Prediction Methods for Time Evolving Dyadic Processes

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

***Abstract*—Stock prices evolve dynamically through time. Capturing their changes is crucial in order to make accurate predictions. In addition, it is well-known that the probability density function of stock prices exhibits heavy tails and there is a large degree of uncertainty in stock price evolution. Building on the aforementioned facts, a robust collaborative Kalman filter is proposed for stock price prediction within the context of time-evolving dyadic processes, where the prediction error is treated as a heavy-tailed noise whose variance is a properly modeled random variable. Variational approximation is exploited to make posterior inference tractable. The proposed model captures the volatility of stock prices through time, yielding more accurate predictions than the state-of-the-art and enabling the consistent tracking of the extreme values of stock prices.**

I. Introduction

Collaborative filtering is a special case of dyadic prediction. It is the process of filtering information by utilizing the assessments of other individuals. It is widely applied to recommender systems, where the goal is to anticipate estimations of user preferences and suggest items as close as possible to users' taste. The input to the prediction system is a dyad, i.e., a user-item pair. For example, user rating systems, such that of Netflix, Amazon, and YouTube utilize their users' preferences for movies, products, or various videos, respectively, in order to make satisfactory recommendations. Another application is stock price prediction, where historical data are utilized in order to make future predictions. The most successful collaborative filtering models are based on matrix factorization (MF) techniques [1], [2]. In MF techniques, users and objects are mapped into the same latent space, such that their interactions are represented as inner products in this joint latent space [3]. However, user and object locations are considered to be fixed through time. That is, a stationary process is assumed that does not take into account a possible change in users' preferences or the dynamic fluctuations of stock prices through time. Tracking data evolution over time is crucial for efficient predictions. Collaborative Kalman filter (CKF) constitutes a dynamic extension of MF methods. In CFK, user and item vectors are allowed to shift spatially in the latent space through time according to a multidimensional Brownian motion [4]. This property enables the model to track the temporal changes that may occur in user preferences. In stock price prediction, the Brownian motion drift parameter is not static, but a dynamic one in order to capture the price volatility. Thus, a dynamic drift parameter for each stock is being learned by the model.

Despite the dynamic properties of the CKF model, the assumption of a fixed variance for the measurement noise (i.e., prediction error) does not capture the uncertainty inherent in stock prices. Moreover, it is assumed that stock price series follow heavy-tailed distributions, because price deviations from the mean value are more likely to occur than those predicted by the Gaussian distribution [5]. Motivated by the work in [6], to address the just described shortcoming, we propose a robust CKF with a heavy-tailed measurement noise, whose variance is treated as a random variable, extending thus the framework proposed in [4]. The aforementioned "uncertainty about uncertainty" assumption endows the CKF with the ability to model the volatility of latent state vectors. As a result, the model tracks consistently the intense fluctuations of stock prices through time and predicts accurately the extreme values. Throughout the paper, the following metaphor is applied: stock prices play the role of users, and the states of the world play the role of items as in [4]. A dynamic probabilistic approach is applied to estimate the latent user and item vector distributions at each time-step. To maintain tractability, a variational inference approximation of posterior densities is applied [7].

The outline of the paper is as follows. In Section II, we review briefly the CKF. In Section III, we introduce the CKF with heavy-tailed measurement noise. Approximate distributions are derived in Section IV by means of variational inference. Experimental results are demonstrated in Section V.

II. Collaborative Kalman Filter

In Kalman filtering, the current state vectors are equal to the previous state ones plus additive noise [4], [8]. Let $\mathbf{y}_n \in \mathbb{R}^p$ and $\mathbf{w}_n \in \mathbb{R}^d$ be the sequences of observed measurement and latent state vectors, respectively, for $n = 1, 2, \ldots, N$. Assume that given $\mathbf{w}_n$, $\mathbf{w}_{n+1}$ is distributed as a multivariate Gaussian probability density function (pdf) with mean vector equal to $\mathbf{w}_n$ and covariance matrix $\alpha\, \mathbf{I}$, i.e., $\mathbf{w}_{n+1} \mid \mathbf{w}_n \sim \mathcal{N}(\mathbf{w}_n, \alpha\mathbf{I})$, where $\mathbf{I}$ is the identity matrix of compatible size and α is a dynamically evolving drift parameter, which reflects the volatility of $\mathbf{w}_{n+1}$.

