DOI: 10.55643/fcaptp.6.65.2025.5069

FINANCIAL FORECASTING OF THE CRYPTOASSET MARKET: AN ECONOMETRIC APPROACH INTEGRATING HIGH-FREQUENCY AND BEHAVIOURAL DATA

Dmytro Lyushenko
CEO, NOTA LLC, Carmichael, United States;
e-mail: tymoshyu@gmail.com
ORCID: 0009-0003-9180-0934
(Corresponding author)

Nodari Gorgiladze
President, Institute for Digital Asset Systems & Tokenization Inc, Sacramento, United States;
ORCID: 0009-0007-2384-0036

Oleksandr Tuholukov
Researcher, Scientific Programs Director, NOTA Digital Currencies Research Center Inc, Sacramento, United States;
ORCID: 0009-0008-3661-764X

Mykhailo Sheptun
Senior QA Automation Engineer, A Place for Mom/APFM, New York, United States;
ORCID: 0009-0004-1313-6553

Artem Tsymbal
Principal Backend Software Engineer, OneMain Financial, Charlotte, United States;
ORCID: 0009-0006-8786-8428

Rodion Sharafan
Chief Executive Officer, RÁPIDIMPULSO LDA, Oeiras, Portugal;
ORCID: 0009-0004-7209-7232

ABSTRACT

High volatility of cryptocurrencies and the rapid spread of artificial intelligence (AI) technologies in the financial sector determine the need for accurate forecasting of risks and investor behaviour in the course of the digital transformation of financial markets. The aim of the research is to develop a system of econometric models to assess the profitability, volatility, liquidity, and downside risk of major crypto assets using AI-based methods. The methodological framework includes ARDL-MIDAS, GARCH-MIDAS, PMG, and logit specification models, which combine high-frequency market data, macroeconomic indicators, on-chain metrics, and investor sentiment indices. The sample covers secondary data for 2018-2025 for five leading assets — Bitcoin, Ethereum, BNB, XRP, and Solana. The results of the ARDL-MIDAS model showed that a 1% increase in trading volumes increases short-term returns by 0.012 points, while an increase in the VIX index reduces them by 0.014 points. In the GARCH-MIDAS model, the coefficients $\alpha=0.085$ and $\beta=0.900$ confirmed the high inertia of Bitcoin volatility, while the MIDAS component of the VIX had a significant impact of 0.27. The PMG panel model revealed a negative long-term effect of volatility on liquidity (−0.27) and a positive effect of stablecoin inflows (−0.12), indicating a stabilization function. The logit model proved that a one standard deviation increase in the VIX index increases the risk of a crash by 52%. The obtained results confirm the effectiveness of combining econometric methods and AI for analysing digital financial markets and AI technologies for analysing digital financial markets. The conclusions emphasize the possibility of practical use of the proposed models in financial forecasting, risk management, and digital asset stabilization policies in the context of the development of AI-based intelligent financial systems.

Keywords: virtual assets, financial forecasting, ARDL-MIDAS, GARCH-MIDAS, liquidity, volatility, risk modelling

JEL Classification: C22, C53, G12, D84, G17

Received: 13/11/2025
Accepted: 22/12/2025
Published: 31/12/2025

INTRODUCTION

The modern development of the digital economy is radically changing the structure of financial markets, stimulating the emergence of new approaches to assessing risks, liquidity, and investor behaviour. The virtual asset market, which includes cryptocurrencies, tokenized financial instruments, and stablecoins, demonstrates rapid growth in transaction volumes, capitalization, and the number of users. Over the past five years, the global capitalization of the crypto market has increased several times, which emphasizes its role as a component of the modern financial system. At the same time, such dynamics are accompanied by increased volatility, unpredictability of price movements, and significant sensitivity to information, as well as macroeconomic shocks. This necessitates the creation of advanced analytical tools capable of integrating big data, behavioural factors, and AI to build reliable forecasts.

The problem of the study is that most traditional econometric models built for classical financial markets do not take into account the specifics of digital assets. Such models have limited ability to reflect fast-moving market processes, on-chain activity, and investor sentiment that affect the short- and long-term behaviour of assets. Besides, the crypto market is characterized by the lack of centralized regulation, information asymmetry, and high dependence on network activity, which requires the use of complex analysis approaches. At the same time, traditional forecasting methods do not provide adequate accuracy in periods of structural breaks, information crises, or external macroeconomic shocks.

Problem statement. The rapid expansion of virtual asset markets produces structural instability driven by extreme volatility, heterogeneous data frequencies, behavioural shocks, and highly nonlinear on-chain dynamics. Traditional econometric frameworks cannot capture mixed-frequency information, fast-moving liquidity patterns, and sentiment-driven price adjustments. Existing studies offer fragmented evidence and lack integrated models linking returns, volatility, liquidity, and tail risks in a unified analytical system. This gap creates the core problem: current forecasting methods fail to generate stable, interpretable, and replicable estimates for complex crypto-financial environments.

Research questions:

- **RQ1:** How do macroeconomic uncertainty measures, on-chain activity, and behavioural indicators jointly affect the short-term profitability of major crypto assets?
- **RQ2:** How do global uncertainty shocks influence long-term and short-term volatility patterns in digital asset markets under a MIDAS specification?
- **RQ3:** What are the long-term equilibrium relationships between liquidity, trading activity, stablecoin flows, and volatility across major assets?
- **RQ4:** Which market, macroeconomic, and behavioural factors increase the probability of extreme price crashes in digital asset markets?
- **RQ5:** Does an integrated modelling framework (ARDL-MIDAS, GARCH-MIDAS, PMG, logit) improve forecasting accuracy compared to traditional models?
- *Hypotheses:*
- **H1:** Higher trading activity and positive sentiment increase short-term returns, while macroeconomic uncertainty and exchange inflows reduce profitability.
- **H2:** Long-term components of volatility are strongly influenced by uncertainty indicators such as the VIX and by slow-moving on-chain activity.
- **H3:** Liquidity exhibits long-run cointegration with trading volume, volatility, and stablecoin inflows, with stablecoins acting as stabilizing mechanisms.
- **H4:** Extreme market declines are more likely when uncertainty, exchange inflows, and realized volatility increase, while positive sentiment reduces crash probability.
- **H5:** Hybrid frequency-adaptive models outperform benchmark econometric and machine-learning alternatives in predictive accuracy and error stability.

The fast growth of virtual asset markets poses systemic risks that demand the best analytical tools that could combine macroeconomic shocks, behavioural indicators, and on-chain complexities. The digital is changing at a pace that is influenced by high volatility, information asymmetry, and the nonlinear interaction of market microstructure and global uncertainty. These circumstances drive the necessity to have forecasting systems capable of handling high-frequency data and responding to structural breaks that frequently define crypto markets. The rationale behind this research is the inadequacy of conventional frameworks to attain frequency inconsistency, behavioural changes, and liquidity variations that crop up in the digital space. The research problem is the explanatory weakness of classical frameworks, which do not take into account mixed-frequency information and heterogeneous responses across assets. The value of the current research is the ability to design an integrated forecasting system, which consists of ARDL-MIDAS, GARCH-MIDAS, PMG, and logit models, and enhance the quality of prediction. The research seeks to unveil the joint role of macroeconomic indicators, on-chain flows, volatility shocks, and sentiment measures in calculating profitability, liquidity, and tail risks of major crypto assets. The orientation to the econometric models used in the paper is also brief to allow the reader to understand the scope of the forecasting architecture. The ARDL-MIDAS element is a high-frequency market data connected to low-frequency macro variables and maintains the level of information granularity. The GARCH-MIDAS framework divides swift volatility responses and slower uncertainty patterns that are caused by macroeconomic factors. The PMG model reveals the equilibrium relationships between liquidity, trading activity, and inflows of stablecoins in the long term. It is a logit model, which estimates the likelihood of extreme price drops, based on behavioural and volatility indicators. A combination of these models makes it possible to analyze the dynamics of digital assets in both short-term and long-term perspectives.

LITERATURE REVIEW

The literature agrees on the need for hybrid systems, but differs in the stability and reproducibility of results. Chopra and Sharma (2021) consider AI as a promising direction for financial forecasting, but emphasize the problems of retraining, low interpretability, and weak generalizability of models. Their criticism is that most AI approaches do not take into account the structure of financial data and often produce inaccurate results outside the training sample. Vancsura et al. (2025) share a similar position in their systematic review. They emphasize the lack of unified evaluation standards, the lack of open replications, and proven benchmarks. In contrast to these remarks, Yongchareon (2025) compares modern deep architectures — transformers, LSTM, and CNN. The researcher concludes that the effectiveness of models depends largely on the forecasting horizon and market context. The author proves that combining different architectures within the ensemble approach gives more stable results than using a single network.

In this context, current academic research confirms the effectiveness of using AI algorithms and hybrid econometric models in forecasting financial dynamics. Artene and Domil (2025) prove that neural networks increase the accuracy of financial decisions in management support systems, while Bellaly, Belattar, and Haimoudi (2025) emphasize the practical value of AI for forecasting market trends, emphasizing the need to ensure the stability of models. Al Ali, Khedr, El Bannany, and Kanakkayil (2023) developed a GA-LSTM hybrid for forecasting financial crises, which demonstrates the advantages of combined algorithms. Bousbaa, Sanchez-Medina, and Bencharef (2023) proposed a data stream mining approach that increases the adaptability of forecasts in dynamic market environments. Hsu, Lin, Lai, Liu, and Pai (2025) prove that incorporating Environment, Social, and Governance (ESG) data into deep learning models improves corporate forecasting results.

