Abstract of "Improving Portfolio Construction through Side Information and Quantified Uncertainty" by Anish R. Shah, Ph.D., Brown University, May 2023.

Motivated largely by the problems of estimation error in investment portfolio optimization, this dissertation develops techniques to improve covariance forecasting and portfolio construction. Covariance forecasting is improved by incorporating contemporaneous side information and by propagating uncertainty of the parameters in factor-modeled covariance to uncertainty of the covariance matrix itself. The latter makes it possible to approach several investment problems explicitly modeling uncertainty. Markowitz mean-variance portfolio optimization becomes robust optimization with an uncertainty set on portfolio utility. Uncertain risk parity considers deviation from the risk budget. Risk under uncertainty and price movement measures the stability of hedges under forecasting uncertainty and shifts in risk exposure caused by dynamic prices.

Abstract of "Improving Portfolio Construction through Side Information and Quantified Uncertainty" by Anish R. Shah, Ph.D., Brown University, May 2023.

Motivated largely by the problems of estimation error in investment portfolio optimization, this dissertation develops techniques to improve covariance forecasting and portfolio construction. Covariance forecasting is improved by incorporating contemporaneous side information and by propagating uncertainty of the parameters in factor-modeled covariance to uncertainty of the covariance matrix itself. The latter makes it possible to approach several investment problems explicitly modeling uncertainty. Markowitz mean-variance portfolio optimization becomes robust optimization with an uncertainty set on portfolio utility. Uncertain risk parity considers deviation from the risk budget. Risk under uncertainty and price movement measures the stability of hedges under forecasting uncertainty and shifts in risk exposure caused by dynamic prices.

# Improving Portfolio Construction through Side Information and Quantified Uncertainty

by
Anish R. Shah
B.Sc., Tufts University, 1991
M.Sc., UC Berkeley, 1994
M.Sc., Brown University, 2003

A dissertation submitted in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy
in the Division of Applied Mathematics at Brown University

Providence, Rhode Island
May 2023

This dissertation by Anish R. Shah is accepted in its present form by
the Division of Applied Mathematics as satisfying the dissertation requirement
for the degree of Doctor of Philosophy.

Date _____ _____

Matthew Harrison, Director

Recommended to the Graduate Council

Date _____ _____

Eli Upfal, Reader
(Computer Science)

Date _____ _____

Petter Kolm, Reader
(NYU Courant)

Approved by the Graduate Council

Date _____ _____

Thomas A. Lewis
Dean of the Graduate School

# VITA

*Anish Shah graduated in 1991 from Tufts University with a B.Sc. double majoring in mathematics and mechanical engineering, in 1994 from the University of California at Berkeley with an M.Sc. in operations research, and in 2003 from Brown University with a M.Sc. in applied mathematics. He has worked in quantitative finance for decades and frequently presents his research in the US and abroad.*

# ACKNOWLEDGEMENTS

*I send my sincere thanks to my dissertation advisor Matt Harrison and readers Petter Kolm and Eli Upfal, to faculty in data science and applied math who made completing the PhD possible after so long away, and to the many applied math professors who made it impossible not to fall in love with such magnificently beautiful material.*

# CONTENTS

**List of Tables** x

**List of Figures** xi

**1 Introduction** 1

**2 Literature review** 3

2.1 Background – stock returns and portfolio selection . . . . . 3

2.1.1 Nature of stock returns . . . . . 3

2.1.2 Portfolio selection . . . . . 4

2.2 Estimation error in portfolio optimization . . . . . 6

2.2.1 Frequentist problems . . . . . 6

2.2.2 Better point-estimates . . . . . 7

2.2.3 Alternative construction methods . . . . . 9

2.2.4 Loss criteria for estimators . . . . . 14

2.3 Research of this dissertation . . . . . 14

**3 Mathematical framework** 16

3.1 Model - observations, latent states, side information . . . . . 16

3.1.1 Gaussian returns . . . . . 18

3.2 Examples . . . . . 19

3.2.1 Factor-modeled covariance . . . . . 19

3.2.2 Multivariate GARCH . . . . . 20

**4 Short-term covariance given instantaneous side information** 23

4.1 Literature review . . . . . 23

4.2 Problem formulation . . . . . 25

4.3 A particular parameterization . . . . . 27

4.4 Practical considerations . . . . . 28

4.5 Abstract . . . 30
4.6 Key Messages . . . 30
4.7 Introduction . . . 30
4.7.1 Background . . . 31
4.7.2 General Structure of Factor Covariance Models . . . 32
4.8 Current Information . . . 32
4.8.1 Portfolio or Individual Security Variance . . . 33
4.8.2 Expected Cross-Sectional Volatility . . . 34
4.8.3 Pairwise Correlation between Securities or Portfolios . . . 34
4.8.4 Average Correlation within a Portfolio . . . 34
4.9 Bayesian Framework for Incorporating Current Information . . . 34
4.9.1 Parameters to Accommodate State . . . 35
4.9.2 Forecasts and Errors . . . 36
4.9.3 Prior . . . 37
4.10 Adapting to Current Market Conditions by Blending Ledoit-Wolf Covariance Reductions . . . 38
4.10.1 Expressed as 1-Factor Models . . . 39
4.10.2 Parametrization . . . 40
4.11 Adapting to Current Market Conditions by Linear and Quadratic Programming . . . 41
4.11.1 Work in a Representation with Uncorrelated Factors . . . 41
4.11.2 Linear Properties . . . 41
4.11.3 Parameterization . . . 42
4.11.4 Observation Noise and Prior . . . 43
4.11.5 MAP Estimate of Parameters . . . 43
4.11.6 Restrictions for Observable Factors . . . 44
4.11.7 Casting Prior onto Orthogonal Factors . . . 44
4.11.8 Reducing Parameters to Prevent Overfitting . . . 45
4.12 Orthogonal Directions that Best Explain Securities . . . 45
4.13 Changing the Time Scale . . . 48
4.14 Expected Cross-Sectional Volatility . . . 50
4.15 Average Correlation and Variance within a Portfolio . . . 52

**5 Uncertain covariance and portfolio optimization under uncertainty 55**
5.1 Literature and mathematical model . . . 56
5.2 Abstract . . . 58
5.3 Introduction . . . 58
5.3.1 Background . . . 59
5.4 Probability model . . . 60
5.5 Uncertain covariance model . . . 62
5.6 Portfolio variance with orthogonal factors . . . 62
5.7 Robust optimization on portfolio utility . . . 67
5.7.1 Indifference region to reduce transaction costs . . . 68
5.8 An example . . . 68
5.9 Summary . . . 72
5.10 Uncertain factor-modeled $\mathrm{E}[\Sigma_{i,j}]$ and $\mathrm{cov}[\Sigma_{i,j}, \Sigma_{m,n}]$ . . . 73
5.11 Incorporating structural uncertainty . . . 76
5.12 Mathematical results . . . 77

**6 Uncertain risk parity 81**
6.1 Literature review . . . 81
6.2 Mathematical model . . . 82
6.3 Abstract . . . 83
6.4 Introduction . . . 83
6.4.1 Literature review . . . 84
6.4.2 Problem statement . . . 85
6.5 Sensitivity to covariance forecast . . . 86
6.5.1 As solution to minimum variance with regularization . . . 86
6.5.2 Non-robustness in risk contribution assignment . . . 87
6.6 Adding uncertainty . . . 87
6.6.1 Risk parity restated with uncertainty . . . 88
6.6.2 Multiple regimes . . . 89
6.7 Uncertain risk decomposition . . . 90
6.8 Factor-modeled covariance uncertainty . . . 91
6.9 An example . . . 92
6.10 Summary . . . 93

6.11 Uncertain covariance between portfolios from securities . . . . . 95
6.11.1 Model-free . . . . . 96
6.11.2 Uncertain covariance modeled . . . . . 96
6.12 Parametric uncertainty covariance of unbiased covariance estimates . 98

**7 Risk under uncertainty and price movement 102**
7.1 Mathematical model . . . . . 103
7.2 Abstract . . . . . 104
7.3 Introduction . . . . . 104
7.3.1 Problem with conventional method . . . . . 105
7.3.2 Contribution and central idea of this paper . . . . . 106
7.4 Conventional risk contributions . . . . . 106
7.4.1 Variance contributions of common partitions . . . . . 107
7.5 Uncertain variance contributions . . . . . 109
7.5.1 Uncertain factor-modeled covariance . . . . . 110
7.5.2 Mean and covariance of contributions . . . . . 111
7.5.3 Computational complexity . . . . . 119
7.6 Uncertain standard deviation contributions . . . . . 119
7.7 Modeling weight dynamics . . . . . 122
7.8 An example . . . . . 124
7.8.1 Portfolio 1: Diversified long, cash benchmark . . . . . 125
7.8.2 Portfolio 2: 44 stock long-short, cash benchmark . . . . . 127
7.9 Summary . . . . . 128
7.10 Mathematical results . . . . . 130
7.11 Mean and variance of factor covariance from uncertain covariance model 132

**8 Conclusion 134**

# LIST OF TABLES


2.1 Robust optimization sets . . . . . 9

5.1 Realized portfolio std dev quantiles by method (annualized) . . . . . 70
5.2 Head-to-head comparison of realized volatility . . . . . 71
5.3 Median expected variance, uncertainty, and realized variance (weekly, not annualized) . . . . . 71
5.4 Expected minus realized annualized standard deviation . . . . . 72
5.5 Z-scores: $\frac{\text{predicted - realized variance}}{\text{uncertainty (std dev) of variance}}$ . . . . . 72

6.1 Experiment 1: Std Dev & Correlation . . . . . 93
6.2 Experiment 2: Std Dev & Correlation . . . . . 94
6.3 Experiment 1: Weight and Variance Contributions . . . . . 94
6.4 Experiment 2: Weight and Variance Contributions . . . . . 95

7.1 Computational complexity of reporting contributions' mean . . . . . 119
7.2 Computational complexity of reporting contributions' variance . . . . 119
7.3 Diversified long—mean and SD of factor exposures contributed by sector 126
7.4 Diversified long—conventional and uncertain risk contributions by factor 127
7.5 Diversified long—ex-ante modeled and realized risk contributions by sector . . . . . 127
7.6 44 stock long-short—mean and SD of factor exposures contributed by side . . . . . 128
7.7 44 stock long-short—mean and SD of factor exposures contributed by sector . . . . . 129
7.8 44 stock long-short—conventional and uncertain risk contributions by factor . . . . . 130
7.9 44 stock long-short—ex-ante modeled and realized risk contributions by sector . . . . . 130

# LIST OF FIGURES

3.1 Model. . . . . 17

# CHAPTER 1

# INTRODUCTION

Quantitative finance is a broad label for the application of mathematical, algorithmic, statistical, or inferential techniques to financial markets. The expertise involved can vary from stochastic processes, PDEs, and optimal control to optimization, econometrics, statistics, machine learning, and computer science.

Attilio Meucci, for example in Meucci [2011], makes a general split between what he calls 'P' quants and 'Q' quants. Repeating his ideas: 'P' quants are concerned with investing money, operate in discrete time and high dimension (e.g., many stocks to invest across), try to predict the future, and use tools such as statistics and optimization. 'Q' quants are concerned with trading, operate in continous time and low dimension, extrapolate the present to price securities that lack market information, and use tools such as stochastic calculus.

This dissertation addresses several 'P' quant problems I encountered over many years in the field. The impetus for much of the work is estimation error's confounding portfolio optimization, the source of enough colossal investment blow-ups, one involving two Nobel laureates and catastrophic loss across financial markets averted only by government intervention[1]. Nassim Nicholas Taleb mentions the problem in his 2012 book Anti-Fragile[2]. First thinking on it in the mid 2000s, after many failed attempts, I began (to my eye) to see clearly perhaps a decade later. The insight also led to solving a couple of related problems. Progressing on the different pieces over the years, I've presented to combined professional/academic audiences in Boston, New York, San Francisco, London, Amsterdam, and globally via video conference.

The four topics in the dissertation are 1–forecasting short-term covariance given instantaneous side information, 2–uncertain covariance and portfolio optimization under uncertainty, 3–uncertain risk parity portfolio construction, and 4–risk under

[1] Long-Term Capital Management. See Jorion [2000].

[2] Taleb [2012]. Taleb retweeted one of my presentations in Aug 2020.

uncertainty and price movement. While all are related, only the first didn't arise from the aforementioned estimation error problem.

The chapters are arranged as follows: Chapter 2 is a literature review. Chapter 3 describes a common mathematical framework. Each of chapters 4 – 7 covers one of the four topics. Chapter 8 is a conclusion.

# CHAPTER 2

# LITERATURE REVIEW

## 2.1 Background – stock returns and portfolio selection

### 2.1.1 Nature of stock returns

Let $S_t$ be the price of a stock at time t. The stock's return $R_t$ is its relative change in value, measured as a fraction or as the change in log of value. Ignoring dividends, $R_t = \left(\frac{d}{dt}S_t\right)/S_t$ or log return $\frac{d}{dt}\log S_t$ in continuous time and $S_t/S_{t-1} - 1$ or (less common) log return $\log S_t - \log S_{t-1}$ in discrete time. The two canonical models are geometric Brownian motion[1]

$$dS_t = \mu_t S_t dt + S_t \sigma_t dW_t \tag{2.1}$$

$$\text{i.e.,} \quad d\log S_t = \left(\mu_t - \sigma_t^2/2\right) dt + \sigma_t dW_t \tag{2.2}$$

and, to allow for mean reversion, Ornstein-Uhlenbeck[2]

$$dS_t = -\theta\left(S_t - \bar{S}\right) dt + \sigma_t dW_t \tag{2.3}$$

Discrete analogs model returns as

$$\text{independent} \quad R_t = \mu_t + \varepsilon_t \tag{2.4}$$

$$\text{or serially correlated via an AR(1)} \quad R_t = \rho R_{t-1} + \varepsilon_t \tag{2.5}$$

where $\varepsilon_t \sim N(0, \sigma_t^2)$ and ignoring that returns usually can't occur below -100%. With multiple stocks, $S, W, \varepsilon$ and discrete $\mu$ are vectors and $\sigma, \theta, \rho$ and continuous $\mu$ matrices.

[1] See section 5.8 of Karatzas and Shreve [1991].

[2] See Barndorff-Nielsen and Shephard [2001] or Meucci [2009].

While theoretically and practically important variants exist[3]—for example, allowing for time dependency and jumps in the randomness—the simpler versions often yet appear out of approximation, ease of manipulation, or the limits of forecasting accuracy.

Unlike a physical system, in addition to not adhering to rules, returns have a mechanism which makes them unpredictable: the market efficiency of investor action. Buying raises a stock's price and selling lowers it, a process that occurs until the opportunity for profiting vanishes. Predictable edges are *arbitraged away*. In particular, this means *the past does not imply the future*. However, covariance persists through structural links; e.g., oil price increases affect oil companies positively and airlines negatively, linking them regardless of whether oil prices can be predicted.

An aspect of returns that can be predicted is *risk premium*, compensation for bearing risk, akin to writing insurance and receiving premiums. Apparently successful strategies are often in truth compensated for bearing risks. When the rare event occurs, profits are disgorged. Finding an apparently profitable strategy, a skillful investor wonders, "Is there a risk I am getting paid for?"

### 2.1.2 Portfolio selection

Portfolio selection determines the amount of capital to invest in each asset. The canonical methods are Markowitz [1952] ('mean-variance' or 'Markowitz') and Kelly [1956] ('Kelly betting'). Markowitz [1952] introduces the notion of efficient portfolios—maximum expected return at their variance and minimum variance at their expected return—and finds them by maximizing expected return minus a scalar times variance.

$$w^*(\mu, \Sigma, \lambda) \equiv {}^{4} \arg\max_{w} \; w^{\mathrm{T}}\mu - \frac{\lambda}{2} w^{\mathrm{T}}\Sigma w \tag{2.6}$$

$$\begin{aligned} \text{where } \mu &\equiv \text{ expected returns} \in \mathbb{R}^n \\ \Sigma &\equiv \text{ covariance of returns} \in \mathbb{R}^{n\times n} \\ \lambda &\equiv \text{ risk aversion} > 0 \\ w &\equiv \text{ security weights} \in \mathbb{R}^n \end{aligned}$$

[3] Beginning with Mandelbrot [1963], Mandelbrot looked at price series' fractal properties. A recent evolution of that line is Dandapani et al. [2021]. Bouchaud [2021] is a clear, short overview of the history and areas of current research.

Kelly [1956] allocates a fraction of the capital present each period to maximize the exponential growth rate or, to same effect, expected log of wealth.

$$\text{exp growth rate} \equiv \lim_{T\to\infty} \frac{1}{T} \log \frac{S_T}{S_0} \tag{2.7}$$

$$\begin{aligned} S_T &= S_0 \prod_{t=1}^{T} \frac{S_t}{S_{t-1}} \\ &= S_0 \prod_{t=1}^{T} \left( \sum_{i=1}^{n} w_i \left[1 + R_{i,t}\right] \right) \end{aligned}$$

$$\mathrm{E}\left[\log \frac{S_T}{S_0}\right] = \sum_{t=1}^{T} \mathrm{E}\left[\log\left(\sum_{i=1}^{n} w_i \left[1 + R_{i,t}\right]\right)\right] \tag{2.8}$$

$$\begin{aligned} \text{where } S_t &\equiv \text{ portfolio value at } t \\ R_{\cdot,t} &\equiv \text{ security returns from } t-1 \text{ to } t \end{aligned}$$

Breiman [1961] and Thorp [1971] prove results about its optimality. Page 122 of Markowitz [1959] shows the two term Taylor expansion of expected exponential growth is approximately mean return - 1/2 variance of return, i.e., Markowitz with a low risk aversion of $\lambda = 1$. Chamberlain [1983] shows that for concave utility and elliptically distributed investment returns, expected utility of a portfolio's (scalar, random) return is a function of mean and variance. Markowitz [2014] is a survey of approximating concave utility with Markowitz.

In practice, Kelly betting is regarded as having too large swings, especially for investing on behalf of a client. Since Markowitz is ubiquitous, this research takes it as a given.

Several areas border the scope of this dissertation. There is literature on utility functions—what utility should be, what properties are advantageous and sensible, how real human beings behave. There is (very technical) literature on optimal control in portfolio selection. Near its head, Merton [1971] proves results on maximizing expected utility when utility is concave and prices are modeled as continuous time Markov stochastic processes. There is literature on alterations to Markowitz. For example, risk measures[5] in place of variance include mean absolute deviation and semi-variance, $\mathrm{E}\left[\min\left(x - target, 0\right)^2\right]$. The motivation for the latter is that, at odds with

[4] Notation: $\equiv$ means a definition.

reason, adding variable upside can decrease a portfolio's utility when risk is measured by variance. To deal with non-Gaussian returns, one approach[6] is to minimize (CVaR) conditional value-at-risk, $\mathrm{E}[x|x \leq F_x^{-1}(c)]$, subject to some minimum $\mathrm{E}[x]$. Though Markowitz considers only the present, the problem becomes multi-period in the presence of transaction costs and market impact.[7]

## 2.2 Estimation error in portfolio optimization

Ideas on estimation error's effect on optimizing Markowitz utility follow three themes: frequentist problems, better point-estimates (Bayesian, robust statistics), and alternative construction methods (Bayesian posterior distribution of utility, robust optimization).

### 2.2.1 Frequentist problems

Though no longer so, it was once surprising that optimizing weights across stocks, which are numerous and similar, based on unbiased estimates of a few parameters (mean and covariance) taken as true—with real differences dominated by noise—yields a portfolio that performs worse than expected and a decision rule that is easily improved.

Jobson and Korkie [1980] derives an approximate distribution of the mean and variance of a portfolio optimized from frequentist estimates and finds that they work poorly with small histories and would be better replaced by Bayesian estimates. Michaud [1989] brought the problem to a broad audience. Broadie [1993] simulates returns from known parameters, infers unbiased mean and covariance, and optimizes the efficient frontier on both to show that optimizing estimates goes badly. A few lines proof of the mechanism is in Shah [2009]. Fan et al. [2008] analyzes properties of factor-modeled[8]vs sample covariance and finds factor-modeling is advantageous when inverse of covariance is of interest as in Markowitz portfolio optimization.

[5]See chapter 9 of Markowitz [1959], Konno and Yamazaki [1991], Chang et al. [2009].

[6]See Rockafellar and Uryasev [2000], Rockafellar and Uryasev [2002].

[7]See Gârleanu and Pedersen [2013].

[8]The standard ways of factor-modeling covariance—time series, cross-sectional, PCA—are described in chapter 9 of Tsay [2010].

Although there are more papers, there is little more to say once it's clear why frequentist estimates don't work well with optimization.

### 2.2.2 Better point-estimates

Fabozzi et al. [2007] gives an overview of the issues, and Fabozzi et al. [2010] is a survey largely on approaches to counter estimation error issues in mean-variance portfolio construction. From the latter:

> There are two standard methods extensively adopted in the literature to deal with estimation errors. The first is the robust estimation methods, such as moment estimation, which can be quite robust to distributional assumptions. The second is the Bayesian approach that is neutral to uncertainty in the sense of Knight (1921) because it assumes a single prior on the portfolio distribution (Garlappi et al. 2007).

#### Bayesian point-estimates

First, two techniques that are widely used in practice: Black and Litterman [1992] converts forecasts of $\mu$ from unbiased to Bayesian by assuming Gaussian observation noise and prior. The computation is the same as a Kalman filter with one noisy observation and no dynamics. Ledoit and Wolf [2003] and other papers by the authors shrink covariance toward a structured form—usually one factor (effectively the first PCA) + a diagonal matrix—via convex combination and derive results for the optimal shrinkage.

Notice that, in real life, $\mu$ and $\Sigma$ are forecasted separately. Again, this is because $\Sigma$ is largely from history and $\mu$ from diverse outside information. The many Bayesian papers that forecast both from history ignore the predictability-erasing mechanism of the market.

Other covariance shrinkage estimators in the style of Ledoit and Wolf are Daniels and Kass [2001][9] and Schäfer and Strimmer [2005]. Continuing in that direction, a set of shrinkage point-estimators of covariance comes from random matrix theory. The general idea: eigenvalues of the sample correlation matrix are biased—upward for

[9] Cites a 1996 working paper by Ledoit.

large eigenvalues and downward for small—so are shrunk toward a point or distribution (Marčenko–Pastur) or discarded below a critical value with their effects diagonalized. Plerou et al. [2002] and Potters et al. [2005] are early papers, Johnstone [2007] a clear introduction, Bun et al. [2017] a survey, and Potters and Bouchaud [2020] a textbook. Ledoit and Wolf [2012], which shrinks eigenvalues toward a distribution, appears in several multivariate ARCH papers mentioned in chapter 4 below.

### Robust statistics

Scutellà and Recchia [2013], a survey on robust optimization, also covers robust statistics[10]and cites the methods of the following two paragraphs.

Robust point-estimates of mean[11]and covariance are used to optimize portfolios in Cavadini et al. [2001][12], Perret-Gentil and Victoria-Feser [2004][13], de Melo Mendes and Leal [2005][14], and Welsch and Zhou [2007][15].

Rather than using robust estimates of parameters, several papers optimize utility defined as a robust function of weighted security returns. Note that these methods are impractical since future expected returns must be forecasted and not inferred from history. Lauprete et al. [2003] minimizes a robust estimate of portfolio variance, the Huber function. DeMiguel and Nogales [2009] explores two ways of optimizing a robust estimate of portfolio variance involving both location and scale: 1) M-estimators via the Huber function, 2) S-estimators via Tukey's biweight function.

[10]See Huber [1981].

[11]Suffers from the usual problem—predicting mean requires outside information.

[12]Robust estimator from Krishnakumar and Ronchetti [1997]

[13]Robust estimator from Rocke [1996]

[14]Robust estimator from Leung and Ng [2004] and a variation on minimum covariance determinant of Rousseeuw [1985]

[15]1-FAST-MCD from Rousseeuw and Driessen [1999], 2-Iterated Bivariate Winsorization from Chilson et al. [2006] with modification of Maronna and Zamar [2002], 3-Fast 2-D Winsorization from Khan et al. [2007] with modification of Maronna and Zamar [2002].

### 2.2.3 Alternative construction methods

#### Robust optimization

Robust optimization finds the best worst case when parameters aren't fixed but fall within a set. The optimization changes from $\arg\max_w \ f_\theta(w)$ to $\max_w \ \min_{\theta \in S} \ f_\theta(w)$. Though the idea is straightforward, the max-min optimization limits the technique to special configurations, thus the literature is largely an exploration of particular cases described in Table 2.1.

Table 2.1: Robust optimization sets

| Reference | Set |
|---|---|
| Rustem et al. [2000] | $(\mu, \Sigma)$ comes from a number of scenarios, $\cup_{i=1\ldots n} (\mu_i, \Sigma_i)$ |
| Halldórsson and Tütüncü [2003] | $\mu^L \leq \mu \leq \mu^U, \ \Sigma^L \leq \Sigma \leq \Sigma^U$ |
| Goldfarb and Iyengar [2003] | $\mu^L \leq \mu \leq \mu^U$<br>$\Sigma = EFE^{\mathrm{T}} + \mathrm{diag}(\varepsilon^2)$ where<br>$\varepsilon_L^2 \leq \varepsilon^2 \leq \varepsilon_U^2$<br>$\{E : E = \bar{E} + \eta, \ \|\eta_{i,\cdot}\|_G \leq \delta_i\}$ norm is wrt some positive definite matrix $G$<br>$\{F : F^{-1} = \bar{F}^{-1} + \eta, \ \text{max eigenval of } \bar{F}^{1/2}\eta\bar{F}^{1/2} \leq \delta_F\}$<br>or $\{F : F = \bar{F} + \eta, \ \|N^{-1/2}\eta N^{-1/2}\| \leq \delta_F\}$ where norm is max eigenval or sum of squared eigenvals |
| Ceria and Stubbs [2006] | $\{\mu : (\mu - \bar{\mu})^{\mathrm{T}} \, \Omega \, (\mu - \bar{\mu}) \leq \delta\}$ |
| Garlappi et al. [2007] | $\mu$ is in the intersection of ellipses on subsets of securities, $\cap_{i=1\ldots n}\{\mu : \left(\mu^{S_i} - \bar{\mu}_i\right)^{\mathrm{T}} \Omega^{S_i} \left(\mu^{S_i} - \bar{\mu}_i\right) \leq \delta_i\}$.<br>$x^{S_i}$ means the $S_i$ coordinates of $x$.<br>Expected returns that come from factors can be put in this form. |
| Lu [2011] | defined jointly on $\mu$ and factor exposures |

Anderson and Cheng [2016] in section 2.2.3 below solves robust optimization with uncertainty in means. Lu [2011] makes the interesting point that separable uncertainty sets, i.e., ones defined as products, are too conservative and operate at a much higher confidence level than reported.

Note that while confidence sets defined separately on each parameter ignore the diversification of error and see every extreme outcome realized at once, jointly defined confidence sets become informationless as the universe grows. For example, imagine Gaussian $\mu$ frequentist estimated as a vector of 0's with identity noise covariance. A confidence set is a sphere of radius $h$. Since a single coordinate outside $h$ breaks the sphere, the critical $h$ for a given confidence level increases with dimension. These problems vanish in a Bayesian setting.

Previously mentioned Scutellà and Recchia [2013] is a survey paper on robust optimization of portfolios, and Bertsimas et al. [2011] on robust optimization in general. Hansen and Sargent [2011] explores robust decision making/control in depth.

### Bayesian posterior

Most of the ideas were introduced in early, universally cited work. Barry [1974] assumes $(\mu, \Sigma)$ is Gaussian-Inverse Wishart whose parameters are the sample estimates of $\mu$ and $\Sigma$. This corresponds[16]to Bayesian inference with the prior $\pi(\mu, \Sigma) \propto |\Sigma|^{-\frac{n+1}{2}}$ and Gaussian returns. The mean of the (multivariate t) predictive distribution of returns is the sample mean, and the variance a scalar multiple of the sample variance. Thus, mean-variance utility is the same as frequentist-inferred under higher risk aversion. Klein and Bawa [1976] considers three priors, two of which lead to different choices than can be had from frequentist estimates. Frost and Savarino [1986] sets the prior's parameters empirically; the prior is Gaussian-Inverse Wishart with all securities having the same mean return, variance, and pairwise correlation.

[16]Derived in Klein and Bawa [1976]

Jorion [1986] starts by evaluating the statistical risk of an estimator in terms of utility under optimized weights,

$$\begin{aligned} U(w,\theta) &\equiv \text{ (random) utility of portfolio } w\text{'s returns under parameters } \theta \\ w^*(\theta) &\equiv \arg\max_w \mathrm{E}\left[U(w,\theta)\right] \\ U^*(\theta) &\equiv U(w^*\left[\theta\right],\theta) \\ r(\theta) &\equiv \text{risk of estimator} \\ &\equiv \mathrm{E}\left[\frac{U^*(\theta) - U\left(w^*\left[\hat{\theta}\right],\theta\right)}{|U^*(\theta)|}\right] \end{aligned} \tag{2.9}$$

and finds frequentist estimates of $\mu$ inadmissible. The paper then explores Stein shrinkage estimators[17], of which a particular case can be framed as an informative prior on $\mu$ (Gaussian centered at a common mean). With variance assumed proportional to the sample variance, the prior's common mean and variance concentration scalar are determined empirically.

Avramov and Zhou [2010] is a survey paper.

Recent work leans toward specifics. Anderson and Cheng [2016] models next period returns as a mixture of normals having inverse-Wishart variance. The models correspond to the length of the current regime: at period $t$, there are $t$ normals, with the $i$'th regarding returns from periods $i \ldots t$ as coming from the same distribution. Since mean returns come from history, it is not practical. In addition to maximizing posterior utility, the paper solves a robust version of the problem ("risk-sensitive optimization" in control theory) that considers uncertainty in mean returns. Kacperczyk and Damien [2011] is unusual in that it considers 'distribution uncertainty', the possibility of returns' following distributions other than Gaussian. Lai et al. [2011], noted here only because the authors are from statistics and applied math departments, looks at the mean and variance of the predictive distribution of returns. Bauder et al. [2021] is likewise a straight Bayesian treatment using the mean and variance of the predictive distribution assuming the series of returns is exchangeable, the future is like the past, and the number of observations exceeds the number of securities. The latter two assumptions make it practically unusable, but published in the last two

[17] See Stein [1956]

years in a credible journal, the paper offers a current snapshot of references.

