A Multi-step Ahead Dyadic Particle Filter for Price Prediction

Myrsini Ntemi and Constantine Kotropoulos
Department of Informatics, Aristotle University of Thessaloniki
Thessaloniki 54124, GREECE
Email: {nmyrsini,costas}@csd.auth.gr

***Abstract*—A very challenging task in financial forecasting is the accurate prediction at multiple future time steps based on historical prices. In this paper, a multi-step ahead dyadic particle filter is proposed for stock price prediction. A dyad corresponds to a latent vector modeling each stock and a latent vector modeling the group of companies in the same category. The stock latent vectors evolve through time according to a geometric Brownian drift parameter, which enables stock latent vectors to capture the erratic price evolution appearing in future time steps. The multi-step ahead framework of the dyadic particle filter and the drift parameter lead to more accurate predictions compared to state-of-the-art techniques.**

***Index Terms*—multi-step ahead prediction, price prediction, particle filtering, Brownian drift**

I. Introduction

Multi-step ahead forecasting is the prediction of a sequence of future time steps [1]. Since financial time series exhibit chaotic patterns, accurate multi-step ahead price prediction is very difficult given only past prices. Contrary to single-step forecasting, multi-step ahead prediction can optimize investment strategies.

A variety of works in multi-step ahead prediction has emerged the last years. Some of them apply co-evolutionary multi-task learning [2]. Recently, Google Cloud AI Research developed a Temporal Fusion Transformer to forecast multi-horizon time series [3]. LSTNet and DeepAR have been successfully applied to multi-step prediction [4] [5]. However, deep neural networks (DNNs) require frequent recalibration and usually suffer from computational infeasibility [6]. Several DNNs require a plethora of heterogeneous data sources, i.e., static metadata, upcoming important dates, historical traffic, etc. Other, assume that some of this heterogeneous information is static and known in the future [5] [7] [8].

Multi-stage prediction methods rely only on historical values and predictions, but they propagate the errors into future predictions [1]. As the length of the prediction steps increases, the prediction error becomes larger. Usually, the diffusion modeling relies on the volatility of the exact previous time step, following a Bayesian context [9] [10].

A multi-step ahead dyadic particle filter (multiT-DPF) is proposed for stock price prediction, extending the dyadic particle filter (DPF) [11] where only the next day opening price is predicted based on the opening price of the current day. Contrary to DPF, multiT-DPF is an online method, which utilizes 10 historical prices and predicts the next 10 ones. A dyad consists of a stock latent vector and a market segment latent vector, modeling the price evolution and market segment evolution, respectively. Here, the stock latent vectors are not static as in [11], but they evolve dynamically through time according to a geometric Brownian motion log drift parameter. The log drift parameter captures volatility relies on the drifts appeared 10 time steps back in time. Hence, the modeling is not equivalent to merely attaching noise to a random variable to ensure price volatility tracking. A Taylor expansion around the previously observed drift parameter approximates the negative log-likelihood of the posterior distribution. The first and second-order derivatives of the negative log-likelihood with respect to the previously observed drift parameter are approximated by novel analytical expressions. The prior covariance matrices of the stock latent vectors are formed by attaching the aforementioned volatility information to the main diagonal of the posterior covariance matrices at the previous time steps. A particle coefficient vector is computed through the product of the diagonal elements of the prior covariance matrices of the latent vectors, which form a dyad. Efficient particles are drawn according to this particle coefficient vector. A small number of particles is shown to be sufficient keeping the computational complexity low. Experiments have demonstrated that multiT-DPF outperforms the state-of-the-art methods in [3], [4], and [5].

To sum up, the novel contribution of the paper is in the proposal and thorough assessment of a multi-step framework for a dyadic particle filter, which ensures computational feasibility since it relies exclusively on the 10 past prices to predict the 10 next ones. Novel approximations of the Brownian log drift parameter are tested which endow multiT-DPF with robust tracking abilities, avoiding thus error propagation.

The multi-step ahead framework of the dyadic particle filter is presented in Section II. Experimental results are demonstrated in Section III, and conclusions are drawn in Section IV.

Acknowledgement: This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH - CREATE - INNOVATE (project code: T1EDK-02474).

