

Forecasting Intraday Trading Volume: A Kalman Filter Approach[☆]

Ran Chen^{a,c}, Yiyong Feng^{b,c}, Daniel P. Palomar^{c,*}

^a*Bank of America Merrill Lynch, New York, NY, USA*

^b*Three Stones Capital Limited, Hong Kong*

^c*Hong Kong University of Science and Technology, Hong Kong*

Abstract

An accurate forecast of intraday volume is a key aspect of algorithmic trading. This manuscript proposes a state-space model to forecast intraday trading volume via the Kalman filter and derives closed-form expectation-maximization (EM) solutions for model calibration. The model is extended to handle outliers in real-time market data by applying a sparse regularization technique. Empirical studies using thirty securities on eight exchanges show that the proposed model substantially outperforms the rolling means (RM) and the state-of-the-art Component Multiplicative Error Model (CMEM) by 64% and 29%, respectively, in volume prediction and by 15% and 9%, respectively, in Volume Weighted Average Price (VWAP) trading.

Keywords: algorithmic trading, EM, intraday trading volume, Kalman filter, Lasso, VWAP

1. Introduction

Trading volume is a key ingredient in many financial and economic theories as well as practical trading implementations. It indicates the liquidity, affects the prices and slippage, and reveals the overall market activities. Over the past decade, constantly evolving technology has had substantial impacts on the financial industry. The widespread development of algorithmic trading, along with the improved accessibility of ultra-high-frequency financial data, has opened up great potential for applications of trading volume. In fact, many trading algorithms require intraday volume forecasts as an important input, and further performance improvement of those strategies strongly depends on the prediction accuracy. As a result, there is growing interest in well-established models for intraday trading volume.

Algorithmic trading refers to the automated execution of trading decisions in the electronic financial markets, with sophisticated order execution strategies to minimize market impact and transaction costs. Market impact refers to an undesired price change caused by executing a trade. Informally, initiating a buy order drives the price up, and initiating a sell order drives it down. Many publications such as Bertsimas and Lo (1998), Almgren et al. (2005), Huberman and Stanzl (2005), Fabozzi et al. (2010) and

Feng et al. (2015) proposed optimal order execution strategies to reduce market impact. The Volume Weighted Average Price (VWAP) is a commonly used benchmark to measure the order execution quality. It is calculated at the end of a trading day as the sum of every transaction price weighted by the corresponding volume percentage. To achieve the VWAP, large orders are split into smaller ones based on the expected intraday volume pattern and are executed sequentially throughout the day. Thus, successful implementation of VWAP trading strategies requires a reliable model for the intraday evolution of volume.

In addition to algorithmic trading, the importance of intraday volume in other contexts has also been widely explored by the existing literature. For example, a large number of publications investigate the relationship between intraday volume and other fundamental financial variables, namely, price (e.g., Admati and Pfleiderer (1988), Lee and Rui (2002), Smirlock and Starks (1988) and Stephan and Whaley (1990)), return volatility (e.g., Chevallier and Sévi (2012), Darrat et al. (2003) and Gwilym et al. (1999)), bid-ask spreads (e.g., Brock and Kleidon (1992), Cai et al. (2004) and Hussain (2011)), and liquidity (e.g., Easley et al. (2012) and Pagano (1989)). Another group of researchers examine how the market activities or new information affect trading volume, and use it to explain a particular intraday shape in trading volume (e.g., Gerety and Mulherin (1992), Kluger and McBride (2011), Lee et al. (1994) and Malinova and Park (2014)). An effective intraday volume model can be potentially used as a building block in these studies, facilitating the analysis of crucial financial variables and market activities.

Despite the fact that intraday volume is an essential variable in the modern financial market, its modeling is

[☆]The main part of this work was conducted at Hong Kong University of Science and Technology. The views expressed in this paper are those of the author and do not reflect the position of Bank of America Merrill Lynch or Three Stones Capital Limited.

*Corresponding author. Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST), Clear Water Bay, Kowloon, Hong Kong.

Email address: palomar@ust.hk (Daniel P. Palomar)

not a trivial task due to the statistically complex features in high-frequency trading volume. There are not many well-established forecasting models in the existing open literature. The classical approach to predict intraday volume is the rolling means (RM), where the volume prediction for a particular time interval is obtained by the average volume traded in the same interval over the past days. This method is easy to implement, but it fails to adequately capture the intraday regularities existing in the volume time series. Brownlees et al. (2011) introduced perhaps the first publicly available intraday volume prediction model. They decompose volume into three components, namely, a daily average component, an intraday periodic component, and an intraday dynamic component, and use the Component Multiplicative Error Model (CMEM) to forecast the three terms. Empirical results using Exchange-Traded Funds (ETFs) SPY, DIA, and QQQ show that model by Brownlees et al. (2011) outperforms the RM approach. Although this work provides insights into the nature of modeling intraday volume, it has certain limitations. For example, the model is highly non-linear with positiveness constraint imposed on the underlying components and the noise terms, which significantly complicates its practical implementation; the parameter estimation for the CMEM is vulnerable under noisy conditions, and it is strongly dependent on the choice of the initial parameters. Satish et al. (2014) adopt a similar volume decomposition idea to Brownlees et al. (2011) and predict intraday volume by autoregressive moving average (ARMA) models. This work outperforms the RM by 29% in volume prediction and by 9% in VWAP replication, but it does not strictly compare the model performance with Brownlees et al. (2011).

This manuscript is in line with the volume decomposition idea of Brownlees et al. (2011), and it has the following contributions to the existing literature. First, we model the logarithm of intraday volume to reduce right skewness in the distribution, remove the positiveness constraint, and simplify the multiplicative model into an additive one. Second, we construct a simple but effective forecasting model based on the Kalman filter to characterize the relationships between volume and its underlying components in a state-space form. Third, we adopt the expectation-maximization (EM) algorithm for the parameter estimation and derive a set of closed-form solutions, which turns out to be numerically robust and computationally efficient. Fourth, we further extend the model by applying the Lasso regularization and present a robust version with consistent performance despite outliers in real-time non-censored data. Finally, empirical studies show that the proposed method significantly outperforms the RM and the state-of-the-art CMEM with an improvement of 64% and 29%, respectively, in volume prediction and of 15% and 9%, respectively, in the VWAP trading replication. These results are obtained from volume data over a wider range of financial products and a broader geographical distribution than those used by prior works, including twelve ETFs and eighteen stocks over the U.S., European

and Asian markets.

This manuscript is organized as follows. Section 2 details the proposed methodology for intraday volume prediction, including the state-space model, the Kalman prediction mechanism, and the model calibration. Section 3 extends the model to an outlier robust version by applying the Lasso regularization. The model effectiveness is also demonstrated with numerical simulations. Section 4 summarizes our empirical studies and compares the proposed models with the benchmarks based on the out-of-sample volume prediction and the VWAP trading replication. Section 5 concludes the manuscript.

2. Proposed Methodology for Intraday Volume Prediction

To track the intraday evolution of trading volume, we denote the day with index $t \in \{1, 2, \dots\}$ and divide each day into I 15-minute intervals (referred to as bins) indexed by $i \in \{1, \dots, I\}$. The intraday volume observations are labeled with the double subscript t, i , or equivalently, the single subscript $\tau = I \times (t - 1) + i$ to represent the time index. In this manuscript, we assume the observed volumes are non-zero for all bins, and we define volume as the number of shares traded normalized by daily outstanding shares:

$$\text{volume}_{t,i} = \frac{\text{shares traded}_{t,i}}{\text{daily outstanding shares}_t}. \quad (1)$$

This ratio is a widely used volume measure in previous studies such as Andersen (1996), Brownlees et al. (2011), Campbell et al. (1992) and Lo and Wang (2000). Since the number of shares outstanding and the number of shares traded both change over time (e.g., due to stock splits), normalization helps to correct this low-frequency variation. Log-volume refers to the natural logarithm of volume. This manuscript models log-volume, but evaluates the predictive performance based on volume.

2.1. State-Space Model

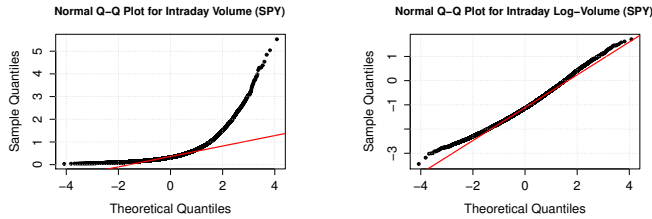
As mentioned in the introduction, Brownlees et al. (2011) decomposes intraday volume into three independent components: a daily average component, an intraday periodic component, and an intraday dynamic component. Specifically, the daily average component illustrates the inter-day trend and adjusts the mean level of the intraday series; the intraday periodic component characterizes the time-of-day seasonal pattern, which typically exhibits a U-shape because the trading activities are higher at the opening and closing time of the market; and the intraday dynamic component refers to the regularities remaining after removing the daily trend and the intraday seasonality. However, the true values of these underlying components are not directly observable, and the observed volume is a combination of these components with a multiplicative

innovation term representing the observation noise. Mathematically,

$$\text{volume}_{t,i} = \text{daily}_t \times \text{intraday periodic}_i \times \text{intraday dynamic}_{t,i} \times \text{noise}_{t,i}. \quad (2)$$

This multiplicative model is statistically complicated in nature. The positiveness constraint is imposed on the three underlying components and the noise terms, thereby making this already non-linear model even more difficult to directly fit to the data.

To make the model more manageable, we use a logarithmic transformation of Equation (2), converting this non-linear model into a simple linear one without the positiveness constraint. Furthermore, Ajinkya and Jain (1989) suggest that trading volume has significantly right-skewed distribution with a thin left tail and a fat right tail. The logarithmic transformation also helps to reduce skewness. Figure 1 (a) and (b) present the normal quantile-quantile (Q-Q) plots of intraday volume and intraday log-volume using the SPY data over a sample period between January 2013 and June 2016. It clearly shows that the intraday log-volume data is much more Gaussian distributed than the intraday volume, making Gaussian noise a reasonable assumption when modeling log-volume.



(a) The Q-Q plot of the intraday volume: the data are highly right-skewed distributed with a thin left tail and a fat right tail.

(b) The Q-Q plot of the intraday log-volume: both the tails are reasonably well specified relative to the Gaussian distribution.

Figure 1: The quantile-quantile (Q-Q) plots of intraday volume and intraday log-volume, using the SPY intraday data from January 2013 to June 2016.

Let $y_{t,i}$ denote the log-volume, η_t denote the logarithm of the daily average component, ϕ_i denote the logarithm of the intraday periodic component, $\mu_{t,i}$ denote the logarithm of the intraday dynamic component, and $v_{t,i}$ denote the market noise in the logarithmic scale. We rewrite Equation (2) as

$$y_{t,i} = \eta_t + \phi_i + \mu_{t,i} + v_{t,i}. \quad (3)$$

With this linear and additive model, the decomposition idea of intraday log-volume can be perfectly presented in a state-space form. Figure 2 shows a graphical representation of a state-space model: each vertical slice represents a time instance; the top node in each slice is the hidden state variable corresponding to the underlying volume components; and the bottom node in each slice is the observed

volume in the market. The Kalman filter, introduced by Kalman et al. (1960), is one efficient method to characterize the state-space model. It operates recursively on streams of observations to produce statistically optimal estimates of the hidden states. In addition, empirical evidence in the log-volume data suggests that the Gaussian assumption is reasonable for the application of Kalman filtering.

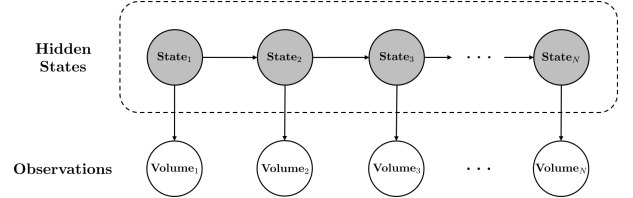


Figure 2: A graphical representation of the state-space model: each vertical slice represents a time instance; the top node in each slice is the hidden state variable corresponding to the underlying volume components; and the bottom node in each slice is the observed volume in the market.