Two techniques that don't forecast returns from history: Davis and Lleo [2013] extends Black and Litterman [1992] and applies the continous time Kalman filter to incorporate information arriving across time.[18] In Meucci [2008] and other papers, Meucci lays out a framework, 'entropy pooling', for updating general distributions with forecast information. The starting point is a prior on the world of interest, $X \sim f_X$. Views are expressed as distributions on functionals, $g(X) \sim f_V$. A distribution $f^*$ is sought[19]that is coherent with the views and has minimum K-L divergence from the prior. This is convexly combined with the prior to express the forecaster's confidence, $f_{forecast} = cf^* + (1-c)f_X$.

### Resampling

Michaud [1998], which is critiqued by Scherer [2002], introduces 'resampling'—repeatedly generating a fixed length sample of a normal distribution under the point-estimated parameters and optimizing using parameters inferred from the sample, finally averaging the output. Particulars aside, it is akin to a softened point-estimate of covariance: Because the maximizer of $w^{\mathrm{T}}\mu - \frac{\lambda}{2}w^{\mathrm{T}}\Sigma w$ is $w^* = \frac{1}{\lambda}\Sigma^{-1}\mu$ and samples $(\hat{\Sigma}, \mu + \hat{\delta})$ and $(\hat{\Sigma}, \mu - \hat{\delta})$ are equally likely, $\mathrm{E}\left[\hat{w}^*\right] = \frac{1}{\lambda}\mathrm{E}\left[\hat{\Sigma}^{-1}\right]\mu$.[20]

### Risk parity

Named by Qian [2005], risk parity is a popular portfolio construction technique that determines weights by making each investment's risk contribution (component in the direction of the net portfolio risk[21]) match the given, typically equally-weighted risk

[18] See Kolm and Ritter [2017], Kolm and Ritter [2020], and Kolm et al. [2021] for other extensions.

[19] Cannot be solved analytically but numerically by reweighting the probabilities on a fixed set of samples.

[20] With linear equality constraints, sampling $\mu$ washes out for the same reason: $\arg\max_w w^{\mathrm{T}}\mu - \frac{\lambda}{2}w^{\mathrm{T}}\Sigma w$ s.t. $Aw = b$ is
$w^* = \frac{1}{\lambda}\left[\Sigma^{-1} - \Sigma^{-1}A^{\mathrm{T}}\left(A\Sigma^{-1}A^{\mathrm{T}}\right)^{-1}A\Sigma^{-1}\right]\mu + \Sigma^{-1}A^{\mathrm{T}}\left(A\Sigma^{-1}A^{\mathrm{T}}\right)^{-1}b,$
so $\mathrm{E}\left[\hat{w}^*\right] = \frac{1}{\lambda}\mathrm{E}\left[\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}A^{\mathrm{T}}\left(A\hat{\Sigma}^{-1}A^{\mathrm{T}}\right)^{-1}A\hat{\Sigma}^{-1}\right]\mu + \mathrm{E}\left[\hat{\Sigma}^{-1}A^{\mathrm{T}}\left(A\hat{\Sigma}^{-1}A^{\mathrm{T}}\right)^{-1}\right]b.$

[21] See Menchero and Davis [2011].

budget.[22]

$$
\begin{aligned}
w^*\left(\Sigma, \sigma^2, f\right) &\equiv \text{any } w > 0 \text{ s.t. } w \odot \Sigma w = \sigma^2 f \\
\text{where } \sigma^2 &\equiv \text{ target variance of portfolio} \\
f &\equiv \text{ fraction of risk budgeted to each security, } f \in \mathbb{R}^n_+ \\
\odot &\equiv \text{ element-by-element multiplication}
\end{aligned} \tag{2.10}
$$

It is typically done at the index or asset class level rather than on individual stocks, e.g., one security is the S&P500 index, another US corporate bonds.

Maillard et al. [2010] shows risk parity is equivalent[23]to minimizing variance with a regularization penalty.

$$
w^*\left(\Sigma, \sigma^2, f\right) = \underset{w>0}{\arg\min} \ \ \frac{1}{2} w^{\mathrm{T}} \Sigma w - \sigma^2 \sum_i f_i \log w_i \tag{2.11}
$$

In addition to much simpler computation, strict convexity makes obvious the existence of a unique solution in each orthant (using $\log|\cdot|$) when $w$ isn't sign constrained.

Maillard et al. [2010] and Bai et al. [2016] explore mathematical properties and alternative formulations. Instead of equalizing risk contributions across investments, Lohre et al. [2014] equalizes them across principal component directions of the covariance matrix, Lassance et al. [2022] across independent component directions, and Roncalli and Weisang [2016] across directions specified by any factor model. Bernardi et al. [2019] looks at risk parity's 'second order risk', the ratio of realized to predicted risk, essentially the bias from estimate-optimized selection.

### Hierarchical risk parity

De Prado [2016] introduces 'hierarchical risk parity' as a way to avoid covariance's low eigenvalue directions (seemingly low risk portfolios) that appear as artifacts under a large universe. The algorithm 1–constructs a notion of distance between securities based on differences in their correlations to all securities, 2–extracts a tree structure through hierarchical clustering, 3–reorders securities according to proximity (a depth

[22]Controlling the scale via full investment $\left(1^{\mathrm{T}} w = 1\right)$ instead of target variance $\sigma^2$ is achieved easily by arbitrarily setting $\sigma^2$ and normalizing the solution, $w^*/1^{\mathrm{T}} w^*$.

[23]Setting the gradient to 0 yields the risk parity equation (2.10).

first search of the tree), and finally, 4–starting from a single group containing all securities equally weighted, repeatedly bisects each group into a finer scale (groups containing 1/2 as many securities) and scales the weights on each half inversely proportional to its relative variance. Cotton [2022b] generalizes the idea.

### 2.2.4 Loss criteria for estimators

In equation (2.9) above, Jorion [1986] evaluates an estimator by the utility of a portfolio built from its estimates. Facing Markowitz utility as in equation (2.6), why do few papers estimate $\{\mu, \Sigma\}$ and set $w^*$ to maximize the performance of $U\left(w^*\left[\hat{\mu}, \hat{\Sigma}\right], \mu, \Sigma\right)$, which is standard in statistics?[24]

As previously mentioned, market efficiency dictates that forecasts for $\mu$ and $\Sigma$ come from different information sets, with $\mu$'s continually changing and expanding. Nevertheless, utility considerations appear implicitly. For example, factor-modeling and the many covariance shrinkage estimators in section 2.2.2 aim to erase low eigenvalue directions of covariance *because* the maximizer of $w^{\mathrm{T}}\mu - \frac{\lambda}{2}w^{\mathrm{T}}\Sigma w$ is $w^* = \frac{1}{\lambda}\Sigma^{-1}\mu$. This wouldn't be important if $w$ were independent of the estimates. The widely popular Black and Litterman [1992] sets the prior on $\mu$ to stabilize the results of optimization. And most papers include empirical results of optimized portfolios.

Papers that explicitly tie the estimation to utility, Jorion [1986] included, typically infer both $\mu$ and $\Sigma$ from history in a traditional framing that doesn't work with markets.[25]However interesting, they can't directly be used in practice.

## 2.3 Research of this dissertation

From this background arise the four problems involving covariance, portfolio construction, and uncertainty that are addressed by this dissertation:

1. Short-term covariance given instantaneous side information
2. Uncertain covariance and portfolio optimization under uncertainty

[24]In this vein, presenting a deep learning application to trading which didn't involve different information sets, (Pelger [2023]) often repeated the importance of making the loss function match one's objectives.

[25]Cotton [2022a] presents many arguments for market information over forecasting from history.

3. Uncertain risk parity
4. Risk under uncertainty and price movement

Additional references specific to each topic are in the corresponding chapters.

# CHAPTER 3

# MATHEMATICAL FRAMEWORK

An abstract mathematical framework serves as starting point for conceptualization and finding approaches to solutions.

## 3.1 Model - observations, latent states, side information

$$Y_t \equiv \text{random stock returns } (Y_t \in \mathbb{R}^n,\ t \in \mathbb{N})$$
$$Z_t \equiv \text{unobserved latent states } (Z_t \in \mathcal{Z},\ t \in \mathbb{N})$$
$$X_t \equiv \text{random side information } (X_t \in \mathcal{X},\ t \in \mathbb{N})$$

Of interest is the distribution given information at time $T$ of future stock returns, $P(Y_T|X_{1:T}, Y_{1:T-1})$ or $P(Y_T|X_{1:T-1}, Y_{1:T-1})$, or latent states, $P(Z_T|X_{1:T}, Y_{1:T-1})$ or $P(Z_T|X_{1:T-1}, Y_{1:T-1})$.

Assume $Y$ and $X$ are independent given $Z$,

$$P(Y_{1:T}, X_{1:T}|Z_{1:T}) = P(Y_{1:T}|Z_{1:T})\, P(X_{1:T}|Z_{1:T}),$$

and conditionally depend on $Z$ only at the same period,

$$P(Y_{1:T}|Z_{1:T}) = \prod_{t=1}^{T} P(Y_t|Z_t)$$
$$P(X_{1:T}|Z_{1:T}) = \prod_{t=1}^{T} P(X_t|Z_t).$$

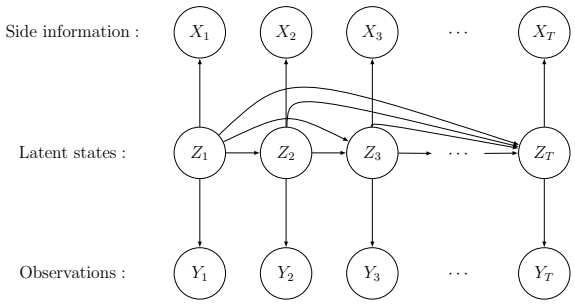


Figure 3.1: Model.

$P(Z_{1:T})$ describes the dynamics of $Z$.

$$\begin{aligned}
P(Y_T|X_{1:T}, Y_{1:T-1}) &= \int P(Y_T, Z_{1:T}|X_{1:T}, Y_{1:T-1})\, dZ_1 \dots dZ_T \\
&= \int P(Y_T|X_{1:T}, Y_{1:T-1}, Z_{1:T})\, P(Z_{1:T}|X_{1:T}, Y_{1:T-1})\, dZ_1 \dots dZ_T \\
&= \int P(Y_T|Z_T)\, P(Z_{1:T}|X_{1:T}, Y_{1:T-1})\, dZ_1 \dots dZ_T \\
&= \frac{1}{P(X_{1:T}, Y_{1:T-1})} \int P(Y_T|Z_T)\, P(X_{1:T}, Y_{1:T-1}|Z_{1:T})\, P(Z_{1:T})\, dZ_1 \dots dZ_T \\
&= \frac{1}{P(X_{1:T}, Y_{1:T-1})} \int \left[\prod_{t=1}^{T} P(X_t|Z_t)\, P(Y_t|Z_t)\right] P(Z_{1:T})\, dZ_1 \dots dZ_T
\end{aligned} \tag{3.1}$$

Similarly,

$$P(Y_T|X_{1:T-1}, Y_{1:T-1}) = \frac{1}{P(X_{1:T-1}, Y_{1:T-1})} \int \left[\prod_{t=1}^{T-1} P(X_t|Z_t)\, P(Y_t|Z_t)\right] P(Y_T|Z_T)\, P(Z_{1:T})\, dZ_1 \dots dZ_T \tag{3.2}$$

$$\begin{aligned}
P(Z_T|X_{1:T}, Y_{1:T-1}) &= \int P(Z_{1:T}|X_{1:T}, Y_{1:T-1})\, dZ_1 \dots dZ_{T-1} \\
&= \frac{P(X_T|Z_T)}{P(X_{1:T}, Y_{1:T-1})} \int \left[\prod_{t=1}^{T-1} P(X_t|Z_t)\, P(Y_t|Z_t)\right] P(Z_{1:T})\, dZ_1 \dots dZ_{T-1}
\end{aligned} \tag{3.3}$$

$$P(Z_T|X_{1:T-1}, Y_{1:T-1}) = \frac{1}{P(X_{1:T-1}, Y_{1:T-1})} \int \left[\prod_{t=1}^{T-1} P(X_t|Z_t)\, P(Y_t|Z_t)\right] P(Z_{1:T})\, dZ_1 \dots dZ_{T-1} \tag{3.4}$$

Covariance of $Y_T|X_{1:T-1}, Y_{1:T-1}$ can be written in terms of the state variables:

$$\begin{aligned}
\operatorname{cov}\left[Y_T|X_{1:T-1}, Y_{1:T-1}\right] &= \mathrm{E}\left[\underbrace{\operatorname{cov}\left[Y_T|X_{1:T-1}, Y_{1:T-1}, Z_T\right]}_{\operatorname{cov}[Y_T|Z_T]}|X_{1:T-1}, Y_{1:T-1}\right] \\
&\quad + \operatorname{cov}\left[\underbrace{\mathrm{E}\left[Y_T|X_{1:T-1}, Y_{1:T-1}, Z_T\right]}_{\mathrm{E}[Y_T|Z_T]}|X_{1:T-1}, Y_{1:T-1}\right] \\
&= \mathrm{E}\left[\operatorname{cov}\left[Y_T|Z_T\right]|X_{1:T-1}, Y_{1:T-1}\right] \\
&\quad + \operatorname{cov}\left[\mathrm{E}\left[Y_T|Z_T\right]|X_{1:T-1}, Y_{1:T-1}\right] \qquad (3.5)
\end{aligned}$$

### 3.1.1 Gaussian returns

$Y_t|Z_t \sim \mathcal{N}\left(\mu_t(Z_t), \Sigma_t(Z_t)\right)$ is a canonical model of stock returns, with $\mu_t$ independent of $\{Y_s, \mu_s, \Sigma_s\}_{s<t}$ through market efficiency. $\mu$ is typically close to 0 and of smaller magnitude than the standard deviation.

## 3.2 Examples

The specific cases below illustrate how the model accommodates standard techniques for working with investment returns.

### 3.2.1 Factor-modeled covariance

This is the standard model throughout this dissertation. Next period's return covariance is point-estimated from history, $\operatorname{cov}[Y_T|X_{1:T-1}, Y_{1:T-1}]$, by assuming returns follow a linear regression model and computing the covariance of the modeled returns with assumptions on independence.

$$
\begin{aligned}
F_t &\equiv \text{factors} \in \mathbb{R}^k \\
\beta &\equiv \text{exposures} \in \mathbb{R}^{n\times k} \\
\varepsilon_t &\equiv \text{idiosyncratic effects} \in \mathbb{R}^n \\
\Omega &\equiv \text{covariance of factors} \in \mathbb{R}^{k\times k} \\
\Lambda &\equiv \text{covariance of idiosyncratic effects} \in \mathbb{R}^{n\times n}\text{, a diagonal matrix}
\end{aligned}
$$

$$
\begin{aligned}
Y_t &= \beta\, F_t + \varepsilon_t \\
F_t|\Omega &\overset{\text{iid}}{\sim} \mathcal{N}(0,\Omega)\text{, ind of } \{\beta,\varepsilon\} \\
\varepsilon_t|\Lambda &\overset{\text{iid}}{\sim} \mathcal{N}(0,\Lambda)\text{, ind of } \{\beta,F\}
\end{aligned}
$$

$$
\begin{aligned}
Z_t &\equiv \{\beta,\Omega,\Lambda,F_t\} \\
P(Z_1,\ldots,Z_T) &= \pi\left(\beta,\Omega,\Lambda\right)\prod_{t=1}^{T} P(F_t|\Omega)
\end{aligned}
$$

$$
\begin{aligned}
Y_t|Z_t &\sim \mathcal{N}\left(\beta F_t,\Lambda\right) \\
Y_t|Z_t \setminus F_t &\sim \mathcal{N}\left(0,\beta\,\Omega\,\beta^{\mathrm{T}}+\Lambda\right) \\
X_t &\equiv \{F_t\} \\
X_t|Z_t &\sim \delta_{F_t}
\end{aligned}
$$

As in equation (3.5) but conditioning on $Z_T \setminus F_T$,

$$\begin{aligned}
\operatorname{cov}\left[Y_T | X_{1:T-1}, Y_{1:T-1}\right] &= \mathrm{E}\left[\underbrace{\operatorname{cov}\left[Y_T | X_{1:T-1}, Y_{1:T-1}, Z_T \setminus F_T\right]}_{\beta\,\Omega\,\beta^{\mathrm{T}}+\Lambda} | X_{1:T-1}, Y_{1:T-1}\right] \\
&\quad + \operatorname{cov}\left[\underbrace{\mathrm{E}\left[Y_T | X_{1:T-1}, Y_{1:T-1}, Z_T \setminus F_T\right]}_{\beta\,\mathrm{E}[F_T | Z_T \setminus F_T]=0} | X_{1:T-1}, Y_{1:T-1}\right] \\
&= \mathrm{E}\left[\beta\,\Omega\,\beta^{\mathrm{T}} + \Lambda | X_{1:T-1}, Y_{1:T-1}\right] \qquad (3.6)
\end{aligned}$$

Versions with more complicated econometrics have non-iid $\varepsilon_t$, e.g., following an ARMA, and $\beta, \Omega$, and $\Lambda$ varying with $t$ according to some dynamics.

### 3.2.2 Multivariate GARCH

Multivariate GARCH(p,q) models time series with heteroskedasticity. The general form:

$$\begin{aligned}
Y_t | \mu_t, \Sigma_t &\sim \mathcal{N}\left(\mu_t, \Sigma_t\right) \\
\Sigma_t &= f\left(Y_{t-p}\, Y_{t-p}^{\mathrm{T}}, \ldots, Y_{t-1}\, Y_{t-1}^{\mathrm{T}}\right) + g\left(\Sigma_{t-q}, \ldots, \Sigma_{t-1}\right)
\end{aligned}$$

$\mu_t$ is usually fixed ($\mu_t = \mu$) or zero. Different assumptions on $f$ and $g$ lead to numerous acronym-named variants[1]. It is also possible to incorporate side information. Among the many versions of GARCH, here are two to show how it fits within the model above.

[1] Silvennoinen and Teräsvirta [2009] is an overview.

**Vanilla Multivariate GARCH(1,1)**

To limit the number of parameters, structure needs to be imposed. One form is GARCH(1,1) in the 'BEKK' model of Engle and Kroner [1995]:

$$
\begin{aligned}
Y_t|\mu,\Sigma_t &\sim \mathcal{N}\left(\mu,\Sigma_t\right) \\
\Sigma_t &= aa^{\mathrm{T}} + \sum_{i=1}^{K} b_i \left(Y_{t-1}-\mu\right)\left(Y_{t-1}-\mu\right)^{\mathrm{T}} b_i^{\mathrm{T}} + \sum_{i=1}^{K} c_i \,\Sigma_{t-1}\, c_i^{\mathrm{T}} \\
&\text{where } a, b_i, c_i \in \mathbb{R}^{n\times n},\ i = 1\ldots K \\
Z_t &= \{\mu, a, b, c, \Sigma_t, Y_t\} \\
P(Z_1,\ldots,Z_T) &= \pi\left(\mu,a,b,c,\Sigma_1\right)\prod_{t=2}^{T} P(\Sigma_t|\Sigma_{t-1},Y_{t-1},a,b,c)\prod_{t=1}^{T} P(Y_t|\mu,\Sigma_t) \\
Y_t|Z_t &\sim \delta_{Y_t} \\
X_t &\equiv \{\ \}
\end{aligned}
$$

**Factor model with GARCH**

GARCH can be applied within a factor model, for example, GARCH(1,1) in Engle et al. [1990] and Vrontos et al. [2003]. From Vrontos et al. [2003]:

$$
\begin{aligned}
F_t &\equiv \text{unobserved factors} \in \mathbb{R}^k \\
\beta &\equiv \text{exposures} \in \mathbb{R}^{n \times k} \\
\Theta_t &\equiv \text{dynamic factor variances} \in \mathbb{R}^k \\
a, b, c &\equiv \text{parameters} \in \mathbb{R}^k
\end{aligned}
$$

$$
\begin{aligned}
Y_t &= \mu + \beta F_t \\
F_t &= \text{diag}\left[\Theta_t\right]^{\frac{1}{2}} \times \text{iid}\, \mathcal{N}\left[0, I_k\right] \\
\Theta_{i,t} &= a_i + b_i F_{i,t-1}^2 + c_i\, \Theta_{i,t-1} \\
F_t | \Theta_t &\sim \mathcal{N}\left(0,\, \text{diag}\left[\Theta_t\right]\right)
\end{aligned}
$$

$$
\begin{aligned}
Z_t &\equiv \{\mu, \beta, a, b, c, \Theta_t, F_t\} \\
P(Z_1, \ldots, Z_T) &= \pi\left(\mu, \beta, a, b, c, \Theta_1\right) P(F_1 | \Theta_1) \\
&\qquad \prod_{t=2}^{T} P(\Theta_t | \Theta_{t-1}, F_{t-1}, a, b, c) P(F_t | \Theta_t)
\end{aligned}
$$

$$
\begin{aligned}
Y_t | Z_t \setminus F_t &\sim \mathcal{N}\left[\mu, \beta\ \text{diag}\left[\Theta_t\right]\ \beta^{\mathrm{T}}\right] \\
X_t &\equiv \{\ \}
\end{aligned}
$$

$$
\begin{aligned}
\text{cov}\left[Y_T | Y_{1:T-1}\right] &= \text{E}\left[\text{cov}\left[Y_T | \mu, \beta, \Theta_T, Y_{1:T-1}\right] | Y_{1:T-1}\right] \\
&\quad + \text{cov}\left[\underbrace{\text{E}\left[Y_T | \mu, \beta, \Theta_T, Y_{1:T-1}\right]}_{\mu} | Y_{1:T-1}\right] \\
&= \text{E}\left[\beta\ \text{diag}\left[\Theta_T\right]\ \beta^{\mathrm{T}} | Y_{1:T-1}\right] + \text{cov}\left[\mu | Y_{1:T-1}\right]
\end{aligned}
$$

# CHAPTER 4

# SHORT-TERM COVARIANCE GIVEN INSTANTANEOUS SIDE INFORMATION

Covariance of stock returns varies over time. Volatility tends to increase abruptly and fall slowly.[1]For example, a chart of the S&P 500 option-implied volatility VIX looks different forward (impenetrable vertical faces, gradual drops) than backward (gradual increases, cliffs). Correlations between securities also shift. However unpredictable, these changes are apparent in sources of side information almost immediately, long before behavior can be inferred from observing the process alone.

This chapter describes a Bayesian framework for incorporating side information to forecast return covariance over a given horizon (e.g., one or two weeks) and frequency (e.g., daily or weekly returns). One particular parameterization is computationally fast, solved by linear or quadratic programming depending on distribution assumptions.

## 4.1 Literature review

Originating from Engle [1982] and Bollerslev [1986], ARCH/GARCH models are a way to capture dynamic volatility in time series. The general idea—updating variance at a point toward recent error between fit and data—has many variants. Engle [2004] is a survey, and Hansen and Lunde [2005] a comparison. Multivariate versions include Bollerslev [1990] and the DCC model of Engle [2002]; Bauwens et al. [2006] is a survey. Engle et al. [2019] improves DCC by adding the random matrix theory shrinkage of Ledoit and Wolf [2012]. Engle et al. [1990] and section 6.2 of Bollerslev et al. [1994]

[1]Section 1.2 of Bollerslev et al. [1994] describes some of the structure present in stock returns.

apply ARCH to factor-modeled covariance. Though often used, like many econometric time series techniques, ARCH type models miss relevant information.

Range-based estimators infer volatility from a shorter, i.e., more recent, history. Parkinson [1980] models log(price) as a Brownian motion and, from the distribution of its high-low[2], works out the ratio between powers of volatility and expected powers of log(intraday high price/low) yielding an estimator with the accuracy of the unbiased time series estimator in 1/5 the observations. The idea can be applied at higher frequency for shorter lag and more localization. Other range estimators, adding the information of price at market open and close, include Garman and Klass [1980], Beckers [1983], Rogers and Satchell [1991], Kunitomo [1992], and Yang and Zhang [2000]. Chou et al. [2010] is a good overview. Correlation between securities can also be inferred, as in Rogers and Zhou [2008] and Chou et al. [2009]. ARCH style models that update volatility with range information are Chou [2005], Brandt and Jones [2006], Hansen et al. [2012], and De Nard et al. [2021] which is Engle et al. [2019] with range information added.

Implied volatility, from market-traded options and indices, e.g., VIX, is available instantaneously and contains the forward-looking, high quality information of market participants. Cross-sectional volatility, the dispersion of returns across securities, reveals correlation and can be computed over any interval. There are also continually updated predictive signals, for example, stock message board postings in Antweiler and Frank [2004] and Das and Chen [2007], news in Mitra et al. [2009] and Atkins et al. [2018], and internet search queries in Dimpfl and Jank [2016] in the ever-growing list of alternative data. Two examples of ARCH style models that update volatility with news signals are Kalev et al. [2004] and Robertson et al. [2007].

The fragmented information from these current information sources relates to particular aspects of covariance, e.g., the variance of a certain stock or the spread across constituents of an index. Forming a forecast covariance matrix requires grafting onto structure. One way is to gather structure from history, then tune levels to the information. Made public in 1998, diBartolomeo and Warrick [2005] updates an orthogonal-factor model of covariance,

$$\Sigma = E \ \mathrm{diag}(\Lambda) \, E^{\mathrm{T}} + \mathrm{diag}(\varepsilon^2), \tag{4.1}$$

[2]From Feller [1951].

with option-implied-volatility-forecasted security variances $\hat{v}$ by applying scalars $\theta$ to the factor variances,

$$\Sigma_\theta = E \operatorname{diag}(\Lambda) \operatorname{diag}(\theta) E^{\mathrm{T}} + \operatorname{diag}(\varepsilon^2), \tag{4.2}$$

and setting their value through constrained L2 regression,

$$\theta^* = \underset{\theta \geq 0}{\arg\min} \sum_i \left( \sum_k E_{ik}^2 \Lambda_k \theta_k + \varepsilon_i^2 - \hat{v}_i \right)^2. \tag{4.3}$$

Mitra et al. [2009] does the same adding signals from quantified news. Shah [2007] formalizes the idea into a Bayesian framework for fitting any factor model to forecasts of functions of covariance and derives a method to scale frequency to correct mismatch in underlying structure, e.g., monthly to daily returns. Shah [2014], the basis of this chapter, is a generalization that has a computationally fast parameterization for certain types of forecasts. Miranyan [2012] introduces scalars into a diagonalized factor model and sets their value by minimizing Frobenius distance to a target full covariance matrix having correlation from history and standard deviations from individual stock volatility forecasts.

## 4.2 Problem formulation

The goal is a forecast of return covariance over some frequency and horizon given history and side information. Two requirements are 1–modeling the present since covariance changes with market conditions and 2–fast computation since the horizon of interest can be short.

Within the mathematical model of chapter 3 (stock returns $Y$, side information $X$, latent states $Z$), assume the covariance takes a known parametric form,

$$\operatorname{cov}[Y_t | S_t] = f(S_t) \text{ where } S_t \text{ is a subset of } Z_t$$

One example is the factor model of 3.2.1,

$$\begin{aligned} \text{cov}\left[Y_t|S_t\right] &= \beta\,\Omega\,\beta^{\mathrm{T}} + \Lambda \\ S_t &\equiv \{\beta, \Omega, \Lambda\} \\ Z_t &\equiv S_t \cup \{F_t\} \\ \text{where } Y_t &\equiv \beta\,F_t + \varepsilon_t \\ F_t &\sim \text{iid } \mathcal{N}\left[0, \Omega\right] \\ \varepsilon_t &\sim \text{iid } \mathcal{N}\left[0, \Lambda\right] \end{aligned} \tag{4.4}$$

Side information contains noisy estimates of functions of covariance built from contemporaneous and recent data,

$$\{G_t, g_t\}_t \text{ where } G_t = g_t\left(\text{cov}\left[Y_t|S_t\right]\right) + \eta_t \tag{4.5}$$

Two practical examples of side information:

1. Current estimates of return variance for individual securities, $\text{cov}\left[Y_t\right]_{i,i}$, or portfolios, $w^{\mathrm{T}}\text{cov}\left[Y_t\right]w$. One source is range-based volatility estimation: assuming a model for security movement, e.g., geometric Brownian motion, variance is inferred from high and low prices over an interval. Another source is implied volatility from continuously traded options. A third is forecasting via any regression, including machine learning[3], fed recent data.

2. Cross-sectional variance of the returns of a portfolio's constituents over an interval, $\sum_i w_i\left(r_i - w^{\mathrm{T}}r\right)^2$. Its expected value $\approx$ interval length$\times w^{\mathrm{T}}\,\text{diag}(\text{cov}\left[Y_t\right]) - w^{\mathrm{T}}\text{cov}\left[Y_t\right]w$. Paired with individual security variance estimates, these capture aspects of correlation.

The covariance forecast is calibrated to current market conditions by fitting parameters to side information and history as in equation (3.4) (repeated here),

$$P(Z_T|X_{1:T-1}, Y_{1:T-1}) = \frac{1}{P(X_{1:T-1}, Y_{1:T-1})}\int\left[\prod_{t=1}^{T-1}P(X_t|Z_t)\,P(Y_t|Z_t)\right]P(Z_{1:T})\,dZ_1\ldots dZ_{T-1}$$

Through both $P(Z_{1:T})$ and changes in side information, slow-moving parameters are determined more by history, and fast-moving parameters by current side information.

