A Simple Method for Predicting Covariance Matrices of Financial Returns

Kasper Johansson Mehmet Giray Ogut Markus Pelger
Thomas Schmelzer Stephen Boyd

June 1, 2023

Abstract

We consider the well-studied problem of predicting the time-varying covariance matrix of a vector of financial returns. Popular methods range from simple predictors like rolling window or exponentially weighted moving average (EWMA) to more sophisticated predictors such as generalized autoregressive conditional heteroscedastic (GARCH) type methods. Building on a specific covariance estimator suggested by Engle in 2002, we propose a relatively simple extension that requires little or no tuning or fitting, is interpretable, and produces results at least as good as MGARCH, a popular extension of GARCH that handles multiple assets. To evaluate predictors we introduce a novel approach, evaluating the regret of the log-likelihood over a time period such as a quarter. This metric allows us to see not only how well a covariance predictor does over all, but also how quickly it reacts to changes in market conditions. Our simple predictor outperforms MGARCH in terms of regret. We also test covariance predictors on downstream applications such as portfolio optimization methods that depend on the covariance matrix. For these applications our simple covariance predictor and MGARCH perform similarly.

Contents

1 Introduction

1.1 Covariance prediction

We consider cross-sections, *e.g.*, a vector time series of n financial returns, denoted $r_t \in \mathbf{R}^n$, $t = 1, 2, \ldots$. (We take r_t to be the return from $t-1$ to t.) We focus on the case where the mean is negligible, so the second moment $\mathbf{E}\, r_t r_t^T \in \mathbf{R}^{n\times n}$ is a very good approximation of the covariance $\mathbf{cov}(r_t) = \mathbf{E}\, r_t r_t^T - (\mathbf{E}\, r_t)(\mathbf{E}\, r_t)^T$. This is the case for most daily or weekly stock, bond, and futures returns, factor returns, and index returns. We focus on a modest number of assets, *i.e.*, n on the order 10–100 or so. Our method is readily extended to the case where these assumptions do not hold, *i.e.*, nonnegligible mean return and large n, but for simplicity we focus on the simpler case, and mention these extensions in §7.

We model these returns as independent random variables with zero mean and covariance $\Sigma_t \in \mathbf{S}^n_{++}$ (the set of symmetric positive semidefinite matrices). We focus on the problem of predicting or estimating Σ_t, based on knowledge of $r_1, \ldots, r_{t-1}$. The prediction is denoted as $\hat{\Sigma}_t$. The predicted volatilities of assets are given by $\hat{\sigma}_t = \mathbf{diag}(\hat{\Sigma}_t)^{1/2} \in \mathbf{R}^n$, where $\mathbf{diag}$ of a matrix is the vector of diagonal entries of the matrix, and the squareroot is elementwise. We denote the predicted correlations as

$$\hat{C}_t = \mathbf{diag}(\hat{\sigma}_t)^{-1}\hat{\Sigma}_t\, \mathbf{diag}(\hat{\sigma}_t)^{-1},$$

where $\mathbf{diag}$ of a vector is a diagonal matrix with diagonal entries given by its arguments.

Covariance estimation comes up in several areas of finance, including Markowitz portfolio construction [Mar52, GK00], risk management [MFE15], and asset pricing [Sha64]. Much attention has been devoted to this problem, and a Nobel Memorial Prize in Economic Sciences was awarded for work directly related to volatility estimation [Eng82].

While it is well known that the tails of financial returns are poorly modeled by a Gaussian distribution, our focus here is on the bulk of the distribution, where the Gaussian assumption is reasonable. For future use, we note that the log-likelihood of an observed return r_t, under the Gaussian distribution $r_t \sim \mathcal{N}(0, \hat{\Sigma}_t)$, is

$$l_t(\hat{\Sigma}_t) = \frac{1}{2}\left(-n\log(2\pi) - \log\det\hat{\Sigma}_t - r_t^T\hat{\Sigma}_t^{-1}r_t\right). \tag{1}$$

Roughly speaking, we seek covariance predictors that achieve large values of log-likelihood on realized returns. We will describe evaluation of covariance predictors in detail in §4.

1.2 Contributions

This paper makes three contributions. First, we propose a new method for predicting the time-varying covariance matrix of a vector of financial returns, building on a specific covariance estimator suggested by Engle [Eng02] in 2002. Our method is a relatively simple extension that requires very little tuning and is readily interpretable. It relies on solving a small convex optimization problem, which can be carried out very quickly and reliably. It performs as well as much more complex methods, on several metrics.

Our second contribution is to propose a new method for evaluating a covariance predictor, by considering the regret of the log-likelihood over some time period such as a quarter or month. This approach allows us to evaluate how quickly a covariance estimator reacts to changes in market conditions.

Our third contribution is an extensive empirical study of covariance predictors. We compare our new method to other popular predictors, including rolling window, EWMA, and GARCH type methods. We find that our method performs slightly better than other predictors. However, even the simplest predictors perform well for practical problems like portfolio optimization.

Everything needed to reproduce our results is available, together with an open source implementation of our proposed covariance predictor, online at

https://github.com/cvxgrp/cov_pred_finance.

1.3 Outline

The outline of this paper is as follows. In §2 we describe some common predictors, including the one that our method builds on. In §3 we introduce our covariance predictor. In §4 we discuss methods for validating covariance predictors that measures both overall performance and reactivity to market changes. We describe the data we use in our numerical experiments in §5, and the results in §6. We describe a few extensions of our method in §7.

2 Some common covariance predictors

In this section we review some common covariance predictors, ranging from simple to complex, with the goal of giving context and fixing our notation. To simplify some formulas, we take $r_\tau = 0$ for $\tau \leq 0$.

2.1 Rolling window

The rolling window predictor with window length or memory M is the average of the last $M \geq n$ outer products,

$$\hat{\Sigma}_t = \alpha_t \sum_{\tau=t-M}^{t-1} r_\tau r_\tau^T, \quad t = 2, 3, \ldots,$$

where $\alpha_t = 1/\min\{t-1, M\}$ is the normalization constant. The rolling window predictor can be evaluated via the recursion

$$\hat{\Sigma}_{t+1} = \frac{\alpha_{t+1}}{\alpha_t}\hat{\Sigma}_t + \alpha_{t+1}(r_t r_t^T - r_{t-M} r_{t-M}^T), \quad t = 2, 3, \ldots,$$

with initialization $\hat{\Sigma}_1 = 0$.

For $t < n$, the rolling window covariance estimate is not full rank. To handle this, as well as to improve the quality of the prediction, we can add regularization or shrinkage, for example by adding a positive multiple of $\mathbf{diag}(\hat{\Sigma}_t)$ to our estimate [LW04, LW03], or appproximating the predicted covariance matrix by a diagonal plus low rank matrix, as described in §7.2. Another option is to simply not use the estimate when $\hat{\Sigma}_t$ is singular.

2.2 EWMA

The exponentially weighted moving average (EWMA) estimator, with forgetting factor $\beta \in (0, 1)$, is

$$\hat{\Sigma}_t = \alpha_t \sum_{\tau=1}^{t-1} \beta^{t-1-\tau} r_\tau r_\tau^T, \quad t = 2, 3, \ldots,$$

where

$$\alpha_t = \left(\sum_{\tau=1}^{t-1} \beta^{t-1-\tau} \right)^{-1} = \frac{1-\beta}{1-\beta^{t-1}}$$

is the normalization constant. The forgetting factor β is usually expressed in terms of the half-life $H = -\log 2/\log \beta$, for which $\beta^H = 1/2$. The EWMA predictor is widely used in practice; see, *e.g.*, [MOW11, OS96].

The EWMA covariance predictor can be computed recursively as

$$\hat{\Sigma}_{t+1} = \frac{\beta - \beta^t}{1-\beta^t} \hat{\Sigma}_t + \frac{1-\beta}{1-\beta^t} r_t r_t^T, \quad t = 2, 3, \ldots,$$

with initialization $\hat{\Sigma}_1 = 0$. Like the rolling window predictor, the EWMA predictor is singular for $t < n$, which can be handled using the same regularization methods described above.

2.3 ARCH

The autoregressive conditional heteroscedastic (ARCH) predictor decomposes the return of a single asset as

$$r_t = \mu + \epsilon_t,$$

where μ is the mean return and ϵ_t is the innovation, and models the innovation as

$$\epsilon_t = \sigma_t z_t, \qquad \sigma_t^2 = \omega + \sum_{\tau=1}^{p} a_\tau \epsilon_{t-\tau}^2,$$

where σ_t is the asset volatility, z_t are independent $\mathcal{N}(0, 1)$ and p is the order of the ARCH predictor, often set to one. (Note that this paper assumes $\mu = 0$.) The model parameters are ω and $a_1, \ldots, a_p$. Estimating these parameters requires solving a nonconvex optimization problem [BB22].

ARCH was introduced by Engle in [Eng82]. It set the foundation for a wide variety of popular volatility and correlation predictors and earned him the 2003 Nobel Memorial Prize in Economic Sciences.

2.4 GARCH

In the generalized ARCH (GARCH) predictor the return is decomposed into a mean and innovation as in §2.3 and the volatility σ_t is modeled as

$$\epsilon_t = \sigma_t z_t, \qquad \sigma_t^2 = \omega + \sum_{\tau=1}^{p} a_\tau \sigma_{t-j}^2 + \sum_{\tau=1}^{q} b_\tau \epsilon_{t-\tau}^2,$$

where again $z_t \sim \mathcal{N}(0,1)$, and p and q (often both set to one in practice) determine the GARCH order. The model parameters are ω, $a_1, \ldots, a_p$, and $b_1, \ldots, b_q$. As with ARCH, estimating the model parameters requires solving a nonconvex optimization problem [BB22].

The extension from ARCH to GARCH allows for a more flexible model of time-varying volatility, driven by both past returns and past volatilities. The GARCH predictor was introduced by Bollerslev [Bol86] as a natural extension of Engle's ARCH predictor. It has become one of the most popular tools for modeling time-varying volatilities.