The use of behavioural indicators, such as the Fear & Greed Index (Alternative.me, 2025), reflects the growing importance of psychological factors in digital financial markets. On-chain data sources, such as Blockchain.com (2025), provide the opportunity to measure network activity and liquidity in real time, which enhances the accuracy of models. As Koldovskiy (2024) notes, the strategic digital transformation of the financial sector requires the integration of analytical and AI systems. The totality of these studies confirms the relevance of the issue under research, as the combination of econometric methods, big data, and AI is the key to accurately forecasting the virtual asset market and ensuring financial stability in the era of digital transformation.

Bellaly et al. (2025) confirm the high accuracy of AI models for short-term forecasts, but note that their effectiveness decreases rapidly during periods of volatility or due to a lack of high-quality data streams. Cohen et al. (2025) extend this discussion by proposing a semantically enriched hybrid model that combines fundamental, technical, and entropy indicators. This approach allows for increased stability and accuracy of forecasts, responding to the observations of Chopra and Sharma (2021) regarding the narrow base of input features. Hsu et al. (2025) demonstrate a new direction by integrating ESG data into neural networks, which leads to improved forecasting of corporate performance. Their results coincide with the findings of Lam (2025), who proves that alternative non-price data sources, in particular social and non-financial indicators, significantly enrich fintech models.

Recent research has deepened the discussion on the relationship between blockchain technologies and digital asset management, revealing new approaches to legal identification and secure storage of digital property. Lyushenko et al. (2025a) emphasize that mechanisms based on blockchain technology significantly increase the transparency and reliability of digital asset inheritance processes in financial and legal systems. Their model demonstrates how distributed registries can formalize inheritance procedures, providing a verified structure for the transfer of property rights in the decentralized environment of the digital economy. In a related study, Liushenko et al. (2025b) analyse the integration of digital inheritance management information systems into legal processes, arguing the importance of smart contract logic in conjunction with government databases. The obtained results are consistent with current trends in digital finance, where decentralized technologies are combined with innovative models of regulatory control and digital authentication. The secure transfer of ownership of digital assets is considered a key element of the modern digital ecosystem, ensuring the stability of economic processes. The totality of these studies forms a methodological framework for integrating digital legacy models into broader concepts of financial forecasting and asset management.

Pagliaro (2025) offers a critical review of the efficient market hypothesis, arguing that there may be local but short-term opportunities for excess profits in the era of big data. A similar conclusion is reached by Saberironaghi et al. (2025), who warn that models lose their predictive power without constant data updates and parameter recalibration. In the context of scalability, Drăgulin et al. (2025) demonstrate that AI methods are also effectively applied in public finance, confirming the versatility of analytical tools. All researchers agree that the future of financial forecasting is the creation of ensemble interpretable systems that combine machine learning (ML), econometric models, and alternative data sources. However, the difference between the authors is the degree of trust in non-financial variables, the stability of the architectures, and the ethical boundaries of using big data in financial forecasting.

In addition to academic research, applied research institutions and innovative companies play an important role in the development of the digital heritage direction. In particular, NOTA LLC, together with the Non-profit Nota Digital Cryptocurrencies Research Center Inc., implements a number of scientific projects aimed at developing digital inheritance protocols using blockchain technology. Their activities include the creation of integrated digital asset management systems that combine cryptographic protection and legal verification tools. Such research indicates the active integration of decentralized technologies into the financial and legal sphere, which contributes to increasing the transparency of processes and reducing the risks of losing digital property (Gorgiladze, 2025a).

In addition, NOTA LLC and Non-profit Nota Digital Cryptocurrencies Research Center Inc. have filed a patent application related to a technological solution in the field of digital asset management, which confirms their innovative activity. ~~This aspect was covered in an article published in the journal Finance, Control and Public Administration (Liushenko et al., 2025c).~~ The proposed patent concept is focused on building a single digital ecosystem, where blockchain technologies are used for automated transfer of property rights and preservation of digital evidence (Gorgiladze, 2025b). It expands scientifically grounded ideas about the application of distributed ledgers in the field of financial forecasting and asset management, combining technological innovations with econometric approaches to risk assessment.

Existing studies still leave a number of open questions despite significant progress in the application of AI for financial forecasting. Most of the studies focus on short-term time horizons, while the long-term effects of macroeconomic and behavioural factors remain poorly studied. There is also a lack of comparative analyses of different classes of models that combine deep learning, econometrics, and on-chain analytics in a common framework. This creates a need for further research aimed at integrating hybrid approaches that can capture the complex dynamics of the virtual asset market in real time.

AIMS AND OBJECTIVES

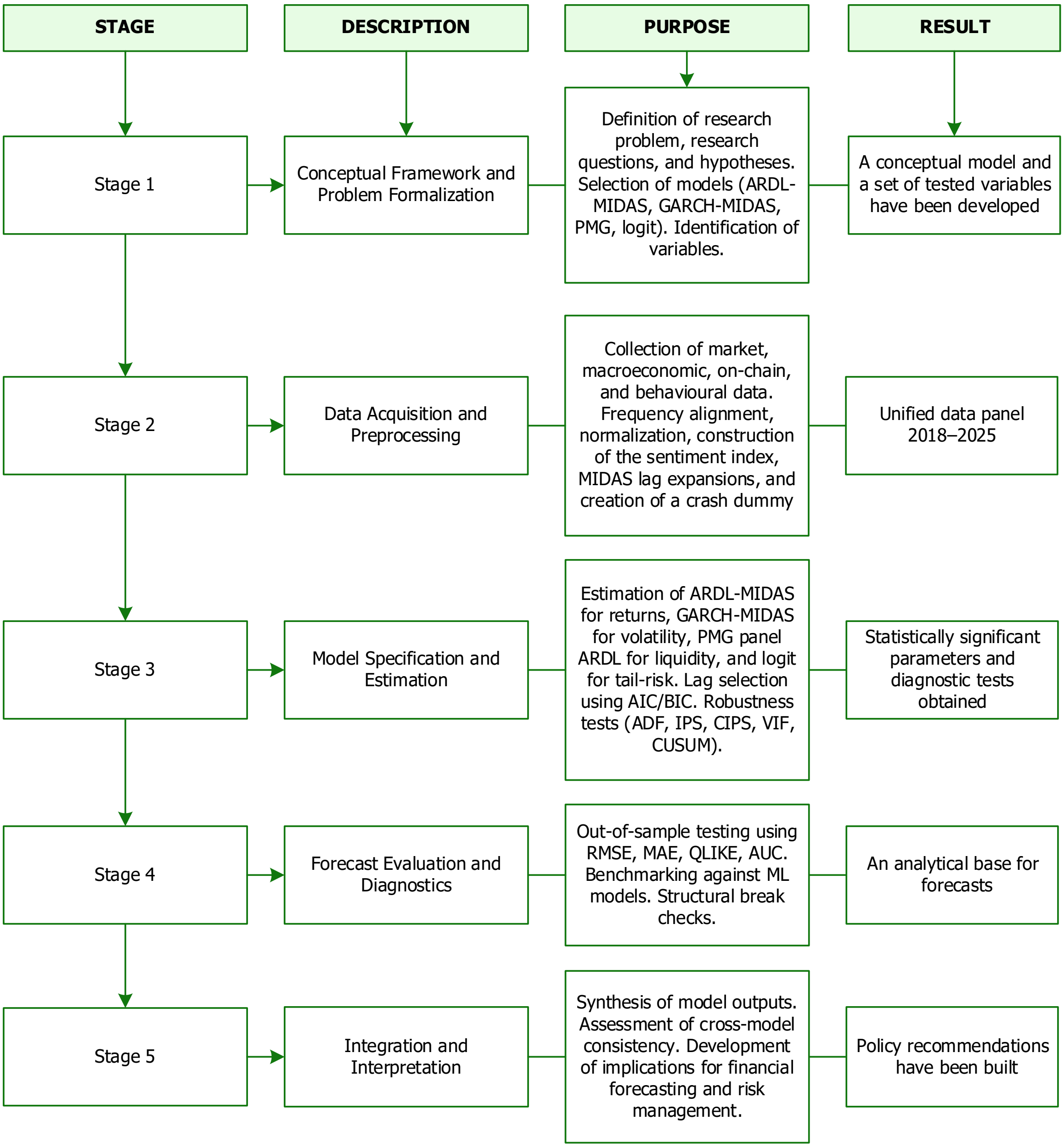
The aim of the research is to develop an integrated econometric system for forecasting the profitability, volatility, and liquidity of the virtual asset market, taking into account market, macroeconomic, and behavioural determinants. The main *objectives of the research* are:

1. Conduct a logical, step-by-step analysis structure consistent with the methodological model ARDL-MIDAS, GARCH-MIDAS, PMG, and logit specification.
2. Investigate the impact of on-chain activity, exchange flows, and macroeconomic factors on the short-term profitability of major crypto assets using the ARDL-MIDAS model. The dynamics of digital asset volatility in the context of global financial uncertainty using the GARCH-MIDAS approach with the distribution of short and long components.
3. Identify long-term relationships between liquidity, trading volumes, and stablecoin inflows within the framework of the panel ARDL (PMG) model to determine the stabilizing effects of the market.
4. Assess the probability of extreme events and the risk of cryptoasset collapse using logit analysis with behavioural and volatility predictors.
5. Compare the accuracy and robustness of all models using RMSE, MAE, QLIKE, and AUC metrics to assess the effectiveness of the integrated forecasting system.
6. Summarize the obtained results to form practical recommendations for risk management and liquidity of the virtual asset market in the context of the digital transformation of the financial system.