[3] e.g., Zhang et al. [2023]

## 4.3 A particular parameterization

The fast computation parameterization of the paper starts from a factor model as in equation (4.4), diagonalizes the covariance between factors, and inserts time-varying variances. The factor covariance matrix $\Omega$ is replaced by $u\ \mathrm{diag}\left[\Theta_t\right]\ u^{\mathrm{T}}$, and the idiosyncratic variance $\Lambda$ by $\gamma_t\,\Lambda$.

$$\begin{aligned}
\mathrm{cov}\left[Y_t|S_t\right] &= \beta\,u\ \mathrm{diag}\left[\Theta_t\right]\ u^{\mathrm{T}}\,\beta^{\mathrm{T}} + \gamma_t\,\Lambda \qquad (4.6)\\
S_t &\equiv \{\beta, u, \Lambda, \Theta_t, \gamma_t\}\\
Z_t &\equiv S_t \cup \{F_t\}\\
\text{where } Y_t &\equiv \beta\,F_t + \varepsilon_t\\
F_t &\equiv u\ \mathrm{diag}\left[\Theta_t\right]^{\frac{1}{2}} \times \text{iid}\ \mathcal{N}\left[0, I_k\right]\\
\varepsilon_t &\equiv \sqrt{\gamma_t} \times \text{iid}\ \mathcal{N}\left[0, \Lambda\right]\\
\Theta_t, \gamma_t &\equiv \text{dynamic variances} \in \mathbb{R}^k,\ \mathbb{R}\\
u, \Lambda &\equiv \text{structure} \in \mathbb{R}^{k\times k},\ \mathbb{R}^{n\times n},\ \Lambda \text{ is diagonal}
\end{aligned}$$

$$\begin{aligned}
\text{So,}\quad Y_t|Z_t &\sim \mathcal{N}\left(\beta F_t, \gamma_t\,\Lambda\right) \qquad (4.7)\\
Y_t|S_t &\sim \mathcal{N}\left(0, \beta\,u\ \mathrm{diag}\left[\Theta_t\right]\ u^{\mathrm{T}}\,\beta^{\mathrm{T}} + \gamma_t\,\Lambda\right) \qquad (4.8)
\end{aligned}$$

Add side information:

$$\begin{aligned}
X_t &\equiv \{F_t, G_t, g_t\left(\cdot\right)\}\\
\text{where } G_t &= g_t\left(\mathrm{cov}\left[Y_t|S_t\right]\right) + \eta_t \quad \in \mathbb{R}^{m_t} \text{ as in equation (4.5)}\\
&= g_t(\beta\,u\ \mathrm{diag}\left[\Theta_t\right]\ u^{\mathrm{T}}\,\beta^{\mathrm{T}} + \gamma_t\Lambda) + \eta_t\\
&= g_t\left(Z_t\right) + \eta_t, \quad \eta_t \sim P_{\eta,t} \text{ ind of everything else}\\
X_t|Z_t &\sim \delta_{F_t}\,P_{\eta,t}(G_t - g_t(Z_t)) \qquad (4.9)
\end{aligned}$$

and a distribution on $Z$:

$$\begin{aligned}
Z_t &\equiv \{\beta, u, \Lambda, \Theta_t, \gamma_t, F_t\}\\
P(Z_1, \ldots, Z_T) &= \delta_{\beta,u,\Lambda} \times \pi\left(\beta, u, \Lambda, \Theta_0, \gamma_0\right) \times (\text{dynamics on } \Theta, \gamma) \times \prod_{t=1}^{T} P(F_t|u, \Theta_t) \qquad (4.10)
\end{aligned}$$

For dynamics, $\Theta_t, \gamma_t$ might be Markov $\sim \phi(\Theta_{t-1}, \gamma_{t-1})$ where $\phi$ is some non-negative distribution, e.g., log-normal with known parameters. Then,

$$P(Z_{1:T}) = \delta_{\beta,u,\Lambda} \times \pi\left(\beta, u, \Lambda, \Theta_0, \gamma_0\right) \times \prod_{t=1}^{T} \phi(\Theta_{t-1}, \gamma_{t-1}) \times \prod_{t=1}^{T} P(F_t | u, \Theta_t)$$

Conditioning on $Z_T \setminus F_T$ as in equation (3.6),

$$\begin{aligned}
\operatorname{cov}\left[Y_T | X_{1:T-1}, Y_{1:T-1}\right] &= \mathrm{E}\left[\underbrace{\operatorname{cov}\left[Y_T | X_{1:T-1}, Y_{1:T-1}, Z_T \setminus F_T\right]}_{\beta\, u\, \operatorname{diag}[\Theta_T]\, u^{\mathrm{T}}\, \beta^{\mathrm{T}} + \gamma_T\, \Lambda} | X_{1:T-1}, Y_{1:T-1}\right] \\
&\quad + \operatorname{cov}\left[\underbrace{\mathrm{E}\left[Y_T | X_{1:T-1}, Y_{1:T-1}, Z_T \setminus F_T\right]}_{\beta\, \mathrm{E}[F_T | Z_T \setminus F_T] = 0} | X_{1:T-1}, Y_{1:T-1}\right] \\
&= \mathrm{E}\left[\beta\, u\, \operatorname{diag}\left[\Theta_T\right]\, u^{\mathrm{T}}\, \beta^{\mathrm{T}} + \gamma_T\, \Lambda | X_{1:T-1}, Y_{1:T-1}\right]
\end{aligned}$$

The distribution of parameters $P(Z_T | X_{1:T-1}, Y_{1:T-1})$ can be computed via equation (3.4) from $P(Y_t|Z_t)$, $P(X_t|Z_t)$, and $P(Z_{1:T})$ (equations 4.7, 4.9, and 4.10). Sampling numerically yields an approximate expectation and likely configurations. However, practical considerations intrude.

## 4.4 Practical considerations

1. Data. Some numbers are missing, wrong, or peculiar and unindicative of the future[4]. Additionally, the universe of securities changes — stocks disappear, merge, or suddenly appear, e.g., after an initial public offering or with the creation of the euro in 1999. Thus, the matrix of relevant historic returns $Y_{1:T}$ is incomplete and dirty. In practice, steps are taken to trim outliers and proxy from similar securities. Admittedly, this could be addressed through an informative prior on ($\beta$, $\Lambda$) with missing entries of $Y_{1:T}$ unobserved.

2. Complicated econometric modeling. For example, $\beta$ is often inferred using outside data, e.g., from SEC financial filings. While making $\beta$, $\Lambda$ time-dependent and rolling the outside data into a prior for each period could accommodate within the formal framework, other issues such as markets trading over different time intervals in Asia and the US are trickier. Because estimation specifics

can be both involved and continually revised, there is an advantage to modularity.

3. Reporting. Firms across the investment industry use a common set of models of established structure to communicate, e.g., for comparison, auditing, and showing portfolio risks without revealing positions. A model built from scratch is less valuable than a standard adapted to current conditions. For example, with the latter, one could see how much risk different well-known factors contribute over several horizons, e.g., 2 days, 2 weeks, and 1 month.

4. Computational speed. Models need to be recomputed quickly under short horizons or regime change, e.g., at the inception of the coronvirus pandemic. Separately, it is likely that, even with full computation, material change occurs in few areas.

For these reasons, the approach taken is to forecast covariance given current information by altering an existing estimate, $\Sigma_t$, using only the most recent observation of current information, $X_t$, rather than the full history, $\{X_{1:t}, Y_{1:t}\}$. Assuming the effects of history and dynamics to be encapsulated in the prior simplifies the problem to a single period. Though a distribution is available, only the MAP estimate of $S_t$ is used in a point-estimate of covariance.

[4] A difference between a market and a physical system

**from Shah [2014] Short-Term Risk and Adapting Covariance to Current Market Conditions**

## 4.5 Abstract

Covariance models of stock returns appear throughout the investment process, e.g., forecasting portfolio risk, hedging, constructing mean-variance optimal portfolios, and algorithmic trading. Typically built from historic time-series, they estimate the past but—because markets and regimes continually change—not the present and future. There are non-time-series, instantaneous ways to infer or predict volatility, e.g., option implied volatility, intra-period trading range, machine learning on alternative data. This paper describes a conceptual framework and algorithm for incorporating these inferences into any linear factor covariance model so that it represents one's beliefs about future behavior instead of echoing the past.

## 4.6 Key Messages

1. Factor covariance models are built from historic series, necessary to infer structure.
2. Thus, their information is stale and ignores much of what's known and relevant.
3. Expressing (often non-time series) current information as quantities relating to covariance, one can inject it into a factor model to forecast the current and future.

## 4.7 Introduction

Covariance models of stock returns appear throughout the investment process, e.g., forecasting portfolio risk, hedging, constructing mean-variance optimal portfolios, and algorithmic trading. They are built from historical data, with more complex models

requiring more data thus reaching farther back in time. Richness in structure is impossible without data.

But the world continually changes. What happened years ago or recently in a different regime poorly represents the present and future. Morevoer, the data entering these models is just a fraction of what's relevant. Many immediate information sources—ignored due to econometric legacy—can be used to infer or predict volatility, e.g., the cross-sectional spread of returns, option implied volatility, intra-period high/low prices, volume, bond spreads, short interest, news/social media and other alternative data. What follows is a framework and algorithm for incorporating this information.

### 4.7.1 Background

The literature on several related threads is rich and well-developed. Engle [1982] introduced ARCH models to forecast volatility of heteroscedastic time-series from history. Countless variants have since been investigated. Parkinson [1980] describes one of many estimators to infer volatility of stock returns from intra-period high/low prices or other intra-period statistics. Antweiler and Frank [2004] forecasted volatility from stock message board postings. Today, alternative data sources are usually the best way to observe or predict essentially every quantity of interest in the world.

On this paper's particular thread, diBartolomeo and Warrick [2005] adjusted an orthogonal factor covariance model to match forward-looking option-implied volatility. Mitra et al. [2009] applied the technique to incorporate an aspect of quantified news. Shah [2007] presented a Bayesian framework to adjust models whose factors are correlated while considering information accuracy. Miranyan [2012] used stock level volatility information to adjust a diagonalized factor model – first creating a target by inflating or deflating rows and columns of modeled covariance to match individual stock volatilities, then scaling the model's factor variances to yield covariance as close (measured by Frobenius distance) to the target as possible.

This paper contributes 1–a general framework to incorporate, considering accuracy, forecasts of any quantity that can be calculated by a covariance matrix; 2–a computationally fast algorithm to incorporate stock, portfolio, and cross-sectional volatility forecasts with consideration of accuracy; 3–a model and formula to scale

covariance to longer/shorter horizons under serial correlation; and 4–a diagonalization of factor covariance that maximizes explained variance of securities linked to factors by exposures.

### 4.7.2 General Structure of Factor Covariance Models

Because there are too many securities to accurately infer all the pairwise entries, models are built on a skeleton of common factors, e.g., market, industry, country, and growth/value spread. Covariance between securities happens through links to these factors. Movement uncorrelated with the factors is treated as uncorrelated across securities and called stock-specific.

The factor model view of covariance:

$$\begin{aligned} \boldsymbol{C} &= (n \times n) \text{ modeled covariance of stock returns} \\ &= \boldsymbol{E}\boldsymbol{F}\boldsymbol{E}^{\mathrm{T}} + \mathrm{diag}(\boldsymbol{\varepsilon}^2) \qquad (4.11) \\ \text{where } n &= \#\text{ of securities} \\ m &= \#\text{ of factors} \\ \boldsymbol{E} &= (n \times m) \text{ security exposures to factors} \\ \boldsymbol{F} &= (m \times m) \text{ covariance between factors} \\ \boldsymbol{\varepsilon}^2 &= (n \times 1) \text{ uncorrelated stock specific variance} \end{aligned}$$

The individual pieces are typically inferred from time-series reaching back 3-5 years. Some might come from stock prices and accounting or econometric data updated perhaps quarterly.

## 4.8 Current Information

Current information here is communicated through *forecasts of quantities related to variance or covariance* over one's horizon. Forecasts can be made exclusively from instantly observable data, e.g., VIX, or involve history, e.g., the volume of news today and every day over the past year. They can be the result of black box machine learning on alternative data.

A quantity forecasted might be the variance of TSLA over the next 3 months or the standard deviation of the S&P500 over the next day. Quantities can be relative—the variance of TSLA over the next 3 months divided by its variance over the past 5 years.

Insufficient to construct the entire covariance, *current information is projected onto the structure of a base covariance model*. More information yields projections that are more descriptive and, over time, dynamic. Less, and the projection resembles the base model. Communicating information in quantities calculated by a covariance matrix makes the process possible. Examples:

### 4.8.1 Portfolio or Individual Security Variance

$$\begin{aligned} \sigma_v^2 &= \text{variance of portfolio } \boldsymbol{v} \\ &= \boldsymbol{v}^{\mathrm{T}} \boldsymbol{C} \boldsymbol{v} \\ \text{where } \boldsymbol{v} &= (n \times 1) \text{ portfolio weights} \end{aligned} \tag{4.12}$$

Useful cases:

1. variance stripped of market (or other) effects[5]
   $\boldsymbol{v}$ = portfolio weights $-\ \beta \times$ market weights

2. tracking error
   $\boldsymbol{v}$ = portfolio weights $-$ benchmark weights

3. The difference in variance between two portfolios, e.g. a stock's variance above the market

[5]Imagine different sources are relevant for predicting market and non-market behavior, e.g., VIX for market, company news for non-market. In a forecast of combined market and non-market behavior, the (very large) market effect would dominate the error. Without separating the two, information in the non-market part would get lost.

### 4.8.2 Expected Cross-Sectional Volatility

The cross-sectional expected squared difference from the average return, weighted by portfolio value

$$\begin{aligned} \sigma^2_{v,cs} &= \text{expected cross-sectional variance}[6]\text{of portfolio } \boldsymbol{v} \\ &\approx \boldsymbol{v}^{\mathrm{T}}\boldsymbol{\omega}^2 - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v} \qquad (4.13) \\ \text{where } \boldsymbol{\omega}^2 &= (n \times 1) \text{ variance of stock returns} \\ &= \operatorname{diag}(\boldsymbol{C}) \qquad (4.14) \end{aligned}$$

### 4.8.3 Pairwise Correlation between Securities or Portfolios

$$\begin{aligned} \rho_{u,v} &= \text{correlation between portfolios } \boldsymbol{u} \text{ and } \boldsymbol{v} \\ &= \frac{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v}}{\sqrt{\left(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{u}\right)\left(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v}\right)}} \qquad (4.15) \end{aligned}$$

### 4.8.4 Average Correlation within a Portfolio

$$\begin{aligned} \bar{\rho}_v &= \text{average correlation}[7]\text{within portfolio } \boldsymbol{v} \\ &= \frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v} - \left(\boldsymbol{v}^2\right)^{\mathrm{T}}\boldsymbol{\omega}^2}{\left(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{\omega}\right)^2 - \left(\boldsymbol{v}^2\right)^{\mathrm{T}}\boldsymbol{\omega}^2} \qquad (4.16) \end{aligned}$$

## 4.9 Bayesian Framework for Incorporating Current Information

Presented in Shah [2007], the framework is as follows: Into the model (1), introduce parameters $\boldsymbol{\theta}$ which will serve to characterize the state of the market. State here

[6] Derivation in 4.14

[7] Derivation in 4.15

resembles regime but can be richer and more granular. Let $\pi(\boldsymbol{\theta})$ be $\boldsymbol{\theta}$'s prior distribution. Current information forecasts, $\hat{\boldsymbol{g}}$, are noisy estimates of quantities, $\boldsymbol{g}$, calculated by the model: $\boldsymbol{g}$ = functions of the covariance matrix under $\boldsymbol{\theta}$.

Written in terms of $\boldsymbol{\theta}$, $\boldsymbol{g} = \boldsymbol{f}(\boldsymbol{\theta})$.

Estimates $\hat{\boldsymbol{g}} = \boldsymbol{g} + \boldsymbol{\delta}$.

Errors, $\boldsymbol{\delta}$, follow some distribution, $q(\boldsymbol{\delta})$.

These give the distribution of $\boldsymbol{\theta}$ given $\hat{\boldsymbol{g}}$:

$$\begin{aligned} p\left(\boldsymbol{\theta}|\hat{\boldsymbol{g}}\right) &= p\left(\hat{\boldsymbol{g}}|\boldsymbol{\theta}\right)\, \pi\left(\boldsymbol{\theta}\right)\, / p\left(\hat{\boldsymbol{g}}\right) \\ &= q\left(\hat{\boldsymbol{g}} - \boldsymbol{g}\right)\, \pi\left(\boldsymbol{\theta}\right)\, / p\left(\hat{\boldsymbol{g}}\right) \\ &= q\left(\hat{\boldsymbol{g}} - \boldsymbol{f}[\boldsymbol{\theta}]\right)\, \pi\left(\boldsymbol{\theta}\right)\, / p\left(\hat{\boldsymbol{g}}\right) \\ &\propto q\left(\hat{\boldsymbol{g}} - \boldsymbol{f}[\boldsymbol{\theta}]\right)\, \pi\left(\boldsymbol{\theta}\right) \end{aligned} \tag{4.17}$$

Now the question can be posed, "How does the world appear considering the forecasts?" The expected value is $\mathrm{E}\left[\boldsymbol{\theta}\right]$ taken under (4.17). The most probable configuration—the Bayesian analog of frequentist statistics' maximum likelihood—is the maximum a posteriori (MAP) estimate:

$$\begin{aligned} \boldsymbol{\theta}^{*} &= \arg\max_{\boldsymbol{\theta}}\ p\left(\boldsymbol{\theta}|\hat{\boldsymbol{g}}\right) \\ &= \arg\max_{\boldsymbol{\theta}}\ \log\left[p\left(\boldsymbol{\theta}|\hat{\boldsymbol{g}}\right)\right] \\ &= \arg\max_{\boldsymbol{\theta}}\ \log\left[q\left(\hat{\boldsymbol{g}} - \boldsymbol{f}[\boldsymbol{\theta}]\right)\right] + \log\left[\pi\left(\boldsymbol{\theta}\right)\right] \end{aligned} \tag{4.18}$$

The first term fits toward observations; the second stabilizes and seeks coherence with prior beliefs. Three questions: 1-How are the parameters introduced? 2-What is the observation error distribution? 3-What is a reasonable prior? Consider an example:

### 4.9.1 Parameters to Accommodate State

Recall the base model (4.11), $\boldsymbol{C} = \boldsymbol{E}\boldsymbol{F}\boldsymbol{E}^{\mathrm{T}} + \mathrm{diag}(\boldsymbol{\varepsilon}^{2})$. Among the many possibilities, for this example, place a scalar on the standard deviation of each factor and on stock-specific volatility.

$$\begin{aligned}
\boldsymbol{C}(\boldsymbol{\theta}) &= (n \times n) \text{ covariance} \\
&= \boldsymbol{E}\left[\operatorname{diag}(\boldsymbol{\theta}_{1..m})\,\boldsymbol{F}\operatorname{diag}(\boldsymbol{\theta}_{1..m})\right]\boldsymbol{E}^{\mathrm{T}} + \boldsymbol{\theta}_{m+1}^{2}\operatorname{diag}\left(\boldsymbol{\varepsilon}^{2}\right) \qquad (4.19)\\
\text{where} \quad \boldsymbol{\theta} &= (m+1 \times 1) \\
\boldsymbol{\theta}_{1..m} &= \text{scales factor std deviations} \\
\boldsymbol{\theta}_{m+1} &= \text{scales security specific std deviation} \\
\text{as before,} \quad \boldsymbol{\omega}^{2}(\boldsymbol{\theta}) &= (n \times 1) \text{ variance of stock returns} \\
&= \operatorname{diag}[\boldsymbol{C}(\boldsymbol{\theta})] \qquad (4.20)
\end{aligned}$$

Now the model can adapt to state.

### 4.9.2 Forecasts and Errors

Ideally, forecasts well outnumber parameters. Redundancy connects optimized parameters to reality. However, even absent data, the prior makes the problem well-posed.

Suppose the information consists of the volatility of two indices, the relative cross-sectional volatility of one, and the volatility of certain individual stocks.

$$\begin{aligned}
\hat{g}_1 &= \text{std dev of S\&P500 from VIX implied vol index} \\
\hat{g}_2 &= \text{std dev of R2000 inferred from intraday high/low} \\
\hat{g}_3 &= \text{relative level (3 day avg / 3 year avg) of daily} \\
&\quad \text{cross-sectional std dev of R2000 returns} \\
\hat{g}_4 &= \text{std dev of MSFT from option implied vol} \\
&\vdots
\end{aligned}$$

These correspond in the model to:

$$
\begin{aligned}
f_1(\boldsymbol{\theta}) &= \sqrt{\boldsymbol{v}_{SP500}^{\mathrm{T}}\, \boldsymbol{C}(\boldsymbol{\theta})\, \boldsymbol{v}_{SP500}} \\
f_2(\boldsymbol{\theta}) &= \sqrt{(\boldsymbol{v}_{R2000}^{\mathrm{T}}\, \boldsymbol{C}(\boldsymbol{\theta})\, \boldsymbol{v}_{R2000}} \\
f_3(\boldsymbol{\theta}) &= \frac{\sqrt{(\boldsymbol{v}_{R2000}^{\mathrm{T}}\, \boldsymbol{\omega}^2(\boldsymbol{\theta}) - \boldsymbol{v}_{R2000}^{\mathrm{T}}\, \boldsymbol{C}(\boldsymbol{\theta})\, \boldsymbol{v}_{R2000}}}{\sqrt{(\boldsymbol{v}_{R2000}^{\mathrm{T}}\, \boldsymbol{\omega}^2 - \boldsymbol{v}_{R2000}^{\mathrm{T}}\, \boldsymbol{C}\, \boldsymbol{v}_{R2000}}} \\
f_4(\boldsymbol{\theta}) &= \sqrt{\boldsymbol{\omega}_{MSFT}^2(\boldsymbol{\theta})} \\
&\vdots
\end{aligned}
$$

Errors can be estimated since quantities are observed repeatedly over time or across a peer group. Depending on the method of inference, econometrically derived estimates also exist.

Suppose in this example they are distributed independent Laplace, which has the advantage of making the parameter selection robust to outliers.

$$q(\hat{\boldsymbol{g}} - \boldsymbol{f}[\boldsymbol{\theta}]) = c \prod_j e^{-\sqrt{2}|\hat{g}_j - f_j(\boldsymbol{\theta})|/\sigma_j} \tag{4.21}$$

$$\log[q(\hat{\boldsymbol{g}} - \boldsymbol{f}[\boldsymbol{\theta}])] = c_2 - \sqrt{2} \sum_j |\hat{g}_j - f_j(\boldsymbol{\theta})|/\sigma_j \tag{4.22}$$

Note that when information sources have correlated errors—e.g., errors in S&P 500 and R2000 volatility forecasts result largely from changes in market volatility—ignoring the dependency gives undue influence to not-independent observations. The correlation can be explicitly considered in the error distribution or ad hoc by inflating error estimates.

### 4.9.3 Prior

A model's factors are generally observable, and their statistics paired with understanding about market behavior—e.g., how certain factor volatilities naturally correlate—can be the basis of informative priors. For the example consider a simple prior, uncorrelated Gaussians around a common center that is a Gaussian centered at 1.

$$\pi(\boldsymbol{\theta}) = c \underbrace{\prod_k e^{-\frac{(\theta_k-\theta_0)^2}{2\lambda_k^2}}}_{\text{parameters}} \underbrace{e^{-\frac{(\theta_0-1)^2}{2\lambda_0^2}}}_{\text{center}} \tag{4.23}$$

$$\log[\pi(\boldsymbol{\theta})] = c_2 - \sum_k \frac{(\theta_k-\theta_0)^2}{2\lambda_k^2} - \frac{(\theta_0-1)^2}{2\lambda_0^2} \tag{4.24}$$

The spread of the center, $\lambda_0$, can come from the standard deviation over time of (the market's volatility out the horizon, say 2 weeks / its volatility as in the model) Spreads $\lambda_1 \ldots \lambda_k$ can come from the same on the factors or simply take the value $\lambda_0$.

Assembling the pieces yields the most likely configuration:

$$\arg\max_{\boldsymbol{\theta},\theta_0} \Big\{ -\sqrt{2} \sum_j |\hat{g}_j - f_j(\boldsymbol{\theta})|/\sigma_j - \sum_k \frac{(\theta_k-\theta_0)^2}{2\lambda_k^2} - \frac{(\theta_0-1)^2}{2\lambda_0^2} \Big\} \tag{4.25}$$

Entering the maximizing values into the parametrized model of equation (4.19) yields a covariance model that represents current information.

## 4.10 Adapting to Current Market Conditions by Blending Ledoit-Wolf Covariance Reductions

The framework doesn't need a particular parameterization of covariance.

To soften outliers in portfolio optimization, Ledoit and Wolf [2004] blend the estimated covariance matrix with highly structured reductions[8]:

1. single index – covariance happens only through a market factor
2. constant correlation – all securities have the same pairwise correlation

A variant is:

3. constant covariance – all securities have the same pairwise correlation and variance

Beyond low dimensional approximations, these can be regarded as market deformations, e.g., during a panic, securities become more alike, and the market heads toward constant covariance.

[8] Derived in 4.15

### 4.10.1 Expressed as 1-Factor Models

One factor covariance is of the form $\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}\sigma^2 + \mathrm{diag}(\boldsymbol{\varepsilon}^2)$.

$$
\begin{aligned}
\text{where } \boldsymbol{e} &= (n \times 1) \text{ exposures to factor} \\
\boldsymbol{\varepsilon}^2 &= (n \times 1) \text{ uncorrelated stock specific variance} \\
\sigma^2 &= \text{variance of factor}
\end{aligned}
$$

$$
\begin{aligned}
\text{let } \boldsymbol{m} &= (n \times 1) \text{ market portfolio weights} \\
\sigma_m^2 &= \text{variance of market portfolio} \\
&= \boldsymbol{m}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{m} \\
\bar{\rho} &= \text{average correlation} \\
\bar{\sigma}^2 &= \text{average variance}
\end{aligned}
$$

The Ledoit-Wolf forms can be expressed as 1-factor models:

1. Single Index

$$
\begin{aligned}
\sigma_\beta^2 &= \sigma_m^2 \\
\boldsymbol{e}_\beta &= \boldsymbol{C}\boldsymbol{m}/\sigma_m^2 \\
\boldsymbol{\varepsilon}_\beta^2 &= \boldsymbol{\omega}^2 - \sigma_m^2\boldsymbol{e}_\beta^2 \\
\boldsymbol{C}_\beta &= \boldsymbol{e}_\beta\boldsymbol{e}_\beta^{\mathrm{T}}\sigma_m^2 + \mathrm{diag}(\boldsymbol{\varepsilon}_\beta^2)
\end{aligned}
$$

2. Constant Correlation

$$
\begin{aligned}
\sigma_{\bar{\rho}}^2 &= \bar{\rho} \\
\boldsymbol{e}_{\bar{\rho}} &= \boldsymbol{\omega} \\
\boldsymbol{\varepsilon}_{\bar{\rho}}^2 &= \boldsymbol{\omega}^2(1 - \bar{\rho}) \\
\boldsymbol{C}_{\bar{\rho}} &= \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}}\bar{\rho} + \mathrm{diag}(\boldsymbol{\omega}^2)(1 - \bar{\rho})
\end{aligned}
$$

3. Constant Covariance

$$\begin{aligned}
\sigma^2_{\bar{\rho},\bar{\sigma}} &= \bar{\rho} \\
\boldsymbol{e}^2_{\bar{\rho},\bar{\sigma}} &= \mathbf{1}\bar{\sigma} \\
\boldsymbol{\varepsilon}^2_{\bar{\rho},\bar{\sigma}} &= \mathbf{1}\bar{\sigma}^2(1-\bar{\rho}) \\
\boldsymbol{C}_{\bar{\rho},\bar{\sigma}} &= \mathbf{1}\mathbf{1}^{\mathrm{T}}\bar{\sigma}^2\bar{\rho} + \boldsymbol{I}\,\bar{\sigma}^2(1-\bar{\rho})
\end{aligned}$$

### 4.10.2 Parametrization

Consider a convex combination of the three reductions $\boldsymbol{C}_{\beta}, \boldsymbol{C}_{\bar{\rho}}, \boldsymbol{C}_{\bar{\rho},\bar{\sigma}}$ and the original $\boldsymbol{C}$:

$$\begin{aligned}
\boldsymbol{C}(\boldsymbol{\theta},\bar{\rho},\bar{\sigma}) &= \theta_1\boldsymbol{C}_{\beta} + \theta_2\boldsymbol{C}_{\bar{\rho}} + \theta_3\boldsymbol{C}_{\bar{\rho},\bar{\sigma}} + \theta_4\boldsymbol{C} \\
&= \theta_1\left[\boldsymbol{e}_{\beta}\boldsymbol{e}_{\beta}^{\mathrm{T}}\sigma_m^2 + \mathrm{diag}(\boldsymbol{\varepsilon}_{\beta}^2)\right] \\
&\quad + \theta_3\left[\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}}\bar{\rho} + \mathrm{diag}(\boldsymbol{\omega}^2)(1-\bar{\rho})\right] \\
&\quad + \theta_3\left[\mathbf{1}\mathbf{1}^{\mathrm{T}}\bar{\sigma}^2\bar{\rho} + \boldsymbol{I}\,\bar{\sigma}^2(1-\bar{\rho})\right] \\
&\quad + \theta_4\left[\boldsymbol{E}\boldsymbol{F}\boldsymbol{E}^{\mathrm{T}} + \mathrm{diag}(\boldsymbol{\varepsilon}^2)\right]
\end{aligned}$$

$$\begin{aligned}
\text{where} \quad & \boldsymbol{\theta}, \bar{\sigma} \geq 0 \\
& 0 \leq \bar{\rho} \leq 1 \\
& \sum_k \theta_k = 1
\end{aligned}$$

The most likely configuration given the prior and forecasts is:

$$\arg\max_{\boldsymbol{\theta},\bar{\rho},\bar{\sigma}} \ \log\left[q\left(\hat{\boldsymbol{g}} - \boldsymbol{f}[\boldsymbol{\theta},\bar{\rho},\bar{\sigma}]\right)\right] + \log\left[\pi\left(\boldsymbol{\theta},\bar{\rho},\bar{\sigma}\right)\right] \tag{4.26}$$

With just five free parameters—one $\theta$ doesn't count since it's determined by the others—sufficient number of observations exist that they likely dominate the prior. Without a prior, i.e. with a non-informative prior, the optimization is on just the error distribution:

$$\arg\max_{\boldsymbol{\theta},\bar{\rho},\bar{\sigma}} \ \log\left[q\left(\hat{\boldsymbol{g}} - \boldsymbol{f}[\boldsymbol{\theta},\bar{\rho},\bar{\sigma}]\right)\right] \tag{4.27}$$

The solution yields a current information covariance model within the parametrized form.

## 4.11 Adapting to Current Market Conditions by Linear and Quadratic Programming

Using a particular set of forecast quantities and assuming Laplace or Gaussian prior and error distributions, an adapting parameterization can be solved by linear or quadratic programming.

