might shape stock market behavior under different types of market stress scenarios.

We have three main contributions. First, we present a methodology that distinctly separates traditional financial metrics from sustainability-related factors, helping to reveal how climate risks influence asset pricing. Second, we focus on the more extreme movements in the market, uncovering patterns that simpler, linear models tend to miss. Third, we account for episodes of pronounced market turbulence—such as during the COVID-19 pandemic—by examining how relationships shift over time. By combining advanced machine learning techniques with transparent interpretation (using Explainable AI), we show how transition risk could be a key driver of stock market swings, especially during periods of heightened uncertainty.

Our analysis is structured as follows. We start by exploring key findings from existing research, placing our study within the wider discussion. Second, we describe our data sources and methodology, including how we build our Transition Risk Index and integrate it into the XGBoost framework. Next, we present our findings, relate them to prior research, and use SHAP-based methods to highlight the critical role of transition risk. Finally, we offer our conclusions, point out limitations, and suggest areas for future exploration, along with practical advice for both market participants and researchers.

Literature review

Transitioning to a low-carbon economy has become both a huge opportunity and a serious threat for financial systems worldwide. Known as “transition risks,” these challenges arise from changing policies, market shifts, and consumer behaviors geared toward fighting climate change. As governments enforce stricter environmental rules, industries that produce a lot of carbon could be stuck with “stranded assets” and weaker credit standing, while more climate-friendly sectors continue to flourish (Semieniuk et al., 2020; Carattini et al., 2023).

Current research shows that transition risks pose a systemic concern for financial markets: companies that emit large amounts of greenhouse gases often face higher credit risk, whereas firms that disclose emissions data and commit to rigorous reduction targets tend to lower their overall risk (Carbone et al., 2022; Alessi & Battiston, 2022). Notably, European companies appear more exposed to these risks than their U.S. counterparts, likely because of Europe’s more stringent climate policies (Reboredo & Ugolini, 2022).

In this shift to a low-carbon future, banks have a key role to play, although their actions have been relatively gradual. Instead of rapidly divesting from high-emission industries, many banks are methodically tweaking their lending portfolios, motivated by increasing pressure from both regulators and shareholders to align financing with global climate goals (Martini et al., 2024). At the same time, stricter climate rules—like bigger carbon taxes and net-zero requirements—can trigger market turbulence by increasing the risk that carbon-intensive assets may suddenly lose value. One potential remedy involves macroprudential measures that include adjustments to bank balance sheets, providing incentives to steer more investment toward low-carbon initiatives (Carattini et al., 2023; Semieniuk et al., 2020).

Moreover, to better manage these transition risks, a variety of tools and frameworks have been put forward. The Carbon Risk Score, for instance, combines emissions data with other climate indicators to offer a more detailed view of each firm's exposure (Reboredo & Ugolini, 2022). Alessi & Battiston (2022) emphasize how the EU Taxonomy gives a standardized reference for assessing how "green" financial assets are, while also highlighting the need to balance promoting green investments with keeping overall financial risks in check during the transition.

Yet, as Semieniuk et al. (2020) explain, the scale of the low-carbon shift is unlike anything we've seen, affecting industries from fossil fuels and manufacturing to finance and beyond. High-emission sectors in particular confront growing risks of stranded assets, whereas renewable energy and green technology are expanding rapidly in both scale and economic impact. Following the Paris Agreement in 2015 (COP21), global investors started paying much closer attention to climate-related financial risks. Recent studies show that European investors, in particular, have begun factoring these risks into their decisions more carefully, while the U.S. seems to be lagging behind. This contrast suggests that differences in climate policies and investor awareness vary across regions (Reboredo & Ugolini, 2022).

Because transition risks cut across various sectors and regions, governments and markets need to respond in a coordinated way. Effective climate policy must work to reduce emissions without causing undue financial instability. Carattini et al. (2023) and Semieniuk et al. (2020) both call for a broader framework that accounts for physical climate threats (like extreme weather) as well as transition risks, such as sudden changes in asset values. While awareness of these issues is on the rise, there are still big gaps in the research. Semieniuk et al. (2020) note that financial risks are often looked at in isolation instead of as part of wider economic shifts. Meanwhile, Carattini et al. (2023) argue that a more holistic approach is crucial to giving policymakers and business leaders the insights they need to direct investments toward green sectors while also keeping financial systems stable.

