A Simple Method for Predicting Covariance Matrices of Financial Returns

Qualifying Examination

Department of Electrical Engineering, Stanford University

Kasper Johansson

September 2023

Outline

Covariance prediction in finance

Our method

Empirical study

Covariance prediction in finance

Financial returns (vector time-series)

- $r_t \in \mathbf{R}^n$ 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.

Financial covariance

- model $r_t \sim \mathcal{N}(0, \Sigma_t)$
- Σ_t is well approximated by $\mathbf{E} r_t r_t^T$, the second moment
- common for most daily, weekly, or monthly return data
- **objective:** given $r_1, \ldots, r_{t-1}$, estimate covariance Σ_t
- denote $\hat{\Sigma}_t$ as the prediction for Σ_t

Evaluating covariance predictors

- 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)

- best constant predictor is $\Sigma^{\text{emp}} = \frac{1}{T}\sum_{t=1}^{T} r_t r_t^T$
- the *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)
- covariance predictors can also be evaluated by the investment performance of portfolio construction methods

Rolling window (RW) covariance predictor

- RW:
$$\hat{\Sigma}_t = \alpha_t \sum_{\tau=t-M}^{t-1} r_\tau r_\tau^T, \quad t = 2, 3, \ldots,$$
- $\alpha_t = 1/\min\{t-1, M\}$ is a normalizing constant
- M is the RW memory

Exponentially weighted moving average (EWMA) predictor

- EWMA:

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

- $\alpha_t = \left(\sum_{\tau=1}^{t-1} \beta^{t-1-\tau}\right)^{-1} = \frac{1-\beta}{1-\beta^{t-1}}$ is a 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 on volatility and covariance prediction
- MGARCH requires solving non-convex optimization problems, and involves many parameters difficult to estimate reliably

Our method

Iterated covariance predictors

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

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

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

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

- can iterate [Barratt and Boyd, 2022]

Iterated EWMA (IEWMA) predictor

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

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

- form $\hat{\Sigma}_t^{(2)}$ as EWMA covariance of $\tilde{r}_t$ using half-life H^{cor}
- two parameters: H^{vol} and H^{cor}

Dynamically weighted prediction combiner

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

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

- create the convex combination

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

  where $\pi_k \geq 0$ and $\sum_{k=1}^{K} \pi_k = 1$
- 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

A novel covariance predictor for financial returns: Combined multiple iterated EWMA (CM-IEWMA)

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

Empirical study

Data set and experimental setup

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

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

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**

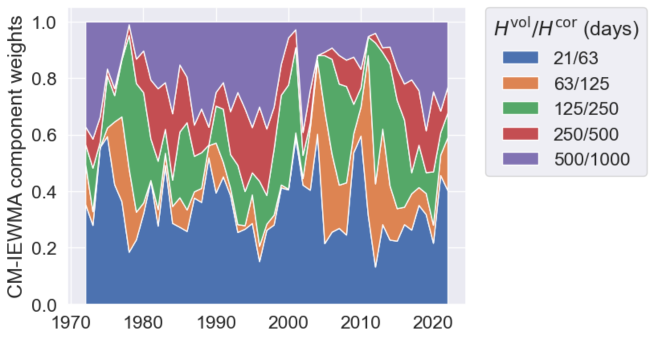
- metrics on the average quarterly regret from 1970–2023
- CM-IEWMA has the lowest average regret, lowest standard deviation, and lowest maximum regret

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-diluted 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
- CM-IEWMA automatically adjusts to market conditions

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

Thank you!

Questions?

Appendix

Extensions and variations

- large universes
  - when n is larger than 100 or so
  - factor models can reduce computational cost and improve log-likelihood and portfolio performance
- smoothing
  - penalize variation in covariance estimate
  - can improve log-likelihood and portfolio performance
  - significantly reduces trading
  - can attain smoothly varying or piecewise constant covariance predictors
- simulating returns
  - our predictor can be used to simulate future returns
  - can generate realistic portfolio scenarios and distributions on metrics

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
- factor returns constructed by several methods, like principal component analysis (PCA), or by hand.
- 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

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-Liebler 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 $r \sim \Sigma$ under the model $\mathcal{N}(0, \hat{\Sigma})$, and can be solved via the expectation maximization algorithm
- results in iterative algorithm of matrix multiplications [1]

[1] this method was suggested and derived by Emmanuel Candès

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 a piecewise constant prediction

Simulating returns

- start with realized returns for periods $1, \ldots, t-1$
- compute $\hat{\Sigma}_t$
- generate sample r_t^{sim} from $\mathcal{N}(0, \hat{\Sigma}_t)$
- find $\hat{\Sigma}_{t+1}$ from $r_1, \ldots, r_{t-1}, r_t^{\text{sim}}$
- generate r_{t+1}^{sim} from $\mathcal{N}(0, \hat{\Sigma}_{t+1})$
- repeat
- generates realistic returns that can be used, for example, to simulate different realizations of portfolio metrics