### 4.11.1 Work in a Representation with Uncorrelated Factors

Assume the covariance model has orthogonal factors:

$$\begin{aligned} \boldsymbol{C} &= \boldsymbol{E}\ \mathrm{diag}(\boldsymbol{\eta}^2)\ \boldsymbol{E}^{\mathrm{T}} + \mathrm{diag}(\boldsymbol{\varepsilon}^2) \\ \text{where } \boldsymbol{\eta}^2 &= (m \times 1) \text{ variance of the factors} \\ \boldsymbol{\omega}^2 &= (n \times 1) \text{ variance of stock returns} \\ &= \boldsymbol{E}^2\boldsymbol{\eta}^2 + \boldsymbol{\varepsilon}^2 \end{aligned}$$

When that isn't the case, make it so by projecting the factors onto an orthogonal basis. One way is multiplying exposures by the square root of the covariance matrix:

$$\begin{aligned} \boldsymbol{F} &= \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}} \text{ eigendecomposed} \\ \boldsymbol{F}^{\frac{1}{2}} &= \boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{Q}^{\mathrm{T}} \\ \boldsymbol{M} &= \boldsymbol{F}^{\frac{1}{2}} \\ \boldsymbol{M}^{\mathrm{T}}\boldsymbol{M} &= \boldsymbol{F} \\ \boldsymbol{E}\boldsymbol{F}\boldsymbol{E}^{\mathrm{T}} &= (\boldsymbol{E}\boldsymbol{M}^{\mathrm{T}})(\boldsymbol{E}\boldsymbol{M}^{\mathrm{T}})^{\mathrm{T}} \\ \boldsymbol{E}_{new} &\leftarrow \boldsymbol{E}\boldsymbol{M}^{\mathrm{T}} \\ \boldsymbol{F}_{new} &\leftarrow \boldsymbol{I} \end{aligned} \tag{4.28}$$

Another way is described in section 4.12.

### 4.11.2 Linear Properties

Forecasting properties that can be written as linear combinations of the parameters under optimization yields an exploitable mathematical structure. Instead of requiring

a general purpose nonlinear optimization algorithm, with appropriate prior and error distributions, the optimization is solved by linear or quadratic programming.

Individual security, portfolio, and expected cross-sectional variance are linear functions of the model variances, $\boldsymbol{\eta}^2$ and $\boldsymbol{\varepsilon}^2$.

1. Portfolio Variance

$$\begin{aligned}\sigma_v^2 &= \boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v} \\ &= \underbrace{\left(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{E}\right)^2}_{\boldsymbol{a}^{\mathrm{T}}}\boldsymbol{\eta}^2 + \underbrace{\left(\boldsymbol{v}^2\right)^{\mathrm{T}}}_{\boldsymbol{b}^{\mathrm{T}}}\boldsymbol{\varepsilon}^2 \\ &= \boldsymbol{a}^{\mathrm{T}}\boldsymbol{\eta}^2 + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\varepsilon}^2\end{aligned}$$

2. Portfolio-weighted Expected Cross-Sectional Variance

$$\begin{aligned}\sigma_{v,cs}^2 &= \boldsymbol{v}^{\mathrm{T}}\boldsymbol{\omega}^2 - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v} \\ &= \boldsymbol{v}^{\mathrm{T}}\left[\boldsymbol{E}^2\boldsymbol{\eta}^2 + \boldsymbol{\varepsilon}^2\right] - \left[\left(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{E}\right)^2\boldsymbol{\eta}^2 + \left(\boldsymbol{v}^2\right)^{\mathrm{T}}\boldsymbol{\varepsilon}^2\right] \\ &= \underbrace{\left[\boldsymbol{v}^{\mathrm{T}}\boldsymbol{E}^2 - \left(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{E}\right)^2\right]}_{\boldsymbol{a}^{\mathrm{T}}}\boldsymbol{\eta}^2 + \underbrace{\left[\boldsymbol{v} - \boldsymbol{v}^2\right]^{\mathrm{T}}}_{\boldsymbol{b}^{\mathrm{T}}}\boldsymbol{\varepsilon}^2 \\ &= \boldsymbol{a}^{\mathrm{T}}\boldsymbol{\eta}^2 + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\varepsilon}^2\end{aligned}$$

On these types of forecasts, the connection between forecasts and model predictions is $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{\eta}^2 + \boldsymbol{B}\boldsymbol{\varepsilon}^2$.

### 4.11.3 Parameterization

Let $\boldsymbol{\theta} \geq 0$ scale factor variances and $\gamma \geq 0$ stock-specific variances.

$$\begin{aligned}\boldsymbol{\eta}^2(\boldsymbol{\theta}) &= \operatorname{diag}\left(\boldsymbol{\eta}^2\right)\boldsymbol{\theta} \\ \boldsymbol{\varepsilon}^2(\gamma) &= \gamma\boldsymbol{\varepsilon}^2\end{aligned}$$

$$\begin{aligned}\text{Then } \boldsymbol{g} &= \boldsymbol{f}(\boldsymbol{\theta}, \gamma) \\ &= \underbrace{\boldsymbol{A}\operatorname{diag}\left(\boldsymbol{\eta}^2\right)}_{\tilde{\boldsymbol{A}}}\boldsymbol{\theta} + \underbrace{\boldsymbol{B}\boldsymbol{\varepsilon}^2}_{\tilde{\boldsymbol{b}}}\gamma \\ &= \tilde{\boldsymbol{A}}\boldsymbol{\theta} + \tilde{\boldsymbol{b}}\gamma \text{ which is linear in } \boldsymbol{\theta} \text{ and } \gamma\end{aligned}$$

Note: if the model begin with correlated factors, adjustments would transfer back via

$$\boldsymbol{F}(\boldsymbol{\theta}) = \boldsymbol{M}^{\mathrm{T}} \operatorname{diag}\left[\boldsymbol{\eta}^2(\boldsymbol{\theta})\right] \boldsymbol{M} \tag{4.29}$$

### 4.11.4 Observation Noise and Prior

Suppose both the noise and $\boldsymbol{\theta}$'s prior are distributed independent Laplace[9], with $\boldsymbol{\theta}$ centered at $\theta_0$. $\theta_0$ and $\gamma$ have non-informative priors.

Noise:

$$q(\hat{\boldsymbol{g}} - \boldsymbol{f}[\boldsymbol{\theta}, \gamma]) = c \prod_j e^{-\sqrt{2}|\hat{g}_j - f_j(\boldsymbol{\theta},\gamma)|/\sigma_j} \tag{4.30}$$

$$\begin{aligned} \log[q(\hat{\boldsymbol{g}} - \boldsymbol{f}[\boldsymbol{\theta}, \gamma])] &= c_2 - \sqrt{2} \sum_j |\hat{g}_j - f_j(\boldsymbol{\theta}, \gamma)|/\sigma_j \\ &= c_2 - \left(\sqrt{2}/\boldsymbol{\sigma}\right)^{\mathrm{T}} |\hat{\boldsymbol{g}} - \boldsymbol{f}(\boldsymbol{\theta}, \gamma)|^{10} \\ &= c_2 - \left(\sqrt{2}/\boldsymbol{\sigma}\right)^{\mathrm{T}} |\hat{\boldsymbol{g}} - \tilde{\boldsymbol{A}}\boldsymbol{\theta} - \tilde{\boldsymbol{b}}\gamma| \end{aligned} \tag{4.31}$$

Prior:

$$\pi(\boldsymbol{\theta}) = c \prod_k e^{-\sqrt{2}|\theta_k - \theta_0|/\lambda_k} \tag{4.32}$$

$$\log\left[\pi(\boldsymbol{\theta})\right] = c_2 - \left(\sqrt{2}/\boldsymbol{\lambda}\right)^{\mathrm{T}} |\hat{\boldsymbol{\theta}} - \mathbf{1}\theta_0| \tag{4.33}$$

### 4.11.5 MAP Estimate of Parameters

Under these distributions and forecasts, the most likely configuration is

$$\arg\max_{\boldsymbol{\theta}, \theta_0, \gamma} \ \log[q(\cdot)] + \log[\pi(\cdot)] \tag{4.34}$$

$$= \arg\min_{\boldsymbol{\theta}, \theta_0, \gamma} \ (1/\boldsymbol{\sigma})^{\mathrm{T}} |\hat{\boldsymbol{g}} - \tilde{\boldsymbol{A}}\boldsymbol{\theta} - \tilde{\boldsymbol{b}}\gamma| + (1/\boldsymbol{\lambda})^{\mathrm{T}} |\hat{\boldsymbol{\theta}} - \mathbf{1}\theta_0|$$

$$\text{where } \boldsymbol{\theta}, \theta_0, \gamma \geq 0 \tag{4.35}$$

[9] but restricted so variance numbers are positive: $\boldsymbol{\theta} \geq 0$, and a variance forecast's error $\leq$ the forecast. These will be imposed as constraints and not properly considered in the distributions.

[10] $|\cdot|$ here and throughout means absolute value taken component by component.

which is solved by linear programming[11].

Following the same steps with Gaussian errors and priors leads to an optimization problem solved by quadratic programming.

### 4.11.6 Restrictions for Observable Factors

Factors, before orthogonalization, are often observable and have sources of information about the future—e.g., oil price and the option implied volatility of oil—that should be represented in the fitted model.

$$\begin{aligned}
\boldsymbol{F}(\boldsymbol{\theta})_{i,i} &= \text{adjusted variance of factor } i \\
&= \left(\boldsymbol{M}^{\mathrm{T}} \operatorname{diag}\left[\boldsymbol{\eta}^2(\boldsymbol{\theta})\right] \boldsymbol{M}\right)_{i,i} \\
&= \sum_k M_{i,k}^2 \, \eta^2(\boldsymbol{\theta})_k \\
&= \sum_k \underbrace{M_{i,k}^2 \, \eta_k^2}_{X_{i,k}} \theta_k \\
&= \boldsymbol{X}_i \boldsymbol{\theta}
\end{aligned} \tag{4.36}$$

Mapping to a linear combination of parameters, a factor's prediction can be incorporated like any forecast or as an optimization constraint.

### 4.11.7 Casting Prior onto Orthogonal Factors

Prior beliefs are best made on factors pre-orthogonalization. Beyond the first few, factors from eigendecomposition are hard to map to identifiable phenomena and change over time.

To avoid using reusing symbols, let $\boldsymbol{\psi}$ = the variance multipliers on the original factors, and $\boldsymbol{\theta}$ the same on the orthogonal factors.

$$\begin{aligned}
\psi_i &= \text{scaled/base variance of original factor } i \\
&= \frac{\boldsymbol{X}_i \boldsymbol{\theta}}{\boldsymbol{X}_i \mathbf{1}} \\
&= \underbrace{\frac{\boldsymbol{X}_i}{\boldsymbol{X}_i \mathbf{1}}}_{\boldsymbol{u}} \boldsymbol{\theta}
\end{aligned}$$

[11] Apply standard trick: min $|x|$ becomes min $\delta$ where $\delta \geq x, \delta \geq -x$

The map is linear, $\boldsymbol{\psi}(\boldsymbol{\theta}) = \boldsymbol{U}\boldsymbol{\theta}$. Change of variables gives the prior on the orthogonal factors.

$$\begin{aligned} \pi(\boldsymbol{\psi}) &= \text{prior on original variance scalars } \boldsymbol{\psi} \\ \text{prior on } \boldsymbol{\theta} &= \pi(\boldsymbol{U}\boldsymbol{\theta})\,|\boldsymbol{U}| \\ &\propto \pi(\boldsymbol{U}\boldsymbol{\theta}) \end{aligned} \tag{4.37}$$

### 4.11.8 Reducing Parameters to Prevent Overfitting

More data relative to the number of parameters provides stability, makes the optimizing configuration less sensitive to individual data points.

Suppose $m' < m$ of the $m$ factor variances were individually scaled and the remaining $(m - m')$ scaled as one. What factors should be among the $m'$? Ideally, the set yields the highest posterior likelihood under its maximizer. However, finding that set is a combinatorial problem where each function evaluation requires solving a linear program.

Consider a heuristic. If forecasts are on portfolio or cross-sectional variance (not on the difference in portfolio variances), entries in $\tilde{\boldsymbol{A}}$ and $\tilde{\boldsymbol{b}}$ are non-negative, and rows sum to the forecasted quantities under the base model. The $i$'th column of $\tilde{\boldsymbol{A}}$'s holds the $i$'th factor's contributions. $\tilde{\boldsymbol{b}}$ holds the contributions from stock-specific variance. The heuristic is to sum the columns of $\tilde{\boldsymbol{A}}$ and take the highest $m'$. To place more weight on accurate forecasts, weight rows by the inverse of the forecast's error standard deviation.

In the opposite direction, for more granularity, stock-specific variance scalars can vary by group, for example, by a company's economic sector.

## 4.12 Orthogonal Directions that Best Explain Securities

Orthogonalizing via eigendecomposing factor covariance, as in equation (4.28), doesn't consider connection to securities. The most important first orthogonal factor could be a direction of high factor variance but zero explanatory power.

What is truly sought is the entirely different goal of directions that best explain securities (as specified by a set of securities and weights).

This can be tailored to one's particular universe. For example, investment style or sector focus might concentrate a portfolio's risk along certain directions. Viewing under a tailored basis could better reveal risks and improve the incorporation of current information.

The orthogonal decomposition that best explains the $\boldsymbol{w}$-weighted variance of securities is $\boldsymbol{F} = \boldsymbol{N}\boldsymbol{N}^{\mathrm{T}}$ where

$$\begin{aligned}
\boldsymbol{w} &= (n \times 1) \text{ weight of securities in the fit} \\
\boldsymbol{E}_{i,\cdot} &= (1 \times m) \; i\text{'th row of } \boldsymbol{E} \\
\boldsymbol{H} &= (m \times m) \; \boldsymbol{w}\text{-weighted second moments of } \boldsymbol{E} \\
&= \sum_i w_i \, \boldsymbol{E}_{i,\cdot}^{\mathrm{T}} \, \boldsymbol{E}_{i,\cdot} \\
\boldsymbol{Q}, \boldsymbol{\Lambda} &= (m \times m) \text{ from eigendecomposition} \\
&\quad \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{F} \\
\tilde{\boldsymbol{Q}} &= (m \times m) \text{ from eigendecomposition} \\
&\quad \tilde{\boldsymbol{Q}}\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{Q}}^{\mathrm{T}} = \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}} \\
\boldsymbol{N} &= (m \times m) \; \boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\tilde{\boldsymbol{Q}}
\end{aligned}$$

*Proof.*

$$\begin{aligned}
\text{Let } \boldsymbol{d} &= (k \times 1) \text{ a combination (direction) of factors} \\
\boldsymbol{\beta} &= (n \times 1) \text{ stock exposures to } \boldsymbol{d} \\
&= \operatorname{cov}\left[\text{stock returns}, \boldsymbol{d} \text{ returns}\right] / \operatorname{var}\left[\boldsymbol{d} \text{ returns}\right] \\
&= \frac{\boldsymbol{E}\boldsymbol{F}\boldsymbol{d}}{\boldsymbol{d}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{d}}
\end{aligned}$$

The variance of stock $i$ explained by $\boldsymbol{d}$ is $\beta_i^2 \operatorname{var}\left[\boldsymbol{d} \text{ returns}\right] = \frac{(\boldsymbol{E}_{i,\cdot}\boldsymbol{F}\boldsymbol{d})^2}{\boldsymbol{d}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{d}}$. The goal is to

find $\boldsymbol{d}$ that maximizes the weighted explained variance.

$$\begin{aligned}
\text{wt explained var} &= \sum_i w_i \frac{(\boldsymbol{E}_{i,\cdot}\boldsymbol{F}\boldsymbol{d})^2}{\boldsymbol{d}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{d}} \\
&= \sum_i w_i \frac{\left(\boldsymbol{d}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{E}_{i,\cdot}^{\mathrm{T}}\right)\left(\boldsymbol{E}_{i,\cdot}\boldsymbol{F}\boldsymbol{d}\right)}{\boldsymbol{d}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{d}} \\
&= \frac{\boldsymbol{d}^{\mathrm{T}}\boldsymbol{F}\left(\sum_i w_i \boldsymbol{E}_{i,\cdot}^{\mathrm{T}}\boldsymbol{E}_{i,\cdot}\right)\boldsymbol{F}\boldsymbol{d}}{\boldsymbol{d}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{d}} \\
&= \frac{\boldsymbol{d}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{H}\boldsymbol{F}\boldsymbol{d}}{\boldsymbol{d}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{d}} \qquad (4.38)
\end{aligned}$$

$$\text{Let } \boldsymbol{v} = \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{d} \qquad (4.39)$$

$$\text{Then } \boldsymbol{d} = \boldsymbol{Q}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{v} \qquad (4.40)$$

$$\begin{aligned}
\text{and } \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_2 &= \left(\boldsymbol{d}_1^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)\left(\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{d}_2\right) \\
&= \boldsymbol{d}_1^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{d}_2 \\
&= \boldsymbol{d}_1^{\mathrm{T}}\boldsymbol{F}\boldsymbol{d}_2 \qquad (4.41) \\
&= \operatorname{cov}\left[\boldsymbol{d}_1, \boldsymbol{d}_2\right] \qquad (4.42)
\end{aligned}$$

$$\begin{aligned}
\text{wt explained var} &= \frac{\boldsymbol{d}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{H}\boldsymbol{F}\boldsymbol{d}}{\boldsymbol{d}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{d}} \\
&= \frac{(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{\mathrm{T}})\boldsymbol{F}\boldsymbol{H}\boldsymbol{F}(\boldsymbol{Q}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{v})}{(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{\mathrm{T}})\boldsymbol{F}(\boldsymbol{Q}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{v})} \\
&= \frac{\boldsymbol{v}^{\mathrm{T}}\left(\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)\boldsymbol{v}}{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}} \qquad (4.43)
\end{aligned}$$

Eigendecomposing the center matrix, $\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}$, into $\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{Q}}^{\mathrm{T}}$ gives $\boldsymbol{v}$ directions that maximize weighted explained variance. They are the columns of $\tilde{\boldsymbol{Q}}$.

Written in terms of $\boldsymbol{d}$, via (4.40), they are the columns of $\boldsymbol{D} \equiv \boldsymbol{Q}\boldsymbol{\Lambda}^{-\frac{1}{2}}\tilde{\boldsymbol{Q}}$.

Because vectors from the eigendecomposition are orthonormal and $\operatorname{cov}\left[\boldsymbol{d}_1, \boldsymbol{d}_2\right] = \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_2$, $\operatorname{cov}\left[\boldsymbol{D}\right] = \boldsymbol{I}$.

$$\begin{aligned}
\text{Let } \tilde{\boldsymbol{E}} &= (n \times m) \text{ exposures to factors in the new basis} \\
&= \boldsymbol{EFD} \qquad (4.44) \\
\boldsymbol{R} &= (m \times m) \text{ exposure of existing factors to it} \\
&= \boldsymbol{IFD} \\
&= \boldsymbol{FD} \qquad (4.45)
\end{aligned}$$

The covariance decomposition is

$$\begin{aligned}
\boldsymbol{F} &= \operatorname{cov}[\boldsymbol{R}] \\
&= (\boldsymbol{FD}) \operatorname{cov}[\boldsymbol{D}] (\boldsymbol{FD})^{\mathrm{T}} \\
&= \left(\boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\tilde{\boldsymbol{Q}}\right) \boldsymbol{I} \left(\boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\tilde{\boldsymbol{Q}}\right)^{\mathrm{T}} \qquad (4.46)
\end{aligned}$$

□

Note: One could also find the solution via the first $m$ entries of the eigendecomposition of the $n \times n$ matrix $\boldsymbol{EFE}^{\mathrm{T}}$ (typically $n \gg m$) with computational complexity $\mathcal{O}(n^3)$ instead of $\mathcal{O}(m^3 + nm^2)$.

## 4.13 Changing the Time Scale

Return frequency of the base and adapted model need not agree. The base model might look at monthly returns with a horizon of 6 months or a year while current information typically applies to short horizons (2 days, 1 week) and return frequencies (1 day, 1 week). Adapting addresses the horizon, but time scale differences remain.

Though both sets of numbers can be reported in annualized terms[12], negative serial correlation in daily returns inflates variance over short intervals relative to long. Not accounting for the serial correlation underpredicts risk. The variance multiplier to correct this is approximately

$$\frac{1 - \rho^2 - 2\rho\left(1 - \rho^{T_1}\right)/T_1}{1 - \rho^2 - 2\rho\left(1 - \rho^{T_2}\right)/T_2}(1 + \mu)^{2(T_1 - T_2)} \qquad (4.47)$$

[12] annualized variance = periods in yr × 1 period variance

$$\begin{aligned}
\text{where } T_1 &= \text{short horizon, e.g., 5 days} \\
T_2 &= \text{long horizon, e.g., 60 days} \\
\rho &= \text{daily serial correlation in returns} \\
\mu &= \text{mean daily return}
\end{aligned}$$

*Proof.*

$$\begin{aligned}
r_t &= \text{return on day t} \\
r_{0\to T} &= \text{cumulative return from time 0 to time } T \\
&= f(r_1, \ldots, r_T) = \prod_{t=1}^{T} (1 + r_t) - 1
\end{aligned}$$

Linearized around its mean,

$$\begin{aligned}
&\approx f|_{r_1,\ldots,r_T=\mu} + \nabla f|_{r_1,\ldots,r_T=\mu} (\boldsymbol{r} - \boldsymbol{\mu}) \\
&= c + (1+\mu)^{T-1} \sum_{t=1}^{T} r_t - \mu
\end{aligned}$$

Taking the variance of both sides,

$$\text{var}\left[r_{0\to T}\right] \approx (1+\mu)^{2T-2} \sum_{1\le j,k\le T} \text{cov}\left[r_j, r_k\right]$$

Suppose 1 day returns follow an AR(1)

$$\begin{aligned}
r_{t+1} &= c + \rho r_t + \varepsilon_{t+1} \\
\text{var}\left[r_t\right] &= \sigma^2
\end{aligned}$$

Then

$$\begin{aligned}
\operatorname{cov}\left[r_j, r_k\right] &= \sigma^2 \rho^{|j-k|} \\
\operatorname{var}\left[r_{0\to T}\right] &\approx (1+\mu)^{2T-2} \sum_{1\le j,k\le T} \operatorname{cov}\left[r_j, r_k\right] \\
&= (1+\mu)^{2T-2} \sum_{1\le j,k\le T} \sigma^2 \rho^{|j-k|} \\
&= (1+\mu)^{2T-2} \sigma^2 T \left[\frac{1-\rho^2-2\rho\left(1-\rho^T\right)/T}{(1-\rho)^2}\right]
\end{aligned}$$

$\frac{\operatorname{var}\left[r_{0\to T_1}\right]/T_1}{\operatorname{var}\left[r_{0\to T_2}\right]/T_2}$ = the ratio of annualized variances

$$\approx \frac{1-\rho^2-2\rho\left(1-\rho^{T_1}\right)/T_1}{1-\rho^2-2\rho\left(1-\rho^{T_2}\right)/T_2}(1+\mu)^{2(T_1-T_2)} \tag{4.48}$$

□

## 4.14 Expected Cross-Sectional Volatility

Portfolio weighted cross-sectional expected squared difference from portfolio return equals portfolio weighted average variance minus the variance of the portfolio plus (negligible) squared difference in means.

$$\begin{aligned}
\sigma^2_{v,cs} &= \text{expected cross-sectional variance of portfolio } \boldsymbol{v} \\
&\approx \boldsymbol{v}^{\mathrm{T}}\boldsymbol{\omega}^2 - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v}
\end{aligned} \tag{4.49}$$

*Proof.*

$$
\begin{aligned}
\boldsymbol{r} &= (n \times 1) \text{ security returns} \\
\boldsymbol{\mu} &= \mathrm{E}\left[\boldsymbol{r}\right] \\
\sum_i v_i &= 1 \\
\sigma^2_{v,cs} &= \mathrm{E}\left[\sum_i v_i \left(r_i - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{r}\right)^2\right] \\
&= \mathrm{E}\left[\sum_i v_i \big( \left[r_i - \mu_i\right] - \boldsymbol{v}^{\mathrm{T}}\left[\boldsymbol{r} - \boldsymbol{\mu}\right] + \left[\mu_i - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{\mu}\right] \big)^2\right] \\
&= \sum_i v_i \,\mathrm{E}\left[\left(r_i - \mu_i\right)^2\right] + \sum_i v_i \,\mathrm{E}\left[\left(\boldsymbol{v}^{\mathrm{T}}\left[\boldsymbol{r} - \boldsymbol{\mu}\right]\right)^2\right] \\
&\quad + \sum_i v_i \,\left(\mu_i - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{\mu}\right)^2 \\
&\quad - 2\mathrm{E}\left[\boldsymbol{v}^{\mathrm{T}}\left(\boldsymbol{r} - \boldsymbol{\mu}\right) \underbrace{\sum_i v_i \left(r_i - \mu_i\right)}_{\boldsymbol{v}^{\mathrm{T}}\left(\boldsymbol{r} - \boldsymbol{\mu}\right)}\right] \\
&\quad + 2\sum_i v_i \underbrace{\mathrm{E}\left[r_i - \mu_i\right]}_{0} \left(\mu_i - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{\mu}\right) \\
&\quad - 2\sum_i v_i \underbrace{\mathrm{E}\left[\boldsymbol{v}^{\mathrm{T}}\left(\boldsymbol{r} - \boldsymbol{\mu}\right)\right]}_{0} \left(\mu_i - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{\mu}\right) \\
&= \underbrace{\sum_i v_i \,\mathrm{E}\left[\left(r_i - \mu_i\right)^2\right]}_{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{\omega}^2} - \underbrace{\mathrm{E}\left[\left(\boldsymbol{v}^{\mathrm{T}}\left[\boldsymbol{r} - \boldsymbol{\mu}\right]\right)^2\right]}_{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v}} \\
&\quad + \underbrace{\sum_i v_i \,\left(\mu_i - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{\mu}\right)^2}_{negligible}{}^{13} \\
&\approx \boldsymbol{v}^{\mathrm{T}}\boldsymbol{\omega}^2 - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v}
\end{aligned}
$$

[13] Both variance and mean return scale with time. Because the mean return term is squared, it vanishes over small time. For example, say the annual variance is $(25\%)^2$ and the annual return is 10%. Over one trading day, the variance is $(25\%)^2/250 = 2.5\%^2$, and the squared mean return is $(10\%/250)^2 = 0.0016\%^2$.

□

## 4.15 Average Correlation and Variance within a Portfolio

$$\begin{aligned}\bar{\rho}_v &= \text{average correlation within portfolio } \boldsymbol{v} \\ &= \frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v} - \left(\boldsymbol{v}^2\right)^{\mathrm{T}}\boldsymbol{\omega}^2}{\left(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{\omega}\right)^2 - \left(\boldsymbol{v}^2\right)^{\mathrm{T}}\boldsymbol{\omega}^2}\end{aligned} \tag{4.50}$$

$$\begin{aligned}\bar{\sigma}_v^2 &= \text{average variance under average correlation} \\ &= \frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v}}{\bar{\rho}_v - (1-\bar{\rho}_v)\,\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}}\end{aligned} \tag{4.51}$$

*Proof.*

$\bar{\rho}$:

$$
\begin{aligned}
\boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v} &= var_v \\
&= \sum_i v_i^2\, var_i + \sum_{j\neq i} v_i v_j\, cov_{i,j} \\
&= \sum_i v_i^2\omega_i^2 + \sum_{j\neq i} v_i v_j \omega_i\omega_j \rho_{i,j} \\
&= \sum_i v_i^2\omega_i^2 + \sum_{j\neq i} v_i v_j \omega_i\omega_j \bar{\rho} \\
&= \sum_i v_i^2\omega_i^2 + \sum_{j\neq i} v_i v_j \omega_i\omega_j \bar{\rho} \\
&\quad + \sum_{j=i} v_i v_j \omega_i\omega_j \bar{\rho} - \sum_{j=i} v_i v_j \omega_i\omega_j \bar{\rho} \\
&= \sum_i v_i^2\omega_i^2 + \bar{\rho}\left(\sum_i v_i\omega_i\right)\left(\sum_j v_j\omega_j\right) \\
&\quad - \bar{\rho}\sum_i v_i^2\omega_i^2 \\
&= (\boldsymbol{v}^2)^{\mathrm{T}}\boldsymbol{\omega}^2 + \bar{\rho}(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{\omega})^2 - \bar{\rho}(\boldsymbol{v}^2)^{\mathrm{T}}\boldsymbol{\omega}^2
\end{aligned}
$$

$$
\begin{aligned}
\bar{\rho} &= \frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v} - (\boldsymbol{v}^2)^{\mathrm{T}}\boldsymbol{\omega}^2}{(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{\omega})^2 - (\boldsymbol{v}^2)^{\mathrm{T}}\boldsymbol{\omega}^2} \\
&= \frac{\text{var} - \text{uncorrelated var}}{\text{perfectly correlated var} - \text{uncorrelated var}}
\end{aligned}
$$

$\bar{\sigma}$:

$$\begin{aligned}
\boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v} &= (\boldsymbol{v}^2)^{\mathrm{T}}\boldsymbol{\omega}^2 + \bar{\rho}(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{\omega})^2 - \bar{\rho}(\boldsymbol{v}^2)^{\mathrm{T}}\boldsymbol{\omega}^2 \\
&= (\boldsymbol{v}^2)^{\mathrm{T}}\left[\bar{\sigma}\mathbf{1}\right]^2 + \bar{\rho}(\boldsymbol{v}^{\mathrm{T}}\left[\bar{\sigma}\mathbf{1}\right])^2 - \bar{\rho}(\boldsymbol{v}^2)^{\mathrm{T}}\left[\bar{\sigma}\mathbf{1}\right]^2 \\
&= \bar{\sigma}^2\left[\underbrace{(\boldsymbol{v}^2)^{\mathrm{T}}\mathbf{1}}_{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}} + \bar{\rho}(\underbrace{\boldsymbol{v}^{\mathrm{T}}\mathbf{1}}_{1})^2 - \bar{\rho}(\boldsymbol{v}^2)^{\mathrm{T}}\mathbf{1}\right] \\
&= \bar{\sigma}^2\left[\bar{\rho} + (1-\bar{\rho})\,\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}\right]
\end{aligned}$$

$$\bar{\sigma}^2 = \frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v}}{\bar{\rho} + (1-\bar{\rho})\,\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}}$$

□

# CHAPTER 5

# UNCERTAIN COVARIANCE AND PORTFOLIO OPTIMIZATION UNDER UNCERTAINTY

The work in this chapter grew out of the problem of estimation error in portfolio optimization. It can also be applied elsewhere in quantitative finance, e.g., chapters 6 and 7, and beyond finance.

Because optimized selection of frequentist estimates finds outliers in observation noise, especially as dimension grows, forecasts need to represent a predictive (according to one's beliefs) distribution where each number with its associated uncertainty is believed and articulated.

