Prediction with Expert Advice for Value at Risk

Raisa Dzhamtyrova
Department of Computer Science
Royal Holloway, University of London
Egham, United Kingdom
Raisa.Dzhamtyrova@rhul.ac.uk

Yuri Kalnishkan
Department of Computer Science
Royal Holloway, University of London
Laboratory of Advanced Combinatorics and Network Applications
Moscow Institute of Physics and Technology
Egham, United Kingdom
Yuri.Kalnishkan@rhul.ac.uk

***Abstract*—We propose to apply the method of online prediction with expert advice for estimation of Value at Risk. We show that in some cases the combination of different methods can produce better results compared to a single model. Our approach is based on Weak Aggregating Algorithm (WAA), which is similar to the Bayesian method, where the prediction is the average over all models based on the likelihood of the available data. WAA provides a theoretical guarantee that the prediction strategy is asymptotically as good as the best expert. We propose two ways of combining predictions of different experts. The first approach combines predictions of normal distribution experts, whereas the second method combines predictions of conventional models that are used to estimate Value at Risk. The experimental results on three stocks show that WAA performs close to or better than the best expert model. In addition, backtesting with Kupiec unconditional coverage test and Christoffersen conditional coverage test shows that WAA is the only method that fails to reject the null hypothesis for all test cases.**

***Index Terms*—online learning, prediction with expert advice, Value at Risk, Weak Aggregating Algorithm**

I. Introduction

In the history of finance, there have been a lot of crises that deeply influenced the global economy. Examples of these crises are the Wall Street crash in 1987, the Japan financial crisis in 1989, the Asian financial crisis in 1997, the subprime mortgage crisis of 2007-2008 and the European debt crisis in 2010. Financial crises and the rise of uncertainty in the financial market emphasize the need of effective risk calculation.

Value at Risk (VaR) measure is one of the most important methods of risk management. The VaR method was introduced in 1994 by J. P. Morgan [6] and became widely used by most financial institutions. J. P. Morgan [13] defines VaR as 'a measure of the maximum possible change in the value of a portfolio of financial instruments over a pre-set horizon'.

There are several conventional methods that are widely used for measuring VaR. Historical Simulation is one of the non-parametric methods for measuring VaR, which assumes that all possible future variations have been experienced in the past and will be repeated in the future [3]. Another approach, known as parametric, is when one estimates volatility of assets' returns in turn to obtain their VaR. Across parametric approaches the conventional methods include Variance–Covariance, Exponential Weighted Moving Average (EWMA) (Section 20.6 in [7]) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) [2].

Some of the procedures to estimate VaR propose the use of quantile regression. The quantile regression approach suggested by Koenker and Basset [11] is one of the methods which models a quantile of the response variable conditional on the explanatory variables. 'It is natural to evaluate a VaR model by a quantile regression method due to its capability of conditional distribution exploration with distribution-free assumption, also allowing for serial correlation and conditional heteroskedasticity' [5]. In [15] a procedure to estimate a conditional quantile model to calculate VaRs for portfolios is presented; this method is found to be comparable with conventional methods in forecasting VaR.

In this paper we will use the same pinball loss function as is used in the quantile regression model. However, we do not try to optimize parameters of some model. Instead our approach combines predictions of different models based on the method of online prediction with expert advice. Contrary to batch mode, where the algorithm is trained on a training set and gives predictions on a test set, in online setting we learn as soon as new observations become available. One may wonder why we don't just use predictions of only one best expert from the beginning and ignore predictions of others. First, sometimes we cannot have enough data to identify the best expert from the start. Second, good performance in the past does not necessary lead to good performance in the future. In addition, previous research shows that combining predictions of multiple regressors often produces better results compared to a single model [14].

Our approach is based on the Weak Aggregating Algorithm (WAA) which was first introduced in [9]. The WAA works as follows: we assign initial weights to experts and at each step the weights of experts are updated according to their performance. The approach is similar to the Bayesian method, where the prediction is the average over all models based on the likelihood of the available data. The WAA gives a guarantee ensuring that the learner's loss is as small as best expert's loss up to an additive term of the form $C\sqrt{T}$.

The WAA was proposed as an alternative to the Aggregating Algorithm (AA), which was first introduced in [16]. The AA gives a guarantee ensuring that the learner's loss is as small as the best expert's loss up to a constant in case of finitely many experts. The AA provides better theoretical guarantees. However, it works with mixable loss functions and is not

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

applicable in our task. An interesting application of the method of prediction with expert advice for the Brier loss function in forecasting football outcomes can be found in [17]; it was shown that the proposed strategy that follows AA is as good as any bookmaker. In a recent paper [1] merging algorithms were proposed for prediction of packs with tight worst case loss upper bounds similar to those for AA; empirical experiments on sports and house price datasets are carried out to study the performance of the new algorithms and compare them against an existing method.

The main contribution of this paper is the application of the method of prediction with expert advice to estimate VaR. To the best of our knowledge, there was no previous research in this area. We propose two ways of applying the framework of prediction with expert advice for prediction of VaR. The first approach is to apply WAA to combine predictions of normal distribution experts, where each expert has particular parameters of standard deviation. We choose to evaluate performance of proposed strategies using stocks' adjusted closing prices of Walmart, WPP inc. and Apple. The experiments show that loss of the WAA is close to or better than the loss of the retrospectively best normal distribution expert. We compare WAA with the model of quantile regression, and experimental results show that in most cases WAA outperforms quantile regression.