Assume that the latent state vector $\mathbf{w}_n$ in the *current* state follows a multivariate normal distribution $\mathbf{w}_n \sim \mathcal{N}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$ with mean vector $\boldsymbol{\mu}_n$ and covariance matrix $\boldsymbol{\Sigma}_n$. To find the posterior distribution of the *next* state $\mathbf{w}_{n+1}$, we marginalize with respect to the current state, obtaining thus $p(\mathbf{w}_{n+1}) = \mathcal{N}(\boldsymbol{\mu}_n, \alpha\mathbf{I} + \boldsymbol{\Sigma}_n)$. In order to exclude negative values, we model α as a geometric Brownian motion, setting $\alpha[t] = e^{a[t]}$, where a is a Brownian motion distributed as [4]:

$$a[t] \sim \mathcal{N}(a[t-\Delta_a^{[t]}], \xi\, \Delta_a^{[t]}) \tag{1}$$

with $t - \Delta_a^{[t]}$ being the time elapsed since $a[t]$ was previously measured and ξ being an extra drift parameter.

Subsequently, after having observed a new measurement, the posterior distribution is given by a multivariate Gaussian pdf with mean vector $\boldsymbol{\mu}_{n+1}$ and covariance matrix $\boldsymbol{\Sigma}_{n+1}$, i.e., $p(\mathbf{w}_{n+1} \mid \mathbf{y}_{n+1}) = \mathcal{N}(\boldsymbol{\mu}_{n+1}, \boldsymbol{\Sigma}_{n+1})$. The posterior of $\mathbf{w}_{n+1}$, then can be used to obtain the prior of the next state $\mathbf{w}_{n+2}$ and so on.

Let $\mathbf{u}_i \in \mathbb{R}^d$ and $\mathbf{w}_j \in \mathbb{R}^d$ denote the user and item vectors, respectively, consisting the dyad i, j. The prediction of each dyad at every time step is carried out by estimating the pdf of the inner product between $\mathbf{u}_i$ and $\mathbf{w}_j$. In Kalman filtering, the posterior distributions of latent user and item vectors are in fact multivariate normal distributions, defined as

$$\begin{aligned} \mathbf{u}_i[t-\Delta_{u_i}^{[t]}] &\sim \mathcal{N}(\boldsymbol{\mu}'_{u_i}[t-\Delta_{u_i}^{[t]}], \boldsymbol{\Sigma}'_{u_i}[t-\Delta_{u_i}^{[t]}]), \quad (2)\\ \mathbf{w}_j[t-\Delta_{w_j}^{[t]}] &\sim \mathcal{N}(\boldsymbol{\mu}'_{w_j}[t-\Delta_{w_j}^{[t]}], \boldsymbol{\Sigma}'_{w_j}[t-\Delta_{w_j}^{[t]}]) \quad (3)\end{aligned}$$

where $t - \Delta_{u_i}^{[t]}$ denotes the time elapsed, since user $\mathbf{u}_i$ was previously observed. Let us omit the subscripts $\mathbf{u}_i$ and $\mathbf{w}_j$, when they can easily be implied from the context. The mean vector $\boldsymbol{\mu}'$ and the covariance matrix $\boldsymbol{\Sigma}'$ are also dynamically evolving quantities. The marginalization of Eq. (2) or Eq. (3) leads to the latent vector prior distributions.