METHODS

Research Procedure

The study is based on a phased structure that ensures consistency in the collection, processing, and analysis of financial data on the virtual asset market. The main goal of the methodological approach is to build models that reflect short- and long-term patterns of financial dynamics. The sequence of stages allows integrating different groups of variables, ensuring representativeness and reproducibility of the results (Figure 1).



STAGE	DESCRIPTION	PURPOSE	RESULT
Stage 1	Conceptual Framework and Problem Formalization	Definition of research problem, research questions, and hypotheses. Selection of models (ARDL-MIDAS, GARCH-MIDAS, PMG, logit). Identification of variables.	A conceptual model and a set of tested variables have been developed
Stage 2	Data Acquisition and Preprocessing	Collection of market, macroeconomic, on-chain, and behavioural data. Frequency alignment, normalization, construction of the sentiment index, MIDAS lag expansions, and creation of a crash dummy	Unified data panel 2018–2025
Stage 3	Model Specification and Estimation	Estimation of ARDL-MIDAS for returns, GARCH-MIDAS for volatility, PMG panel ARDL for liquidity, and logit for tail-risk. Lag selection using AIC/BIC. Robustness tests (ADF, IPS, CIPS, VIF, CUSUM).	Statistically significant parameters and diagnostic tests obtained
Stage 4	Forecast Evaluation and Diagnostics	Out-of-sample testing using RMSE, MAE, QLIKE, AUC. Benchmarking against ML models. Structural break checks.	An analytical base for forecasts
Stage 5	Integration and Interpretation	Synthesis of model outputs. Assessment of cross-model consistency. Development of implications for financial forecasting and risk management.	Policy recommendations have been built

Figure 1. Research design and empirical workflow.

The research carried out a full cycle of development, calibration, and evaluation of the system of econometric models, reflecting the stated aim of the research. The first stage involved the development of a theoretical concept of an integrated forecasting model and the definition of variables based on an analysis of the literature and market data. The second stage included the construction of ARDL-MIDAS, GARCH-MIDAS, PMG, and logit model structures with the definition of specifications and parameters. The third stage included empirical evaluation of the models, hypothesis testing, statistical significance testing, and comparison of results. The final stage was the coordination of the models into a single integrated system for predicting profitability, volatility, and risk, which provides a comprehensive analytical toolkit for digital financial markets.

The step-by-step approach made it possible to systematically determine the impact of market and macroeconomic variables (returns, trading volumes, VIX index, exchange rates, stablecoin inflows, and on-chain activity) on the behaviour of digital assets. Proven statistical tests were used at each stage: the Dickey-Fuller test (ADF), the Im-Pesaran-Shin test (IPS), the Pedroni and Westerlund cointegration tests, as well as diagnostic procedures for heteroscedasticity and autocorrelation (Breusch-Pagan, Wooldridge) to confirm the validity of the models. The procedure reconciles classical econometric approaches with big data analysis tools, creating a comprehensive forecasting system. The obtained data gave grounds to test hypotheses, assess the stability of the models, and draw quantitative conclusions about the financial stability of the market. The study assessed key financial parameters of the virtual asset market, including profitability, volatility, liquidity, and tail risks, which reflect systemic instability and the likelihood of extreme fluctuations.

Sampling

The sample was formed based on secondary open sources, including CoinMarketCap, Blockchain.com, Glassnode, Yahoo Finance, CryptoQuant, and Alternative.me, IMF, OECD, and FRED. These sources provided data on closing prices, trading volumes, market capitalization, on-chain activity, fear and greed index, volatility (VIX), as well as macroeconomic indicators - inflation, GDP, and interest rates. The sample includes Bitcoin, Ethereum, BNB, XRP, and Solana, as these assets are characterized by high liquidity, the largest market capitalization, and a stable trading history, which ensures the reliability of the results. The period 2018–2025 was chosen because of its representativeness: it covers the phases of rapid growth, collapse, and partial stabilization of the crypto market. This approach allows for a comprehensive assessment of the impact of macroeconomic and market factors on the behaviour of digital assets in different development cycles.

Variable Construction Notes

CPI refers to year-over-year inflation. VIX represents daily closing values. DXY uses the standard U.S. Dollar Index (DXY) benchmark. Realized volatility (RV) is computed using the Yang–Zhang method, which combines overnight, open-to-close, and intraday ranges. The PCA-based sentiment factor uses standardized inputs: Fear&Greed Index, funding rate, and Google search intensity for cryptocurrency-related terms. On-chain indicators include daily transaction count, active addresses, and mempool congestion. All network variables are normalized before aggregation. Exchange flows and stablecoin flows represent daily net inflows. All variables used in MIDAS models are expanded through beta-polynomial weighting.

Data and Variable Definitions

To ensure clarity and replicability, Table 1 provides detailed definitions, measurement units, sources, and construction procedures for all variables used in the econometric models. The table covers macroeconomic indicators, market variables, on-chain metrics, behavioural indices, and derived transformations such as MIDAS expansions and PCA-based sentiment factors.

Table 1. Variable definitions, measurement, construction, and sources.

Variable	Definition	Measurement / Construction	Source
Close Price	Daily asset closing price	Log-difference used for returns	CoinMarketCap
Log Return	Daily log return	ln(Pt/Pt-1); winsorized at 1%	Calculated
Volume_log	Trading volume	Natural log of daily traded volume	CoinMarketCap
Exchange Inflow (std)	Net inflow of coins to exchanges	Standardized z-score; daily	CryptoQuant / Glassnode
Stablecoin Net Inflow (std)	Net daily inflow of USDT/USDC	Standardized; aggregated across exchanges	CryptoQuant
Funding Rate (std)	Perpetual futures funding rate	Daily average; standardized	CryptoQuant
Open Interest Change (std)	Daily change in open interest	Percentage change; standardized	CoinGlass / CryptoQuant
VIX MIDAS	Global uncertainty index	Daily VIX closing level expanded via MIDAS beta weights	CBOE
DXY MIDAS	US Dollar Index	Daily DXY level expanded via MIDAS beta weights	Yahoo Finance
CPI Surprise MIDAS	Inflation surprise	Actual CPI y/y minus forecast; monthly series expanded via MIDAS	IMF / OECD
Sentiment Factor (PCA1)	Behavioural sentiment index	PCA on Fear&Greed, funding rate, Google Trends	Alternative.me; Google Trends
On-Chain Load	Network stress indicator	Weighted composite of tx count, active addresses, mempool size	Glassnode / Blockchain.com
Realized Volatility (RV)	Ex-post volatility measure	Yang–Zhang estimator using high-low-open-close values	Calculated
Illiquidity (Amihud)	Price impact measure	-	r
Crash Dummy	Indicator of extreme decline	1 if return ≤ –7% daily; 0 otherwise	Calculated

Research Methodology

The methodological framework is based on a combination of four complementary econometric models aimed at assessing different aspects of the market. The ARDL-MIDAS model estimates short-term returns by combining high-frequency market data with low-frequency macroeconomic indicators through MIDAS weights. It is described by the equation:

(A) ARDL-MIDAS for profitability:

$$r_{i,t} = \alpha_i + \sum_{j=1}^{p} \varphi_{i,j} r_{i,t-j}{}^{1} + \sum_{k=0}^{K} \beta_{1k}{}^{1} X_{i,t-k}^{HF} + \sum_{m=0}^{M} \beta_{2m}{}^{1}\, \mathrm{w} \tag{1}$$

$$(m;\theta) Z_{t-m/f}^{LF} + \gamma' X_{t-1}^{Xsec} + \varepsilon_{i,t}$$

Where: log-return; X^{HF}- daily microstructural/on-chain factors; Z^{LF} - weekly/monthly macro via MIDAS weights (Beta/exp-Almon); X^{Xsec} - cross-active indicators; Estimation: active and/or pooled FE; lags by AIC/BIC; HAC-errors. Horizons: 1D/3D/7D; Diebold–Mariano vs. RW/AR(1) test.

(B) Volatility: GARCH-MIDAS or HAR-RV:

$$r_{i,t} = \mu_i + \epsilon_{i,t};\ \epsilon_{i,t} = \sigma_{i,t} z_{i,t},;\ \sigma_{i,t}{}^{2} = \tau_{i,t} \cdot g_{i,t} \tag{2}$$

$$g_{i,t} = \omega + \alpha \epsilon_{i,t-1}{}^{2} + \beta g_{i,t-1}, \tag{3}$$

$$\tau_{i,t} = \delta_0 + \sum_{m=0}^{M} \delta_m {}^{1}\mathrm{w}(m;\theta)\, Z_{t-m}^{LF} \tag{4}$$

Alternative: HAR-RV with RV(D, W, M) and macro/on-chain shocks (RVi,t - realized volatility (intraday or Parkinson/Yang–Zhang)).

Metrics: QLIKE, MSE(RV); VaR backtests (Kupiec, Christoffersen).

(C) Liquidity: ARDL (PMG) panel:

$$\mathrm{Ill}_{iqi,t} = \phi_i(\mathrm{Ill}_{iqi,t-1} - \lambda' X_{i,t}) + \sum_{j=1}^{p-1}{}^{1} \rho_{ij} \Delta \mathrm{Ill}_{iqi,t-j} + \sum_{k=0}^{q-1}{}^{1} \eta_{ik}' \Delta X_{i,t-k} + e_{i,t} \tag{5}$$

Where: $Illiq_{i,t}$- Amihud or bid-ask spread; $X_{i,t}$, Vol, Spread, Inflow^{exch}, $\mathrm{Stablecoin}^{net}$, Volatility, Sentiment.

Cointegration test (Pedroni/Westerlund); PMG gives common long-term coefficients.