2.5 Multivariate GARCH

More recently, multivariate extensions of the GARCH predictor have been proposed; see, *e.g.*, [BEW88, Bol86, EK95, VDP03, AC97, vdW02, Bol90, Eng02]. There are several ways of extending the GARCH predictor to a multivariate or vector setting. The most commonly used multivariate GARCH (MGARCH) predictor is the dynamic conditional correlation (DCC) predictor. The DCC GARCH predictor expresses the covariance matrix at time t as

$$\Sigma_t = D_t R_t D_t,$$

where D_t is a diagonal matrix of standard deviations, and R_t is the correlation matrix of the standardized returns. DCC GARCH then models the diagonal elements of D_t as separate univariate GARCH processes, and the covariance matrix corresponding to R_t as a constrained multivariate GARCH process. The DCC GARCH predictor was introduced in [Eng02] and has become a popular alternative amongst MGARCH predictors due to its interpretability.

Several other MGARCH predictors have been proposed in the literature. The most straightforward generalization is the VEC predictor, where the covariance matrix is vectorized and each element is modeled as a GARCH process with dependencies on all other elements [BEW88]. However, this extension requires estimating $n(n+1)(n(n+1)+1)/2$ parameters for an n-dimensional covariance matrix, which can be impractical even for a few assets.

Following the VEC extension of GARCH, MGARCH approaches have been proposed in two lines of development [ST09]. The first line involves models that impose restrictions on the parameters of the VEC predictor, including DVEC [Bol86], BEKK [EK95], FF-MGARCH [VDP03], O-GARCH [AC97], and GO-GARCH [vdW02], to name some. However, these predictors have been shown to be hard to fit and can yield inconsistent estimates [BBP03]. (These inconsistencies may not have much practical impact.)

The second line of extensions is the idea of modelling conditional covariances through separate estimates of conditional variances and correlations [ST09]. In [Bol90] Bollerslev introduced the constant conditional correlation predictor (CCC) where the individual asset volatilities are modeled as separate GARCH processes, while the correlation matrix is assumed constant and equal to the unconditional correlation matrix. This predictor was later extended to the dynamic conditional correlation (DCC) predictor where the correlation matrix is allowed to change over time [Eng02]. Variants of the DCC predictor are widely used in finance, where it is also often used in combination with EWMA estimates. Conditional correlation predictors are easier to estimate than other multivariate GARCH predictor, and their parameters are more interpretable. However, these models still require solving nonconvex optimization problems, in a two-step estimation procedure.

For a more detailed discussion of MGARCH predictors we refer the reader to [ST09, BLR06].

2.6 Iterated EWMA

Iterated EWMA (IEWMA) was proposed by [Eng02] and is analogous to DCC GARCH but with EWMA estimates of the volatilities and correlations instead of GARCH. Engle proposed IEWMA as an efficient alternative to the DCC GARCH predictor, although he did not refer to it as IEWMA; we use this term to emphasize its connection to iterated whitening, as proposed in [BB22]. Specifically, IEWMA can be viewed as an iterated whitener, where we first use a diagonal whitener (which estimates the volatilities) and then a full matrix whitener (which estimates the correlations).

First we form an estimate of the volatilities $\hat{\sigma}_t$ using separate EWMA predictors for each squared asset, *i.e.*, entry of r_t^2 (where the square is elementwise), and taking the squareroot. We denote the half-life of these volatility estimates as H^{vol}. We then form the marginally standardized returns as

$$\tilde{r}_t = D_t^{-1} r_t, \tag{2}$$

where $D_t = \mathbf{diag}(\hat{\sigma}_t)$. These vectors should have entries with standard deviation near one.

Then we form a EWMA estimate of the covariance of $\tilde{r}_t$, which we denote as $\tilde{R}_t$, using half-life H^{cor} for this EWMA estimate. (We use the superscript 'cor' since the diagonal entries of $\tilde{R}_t$ should be near one, so $\tilde{R}_t$ is close to a correlation matrix.) From $\tilde{R}_t$ we form its associated correlation matrix R_t, *i.e.*, we scale $\tilde{R}_t$ on the left and right by a diagonal matrix with entries $(\tilde{R}_t)_{ii}^{-1/2}$. Since the diagonal entries of $\tilde{R}_t$ should be near one, $\tilde{R}_t$ and R_t are not too different.

Our IEWMA covariance predictor is

$$\hat{\Sigma}_t = D_t R_t D_t, \quad t = 2, 3, \ldots.$$

This is the covariance predictor proposed by Engle in [Eng02]; replacing R_t with $\tilde{R}_t$ we obtain the iterated whitener proposed by Barratt and Boyd in [BB22]. As mentioned above, they are typically quite close.

It is common to choose the volatility half-life H^{vol} to be smaller than the correlation half-life H^{cor}. The intuition here is that we can average over fewer past samples when we predict the n volatilities $\hat{\sigma}_t$, but need more past samples to reliably estimate the $n(n-1)/2$ off-diagonal entries of R_t. Numerical experiments on real return data confirm that choosing a faster volatility half-life than correlation half-life yields better estimates.

3 Combined multiple iterated EWMAs

In this section we introduce a novel covariance predictor, which we call combined multiple iterated EWMAs, for which we use the acronym CM-IEWMA. The CM-IEWMA predictor is constructed from a modest number of IEWMA predictors, with different pairs of half-lives, which are combined using dynamically varying weights that are based on recent performance.

The CM-IEWMA predictor is motivated by the idea that different pairs of half-lives may work better for different market conditions. For example, short half-lives may perform better in volatile markets, while long half-lives may perform better for calm markets where conditions are changing slowly.

3.1 Dynamically weighted prediction combiner

We first describe the idea in a general setting. We start with K different covariance predictors, denoted $\hat{\Sigma}_t^{(k)}$, $k = 1, \ldots, K$. These could be any of the predictors described above, or predictors of the same type with different parameter values, *e.g.*, half-life (for EWMA) or pairs of half-lives (for IEWMA). In some contexts these different predictors are referred to as a set of K experts [HTF09, JJ94].

We denote the Cholesky factorizations of the associated precision matrices $(\hat{\Sigma}_t^{(k)})^{-1}$ as $\hat{L}_t^{(k)}$, *i.e.*,

$$\left(\hat{\Sigma}_t^{(k)}\right)^{-1} = \hat{L}_t^{(k)}(\hat{L}_t^{(k)})^T, \quad k = 1, \ldots, K,$$

where $\hat{L}_t^{(k)}$ are lower triangular with positive diagonal entries. We will combine these Cholesky factors with nonnegative weights $w_1, \ldots, w_K$ that sum to one, to obtain

$$\hat{L}_t = \sum_{k=1}^{K} w_k \hat{L}_t^{(k)}. \tag{3}$$

From this we recover the weighted combined predictor

$$\hat{\Sigma}_t = \left(\hat{L}_t \hat{L}_t^T\right)^{-1}. \tag{4}$$

We will see below why we combine the Cholesky factors of the precision matrices, and not the covariance or precision matrices themselves.

3.2 Choosing the weights via convex optimization

The log-likelihood (1) can be expressed in terms of the Cholesky factor of the precision matrix $\hat{L}_t$ as

$$l_t(\hat{\Sigma}_t) = -(n/2)\log(2\pi) + \sum_{i=1}^{n} \log \hat{L}_{t,ii} - (1/2)\|\hat{L}_t^T r_t\|_2^2,$$

which is a concave function of $\hat{L}_t$ [BV04, BB22]. This implies that it is a concave function of the weights $w \in \mathbf{R}_+^K$.

We choose the weights at time t as the solution of the convex optimization problem

$$\begin{array}{ll} \text{maximize} & \sum_{j=1}^{N} \left(\sum_{i=1}^{n} \log \hat{L}_{t-j,ii} - (1/2)\|\hat{L}_{t-j}^T r_{t-j}\|_2^2 \right) \\ \text{subject to} & \hat{L}_\tau = \sum_{j=1}^{K} w_j \hat{L}_\tau^{(j)}, \quad \tau = t-1, \ldots, t-N \\ & w \geq 0, \quad \mathbf{1}^T w = 1, \end{array} \tag{5}$$

with variables $w_1, \ldots, w_K$, where N is the look-back, $\mathbf{1}$ denotes the vector with entries one, and $\geq$ between vectors means entrywise. The covariance predictor is then recovered using (3) and (4).

In words: we choose the (mixture) weights so as to maximize the average log-likelihood of the combined prediction over the trailing N periods. The problem (5) is convex, and can be solved very quickly and reliably by many methods [BV04].

We mention several extensions of the weight problem (5). First, we can add one prediction which is diagonal, using any estimates of the volatilities (including constant). This gives us shrinkage, automatically chosen. We can also add a constraint or objective term that encourages the weights to vary smoothly over time.

The CM-IEWMA predictor is a special case of the dynamically weighted prediction combiner described above, where the K predictions are each IEWMA, with different pairs of half-lives H^{vol} and H^{cor}.

4 Evaluating covariance predictors

In this section we discuss evaluation metrics for covariance predictors. The first two metrics are based on a statistical measure, the log-likelihood. The remaining metrics judge a covariance predictor by the performance of a portfolio using a method that depends on a covariance matrix.

4.1 Log-likelihood

A natural way of judging a covariance predictor is via its average log-likelihood on realized returns,

$$\frac{1}{2T}\sum_{t=1}^{T}\Big(-n\log(2\pi) - \log\det\hat{\Sigma}_t - r_t^T\hat{\Sigma}_t^{-1} r_t \Big),$$

with larger values being better. This metric can be used to compare different predictors.

To understand the performance of a covariance predictor over time and changing market conditions, we can examine the average log-likelihood over periods such as quarters, and look at the distribution of quarterly average log-likelihood values. We are particularly interested in poor, *i.e.*, low values.

