Intraday Volatility-Volume Joint Modeling and Forecasting: A State-space Approach

Shengjie Xiu and Daniel P. Palomar Department of Electronic and Computer Engineering The Hong Kong University of Science and Technology Clear Water Bay, Hong Kong sxiu@connect.ust.hk, palomar@ust.hk

Abstract—Intraday volatility is a crucial indicator for shortterm price movements of financial assets, playing a pivotal role in risk management and other financial applications in the era of high-frequency algorithmic trading. However, existing methods fail to capture complex intraday volatility patterns, leading to inadequate forecast capabilities. To address these shortcomings, we propose a novel multiplicative component framework that leverages the close relationship between volatility and trading volume to analyze intraday volatility. We further introduce a state-space approach to implement this framework for modeling and forecasting. Empirical experiments demonstrate that our proposed method outperforms traditional models by providing more comprehensive insights and achieving higher forecasting accuracy for intraday volatility.

Index Terms—Intraday volatility modeling and forecasting, volatility-volume relation, state-space model

I. INTRODUCTION

In the last half-century, stock market volatility has been extensively discussed. Volatility, usually measured by the standard deviation of stock returns, is a crucial factor as it characterizes the level of risk associated with an asset. With the emergence of algorithmic trading, financial assets are now subject to substantial price fluctuations within minutes. Therefore, intraday volatility has become increasingly important to academia and the financial industry. Accurate intraday volatility modeling and forecasting assist traders in assessing their risk exposure and identifying profitable trading opportunities [1]. They are also fundamental in pricing options and other derivatives [2], [3].

In the literature, daily volatility modeling has been well researched, with three classes of models receiving significant attention: historical volatility [4], ARCH-GARCH class conditional volatility (e.g. [5], [6]), and stochastic volatility (e.g. [7]). However, intraday volatility modeling and forecasting pose greater challenges. Although ARCH class models can be applied directly to intraday volatility [8], [9], they have been criticized for their inability to capture complex structures, such as intraday regularities. To address this limitation, [10] proposes filtering out the intraday seasonal pattern before applying GARCH models. Alternatively, [11] suggests representing intraday volatility as a product of daily, intraday seasonal, and intraday dynamics components and developing

a multiplicative component GARCH (mcsGARCH). However, mcsGARCH is impractical in practice due to the potential error accumulation in multistep estimation. Therefore, intraday volatility modeling and forecasting remain under-explored research areas.

In recent years, scholars have been exploring the role of trading volume in advancing the understanding of intraday volatility. The positive relationship between volatility and volume is widely acknowledged, with two established theoretical explanations. The sequential information arrival hypothesis (SIAH) assumes that traders react to new information sequentially, thus, lagged values of volatility and volume can predict current ones [12], [13]. The other explanation for the positive correlation is the mixture of distributions hypothesis (MDH), which posits that stock price and volume are jointly dependent on common information flow [14], [15]. While incorporating trading volume information may be beneficial, how to effectively model and exploit the intraday volatility-volume relationship remains an open question.

Our study contributes to the existing literature in the following ways:

- We propose a new multiplicative component framework for intraday volatility that accounts for the volatilityvolume relationship.
- We develop a state-space approach for intraday volatility modeling and forecasting.
- We conduct extensive empirical experiments that demonstrate our proposed model's superior financial interpretation and forecasting performance.

II. PRELIMINARIES AND MOTIVATION

A. Intraday Volatility

In this study, we use 15-minute intervals to model intraday volatility, where each trading day is indexed by $t \in \{1, 2, ..., N\}$ and each bin within a day is indexed by $n \in \{1, ..., N\}$. We denote the asset price at the end of bin n of day t by $P_{t,n}$ and compute the return in every 15 minutes as

$$r_{t,n} = \begin{cases} \log\left(P_{t,n}/P_{t,n-1}\right) & n \ge 1\\ \log\left(P_{t,1}/P_{t-1,N}\right) & n = 0. \end{cases}$$
(1)

This work was supported by the Hong Kong GRF 16207820 research grant.

The R package intradayModel will support the proposed algorithm.

