OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

Horizon-Dependent Demand Forecasting in High-Frequency Cryptocurrency Markets

Dr. Subhajyoti Pal
Department of Mathematics, Jhargram Raj College Jhargram, West Bengal, India

Abstract
Short-horizon demand forecasting is an important input to operational decision making in con- tinuously operating electronic markets.While a large literature has examined high-frequency price dynamics in cryptocurrency markets, comparatively less attention has been given to fore- casting transactional demand at very short horizons and to understanding how predictive per- formance changes as the forecasting horizon increases.This paper studies horizon-dependent demand forecasting using one year of minute-level data for three major cryptocurrencies:Bit- coin, Ethereum and Binance Coin.Baseline forecasting models based on historical demand and calendar effects are compared with augmented models that incorporate simple system state in- dicatorsderivedfromcontemporaneouspriceinformation.Forecastingperformanceisevaluated at one-minute, five-minute and fifteen-minute horizons using standard accuracy metrics.The empiricalresultsshowthatsystemstateindicatorsprovideconsistentimprovementsinforecast- ingaccuracyatallhorizons,butthatthemagnitudeofimprovementdeclinessystematically as the horizon lengthens. The findings highlight the importance of horizon-aware evaluation in short-term forecasting and suggest that parsimonious state indicators can offer practical valuein high-frequency demand prediction tasks.
Keywords:*Demandforecasting;Cryptocurrencymarkets;High-frequencydata*
MSC2020:*62M10,90B05,91B84*

Date of Submission: 15-12-2025 Date of acceptance: 31-12-2025

I. Introduction

Short-horizon demand forecasting is a central component of operational decision-making in elec- tronic systems that operate continuously and respond to rapidly evolving activity.In such envi- ronments, forecasts are required at granular time scales, often measured in minutes rather than hours or days and are repeatedly updated as new information becomes available.Even modest improvements in short-horizon predictive accuracy can translate into meaningful operational gains when forecasts are used frequently for monitoring, capacity allocation and real-time control.

A substantial forecasting literature documents that high-frequency activity exhibits strong tempo- ral regularities, including intraday seasonality, short-term persistence and calendar effects.Clas- sical forecasting frameworks emphasize the effectiveness of history-based predictors and temporal indicators for short-horizon forecasting tasks [1,2].Empirical studies further show that recent observationsoftenexplainasignificantportionofnear-termvariability, motivatingbaselinemodels built around lagged demand and moving-average structures [3,4].

At the same time, high-frequency systems are inherently state dependent.The contemporaneous operating condition of a system—reflected in short-term variability, dispersion and intensity ofobservedsignals—capturesinformationabouttheinteractionofheterogeneousparticipantsandthecurrentbalancebetweensupplyanddemand.Infinancialmarkets,suchstateinformationhas beenshowntoevolvegraduallyandtoinfluenceshort-horizondynamicsbeyondwhatiscaptured by historical averages alone [10,7,8].

Recent advances in empirical forecasting increasingly combine traditional time-series features with machine learning techniques to exploit nonlinear relationships and interaction effects.Tree-based ensemble methods, such as random forests and gradient boosting, have demonstrated strong per- formanceacrossawiderangeofappliedforecastingproblemsduetotheirflexibilityandrobustness [5,6]. However,large-scaleempiricalcomparisonsemphasizethatcarefulout-of-sampleevaluation remainsessentialtoensurethatreportedimprovementsreflectgenuinepredictivevalueratherthan over-fitting [4].

Cryptocurrencymarketsprovideanaturallaboratoryforstudyingshort-horizondemandforecasting under continuous operation.These markets trade on a 24/7 basis, exhibit substantial trading activityandexperiencerapidtransitionsbetweenroutineandstressedconditions.Unliketraditional financial markets, they do not exhibit opening or closing effects, allowing intraday dynamics to be analyzed uniformly across the entire day.Existing empirical studies on cryptocurrencies focus primarily on price behavior, volatility and market efficiency [11,12,13,14,15], while transactional demand has received comparatively less attention.