The second approach is to combine predictions of several conventional methods for estimating VaR, such as Historical Simulation, Variance–Covariance, EWMA and GARCH. The experiments illustrate that combining predictions of different experts sometimes could provide better results compared to the single retrospectively best model.

We run backtesting of all methods by using Kupiec [12] unconditional coverage test and Christoffersen [4] conditional coverage test to do the Backtesting on VaR. WAA with normal distribution experts is the only method that fails to reject the null hypothesis for all test cases. WAA, which combines predictions of four conventional models, seem to reject the null hypothesis in situations when most of these models reject the null hypothesis.

II. Framework

We consider a game $\mathfrak{G}$, where space of outcomes $\Omega = \mathbb{R}$ and decision space $\Gamma = \mathbb{R}$, where for any $y \in \Omega$ and $\gamma \in \Gamma$ we define the pinball loss for $\alpha \in (0,1)$

$$\lambda(y,\gamma) = \begin{cases} \alpha(y-\gamma), & \text{if } y \geq \gamma \\ (1-\alpha)(\gamma - y), & \text{if } y < \gamma \end{cases}. \tag{1}$$

When N days is the time horizon and $1-\alpha$ is the confidence level, $\mathrm{VaR}_{1-\alpha}$ is the loss corresponding to the α-quantile of the distribution of the gain in the value of the portfolio over the next N days (Chapter 21.1 in [8]). We consider outcomes to be returns of some stocks or portfolios. Let outcomes have a cumulative distribution function $F_Y(z) = \Pr(Y \leq z)$. Because VaR is conventionally reported as a positive number, we define

$$\mathrm{VaR}_{1-\alpha} = -\inf\{z : F_Y(z) \geq \alpha\} \tag{2}$$

as the negative α-quantile of Y. Then the problem of VaR estimation is equivalent to the problem of prediction of α-quantile of returns. This problem can be solved by applying the quantile regression.

Letting x_t denote a sequence of signals, suppose y_t is a random sample on the regression process $u_t = y_t - x_t\beta$. The α-th quantile regression, $0 < \alpha < 1$, is defined as any solution to the minimization problem:

$$\min_{b \in \mathbb{R}^n} \sum_{t: y_t > x_t b} \alpha |y_t - x_t b| + \sum_{t: y_t < x_t b} (1-\alpha)|y_t - x_t b|.$$

The least absolute error estimator is the regression median, i.e., the quantile regression for $\alpha = 1/2$ ([11]). The loss function (2) is appropriate for quantile regression because on average it is minimized by the α-th quantile. Namely, if Y is a real-valued random variable with a cumulative distribution function $F_Y(z) = \Pr(Y \leq z)$, then the expectation $\mathbb{E}\lambda(Y,\gamma)$ is minimized by $\gamma = \inf\{z : F_Y(z) \geq \alpha\}$ (see Section 1.3 in [10] for a discussion).

In the framework of prediction with expert advice the learner has access to predictions $\xi_t(1),\ \xi_t(2), \dots, \xi_t(N)$ at time t generated by experts $\mathcal{E}_1,\ \mathcal{E}_2, \dots, \mathcal{E}_N$ that try to predict elements of the same sequence.

Learner works according to the following protocol:

Protocol 1.

FOR $t = 1, 2, \dots$
  learner reads experts' predictions $\gamma_t^1,\ \gamma_t^2, \dots, \gamma_t^N \in \Gamma$
  learner outputs $\gamma_t \in \Gamma$
  nature announces $y_t \in \Omega$
  learner suffers loss $\lambda(y_t, \gamma_t)$
END FOR

The performance of a learner is measured by the cumulative loss.

Let us denote L_T^i the cumulative loss of expert $\mathcal{E}_i$ at step T:

$$\begin{aligned} L_T^i &:= \sum_{t=1}^{T} \lambda(y_t, \xi_t(i)) \\ &= \sum_{\substack{t=1,\dots,T: \\ y_t > \xi_t(i)}} \alpha |y_t - \xi_t(i)| + \sum_{\substack{t=1,\dots,T: \\ y_t < \xi_t(i)}} (1-\alpha)|y_t - \xi_t(i)|. \end{aligned}$$

The cumulative loss of the learner at step T is:

$$\begin{aligned} L_T &:= \sum_{t=1}^{T} \lambda(y_t, \gamma_t) \\ &= \sum_{\substack{t=1,\dots,T: \\ y_t > \gamma_t}} \alpha |y_t - \gamma_t| + \sum_{\substack{t=1,\dots,T: \\ y_t < \gamma_t}} (1-\alpha)|y_t - \gamma_t|. \end{aligned}$$

III. Weak Aggregating Algorithm

In the framework of prediction with expert advice we have access to experts' predictions at each time step and the learner has to make a prediction based on experts' past performance. We use an approach based on the WAA since a pinball loss function $\lambda(y,\gamma)$ is convex in γ. The WAA accepts N initial normalised weights $q_1, q_2, \dots, q_N \in [0,1]$ such that $\sum_{i=1}^N q_i = 1$ and a positive number c as parameters. The parameter c has a meaning of the learning rate in the theory of prediction with expert advice. The choice of the initial weights $q_i,\ i = 1, \dots, N$ might contain our prior knowledge about experts. When no prior information is available, the common choice is to assign equal initial weights.

The pseudo-code of WAA is given below:

Protocol 2.