In the model specified in Equation (3), some components do not depend on the index i whereas others do not depend on the index t , which complicates the application of the Kalman equations. To simplify notation with a common single subscript $\tau = I \times (t - 1) + i$, we repeat η_t I times for bins $i = 1 \dots I$ of day t , and name the new series η_τ ; similarly, we rewrite ϕ_i as ϕ_τ by repeating the whole sequence $\{\phi_1 \dots \phi_I\}$ for every trading day $t = 1, 2, \dots$ so that all three components have the same length as the intraday log-volume series. The resulting space-state model is

$$\mathbf{x}_{\tau+1} = \mathbf{A}_\tau \mathbf{x}_\tau + \mathbf{w}_\tau, \quad (4)$$

$$y_\tau = \mathbf{C} \mathbf{x}_\tau + \phi_\tau + v_\tau, \quad (5)$$

for $\tau = 1, 2, \dots$, where:

- $\mathbf{x}_\tau = [\eta_\tau \ \mu_\tau]^\top$ is the hidden state vector containing the daily average part and the intraday dynamic part. It determines the deseasonalized volume but cannot be directly observed;
- $\mathbf{A}_\tau = \begin{bmatrix} a_\tau^\eta & 0 \\ 0 & a^\mu \end{bmatrix}$ is the state transition matrix with $a_\tau^\eta = \begin{cases} a^\eta & \tau = kI, k = 1, 2, \dots; \\ 1 & \text{otherwise} \end{cases}$;
- $\mathbf{C} = [1 \ 1]$ is the observation matrix;
- $\mathbf{w}_\tau = [\varepsilon_\tau^\eta \ \varepsilon_\tau^\mu]^\top \sim \mathcal{N}(0, \mathbf{Q}_\tau)$ represents the i.i.d. Gaussian noise in the state transition, with diagonal covariance matrix $\mathbf{Q}_\tau = \begin{bmatrix} (\sigma_\tau^\eta)^2 & 0 \\ 0 & (\sigma^\mu)^2 \end{bmatrix}$ and $(\sigma_\tau^\eta)^2 = \begin{cases} (\sigma^\eta)^2 & \tau = kI, k = 1, 2, \dots; \\ 0 & \text{otherwise} \end{cases}$;

- $v_\tau \sim \mathcal{N}(0, r)$ is the i.i.d. Gaussian noise in the observation;
- \mathbf{x}_1 is the initial state, and it is assumed to follow $\mathcal{N}(\boldsymbol{\pi}_1, \boldsymbol{\Sigma}_1)$ distribution.

Importantly, η_τ is a time series with piece constant values, that is, it is the same for all the bins in the same day and changes from day by day, therefore, $a_\tau^\eta = 1$ and $(\sigma_\tau^\eta)^2 = 0$ within one trading day. The observation equation (5) is the same as Equations (3) and (4) that models the evolution of the underlying volume components over time. In the proposed model, $\boldsymbol{\pi}_1$, $\boldsymbol{\Sigma}_1$, \mathbf{A}_τ , \mathbf{Q}_τ , r , and the seasonality $\boldsymbol{\phi} = [\phi_1 \dots \phi_I]^\top$ are treated as the unknown parameters. We group these parameters of interest into a new vector

$$\boldsymbol{\theta} = (\boldsymbol{\pi}_1; \boldsymbol{\Sigma}_1; a^\eta; a^\mu; (\sigma^\eta)^2; (\sigma^\mu)^2; r; \boldsymbol{\phi}), \quad (6)$$

and we will estimate $\boldsymbol{\theta}$ during the model calibration.

Figure 3 presents a general block diagram of the proposed method for intraday log-volume prediction, consisting of a prediction part and a model calibration part that will be specified in the remainder of this section.

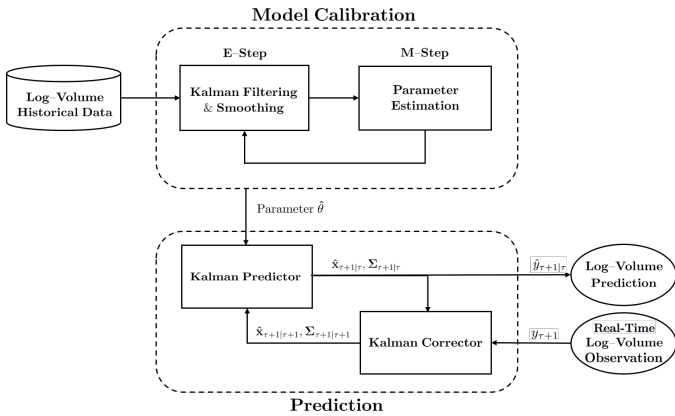


Figure 3: A block diagram of the proposed method, consisting of a prediction part and a model calibration part.

2.2. Prediction: the Kalman Filtering

The prediction problem is indeed to model the distribution of hidden state $\mathbf{x}_{\tau+1}$ conditional on all the log-volume observations available up to time τ (information set denoted as \mathcal{F}_τ). With the assumption of Gaussian noise \mathbf{w}_τ and v_τ in Equations (4) and (5), the hidden state $\mathbf{x}_{\tau+1}$ follows a Gaussian distribution as well, and it is only necessary to characterize the conditional mean and the conditional covariance:

$$\hat{\mathbf{x}}_{\tau+1|\tau} \triangleq \mathbb{E}[\mathbf{x}_{\tau+1} | \mathcal{F}_\tau], \quad (7)$$

$$\boldsymbol{\Sigma}_{\tau+1|\tau} \triangleq \text{Cov}[\mathbf{x}_{\tau+1} | \mathcal{F}_\tau]. \quad (8)$$

Kalman filtering is an online algorithm to precisely estimate the mean and covariance matrix. Assume the param-

eters in $\boldsymbol{\theta}$ are known, Algorithm 1 outlines the prediction-correction feedback control mechanism of the Kalman filter. At time τ , in the prediction steps shown at lines 2 to 3, the Kalman filter uses the state estimate $(\hat{\mathbf{x}}_{\tau|\tau}, \boldsymbol{\Sigma}_{\tau|\tau})$ to predict the conditional mean and covariance $(\hat{\mathbf{x}}_{\tau+1|\tau}, \boldsymbol{\Sigma}_{\tau+1|\tau})$ for time $\tau + 1$. This predicted state estimate is known as the *a priori* state estimate because it does not include any information from time $\tau + 1$. Once the actual log-volume $y_{\tau+1}$ is observed at $\tau + 1$, in the correction steps shown at lines 4 to 6, the predicted state estimate $(\hat{\mathbf{x}}_{\tau+1|\tau}, \boldsymbol{\Sigma}_{\tau+1|\tau})$ is refined into $(\hat{\mathbf{x}}_{\tau+1|\tau+1}, \boldsymbol{\Sigma}_{\tau+1|\tau+1})$ by incorporating the latest observation. The conditional mean and covariance are corrected through a linear combination of the *a priori* estimate and a weighted difference between the predicted and the observed log-volume, where the weight, also known as the optimal Kalman gain, is computed at line 4. The corrected estimate is termed the *a posteriori* state estimate, and it is a statistically optimal estimate for $\mathbf{x}_{\tau+1}$.

Algorithm 1 The Kalman filtering algorithm.

Require: the parameters in $\boldsymbol{\theta}$ are known

- 1: **for** $\tau = 1, 2, \dots$, **do**
 - 2: predict mean: $\hat{\mathbf{x}}_{\tau+1|\tau} = \mathbf{A}_\tau \hat{\mathbf{x}}_{\tau|\tau}$
 - 3: predict covariance: $\boldsymbol{\Sigma}_{\tau+1|\tau} = \mathbf{A}_\tau \boldsymbol{\Sigma}_{\tau|\tau} \mathbf{A}_\tau^\top + \mathbf{Q}_\tau$
 - 4: compute Kalman gain:
 $\mathbf{K}_{\tau+1} = \boldsymbol{\Sigma}_{\tau+1|\tau} \mathbf{C}^\top (\mathbf{C} \boldsymbol{\Sigma}_{\tau+1|\tau} \mathbf{C}^\top + r)^{-1}$
 - 5: correct conditional mean:
 $\hat{\mathbf{x}}_{\tau+1|\tau+1} = \hat{\mathbf{x}}_{\tau+1|\tau} + \mathbf{K}_{\tau+1} (y_{\tau+1} - \hat{\phi}_{\tau+1} - \mathbf{C} \hat{\mathbf{x}}_{\tau+1|\tau})$
 - 6: correct conditional covariance:
 $\boldsymbol{\Sigma}_{\tau+1|\tau+1} = \boldsymbol{\Sigma}_{\tau+1|\tau} - \mathbf{K}_{\tau+1} \mathbf{C} \boldsymbol{\Sigma}_{\tau+1|\tau}$
 - 7: **end for**
-

Typically, the two steps are performed for each bin to forecast the log-volume one-bin-ahead (referred to as dynamic prediction), with the prediction steps advancing the state estimate until the next volume observation, and the correction steps incorporating the new observation. If multi-bin-ahead volume forecasts (referred to as static prediction) are required or a certain number of observations are unavailable, however, the correction steps in between are skipped and multiple prediction steps are performed in succession. For $h = 1, 2, \dots$, with the predicted states $\hat{\mathbf{x}}_{\tau+h|\tau} = [\hat{\eta}_{\tau+h|\tau} \ \hat{\mu}_{\tau+h|\tau}]^\top$ and the estimated intraday periodic component $\hat{\phi}_{\tau+h}$, we can construct the h -step ahead log-volume forecast by:

$$\begin{aligned} \hat{y}_{\tau+h|\tau} &= \hat{\eta}_{\tau+h|\tau} + \hat{\mu}_{\tau+h|\tau} + \hat{\phi}_{\tau+h} \\ &= \mathbf{C} \hat{\mathbf{x}}_{\tau+h|\tau} + \hat{\phi}_{\tau+h}. \end{aligned} \quad (9)$$

If necessary, this can be easily converted back into the linear scale for further applications.

2.3. Model Calibration

The Kalman prediction presented in Algorithm 1 requires knowledge of the parameters in $\boldsymbol{\theta}$, but in practice they have to be estimated, typically referred to as model

calibration. Assume the log-volume historical data used for model calibration are $\{y_\tau\}_{\tau=1}^N$, containing T trading days and $N = T \times I$ observed intraday bins, and $\{\mathbf{x}_\tau\}_{\tau=1}^N$ are the corresponding hidden states. We wish to estimate θ defined in Equation (6) to fit the state-space model (4)-(5) to the training data $\{y_\tau\}_{\tau=1}^N$. This is achieved by combining the Kalman filtering, the Kalman smoothing, and the EM algorithm. The upper part of Figure 3 shows the model calibration mechanism.

2.3.1. The Kalman Smoothing

Although our main problem of interest is to obtain the future volume prediction by modeling the distribution of $\mathbf{x}_{\tau+1}$ conditional on \mathcal{F}_τ , the inference of the past states $\mathbf{x}_1 \dots \mathbf{x}_N$ conditional on all the observations in the training set (denoted as \mathcal{F}_N) is a necessary step in model calibration because it provides more accurate information of the unobservable states. This process is known as the Kalman smoothing. As illustrated in Figure 3, it serves as one building block within model calibration to produce smoothed states estimates. But once the parameters are well-estimated, only the Kalman filtering is necessary for the predictions. The Kalman smoothing uses the outputs from the filtering algorithm and infers $(\hat{\mathbf{x}}_{\tau|N}, \Sigma_{\tau|N})$, where

$$\hat{\mathbf{x}}_{\tau|N} \triangleq E[\mathbf{x}_\tau | \mathcal{F}_N], \quad (10)$$

$$\Sigma_{\tau|N} \triangleq \text{Cov}[\mathbf{x}_\tau | \mathcal{F}_N], \quad (11)$$

are the mean and covariance of \mathbf{x}_τ conditional on all the observations $\{y_\tau\}_{\tau=1}^N$. This estimation is achieved by starting from the last time step and proceeding backwards in time using Algorithm 2.

Algorithm 2 The Kalman smoothing algorithm.

Require: the parameters in θ are known; $\hat{\mathbf{x}}_{N|N}$ and $\Sigma_{N|N}$ are computed

- 1: **for** $\tau = N - 1, \dots, 2, 1$ **do**
 - 2: $\mathbf{L}_\tau = \Sigma_{\tau|\tau} \mathbf{A}_\tau^\top \Sigma_{\tau+1|N}^{-1}$
 - 3: $\hat{\mathbf{x}}_{\tau|N} = \hat{\mathbf{x}}_{\tau|\tau} + \mathbf{L}_\tau (\hat{\mathbf{x}}_{\tau+1|N} - \hat{\mathbf{x}}_{\tau+1|\tau})$
 - 4: $\Sigma_{\tau|N} = \Sigma_{\tau|\tau} + \mathbf{L}_\tau (\Sigma_{\tau+1|N} - \Sigma_{\tau+1|\tau}) \mathbf{L}_\tau^\top$
 - 5: **end for**
-

2.3.2. Parameter Estimation

Once we have the estimates for $\{\hat{\mathbf{x}}_{1|N} \dots \hat{\mathbf{x}}_{N|N}\}$ and $\{\Sigma_{1|N} \dots \Sigma_{N|N}\}$, we proceed to the actual estimation of the parameters in θ . We apply the EM algorithm to estimate parameters for the proposed Kalman filter model. The EM extends the maximum likelihood estimation to cases where hidden states are involved (e.g., Shumway and Stoffer (1982)). It is an iterative method of finding maximum likelihood estimates of parameters. The EM iteration alternates between performing an E-step (i.e., Expectation step), which constructs a global convex lower bound of the expectation of log-likelihood using the current estimation

of parameters:

$$\mathcal{Q}(\theta | \hat{\theta}^{(j)}) = E_{\{\mathbf{x}\}|\{y\}, \hat{\theta}^{(j)}} \left[\log P(\{\mathbf{x}_\tau\}_{\tau=1}^N, \{y_\tau\}_{\tau=1}^N) \right], \quad (12)$$

and an M-step (i.e., Maximization step), which computes parameters to maximize the lower bound found in previous step:

$$\hat{\theta}^{(j+1)} = \arg \max_{\theta} \mathcal{Q}(\theta | \hat{\theta}^{(j)}). \quad (13)$$

These parameters are then used to determine $\mathcal{Q}(\theta | \hat{\theta}^{(j+1)})$ in the next E-step. The derivation of Equations (12) and (13) are provided in Appendix A.