4.2 Log-likelihood regret

Recall that the empirical covariance

$$\Sigma^{\mathrm{emp}} = \frac{1}{T}\sum_{t=1}^{T} r_t r_t^T$$

maximizes the average log-likelihood, with value

$$\frac{1}{2}\big(-n(\log(2\pi)+1) - \log\det\Sigma^{\mathrm{emp}}\big).$$

For any other constant $\Sigma \in \mathbf{S}_{++}^n$, the log-likelihood is lower than the log-likelihood of Σ^{emp}. We define the *average log-likelihood regret* as the average log-likelihood of the (constant) empirical covariance, minus the average log-likelihood of the covariance predictor. The regret is a measure of how much the covariance predictor $\hat{\Sigma}_t$, $t = 1, \ldots, T$, underperforms the best possible constant covariance predictor (*i.e.*, the sample covariance matrix). The term regret comes from the field of online optimization; see, *e.g.*, [Zin03, MSJR16, HAK07, Haz16].

We want our covariance predictor to have small regret. The regret is typically positive, but it can be negative, *i.e.*, our time-varying covariance can have higher log-likelihood than the best constant one. The regret is not any more useful than the log-likelihood when comparing predictors over one time interval, since it simply adds a constant and switches the sign. But it is interesting when we compute the regret over multiple periods, like months or quarters. The regret over multiple quarters removes the effect of the log-likelihood of the empirical covariance varying due to changing market conditions, and allows us to assess how well the covariance predictor adapts.

4.3 Portfolio performance

We can also judge the performance of a covariance predictor by the investment performance of portfolio construction methods that depend on the estimated covariance matrix. As with log-likelihood or log-likelihood regret, we can examine the portfolio performance in periods such as quarters, to see how evenly the performance is spread over time. One metric of interest is how close the ex-ante and realized portfolio volatility are. The statistical metrics described above are agnostic to the portfolio; with specific real portfolios we can see how well our covariance predictors predict portfolio volatility.

We will assess a covariance predictor using five simple portfolio construction methods. The first is an equally weighted (or $1/n$) portfolio, which does not by itself depend on the covariance, but does when we adjust it with cash to achieve a given ex-ante risk. The second, third, and fourth portfolios depend only on the covariance matrix. They are minimum variance, risk parity, and maximum diversification portfolios. For an in depth discussion of these portfolios, see [Bra15]. The last portfolio we consider is a mean-variance portfolio, using a very simple mean estimator.

For each portfolio we look at four metrics: realized return, volatility, Sharpe ratio, and maximum drawdown of the portfolio. The returns, volatilities, and Sharpe ratios are reported in annualized values. The Sharpe ratio is defined as the ratio of the excess return (over the risk-free rate), divided by the volatility of the excess return,

$$\frac{\frac{1}{T}\sum_{\tau=1}^{T}(r_t^p - r_t^{\mathrm{rf}})}{\left(\frac{1}{T}\sum_{\tau=1}^{T}\left(r_t^p - \frac{1}{T}\sum_{\tau=1}^{T} r_t^p\right)^2\right)^{1/2}},$$

where r_t^p and r_t^{rf} are the portfolio and risk-free returns at time t, and T is the number of time-steps in the evaluation period. The maximum drawdown is defined as

$$\max_{1\leq t_1<t_2\leq T} \frac{V_{t_1}^p}{V_{t_2}^p} - 1,$$

where V_t^p is the portfolio value at time t (with returns re-invested).

In addition to portfolio performance, we can also examine how well the covariance prediction captures the portfolio volatility. We compare the realized portfolio volatility,

$$\left(\frac{1}{T}\sum_{t=1}^{T}(r_t^T w_t)^2\right)^{1/2},$$

to the predicted or ex-ante portfolio volatility,

$$\left(\frac{1}{T}\sum_{t=1}^{T} w_t^T \hat{\Sigma}_t w_t\right)^{1/2},$$

where w_t is the portfolio weights (which sum to one). This directly measures the ability of the estimated covariance matrix to predict portfolio risk.

Equal weight portfolio. We take the equal weight or $1/n$ portfolio with $w = (1/n)\mathbf{1}$. This portfolio does not depend on the covariance $\hat{\Sigma}_t$, but when we mix it with cash, as described below, it will.

Minimum variance portfolio. The (constrained) minimum variance portfolio is the solution of the convex optimization problem

$$\begin{array}{ll} \text{minimize} & w^T\hat{\Sigma}_t w \\ \text{subject to} & w^T\mathbf{1} = 1, \quad \|w\|_1 \leq L_{\max}, \quad w_{\min} \leq w \leq w_{\max} \end{array}$$

with variable w, where $L_{\max} \geq 1$ is a leverage limit, and $w_{\min}$ and $w_{\max}$ are lower and upper bounds on the weights, respectively.

Risk-parity portfolio. The portfolio return volatility $\sigma(w) = (w^T \hat{\Sigma}_t w)^{1/2}$ can be broken down into a sum of volatilities (risks) associated with each asset as

$$\frac{\partial \log \sigma(w)}{\partial w_i} = \frac{\partial \sigma(w)}{\sigma(w)} \frac{w_i}{\partial w_i} = \frac{w_i (\hat{\Sigma}_t w)_i}{w^T \hat{\Sigma}_t w}, \quad i = 1, \ldots, n.$$

The risk parity portfolio is the one for which these volatility attributions are equal [Qia11]. This portfolio can be found by solving the convex optimization problem [BV],

$$\text{minimize} \quad (1/2) x^T \hat{\Sigma}_t x - \sum_{i=1}^{n} (1/n) \log x_i,$$

with variable x, and then taking $w = x^\star / (\mathbf{1}^T x^\star)$.

Maximum diversification portfolio. The diversification ratio of a long-only portfolio (*i.e.*, one with $w \geq 0$) is defined as

$$D(w) = \frac{\hat{\sigma}_t^T w}{(w^T \hat{\Sigma}_t w)^{1/2}}.$$

The diversification ratio tells us how much higher the portfolio volatility would be if all assets were perfectly correlated. The maximum diversification portfolio is the portfolio w that maximizes $D(w)$, possibly subject to constraints [CC08]. Like the risk-parity portfolio, the maximum diversification portfolio can be found via convex optimization. We let $x^\star$ denote the solution of the convex optimization problem [BV]

$$\begin{array}{ll} \text{minimize} & x^T \hat{\Sigma}_t x \\ \text{subject to} & \hat{\sigma}_t^T x = 1, \quad x \geq 0, \end{array}$$

with variable x. The maximum diversification portfolio is $w = x^\star / \mathbf{1}^T x^\star$.

Volatility control with cash. We mix each of the four portfolios described above with cash to achieve a target value of ex-ante volatility σ^{tar}. To do this we start with the portfolio weight vector w_t, and compute its ex-ante volatility $\sigma_t = (w_t^T \hat{\Sigma}_t w_t)^{1/2}$. Then we add a cash component so that the overall ex-ante volatility equals our target, *i.e.*, we use the $(n+1)$ weights (with the last component denoting cash)

$$\left[\begin{array}{c} \theta w_t \\ (1-\theta) \end{array} \right], \qquad \theta = \frac{\sigma^{\text{tar}}}{\sigma_t}.$$

This portfolio will have ex-ante volatility σ^{tar}. Note that the cash weight can be either positive (when it dilutes the portfolio volatility) or negative (when it leverages the portfolio volatility to the desired level). The target volatility σ^{tar} should be chosen so as to avoid portfolios that are either too diluted or too leveraged.

Mean variance portfolio. The last portfolio we consider is a basic mean-variance portfolio, defined as the solution of the convex optimization problem

$$\begin{array}{ll} \text{maximize} & \hat{r}_t^T w \\ \text{subject to} & w^T \hat{\Sigma}_t w \leq \sigma^{\text{tar}} \\ & w^T \mathbf{1} = 1, \quad \|w_{1:n}\|_1 \leq L_{\max}, \quad w_{\min} \leq w \leq w_{\max} \end{array}$$

with variable w, where $\hat{r}_t$ is the predicted mean return vector at time t. The subvector $w_{1:n}$ gives the weights of the non-cash assets. This portfolio does not need cash dilution, since it includes cash in its construction. (If σ^{tar} is chosen appropriately, it will have ex-ante risk σ^{tar}.) The mean-variance portfolio depends not only on a covariance estimate, but also a return estimate. For this we use one of the simplest possible return estimates, a EWMA of the realized returns.

5 Data sets and experimental setup

We illustrate our methods on three different data sets: a set of 49 industry portfolios, a set of 25 stocks, and a set of 5 factor returns, each augmented with cash (with the historical risk-free interest rate). For each data set we show results for six covariance predictors. Everything needed to reproduce the results is available online at

https://github.com/cvxgrp/cov_pred_finance.

5.1 Data sets

Industry portfolios. The first data set consists of the daily returns of a universe of $n = 49$ daily traded industry portfolios, shown in table 1, along with cash. The data set spans July 1st 1969 to December 30, 2022, for a total of 13496 (trading) days. The data was obtained from the Kenneth French Data Library [Fre].

Stocks. The second data set consists of the daily returns of $n = 25$ stocks and cash. The stocks were chosen to be the 25 largest stocks in the S&P 500 at the beginning of 2010, listed in table 2. This data set spans January 4th 2010 to December 30, 2022, for a total of 3272 (trading) days. The stock data was attained through the Wharton Research Data Services (WRDS) portal [WRD23].

Factor returns. The third data set consists of daily returns of the five Fama-French factors taken from the Kenneth French Data Library [Fre], shown in table 3. The data set spans July 1st 1963 to December 30, 2022, for a total 14979 (trading) days.

5.2 Six covariance predictors

For each data set we evaluate six covariance predictors, described below.

Table 1: Industry portfolios.

Agriculture	Food products	Candy & soda
Beer & liquor	Tobacco products	Recreation
Entertainment	Printing and publishing	Consumer goods
Apparel	Healthcare	Medical equipment
Pharmaceutical products	Chemicals	Rubber and plastic products
Textiles	Construction materials	Construction
Steel works etc.	Fabricated products	Machinery
Electrical equipment	Automobiles and trucks	Aircraft
Shipbuilding, railroad equipment	Defense	Precious metals
Non-metallic and industrial metal mining	Coal	Petroleum and natural gas
Utilities	Communication	Personal services
Business services	Computers	Computer software
Electronic equipment	Measuring and control equipment	Business supplies
Shipping containers	Transportation	Wholesale
Retail	Restaurants, hotels, motels	Banking
Insurance	Real estate	Trading
Other		

- Rolling window estimates with 500-, 250-, and, 125-day windows for the industry, stock, and factor data sets, respectively, denoted RW in plots and tables.