(1) $l_1^i := 0,\ i = 1, 2, \dots, N$
(2) FOR $t = 1,\ 2, \dots, T$
(3) $\quad \beta_t = e^{-c/\sqrt{t}}$
(4) $\quad w_t^i := q_i \beta_t^{l_t^i},\ i = 1, 2, \dots, N$
(5) $\quad p_t^i := \frac{w_t^i}{\sum_{j=1}^N w_t^j},\ i = 1, 2, \dots, N$
(6) $\quad$ read experts' predictions $\gamma_t^1,\ \gamma_t^2, \dots, \gamma_t^N$
(7) $\quad$ output prediction $\gamma_t = \sum_{i=1}^N \gamma_t^i p_t^i$
(8) $\quad$ observe y_t
(9) $\quad l_{t+1}^i = l_t^i + \lambda(y_t, \gamma_t^i),\ i = 1, 2, \dots, N$
(10) END FOR

The variable l_t^i stores the loss of the i-th expert $\mathcal{E}_i$, i.e., after trial t we have $l_{t+1}^i = L_t^{\mathcal{E}_i}$. The values w_t^i are weights assigned to the experts during the work of the algorithm; they depend on the loss suffered by the experts and the initial weights q_i . The values p_t^i are obtained by normalising w_t^i.

For bounded games WAA provides the following theoretical guarantee on the cumulative loss of a learner.

Lemma 1. (Lemma 11 in [9]) For every $L > 0$, every game $\langle \Omega, \Gamma, \lambda \rangle$ such that $|\Omega| < \infty$ and $\lambda(y,\gamma) \leq L$ for all $y \in \Omega$ and $\gamma \in \Gamma$ and every $N = 1, 2, \dots,$ for every merging strategy for N experts that follows the WAA with initial weights $q_1, q_2, \dots, q_N \in [0,1]$ such that $\sum_{i=1}^N q_i = 1$ and $c > 0$ the bound

$$L_T \leq L_T^{\mathcal{E}_i} + \sqrt{T}\left(\frac{1}{c}\ln\frac{1}{q_i} + cL^2\right)$$

is guaranteed for every $T = 1, 2, \dots$ and every $i = 1, 2, \dots, N$.

The theoretical bound of WAA depends on the maximum value of loss L. However, for many tasks, predicted outcomes are bounded. Therefore, it is possible to have a sensible estimate for the maximum loss based on the historical information.

After taking equal initial weights $q_1 = q_2 = \dots = q_N = 1/N$ in the WAA, the additive term reduces to $(cL^2 + (\ln N)/c)\sqrt{T}$. When $c = \sqrt{\ln N}/L$, this expression reaches its minimum. We get the following corollary.

Corollary 1. (Corollary 14 in [9]) Under the conditions of Lemma 1, there is a merging strategy such that the bound

$$L_T \leq L_T^{\mathcal{E}_i} + 2L\sqrt{T \ln N}$$

is guaranteed.

IV. Experiments

In this section, we apply WAA to the problem of prediction of VaR using three stocks of Walmart, Apple and WPP inc. We use daily adjusted closing prices from January 2011 to December 2018 that are downloaded from Yahoo Finance.[1]

A. WAA for normal distributions

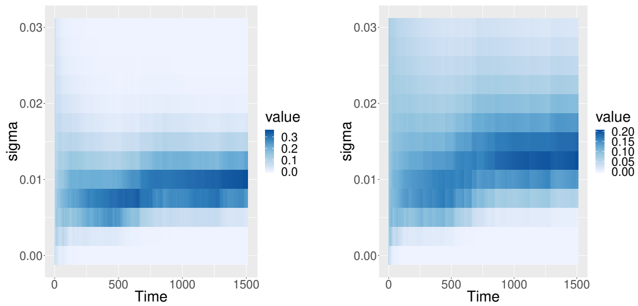
First, as a proof of concept, we apply WAA to normal distribution experts. We assume that stock investment's returns are normally distributed around the mean of a normal probability distribution. The volatility σ of a stock is a measure of our uncertainty about the returns provided by the stock. Each expert $\mathcal{E}_i$ predicts according to $\mathcal{N}(0, \sigma_i^2),\ i = 1, \dots, N$. We pick σ_i to be in a range from 0 to 0.03 with a step 0.0025, and hence we have $N = 13$ normal distribution experts. We assign equal initial weights $q_i = 1/N,\ i = 1, \dots, N$ and follow Protocol 2. We estimate the constant of WAA to be $c = 200$, using the formula $c = \sqrt{\ln N}/L$ which achieves the minimum of the additive term of the loss bound. We estimate L as the maximum loss over the first 500 observations. We test the performance of WAA using dataset without the first 500 observations. Figure 1 shows the weights update for experts $\mathcal{E}_i,\ i = 1, \dots, N$ for Walmart. We can see from the graph that, for significance level $\alpha = 0.05$, expert with $\sigma = 0.01$ has the largest weights at the end of the period. It corresponds with Figure 4, where the same expert has the lowest total pinball loss. It shows that WAA converges to the best expert by updating weights of experts online based on their performance. For significance level $\alpha = 0.01$ expert with $\sigma = 0.0125$ has the largest weight at the end of the period and the lowest total loss. However, for $\alpha = 0.01$ losses of several experts are close to the loss of best expert, and as a result, their weights are also close to each other. A similar picture can be seen for WPP inc. at Figures 2, 5, and for Apple at Figures 3, 6. Tables I, II summarise losses of normal distribution experts and WAA for $\alpha = 0.05$ and $\alpha = 0.01$ respectively. We can see from the table that losses of WAA are very close to the loss of best normal distribution expert. For example, for WPP inc. losses of WAA are lower than losses of best experts.

