A Novel Algorithmic Trading Strategy Using Data-Driven Innovation Volatility

You Liang
Department of Mathematics
Ryerson University
Toronto, Canada
you.liang@ryerson.ca

Aerambamoorthy Thavaneswaran
Department of Statistics
University of Manitoba
Winnipeg, Canada
thavane@umanitoba.ca

Md. Erfanul Hoque
Department of Statistics
University of Manitoba
Winnipeg, Canada
hoqueme@myumanitoba.ca

***Abstract*—The explosion of algorithmic trading has been one of the most prominent recent trends in the finance industry. Regularized estimating functions including Kalman filtering (KF) allow dynamic data scientists and algo traders to enhance the predictive power of statistical models and improve trading strategies. Recently there has been a growing interest in using KF in pairs trading. However, a major drawback is that the innovation volatility estimate calculated by using a KF algorithm is always affected by the initial values and outliers. A simple yet effective data-driven approach to estimate the innovation volatility with some robustness properties is presented in this paper. The results show that the performance of the trading strategy based on the data-driven innovation volatility forecast (DDIVF) is better than the commonly used KF-based innovation volatility forecast (KFIVF). Autocorrelations of the absolute values of the innovations in multiple trading are used to demonstrate that the innovations are non-normal with time-varying volatility. We describe and analyze experiments on three cointegrated exchange-traded funds (ETFs) and explain how our approach can improve the performance of the trading strategies. A proposed novel trading strategy for multiple trading with robustness to initial values and to the volatile stock market is also discussed in some detail by using a training sample and a test sample.**

***Index Terms*—Pairs Trading, Multiple Trading, Kalman Filter, Data-Driven Volatility, Robustness, Volatile Market**

I. Introduction

Supervised learning ([1]) is the most widely utilized form of machine learning. Its goal is to predict the response from the associated features. Regularization ([2]) puts extra constraints on a machine learning model, and these constraints and penalties are designed to encode specific kinds of prior knowledge. Consider the linear regression model

$$\boldsymbol{y} = \boldsymbol{A\theta} + \boldsymbol{\epsilon} \tag{1}$$

with no intercept, where $\boldsymbol{y}$ is the $p \times 1$ vector of responses, $\mathbf{A}$ is the $p \times m$ matrix of standardized features, and $\boldsymbol{\epsilon}$ is the $p \times 1$ vector of independent and identically distributed (i.i.d) normal errors. The regularized least squares estimate minimizes the sum of the objectives

$$J_1 = ||\boldsymbol{y} - \boldsymbol{A\theta}||_2^2, J_2 = p_\lambda(\boldsymbol{\theta}),$$

where J_1 is the minimization objective for the regression residuals, J_2 is the minimization objective for the prior information, and λ is a tuning parameter. Lasso estimates are viewed as L_1-penalized least squares estimates with penalty $J_2 = \lambda||\boldsymbol{\theta}||_1$.

978-1-7281-2547-3/20/$31.00 ©2020 Crown

As an extension to the dynamic setup, we consider the filtering problems and state space models ([3]). Kalman filtering (KF), non-normal filtering and partially Bayes maximum informative filtering have algorithmic trading applications in quantitative finance ([4], [5]). In dynamical learning, KF offers a computationally efficient recursive procedure to learn the dynamical systems using prior knowledge.

First the information based estimating function (EF) approach is introduced to study the robust filtering problems. Consider a probability space $(\Omega, \mathcal{F}, P)$, on which $\boldsymbol{y}$ and $\boldsymbol{\theta}$ are jointly distributed random variables, and $\boldsymbol{\theta}$ is real valued. An EF for $\boldsymbol{\theta}$ is a real valued function, denoted by $\boldsymbol{g}(\boldsymbol{y};\boldsymbol{\theta})$, and it is unbiased if $E[\boldsymbol{g}(\boldsymbol{y};\boldsymbol{\theta})] = \mathbf{0}$. The information matrix associated with $\boldsymbol{g}$ is defined by

$$I_{\boldsymbol{g}} = E[\boldsymbol{gg}'] = (E[\partial\boldsymbol{g}/\partial\boldsymbol{\theta}])'(E[\boldsymbol{gg}'])^{-1}(E[\partial\boldsymbol{g}/\partial\boldsymbol{\theta}]).$$

[6]–[8] illustrate the combined EFs approach in a number of linear non-Gaussian process filtering problems in the scalar case. Filters are obtained as the solutions of maximum informative estimating equations. Recently, [9] introduced the penalized EF approach by including a penalty in the combined (linear and quadratic) EF obtained in [10] and studied the penalized estimate of $\boldsymbol{\theta}$ for logarithmic autoregressive conditional duration models. However, the resulting combined EF with a penalty added in [9] is biased. In this paper, an *unbiased* Bayesian regularized EF is defined as a combination of the optimal EF based on the observed process and the optimal EF of the prior process. Consider a dynamic model for a filtering problem:

$$y_t = A_t\theta_t + \epsilon_t,$$

where ϵ_t is an independent sequence of zero mean random variables with density $f(\cdot)$, y_t is an observed sequence of variables, and $\theta_t|\theta_{t-1}$ has density $\lambda_t(\cdot)$. For example, ϵ_t might have a heavy-tailed distribution such as the Laplace distribution or the Cauchy distribution. In this case, it is impossible to obtain a simple recursive relation for the posterior mean. However, we can take Godambe's formulation (see [8]) as a starting point and investigate a combination of orthogonal EFs. It can be shown that if θ_{t-1} were known the optimal regularized filtered estimate $\hat{\theta}_t$ is obtained as the solution of the combined EFs,

$$\frac{\partial \ln f(y_t - A_t\theta_t)}{\partial \theta_t} + \frac{\partial \ln \lambda_t(\theta_t|\theta_{t-1})}{\partial \theta_t} = 0. \quad (2)$$

If f and λ are normal densities then the filtered estimate $\hat{\theta}_t$ turns out to be posterior mean as in KF for state space models. It can also be shown that the optimal linear predictor could also be obtained as a solution of the the unbiased filtering equation

$$-\frac{A_t(y_t - A_t\theta_t)}{\sigma_\eta^2} + \frac{\theta_t - \mu_t}{\sigma^2} = 0, \quad (3)$$

where μ_t is the conditional mean of θ_t. Moreover, if f is the normal density and $\lambda_t(\cdot)$ is the density of a symmetric distribution with mean zero, sign correlation ρ (see [11]) and variance σ^2, the optimal linear predictor could be obtained as a solution of the unbiased filtering equation

$$-\frac{A_t(y_t - A_t\theta_t)}{\sigma_\eta^2} + \frac{\text{sign}(\theta_t - \mu_t)(|\theta_t - \mu_t| - \rho\sigma)}{\sigma^2(1-\rho^2)} = 0, \quad (4)$$

which can be rewritten as

$$-\frac{A_t(y_t - A_t\theta_t)}{\sigma_\eta^2} + \frac{\theta_t - \mu_t}{\sigma^2(1-\rho^2)} - \frac{\rho}{\sigma(1-\rho^2)}\text{sign}(\theta_t - \mu_t) = 0.$$

As a special case, if we further assume that $\lambda_t(\cdot)$ is the density of a Laplace distribution then we can interpret the dynamic Lasso estimate as a posterior mode for each t. In general, (4) can be interpreted as a generalized unbiased filtering equation with symmetric priors and it is more informative than the least squares EF. The corresponding estimate can be interpreted as a dynamic generalized Bayesian least square estimate.