- EWMA predictors with 250-, 125-, and, 63-day half-lives, for the industry, stock, and factor data sets, respectively, denoted EWMA.

- IEWMA predictors with half-lives (in days) $H^{\mathrm{vol}}/H^{\mathrm{cor}}$ of 125/250, 63/125, and 21/63 for the three data sets, respectively, denoted IEWMA.

- DCC MGARCH predictor, denoted MGARCH, with parameters re-estimated annually using the `rmgarch` package in R [Gha19].

- CM-IEWMA predictor with $K = 5$ IEWMA predictors and a lookback of $N = 10$ days, with half-lives shown in table 4. For each of the fastest IEWMA predictors we regularize the covariance estimate by increasing the diagonal entries by 5%.

- Prescient predictor, *i.e.*, the empirical covariance for the quarter the day is in. This predictor maximizes log-likelihood for each quarter, and achieves zero regret. It is of course not implementable, and meant only to show a bound on performance with which to compare our implementable predictors.

All the parameters above (*e.g.*, half-lives) are chosen as reasonable values that give good overall performance for each predictor. The results are not sensitive to these choices.

For our experiments we use the first two years (500 data points) of each data set to train the MGARCH predictor and initialize the other predictors. (After this initial MGARCH fit, we re-estimate its parameters annually.) Hence, the evaluation period for our experiments below ranges from June 24th 1971 to December 30, 2022, for the industry portfolios, from

Table 2: List of companies and their tickers.

Ticker	Company Name
XON	Exxon Mobil
WMT	Walmart
AAPL	Apple Inc.
PG	Procter & Gamble
JNJ	Johnson & Johnson
CHL	China Mobile
IBM	IBM
SBC	AT&T
GE	General Electric
CHV	Chevron
PFE	Pfizer
NOB	Noble
NCB	NCR
KO	Coca-Cola
ORCL	Oracle Corporation
HWP	Hewlett-Packard
INTC	Intel Corporation
MRK	Merck & Co.
PEP	PepsiCo
BEL	Becton, Dickinson and Company
ABT	Abbott Laboratories
SLB	Schlumberger
P	Pandora Media
PA	Pan American Silver
MCD	McDonald's

Table 3: The five Fama-French factors.

Factor	**Description**
MKT-Rf	market excess return over risk-free rate
SMB	small stocks minus big stocks
HML	high book-to-market stocks minus low book-to-market stocks
RMW	stocks with high operating profitability minus stocks with low operating profitability
CMA	stocks with conservative investment policies minus stocks with aggressive investment policies

Table 4: Half-lives for CM-IEWMA predictors, given as $H^{\text{vol}}/H^{\text{cor}}$, in days.

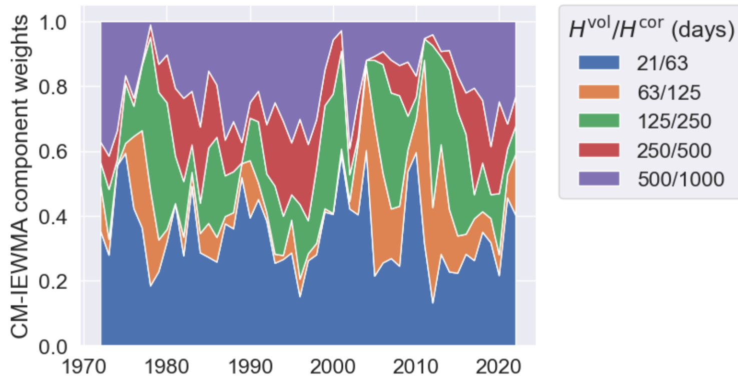
Data set	**Half-lives**				
Industries	21/63	63/125	125/250	250/500	500/1000
Stocks	10/21	21/63	63/125	125/250	250/500
Factors	5/10	10/21	21/63	63/125	125/250

December 28, 2011, to December 30, 2022, for the stock portfolios, and from June 28th 1965 to December 30, 2022, for the factor portfolios.

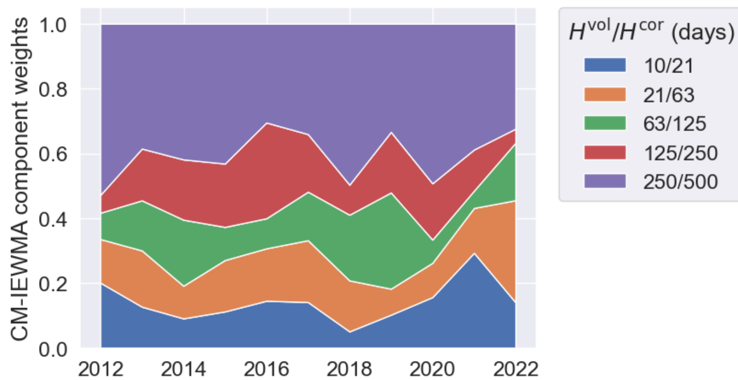
6 Results

6.1 CM-IEWMA component weights

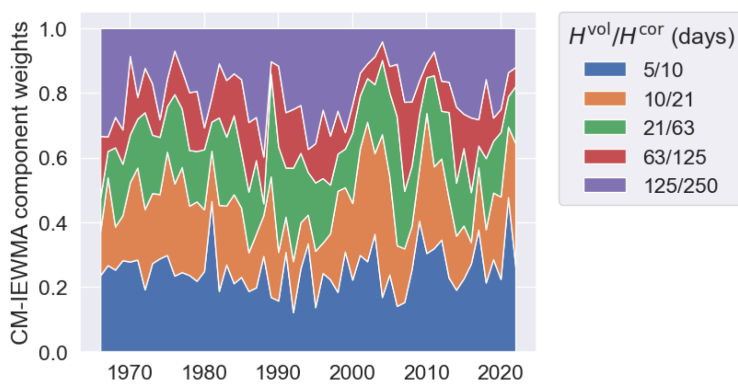
Figure 1 shows the weights attributed to each of the five components of the CM-IEWMA predictors, averaged yearly, for the three data sets. We can see how the predictor adapts the weights depending on market conditions. Substantial weight is put on the slower (longer half-life) IEWMAs most years. During and following volatile periods like the 2000 dot.com bubble or 2008 market crash, we see a significant increase in weight on the faster IEWMAs. We can illustrate these changes in weights in response to market conditions via the effective half-life of the CM-IEWMA, defined as the weighted average of the five (longer) half-lives, shown in figure 2, averaged yearly.



(a) Industry data set.

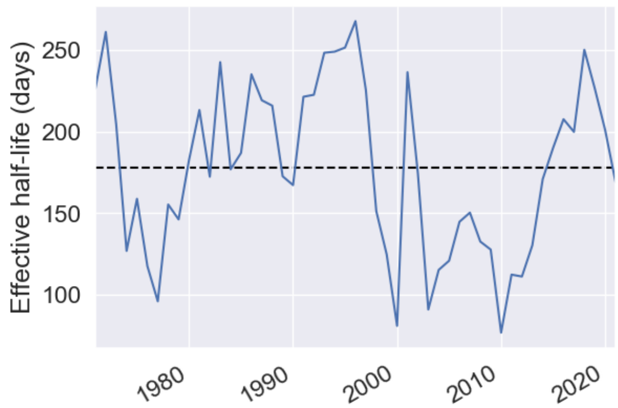


(b) Stock data set.

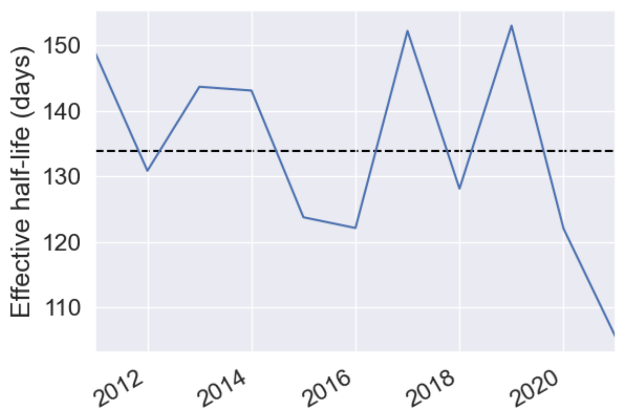


(c) Factor data set.

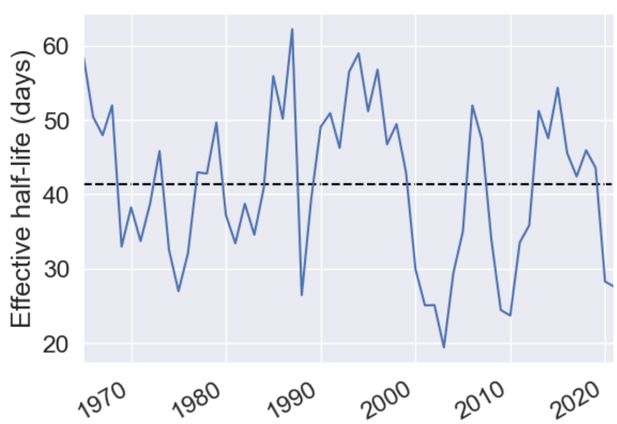
Figure 1: Weights of the various IEWMA components in the CM-IEWMA predictors on three data sets. The IEWMA components are represented as $H^{\text{vol}}/H^{\text{cor}}$ for the volatility and correlation half-lives, respectively.



(a) Industry data set.



(b) Stock data set.



(c) Factor data set.

Figure 2: Effective half-lives of the CM-IEWMA predictor on three data sets.