On the other hand, XGBoost has quickly become a go-to tool in machine learning research since Chen and Guestrin (2016) introduced it as a scalable tree-boosting framework. Its popularity lies in its efficient handling of large, complex datasets—thanks to parallel computing and out-of-core processing—along with built-in regularization that helps prevent overfitting and ensures stronger generalization. XGBoost has also shown itself to be a top performer in time series forecasting in many financial applications, largely because it can capture complicated, nonlinear relationships and handle big sets of features effectively (Tan et al., 2023; Xu et al., 2022). Studies that pair advanced feature engineering with XGBoost highlight how gradient boosting can outperform traditional statistical models in predicting things like stock prices and retail demand (Esmailizade et al., 2024; Xu et al., 2022). By incrementally improving weaker models (weak learners), XGBoost reduces overfitting and handles missing data well, offering both accuracy and computational speed for financial time series (Tan et al., 2023).

Methodology

In this study, we collected daily data from January 2005 to December 2023, resulting in a total of 4,955 observations. Our dataset includes various macroeconomic and financial indicators, the STOXX600 index (sourced from Datastream by Refinitiv LSEG), and the Transition Risk Index. The entire analysis was conducted using Python, utilizing libraries such as pandas, scikit-learn, XGBoost, and SHAP for data processing, machine learning modeling, and interpretability. Table 1 lists all the independent variables that were initially chosen. Prior to analysis, we pre-processed each dataset to standardize the date format and ensure that all time points lined up consistently.

Table 1. List of independent variables

Acronym	Full Name
EURIBOR 3M	3-month Euro Interbank Offered Rate
USD/EUR	U.S. Dollar to Euro Exchange Rate
EEX	European Energy Exchange
Gold Spot LDN	London Gold Spot Price
MSCI World	MSCI World Index
S&P 500	Standard & Poor's 500 Index
NASDAQ 100	NASDAQ 100 Index
Hang Seng	Hang Seng Index
US 10-Year Gov Bonds	U.S. 10-Year Government Bond Yield
EUR/GBP	Euro to British Pound Exchange Rate
EUR/CNY	Euro to Chinese Yuan Exchange Rate
Transition Risk Index	Index created by by Bua et al. (2024)

Source: Authors' own research.

We used the Transition Risk Index, created by Bua et al. (2024) using the text-based methodology described in their paper. As Bua et al. (2024) mentioned, this process involved applying natural language processing (NLP) techniques to financial news from Reuters, spanning 2005–2023, and comparing each news item to a specially curated climate risk vocabulary. By doing so, the index is able to capture and quantify transition risk based on shifts in regulations, technology, and market sentiment related to moving toward a low-carbon economy. In practical terms, Bua et al. (2024) used the Transition Risk Index through cosine similarity between daily news reports and our defined climate risk vocabulary, then used an autoregressive filter to pick up unexpected spikes or "shocks" in transition risk levels.

We selected the STOXX600 as the dependent variable for our analysis because it is one of the most widely recognised benchmarks for European equity markets, representing companies of various sizes across 17 countries. Its broad coverage makes it particularly valuable for examining how transition risk affects market valuations, as it reflects Europe's diverse economic landscape and evolving regulatory environment. This is especially relevant given the region's ambitious climate policies and increasing investor focus on sustainability, both of which can significantly influence market sentiment and, consequently, the performance of key indices like the STOXX600 index.

To handle missing data, we chose the K-Nearest Neighbors (KNN) algorithm because it accommodates complex relationships among variables while preserving their underlying structure. Since time series data often suffer from non-stationarity, we tested for stationarity in the main variables using both the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) methods. The dependent variable—STOXX600—was log-transformed to stabilize its variance and improve stationarity.

Next, to capture both short- and long-range signals within the data, we created up to five lagged versions of each variable and calculated moving averages with window sizes ranging from 3 to 40 periods. Lagged variables help model the natural autocorrelation seen in time series data, while moving averages smooth out day-to-day "noise" so that broader trends are clearer. We then divided the data into training and testing sets in an 80-20 split, keeping the original chronological order intact. This setup ensured that STOXX600 was correctly aligned with the features across their respective time offsets.

