Advanced Empirical Finance: Topics and Data Science

Stefan Voigt

Spring 2024

University of Copenhagen and Danish Finance Institute

Volatility estimation

Scope – what can we say about Σ?

Definition and Relevance

Definitions

- Any time series can be decomposed into a predictable part an unpredictable part

$$r_t = E\left(r_t|F_{t-1}\right) + \varepsilon_t$$

- In the last part of the lecture we focused on the conditional mean $E\left(r_t|F_{t-1}\right)$
- This part: The conditional variance of r_t

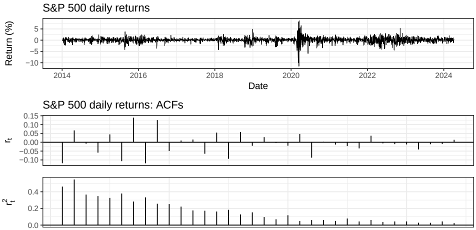
$$\sigma_t^2 = \text{Var}\left(r_t|F_{t-1}\right) = E\left(\varepsilon_t^2|F_{t-1}\right)$$

Recap

- $\hat{\Sigma}_t$ is relevant for the *portfolio choice problem*
- $\hat{\sigma}_t^2$ is also a crucial component for *option pricing* - recall Black-Scholes
- Obvious problem: σ_t^2 is not observable

What do we know about σ_t^2?

```r
prices <- tidyquant::tq_get("SPY", get = "stock.prices")
returns <- prices |>
  mutate(log_price = log(adjusted),
         return = log_price - lag(log_price)) |>
  select(symbol, date, return) |>
  drop_na()
```



- Returns are uncorrelated but *not* independent!
- Volatility clustering & persistence

Conditional Heteroscedastic Models

- The manner under which σ_t^2 evolves over time distinguishes one volatility model from another
- I will focus on conditional heteroscedastic models that use an exact function to govern the evolution of σ_t^2, but there are also model with stochastic components (stochastic volatility)

1. ARCH (autoregressive conditional heteroscedasticity)
2. GARCH (generalized ARCH)

Volatility model building[1]

Building a volatility model for an asset return series consists of four steps:

1. Specify a mean equation (e.g., multi-factor model) or test for serial dependence in the data and, if necessary, build an econometric model (e.g., an ARMA model) for the return series to remove any linear dependence
2. Use the residuals of the mean equation to test for autoregressive conditional heteroskedasticity (ARCH) effects
3. Specify a volatility model if ARCH effects are statistically significant, and perform a joint estimation of the mean and volatility equations
4. Check the fitted model carefully and refine it if necessary

[1] Consult Tsay (2010) for a textbook treatment

The ARCH model (Engle, 1982)

- Describes the dependence of ε_t by an ARCH(m) model

$$\varepsilon_t = \sigma_t z_t \qquad \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_m \varepsilon_{t-m}^2$$

- where z_t is a sequence of independent and identically distributed (iid) random variables with mean zero and unit variance
- (common assumption for MLE: Gaussian or t-distributed z_t)
- Intuition: large past shocks ε_{t-1}^2 imply higher conditional variance (clustering)

Parameter considerations

- The sequence of parameters α_i must satisfy some conditions to guarantee positivity of σ_t^2: $\alpha_0 \geq 0$, and $\alpha_i > 0 \forall i > 0$
- The coefficients α_i must satisfy some regularity conditions to ensure that the unconditional variance of ε_t is finite / σ_t^2 is stationary (all roots to k of $1 - \alpha_1 k - \ldots - \alpha_m k^p = 0$ are outside the unit circle)
- The unconditional variance is given by $\sigma_\varepsilon^2 = \frac{\alpha_0}{1 - \sum_{i=1}^m \alpha_i}$

The GARCH model (Bollerslev, 1986)

- Problem of the ARCH model: typical financial time series often require highly parameterized ARCH models
- The GARCH(p,q) model is given by

$$\varepsilon_t = \sigma_t z_t \qquad \sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2$$

- Any stationary GARCH(p,q) model can be written as an ARCH(∞) model

Parameter considerations

- positive σ_t^2 is ensured by $\alpha_0 \geq 0, \alpha_j \geq 0, \beta_i \geq 0$
- Similar conditions for stationarity as for ARCH: (all roots of $1 - \alpha_1 k - \ldots - \alpha_p k^p - \beta_1 k - \ldots - \beta_q k^q = 0$ are outside the unit circle)

Forecasting

- Consider a GARCH(1, 1) model
- At time t, the 1-step ahead forecast is

$$\sigma_t^2(1) = E(\sigma_{t+1}^2 | F_t) = \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma_t^2$$

- Note: You can rewrite the GARCH(1, 1) equation with $z_t = \varepsilon_t / \sigma_t$ as

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 \sigma_t^2 z_t^2 + \beta_1 \sigma_t^2 = \alpha_0 + (\alpha_1 + \beta_1)\sigma_t^2 + \alpha_1 \sigma_t^2 (z_t^2 - 1)$$

- Then, because $E(z_t^2 - 1) = 0$, repeated substitution yields

$$\sigma_t^2(l) = E(\sigma_{t+l}^2 | F_t) = \alpha_0 + (\alpha_1 + \beta_1)\sigma_t^2(l-1)$$