Standard Bayesian techniques to infer covariance of time series are unwieldy, perhaps ill-equipped to deal with stock returns' high dimension and non-stationarity, and hard to visualize. Meanwhile, modeling covariance through factors,

$$\Sigma_t \equiv \mathrm{cov}\left[Y_t|\beta,\Omega,\Lambda\right] = \beta\,\Omega\,\beta^{\mathrm{T}} + \Lambda,$$

is ubiquitous in the investment world, and versions with parameters occupying a range rather than point-estimated appear in the robust optimization literature[1].

To users of factor models, the covariance matrix is of interest, but so are the values of each parameter; indeed, a slew of diagnostic reports are generated from them. For example, the portfolio averaged exposures, $w^{\mathrm{T}}\beta$, reveal the bets, intentional and unintentional, taken by the manager. Incorporating uncertainty while retaining the factor model structure yields information not available through alternative Bayesian methods in addition to reducing computation.

Parameters change from point-estimated to having a mean and covariance, making

[1] e.g., Goldfarb and Iyengar [2003].

$\Sigma_t$ random. Formulae are derived for $\mathrm{E}[\Sigma_t]$ and $\mathrm{cov}[\Sigma_t]$ under assumptions on independence and distributional shape. Note that, because of convexity, expected portfolio variance exceeds the point-estimate assembled from parameter means[2], $w^{\mathrm{T}}\,\mathrm{E}[\Sigma_t]\,w \geq w^{\mathrm{T}}\hat{\Sigma}_t w$.

The second part of the research uses the result to improve portfolio construction. Given Markowitz utility as in equation (2.6),

$$U_\lambda(w) = w^{\mathrm{T}}\mu - \frac{\lambda}{2}w^{\mathrm{T}}\Sigma w$$

$$\text{So,}\quad \mathrm{var}[U(w)] = \mathrm{var}\left[w^{\mathrm{T}}\mu\right] + \frac{\lambda^2}{4}\mathrm{var}\left[w^{\mathrm{T}}\Sigma w\right] - \lambda\,\mathrm{cov}\left[w^{\mathrm{T}}\mu, w^{\mathrm{T}}\Sigma w\right] \tag{5.1}$$

$$\approx \mathrm{var}\left[w^{\mathrm{T}}\mu\right] + \frac{\lambda^2}{4}\mathrm{var}\left[w^{\mathrm{T}}\Sigma w\right] \tag{5.2}$$

The first term comes from the forecaster, the second from the uncertain covariance model. Maximizing expected utility with a penalty on uncertainty,

$$w^* \equiv \arg\max_w\ \mathrm{E}[U(w)] - \kappa \times \mathrm{stdev}[U(w)]\,, \tag{5.3}$$

is effectively robust optimization with the uncertainty set defined on utility and is a way to mitigate the estimation error problems encountered in portfolio optimization.

## 5.1 Literature and mathematical model

Section 2.2.1 'Frequentist problems' in the literature review of chapter 2 outlines the motivation for this work. The alternative construction methods covered in section 2.2.3 describe creating a predictive distribution of covariance (under 'Bayesian posterior') and adding uncertainty to the parameters of a factor model (under 'Robust optimization'). This work bridges the two.

For efficiency in the repetitive computations of optimization, the paper that is the basis of this chapter, Shah [2015], uses a diagonalized factor model similar to that of equation (4.6),

$$\mathrm{cov}[Y_t|\beta,\Theta,\gamma,\Lambda] \equiv \beta\ \mathrm{diag}[\Theta]\ \beta^{\mathrm{T}} + \gamma\,\Lambda, \tag{5.4}$$

[2] $\hat{\Sigma}_t = \mathrm{E}[\beta]\,\mathrm{E}[\Omega]\,\mathrm{E}[\beta]^{\mathrm{T}} + \mathrm{E}[\Lambda]$

with different variable names,

$$\begin{aligned}\Sigma_t &\equiv \operatorname{cov}\left[Y_t | Z_t\right] \\ &\equiv E_t \operatorname{diag}\left(\theta_{t,1:K}\right) E_t^{\mathrm{T}} + \theta_{t,K+1} \operatorname{diag}(\varepsilon_t^2) \qquad (5.5) \\ \text{where } Z_t &\equiv \{E_t, \theta_t, \varepsilon_t^2\} \qquad (5.6)\end{aligned}$$

In the mathematical framework of chapter 3, we are interested in the mean and variance of $\Sigma_t | X_t$ where the side information $X_t = \{\text{mean and variance of } Z_t\}$ comes from a forecaster. $X_t$ encapsulates all the information of $Y_{1:t-1}, X_{1:t-1}, Z_{1:t-1}$, i.e., forecasts can't be improved. The act of forecasting is a separate earlier procedure whose methods are left to the forecaster and free to evolve.

Assumptions on distributional shape (Gaussian $E$) and independence ($E, \theta, \varepsilon^2$ jointly independent, the elements of $\varepsilon^2$ independent) add the information to make computation possible.

**from Shah [2015] Uncertain Covariance Models and Uncertainty-Penalized Portfolio Optimization**

## 5.2 Abstract

Covariance appears throughout investment management, e.g., risk reporting, portfolio optimization, risk parity, smart beta, algorithmic trading, and hedging. It is often represented via multi-factor model. The imposed structure—comovement through sensitivity to common factors, diagonal residuals—offers advantages such as less data for estimation, easy proxying, and removing transient behavior and low eigenvalue directions. Nevertheless, parameter values are inferred and imperfect. Though it's common to ignore the error and proceed from point estimates, the spread of possible values is considered in Bayesian techniques and robust optimization. The research in this paper connects to both: Forecasts of mean and variance of the factor model parameters propagate to forecast the mean and variance of the covariance matrix. Using the result, optimization is performed to maximize portfolio utility a given number of standard deviations below the mean.

## 5.3 Introduction

Covariance between security returns, whose behavior is non-stationary and dimension high relative to the length of history, is often represented via factor model. Connections between securities happen through *exposure* to common factors, e.g., market,

industry, principal components. Movement uncorrelated with factors, called *idiosyncratic*, is assumed uncorrelated across securities. The standard form:

$$\begin{aligned} \boldsymbol{\Sigma} &= (N \times N) \text{ factor-modeled covariance of stock returns} \\ &= \boldsymbol{E}\boldsymbol{F}\boldsymbol{E}^{\mathrm{T}} + \text{diag}\left[\boldsymbol{\varepsilon}^2\right] \qquad (5.7) \\ \text{where } N &= \#\text{ of securities} \\ K &= \#\text{ of factors} \\ \boldsymbol{E} &= (N \times K) \text{ exposure to factors} \\ \boldsymbol{F} &= (K \times K) \text{ covariance between factors} \\ \boldsymbol{\varepsilon}^2 &= (N \times 1) \text{ idiosyncratic variance} \end{aligned}$$

Parameter values, often treated as known[3], are estimates. Examples where they aren't point-estimated but occupy a range exist in the robust optimization literature, e.g., Goldfarb and Iyengar [2003] and Lu [2011]. In this paper, rather than falling in a range, parameters are regarded as following a distribution whose mean and covariance are given by the forecaster. With assumptions on independence and distributional shape, the information implies the mean and covariance of $\boldsymbol{\Sigma}$.

The second part of paper enlists the result to compute the variance of portfolio utility and maximize utility a number of standard deviations below the mean, akin to robust optimization with the uncertainty set a function of portfolio weights and defined on portfolio utility.

### 5.3.1 Background

Within the large literature on portfolio optimization written since Markowitz [1952] introduced the notion of an efficient frontier, this paper connects to work on estimation error and modeling covariance uncertainty.

Topics around estimation error fall into 3 categories: 1–frequentist problems, 2–improving point estimates by Bayesian and other techniques, 3–optimizing beyond point estimates. For the first, early especially insightful papers are Jobson and Korkie [1980] and Broadie [1993]. For the second, on converting frequentist to Bayes estimates are Jorion [1986], Frost and Savarino [1986], and Black and Litterman

[3] e.g., commercial risk models from MSCI Barra, Axioma, and Northfield

[1992]. Robust statistics, as in Huber [1981], is used to infer point estimated inputs in Cavadini et al. [2001]. In its essence Bayesian, Ledoit and Wolf [2003] collapses a full covariance matrix toward a single factor model projection. Forms of shrinkage also appear in Fan et al. [2013], Abadir et al. [2014], Lam [2016], Ledoit and Wolf [2017], Goldberg et al. [2018], De Nard et al. [2019], and Engle et al. [2019]. Michaud [1998], which is critiqued by Scherer [2002], introduces 'resampling'—repeatedly generating a fixed length sample of a normal distribution under the estimated parameters and optimizing using parameters inferred from the sample, then averaging the output. Particulars aside, it is akin to optimizing with a softened covariance matrix. For the third category, Goldfarb and Iyengar [2003] explores portfolio construction by robust optimization (no relation to robust statistics), associating a range with each point and finding the point with the best worst case. Its sophisticated max of min optimization restricts application to particular structures of utility functions and constraints. A more recent survey is Bertsimas et al. [2011]. Lu [2011] looks at joint uncertainty in return and variance. DeMiguel and Nogales [2009] and Yang et al. [2015] maximize portfolio utility as measured by robust statistics.

On covariance uncertainty, standard Bayesian time-series techniques, though unwieldy and requiring simulation, can treat covariance beyond point estimates. Models also appear in the robust optimization literature. Halldórsson and Tütüncü [2003] defines the range as all positive semi-definite matrices that fall within bounds on each entry. Goldfarb and Iyengar [2003] operates under the factor structure of equation (5.7) and bounds the separate pieces.

## 5.4 Probability model

As a foundation for the randomness implied by forecasts, imagine embedding equation (5.7) in a probability space $(\Omega, \mathcal{F}, P)$ and that covariance conditionally follows a factor

model.

$$\boldsymbol{\Sigma} = \operatorname{cov}\left[\boldsymbol{r}|\boldsymbol{E},\boldsymbol{F},\boldsymbol{\varepsilon}^2\right] \tag{5.8}$$

$$= \boldsymbol{E}\boldsymbol{F}\boldsymbol{E}^{\mathrm{T}} + \operatorname{diag}\left[\boldsymbol{\varepsilon}^2\right] \tag{5.9}$$

$$\begin{aligned}
\text{where } \boldsymbol{r} &= \text{security returns} \\
\boldsymbol{r}_F &= \text{factor returns} \\
\boldsymbol{r}_\varepsilon &= \text{idiosyncratic returns} \\
\boldsymbol{r} &= \boldsymbol{E}\,\boldsymbol{r}_F + \boldsymbol{r}_\varepsilon
\end{aligned}$$

Note that the unconditional covariance is expected factor-modeled covariance plus (relatively small) variance in mean.

$$\begin{aligned}
\operatorname{cov}[\boldsymbol{r}] &= \mathrm{E}\left[\operatorname{cov}\left[\boldsymbol{r}|\boldsymbol{E},\boldsymbol{F},\boldsymbol{\varepsilon}^2\right]\right] + \operatorname{cov}\left[\mathrm{E}\left[\boldsymbol{r}|\boldsymbol{E},\boldsymbol{F},\boldsymbol{\varepsilon}^2\right]\right] \\
&= \mathrm{E}[\boldsymbol{\Sigma}] + \operatorname{cov}\left[\mathrm{E}\left[\boldsymbol{E}\,\boldsymbol{r}_F + \boldsymbol{r}_\varepsilon|\boldsymbol{E},\boldsymbol{F},\boldsymbol{\varepsilon}^2\right]\right]
\end{aligned} \tag{5.10}$$

Returning to equation (5.9), writing the dependence on $\omega$ highlights that these are random variables:

$$\text{for } \omega \in \Omega,\ \boldsymbol{\Sigma}(\omega) = \boldsymbol{E}(\omega)\boldsymbol{F}(\omega)\boldsymbol{E}^{\mathrm{T}}(\omega) + \operatorname{diag}\left[\boldsymbol{\varepsilon}^2(\omega)\right]. \tag{5.11}$$

*The probability measure $P$ represents the forecaster's beliefs.* For example, a user of point-estimates effectively puts all the mass on

$$\{w : \boldsymbol{E}(\omega) = \hat{\boldsymbol{E}}, \boldsymbol{F}(\omega) = \hat{\boldsymbol{F}}, \boldsymbol{\varepsilon}^2(\omega) = \hat{\boldsymbol{\varepsilon}}^2\},$$

in which case,

$$\mathrm{E}_P[\boldsymbol{\Sigma}] = \hat{\boldsymbol{E}}\hat{\boldsymbol{F}}\hat{\boldsymbol{E}}^{\mathrm{T}} + \operatorname{diag}\left[\hat{\boldsymbol{\varepsilon}}^2\right], \text{ and } \operatorname{cov}_P[\boldsymbol{\Sigma}] = \mathbf{0}.$$

When parameter values aren't fixed, $P$ isn't a point mass. The first half of this paper is interested in the implications on $\boldsymbol{\Sigma}$, in particular, $\mathrm{E}_P[\boldsymbol{\Sigma}]$ and $\operatorname{cov}_P[\boldsymbol{\Sigma}]$.

Though notation will be dropped for brevity, this is the conceptual structure for $\boldsymbol{\Sigma}$, and expectation and variance are henceforth with respect to $P$.

## 5.5 Uncertain covariance model

Let covariance follow the factor model of equation (5.7) with a parameter $\gamma$ added to connect the level of idiosyncratic variance to $\boldsymbol{F}$.

$$\boldsymbol{\Sigma} = \text{cov}\left[\boldsymbol{r}|\boldsymbol{E}, \boldsymbol{F}, \boldsymbol{\varepsilon}^2, \gamma\right] \tag{5.12}$$

$$= \boldsymbol{E}\boldsymbol{F}\boldsymbol{E}^{\mathrm{T}} + \gamma\,\text{diag}\left[\boldsymbol{\varepsilon}^2\right] \tag{5.13}$$

The goal is to compute $\text{E}\left[\boldsymbol{\Sigma}\right]$ and $\text{cov}\left[\boldsymbol{\Sigma}\right]$ from forecasts of means and covariances of $\boldsymbol{E}, \boldsymbol{\varepsilon}^2$ and $(\boldsymbol{F}, \gamma)$. This is possible through three assumptions:

1. $\boldsymbol{E}$, $\boldsymbol{\varepsilon}^2$, and $(\boldsymbol{F}, \gamma)$ are jointly independent.

2. $\boldsymbol{E}$ is Gaussian.

3. The elements of $\boldsymbol{\varepsilon}^2$ are independent.

The first, perhaps contentious, is consistent with the separation of uncertainty sets in the robust optimization literature[4]. The second need hold only at the portfolio level for portfolio variance computations. The third could be relaxed if correlations between idiosyncratic variances were provided.

Formulas for $\text{E}\left[\boldsymbol{\Sigma}\right]$ and $\text{cov}\left[\boldsymbol{\Sigma}\right]$ are derived in section 5.10. From equation (5.56),

$$\text{E}\left[\boldsymbol{\Sigma}\right] = \hat{\boldsymbol{E}}\hat{\boldsymbol{F}}\hat{\boldsymbol{E}}^{\mathrm{T}} + \hat{\gamma}\ \text{diag}\left[\hat{\boldsymbol{\varepsilon}}^2\right] + \sum_{k,l} \hat{F}_{k,l} \underbrace{\text{cov}\left[E_{\cdot,k}, E_{\cdot,l}\right]}_{N\times N}, \tag{5.14}$$

which equals the typical point-estimated covariance plus a term. The term raises the eigenvalues by adding $K(K+1)/2$ symmetric matrices whose sum is positive semi-definite. For instance, with $\boldsymbol{E}$ uncorrelated across securities, each of the matrices is diagonal.

## 5.6 Portfolio variance with orthogonal factors

Portfolio variance is the covariance quantity of interest in portfolio optimization, and less computation is incurred under the special case of orthogonal factors with random variances. Formulas are derived below for use in optimization.

[4] See Goldfarb and Iyengar [2003].

The structured form divides the factor covariance matrix $\boldsymbol{F}$ of equation (5.13) into $\boldsymbol{M}\,\mathrm{diag}\,[\theta_1,\ldots,\theta_K]\,\boldsymbol{M}^{\mathrm{T}}$ where $\boldsymbol{M}$ is fixed and $\mathrm{E}\,[\boldsymbol{F}] = \boldsymbol{M}\boldsymbol{M}^{\mathrm{T}}$. $\boldsymbol{M}$ expresses the original factors as linear combinations of orthonormal factors and limits the deformations of $\boldsymbol{F}$. Possible choices for $\boldsymbol{M}$ are infinite, e.g., $\mathrm{E}\,[\boldsymbol{F}]^{\frac{1}{2}}$ $\times$ any rotation[5].

$$\boldsymbol{\Sigma} = \boldsymbol{E}\left(\overbrace{\boldsymbol{M}\,\mathrm{diag}\,[\theta_1,\ldots,\theta_K]\,\boldsymbol{M}^{\mathrm{T}}}^{\boldsymbol{F}}\right)\boldsymbol{E}^{\mathrm{T}} + \theta_{K+1}{}^{6}\,\mathrm{diag}\left[\boldsymbol{\varepsilon}^2\right] \tag{5.15}$$

where $\boldsymbol{M} = (K \times K)$

$\boldsymbol{\theta} = (K+1 \times 1)$

Estimates of the mean and variance of the parameters are supplied by the forecaster.

$$\begin{aligned}
\hat{\boldsymbol{E}} &= (N \times K) & \hat{\boldsymbol{E}} &= \mathrm{E}\,[\boldsymbol{E}] \\
\hat{\boldsymbol{X}} &= (N \times K \times N \times K) & \hat{X}_{m,i,n,j} &= \mathrm{cov}\,[E_{m,i}, E_{n,j}] \\
\hat{\boldsymbol{\varepsilon}}^2 &= (N \times 1) & \hat{\boldsymbol{\varepsilon}}^2 &= \mathrm{E}\left[\boldsymbol{\varepsilon}^2\right] \\
\hat{\boldsymbol{\omega}} &= (N \times 1) & \hat{\omega}_n &= \mathrm{var}\left[\varepsilon_n^2\right] \\
\hat{\boldsymbol{\theta}} &= (K+1 \times 1) & \hat{\boldsymbol{\theta}} &= \mathrm{E}\,[\boldsymbol{\theta}] \\
\hat{\boldsymbol{\Omega}} &= (K+1 \times K+1) & \hat{\Omega}_{i,j} &= \mathrm{cov}\,[\theta_i, \theta_j]
\end{aligned}$$

As a pre-processing step, $\boldsymbol{M}$ is subsumed into $\boldsymbol{E}$,

$$\boldsymbol{E} \leftarrow \boldsymbol{E}\boldsymbol{M} \tag{5.16}$$

$$\hat{\boldsymbol{E}} \leftarrow \hat{\boldsymbol{E}}\boldsymbol{M} \tag{5.17}$$

$$\hat{\boldsymbol{X}}_{n,\cdot,m,\cdot} \leftarrow \boldsymbol{M}^{\mathrm{T}}\,\hat{\boldsymbol{X}}_{n,\cdot,m,\cdot}\,\boldsymbol{M} \tag{5.18}$$

and equation (5.15) becomes

$$\boldsymbol{\Sigma} = \boldsymbol{E}\,\mathrm{diag}\,[\theta_1 \ldots \theta_K]\,\boldsymbol{E}^{\mathrm{T}} + \theta_{K+1}\,\mathrm{diag}\left[\boldsymbol{\varepsilon}^2\right] \tag{5.19}$$

[5] If $\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{I}, (\boldsymbol{M}\boldsymbol{Q})\,(\boldsymbol{M}\boldsymbol{Q})^{\mathrm{T}} = \boldsymbol{M}\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{M}^{\mathrm{T}} = \boldsymbol{M}\boldsymbol{M}^{\mathrm{T}}$.

[6] $\gamma$ of equation (5.13) has been renamed $\theta_{K+1}$ to simplify notation.

**Proposition 5.1.** *Mean and variance of portfolio variance under orthogonal factors*
*If*

1. $\boldsymbol{\Sigma} = \boldsymbol{E} \operatorname{diag}\left[\theta_1 \dots \theta_K\right] \boldsymbol{E}^T + \theta_{K+1} \operatorname{diag}\left[\boldsymbol{\varepsilon}^2\right]$
2. $\boldsymbol{E}, \boldsymbol{\varepsilon}^2$, *and* $\boldsymbol{\theta}$ *are jointly independent*
3. *The elements of* $\boldsymbol{\varepsilon}^2$ *are independent*
4. $\boldsymbol{E}$ *is Gaussian*
5. *Mean and covariance of* $\boldsymbol{E}, \boldsymbol{\varepsilon}^2$, *and* $\boldsymbol{\theta}$ *are* $\hat{\boldsymbol{E}}, \hat{\boldsymbol{X}}, \hat{\boldsymbol{\varepsilon}}^2, \hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Omega}}$ *as defined above*

$$
\begin{aligned}
\boldsymbol{w} &= (N \times 1) \text{ portfolio weights or holdings} \\
\sigma_{\boldsymbol{w}}^2 &= \text{variance of portfolio } \boldsymbol{w} \\
\mathrm{E}\left[\sigma_{\boldsymbol{w}}^2\right] &= \boldsymbol{a}^T \hat{\boldsymbol{\theta}} && (5.20) \\
\operatorname{var}\left[\sigma_{\boldsymbol{w}}^2\right] &= \boldsymbol{a}^T \hat{\boldsymbol{\Omega}}\, \boldsymbol{a} + \hat{\boldsymbol{\theta}}^T \boldsymbol{B}\, \hat{\boldsymbol{\theta}} + \sum_{i,j} \hat{\Omega}_{i,j} B_{i,j} && (5.21)
\end{aligned}
$$

$$
\text{where } a_i = \begin{cases} \hat{\mu}_i^2 + \hat{\Psi}_{i,i} & 1 \le i \le K \\ \sum_{n=1}^{N} w_n^2 \hat{\varepsilon}_n^2 & i = K+1 \end{cases} \tag{5.22}
$$

$$
B_{i,j} = \begin{cases} 2\ \hat{\Psi}_{i,j}^2 + 4\, \hat{\mu}_i\, \hat{\mu}_j\, \hat{\Psi}_{i,j} & 1 \le i, j \le K \\ \sum_{n=1}^{N} w_n^4\, \hat{\omega}_n & i = j = K+1 \\ 0 & \text{otherwise} \end{cases} \tag{5.23}
$$

$$
\begin{aligned}
\boldsymbol{e} &= \text{exposures of portfolio } \boldsymbol{w} \\
\hat{\boldsymbol{\mu}}\left(\boldsymbol{w}\right) &= \mathrm{E}\left[\boldsymbol{e}\right] \\
&= \hat{\boldsymbol{E}}^T \boldsymbol{w} && (5.24) \\
\hat{\boldsymbol{\Psi}}\left(\boldsymbol{w}\right) &= \operatorname{cov}\left[\boldsymbol{e}\right] \\
\hat{\boldsymbol{\Psi}}_{i,j}\left(\boldsymbol{w}\right) &= \operatorname{cov}\left[e_i, e_j\right] \\
&= \boldsymbol{w}^T \hat{\boldsymbol{X}}_{\cdot,i,\cdot,j} \boldsymbol{w} && (5.25)
\end{aligned}
$$

*Proof.*

$$\sigma_{\boldsymbol{w}}^2 = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{w} \tag{5.26}$$

$$= \boldsymbol{w}^{\mathrm{T}} \left( \boldsymbol{E} \operatorname{diag} \left[ \theta_1 \dots \theta_K \right] \boldsymbol{E}^{\mathrm{T}} + \theta_{K+1} \operatorname{diag} \left[ \boldsymbol{\varepsilon}^2 \right] \right) \boldsymbol{w} \tag{5.27}$$

$$= \sum_{i=1}^{K} \theta_i \left( \boldsymbol{w}^{\mathrm{T}} \boldsymbol{E} \right)_i^2 + \theta_{K+1} \sum_{n=1}^{N} w_n^2 \, \varepsilon_n^2 \tag{5.28}$$

This can be written as the inner product

$$\sigma_{\boldsymbol{w}}^2 = \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{\nu} \tag{5.29}$$

$$\text{where } \boldsymbol{e}(\boldsymbol{w}) = \text{exposures of portfolio } \boldsymbol{w}$$

$$= \boldsymbol{E}^{\mathrm{T}} \boldsymbol{w} \tag{5.30}$$

$$\text{and } \nu_i(\boldsymbol{w}) = \begin{cases} e_i^2 & 1 \le i \le K \\ \sum_{n=1}^{N} w_n^2 \, \varepsilon_n^2 & i = K+1 \end{cases} \tag{5.31}$$

$\boldsymbol{\theta}$ and $\boldsymbol{\nu}$ are independent because $\boldsymbol{\theta}$ is independent of $\boldsymbol{E}, \boldsymbol{\varepsilon}$.

Mean and variance of portfolio variance are the expected value and variance of equation (5.29).

$$\mathrm{E}\left[\sigma_{\boldsymbol{w}}^2\right] = \mathrm{E}\left[\boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{\nu}\right] \tag{5.32}$$

$$= \mathrm{E}\left[\boldsymbol{\theta}\right]^{\mathrm{T}} \mathrm{E}\left[\boldsymbol{\nu}\right] \quad \text{since } \boldsymbol{\theta} \text{ ind. of } \boldsymbol{\nu} \tag{5.33}$$

$$\operatorname{var}\left[\sigma_{\boldsymbol{w}}^2\right] = \operatorname{var}\left[\boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{\nu}\right] \tag{5.34}$$

$$= \mathrm{E}\left[\boldsymbol{\nu}\right]^{\mathrm{T}} \operatorname{cov}\left[\boldsymbol{\theta}\right] \mathrm{E}\left[\boldsymbol{\nu}\right] + \mathrm{E}\left[\boldsymbol{\theta}\right]^{\mathrm{T}} \operatorname{cov}\left[\boldsymbol{\nu}\right] \mathrm{E}\left[\boldsymbol{\theta}\right]$$

$$+ \sum_{i,j} \left(\operatorname{cov}\left[\boldsymbol{\theta}\right]\right)_{i,j} \left(\operatorname{cov}\left[\boldsymbol{\nu}\right]\right)_{i,j} \quad \text{by Lemma 5.1} \tag{5.35}$$

Values are computed via equations (5.33) and (5.35) from the mean and covariance of $\boldsymbol{\theta}$ and of $\boldsymbol{\nu}$. $\boldsymbol{\theta}$ is covered by the forecasts $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\Omega}}$. What remains is to work out $\mathrm{E}\left[\boldsymbol{\nu}\right]$ and $\operatorname{cov}\left[\boldsymbol{\nu}\right]$.

Step one is computing the mean and covariance of portfolio exposures.

$$\text{Let } \hat{\boldsymbol{\mu}}(\boldsymbol{w}) = \mathrm{E}[\boldsymbol{e}] \tag{5.36}$$

$$= \hat{\boldsymbol{E}}^{\mathrm{T}} \boldsymbol{w} \tag{5.37}$$

$$\text{Let } \hat{\boldsymbol{\Psi}}(\boldsymbol{w}) = \mathrm{cov}[\boldsymbol{e}] \tag{5.38}$$

$$\hat{\boldsymbol{\Psi}}_{i,j}(\boldsymbol{w}) = \mathrm{cov}[e_i, e_j]$$

$$= \boldsymbol{w}^{\mathrm{T}} \hat{\boldsymbol{X}}_{\cdot,i,\cdot,j} \boldsymbol{w} \tag{5.39}$$

The expectation of $\boldsymbol{\nu}$ defined in equation (5.31) is a function of $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Psi}}$, and the forecast $\hat{\boldsymbol{\varepsilon}}^2$:

$$\mathrm{E}[\nu_i] = \begin{cases} \mathrm{E}[e_i^2] & 1 \le i \le K \\ \mathrm{E}\left[\sum_{n=1}^{N} w_n^2 \varepsilon_n^2\right] & i = K+1 \end{cases} \tag{5.40}$$

$$= \begin{cases} \mathrm{E}^2[e_i] + \mathrm{var}[e_i] & 1 \le i \le K \\ \sum_{n=1}^{N} w_n^2 \, \mathrm{E}[\varepsilon_n^2] & i = K+1 \end{cases} \tag{5.41}$$

$$= \begin{cases} \hat{\mu}_i^2 + \Psi_{i,i} & 1 \le i \le K \\ \sum_{n=1}^{N} w_n^2 \, \hat{\varepsilon}_n^2 & i = K+1 \end{cases} \tag{5.42}$$

Note that equation (5.41) contains an 'uncertainty correction' ($+ \mathrm{var}[e_i]$) arising from nonlinearity, $\mathrm{E}[y^2] = \mathrm{E}^2[y] + \mathrm{var}[y]$. Exposure variance adds to a portfolio's expected variance, i.e., the common practice of squaring a point estimate of exposure underestimates risk when uncertainty in exposures doesn't diversify away, for example, in concentrated portfolios or with exposure error correlated across securities.