An additional dimension that remains under-explored is the role of the forecasting horizon itself. Forecastingdemandoneminuteaheaddiffersfundamentallyfromforecastingcumulativedemand over longer horizons such as five or fifteen minutes.Temporal aggregation alters the signal-to-noise ratio,therelevanceofcontemporaneousinformationandtheeffectivepersistenceofdemand.As a result, evaluating forecasting models at a single horizon provides only a partial view of their operational usefulness.

Theobjectiveofthisstudyistoexaminewhethersimpleandeasilycomputablesystemstate indicators improve short-horizon demand forecasting performance and to characterize how their predictive contribution varies systematically across forecasting horizons.Using one year of minute-level data for three highly liquid cryptocurrency assets—Bitcoin, Ethereum and Binance Coin—we compare baseline history-based models with augmented models that incorporate contemporaneous state information.The analysis is conducted across one-minute, five-minute and fifteen-minutehorizonsusinglinearregressionandnonlinearensemblemethods.

The contribution of this paper is empirical and methodological.First, it provides evidence that parsimonioussystemstateindicatorsyieldconsistentreductionsinforecastingerroracrossmultiple model classes.Second, it documents a clear horizon-dependent decay in the magnitude of these improvements, highlighting the importance of horizon-aware evaluation.Finally, the study adoptsa controlled, model-agnostic comparison framework that emphasizes robustness and operational interpretability rather than aggressive optimization.

The remainder of the paper is organized as follows.Section 2 describes the data and preprocessing pipeline.Section 3 formalizes the forecasting setup and feature construction.Section 4 outlines the forecasting models and experimental design.Section 5 presents and interprets the multi-horizon empiricalresults.Section6discussesoperationalimplicationsandlimitations.Section7concludes the paper.

II. Data Description

The empirical analysis in this study is based on one year of high-frequency data obtained from the cryptocurrency market, focusing on three major digital assets:Bitcoin (BTC), Ethereum (ETH) andBinanceCoin(BNB).Theseassetswereselectedduetotheirsustainedliquidity,consistently hightradingvolumesandcentralrolewithinthebroadercryptocurrencyecosystem.Together, theyprovidearepresentativeandreliablesettingforexaminingshort-horizondemandforecasting incontinuouslyoperatingelectronicmarkets.

The raw data were collected from the Binance exchange using the official Binance application programming interface (API), specifically through the *klines* endpoint.Binance klines provide aggregated market information at fixed time intervals and are widely used in empirical studies due to their transparency, availability and consistency.Each kline record corresponds to a predefined intervaland summarizes market activitywithin that interval usingstandard price and volume fields.

In this study, one-minute klines were used as the base temporal resolution.The one-minute fre- quency represents a practical compromise between capturing short-horizon market dynamics and maintaining computational tractability for large-scale empirical analysis.At this resolution, thedata retain meaningful intraday structure while avoiding the excessive noise and microstructural artifacts often present at sub-minute frequencies.

Cryptocurrency markets operate continuously without centralized trading hours, overnight clo- sures, or formal opening and closing periods.This continuous operation distinguishes them from traditional equity and futures markets and eliminates the need to account for market open or close effects.As a result, all minutes within the day are treated symmetrically, allowing the analysis to focus entirely on intrinsic demand dynamics rather than institutional trading schedules.

Each raw observation contains a timestamp identifying the start of the one-minute interval, an asset identifier and the standard open, high, low, close and volume (OHLCV) fields.The *open* and *close* prices represent the first and last traded prices within the minute, while the *high* and *low* prices record the maximum and minimum traded prices during that interval. The *volume* variable records the total traded quantity executed within the minute and serves as the primary proxy for transactional demand.

Transactional demand is proxied by traded volume throughout the analysis.At very short horizons, volume provides a direct and operationally meaningful measure of executed activity and reflects the intensity of participation by market participants.Interpreting volume as demand aligns the analysis with broader operational concepts such as system load, throughput and activity intensityin electronic platforms.

Rawdataforeachassetwerecollectedseparatelyandsubsequentlycombinedintoaunifieddataset using an asset identifier.Before feature construction, all observations were sorted strictly by as-set and timestamp to ensure correct temporal ordering.This step is essential in high-frequency forecasting to avoid information leakage and to preserve the causal structure of the data.