  Therefore, for $l \to \infty$

$$\sigma_t^2(l) \to \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

Model estimation

- Log likelihood function under the normality assumption for ε_t

$$\begin{aligned}\log L(\theta) &= \log f\left(r_1|\theta\right) + \log f\left(r_2|r_1,\theta\right) + \ldots + \log f\left(r_T|r_{t-1},\ldots,r_1,\theta\right)\\ &= -\frac{T}{2}\log(2\pi) - \sum_{t=1}^{T}\left(\log(\sigma_t^2) + \frac{(r_t-\mu_t)^2}{2\sigma_t^2}\right)\end{aligned}$$

  where σ_t^2 and μ_t^2 are computed recursively
- Similar of course for different likelihood specifications, e.g. student t (consult Tsay (2010) for specific estimation issues)

Maximum likelihood estimation

- Consult the exercises for hands-on examples

1. Define (negative) log-likelihood function
2. Use `optim` (R) or `minimize` (Python) to find maximum likelihood estimator (in some cases, constrained optimization is required)
3. Retrieve information matrix (`hessian = TRUE` (R) or (`hess_inv` (Python))
4. Compute standard errors

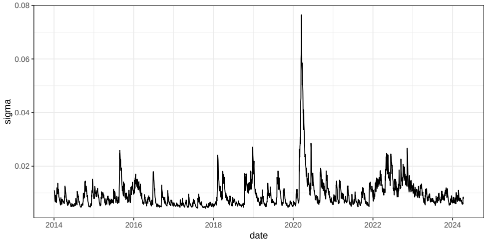
```
#R
fit <- optim( ... ) # fitted mle object
sqrt(diag(solve(fit$hessian)))
```

(G)ARCH in R: `rugarch`

```r
library(rugarch)
model.spec <- ugarchspec(variance.model = list(model = 'sGARCH' ,
                                               garchOrder = c(1 , 1)),
                         mean.model = list(armaOrder = c(0 , 0)))
model.fit <- ugarchfit(spec = model.spec , data = returns |> pull(return))
```

Implementation in R

```r
tibble(returns, sigma = sigma(model.fit)) |> ggplot(aes(x = date, y = sigma)) + geom_line()
```



Small example: Value-at-risk

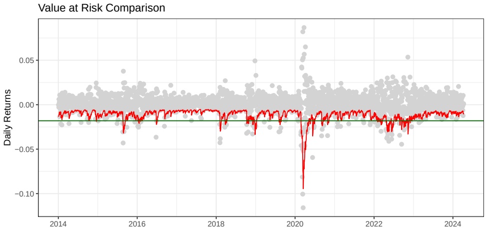
- We define the VaR (Value-at-risk) of a financial position over a given time horizon l with tail probability p as

$$p = \Pr(L(l) \geq VaR) = 1 - \Pr(L(l) \leq VaR)$$

  where L is a loss function, e.g. in terms of USD
- probability that the position holder would encounter a loss greater than or equal to VaR over the time horizon l is p

Small example: Value-at-risk

```
ggplot(returns, aes(y = return , x = date , geom = 'point')) +
  geom_point(colour = 'lightgrey' , size = 2) +
  geom_line(aes(y = model.fit@fit$sigma*(qdist(distribution = 'std',
                shape = (fitdist(distribution = 'std', x = returns |> pull(return))$pars)[3], p = 0.05))) ,
          colour = 'red') +
  geom_hline(yintercept = sd(returns |> pull(return))*qnorm(0.05),
             colour = 'darkgreen') +
 labs(x = '' , y = 'Daily Returns' , title = 'Value at Risk Comparison')
```



- The two lines indicate the unconditional $p = 0.05$ VaR and the VaR implied by a Garch$(1, 1)$ model

Multivariate volatility estimation

Multivariate volatility estimation

- Covariances are relevant for portfolio theory (and therefore asset pricing, capital budgeting, investment decisions)
- Requires estimation of Σ_t instead of just σ_t^2

Problems for estimates $\hat{\Sigma}_t$

1. The curse of dimensionality (too many parameters)
2. Time variation (volatilities and correlations may change over time)
3. Frequency (minute, daily, monthly data?)

Discuss: How would you compute minimum variance portfolio weights for a universe of 1000 assets?

General Framework

- We consider the N-dimensional return series

$$r_t = E\left(r_t | F_t\right) + \varepsilon_t$$

  where ε_t is a $(N \times 1)$ vector of error terms with (conditional) covariance matrix Σ_t
- Exploiting the symmetry property of Σ_t we can write

$$\Sigma_t = D_t Q_t D_t$$

  where Q_t is the $(N \times N)$ conditional correlation matrix of ε_t and $D_t = \operatorname{diag}\left(\sigma_{11,t}, \dots, \sigma_{NN,t}\right)$ is a $(N \times N)$ diagonal matrix
- Because of the symmetry of Σ_t it is sufficient to only consider $N(N+1)/2$ unique elements