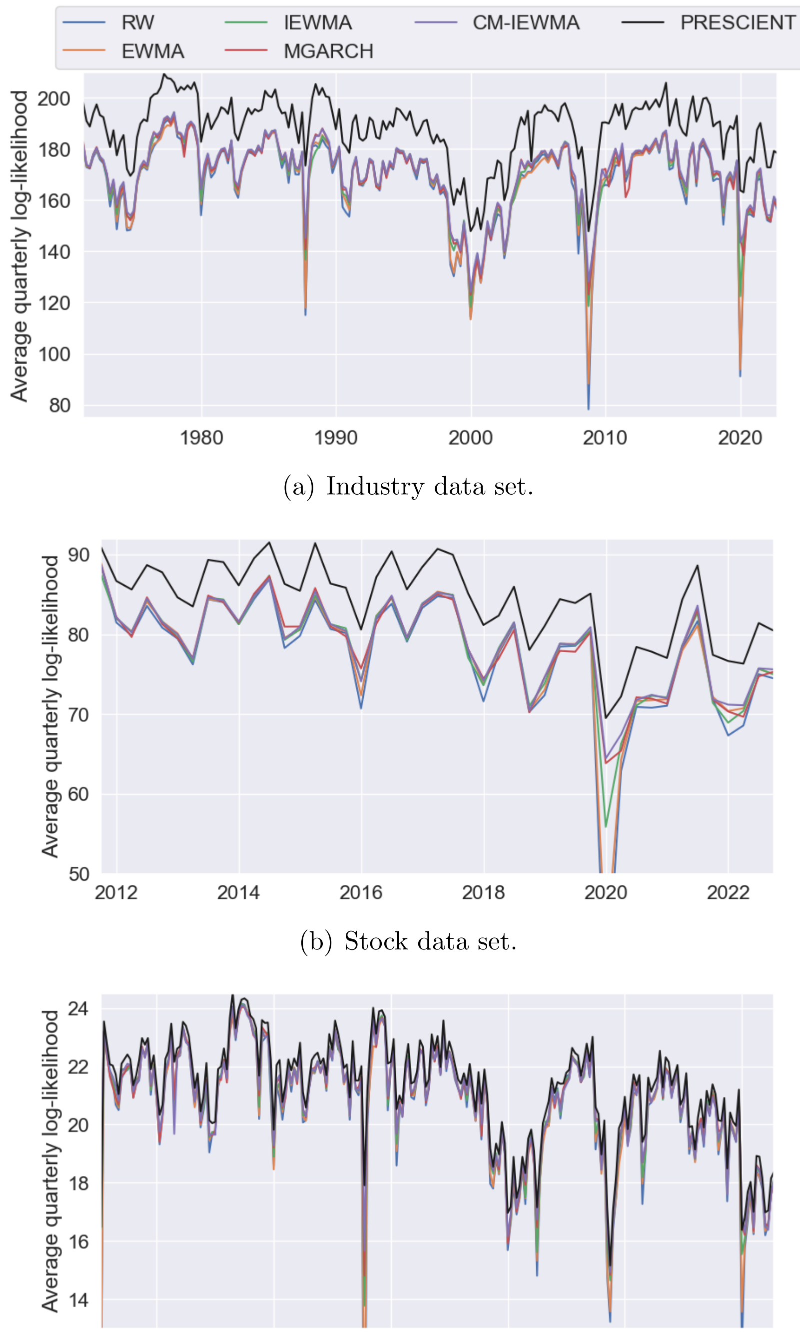
6.2 Log-likelihood and log-likelihood regret

Figure 3 shows the average quarterly log-likelihood for the different covariance predictors over the evaluation period. Not surprisingly, the prescient predictor does substantially better than the others. The different predictors follow similar trends, with even the prescient predictor experiencing a drop in log-likelihood during market turbulence. Close inspection shows that the CM-IEWMA and MGARCH predictors almost always have the highest log-likelihood in each quarter.

Figure 4 shows the average quarterly log-likelihood regret for the different covariance predictors over the evaluation period. Clearly, CM-IEWMA and MGARCH perform best in volatile markets. Figure 5 illustrates the difference between CM-IEWMA and MGARCH. As seen, CM-IEWMA consistently has lower regret on the industry and stock data sets, while they perform similar on the factor data. More precisely, CM-IEWMA has lower regret than MGARCH in 87% of the quarters for the industry data, 71% for the stock data, and 51% for the factor data.

Table 5 illustrates the differences in regret further, by showing the average, standard deviation, and the maximum of the average quarterly regret. As we can see, the average quarterly regret is lower for CM-IEWMA than for the other predictors. The regret is also more stable for CM-IEWMA, as the standard deviation is lower. Finally, the maximum average quarterly regret is also significantly lower for CM-IEWMA than for the other predictors. These results are most prominent on the industry and stock data, while MGARCH does similar on the factor data.

Figure 6 gives a final illustration of these results, by showing the cumulative distribution functions of the average quarterly regret for the different covariance predictors. Clearly, CM-IEWMA has the lowest regret on the industry and stock data set, and MGARCH does similar on the factor data.

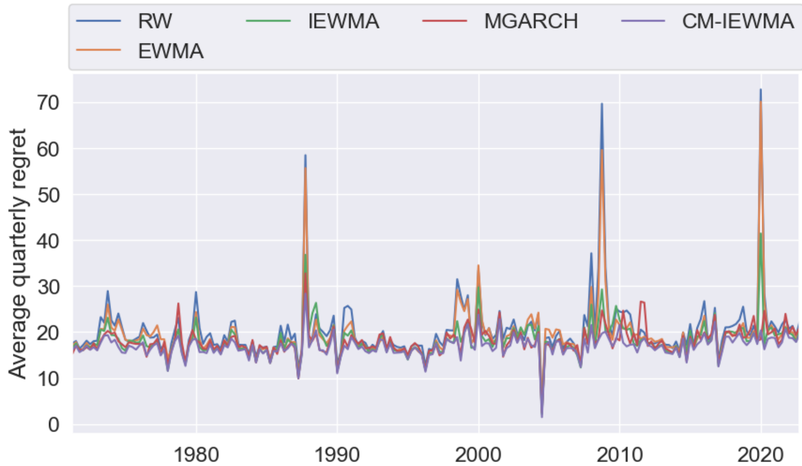


(a) Industry data set.

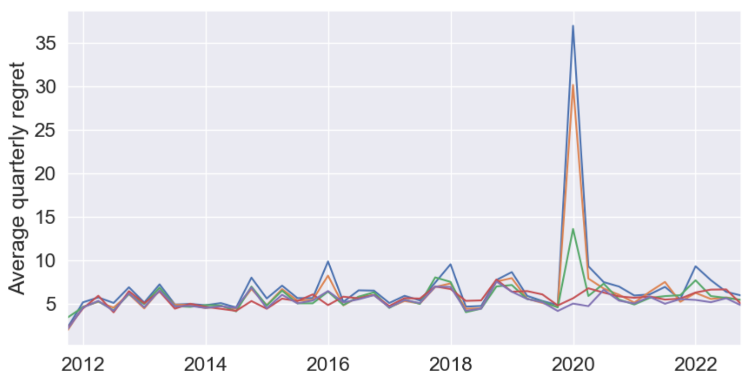
(b) Stock data set.

(c) Factor data set.

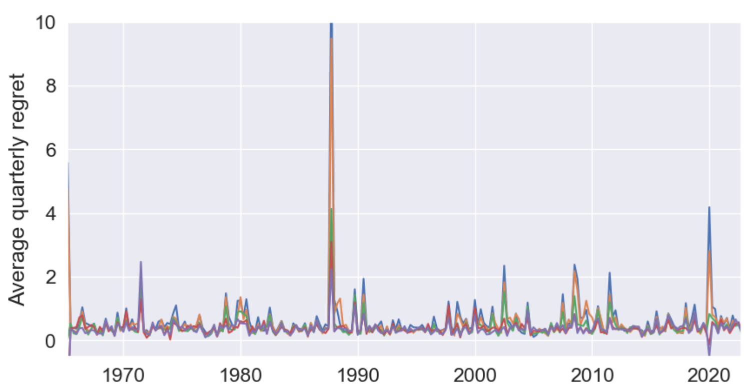
Figure 3: The log-likelihood, averaged quarterly, for six covariance predictors over the evaluation periods for three data sets.



(a) Industry data set.

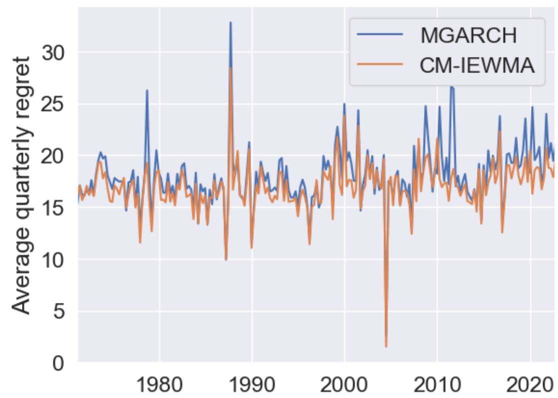


(b) Stock data set.

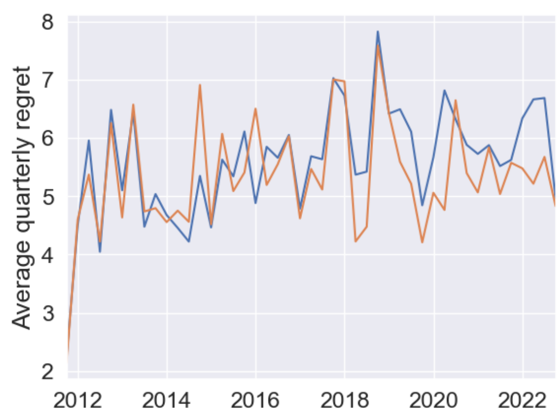


(c) Factor data set.

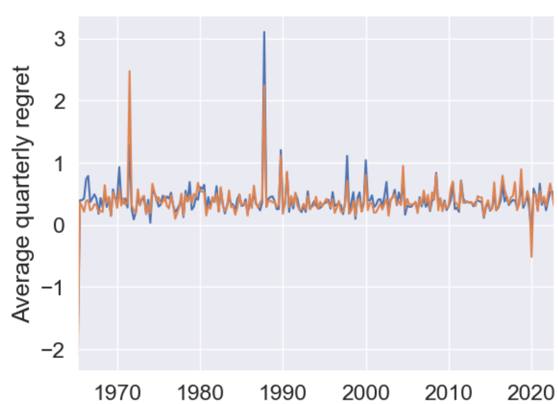
Figure 4: The regret, averaged quarterly, for five covariance predictors over the evaluation periods for three data sets.