We also compare the performance of WAA with quantile regression model (QR). QR was trained in online mode using sliding window of the length 500. We can see from Tables I, II that in most cases losses of WAA are lower than losses of QR. Tables III, IV show the actual exceptions of VaR, i.e., the number of times when stock's losses exceed VaR, for each method for $\alpha = 0.05$ and $\alpha = 0.01$ respectively. It seems that WAA tends to underestimate VaR a little, while QR overestimates VaR.

This approach shows that it is reasonable to apply WAA in the considered setting. WAA converges to the best expert by updating experts' weights online. In addition, best experts might be different for different significance levels. It shows that the single retrospectively best model might not perform well in the future, and it is reasonable to apply the mixture of

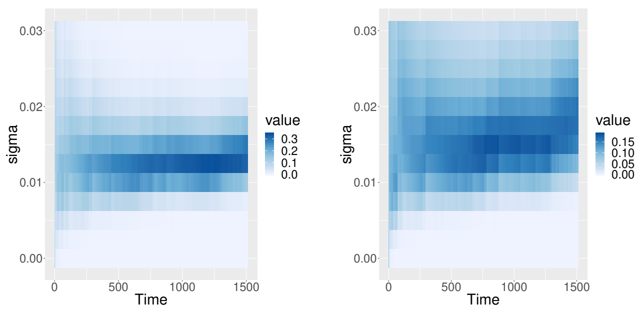
[1]The code written in R is available at https://github.com/RaisaDZ/VaR.

models instead. The performance of WAA is close to or better than the best normal distribution expert, and in most cases it outperforms the model of QR.



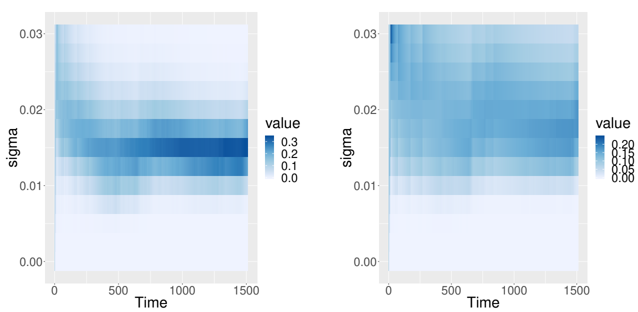
(a) $\alpha = 0.05$ (b) $\alpha = 0.01$

Fig. 1. Weights update for Walmart



(a) $\alpha = 0.05$ (b) $\alpha = 0.01$

Fig. 2. Weights update for WPP inc.

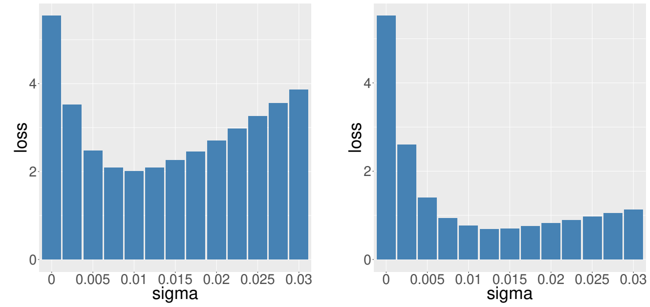


(a) $\alpha = 0.05$ (b) $\alpha = 0.01$

Fig. 3. Weights update for Apple

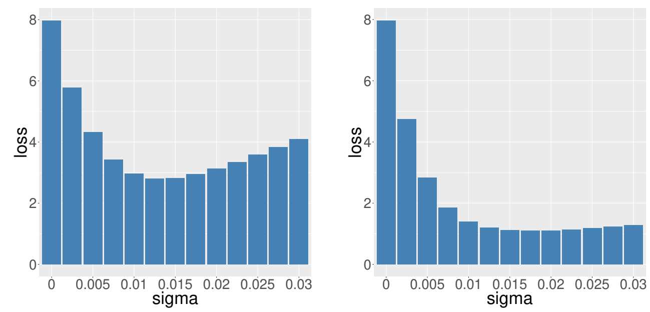
B. *WAA for conventional models*

In this section, we use WAA with four conventional models that are widely used to calculate VaR: Historical Simulation, Variance–Covariance, EWMA and GARCH. ‘Historical simulation is one popular way of estimating VaR. It involves using past data as a guide to what will happen in the future’ (Section 21.2 in [8]). Suppose that we want to calculate $\mathrm{VaR}_{1-\alpha}$ for a stock, and data are collected on movements in the market variables over the most recent N days. This provides



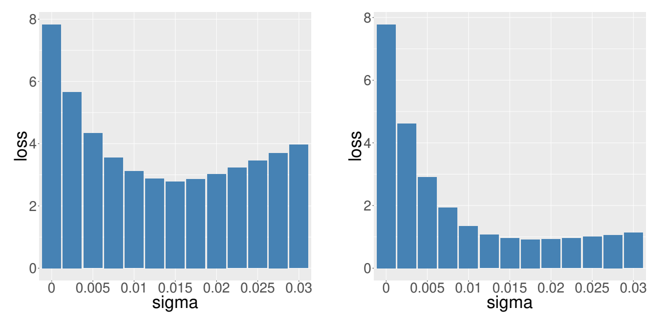
(a) $\alpha = 0.05$ (b) $\alpha = 0.01$

Fig. 4. Losses of normal distribution experts for Walmart



(a) $\alpha = 0.05$ (b) $\alpha = 0.01$

Fig. 5. Losses of normal distribution experts for WPP inc.