II. MULTI-STEP AHEAD DPF

A. Latent state structure

Kalman filtering is applied to construct the latent state of multiT-DPF through a probabilistic approach. Let $\mathfrak{t}$ be a time instant and $t = \mathrm{mod}(\mathfrak{t}, d) + 1$ be a time step in the event $k = 0, 1, \ldots, [\frac{N}{d}] - 1$, where N denotes the total number of prices. At every event k, d prices are given as input to multiT-DPF and d predictions are provided in the output. At the current event k, the latent state is constructed by d dyads i, j formed by a stock latent vector $\mathbf{s}_i^{(k)}[t] \in \mathbb{R}^n$ and a segment latent vector $\mathbf{m}_j^{(k)}[t] \in \mathbb{R}^n$. The state model of multiT-DPF is formed by the multivariate Gaussian prior distributions of these latent vectors [12]

$$\begin{aligned} \mathbf{s}_i^{(k)}[t] &\sim \mathcal{N}\big(\boldsymbol{\mu}_{\mathbf{s}_i}^{(k)}[t], \boldsymbol{\Sigma}_{\mathbf{s}_i}^{(k)}[t]\big) \\ \mathbf{m}_j^{(k)}[t] &\sim \mathcal{N}\big(\boldsymbol{\mu}_{\mathbf{m}_j}^{(k)}[t], \boldsymbol{\Sigma}_{\mathbf{m}_j}^{(k)}[t]\big) \end{aligned} \tag{1}$$

where $\boldsymbol{\mu}_{\mathbf{s}_i}^{(k)}, \boldsymbol{\mu}_{\mathbf{m}_j}^{(k)} \in \mathbb{R}^n$ are the prior mean vectors and $\boldsymbol{\Sigma}_{\mathbf{s}_i}^{(k)}, \boldsymbol{\Sigma}_{\mathbf{m}_j}^{(k)} \in \mathbb{R}^{n \times n}$ are the prior covariance matrices. Every prior mean vector at event k is the respective posterior one at the previous event $k-1$, i.e., $\boldsymbol{\mu}_{\mathbf{s}_i}^{k}[t] = \boldsymbol{\mu}_{\mathbf{s}_i}^{\prime(k-1)}[t]$. Here, $\boldsymbol{\Sigma}_{\mathbf{s}_i}^{(k)}[t]$ is not considered fixed as in [11], but it evolves according to a geometric Brownian motion

$$\boldsymbol{\Sigma}_{\mathbf{s}_i}^{(k)}[t] = \boldsymbol{\Sigma}_{\mathbf{s}_i}^{\prime(k-1)}[t] + \big(d\, \alpha_{\mathbf{s}_i}^{(k)}[t]\big)\mathbf{I} \tag{2}$$

where $\boldsymbol{\Sigma}^{\prime(k-1)}[t]$ is the posterior covariance matrix at event $k-1$, $\mathbf{I} \in \mathbb{R}^{n \times n}$ is the identity matrix, and $\alpha_{\mathbf{s}_i}^{(k)}[t]$ is the drift value of the Brownian motion of $\mathbf{s}_i$ at event k. Equivalent expressions hold for $\boldsymbol{\mu}_{\mathbf{m}_j}^{(k)}$. At the current event k, the posterior probability density functions (pdfs) of the latent vectors are also multivariate Gaussians, i.e., $\mathbf{s}_i^{\prime(k)}[t] \sim \mathcal{N}\big(\boldsymbol{\mu}_{\mathbf{s}_i}^{\prime(k)}[t], \boldsymbol{\Sigma}_{\mathbf{s}_i}^{\prime(k)}[t]\big)$. The respective expression holds for $\mathbf{m}_j^{\prime(k)}[t]$. After having observed the past d prices $y^{(k-1)}[t], t = 1, ..., d$ at $k-1$, variational inference is applied to approximate the true joint posterior distribution $p\big(\mathbf{s}_i^{(k)}[t], \mathbf{m}_j^{(k)}[t]\big)$. This is accomplished by a factorized distribution, i.e., $q\big(\mathbf{s}_i^{(k)}[t], \mathbf{m}_j^{(k)}[t]\big) \approx q\big(\mathbf{s}_i^{(k)}[t]\big)\, q\big(\mathbf{m}_j^{(k)}[t]\big)$, where $q(\cdot)$ is an approximate distribution [13]. The objective is to minimize the Kullback-Leibler (KL) divergence $KL(q||p) = \mathbb{E}_q\big[\log \frac{q}{p}\big]$, where $p(\cdot)$ is the true posterior distribution and $\mathbb{E}_q[\cdot]$ is the expectation with respect to $q(\cdot)$. Then, a coordinate ascent update is applied to obtain the optimal parameters of $q(\cdot)$ which are [12] [14] [15]:

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{s}_i}^{\prime(k)}[t] &= \left(\big[\boldsymbol{\Sigma}_{\mathbf{s}_i}^{(k)}\big]^{-1}[t] + \frac{\boldsymbol{\mu}_{\mathbf{m}_j}^{\prime(k)}[t] \big[\boldsymbol{\mu}_{\mathbf{m}_j}^{\prime(k)}\big]^T[t] + \boldsymbol{\Sigma}_{\mathbf{m}_j}^{\prime(k)}[t]}{\sigma_{ij}^2} \right)^{-1} \\ \boldsymbol{\mu}_{\mathbf{s}_i}^{\prime(k)}[t] &= \boldsymbol{\Sigma}_{\mathbf{s}_i}^{\prime(k)}[t] \left(\frac{y^{(k-1)}[t] \boldsymbol{\mu}_{\mathbf{m}_j}^{\prime(k)}[t]}{\sigma_{ij}^2} + \big[\boldsymbol{\Sigma}_{\mathbf{s}_i}^{(k)}\big]^{-1}[t] \boldsymbol{\mu}_{\mathbf{s}_i}^{(k)}[t] \right) \end{aligned} \tag{3}$$

where $y^{(k-1)}[t]$ is the price of the previous event $k-1$ at the respective time step t and σ_{ij}^2 is the variance. Similar updates are considered for $\boldsymbol{\Sigma}_{\mathbf{m}_j}^{\prime(k)}, \boldsymbol{\mu}_{\mathbf{m}_j}^{\prime(k)}[t]$.

B. Volatility

In Eq. (2), $\alpha_{\mathbf{s}_i}^{(k)}[t]$ defines the volatility of the prior latent vector $\boldsymbol{\mu}_{\mathbf{s}_i}^{(k)}[t]$, where $\alpha_{\mathbf{s}_i}^{(k)}[t] = e^{a_{\mathbf{s}_i}^{(k)}[t]}$ is a geometric Brownian motion. Let $a_{\mathbf{s}_i}^{(k)}[t] \sim \mathcal{N}\big(a_{\mathbf{s}_i}^{(k-1)}[t], \gamma\ d\big)$ be the Brownian log drift value, where γ is an additional drift parameter [12]. A second-order Taylor expansion about its value at the previous event $k-1$ at $t = 1, ..., d$, $a_{\mathbf{s}_i}^{(k-1)}[t]$ is applied to compute this log drift value as [12]

$$\begin{aligned} f(a_{\mathbf{s}_i}^{(k)}[t]) \approx\ & f(a_{\mathbf{s}_i}^{(k-1)}[t]) + \big(a_{\mathbf{s}_i}^{(k)}[t] - a_{\mathbf{s}_i}^{(k-1)}[t]\big) \cdot \\ & \dot{f}(a_{\mathbf{s}_i}^{(k-1)}[t]) + \frac{1}{2}\big(a_{\mathbf{s}_i}^{(k)}[t] - a_{\mathbf{s}_i}^{(k-1)}[t]\big)^2 \ddot{f}(a_{\mathbf{s}_i}^{(k-1)}[t]) \end{aligned} \tag{4}$$

where $f(\cdot) = -\ln p(\mathbf{s}_i^{(k)}[t], a_{\mathbf{s}_i}^{(k)}[t])$ denotes the negative log-likelihood, and $\dot{f}()$, $\ddot{f}()$ are its first-order and second-order derivative w.r.t. $a_{\mathbf{s}_i}^{(k-1)}[t]$, respectively. $f(\cdot)$ should be minimized w.r.t. $a_{\mathbf{s}_i}^{(k)}[t]$ The optimal $a_{\mathbf{s}_i}^{(k)}[t]$ is [12]

$$a_{\mathbf{s}_i}^{(k)}[t] = a_{\mathbf{s}_i}^{(k-1)}[t] - \frac{\dot{f}(a_{\mathbf{s}_i}^{(k-1)}[t])}{\ddot{f}(a_{\mathbf{s}_i}^{(k-1)}[t])} \tag{5}$$

If $\boldsymbol{\Sigma}_{\mathbf{s}_i}^{\prime(k-1)}[t] = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}^T$, with $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_n)$ denoting the diagonal matrix of the eigenvalues of the prior covariance matrix at $k-1$ and $\mathbf{H} \in \mathbb{R}^{n \times n}$ denoting the matrix whose columns are the associated eigenvectors. Let $\mathbf{M} = \mathbf{H}^T \boldsymbol{\Sigma}_{\mathbf{s}_i}^{\prime(k)}[t]\mathbf{H}$. Let also $\tilde{\boldsymbol{\Lambda}} = \mathrm{diag}(\tilde{\lambda}_n)$ with $\tilde{\lambda}_n = (\lambda_n + e^{a_{\mathbf{s}_i}^{(k)}[t]} d)$, b_n be the nth element of $\mathbf{b} = (\mathbf{M}\tilde{\boldsymbol{\Lambda}}^{-1} - \mathbf{I})\mathbf{H}^T \boldsymbol{\mu}_{\mathbf{s}_i}^{\prime(k-1)}[t]$, and $\varphi_n = \frac{e^{a_{\mathbf{s}_i}^{(k-1)}[t]} d}{\lambda_n + e^{a_{\mathbf{s}_i}^{(k-1)}[t]} d}$. Eqs. (6) and (7) correspond to the proposed approximation of first and second order derivatives $\dot{f}(a_{s_i}[t-1])$ and $\ddot{f}(a_{s_i}[t-1])$, respectively, i.e.,