(a) Industry data set.



(b) Stock data set.



(c) Factor data set.

Figure 5: The regret for MGARCH and CM-IEWMA, averaged quarterly, over the evaluation periods for three data sets.

Table 5: Metrics on the average quarterly regret for six covariance predictors on three data sets.

Predictor	Average	Std. dev.	Max
RW	20.4	6.9	72.8
EWMA	19.4	6.2	70.1
IEWMA	18.2	3.6	41.4
MGARCH	17.9	3.0	32.8
CM-IEWMA	**16.9**	**2.4**	**28.4**
PRESCIENT	0.0	0.0	0.0

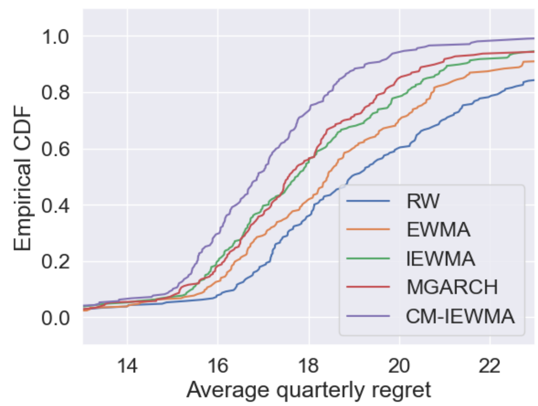
(a): Industry data set.

Predictor	Average	Std. dev.	Max
RW	7.0	4.8	37.0
EWMA	6.2	3.8	30.2
IEWMA	5.8	1.6	13.6
MGARCH	5.6	**1.0**	7.8
CM-IEWMA	**5.3**	**1.0**	**7.6**
PRESCIENT	0.0	0.0	0.0

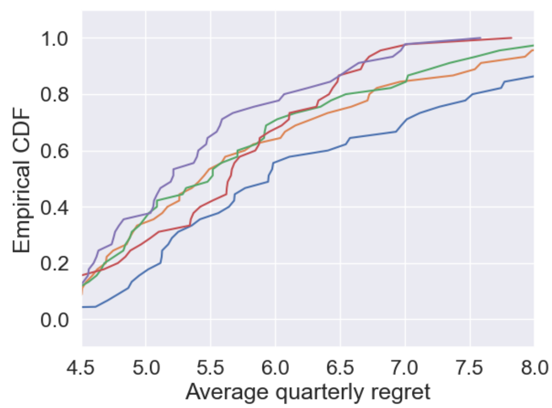
(b): Stock data set.

Predictor	Average	Std. dev.	Max
RW	0.6	0.9	12.2
EWMA	0.6	0.7	9.5
IEWMA	**0.4**	**0.3**	4.1
MGARCH	**0.4**	**0.3**	3.1
CM-IEWMA	**0.4**	**0.3**	**2.9**
PRESCIENT	0.0	0.0	0.0

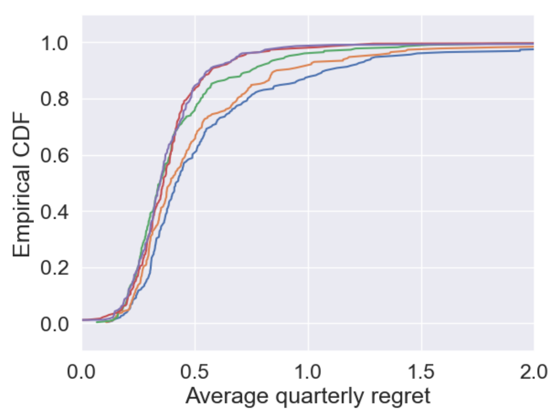
(c): Factor data set.



(a) Industry data.



(b) Stock data set.



(c) Factor data set.

Figure 6: Cumulative distribution functions of average quarterly regret for five covariance predictors on three data sets.

6.3 Portfolio performance

In this section we evaluate the covariance predictors on the portfolios described in §4.3. In the minimum variance and mean-variance portfolios, we use $L_{\max} = 1.6$ (which corresponds to 130:30 long:short), $w_{\min} = -0.1$ and $w_{\max} = 0.15$ for the industry and stock return portfolios, and $w_{\min} = -0.3$ and $w_{\max} = 0.4$ for the factor return portfolio. We use target (annualized) volatilities of 5%, 10%, and 2% for the industry, stock, and factor return portfolios, respectively. For the mean-variance portfolio, our estimated return is the 63-day half-life EWMA of the trailing realized returns.

Equal weight portfolio. Table 6 shows the metrics for the equal weight portfolio. All predictors track the volatility targets well. MGARCH attains the highest Sharpe ratios, although for the industry and factor data sets, the results are very close. The drawdowns are also very similar for all predictors, but MGARCH and CM-IEWMA seem slightly better than the rest.

Minimum variance portfolio. Table 7 shows the metrics for the minimum variance portfolio. For the factor data set, the results are almost identical for all predictors. On the industry and stock data sets, the three EWMA-based predictors follow the volatility target fairly well, while RW and MGARCH underestimate volatility. CM-IEWMA and MGARCH both attain a high Sharpe ratio. However, we note that the high Sharpe ratio for MGARCH, as compared to the other methods, is a consequence of the high volatility. Finally, CM-IEWMA seems to consistently attain a lower drawdown than the other methods, although the other EWMA-based approaches also do well. The EWMA-based predictors seem to track the volatility target well, while RW and MGARCH underestimate the volatility on the industry and stock data sets. On the factor data set all predictors track the volatility target well.

Risk parity portfolio. The results for the risk-parity portfolio are shown in table 8. Overall the results are similar for the various predictors. There is very little that separates the predictors on the industry data set. On the stock data, CM-IEWMA and MGARCH attain the highest Sharpe ratios and lowest drawdowns. On the factor data set, MGARCH has the best overall performance.

Maximum diversification portfolio. The maximum diversification portfolio results are illustrated in table 9. On the industry and stock data sets, CM-IEWMA and MGARCH do best in terms of Sharpe ratio, drawdown, and tracking the volatility target. On the factor data set, MGARCH does best overall.

Mean variance portfolio. The results for the mean-variance portfolio are given in table 10. On the industry data set all predictors underestimate volatility. The results are similar across predictors, with CM-IEWMA and MGARCH performing slightly better than

Table 6: Metrics for the equal weight portfolio performance for six covariance predictors over the evaluation periods on three data sets.

Predictor	Return	Risk	Sharpe	Drawdown
RW	2.2%	5.4%	0.4	16%
EWMA	2.2%	5.1%	0.4	15%
IEWMA	2.2%	5.1%	0.4	15%
MGARCH	2.4%	5.1%	**0.5**	14%
CM-IEWMA	2.3%	5.0%	**0.5**	**13%**
PRESCIENT	4.3%	4.9%	0.9	8%

(a): Industry data set.

Predictor	Return	Risk	Sharpe	Drawdown
RW	6.8%	10.6%	0.6	23%
EWMA	6.4%	10.0%	0.6	21%
IEWMA	6.7%	10.1%	0.7	20%
MGARCH	7.2%	9.4%	**0.8**	**15%**
CM-IEWMA	6.8%	9.6%	0.7	17%
PRESCIENT	12.8%	9.9%	1.3	10%

(b): Stock data set.

Predictor	Return	Risk	Sharpe	Drawdown
RW	2.9%	2.1%	1.4	15%
EWMA	2.9%	2.0%	1.4	15%
IEWMA	3.0%	2.0%	1.5	14%
MGARCH	3.2%	2.0%	**1.6**	**12%**
CM-IEWMA	2.9%	2.1%	1.4	15%
PRESCIENT	3.3%	2.0%	1.7	12%

(c): Factor data set.

Table 7: Metrics for the minimum variance portfolio performance for six covariance predictors over the evaluation periods on three data sets.

Predictor	Return	Risk	Sharpe	Drawdown
RW	3.1%	5.8%	0.5	23%
EWMA	3.1%	5.4%	0.6	**19%**
IEWMA	3.3%	5.5%	0.6	**19%**
MGARCH	4.3%	6.1%	**0.7**	20%
CM-IEWMA	3.5%	5.3%	**0.7**	20%
PRESCIENT	3.8%	5.0%	0.8	13%

(a): Industry data set.

Predictor	Return	Risk	Sharpe	Drawdown
RW	9.7%	12.0%	0.8	23%
EWMA	8.9%	11.1%	0.8	20%
IEWMA	9.7%	11.3%	**0.9**	19%
MGARCH	11.3%	12.3%	**0.9**	18%
CM-IEWMA	9.1%	11.0%	0.8	**15%**
PRESCIENT	15.6%	10.0%	1.6	10%

(b): Stock data set.

Predictor	Return	Risk	Sharpe	Drawdown
RW	1.3%	2.2%	0.6	20%
EWMA	1.4%	2.1%	0.7	18%
IEWMA	1.2%	2.1%	0.6	17%
MGARCH	1.8%	2.1%	**0.9**	**15%**
CM-IEWMA	1.2%	2.1%	0.5	21%
PRESCIENT	1.0%	2.0%	0.5	22%

(c): Factor data set.

Table 8: Metrics for the risk parity portfolio performance for six covariance predictors over the evaluation periods on three data sets.

Predictor	Return	Risk	Sharpe	Drawdown
RW	2.4%	5.4%	**0.5**	16%
EWMA	2.4%	5.1%	**0.5**	15%
IEWMA	2.5%	5.1%	**0.5**	14%
MGARCH	2.7%	5.1%	**0.5**	14%
CM-IEWMA	2.5%	5.0%	**0.5**	**13%**
PRESCIENT	4.7%	4.9%	1.0	8%

(a): Industry data set.