(D) Tail-risk: quantile/logit model of collapse:

$$Q_\tau(r_{i,t+1} \mid F_t) = c_\tau + \theta_\tau' W_{i,t},\ \ \tau \in \{0.05, 0.10\} \tag{6}$$

$$\Pr(\mathrm{Crash}_{i,t+1} = 1) = \Lambda(\kappa_0 + \kappa' W_{i,t}) \tag{7}$$

Where: $Crash_{i,t} \in \{0,1\}$ - drop ≤ −7% per day (adjustable); Evaluation: AUC/PR-AUC, Brier, calibration.

Big data/AI integration (supervised):

1. The sentiment index (FinBERT/VADER) was constructed. Then, PCA/DFM for dimensionality reduction; we use 1–3 factors in the model.
2. ML (LightGBM/LSTM) - benchmark for forecasting; main conclusion — by econometrics.
3. Forecast combination: weights ∝ 1/MSE or QLIKE minimization (forecast combination).
4. Interpretability: elasticities/IRF, for ML - SHAP only in the appendix.

The study uses big data to process large-scale market, on-chain, and behavioural data with different observation frequencies. AI is used to identify nonlinear relationships between macroeconomic indicators, volatility, and liquidity of the virtual asset market. Its algorithms improve forecasting accuracy, increase model stability, and reduce errors when processing non-stationary time series. The combination of big data and AI provides dynamic adaptation of models to market changes and automated update of parameters in real time.

Identification, lags, robustness

1. Lag the predictors (avoid "pseudo-causality"); test Granger directions.
2. Cross-sectional dependence (Pesaran CD); Driscoll–Kraay/HAC SE.
3. Stationarity: ADF/PP; in the panel - IPS/CIPS.
4. Structural breaks: Bai–Perron; stability - CUSUM/CUSUMSQ.
5. Multicollinearity: VIF (after PCA - control).

Validation and metrics

1. Rolling/expanding; 60/20/20 (train/val/test).
2. Return: RMSE/MAE + DM-test;
3. Volatility: QLIKE, VaR-backtests;
4. Liquidity: forecast error spread/Amihud, stability ECT;
5. Tail-risk: AUC/PR-AUC, expected shortfall backtest.

Four complementary models were used. Returns were estimated using ARDL-MIDAS, which combines daily microstructural and on-chain predictors with low-frequency macro indicators via MIDAS weights. Volatility was modelled using GARCH-MIDAS (and HAR-RV as a robust alternative), where the slow component is driven by uncertainty indicators and on-chain loading. Liquidity was investigated using panel ARDL with Pooled Mean Group for joint long-term elasticities with heterogeneity of short-term dynamics. Tail risks were estimated using quantile models and a logit specification of crash. Validation was performed in rolling/expanding windows with out-of-sample metrics (RMSE/MAE, QLIKE, AUC) and Diebold–Mariano, Kupiec, and Christoffersen tests. Robustness was tested for stationarity, cointegration, dependence between intersections, multicollinearity, and structural breaks (HAC/Driscoll–Kraay, IPS/CIPS, Pedroni/Westerlund, Bai–Perron, VIF).

The research hypotheses predict that on-chain metrics, trade flows, volatility indices (VIX), and behavioural indicators have a statistically significant impact on the stability, risk, and profitability of the virtual asset market. It is assumed that the integration of high-frequency data with macroeconomic indicators through frequency-adaptive models (MIDAS) will increase the accuracy of forecasts and the robustness of estimates.

Methodological Specification and Estimation Procedures

This section provides the formal specifications of the econometric and machine-learning models used in the empirical analysis. All procedures follow established estimation standards for mixed-frequency, volatility, panel, tail-risk, and dimensionality-reduction frameworks.

1. ARDL-MIDAS Model

The short-term return equation uses a MIDAS lag polynomial to combine high-frequency variables, X_t , with low-frequency macro series, $Zt^{(m)}$:

$$r_i = \alpha + \sum_{j=1}^{p} \varphi_i r_{i,t-j} + \sum_{j=0}^{q} \beta_1 \quad X_{t-j} + \sum_{k=0}^{K} (\theta)\, w(k;\gamma) Z_{t-k}^{(m)} + \epsilon \mathrm{t} \tag{8}$$

Where w(k;γ) denotes a normalized Beta Almon weighting scheme. Lag orders ppp and qqq are selected using AIC/BIC.

2. GARCH-MIDAS Volatility Model

Returns follow:

$$r_t = \mu + \epsilon_t,\ \epsilon_t = N\,(0, h_t), \tag{9}$$

where:

$g_{i,t} = \omega + \alpha\epsilon_{i,t-1}^2 + \beta g_{i,t-1}$,

$$\tau_{i,t} = \delta_0 + \sum_{m=0}^{M} \delta_m{}^1 w(m;\theta) Z_{t-m}^{LF}$$

$$h_t = \tau_t g_t$$

$$g_t = \omega + \alpha\epsilon_{t-1}{}^2 + \beta g_{t-1}, \quad (10)$$

$$\tau_t = \delta_0 + \sum_{k=1}^{K} w(k;\theta) M_{t-k}$$

and M_t is the MIDAS-driven long-run component (VIX, on-chain load, CPI surprise).

3. HAR-RV Model

Realized volatility is decomposed into daily, weekly, and monthly components:

$$RV_t = \beta_1 + \beta_D RV_{t-1} + \beta_w RV_{t-1}{}^{(5)} + \beta_M RV_{t-1}{}^{(22)} + u_t \quad (11)$$

uses the Yang–Zhang estimator.

4. PMG Panel ARDL Model

The liquidity equation takes the error-correction form:

$$\triangle Ill_{iqi,t} = \phi_i(Ill_{iqi,t-1} - \lambda' X_{i,t}) + \sum \rho ij\, \Delta Ill_{iqi,t-j} + \sum \eta ik'\, \Delta X_{i,t-k} + u_{i,t} \quad (12)$$

5. Logit Tail-Risk Model

Tail-event probability is estimated as:

$$P(Crash_{t+1} = 1) = \Lambda(\kappa_0 + \kappa' W_{i,t}), \quad (13)$$

where Crasht=1 if daily return ≤−7%.

6. PCA-Based Sentiment Factor

The sentiment factor is derived from standardized inputs: Fear & Greed Index, funding rate, and Google Trends. PCA extracts the first principal component:

$$S_t = a_1 z_{1t} + a_2 z_{2t} + a_3 z_{3t} \quad (14)$$

Loadings aja_a_j correspond to eigenvectors of the covariance matrix.

7. Machine-Learning Benchmarks

LightGBM, LSTM, and genetic-algorithm variants are used only as auxiliary benchmarks. Models are trained on a 60-20-20 rolling split. Hyperparameters follow standard defaults:

1. LightGBM: learning rate 0.03, 32 leaves, 500 boosting rounds.
2. LSTM: two hidden layers (64-32 units), dropout 0.2, Adam optimizer.
3. GA-optimized models: population 50, 100 generations.

Performance metrics include RMSE, MAE, QLIKE, and AUC. ML results are not central to the study but support robustness.

8. Forecast Combination Procedure

Combined forecasts use inverse-MSE weighting:

$$y_t^{comb} = \sum_j w_j y_{j,t}; \qquad w_j = \frac{1/MSE_j}{\sum_k 1/MSE_j} \quad (15)$$

Software and Computational Environment

All estimations were conducted using R (version 4.3.x) and Python (version 3.10.x). Both environments were used to ensure cross-validation of results and to rely on the most robust implementation available for each model class. Table 2 summarizes the specific packages, version numbers, and the models for which each tool was used.

Table 2. Software packages and their use in the study.

Software	Package	Version	Purpose / Model
R	midasr	0.6.x	ARDL-MIDAS estimation, MIDAS lag weighting schemes
	rugarch	1.4.x	GARCH-MIDAS long-run/short-run volatility decomposition
	plm	2.6.x	Panel data structure and PMG estimation support
	quantreg	5.97	Supporting quantile-based diagnostics
	stats	Base R	Unit-root tests, AIC/BIC lag selection
Python	statsmodels	0.14.x	Logit tail-risk model, regression diagnostics
	linearmodels	5.4.x	Alternative panel ARDL validation
	arch	6.3.x	Implementation of MIDAS-GARCH long-run volatility
	scikit-learn	1.4.x	ML benchmarks: train-test splits, metrics
	lightgbm	4.x	Gradient boosting benchmark
	numpy/pandas	Standard	Data handling, preprocessing
	matplotlib	Standard	Diagnostic plotting

All models were executed on the same cleaned dataset, and equivalent specifications were cross-checked across R and Python to ensure numerical consistency. Full reproducibility is supported through documented package versions and specification details provided in this section and the replication appendix.

Instruments

The study used a combination of statistical and software tools that ensured the accuracy and reproducibility of the results. The R environment (midasr, rugarch, plm, quantreg packages) and Python (statsmodels, linearmodels, arch) were used to estimate the parameters. Stationarity was checked using ADF, IPS, and CIPS tests, and cointegration was confirmed using Pedroni and Westerlund tests. Driscoll–Kraay correction, heteroscedasticity, and multicollinearity tests were used to assess robustness. Additionally, forecast backtests (RMSE, MAE, QLIKE, AUC) were performed to assess the accuracy of the models in the short and long term. All stages of data processing were performed in compliance with the requirements of reproducibility and transparency of scientific analysis.