III. Collaborative Kalman Filter with heavy-tailed measurement noise

The rate of change of stock prices is not constant over time and an exact pattern cannot be identified. Accordingly, the main goal is to model the time varying volatility of prices. We propose a dynamic model, which is capable of capturing this volatility by means of a combination of Brownian motion and an uncertain variance of the prediction error, which is treated as a random variable. If y_{ij} is the measurement (i.e., stock market price) for the dyad i, j, the following state-space model is assumed:

$$\mathbf{u}_i[t] \sim \mathcal{N}(\boldsymbol{\mu}_{u_i}[t], \boldsymbol{\Sigma}_{u_i}[t]),\ \mathbf{w}_j[t] \sim \mathcal{N}(\boldsymbol{\mu}_{w_j}[t], \boldsymbol{\Sigma}_{w_j}[t]) \tag{4}$$

$$y_{ij}[t] \mid \mathbf{u}_i, \mathbf{w}_j, r[t] \sim \mathcal{N}(\mathbf{u}_i[t]^T \mathbf{w}_j[t], r[t]). \tag{5}$$

Eq. (4) refers to the state model, consisted of the prior distributions of user $\mathbf{u}_i$ and item $\mathbf{w}_j$ at time t with prior mean vector $\boldsymbol{\mu}$ and prior covariance matrix $\boldsymbol{\Sigma}$,

$$\begin{aligned} \boldsymbol{\mu}[t] &= \boldsymbol{\mu}'[t-\Delta^{[t]}] \quad (6)\\ \boldsymbol{\Sigma}[t] &= \boldsymbol{\Sigma}'[t-\Delta^{[t]}] + \Delta^{[t]}\alpha[t]\,\mathbf{I} \quad (7)\end{aligned}$$

where $\boldsymbol{\mu}'[], \boldsymbol{\Sigma}'[]$ are the posterior parameters at $t - \Delta^{[t]}$, and α is the geometric Brownian motion drift parameter, which captures the volatility of latent variables at one unit of time. Eq. (5) is the measurement model, which asserts that the distribution of y_{ij} given the inner product between $\mathbf{u}_i$ and $\mathbf{w}_j$ is a Gaussian distribution with mean equal to the inner product of the latent vectors of user $\mathbf{u}_i$ and item $\mathbf{w}_j$ at time t and variance equal to $r[t]$.

The heavy-tailed behaviour of y_{ij} can be captured by treating $r[t]$ as a random variable. Generally, the covariance matrix of a random Gaussian measurement vector of size d with a known mean could be considered as random matrix, whose prior distribution is the inverse Wishart distribution $\mathbf{R}[t] \sim W^{-1}(\boldsymbol{\Psi}, v)$, where $\boldsymbol{\Psi} \in \mathbb{R}^{d\times d}$ is the scale matrix and $v > d-1$ are the degrees of freedom [6]. The inverse Wishart distribution is the conjugate prior for the covariance matrix of a jointly Gaussian random vector. Here, we are interested in the univariate case. Capitalizing on the fact that the conjugate prior of the variance of the Gaussian random variable $y_{ij}[t]$ is the inverse Gamma distribution, the random variable $r[t]$ at time step t, is modeled as:

$$r[t] \sim IG(\eta[t], \theta[t]) \tag{8}$$

where $\eta[t]$ and $\theta[t]$ are the shape and scale parameter, respectively, which are referred to as prior hyperparameters. The conjugacy ensures that the posterior distribution is of the same form as the prior, allowing for a less complex analysis. The posterior hyperparameters at time t are computed as:

$$\begin{aligned} \eta[t] &= \eta[t-\Delta^{[t]}] + \frac{|\mathcal{T}|}{2} \quad (9)\\ \theta[t] &= \theta[t-\Delta^{[t]}] + \frac{\sum_{t\in\mathcal{T}}(y_{ij}[t] - \mu_y[t])^2}{2} \quad (10)\end{aligned}$$

where $\mu_y[t] = \frac{\sum_{t\in\mathcal{T}} y_{ij}[t]}{|\mathcal{T}|}$ is the mean value of the past observations included in a sliding window $\mathcal{T}$ over the time series and $|\mathcal{T}|$ denotes the window length.