After this step, we carried out a systematic feature selection process using Pearson correlation to identify the most useful predictors while avoiding redundancy or overfitting. During this step, we chose to exclude the MSCI World, S&P 500, and NASDAQ 100 indices because of their strong correlations (0.97–0.99) with one another, which introduced multicollinearity. In particular, the NASDAQ 100 and S&P 500 exhibited an almost perfect correlation (0.99), meaning they offered nearly identical information. Similarly, the MSCI World index was highly correlated with both, indicating that all three effectively captured the same overarching market dynamics. The results are presented in Figure 1 down below.

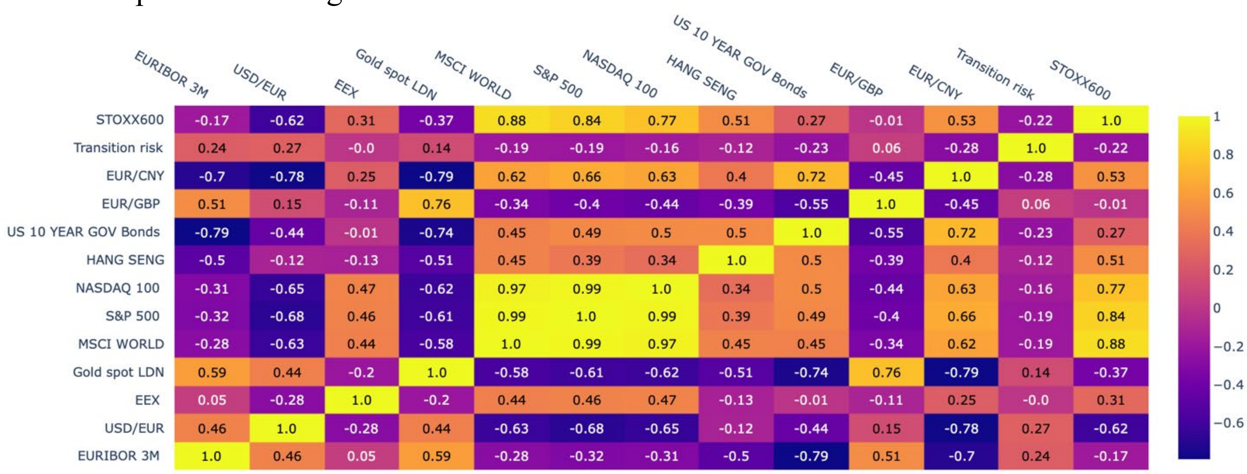

Figure 1. Correlation heatmaps of the independent variables

Source: Authors' own research.

For our STOXX600 price-forecasting study, we chose an XGBoost regression model (Chen and Guestrin, 2016) because it excels at handling complex interactions among variables and provides built-in tools for assessing feature importance. We first trained the model using each variable, along with its lagged values and moving averages. Whichever version—original, lagged, or moving average—delivered the lowest Mean Squared Error was deemed the most useful and included in the final model. This way, if a moving average showed stronger predictive power than the original variable or its lags, it took priority. The end goal was to focus on features that boosted accuracy while also acknowledging the inherent time-dependent patterns in financial markets. After we narrowed down the best set of features, we fine-tuned the final model using XGBoost.

We choose XGBoost for its strong performance in modeling non-linear relationships, its ability to capture intricate interactions, and its built-in regularization to avoid overfitting—essential for our high-dimensional dataset. We concentrated on minimizing negative Mean Squared Error, and cross-validation confirmed that the model would generalize well to new data. The topperforming hyperparameters were then locked in, and the final model was retrained with these optimal settings.

To efficiently explore the hyperparameter space and manage the risk of overfitting, we employed Randomized Search Cross-Validation, which is especially helpful given the multiple lagged variables and moving averages in our dataset. We tested the final XGBoost model on our hold-out set using Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2). These metrics offered a well-rounded view of the model's accuracy and how well it explained the variation in the data. For deeper insight into how the model arrived at its predictions, we used SHAP values, which clearly show the positive or negative influence of each feature on the model's output.