To measure intraday volatility, we use the sum of higherfrequency squared returns (realized volatility), expressed as

$$\sigma_{t,n} = \sqrt{\sum_{j=1}^{1/\Delta} \hat{r}_{t,n-1+j\Delta}^2},\tag{2}$$

where $\Delta \in (0, 1]$ is a possibly shorter horizon standardized by 15 minutes, for example, $\Delta = 1/3$ stands for 5-minute intervals. $\hat{r}_{t,n-1+j\Delta}$ is the higher-frequency return satisfying (1) and $r_{t,n} = \sum_{j=1}^{1/\Delta} \hat{r}_{t,n-1+j\Delta}$. As shown in the top panel of Figure 1, the time series of intraday volatility exhibits volatility clustering and persistence across different time scales, which are crucial for accurate forecasting.



Fig. 1. Intraday volatility and volume series of AMZN stock in log scale.

B. Multiplicative Component GARCH

Engle and Sokalska propose the mcsGARCH model for intraday financial returns [11], which has the expression

$$r_{t,n} = \hat{\sigma}_{t,n} \varepsilon_{t,n}, \quad \text{and} \quad \varepsilon_{t,n} \sim N(0,1),$$
 (3)

where the conditional variance is decomposed into three components:

$$\hat{\sigma}_{t,n}^2 = \text{daily}_t \times \text{seasonal}_n \times \text{intraday dynamics}_{t,n}.$$
 (4)

The daily component is the systematic volatility variation across different trading days and adjusts the average level of intraday volatility. It can be obtained from the volatility forecasts such as the multifactor risk model or daily GARCH. The seasonal component is the average volatility within the same bin across different days. It exhibits a U-shaped pattern due to the higher trading activity at the opening and closing hours. The intraday dynamics component captures the intraday variation of volatility, such as due to unexpected news or sudden changes in market sentiment.

While the mcsGARCH has reasonable financial interpretations, it also has several practical limitations that deteriorate its forecasting ability: 1) Errors in the one-day-ahead prediction of the daily component can propagate and cause errors in the intraday GARCH model [16]. 2) Intraday GARCH cannot capture the longer-lasting effects of market shocks [4]. To address these issues, incorporating new sources of information, such as trading volume, is necessary.

C. Mixture of Distribution Hypothesis

The positive relationship between daily volatility and trading volume has been extensively investigated in New York Stock Exchange common stocks [17] and nine national markets [18]. Similarly, the positive correlation can be observed in intraday data [19], [20], which can be identified in Figure 1. The widely researched MDH hypothesis builds a bivariate normal mixture model of daily price change ΔP and trading volume V, expressed as

$$\Delta P = \sigma_1 \sqrt{IZ_1}$$

$$V = \mu_2 I + \sigma_2 \sqrt{IZ_2},$$
(5)

where Z_1 and Z_2 are N(0,1) random variables. This model directly and positively links volatility and volume with the information flow I.

The MDH hypothesis, while appealing from an information flow standpoint, oversimplifies the intraday relationship as one latent variable I and suffers two significant limitations: 1) It fails to fully capture the volatility and volume intraday dynamics, including unexpected movements [20]. 2) It cannot account for the long-term variations in daily volatility and volume [21]. Information flow is associated only with medium-term volatility persistence, highlighting the need for a more sophisticated modeling approach to incorporate volume information in intraday volatility modeling.

III. PROPOSED STATE-SPACE APPROACH

A. New Multiplicative Component Framework

We propose a novel framework for intraday volatility modeling that decomposes it into multiple components: seasonal, daily (long-term), information flow (medium-term), unexpected volatility (short-term), and noise components. Information flow and unexpected volatility correspond to "good" and "bad" intraday volatilities, respectively. "Good" volatility is directional, persistent, relatively easy to anticipate, and accompanied by sufficiently high volume. "Bad" volatility is erratic, relatively difficult to predict, and less associated with volume [20]. Intraday volume is decomposed similarly but without the unexpected volatility component.