$\mathrm{cov}[\nu_i, \nu_j]$ separates into three cases:

1. $i \le K, j \le K$:

$$\mathrm{cov}[\nu_i, \nu_j] = \mathrm{cov}\left[e_i^2, e_j^2\right] \tag{5.43}$$

$$= 2\left(\mathrm{cov}[e_i, e_j]\right)^2 + 4\,\mathrm{E}[e_i]\,\mathrm{E}[e_j]\,\mathrm{cov}[e_i, e_j] \tag{5.44}$$

by Corollary 7.1 since $\boldsymbol{e}$ Gaussian

$$= 2\,\hat{\Psi}_{i,j}^2 + 4\,\hat{\mu}_i \hat{\mu}_j \hat{\Psi}_{i,j} \tag{5.45}$$

2. $i = j = K + 1$:

$$\operatorname{cov}[\nu_i, \nu_j] = \operatorname{var}\left[\sum_{n=1}^{N} w_n^2 \, \varepsilon_n^2\right] \tag{5.46}$$

$$= \sum_{n=1}^{N} w_n^4 \operatorname{var}\left[\varepsilon_n^2\right] \quad \text{since elements of } \boldsymbol{\varepsilon} \text{ ind.} \tag{5.47}$$

$$= \sum_{n=1}^{N} w_n^4 \, \hat{\omega}_n \tag{5.48}$$

3. $i \leq K, \; j = K + 1$:

$$\operatorname{cov}[\nu_i, \nu_j] = \operatorname{cov}\left[e_i^2, \sum_{n=1}^{N} w_n^2 \, \varepsilon_n^2\right] = 0 \quad \text{since } \boldsymbol{E} \text{ ind. of } \boldsymbol{\varepsilon} \tag{5.49}$$

□

## 5.7 Robust optimization on portfolio utility

Vanilla Markowitz utility is

$$U_\lambda(w) = w^{\mathrm{T}}\alpha - \frac{\lambda}{2} w^{\mathrm{T}}\Sigma w \tag{5.50}$$

$$\text{So, } \operatorname{var}[U(w)] = \operatorname{var}\left[w^{\mathrm{T}}\alpha\right] + \frac{\lambda^2}{4}\operatorname{var}\left[w^{\mathrm{T}}\Sigma w\right] - \lambda \operatorname{cov}\left[w^{\mathrm{T}}\alpha, w^{\mathrm{T}}\Sigma w\right] \tag{5.51}$$

The first term comes from the forecaster[7], the second from the uncertain covariance model; assuming the third negligible yields

$$\operatorname{var}[U(w)] \approx \operatorname{var}\left[w^{\mathrm{T}}\alpha\right] + \frac{\lambda^2}{4}\operatorname{var}\left[w^{\mathrm{T}}\Sigma w\right]. \tag{5.52}$$

Maximizing expected utility with a penalty on uncertainty,

$$w^* \equiv \arg\max_w \, \mathrm{E}[U(w)] - \kappa \times \operatorname{stdev}[U(w)], \tag{5.53}$$

is effectively robust optimization with the uncertainty set defined—not on separate parameters as in conventional robust optimization—but on utility. It extends $\kappa$ standard deviations below the mean and is a function of portfolio weights $w$. Cast in

[7] $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{w}$ where $\boldsymbol{\Lambda}$ represents commonality in bets.

robust optimization language:

$$w^* \equiv \arg\max_{w} \min_{u_w \in S_w} u_w$$
$$\text{where } S_w = \{\mathrm{E}\left[U(w)\right] - c \times \mathrm{stdev}\left[U(w)\right] \mid c \leq \kappa\} \tag{5.54}$$

The maximin optimization becomes a simple max since the inner minimum is known when the uncertainty set is defined on the quantity being maximized (utility).

This formulation's success mitigating estimation error problems[8]is tested in section 5.8.

### 5.7.1 Indifference region to reduce transaction costs

The literature on managing transaction costs, e.g., Michaud [1998], mentions an indifference region to prevent nuisance trading. Rebalancing can be skipped if the improvement is judged negligible.

Such a rule can be defined from mean and standard deviation of portfolio utility. The portfolios are indifferent if their utility ranges overlap, i.e., if the upper range of the existing exceeds the lower range of the changed,

$$\mathrm{E}\left[U_0\right] + c \times \mathrm{stdev}\left[U_0\right] \;>\; \mathrm{E}\left[U\right] - c \times \mathrm{stdev}\left[U\right]. \tag{5.55}$$

## 5.8 An example

Data are 106 periods of weekly Wed–Wed returns from Wed Dec 20, 2017 through Tue Dec 31, 2019 for the 927 securities with complete series in (IWB) iShares Russell 1000 ETF as of Oct 1, 2020. The first 80 periods, 12/20/2017–7/3/2019, are the training set. The final 26 periods, 7/3/2019–12/31/2019, are the testing set.

An orthogonal model of section 5.6 is constructed from the training set as follows:

1. 5 principal components are inferred from return covariance with securities equally weighted. This yields $\boldsymbol{M}$ (an identity matrix), $\hat{\boldsymbol{\theta}}$[9], and from OLS regression, a first estimate of exposures and idiosyncratic variances, $\hat{\boldsymbol{E}}_{ols}$ and $\hat{\boldsymbol{\varepsilon}}^2_{ols}$.

[8] Taleb [2012] writes that the problem with 'Portfolio theory, mean-variance, etc.' is 'Assuming knowledge of the parameters, not integrating models across parameters, relying on (very unstable) correlations'.

[9] $\hat{\theta}_{K+1}$, the mean scalar on idiosyncratic variance, is 1.

2. Using the universe median and variance of the OLS estimates as a prior—exposures Gaussian, idiosyncratic variance inverse gamma—$\boldsymbol{E}$ and $\boldsymbol{\varepsilon}^2$ are inferred through Bayesian regression, yielding $\hat{\boldsymbol{E}}$, $\hat{\boldsymbol{\varepsilon}}^2$, $\hat{\boldsymbol{X}}$, and $\hat{\boldsymbol{\omega}}$. ($\hat{\boldsymbol{X}}$ assumes no exposure covariance between securities.)

3. $\hat{\boldsymbol{\Omega}}$, the covariance of factor variances and idiosyncratic variance scalar, is built in two steps. The $K$ rows and columns corresponding to factors are the standard error.[10] For idiosyncratic variance, a measure of its level is the cross-sectional median of each period's squared OLS regression residuals, divided by the mean across time for normalization. Its uncertainty is the variance of this series. The level is assumed uncorrelated with the factor variances.

The experiment selects 50 stocks at random and constructs fully-invested, long-only minimum variance portfolios eight ways:

**Equally-weighted**

Two percent in each of the 50 securities.

**OLS** (Conventional)

Minimum variance portfolio under covariance point-estimated from OLS exposures and idiosyncratic variances.

**Bayes** (Conventional)

Minimum variance portfolio under covariance point-estimated from Bayes posterior mean exposures and idiosyncratic variances.

$\kappa = 0$ Expected variance

Minimum expected[11] variance portfolio but no aversion to uncertainty.

[10] For the usual unbiased variance estimate of Gaussian $x, y$ from $n$ points, $\operatorname{cov}\left[\hat{\sigma}_x^2, \hat{\sigma}_y^2\right] = 2\sigma_{x,y}^2/(n-1)$. Standard error replaces $\sigma_{x,y}$ with $\hat{\sigma}_{x,y}$.

[11] Adds the variance term in equation (5.41) to the Bayes point estimate.

$\kappa > 0$ Expected variance with uncertainty aversion

For each $\kappa$ in $\{0.1, 1, 10, 100\}$, a local minimizer of E [variance]+$\kappa\times$stdev [variance] reached by searching from the $\kappa = 0$ portfolio.

Realized volatility is measured on the testing set. The experiment is repeated 10,000 times.[12]

Table 5.1 shows realized volatility quantiles for each method. Equally-weighted was dominated by OLS, which was dominated by Bayes, which was dominated by the uncertainty methods, whose results improved uniformly with increasing $\kappa$.

Table 5.1: Realized portfolio std dev quantiles by method (annualized)

| | Quantile | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0.02 | 0.05 | 0.25 | 0.50 | 0.75 | 0.95 | 0.98 |
| Equal wt | 10.86 | 11.41 | 12.78 | 13.79 | 14.84 | 16.47 | 17.23 |
| OLS | 6.41 | 6.78 | 7.72 | 8.50 | 9.46 | 11.43 | 12.36 |
| Bayes | 6.29 | 6.64 | 7.51 | 8.21 | 9.07 | 10.81 | 11.61 |
| $\kappa = 0$ | 6.29 | 6.63 | 7.49 | 8.19 | 9.03 | 10.73 | 11.52 |
| $\kappa = 0.1$ | 6.28 | 6.62 | 7.48 | 8.18 | 9.01 | 10.70 | 11.48 |
| $\kappa = 1$ | 6.26 | 6.59 | 7.43 | 8.09 | 8.88 | 10.47 | 11.22 |
| $\kappa = 10$ | 6.21 | 6.51 | 7.24 | 7.82 | 8.51 | 9.78 | 10.51 |
| $\kappa = 100$ | 6.17 | 6.47 | 7.15 | 7.70 | 8.32 | 9.49 | 10.16 |

Table 5.2 compares methods head-to-head. Values are the fraction of experiments where the column method realized lower volatility than the row method. Uncertain methods beat each certain method in more than 80% of experiments. Increasing uncertainty consideration almost uniformly improved winning percentage through $\kappa = 10$. $\kappa = 100$ won 80% of experiments against $\kappa = 10$, but the fraction against other methods fell slightly.

Table 5.3 shows median expected variance[13], uncertainty, and realized variance. Uncertainty falls and expected variance rises with increasing $\kappa$, a mathematical truth from the optimization objectives. Repeating the information from table 5.1: *realized*

[12]Total run time for 10,000 $\times$ 8 portfolios is under 20 minutes using the SLSQP solver in scipy.optimize.minimize Virtanen et al. [2020] on a circa 2015 laptop with 4th Generation Intel i7 CPU (i7-4710HQ @ 2.5GHz).

[13]Expected variance and uncertainty are via the uncertain covariance model.

Table 5.2: Head-to-head comparison of realized volatility

| | Fraction of tests method had lower realized volatility | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | Bayes | $\kappa = 0$ | $\kappa = 0.1$ | $\kappa = 1$ | $\kappa = 10$ | $\kappa = 100$ |
| Equal wt | 0.997 | 0.999 | 0.999 | 0.999 | 1.000 | 1.000 | 1.000 |
| OLS | | 0.890 | 0.903 | 0.906 | 0.913 | 0.903 | 0.894 |
| Bayes | | | 0.808 | 0.824 | 0.842 | 0.850 | 0.847 |
| $\kappa = 0$ | | | | 0.842 | 0.848 | 0.852 | 0.846 |
| $\kappa = 0.1$ | | | | | 0.848 | 0.850 | 0.845 |
| $\kappa = 1$ | | | | | | 0.842 | 0.837 |
| $\kappa = 10$ | | | | | | | 0.798 |

variance fell. As with all the results, the cause could be simple regularization or, as tables 5.4 and 5.5 suggest, uncertainty modeling's skill.

Table 5.3: Median expected variance, uncertainty, and realized variance (weekly, not annualized)

| Median | Equal wt | OLS | Bayes | $\kappa = 0$ | $\kappa = 0.1$ | $\kappa = 1$ | $\kappa = 10$ | $\kappa = 100$ |
|---|---|---|---|---|---|---|---|---|
| Expected variance | 4.99 | 1.72 | 1.70 | 1.70 | 1.70 | 1.71 | 1.74 | 1.79 |
| Uncertainty, sd(var) | 0.77 | 0.25 | 0.24 | 0.24 | 0.24 | 0.24 | 0.22 | 0.22 |
| Realized variance | 3.66 | 1.39 | 1.30 | 1.29 | 1.29 | 1.26 | 1.18 | 1.14 |

Table 5.4 shows quantiles of the uncertain covariance model's prediction bias. Higher $\kappa$ overpredicted more but had a narrower error range and smaller errors of underprediction.

Table 5.5 shows quantiled z-scored predictions: (expected - realized variance) / (stdev of variance). The range between low and high quantile shrank slightly from $\sim 8$ with $\kappa = 0$ to $\sim 6$ with $\kappa = 100$. Note that the opposite would occur without an increase in accuracy since higher $\kappa$ corresponds to lower uncertainty in the denominator as well as higher expected variance. That the range didn't collapse or blow up suggests uncertainty predictions are on the correct scale and not naive regularization.

Table 5.4: Expected minus realized annualized standard deviation

| | Quantile | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0.02 | 0.05 | 0.25 | 0.50 | 0.75 | 0.95 | 0.98 |
| Equal wt | -0.68 | -0.03 | 1.41 | 2.30 | 3.20 | 4.41 | 4.85 |
| OLS | -3.21 | -2.36 | -0.10 | 0.99 | 1.90 | 3.05 | 3.56 |
| Bayes | -2.52 | -1.71 | 0.24 | 1.23 | 2.07 | 3.17 | 3.66 |
| $\kappa = 0$ | -2.42 | -1.61 | 0.29 | 1.25 | 2.09 | 3.17 | 3.65 |
| $\kappa = 0.1$ | -2.37 | -1.57 | 0.31 | 1.26 | 2.09 | 3.18 | 3.65 |
| $\kappa = 1$ | -2.06 | -1.28 | 0.45 | 1.35 | 2.15 | 3.22 | 3.68 |
| $\kappa = 10$ | -0.97 | -0.41 | 0.94 | 1.70 | 2.43 | 3.41 | 3.86 |
| $\kappa = 100$ | -0.46 | 0.09 | 1.25 | 1.97 | 2.66 | 3.59 | 4.01 |

Table 5.5: Z-scores: $\frac{\text{predicted - realized variance}}{\text{uncertainty (std dev) of variance}}$

| | Quantile | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0.02 | 0.05 | 0.25 | 0.50 | 0.75 | 0.95 | 0.98 |
| Equal wt | -0.56 | -0.03 | 1.09 | 1.72 | 2.31 | 3.04 | 3.29 |
| OLS | -5.12 | -3.61 | -0.15 | 1.36 | 2.42 | 3.53 | 3.96 |
| Bayes | -4.04 | -2.69 | 0.36 | 1.70 | 2.70 | 3.78 | 4.16 |
| $\kappa = 0$ | -3.91 | -2.54 | 0.43 | 1.74 | 2.74 | 3.82 | 4.20 |
| $\kappa = 0.1$ | -3.86 | -2.48 | 0.45 | 1.77 | 2.75 | 3.83 | 4.22 |
| $\kappa = 1$ | -3.38 | -2.08 | 0.67 | 1.90 | 2.88 | 3.94 | 4.31 |
| $\kappa = 10$ | -1.70 | -0.71 | 1.47 | 2.53 | 3.37 | 4.38 | 4.70 |
| $\kappa = 100$ | -0.80 | 0.14 | 2.00 | 2.95 | 3.74 | 4.67 | 4.99 |

## 5.9 Summary

The framework in this paper is a practical and mathematically tight way to model covariance with uncertainty. Uncertainty about individual pieces—for example, that the market beta of stock X is more reliably forecasted than stock Y's—propagates to the portfolio level, and a portfolio variance forecast, no longer a fixed number, has a mean and variance.

Models can be constructed for a broad set of securities and easily shared. The method requires no simulation and is computationally fast, and optimization can occur beyond agreeably structured cases. The method is also easily extended to incorporate uncertainty about future regimes (section 5.11).

Among the many applications, one can, for example, choose the most reliably predicted from possible hedging securities. Or articulating the issue of estimation error, a computer can penalize uncertainty during optimization to find portfolios believed more likely to perform as expected.

## 5.10 Uncertain factor-modeled $\mathrm{E}\left[\Sigma_{i,j}\right]$ and $\mathrm{cov}\left[\Sigma_{i,j}, \Sigma_{m,n}\right]$

**Proposition 5.2.** *Mean and covariance of entries of* $\boldsymbol{\Sigma}$

*If*

1. $\boldsymbol{\Sigma} = \boldsymbol{E}\boldsymbol{F}\boldsymbol{E}^T + \gamma \operatorname{diag}\left[\boldsymbol{\varepsilon}^2\right]$
2. $\boldsymbol{E}, \boldsymbol{\varepsilon}^2$, *and* $(\boldsymbol{F}, \gamma)$ *are jointly independent*
3. *The elements of* $\boldsymbol{\varepsilon}^2$ *are independent*
4. $\boldsymbol{E}$ *is Gaussian*
5. *Mean and covariance of* $\boldsymbol{E}, \boldsymbol{\varepsilon}^2$, *and* $(\boldsymbol{F}, \gamma)$ *are:*

*Exposures*

$$\hat{\boldsymbol{E}} = (N \times K) \qquad \hat{\boldsymbol{E}} = \mathrm{E}\left[\boldsymbol{E}\right]$$
$$\hat{\boldsymbol{X}} = (N \times K \times N \times K) \qquad \hat{X}_{m,i,n,j} = \operatorname{cov}\left[E_{m,i}, E_{n,j}\right]$$

*Idiosyncratic effects*

$$\hat{\boldsymbol{\varepsilon}}^2 = (N \times 1) \qquad \hat{\boldsymbol{\varepsilon}}^2 = \mathrm{E}\left[\boldsymbol{\varepsilon}^2\right]$$
$$\hat{\boldsymbol{\omega}} = (N \times 1) \qquad \hat{\omega}_n = \operatorname{var}\left[\varepsilon_n^2\right]$$

*Factor structure*

$$\hat{\boldsymbol{F}} = (K \times K) \qquad \hat{\boldsymbol{F}} = \mathrm{E}\left[\boldsymbol{F}\right]$$
$$\hat{\boldsymbol{H}} = (K \times K \times K \times K) \qquad \hat{H}_{i,j,k,l} = \operatorname{cov}\left[F_{i,j}, F_{k,l}\right]$$
$$\hat{\gamma} = (scalar) \qquad \hat{\gamma} = \mathrm{E}\left[\gamma\right]$$
$$\hat{\eta} = (scalar) \qquad \hat{\eta} = \operatorname{var}\left[\gamma\right]$$
$$\hat{\boldsymbol{h}} = (K \times K) \qquad \hat{h}_{i,j} = \operatorname{cov}\left[F_{i,j}, \gamma\right]$$

*Then mean and covariance of the covariance matrix entries are*

$$\mathrm{E}\left[\boldsymbol{\Sigma}\right] = \hat{\boldsymbol{E}}\hat{\boldsymbol{F}}\hat{\boldsymbol{E}}^{T} + \hat{\gamma}\ \mathrm{diag}\left[\hat{\boldsymbol{\varepsilon}}^2\right] + \sum_{k,l} \hat{F}_{k,l}\ \hat{\boldsymbol{X}}_{\cdot,k,\cdot,l} \tag{5.56}$$

$$\begin{aligned}
\mathrm{cov}\left[\Sigma_{i,j}, \Sigma_{m,n}\right] = \sum_{k,l,s,t} \Big[ & \hat{E}_{i,k}\hat{E}_{m,s}\hat{X}_{j,l,n,t} + \hat{E}_{j,l}\hat{E}_{n,t}\hat{X}_{i,k,m,s} \\
& + \hat{X}_{i,k,m,s}\hat{X}_{j,l,n,t} \\
& + \hat{E}_{i,k}\hat{E}_{n,t}\hat{X}_{j,l,m,s} + \hat{E}_{j,l}\hat{E}_{m,s}\hat{X}_{i,k,n,t} \\
& + \hat{X}_{i,k,n,t}\hat{X}_{j,l,m,s}\Big] \left(\hat{F}_{k,l}\hat{F}_{s,t} + \hat{H}_{k,l,s,t}\right) \\
+ \sum_{k,l,s,t} & \left[\hat{E}_{i,k}\hat{E}_{j,l} + \hat{X}_{i,k,j,l}\right] \left[\hat{E}_{m,s}\hat{E}_{n,t} + \hat{X}_{m,s,n,t}\right] \hat{H}_{k,l,s,t} \\
+ 1_{\{i=j\}}\, \hat{\varepsilon}_i^2 & \sum_{s,t} \left[\hat{E}_{m,s}\hat{E}_{n,t} + \hat{X}_{m,s,n,t}\right] \hat{h}_{s,t} \\
+ 1_{\{m=n\}}\, \hat{\varepsilon}_m^2 & \sum_{k,l} \left[\hat{E}_{i,k}\hat{E}_{j,l} + \hat{X}_{i,k,j,l}\right] \hat{h}_{k,l} \\
+ 1_{\{i=j\}} 1_{\{m=n\}}\, & \hat{\varepsilon}_i^2\, \hat{\varepsilon}_m^2\, \hat{\eta} \\
+ 1_{\{i=j=m=n\}}\, & \hat{\omega}_i \left(\hat{\gamma}^2 + \hat{\eta}\right)
\end{aligned} \tag{5.57}$$

*Proof.*

$$\Sigma_{i,j} = \sum_{k,l} E_{i,k}E_{j,l}F_{k,l}\ + 1_{\{i=j\}}\, \varepsilon_i^2\, \gamma \tag{5.58}$$

$$\mathrm{E}\left[\Sigma_{i,j}\right] = \sum_{k,l} \mathrm{E}\left[E_{i,k}E_{j,l}\right] \mathrm{E}\left[F_{k,l}\right]\ + 1_{\{i=j\}}\, \mathrm{E}\left[\varepsilon_i^2\right] \mathrm{E}\left[\gamma\right] \tag{5.59}$$

$$\begin{aligned}
&= \sum_{k,l} \big(\mathrm{E}\left[E_{i,k}\right] \mathrm{E}\left[E_{j,l}\right] + \mathrm{cov}\left[E_{i,k}, E_{j,l}\right]\big) \mathrm{E}\left[F_{k,l}\right] \\
&\qquad + 1_{\{i=j\}}\, \mathrm{E}\left[\varepsilon_i^2\right] \mathrm{E}\left[\gamma\right]
\end{aligned} \tag{5.60}$$

$$= \sum_{k,l} \left[\hat{E}_{i,k}\hat{E}_{j,l} + \hat{X}_{i,k,j,l}\right] \hat{F}_{k,l}\ + 1_{\{i=j\}}\, \hat{\varepsilon}_i^2\, \hat{\gamma} \tag{5.61}$$

$$\text{so } \mathrm{E}\left[\boldsymbol{\Sigma}\right] = \hat{\boldsymbol{E}}\hat{\boldsymbol{F}}\hat{\boldsymbol{E}}^{\mathrm{T}} + \hat{\gamma}\ \mathrm{diag}\left[\hat{\boldsymbol{\varepsilon}}^2\right] + \sum_{k,l} \hat{F}_{k,l}\ \underbrace{\hat{\boldsymbol{X}}_{\cdot,k}(\cdot,l)}_{N\times N} \tag{5.62}$$

$$\operatorname{cov}\left[\Sigma_{i,j},\Sigma_{m,n}\right]=\underbrace{\operatorname{cov}\left[\mathrm{E}\left[\Sigma_{i,j}|\boldsymbol{F},\gamma\right],\mathrm{E}\left[\Sigma_{m,n}|\boldsymbol{F},\gamma\right]\right]}_{①}+\underbrace{\mathrm{E}\left[\operatorname{cov}\left[\Sigma_{i,j},\Sigma_{m,n}|\boldsymbol{F},\gamma\right]\right]}_{②} \tag{5.63}$$

$$\begin{aligned}① = cov\Big[&\sum_{k,l}\mathrm{E}\left[E_{i,k}E_{j,l}\right]F_{k,l}+1_{\{i=j\}}\mathrm{E}\left[\varepsilon_i^2\right]\gamma,\\&\sum_{s,t}\mathrm{E}\left[E_{m,s}E_{n,t}\right]F_{s,t}+1_{\{m=n\}}\mathrm{E}\left[\varepsilon_m^2\right]\gamma\Big]\end{aligned} \tag{5.64}$$

$$\begin{aligned}&=\sum_{k,l,s,t}\mathrm{E}\left[E_{i,k}E_{j,l}\right]\mathrm{E}\left[E_{m,s}E_{n,t}\right]\operatorname{cov}\left[F_{k,l},F_{s,t}\right]\\&+1_{\{i=j\}}\mathrm{E}\left[\varepsilon_i^2\right]\sum_{s,t}\mathrm{E}\left[E_{m,s}E_{n,t}\right]\operatorname{cov}\left[F_{s,t},\gamma\right]\\&+1_{\{m=n\}}\mathrm{E}\left[\varepsilon_m^2\right]\sum_{k,l}\mathrm{E}\left[E_{i,k}E_{j,l}\right]\operatorname{cov}\left[F_{k,l},\gamma\right]\\&+1_{\{i=j\}}1_{\{m=n\}}\mathrm{E}\left[\varepsilon_i^2\right]\mathrm{E}\left[\varepsilon_m^2\right]\operatorname{var}\left[\gamma\right]\end{aligned} \tag{5.65}$$

$$\begin{aligned}&=\sum_{k,l,s,t}\left[\hat{E}_{i,k}\hat{E}_{j,l}+\hat{X}_{i,k,j,l}\right]\left[\hat{E}_{m,s}\hat{E}_{n,t}+\hat{X}_{m,s,n,t}\right]\hat{H}_{k,l,s,t}\\&+1_{\{i=j\}}\hat{\varepsilon}_i^2\sum_{s,t}\left[\hat{E}_{m,s}\hat{E}_{n,t}+\hat{X}_{m,s,n,t}\right]\hat{h}_{s,t}\\&+1_{\{m=n\}}\hat{\varepsilon}_m^2\sum_{k,l}\left[\hat{E}_{i,k}\hat{E}_{j,l}+\hat{X}_{i,k,j,l}\right]\hat{h}_{k,l}\\&+1_{\{i=j\}}1_{\{m=n\}}\hat{\varepsilon}_i^2\,\hat{\varepsilon}_m^2\,\hat{\eta}\end{aligned} \tag{5.66}$$

$$
\begin{aligned}
② &= \mathrm{E}\Big[cov\Big(\sum_{k,l} E_{i,k}E_{j,l}F_{k,l} \ + 1_{\{i=j\}}\,\varepsilon_i^2\,\gamma, \\
&\qquad \sum_{s,t} E_{m,s}E_{n,t}F_{s,t} \ + 1_{\{m=n\}}\,\varepsilon_m^2\,\gamma \,\Big|\, \boldsymbol{F}, \gamma\Big)\Big] && (5.67)\\
&= \mathrm{E}\Big[\sum_{k,l,s,t} \mathrm{cov}\left[E_{i,k}E_{j,l}, E_{m,s}E_{n,t}\right] F_{k,l}F_{s,t} \ + 1_{\{i=j=m=n\}}\mathrm{var}\left[\varepsilon_i^2\right]\gamma^2\Big] && (5.68)\\
&= \sum_{k,l,s,t} \mathrm{cov}\left[E_{i,k}E_{j,l}, E_{m,s}E_{n,t}\right] \mathrm{E}\left[F_{k,l}F_{s,t}\right] \\
&\qquad + 1_{\{i=j=m=n\}}\mathrm{var}\left[\varepsilon_i^2\right] \mathrm{E}\left[\gamma^2\right] && (5.69)\\
&= \sum_{k,l,s,t} \underbrace{\mathrm{cov}\left[E_{i,k}E_{j,l}, E_{m,s}E_{n,t}\right]}_{③} \left(\hat{F}_{k,l}\hat{F}_{s,t} + \hat{H}_{k,l,s,t}\right) \\
&\qquad + 1_{\{i=j=m=n\}}\,\hat{\omega}_i\left(\hat{\gamma}^2 + \hat{\eta}\right) && (5.70)
\end{aligned}
$$

$$
\begin{aligned}
③ &= \mathrm{cov}\left[E_{i,k}E_{j,l}, E_{m,s}E_{n,t}\right] \\
&= \hat{E}_{i,k}\hat{E}_{m,s}\hat{X}_{j,l,n,t} + \hat{E}_{j,l}\hat{E}_{n,t}\hat{X}_{i,k,m,s} + \hat{X}_{i,k,m,s}\hat{X}_{j,l,n,t} \\
&\quad + \hat{E}_{i,k}\hat{E}_{n,t}\hat{X}_{j,l,m,s} + \hat{E}_{j,l}\hat{E}_{m,s}\hat{X}_{i,k,n,t} + \hat{X}_{i,k,n,t}\hat{X}_{j,l,m,s} && (5.71)\\
&\quad \text{via Lemma 7.1 since } \boldsymbol{E} \text{ Gaussian}
\end{aligned}
$$

□

## 5.11 Incorporating structural uncertainty

Uncertainty so far has taken the flavor of estimation error, ambiguity in the value of parameters on an assumed-correct skeleton. A different, macro scale uncertainty exists in how events might unfold, the possibility of alternative regimes whose centers (configuration of factor variances) or rules (factor structures) differ by more than small perturbations so need to be modeled separately. Considering the possibilities is a view into a portfolio's robustness.

Suppose the future is a random draw from $S$ scenarios, each described by a separate uncertain covariance model. Unconditional mean and variance of portfolio variance are assembled from forecasts for each scenario.