When tuning hyperparameters, we focused on several core settings. The number of estimators (n_estimators) ranged from 10 to 500, and the maximum depth of the trees (max_depth) ranged from 1 to 15, controlling how complex the model could get. We tested learning rates (learning_rate) between 0.01 and 0.3, striking a balance between how fast the model trains and how accurate it is. To combat overfitting, we randomly chose subsample and colsample_bytree values between 0.5 and 0.8, limiting how much data and how many features each tree could see. We also tweaked L1 regularization (reg_alpha) from 0 to 5 to introduce feature sparsity, and L2 regularization (reg_lambda) between 1 and 5 to rein in overly complex models. To ensure our findings held up, we used a 5-fold cross-validation strategy with 50 different hyperparameter configurations, and chose the best one based on negative MSE.

Results and discussions

The XGBoost model has delivered highly accurate predictions for the STOXX 600 index price, as reflected in its performance metrics. The low MAE of 0.01298 suggests that the model's predictions closely align with actual values, while the MSE of 0.00035 indicates minimal variance in errors. The R-squared value of 0.9927 demonstrates that the model explains 99.27% of the variation in the index price, reinforcing its reliability. These results suggest that the model effectively captures market trends and patterns, making it a valuable tool for forecasting price movements with confidence.

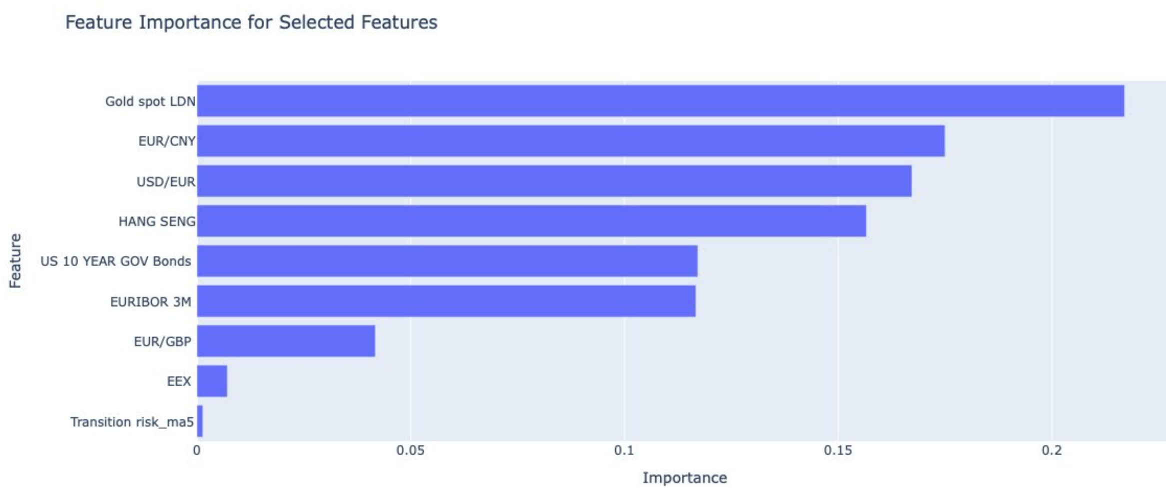

Figure 2. Feature Importance of the Champion XGBoost Model

Source: Authors' own research.

The predictive performance of the XGBoost model was examined using feature importance metrics, offering insights into the relative contribution of each variable to the model's forecasting capability. As illustrated in the Feature Importance of the Champion XGBoost Model Plot (see Figure 2), the Gold Spot LDN variable emerged as the most influential predictor, followed closely by major currency exchange rates such as EUR/CNY and USD/EUR. This finding suggests that fluctuations in gold prices and exchange rate dynamics play a critical role in shaping market expectations for the STOXX600 index. Additionally, the HANG SENG index and US 10-Year Government Bonds exhibited substantial predictive power, further reinforcing the notion that European stock market movements are closely interconnected with global financial trends and macroeconomic indicators.

By contrast, Transition Risk (Transition_risk_ma5) demonstrated notably lower importance in the model, ranking at the bottom among the selected predictors. This result implies that transition risk does not exert a strong direct influence on STOXX600 Index price dynamics relative to conventional financial variables. While transition risk remains a relevant concept in sustainability-driven investment decisions, its role in short-term market predictions appears to be overshadowed by more traditional economic and financial indicators.

To complement the feature importance analysis, SHAP values were utilised to assess the direction and magnitude of each feature's impact on the model's output. The SHAP analyses for the Champion XGBoost Model plot (Figure 3) highlights significant variation in feature effects, particularly for Gold Spot LDN, USD/EUR, EUR/CNY, and HANG SENG, which exhibited the largest dispersion of SHAP values. The colour gradient in the plot indicates the relative magnitude of feature values, with red denoting higher values and blue representing lower values. Notably, the distribution of SHAP values suggests that these variables exert both positive and negative influences on market forecasts, depending on their respective levels at any given time.