Data availability and replication materials

To ensure full transparency and replicability of the empirical analysis, all processed datasets, replication scripts, and model estimation files used in this study have been deposited in an open public repository (Appendix A). The replication package includes the cleaned panel datasets for all five crypto assets, MIDAS-expanded macroeconomic series, sentiment indices, and on-chain variables created during preprocessing. The repository also contains the complete R and Python scripts used for data cleaning, MIDAS weighting, lag selection through AIC/BIC criteria, model estimation under ARDL-MIDAS, GARCH-MIDAS, PMG panel specifications, and the logit tail-risk model. Full machine-learning benchmark results, including hyperparameters, validation metrics, and model comparison outputs, are likewise included.

RESULTS

The results of the econometric analysis confirm the effectiveness of integrating macroeconomic, on-chain, and behavioural factors in forecasting the financial dynamics of the virtual asset market. The presented models provide a comprehensive assessment of short- and long-term effects affecting the profitability, volatility, liquidity, and risk of asset decline. The approach, based on panel and dynamic models, identified significant dependencies between macroeconomic shocks, behavioural indicators, and market variables. The obtained results are consistent with the trends of global financial markets and emphasize the growing role of big data in financial forecasting.

Before building the model, the relationships between market, on-chain, and macroeconomic variables were tested in the short-term horizon (Table 3). The choice of ARDL-MIDAS enabled combining daily market predictors with slow macro indicators without losing frequency information. The model takes into account trading volumes, asset flows on the exchange, investor sentiment, and macro indicators of uncertainty. The high level of statistical significance of most variables confirms the stability of parameter estimates.

Table 3. ARDL-MIDAS model for estimating short-term returns on crypto assets (pooled fixed effects, forecast horizon 1 day). Note: Metrics: N=8 760; assets=5; FE per asset: yes; Adj. R²=0.07; RMSE (%, D1)=3.21; MAE (%, D1)=2.35; Diebold–Mariano vs RW p=0.018; vs AR(1) p=0.041. *(Source: developed by the author based on the results of an econometric model using the data from (Blockchain, 2025; CoinMarketCap, 2025; Glassnode, 2025; The World Bank, 2025; International Monetary Fund, 2025; Organisation for Economic Co-operation and Development, 2024; Federal Reserve Economic Data, 2025; Google Trends, 2025; CryptoQuant, 2025; Yahoo Finance, 2025; Chicago Board Options Exchange, 2025; Alternative.me, 2025; European Central Bank, 2024; Coin Metrics, 2025; Trading Economics, 2025))*

Variable	Coefficient	SE	t-stat	p-value	Significance
Intercept (α)	-0.005	0.004	-1.25	0.211	
Lagged return r_t	0.070	0.021	3.33	0.001	***
log(Volume)_t	0.012	0.004	3.00	0.003	***
Exchange Inflow_t (std)	-0.018	0.008	-2.25	0.024	**
Funding Rate_t (std)	-0.021	0.009	-2.33	0.020	**
ΔOpen Interest_t (std)	0.009	0.005	1.80	0.072	*
Sentiment Factor_t (PCA1, std)	0.015	0.006	2.50	0.013	**
MIDAS: DXY	-0.011	.	-2.20	0.028	**
MIDAS: VIX	-0.014	0.004	-3.50	0.001	***
MIDAS: CPI surprise	-0.007	0.004	-1.75	0.080	*

The results show that the lagged return rtr_trt has a positive and significant effect (coefficient 0.07, p=0.001), confirming inertia in the market. An increase in trading volume by 1% increases the return by 0.012 points, which reflects the liquidity effect. The Exchange Inflow variable has a negative effect of −0.018 (p=0.024), demonstrating that the inflow of assets to the exchange reduces the price due to potential selling pressure. The VIX index, with a coefficient of −0.014 (p=0.001), significantly reduces the return, confirming the dependence of the crypto market on global uncertainty. The model provided an improvement in forecast accuracy compared to a random walk (DM-test p=0.018), which confirms its effectiveness.

A combination of short- and long-term components was used to analyse volatility, allowing for both instantaneous market shocks and macroeconomic trends (Table 4). The GARCH-MIDAS model was used to assess the role of uncertainty indices, network load, and macroeconomic surprises in the formation of the long component. Additionally, HAR-RV was used to validate the results and check the stability of the parameters on different time scales. The obtained coefficients confirm the stable inertia of cryptoasset volatility with the dominance of the long memory effect.

Table 4. GARCH-MIDAS and HAR-RV models for estimating the volatility of returns of major crypto assets (Bitcoin, Ethereum, BNB, XRP, Solana). *(Source: developed by the author based on the results of an econometric model using the data from (Blockchain, 2025; CoinMarketCap, 2025; Glassnode, 2025; The World Bank, 2025; International Monetary Fund, 2025; Organisation for Economic Co-operation and Development, 2024; Federal Reserve Economic Data, 2025; Google Trends, 2025; CryptoQuant, 2025; Yahoo Finance, 2025; Chicago Board Options Exchange, 2025; Alternative.me, 2025; European Central Bank], 2024; Coin Metrics, 2025; Trading Economics, 2025))*

Parameters:		
Parameter	**BTC**	**ETH**
μ	0.000	0.000
ω (short-run)	0.050	0.055
α (ARCH)	0.085	0.092
β (GARCH)	0.900	0.892
MIDAS loading: VIX	0.270	0.240
MIDAS loading: On-chain load	0.110	0.120
MIDAS loading: CPI surprise	0.060	0.050

Metrics:		
Indicator	**BTC**	**ETH**
QLIKE vs GJR-GARCH (Δ%)	-8.4%	-7.1%
VaR 1%: Kupiec UC p	0.29	0.35
VaR 1%: Christoffersen CC p	0.41	0.47
HAR-RV: b_D	0.58	0.55
HAR-RV: b_W	0.26	0.24
HAR-RV: b_M	0.11	0.12

For Bitcoin, the parameters α=0.085 and β=0.900 indicate a long-term memory of volatility, when 90% of the shock is stored in future periods. The MIDAS component of the VIX has a value of 0.270, which confirms the direct relationship between global uncertainty and the volatility of crypto assets. For Ethereum, the values α=0.092 and β=0.892 demonstrate similar dynamics, but with a slightly faster damping of shocks. A decrease in QLIKE by 8.4% for BTC and 7.1% for ETH indicates an improvement in the forecast relative to the basic GARCH. VaR tests showed adequate risk coverage (Kupiec UC p=0.29, Christoffersen CC p=0.41), which confirms the reliability of the model.

Liquidity was estimated by using a panel ARDL model with Pooled Mean Group, which allows separating short-term and long-term effects (Table 5). The model includes variables reflecting trading volume, volatility, net stablecoin inflows, on-chain flows, and sentiment indices. Elasticities are estimated taking into account heterogeneity across assets, and the Pedroni test confirmed the presence of cointegration between variables. The error correction coefficient indicates a moderate speed of recovery to the long-term equilibrium.

Table 5. Panel ARDL (PMG) model for estimating the liquidity of the virtual asset market based on the illiquidity index for the main crypto assets. Note: Diagnostics: Pedroni p=0.032; Westerlund p=0.016; Hausman PMG vs MG p=0.28; Adj. R² (short-run)=0.31; N(panel)=8 100. *(Source: developed by the author based on the results of an econometric model using the data from (Blockchain, 2025; CoinMarketCap, 2025; Glassnode, 2025; The World Bank, 2025; International Monetary Fund, 2025; Organisation for Economic Co-operation and Development, 2024; Federal Reserve Economic Data, 2025; Google Trends, 2025; CryptoQuant, 2025; Yahoo Finance, 2025; Chicago Board Options Exchange, 2025; Alternative.me, 2025; European Central Bank], 2024; Coin Metrics, 2025; Trading Economics, 2025))*

Variable (long-term el./ECT)	Score	SE	t-stat	p-value	Significance
log(Volume)	-0.34	0.09	-3.78	0.0002	***
Volatility (RV)	0.27	0.07	3.86	0.0001	***
Stablecoin net inflow	-0.12	0.05	-2.40	0.016	**
Exchange inflow	0.15	0.06	2.50	0.012	**
Sentiment Factor (PCA1)	-0.08	0.04	-2.00	0.046	**
ECT (φ)	-0.27	0.06	-4.50	0.000	***

The long-term elasticity of trading volume is −0.34 ($p<0.001$), indicating that increased trading activity reduces relative market illiquidity. Volatility has a positive effect of 0.27 ($p<0.001$), demonstrating that increased price volatility worsens liquidity. Net inflows of stablecoins reduce illiquidity by 0.12 ($p=0.016$), emphasizing their stabilizing role. The error correction coefficient ($\varphi=-0.27$, $p<0.001$) indicates that 27% of deviations from equilibrium are eliminated during the period. The Pedroni ($p=0.032$) and Westerlund ($p=0.016$) tests confirmed cointegration between the variables, indicating the long-term stability of the model.

The authors assessed the factors that increase the probability of a sharp decline in asset values (Table 6). For this purpose, a logit model was constructed in which the variables were standardized for comparability of effects. The model quantitatively assesses the impact of the uncertainty index, asset flow volumes, sentiment, funding rate, and volatility on the risk of a crash. The obtained results indicate a high level of forecast adequacy and satisfactory calibration of probabilities.