(a) $\alpha = 0.05$ (b) $\alpha = 0.01$

Fig. 6. Losses of normal distribution experts for Apple

$N-1$ alternative scenarios for what can happen between today and tomorrow. The estimate of $\mathrm{VaR}_{1-\alpha}$ is the negative α-quantile (2) of returns based on $N-1$ historical scenarios. The Variance-Covariance method is one of the parametric methods which estimates the volatility of returns based on the normal distribution assumption. Then VaR is calculated as the α-quantile of the normal distribution with zero mean and the estimated volatility. The exponentially weighted moving average (EWMA) is another parametric method, where the estimate of the volatility σ_t for day t is given by the formula

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1-\lambda) u_{t-1}^2,$$

where $0 < \lambda < 1$ is a constant, σ_{t-1} is the volatility estimate at the end of day $t-2$ of the volatility for day $t-1$ and u_{t-1}

TABLE I

TOTAL LOSSES OF NORMAL DISTRIBUTION EXPERTS FOR $\alpha = 0.05$.

sigma	WMT	WPP	AAPL
0	5.545	7.974	7.834
0.0025	3.515	5.775	5.655
0.005	2.478	4.329	4.337
0.0075	2.083	3.427	3.561
0.01	**2.007**	2.975	3.113
0.0125	2.088	**2.811**	2.876
0.015	2.252	2.828	**2.788**
0.0175	2.450	2.948	2.865
0.02	2.700	3.130	3.023
0.0225	2.975	3.346	3.228
0.025	3.262	3.587	3.453
0.0275	3.556	3.838	3.702
0.03	3.857	4.094	3.968
WAA	2.013	2.806	2.834
QR	2.089	2.851	2.761

TABLE II

TOTAL LOSSES OF NORMAL DISTRIBUTION EXPERTS FOR $\alpha = 0.01$.

sigma	WMT	WPP	AAPL
0	5.523	7.970	7.782
0.0025	2.604	4.745	4.619
0.005	1.397	2.838	2.910
0.0075	0.939	1.854	1.935
0.01	0.763	1.397	1.344
0.0125	**0.688**	1.199	1.066
0.015	0.702	1.117	0.959
0.0175	0.751	**1.099**	**0.919**
0.02	0.817	1.103	0.923
0.0225	0.894	1.128	0.953
0.025	0.970	1.177	1.001
0.0275	1.046	1.230	1.059
0.03	1.122	1.283	1.131
WAA	0.705	1.085	0.930
QR	0.796	1.181	1.080

TABLE III

ACTUAL EXCEPTIONS OF NORMAL DISTRIBUTION EXPERTS FOR $\alpha = 0.05$.

	expected = 75.5		
sigma	WMT	WPP	AAPL
0	711	720	721
0.0025	439	501	492
0.005	227	360	320
0.0075	123	234	219
0.01	74	143	155
0.0125	43	90	115
0.015	31	58	79
0.0175	20	37	42
0.02	10	28	29
0.0225	7	19	23
0.025	5	15	18
0.0275	3	14	12
0.03	2	11	10
WAA	72	73	63
QR	92	86	85

is the most recent daily percentage change in returns (Section 22.2 in [8]). Finally, the GARCH(p, q) estimates the volatility σ_t for day t as

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2.$$

TABLE IV

ACTUAL EXCEPTIONS OF NORMAL DISTRIBUTION EXPERTS FOR $\alpha = 0.01$.

	expected = 15.1		
sigma	WMT	WPP	AAPL
0	711	720	721
0.0025	339	434	407
0.005	140	251	226
0.0075	63	129	144
0.01	33	64	91
0.0125	20	36	42
0.015	9	23	26
0.0175	5	15	19
0.02	2	14	11
0.0225	2	8	8
0.025	2	6	6
0.0275	2	6	4
0.03	2	6	2
WAA	9	14	12
QR	22	32	28

In our experiments we use GARCH(1, 1) which is based on the most recent volatility estimates and the most recent returns' changes.

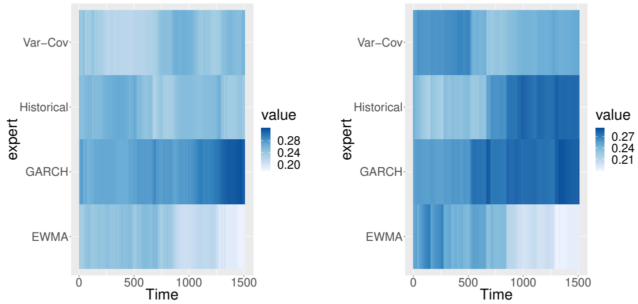
We train these models using a sliding window of length 500, and then apply WAA using forecasts of these models to predict a one-step ahead forecast. We re-train all models except GARCH(1, 1) after each new observation becomes available, for GARCH(1, 1) we do it after each 50 steps due to computational complexity of this method. We start with equal initial weights of each model and then update their weights according to their current performance.

Figures 7, 8, 9 illustrate weights of each model depending on the current time step. Figure 10 with the corresponding Tables V, VI show total losses of each model and WAA for $\alpha = 0.05$ and $\alpha = 0.01$ respectively. We can see from the graphs that in most cases GARCH(1, 1) obtains the largest weights which indicates that it suffers smaller losses compared to other models. However, it changes for $\alpha = 0.01$ for WPP inc., where the largest weights are acquired by Historical Simulation model. It shows that sometimes we cannot use the past information to evaluate the best model. The retrospectively best model can perform worse in the future as an underlying nature of data generating can change. In addition, different models can perform better for different significance levels of VaR.