For the proposed model, the EM algorithm is efficient to implement and numerically stable. In the E-step, the implementation does not require heavy preparatory analytical work, because the expected log-likelihood \mathcal{Q} has three sufficient statistics:

$$\hat{\mathbf{x}}_\tau = E[\mathbf{x}_\tau | \{y_\tau\}_{\tau=1}^N, \hat{\theta}^{(j)}], \quad (14)$$

$$\mathbf{P}_\tau = E[\mathbf{x}_\tau \mathbf{x}_\tau^\top | \{y_\tau\}_{\tau=1}^N, \hat{\theta}^{(j)}], \quad (15)$$

$$\mathbf{P}_{\tau, \tau-1} = E[\mathbf{x}_\tau \mathbf{x}_{\tau-1}^\top | \{y_\tau\}_{\tau=1}^N, \hat{\theta}^{(j)}], \quad (16)$$

which can be easily calculated from the Kalman filtering and smoothing. In the M-step, we derive a set of closed-form solutions which greatly simplifies the parameter estimation process. The optimal closed-form updates of $\hat{\theta}^{(j+1)}$ are as follows (see Appendix A for the derivation):

$$\pi_1^{(j+1)} = \hat{\mathbf{x}}_1, \quad (17)$$

$$\Sigma_1^{(j+1)} = \mathbf{P}_1 - \hat{\mathbf{x}}_1 \hat{\mathbf{x}}_1^\top, \quad (18)$$

$$(a^\eta)^{(j+1)} = \left[\sum_{\tau=kI+1} \mathbf{P}_{\tau, \tau-1}^{(1,1)} \right] \left[\sum_{\tau=kI+1} \mathbf{P}_{\tau-1}^{(1,1)} \right]^{-1}, \quad (19)$$

$$(a^\mu)^{(j+1)} = \left[\sum_{\tau=2}^N \mathbf{P}_{\tau, \tau-1}^{(2,2)} \right] \left[\sum_{\tau=2}^N \mathbf{P}_{\tau-1}^{(2,2)} \right]^{-1}, \quad (20)$$

$$\begin{aligned} [(\sigma^\eta)^2]^{(j+1)} &= \frac{1}{T-1} \sum_{\tau=kI+1} \left\{ \mathbf{P}_\tau^{(1,1)} + [(a^\eta)^{(j+1)}]^2 \mathbf{P}_{\tau-1}^{(1,1)} \right. \\ &\quad \left. - 2(a^\eta)^{(j+1)} \mathbf{P}_{\tau, \tau-1}^{(1,1)} \right\}, \end{aligned} \quad (21)$$

$$\begin{aligned} [(\sigma^\mu)^2]^{(j+1)} &= \frac{1}{N-1} \sum_{\tau=2}^N \left\{ \mathbf{P}_\tau^{(2,2)} + [(a^\mu)^{(j+1)}]^2 \mathbf{P}_{\tau-1}^{(2,2)} \right. \\ &\quad \left. - 2(a^\mu)^{(j+1)} \mathbf{P}_{\tau, \tau-1}^{(2,2)} \right\}, \end{aligned} \quad (22)$$

$$\begin{aligned} r^{(j+1)} &= \frac{1}{N} \sum_{\tau=1}^N \left[y_\tau^2 + \mathbf{C} \mathbf{P}_\tau \mathbf{C}^\top - 2y_\tau \mathbf{C} \hat{\mathbf{x}}_\tau + (\phi_\tau^{(j+1)})^2 \right. \\ &\quad \left. - 2y_\tau \phi_\tau^{(j+1)} + 2\phi_\tau^{(j+1)} \mathbf{C} \hat{\mathbf{x}}_\tau \right], \end{aligned} \quad (23)$$

$$\phi_i^{(j+1)} = \frac{1}{T} \sum_{t=1}^T (y_{t,i} - \mathbf{C}\hat{\mathbf{x}}_{t,i}). \quad (24)$$

Algorithm 3 outlines the EM algorithm used to estimate the model parameters, and the corresponding derivation is provided in Appendix A.

Algorithm 3 The EM algorithm for the proposed Kalman filter model.

```

1: initial values of parameters in  $\hat{\boldsymbol{\theta}}^{(0)}$  are known
2: iteration:  $j \leftarrow 0$ 
3: repeat
4:   for all  $\tau = N, N-1, \dots, 1$  do
5:      $\hat{\mathbf{x}}_\tau = \hat{\mathbf{x}}_{\tau|N}$ 
6:      $\mathbf{P}_\tau = \boldsymbol{\Sigma}_{\tau|N} + \hat{\mathbf{x}}_{\tau|N} \hat{\mathbf{x}}_{\tau|N}^\top$ 
7:      $\boldsymbol{\Sigma}_{\tau, \tau-1|N} = \boldsymbol{\Sigma}_{\tau|\tau} \mathbf{L}_{\tau-1}^\top$ 
         $\quad + \mathbf{L}_\tau (\boldsymbol{\Sigma}_{\tau+1, \tau|N} - \mathbf{A}_\tau \boldsymbol{\Sigma}_{\tau|\tau}) \mathbf{L}_{\tau-1}^\top$ 
8:      $\mathbf{P}_{\tau, \tau-1} = \boldsymbol{\Sigma}_{\tau, \tau-1|N} + \hat{\mathbf{x}}_{\tau|N} \hat{\mathbf{x}}_{\tau-1|N}^\top$ 
9:   end for
10:   $\boldsymbol{\pi}_1^{(j+1)} = \hat{\mathbf{x}}_1$ 
11:   $\boldsymbol{\Sigma}_1^{(j+1)} = \mathbf{P}_1 - \hat{\mathbf{x}}_1 \hat{\mathbf{x}}_1^\top$ 
12:   $(a^\eta)^{(j+1)} = \left[ \sum_{\tau=kI+1}^N \mathbf{P}_{\tau, \tau-1}^{(1,1)} \right] \left[ \sum_{\tau=kI+1}^N \mathbf{P}_{\tau-1}^{(1,1)} \right]^{-1}$ 
13:   $(a^\mu)^{(j+1)} = \left[ \sum_{\tau=2}^N \mathbf{P}_{\tau, \tau-1}^{(2,2)} \right] \left[ \sum_{\tau=2}^N \mathbf{P}_{\tau-1}^{(2,2)} \right]^{-1}$ 
14:   $[(\sigma^\eta)^2]^{(j+1)} = \frac{1}{T-1} \sum_{\tau=kI+1}^N \left\{ \mathbf{P}_\tau^{(1,1)} \right.$ 
         $\quad \left. + \left[ (a^\eta)^{(j+1)} \right]^2 \mathbf{P}_{\tau-1}^{(1,1)} - 2(a^\eta)^{(j+1)} \mathbf{P}_{\tau, \tau-1}^{(1,1)} \right\}$ 
15:   $[(\sigma^\mu)^2]^{(j+1)} = \frac{1}{N-1} \sum_{\tau=2}^N \left\{ \mathbf{P}_\tau^{(2,2)} + \left[ (a^\mu)^{(j+1)} \right]^2 \mathbf{P}_{\tau-1}^{(2,2)} \right.$ 
         $\quad \left. - 2(a^\mu)^{(j+1)} \mathbf{P}_{\tau, \tau-1}^{(2,2)} \right\}$ 
16:   $r^{(j+1)} = \frac{1}{N} \sum_{\tau=1}^N \left[ y_\tau^2 + \mathbf{C} \mathbf{P}_\tau \mathbf{C}^\top - 2y_\tau \mathbf{C} \hat{\mathbf{x}}_\tau + \left( \phi_\tau^{(j+1)} \right)^2 \right.$ 
         $\quad \left. - 2y_\tau \phi_\tau^{(j+1)} + 2\phi_\tau^{(j+1)} \mathbf{C} \hat{\mathbf{x}}_\tau \right]$ 
17:   $\phi_i^{(j+1)} = \frac{1}{T} \sum_{t=1}^T (y_{t,i} - \mathbf{C}\hat{\mathbf{x}}_{t,i})$ 
18: until convergence

```

2.3.3. Numerical Illustration on the Convergence of the EM

Recall that the parameter estimation method of the benchmark CMEM is highly sensitive to the choice of initial parameters, making the model difficult to be used in practice. To evaluate the performance of the EM, we conduct synthetic experiments with known actual parameter values. Figure 4 shows examples when different initial values are assigned to estimate model parameters. From these experiments we can conclude that the parameters are able to converge within a few iterations, and the estimated results are robust to the choice of initial parameters.

3. Outlier Robust Intraday Volume Prediction

In this section, we present a variation on the standard Kalman filter, which is designed to be more robust to the

observation noise. The Kalman filter model defined in Equations (4) and (5) assumes that all the noise terms follow a Gaussian distribution, but real-time trading volume data is expected to have outliers that deviate notably from the curated data. Therefore, it is beneficial to design a robust model that can avoid overfitting to the irrelevant information and perform consistently well even under extremely noisy conditions.

3.1. The Robust Kalman Filter via the Lasso Regularization

We assume a similar model for the log-volume y_τ to Equation (3), but with an additional noise term z_τ :

$$y_\tau = \eta_\tau + \phi_\tau + \mu_\tau + v_\tau + z_\tau, \quad (25)$$

where z_τ is a sparse outlier component, i.e., most of the time is zero with a few exceptions when the value can be very large. Compared to Equations (4) and (5), the robust state-space model is refined as

$$\mathbf{x}_{\tau+1} = \mathbf{A}_\tau \mathbf{x}_\tau + \mathbf{w}_\tau, \quad (26)$$

$$y_\tau = \mathbf{C} \mathbf{x}_\tau + \phi_\tau + v_\tau + z_\tau, \quad (27)$$

where the only difference is the term z_τ .

To handle the additional sparse noise term, the Kalman correction step needs to be modified. In the standard correction step, the *a posteriori* state estimate $\hat{\mathbf{x}}_{\tau+1|\tau+1}$ in Algorithm 1 can be interpreted as the result of solving the quadratic minimization problem:

$$\begin{aligned} \min_{\mathbf{x}_{\tau+1}} \quad & (\mathbf{x}_{\tau+1} - \hat{\mathbf{x}}_{\tau+1|\tau})^\top \boldsymbol{\Sigma}_{\tau+1|\tau}^{-1} (\mathbf{x}_{\tau+1} - \hat{\mathbf{x}}_{\tau+1|\tau}) \\ & + v_{\tau+1}^\top r^{-1} v_{\tau+1} \\ \text{s.t.} \quad & y_{\tau+1} = \mathbf{C} \mathbf{x}_{\tau+1} + \phi_{\tau+1} + v_{\tau+1}, \end{aligned} \quad (28)$$

where the first term in the objective is a loss term corresponding to the latest observation noise, and the second is a loss term associated with the *a priori* state estimate. In the robust Kalman filter, we apply the Lasso regularization introduced by Tibshirani (1996) to the Kalman filter and solve the following convex minimization problem with an l_1 -norm term to handle the sparse noise:

$$\begin{aligned} \min_{\mathbf{x}_{\tau+1}, z_{\tau+1}} \quad & (\mathbf{x}_{\tau+1} - \hat{\mathbf{x}}_{\tau+1|\tau})^\top \boldsymbol{\Sigma}_{\tau+1|\tau}^{-1} (\mathbf{x}_{\tau+1} - \hat{\mathbf{x}}_{\tau+1|\tau}) \\ & + v_{\tau+1}^\top r^{-1} v_{\tau+1} + \lambda |z_{\tau+1}| \\ \text{s.t.} \quad & y_{\tau+1} = \mathbf{C} \mathbf{x}_{\tau+1} + \phi_{\tau+1} + v_{\tau+1} + z_{\tau+1}. \end{aligned} \quad (29)$$