$$\begin{aligned} \dot{f}(a_{\mathbf{s}_i}^{(k-1)}[t]) = & -\frac{1}{\gamma d}(a_{\mathbf{s}_i}^{(k)}[t] - a_{\mathbf{s}_i}^{(k-1)}[t]) + \frac{1}{2}\sum_n \varphi_n \cdot \\ & \Big[\big(1 - \frac{b_n^2}{\tilde{\lambda}_n}\big)\big(1 + \frac{2M_{nn}}{M_{nn} - \tilde{\lambda}_n}\big)\Big] \end{aligned} \tag{6}$$

$$\begin{aligned} \ddot{f}(a_{\mathbf{s}_i}^{(k-1)}[t]) = & \frac{1}{\gamma d} + \frac{1}{2}\sum_n \varphi_n(1 - \varphi_n) + \frac{1}{2}\sum_n \frac{b_n^2 \varphi_n}{\tilde{\lambda}_n} \cdot \\ & \Big[1 - 2(1 - \varphi_n)\frac{M_{nn}}{M_{nn} - \tilde{\lambda}_n}\Big] + \sum_n \varphi_n b_n \frac{M_{nn}}{M_{nn} - \tilde{\lambda}_n} \cdot \\ & \Big[1 - \varphi_n\big(\frac{M_{nn}}{M_{nn} - \tilde{\lambda}_n}\big)\Big] \end{aligned} \tag{7}$$

Unique drifts allow the respective stock latent vectors $\mathbf{s}_i^{(k)}[t]$ to capture price volatility efficiently in every event. This information is then attached to particle coefficient vector. A formal proof can be found in [16].

C. Particle filtering

Let $\mathbf{y}_{(r)}^{(k)}[t] \in \mathbb{R}^n, r = 1, ..., x$ denote the observation vector, where the rth observation is its rth component, which is computed for every $t = 1, ..., d$ at event k. In order to estimate a price in event k, the averages of x observation vectors are computed. Let $\mathbf{c}^{(r)}[t] \in \mathbb{R}^n$ be a coefficient vector with

components

$$c_{(r)l}^{(k)}[t] = \left[\sigma_{\mathbf{s}_{i,l}}^{(k)}[t]\right]^2 \cdot \left[\sigma_{\mathbf{m}_{j,l}}^{(k)}[t]\right]^2 \cdot u^{(r)}[t], \quad l = 1, 2, \ldots, n \quad (8)$$

where $\left[\sigma_{\mathbf{s}_{i,l}}^{(k)}[t]\right]^2$, $\left[\sigma_{\mathbf{m}_{j,l}}^{(k)}[t]\right]^2$ are the lth diagonal elements of the prior covariance matrices $\mathbf{\Sigma}_{\mathbf{s}_i}^{(k)}[t], \mathbf{\Sigma}_{\mathbf{m}_j}^{(k)}[t]$, respectively, and $u^{(r)}[t] \sim \mathcal{N}\big(0, 1\big)$ [15]. The particles are generated according to these coefficients, i.e.,

$$\mathbf{o}_{(r)}^{(k)}[t] = \mathbf{o}_{(r)}^{(k-1)}[t] + \mathbf{c}_{(r)}^{(k-1)}[t], r = 1, \ldots, x. \quad (9)$$

Here, x particles are sampled from the importance distribution $p(\mathbf{o}_{(r)}^{(k)}[t]|\mathbf{o}_{(r)}^{(k-1)}[t])$ [17]. This is the sequential importance sampling (SIS) step. At event k and time step $t = 1, \ldots, d$ the observation model of multiT-DPF is

$$\mathbf{y}_{(r)}^{(k)}[t] = \mathbf{o}_{(r)}^{(k)}[t] + \mathbf{c}_{(r)}^{(k)}[t], \; r = 1, \ldots, x. \quad (10)$$