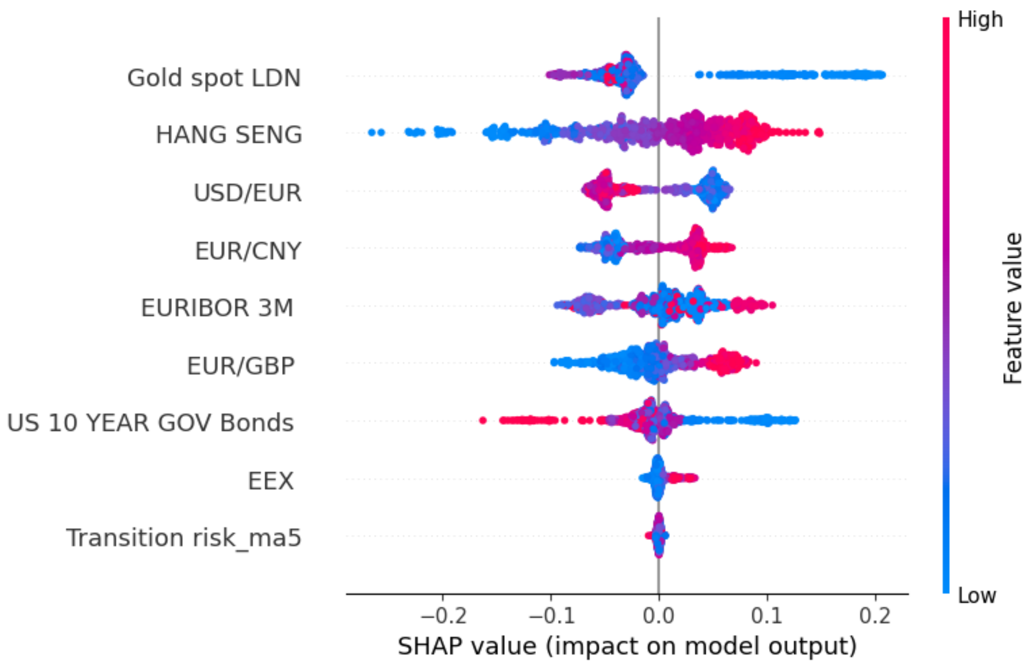

Figure 3. SHAP analyses for the Champion XGBoost Model

Source: Authors' own research.

The Transition Risk (Transition_risk_ma5) variable, however, displayed a concentration of SHAP values around zero, further substantiating its limited influence in the predictive model. This result aligns with the feature importance analysis, reinforcing the conclusion that transition risk—at least in its current measurement—has a marginal impact on STOXX600 price fluctuations.

The findings from both feature importance and SHAP analyses underscore the dominance of traditional macroeconomic and financial variables in predicting European stock market movements. The prominence of gold prices, exchange rate fluctuations, and global financial indicators suggest that market participants respond primarily to well-established economic signals rather than transition risk-related factors.

The relatively modest predictive influence of transition risk in this study raises important questions about its role in forecasting. Although it is broadly recognized as a long-term structural driver in financial markets, its muted impact in shorter-term models may stem from several factors. First, the transition risk indicator used may not fully capture the nuances of shifting policies, emerging technologies, and changing market dynamics. Second, transition risk often plays out over longer horizons, making it less visible in short- to medium-term forecasts focused on rapid price movements. Finally, investors may already be factoring these risks into asset prices, dispersing their effects across different securities rather than creating a pronounced signal within a single stock market index.

Conclusion

This study set out to see whether adding a Transition Risk quantification into an XGBoost model could significantly boost how well we can predict the STOXX600 index. Based on our results, while the transition risk measure captures certain sustainability-related details, its overall impact on near-term market forecasts seems relatively small compared to traditional macroeconomic and financial factors.

Our findings show that well-established indicators—like exchange rates, gold prices, and major equity indices—still play the biggest role in predicting STOXX600 index movements. By contrast, the transition risk index was less influential when we looked at both feature importance and SHAP analyses. This suggests that, at least in the short-to-medium term, the effects of transition risk might be overshadowed by more immediate macro-financial signals.

