

# Global Market Contagion Intelligence System for Systemic Risk Analysis and Forecasting

Kriti Karunam

Mahatma Gandhi Institute of Technology  
Email: kkriti\_csb243222@mgit.ac.in

Dr. A Nagesh

Mahatma Gandhi Institute of Technology  
Email: anagesh\_cse@mgit.ac.in

**Abstract**—Global financial markets are increasingly interconnected, enabling localized shocks to propagate rapidly across regions. Traditional indicators such as volatility and correlation measure uncertainty and co-movement but do not explain how risk originates, how fast it spreads, or which market is likely driving the disturbance. This new paper presents a Global Market Contagion Intelligence System that quantifies systemic risk, identifies probable contagion sources, estimates transmission lag, and forecasts next-day risk for India-centric cross-market monitoring. The framework combines Vector Autoregression (VAR), Granger causality, a composite Systemic Risk Index, and XGBoost forecasting inside a FastAPI–React architecture. Experimental output indicates a current risk score of 39.85, identifies the U.S. S&P 500 as the dominant contagion source, estimates a one-day lag, and forecasts next-day risk at 41.83. The novelty of the proposed system lies in its integrated framework that combines causality analysis, risk quantification, and machine learning-based forecasting into a single interpretable pipeline. The proposed platform is interpretable, modular, and scalable for future research and real-time deployment.

**Keywords** — *financial contagion, systemic risk, VAR, Granger causality, XGBoost, spillovers, India, dashboard analytics*

## I. INTRODUCTION

Global financial markets operate as an interconnected system in which shocks are transmitted through capital flows, investor sentiment, trade channels, and macroeconomic linkages. While such integration improves efficiency during normal periods, it also creates vulnerability during crises, when localized disturbances can rapidly propagate across geographies. This phenomenon is commonly described as financial contagion and is closely associated with systemic risk.

The importance of contagion analysis has been demonstrated repeatedly by major disruptions such as the 2008 financial crisis and the COVID-19 market shock. In both cases, stress was not confined to the originating region; instead, it spread across markets in a way that revealed deep interdependence. These events exposed a practical weakness in traditional risk analysis, which often relies on volatility, correlation, or stand-alone market indicators. Such measures are informative, but they do not answer questions about timing, direction, or source.

The objective of this work is therefore not only to measure market stress but to explain how stress propagates and to estimate what may happen next. The proposed Global Market Contagion Intelligence System is designed to quantify India-centric systemic risk, identify the most probable source of contagion, estimate transmission lag, and forecast next-day risk using a hybrid analytical framework.

### I-A. Key Contributions

The novelty of this work lies not in proposing new individual models, but in designing an integrated system that combines econometric analysis, machine learning, and risk aggregation into a unified framework for financial contagion analysis. The key contributions are:

- 1. Integrated Contagion Framework:** A unified pipeline combining VAR, Granger causality, and XGBoost to simultaneously estimate contagion timing, direction, and future risk.
- 2. Multi-Dimensional Risk Interpretation:** Unlike traditional approaches that rely on a single metric, the system decomposes systemic risk into interpretable

components such as volatility, correlation, and drawdowns.

- 3. India-Centric Contagion Modeling:** The framework specifically models global-to-India spillovers, enabling targeted analysis of emerging market vulnerability.
- 4. End-to-End Decision System:** The system moves beyond analysis by integrating prediction, visualization, and real-time interpretation in a deployable architecture.
- 5. Explainable Risk Forecasting:** By combining statistical causality with machine learning, the system provides both predictive accuracy and interpretability.