Then, systematic resampling is applied within ϱ resampling loops to prevent particle degeneracy [18]. Let

$$\rho_{(r)}^{(k)} = \frac{(r-1) + \tilde{\rho}}{x}, \; \tilde{\rho} \sim \mathcal{U}\big(0, 1\big) \quad (11)$$

where $\rho_{(r)}^{(k)} \in \left[\frac{r-1}{x}, \frac{r}{x}\right)$. Then, particle weights are

$$\delta_{(r)}^{(k)}[t] \sim \mathcal{N}\big(y^{(k-1)}[t], v_{(r)}^{(k)}[t]\big), \; r = 1, \ldots, x \quad (12)$$

with $v_{(r)}^{(k)}[t]$ distributed as a truncated normal pdf $\mathcal{TN}\big(\bar{y}_{(r)}^{(k)}[t], \sigma_\delta^2\big)$, where $\bar{y}_{(r)}^{(k)}[t] = \frac{1}{n}\mathbf{1}_{n\times 1}^T \mathbf{y}_{(r)}^{(k)}[t]$ is the average value of the r elements of the observation vector in Eq.(10). A threshold $\vartheta = \sum_{r=1}^{x} \tilde{\psi}_{(r)}^{(k)}$, is defined, where $\tilde{\psi}_{(r)}^{(k)} = \psi_{(r)}^{(k)} / \sum_{\xi=1}^{x} \psi^{(\xi)}$, and $\psi_{(r)}^{(k)} = \exp\big(-\delta_{(r)}^{(k)} / \max_{\xi}\{|\delta^{(\xi)}|\}\big)$ [11]. If $\rho_{(r)}^{(k)} < \vartheta$ a new observation vector is generated according to (10), otherwise the current $\mathbf{y}_{(r)}^{(k)}[t]$ is used. Then roughening is applied to ensure observation vectors variance as $\mathbf{y}_{(r)}^{(k)}[t] \leftarrow \mathbf{y}_{(r)}^{(k)}[t] + \boldsymbol{\kappa}[t]$, where $\boldsymbol{\kappa}[t] \sim \mathcal{N}\big(\mathbf{0}, \sigma_{\boldsymbol{\kappa}}^2 \mathbf{I}\big)$. Finally, the estimation of d prices for the event k and time steps $t = 1, ..., d$ is given by computing the average across x observation vectors and their n components at every resampling loop as

$$\hat{y}^{(k)}[t] = \frac{1}{n \cdot x} \sum_{l=1}^{n} \sum_{r=1}^{x} y_{(r)l}^{(k)}[t] \quad (13)$$

where $y_{(r)l}^{(k)}[t]$ denotes the lth element of $\mathbf{y}_{(r)}^{(k)}[t]$. Within a current event k, d log drifts $a_{\mathbf{s}_i}^{(k)}$, d prior and posterior mean vectors $\boldsymbol{\mu}_{\mathbf{s}_i}^{(k)}$ and $\boldsymbol{\mu}_{\mathbf{s}_i}^{\prime(k)}$, d prior and posterior covariance matrices $\mathbf{\Sigma}_{\mathbf{s}_i}^{(k)}$ and $\mathbf{\Sigma}_{\mathbf{s}_i}^{\prime(k)}$, x particle coefficient vectors are computed for every $t = 1, 2, \ldots, d$, based exclusively on the d prices of the previous event $k-1$. Then, d price predictions are computed for $k+1$.

III. Experimental results

Experiments have demonstrated that the proposed multi-step framework is efficient and it keeps the computational complexity low. They are conducted on opening daily prices within $1961 - 2018$ from Coca-Cola, Pepsi, Novartis, Roche, Pfizer, BP, Shell, and Posco stocks and the prediction performance of multiT-DPF is compared to that of DPF [11] and to that of the deep learning algorithms in [3], [4], [5]. multiT-DPF is an online algorithm which predicts the $d = 10$ next prices given the past 10 ones. Time steps corresponds to days. For the deep learning algorithms 70% of data are used as training set and 30% as test set. To ensure a fair comparison, multiT-DPF is applied on the same test set and the prediction horizon is $d = 10$, as well. DPF is applied on the same test set, as well. Since DPF is a single-step prediction algorithm, only the dth prediction is recorded. Thus, multiT-DPF and DPF predictions are made comparable. The RMSE related to these predictions is calculated. Here, the latent dimension is $n = 5$, the number of particles is $x = 20$, the number of resampling loops is $\varrho = 100$, $\sigma_{\boldsymbol{\kappa}}^2 = 0.01$, $\sigma_\delta^2 \sim \mathcal{TN}\big(0, 0.03\big)$, $\sigma_{ij}^2 = 0.01$, the additional drift parameter is $\gamma = 8 \times 10^{-3}$, and the initial drift value of the Brownian motion of $\mathbf{s}_i$ is $a_{\mathbf{s}_i}[t_0] = -6.5$. The respective log drift parameter is fixed and common across all segment vectors j, i.e., $a_{\mathbf{m}_j}[t] = -18$.