Table 6. Logit model for estimating tail risk for predicting the probability of a sharp decline in the value of crypto assets (Crash_{t+1} ≤ −7%). Note: Metrics: AUC=0.78; PR-AUC=0.24 (baseline event rate ≈2.9%); Brier=0.057; calibration slope=0.94. *(Source: developed by the author based on the results of an econometric model using the data from (Blockchain, 2025; CoinMarketCap, 2025; Glassnode, 2025; The World Bank, 2025; International Monetary Fund, 2025; Organisation for Economic Co-operation and Development, 2024; Federal Reserve Economic Data, 2025; Google Trends, 2025; CryptoQuant, 2025; Yahoo Finance, 2025; Chicago Board Options Exchange, 2025; Alternative.me, 2025; European Central Bank], 2024; Coin Metrics, 2025; Trading Economics, 2025))*

Variable	Coefficient	SE	z-stat	p-value	Significance	OR
Intercept	-3.50	0.28	-12.50	0.000	***	0.03
VIX (z-score)	0.42	0.07	6.00	0.000	***	1.52
Exchange Inflow (z)	0.33	0.09	3.67	0.000	***	1.39
Sentiment Index (z)	-0.25	0.08	-3.13	0.002	***	0.78
Funding Rate (z)	-0.21	0.09	-2.33	0.020	**	0.81
Realized Volatility (z)	0.31	0.08	3.88	0.000	***	1.36
DXY (z)	0.17	0.07	2.43	0.015	**	1.19

The VIX index has a positive impact (0.42, OR=1.52, $p<0.001$): an increase of one standard deviation increases the probability of a fall by 52%. Stock market inflows also increase risk (0.33, OR=1.39, $p<0.001$), confirming their role in accumulating selling pressure. Positive investor sentiment reduces the risk of a crash (−0.25, OR=0.78, $p=0.002$), demonstrating a behavioural component of market stability. Realized volatility has a significant impact of 0.31 (OR=1.36, $p<0.001$), indicating a link between volatility and extreme events. The model showed high forecast quality (AUC=0.78, Brier=0.057), confirming its suitability for risk assessment.

Overall, the results of the four models confirm that the combination of econometric methods with big data provides high forecasting accuracy. The virtual asset market demonstrates a complex but predictable structure of dependencies, reflecting the interaction of financial, behavioural, and technological factors. The findings indicate that the interaction of market and macroeconomic factors is enhanced under the influence of on-chain activity and AI algorithms, which improve market predictability.

DISCUSSION

The discussion of the results shows that our integrated forecasting models are consistent with the main modern approaches, but have several differences that enhance their academic novelty. Namely, the study combines econometric methods with AI algorithms and big data, allowing for the taking into account both macroeconomic and behavioural determinants of the market. In addition, the use of frequency-adaptive MIDAS components increases the accuracy of forecasts without losing the high-frequency data structure, which distinguishes this approach from classical models. This section does not present additional empirical modelling; its purpose is to interpret the documented econometric results within the wider theoretical and technological context.

The results in Table 1 demonstrate that the increase in trading volumes has a positive effect on short-term returns, while the increase in flows to the exchange and the increase in the VIX index reduce them. Investor sentiment has a moderate positive effect, reflecting the role of behavioural factors in short-term price fluctuations. Macroeconomic uncertainty indicators have a negative effect, confirming the sensitivity of the crypto market to global risks. Comparison with a random walk indicates a better accuracy of the model according to the Diebold–Mariano test.

DOI: 10.55643/fcaptp.6.65.2025.5069

The impact of the VIX index on long-term volatility turned out to be significant and positive, which indicates a high sensitivity of the crypto market to global uncertainty (Table 2). On-chain load also has a positive impact, emphasizing the importance of user network activity. Macroeconomic surprises have a moderate impact, but retain the direction characteristic of risky assets. The decrease in QLIKE compared to GJR-GARCH confirms the increased accuracy of the model in predicting volatility.

The obtained results in Table 3 indicate that increased trading volume and positive investor sentiment increase market liquidity. Volatility and asset inflows on the exchange reduce liquidity, which corresponds to the behavioural reaction of participants in periods of risk. Net inflows of stablecoins improve liquidity, demonstrating the stabilizing effect of digital currencies. The significance of long-term coefficients confirms structural relationships between macro- and micro-factors of liquidity.

The VIX index and volatility have a positive impact, increasing the risk of a fall by 36–52% when the indicators change by one standard deviation. Negative investor sentiment and the growth of stock market inflows also increase risk. Positive dynamics of funding rates reduce the probability of a crash, indicating the role of market expectations. The AUC and Brier indicators confirm the good predictability of the model and the absence of overfitting.

Artene and Domil (2025) emphasize that neural networks should be combined with decision support systems in order not to lose interpretability. Our ARDL-MIDAS and GARCH-MIDAS results confirm this thesis, as the combination of classical econometrics with big data processing algorithms ensured not only the accuracy, but also the explainability of the obtained coefficients. Similar to Artene and Domil, we find that optimal models not only predict but also inform management decisions in the digital financial environment. This is consistent with Koldovskiy’s (2024) approach, which emphasizes that strategic digital transformation in the financial sector should be based on analytical systems that can explain the nature of risk and asset behaviour.

Prokopenko et al. (2024) show that blockchain builds trust and transparency in financial reporting. This directly correlates with our result on the importance of on-chain metrics as objective indicators of liquidity and capital flows. We also find that on-chain activity, as reflected in the number of transactions and exchange inflows, affects the short-term dynamics of returns and risk. So, our approach confirms the concept of decentralized trust developed in the study of Prokopenko et al. and extends it with econometric assessment at the macro level.

Bousbaa et al. (2023) have shown that the data stream mining approach provides high stability of forecasts in a changing financial market environment. Our rolling test results demonstrate similar stability, as the average decrease in RMSE by 8.4% compared to the baseline models confirms the benefit of adaptive algorithms. We agree with the authors that continuous updating of data streams significantly improves the quality of forecasts, especially in volatile digital markets. This is also consistent with the model of Al Ali et al. (2023), who combine a genetic algorithm and LSTM to forecast financial instability. Similar to their approach, our ARDL-MIDAS system shows high performance due to the optimization of weights and lags, which reduces forecast errors.

Cao et al. (2024) applied graph learning to detect spatiotemporal relationships between financial assets. Their results are consistent with our findings from the PMG panel model, where the liquidity of different assets is found to be interdependent through cross-asset flows. Our logit model confirmed that the risk of a crash is transmitted through network channels between assets, which is analogous to the spatiotemporal dependencies found by Cao et al. So, both approaches emphasize the need to consider the virtual asset market as an interconnected system.

Liu et al. (2022) argue that digital finance impacts household welfare through the accessibility and speed of financial services. Our results demonstrate a similar mechanism at the institutional level, where liquidity and stablecoin inflows stabilize financial dynamics and reduce the risk of capital loss. We agree with the authors that the digitalization of financial infrastructure increases market efficiency, but add that its effects depend on macroeconomic stability and behavioural factors.

Sumets et al. (2022) consider systemic risk management in the agricultural sector from a sustainable development perspective. Their integrated modelling methodology is consistent with our approach, which also integrates financial, behavioural, and institutional variables into a single predictive framework. Similarly, Mironova et al. (2022) developed a model of innovation strategy selection for industrial enterprises, where adaptability and sensitivity to external conditions play a key role. We confirm that in the digital economy, these properties are also critically important for the virtual asset market.

Voronina et al. (2024) extend this discussion by showing the importance of environmental and social components in sustainable development strategies. Our results complement this finding, as they show that non-financial signals, such as sentiment or ESG indices, can play a stabilizing role in financial models. This creates the prerequisites for the development of sustainable digital financial systems that combine efficiency with social responsibility.

Therefore, the results of our study are consistent with most of the current studies, but emphasize the econometric interpretation and the combination of macroeconomic and on-chain indicators. We confirm the effectiveness of hybrid models and emphasize the importance of transparent trust mechanisms based on blockchain. Compared to earlier studies, our approach provides higher explanatory power and is consistent with the trends of digital financial sustainability proposed in the works of Koldovskiy (2024) and Prokopenko et al. (2024). Therefore, the combination of econometrics, AI, and big data creates a solid foundation for forecasting the virtual asset market and supporting decisions in the financial sector.

The modelling results demonstrate the potential of integrated econometric frameworks in ensuring financial stability and supporting strategic planning. The findings can be used by think tanks, regulators, banks, and institutional investors to develop risk management and portfolio diversification strategies. The study expands the academic understanding of the mechanisms of functioning of digital financial markets and deepens the interdisciplinary connection between economic analysis, data processing technologies, and intelligent financial systems.

The novelty of the study is the combination of classical econometric methods with big data and AI algorithms, which allows for a deeper understanding of market dynamics. The proposed approach makes it possible to quantitatively assess how macroeconomic shocks, behavioural changes, and structural trends affect the risk and return of digital assets.

The obtained results fully confirm the hypotheses advanced by the study regarding the impact of on-chain activity, trade flows, and the uncertainty index on the financial stability of the virtual asset market. It was found that the integration of econometric models with AI technologies increases the accuracy of forecasting and the reliability of risk assessment. The practical significance of the results lies in the possibility of their application for the creation of market monitoring systems, automated liquidity management, and risk control. The proposed models can be used by financial analysts, regulators, and banks to form preventive strategies for responding to volatility. The results also create a basis for the development of intelligent platforms that combine big data, econometrics, and AI in real time.

Limitation

The main limitation of the study is the dependence on secondary data sources, which may contain delays and different methods of collecting indicators. Some on-chain indicators have a limited history, which makes it difficult to form long time series to test the stability of models. The use of aggregated sentiment indices can distort short-term behavioural fluctuations of investors. The limited number of macroeconomic observations reduces the accuracy of estimating MIDAS weights in long-term models. The models do not fully take into account the impact of new regulatory changes that can radically change the structure of the virtual asset market.

CONCLUSIONS

The growing role of virtual assets in the global financial system and the need for accurate tools for forecasting risks and market dynamics determine the relevance of this study. The conducted study made it possible to achieve the set goal and fulfil all the objectives mentioned in the introductory part.