Mattingley and Boyd (2010) introduced an equivalent but more computationally efficient way by setting $e_\tau = y_\tau - \phi_\tau - \mathbf{C} \hat{\mathbf{x}}_{\tau|\tau-1}$ and in each time solving

$$\min_{z_{\tau+1}} (e_{\tau+1} - z_{\tau+1})^\top W_{\tau+1} (e_{\tau+1} - z_{\tau+1}) + \lambda |z_{\tau+1}|, \quad (30)$$

where $W_{\tau+1} = (\mathbf{C} \boldsymbol{\Sigma}_{\tau+1|\tau} \mathbf{C}^\top + r)^{-1}$, and λ is the regularization parameter to be adjusted suitably. Summarizing,

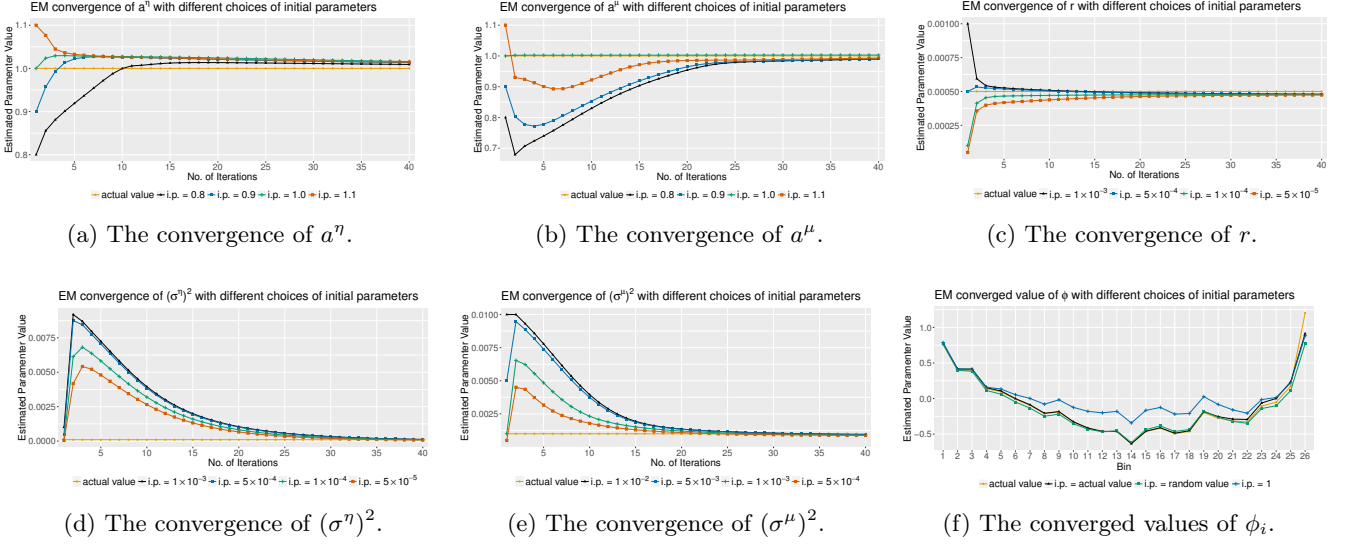


Figure 4: Numerical illustration on the convergence of the EM algorithm with different choices of initial parameters. These results are from synthetic experiments with known actual parameter values.

the correction step at line 5 of Algorithm 1 is replaced with

$$e_{\tau+1} = y_{\tau+1} - \hat{\phi}_{\tau+1} - \mathbf{C}\hat{\mathbf{x}}_{\tau+1|\tau}, \quad (31)$$

$$\hat{\mathbf{x}}_{\tau+1|\tau+1} = \hat{\mathbf{x}}_{\tau+1|\tau} + \mathbf{K}_{\tau+1} (e_{\tau+1} - z_{\tau+1}^*), \quad (32)$$

where z_τ^* is the solution of problem (30), which can be solved analytically because W_τ is a scalar for the proposed model, and the optimal solution of z_τ^* is given by

$$z_\tau^* = \begin{cases} e_\tau - \frac{\lambda}{2W_\tau} & \left(e_\tau > \frac{\lambda}{2W_\tau} \right) \\ 0 & \left(-\frac{\lambda}{2W_\tau} \leq e_\tau \leq \frac{\lambda}{2W_\tau} \right) \\ e_\tau + \frac{\lambda}{2W_\tau} & \left(e_\tau < -\frac{\lambda}{2W_\tau} \right) \end{cases}, \quad (33)$$

and

$$e_\tau - z_\tau^* = \begin{cases} \frac{\lambda}{2W_\tau} & \left(e_\tau > \frac{\lambda}{2W_\tau} \right) \\ e_\tau & \left(-\frac{\lambda}{2W_\tau} \leq e_\tau \leq \frac{\lambda}{2W_\tau} \right) \\ -\frac{\lambda}{2W_\tau} & \left(e_\tau < -\frac{\lambda}{2W_\tau} \right) \end{cases}. \quad (34)$$

As presented in Equation (34), the Lasso regularization gives sparse solutions and generates a threshold $\frac{\lambda}{2W_\tau}$ to truncate the outliers. The quantity $\frac{\lambda}{2W_\tau}$ can be interpreted as a width of a distribution of e_τ without the outlier z_τ .

3.2. Model Calibration of the Robust Kalman Filter

The EM method for the standard Kalman filter model can be applied to the robust model with some modifications. Let $z_1^* \dots z_N^*$ denote the solutions of problem (30) calculated in the E-step, the estimations of parameter r and ϕ_i in the M-step in Algorithm 3 are replaced with

$$r^{(j+1)} = \frac{1}{N} \sum_{\tau=1}^N \left[y_\tau^2 + \mathbf{C}\mathbf{P}_\tau \mathbf{C}^\top - 2y_\tau \mathbf{C}\hat{\mathbf{x}}_\tau + \left(\phi_\tau^{(j+1)} \right)^2 \right.$$

$$\left. - 2y_\tau \phi_\tau^{(j+1)} + 2\phi_\tau^{(j+1)} \mathbf{C}\hat{\mathbf{x}}_\tau + (z_\tau^*)^2 - 2z_\tau^* y_\tau + 2z_\tau^* \mathbf{C}\hat{\mathbf{x}}_\tau + 2z_\tau^* \phi_\tau^{(j+1)} \right], \quad (35)$$

$$\phi_i^{(j+1)} = \frac{1}{T} \sum_{t=1}^T (y_{t,i} - \mathbf{C}\hat{\mathbf{x}}_{t,i} - z_{t,i}^*). \quad (36)$$

3.3. Model Validation with Numerical Simulations

We now provide a numerical example to validate the effectiveness of the robust model under noisy conditions. Real-time trading volume data is expected to have outliers, as opposed to the curated historical data we use in our experiments, that is why we need to artificially introduce outliers. The tickers we use for numerical simulations are SPY, DIA, and IBM. For each ticker, we randomly select 10% of the total number of bins to add synthetic outliers so that the contaminated volume data will have significantly more samples with extreme values. To illustrate the impact of different levels of outliers, for each ticker, three contaminated data series are generated with small, medium, and large scales of outliers respectively. We train the models and make out-of-sample predictions using the contaminated data, but compare the predicted volume with the clean data to evaluate model performance. The mean absolute percentage error (MAPE) is used to measure the volume prediction error. Let M be the total number of bins in the out-of-sample set, mathematically, the MAPE is expressed as

$$\text{MAPE} = \frac{1}{M} \sum_{\tau=1}^M \frac{|\text{volume}_\tau - \text{predicted volume}_\tau|}{\text{volume}_\tau}. \quad (37)$$

Table 1 summarizes the out-of-sample MAPE given by different models. Under a moderate condition with small-scale outliers, the predictive capability of the robust

Kalman filter is comparable with the standard Kalman filter, and both models outperform the benchmark CMEM and RM. But as the noise scale increases, the differences become evident. The performances of the Kalman filter, the CMEM and the RM degrade much faster than the robust Kalman filter. These results demonstrate that the robust Kalman filter is much less sensitive to the extreme outliers and provides a more robust estimation. Another fact worth mentioning is that when the noise scale is large, the CMEM fails to find solutions during the parameter estimation (indicated as N/A in the table), raising some concerns regarding the feasibility of the model.

4. Empirical Studies

4.1. Data and Set-Up

Our empirical studies analyze intraday volume of thirty well-diversified securities, including twelve ETFs and eighteen stocks traded on major U.S., European, and Asian markets. The tickers are summarized in Table 2. The data covers the period between January 2013 and June 2016, excluding half-day trading days. Each trading day is divided into I 15-minute bins (note that the total number of bins per day I varies from market to market, depending on the trading hours of the stock exchange). Volume is computed as Equation (1), and the last recorded transaction price in each bin is used as the reference price. All the historical data used in the analysis are obtained from the financial data provider Bloomberg.

Data between June 2015 and June 2016 are considered as the out-of-sample portion and used to evaluate model performance. This portion contains $D = 250$ trading days or equivalently $M = D \times I$ bins. A standard rolling window forecasting scheme is employed. For a given training length N , each bin in the out-of-sample set is predicted using the parameters estimated from data in the previous N bins. The optimal training data length N^* and the optimal Lasso regularization parameter λ^* are determined by the cross-validation method. Specifically, data between January 2015 and May 2015 are considered as a cross-validation set. We repeat performing the rolling window forecasting on the cross-validation set using different values of N and λ , and choose the pair that gives the minimum predictive error rate as the optimal pair.

4.2. Intraday Volume Prediction Out-of-Sample Results

The numerical experiments consist of two types of intraday volume predictions, namely, the static prediction and the dynamic prediction. The static prediction refers to forecasting all the I bins of day $t + 1$ using information up to day t only. The dynamic prediction stands for the one-bin-ahead forecasting, where the volume at one particular bin is predicted based on all the information up to the previous bin. The predictive performances of different models are evaluated with the MAPE. We also calculate

the relative improvement with respect to a given benchmark model, defined as

$$\text{improvement}_i = 100 \times \frac{\text{error}_{\text{benchmark}} - \text{error}_{\text{model } i}}{\text{error}_{\text{benchmark}}}, \quad (38)$$

where the MAPE is used as the error measurement in the above equation.

Table 3 presents the out-of-sample volume prediction MAPE given by different models, where the best model for a given type of prediction is highlighted in boldface. In terms of the dynamic prediction, the proposed Kalman filter models significantly outperform both the benchmarks, RM and CMEM, in all thirty of the securities. As for the static prediction, the proposed Kalman filter models outperform the benchmarks in twenty-nine securities. When comparing the two proposed models, since the historical data used in the experiments has been curated, both models give very comparable results for all cases, but the robust Kalman filter slightly outperforming the other on average.

Figure 5 (a) and (b) show the relative improvements of the dynamic and the static predictions given by the robust Kalman filter, the standard Kalman filter, and the CMEM over the benchmark RM. Figure 5 (c) and (d) show improvements of the dynamic volume predictions of the robust Kalman filter and the standard Kalman filter over the dynamic CMEM, as well as the static volume predictions over the static CMEM.

Overall, the robust Kalman filter with dynamic prediction performs the best in the empirical studies. It gives an average MAPE of 0.46, or equivalently, an improvement of 64% over the RM, and of 29% over the dynamic CMEM.

4.3. VWAP Replication Out-of-Sample Results

We also evaluate the performance of different methods on minimizing the VWAP order execution risk. The concept of VWAP is an average of intraday transaction prices weighted by the corresponding intraday volume percentages:

$$\begin{aligned} \text{VWAP}_t &= \frac{\sum_{i=1}^I \text{volume}_{t,i} \times \text{price}_{t,i}}{\sum_{i=1}^I \text{volume}_{t,i}} \\ &= \sum_{i=1}^I w_{t,i} \times \text{price}_{t,i}, \end{aligned} \quad (39)$$

where $\text{price}_{t,i}$ is the last recorded transaction price within the bin i of day t , acting as a proxy for the VWAP of that bin, and $w_{t,i} = \frac{\text{volume}_{t,i}}{\sum_{i=1}^I \text{volume}_{t,i}}$ is the corresponding volume weight of this interval. Two types of VWAP trading replication strategies have been introduced by Białkowski et al. (2008). The static VWAP replication strategy assumes that the order slicing is set before the market opening and

Table 1: Comparison of out-of-sample MAPE for the contaminated data. Different levels of synthetic outliers are introduced into the intraday volume data for ticker SPY, DIA, IBM. The best model for a given type of prediction is highlighted in boldface. N/A indicates that the model fails during numerical simulations.