Similar to the previous experiments, losses of WAA are very close to the loss of the best performing expert. In most of the cases the best expert is GARCH(1, 1), and WAA follows its predictions. However, for $\alpha = 0.01$ for WPP inc. the best expert changes. It again illustrates that the retrospectively best model could change with time, and one should be cautious about choosing the single retrospectively best model for future forecasts.

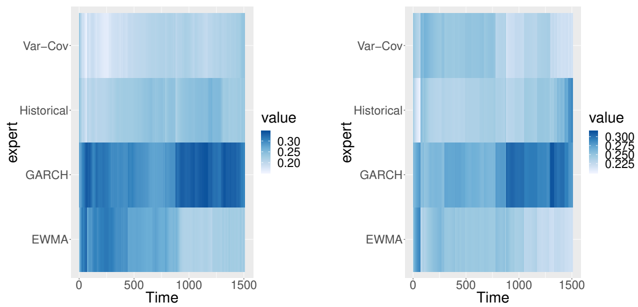
C. Backtesting

First, we introduce the Kupiec unconditional coverage test, which is also known as the proportion of failures test. The most common way to test the performance of VaR models is to count the number of exceptions (failures), i.e., the number



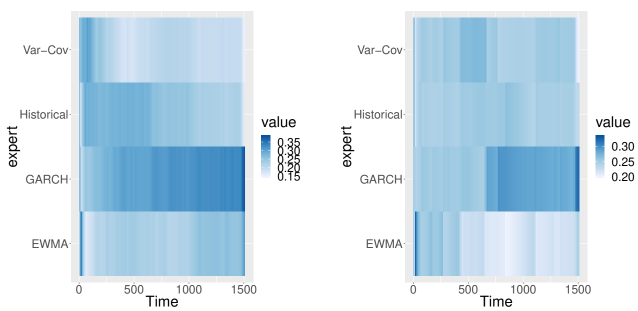
(a) $\alpha = 0.05$ (b) $\alpha = 0.01$

Fig. 7. Weights update for Walmart



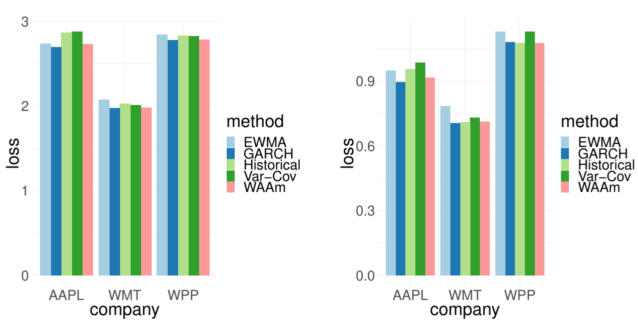
(a) $\alpha = 0.05$ (b) $\alpha = 0.01$

Fig. 8. Weights update for WPP inc.



(a) $\alpha = 0.05$ (b) $\alpha = 0.01$

Fig. 9. Weights update for Apple



(a) $\alpha = 0.05$ (b) $\alpha = 0.01$

Fig. 10. Total losses of methods

TABLE V
TOTAL LOSSES OF METHODS FOR $\alpha = 0.05$.

Method	WMT	WPP	AAPL
Historical	2.031	2.829	2.867
Var-Cov	2.012	2.827	2.880
EWMA	2.077	2.845	2.734
GARCH(1, 1)	**1.978**	**2.781**	**2.695**
WAAm	1.983	2.782	2.733

TABLE VI
TOTAL LOSSES OF METHODS FOR $\alpha = 0.01$.

Method	WMT	WPP	AAPL
Historical	0.711	1.076	0.956
Var-Cov	0.731	1.129	0.986
EWMA	0.786	1.130	0.948
GARCH(1, 1)	**0.706**	1.081	**0.896**
WAAm	0.713	**1.075**	0.917

of times when stock's losses exceed VaR. Denoting m to be the number of exceptions, we define the failure rate during the time horizon T as m/T. The Kupiec unconditional coverage test measures whether the number of exceptions is consistent with the confidence level. The null hypothesis H_0 is

$$H_0 : \alpha = \hat{\alpha} = m/T,$$

where $\hat{\alpha}$ is the observed failure rate and α is the significance level of $\mathrm{VaR}_{1-\alpha}$. According to [12] the test statistics takes the form of a likelihood ratio test:

$$\mathrm{LR_{UC}} = -2\ln\left(\frac{(1-p)^{T-m}p^m}{(1-m/T)^{T-m}(m/T)^m}\right).$$

This statistic is asymptotically distributed as a chi-square variable with 1 degree of freedom. The VaR model fails the test if this likelihood ratio exceeds a critical value. The critical value depends on the test confidence level.

The Kupiec unconditional coverage test focuses only on the number of exceptions. However, we would like to test whether these exceptions were evenly spread over time. The null hypothesis H_0 for Christoffersen conditional coverage test is that the probability of observing an exception on a particular day does not depend on whether an exception occurred. The test statistic for independence is given by

$$\mathrm{LR_{CCI}} = -2\ln\left(\frac{(1-\pi)^{n_{00}+n_{10}}\pi^{n_{01}+n_{11}}}{(1-\pi_0)^{n_{00}}\pi_0^{n_{01}}(1-\pi_1)^{n_{10}}\pi_1^{n_{11}}}\right),$$

where n_{00} is the number of periods with no failures followed by a period with no failures,
n_{10} is the number of periods with failures followed by a period with no failures,
n_{01} is the number of periods with no failures followed by a period with failures,
n_{11} is the number of periods with failures followed by a period with failures,

and π_i is the probability of having a failure conditional on the previous period:

$$\pi_0 = \frac{n_{01}}{n_{00}+n_{01}}, \ \pi_1 = \frac{n_{11}}{n_{10}+n_{11}}$$
$$\text{and } \pi = \frac{n_{01}+n_{11}}{n_{00}+n_{01}+n_{10}+n_{11}}.$$

This statistic is asymptotically distributed as a chi-square with 1 degree of freedom. The Christoffersen conditional coverage test is a combination of this statistic with the frequency unconditional coverage test:

$$\mathrm{LR}_{\mathrm{CC}} = \mathrm{LR}_{\mathrm{CCI}} + \mathrm{LR}_{\mathrm{UC}}.$$

This test is asymptotically distributed as a chi-square variable with 2 degrees of freedom.