Algorithmic trading ([12]–[16]) uses a computer program that follows a defined set of instructions (an algorithm) to place a trade and can generate profits at a speed and frequency that is impossible for a human trader. It is rarely in the best interest of investment managers to share profitable trading strategies with the public, so most trading strategies including pairs trading remained a secret of the investors until the introduction of online trading. Pairs trading is a trading strategy used to exploit financial markets that are out of equilibrium. The strategy involves identifying two securities whose prices tend to move together in the long term. Upon divergence, the cheaper security is bought long and the more expensive one is sold short. When prices converge back to their historical equilibrium, the trade is closed and a profit collected. Pairs trading has been introduced to the academic community through [17]. The idea behind a pair (of stocks, bonds, foreign exchanges, commodities, etc.) is closely linked to the statistical concept of cointegration. If a linear combination of a collection of nonstationary time series is stationary, then the collection is said to be cointegrated. For cointegrated prices $P_{1,t}$ and $P_{2,t}$, the difference or spread of two prices, $\epsilon_t = P_{1,t} - \theta_0 - \theta_1 P_{2,t}$, is stationary, which suggests that ϵ_t is perturbed around an equilibrium value. In pairs trading, the regression coefficient θ_1 is called the hedge ratio, and it describes the amount of one security to purchase or sell for every unit of the other security. In a non-dynamic setting, we use in-sample data to obtain offline estimates $\hat{\theta}_0$ and $\hat{\theta}_1$ for regression coefficients θ_0 and θ_1, and calculate an estimate $\hat{\sigma}$ of the standard deviation of ϵ_t. In the trading period, the z-score z_t is computed as

$$z_t = \nu_t/\hat{\sigma} = (P_{1,t} - \hat{\theta}_0 - \hat{\theta}_1 P_{2,t})/\hat{\sigma},$$

which is used to generate trading signals. Further studies describe the price difference of a pair in the state space formulation ([18], [19]). In order to incorporate the time varying regression coefficients ([4], [5]), and extend pairs trading to multiple trading ([20]), the linear state space model or dynamic linear model can be used. The state space model employs a random walk as the state equation:

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{v}_t, \quad (5)$$

where $\boldsymbol{\theta}_t$ is the m-dimensional state vector at time t, and $\boldsymbol{v}_t$ is i.i.d with mean zero and covariance matrix $\Sigma_{\boldsymbol{v}}$. An observed process y_t can be described by an observation equation:

$$y_t = \boldsymbol{A}_t\boldsymbol{\theta}_t + \epsilon_t, \quad (6)$$

where $\boldsymbol{A}_t$ is a m-dimensional feature, and the observational noise ϵ_t is i.i.d with mean zero and variance σ_ϵ^2. A primary aim of the analysis is to produce dynamic filtered estimates, $\widehat{\boldsymbol{\theta}}_{t|t} = E[\boldsymbol{\theta}_t|\mathcal{F}_t^y]$, for the hedge ratio $\boldsymbol{\theta}_t$ to hedge the risk exposure of the stock price movement, given the data $\mathcal{F}_t^y = \{y_1, \ldots, y_t\}$ up to time t. Using the filtered estimate $\widehat{\boldsymbol{\theta}}_{t-1|t-1}$, $\nu_t = y_t - \boldsymbol{A}_t\widehat{\boldsymbol{\theta}}_{t-1|t-1}$ is called the innovation at time t. The innovation sequence ν_t and its time varying volatility are used to generate trading signals in algorithmic trading.

Recently there has been a growing interest in pairs trading and multiple trading based on Kalman filtering. In the literature [4], [5], [20] among others, very small initial values of the KF are used. Trading profit is sensitive to initial values, and it decreases sharply when initial values are slightly increased. In this paper, a novel data-driven robust filtering algorithm based on regularized EFs is proposed for multiple trading, which does not need to assume very small initial values. It is shown that the commonly used square root of the innovation variance is not an appropriate estimator of the innovation volatility (see [11] for details). A data-driven trading strategy based on joint forecasts of volatility and stock price is studied in [21]. The data-driven generalized exponential weighted moving average (DD-EWMA) volatility forecasting model proposed in [11] is used to forecast the innovation volatility directly in this paper. The data-driven innovation volatility forecast (DDIVF) provides accurate dynamic interval forecasts of innovations and can be used to generate the trading signals appropriately. Let the conditional variance of the innovation ν_t, based on the past data up to time $t-1$, be σ_t^2. The DD-EWMA volatility forecasting model for innovations is given by

$$\hat{\sigma}_t = (1-\alpha)\,\hat{\sigma}_{t-1} + \alpha\frac{|\nu_{t-1} - \bar{\nu}|}{\hat{\rho}_\nu}, \quad 0 < \alpha < 1, \quad (7)$$

where α is the smoothing constant, and $\hat{\rho}_\nu$ is the sample sign correlation of the innovation sequence, defined as $\text{Corr}(\nu_t - \bar{\nu}, \text{sgn}(\nu_t - \bar{\nu}))$. Model (7) is data-driven in the sense that the optimal value of α is obtained by minimizing the one-step ahead forecast error sum of squares (FESS), and the sample sign correlation $\hat{\rho}_\nu$ is used to identify the conditional distribution of ν_t. In this paper, this model is used and extended to study the volatility forecasts for innovation and improve the stability of filtering algorithm.

The remainder of this paper is organized as follows. In Section II, a maximum informative filtering algorithm is proposed with DDIVF. In Section III, a data-driven multiple trading strategy using maximum informative filtered hedge ratios and DDIVF is proposed. Trading strategies constructed using the DDIVF performs better than the commonly used KF-based innovation volatility forecast (KFIVF). The robustness of these two strategies are analyzed and compared using a training sample and a test sample. The trading strategy using DDIVF is robust to a wide range of initial values, and robust to the volatile stock market, since the time varying innovation volatility is properly investigated. Finally, Section IV provides conclusions.