Missing or irregular intervals were handled prior to feature engineering to preserve a uniform one- minute grid.Maintaining a regular temporal structure is particularly important when constructing rolling-window statistics and lagged variables, as irregular spacing can introduce artificial disconti- nuities and distort derived indicators.

To capture systematic temporal patterns in demand, two calendar variables were constructed di- rectlyfromthetimestamp.Thevariable*minuteofday*recordsthepositionofeachobservation

within the daily cycle, measured as the number of minutes elapsed since midnight.The variable*day of week* identifies the day of the week associated with each observation.Although cryptocur- rency markets operate continuously, empirical evidence suggests that intraday and weekly regular- itiesmaystillariseduetohumanbehavior,institutionalparticipationandregionalactivitycycles.

All subsequent variables used in the empirical analysis were constructed exclusively from informa- tion available up to the forecasting origin. No future information was used at any stage of feature constructionortargetdefinition.Thisdesignensuresthattheforecastingexercisereflectsarealistic operational setting and that reported performance measures correspond to genuine out-of-sample predictions.

III. Forecasting Variables and System State Indicators

The forecasting framework employed in this study is based on transforming raw minute-level mar- ket observations into a structured set of explanatory variables suitable for short-horizon demand prediction.Each row of the final modeling dataset corresponds to a specific asset and minute and all variables are computed using information available up to that time.This section provides a de- tailed description of the forecasting targets, baseline demand variables and system state indicators used in the empirical analysis.

Theprimaryobjectiveistoforecastnear-futuretransactionaldemandovermultiplehorizons. Let $V_t$denote the observed traded volume at minute t.Three horizon-dependent forecasting targets are constructed.The variable *target volume next min* corresponds to V_{t+1}, the demand observed in the immediately subsequent minute.The variables *target volume next 5min* and *tar- get volume next 15min* are defined as the cumulative traded volume over the next five and fifteen minutes, respectively. Using cumulative volume rather than averages aligns the targets with oper- ational interpretations of total expected load over short planning windows.

Baselinedemandpredictorsaredesignedtocaptureshort-termpersistence,localtrendsandsystem- atic temporal patterns. The variable *vol lag 1* represents the observed demand in the immediately precedingminuteandservesasthemostbasicpersistence-basedpredictor.High-frequencydemand is known to exhibit strong short-term dependence, making lagged volume a natural starting point for forecasting.

To summarize recent demand trends while smoothing high-frequency noise, two moving-average variables are constructed.The variable *vol ma 5* represents the average traded volume over the previous five minutes, while *vol ma 15* captures the average demand over the previous fifteen minutes.These variables provide information about local demand intensity and help distinguish transient fluctuations from more persistent changes in activity.

Calendar variables complement historical demand measures by capturing systematic temporal ef- fects.The variable *minute of day* encodes intraday position, allowing the model to account for recurring daily patterns in demand.The variable *day of week* captures potential weekly regulari- ties.Together, these variables form the core of the baseline forecasting specification.

In addition to baseline demand features, several system state indicators are constructed from con- temporaneous price information.These variables are intended to capture the current operating condition of the market, reflecting short-term variability, dispersion and intensity that may influ- ence near-future demand.

Thevariable*return1m*measurestherelativechangeintheclosingpricebetweenconsecutive minutes.This one-minute return serves as a proxy for short-term price movement intensity and reflectsthearrivalofnewinformationorshiftsintradingbehavior.Periodsoflargeabsolutereturns often coincide with heightened activity and increased participation.

Price dispersion within a minute is captured by the variable *hl range*, defined as the relative differ- ence between the high and low prices during the interval.This measure summarizes intraminute volatility and reflects the degree of price fluctuation observed within a very short time window.

The variable *co range* measures the relative difference between the closing and opening prices within a minute.Unlike *hl range*, which captures dispersion, this variable emphasizes directional movement and provides information about the net price change over the interval.

Tosummarizerecentvolatilityconditionsbeyondasingleminute,arealizedvolatilitymeasure is computed over a rolling five-minute window.The variable *realized vol 5m* aggregates squared short-termpricechangesoverthepreceedingfiveminutesandprovidesacompactmeasureofrecent market instability.Elevated realized volatility indicates departures from routine conditions andmay signal increased uncertainty in near-term demand.

