A Simple Method for Predicting Covariance Matrices of Financial Returns

Kasper Johansson Mehmet Ogut Markus Pelger
Thomas Schmelzer Stephen Boyd

Stanford University

Fidelity Investments seminar series, March 21, 2024

Outline

Covariance prediction in finance

Evaluating covariance predictors

Iterated methods

Our method

Empirical study

Extensions and variations

Contributions

- a simple and effective method for predicting covariance matrices of financial returns
- a new method for evaluating a covariance predictor over changing market conditions
- extensive empirical study on several large data sets
- open-source implementation in Python:
  `https://github.com/cvxgrp/cov_pred_finance`

Covariance prediction in finance

Financial returns

- $r_t \in \mathbf{R}^n$ is the vector of n financial asset returns over period t
- $t = 1, \ldots, T$ are the time periods
- could be days, weeks, months, etc.
- $(r_t)_i$ is the return of asset i over period t
- assets could be bonds, stocks, factors, etc.

Gaussian model

model: $r_t \sim \mathcal{N}(0, \Sigma_t)$

- can demean return data if needed
- for most daily, weekly, or monthly return data

$$\Sigma_t = \mathbf{E} r_t r_t^T - (\mathbf{E} r_t)(\mathbf{E} r_t)^T \approx \mathbf{E} r_t r_t^T$$

objective: find estimate $\hat{\Sigma}_t$ of Σ_t, based on $r_1, \ldots, r_{t-1}$

Rolling window (RW) covariance predictor

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

- $\alpha_t = 1/\min\{t-1, M\}$ is the normalizing constant
- M is the RW memory

Exponentially weighted moving average (EWMA) predictor

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

- $\alpha_t = \left(\sum_{\tau=1}^{t-1} \beta^{t-1-\tau}\right)^{-1} = \frac{1-\beta}{1-\beta^{t-1}}$ is the normalizing constant
- $\beta \in (0, 1)$ is the forgetting factor, often expressed in terms of the half-life $H = -\log 2 / \log \beta$

Some more complex predictors

- generalized autoregressive conditional heteroskedasticity (GARCH)
  - introduced in the 1980s [Bollerslev, 1986]
  - models univariate volatility
  - Nobel memorial prize awarded for related work [Engle, 1982]
- MGARCH: multivariate extension of GARCH
- currently considered state-of-the-art for volatility and covariance prediction
- MGARCH requires solving non-convex optimization problems, and involves many parameters difficult to estimate reliably

Evaluating covariance predictors

Mean-squared error

- mean squared error (MSE) of predictions $\hat{\Sigma}_1, \dots, \hat{\Sigma}_T$

$$\frac{1}{T}\sum_{t=1}^{T} \|r_t r_t^T - \hat{\Sigma}_t\|_F^2,$$

  (smaller values are better)
- commonly used in the literature [Patton, 2011]
- MSE best constant predictor is $\Sigma^{\text{emp}} = \frac{1}{T}\sum_{t=1}^{T} r_t r_t^T$

Log-likelihood

- predictions $\hat{\Sigma}_1, \ldots, \hat{\Sigma}_T$ evaluated on average log-likelihood

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

  (larger values are better)
- closely related to (Gaussian) quasi-likelihood (QLIKE) [Patton, 2011; Patton and Sheppard, 2009; Laurent et al., 2013]
- log-likelihood best constant predictor is $\Sigma^{\text{emp}} = \frac{1}{T}\sum_{t=1}^{T} r_t r_t^T$

Log-likelihood regret

- **log-likelihood regret** is the difference between the log-likelihood of the best constant predictor and that of the predictors $\hat{\Sigma}_1, \ldots, \hat{\Sigma}_T$ (smaller values are better)
- useful when we compute the regret over multiple periods, like months or quarters
- the regret over multiple periods removes the effect of the log-likelihood of the empirical covariance varying due to changing market conditions

Portfolio performance

- can evaluate covariance predictor by investment performance
- for example the minimum variance portfolio

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

  with variable w (portfolio weight vector)
- other portfolios: risk-parity, max diversification
- performance metrics: realized return, volatility, Sharpe ratio, max drawdown ...