IV. Variational Inference

The next step is the calculation of posterior distribution at time step t

$$p(\mathbf{u}_i[t], \mathbf{w}_j[t], y_{ij}[t] \mid r[t]). \tag{11}$$

As a result of interdependencies, an analytical solution is not attainable. An approximate solution can be obtained through variational inference [9]. The approximate solution is obtained through a factorized distribution,

$$\begin{aligned} &p(\mathbf{u}_i[t], \mathbf{w}_j[t], y_{ij}[t] \mid r[t]) \approx \\ &\quad q(\mathbf{u}_i[t])\, q(\mathbf{w}_j[t])\, q(y_{ij}[t])\, q(r[t]) \quad (12)\end{aligned}$$

where $q(\cdot)$ is the approximate posterior distribution. In Eq. (12), statistical independence is assumed between the variables. At time step t, the Kullback-Leibler (KL) divergence between the approximate distribution $q(\cdot)$ and the true posterior distribution $p(\cdot)$ should be minimized. The KL divergence is given by,

$$KL(q||p) = \mathbb{E}_q\big[\log \frac{q}{p}\big] \tag{13}$$

where $\mathbb{E}_q[]$ denotes expectation with respect to pdf q. Practically, the minimization of KL divergence is infeasible. However, a function which is equal to KL up to an additive constant, can be utilized. Such function is known as the *evidence lower bound* function $\mathcal{L}$, and it can be obtained by applying Jensen's inequality to the log probabilities [7]. The minimization of KL divergence is equivalent to maximizing $\mathcal{L}$. This can be achieved by finding the closest distribution family to the true posterior distribution. For simplicity, the *mean-field variational family* is chosen. The objective is to maximize the following function,

$$\mathcal{L} = \mathbb{E}_q[\log p(\mathbf{u}_i[t], \mathbf{w}_j[t], y_{ij}[t], r[t]) - \mathbb{E}_q[\log q] \tag{14}$$

where the first term on the right-hand side is the expected joint log-likelihood, while the second term is the entropy of the approximate distribution.

A. Calculation of variational distribution

The optimization procedure is an iterative one. Only one variational factor is optimized at a time, while keeping the other variational factors fixed:

$$q_\ell^* \propto \exp\{\mathbb{E}_{-\ell}[\log\, p(\cdot)]\} \tag{15}$$

where $\ell \in \{\mathbf{u}_i, \mathbf{w}_j, y_{ij}, r\}$. In Eq. (15), q_ℓ^* represents the optimal solution, which is proportional to the exponentiated expected log true posterior holding out the distribution of the variable ℓ of interest. At each step, Eq. (15) is recomputed. Thus, the optimal approximate distribution of y_{ij} at time t is a truncated normal distribution [4]:

$$q^*(y_{ij}[t]) = \mathcal{TN}(\boldsymbol{\mu}_{u_i}^T[t]\boldsymbol{\mu}_{w_j}[t], r[t]) \tag{16}$$

where the mean is the inner product between the prior mean (latent) vectors of user i, $\mathbf{u}_i$, and item j, $\mathbf{w}_j$, at time t, and variance equal to $r[t]$.

The respective optimal approximate distributions of the latent vectors of user $\mathbf{u}_i$ and item $\mathbf{w}_j$ at time step t are found to be multivariate Gaussian distributions:

$$q^*(\mathbf{u}_i[t]) = \mathcal{N}(\boldsymbol{\mu}'_{u_i}[t], \boldsymbol{\Sigma}'_{u_i}[t]) \tag{17}$$
$$q^*(\mathbf{w}_j[t]) = \mathcal{N}(\boldsymbol{\mu}'_{w_j}[t], \boldsymbol{\Sigma}'_{w_j}[t]). \tag{18}$$

In order to infer the geometric Brownian motion, individual drift parameters a for each user $\mathbf{u}_i$ and item $\mathbf{w}_j$ are assumed and a point estimation procedure is exploited. The updates of the drift parameters are found by approximating the relevant terms in the likelihood with a second-order Taylor expansion about the last inferred value $a_{u_i}[t-\Delta_{a_{u_i}}^{[t]}]$. Let us assume the second-order Taylor approximation of $f(a_{u_i}[t])$:

$$\begin{aligned} f(a_{u_i}[t]) &\approx f(a_{u_i}[t-\Delta_{a_{u_i}}^{[t]}]) \\ &+ \big(a_{u_i}[t] - a_{u_i}[t-\Delta_{a_{u_i}}^{[t]}]\big) f'(a_{u_i}[t-\Delta_{a_{u_i}}^{[t]}]) \\ &+ \frac{1}{2}\big(a_{u_i}[t] - a_{u_i}[t-\Delta_{a_{u_i}}^{[t]}]\big)^2 f''(a_{u_i}[t-\Delta_{a_{u_i}}^{[t]}]) \end{aligned} \tag{19}$$