The curse of dimensionality

- Consider the case with a $(T \times N)$ matrix of returns R where $T < N$
- Then the sample variance covariance matrix is

$$S = \frac{1}{T-1} R' \left(I_T - \frac{1}{T} \iota\iota' \right) R$$

- The rank of the matrix $R'R$ is $\min(T, N) = T$
- (intuition: $R'R$ qua linear transformation sends $\mathbb{R}^N$ into a subspace with dimension at most T)
- This implies that $R'R$ is singular for $N \geq T$ and thus

$$\exists w \in \mathbb{R}^N : w'Sw = 0$$

- In other words, S is not invertible and there is no unique efficient portfolio
- More structure is required to overcome this issue (e.g. based on factor structure)

Linear shrinkage (Ledoit Wolf, 2003, 2004)

- The sample covariance matrix is unbiased and the maximum likelihood estimator under normality
- In the context of mean estimation, Stein (1956) showed that a better estimator than the sample mean can be constructed by shrinking the sample mean to a target vector, that is, by using a linear combination of the sample mean and the target vector
- Better in the sense of the means squared error (MSE)
- Ledoit and Wolf (2003, 2004) apply the same thinking to Σ
- They suggest a class of linear combinations of some target matrix F (identity, single factor or equicorrelation matrix) and the sample variance covariance matrix $\hat{\Sigma}$

$$\hat{\Sigma}^{LW} = \alpha F + (1 - \alpha)\hat{\Sigma} \text{ for } 0 \leq \alpha \leq 1$$

The shrinkage target (Ledoit Wolf, 2004)

- We already discussed (multi) factor models
- Constant correlation model: all the (pairwise) correlations are assumed identical.
- Estimator of the common constant correlation: average of all the sample correlations

The shrinkage constant

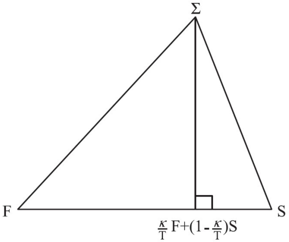


Fig. 1. Geometric interpretation of Theorem 1. The notion of orthogonality among N-dimensional symmetric matrices is defined by the inner product associated with the Frobenius norm.

- The optimal linear combination depends on unknown population quantities and must, therefore, be thought of as an ideal (or “oracle”) but infeasible estimator
- Ledoit and Wolf derive feasible estimators (in an exercise you are asked to implement the estimator)

The optimal shrinkage intensity (Ledoit-Wolf 2003, 2004)

- To define the MSE for a $(N \times N)$ matrice A, we use the Frobenius norm

$$||A||_F^2 = < A, A > := \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} A_{i,j}^2 = \frac{1}{N} tr(A'A)$$

- The scaling $\frac{1}{N}$ is just convenience to ensure that $||I_N||_F = 1$
- The scaled mean squared error is

$$E\left(||\Sigma^{LW} - \Sigma||_F^2\right) \text{ where } \Sigma^{LW} = \alpha F + (1 - \alpha)\hat{\Sigma}$$

- I leave it as an exercise to study the solution to this problem
- Let π the sum of asymptotic variances of the entries of the sample covariance matrix scaled by $\sqrt{T}$, ρ, the scaled asymptotic covariances of the entries of the single-index covariance and γ the bias (misspecification) of the model
- Ledoit and Wolf (2003) show that

$$\alpha^* = \frac{1}{T} \frac{\pi - \rho}{\gamma}$$

Example (AAPL, MMM and BA)

- 0.407 is the average sample correlation $\bar{r} = \left(\frac{2}{(N-1)N} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{\hat{\Sigma}_{ij}}{\sqrt{\hat{\Sigma}_{ii}\hat{\Sigma}_{jj}}} \right)$
- Define the sample constant correlation matrix F by means of the sample variances and the average sample correlation: $f_{ii} = \hat{\Sigma}_{ii}$ and $f_{ij} = \bar{r}\sqrt{\hat{\Sigma}_{ii}\hat{\Sigma}_{jj}}$

	AAPL	MMM	BA	AAPL	MMM	BA	AAPL	MMM	BA
AAPL	3.16	1.08	1.79	1.000	0.409	0.408	3.16	1.07	1.79
MMM	1.08	2.21	1.48	0.409	1.000	0.403	1.07	2.21	1.50
BA	1.79	1.48	6.13	0.408	0.403	1.000	1.79	1.50	6.13

Dynamic multivariate models

- The generalization of univariate GARCH models to the multivariate domain is conceptually simple
- The literature on the dynamics governing Σ_t may be broadly divided into direct multivariate extensions, factor models, linear combination of univariate GARCH models (Generalized Orthogonal GARCH) and nonlinear combination of univariate GARCH models (the broader class of Dynamic Correlation models)
- The need to invert the covariance matrix in many MGARCH parameterizations introduces estimation problems for large systems, as the eigenvalues of the covariance matrix decrease exponentially fast toward zero, even when the covariance is not singular

Constant conditional correlation (CCC)