An interaction term, *demandpressure*, is constructed to capture situations in which elevated trading activity coincides with heightened volatility.This variable combines contemporaneous volume with recentvolatilityandisdesignedtoidentifystressedmarketconditionsinwhichforecastingerrors maybemoreconsequentialfromanoperationalperspective.

All system state indicators are computed using minimal and readily available information derived from price series.No order book data or proprietary indicators are required, emphasizing the generality and ease of implementation of the proposed framework.

All explanatory variables are aligned so that only information available up to time t is used to forecast demand at future horizons. This strict alignment ensures that the forecasting exercise re- mains fully out-of-sample and avoids look-ahead bias.The resulting dataset provides a transparent and reproducible mapping from raw market observations to the forecasting variables used in the empirical analysis.

By separating baseline demand features from system state indicators, the framework allows a controlled comparison between history-based forecasting and state-aware forecasting approaches. This structure makes it possible to isolate the incremental predictive contribution of system state information and to assess how its relevance varies systematically across forecasting horizons.

Forecasting Models and Experimental Design

This section describes the forecasting models employed in the empirical analysis and outlines the ex- perimental design used to evaluate their out-of-sample performance.The objective is not to propose novel forecasting algorithms, but to assess, in a controlled and transparent manner, whether incor- porating system state indicators improves short-horizon demand forecasts across different modelclasses and forecasting horizons.

Three forecasting models are considered:linear regression, random forest regression and gradi-ent boosting regression.These models represent increasing levels of functional flexibility and are widely used in applied forecasting and empirical modeling.Examining performance across these model classes allows the analysis to distinguish between improvements attributable to additional explanatory information and those driven purely by model complexity.

For each model class, two specifications are estimated.The baseline specification includes onlyhistorical demand variables and calendar effects, while the augmented specification additionally incorporates system state indicators derived from contemporaneous price information.This paired designensuresthatanyobservedimprovementinforecastingaccuracycanbeattributeddirectlyto the inclusion of system state variables rather than differences in estimation procedures or sample composition.

Atime-orderedtrain–testsplitisusedtoevaluateout-of-sampleperformance.Specifically, thefirst 80% of the observations for each asset are used for model estimation,while the remaining 20%are reserved for evaluation. This approach avoids look-ahead bias and reflects realistic forecasting conditions in which future demand must be predicted using only information available at the forecasting origin.

Forecastingtargetsareconstructedformultiplehorizons.Theone-minutetargetcorrespondstodemand observed in the immediately subsequent minute. The five-minute and fifteen-minute targets aredefinedascumulativedemandoverthenextfiveandfifteenminutes,respectively.Usingcumulative volume rather than averages aligns the targets with operational interpretations of expected system load over short planning windows.

Linear regression is employed as a benchmark forecasting model due to its transparency, inter-pretabilityandwidespreaduseinoperationalsettings.Despiteitssimplicity, linearregressionoften performs competitively in short-horizon forecasting tasks, particularly when strong persistence and seasonal effects are present.

Formally, the linear regression model specifies the forecasted demand as a linear combination of explanatoryvariablesavailableattime$t$.Inthebaselinespecification,thesevariablesincludelagged demand, moving averages and calendar indicators.The augmented specification extends this setby adding system state indicators such as short-term returns, price ranges and realized volatility measures.

Thelinearmodelprovidesausefulreferencepointforassessingtheincrementalvalueofsystemstate information.Because linear regression cannot capture nonlinear interactions or regime-dependent effects, any improvement observed in this model indicates that system state indicators contain information that is not redundant with historical demand patterns alone.

Randomforestregressionisincludedasanonlinearforecastingapproachcapableofcapturingcom- plexrelationshipsamongexplanatoryvariables.Arandomforestconsistsofanensembleofdecision trees,eachtrainedonabootstrapsampleofthedata.Ateachsplitinatree,arandomsubset of explanatory variables is considered, introducing diversity among trees and reducing correlation within the ensemble.

The final random forest prediction is obtained by averaging predictions across all trees in the ensemble.Thisaggregationreducesvarianceandenhancesrobustness,makingrandomforestspar- ticularly well suited for high-frequency forecasting tasks characterized by noisy and heterogeneous data.