where $f()$ is chosen as the negative log-likelihood and $f'()$ and $f''()$ denote its first-order and second-order derivative at the argument inside parenthesis, respectively. Seeking to optimize the objective function Eq. (19) with respect to $a_{u_i}[t]$, the following updates are obtained [4]:

$$a_{u_i}[t] = a_{u_i}[t-\Delta_{a_{u_i}}^{[t]}] - \frac{f'(a[t-\Delta_{a_{u_i}}^{[t]}])}{f''(a[t-\Delta_{a_{u_i}}^{[t]}])}. \tag{20}$$

B. Updating the approximate distributions

Since the model has dynamic behaviour, the approximate distribution $q(\cdot)$ of each variable should evolve as well. A coordinate ascent update is applied in order to find the optimal parameters of each approximate distribution $q(\cdot)$, which ensures a global maximum of $\mathcal{L}$.

The following analytical result holds [4]:

$$\mathbb{E}_q\big[y_{ij}[t]\big] = \mathbb{E}_q\big[\mathbf{u}_i\big]^T\,\mathbb{E}_q\big[\mathbf{w}_j\big] = \boldsymbol{\mu}_{u_i}^T[t]\,\boldsymbol{\mu}_{w_j}[t] \tag{21}$$

where $\boldsymbol{\mu}_{u_i}[t], \boldsymbol{\mu}_{w_j}[t]$ are the prior mean vectors of $\mathbf{u}_i$ and $\mathbf{w}_j$ at time t, respectively.

The variational update for the parameters of the approximate $q(\mathbf{u}_i)$ for user i at time step t are given by [4]:

$$\boldsymbol{\Sigma}'_{u_i}[t] = \left(\boldsymbol{\Sigma}_{u_i}^{-1}[t] + \frac{\boldsymbol{\mu}'_{w_j}[t]\,\boldsymbol{\mu}'^T_{w_j}[t] + \boldsymbol{\Sigma}'_{w_j}[t]}{r[t]}\right)^{-1} \tag{22}$$

$$\boldsymbol{\mu}'_{u_i}[t] = \boldsymbol{\Sigma}'_{u_i}[t]\left(\frac{\mathbb{E}_q[y_{ij}[t]]\boldsymbol{\mu}'_{w_j}[t]}{r[t]} + \boldsymbol{\Sigma}_{u_i}^{-1}[t]\,\boldsymbol{\mu}_{u_i}[t]\right) \tag{23}$$

Similarly, the variational update for the parameters of the approximate $q(\mathbf{w}_j)$ for item j at time step t is:

$$\boldsymbol{\Sigma}'_{w_j}[t] = \left(\boldsymbol{\Sigma}_{w_j}^{-1}[t] + \frac{\boldsymbol{\mu}'_{u_i}[t]\,\boldsymbol{\mu}'^T_{u_i}[t] + \boldsymbol{\Sigma}'_{u_i}[t]}{r[t]}\right)^{-1} \tag{24}$$

$$\boldsymbol{\mu}'_{w_j}[t] = \boldsymbol{\Sigma}'_{w_j}[t]\left(\frac{\mathbb{E}_q[y_{ij}[t]]\boldsymbol{\mu}'_{u_i}[t]}{r[t]} + \boldsymbol{\Sigma}_{w_j}^{-1}[t]\,\boldsymbol{\mu}_{w_j}[t]\right) \tag{25}$$