II. Methods

We consider multiple trading for stocks with multiple cointegrations and construct a novel trading strategy using dynamic maximum informative filtering. Consider m asset prices $P_{1,t}, P_{2,t}, \dots, P_{m,t}$ with a multiple cointegrated relationship. The state space model (5) - (6) is used where $\boldsymbol{\theta}_t = (\theta_{0,t}, \theta_{1,t}, \dots, \theta_{m-1,t})'$, $y_t = P_{1,t}$ and $\boldsymbol{A}_t = (1, P_{2,t}, \dots, P_{m,t})$. In addition, it is assumed for simplicity that $\boldsymbol{\theta}_0, \boldsymbol{v}_t$ and ϵ_t are uncorrelated.

A. Data-Driven Maximum Informative Filters Using Estimating Functions

For model (5) - (6), let $\widehat{\boldsymbol{\theta}}_{t-1|t-1} = \mathrm{E}[\boldsymbol{\theta}_{t-1}|\mathcal{F}^y_{t-1}]$ and $I^{-1}_{t-1|t-1} = P_{t-1|t-1} = \mathrm{Var}(\boldsymbol{\theta}_{t-1} - \widehat{\boldsymbol{\theta}}_{t-1|t-1}|\mathcal{F}^y_{t-1})$. Based on the non-Gaussian maximum informative filter provided in [8], we consider the combination of two elementary EFs:

$$\boldsymbol{g}_{1t} = \boldsymbol{\theta}_t - \mathrm{E}[\boldsymbol{\theta}_t|\mathcal{F}^{\boldsymbol{y}}_{t-1}] = \boldsymbol{\theta}_t - \widehat{\boldsymbol{\theta}}_{t-1|t-1},$$

and

$$g_{2t} = \nu_t = y_t - \mathrm{E}[\boldsymbol{y}_t|\mathcal{F}^y_{t-1}] = y_t - \boldsymbol{A}_t\widehat{\boldsymbol{\theta}}_{t-1|t-1},$$

where $\nu_t = y_t - \widehat{y}_{t|t-1}$ is the innovation or forecast error of y_t. The "optimal" combination in the class of linear combinations of $\boldsymbol{g}_{1t}$ and ν_t is given by

$$\boldsymbol{\theta}_t - \widehat{\boldsymbol{\theta}}_{t-1|t-1} - \frac{\mathrm{Cov}(\boldsymbol{g}_{1t}, \nu_t|\mathcal{F}^y_{t-1})}{\mathrm{Var}(\nu_t|\mathcal{F}^y_{t-1})}(y_t - \boldsymbol{A}_t\widehat{\boldsymbol{\theta}}_{t-1|t-1}). \quad (8)$$

This yields the "optimal" estimate of $\boldsymbol{\theta}_t$ as

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1|t-1} + (I^{-1}_{t-1|t-1} + \Sigma_{\mathbf{v}})\boldsymbol{A}'_t Q_t^{-1}(y_t - \boldsymbol{A}_t\widehat{\boldsymbol{\theta}}_{t-1|t-1}), \quad (9)$$

where the innovation variance is given by

$$Q_t = \mathrm{Var}(\nu_t|\mathcal{F}^y_{t-1}) = \mathbf{A}_t(I^{-1}_{t-1|t-1} + \Sigma_{\mathbf{v}})\mathbf{A}'_t + \sigma^2_\epsilon.$$

In most of the applications including pairs trading and risk forecasting, the filtered estimate $\widehat{\boldsymbol{\theta}}_{t-1|t-1}$ for the state variable, the innovation ν_t and the innovation volatility $\sqrt{Q_t}$ are used. However, $\sqrt{Q_t}$ is not an appropriate estimate of the innovation volatility. Therefore, DD-EWMA volatility forecasting model (7) is used to obtain DDIVF, and Algorithm 1 illustrates the details of DDIVF calculation. Based on the past k innovations $\nu_{t-k}, \dots, \nu_{t-1}$, sample sign correlation $\hat{\rho}_\nu$ and volatility estimate $|\nu_s - \bar{\nu}|/\hat{\rho}_\nu, s = t-k, \cdots, t-1$ are calculated. The smoothed value S_s of the volatility estimate is calculated recursively. The optimal smoothing constant α_{opt} is determined by minimizing the one-step ahead FESS. Using the optimal value α_{opt}, we calculate the smoothed value S_s recursively. Finally, S_{t-1} is computed, and used as the volatility forecast $\hat{\sigma}_t$ for ν_t.

Algorithm 1 Dynamic DD-EWMA volatility forecasts of innovation

Require: Predicted errors $\nu_s, s = t-k, \cdots, t-1$

$\hat{\rho}_\nu \leftarrow \mathrm{Corr}(\nu_s - \bar{\nu}, \mathrm{sign}(\nu_s - \bar{\nu}))$

$V_s \leftarrow |\nu_s - \bar{\nu}|/\hat{\rho}_\nu$ {Compute estimated volatility}

$S_{t-k-1} \leftarrow \bar{V}_l$ {Initial volatility forecast using first l observations}

$\alpha \leftarrow (0.01, 0.5)$ *by* 0.01{Set a range for α}

$S_s \leftarrow \alpha * V_s + (1-\alpha) * S_{s-1}, s = t-k, \dots, t-1$

$\alpha_{opt} \leftarrow \min_\alpha \sum_{s=t-k+l}^{t-1}(V_s - S_{s-1})^2$ {Determine optimal α by minimizing FESS}

$S_s \leftarrow \alpha_{opt} * V_t + (1-\alpha_{opt}) * S_{s-1}, s = t-k, \dots, t-1$

$\hat{\sigma}_t \leftarrow S_{t-1}$ {Calculate one-step-ahead DDIVF based on k observations}

return $\alpha_{opt}, \hat{\sigma}_t$

It follows from [6] and [8] that the information matrix associated with the combined EF (8) is maximal in the class of linear combination of $\boldsymbol{g}_{1t}$ and ν_t. When Gaussian assumptions hold for $\boldsymbol{v}_t$ and ϵ_t, the linear optimal filter (9) turns out to be KF. The point estimation can be regarded as recursive if $\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t|t} = \mathrm{E}[\boldsymbol{\theta}_t|\mathcal{F}^y_t]$. We cannot conclude this in general, though we conclude that $\widehat{\boldsymbol{\theta}}_t$ is $\mathcal{F}^y_t$-measurable and $\mathrm{E}[\widehat{\boldsymbol{\theta}}_t|\mathcal{F}^y_{t-1}] = \mathrm{E}[\boldsymbol{\theta}_t|\mathcal{F}^y_{t-1}]$. The following dynamic maximum informative filtering algorithm is used to calculate dynamic hedge ratios recursively. The algorithm updates $I_{t|t}$, which is computationally simpler than updating the covariance matrix $P_{t|t}$. The updating formulas for $I_{t|t}$ is equivalent to the updating formulas for $P_{t|t}$, which is given by $P_{t|t} = (I - I^{-1}_{t|t-1}\boldsymbol{A}'_t Q_t^{-1}\boldsymbol{A}_t)P_{t|t-1}$.