Predictor	Return	Risk	Sharpe	Drawdown
RW	7.4%	10.8%	0.7	22%
EWMA	6.8%	10.1%	0.7	21%
IEWMA	7.2%	10.2%	0.7	20%
MGARCH	7.9%	9.7%	**0.8**	**15%**
CM-IEWMA	7.4%	9.7%	**0.8**	16%
PRESCIENT	14.3%	9.9%	1.5	9%

(b): Stock data set.

Predictor	Return	Risk	Sharpe	Drawdown
RW	1.6%	2.1%	0.7	19%
EWMA	1.7%	2.1%	0.8	18%
IEWMA	1.6%	2.1%	0.8	18%
MGARCH	2.0%	2.1%	**1.0**	**16%**
CM-IEWMA	1.5%	2.1%	0.7	17%
PRESCIENT	1.4%	2.0%	0.7	17%

(c): Factor data set.

Table 9: Metrics for the maximum diversification portfolio performance for six covariance predictors over the evaluation periods on three data sets.

Predictor	Return	Risk	Sharpe	Drawdown
RW	2.1%	5.5%	0.4	16%
EWMA	2.1%	5.1%	0.4	16%
IEWMA	2.2%	5.2%	0.4	14%
MGARCH	2.5%	5.1%	**0.5**	**12%**
CM-IEWMA	2.3%	5.0%	**0.5**	**12%**
PRESCIENT	3.8%	5.0%	0.8	10%

(a): Industry data set.

Predictor	Return	Risk	Sharpe	Drawdown
RW	8.4%	11.2%	0.8	22%
EWMA	7.9%	10.4%	0.8	21%
IEWMA	8.2%	10.4%	0.8	20%
MGARCH	10.0%	9.8%	**1.0**	**15%**
CM-IEWMA	8.8%	10.0%	0.9	16%
PRESCIENT	13.5%	9.9%	1.4	11%

(b): Stock data set.

Predictor	Return	Risk	Sharpe	Drawdown
RW	1.4%	2.2%	0.7	19%
EWMA	1.5%	2.1%	0.7	19%
IEWMA	1.4%	2.1%	0.7	19%
MGARCH	2.0%	2.1%	**1.0**	**16%**
CM-IEWMA	1.4%	2.1%	0.7	18%
PRESCIENT	1.3%	2.0%	0.7	18%

(c): Factor data set.

the rest in terms of Sharpe ratio and drawdown. On the stock data set, the Sharpe ratios are low, likely indicating the difficulty of predicting stock returns in recent years; IEWMA and CM-IEWMA seem to do best overall. On the factor data set, MGARCH seems to do best.

Table 10: Metrics for the mean variance portfolio performance for six covariance predictors over the evaluation periods on three data sets.

Predictor	Return	Risk	Sharpe	Drawdown
RW	7.1%	7.1%	1.0	25%
EWMA	6.8%	6.7%	1.0	23%
IEWMA	7.3%	6.5%	1.1	20%
MGARCH	8.6%	6.6%	**1.3**	18%
CM-IEWMA	7.4%	6.2%	1.2	**16%**
PRESCIENT	3.4%	4.6%	0.7	18%

(a): Industry data set.

Predictor	Return	Risk	Sharpe	Drawdown
RW	2.9%	11.7%	0.2	27%
EWMA	2.6%	11.2%	0.2	25%
IEWMA	3.2%	10.9%	**0.3**	23%
MGARCH	2.3%	11.1%	0.2	21%
CM-IEWMA	3.5%	10.9%	**0.3**	**19%**
PRESCIENT	5.6%	9.3%	0.6	16%

(b): Stock data set.

Predictor	Return	Risk	Sharpe	Drawdown
RW	1.6%	2.1%	0.7	19%
EWMA	1.7%	2.1%	0.8	18%
IEWMA	1.6%	2.1%	0.8	18%
MGARCH	2.0%	2.1%	**1.0**	**16%**
CM-IEWMA	1.5%	2.1%	0.7	17%
PRESCIENT	1.4%	2.0%	0.7	17%

(c): Factor data set.

6.4 Summary

In terms of log-likelihood and regret, CM-IEWMA performs best, followed by MGARCH, which performs better than the simpler covariance predictors.

In downstream portfolio optimization experiments, CM-IEWMA and MGARCH again perform better than the other predictors, although in many cases not significantly. In these experiments there is more variation in the results, partly explained by the difference between our prediction (of a covariance matrix) and our metrics (such as return, risk, drawdown). Even the simplest covariance predictors do a reasonable job of predicting the portfolio risk.

7 Extensions and variations

We mention here several extensions of the proposed method.

7.1 Nonnegligible means

In this paper we have assumed that the mean returns are negligible, at least for the purposes of computing a covariance. In some circumstances this may not be a good assumption, *e.g.*, when working with monthly or quarterly returns. There are many ways to extend the methods described above to handle nonnegligible means; here we outline the most straightforward extension. For more discussion in the context of iterated covariance and mean predictors, see [BB22]. When we take into account nonnegligible means, we will end up with not only a covariance estimate $\hat{\Sigma}_t$, but also a mean estimate $\hat{\mu}_t$, which obviously can be useful in downstream applications such as portfolio construction.

EWMA and IEWMA. EWMA estimates are readily extended to handle nonnegligible means. Given a vector time series $x_t \in \mathbf{R}^n$, we estimate its mean as a EWMA of x_t. We subtract this EWMA mean estimate from x_t to form $\tilde{x}_t$. To obtain a covariance estimate, we form the EWMA of $\tilde{x}_t\tilde{x}_t^T$, using the same half-life as for the mean.

We now explain how this extension affects our method. In the first step of our covariance prediction method, we form a EWMA estimate of the mean and standard deviation of each asset separately, denoted $\omega_{t,i}$ and $d_{t,i}$. We then form the marginally standardized returns as $\tilde{r}_t = D_t^{-1}(r_t - \omega_t)$, where $D_t = \mathbf{diag}(d_{t,1}, \ldots, d_{t,n})$ (which can be compared to (2)). Now we form a EWMA estimate of the mean and covariance of $\tilde{r}_t$. (We expect the means to be small, and the diagonal entries of the covariance to be near one.) We denote these as $\tilde{\mu}_t$ and $\tilde{\Sigma}_t$. To combine these first and second stage EWMA estimates, we can use the method suggested via iterated whitening [BB22], based on

$$r_t = \omega_t + D_t\tilde{r}_t, \qquad \tilde{r}_t \sim \mathcal{N}(\tilde{\mu}_t, \tilde{\Sigma}_t).$$

This gives

$$\hat{\mu}_t = \omega_t + D_t\tilde{\mu}_t, \qquad \hat{\Sigma}_t = D_t\tilde{\Sigma}_tD_t.$$

(It is also possible to modify $\tilde{\Sigma}_t$ to be a correlation matrix, as in §2.6.)

Combining predictions. Taking a nonzero mean into account, the log-likelihood (1) is

$$-(n/2)\log(2\pi) + \sum_{i=1}^{n} \log \hat{L}_{t,ii} - (1/2)\|\hat{L}_t^T(r_t - \hat{\mu}_t)\|_2^2,$$

where $\hat{L}_t$ is the Cholesky factor of $\hat{\Sigma}_t^{-1}$ and $\hat{\mu}_t$ the mean prediction. With the change of variables $\hat{\nu}_t = \hat{L}_t^T \hat{\mu}_t$ the log-likelihood becomes

$$-(n/2)\log(2\pi) + \sum_{i=1}^{n} \log \hat{L}_{t,ii} - (1/2)\|\hat{L}_t^T r_t - \hat{\nu}_t\|_,^2$$

which is jointly concave in $\hat{L}_t^T$ and $\hat{\nu}_t$. With K expert mean and covariance predictors $\hat{L}_t^{(k)}$ and $\hat{\mu}_t^{(k)}$, $k = 1, \dots, K$ we can define the convex combinations

$$\hat{L}_t = \sum_{k=1}^{K} w_k \hat{L}_t^{(k)}, \qquad \hat{\nu}_t = \sum_{k=1}^{K} w_k \hat{\nu}_t^{(k)},$$

where $\hat{\nu}_t^{(k)} = (\hat{L}_t^{(k)})^T \hat{\mu}_t^{(k)}$, and extend the CM-IEWMA of §3 to jointly predict the mean and covariance. We recover the final predicted mean and covariance as

$$\hat{\mu}_t = \hat{L}_t^{-T} \hat{\nu}_t, \qquad \hat{\Sigma}_t = \left(\hat{L}_t \hat{L}_t^T\right)^{-1}.$$

7.2 Large universes

The methods described above can be adapted to handle large universes of assets, say n larger than 100 or so, in two related ways, which we describe here. Both methods end up modeling $\hat{\Sigma}_t$ as a low rank plus diagonal matrix, in so-called factor form.

Traditional factor analysis and model. In practice most return covariance matrices for large universes are constructed from factors, with the model

$$r_t = F_t f_t + z_t, \quad t = 1, 2, \dots,$$

where $F_t \in \mathbf{R}^{n \times r}$ is the factor exposure matrix, $f_t \in \mathbf{R}^r$ is the factor return vector, $z_t \in \mathbf{R}^n$ is the idiosyncratic return, and r is the number of factors, typically much smaller than n. The factor returns are constructed or found by several methods, such as principal component analysis, or by hand; see, *e.g.*, [BN08, Bai03, LP20a, LP20b, PX22b, PX22a, FF93, FF92]. Thus we assume that the factor returns are known. Given the factor returns, the rows of the factor exposure matrix are typically found by least squares regression over a rolling or exponentially weighted window [Coc09]. The idiosyncratic returns z_t are then found as the residuals in this least squares fit. The factor returns f_t are modeled as $\mathcal{N}(0, \Sigma_t^{\mathrm{f}})$, and the idiosyncratic returns z_t are modeled as $\mathcal{N}(0, E_t)$, where E_t is diagonal. It is also assumed that the factor returns and idiosyncratic returns are independent across time and of each other. We end up with a covariance matrix in factor form, *i.e.*, rank r plus diagonal,

$$\Sigma_t = F_t \Sigma_t^{\mathrm{f}} F_t^T + E_t.$$

The factor model has parameters F_t, Σ_t^{f}, and E_t, which all together include $nr+r(r+1)/2+n$ scalar parameters. This can be substantially fewer than the number of scalar parameters in a generic covariance matrix, which is $n(n+1)/2$. (The smaller number of parameters is not the only reason for using a factor model.)