Ticker	Outliers	Dynamic Volume Prediction			Static Volume Prediction			
		Robust Kalman Filter	Kalman Filter	CMEM	Robust Kalman Filter	Kalman Filter	CMEM	RM
SPY	No	0.24	0.24	0.26	0.36	0.36	0.39	0.41
	Small	0.27	0.29	0.48	0.37	0.37	0.56	0.54
	Medium	0.27	0.30	N/A	0.38	0.39	N/A	0.61
	Large	0.29	0.33	N/A	0.40	0.44	N/A	0.68
DIA	No	0.38	0.38	0.45	0.51	0.51	0.64	0.73
	Small	0.39	0.40	0.79	0.51	0.52	0.90	0.91
	Medium	0.40	0.43	1.13	0.53	0.54	1.12	1.15
	Large	0.43	0.48	1.18	0.56	0.61	1.21	1.32
IBM	No	0.24	0.24	0.28	0.35	0.35	0.38	0.42
	Small	0.26	0.28	N/A	0.35	0.36	N/A	0.56
	Medium	0.27	0.30	N/A	0.36	0.37	N/A	0.58
	Large	0.28	0.32	N/A	0.39	0.43	N/A	0.64

Table 2: A summary of the intraday volume data used in our empirical studies.

Country	Exchange	Ticker	Name	Mean	Std	Q5	Q95
U.S.	NYSE	C	Citigroup	0.0386	0.0402	0.0101	0.1084
U.S.	NYSE	IBM	IBM	0.0815	0.1031	0.0155	0.2420
U.S.	NYSE	KO	Coca-Cola	0.0258	0.0222	0.0070	0.0667
U.S.	NYSE	WMT	Wal-Mart	0.0133	0.0139	0.0040	0.0332
U.S.	NASDAQ	AAPL	Apple	0.0110	0.0098	0.0035	0.0277
U.S.	NASDAQ	FB	Facebook	0.0079	0.0084	0.0023	0.0199
U.S.	NYSEArca	SPY	SPDR S&P 500 ETF	0.4305	0.3723	0.1200	1.1115
U.S.	NYSEArca	DIA	SPDR Dow Jones Industrial Average ETF	0.3109	0.3046	0.0615	0.8650
U.S.	NYSEArca	XOP	SPDR S&P Oil & Gas Explor & Prodtn ETF	0.2591	0.2386	0.0603	0.7168
U.S.	NYSEArca	OIH	VanEck Vectors Oil Services ETF	0.2367	0.2983	0.0280	0.7299
U.S.	NasdaqGM	QQQ	PowerShares QQQ ETF	1.2281	1.1784	0.2277	3.3477
U.S.	NasdaqGM	IBB	iShares Nasdaq Biotechnology ETF	0.6595	0.7221	0.0896	1.9531
France	EPA	OR	L'Oreal	0.0129	0.0168	0.0026	0.0331
France	EPA	BNP	BNP Paribas	0.0096	0.0091	0.0020	0.0250
U.K.	LON	ULVR	Unilever	0.0027	0.0025	0.0006	0.0068
U.K.	LON	MKS	Marks and Spencer	0.0085	0.0071	0.0022	0.0205
Germany	ETR	ADS	Adidas	0.0044	0.0039	0.0010	0.0111
Netherlands	Amsterdam	PHIA	Philips	0.0058	0.0073	0.0009	0.0162
Japan	TYO	6758JP	Sony	0.0040	0.0048	0.0007	0.0114
Japan	TYO	7731JP	Nikon	0.0063	0.0082	0.0007	0.0203
Japan	TYO	8604JP	Nomura Holdings	0.0038	0.0041	0.0009	0.0113
Japan	TYO	1306JP	Topix ETF	0.0671	0.0912	0.0133	0.2027
Japan	TYO	1321JP	Nikkei 225 ETF	0.0478	0.0553	0.0104	0.1456
Japan	TYO	1330JP	Nikko ETIF 225	0.0530	0.0675	0.0095	0.1661
Japan	TYO	1570JP	Topix-1 Nikkei 225 Leveraged ETF	0.0746	0.1007	0.0014	0.2623
Hong Kong	HKEX	5HK	HSBC Holdings	0.1624	0.1738	0.0158	0.4710
Hong Kong	HKEX	700HK	Tencent Holdings	0.0162	0.0228	0.0013	0.0565
Hong Kong	HKEX	941HK	China Mobile	0.0282	0.0407	0.0023	0.0962
Hong Kong	HKEX	2800HK	Tracker Fund of Hong Kong ETF	0.0262	0.0435	0.0009	0.1029
Hong Kong	HKEX	2823HK	A50 China Tracker ETF	2.0580	2.4431	0.2745	6.4978

Table 3: The out-of-sample volume prediction MAPE results. The best model for a given type of prediction is highlighted with boldface.

Ticker	Dynamic Volume Prediction						Static Volume Prediction							
	Robust		Kalman		CMEM		Robust		Kalman		CMEM		RM	
	Kalman	Filter	Filter				Kalman	Filter	Filter					
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
AAPL	0.21	0.17	0.21	0.17	0.23	0.20	0.32	0.27	0.32	0.27	0.36	0.32	0.44	0.40
FB	0.23	0.19	0.23	0.19	0.25	0.24	0.34	0.29	0.34	0.29	0.38	0.37	0.47	0.43
C	0.24	0.21	0.25	0.21	0.27	0.25	0.35	0.29	0.35	0.29	0.37	0.33	0.41	0.36
IBM	0.24	0.20	0.24	0.21	0.27	0.28	0.35	0.29	0.35	0.30	0.38	0.35	0.42	0.41
KO	0.25	0.21	0.25	0.22	0.29	0.27	0.35	0.30	0.35	0.31	0.39	0.36	0.41	0.39
WMT	0.23	0.19	0.23	0.19	0.24	0.21	0.33	0.27	0.33	0.29	0.31	0.26	0.33	0.24
SPY	0.24	0.19	0.24	0.19	0.26	0.22	0.36	0.31	0.36	0.31	0.39	0.35	0.41	0.38
DIA	0.38	0.35	0.38	0.36	0.45	0.47	0.51	0.51	0.51	0.51	0.64	0.72	0.73	0.87
QQQ	0.30	0.26	0.30	0.27	0.33	0.31	0.42	0.41	0.41	0.39	0.45	0.44	0.45	0.44
IBB	0.42	0.41	0.42	0.42	0.49	0.54	0.58	0.61	0.58	0.61	0.72	0.84	0.73	0.89
XOP	0.35	0.31	0.35	0.32	0.41	0.40	0.43	0.42	0.43	0.41	0.53	0.55	0.71	0.84
OIH	0.40	0.39	0.41	0.41	0.50	0.58	0.52	0.57	0.54	0.63	0.64	0.77	0.69	0.83
ADS	0.42	0.45	0.42	0.46	0.57	0.72	0.59	0.67	0.59	0.67	0.87	1.07	1.08	1.38
PHIA	0.40	0.40	0.41	0.41	0.50	0.57	0.53	0.57	0.53	0.57	0.65	0.75	0.70	0.81
OR	0.39	0.40	0.40	0.42	0.48	0.56	0.52	0.58	0.52	0.59	0.61	0.74	0.65	0.79
BNP	0.35	0.34	0.35	0.35	0.42	0.45	0.47	0.48	0.47	0.48	0.63	0.69	0.75	0.85
ULVR	0.43	0.46	0.43	0.46	0.53	0.64	0.52	0.59	0.52	0.59	0.68	0.82	0.73	0.92
MKS	0.70	8.07	0.69	8.06	0.89	10.47	0.89	12.41	0.86	11.87	1.18	17.03	1.20	16.62
5HK	0.41	0.43	0.41	0.44	0.49	0.60	0.58	0.66	0.59	0.67	0.72	1.02	0.83	1.26
700HK	0.32	0.29	0.32	0.29	0.37	0.37	0.46	0.47	0.46	0.48	0.57	0.63	0.59	0.69
941HK	0.31	0.29	0.31	0.29	0.36	0.37	0.42	0.42	0.42	0.42	0.53	0.60	0.79	0.94
6758JP	0.25	0.23	0.26	0.24	0.31	0.29	0.38	0.39	0.37	0.36	0.47	0.47	0.63	0.69
7731JP	0.33	0.32	0.33	0.33	0.39	0.45	0.48	0.51	0.48	0.49	0.60	0.67	0.68	0.81
8604JP	0.26	0.22	0.26	0.22	0.30	0.28	0.39	0.35	0.39	0.35	0.48	0.44	0.59	0.58
2800HK	1.94	9.12	1.96	9.52	3.53	16.76	2.20	10.86	2.28	12.07	4.15	21.21	4.45	21.28
2823HK	0.65	2.43	0.65	2.36	0.86	4.17	0.87	3.16	0.90	3.25	1.39	6.32	1.72	7.85
1306JP	0.96	15.50	0.93	14.17	1.57	29.18	1.42	26.66	1.36	23.43	2.43	52.30	3.21	56.93
1321JP	0.72	1.79	0.71	1.52	1.11	3.03	0.95	2.11	0.93	1.91	1.61	4.10	2.22	6.09
1330JP	1.28	5.40	1.27	5.39	2.45	9.45	1.53	6.33	1.52	6.22	3.29	13.13	9.89	39.72
1570JP	0.37	0.37	0.37	0.37	0.45	0.51	0.48	0.54	0.48	0.54	0.62	0.78	1.37	1.59
Average	0.46	1.65	0.47	1.62	0.65	2.76	0.61	2.41	0.62	2.32	0.90	4.28	1.28	5.54

it is not revised during the day. It is implemented as

$$\hat{w}_{t,i}^{(s)} = \frac{\widehat{\text{volume}}_{t,i}^{(s)}}{\sum_{i=1}^I \widehat{\text{volume}}_{t,i}^{(s)}}, \quad (40)$$

where $\widehat{\text{volume}}_{t,i}^{(s)}$ denotes the static volume predictions. In the dynamic VWAP replication strategy scenario, on the other hand, order slicing is revised at each new bin as a new intraday volume is gradually observed. It is implemented using slices with weights:

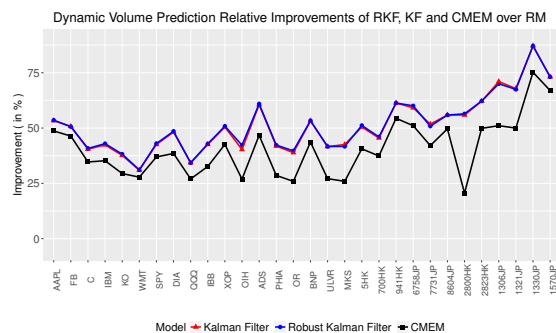
$$\hat{w}_{t,i}^{(d)} = \begin{cases} \frac{\widehat{\text{volume}}_{t,i}^{(d)}}{\sum_{j=i}^I \widehat{\text{volume}}_{t,j}^{(d)}} \left(1 - \sum_{j=1}^{i-1} \hat{w}_{t,j}^{(d)}\right), & i = 1, \dots, I-1 \\ \left(1 - \sum_{j=1}^{I-1} \hat{w}_{t,j}^{(d)}\right), & i = I \end{cases} \quad (41)$$

where $\widehat{\text{volume}}_{t,i}^{(d)}$ denotes the dynamic volume predictions. This weight is given by the proportion of predicted volume in bin i with respect to the sum of the remaining

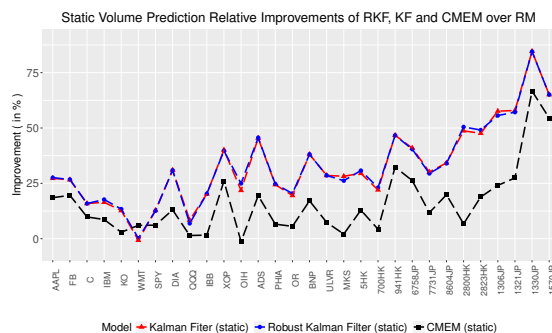
predicted volumes, multiplied by the slice proportion left to be traded. We adopt the VWAP tracking error to measure the VWAP order execution risk. Let D be the total number of days in the out-of-sample data set, the VWAP tracking error is defined as

$$\text{VWAP}^{\text{TE}} = \frac{1}{D} \sum_{t=1}^D \frac{|\text{VWAP}_t - \text{replicated VWAP}_t|}{\text{VWAP}_t}. \quad (42)$$

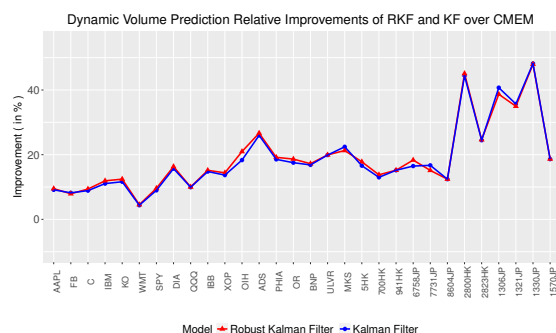
Table 4 presents the VWAP tracking error rates given by different methods. The errors are expressed in basis point (bps), where 1 bps = 0.01%, and the best model for a given type of VWAP tracking is highlighted in boldface. When comparing the proposed Kalman filter models with the RM, the proposed models with the dynamic VWAP strategy outperform the RM in all thirty of the securities, among which half of the securities have gains greater than 1 bps, and the proposed models with the static VWAP strategy give a lower execution error than the



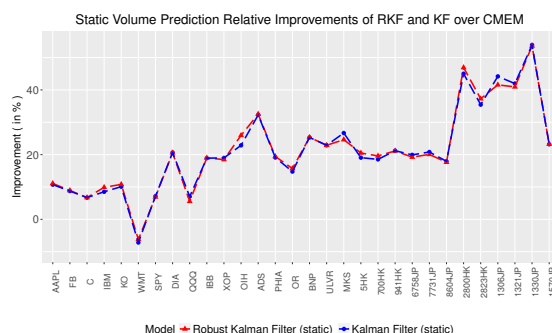
(a) Dynamic volume prediction relative improvements when the RM is considered as the benchmark.