- Bollerslev (1990)
- In the constant conditional correlation (CCC) model of Bollerslev (1990), the covariance matrix can be decomposed into

$$\Sigma_t = D_t R D_t = \rho_{ij}\sqrt{h_{ijt}h_{ijt}}$$

  where $D_t = \text{diag}(\sqrt{h_{11t}}, \dots, \sqrt{h_{NNt}})$ and R is a positive definite correlation matrix
- Maximum likelihood estimation possible in the multivariate normal case (see Eq. 50 in Boudt et al., 2019)

Obvious extension: DCC

- Covariance Targeting (Engle (2002))

$$\Sigma_t = D_t \mathbf{R_t} D_t = \rho_{ij}\sqrt{h_{ijt}h_{ijt}}$$

- Most popular model of dynamic conditional correlations with proxy process Q_t

$$Q_t = \bar{Q} + \alpha\left(z_t z_t' - \bar{Q}\right) + \beta\left(Q_{t-1} - \bar{Q}\right)$$

- The correlation matrix R_t is then obtained as

$$R_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2}$$

Large-scale portfolio allocation under model uncertainty and transaction costs

Brief Recap: Mean-Variance Portfolio Optimization and ...

- At time t, we solve the problem

$$\max_{\omega_{t+1}} \ \omega'_{t+1}\mu_t - \frac{\gamma}{2}\omega'_{t+1}\,\Sigma_t\,\omega_{t+1} \quad s.t. \quad \omega'_{t+1}\,\iota = 1,$$

  where Σ_t denotes the $N \times N$ asset return covariance, μ_t the mean vector, γ the relative risk aversion, and ι a vector of ones

- Optimal portfolio weights:

$$\omega^*_{t+1} = \frac{1}{\iota'\Sigma_t^{-1}\iota}\Sigma_t^{-1}\iota + \frac{1}{\gamma}\left(\Sigma_t^{-1} - \frac{1}{\iota'\Sigma_t^{-1}\iota}\Sigma_t^{-1}\iota\iota'\Sigma_t^{-1}\right)\mu_t$$

... the classical problem

- Noisy estimates of μ_t over short horizons
- Solution (?): Fixing μ_t (e.g. $\mu_t = 0$) or using long horizons
- $(N \times N)$ matrix Σ_t contains $N(N-1)/2$ distinct elements $\Rightarrow$ Estimation error
- For $N = 500$: 124,750 elements
- Sample covariance only positive definite if $T \geq N$

$\Rightarrow$ Plugging in sample estimates $\hat{\mu}_t$ and $\hat{\Sigma}_t$ performs badly!

Solution: Regularizing Σ ...

- Shrinking the covariance matrix (Ledoit and Wolf, 2004):

$$\Sigma_t = \hat{\alpha} F_t + (1 - \hat{\alpha}) S_t,$$

  where S_t is the sample variance-covariance matrix and F_t denotes the equi-correlation matrix

Alternative approaches:

- Restricting portfolio weights (*Jagannathan and Ma, 2003*; Fan et al, 2012)
- Factor models (MacKinlay Pastor, 2000; Fan et al, 2008)
- Parsimonious high-dimensional parametric volatility models (Engle Ledoit Wolf, 2017)
- Imposing priors (Barry, 1974; Brown, 1979; Jorion, 1985, 1986; Black Litterman, 1992; Pastor Stambaugh, 2000)

... or Using High-Frequency Data

- Idea: Constructing a daily covariance estimate using HF data of the same day
- E.g., Liu (2009), Hautsch et al (2012), Hautsch et al (2015), Lunde et al (2015), Bollerslev et al (2016), ...

But ...

Frictions

- No consideration of transaction costs!
- Re-balancing is costly!

Estimation

- Parameters are uncertain!
- Model ambiguitiy
- Predictive ability of models is time-varying!
- De Miguel et al (2009): Out of sample, the $1/N$ portfolio cannot be outperformed in high dimensions!

Portfolio optimization under transaction costs

Objective function of the investor at time t

- observe past returns: $R_t = \left(r_1', \dots, r_t'\right)' \in \mathbb{R}^{t \times N}$,
- assumptions on underlying (predictive) model M

(Myopic) portfolio optimization problem

$$\omega_{t+1}^* := \arg \max_{\omega \in \mathbb{R}^N, \iota'\omega = 1} E\left(U_\gamma\left(\omega'\left(1 + r_{t+1}\right) - v_t(\omega)\right) | M, F_t\right)$$

$$\arg \max_{\omega \in \mathbb{R}^N, \iota'\omega = 1} \int U_\gamma\left(\omega'\left(1 + r_{t+1}\right) - v_t(\omega)\right) p_t(r_{t+1}|M) \mathrm{d}r_{t+1},$$

where $p_t(r_{t+1}|M) := p_t(r_{t+1}|M, F_t)$ denotes the predictive return distribution and $v_t(\omega)$ the transaction costs

A closed-form solution for Gaussian returns and quadratic transaction costs

- Initial Markowitz (1952) approach with Gaussian returns: $p_t(r_{t+1}|M) = N(\mu, \Sigma)$
- Can be solved by maximizing certainty equivalent

$$\omega^*_{t+1} = \arg \max_{\omega \in \mathbb{R}^N, \iota'\omega=1} \omega'\mu - \nu_t(\omega) - \frac{\gamma}{2}\omega'\Sigma\omega.$$

- Suppose quadratic transaction costs:

$$\nu\left(\omega_{t+1}, \omega_{t^+}, \beta\right) := \nu_t\left(\omega_{t+1}, \beta\right) = \frac{\beta}{2}\left(\omega_{t+1} - \omega_{t^+}\right)'\left(\omega_{t+1} - \omega_{t^+}\right),$$

  with cost parameter $\beta > 0$
- Note: ω_{t^+} differs mechanically from ω_t
- $\omega_{t^+} := \omega_t \circ (1 + r_t)/\iota'(\omega_t \circ (1 + r_t))$.