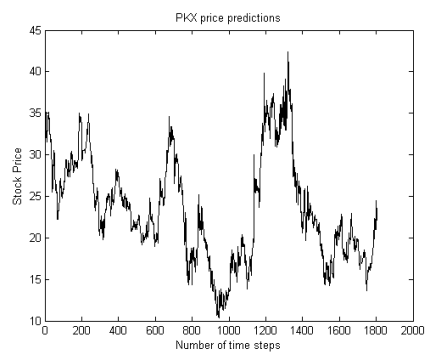

Fig. 1: Historical Posco prices.

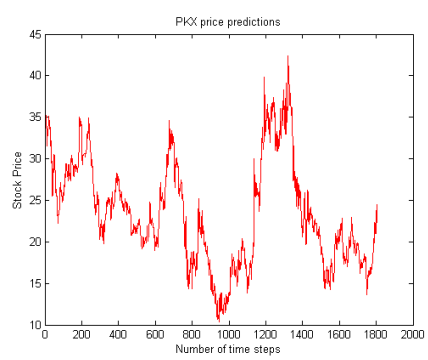

Fig. 2: Predicted Posco prices.

TABLE I: Prediction performance comparison.

Stock	**RMSE** (USD)					number of prices	Price Range (USD)
	multiT-DPF	**DPF** [11]	**train / test** [3]	**train / test** [4]	**train / test** [5]	train/test	
BP	**0.8494**	1.2736	1.6690 / 1.8103	2.5413 / 3.1994	2.1176 / 2.9491	7,361 / 3,154	27.25 - 147.125
Coca-Cola	**0.0391**	0.8066	0.8001 / 0.9794	1.1086 / 1.1317	1.4794 / 1.5188	9,990/4,281	28.875 - 155.75
Novartis	**1.1246**	1.8501	1.8791 / 1.9143	2.9644 / 3.4152	2.8732 / 3.0641	3,847 / 1,649	27.313 - 106.73
Pepsi	**0.0895**	0.8934	1.4229 / 1.5026	1.6132 / 1.7266	1.7067 / 1.8613	8,170 / 3,501	0.544 - 121.38
Pfizer	**0.0621**	1.0055	0.9709 / 1.2648	1.4174 / 1.5478	1.3279 / 1.4116	8,170 / 3,501	11.84 - 149.187
Posco	**1.9089**	2.5007	2.5976 / 2.9129	4.0122 / 5.5944	3.5084 / 3.8871	4,213 / 1,806	10.3750 - 42.3750
Roche	**1.0022**	1.4113	1.2699 / 1.4017	1.2938 / 1.3661	1.3144 / 1.4006	2,654 / 1,137	9.5313 - 38.6
Shell	3.0098	**2.8209**	2.7462 / 2.9316	2.9862 / 3.2684	3.1461 / 3.6806	2,316 / 992	36.75 - 88.09

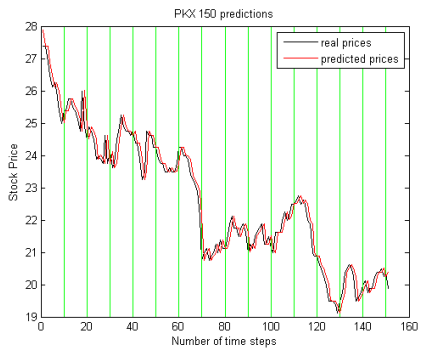

Fig. 3: Zoom in 150 Posco prices.

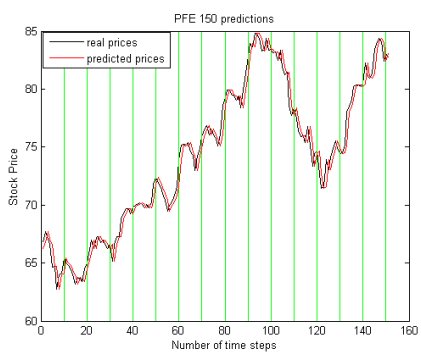

Fig. 4: Zoom in 150 Pfizer prices.

Figures 1 and 2 illustrate the historical and predicted stock prices of Posco, respectively. Figures 3 and 4 depict a zoom in 150 prices of Posco and Pfizer, respectively. The green vertical lines define the events of length $d = 10$ time steps. As can be seen, multiT-DPF captures efficiently price evolution within every event, without losing tracking through time. This fact indicates that the prediction error is not propagated. The overall prediction performance is summarized in Table I. The root-mean-square error (RMSE) was utilized as figure of merit. The first column refers to each stock, the second column refers to the performance of the proposed multiT-DPF, the third column indicates the performance of DPF [11], the fourth column indicates the performance of [3], the fifth column refers to the performance of [4], the sixth column summarizes the performance of [5], the seventh column refers to the number of train and test prices, and the price range of each stock in USD is shown in the eighth column.