Therefore, our framework is expressed as

$$\sigma_{t,n} = d_t \times s_n \times I_{t,n} \times u_{t,n} \times \varepsilon_{t,n}$$

$$V_{t,n} = d_t^* \times s_n^* \times I_{t,n} \times \varepsilon_{t,n}^*,$$
(6)

where

- d_t (d_t^*) is the daily volatility (volume) component;
- s_n (s_n^*) is the intraday seasonal volatility (volume) component;
- $I_{t,n}$ is the shared information flow;
- $u_{t,n}$ is the unexpected volatility component;
- $\varepsilon_{t,n}$ ($\varepsilon_{t,n}^*$) is the log-normal noise term in volatility (volume).

Specifically, the intraday seasonal components are estimated separately based on historical averages.

Our proposed model for intraday volatility modeling offers several improvements over the traditional mcsGARCH model. First, incorporating information flow can better extract medium-term volatility persistence with the knowledge of trading volume. Second, we gain greater flexibility in capturing intraday dynamics by modeling it as a stochastic process instead of a deterministic function. Third, our state-space approach, introduced in Section III-B, enables us to extract all components simultaneously.

Compared to the MDH hypothesis, our framework provides a more precise identification of the information flow by filtering out components that cannot be well captured, such as daily patterns with long-term differences and unexpected volatility. These improvements enable us to identify volatility persistence with different frequencies, leading to richer interpretations of market microstructure.

B. State-space Approach

To facilitate component extraction in the model (6), we propose a state-space formulation that employs the Kalman filter and smoother. Notably, a key issue with intraday data is its heavy-tailed distribution. As demonstrated by the quantilequantile plot (Q-Q plot) in Figure 2a, intraday volatility deviates significantly from the straight line, indicating a non-Gaussian distribution. This finding is consistent with previous studies of intraday volume [22]. To address this issue, we apply logarithmic transformations to obtain a more Gaussianlike distribution, as shown in Figure 2b [23], [24].



Fig. 2. Q-Q plot of intraday volatility and log-volatility of AAPL stock.

Therefore, we reformulate the model (6) as

$$\ln \sigma_{t,n} = \ln d_t + \ln s_i + \ln I_{t,i} + \ln u_{t,i} + \ln \varepsilon_{t,i} \ln V_{t,n} = \ln d_t^* + \ln s_i^* + \ln I_{t,i} + \ln \varepsilon_{t,i}^*.$$
(7)

We make several assumptions to specify the evolution of each latent variable in the model (7). Firstly, the daily components change only during the close of markets and remain constant during trading hours, while the seasonal components are assumed to be unchanged throughout. Additionally, the information flow and unexpected volatility components are modeled as autoregressive processes with order 1 (AR(1)), as both the informed trades and uninformed trades are highly history dependent (e.g., [25]).

To accommodate these assumptions, we propose a statespace model inspired by prior work [22]. For ease of notation, we index the subscript (t, n) as $\tau = N \times (t - 1) + n, \tau =$ $1, 2, \ldots, \mathcal{T}$. Subsequently, we can formulate the state-space model as follows,

$$\mathbf{y}_{\tau} = \mathbf{W}\mathbf{x}_{\tau} + \mathbf{s}_{\tau} + \boldsymbol{\varepsilon}_{\tau}, \\ \mathbf{x}_{\tau+1} = \mathbf{F}_{\tau}\mathbf{x}_{\tau} + \boldsymbol{\eta}_{\tau},$$
(8)

where

- $\mathbf{y}_{\tau} = [\ln \sigma_{\tau}, \ln V_{\tau}]^{\mathsf{T}}$ is the observation vector of intraday log-volatility and log-volume;
- $\mathbf{x}_{\tau} = [\ln d_{\tau}, \ln d_{\tau}^*, \ln I_{\tau}, \ln u_{\tau}]^{\mathsf{T}}$ is the state vector containing daily, information flow, and unexpected volatility components;
- W is the observation matrix:

$$\mathbf{W} = \left[\begin{array}{rrrr} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{array} \right];$$

- $\mathbf{s}_{\tau} = [\ln s_{\tau}, \ln s_{\tau}^*]^{\mathsf{T}}$ is the vector of intraday seasonal components corresponding to specific bins;
- \mathbf{F}_{τ} is the transition matrix

$$\mathbf{F}_{\tau} = \begin{cases} \operatorname{diag}\left(a_{d}, a_{d}^{*}, a_{I}, a_{u}\right) & \tau = kN\\ \operatorname{diag}\left(1, 1, a_{I}, a_{u}\right) & \text{otherwise}; \end{cases}$$