(b) Static volume prediction relative improvements when the RM is considered as the benchmark.



(c) Dynamic volume prediction relative improvements when the dynamic CMEM is considered as the benchmark.



(d) Static volume prediction relative improvements when the static CMEM is considered as the benchmark.

Figure 5: The out-of-sample volume prediction relative improvements of the robust Kalman filter and the standard Kalman filter over the benchmark RM and the state-of-the-art CMEM.

RM in twenty-six securities, with errors in seven securities being reduced by more than 1 bps. When comparing with the CMEM, the performances of the Kalman filter models remain good. For the dynamic strategy, the Kalman filter models outperform in twenty-eight of the thirty securities, among which five securities have gains greater than 1 bps. For the static strategy, the Kalman filter models show a lower execution error than the CMEM in all thirty of the securities, with errors in ten being improved by more than 1 bps. When comparing the two proposed models, both models give very comparable results for all cases, but the robust Kalman filter performs slightly better on average. Note that a gain with greater than 1 bps is often considered as a significant improvement especially for big brokers. If a broker has a daily turnover of 10 billion USD, then saving 1 bps in execution means earning 1 million USD ($10 \text{ billion} \times 1 \text{ bps} = 1 \text{ million}$).

We also calculate the VWAP tracking relative improvement with respect to a given benchmark model, as in Equation (38), and the VWAP tracking error is used as the error measurement. Figure 6 (a) and (b) show the dynamic and static VWAP tracking relative improvements of the robust Kalman filter, the standard Kalman filter and the CMEM model, when the RM is considered as the benchmark. Figure 6 (c) and (d) show the the dynamic VWAP

tracking improvements of the robust Kalman filter and the standard Kalman filter over the dynamic CMEM, and the static VWAP tracking over the static CMEM.

These results show that the robust Kalman filter with a dynamic VWAP strategy performs the best in the empirical studies. It gives an average VWAP tracking error of 6.38 bps, or equivalently, an improvement of 15% when compared with the RM, and of 9% when compared with the dynamic CMEM.

5. Conclusion

Intraday trading volume is an important variable in the modern financial market because many algorithmic trading strategies require an accurate forecast of it to improve their performance. This manuscript adopts the idea of decomposing volume into three components (i.e., daily, intraday periodic, and intraday dynamic components) and proposes a new methodology based on the Kalman filter to forecast intraday volume from a hidden state perspective. Since real-time intraday market data may contain outliers, we further extend the model by applying the Lasso regularization to derive a robust version with consistent performance under extremely noisy conditions. The model parameters can be effectively estimated by the EM algorithm

Table 4: The out-of-sample VWAP tracking error (in basis point) results. The best model for a given type of VWAP tracking is highlighted with boldface.

Ticker	Dynamic VWAP Tracking						Static VWAP Tracking							
	Robust Kalman Filter		Kalman Filter		CMEM		Robust Kalman Filter		Kalman Filter		CMEM		RM	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
AAPL	4.93	7.74	4.87	7.71	5.58	10.32	4.99	7.18	4.97	7.06	5.88	9.36	5.84	8.26
FB	5.58	5.57	5.52	5.35	6.50	8.17	6.16	6.81	6.17	6.65	7.06	8.14	6.96	8.15
C	5.28	7.42	5.27	7.54	6.24	9.80	5.59	8.31	5.59	8.36	6.77	10.53	6.02	9.77
IBM	4.02	7.16	4.06	7.79	4.65	8.60	4.16	6.72	4.23	6.92	4.81	7.91	4.77	7.76
KO	2.80	3.63	2.77	3.60	3.16	4.84	2.87	3.43	2.85	3.41	3.34	4.78	2.88	3.78
WMT	3.66	10.35	3.69	10.92	4.46	13.44	3.75	8.99	3.71	8.96	4.64	11.68	4.02	11.56
SPY	2.61	2.90	2.70	3.38	3.81	6.14	2.70	3.08	2.89	3.41	3.75	5.04	3.02	3.69
DIA	3.40	4.10	3.43	4.16	4.21	7.01	3.56	4.41	3.57	4.38	4.36	6.72	3.74	4.69
QQQ	3.81	4.51	3.82	4.97	5.07	7.75	4.02	4.40	3.78	4.87	4.78	6.72	4.49	5.68
IBB	11.88	18.25	11.89	18.59	13.24	23.59	12.48	17.37	12.55	17.50	15.29	23.34	14.58	20.12
XOP	9.74	11.47	9.54	11.30	9.30	11.84	10.62	11.43	10.59	11.40	10.72	11.00	9.96	11.03
OIH	8.45	8.40	8.49	8.61	8.46	9.43	8.99	8.93	9.09	8.90	9.62	10.34	9.07	9.88
ADS	7.87	14.58	7.88	14.35	8.48	17.06	8.88	15.48	8.90	15.45	10.01	17.13	10.88	17.43
PHIA	6.20	9.12	6.20	9.26	6.35	10.95	6.75	9.83	6.76	9.78	7.41	11.48	7.30	10.87
OR	6.49	12.78	6.50	12.81	7.15	15.48	6.78	11.11	6.81	11.29	7.55	12.97	7.31	12.49
BNP	7.51	11.88	7.51	11.70	7.82	14.04	8.25	12.41	8.24	12.53	8.43	14.52	8.78	15.03
ULVR	4.83	6.54	4.84	6.52	4.70	7.17	5.40	6.67	5.37	6.75	5.66	7.50	5.27	6.80
MKS	5.68	6.78	5.55	6.63	5.81	6.13	6.21	7.23	6.20	7.22	6.81	7.59	7.17	7.83
5HK	4.81	6.90	4.97	7.10	5.37	8.51	4.94	6.66	5.34	7.79	5.58	8.54	6.04	9.30
700HK	5.14	6.04	5.39	6.27	5.48	6.46	5.58	6.38	5.96	6.57	6.59	7.26	6.47	7.46
941HK	4.97	6.83	5.08	7.06	5.74	9.04	5.42	7.97	5.50	8.06	6.64	10.00	6.84	10.34
6758JP	6.42	6.28	6.64	6.88	8.24	10.52	6.59	6.73	6.64	6.63	8.78	11.03	8.08	8.83
7731JP	6.06	6.88	6.01	6.81	6.45	7.58	6.80	7.79	6.73	7.73	7.52	8.82	7.11	8.17
8604JP	7.92	12.07	7.98	12.26	9.16	14.72	7.94	10.83	8.05	10.87	9.42	13.92	8.67	13.28
2800HK	5.28	5.73	5.29	5.68	5.89	7.72	5.70	6.50	5.69	6.50	6.49	8.57	6.17	7.91
2823HK	10.18	14.70	10.38	15.25	10.42	19.04	11.85	16.75	12.08	17.62	12.13	21.23	12.97	22.64
1306JP	7.26	8.91	7.06	8.99	7.95	11.79	7.70	9.09	7.92	9.74	8.64	12.36	7.21	9.81
1321JP	8.28	11.98	8.22	11.44	9.14	13.58	8.77	10.60	8.78	10.61	9.90	13.76	9.47	13.80
1330JP	9.18	9.48	8.99	9.56	9.85	11.46	10.33	10.86	10.25	10.78	10.28	11.58	9.86	11.53
1570JP	11.11	16.93	10.91	16.72	11.62	20.57	11.65	15.02	11.47	14.96	12.34	20.81	13.43	22.56
Average	6.38	8.86	6.39	8.97	7.01	11.09	6.85	8.98	6.89	9.09	7.71	11.16	7.48	10.68

in closed-form. The proposed methodology significantly reduces the forecast complexity and works considerably better in minimizing the volume prediction error and the VWAP order execution risk. The empirical results are obtained through extensive experiments on the intraday volume data from twelve ETFs and eighteen stocks over the U.S., European, and Asian markets. The out-of-sample experiments illustrate that the proposed model substantially outperforms the benchmark RM and the state-of-the-art CMEM with an improvement of 64% and 29%, respectively, in volume prediction and of 15% and 9%, respectively, in VWAP trading replication.

Appendix A. Derivation of the EM Algorithm Closed-Form Solutions

This section provides a derivation of Algorithm 3. Let $\{y_\tau\}_{\tau=1}^N$ denote the training data set, consisting of N observed log-volume historical data, and let $\{\mathbf{x}_\tau\}_{\tau=1}^N$ denote

the corresponding unobservable states. We wish to estimate θ of the proposed Kalman filter model, defined in Equation (6), via the EM algorithm.

Appendix A.1. Jointly Likelihood Function

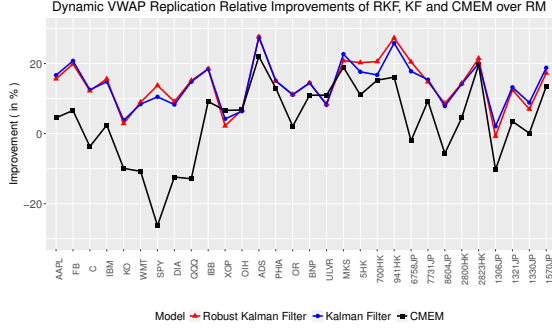
Recall that the proposed Kalman filter model is defined as

$$\mathbf{x}_{\tau+1} = \mathbf{A}_\tau \mathbf{x}_\tau + \mathbf{w}_\tau, \quad (\text{A.1})$$

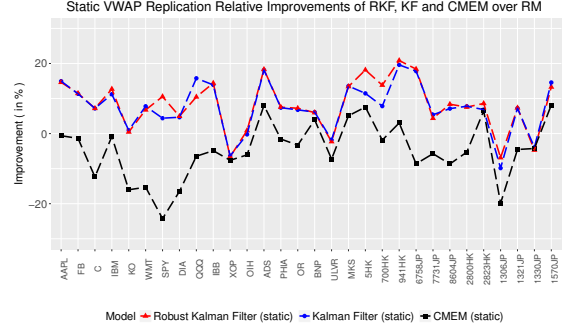
$$y_\tau = \mathbf{C} \mathbf{x}_\tau + \phi_\tau + v_\tau \quad (\text{A.2})$$

where the noise terms \mathbf{w}_τ and v_τ are assumed to be Gaussian. We can write the conditional probabilities for the hidden state and the observation as

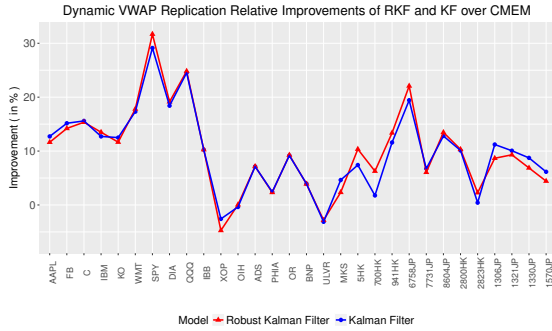
$$P(y_\tau | \mathbf{x}_\tau) = \frac{\exp\left(-\frac{(y_\tau - \phi_\tau - \mathbf{C} \mathbf{x}_\tau)^2}{2r}\right)}{\sqrt{2\pi r}}, \quad (\text{A.3})$$



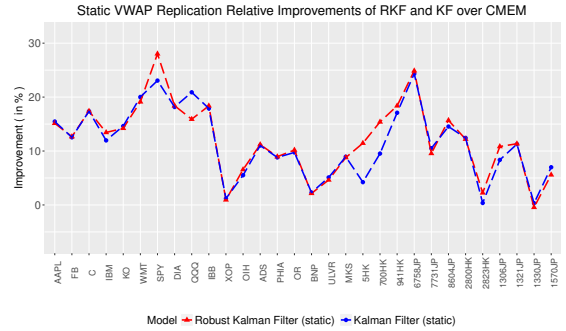
(a) Dynamic VWAP tracking relative improvements when the RM is considered as the benchmark.