Assume mean returns are the same in each scenario.

$$\begin{aligned}
\text{Let} \quad \pi_s &= \text{probability of scenario } s, \quad s = 1 \dots S \\
m_s(\boldsymbol{w}) &= \mathrm{E}\left[\sigma^2_{\boldsymbol{w}} | \text{ scenario } s\right] \\
v_s(\boldsymbol{w}) &= \mathrm{var}\left[\sigma^2_{\boldsymbol{w}} | \text{ scenario } s\right]
\end{aligned}$$

$$\begin{aligned}
\text{Then } \mathrm{E}\left[\sigma^2_{\boldsymbol{w}}\right] &= \mathrm{E}\left[\mathrm{E}\left[\sigma^2_{\boldsymbol{w}} | s\right]\right] \\
&= \sum_s \pi_s \, m_s && (5.72) \\
\text{and } \mathrm{var}\left[\sigma^2_{\boldsymbol{w}}\right] &= \mathrm{E}\left[\mathrm{var}\left[\sigma^2_{\boldsymbol{w}} | s\right]\right] + \mathrm{var}\left[\mathrm{E}\left[\sigma^2_{\boldsymbol{w}} | s\right]\right] && (5.73) \\
&= \mathrm{E}\left[\mathrm{var}\left[\sigma^2_{\boldsymbol{w}} | s\right]\right] + \mathrm{E}\left[\mathrm{E}^2\left[\sigma^2_{\boldsymbol{w}} | s\right]\right] - \mathrm{E}^2\left[\mathrm{E}\left[\sigma^2_{\boldsymbol{w}} | s\right]\right] && (5.74) \\
&= \sum_s \pi_s v_s + \sum_s \pi_s m_s^2 - \left[\sum_s \pi_s m_s\right]^2 && (5.75)
\end{aligned}$$

## 5.12 Mathematical results

**Lemma 5.1.** *Variance of inner product of independent variables*
*For* $\boldsymbol{a}, \boldsymbol{b}$ *independent,*

$$\begin{aligned}
\mathrm{var}\left[\boldsymbol{a}^T \boldsymbol{b}\right] &= \mathrm{E}\left[\boldsymbol{b}\right]^T \mathrm{cov}\left[\boldsymbol{a}\right] \mathrm{E}\left[\boldsymbol{b}\right] \\
&\quad + \mathrm{E}\left[\boldsymbol{a}\right]^T \mathrm{cov}\left[\boldsymbol{b}\right] \mathrm{E}\left[\boldsymbol{a}\right] \; + \sum_{i,j} \left(\mathrm{cov}\left[\boldsymbol{a}\right]\right)_{i,j} \left(\mathrm{cov}\left[\boldsymbol{b}\right]\right)_{i,j} && (5.76)
\end{aligned}$$

*Proof.*

$$\begin{aligned}
\operatorname{var}\left[\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}\right] &= \operatorname{var}\left[\mathrm{E}\left[\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}|\boldsymbol{a}\right]\right] + \mathrm{E}\left[\operatorname{var}\left[\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}|\boldsymbol{a}\right]\right] && (5.77)\\
&= \operatorname{var}\left[\boldsymbol{a}^{\mathrm{T}}\mathrm{E}\left[\boldsymbol{b}\right]\right] + \mathrm{E}\left[\boldsymbol{a}^{\mathrm{T}}\operatorname{cov}\left[\boldsymbol{b}\right]\boldsymbol{a}\right] \text{ since } \boldsymbol{a} \text{ ind. of } \boldsymbol{b} && (5.78)\\
&= \mathrm{E}\left[\boldsymbol{b}\right]^{\mathrm{T}}\operatorname{cov}\left[\boldsymbol{a}\right]\mathrm{E}\left[\boldsymbol{b}\right] + \mathrm{E}\left[\boldsymbol{a}^{\mathrm{T}}\operatorname{cov}\left[\boldsymbol{b}\right]\boldsymbol{a}\right] && (5.79)\\
&= \mathrm{E}\left[\boldsymbol{b}\right]^{\mathrm{T}}\operatorname{cov}\left[\boldsymbol{a}\right]\mathrm{E}\left[\boldsymbol{b}\right] + \mathrm{E}\left[\sum_{i,j} a_i a_j \left(\operatorname{cov}\left[\boldsymbol{b}\right]\right)_{i,j}\right] && (5.80)\\
&= \mathrm{E}\left[\boldsymbol{b}\right]^{\mathrm{T}}\operatorname{cov}\left[\boldsymbol{a}\right]\mathrm{E}\left[\boldsymbol{b}\right] + \sum_{i,j}\mathrm{E}\left[a_i a_j\right]\left(\operatorname{cov}\left[\boldsymbol{b}\right]\right)_{i,j} && (5.81)\\
&= \mathrm{E}\left[\boldsymbol{b}\right]^{\mathrm{T}}\operatorname{cov}\left[\boldsymbol{a}\right]\mathrm{E}\left[\boldsymbol{b}\right]\\
&\quad + \sum_{i,j}\left(\mathrm{E}\left[a_i\right]\mathrm{E}\left[a_j\right] + \operatorname{cov}\left[a_i, a_j\right]\right)\left(\operatorname{cov}\left[\boldsymbol{b}\right]\right)_{i,j} && (5.82)\\
&= \mathrm{E}\left[\boldsymbol{b}\right]^{\mathrm{T}}\operatorname{cov}\left[\boldsymbol{a}\right]\mathrm{E}\left[\boldsymbol{b}\right]\\
&\quad + \mathrm{E}\left[\boldsymbol{a}\right]^{\mathrm{T}}\operatorname{cov}\left[\boldsymbol{b}\right]\mathrm{E}\left[\boldsymbol{a}\right] \; + \sum_{i,j}\left(\operatorname{cov}\left[\boldsymbol{a}\right]\right)_{i,j}\left(\operatorname{cov}\left[\boldsymbol{b}\right]\right)_{i,j} && (5.83)
\end{aligned}$$

□

**Lemma 5.2.**

$$For \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

$$\begin{aligned} \mathrm{E}[x_1 x_2 x_3 x_4] &= \mu_1\mu_2\mu_3\mu_4 \\ &\quad + \mu_1\mu_2\sigma_{3,4} + \mu_3\mu_4\sigma_{1,2} + \sigma_{1,2}\sigma_{3,4} \\ &\quad + \mu_1\mu_3\sigma_{2,4} + \mu_2\mu_4\sigma_{1,3} + \sigma_{1,3}\sigma_{2,4} \\ &\quad + \mu_1\mu_4\sigma_{2,3} + \mu_2\mu_3\sigma_{1,4} + \sigma_{1,4}\sigma_{2,3} \end{aligned} \tag{5.84}$$

$$\begin{aligned} \mathrm{cov}[x_1 x_2, x_3 x_4] &= \mu_1\mu_3\sigma_{2,4} + \mu_2\mu_4\sigma_{1,3} + \sigma_{1,3}\sigma_{2,4} \\ &\quad + \mu_1\mu_4\sigma_{2,3} + \mu_2\mu_3\sigma_{1,4} + \sigma_{1,4}\sigma_{2,3} \end{aligned} \tag{5.85}$$

$$\begin{aligned} &= (\mu_1\mu_3 + \sigma_{1,3})(\mu_2\mu_4 + \sigma_{2,4}) \\ &\quad + (\mu_1\mu_4 + \sigma_{1,4})(\mu_2\mu_3 + \sigma_{2,3}) - 2\mu_1\mu_3\mu_2\mu_4 \end{aligned} \tag{5.86}$$

*Proof.*

$$\begin{aligned}
\mathrm{E}\left[x_1x_2x_3x_4\right] &= \mathrm{E}\left[\prod_{i=1}^{4}\left[\mu_i + (x_i - \mu_i)\right]\right] \\
&= \mu_1\mu_2\mu_3\mu_4 \\
&\quad + \mu_1\mu_2\sigma_{3,4} + \mu_3\mu_4\sigma_{1,2} \\
&\quad + \mu_1\mu_3\sigma_{2,4} + \mu_2\mu_4\sigma_{1,3} \\
&\quad + \mu_1\mu_4\sigma_{2,3} + \mu_2\mu_3\sigma_{1,4} \\
&\quad + \mathrm{E}\left[\prod_{i=1}^{4}(x_i - \mu_i)\right] \\
&\quad \text{since odd moments are 0}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\left[\prod_{i=1}^{4}(x_i - \mu_i)\right] &= \sigma_{1,2}\sigma_{3,4} + \sigma_{1,3}\sigma_{2,4} + \sigma_{1,4}\sigma_{2,3} \\
&\quad \text{by Isserlis' theorem since 0 mean Gaussians}
\end{aligned}$$

$$\mathrm{cov}\left[x_1x_2, x_3x_4\right] = \mathrm{E}\left[x_1x_2x_3x_4\right] - \underbrace{\mathrm{E}\left[x_1x_2\right]}_{(\mu_1\mu_2+\sigma_{12})}\underbrace{\mathrm{E}\left[x_3x_4\right]}_{(\mu_3\mu_4+\sigma_{34})}$$

□

**Corollary 5.1.** *Covariance of squared Gaussians*

*For* $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right), \mathrm{cov}\left[x_1^2, x_2^2\right] = 4\mu_1\mu_2\sigma_{1,2} + 2\sigma_{1,2}^2$

---

# CHAPTER 6

# UNCERTAIN RISK PARITY

Introduced in the alternative portfolio construction methods of section 2.2.3, risk parity determines portfolio weights using covariance alone. It is typically done at the index or asset class level rather than on individual stocks, e.g., one security is the S&P 500 index, another is a corporate bond index. Risk parity finds the unique solution where each security matches its assigned risk budget (repeating equation 2.10),

$$w^*\left(\Sigma, \sigma^2, f\right) \equiv w > 0 \text{ s.t. } w \odot \Sigma w = \sigma^2 f \tag{6.1}$$

$$\begin{aligned} \text{where } \sigma^2 &\equiv \text{ target variance of portfolio} \\ f &\equiv \text{ fraction of risk budgeted to each security, } f \in \mathbb{R}_+^n \\ \odot &\equiv \text{ element-by-element multiplication} \end{aligned}$$

which happens to be equivalent to minimizing variance with a regularization penalty against small weights (repeating equation 2.11),

$$w^*\left(\Sigma, \sigma^2, f\right) = \underset{w>0}{\arg\min} \ \frac{1}{2} w^{\mathrm{T}} \Sigma w - \sigma^2 \sum_i f_i \log w_i \tag{6.2}$$

Working from a point-estimate of covariance leaves the method vulnerable to estimation error.

## 6.1 Literature review

Risk parity without uncertainty is covered in section 2.2.3 under 'Risk parity'. Starting from the regularized minimum variance formulation of equation (6.2) with the penalty moved to constraints, Costa and Kwon [2019] minimizes the worst case variance where the covariance matrix comes from a Markov chain, each state having factor-modeled covariance with elliptical confidence sets on exposures and bounds on stock-specific variance. Costa and Kwon [2020a] minimizes the worst case error between variance risk contributions and budgeted risk when covariance falls within box

constraints. Costa and Kwon [2020b] adds a return forecast term having an elliptical uncertainty set. The authors' collected work on risk parity is in a dissertation [Costa, 2021]. Kapsos et al. [2018] considers robust optimization of risk parity under several uncertainty scenarios. Kim and Kim [2021] applies the resampling idea of Michaud [1998] with covariance assumed Wishart centered at the unbiased time series estimate and the final risk parity portfolio as the average of risk parity portfolios constructed from sampled matrices inversely weighted by sampled matrix condition number. The basis of this chapter, Shah [2021a], looks at the difference between variance contributions and budgeted risk as in Costa and Kwon [2020a], but models covariance as a distribution and minimizes functions of mean squared error.

## 6.2 Mathematical model

In the form of equation (6.1), the goal is a solution to $n$ equations that are quadratic in the decision variable $w$ and linear in the data $\Sigma$. Normally $\Sigma$ is replaced by a point-estimate, and there is a unique solution since $\Sigma$ is regarded as fixed. When $\Sigma$ is random, there isn't a solution.

The paper develops an approach given $\mathrm{E}\left[\Sigma\right]$ and $\mathrm{cov}\left[\Sigma\right]$. These could be supplied directly or, since risk parity is generally at the index level, derived from portfolios of securities modeled by an uncertain covariance model as in chapter 5.

In the latter case, as in equation (5.4), the covariance of individual stock returns is expressed in terms of state variables,

$$\mathrm{cov}\left[Y_t|\beta,\Theta,\gamma,\Lambda\right] \equiv \beta\ \mathrm{diag}\left[\Theta\right]\ \beta^{\mathrm{T}} + \gamma\,\Lambda$$

If each row of matrix $p$ represents the portfolio weights of a different index,

$$\begin{aligned}\Sigma_t &\equiv \mathrm{cov}\left[p\,Y_t|\beta,\Theta,\gamma,\Lambda\right] &(6.3)\\ &\equiv p\left(\beta\ \mathrm{diag}\left[\Theta\right]\ \beta^{\mathrm{T}} + \gamma\,\Lambda\right)\ p^{\mathrm{T}} &(6.4)\end{aligned}$$

$\mathrm{E}\left[\Sigma\right]$ and $\mathrm{cov}\left[\Sigma\right]$, after assumptions on independence and distributional shape, are functions of mean and covariance of the parameters.

**from Shah [2021a] Uncertain Risk Parity**

## 6.3 Abstract

Risk parity is a portfolio construction technique that scales sections of a portfolio—e.g., stocks, bonds, currencies, commodities—so that forecasted contributions to net portfolio risk match the budget. Because risks are measured from a point-estimate of covariance, the method is subject to problems of estimation error. This paper treats covariance as uncertain in order to find a risk parity weighting that doesn't count on perfectly optimized hedges and is robust to changes in regime.

Separately, of general interest are the uncertain risk contributions calculated en route. Reporting a portfolio's uncertain risk decomposition puts a band around numbers and reveals fragility. For example, market could seem hedged in a long-short portfolio but surface as the biggest risk when parameters are considered across their error range.

## 6.4 Introduction

Risk parity, named by Qian [2005], is a method of portfolio construction that, from risk alone, scales sections of a portfolio—e.g., stocks, levered bonds, currencies, commodities—so each section's forecasted contribution to net portfolio risk matches the (typically equally-weighted) risk budget. The idea without considering correlation dates to 1996 in Bridgewater All Weather Fund's weighting investments by inverse of volatility.

One motivation, as in Qian [2011], is capturing the risk premium across a diversified basket of asset classes. Another is correcting problems of the canonical example of risk-only portfolio construction, minimum variance, whose composition is sensitive to inputs and can be concentrated.

Absent bounds, minimum variance is the point where marginal variances for investments are equal. When risk parity equalizes risk contributions—proportional to

*marginal variance* × *weight*[1]—small positions grow, large ones shrink, and the portfolio is less concentrated and less sensitive to inputs. Yet, operating from a single estimated covariance matrix, it remains vulnerable to estimation error.

While this paper investigates risk parity under uncertainty, risk parity itself can be explicitly cast as an approach to minimum variance under uncertainty. Maillard et al. [2010] shows that risk parity is the solution to minimum variance with a regularization penalty. The penalty can be regarded as Bayesian tempering data with a prior. From this perspective, de Jong [2018] sees risk parity as an intermediate optimal minimum risk portfolio on the spectrum from minimum variance (no uncertainty in estimates, solution dominated by data) to equal-weighted (complete uncertainty, solution dominated by non-informative prior).

This paper contributes four ideas to the literature: 1–justification for estimation error concerns in measuring risk contributions, 2–a framework and algorithm to solve risk parity with uncertainty, 3–a formula to measure risk contributions (decompose risk) with uncertainty, and 4–a way to model covariance uncertainty between portfolios from the securities within.

### 6.4.1 Literature review

Maillard et al. [2010] and Bai et al. [2016] explore mathematical properties and alternative formulations of risk parity. Instead of equalizing risk contributions across investments, Lohre et al. [2014] equalizes them across principal component directions of the covariance matrix, and Roncalli and Weisang [2016] across directions specified by any factor model. Bernardi et al. [2019] looks at risk parity's 'second order risk', the ratio of realized to predicted risk, essentially the bias from estimate-optimized selection. Closest to this paper, Costa and Kwon [2019] uses factor-modeled covariance and solves risk parity under uncertainty and regime switching through robust (minimax) optimization with elliptical confidence sets on exposures and bounds on stock-specific variance. On risk parity but not related is the large literature on empirical performance in both return and volatility. This paper concerns construction, volatility but not return, and individual volatility contributions as much as the combined portfolio's.

[1] See equation (6.5)

### 6.4.2 Problem statement

Budgeted fractions of total portfolio risk are assigned to $N$ investments, each of which can itself be a portfolio. Using an estimate of covariance to measure risk, risk parity finds a weighting that meets the budget. The inputs and decision variables are:

$$\begin{aligned}
\boldsymbol{f} &= (N \times 1) \text{ budgeted fractions of portfolio risk} \\
&\quad \boldsymbol{f} > 0, \ \sum_i f_i = 1 \\
\boldsymbol{C} &= (N \times N) \text{ covariance between investments} \\
\boldsymbol{w} &= (N \times 1) \text{ weights, usually constrained} \geq \boldsymbol{0}
\end{aligned}$$

One other possible input is target portfolio variance, $\sigma^2$, to fix the scale: $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{w} = \sigma^2$. Otherwise, scale is set by forcing weights to sum to 1, $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{1} = 1$, which without uncertainty is the same as picking any target variance and dividing the solution by the sum of its weights.

Budgets can be enforced on contributions to portfolio standard deviation or variance, arriving at the same result.[2] The derivation is repeated here because both forms appear in what follows.

Contributions to standard deviation are calculated by Euler's theorem applied to portfolio standard deviation, which is homogenous of degree 1 in the weights[3].

$$\begin{aligned}
s_i &= \text{std dev contribution of i'th investment} \\
&= w_i \frac{\partial}{\partial w_i} (\text{portfolio std dev}) \\
&= w_i \frac{\partial}{\partial w_i} \sqrt{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{w}} \\
&= \frac{w_i}{2\sqrt{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{w}}} \frac{\partial}{\partial w_i} \boldsymbol{w}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{w} \qquad (6.5) \\
&= w_i \left(\boldsymbol{C}\boldsymbol{w}\right)_i / \sqrt{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{w}} \qquad (6.6)
\end{aligned}$$

So risk parity in standard deviation is

$$\sigma f_i = w_i \left(\boldsymbol{C}\boldsymbol{w}\right)_i / \sqrt{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{w}} \quad i = 1..N, \ \boldsymbol{w} > 0 \qquad (6.7)$$

$$\text{where } \sigma\boldsymbol{f} = (N \times 1) \text{ standard deviation budgets}$$

[2] See section 3.2.3 of Qian et al. [2007]

[3] See Maillard et al. [2010]

Contributions to variance are covariance with the portfolio.

$$\begin{aligned} v_i &= \text{variance contribution of i'th investment} \\ &= \begin{bmatrix} 0 \cdots 0 & w_i & 0 \cdots 0 \end{bmatrix} \boldsymbol{C}\boldsymbol{w} \\ &= w_i \left(\boldsymbol{C}\boldsymbol{w}\right)_i \end{aligned} \tag{6.8}$$

Risk parity in variance is

$$\begin{gathered} \sigma^2 f_i = w_i \left(\boldsymbol{C}\boldsymbol{w}\right)_i \quad i = 1..N,\ \boldsymbol{w} > 0 \\ \text{where } \sigma^2 \boldsymbol{f} = (N \times 1) \text{ variance budgets} \end{gathered} \tag{6.9}$$

Since $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{w} = \sigma^2$, equations (6.7) and (6.9) are identical.

## 6.5 Sensitivity to covariance forecast

Emerging entirely from a point-estimate of covariance taken as true, risk parity could be sensitive to estimation error. Here are two arguments the concern is valid.

### 6.5.1 As solution to minimum variance with regularization

Maillard et al. [2010] shows that risk parity is the solution to minimum variance with a regularization penalty against small positions.

$$\underset{\boldsymbol{w}>0}{\arg\min} \ \ \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{w} - \sigma^2 \sum_i f_i \log(w_i) \tag{6.10}$$

Setting the gradient to 0 yields $\left(\boldsymbol{C}\boldsymbol{w}\right)_i - \sigma^2 f_i / w_i = 0$ for each $i$, so $w_i \left(\boldsymbol{C}\boldsymbol{w}\right)_i = \sigma^2 f_i$, the risk parity equations (6.9).

Since it can be cast as optimization, risk parity might encounter issues described in the large literature on optimization's bias and estimation error problems.

The optimization characterization is also the easiest way to compute risk parity, minimizing a smooth, strictly convex (when restricted to any orthant, $log\,|w_i|$ in place of $log\,w_i$) function that has a unique solution.

### 6.5.2 Non-robustness in risk contribution assignment

Menchero and Davis [2011] reach the risk contribution assignment of (6.6) by measuring behavior in the direction of the net portfolio.

$$
\begin{aligned}
\sigma_i &= \text{standard dev of i'th investment} \\
\rho_{i,w} &= \text{correlation of i'th investment and portfolio} \\
\sigma_i\,\rho_{i,w} &= \text{std dev of i'th investment} \times \text{corr with portfolio} \\
&= \sigma_i\ \mathrm{cov}\,(i,w)\,/\,(\sigma_i\ \sigma) && (6.11) \\
&= w_i\,(\boldsymbol{C}\boldsymbol{w})_i\,/\sqrt{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{w}} && (6.12) \\
&= s_i
\end{aligned}
$$

What about behavior orthogonal to the portfolio?

$$
\begin{aligned}
s_i^{\perp} &= \text{length of orthogonal behavior of i'th investment} \\
&= \sigma_i\sqrt{1-\rho_{i,w}^2} && (6.13)
\end{aligned}
$$

By definition netting to zero across holdings, *risks orthogonal to the net portfolio are not counted* however consequential they might be.

As an extreme case, imagine a market-neutral equity portfolio broken into long and short. Market is the greatest risk within each side and, perceived as net hedged, not counted in risk contributions. However, since estimates are imperfect and the true beta is not exactly zero, the imperfect alignment can make market the portfolio's biggest real risk. The phenomenon occurs, to different degrees, in every hedge in any portfolio.

Thus, in conventional risk parity—particularly since it's the product of optimization and optimization finds perceived hedges—the most important real risk contributions can go unmeasured.

## 6.6 Adding uncertainty

To counter sensitivity to a particular estimate, especially linear combinations that seem to annihilate a direction of risk, covariance is no longer regarded as fixed and

known. Instead, entries have a mean and covariance: $\mathrm{E}\left[C_{ij}\right]$ and $\mathrm{cov}\left[C_{i,j}, C_{k,l}\right]$ for every $i, j, k, l$.

Equation (6.8) gives an investment's variance contribution, $v_i = w_i \left(\boldsymbol{C}\boldsymbol{w}\right)_i = w_i \sum_j C_{i,j}\, w_j$ . A portfolio's variance is the sum of the contributions, $v_P = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{w} = \sum_i v_i$ . Written in terms of the entries of $\boldsymbol{C}$, means and covariances are

$$\mathrm{E}\left[v_i\right] = w_i \sum_j \mathrm{E}\left[C_{i,j}\right] w_j \tag{6.14}$$

$$\mathrm{E}\left[v_P\right] = \sum_i \mathrm{E}\left[v_i\right] \tag{6.15}$$

$$\mathrm{cov}\left[v_i, v_j\right] = w_i w_j \sum_{k,l} w_k w_l \,\mathrm{cov}\left[C_{i,k}, C_{j,l}\right] \tag{6.16}$$

$$\mathrm{cov}\left[v_i, v_P\right] = \sum_j \mathrm{cov}\left[v_i, v_j\right] \tag{6.17}$$

$$\mathrm{var}\left[v_P\right] = \sum_i \mathrm{cov}\left[v_i, v_P\right] \tag{6.18}$$

### 6.6.1 Risk parity restated with uncertainty

Variability in uncertain quantities makes meeting targets, e.g., equation (6.9), impossible. A loss function assigns costs to deviations. Risk parity weights are found by minimizing expected loss.

Uncertainty also means that portfolio variance is random, so there is a difference between aiming toward fixed *target contributions* (target fraction × target portfolio variance) and *target fractions* of random portfolio variance (target fraction × random portfolio variance) with scale set by budget constraint or portfolio variance itself aimed toward a target. The deviations for each case:

1. Target contributions

$$\begin{aligned}\delta_i &= \text{investment } i\text{'s deviation from budgeted variance, } i = 1 \ldots N \\ &= v_i - f_i\, \sigma^2\end{aligned} \tag{6.19}$$

2. Target fractions

$$\begin{aligned}\delta_i &= \text{investment } i\text{'s deviation from budgeted fraction of variance, } i = 1 \ldots N \\ &= v_i - f_i\, v_P\end{aligned} \tag{6.20}$$

normalized by a portfolio variance target

$$\delta_{N+1} = v_P - \sigma^2 \tag{6.21}$$

or a budget constraint

$$\sum_i w_i = 1 \tag{6.22}$$

Measuring loss by summed mean-squared error yields

$$\begin{aligned} l(\boldsymbol{w}) &= \text{risk parity loss under } \boldsymbol{w} \\ &= \sum_i \mathrm{E}\left[\delta_i^2\right] && (6.23) \\ &= \sum_i \mathrm{E}^2\left[\delta_i\right] + \mathrm{var}\left[\delta_i\right] && (6.24) \end{aligned}$$

With no uncertainty, the variance terms are zero, and the solution is the same as standard risk parity.

Loss is computed from the quantities in equations (6.14) – (6.18):

Target contributions

$$\mathrm{E}\left[\delta_i\right] = \mathrm{E}\left[v_i\right] - f_i\,\sigma^2 \tag{6.25}$$

$$\mathrm{var}\left[\delta_i\right] = \mathrm{var}\left[v_i\right] \tag{6.26}$$

Target fractions

$$\mathrm{E}\left[\delta_i\right] = \mathrm{E}\left[v_i\right] - f_i\mathrm{E}\left[v_P\right] \tag{6.27}$$

$$\mathrm{var}\left[\delta_i\right] = \mathrm{var}\left[v_i\right] + f_i^2\mathrm{var}\left[v_P\right] - 2f_i\,\mathrm{cov}\left[v_i, v_P\right] \tag{6.28}$$

$$\mathrm{E}\left[\delta_{N+1}\right] = \mathrm{E}\left[v_P\right] - \sigma^2 \tag{6.29}$$

$$\mathrm{var}\left[\delta_{N+1}\right] = \mathrm{var}\left[v_P\right] \tag{6.30}$$

Both characterizations are 4th degree polynomials of portfolio weights, whose (local) minimizer can be found using nonlinear optimization techniques.

### 6.6.2 Multiple regimes

In determining allocations, regimes—e.g., risk-on vs. risk-off—and scenarios naturally enter. Rather than representing variation in the entries, each has its own covariance

estimate to measure loss. With a probability assigned to each scenario, the distribution of $\boldsymbol{C}$ is a mixture. Combined loss is the probability weighted average of the losses.

$$\begin{aligned} l(\boldsymbol{w}) &= \text{risk parity loss under } \boldsymbol{w} \\ &= \sum_i \mathrm{E}\left[\delta_i^2\right] \qquad (6.31) \\ &= \sum_i \mathrm{E}\left[\mathrm{E}\left[\delta_i^2 | \text{regime's estimates}\right]\right] \qquad (6.32) \end{aligned}$$

## 6.7 Uncertain risk decomposition

A risk decomposition report attributes a portfolio's risk to sources, which are none other than risk parity's contributions. Applying the same machinery measures the effect of quantities' varying within their uncertainty. It surfaces the risks buried in fragile hedges.

Equation (6.8) is the convention for determining an investment's risk contribution to variance, $v_i$, and divided by portfolio standard deviation as in equation (6.6), standard deviation.

With uncertainty added, the single number generated from a point-estimate is replaced by the more informative $\mathrm{E}\left[v_i\right] \pm stdev[v_i]$ from equations (6.14) and (6.16).

For groups of securities—e.g., to summarize by industry, capitalization, or style—the risk contribution of a weighted basket, $\sum_i a_i v_i$, has mean $\sum_i a_i \mathrm{E}\left[v_i\right]$ and variance $\sum_{i,j} a_i a_j \mathrm{cov}\left[v_i, v_j\right]$.

Note that with covariance modeled by factors as in section 6.8, uncertainty affects even the expected contribution, as in equation (6.41) in section 6.11.2 (repeated here):

$$\mathrm{E}\left[C_{i,j}\right] = \sum_{k=1\ldots K} \left[\hat{E}_{i,k}\hat{E}_{j,k} + \underbrace{\hat{X}_{i,k}(j,k)}_{\text{from uncertainty}}\right] \hat{\theta}_k + 1_{\{i=j\}}\, \hat{\varepsilon}_i^2\, \hat{\theta}_{K+1} \qquad (6.33)$$

This happens because the map from exposures to risk is nonlinear. Recall the earlier long-short market neutral example. The additional term is the average consequence of beta taking values in the range of uncertainty around its mean of zero, all of which involve risk (while the mean has none).

## 6.8 Factor-modeled covariance uncertainty

The estimates needed for uncertain risk parity—mean and covariance of the covariance matrix entries—can be generated assorted ways. For example, parametrically as in section 6.12, one could use the covariance of estimated covariance of samples assumed Gaussian[4]. Or nonparametrically, one could repeatedly sample data and estimate a matrix, then compute moments across the estimates. Instead of sampling, the same could be done across a modeled set of possible covariance matrices, perhaps assigning probabilities to differently structured versions, e.g., full covariance, diagonalized, single index, constant correlation.

A useful framework for expressing covariance—especially among investments whose composition changes over time, e.g., S&P500's shift from industrials to financials to tech, thus not properly seen by history—is a linear factor model. To a factor model, a large universe does not present a problem. Covariance between investments that are themselves portfolios can be built up from their factor-modeled individual holdings to create forecasts based on present composition.

A conventional factor model expresses covariance as

$$\boldsymbol{C} = \boldsymbol{E}\boldsymbol{F}\boldsymbol{E}^{\mathrm{T}} + \operatorname{diag}\left[\boldsymbol{\varepsilon}^2\right] \tag{6.34}$$

$$\begin{aligned} \text{where} \quad \boldsymbol{E} &= (N \times K) \text{ exposures} \\ \boldsymbol{F} &= (K \times K) \text{ covariance between factors} \\ \boldsymbol{\varepsilon}^2 &= (N \times 1) \text{ stock-specific risk} \end{aligned}$$

The uncertain form of Shah [2015], with parameters no longer fixed, is

$$\boldsymbol{C} = \boldsymbol{E}\left(\overbrace{\boldsymbol{M}\operatorname{diag}\left[\theta_1, \ldots, \theta_k\right]\boldsymbol{M}^{\mathrm{T}}}^{\boldsymbol{F}}\right)\boldsymbol{E}^{\mathrm{T}} + \theta_{K+1}\operatorname{diag}\left[\boldsymbol{\varepsilon}^2\right] \tag{6.35}$$

$$\begin{aligned} \text{where} \quad \theta &= (K+1 \times 1) \\ \boldsymbol{M} &= (K \times K) \end{aligned} \tag{6.36}$$

As in that paper, $\boldsymbol{E}, \boldsymbol{M}, \theta$, and $\boldsymbol{\varepsilon}^2$ are assumed jointly independent, and $\boldsymbol{E}$ redefined to subsume $\boldsymbol{M}$: $\boldsymbol{E} \leftarrow \boldsymbol{E}\boldsymbol{M}$.

[4]For the usual unbiased estimates of covariance made from T samples of an i.i.d. multivariate Gaussian $\boldsymbol{x}$, $\operatorname{cov}\left[\hat{\sigma}_{x_1x_2}, \hat{\sigma}_{x_3x_4}\right] = \left(\sigma_{x_1,x_3}\,\sigma_{x_2,x_4} + \sigma_{x_1,x_4}\,\sigma_{x_2,x_3}\right)/(T-1)$

$$\text{Then } \boldsymbol{C} = \boldsymbol{E}\,\text{diag}\left[\theta_1 \dots \theta_k\right]\boldsymbol{E}^{\mathrm{T}} + \theta_{K+1}\,\text{diag}\left[\boldsymbol{\varepsilon}^2\right] \tag{6.37}$$

$$\text{and } C_{i,j} = \sum_{k=1\dots K} E_{i,k}E_{j,k}\theta_k + 1_{\{i=j\}}\,\varepsilon_i^2\,\theta_{K+1} \tag{6.38}$$

Given estimates of the mean and covariance of the pieces $(\boldsymbol{E}, \theta, \boldsymbol{\varepsilon}^2)$, equations (6.46) and (6.47) in section 6.11 yield $\mathrm{E}\left[C_{p_1,p_2}\right]$ and $\mathrm{cov}\left[C_{p_1,p_2}, C_{p_3,p_4}\right]$ for any portfolios $p_1, p_2, p_3, p_4$.

## 6.9 An example

Python code for this example is on the author's github:
`https://github.com/AnishRSIGM/UncertainRiskParity`.

Conventional and uncertain risk-parity portfolios are constructed from 4 indices proxied by ETFs: AGG-iShares Core US Aggregate Bond, EEM-iShares MSCI Emerging Markets, EFA-iShares MSCI EAFE, SPY-S&P500. From 160 weeks of Wed–Wed weekly returns, 2017 Aug 23 – 2020 Sept 16, 75% (120 periods) are used for estimation (training set), and 25% (40 periods) for out-of-sample performance (testing set).