## II. PROBLEM STATEMENT AND LITERATURE BACKGROUND

A major challenge in financial risk analysis is separating genuine contagion from simple co-movement. Markets can move together due to shared external shocks, lead-lag relationships, or a common latent factor. Forbes and Rigobon [11] formalized this distinction, arguing that apparent contagion during crises is often pre-existing interdependence amplified by volatility, while Dungey et al. [12] noted that no single approach captures all dimensions of cross-market transmission. Existing systems often provide only partial answers: correlation-based methods reveal association without direction, volatility indices quantify uncertainty without transmission pathways, and forecasting models may offer predictions without interpretable causal evidence. Billio et al. [7] proposed econometric connectedness measures to capture systemic risk, highlighting that pairwise correlations alone are insufficient to detect propagation channels.

Econometric research provides the core building blocks for contagion analysis. Sims [1] introduced VAR for interdependent time-series modeling, Granger [2] proposed a predictive causality framework, and Diebold and Yilmaz [3] extended these ideas to spillover measurement. Engle's [4] ARCH work highlighted volatility clustering, Mantegna [8] demonstrated that markets exhibit hierarchical clustering structures, and Lütkepohl [6] formalized multivariate time-series tools for interpreting lagged relationships. On the machine learning side, XGBoost captures non-linear structure with computational efficiency [5], Gu et al. [9]

showed that ML methods systematically outperform traditional factor models in asset pricing, and Kritzman et al. [10] demonstrated that principal components of asset returns serve as compact systemic risk indicators — motivating multi-dimensional aggregation approaches.

Despite these advances, existing literature often emphasizes either explanatory econometrics or predictive learning rather than an integrated user-facing system. The present work contributes by combining both perspectives with a composite risk index and dashboard-based visualization, supporting not just model estimation but also interpretation and decision support.

Table I. Condensed literature survey summary.

Work	Key contribution	Limitation
Sims (1980)	VAR for lagged interdependence	Assumes linearity
Granger (1969)	Directional predictive testing	Not true causation
Diebold and Yilmaz (2012)	Spillover measurement	Sensitive to specification
Engle (1982)	Volatility clustering models	No directional contagion
Chen and Guestrin (2016)	XGBoost for scalable forecasting	Requires feature engineering
Lütkepohl (2005)	Multivariate time-series interpretation	Needs careful assumptions

### III. PROPOSED SYSTEM AND METHODOLOGY

The proposed system follows a modular pipeline: data collection, preprocessing, statistical contagion analysis, risk index construction, forecasting, and dashboard visualization. Each stage is designed to solve a specific weakness of existing approaches. Preprocessing creates stable and comparable inputs, VAR measures lagged dependence, Granger causality identifies direction, the composite index summarizes overall stress, and XGBoost extends the system from explanation to short-horizon prediction.

Historical market data are collected for India and selected global markets together with supporting indicators such as volatility indices, exchange rates, and macro-financial proxies.

#### III-A. Log Return Computation

Because raw price levels can be non-stationary and unsuitable for direct inference, the system converts prices into log returns before model fitting. Let  $P_t^{(i)}$  denote the closing price of market  $i$  at time  $t$ . The log return is defined as:

$$r_t^{(i)} = \ln \left( \frac{P_t^{(i)}}{P_{t-1}^{(i)}} \right)$$

This transformation stabilizes variance and renders prices stationary, satisfying the assumption required for VAR estimation.

#### III-B. VAR(p) Model — Full Formulation

In a VAR(p) process, every market depends on its own history and on the histories of the other markets in the system. A  $K$ -dimensional VAR of order  $p$  is specified as:

$$Y_t = \mu + \sum_{j=1}^p A_j Y_{t-j} + \varepsilon_t$$

where:

- $Y_t = [r_t^{(1)}, r_t^{(2)}, \dots, r_t^{(K)}]^T \in \mathbb{R}^K$  is the vector of log returns at time  $t$
- $\mu \in \mathbb{R}^K$  is the intercept vector
- $A_j \in \mathbb{R}^{K \times K}$  is the coefficient matrix at lag  $j$
- $\varepsilon_t \sim \mathcal{N}(\mathbf{0}, \Sigma)$  is white noise with covariance  $\Sigma$

The lag order  $p$  is selected by minimizing the Akaike Information Criterion:

$$AIC(p) = -2\ln(\hat{L}) + 2(Kp + K^2p)$$

where  $\hat{L}$  is the maximized log-likelihood of the VAR model.