Volatility control with cash

to more easily compare portfolio performance across different covariance predictors, we mix each portfolio with cash to attain ex-ante volatility target σ^{tar}

1. start with portfolio weight w_t
2. compute ex-ante volatility $\sigma_t = \sqrt{w_t^T \hat{\Sigma}_t w_t}$
3. add a cash component to attain the new $n+1$ weight vector

$$\begin{bmatrix} \theta w_t \\ (1-\theta) \end{bmatrix}, \qquad \theta = \frac{\sigma^{\text{tar}}}{\sigma_t}$$

Iterated methods

Iterated covariance predictors

1. form initial estimate $\hat{\Sigma}_t^{(1)}$ of Σ_t
2. form "whitened" returns

$$\tilde{r}_t = \left(\hat{\Sigma}_t^{(1)}\right)^{-1/2} r_t, \quad t = 1, \ldots, T$$

3. form estimate $\hat{\Sigma}_t^{(2)}$ of covariance of $\tilde{r}_t$
4. final estimate

$$\hat{\Sigma}_t = \left(\hat{\Sigma}_t^{(1)}\right)^{1/2} \hat{\Sigma}_t^{(2)} \left(\hat{\Sigma}_t^{(1)}\right)^{1/2}$$

- variation: let $\hat{\Sigma}_t^{(2)}$ be correlation matrix of $\tilde{r}_t$ [Engle, 2002]
- can iterate [Barratt and Boyd, 2022]

Iterated EWMA (IEWMA) predictor

1. $\Sigma_t^{(1)}$ is diagonal matrix of variances of r_t
2. form $\left(\hat{\Sigma}_t^{(1)}\right)_{ii}$ as EWMA of $(r_t)_i^2$ using half-life H^{vol}
3. volatility adjusted returns

$$\tilde{r}_t = \left(\hat{\Sigma}_t^{(1)}\right)^{-1/2} r_t, \quad t = 1, \dots, T$$

4. form $\hat{\Sigma}_t^{(2)}$ as EWMA covariance of $\tilde{r}_t$ using half-life H^{cor}

- two parameters: H^{vol} and H^{cor}
- proposed in [Engle, 2002]

Our method

Dynamically weighted prediction combiner

1. start with K covariance predictors $\hat{\Sigma}_t^{(k)}, \quad k = 1, \ldots, K$
2. Cholesky factorizations of associated precision matrices

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

3. create convex combination

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

   where $\pi_k \geq 0$ and $\sum_{k=1}^{K} \pi_k = 1$
4. recover covariance predictor as $\hat{\Sigma}_t = \left(\hat{L}_t \hat{L}_t^T\right)^{-1}$

Choosing the weights via convex optimization

- choose weights π at time t to maximize log-likelihood over past N time-steps

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

- convex problem that can be solved quickly and reliably by many methods

Combined multiple iterated EWMA (CM-IEWMA)

1. choose K half-life pairs H_k^{vol} and H_k^{cor}, $k = 1, \ldots, K$
2. form the K IEWMA predictors $\hat{\Sigma}_t^{(k)}$ for these half-life pairs
3. combine the IEWMAs using the dynamically weighted prediction combiner to get the prediction $\hat{\Sigma}_t = \left(\hat{L}_t \hat{L}_t^T\right)^{-1}$

- parameters: half-life pairs and lookback N

Empirical study

Data set and experimental setup

- data: $n = 49$ daily industry portfolio returns 1970–2023, $T = 13{,}496$ trading days
- compare six covariance predictors
  - RW with a 500-day window
  - EWMA with 250-day half-life
  - IEWMA with half-lives $H^{\text{vol}}/H^{\text{cor}}$ of 125/250 (in days)
  - MGARCH with parameters re-estimated annually
  - CM-IEWMA with $K = 5$ predictors with half-lives (in days):

H^{vol}	21	63	125	250	500
H^{cor}	63	125	250	500	1000

- results on other data sets like stocks and factors are qualitatively similar

Mean-squared error

Predictor	Average/10^{-4}	Std. Dev./10^{-3}	Max/10^{-2}
RW	7.6	4.0	3.9
EWMA	7.5	4.0	3.9
IEWMA	7.4	3.9	3.9
MGARCH	**6.8**	**3.6**	**3.8**
CM-IEWMA	6.9	**3.6**	**3.8**

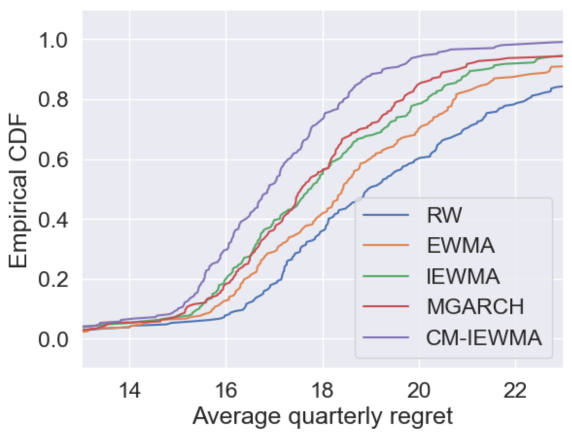
- metrics on quarterly MSE, over 212 quarters
- CM-IEWMA and MGARCH perform best