(b) Static VWAP tracking relative improvements when the RM is considered as the benchmark.



(c) Dynamic VWAP tracking relative improvements when the dynamic CMEM is considered as the benchmark.



(d) Static VWAP tracking relative improvements when the static CMEM is considered as the benchmark.

Figure 6: The out-of-sample VWAP tracking relative improvements of the robust Kalman filter and the standard Kalman filter over the RM and the state-of-the-art CMEM.

$$P(\mathbf{x}_\tau | \mathbf{x}_{\tau-1}, \tau \neq kI + 1) = \frac{\exp\left(-\frac{(\mu_\tau - a^\mu \mu_{\tau-1})^2}{2(\sigma^\mu)^2}\right)}{\sqrt{2\pi}(\sigma^\mu)^2}, \quad (\text{A.4})$$

$$P(\mathbf{x}_\tau | \mathbf{x}_{\tau-1}, \tau = kI + 1) = \frac{\exp\left(\frac{(\eta_\tau - a^\eta \eta_{\tau-1})^2}{2(\sigma^\eta)^2} - \frac{(\mu_\tau - a^\mu \mu_{\tau-1})^2}{2(\sigma^\mu)^2}\right)}{\sqrt{2\pi}(\sigma^\eta)^2(\sigma^\mu)^2} \quad (\text{A.5})$$

where $k = 1, 2, \dots, N-1$. Assume the initial state is a Gaussian random variable with density

$$P(\mathbf{x}_1) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}_1 - \boldsymbol{\pi}_1)^\top \boldsymbol{\Sigma}_1^{-1}(\mathbf{x}_1 - \boldsymbol{\pi}_1)\right)}{\sqrt{2\pi}|\boldsymbol{\Sigma}_1|}. \quad (\text{A.6})$$

By the Markov property implicit in this model, the joint probability of the observed log-volume series $\{y_\tau\}_{\tau=1}^N$ and the hidden states $\{\mathbf{x}_\tau\}_{\tau=1}^N$ is

$$P(\{\mathbf{x}_\tau\}_{\tau=1}^N, \{y_\tau\}_{\tau=1}^N) = P(\mathbf{x}_1) \prod_{\tau=2}^N P(\mathbf{x}_\tau | \mathbf{x}_{\tau-1}) \prod_{\tau=1}^N P(y_\tau | \mathbf{x}_\tau). \quad (\text{A.7})$$

Thus, the joint log-likelihood function can be expressed as

$$\begin{aligned} \log P(\{\mathbf{x}_\tau\}_{\tau=1}^N, \{y_\tau\}_{\tau=1}^N) &= -\sum_{\tau=1}^N \left(\frac{(y_\tau - \phi_\tau - \mathbf{C}\mathbf{x}_\tau)^2}{2r} \right) - \frac{N}{2} \log(r) \\ &\quad - \sum_{\tau=2}^N \left(\frac{(\mu_\tau - a^\mu \mu_{\tau-1})^2}{2(\sigma^\mu)^2} \right) - \frac{N-1}{2} \log(\sigma^\mu)^2 \\ &\quad - \sum_{\tau=kI+1}^N \left(\frac{(\eta_\tau - a^\eta \eta_{\tau-1})^2}{2(\sigma^\eta)^2} \right) - \frac{T-1}{2} \log(\sigma^\eta)^2 \\ &\quad - \frac{1}{2}(\mathbf{x}_1 - \boldsymbol{\pi}_1)^\top \boldsymbol{\Sigma}_1^{-1}(\mathbf{x}_1 - \boldsymbol{\pi}_1) - \frac{1}{2} \log|\boldsymbol{\Sigma}_1| \\ &\quad - \frac{2N+T}{2} \log(2\pi). \end{aligned} \quad (\text{A.8})$$

Explicitly maximizing Equation (A.8) is difficult. We instead repeatedly construct a lower-bound on the log-likelihood function using Jensen's inequality (E-step) and optimize that lower-bound (M-step).

Appendix A.2. The E-step

The E-step calculates the expected value of the log-likelihood function with respect to the conditional distri-

bution of $\{\mathbf{x}_\tau\}_{\tau=1}^N$ given $\{y_\tau\}_{\tau=1}^N$ under the current estimate of the parameters $\hat{\boldsymbol{\theta}}^{(j)}$:

$$\mathcal{Q}(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(j)}) = E_{\{\mathbf{x}\}|\{y\}, \hat{\boldsymbol{\theta}}^{(j)}} \left[\log P \left(\{\mathbf{x}_\tau\}_{\tau=1}^N, \{y_\tau\}_{\tau=1}^N \right) \right]. \quad (\text{A.9})$$

We calculate \mathcal{Q} part by part:

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(j)}) = & -E_1 - E_2 - E_3 - E_4 \\ & - \frac{N}{2} \log(r) - \frac{N-1}{2} \log(\sigma^\mu)^2 \\ & - \frac{T-1}{2} \log(\sigma^\eta)^2 - \frac{1}{2} \log |\boldsymbol{\Sigma}_1| \\ & - \frac{2N+T}{2} \log(2\pi), \end{aligned} \quad (\text{A.10})$$

where

$$\begin{aligned} E_1 = & E \left\{ \sum_{\tau=1}^N \left[\frac{(y_\tau - \phi_\tau - \mathbf{C}\mathbf{x}_\tau)^2}{2r} \right] | \{y_\tau\}_{\tau=1}^N, \hat{\boldsymbol{\theta}}^{(j)} \right\} \\ = & \frac{1}{2r} \sum_{\tau=1}^N \left[y_\tau^2 + \mathbf{C}\mathbf{P}_\tau \mathbf{C}^\top - 2y_\tau \mathbf{C}\hat{\mathbf{x}}_\tau + \phi_\tau^2 \right. \\ & \left. - 2y_\tau \phi_\tau + 2\phi_\tau \mathbf{C}\hat{\mathbf{x}}_\tau \right], \end{aligned} \quad (\text{A.11})$$

$$\begin{aligned} E_2 = & E \left\{ \sum_{\tau=2}^N \left[\frac{1}{2(\sigma^\mu)^2} (\mu_\tau - a^\mu \mu_{\tau-1})^2 \right] | \{y_\tau\}_{\tau=1}^N, \hat{\boldsymbol{\theta}}^{(j)} \right\} \\ = & \sum_{\tau=2}^N \frac{1}{2(\sigma^\mu)^2} \left[\mathbf{P}_\tau^{(2,2)} + (a^\mu)^2 \mathbf{P}_{\tau-1}^{(2,2)} - 2a^\mu \mathbf{P}_{\tau,\tau-1}^{(2,2)} \right], \end{aligned} \quad (\text{A.12})$$

$$\begin{aligned} E_3 = & E \left\{ \sum_{\tau=kI+1}^N \left[\frac{1}{2(\sigma^\eta)^2} (\eta_\tau - a^\eta \eta_{\tau-1})^2 \right] | \{y_\tau\}_{\tau=1}^N, \hat{\boldsymbol{\theta}}^{(j)} \right\} \\ = & \sum_{\tau=kI+1}^N \frac{1}{2(\sigma^\eta)^2} \left[\mathbf{P}_\tau^{(1,1)} + (a^\eta)^2 \mathbf{P}_{\tau-1}^{(1,1)} - 2a^\eta \mathbf{P}_{\tau,\tau-1}^{(1,1)} \right], \end{aligned} \quad (\text{A.13})$$

$$\begin{aligned} E_4 = & E \left\{ \frac{1}{2} (\mathbf{x}_1 - \boldsymbol{\pi}_1)^\top \boldsymbol{\Sigma}_1^{-1} (\mathbf{x}_1 - \boldsymbol{\pi}_1) | \{y_\tau\}_{\tau=1}^N, \hat{\boldsymbol{\theta}}^{(j)} \right\} \\ = & \frac{1}{2} \left[\text{tr}(\boldsymbol{\Sigma}_1^{-1} \mathbf{P}_1) - 2\boldsymbol{\pi}_1^\top \boldsymbol{\Sigma}_1^{-1} \hat{\mathbf{x}}_1 + \boldsymbol{\pi}_1^\top \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\pi}_1 \right], \end{aligned} \quad (\text{A.14})$$

and

$$\hat{\mathbf{x}}_\tau = E \left[\mathbf{x}_\tau | \{y_\tau\}_{\tau=1}^N, \hat{\boldsymbol{\theta}}^{(j)} \right], \quad (\text{A.15})$$

$$\mathbf{P}_\tau = E \left[\mathbf{x}_\tau \mathbf{x}_\tau^\top | \{y_\tau\}_{\tau=1}^N, \hat{\boldsymbol{\theta}}^{(j)} \right], \quad (\text{A.16})$$

$$\mathbf{P}_{\tau,\tau-1} = E \left[\mathbf{x}_\tau \mathbf{x}_{\tau-1}^\top | \{y_\tau\}_{\tau=1}^N, \hat{\boldsymbol{\theta}}^{(j)} \right], \quad (\text{A.17})$$

are three sufficient statistics of \mathcal{Q} . Fortunately, the first two are the quantities that can be directly calculated through the Kalman filtering and smoothing described in Algo-

rithms 1 and 2:

$$\hat{\mathbf{x}}_\tau = \hat{\mathbf{x}}_{\tau|N}, \quad (\text{A.18})$$

$$\mathbf{P}_\tau = \boldsymbol{\Sigma}_{\tau|N} + \hat{\mathbf{x}}_{\tau|N} \hat{\mathbf{x}}_{\tau|N}^\top. \quad (\text{A.19})$$

The third quantity $\mathbf{P}_{\tau,\tau-1}$ can be obtained by defining

$$\begin{aligned} \boldsymbol{\Sigma}_{\tau,\tau-1|N} = & \boldsymbol{\Sigma}_{\tau|\tau} \mathbf{L}_{\tau-1}^\top \\ & + \mathbf{L}_\tau (\boldsymbol{\Sigma}_{\tau+1,\tau|N} - \mathbf{A}_\tau \boldsymbol{\Sigma}_{\tau|\tau}) \mathbf{L}_{\tau-1}^\top, \end{aligned} \quad (\text{A.20})$$

which is initialized by

$$\boldsymbol{\Sigma}_{N,N-1|N} = (\mathbf{I} - \mathbf{K}_N \mathbf{C}) \mathbf{A}_{N-1} \boldsymbol{\Sigma}_{N-1|N-1}. \quad (\text{A.21})$$

Thus, the last sufficient statistic is

$$\mathbf{P}_{\tau,\tau-1} = \boldsymbol{\Sigma}_{\tau,\tau-1|N} + \hat{\mathbf{x}}_{\tau|N} \hat{\mathbf{x}}_{\tau-1|N}^\top. \quad (\text{A.22})$$

Appendix A.3. The M-Step

The M-step finds the parameters that maximize Equation (A.10):

$$\hat{\boldsymbol{\theta}}^{(j+1)} = \arg \max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(j)}), \quad (\text{A.23})$$

where $\boldsymbol{\theta}$ is the parameter vector of the proposed model, which includes all the unknown parameters $\boldsymbol{\pi}_1$, $\boldsymbol{\Sigma}_1$, \mathbf{A}_τ , \mathbf{Q}_τ , r and the seasonality ϕ_τ . Each of these is estimated by taking the corresponding partial derivative of the expected log-likelihood and setting the derivative to zero.