In this section, we perform backtesting of all considered methods by running Kupiec unconditional coverage test and Christoffersen conditional coverage test. Tables VII, VIII and IX show results for $\alpha = 0.05$ for Walmart, WPP inc. and Apple respectively. Tables X, XI and XII illustrate results for $\alpha = 0.01$. UCD and CCD denotes unconditional and conditional decisions respectively: Reject (Reject $H0$), Fail (Fail to Reject $H0$). We denote WAAn the method considered in Section IV-A, and WAAm is the method from Section IV-B. We can see from the tables that WAA for normal distribution experts (WAAn) is the only method that fails to reject the null hypothesis H_0. The second best performing model seems to be GARCH(1, 1) as it rejects the only test case for WPP inc. with significance level $\alpha = 0.01$. In Table X we can see that all methods reject the null hypothesis except WAAn. WAA for conventional model experts (WAAm) sometimes rejects the null hypothesis. It happens in situations when most of models that are used in WAAm reject the null hypothesis

Figures 11, 12 and 13 illustrate daily returns (in percent) of each company and VaR for WAAn and WAAm. The behavior of VaR for WAAn is smooth because WAAn uses predictions of constant normal distribution experts. VaR of WAAm has more spikes because it uses predictions of methods such as Historical Simulation, Variance-Covariance, EWMA and GARCH(1, 1) which have more fluctuations in their predictions.

TABLE VII
WALMART, $\alpha = 0.05$, EXPECTED = 75.5.

Method	Actual	Luc	Lcc	UCD	CCD
Historical	95	0.0266	0.0398	Reject	Reject
Var-Cov	58	0.0315	0.0869	Reject	Fail
EWMA	69	0.4364	0.4433	Fail	Fail
GARCH	69	0.4364	0.6582	Fail	Fail
QR	92	0.0592	0.0618	Fail	Fail
WAAn	72	0.6772	0.8733	Fail	Fail
WAAm	64	0.1637	0.1609	Fail	Fail

TABLE VIII
WPP INC., $\alpha = 0.05$, EXPECTED = 75.5.

Method	Actual	Luc	Lcc	UCD	CCD
Historical	84	0.3238	0.0056	Fail	Reject
Var-Cov	60	0.0580	0.0192	Fail	Reject
EWMA	74	0.8590	0.0387	Fail	Reject
GARCH	78	0.7690	0.3462	Fail	Fail
QR	86	0.2247	0.0978	Fail	Fail
WAAn	73	0.7667	0.0891	Fail	Fail
WAAm	67	0.3066	0.0218	Fail	Reject

TABLE IX
APPLE, $\alpha = 0.05$, EXPECTED = 75.5.

Method	Actual	Luc	Lcc	UCD	CCD
Historical	85	0.2711	0.0005	Fail	Reject
Var-Cov	72	0.6772	0.0020	Fail	Reject
EWMA	66	0.2521	0.0160	Fail	Reject
GARCH	82	0.4488	0.3711	Fail	Fail
QR	85	0.2711	0.3277	Fail	Fail
WAAn	63	0.1291	0.0536	Fail	Fail
WAAm	72	0.6772	0.1831	Fail	Fail

TABLE X
WALMART, $\alpha = 0.01$, EXPECTED = 15.1.

Method	Actual	Luc	Lcc	UCD	CCD
Historical	17	0.6300	0.7336	Fail	Fail
Var-Cov	30	0.0007	0.0028	Reject	Reject
EWMA	35	0.0000	0.0001	Reject	Reject
GARCH	20	0.2273	0.2594	Fail	Fail
QR	22	0.0948	0.1789	Fail	Fail
WAAn	9	0.0880	0.2211	Fail	Fail
WAAm	24	0.0340	0.0737	Reject	Fail

TABLE XI
WPP INC., $\alpha = 0.01$, EXPECTED = 15.1.

Method	Actual	Luc	Lcc	UCD	CCD
Historical	18	0.4666	0.0437	Fail	Reject
Var-Cov	26	0.0106	0.0082	Reject	Reject
EWMA	30	0.0007	0.0028	Reject	Reject
GARCH	26	0.0106	0.0241	Reject	Reject
QR	32	0.0001	0.0001	Reject	Reject
WAAn	14	0.7733	0.2813	Fail	Fail
WAAm	24	0.0340	0.0737	Reject	Fail

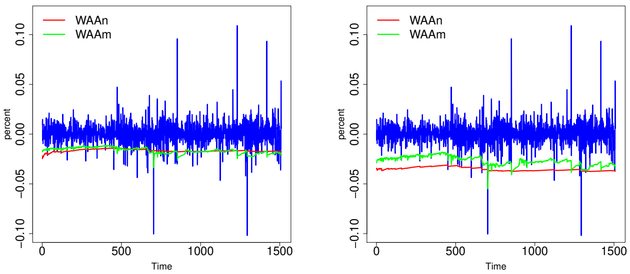
TABLE XII
APPLE, $\alpha = 0.01$, EXPECTED = 15.1.