#### III-C. Granger Causality — F-Test Formulation

Let  $r_t^{(X)}$  and  $r_t^{(Y)}$  represent returns of markets  $X$  and  $Y$  respectively. Market  $X$  Granger-causes  $Y$  if:

$$SSR_{\text{restricted}} > SSR_{\text{unrestricted}}$$

Formally, the F-statistic is:

$$F = \frac{(SSR_R - SSR_U)/p}{SSR_U/(T - 2p - 1)} \sim F(p, T - 2p - 1)$$

where:

- $SSR_R$ : sum of squared residuals of the restricted model (no lags of  $X$ )
- $SSR_U$ : sum of squared residuals of the unrestricted model (lags of both  $X$  and  $Y$ )
- $T$ : total observations;  $p$ : lag order

The null hypothesis  $H_0$ :  $X$  does not Granger-cause  $Y$  is rejected when  $p$ -value  $< 0.05$ .

#### III-D. Composite Systemic Risk Index — Weighted Aggregation

The composite risk index at time  $t$  is:

$$\mathcal{R}_t = \sum_{i=1}^N w_i \cdot \tilde{X}_{i,t}, \quad \sum_{i=1}^N w_i = 1$$

where each feature  $\tilde{X}_{i,t}$  is min-max normalized:

$$\tilde{X}_{i,t} = \frac{X_{i,t} - \min_{\tau} (X_{i,\tau})}{\max_{\tau} (X_{i,\tau}) - \min_{\tau} (X_{i,\tau})}$$

The component features and weights used in this work are:

Table II. component features and weights used in Composite Systemic Risk

Feature $X_i$	Weight $w_i$
Rolling volatility $\sigma_t^{(l)}$	0.30
Cross-market correlation $\rho_{I,US}$	0.15
India drawdown $D_t^{(l)}$	0.15
Global VIX index	0.10
Global market stress $\mathcal{G}_t$	0.30

**III-E. XGBoost Objective Function**

XGBoost minimizes a regularized loss over  $M$  additive trees:

$$\mathcal{L} = \sum_{t=1}^T \ell(\mathcal{R}_t, \widehat{\mathcal{R}}_t) + \sum_{m=1}^M \Omega(f_m)$$

where  $\ell(\cdot)$  is the mean squared error loss and the regularization term penalizes tree complexity:

$$\Omega(f_m) = \gamma T_m + \frac{1}{2} \lambda \|\mathbf{w}_m\|^2$$

Here  $T_m$  is the number of leaves in tree  $m$ ,  $\mathbf{w}_m$  are leaf weights,  $\gamma$  controls minimum gain for a split, and  $\lambda$  is the L2 penalty on leaf scores.

The methodological advantage of this design is integration. Timing, source, magnitude, and forecast are not produced by independent disconnected tools; they are generated by one coordinated pipeline, which makes the final system more interpretable and more useful for practical monitoring.