• $\boldsymbol{\varepsilon}_{\tau} \sim N(0, \mathbf{R})$ is the i.i.d. Gaussian noise with

$$\mathbf{R} = \operatorname{diag}\left(r_{\sigma}^2, r_v^2\right);$$

• $\eta_{\tau} \sim N(0, \mathbf{Q}_{\tau})$ is the i.i.d. Gaussian noise with

$$\mathbf{Q}_{\tau} = \begin{cases} \operatorname{diag}\left(q_d^2, (q_d^*)^2, q_I^2, q_u^2\right) & \tau = kN\\ \operatorname{diag}\left(0, 0, q_I^2, q_u^2\right) & \text{otherwise} \end{cases}$$

The parameters in model (8) include

$$\boldsymbol{\theta} = (a_d, a_d^*, a_I, a_u, r_\sigma, r_v, q_d, q_d^*, q_I, q_u),$$

and they are estimated using the expectation-maximization (EM) algorithm [26].

1) Modeling Procedure: Once the parameters are estimated and fixed, we can extract the hidden components from the observed intraday volatility signal. The Kalman smoother is utilized to estimate the optimal decomposition, the latent variable $\mathbf{x}_{\tau|\tau}$, which is the optimal estimate of the state conditioned on all data from $\tau = 1$ to $\tau = \tau$.

2) Forecasting Procedure: In addition to component extraction, the proposed state-space model also allows for intraday volatility forecasting. This study focuses on one-bin-ahead forecasting, where the volatility at a specific bin is predicted based on all the information up to the last bin. Specifically, we use the Kalman filter to generate the one-bin-ahead state prediction, denoted as $\mathbf{x}_{\tau+1|\tau}$, and then obtain the volatility prediction by adding up the components:

$$\ln \sigma_{\tau+1|\tau} = \ln d_{\tau+1|\tau} + \ln s_{\tau+1} + \ln I_{\tau+1|\tau} + \ln u_{\tau+1|\tau}.$$
 (9)
IV. Experiment

A. Experiment Set-Up

This empirical study examines the intraday price and volume data of S&P 500 stocks from January 2019 to June 2021. The stocks are divided into five batches based on alphabetical order. Each trading day is divided into N 15-minute bins, and the actual volatility is calculated using 5-minute price data, as shown in Equation (2). To evaluate the modeling and forecasting performance, we split each stock's data into a training set to estimate the model parameters and an out-ofsample testing set. Table II provides further details about the dataset.

B. Modeling Results

The volatility components of the sample data are presented in Figure 3. The daily component captures the average level of daily volatility and demonstrates long-term persistence, contributing to the volatility clustering across trading days. The information flow captures most intraday patterns, while the unexpected volatility component is less informative with a small magnitude.



Fig. 3. Decomposition of intraday volatility of ABT stock in log scale.

Table I summarizes the estimated parameters from S&P 500, indicating a long-term daily volatility persistence with a large parameter a_d . The information flow has a smaller parameter a_I , supporting the presence of autocorrelation in information flow and medium-term volatility persistence. The unexpected volatility component has the smallest parameter a_u . However, it indicates the existence of some intraday patterns that cannot be explained by the joint volatility-volume movement, aligning with the critics in [20].

 TABLE I

 Estimated parameters of volatility components in S&P 500.

Parameter	Mean	Standard error	t-ratio
a_d	0.833	0.00503	165.462
a_I	0.686	0.00375	182.791
a_u	0.449	0.01708	26.299

To further confirm the features of each component, we perform spectral analysis through Fourier Transform. As illustrated in Figure 4, the daily component dominates the low-frequency spectrum, while the information flow has a higher frequency, responsible for long-term and medium-term volatility persistence, respectively. Conversely, the unexpected volatility component has a flatter spectrum, indicating its higher randomness and tendency to capture unexpected volatility changes.



Fig. 4. Frequency spectrum of volatility components of ABT stock after smoothing.