F-tests are applied to evaluate the statistical significance of MSE differences. The F statistic is $F_{1\iota} = \frac{\pi_1^2}{\pi_\iota^2}$, where the subscript 1 refer to multiT-DPF and $\iota = 2, 3, 4, 5$ refer to DPF [11], [3], [4], and [5], respectively. For each method $\pi^2 = \frac{1}{N-1}\sum_{l=1}^{N}(\hat{y}_l - \bar{y})^2$ is defined, where $\hat{y}$ is the predicted price, $\bar{y} = \frac{1}{N}\sum_{l=1}^{N}\hat{y}$ is the mean of predicted prices, N is the size of the test set appearing in Table I, and the significance level of the F test is $\beta = 5\%$. The null hypothesis between multiT-DPF and every other method ι is $H_0 : \pi_1^2 = \pi_\iota^2$, which is rejected if $F_{1\iota} < F_{\beta/2}$ or $F_{1\iota} > F_{1-\beta/2}$, where $F_{\beta/2} = F(\beta/2, N-1, N-1)$ and $F_{1-\beta/2} = F(1-\beta/2, N-1, N-1)$ are the critical values of the F distribution with significance level equal to the subscript and $N - 1$ are the degrees of freedom.

TABLE II: F-test multiT-DPF against the methods in [11], [3], [4], and [5].

Stock	$F_{\beta/2}$	$F_{1-\beta/2}$	F_{12}	F_{13}	F_{14}	F_{15}
BP	0.9431	1.0603	**0.7016**	**0.6724**	**0.5102**	**0.5391**
Coca-Cola	0.9510	1.0516	**0.1088**	**0.0967**	**0.0238**	**0.2089**
Novartis	0.9221	1.0844	**0.7208**	**0.7001**	**0.4122**	**0.3855**
Pepsi	0.9459	1.0572	**0.0901**	**0.0826**	**0.0693**	**0.0201**
Pfizer	0.9459	1.0572	**0.1055**	**0.0993**	**0.0621**	**0.0287**
Posco	0.9255	1.0805	**0.4394**	**0.4064**	**0.0589**	**0.1993**
Roche	0.9070	1.1025	**0.7402**	**0.7695**	**0.7008**	**0.7674**
Shell	0.9007	1.1102	1.0105	1.0012	**1.1941**	**0.9001**

For Shell stock, every event exhibits major price variations, i.e., sharp price jumps, with respect to the previous event. Since multiT-DPF utilized only the information of the previous event, it lost price tracking. multiT-DPF outperformed all other methods in 7 out of 8 stocks. In these cases, the MSE differences are statistical significant as can be seen in Table II. Especially in case of BP, Coca-Cola, Pepsi, and Pfizer the RMSE is remarkably better.

The coefficient of determination $R^2 \in [0, 1]$ measures how well each model fits the data. The prediction performance is assessed w.r.t. the coefficient of determination in Table III. The findings concur with the assessment of statistical significance of MSE differences.

TABLE III: Coefficient of determination.

Stock	R^2				
	multiT-DPF	DPF [11]	[3]	[4]	[5]
BP	**0.9615**	0.9208	0.9193	0.8497	0.8561
Coca-Cola	**0.9950**	0.9001	0.8944	0.8892	0.8843
Novartis	**0.9466**	0.9195	0.9207	0.8065	0.8171
Pepsi	**0.9899**	0.9199	0.9281	0.9190	0.9055
Pfizer	**0.9815**	0.9600	0.9575	0.9508	0.9599
Posco	**0.9405**	0.8595	0.8709	0.8411	0.8564
Roche	**0.9841**	0.9602	0.9660	0.9697	0.9657
Shell	0.9008	**0.9197**	0.9119	0.8998	0.8901

To examine the impact of the proposed approximation of the Brownian log drift parameter, a comparison is conducted between multiT-DPF using the proposed approximations and multiT-DPF using the approximations proposed in [12]. Both versions of the algorithm were tested on the whole price time series. The results are summarized in Table IV.

TABLE IV: RMSE of multiT-DPF when the approximations of Eqs. (6) and (7) are used against the RMSE of multiT-DPF when approximations proposed in [12] are employed.

Stock	**RMSE** (USD)		
	multiT-DPF using (6) and (7)	**multiT-DPF** with approximations proposed in [12]	Number of prices
BP	**0.7406**	1.5694	10,515
Coca-Cola	**0.4807**	1.6388	14,271
Novartis	**1.0001**	1.2615	5,496
Pepsi	**0.9710**	1.6141	11,671
Pfizer	**0.5100**	1.6276	11,671
Posco	**1.8011**	2.2491	6,019
Roche	**1.1096**	2.4954	3,791
Shell	**3.0607**	3.7095	3,308

As can be seen, multiT-DPF employing (6) and (7) performed best. This fact indicates that the proposed approximation of the Brownian log drift parameter plays a critical role in the performance of multiT-DPF, enhancing the efficacy of particle coefficients and the overall robustness of the filter. The introduced multi-step framework along with the proposed approximation for the log drift parameter led multiT-DPF to capture consistently the chaotic behaviour of stock prices through time.