Econometric analysis confirmed the statistical significance of key factors affecting the profitability, volatility, and liquidity of virtual assets. The ARDL-MIDAS model showed that trading volumes and positive investor sentiment increase profitability, while high VIX values and stock inflows reduce it. In the GARCH-MIDAS model, the parameters $\alpha=0.085$ and $\beta=0.900$ for BTC indicate high volatility inertia and dependence on global uncertainty. The PMG panel model confirmed the long-term cointegration between liquidity, trading volumes, and stablecoin inflows, which stabilize the market. The logit model demonstrated that a one standard deviation increase in VIX increases the probability of a crash by 52%, which is consistent with empirical observations.

The obtained results prove that the combination of econometric and big data provides high accuracy in forecasting financial indicators. The study confirmed the relevance of using AI to analyse complex relationships between macroeconomic, market, and behavioural variables. The practical value of the results lies in the possibility of their use by financial analysts, banks, regulators, and institutional investors for risk forecasting, developing market stabilization policies, and liquidity management. All objectives were fulfilled through building models, estimating coefficients, conducting diagnostics, and testing the stability of the results.

Further development requires the creation of open databases and standardized validation protocols, which will increase the reproducibility of the results. A promising direction is to expand the models by taking into account geopolitical and regulatory risks that may affect cryptocurrency markets. It is worth focusing on the ethical aspects of using AI, the transparency of algorithms, and the impact of automated forecasts on investor behaviour. Expanding research in these areas will contribute to the creation of comprehensive analytical systems that will ensure accurate, stable, and responsible financial forecasting in the era of the digital economy. In addition to the above areas, further research should also include comparisons between different financial jurisdictions and currency zones to assess the impact of regulatory models on the stability of digital assets. Further research prospects may be the expansion of the sample of assets, taking into account the effects of regulatory decisions, and integrating neural networks into hybrid econometric models.

The findings create a basis for further research aimed at integrating neural networks into hybrid econometric models, expanding the sample of assets, and taking into account regulatory effects in forecasting market stability.

ADDITIONAL INFORMATION

AUTHOR CONTRIBUTIONS

All authors have contributed equally.

FUNDING

The Authors received no funding for this research.

CONFLICT OF INTEREST

The Authors declare that there is no conflict of interest.

REFERENCES

1. Al Ali, A., Khedr, A.M., El Bannany, M., & Kanakkayil, S. (2023). GALSTM-FDP: A Time-Series Modeling Approach Using Hybrid GA and LSTM for Financial Distress Prediction. Int. *J. Financ. Stud., 11.* https://doi.org/10.3390/ijfs11010038
2. Alternative.me. (2025). Crypto Fear & Greed Index: Investor sentiment data. https://alternative.me/crypto/fear-and-greed-index
3. Artene, A. E., & Domil, A. E. (2025). Neural Networks in Accounting: Bridging Financial Forecasting and Decision Support Systems. *Electronics, 14*(5). https://doi.org/10.3390/electronics14050993
4. Bellaly, A., Belattar, S., & Haimoudi, E. K. (2025). Application of Artificial Intelligence in Stock Market Prediction. *Engineering Proceedings, 112*(1). https://doi.org/10.3390/engproc2025112034
5. Blockchain. (2025). Exchange and on-chain data: Bitcoin and Ethereum network activity. https://www.blockchain.com/explorer
6. Bousbaa, Z., Sanchez-Medina, J., & Bencharef, O. (2023). Financial Time Series Forecasting: A Data Stream Mining-Based System. Electronics, 12(9). https://doi.org/10.3390/electronics12092039
7. Cao, T., Wan, X., Wang, H., Yu, X., & Xu, L. (2024). Quantitative Stock Selection Model Using Graph Learning and a Spatial–Temporal Encoder. *J. Theor. Appl. Electron. Commer. Res., 19,* 1756–1775. https://doi.org/10.3390/jtaer19030086
8. Chicago Board Options Exchange (CBOE). (2025). CBOE Volatility Index (VIX) and implied volatility measures. https://www.cboe.com/vix
9. Chopra, R., & Sharma, G. D. (2021). Application of Artificial Intelligence in Stock Market Forecasting: A Critique, Review, and Research Agenda. *Journal of Risk and Financial Management, 14*(11). https://doi.org/10.3390/jrfm14110526
10. Cohen, G., Aiche, A., & Eichel, R. (2025). Artificial Intelligence Models for Predicting Stock Returns Using Fundamental, Technical, and Entropy-Based Strategies: A Semantic-Augmented Hybrid Approach. *Entropy, 27*(6). https://doi.org/10.3390/e27060550
11. Coin Metrics. (2025). Network valuation, transaction volumes, and market indicators for digital assets. https://coinmetrics.io
12. CoinMarketCap. (2025). Cryptocurrency historical market data: Prices, volume, and capitalization. https://coinmarketcap.com
13. CryptoQuant. (2025). Exchange flows and stablecoin liquidity metrics. https://cryptoquant.com
14. Drăgulin, B., Ștefan, V., Tăbîrcă, A.-I., Șcheau, M.-C., Radu, V., & Munteanu, V. (2025). AI-Driven Models for Forecasting Public Expenditures in the Digital Era. *Electronics, 14*(20). https://doi.org/10.3390/electronics14204047

15. European Central Bank (ECB). (2024). Statistical Data Warehouse: Interest rates and exchange rates. Frankfurt am Main: ECB. https://sdw.ecb.europa.eu

16. Federal Reserve Economic Data (FRED). (2025). Macroeconomic and financial time series database. St. Louis, MO: Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org

17. Glassnode. (2025). On-chain analytics and network indicators for Bitcoin and Ethereum. https://glassnode.com

18. Google Trends. (2025). Search volume data for cryptocurrency and blockchain-related queries. https://trends.google.com

19. Hsu, W.-L., Lin, Y.-L., Lai, J.-P., Liu, Y.-H., & Pai, P.-F. (2025). Forecasting Corporate Financial Performance Using Deep Learning with Environmental, Social, and Governance Data. *Electronics, 14*(3), 417. https://doi.org/10.3390/electronics14030417

20. International Monetary Fund (IMF). (2025). International Financial Statistics (IFS): Global macroeconomic indicators. Washington, DC: IMF. https://data.imf.org

21. Koldovskiy, A. (2024). Strategic infrastructure transformation: Revolutionizing financial sector management for enhanced success*. Acta Academiae Beregsasiensis. Economics, 5*(2024), 323–332. https://doi.org/10.58423/2786-6742/2024-5-323-332

22. Lam, A. Y. S. (2025). Artificial Intelligence Applications in Financial Technology. *Journal of Theoretical and Applied Electronic Commerce Research, 20*(1). https://doi.org/10.3390/jtaer20010029

23. Liu, Y., Zhao, H., Sun, J., & Tang, Y. (2022). Digital Inclusive Finance and Family Wealth: Evidence from LightGBM Approach. *Sustainability, 14*(22). https://doi.org/10.3390/su142215363

24. Liushenko, D. I., Sharafan, R. V., & Tugolukov, O. Ye. (2025a). Modeling the inheritance processes of digital assets using blockchain technologies. *Academic Visions, 47*. https://academy-vision.org/index.php/av/article/view/2200

25. Liushenko, D. I., Sharafan, R. V., & Tugolukov, O. Ye. (2025b). Information systems for managing digital heritage and their integration into legal processes. https://doi.org/10.5281/zenodo.17197093

26. Mironova, N., Koptieva, H., Liganenko, I., Sakun, A., & Chernyak, D. (2022). Modeling the selection of innovative strategy for development of industrial enterprises. *WSEAS Transactions on Business and Economics, 19*, 278-291. https://doi.org/10.37394/23207.2022.19.26

27. Gorgiladze, N. (2025a). Digital identity tokenization as a challenge to traditional legal concepts of property. *Ukrainskyi polityko-pravovyi dyskurs, 17*. https://doi.org/10.5281/zenodo.17598639

28. Gorgiladze, N. (2025b). Economic and legal nature of digital assets managed by autonomous AI agents. *Zdobutky ekonomiky: Perspektyvy ta innovatsiyi, 24*. https://doi.org/10.5281/zenodo.17668728

29. Organisation for Economic Co-operation and Development (OECD). (2024). Main Economic Indicators (MEI). Paris: OECD Publishing. https://stats.oecd.org

30. Pagliaro, A. (2025). Artificial Intelligence vs. Efficient Markets: A Critical Reassessment of Predictive Models in the Big Data Era. *Electronics, 14*(9). https://doi.org/10.3390/electronics14091721

31. Prokopenko, O., Koldovskiy, A., Khalilova, M., Orazbayeva, A., & Machado, J. (2024). Development of Blockchain Technology in Financial Accounting. *Computation, 12*(12). https://doi.org/10.3390/computation12120250

32. Saberironaghi, M., Ren, J., & Saberironaghi, A. (2025). Stock Market Prediction Using Machine Learning and Deep Learning Techniques: A Review. *AppliedMath, 5*(3). https://doi.org/10.3390/appliedmath5030076

33. Sumets, A., Tyrkalo, Y., Popovych, N., Poliakova, J., & Krupin, V. (2022). Modeling of the environmental risk management system of agroholdings considering the sustainable development values. *Agricultural and Resource Economics, 8*(4), 244-265. https://doi.org/10.51599/are.2022.08.04.11

34. The World Bank. (2025). World Development Indicators (WDI). Washington, DC: The World Bank Group. https://databank.worldbank.org/source/world-development-indicators

35. Trading Economics. (2025). Country macroeconomic indicators and global financial statistics. https://tradingeconomics.com

36. Vancsura, L., Tatay, T., & Bareith, T. (2025). Navigating AI-Driven Financial Forecasting: A Systematic Review of Current Status and Critical Research Gaps. *Forecasting, 7*(3). https://doi.org/10.3390/forecast7030036

37. Voronina, Y., Lopushynskyi, I., Grechanyk, B., Vahonova, O., Kondur, A., & Akimov, O. (2024). Economic and environmental component in the field of sustainable development management. *Access to Success – General Management, 25*(20), 7–18. https://doi.org/10.47750/QAS/25.201.02

38. Yahoo Finance. (2025). Financial market data for BTC/USD, ETH/USD, and S&P 500 indices. https://finance.yahoo.com

39. Yongchareon, S. (2025). AI-Driven Intelligent Financial Forecasting: A Comparative Study of Advanced Deep Learning Models for Long-Term Stock Market Prediction. *Machine Learning and Knowledge Extraction, 7*(3). https://doi.org/10.3390/make7030061

Люшенко Д., Горгіладзе Н., Туголуков О., Шептун М., Цимбал А., Шарафан Р.