- Initial state mean:

$$\frac{\partial \mathcal{Q}}{\partial \boldsymbol{\pi}_1} = \frac{\partial (-E_4)}{\partial \boldsymbol{\pi}_1} = \boldsymbol{\Sigma}_1^{-1} (\hat{\mathbf{x}}_1 - \boldsymbol{\pi}_1) = \mathbf{0}; \quad (\text{A.24})$$

- Initial state covariance:

$$\begin{aligned} \frac{\partial \mathcal{Q}}{\partial \boldsymbol{\Sigma}_1^{-1}} = & \frac{\partial (-E_4 - \frac{1}{2} \log |\boldsymbol{\Sigma}_1|)}{\partial \boldsymbol{\Sigma}_1^{-1}} \\ = & \frac{1}{2} \boldsymbol{\Sigma}_1 - \frac{1}{2} (\mathbf{P}_1 - \hat{\mathbf{x}}_1 \boldsymbol{\pi}_1^\top - \boldsymbol{\pi}_1 \hat{\mathbf{x}}_1^\top + \boldsymbol{\pi}_1 \boldsymbol{\pi}_1^\top) \\ = & \mathbf{0}; \end{aligned} \quad (\text{A.25})$$

- State dynamic matrix $\mathbf{A}_\tau = \begin{bmatrix} a_\tau^\eta & 0 \\ 0 & a^\mu \end{bmatrix}$ with

$$a_\tau^\eta = \begin{cases} a^\eta & \tau = kI, k = 1, 2, \dots \\ 1 & \text{otherwise} \end{cases}$$

$$\begin{aligned} \frac{\partial \mathcal{Q}}{\partial a^\eta} = & \frac{\partial (-E_3)}{\partial a^\eta} \\ = & - \sum_{\tau \in D} \frac{1}{(\sigma^\eta)^2} \mathbf{P}_{\tau,\tau-1}^{(1,1)} + \sum_{\tau \in D} \frac{1}{(\sigma^\eta)^2} a^\eta \mathbf{P}_{\tau-1}^{(1,1)} \\ = & \mathbf{0}, \end{aligned} \quad (\text{A.26})$$

$$\frac{\partial \mathcal{Q}}{\partial a^\mu} = \frac{\partial (-E_2)}{\partial a^\mu}$$

$$\begin{aligned}
&= - \sum_{\tau=2}^N \frac{1}{(\sigma^\mu)^2} \mathbf{P}_{\tau, \tau-1}^{(2,2)} + \sum_{\tau=2}^N \frac{1}{(\sigma^\mu)^2} a^\mu \mathbf{P}_{\tau-1}^{(2,2)} \\
&= \mathbf{0}; \tag{A.27}
\end{aligned}$$

- State noise covariance $\mathbf{Q}_\tau = \begin{bmatrix} (\sigma^\eta)^2 & 0 \\ 0 & (\sigma^\mu)^2 \end{bmatrix}$ with
$$(\sigma^\eta)^2 = \begin{cases} (\sigma^\eta)^2 & \tau = kI, k = 1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned}
\frac{\partial \mathcal{Q}}{\partial [(\sigma^\eta)^2]^{-1}} &= \frac{\partial \{-E_3 - \frac{T-1}{2} [\log (\sigma^\eta)^2]\}}{\partial [(\sigma^\eta)^2]^{-1}} \\
&= \frac{T-1}{2} (\sigma^\eta)^2 - \frac{1}{2} \sum_{\tau \in D} \left[\mathbf{P}_\tau^{(1,1)} \right. \\
&\quad \left. + (a^\eta)^2 \mathbf{P}_{\tau-1}^{(1,1)} - 2a^\eta \mathbf{P}_{\tau, \tau-1}^{(1,1)} \right] \\
&= \mathbf{0}, \tag{A.28}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathcal{Q}}{\partial [(\sigma^\mu)^2]^{-1}} &= \frac{\partial \{-E_2 - \frac{N-1}{2} \log [(\sigma^\mu)^2]\}}{\partial [(\sigma^\mu)^2]^{-1}} \\
&= \frac{N-1}{2} (\sigma^\mu)^2 - \frac{1}{2} \sum_{\tau=2}^N \left[\mathbf{P}_\tau^{(2,2)} \right. \\
&\quad \left. + (a^\mu)^2 \mathbf{P}_{\tau-1}^{(2,2)} - 2a^\mu \mathbf{P}_{\tau, \tau-1}^{(2,2)} \right] \\
&= \mathbf{0}; \tag{A.29}
\end{aligned}$$

- Observation noise covariance r :

$$\begin{aligned}
\frac{\partial \mathcal{Q}}{\partial r^{-1}} &= \frac{\partial [-E_1 - \frac{N}{2} \log (r)]}{\partial r^{-1}} \\
&= \frac{N}{2} r - \frac{1}{2} \sum_{\tau=1}^N \left(y_\tau^2 + \mathbf{C} \mathbf{P}_\tau \mathbf{C}^\top - 2y_\tau \mathbf{C} \hat{\mathbf{x}}_\tau \right. \\
&\quad \left. + \phi_\tau^2 - 2y_\tau \phi_\tau + 2\phi_\tau \mathbf{C} \hat{\mathbf{x}}_\tau \right) \\
&= \mathbf{0}; \tag{A.30}
\end{aligned}$$

- Intraday periodic component ϕ_τ , or equivalently ϕ_i :

$$\begin{aligned}
\frac{\partial \mathcal{Q}}{\partial \phi_\tau} &= \frac{\partial (-E_1)}{\partial \phi_\tau} = \frac{1}{r} \sum_{\tau=1}^N \left(\phi_\tau - y_\tau + \mathbf{C} \hat{\mathbf{x}}_\tau \right) \\
&= \frac{1}{r} \left\{ T \sum_{i=1}^I \phi_i - \sum_{i=1}^I \sum_{t=1}^T (y_{t,i} - \mathbf{C} \hat{\mathbf{x}}_{t,i}) \right\} \\
&= \mathbf{0}. \tag{A.31}
\end{aligned}$$

Solving Equations (A.24)-(A.31) results in the optimal closed-form updates of $\hat{\boldsymbol{\theta}}^{(j+1)}$:

$$\boldsymbol{\pi}_1^{(j+1)} = \hat{\mathbf{x}}_1, \tag{A.32}$$

$$\boldsymbol{\Sigma}_1^{(j+1)} = \mathbf{P}_1 - \hat{\mathbf{x}}_1 \hat{\mathbf{x}}_1^\top, \tag{A.33}$$

$$(a^\eta)^{(j+1)} = \left[\sum_{\tau=kI+1} \mathbf{P}_{\tau, \tau-1}^{(1,1)} \right] \left[\sum_{\tau=kI+1} \mathbf{P}_{\tau-1}^{(1,1)} \right]^{-1}, \tag{A.34}$$

$$(a^\mu)^{(j+1)} = \left[\sum_{\tau=2}^N \mathbf{P}_{\tau, \tau-1}^{(2,2)} \right] \left[\sum_{\tau=2}^N \mathbf{P}_{\tau-1}^{(2,2)} \right]^{-1}, \tag{A.35}$$

$$\begin{aligned}
[(\sigma^\eta)^2]^{(j+1)} &= \frac{1}{T-1} \sum_{\tau=kI+1} \left\{ \mathbf{P}_\tau^{(1,1)} + [(a^\eta)^{(j+1)}]^2 \mathbf{P}_{\tau-1}^{(1,1)} \right. \\
&\quad \left. - 2(a^\eta)^{(j+1)} \mathbf{P}_{\tau, \tau-1}^{(1,1)} \right\}, \tag{A.36}
\end{aligned}$$

$$\begin{aligned}
[(\sigma^\mu)^2]^{(j+1)} &= \frac{1}{N-1} \sum_{\tau=2}^N \left\{ \mathbf{P}_\tau^{(2,2)} + [(a^\mu)^{(j+1)}]^2 \mathbf{P}_{\tau-1}^{(2,2)} \right. \\
&\quad \left. - 2(a^\mu)^{(j+1)} \mathbf{P}_{\tau, \tau-1}^{(2,2)} \right\}, \tag{A.37}
\end{aligned}$$

$$\begin{aligned}
r^{(j+1)} &= \frac{1}{N} \sum_{\tau=1}^N \left[y_\tau^2 + \mathbf{C} \mathbf{P}_\tau \mathbf{C}^\top - 2y_\tau \mathbf{C} \hat{\mathbf{x}}_\tau + (\phi_\tau^{(j+1)})^2 \right. \\
&\quad \left. - 2y_\tau \phi_\tau^{(j+1)} + 2\phi_\tau^{(j+1)} \mathbf{C} \hat{\mathbf{x}}_\tau \right], \tag{A.38}
\end{aligned}$$

$$\phi_i^{(j+1)} = \frac{1}{T} \sum_{t=1}^T (y_{t,i} - \mathbf{C} \hat{\mathbf{x}}_{t,i}). \tag{A.39}$$

References

- Admati, A.R., Pfleiderer, P., 1988. A theory of intraday patterns: volume and price variability. *Review of Financial studies* 1, 3–40.
- Ajinkya, B.B., Jain, P.C., 1989. The behavior of daily stock market trading volume. *Journal of accounting and economics* 11, 331–359.
- Almgren, R., Thum, C., Hauptmann, E., Li, H., 2005. Direct estimation of equity market impact. *Risk* 18, 58–62.
- Andersen, T.G., 1996. Return volatility and trading volume: an information flow interpretation of stochastic volatility. *The Journal of Finance* 51, 169–204.
- Bertsimas, D., Lo, A.W., 1998. Optimal control of execution costs. *Journal of Financial Markets* 1, 1–50.
- Białkowski, J., Darolles, S., Le Fol, G., 2008. Improving vwap strategies: a dynamic volume approach. *Journal of Banking and Finance* 32, 1709–1722.
- Brock, W.A., Kleidon, A.W., 1992. Periodic market closure and trading volume: a model of intraday bids and asks. *Journal of Economic Dynamics and Control* 16, 451–489.
- Brownlees, C.T., Cipollini, F., Gallo, G.M., 2011. Intra-daily volume modeling and prediction for algorithmic trading. *Journal of Financial Econometrics* 9, 489–518.
- Cai, C.X., Hudson, R., Keasey, K., 2004. Intra day bid-ask spreads, trading volume and volatility: recent empirical evidence from the london stock exchange. *Journal of Business Finance & Accounting* 31, 647–676.
- Campbell, J.Y., Grossman, S.J., Wang, J., 1992. Trading volume and serial correlation in stock returns. Technical Report. National Bureau of Economic Research.
- Chevallier, J., Sévi, B., 2012. On the volatility–volume relationship in energy futures markets using intraday data. *Energy Economics* 34, 1896–1909.
- Darrat, A.F., Rahman, S., Zhong, M., 2003. Intraday trading volume and return volatility of the djia stocks: a note. *Journal of Banking and Finance* 27, 2035–2043.
- Easley, D., López de Prado, M.M., O'Hara, M., 2012. Flow toxicity and liquidity in a high-frequency world. *The Review of Financial Studies* 25, 1457–1493.

- Fabozzi, F.J., Focardi, S.M., Kolm, P.N., 2010. Quantitative equity investing: techniques and strategies. John Wiley and Sons.
- Feng, Y., Palomar, D.P., Rubio, F., 2015. Robust optimization of order execution. *IEEE Transactions on Signal Processing* 63, 907–920.
- Gerety, M.S., Mulherin, J.H., 1992. Trading halts and market activity: an analysis of volume at the open and the close. *The Journal of Finance* 47, 1765–1784.
- Gwilym, O.A., McMillan, D., Speight, A., 1999. The intraday relationship between volume and volatility in liffe futures markets. *Applied Financial Economics* 9, 593–604.
- Huberman, G., Stanzl, W., 2005. Optimal liquidity trading. *Review of Finance* 9, 165–200.
- Hussain, S.M., 2011. The intraday behaviour of bid-ask spreads, trading volume and return volatility: evidence from dax30. *International Journal of Economics and Finance* 3, 23.
- Kalman, R.E., et al., 1960. A new approach to linear filtering and prediction problems. *Journal of basic Engineering* 82, 35–45.
- Kluger, B.D., McBride, M.E., 2011. Intraday trading patterns in an intelligent autonomous agent-based stock market. *Journal of Economic Behavior & Organization* 79, 226–245.
- Lee, B.S., Rui, O.M., 2002. The dynamic relationship between stock returns and trading volume: domestic and cross-country evidence. *Journal of Banking and Finance* 26, 51–78.
- Lee, C., Ready, M.J., Seguin, P.J., 1994. Volume, volatility, and new york stock exchange trading halts. *The Journal of Finance* 49, 183–214.
- Lo, A.W., Wang, J., 2000. Trading volume: definitions, data analysis, and implications of portfolio theory. *Review of Financial Studies* 13, 257–300.
- Malinova, K., Park, A., 2014. The impact of competition and information on intraday trading. *Journal of Banking & Finance* 44, 55–71.
- Mattingley, J., Boyd, S., 2010. Real-time convex optimization in signal processing. *IEEE Signal processing magazine* 27, 50–61.
- Pagano, M., 1989. Trading volume and asset liquidity. *The Quarterly Journal of Economics* 104, 255–274.
- Satish, V., Saxena, A., Palmer, M., 2014. Predicting intraday trading volume and volume percentages. *The Journal of Trading* 9, 15–25.
- Shumway, R.H., Stoffer, D.S., 1982. An approach to time series smoothing and forecasting using the em algorithm. *Journal of time series analysis* 3, 253–264.
- Smirlock, M., Starks, L., 1988. An empirical analysis of the stock price-volume relationship. *Journal of Banking & Finance* 12, 31–41.
- Stephan, J.A., Whaley, R.E., 1990. Intraday price change and trading volume relations in the stock and stock option markets. *The Journal of Finance* 45, 191–220.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58, 267–288.