In short, transition risk doesn't currently appear to strongly affect STOXX600 index pricing on a shorter time horizon. That said, as climate policies and investor attitudes evolve, transition risk could become more significant down the line. Refined transition risk measures and richer data sources remain promising avenues for future research—both for theoretical understanding and for practical, sustainability-oriented investment strategies.

Despite these insights, a few challenges emerged. First, our Transition Risk Index may not fully capture how quickly climate policies, technologies, and investor sentiment can shift. Using more diverse textual sources or refining our natural language processing approach might make the index more responsive to these rapid changes. Second, while our dataset covered major economic upheavals, it still might miss some longer-term effects where transition risk has a bigger impact. Future research could look into sector-level or bond-market responses, which might be more sensitive to climate-related policy changes.

We also encountered some technical difficulties in merging different data sources and syncing up multiple time series, emphasizing the need for robust data-handling methods. Adding real-time news feeds, regulatory statements, or corporate reports could yield more detailed transition risk signals. Finally, refining feature engineering—especially by addressing non-linearities and how factors interact—may help clarify how sustainability risks and traditional market drivers work together under various market conditions.

In conclusion, while transition risk has a relatively small impact on short-term STOXX600 index price predictions compared to traditional macro-financial factors, its importance could grow over time. As climate policies evolve and investors place greater emphasis on sustainability, more refined transition risk measures—enhanced with richer data and improved modeling techniques—may better capture these dynamics. Future research exploring sector-specific trends, bond market reactions, and real-time news or corporate disclosures could provide deeper insights into how transition risk influences machine learning forecasting algorithms.

References

Alessi, L., & Battiston, S. (2022). Two sides of the same coin: Green Taxonomy alignment versus transition risk in financial portfolios. *International Review of Financial Analysis, 84*, 102319.

Bua, G., Kapp, D., Ramella, F., & Rognone, L. (2024). Transition versus physical climate risk pricing in European financial markets: A text-based approach. The European Journal of Finance, 30(17), 2076–2110

Carattini, S., Heutel, G., & Melkadze, G. (2023). Climate policy, financial frictions, and transition risk. *Review of Economic Dynamics, 51*, 778–794.

Carbone, S., Giuzio, M., Kapadia, S., Krämer, J. S., Nyholm, K., & Vozian, K. (2022). The low-carbon transition, climate commitments, and firm credit risk. *Sveriges Riksbank Working Paper Series, No. 409*. Sveriges Riksbank.

Chen, T. and Guestrin, C. (2016) 'XGBoost: A scalable tree boosting system', Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), 13–17 August, San Francisco, CA, USA, ACM, pp. 785–794.

Esmailizade, S., Ebrahimi, A., Soltani, H., Sam, A. and Rahimi, M. (2024) 'Machine Learning Approaches for Retail Forecasting: A Study on XGBoost and Time-Series Models', *TechRxiv*.

Martini, F., Sautner, Z., Steffen, S., & Theunisz, C. (2024). Climate transition risks of banks. *SSRN Electronic Journal*.

Meinerding, C., Schüler, Y. S., & Zhang, P. (2024). Shocks to transition risk. *SSRN Electronic Journal*.

Reboredo, J. C., & Ugolini, A. (2022). Climate transition risk, profitability, and stock prices. *International Review of Financial Analysis, 83*, 102271.

Semieniuk, G., Campiglio, E., Mercure, J.-F., Volz, U., & Edwards, N. R. (2020). Low-carbon transition risks for finance. *WIREs Climate Change, 12*(1), e678.

Tan, B., Gan, Z. and Wu, Y. (2023) 'The measurement and early warning of daily financial stability index based on XGBoost and SHAP: Evidence from China', *Expert Systems with Applications*, 227, p. 120375.

Xu, J., He, J., Gu, J., Wu, H., Wang, L., Zhu, Y., Wang, T., He, X. and Zhou, Z. (2022) 'Financial Time Series Prediction Based on XGBoost and Generative Adversarial Networks', *International Journal of Circuits, Systems and Signal Processing*, 16, pp. 637–647.

DOI: 10.2478/picbe-2025-0097, pp. 1226-1234, ISSN 2558-9652 |
Proceedings of the 19th International Conference on Business Excellence 2025