Optimization problem[2]

- Proof left for exercise

$$\begin{aligned}\omega^*_{t+1} &:= \arg \max_{\omega \in \mathbb{R}^N, \iota'\omega=1} \omega'\mu - \nu_t(\omega, \omega_{t^+}, \beta) - \frac{\gamma}{2}\omega'\Sigma\omega \\ &= \arg \max_{\omega \in \mathbb{R}^N,\, \iota'\omega=1} \omega'\mu^* - \frac{\gamma}{2}\omega'\Sigma^*\omega,\end{aligned}$$

where

$$\mu^* := \mu + \beta\omega_{t^+} \quad \text{and} \quad \Sigma^* := \Sigma + \frac{\beta}{\gamma} I_N$$

- As a result:

$$\omega^*_{t+1} = \frac{1}{\gamma}\left(\Sigma^{*-1} - \frac{1}{\iota'\Sigma^{*-1}\iota}\Sigma^{*-1}\iota\iota'\Sigma^{*-1}\right)\mu^* + \frac{1}{\iota'\Sigma^{*-1}\iota}\Sigma^{*-1}\iota$$

[2] All proofs can be found in Hautsch et al (2020)

What happens if $\beta \to \infty$?

- Define $A(\Sigma) := \left(\Sigma^{-1} - \frac{1}{\iota'\Sigma^{-1}\iota}\Sigma^{-1}\iota\iota'\Sigma^{-1}\right)$.
- Proposition. Let ω_0 be the initial allocation. Then,

$$\omega_T^* = \sum_{i=0}^{T-1}\left(\frac{\beta}{\gamma}A(\Sigma^*)\right)^i \omega(\mu,\Sigma^*) + \left(\frac{\beta}{\gamma}A(\Sigma^*)\right)^T \omega_0,$$

  where $\omega(\mu,\Sigma^*)$ denotes the mean-variance efficient allocation.
- Proposition.

$$\lim_{\beta\to\infty}\omega_{t+1}^* = \left(I_N - \frac{1}{N}\iota\iota'\right)\omega_{t^+} + \frac{1}{N}\iota = \omega_{t^+}$$

Efficieny of the portfolio

- Proposition. $\exists \beta^* > 0 \; \forall \tilde{\beta} \in [0, \beta^*) : \left\| \left(\frac{\tilde{\beta}}{\gamma} A(\Sigma^*) \right) \right\|_F < 1$, where $\| \cdot \|_F$ denotes the Frobenius norm $\|A\|_F := \sqrt{\sum_{i=1}^{N} \sum_{i=1}^{N} a_{i,j}^2}$
- For $\beta < \beta^*$ and $T \to \infty$, we have

$$\sum_{i=0}^{T} \left(\frac{\beta}{\gamma} A(\Sigma^*) \right)^i \to \left(I_N - \frac{\beta}{\gamma} A(\Sigma^*) \right)^{-1}$$

  and $\lim_{i \to \infty} \left(\frac{\beta}{\gamma} A(\Sigma^*) \right)^i = 0$.
- Proposition. For $T \to \infty$ and $\beta < \beta^*$ the series ω_T converges to a unique fix-point given by

$$\omega_\infty = \left(I_N - \frac{\beta}{\gamma} A(\Sigma^*) \right)^{-1} \omega(\mu, \Sigma^*) = \omega(\mu, \Sigma)$$

Intuition behind shrinkage

- For asset-specific transaction costs

$$\left(\omega_{t+1} - \omega_{t^+}\right)' B \left(\omega_{t+1} - \omega_{t^+}\right)$$

  where B is positive definite, the linear shrinkage estimator for Σ of Ledoit and Wolf (2003, 2004) can be reconciled.
- Lemma. Assume a regime of high volatility with $\Sigma^h = (1 + h)\Sigma$, where $h > 0$ and Σ is the asset return covariance matrix during calm periods. Assume $\mu = 0$. Then, the optimal weight ω^*_{t+1} is equivalent to the optimal portfolio based on Σ and (reduced) transaction costs $\frac{\beta}{1+h} < \beta$.