In the context of short-horizon demand forecasting, random forests can capture nonlinear depen- dencies and interaction effects between historical demand variables and system state indicators. For example, the impact of short-term volatility on future demand may depend on the prevailing level of activity, a relationship that is difficult to represent using linear specifications.

Gradient boosting regression is employed as a second nonlinear forecasting approach that comple- mentstherandomforestmodel.Whilebothmethodsaretree-basedensembles,gradientboosting

differs fundamentally in how individual trees are constructed and combined. In gradient boosting, trees are added sequentially rather than independently.

Each new tree in the gradient boosting model is trained to predict the residual errors of the cur- rent ensemble, allowing the model to gradually improve predictive performance through iterative refinement.This sequential structure enables gradient boosting to capture complex nonlinear rela- tionships and subtle interaction effects among explanatory variables.

Although gradient boosting models are powerful, they are also more sensitive to over-fitting if not carefullyregularized.Tomitigatethisrisk, modelcomplexityiscontrolledthroughalimitednumber of trees, shallow tree depth and the use of shrinkage parameters.These constraints ensure that observedperformancegainsreflectgenuinepredictiveinformationratherthanexcessiveadaptation to noise.

Using both random forest and gradient boosting models allows the analysis to assess whether improvements from system state indicators persist across different nonlinear modeling strategies. Consistent gains across these approaches provide stronger evidence that the indicators capture meaningful information rather than model-specific artifacts.

Forecast accuracy is evaluated using two standard metrics:mean absolute error (MAE) and root mean squared error (RMSE). These metrics provide complementary perspectives on predictive performance and are widely used in empirical forecasting studies.

Mean absolute error is defined as the average absolute difference between observed demand and predicted demand over the evaluation sample.MAE measures the typical magnitude of forecasting errors and is robust to extreme deviations.As such, it provides an intuitive and interpretable measure of average forecast accuracy.

Root mean squared error is defined as the square root of the average squared difference between observed and predicted demand.RMSE penalizes larger errors more heavily than MAE and is therefore sensitive to occasional large deviations.In high-frequency settings, where extreme demand spikesmayoccur,RMSEprovidesvaluableinformationabouttailriskinforecastingerrors.

Evaluating both MAE and RMSE allows the analysis to distinguish between improvements in typi- calforecastingperformanceandreductionsinlarge,potentiallycostlyerrors.Consistentreductions in both metrics indicate robust improvements across the entire error distribution.

Together, the forecasting models and experimental design provide a structured and transparent framework for assessing the predictive value of system state indicators across multiple horizons.By combining linear and nonlinear models, multiple forecasting horizons and complementary ac- curacymetrics,theanalysisoffersacomprehensiveevaluationofshort-horizondemandforecasting performance in continuously operating electronic markets.

IV. Empirical Results

This section reports out-of-sample forecasting performance for baseline and augmented models across multiple forecasting horizons.Forecast accuracy is evaluated using mean absolute error (MAE)androotmeansquarederror(RMSE).ResultsaresummarizedinTable1.

For clarity, subscripts B and A are used to distinguish between baseline and augmented model specifications.The baseline specification (B) includes only historical demand variables and calendar effects,whiletheaugmentedspecification(A)additionallyincorporatessystemstateindicators derived from contemporaneous price information.Reported percentage improvements are computed relative to the baseline model.

Table1:Out-of-sampleforecastingperformanceacrosshorizons

Horizon	Model	MAE_B	$RMSE_B$	MAE_A	$RMSE_A$	ΔMAE(%)	ΔRMSE(%)
1-min	Linear	224.72	555.22	217.45	542.73	3.24	2.25
1-min	RandomForest	230.27	555.96	218.47	530.80	5.12	4.53
1-min	GradientBoosting	229.25	553.19	217.81	528.19	4.99	4.52
5-min	Linear	507.19	1152.52	483.39	1092.15	4.69	5.24
5-min	RandomForest	506.36	1137.30	451.55	1008.86	10.82	11.29
5-min	GradientBoosting	507.22	1136.53	453.48	1003.22	10.60	11.73
15-min	Linear	570.09	1225.41	565.47	1213.53	0.81	0.97
15-min	RandomForest	573.89	1240.08	566.13	1225.11	1.35	1.21
15-min	GradientBoosting	576.13	1229.18	569.90	1210.09	1.08	1.55

Quantitatively, the results show that incorporating system state indicators leads to consistent reductions in both MAE and RMSE at all horizons and across all model classes.At the one-minute horizon,improvementsrangebetweenapproximately3%and5%,indicatingthatcontemporaneous state information provides incremental predictive value beyond historical demand patterns.