The innovation ν_t, and the standard deviation $\sqrt{Q_t}$ or the DDIVF $\hat{\sigma}_t$ of can be used to construct the signals for a trading strategy at each time t. [5] discussed a pairs trading strategy, and [20] proposed a multiple trading strategy using $\sqrt{Q_t}$. The first two values of ν_t are relatively large since the filter needs a few iterations before stabilization. Without outliers, ν_t follows a normal distribution approximately. Hence, the dynamic z-score z_t is computed as

$$z_t = \nu_t/\sqrt{Q_t}, \quad (10)$$

and the z-scores will be compared with a threshold value p to generate trading signals. The strategies using ν_t and $\sqrt{Q_t}$

Algorithm 2 Dynamic maximum informative filtered hedge ratios

Require: Data: adjusted closing stock prices $P_{1,t}, P_{2,t}, \ldots, P_{m,t}, t = 1, \ldots, n$
Let $y_t = P_{1,t}, \boldsymbol{A}_t = (1, P_{2,t}, \ldots, P_{m,t})$
Initialization: initial state $\boldsymbol{\theta}_0$, initial error covariance matrix $I_{0|0} = \Sigma_0^{-1}$, constant error covariance matrix $\Sigma_{\mathbf{v}}$, constant innovation variance σ_ϵ^2
for $t \leftarrow 1, \ldots, n$ **do**
  Prediction: Based on data available at $t-1$:
  $\widehat{\boldsymbol{\theta}}_{t|t-1} \leftarrow \widehat{\boldsymbol{\theta}}_{t-1|t-1}; I_{t|t-1}^{-1} \leftarrow I_{t-1|t-1}^{-1} + \Sigma_{\mathbf{v}}; \hat{y}_{t|t-1} \leftarrow \boldsymbol{A}_t \widehat{\boldsymbol{\theta}}_{t|t-1}$
  Update: Inference about $\boldsymbol{\theta}_t$ is updated using the observation y_t at time t
  $\nu_t \leftarrow y_t - \hat{y}_{t|t-1}; Q_t \leftarrow \boldsymbol{A}_t I_{t|t-1}^{-1} \boldsymbol{A}_t' + \sigma_\epsilon^2$
  DDIVF $\hat{\sigma}_t$ is calculated based on $\nu_{t-k}, \ldots, \nu_{t-1}$ using Algorithm 1
  $\widehat{\boldsymbol{\theta}}_{t|t} \leftarrow \widehat{\boldsymbol{\theta}}_{t|t-1} + I_{t|t-1}^{-1} \boldsymbol{A}_t' Q_t^{-1} \nu_t; I_{t|t} = I_{t|t-1} + \frac{1}{\sigma_\epsilon^2} \boldsymbol{A}_t' \boldsymbol{A}_t$
end for
return ν_t, Q_t, $\hat{\sigma}_t$

require the initial values of $\Sigma_{\boldsymbol{v}}$ and σ_ϵ^2 to be very small since increased initial values will cause the problem of volatility clustering of ν_t and it is not appropriate to directly use $\sqrt{Q_t}$ as the volatility estimate.

B. A Novel Data-Driven Robust Multiple Trading Strategy

In the literature [4], [5], [20], trading profit is sensitive to initial values. Therefore, a robust multiple trading strategy using $\widehat{\boldsymbol{\theta}}_{t-1|t-1}$, ν_t and DDIVF $\hat{\sigma}_t$ is proposed and compared with the multiple trading strategy using $\sqrt{Q_t}$ to demonstrate the profitability and robustness of proposed strategy. The dynamic robust z-score z_t is computed as

$$z_t = \nu_t / \hat{\sigma}_t, \tag{11}$$

and the z-scores will be compared with a threshold value p to generate trading signals. The strategy using equation (10) always requires very small initial values of $\Sigma_{\boldsymbol{v}}$ and σ_ϵ^2 due to the convergence issue and to guarantee a successful trading strategy. However, the proposed strategy using equation (11) doesn't require such assumptions for $\Sigma_{\boldsymbol{v}}$ and σ_ϵ^2. The spread $\nu_t = P_{1,t} - \hat{\theta}_{0,t-1|t-1} - \hat{\theta}_{1,t-1|t-1} P_{2,t} - \cdots - \hat{\theta}_{m-1,t-1|t-1} P_{m,t}$ is modelled as a mean-reverting process. Upon divergence, the cheaper security (or linear combination of securities) is bought long and the more expensive security (or linear combination of securities) is sold short. When the prices converge back to their historical equilibrium, the trade is closed and a profit is collected. Algorithm 3 generates the trading signals s_t, where sells are represented as $s_t = -1$, buys as $s_t = 1$, and no signal as $s_t = 0$. The buy signal is generated when z_t crosses a threshold p from above; the sell signal is generated when z_t crosses p from below. Then trading positions are determined using s_t, and the profit of holding those positions is further computed. Finally, annualized Sharpe ratios (ASRs) are calculated for a range of values of p. The optimal threshold value, p_{opt}, is determined by maximizing the ASR; the optimal cumulative profit is computed from optimal signals and positions generated by using p_{opt}.