Empirical Illustration

- $N = 308$ assets; 06/2007 - 03/2017 with daily readjustments
- Σ_t estimated by

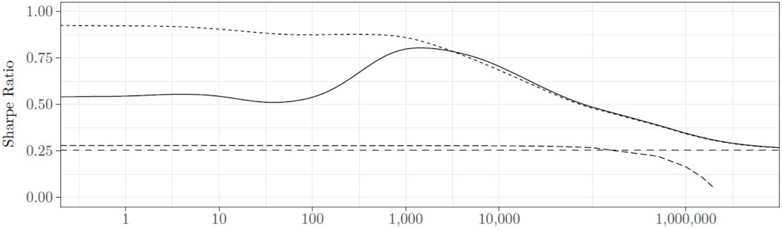
1. Sample covariance
2. Ledoit/Wolf (2004) shrinkage estimator
3. Rolling window of length $h = 500$ days

- μ_t is fixed to zero. Initial portfolios weights: $\frac{1}{N}\iota$. $\gamma = 4$.
- Yields realized portfolio returns net of transaction costs

$$r_t^{\beta} - \beta\left(\omega_{t+1}^{\beta} - \omega_{t^+}^{\beta}\right)'\left(\omega_{t+1}^{\beta} - \omega_{t^+}^{\beta}\right)$$

with $r_t^{\beta} := r_t'\omega_t^{\beta}$.

Realized Sharpe ratios with quadratic transaction costs



- Sample (solid line), LW (dotted line), Equal (dashed)
- X-axis denotes transaction cost parameter β in basis points

L_1 Transaction Costs

- Usual assumption in the literature: L_1 transaction costs

$$v_t\left(\omega, \omega_{t^+}, \beta\right) = \beta||\omega - \omega_{t^+}||_1 = \beta \sum_{i=1}^{N} |\omega_{i,t+1} - \omega_{i,t^+}|.$$

- For $\mu = 0$ and $\Delta := \omega_{t+1} - \omega_{t^+}$, we have

$$\omega^*_{t+1} = \arg \min_{\Delta \in \mathbb{R}^N,\ \iota'\Delta = 0} \gamma \Delta' \Sigma \omega_{t^+} + \frac{\gamma}{2} \Delta' \Sigma \Delta + \beta ||\Delta||_1,$$

- Solving for ω^*_{t+1} yields

$$\omega^*_{t+1} = \left(1 + \frac{\beta}{\gamma} \iota' \Sigma^{-1} \tilde{g}\right) \omega_{\mathrm{mvp}} - \frac{\beta}{\gamma} \Sigma^{-1} \tilde{g},$$

  where $\tilde{g}$ is the vector of sub-derivatives of $||\omega_{t+1} - \omega_{t^+}||_1$ and $\omega_{\mathrm{mvp}} := \frac{1}{\iota'\Sigma^{-1}\iota}\Sigma^{-1}\iota$ are the weights of the GMV portfolio.

Recall Jagannathan and Ma?

- Proposition. For $\mu = 0$, the underlying problem is equivalent to

$$\omega^*_{t+1} = \arg \min_{\omega \in \mathbb{R}^N,\, \iota'\omega = 1} \omega' \Sigma_{\frac{\beta}{\gamma}} \omega,$$

  where $\Sigma_{\frac{\beta}{\gamma}} = \Sigma + \frac{\beta}{\gamma}\left(g^* \iota' + \iota g^{*'}\right)$ and g^* is the subgradient of $||\omega^*_{t+1} - \omega_{t^+}||_1$.

What about parameter uncertainty?

- Recall that the optimal portfolio (for now: $v = 0$) is computed as the solution to

$$\begin{aligned} \omega^*_{t+1} := & \arg\max_{\iota'\omega=1} E\left(U_\gamma\left(\omega'\left(1 + r_{t+1}\right)\right) | M, F_t\right) \\ & \arg\max_{\iota'\omega=1} \int U_\gamma\left(\omega'\left(1 + r_{t+1}\right)\right) p_t(r_{t+1}|M) dr_{t+1} \\ & \arg\max_{\iota'\omega=1} \int \int_\mu \int_\Sigma U_\gamma\left(\omega'\left(1 + r_{t+1}\right)\right) p_t(r_{t+1}|\mu, \Sigma, M) d\mu d\Sigma dr_{t+1}, \end{aligned}$$

  where $p_t(r_{t+1}|M) := p_t(r_{t+1}|M, F_t)$ denotes the predictive return distribution
- Unlike the conditional distribution, the Bayesian predictive distribution accounts for estimation errors by integrating over the unknown parameter space
- The problem above refers to **Bayesian Portfolio Allocation**

1. Account for estimation risk and model uncertainty
2. Elicit economically meaningful prior beliefs about asset pricing models (distribution of future returns)
3. Highly flexible due to Bayesian computation techniques

Bayesian asset allocation[3]

- So far we analysed a two step procedure

1. Estimate the parameters $\hat{\theta}$ (e.g. $\hat{\mu}_t$ and $\hat{\Sigma}_t$)
2. Maximize the expected utility conditional on the estimated parameters being the true ones

$$\max_{\omega} = E\left(U(\omega)|\theta = \hat{\theta}\right)$$

- Optimally, however, the investor adjusts for estimation uncertainty
- Thus, portfolio selection based on $p_t(r_{t+1}|M)$ instead of $p_t(r_{t+1}|\hat{\theta}, M)$

[3]Consult Avramov and Zhou (2010)