Log-likelihood regret

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**

- metrics on quarterly regret
- CM-IEWMA performs best

Log-likelihood regret continued



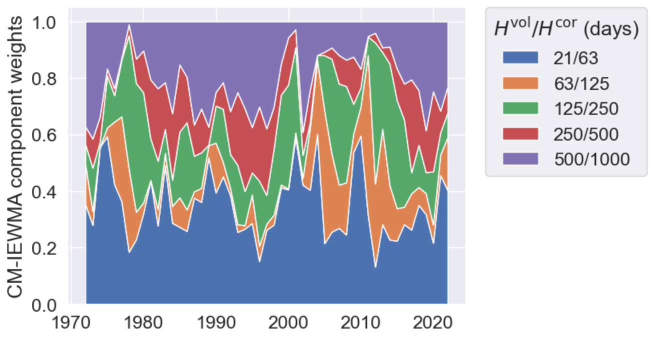
- empirical CDF of quarterly regret (higher is better)

Minimum variance portfolio performance metrics

Predictor	Return	Risk	Sharpe
RW	3.1%	5.8%	0.5
EWMA	3.1%	5.4%	0.6
IEWMA	3.3%	5.5%	0.6
MGARCH	4.3%	6.1%	**0.7**
CM-IEWMA	3.5%	5.3%	**0.7**

- minimum variance portfolios cash-adjusted to 5% risk target
- similar performance across predictors
- CM-IEWMA estimates risk better than the other predictors

CM-IEWMA component weights π



- average weight π_i, $i = 1, \ldots, 5$ on the five predictors each year
- substantial weight is put on the slower (longer half-life) IEWMAs most years
- during and following volatile periods we see a significant increase in weight on the faster IEWMAs

Extensions and variations

Some practical extensions and variations

- realized covariance
  - uses intraperiod returns
- large universes
  - when n is larger than 100 or so
- smoothing
  - penalize variation in covariance estimate

Realized covariance

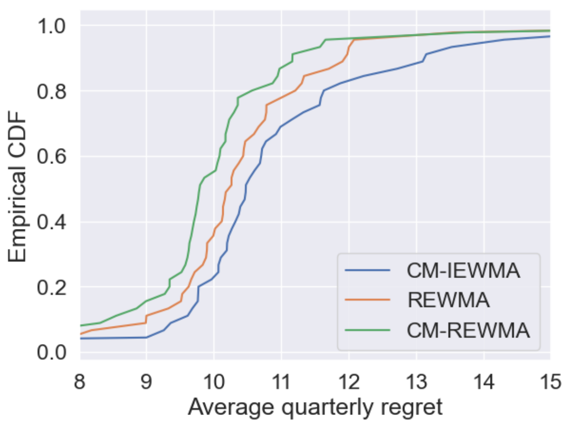
- $r_t \in \mathbf{R}^{n\times m}$ return matrix at time t, with columns that are m intraperiod return vectors
- $C_t = r_t r_t^T$ realized covariance at time t
- realized EWMA (REWMA):

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

- CM-REWMA combines REWMAs with different half-lives

Realized covariance empirical results

- $n = 39$ stocks and $m = 77$ intraperiod returns, January 2 2004 to December 30 2016
- CM-IEWMA gives improvement here too



Large universes

- in practice, the number of assets n can be very large
- we describe two closely related methods for large universes
  - traditional factor model
  - fitting a factor model to a (given) covariance matrix
- computational cost of portfolio optimization reduced from $\mathcal{O}(n^3)$ to $\mathcal{O}(nk^2)$ when using a k-factor model [Boyd and Vandenberghe, 2004]

Traditional factor model

- model: $r_t = F_t f_t + z_t, \quad t = 1, 2, \ldots,$
  - $F_t \in \mathbf{R}^{n \times k}$ factor loadings
  - $f_t \in \mathbf{R}^k$ factor returns
  - $z_t \in \mathbf{R}^n$ idiosyncratic return
- we end up with covariance of low-rank plus diagonal form

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

  - Σ_t^{f} factor return covariance
  - E_t diagonal matrix of idiosyncratic variances
- never have to store $n \times n$ covariance