C. Forecasting Results

1) Competing Methods: This study presents three groups of candidate methods for intraday volatility modeling discussed in the literature [4], [27]. The first group includes traditional time-series models such as random walk, moving average (MA), and exponential weighted moving average (EWMA). The second group removes seasonal effects first and implements classical volatility models, including ARCH and GARCH models [10]. Finally, the third group is the specific-designed models for intraday volatility, i.e., the multiplicative component GARCH (mcsGARCH) [11].

2) Evaluation Measures: We evaluate the one-bin-ahead forecasting performance in the out-of-sample testing set, which consists of M bins. Mean absolute percent error (MAPE) and Theil-U statistic are used as indicators of the forecast accuracy [4]. They are defined as

$$MAPE = \frac{1}{M} \sum_{i=1}^{M} \frac{|\hat{\sigma}_i - \sigma_i|}{\sigma_i}, \text{ Theil-U} = \frac{\sum_{i=1}^{M} (\hat{\sigma}_i - \sigma_i)^2}{\sum_{i=1}^{M} (\sigma_{i-1} - \sigma_i)^2}.$$
(10)

Specifically, the Theil-U statistic is the prediction error standardized by random walk error.

The results are summarized in Table II. Our proposed method demonstrated the highest level of accuracy when compared to the other competing methods, as indicated by both measures. While the ARCH and GARCH models have lower accuracy than traditional time-series models concerning MAPE, they outperform the latter concerning Theil-U. This can be attributed to the ARCH-GARCH class models' sensitivity to unexpected movements, compared to traditional methods based on historical smoothing that typically have a lag in response. The mcsGARCH performs worst likely due to its multistep estimation strategy. Overall, our proposed method demonstrates its efficacy for accurate intraday volatility forecasting.

Stocks	A - CMCSA		CME - GILD		GIS - META		MGM - ROP		ROST - ZTS	
Training Period	Jan - May, 2019		Jul - Nov, 2019		Jan - May, 2020		Jul - Nov, 2020		Jan - May, 2021	
Forecast Period	Jun, 2019		Dec, 2019		Jun, 2020		Dec, 2020		Jun, 2021	
Measure	MAPE	Theil-U								
Proposed	0.547	0.476	0.591	0.482	0.512	0.535	0.550	0.470	0.568	0.474
Random walk	0.752	1.000	0.839	1.000	0.703	1.000	0.745	1.000	0.768	1.000
MA(5)	0.722	0.796	0.860	0.802	0.644	0.760	0.704	0.794	0.724	0.800
EWMA(5)	0.729	0.734	0.884	0.740	0.654	0.712	0.713	0.738	0.734	0.740
ARCH(5)	0.821	0.578	1.067	0.702	0.786	0.844	0.915	0.695	1.162	0.874
GARCH(3,2)	0.788	0.543	0.968	0.616	0.731	0.670	0.833	0.605	0.950	0.649
mcsGARCH	0.926	2.002	0.922	1.871	0.889	2.200	0.910	2.035	0.922	1.939

TABLE IIForecasting performance in S&P 500.

V. CONCLUSION

In this paper, we introduced a novel multiplicative framework for intraday volatility modeling that decomposes volatility into long-, medium-, and short-term components. Our proposed state-space approach enables the extraction of hidden components and has demonstrated improved forecast accuracy compared to existing methods, as shown through empirical experiments on S&P 500 stocks. Furthermore, this work offers valuable tools and insights for future research in microeconomics, such as identifying which latent demands are translated into intraday prices and volumes, as well as in macroeconomics, particularly in studying the relationship between systematic risk and global information flow.