IV. Conclusions

A multi-step ahead dyadic particle filter, namely multiT-DPF has been proposed for stock price prediction. The introduced Brownian log drift parameter empowers the latent vectors to track closely the price evolution through time and ensures high prediction performance. The multi-step framework ensures computational feasibility and leads to efficient price prediction despite the fact that the given information lies 10 time steps back in the past. multiT-DPF performed impressively better than the state-of-the-art methods in 7 out of 8 stocks. Experiments have shown that the proposed approximation of the Brownian log drift parameter enhanced strongly the performance of the particle filter.

The proposed multi-step ahead dyadic particle filter is an online method, which relies on historical stock prices. External information (e.g., user information) is not regarded, because such information is scarce. By mining historical price time-series, model simplicity is promoted, guaranteeing successful price inference. However, the dyadic collaborative filtering can accommodate users and stocks, if such data are available.

Furthermore, the proposed method can be tested for multi-step ahead prediction of room rates at a destination across hotels with the same stars as well as flight prices for a destination across various passenger classes.

References

[1] H. Cheng, P. N. Tan, J. Gao, and J. Scripps, "Multistep-ahead time series prediction," in *Pacific-Asia Conf. on Knowledge Discovery and Data Mining*. Springer, 2006, pp. 765–774.

[2] R. Chandra, Y. S. Ong, and C. K. Goh, "Co-evolutionary multi-task learning with predictive recurrence for multi-step chaotic time series prediction," *Neurocomputing*, vol. 243, pp. 21–34, 2017.

[3] B. Lim, S. O. Arik, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *arXiv preprint arXiv:1912.09363*, 2019.

[4] G. Lai, W. C. Chang, Y. Yang, and H. Liu, "Modeling long-and short-term temporal patterns with deep neural networks," in *Proc. 41st Int. ACM SIGIR Conf. Research & Development in Information Retrieval*, 2018, pp. 95–104.

[5] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, "DeepAR: Probabilistic forecasting with autoregressive recurrent networks," *Int. Journal of Forecasting*, vol. 36, no. 3, pp. 1181–1191, 2020.

[6] Z. Z. Htike, "Multi-horizon ternary time series forecasting," in *Proc. 2013 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, 2013, pp. 337–342.

[7] S. S. Rangapuram, M. W. Seeger, J. Gasthaus, L. Stella, Y. Wang, and T. Januschowski, "Deep state space models for time series forecasting," in *Advances in Neural Information Processing Systems*, 2018, pp. 7785–7794.

[8] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y. X. Wang, and X. Yan, "Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting," in *Advances in Neural Information Processing Systems*, 2019, pp. 5243–5253.

[9] J. Murphy and S. Godsill, "Bayesian parameter estimation of jump-Langevin systems for trend following in finance," in *Proc. 40th IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2015, pp. 4125–4129.

[10] R. S. Tsay, "Financial Time Series," *Wiley StatsRef: Statistics Reference Online*, pp. 1–23, 2014.

[11] M. Ntemi and C. Kotropoulos, "A dyadic particle filter for price prediction," in *Proc. 27th European Signal Processing Conf.(EUSIPCO)*, 2019, pp. 2419–2423.

[12] S. Gultekin and J. Paisley, "A collaborative Kalman filter for time-evolving dyadic processes," in *Proc. 2014 IEEE Int. Conf. Data Mining*, 2014, pp. 30–37.

[13] M. J. Wainwright and M. I. Jordan, "Graphical Models, Exponential Families, and Variational Inference," *Foundations and Trends® in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.

[14] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of American Statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[15] M. Ntemi and C. Kotropoulos, "A dynamic dyadic particle filter for price prediction," *Signal Processing*, vol. 167, 2020, https://doi.org/10.1016/j.sigpro.2019.107334.

[16] ——, "A jump-diffusion particle filter for price prediction," *Signal Processing*, vol. 183, 2021, https://doi.org/10.1016/j.sigpro.2021.107994.

[17] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," *Handbook of Nonlinear Filtering*, vol. 12, no. 656-704, p. 3, 2009.

[18] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Population Monte Carlo schemes with reduced path degeneracy," in *Proc. 7th IEEE Int. Workshop Computational Advances Multi-Sensor Adaptive Processing*, 2017, pp. 1–5.