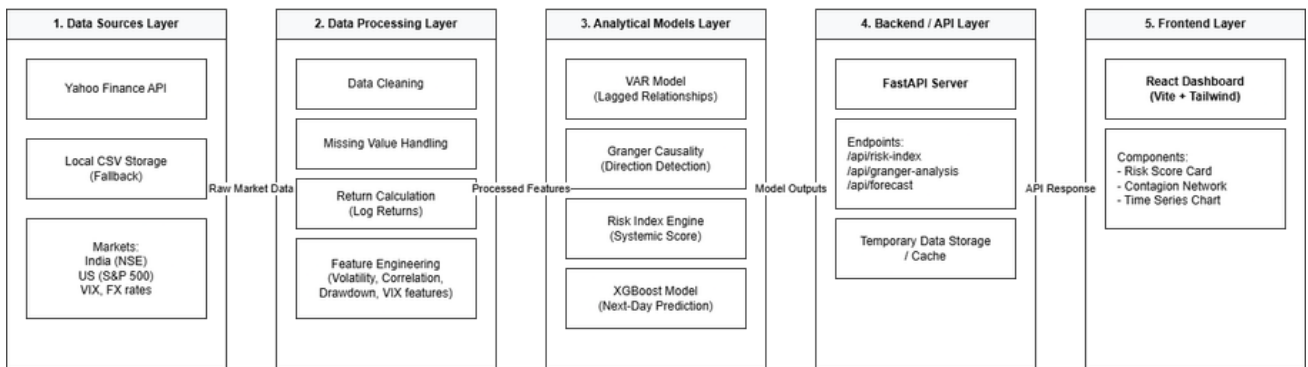


Fig. 1. End-to-end pipeline of the proposed Global Market Contagion Intelligence System. The pipeline proceeds from raw data ingestion through preprocessing, VAR/Granger analysis, composite risk index construction, XGBoost forecasting, and FastAPI-React dashboard delivery.

**IV. SYSTEM DESIGN, UML, AND IMPLEMENTATION**

**IV-A. SYSTEM ARCHITECTURE**

The implementation is organized into four major layers: a data layer, an analytics layer, an API layer, and a user-interface layer. This separation of concerns improves scalability and maintainability. New indicators can be added at the data layer without redesigning the frontend, and alternative forecasting models can be inserted into the analytics layer without changing the dashboard structure.

From a software perspective, Python is used for data handling and model computation, FastAPI provides low-latency service endpoints, React and Vite support the interactive dashboard, and Tailwind CSS maintains visual consistency. This stack was chosen because it supports both rapid prototyping and modular deployment.

Table III. Implementation layers and responsibilities.

Layer	Main responsibility
Data collection and preprocessing	Load historical series, align timestamps, compute returns and features
Analytics core	Run VAR, Granger causality, risk index, and XGBoost models
FastAPI service	Expose outputs through API endpoints.
React dashboard	Present cards, tables, heatmaps, charts, and network views
Scalability support	Allow future extension to more markets, indicators, and models

The UML representation highlights how the user interacts with a unified system rather than with isolated statistical models. The user requests analysis through the dashboard, the backend executes data preparation and model inference, and the resulting outputs are returned as integrated risk intelligence. This is significant because it converts a technically fragmented workflow into a single explainable application.

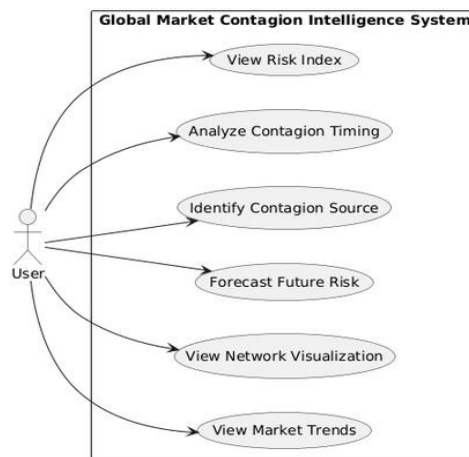


Fig. 2. UML use case diagram depicting analyst interaction with the dashboard. Primary actors include the Risk Analyst and System Administrator. Core use cases include risk score retrieval, contagion source identification, and forecast visualization.