At the five-minute horizon, the magnitude of improvement increases substantially for nonlinear models, with error reductions exceeding 10%. This suggests that short-term aggregation amplifies the usefulness of stateindicators whencombined withflexible models capableof capturingnonlinear interactions.

At the fifteen-minute horizon,improvements remain positive but are markedly smaller,falling below2%inmostcases.Thispatternindicatesaclearhorizon-dependentdecayinthepredictiverelevance ofcontemporaneousstateinformation,consistentwithincreasingdominanceofcumulativenoise and longer-term persistence effects.

Qualitatively, these findings suggest that system state indicators are most valuable when forecasts are required at very short horizons or modest aggregation windows. As the horizon lengthens, the influence of immediate system conditions diminishes and demand dynamics become increasingly governed by broader temporal structure.

Tofurtherillustratethehorizon-dependentbehaviorobservedinTable1,agraphicalrepresentation of forecast error improvements is provided.The figure summarizes percentage reductions in MAEacross forecasting horizons for each model class, highlighting how the contribution of system state indicatorsvariesasthehorizonlengthens.

Figure 1 provides a visual summary of the horizon-dependent effect of system state indicators.Im- provements are largest at short horizons and decline as the forecasting window increases.Nonlinear modelsexhibitsubstantiallylargergainsatthefive-minutehorizon,indicatingtheirabilitytoex- ploit interactions between historical demand and system state information.At the fifteen-minute horizon, improvements remain positive but are markedly smaller, consistent with a decay in the relevance of contemporaneous state conditions.

WhileMAEsummarizestypicalforecasterrors,RMSEplacesgreaterweightonlargedeviations andisthereforesensitivetoextremeforecastingerrors.Toexaminewhetherthehorizon-dependent patternsobservedforMAEalsoholdwhenlargererrorsarepenalizedmoreheavily,acorresponding

10

5

0

1 5 15

ForecastingHorizon(minutes)

Figure1:PercentagereductioninMAEfrombaselinetoaugmentedmodelsacrossforecasting horizons

analysisbasedonRMSEispresentedbelow.

10

5

0
1 5 15
ForecastingHorizon(minutes)
Figure 2:Percentage reduction in RMSE ~~from ba~~seline to augmented models across forecastinghorizons

Figure 2 confirms that the horizon-dependent effects observed for MAE are also present when fore-castingaccuracyisevaluatedusingRMSE.Improvementsareagainlargestatshortandintermedi-atehorizons,particularlyfornonlinearmodels,indicatingthatsystemstateindicatorshelpreduce large forecasting errors during periods of elevated activity.At the fifteen-minute horizon, RMSE improvementsremainpositivebutarecomparativelysmall,suggestingthatextremedeviations become increasingly dominated by longer-term demand dynamics rather than contemporaneoussystem conditions.

V. Discussion and Limitations

Thisstudyexaminedhorizon-dependentdemandforecastinginhigh-frequencycryptocurrencymar-ketswithaparticularfocusontheroleofsimplesystemstateindicators.Theempiricalresults

demonstratethatincorporatingcontemporaneousstateinformationyieldsconsistentimprovements inforecastingaccuracyacrossmultiplemodelclassesandforecastinghorizons.Atthesametime, themagnitudeoftheseimprovementsvariessystematicallywiththeforecasthorizon,highlighting theimportanceofhorizon-awareevaluationinshort-termdemandprediction.

From an operational perspective, the findings suggest that system state indicators are most valuable whenforecastsarerequiredatveryshorthorizonsormodestaggregationwindows.Attheone- minute horizon, improvements are modest but persistent, reflecting the inherently noisy nature ofhigh-frequency demand.At the five-minute horizon, gains become substantially larger for nonlinear models, indicating that short-term aggregation allows system state information to interact more effectively with historical demand patterns.At the fifteen-minute horizon, improvements remain positivebutaremarkedlysmaller,suggestingadiminishinginfluenceofcontemporaneousconditions aslonger-termdynamicsbegintodominate.