Method	Actual	Luc	Lcc	UCD	CCD
Historical	21	0.1496	0.2049	Fail	Fail
Var-Cov	24	0.0340	0.0737	Reject	Fail
EWMA	22	0.0948	0.1789	Fail	Fail
GARCH	15	0.9793	0.8599	Fail	Fail
QR	28	0.0029	0.0032	Reject	Reject
WAAn	12	0.4057	0.6428	Fail	Fail
WAAm	19	0.3322	0.4904	Fail	Fail

V. CONCLUSIONS

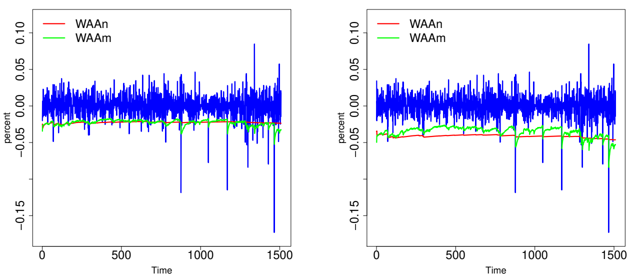
We proposed two ways of applying the framework of prediction with expert advice for calculating VaR. The first approach is to apply WAA with normal distribution experts. The experiments show that WAA converges to the best expert by updating weights of experts online based on their current performance, and its loss is close to or better than the loss of the best expert. WAA also outperforms the quantile regression model that is built using sliding window.

The second approach is to combine predictions of different



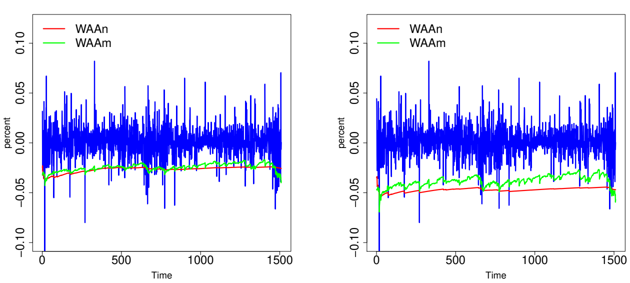
(a) $\alpha = 0.05$ (b) $\alpha = 0.01$

Fig. 11. VaR for Walmart



(a) $\alpha = 0.05$ (b) $\alpha = 0.01$

Fig. 12. VaR for WPP inc.



(a) $\alpha = 0.05$ (b) $\alpha = 0.01$

Fig. 13. VaR for Apple

methods: Historical Simulation, Variance–Covariance, EWMA and GARCH(1, 1). Similar to the previous experiments, losses of WAA are very close to the loss of best performing model, and sometimes WAA shows a better performance. The experiments illustrate that the retrospectively best model could change with time, and combining predictions of different experts could provide better results.

We compare performances of all different methods of prediction of VaR by running Kupiec unconditional coverage test and Christoffersen conditional coverage test. WAA for normal distribution experts is the only method which fails to reject the null hypothesis for both unconditional and conditional coverage tests.

REFERENCES

[1] D. Adamskiy, A. Bellotti, R. Dzhamtyrova, and Y. Kalnishkan. Aggregating algorithm for prediction of packs. *Machine Learning*, 2019.

[2] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, pages 31(3):307–327, 1986.

[3] J. Butler and B. Schachter. Estimating value-at-risk with a precision measure by combining kernel estimation with historical simulation. *Review of Derivatives Research*, pages 1(4):371–390, 1996.

[4] P. Christoffersen. Evaluating interval forecasts. *International Economic Review*, page 39(4):841, 1998.

[5] W. P. Gaglianone, L. R. Lima, and O. Linton. Evaluating value-at-risk models via quantile regressions. *Banco Central Do Brasil*, 2008.

[6] T. Guldimann. Riskmetrics TM. 1995.

[7] C. J. Hull. Fundamentals of futures and options markets. *Fundamentals of Futures and Options Markets*, 2014.

[8] J. C. Hull. *Options, Futures, and Other Derivatives*. Prentice Hall, 6th edition, 2006.

[9] Y. Kalnishkan and M. Vyugin. The weak aggregating algorithm and weak mixability. *Journal of Computer and System Sciences*, pages 74: 1228–1244, 2008.

[10] R. Koenker. Quantile regression. *Cambridge, UK: Cambridge Univ. Press*, 2005.

[11] R. Koenker and G. Bassett. Regression quantiles. *Econometrica*, pages 46: 33–50, 1978.

[12] P. Kupiec. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, pages 3(2):73–84, 1995.

[13] J. P. Morgan. Riskmetrics technical document. 1996.

[14] L. Rokach. Ensemble-based classifiers. *Artificial Intelligence Review*, pages 33:1–39, 2010.

[15] J. W. Taylor. A quantile regression approach to estimating the distribution of multiperiod returns. *Journal of Derivatives*, pages 7: 64–78, 1999.

[16] V. Vovk. Aggregating strategies. In *Proceedings of the 3rd Annual Workshop on Computational Learning Theory*, pages 371–383, San Mateo, CA, 1990. Morgan Kaufmann.

[17] V. Vovk and F. Zhdanov. Prediction with expert advice for the Brier game. *Journal of Machine Learning Research*, 10:2445–2471, 2009.