C. Approximate posterior of the heavy-tailed distribution of r[t]

The approximate posterior distribution $q(r[t])$ is of the same form as the prior distribution of $r[t]$. That is, $q(r[t])$ is also an inverse Gamma distribution, i.e., $q(r[t]) \approx IG\big(\eta[t], \theta[t]\big)$. In order Eq. (5) to hold, the noise statistics should be updated given the latent states by means of

$$\lambda_{u_i}[t] = \frac{c\,r[t] + s_{u_i}[t]}{c+1}, \quad \lambda_{w_j}[t] = \frac{c\,r[t] + s_{w_j}[t]}{c+1} \tag{26}$$

where $c > d-1$, with d representing the latent vector dimension, and $s_{u_i}[t]$ and $s_{w_j}[t]$ are the following sufficient statistics of prediction error variance:

$$s_{u_i}[t] = (y_{ij}[t] - \mathbf{h}^T\boldsymbol{\mu}'_{u_i}[t])^2 + \mathbf{h}^T\boldsymbol{\Sigma}'_{u_i}[t]\mathbf{h} \tag{27}$$
$$s_{w_j}[t] = (y_{ij}[t] - \mathbf{h}^T\boldsymbol{\mu}'_{w_j}[t])^2 + \mathbf{h}^T\boldsymbol{\Sigma}'_{w_j}[t]\mathbf{h} \tag{28}$$

with $\mathbf{h}_{d\times 1} \sim \mathcal{TN}(\mathbf{0}, \mathbf{I})$.

Next, the latent state vectors are updated given the noise statistics. That is, the posterior mean vectors and posterior covariance matrices of u_i and w_j found in Section IV-B, are updated for the next time step $t + \Delta^{[t]}$. Firstly, the Kalman gain vectors of u_i and w_j are derived, which are row vectors, i.e.:

$$\mathbf{k}_{u_i}[t] = \frac{\mathbf{h}^T \boldsymbol{\Sigma}'_{u_i}[t]}{\mathbf{h}^T \boldsymbol{\Sigma}'_{u_i}[t]\mathbf{h} + \lambda_{u_i}[t]} \tag{29}$$

$$\mathbf{k}_{w_j}[t] = \frac{\mathbf{h}^T \boldsymbol{\Sigma}'_{w_j}[t]}{\mathbf{h}^T \boldsymbol{\Sigma}'_{w_j}[t]\mathbf{h} + \lambda_{w_j}[t]}. \tag{30}$$

The Kalman gain vectors are utilized to update the posterior mean vector and the covariance matrix for user u_i and item w_j as follows:

$$\boldsymbol{\mu}_{u_i}[t + \Delta^{[t]}] = \boldsymbol{\mu}'_{u_i}[t] + (y_{ij}[t] - \mathbf{h}^T \boldsymbol{\mu}'_{u_i}[t])\, \mathbf{k}^T_{u_i}[t] \tag{31}$$

$$\boldsymbol{\Sigma}_{u_i}[t + \Delta^{[t]}] = \lambda_{u_i}[t]\, \mathbf{k}^T_{u_i}[t]\mathbf{k}_{u_i}[t] + (\mathbf{I} - \mathbf{h}\mathbf{k}_{u_i}[t])^T \cdot \boldsymbol{\Sigma}'_{u_i}[t]\, (\mathbf{I} - \mathbf{h}\mathbf{k}_{u_i}[t]). \tag{32}$$

Similarly:

$$\boldsymbol{\mu}_{w_j}[t + \Delta^{[t]}] = \boldsymbol{\mu}'_{w_j}[t] + (y_{ij}[t] - \mathbf{h}^T \boldsymbol{\mu}'_{w_j}[t])\, \mathbf{k}^T_{w_j}[t] \tag{33}$$

$$\boldsymbol{\Sigma}_{w_j}[t + \Delta^{[t]}] = \lambda_{w_j}[t]\, \mathbf{k}^T_{w_j}[t]\mathbf{k}_{w_j}[t] + (\mathbf{I} - \mathbf{h}\mathbf{k}_{w_j}[t])^T \cdot \boldsymbol{\Sigma}'_{w_j}[t]\, (\mathbf{I} - \mathbf{h}\mathbf{k}_{w_j}[t]). \tag{34}$$