The observed horizon-dependent decay in predictive improvement has important implications for the design of forecasting systems in continuously operating electronic markets.Features that are highly informative at very short horizons may lose relevance as the forecasting window expands, even when the same underlying data and modeling framework are used.This finding cautions against evaluating forecasting models at a single horizon and extrapolating conclusions to other operational contexts without explicit empirical verification.

An important insight from the empirical results is the differential behavior of linear and nonlinear models.While all model classes benefit from the inclusion of system state indicators, nonlinear ensemble methods exhibit substantially larger gains, particularly at the five-minute horizon.This patternsuggeststhatthepredictivecontributionofstatevariablesisnotpurelyadditivebutarisesin partthroughinteractionswithhistoricaldemandandcalendareffects.Suchinteractionsaredifficult to capture using linear specifications but are naturally accommodated by tree-based ensemble models.

Despite these gains, it is noteworthy that even the most flexible models do not eliminate fore- casting error at short horizons.High-frequency demand remains inherently volatile, reflecting the decentralized nature of participation, rapid information arrival and heterogeneous trading motives in cryptocurrency markets.The results therefore underscore the limits of predictability in such environments and highlight the need for realistic expectations regarding achievable forecasting ac- curacy.

The simplicity of the system state indicators employed in this study is both a strength and a limitation.On the one hand, all indicators are derived from readily available price and volume information and can be computed in real time without access to proprietary order book data.This makes the proposed framework broadly applicable and easy to implement in practical settings.On the other hand, more detailed micro-structure information, such as depth, order flow imbalance, or queue dynamics, may contain additional predictive signals that are not captured by the indicators considered here.

Another limitation of the present analysis is its focus on three highly liquid cryptocurrency assets. While Bitcoin, Ethereum and Binance Coin provide a robust test environment, the results may not generalize directly to less liquid assets or markets with different participation structures.In thinner markets, demand dynamics may be more irregular and the relative importance of state indicators may differ substantially.

The empirical evaluation adopts a deliberately conservative modeling approach.Model hyperpa- rametersarenotaggressivelyoptimizedandtheanalysisemphasizesrobustnessandinterpretability over maximal predictive performance.While this choice strengthens the credibility of the reported results, it may understate the performance achievable through more elaborate tuning or adaptive modeling strategies.

Finally,thestudyispurelypredictiveinnatureanddoesnotattempttoestablishcausalrela- tionships between system state indicators and future demand. The observed improvements should thereforebeinterpretedasempiricalregularitiesratherthanstructuraleffects.Understanding the mechanisms through which contemporaneous market conditions influence near-future demand remainsanimportantdirectionforfutureresearch.

Taken together, the discussion highlights that system state indicators offer a practical and robust enhancement to short-horizon demand forecasting, but that their usefulness is inherently horizon dependent and subject to structural limitations.These findings provide a foundation for more detailed investigations into high-frequency predictability and operational forecasting in electronic markets.

VI. Conclusion

This paper investigated horizon-dependent demand forecasting in high-frequency cryptocurrency markets, with particular emphasis on the role of simple system state indicators derived from con- temporaneous price information.Using one year of minute-level data for Bitcoin, Ethereum and Binance Coin, the study compared baseline forecasting models based on historical demand and cal- endareffectswithaugmentedmodelsthatincorporateshort-termvariabilityandactivity measures.

The empirical results demonstrate that system state indicators provide consistent improvementsin forecasting accuracy across linear regression, random forest and gradient boosting models.Im- portantly, these improvements persist across multiple forecasting horizons, but their magnitude declines systematically as the horizon lengthens.This horizon-dependent decay highlights the im- portance of aligning forecasting models with the operational time scales at which decisions are made.

The findings contribute to the forecasting literature by emphasizing that predictive relationships observed at one horizon cannot be assumed to hold uniformly at others.Even within a fixed modelingframeworkanddataset,therelevanceofexplanatoryvariablesmaychangesubstantiallyas temporalaggregationincreases.Thisinsightreinforcestheneedforexplicitmulti-horizonevaluation in short-term forecasting studies.