Algorithm 3 Robust Multiple Trading using rolling DDIVF and optimal signals

Require: Data: $P_{1,t}, P_{2,t}, \ldots, P_{m,t}, t = 1, \ldots, n$; rolling window size k
Let $y_t = P_{1,t}, \boldsymbol{A}_t = (1, P_{2,t}, \ldots, P_{m,t})$; ν_t and $\widehat{\boldsymbol{\theta}}_{t-1|t-1}$ obtained from Algorithm 2
$\hat{\sigma}_t, t = k+1, \ldots, n$ is obtained by using Algorithm 1 using a rolling approach based on $\nu_t, t = 1, \ldots, n$. Each rolling window size is k
$z_t \leftarrow \nu_t / \hat{\sigma}_t, t = k+1, \ldots, n$
Generate trading signals s_t:
for $t \leftarrow k+2, \ldots, n$ **do**
  If $z_{t-1} < p$ & $z_t > p$, then $s_t \leftarrow -1$; If $z_{t-1} > -p$ & $z_t < -p$, then $s_t \leftarrow 1$; Else $s_t \leftarrow 0$
  $position.A_t \leftarrow -1000 * \widehat{\boldsymbol{\theta}}_{t-1|t-1} * s_t$; $position.y_t \leftarrow 1000 * s_t$
  $profit.A_t \leftarrow (\boldsymbol{A}_t - \boldsymbol{A}_{t-1}) * position.A_t$; $profit.y_t \leftarrow position.y_t * (y_t - y_{t-1})$
  $profit_t \leftarrow profit.A_t + profit.y_t$
end for
Calculate the ASR as $SR(p) = \sqrt{252} * mean(profit_t)/sd(profit_t)$
Determine the optimal value of p, p_{opt} , by maximizing $SR(p)$
Obtain the cumulative profit $cumsum(profit_t)$ using p_{opt}
return p_{opt}, cumulative profit

III. Results

In this section we test the trading strategies constructed using DDIVF against those constructed using KFIVF ($\sqrt{Q_t}$ from KF algorithm), and explore the robustness of the profit and ASR. The proposed method and algorithms are tested with the adjusted closing prices of three exchange-traded funds (ETFs) downloaded from **Yahoo Finance** for the period from 2017-02-01 to 2020-03-15: iShares MSCI Australia ETF (EWA) and iShares MSCI Canada ETF (EWC), and iShares North American Natural Resources ETF (IGE). From the whole period of time, we selected the training sample from 2017-02-01 to 2019-03-14 and the test sample from 2019-03-15 to 2020-03-15. We use EWA and EWC to illustrate the proposed robust trading strategy. The price movements of the these two stocks are visualized in Fig. 1 for the training sample and the test sample, respectively. The training sample is used to obtain p_{opt} for each chosen values of σ_ϵ^2 and δ in Table III. Then the test sample is used to test the profitability and robustness of the proposed robust strategy. It is known that March 2020 was a historically volatile month for the stock market. The proposed strategy is demonstrated to be profitable and robust during this period until these two stocks did not exhibit the cointegration relationship since April 2020.

The cointegration of the two stocks are regularly checked by Engle-Granger test and Johansen test over time.

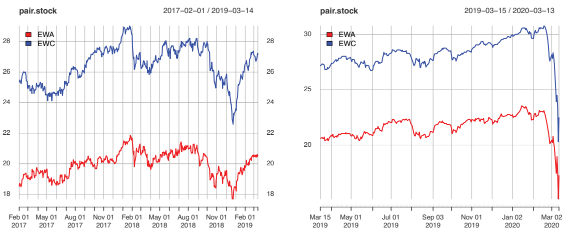

Fig. 1. Daily adjusted closing prices of EWA and EWC

We consider the state space model (5) - (6) where $\boldsymbol{\theta}_t = (\theta_{0,t}, \theta_{1,t}), P_{1,t} = \theta_{0,t} + \theta_{1,t}P_{2,t} + \epsilon_t$. $P_{1,t}$ is EWC, and $P_{2,t}$ is EWA. The initial state, $\boldsymbol{\theta}_0 = \mathbf{0}$, and the initial information matrix I_0 is chosen such that I_0^{-1} is a zero matrix. The innovation covariance $\sigma_\epsilon^2 = 0.001$, and the error covariance matrix $\Sigma_\mathbf{v}$ is a 3×3 diagonal matrix with elements $\delta/(1-\delta)$, where $\delta = 0.0001$. The dynamics of the filtered intercept (green) and hedge ratio for EWA (blue) are shown in Fig. 2. The small initial values of the innovation covariance and the error covariance matrix will guarantee a successful trading strategy by using ν_t and Q_t as in [5] and [20]. However, this condition is not required for the proposed robust trading strategy using DDIVF. We will discuss this in the following sections.

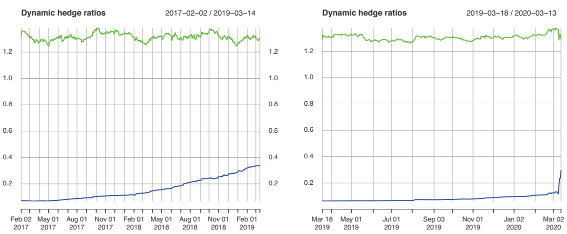

Fig. 2. Filtered $\hat{\theta}_{0,t|t}$ (green) and $\hat{\theta}_{1,t|t}$ (blue)

A. *Robust Multiple trading strategy using ν_t and DDIVF*

A rolling window approach is first applied to the training sample to forecast the volatility of ν_t and obtain p_{opt} to be used for the test sample. The selected data covers 532 days, with 432 overlapping rolling windows of size 100 days. Each window of size 100 is used to calculate a one-day-ahead DDIVF using Algorithm 1, and the corresponding z-score using (11). For example, $\nu_1, \ldots, \nu_{100}$ are used to calculate the volatility forecast $\hat{\sigma}_{101}$ for ν_{101}, and $z_{101} = \nu_{101}/\hat{\sigma}_{101}$. Then the z-scores are compared with a threshold value p to generate trading signals. Using Algorithm 3, the range of p is chosen as $(0.1, 2)$ with an increment of 0.01. The optimal value is determined as $p_{opt} = 1.68$. The corresponding optimal trading signals are visualized in Fig. 4, where the red lines are the bounds calculated by $p_{opt}\hat{\sigma}_t$. Fig. 3 compares DDIVF and KFIVF, and it is shown that KFIVF is very similar to the average of DDIVF and is not able to capture the time varying innovation volatility. Therefore the bounds calculated by $p_{opt}\sqrt{Q_t}$ where $p_{opt} = 1.48$ from the traditional pairs trading is not providing the online guide for signals, as shown in Fig. 5. We calculate the positions in each asset according to the spread and signals using $p_{opt} = 1.68$. The look-ahead bias is eliminated by lagging the signals. Each trade consists of 1,000 units of the spread. The estimated profit is the sum of the price differences multiplied by the corresponding positions in each asset. The cumulative profit of the robust pairs trading is \$7024.441.