Example: Bayesian Portfolio Allocation

- Specify the prior distribution of the parameters as diffuse

$$\pi(\theta) = p(\mu_t|\Sigma_t) \times \pi(\Sigma_t) \propto |\Sigma_t|^{-\frac{N+1}{2}}$$

- Then, assuming that the returns are jointly normally distributed, the posterior distribution is given by

$$p(\mu_t|\Sigma_t, D_t) \propto |\Sigma_t|^{-1/2} \exp\left(-\frac{1}{2} tr\left(T(\mu_t - \hat{\mu}_t)\right)(\mu_t - \hat{\mu}_t)'\Sigma_t^{-1}\right)$$

and

$$\pi(\Sigma_t) \propto |\Sigma_t|^{-(T+N)/2} \exp\left(-\frac{1}{2} tr\Sigma_t^{-1}\left(T\hat{\Sigma}_t\right)\right)$$

- The predictive distribution corresponds to a multivariate t distribution with $T - N$ degrees of freedom
- Under the diffuse prior the mean-variance efficient portfolio is

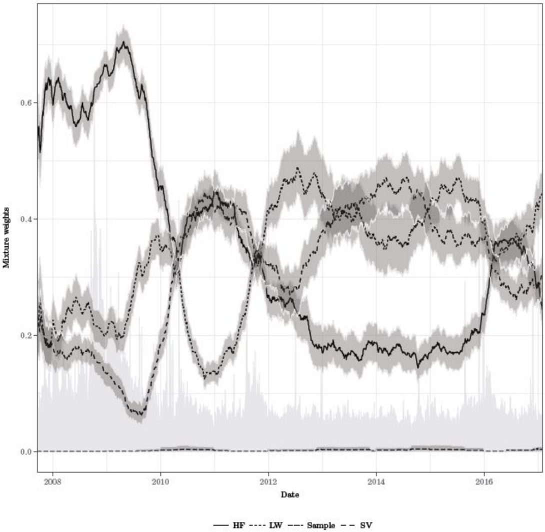
$$\omega^{\text{Bayes}} = \frac{1}{\gamma}\left(\frac{T - N - 2}{T + 1}\right)\hat{\Sigma}_t^{-1}\hat{\mu}_t$$

- Intuition: Invest less in the risky asset and more in the risk-free asset

Extensions and take aways

- Bayesian computation allows to sample from (almost) arbitrary posterior predictive distributions
- Optimal allocation can be derived numerically
- Turnover regularization is closely related to parameter shrinkage
- Optimal portfolios are shifted towards current holdings
- This effect is different from *statistical* approaches: Shrinkage intensity is an exogenous parameter!

Application: Hautsch and Voigt (2019)



Application: Hautsch and Voigt (2019)

	$\hat{\mu}$	$\hat{\sigma}$	SR	CE	TO	pc	sp	% Trade
$\omega^{\text{SV, no TC}}$	-58.9 [-62.7,-55.8]	16.12 [15.72,16.52]	-	-	55.73 [53.52,58.16]	31.102 [30.279,31.942]	2.41 [2.12,2.78]	$>$ 99.95
$\omega^{\text{HF, no TC}}$	$<$ -99	15.97 [15.52,16.33]	-	-	245.4 [239.29,250.43]	56.791 [55.7,57.892]	4.01 [3.83,4.65]	$>$ 99.95
ω^{Sample}	5.4 [3.8,7.2]	16.50 [15.87,17.16]	0.326 [0.224,0.446]	3.993 [2.367,5.908]	0.19 [0.18,0.20]	42.427 [37.75,47.465]	1.21 [0.98,1.38]	18.45 [18.20,19.74]
ω^{LW}	5.6 [4.5,6.6]	14.72 [14.28,15.06]	0.380 [0.307,0.449]	4.501 [3.412,5.497]	0.88 [0.83,0.93]	24.405 [22.784,26.427]	0.98 [0.65,1.11]	11.95 [10.92,12.96]
ω^{HF}	6.5 [5.9,6.9]	16.71 [16.52,16.93]	0.388 [0.354,0.414]	5.081 [4.488,5.518]	0.14 [0.09,0.21]	17.165 [15.937,18.062]	0.48 [0.21,0.61]	14.44 [13.41,15.54]
ω^{SV}	6.6 [5.9,7.0]	16.76 [16.59,16.99]	0.391 [0.354,0.417]	5.153 [4.489,5.589]	0.32 [0.21,0.41]	17.860 [16.086,18.827]	0.71 [0.68,0.92]	14.13 [13.11,15.20]
$\omega^{\text{Comb.}}$	6.5 [5.0,7.8]	14.86 [14.45,15.39]	0.438 [0.342,0.52]	5.401 [3.877,6.693]	0.49 [0.46,0.53]	20.609 [18.483,23.115]	0.81 [0.54,0.92]	13.70 [12.53,14.83]
ω^{Naive}	5.2 [4.9,5.5]	23.5 [23.14,23.84]	0.224 [0.204,0.234]	2.399 [2.035,2.733]	1.10 [1.09,1.12]	0.4 [0.4,0.4]	-	$>$ 99.95
$\omega^{\text{Naive, 2m}}$	5.9 [5.7,6.2]	23.03 [22.68,23.36]	0.258 [0.245,0.273]	3.294 [2.991,3.632]	0.15 [0.14,0.16]	0.403 [0.403,0.404]	-	1.66 [1.63,1.66]
ω^{mvp}	-59.4 [-62.8,-56.1]	14.11 [13.73,14.55]	-	-	53.49 [52.25,54.64]	45.65 [42.864,48.28]	3.33 [3.24, 3.42]	$>$ 99.95
$\omega^{\text{mvp, no s.}}$	2.8 [2.0,3.6]	13.1 [12.97, 13.29]	0.213 [0.149, 0.279]	1.935 [1.092, 2.794]	3.53 [3.34, 3.69]	10.616 [9.161, 11.548]	-	$>$ 99.95
$\omega^{\text{kz, 3f}}$	$<$ -99	$>$ 50	-	-	941.9 [175.51,2339.7]	$>$ 1000	5.98 [3.72, 8.8]	$>$ 99.95
ω^{tz}	$<$ -99	$>$ 50	-	-	$>$ 1000	$>$ 1000	28.09 [17.31, 57.51]	$>$ 99.95
ω^{bs}	$<$ -99	$>$ 50	-	-	$>$ 1000	$>$ 1000	124.7 [39.8, 228.14]	$>$ 99.95
ω^{ϑ}	-24.28 [-26.28, -22.15]	12.04 [11.78, 12.34]	-	-	25.846 [25.43, 26.31]	18.692 [18.05, 19.28]	1.499 [1.40, 1.58]	$>$ 99.95
ω^{Market}	6.2	21.19	0.292	3.958	-	-	-	-