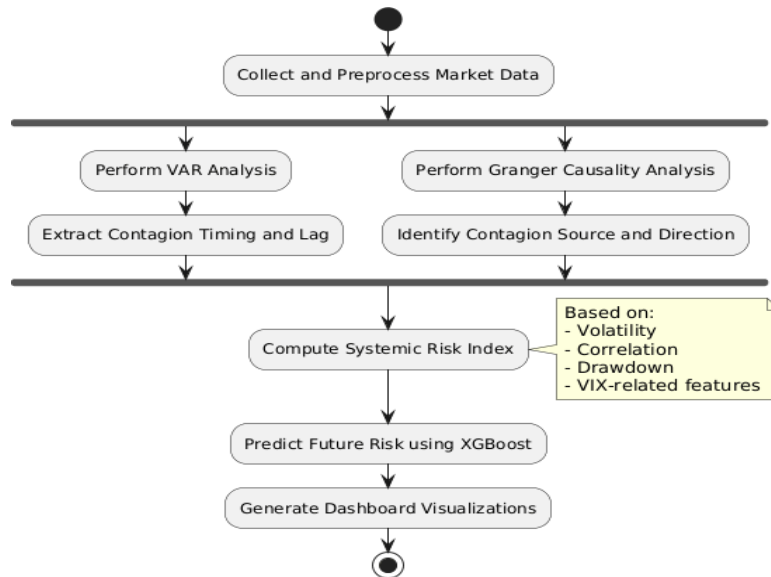


Fig. 3. Activity diagram illustrating the end-to-end workflow from data collection to dashboard rendering, including conditional branches for stationarity testing and model retraining.

At the implementation level, the frontend is composed of reusable pages and widgets such as dashboard cards, chart components, loading indicators, and network views. The backend uses route and service layers for risk computation, contagion analysis, and forecasting. This modular structure helps the system scale from an academic prototype to a broader risk intelligence platform.

Table IV. Major implementation components used in the platform.

Component group	Examples / role
Frontend	Dashboard page, metric cards, chart panels, network view, heatmap
Backend	Risk index endpoint, contagion endpoint, forecast endpoint, data services
Core libraries	pandas, numpy, statsmodels, scikit-learn, xgboost
User experience goal	Readable, responsive, and interpretable presentation of complex outputs

#### IV-B. Dataset Description

Daily financial market data is collected via **Yahoo Finance** with a local CSV fallback. The analysis focuses on India using **Nifty 50**, **S&P 500**, **USD/INR**, and **CBOE Volatility Index**. Additional proxies from the UK, Japan, Germany, Hong Kong, and China support macro stress and network analysis. Data is cleaned, aligned, converted to daily percentage or log returns, and used to derive features like volatility, drawdown, global correlation, VIX indicators, and market stress scores.

Table V. Data Description.

Item	Description
Data source	Yahoo Finance API, local CSV fallback
Frequency	Daily
Target market	India
Core analysis variables	India, US, USD/INR, Global VIX
Additional markets	UK, Japan, Germany, Hong Kong, China
Derived features	Returns, volatility, drawdown, correlation, VIX-based indicators
Target variable	Next-day risk index

#### IV-C. Model Contribution Analysis (Ablation Study)

Table VI. Ablation study: output capability by model configuration.

Configuration	Components	Output Capability	Limitation
VAR Only	VAR	Lag detection	No direction or prediction
VAR + Granger	VAR + Granger	Timing + direction	No future prediction
Risk Index Only	Aggregation	Overall risk level	No causality insight
XGBoost Only	ML	Prediction	No interpretability
<b>Proposed System</b>	<b>All combined</b>	<b>Timing + Direction + Risk + Forecast</b>	—

The ablation analysis demonstrates that individual models provide only partial insights. VAR captures lagged dependencies but lacks directionality. Granger identifies direction but does not quantify risk. XGBoost predicts future risk but lacks interpretability. The proposed system integrates all components, enabling a comprehensive understanding of contagion across multiple dimensions.

#### V. RESULTS AND DISCUSSION

The current implementation produces a coherent set of outputs that together describe the present market state and near-term expectation. The dashboard reports a systemic risk score of 39.85, identifies the U.S. S&P 500 as the dominant contagion source for India, estimates a transmission lag of one trading day, and forecasts next-day risk at 41.83.