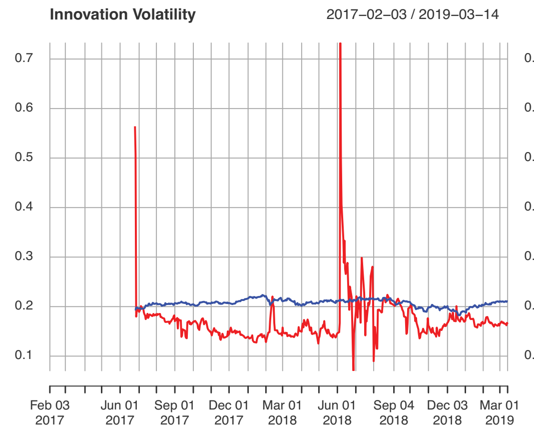

Fig. 3. DDIVF $\hat{\sigma}_t$ vs. KFIVF $\sqrt{Q_t}$: 2017-02-01 to 2019-03-14

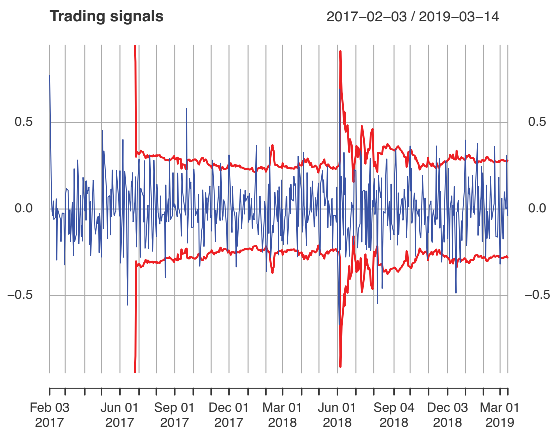

Fig. 4. Robust pairs trading using $\hat{\sigma}_t$: 2017-02-01 to 2019-03-14

The pair EWA and IGE is used to construct another robust pairs trading strategy, where IGE is the response variable and EWA is the feature. EWA, EWC and IGE are used

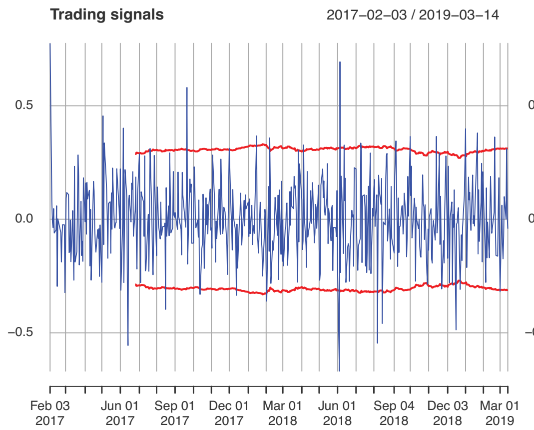

Fig. 5. Pairs trading using $\sqrt{Q_t}$: 2017-02-01 to 2019-03-14

to construct a robust multiple trading strategy with IGE as the response variable and EWA and EWC as features. The summary statistics of innovation ν_t are provided in Table I. The results of all trading strategies are summarized in Table II. Table II compares each robust multiple trading strategy with the traditional multiple trading strategy and the buy and hold strategy. For each collection of stocks, the optimal threshold p_{opt}, cumulative profit and ASR for the robust one is provided in columns 2 to 4, respectively. Those values for the traditional one are provided in columns 5 to 7, respectively. The last column provides the profit from the buy and hold strategy. Each data-driven robust multiple trading strategy outperforms (with a higher profit and ASR) than the corresponding traditional one and the buy and hold without transaction costs. The value p_{opt} is essential to guarantee an appropriate number of signals since we trade at the daily adjusted closing price with no market impact or slippage, and trade for free. The most profitable strategy in theory is not necessarily the best one in practice since (1) cost is increased with the increased number of stocks; (2) strategy with relatively smaller p_{opt} will generate more signals with a higher transaction cost. In practice, the used R code for trading algorithms can be modified to fit into a live trading platform such as Zorro (see [5]).

TABLE I
SUMMARY STATISTICS OF ν_t: 2017-02-01 TO 2019-03-14

	$\hat{\rho}_\nu$	$\hat{\rho}_\nu^*$	acf-ν_t	acf-$\lvert\nu_t\rvert$	acf-ν_t^2
EWA/EWC	0.179	0.801	-0.0362	0.0023	-0.0233
EWA/IGE	0.191	0.789	0.0262	0.0041	0.0119
EWA/EWC/IGE	0.178	0.793	-0.0299	-0.0048	0.0183

$\hat{\rho}_\nu$: sample sign correlation of ν_t; $\hat{\rho}_\nu^*$: sample sign correlation of ν_t without outliers; acf: Lag 1 sample autocorrelation

B. Stability Analysis of Initial Values for Pairs Trading

It follows from Table II that the data-driven robust multiple strategies are profitable. Each one outperforms the corresponding traditional multiple trading strategy and the buy and hold strategy without transaction costs. A stability analysis is further conducted for the initial values of pairs trading using EWA and EWC for the training sample. The robust pairs trading strategy using DDIVF is stable when the initial value of σ_ϵ^2 and the error covariance matrix $\Sigma_{\boldsymbol{v}}$ are increased. However, the traditional pairs trading strategy requires very small initial values. In the following Table III, we compare the optimal p_{opt}, profit and ASR between the proposed robust pairs trading strategy and the traditional one, according to various initial values of σ_ϵ^2 and δ. The traditional pairs trading strategy fails gradually when the initial values are increased, however; our strategy is robust to a wide range of initial values. The traditional pairs trading works well only when $\delta = 0.0001$ and $\sigma_\epsilon^2 = 0.001$. However, our proposed robust data-driven pairs trading strategy works consistently well for various initial values. The robust data-driven pairs trading strategy works consistently well for various initial values. The results of cumulative profits and ASRs are stable when σ_ϵ^2 ranges from 0.001 to 10, and δ ranges from 0.0001 to 0.05, as shown in Table III.

TABLE II
OPTIMAL ROBUST AND TRADITIONAL MULTIPLE TRADING STRATEGIES: 2017-02-01 TO 2019-03-14

	Robust multiple trading			Traditional multiple trading			
	p_{opt}	ASR	Profit	p_{opt}	ASR	Profit	B/H Profit
EWA/EWC	1.68	1.694	7024.441	1.48	1.520	6319.299	3798.899
EWA/IGE	1.73	1.113	8105.911	1.5	-0.013	-98.021	2555.466
EWA/EWC/IGE	1.85	1.896	10645.81	1.08	1.472	8530.784	4768.198