We can easily use the methods described in this paper with a factor model. We simply predict the factor return covariance $\hat{\Sigma}_t^{\mathrm{f}}$ and the idiosyncratic variances $\hat{E}_t$ using the methods described in this paper, and then form the covariance estimate

$$\hat{\Sigma}_t = F_t \hat{\Sigma}_t^{\mathrm{f}} F_t^T + \hat{E}_t.$$

Factor form regularization. Another approach yields the same factor form of the covariance prediction, but directly from the return data, without the need for pre-defined factor returns. We do this by modifying the IEWMA estimates to each produce a factor model, *i.e.*, diagonal plus low rank, as follows. Let R_t denote the correlation prediction derived from the EWMA estimate of $\tilde{R}_t$; see §2.6. To form a rank r plus diagonal approximation of R_t we proceed as follows. First we compute its eigendecomposition $R_t = \sum_{i=1}^n \lambda_i q_i q_i^T$, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. We then form the rank r plus diagonal approximation of this correlation matrix

$$\widehat{R}_t = \sum_{i=1}^r \lambda_i q_i q_i^T + E_t,$$

where the diagonal matrix E_t is chosen so that the diagonal entries of $\widehat{R}_t$ are one, *i.e.*, it is also a correlation matrix. (It can be shown that these are all nonnegative.) Now we simply use $\widehat{R}_t$ in place of R_t. Replacing R_t with its rank r plus diagonal approximation can be considered a sophisticated form of regularization or shrinkage.

Thus, our modified IEWMA predictors all have rank r plus diagonal structure. After combining them exactly as described above, we can optionally approximate our final predicted covariance matrix as low rank plus diagonal.

7.3 Generative mode

Our model can be used to simulate future returns, when seeded by past realized ones. To do this, we start with realized returns for periods $1,\dots,t-1$, and compute $\hat{\Sigma}_t$ using our method. Then we generate or sample r_t^{sim} from $\mathcal{N}(0,\hat{\Sigma}_t)$. We then find $\hat{\Sigma}_{t+1}$ using the returns $r_1,\dots,r_{t-1},r_t^{\mathrm{sim}}$. We generate r_{t+1}^{sim} by sampling from $\mathcal{N}(0,\hat{\Sigma}_{t+1})$. This continues.

This simple method generates realistic return data in the short term. Of course, it does not include shocks or rapid changes in the return statistics that we would see in real data, but the generative return method has several practical applications. To mention just one, we can simulate 100 (say) different realizations over the next quarter (say), and use these to compute 100 performance metrics for our portfolio. This gives us a distribution of the performance metric that we might see over the next quarter.

7.4 Smooth covariance predictions

We mention here a secondary objective for a covariance prediction $\hat{\Sigma}_t$, which is that it vary smoothly across time. To some extent this happens naturally, since whatever method is used to form $\hat{\Sigma}_t$ from $r_1, \ldots, r_{t-1}$ is likely to yield a similar prediction $\hat{\Sigma}_{t+1}$ from $r_1, \ldots, r_t$. It is also possible to further smooth the predictions over time, perhaps trading off some performance, *e.g.*, in log-likelihood regret.

We have already mentioned that the weight optimization problem (5) can be modified to encourage smoothness of the weights over time. We can also directly smooth the prediction $\hat{\Sigma}_t$, to get a smooth version $\hat{\Sigma}_t^{\mathrm{sm}}$. A very simple approach is to let $\hat{\Sigma}_t^{\mathrm{sm}}$ be a EWMA of $\hat{\Sigma}_t$, with a half-life chosen as a trade-off between smoother predictions and performance. This EWMA post-processing is equivalent to choosing $\hat{\Sigma}_t^{\mathrm{sm}}$ to minimize

$$\left\|\hat{\Sigma}_t^{\mathrm{sm}} - \hat{\Sigma}_t\right\|_F^2 + \lambda \left\|\hat{\Sigma}_t^{\mathrm{sm}} - \hat{\Sigma}_{t-1}^{\mathrm{sm}}\right\|_F^2,$$

where λ is a positive regularization parameter used to control the trade off between smoothness and performance. Here the first term is a loss, and the second is a regularizer that encourages smoothly varying covariance predictions.

We can create more sophisticated smoothing methods by changing the loss or the regularizer in this optimization formulation of smoothing. For example we can use the Kullback-Liebler (KL) divergence as a loss. With regularizer $\lambda\|\hat{\Sigma}_t^{\mathrm{sm}} - \hat{\Sigma}_{t-1}^{\mathrm{sm}}\|_F$ (no square in this case), we obtain a piecewise constant prediction, which roughly speaking only updates the prediction when needed. This is a convex optimization problem which can be solved quickly and reliably [BV04].

8 Conclusions

We have introduced a simple method for predicting covariance matrices of financial returns. Our method combines well known ideas such as EWMA, first estimating volatilities and then correlations, and dynamically combining multiple predictions. The method relies on solving a small convex optimization problem (to find the weights used in the combining), which is extremely fast and reliable. The proposed predictor requires little or no tuning or fitting, is interpretable, and produces results better than the popular EWMA estimate, and comparable to MGARCH. Given its interpretability, light weight, and good practical performance, we see it as a practical choice for many applications that require predictions of the covariance of financial returns.

Acknowledgments

Stephen Boyd acknowledges many helpful in-depth discussions with his colleagues Trevor Hastie, Rob Tibshirani, Emmanuel Candes, Mykel Kochenderfer, Kunal Menda, Misha Van Beek, Ron Kahn, and Gabriel Maher. The authors thank Ron Kahn for detailed comments and suggestions on an early draft.

References

[AC97] C. Alexander and A. Chibumba. Multivariate orthogonal factor GARCH. *University of Sussex, Mimeo*, 1997.

[Bai03] J. Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171, 2003.

[BB22] S. Barratt and S. Boyd. Covariance prediction via convex optimization. *Optimization and Engineering*, 2022.

[BBP03] C. Brooks, S. Burke, and G. Persand. Multivariate GARCH models: software choice and estimation issues. *Journal of Applied Econometrics*, 18(6):725–734, 2003.

[BEW88] T. Bollerslev, R. Engle, and J. Wooldridge. A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116–131, 1988.

[BLR06] L. Bauwens, S. Laurent, and J. Rombouts. Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109, 2006.

[BN08] J. Bai and S. Ng. Large dimensional factor analysis. *Foundations and Trends® in Econometrics*, 3(2):89–163, 2008.

[Bol86] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[Bol90] T. Bollerslev. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics*, 72(3):498–505, 1990.

[Bra15] M. Braga. *Risk-based approaches to asset allocation: concepts and practical applications*. Springer, 2015.

[BV] S. Boyd and L. Vandenberghe. Convex optimization additional exercises. `https://github.com/cvxgrp/cvxbook_additional_exercises`.

[BV04] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[CC08] Y. Choueifaty and Y. Coignard. Toward maximum diversification. *The Journal of Portfolio Management*, 35(1):40–51, 2008.

[Coc09] J. Cochrane. *Asset pricing: Revised edition*. Princeton university press, 2009.

[EK95] R. Engle and K. Kroner. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150, 1995.

[Eng82] R. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

[Eng02] R. Engle. Dynamic conditional correlation. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.

[FF92] E. Fama and K. French. The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465, 1992.

[FF93] E. Fama and K. French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

[Fre] K. French. Kenneth french data library. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Research.

[Gha19] A. Ghalanos. rmgarch: Multivariate garch models. *R package version*, pages 1–3, 2019.

[GK00] R. Grinold and R. Kahn. Active portfolio management. 2000.

[HAK07] E. Hazan, A. Agarwal, and S. Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

[Haz16] E. Hazan. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

[HTF09] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

[JJ94] M. Jordan and R. Jacobs. Hierarchical mixtures of experts and the em algorithm. *Neural computation*, 6(2):181–214, 1994.

[LP20a] M. Lettau and M. Pelger. Estimating latent asset-pricing factors. *Journal of Econometrics*, 218(1):1–31, 2020.

[LP20b] M. Lettau and M. Pelger. Factors that fit the time series and cross-section of stock returns. *The Review of Financial Studies*, 33(5):2274–2325, 2020.

[LW03] O. Ledoit and M. Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5):603–621, 2003.

[LW04] O. Ledoit and M. Wolf. Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004.

[Mar52] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[MFE15] A. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press, 2015.

[MOW11] J. Menchero, D. Orr, and J. Wang. *The Barra US equity model (USE4), methodology notes*. MSCI Barra, 2011.

[MSJR16] A. Mokhtari, S. Shahrampour, A. Jadbabaie, and A. Ribeiro. Online optimization in dynamic environments: Improved regret rates for strongly convex problems. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 7195–7201, 2016.

[OS96] J. Ongerstaey and M. Spencer. *Riskmetrics: Technical Document*. JP Morgan and Reuters, 1996.

[PX22a] M. Pelger and R. Xiong. Interpretable sparse proximate factors for large dimensions. *Journal of Business & Economic Statistics*, 40(4):1642–1664, 2022.

[PX22b] M. Pelger and R. Xiong. State-varying factor models of large dimensions. *Journal of Business & Economic Statistics*, 40(3):1315–1333, 2022.

[Qia11] E. Qian. Risk parity and diversification. *The Journal of Investing*, 20(1):119–127, 2011.

[Sha64] W. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964.

[ST09] A. Silvennoinen and T. Teräsvirta. Multivariate GARCH models. In *Handbook of financial time series*, pages 201–229. Springer, 2009.

[VDP03] I. Vrontos, P. Dellaportas, and D. Politis. A full-factor multivariate GARCH model. *The Econometrics Journal*, 6(2):312–334, 2003.

[vdW02] R. van der Weide. GO-GARCH: A multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics*, 17(5):549–564, 2002.

[WRD23] Wharton research data services. https://wrds-web.wharton.upenn.edu/wrds/, 2023.

[Zin03] M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. 2, 2003.