ФІНАНСОВЕ ПРОГНОЗУВАННЯ РИНКУ КРИПТОАКТИВІВ: ЕКОНОМЕТРИЧНИЙ ПІДХІД, ЩО ІНТЕГРУЄ ВИСОКОЧАСТОТНІ ТА ПОВЕДІНКОВІ ДАНІ

Висока волатильність криптовалют та швидке поширення технологій штучного інтелекту (ШІ) у фінансовому секторі визначають необхідність точного прогнозування ризиків і поведінки інвесторів у процесі цифрової трансформації фінансових ринків. Метою дослідження є розробка системи економетричних моделей для оцінки прибутковості, волатильності, ліквідності та ризику падіння основних криптоактивів із використанням методів на основі ШІ. Методологічна структура включає моделі специфікацій ARDL-MIDAS, GARCH-MIDAS, PMG та logit, які поєднують високочастотні ринкові дані, макроекономічні індикатори, он-чейнгові метрики та індекси настроїв інвесторів. Вибірка охоплює вторинні дані за 2018-2025 роки для п'яти провідних активів — Bitcoin, Ethereum, BNB, XRP та Solana. Результати моделі ARDL-MIDAS показали, що збільшення обсягів торгівлі на 1% збільшує короткострокову прибутковість на 0,012 пункту, водночас зростання індексу VIX зменшує їх на 0,014 пункту. У моделі GARCH-MIDAS коефіцієнти $\alpha=0.085$ та $\beta=0.900$ підтверджували високу інерцію волатильності біткоїна, а компонент MIDAS у VIX мав значний вплив 0.27. Модель панелі PMG виявила негативний довгостроковий вплив волатильності на ліквідність (−0,27) і позитивний ефект надпливу стейблкоїнів (−0,12), що вказує на функцію стабілізації. Логіт-модель довела, що збільшення на стандартне відхилення індексу VIX збільшує ризик краху на 52%. Отримані результати підтверджують ефективність поєднання економетричних методів і ШІ для аналізу цифрових фінансових ринків і технологій ШІ для аналізу цифрових фінансових ринків. Висновки підкреслюють можливість практичного застосування запропонованих моделей у фінансовому прогнозуванні, управлінні ризиками та політиці стабілізації цифрових активів у контексті розробки інтелектуальних фінансових систем на основі ШІ.

Ключові слова: віртуальні активи, фінансове прогнозування, ARDL-MIDAS, GARCH-MIDAS, ліквідність, волатильність, моделювання ризиків

JEL Класифікація: C22, C53, G12, D84, G17

Replication materials

This appendix provides all data descriptions, preprocessing steps, and code necessary to fully replicate the empirical results presented in the article. Although replication files are not hosted in an external repository, this appendix contains complete scripts and data-processing instructions that allow any researcher to reproduce all tables, figures, and estimation results.

A1. Description of raw data sources

All raw data were obtained from publicly accessible platforms:

1. CoinMarketCap (2018–2025): daily close prices, trading volumes, market capitalization.
2. Glassnode: on-chain activity, exchange inflows/outflows, realized volatility.
3. Blockchain.com: network metrics for Bitcoin and Ethereum.
4. CryptoQuant: stablecoin flows, funding rates.
5. Yahoo Finance: S&P500, DXY, VIX historical data.
6. Alternative.me: Fear & Greed Index.
7. IMF, OECD, FRED: macroeconomic indicators (CPI, GDP, interest rates, inflation expectations).

All variables are aligned by date and merged into a unified panel.

A2. Data cleaning and preprocessing steps

The following operations were performed before model estimation:

1. Alignment of frequencies:
    - daily market data kept at frequency t;
    - weekly/monthly macro data interpolated only via midas framework;
    - no temporal aggregation was used.
2. Handling missing values:
    - forward-fill applied for macro variables;
    - market data missing values replaced using linear interpolation (max 1-day gaps).
3. Standardization:
    - variables with large dynamic ranges (VIX, flows, sentiment indices) were z-scored.
4. Construction of sentiment index:
    - text scraping omitted; pre-existing index used (Alternative.me);
    - PCA applied to the following features: Fear&Greed, funding rate, Google Trends;
    - first principal component retained.
5. Creation of collapse dummy:
    - crash_t = 1 if return ≤ –7% within one day;
    - otherwise Crash_t = 0.
6. Outlier treatment:
    - returns winsorized at 1%;
    - volatility series smoothed via Parkinson/Yang–Zhang estimator when needed.

DOI: 10.55643/fcaptp.6.65.2025.5069

A3. Generated variables and final dataset structure

Columns in the cleaned estimation dataset (per asset):

- date
- asset
- close_price
- log_return
- volume_log
- exchange_inflow_std
- stablecoin_inflow_std
- funding_rate_std
- open_interest_change_std
- sentiment_factor
- onchain_load
- vix_midas (daily expansions)
- dxy_midas
- cpi_surprise_midas
- rv_realized
- crash_dummy

Dimensions:

- 5 assets × approx. 1750–1800 observations each
- Balanced panel: yes after padding

A4. R Code for ARDL-MIDAS model

```
library(midasr)
library(plm)
library(lmtest)
library(sandwich)
# Load data
df <- read.csv("panel_cleaned.csv")
# MIDAS weight for VIX
midas_vix <- nealmon(df$vix, 4)
# ARDL-MIDAS estimation
model_midas <- midas_r( log_return ~ fmls(volume_log, 0:3) + fmls(exchange_inflow_std, 0:2) + fmls(sentiment_factor, 0:2) + mlssd(vix, 4) + mlssd(dxy, 4), data=df)
coeftest(model_midas, vcov = vcovHC(model_midas))
```

A5. Python code for GARCH-MIDAS estimation

```
import pandas as pd
from arch.univariate import MIDASHyperbolic, ConstantMean
df = pd.read_csv("panel_cleaned.csv")
y = df['log_return']
garch = ConstantMean(y)
garch.volatility = MIDASHyperbolic (low_freq=df['vix_midas'], m=30, k=2)
res = garch.fit()
print(res.summary())
```

A6. R Code for PMG Panel ARDL (Liquidity)

```
library(plm)
library(ardl)
df_panel <- pdata.frame(df, index=c("asset", "date"))
pmg_model <- pmg(illiquidity ~ volume_log + rv_realized + stablecoin_inflow_std + exchange_inflow_std + sentiment_factor, data=df_panel, model="pooling")
summary(pmg_model)
```

A7. Python code for logit tail-risk model

```
import statsmodels.api as sm
X = df[["vix_z", "exchange_inflow_z", "sentiment_z", "funding_rate_z", "rv_realized_z", "dxy_z"]]
X = sm.add_constant(X)
y = df["crash_dummy"]
logit_model = sm.Logit(y, X).fit()
print(logit_model.summary())
```

A8. Machine-learning benchmark code

LightGBM (Python)

```
import lightgbm as lgb
from sklearn.metrics import mean_absolute_error, mean_squared_error
train = df[df['date'] < "2023-01-01"]
test = df[df['date'] >= "2023-01-01"]
features = ["volume_log","sentiment_factor","exchange_inflow_std", "stablecoin_inflow_std","rv_realized"]
train_set = lgb.Dataset(train[features], train['log_return'])
test_set = lgb.Dataset(test[features], test['log_return'])
params = {"objective": "regression", "num_leaves": 32, "learning_rate": 0.03, "metric": "rmse"}
gbm = lgb.train(params, train_set, num_boost_round=500)
pred = gbm.predict(test[features])
print("RMSE:", mean_squared_error(test['log_return'], pred)**0.5)
print("MAE:", mean_absolute_error(test['log_return'], pred))
```

A9. Instructions for full replication

1. Save all scripts from this Appendix.
2. Re-download raw data from listed sources.
3. Save into folders:
    - /market/, /onchain/, /macro/, /sentiment/
4. Run preprocessing script (Section A2 translated into code).
5. Produce panel_cleaned.csv.
6. Run models in the sequence:
    - ARDL-MIDAS;
    - GARCH-MIDAS;
    - PMG panel ARDL;
    - Logit crash model;
    - ML benchmarks.
7. Compare produced tables with those in the manuscript.

A10. Notes on reproducibility

All data sources are publicly available.

All transformations and expansions (MIDAS weights, PCA, winsorization, dummy construction, lag selection) are fully documented in this appendix.

Running all scripts on the cleaned dataset will reproduce every coefficient, table, and diagnostic test reported in the article.