ASR: annualized Sharpe ratio; B/H: buy and hold

TABLE III
STABILITY ANALYSIS OF INITIAL VALUES: 2017-02-01 TO 2019-03-14

		Robust pairs trading			Traditional pairs trading		
σ_ϵ^2	δ	p_{opt}	ASR	Profit	p_{opt}	ASR	Profit
0.001	0.0001	1.68	1.694	7024.441	1.48	1.520	6319.299
0.001	0.001	1.65	1.694	7023.8	0.47	1.520	6317.67
0.001	0.005	1.66	1.694	7023.737	0.21	1.638	7304.173
0.001	0.01	1.66	1.694	7023.729	0.15	1.520	6317.493
0.001	0.05	1.66	1.694	7023.723	0.1	0.477	1066.272
0.01	0.0001	1.72	2.075	8085.982	1.21	1.681	6991.076
0.01	0.001	1.68	1.694	7024.441	0.34	1.509	6787.412
0.01	0.005	1.65	1.694	7023.877	0.21	1.520	6317.862
0.01	0.01	1.65	1.694	7023.8	0.15	1.520	6317.668
0.01	0.05	1.66	1.694	7023.737	0.1	0.477	1066.279
0.05	0.0001	1.57	1.439	5646.207	0.81	1.910	8210.672
0.05	0.001	1.65	1.841	7632.649	0.41	1.434	5962.275
0.05	0.005	1.68	1.694	7024.438	0.21	1.504	6252.573
0.05	0.01	1.12	1.739	7803.182	0.15	1.504	6251.728
0.05	0.05	1.65	1.694	7023.796	0.1	0.477	1066.311
0.1	0.05	1.65	1.694	7023.87	0.1	0.477	1066.351
0.5	0.05	1.68	1.694	7024.409	0.1	0.477	1066.653
1	0.05	1.65	1.628	6754.427	0.1	0.477	1066.995
2	0.05	1.64	1.841	7632.201	0.1	0.477	1067.584
5	0.05	1.72	2.075	8085.885	0.1	0.404	906.864
6	0.05	1.71	2.109	8208.29	0.1	0.404	906.802
7	0.05	1.72	2.052	7994.416	0.1	0.230	516.504
8	0.05	1.69	2.052	7994.204	0.1	0.230	516.456
9	0.05	1.69	2.172	8459.779	NA	NA	NA
10	0.05	1.68	1.763	6891.977	NA	NA	NA

ASR: annualized Sharpe ratio

C. Stability Analysis of Volatile Market

Considering that the duration of the test trading period is 252 days, the rolling window size to forecast innovation volatility is selected to be 50 days, and 172 overlapping rolling windows are used. For $\sigma_\epsilon^2 = 0.001$, $\delta = 0.0001$ and $p_{opt} = 1.68$ from the training sample, trading signals from the robust strategy for the test sample are visualized in Fig. 7. Fig. 6 compared DDIVF (red line) and KFIVF (blue line), and it is shown that KFIVF is very similar to the average of DDIVF and is not able to capture the time varying innovation volatility in the test period. Therefore, for the traditional strategy, the bounds calculated by $p_{opt}\sqrt{Q_t}$ where $p_{opt} = 1.48$ from training sample is not able to provide appropriate trading signals based on the test sample dynamically, as shown in Fig. 8.

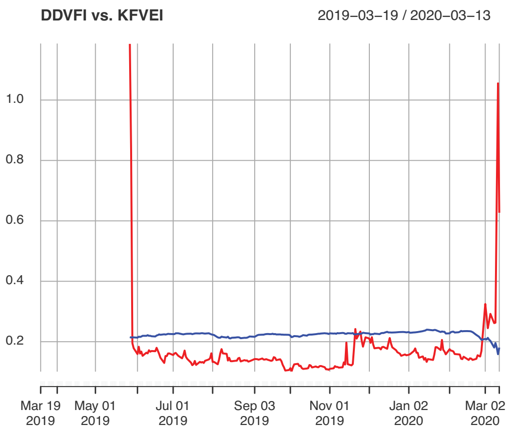

Fig. 6. DDIVF $\hat{\sigma}_t$ vs. KFIVF $\sqrt{Q_t}$

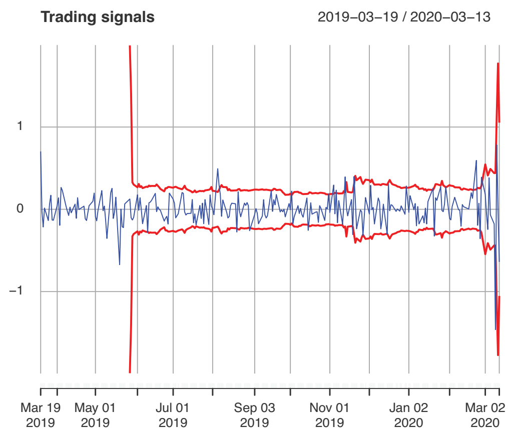

Fig. 7. Robust pairs trading using $\hat{\sigma}_t$

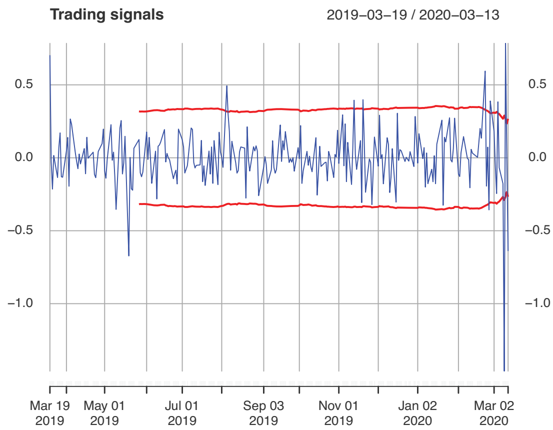

Fig. 8. Trading signals in volatile market: robust pairs trading vs. traditional pairs trading

TABLE IV
STABILITY ANALYSIS OF VOLATILE MARKET: 2019-03-15 TO 2020-03-15

		Robust pairs trading			Traditional pairs trading		
σ_ϵ^2	δ	p_{opt}	ASR	Profit	p_{opt}	ASR	Profit
0.001	0.0001	1.68	2.177	4548.931	1.48	9.408	3415.201
0.001	0.001	1.65	1.748	3661.748	0.47	9.412	3416.024
0.001	0.005	1.66	2.177	4549.977	0.21	9.413	3416.108
0.001	0.01	1.66	2.177	4549.989	0.15	9.413	3416.118
0.001	0.05	1.66	2.177	4549.999	0.1	18.225	1538.643
0.01	0.0001	1.72	2.108	4169.427	1.21	4.546	5523.647
0.01	0.001	1.68	2.177	4548.932	0.34	2.891	6002.732
0.01	0.005	1.65	1.748	3661.637	0.21	9.412	3415.923
0.01	0.01	1.65	1.748	3661.749	0.15	9.413	3416.025
0.01	0.05	1.66	2.177	4549.979	0.1	18.224	1538.627
0.05	0.0001	1.57	1.129	2373.655	0.81	1.799	2269.826
0.05	0.001	1.65	1.912	3999.317	0.41	4.547	5524.517
0.05	0.005	1.68	2.177	4548.936	0.21	9.409	3415.205
0.05	0.01	1.12	1.276	3104.549	0.15	9.411	3415.638
0.05	0.05	1.65	1.748	3661.754	0.1	18.219	1538.557
0.1	0.05	1.65	1.748	3661.647	0.1	18.213	1538.47
0.5	0.05	1.68	2.177	4548.979	0.1	18.174	1537.849
1	0.05	1.65	2.177	4548.15	0.1	18.135	1537.22
2	0.05	1.64	1.912	3999.732	0.1	18.081	1536.338
5	0.05	1.72	2.107	4169.42	0.1	18.004	1535.427
6	0.05	1.71	2.108	4169.496	0.1	17.989	1535.444
7	0.05	1.72	2.108	4169.689	0.1	1.571	226.6097
8	0.05	1.69	1.913	3999.329	0.1	1.572	226.8437
9	0.05	1.69	1.913	3999.679	NA	NA	NA
10	0.05	1.68	1.912	4000.079	NA	NA	NA