From a practical standpoint, the results suggest that incorporating parsimonious system state indicators can enhance real-time monitoring and short-term planning in continuously operating electronic markets.Because the indicators considered in this study are easy to compute and rely only on readily available data, they offer a low-cost improvement to existing forecasting systems.

The analysis also points toward several promising directions for future research within a coherent andfocusedresearchprogram.Onenaturalextensionistomovebeyondprice-basedstateindicators andexaminetheroleoforder-flowinformation.Forexample,futureworkmayinvestigatewhether order-flowimbalanceservesasapredictorofshort-termpricemovementanddemandandhowsuch imbalance measures compare when weighted by liquidity rather than treated uniformly.

Related questions include the interaction between order-flow imbalance and market frictions such as bid–ask spread and slippage.Understanding how these factors influence the decay of intraday predictivesignalsmayprovidedeeperinsightintothelimitsofshort-horizonforecastingandthe sustainabilityofintradayalpha.

Anotherimportantavenueconcernsregimedependence.Theprofitabilityandpredictivevalue of imbalance-based signals may vary across market regimes characterized by differing volatility, liquidity,orparticipationintensity.Systematicallyidentifyingandmodelingsuchregimesmay help explain when and why short-horizon predictability emerges or disappears.

By addressing these questions within a unified empirical framework and using a consistent dataset, futurestudiescanbuildacoherentsequenceofrelatedcontributionsthatprogressivelydeepenour understanding of high-frequency market dynamics.Such a programmatic approach allows indi- vidual studies to remain focused while collectively advancing knowledge in a well-defined research area.

In summary, this paper provides empirical evidence that system state indicators enhance short- horizon demand forecasting in high-frequency cryptocurrency markets, but that their predictive contribution is inherently horizon dependent.The results clarify both the potential and the limi- tationsofstate-awareforecastingandlaythegroundworkforfutureresearchonorderflow,regime dependence and intraday predictability in electronic markets.

Conflict of Interest: Theauthordeclaresnoconflictofinterest.

References

[1] Hyndman,R.J.,&Athanasopoulos,G.(2021).*Forecasting:PrinciplesandPractice* (3rded.). OTexts, Melbourne.

[2] Armstrong,J. S. (2001). *PrinciplesofForecasting:AHandbookforResearchersandPracti- tioners*. Springer.

[3] Tashman,L.J.(2000).Out-of-sampletestsofforecastingaccuracy:Ananalysisandreview. *InternationalJournalofForecasting*,16,437–450.

[4] Makridakis,S.,Spiliotis,E.,&Assimakopoulos,V.(2018).Statisticaland machine learning forecasting methods:Concerns and ways forward. *PLOS ONE*, 13(3), e0194889.

[5] Breiman,L.(2001).Randomforests.*MachineLearning*,45,5–32.

[6] Friedman,J.H.(2001).Greedyfunctionapproximation:Agradientboostingmachine.*Annals of Statistics*, 29, 1189–1232.

[7] Engle, R. F. (2002). Dynamic conditional correlation:A simple class of multivariate GARCH models. *Journal of Business & Economic Statistics*, 20(3), 339–350.

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

[9] Taylor,S.J.(2008).*ModellingFinancialTimeSeries*.WorldScientific.

[10] O'Hara,M.(1995).*MarketMicrostructureTheory*.BlackwellPublishers.

[11] Urquhart,A.(2016).TheinefficiencyofBitcoin.*EconomicsLetters*,148,80–82.

[12] Katsiampa,P.(2017).VolatilityestimationforBitcoin:AcomparisonofGARCHmodels. *EconomicsLetters*,158,3–6.

[13] Bouri, E., Molnár,P., Azzi, G., Roubaud, D., & Hagfors, L. I. (2017). On the hedge and safe havenpropertiesofBitcoin:Isitreallymorethanadiversifier? *FinanceResearchLetters*,20, 192–198.

[14] Kristoufek,L.(2015).WhatarethemaindriversoftheBitcoinprice? Evidencefromwavelet coherence analysis. *PLOS ONE*, 10(4), e0123923.

[15] Taylor,S.J.(2008).*ModellingFinancialTimeSeries*.WorldScientific.