V. Experimental Results

Experiments for stock price prediction were conducted on the time series consisting of opening values from AMEX, NASDAQ, and NYSE exchange, spanning the period 1962-2018. The latent state vectors $\mathbf{u}_i$ and $\mathbf{w}_j$ have size $d = 5$. Each stock is assigned a latent state vector $\mathbf{u}_i$, which plays the role of user. The latent state vector for items is common across all the stocks, i.e., $\mathbf{w}_j = \mathbf{w}_1$, which is termed as "state-of-the-world" (SOW) vector. The uncertainty about the stock prices is modeled by a random variance $r[t]$, which obeys a heavy-tailed distribution learned by the algorithm. The model learns a drift parameter $a_{\mathbf{u}_i}$ for each stock. The drift parameter for the SOW vector was set to $a_w = -11.7$ [4]. The extra drift parameter ξ was set to $\xi = 5 \times 10^{-2}$. The size of the sliding window was set to $\mathcal{T} = 20$ past prices. Parameter c in Eq. (26) was set to 8.

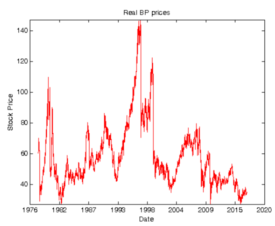

Fig. 1: Real BP prices.

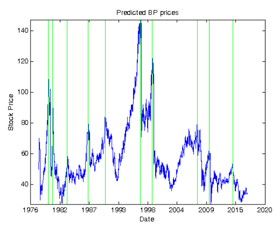

Fig. 2: Predicted BP prices.

Experimental results on BP, Coca-Cola, Pfizer, and Posco stock prices are shown in Figures 1–8. Figures 1, 3, 5, and 7 depict the historical stock prices of BP, Pfizer, Coca-Cola, and Posco companies, respectively. Figures 2, 4, 6, and 8 show the respective predicted values. At each time step, the predicted opening value is very close to the real opening stock value. That is, the predicted prices are approximately identical to the historical prices. More specifically, the green vertical lines overlaid in Figures 2, 4, 6, and 8 mark the peaks found in the historical prices. They are almost perfectly predicted by the proposed algorithm, demonstrating the accuracy of the algorithm. The dynamic properties of the model led it to capture the intense fluctuations and the overall volatility of stock prices through time.

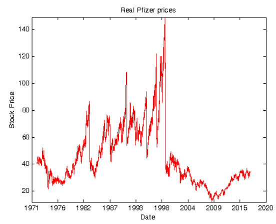

Fig. 3: Real Pfizer prices.

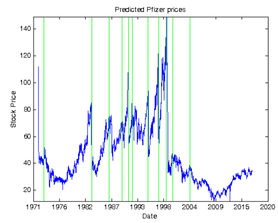

Fig. 4: Predicted Pfizer prices

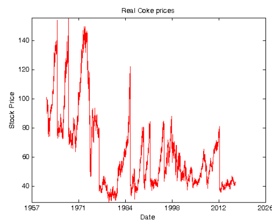

Fig. 5: Real Coca-Cola prices

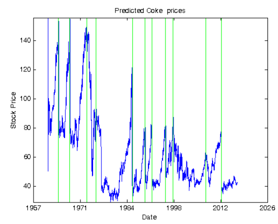

Fig. 6: Predicted Coca-Cola prices

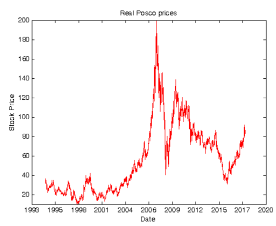

Fig. 7: Real Posco prices

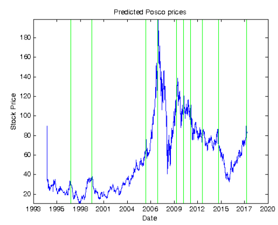

Fig. 8: Predicted Posco prices

The stock price prediction performance is summarized in Table I. In particular, the root-mean-square error (RMSE) for

TABLE I: Stock prices prediction performance.