ASR: annualized Sharpe ratio

In Table IV, for each value of p_{opt} obtained from the training trading period, profit and ASR between the proposed robust strategy and the traditional one are compared, according to various initial values of σ_ϵ^2 and δ during the volatile test trading period. The traditional pairs trading strategy fails even with the very small initial values, however; the robust data-driven pairs trading strategy works consistently well for various initial values. The results of cumulative profits and ASRs are stable when σ_ϵ^2 ranges from 0.001 to 10, and δ ranges from 0.0001 to 0.05, as shown in Table IV. It is noted that the ASR is around 2 (shown in column 4 in Table IV) for most of the cases, which is an ideal value for daily trading.

The profit of buy and hold strategy during this test period is $ -8429.348. Each proposed robust strategy with certain initial values is much more profitable than the buy and hold strategy for the same test period, with a profit shown in column 5 in Table IV.

IV. Conclusion

This paper presents a maximum informative estimating function based robust filter, and a robust multiple trading strategy based on DDIVF using DD-EWMA forecast model for volatility. The driving idea, unlike the existing work, is that the filtering algorithm is obtained using the partially Bayes EF approach (see [8]). Data-driven robust trading strategies have been evaluated through some experiments and it is shown that the proposed robust trading strategy using DDIVF outperforms (i.e., has a larger profit and much more robust to initial values) the trading strategy using KFIVF. Moreover, the robustness of the proposed method and algorithms are further tested in the volatile market.

Acknowledgement

The first author acknowledges the Faculty of Science start-up grant from Ryerson University. The second author acknowledge the Discovery grant from Natural Sciences and Engineering Research Council (NSERC).

References

[1] B. Efron and T. Hastie, Computer age statistical inference. Cambridge University Press, 2016.

[2] M. Jahja, D. Farrow, R. Rosenfeld, and R. J. Tibshirani, “Kalman filter, sensor fusion, and constrained regression: Equivalences and insights”, In 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 2019.

[3] J. Durbin and S. J. Koopman, Time Series Analysis by State Space Methods. Oxford Statistical Science Series, Oxford, 2001.

[4] E. Chan, Algorithmic Trading: Winning Strategies and Their Rationale. Hoboken, New Jersey, John Wiley and Sons, 2013.

[5] K. Longmore, Kalman Filter Example: Pairs Trading in R. Robot Wealth. https://robotwealth.com/kalman-filter-pairs-trading-r/, September, 2019.

[6] M. E. Thompson and A. Thavaneswaran, “Filtering via estimating functions”, Applied Mathematics Letters, vol. 2(5), pp. 6167, 1999.

[7] M. E. Thompson, “Dynamic data science and official statistics”, The Canadian Journal of Statistics, vol. 46 (1) pp. 10 - 23, 2018.

[8] A. Thavaneswaran and M. E. Thompson, “Nonnormal filtering via estimating functions”, In N. Balakrishnan, editor, Aspects of Probability and Statistics, pp. 173 183, CRC Press, London, 2019.

[9] Y. Zhang, J. Zou, N. Ravishanker and A. Thavaneswaran, “Modeling financial durations using penalized estimating functions”, Computational Statistics & Data Analysis, vol. 131, pp. 145-158, 2019.

[10] A. Thavaneswaran, N. Ravishanker, and Y. Liang. “Generalized duration models and optimal estimation using estimating functions”, Annals of the Institute of Statistical Mathematics, vol. 67(1), pp. 129156, 2015.

[11] A. Thavaneswaran, A. Paseka and J. Frank, “Generalized Value at Risk Forecasting”, Communications in Statistics Theory and Methods, pp. 1-8, 2019.

[12] A. Arratia, Computational Finance: An Introductory Course with R. Atlantis Press, 2014.

[13] Á. Cartea, S. Jaimungal, and J. Penalva, Algorithmic and High-Frequency Trading (Mathematics, Finance and Risk). Cambridge University Press, 2015.

[14] C. Conlan, Automated Trading with R: Quantitative Research and Platform Development. Apress, 2016.

[15] M. L. Halls-Moore, Advanced Algorithmic Trading. Available at https://www.quantstart.com/advanced-algorithmic-trading-ebook/, 2017.

[16] A. Thavaneswaran, Y. Liang, Z. Zhu and R. K. Thulasiram, “Novel Data Driven Fuzzy Algorithmic Volatility Forecasting Models with Applications to Algorithmic Trading”, In proceeding of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Glasgow, UK, July, 2020.

[17] E. G. Gatev, W. N. Goetzmann, and K. G. Rouwenhorst, “Pairs trading: Performance of a relative-value arbitrage rule. Review of Financial Studies”, vol. 19(3), pp. 797 827, 2006.

[18] R. J. Elliott, J. van der Hoek and W. P. Malcolm, “Pairs trading”, Quantitative Finance, vol. 5, pp. 271 - 276, 2005.

[19] C. Ed. de Moura, A. H. Pizzinga, and J. P. Zubelli, “A pairs trading strategy based on linear state space models and the kalman filter”, Quantitative Finance, vol. 16, pp. 1559 - 1573, 2016.

[20] Y. Liang, A. Thavaneswaran, N. Yu, M. E. Hoque, and R. K. Thulasiram, “Dynamic data science applications in optimal profit algorithmic trading”, pp. 1294-1299, In proceedings (workshop) of IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC 2020), 2020

[21] Y. Liang, A. Thavaneswaran, A. Paseka, Z. Zhu, and R. K. Thulasiram, “A novel data-driven algorithmic trading strategy using joint forecasts of volatility and stock price”, pp. 293-302, In proceedings (Symposia) of IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC 2020), 2020.