Fitting a factor model to a covariance matrix

- given covariance Σ
- find one in factor form, $\hat{\Sigma} = FF^T + E$, such that the Kullback-Leibler divergence between $\mathcal{N}(0, \Sigma)$ and $\mathcal{N}(0, \hat{\Sigma})$,

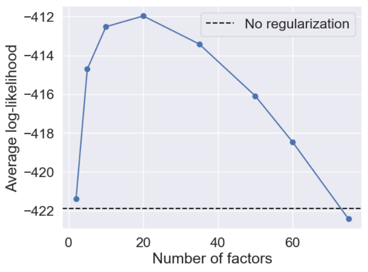
$$\mathcal{K}(\Sigma, \hat{\Sigma}) = \frac{1}{2}\left(\log \frac{\det \hat{\Sigma}}{\det \Sigma} - n + \mathbf{Tr}\, \hat{\Sigma}^{-1}\Sigma\right)$$

  is minimized
- equivalent to maximizing the expected log-likelihood of $r \sim N(0, \Sigma)$ under the model $\mathcal{N}(0, \hat{\Sigma})$
- can be solved via the expectation maximization algorithm (suggested and derived by Emmanuel Candès)

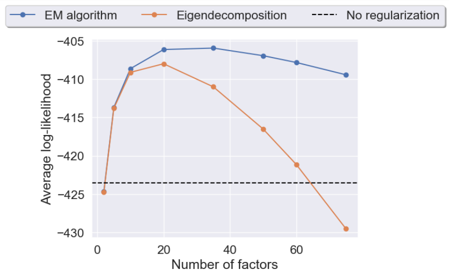
Large universes: empirical setup

- 238 US stocks over 5787 trading days
- traditional factor model
  - create factor model using PCA on two years of data, refitted annually
  - we use k factors and use the CM-IEWMA with half-lives (in days) $H^{\text{vol}}/H^{\text{cor}}$ of $\lceil k/2 \rceil / k, k/3k$, and $3k/6k$, to compute the factor covariance
- fitting factor model to covariance
  - use CM-IEWMA directly with half-lives (in days) $H^{\text{vol}}/H^{\text{cor}}$ of 63/125, 125/250, 250/500, and 500/1000
  - approximate CM-IEWMA predictor using factor model

Large universes: empirical results



traditional factor model



fitting factor model to covariance

Smooth covariance predictions

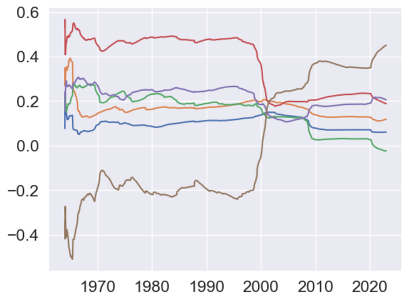
- given predictions $\hat{\Sigma}_t$, $t = 1, 2, \ldots$,
- let $\hat{\Sigma}_t^{\text{sm}}$ be the EWMA of $\hat{\Sigma}_t$
  - equivalent to minimizing

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

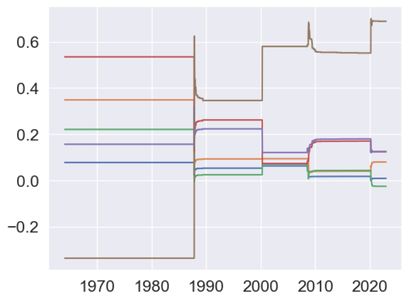
    where λ is a smoothing parameter
  - yields smooth covariance predictions
- with regularizer $\lambda \| \hat{\Sigma}_t^{\text{sm}} - \hat{\Sigma}_{t-1}^{\text{sm}} \|_F$, we obtain piecewise constant predictions
- smoothing can lead to reduced trading and improved portfolio performance

Smooth covariance predictions empirical results

- minimum variance portfolios on five Fama-French factor returns
- portfolio weights for smooth and piecewise constant covariances



smooth



piecewise constant

Conclusions

- introduced a covariance predictor for financial returns
- relies on solving a small convex optimization problem
- requires little or no tuning or fitting
- interpretable, lightweight, and practically effective
- outperforms popular EWMA and is comparable to MGARCH

Try it out!

```
from cvx.covariance.combination import from_ewmas
halflife_pairs = [(10, 21), (21, 63), (63, 125)]
combinator = from_ewmas(returns, halflife_pairs)
covariances = {}
for predictor in combinator.solve(window=10):
    covariances[predictor.time] = predictor.covariance
```

https://github.com/cvxgrp/cov_pred_finance

Thank you!

Questions?