A one-day lag suggests relatively fast cross-market spillover into India once the leading external market moves. The identification of the U.S. S&P 500 as the strongest source is plausible because U.S. market movements often influence global risk sentiment. The forecasted increase from 39.85 to 41.83 indicates a mild short-term rise in systemic stress, which demonstrates the practical value of adding machine learning to the explanatory econometric framework.

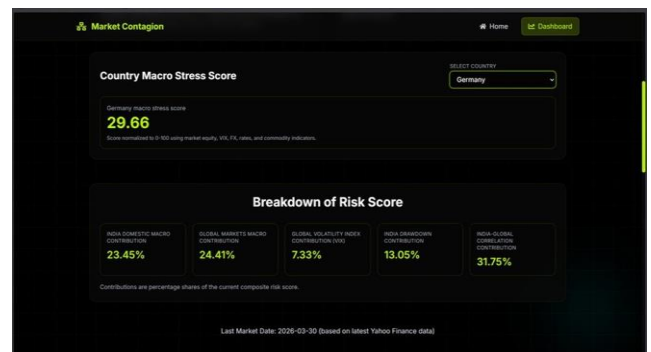
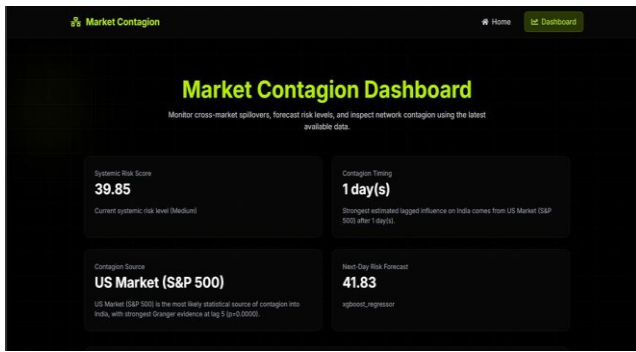


Fig. 4. Dashboard risk score breakdown showing component contributions to the current Systemic Risk Index (score: 39.85). Components include Germany Macro Stress (29.66), India-Global Correlation (31.75%), Global Markets (24.41%), India Domestic (23.45%), Drawdown (13.05%), and VIX (7.33%).

Table VII. Dashboard result metrics produced by the proposed system.

Metric	Observed value	Meaning
Systemic Risk Score	39.85	Current aggregate stress level based on the composite index
Dominant contagion source	U.S. S&P 500	Strongest estimated external driver of Indian market movement
Transmission lag	1 day	Approximate delay before spillover becomes visible in India
Next-day risk forecast	41.83	Predicted near-term increase in systemic risk

The results also demonstrate how the system design solves the original problem statement. VAR provides timing evidence rather than only co-movement, Granger causality adds directional interpretation, the risk index compresses multiple stress indicators into a usable score, and XGBoost forecasts what is likely to happen next. The dashboard then makes these technical outputs understandable through cards, tables, and visual summaries.

Table VIII. Core result metrics produced by the proposed system.

Metric	Value
Mean Absolute Error (MAE)	1.771506
Root Mean Squared Error (RMSE)	3.101222
Mean Absolute Percentage Error (MAPE)	8.784838
Directional Accuracy	56.16
Granger Causality p-value (US→India)	$2.06 \times 10^{-49}$
Optimal VAR Lag Order (AIC)	5

While directional accuracy of 56.16% marginally exceeds random baseline (50%), this is consistent with the efficient market hypothesis and reflects the inherent difficulty of short-horizon financial forecasting.

## VI. PRACTICAL ADVANTAGES AND CURRENT LIMITATIONS

A major strength of the proposed system is that it links multiple analytical viewpoints in a single pipeline. Econometric modules explain timing and source, the composite index provides a compact operational summary, and machine learning adds forward-looking capability. This makes the framework useful for users who need more than a descriptive dashboard. Instead of separately examining volatility charts, correlation matrices, and predictive outputs, the user receives an integrated interpretation of systemic stress.