REFERENCES

- K. Northington, Volatility-based technical analysis: strategies for trading the invisible. John Wiley & Sons, 2009.
- [2] J. A. Stephan and R. E. Whaley, "Intraday price change and trading volume relations in the stock and stock option markets," *The Journal of Finance*, vol. 45, no. 1, pp. 191–220, 1990.
- [3] B. Feunou and C. Okou, "Good volatility, bad volatility, and option pricing," *Journal of financial and quantitative analysis*, vol. 54, no. 2, pp. 695–727, 2019.
- [4] S.-H. Poon and C. W. J. Granger, "Forecasting volatility in financial markets: A review," *Journal of economic literature*, vol. 41, no. 2, pp. 478–539, 2003.
- [5] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica: Journal of the econometric society*, pp. 987–1007, 1982.
- [6] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, vol. 31, no. 3, pp. 307–327, 1986.
- [7] N. Shephard and T. G. Andersen, "Stochastic volatility: origins and overview," in *Handbook of financial time series*. Springer, 2009, pp. 233–254.
- [8] S. J. Taylor and X. Xu, "The incremental volatility information in one million foreign exchange quotations," *Journal of Empirical Finance*, vol. 4, no. 4, pp. 317–340, 1997.
- [9] T. G. Andersen and T. Bollerslev, "Intraday periodicity and volatility persistence in financial markets," *Journal of empirical finance*, vol. 4, no. 2-3, pp. 115–158, 1997.
- [10] M. Martens, Y.-C. Chang, and S. J. Taylor, "A comparison of seasonal adjustment methods when forecasting intraday volatility," *Journal of Financial Research*, vol. 25, no. 2, pp. 283–299, 2002.
- [11] R. F. Engle and M. E. Sokalska, "Forecasting intraday volatility in the US equity market. Multiplicative component GARCH," *Journal of Financial Econometrics*, vol. 10, no. 1, pp. 54–83, 2012.

- [12] T. E. Copeland, "A model of asset trading under the assumption of sequential information arrival," *The Journal of Finance*, vol. 31, no. 4, pp. 1149–1168, 1976.
- [13] R. H. Jennings, L. T. Starks, and J. C. Fellingham, "An equilibrium model of asset trading with sequential information arrival," *The Journal* of Finance, vol. 36, no. 1, pp. 143–161, 1981.
- [14] P. K. Clark, "A subordinated stochastic process model with finite variance for speculative prices," *Econometrica: journal of the Econometric Society*, pp. 135–155, 1973.
- [15] G. E. Tauchen and M. Pitts, "The price variability-volume relationship on speculative markets," *Econometrica: Journal of the Econometric Society*, pp. 485–505, 1983.
- [16] H. Zhang and A. Dufour, "Modeling intraday volatility of European bond markets: A data filtering application," *International Review of Financial Analysis*, vol. 63, pp. 131–146, 2019.
- [17] T. W. Epps and M. L. Epps, "The stochastic dependence of security price changes and transaction volumes: Implications for the mixtureof-distributions hypothesis," *Econometrica: Journal of the Econometric Society*, pp. 305–321, 1976.
- [18] G.-m. Chen, M. Firth, and O. M. Rui, "The dynamic relation between stock returns, trading volume, and volatility," *Financial Review*, vol. 36, no. 3, pp. 153–174, 2001.
- [19] O. A. Gwilym, D. McMillan, and A. Speight, "The intraday relationship between volume and volatility in LIFFE futures markets," *Applied Financial Economics*, vol. 9, no. 6, pp. 593–604, 1999.
- [20] P. Giot, S. Laurent, and M. Petitjean, "Trading activity, realized volatility and jumps," *Journal of Empirical Finance*, vol. 17, no. 1, pp. 168–175, 2010.
- [21] S. Darolles, G. Le Fol, and G. Mero, "Mixture of distribution hypothesis: Analyzing daily liquidity frictions and information flows," *Journal of Econometrics*, vol. 201, no. 2, pp. 367–383, 2017.
- [22] R. Chen, Y. Feng, and D. Palomar, "Forecasting intraday trading volume: a Kalman filter approach," *Available at SSRN 3101695*, 2016.
- [23] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys, "Modeling and forecasting realized volatility," *Econometrica*, vol. 71, no. 2, pp. 579–625, 2003.
- [24] S. Alizadeh, M. W. Brandt, and F. X. Diebold, "Range-based estimation of stochastic volatility models," *The Journal of Finance*, vol. 57, no. 3, pp. 1047–1091, 2002.
- [25] D. Easley, N. M. Kiefer, and M. O'Hara, "One day in the life of a very common stock," *The Review of Financial Studies*, vol. 10, no. 3, pp. 805–835, 1997.
- [26] J. Durbin and S. J. Koopman, *Time series analysis by state space methods*. OUP Oxford, 2012, vol. 38.
- [27] J. Yu, "Forecasting volatility in the New Zealand stock market," Applied Financial Economics, vol. 12, no. 3, pp. 193–202, 2002.