Stock	**RMSE CKF** (USD)	**RMSE RCKF** (USD)	**Price range** (USD)
BP	8.3794	**2.2650**	27.25-147.125
Coca-Cola	3.5529	**3.0617**	28.875-155.75
Pfizer	4.176	**2.3594**	11.84-149.187
Posco	3.1148	**2.5992**	10.375-200.37
Average	4.8058	**2.5713**	19.59-163.11

various stocks is reported. The second column refers to the performance of the CKF [4], which was implemented from scratch. The setting of parameters $a_{\mathbf{u}_i}$, a_w and ξ in CKF is identical to that used in the proposed model in order to examine the impact of random variable $r[t]$ modeling. In CKF, $\sqrt{r[t]}$, i.e., the standard deviation of the prediction error is fixed and equal to 0.01. The RMSE of the proposed model, abbreviated as RCKF, is shown in the third column. The RMSE of RCKF is less than that of CKF and very low compared to the range of the stock prices. The superior performance of RCKF is attributed to the random variance $r[t]$, which is properly modeled in the proposed model.

F-tests were applied to each stock in order to examine whether the differences in RMSE are statistically significant. The test statistic is defined as $F = \frac{\sigma_1^2}{\sigma_2^2}$, where the subscripts 1 and 2 refer to the CKF and the proposed RCKF model, respectively. For each model, $\sigma^2 = \frac{1}{N-1}\sum_{l=1}^{N}(\hat{y}_l - \bar{y})^2$, where $\hat{y}_l$ is the predicted price, $\bar{y} = \frac{1}{N}\sum_{l=1}^{N}\hat{y}_l$ is the mean of predicted stock prices, and N is the number of observations, i.e., $N = 9,837$ for BP, $N = 13,912$ for Coca-Cola, $N = 5,877$ for Posco, and $N = 11,313$ for Pfizer stock prices. Let $\beta = 5\%$ be the significance level of the F test. The null hypothesis is $H_0 : \sigma_1^2 = \sigma_2^2$ which is rejected if $F < F_{1-\beta/2}$ or $F > F_{\beta/2}$, where $F_{1-\beta/2} = F(1-\beta/2, N-1, N-1)$ and $F_{\beta/2} = F(\beta/2, N-1, N-1)$ are the critical values of the F distribution with $N-1$ degrees of freedom and significance level equal to the subscript. Table II gathers the critical values and the F statistic. The F test indicates that there is evidence to reject the null hypothesis of equal variances for CKF and the proposed RCKF at the 0.05 level of significance for BP, Pfizer, and Posco stocks, but not for Coca-Cola.

TABLE II: F-test.

Stock	$F_{1-\beta/2}$	$F_{\beta/2}$	F
BP	0.9618	1.0397	3.2044
Coca-Cola	0.9673	1.0338	1.0279
Pfizer	0.9638	1.0375	4.3565
Posco	0.9501	1.0525	1.7525

VI. Conclusion

A robust collaborative Kalman filter for time-evolving dyadic processes has been proposed, which employs a heavy-tailed measurement noise. The aforementioned modification partially rectifies the over-simplistic assumption that stock prices are treated as Gaussian random variables. The latent state vectors are shown to be able to track the volatility of stock prices through time. The experimental results have demonstrated that the efficient modeling of the measurement noise variance pays off, yielding a more accurate stock price prediction than the original collaborative Kalman filter.

Future research would focus on nonlinear stock price models, where extended Kalman filters, unscented Kalman filters, or particle filters would be more suitable. Furthermore, normalized innovation squared and normalized deviation squared consistency tests would reinforce the detection of inconsistent filter behavior and the loss of tracking for either linear or non-linear Kalman filters. Besides stock price prediction, the proposed method has been successfully applied to Netflix movie rating prediction.

References

[1] A. Mnih and R. R. Salakhutdinov, “Probabilistic matrix factorization,” in *Advances in Neural Information Processing Systems*, 2008, pp. 1257–1264.

[2] R. Salakhutdinov and A. Mnih, “Bayesian probabilistic matrix factorization using Markov chain Monte Carlo,” in *Proc. 25th Int. Conf. Machine Learning*. ACM, 2008, pp. 880–887.

[3] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, vol. 42, no. 8, 2009.

[4] 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.

[5] E. F. Fama, “Mandelbrot and the stable Paretian hypothesis,” *The Journal of Business*, vol. 36, no. 4, pp. 420–429, 1963.

[6] G. Agamennoni, J. I. Nieto, and E. M. Nebot, “Approximate inference in state-space models with heavy-tailed noise,” *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5024–5037, 2012.

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

[8] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[9] 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.