The design also improves usability. FastAPI standardizes model outputs for frontend consumption, while the React dashboard converts technical outputs such as lag structures,

p-values, and forecast values into cards, tables, and visual summaries.

While the system demonstrates practical utility, several limitations should be acknowledged. First, the current implementation uses a fixed set of markets and does not incorporate dynamic network topology changes. Second, XGBoost is inherently a static-input model; sequential dynamics in risk propagation would be better modeled by LSTM or temporal convolutional networks. Third, the composite risk index weights ( $w_i$ ) are heuristically assigned; a data-driven approach such as principal component analysis (PCA) or entropy-based weighting would improve objectivity. Finally, the absence of a holdout crisis window (e.g., COVID-19 March 2020 or 2008 GFC) limits the out-of-sample generalizability claim.

Table IX. Practical strengths of the current system and realistic extension points.

Aspect	Current benefit	Next improvement
Hybrid analytics	Combines timing, source, and forecast in one platform	Add deeper model diagnostics and robustness tests
Interpretability	Dashboard makes outputs readable for non-specialists	Add explainability panels and downloadable reports
Scalability	Modular API and UI support extension	Integrate more markets and real-time streaming
Forecasting	Short-horizon prediction adds proactive monitoring	Compare against LSTM/GRU and crisis-specific baselines

## VII. CONCLUSION AND FUTURE SCOPE

The Global Market Contagion Intelligence System presents a structured, interpretable, and scalable approach to systemic risk analysis. By combining VAR, Granger causality, a composite risk index, XGBoost forecasting, and dashboard visualization, the system addresses key shortcomings of traditional methods that rely mainly on volatility or correlation. It provides timing, source, magnitude, and forecast in one integrated platform.

Future work can extend the system through wider geographic coverage, real-time data integration, richer graph-based visualization, and advanced sequence models such as LSTM or GRU. Cloud deployment, user-defined filters, downloadable reports, and more detailed market-level attribution can further strengthen the platform as a production-ready financial intelligence tool.

## REFERENCES

- [1] C. A. Sims, "Macroeconomics and Reality," *Econometrica*, vol. 48, no. 1, pp. 1–48, 1980.
- [2] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.
- [3] F. X. Diebold and K. Yilmaz, "Better to give than to receive: Predictive directional measurement of volatility spillovers," *Int. J. Forecasting*, vol. 28, no. 1, pp. 57–66, 2012.
- [4] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.
- [5] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. KDD*, 2016, pp. 785–794.
- [6] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*. Berlin, Germany: Springer, 2005.
- [7] M. Billio, M. Getmansky, A. W. Lo, and L. Pelizzon, "Econometric measures of connectedness and systemic risk in the finance and insurance sectors," *J. Financial Economics*, vol. 104, no. 3, pp. 535–559, 2012.
- [8] R. N. Mantegna, "Hierarchical structure in financial markets," *Eur. Phys. J. B*, vol. 11, pp. 193–197, 1999.
- [9] S. Gu, B. Kelly, and D. Xiu, "Empirical asset pricing via machine learning," *Rev. Financial Studies*, vol. 33, no. 5, pp. 2223–2273, 2020.
- [10] M. Kritzman, Y. Li, S. Page, and R. Rigobon, "Principal components as a measure of systemic risk," *J. Portfolio Management*, vol. 37, no. 4, pp. 112–126, 2011.
- [11] K. Forbes and R. Rigobon, "No contagion, only interdependence: Measuring stock market comovements," *J. Finance*, vol. 57, no. 5, pp. 2223–2261, 2002.
- [12] M. Dungey, R. Fry, B. González-Hermosillo, and V. L. Martin, "Empirical modelling of contagion: A review of methodologies," *Quantitative Finance*, vol. 5, no. 1, pp. 9–24, 2005.