Table 2: Annualized averages of the bootstrapped out-of-sample portfolio performances based on 1904 trading days. Transaction costs are proportional to the L_1 norm of re-balancing (as of (12)). $\hat{\mu}$ is the annualized portfolio return in percent, $\hat{\sigma}$ is the annualized standard deviation in percent, SR denotes the (annualized) out-of-sample Sharpe ratio of the individual strategies, CE is the Certainty Equivalent for an investor with power utility function and risk-aversion factor $\gamma = 4$, TO is the average turnover in percent, pc is the average weight concentration (L_2 norm of the portfolio weights) and sp is the average proportion of the sum of negative portfolio weights. *% Trade* is the fraction of days with trading activity more than 0.001%.

Parametric Portfolio Choice

Stock characteristics and optimal portfolio choice

- Idea: parametrize weights as a function of the characteristics such that we maximize expected utility
- feasible for large portfolio dimensions (such as the entire CRSP universe)
- proposed by Brand et al. (2009) in their influential paper *Parametric Portfolio Policies: Exploiting Characteristics in the Cross Section of Equity Returns*.

Basic Idea

- At each date t we have N_t stocks in the investment universe, where each stock i has return of $r_{i,t+1}$ and is associated with a vector of firm characteristics $x_{i,t}$ such as time-series momentum or the market capitalization
- Investors problem is to choose portfolio weights $w_{i,t}$ to maximize the expected utility of the portfolio return

$$\max_{w} E_t\left(U(r_{p,t+1})\right) = E_t\left[U\left(\sum_{i=1}^{N_t} w_{i,t} r_{i,t+1}\right)\right]$$

Where do the stock characteristics show up?

- Parametrize the optimal portfolio weights as a function of $x_{i,t}$ with the following linear specification for the portfolio weights:

$$w_{i,t} = \bar{w}_{i,t} + \frac{1}{N_t}\theta' \hat{x}_{i,t}$$

- here, $\bar{w}_{i,t}$ is the weight of a benchmark portfolio (value-weighted or naive portfolio), θ is a vector of coefficients which we are going to estimate and $\hat{x}_{i,t}$ are the characteristics of stock i, cross-sectionally standardized to have zero mean and unit standard deviation
- Think of the portfolio strategy as a form of active portfolio management relative to a performance benchmark: Deviations from the benchmark portfolio are derived from the individual stock characteristics
- By construction the weights sum up to one as $\sum_{i=1}^{N_t} \hat{x}_{i,t} = 0$ due to the standardization. Note that the coefficients are constant across assets and through time
- Implicit assumption is that the characteristics fully capture all aspects of the joint distribution of returns that are relevant for forming optimal portfolios

Optimal Choice of θ

$$E_t\left(U(r_{p,t+1})\right) = \frac{1}{T}\sum_{t=0}^{T-1} U\left(\sum_{i=1}^{N_t}\left(\bar{w}_{i,t} + \frac{1}{N_t}\theta' \hat{x}_{i,t}\right) r_{i,t+1}\right)$$

- The allocation strategy is simple because the number of parameters to estimate is small
- Instead of a tedious specification of the N_t dimensional vector of expected returns and the $N_t(N_t + 1)/2$ free elements of the variance covariance, all we need to focus on in our application is the vector θ

What about short-selling constraints?

- To ensure portfolio constraints via the parametrization is not straightforward but in our case we simply renormalize the portfolio weights before returning them by computing

$$w_{i,t}^{+} = \frac{\max(0, w_{i,t})}{\sum_{j=1}^{N_t} \max(0, w_{i,t})}$$