The data are examined split two ways. Experiment 1 cuts past and future—the first 120 weeks are for training, the last 40 testing. Experiment 2 separates observations within the interval—the first 3 of every 4 weeks are for training, the 4th for testing. The experiments are unusually different because markets became more volatile and tightly correlated after the pandemic hit in late February 2020. Standard deviation and correlation of weekly returns are in tables 6.1 and 6.2. The difference between training and test volatility for AGG shows bonds experienced periods of extreme returns, which fall in the test set of experiment 1 and training set of experiment 2.

Portfolios are constructed from parameters estimated on the training set and evaluated on the testing set. Expected covariance is taken as the sample covariance over the training set. Covariance of the covariance is estimated ad-hoc by cutting the 120 training periods every 10 periods, calculating the covariance over each group, and finally, the covariance across the matrices. The results are in tables 6.3 and 6.4.

Table 6.1: Experiment 1: Std Dev & Correlation

Training – 8/23/2017–12/11/2019

| | Weekly SD | AGG | EEM | EFA | SPY |
|---|---|---|---|---|---|
| AGG | 0.41 | 1.00 | -0.19 | -0.19 | -0.19 |
| EEM | 2.52 | -0.19 | 1.00 | 0.89 | 0.78 |
| EFA | 1.73 | -0.19 | 0.89 | 1.00 | 0.87 |
| SPY | 1.83 | -0.19 | 0.78 | 0.87 | 1.00 |

Testing – 12/11/2019–9/16/2020

| | Weekly SD | AGG | EEM | EFA | SPY |
|---|---|---|---|---|---|
| AGG | 1.75 | 1.00 | 0.69 | 0.74 | 0.53 |
| EEM | 4.82 | 0.69 | 1.00 | 0.93 | 0.84 |
| EFA | 4.52 | 0.74 | 0.93 | 1.00 | 0.89 |
| SPY | 4.18 | 0.53 | 0.84 | 0.89 | 1.00 |

Notice in uncertain risk parity, the training variance contribution has a mean and standard deviation instead of taking a single value.

Results reflect the historic events. In experiment 1, the pandemic is entirely in the testing set, and predicted contributions and their standard deviation are low. In experiment 2, much of the pandemic is in the training set, and predicted contributions and, more so, their standard deviations are high.

The example is a simplified illustration. In practice, forecasts go through complex procedures of conditioning, forecasting, and points of manual intervention. Using the linked Python code, readers can construct both conventional and uncertain risk parity portfolios from their conditioned forecasts and compare results.

## 6.10 Summary

Risk parity is a standard method of portfolio construction. Because it is an optimization and optimized alignments can conceal latent risks, there is a strong conceptual argument for working from more than point-estimates of covariance. Considering first and second moments, uncertain risk parity is solved using a loss function that weighs the variation around estimates. A secondary benefit of the machinery is a robust view of portfolio risk, uncertain risk decomposition, that applies to all portfolios regardless

Table 6.2: Experiment 2: Std Dev & Correlation
Training – 8/23/2017–9/16/2019 excluding every 4th period

| | Weekly SD | AGG | EEM | EFA | SPY |
|---|---|---|---|---|---|
| AGG | 1.06 | 1.00 | 0.48 | 0.59 | 0.38 |
| EEM | 3.44 | 0.48 | 1.00 | 0.91 | 0.82 |
| EFA | 2.89 | 0.59 | 0.91 | 1.00 | 0.88 |
| SPY | 2.80 | 0.38 | 0.82 | 0.88 | 1.00 |

Testing – 8/23/2017–9/16/2019 every 4th period

| | Weekly SD | AGG | EEM | EFA | SPY |
|---|---|---|---|---|---|
| AGG | 0.39 | 1.00 | -0.07 | -0.14 | -0.05 |
| EEM | 2.53 | -0.07 | 1.00 | 0.86 | 0.79 |
| EFA | 1.96 | -0.14 | 0.86 | 1.00 | 0.91 |
| SPY | 1.91 | -0.05 | 0.79 | 0.91 | 1.00 |

Table 6.3: Experiment 1: Weight and Variance Contributions
Training – 8/23/2017–12/11/2019
Testing – 12/11/2019–9/16/2020

Conventional Risk Parity

| | AGG | EEM | EFA | SPY |
|---|---|---|---|---|
| weight | 0.75 | 0.07 | 0.09 | 0.09 |
| $v$ training | 0.06 | 0.06 | 0.06 | 0.06 |
| $v$ testing | 2.67 | 0.62 | 0.86 | 0.66 |

Uncertain Risk Parity

| | AGG | EEM | EFA | SPY |
|---|---|---|---|---|
| weight | 0.79 | 0.06 | 0.08 | 0.07 |
| $\mathrm{E}[v]$ training | 0.08 | 0.05 | 0.05 | 0.04 |
| $SD[v]$ training | 0.08 | 0.67 | 0.54 | 0.61 |
| $v$ testing | 2.77 | 0.52 | 0.72 | 0.45 |

of construction. To accommodate the changes in composition that occur over time, a factor model, in uncertain form, can be enlisted to model uncertain covariance between portfolios.

Table 6.4: Experiment 2: Weight and Variance Contributions
Training – 8/23/2017–9/16/2019 ex every 4th period
Testing – 8/23/2017–9/16/2019 every 4th period

Conventional Risk Parity

| | AGG | EEM | EFA | SPY |
|---|---|---|---|---|
| weight | 0.54 | 0.13 | 0.15 | 0.17 |
| $v$ training | 0.71 | 0.71 | 0.71 | 0.71 |
| $v$ testing | 0.03 | 0.29 | 0.26 | 0.28 |

Uncertain Risk Parity

| | AGG | EEM | EFA | SPY |
|---|---|---|---|---|
| weight | 0.44 | 0.16 | 0.17 | 0.22 |
| $\mathrm{E}[v]$ training | 0.60 | 0.99 | 0.92 | 1.04 |
| $SD[v]$ training | 4.75 | 11.31 | 10.83 | 8.27 |
| $v$ testing | 0.01 | 0.43 | 0.36 | 0.44 |

## 6.11 Uncertain covariance between portfolios from securities

Mean and covariance of portfolio covariance can be computed from holdings and the mean and covariance of security covariance.

$$\begin{aligned} \text{Let } \boldsymbol{W} &= (N \times M) \text{ holdings of } N \text{ portfolios} \\ \boldsymbol{H} &= (M \times M) \text{ covariance between } M \text{ securities} \\ \boldsymbol{C} &= (N \times N) \text{ covariance between } N \text{ portfolios} \end{aligned}$$

### 6.11.1 Model-free

Without assuming an underlying model, computations are sums of 1st and 2nd moments.

$$\mathrm{E}\left[C_{i,j}\right] = \sum_{a,b} W_{i,a} W_{j,b} \,\mathrm{E}\left[H_{a,b}\right] \tag{6.39}$$

$$\begin{aligned}\mathrm{cov}\left[C_{i,j}, C_{m,n}\right] &= \mathrm{cov}\left[\sum_{a,b} W_{i,a} W_{j,b}\, H_{a,b}, \sum_{c,d} W_{m,c} W_{n,d}\, H_{c,d}\right] \\ &= \sum_{a,b,c,d} W_{i,a} W_{j,b} W_{m,c} W_{n,d} \,\mathrm{cov}\left[H_{a,b}, H_{c,d}\right]\end{aligned} \tag{6.40}$$

There is one problem—the dimension of $\mathrm{cov}\left[\boldsymbol{H}\right]$ is $M^4$, quickly occupying a tremendous amount of memory. A universe of 1000 stocks, accounting for symmetries, would involve over $\frac{\left(M^2/2\right)^2}{2} = 125$ billion numbers.

### 6.11.2 Uncertain covariance modeled

Space complexity issues can be avoided by imposing structure. The uncertain covariance model of Shah [2015] is as follows:

1. $\boldsymbol{H} = \boldsymbol{E}\,\mathrm{diag}\left[\theta_1 \ldots \theta_K\right] \boldsymbol{E}^{\mathrm{T}} + \theta_{K+1}\,\mathrm{diag}\left[\boldsymbol{\varepsilon}^2\right]$
2. $\boldsymbol{E}, \theta$, and $\boldsymbol{\varepsilon}^2$ are jointly independent
3. The elements of $\boldsymbol{\varepsilon}^2$ are independent
4. $\boldsymbol{E}$ is Gaussian
5. Mean and covariance of $\boldsymbol{E}, \theta$, and $\boldsymbol{\varepsilon}^2$ are

$$\begin{aligned}
\hat{\boldsymbol{E}} &= (M \times K) & \hat{\boldsymbol{E}} &= \mathrm{E}\left[\boldsymbol{E}\right] \\
\hat{\boldsymbol{X}} &= (M \times K \times M \times K) & \hat{\boldsymbol{X}}_{a,i,b,j} &= \mathrm{cov}\left[\boldsymbol{E}_{a,i}, \boldsymbol{E}_{b,j}\right] \\
\hat{\boldsymbol{\varepsilon}}^2 &= (M \times 1) & \hat{\boldsymbol{\varepsilon}}^2 &= \mathrm{E}\left[\boldsymbol{\varepsilon}^2\right] \\
\hat{\boldsymbol{\omega}} &= (M \times 1) & \hat{\omega}_a &= \mathrm{var}\left[\varepsilon_a^2\right] \\
\hat{\boldsymbol{\theta}} &= (K+1 \times 1) & \hat{\boldsymbol{\theta}} &= \mathrm{E}\left[\boldsymbol{\theta}\right] \\
\hat{\boldsymbol{\Omega}} &= (K+1 \times K+1) & \hat{\Omega}_{i,j} &= \mathrm{cov}\left[\theta_i, \theta_j\right]
\end{aligned}$$

Then

$$\mathrm{E}\left[H_{a,b}\right] = \sum_{k=1\ldots K} \left[\hat{E}_{a,k}\hat{E}_{b,k} + \hat{X}_{a,k,b,k}\right] \hat{\theta}_k + 1_{\{a=b\}}\, \hat{\varepsilon}_a^2\, \hat{\theta}_{K+1} \tag{6.41}$$

$$\mathrm{E}\left[\boldsymbol{H}\right] = \hat{\boldsymbol{E}}\, \mathrm{diag}\left[\hat{\theta}_1 \ldots \hat{\theta}_K\right] \hat{\boldsymbol{E}}^{\mathrm{T}} + \sum_{k=1\ldots K} \hat{\theta}_k\, \hat{\boldsymbol{X}}_{\cdot,k,\cdot,k} + \hat{\theta}_{K+1}\ \mathrm{diag}\left[\hat{\boldsymbol{\varepsilon}}^2\right] \tag{6.42}$$

$$\begin{aligned}
\mathrm{cov}\left[H_{a,b}, H_{c,d}\right] = & \sum_{1\le k,l\le K} \Big[\hat{E}_{a,k}\hat{E}_{c,l}\hat{X}_{b,k,d,l} + \hat{E}_{b,k}\hat{E}_{d,l}\hat{X}_{a,k,c,l} \\
& \qquad + \hat{X}_{a,k,c,l}\hat{X}_{b,k,d,l} + \hat{E}_{a,k}\hat{E}_{d,l}\hat{X}_{b,k,c,l} \\
& \qquad + \hat{E}_{b,k}\hat{E}_{c,l}\hat{X}_{a,k,d,l} + \hat{X}_{a,k,d,l}\hat{X}_{b,k,c,l}\Big] \left(\hat{\theta}_k\hat{\theta}_l + \hat{\Omega}_{k,l}\right) \\
& + \sum_{1\le k,l\le K} \Big[\hat{E}_{a,k}\hat{E}_{b,k}\hat{E}_{c,l}\hat{E}_{d,l} + \hat{E}_{a,k}\hat{E}_{b,k}\hat{X}_{c,l,d,l} \\
& \qquad + \hat{E}_{c,l}\hat{E}_{d,l}\hat{X}_{a,k,b,k} + \hat{X}_{a,k,b,k}\hat{X}_{c,l,d,l}\Big] \hat{\Omega}_{k,l} \\
& + 1_{\{a=b\}}\, \hat{\varepsilon}_a^2 \sum_{k=1\ldots K} \left[\hat{E}_{c,k}\hat{E}_{d,k} + \hat{X}_{c,k,d,k}\right] \hat{\Omega}_{k,K+1} \\
& + 1_{\{c=d\}}\, \hat{\varepsilon}_c^2 \sum_{k=1\ldots K} \left[\hat{E}_{a,k}\hat{E}_{b,k} + \hat{X}_{a,k,b,k}\right] \hat{\Omega}_{k,K+1} \\
& + 1_{\{a=b\}} 1_{\{c=d\}}\, \hat{\varepsilon}_a^2 \hat{\varepsilon}_c^2 \hat{\Omega}_{K+1,K+1} \\
& + 1_{\{a=b=c=d\}}\, \hat{\omega}_a \left(\hat{\theta}_{K+1}^2 + \hat{\Omega}_{K+1,K+1}\right)
\end{aligned} \tag{6.43}$$

Using $\boldsymbol{C} = \boldsymbol{W}\boldsymbol{H}\boldsymbol{W}^{\mathrm{T}}$ with this structured form and applying formulas (6.39) and (6.40) to equations (6.42) and (6.43) yields the expected value and covariance of

portfolio covariance.

$$\begin{aligned}
\text{Let } \boldsymbol{G} &= (N \times K) \text{ factor exposures of portfolios} \\
\hat{\boldsymbol{G}} &= \mathrm{E}\left[\boldsymbol{G}\right] \\
&= \boldsymbol{W}\hat{\boldsymbol{E}} \qquad (6.44) \\
\hat{\boldsymbol{Y}} &= (N \times K \times N \times K) \ \mathrm{cov}\left[\boldsymbol{G}\right] \\
\hat{\boldsymbol{Y}}_{m,i,n,j} &= \mathrm{cov}\left[\boldsymbol{G}_{m,i}, \boldsymbol{G}_{n,j}\right] \\
\hat{\boldsymbol{Y}}_{\cdot,i,\cdot,j} &= \boldsymbol{W}\hat{\boldsymbol{X}}_{\cdot,i,\cdot,j}\boldsymbol{W}^{\mathrm{T}} \qquad (6.45)
\end{aligned}$$

$$\mathrm{E}\left[\boldsymbol{C}\right] = \hat{\boldsymbol{G}}\,\mathrm{diag}\left[\hat{\theta}_1 \ldots \hat{\theta}_K\right]\hat{\boldsymbol{G}}^{\mathrm{T}} + \sum_{k=1\ldots K}\hat{\theta}_k\,\hat{\boldsymbol{Y}}_{\cdot,k,\cdot,k} + \hat{\theta}_{K+1}\,\boldsymbol{W}\,\mathrm{diag}\left[\hat{\boldsymbol{\varepsilon}}^2\right]\boldsymbol{W}^{\mathrm{T}} \qquad (6.46)$$

$$\begin{aligned}
\mathrm{cov}\left[C_{i,j}, C_{m,n}\right] = &\sum_{1\le k,l\le K}\Big[\hat{G}_{i,k}\hat{G}_{m,l}\hat{Y}_{j,k,n,l} + \hat{G}_{j,k}\hat{G}_{n,l}\hat{Y}_{i,k,m,l} \\
&\quad + \hat{Y}_{i,k,m,l}\hat{Y}_{j,k,n,l} + \hat{G}_{i,k}\hat{G}_{n,l}\hat{Y}_{j,k,m,l} \\
&\quad + \hat{G}_{j,k}\hat{G}_{m,l}\hat{Y}_{i,k,n,l} + \hat{Y}_{i,k,n,l}\hat{Y}_{j,k,m,l}\Big]\left(\hat{\theta}_k\hat{\theta}_l + \hat{\Omega}_{k,l}\right) \\
&+ \sum_{1\le k,l\le K}\Big[\hat{G}_{i,k}\hat{G}_{j,k}\hat{G}_{m,l}\hat{G}_{n,l} + \hat{G}_{i,k}\hat{G}_{j,k}\hat{Y}_{m,l,n,l} \\
&\quad + \hat{G}_{m,l}\hat{G}_{n,l}\hat{Y}_{i,k,j,k} + \hat{Y}_{i,k,j,k}\hat{Y}_{m,l,n,l}\Big]\hat{\Omega}_{k,l} \\
&+ \left(\sum_a W_{i,a}W_{j,a}\,\hat{\varepsilon}_a^2\right)\left(\sum_{k=1\ldots K}\left[\hat{G}_{m,k}\hat{G}_{n,k} + \hat{Y}_{m,k,n,k}\right]\hat{\Omega}_{k,K+1}\right) \\
&+ \left(\sum_a W_{m,a}W_{n,a}\,\hat{\varepsilon}_a^2\right)\left(\sum_{k=1\ldots K}\left[\hat{G}_{i,k}\hat{G}_{j,k} + \hat{Y}_{i,k,j,k}\right]\hat{\Omega}_{k,K+1}\right) \\
&+ \left(\sum_a W_{i,a}W_{j,a}\,\hat{\varepsilon}_a^2\right)\left(\sum_a W_{m,a}W_{n,a}\,\hat{\varepsilon}_a^2\right)\hat{\Omega}_{K+1,K+1} \\
&+ \left(\sum_a W_{i,a}W_{j,a}W_{m,a}W_{n,a}\,\hat{\omega}_a\right)\left(\hat{\theta}_{K+1}^2 + \hat{\Omega}_{K+1,K+1}\right) \qquad (6.47)
\end{aligned}$$

## 6.12 Parametric uncertainty covariance of unbiased covariance estimates

**Lemma 6.1.** *Parametric estimate of covariance uncertainty*

*If* $\hat{\sigma}_{ab}$ *and* $\hat{\sigma}_{cd}$ *are the usual unbiased estimates of covariance made from* $N$ *samples of an i.i.d. multivariate Gaussian,* $\left\{ \begin{pmatrix} a_i \\ b_i \\ c_i \\ d_i \end{pmatrix} \right\}_{i=1..N}$ ,

*then* $\operatorname{cov}\left[\hat{\sigma}_{ab}, \hat{\sigma}_{cd}\right] = \left(\sigma_{a,c}\,\sigma_{b,d} + \sigma_{a,d}\,\sigma_{b,c}\right)/(N-1)$

$$\begin{aligned} \textit{where } \hat{\sigma}_{ab} &= \frac{1}{N-1}\sum_i \left[a_i - \frac{1}{N}\sum_k a_k\right]\left[b_i - \frac{1}{N}\sum_l b_l\right] \\ &= \frac{1}{N-1}\left[\sum_i a_i b_i - \frac{1}{N}\sum_k a_k \sum_l b_l\right] \end{aligned}$$

*Proof.* Since $\mathrm{E}\left[a_i\right] = \mathrm{E}\left[\frac{1}{N}\sum_k a_k\right]$, the means are irrelevant so can be assumed to be 0.

$$\operatorname{cov}\left[\hat{\sigma}_{ab}, \hat{\sigma}_{cd}\right] = \mathrm{E}\left[\hat{\sigma}_{ab}\hat{\sigma}_{cd}\right] - \mathrm{E}\left[\hat{\sigma}_{ab}\right]\mathrm{E}\left[\hat{\sigma}_{cd}\right] \tag{6.48}$$

$$= \mathrm{E}\left[\hat{\sigma}_{ab}\hat{\sigma}_{cd}\right] - \sigma_{ab}\,\sigma_{cd} \tag{6.49}$$

$$(N-1)^2\,\mathrm{E}\left[\hat{\sigma}_{ab}\hat{\sigma}_{cd}\right] = \mathrm{E}\left[\left(\sum_i a_i b_i - \frac{1}{N}\sum_k a_k \sum_l b_l\right)\left(\sum_j c_j d_j - \frac{1}{N}\sum_m c_m \sum_n d_n\right)\right] \tag{6.50}$$

$$\begin{aligned} &= \underbrace{\mathrm{E}\left[\sum_i a_i b_i \sum_j c_j d_j\right]}_{①} - \frac{1}{N}\underbrace{\mathrm{E}\left[\sum_i a_i b_i \sum_m c_m \sum_n d_n\right]}_{②} \\ &\quad - \frac{1}{N}\underbrace{\mathrm{E}\left[\sum_j c_j d_j \sum_k a_k \sum_l b_l\right]}_{③} \\ &\quad + \frac{1}{N^2}\underbrace{\mathrm{E}\left[\sum_k a_k \sum_l b_l \sum_m c_m \sum_n d_n\right]}_{④} \end{aligned} \tag{6.51}$$

$$② = \mathrm{E}\left[\sum_{i,m,n} a_i b_i c_m d_n\right] \tag{6.52}$$

$$= \underbrace{\mathrm{E}\left[\sum_{i,m,n\neq m} a_i b_i c_m d_n\right]}_{=0} + \mathrm{E}\left[\sum_{i,m} a_i b_i c_m d_m\right] \tag{6.53}$$

since first and third moments are zero

$$= \mathrm{E}\left[\sum_i a_i b_i \sum_m c_m d_m\right] \tag{6.54}$$

$$= ① \tag{6.55}$$

$$\text{Likewise, } ③ = ① \tag{6.56}$$

$$① = \mathrm{E}\left[\sum_{i,j} a_i b_i c_j d_j\right] \tag{6.57}$$

$$= \mathrm{E}\left[\sum_{i,j\neq i} a_i b_i c_j d_j + \sum_i a_i b_i c_i d_i\right] \tag{6.58}$$

$$= \sum_{i,j\neq i} \mathrm{E}\left[a_i b_i\right] \mathrm{E}\left[c_j d_j\right] + \sum_i \mathrm{E}\left[a_i b_i c_i d_i\right] \tag{6.59}$$

$$= \sum_{i,j\neq i} \sigma_{a,b}\, \sigma_{c,d} + \sum_i \mathrm{E}\left[a_i b_i c_i d_i\right] \tag{6.60}$$

$$= N(N-1)\sigma_{a,b}\, \sigma_{c,d} + \sum_i \mathrm{E}\left[a_i b_i c_i d_i\right] \tag{6.61}$$

$$= N(N-1)\sigma_{a,b}\, \sigma_{c,d} + \sum_i \sigma_{a,b}\, \sigma_{c,d} + \sigma_{a,c}\, \sigma_{b,d} + \sigma_{a,d}\, \sigma_{b,c} \tag{6.62}$$

by Isserlis' theorem Isserlis [1918]

$$= N^2\sigma_{a,b}\, \sigma_{c,d} + N\sigma_{a,c}\, \sigma_{b,d} + N\sigma_{a,d}\, \sigma_{b,c} \tag{6.63}$$

$$④ = \mathrm{E}\left[\sum_{k,l,m,n} a_k b_l c_m d_n\right] \tag{6.64}$$

$$= \mathrm{E}\left[\sum_{k,m\neq k} a_k b_k c_m d_m\right] + \mathrm{E}\left[\sum_{k,m\neq k} a_k b_m c_k d_m\right] + \mathrm{E}\left[\sum_{k,m\neq k} a_k b_m c_m d_k\right]$$
$$+ \mathrm{E}\left[\sum_{k} a_k b_k c_k d_k\right] \qquad \text{since first and third moments are zero} \tag{6.65}$$

$$= N(N-1)\left(\sigma_{a,b}\,\sigma_{c,d} + \sigma_{a,c}\,\sigma_{b,d} + \sigma_{a,d}\,\sigma_{b,c}\right)$$
$$+ \sum_{k} \sigma_{a,b}\,\sigma_{c,d} + \sigma_{a,c}\,\sigma_{b,d} + \sigma_{a,d}\,\sigma_{b,c} \tag{6.66}$$

$$= N^2\left(\sigma_{a,b}\,\sigma_{c,d} + \sigma_{a,c}\,\sigma_{b,d} + \sigma_{a,d}\,\sigma_{b,c}\right) \tag{6.67}$$

Substituting back into equations (6.49) and (6.51)

$$(N-1)^2\,\mathrm{E}\left[\hat{\sigma}_{ab}\hat{\sigma}_{cd}\right] = \frac{N-2}{N}\left(N^2\sigma_{a,b}\,\sigma_{c,d} + N\sigma_{a,c}\,\sigma_{b,d} + N\sigma_{a,d}\,\sigma_{b,c}\right)$$
$$+ \left(\sigma_{a,b}\,\sigma_{c,d} + \sigma_{a,c}\,\sigma_{b,d} + \sigma_{a,d}\,\sigma_{b,c}\right) \tag{6.68}$$

$$= (N-1)^2\sigma_{a,b}\,\sigma_{c,d} + (N-1)\left(\sigma_{a,c}\,\sigma_{b,d} + \sigma_{a,d}\,\sigma_{b,c}\right) \tag{6.69}$$

$$\mathrm{cov}\left[\hat{\sigma}_{ab}, \hat{\sigma}_{cd}\right] = \mathrm{E}\left[\hat{\sigma}_{ab}\hat{\sigma}_{cd}\right] - \sigma_{ab}\,\sigma_{cd}$$
$$= \frac{\sigma_{a,c}\,\sigma_{b,d} + \sigma_{a,d}\,\sigma_{b,c}}{N-1} \tag{6.70}$$

□

# CHAPTER 7

# RISK UNDER UNCERTAINTY AND PRICE MOVEMENT

Risk decomposition separates a portfolio's variance or standard deviation into sources. It is a standard tool for analyzing the risks of an investment portfolio. The portfolio is divided into notional parts—e.g., individual securities, holdings by sector or region, factor exposures—whose contributions to net risk are estimated and reported. The procedure is described in Menchero and Davis [2011] and O'Cinneide [2018].

For risk measured in variance, a part's contribution is its covariance with whole; in standard deviation, this number is divided by portfolio standard deviation. For example, for the simple case of measuring security contributions to portfolio risk against a risk-free benchmark,

$$\begin{aligned} v \equiv&\ \text{security contributions to portfolio variance} \in \mathbb{R}^n \\ \equiv&\ w \odot \Sigma w \qquad (7.1) \\ s \equiv&\ \text{security contributions to portfolio std dev} \in \mathbb{R}^n \\ \equiv&\ v/\sqrt{w^{\mathrm{T}}\Sigma w} \qquad (7.2) \\ \text{where } \odot \equiv&\ \text{element-by-element multiplication} \end{aligned}$$

Convention regards the inputs—portfolio weights and covariance—as fixed and known, but composition changes with price and estimates have errors. In equation (7.1), neither $w$ nor $\Sigma$ is actually fixed.

Section 6 of O'Cinneide [2018] considers perturbations of $\Sigma$. The machinery developed in chapter 5 of this dissertation makes it possible to deform $\Sigma$ according to beliefs and compute $\mathrm{E}[\Sigma]$ and $\mathrm{cov}[\Sigma]$. The paper underlying this chapter, Shah [2020], additionally lets security weights vary and, assuming weight movement independent of parameter uncertainty, computes risk contributions' mean and variance.

Evaluating across the range of possibility measures risks more accurately and

surfaces latent fragility in hedging. Presentations of the paper have been well-received by a broad audience. Soon after giving an online talk, I was invited to present to professional/academic associations in New York[1], Amsterdam[2], and London[3].

## 7.1 Mathematical model

As in the two preceding chapters, covariance is modeled via factors, this time not orthogonal,

$$\begin{aligned}\Sigma &\equiv \operatorname{cov}\left[Y|E,F,\gamma,\varepsilon^2\right] \\ &\equiv EFE^{\mathrm{T}} + \gamma\operatorname{diag}(\varepsilon^2)\end{aligned} \tag{7.3}$$

In the mathematical framework of chapter 3, we are interested in the center and spread of functions $f\left(\Sigma_t, w_t\right)|X_t$ where $Z_t = \{E_t, F_t, \gamma_t, \varepsilon_t^2, w_t\}$ and the side information $X_t = \{\text{mean and variance of } Z_t\}$ encapsulates all the information of $Y_{1:t-1}, X_{1:t-1}, Z_{1:t-1}$.

[1] Shah [2021c]

[2] Shah [2021b]

[3] Shah [2022]

**from Shah [2020] Risk under Uncertainty and Price Movement**

## 7.2 Abstract

Risk decomposition is a standard tool for analyzing investment portfolio risk. The portfolio is divided into notional parts—e.g., individual securities, holdings by sector or region, factor exposures—whose contributions to net risk are estimated and reported. Convention regards the inputs—portfolio weights and covariance—as fixed and known, but portfolio composition changes with price movement and estimates have errors. Since behavior only in the direction of net risk is counted, hedged effects are invisible regardless of size. What if numbers aren't exact? Hedge instability manifests in proportion to underlying gross (not net) exposure, akin to leverage. For example, in a market-neutral portfolio, market is the largest risk in each side, conventionally uncounted since hedged, and extremely consequential under small deviations. To solve the problem, this paper models weights and parameters as uncertain. Evaluating a portfolio across the range of possibility measures risks better and surfaces latent fragility. No longer point-estimated, contributions are reported with a center and spread.

## 7.3 Introduction

Risk decomposition is a tool used by portfolio managers, risk managers, allocators, and others to examine a portfolio's risk profile. It breaks net risk into contributions from the portfolio's notional parts, which are defined according to the perspective sought. Parts can be concrete—e.g., individual securities; holdings grouped by sector, geographical region, or separately managed subportfolio—or abstract, e.g., exposures to risk factors and idiosyncratic effects.

Knowledge of how parts contribute to or hedge portfolio volatility guides decisions throughout the investment process. A portfolio manager can see over/undersized and hidden bets, for example, that an investment tilt adds less risk than thought so can

be increased. Risk managers and allocators use the information to keep portfolios on mandate, analyze changes, and compute metrics, e.g., parametric VaR contributions. The information enters performance measurement in attribution computations. Client management and sales teams communicate and certify the nature of portfolios through risk contribution reports produced by commercial investment platforms. Beyond reporting, risk contributions are the quantities equalized in risk parity portfolio construction.

### 7.3.1 Problem with conventional method

Numbers are computed from portfolio weights and security covariance, both regarded as fixed and known. However, the composition of a live portfolio shifts continuously with price movement, and inferred quantities are imperfect.

While many tasks rely on point estimates, the problem in this setting is articulated in Shah [2021a], which investigates risk parity under parameter uncertainty. The argument, expanded to include changes in composition, is as follows.

Menchero and Davis [2011], in giving a geometric interpretation of risk contributions, shows that behavior *in the direction of net risk* is what's counted. Behavior in all other directions cancels out by definition. At the same time, the gross exposure of these uncounted hedged behaviors can be large—picture the market effect contained in each side of a market-neutral[4]. If the system experiences any small deviation—through price movement or gap between estimated parameter and reality—the hedging balance is thrown off, and the risk of the gross exposure manifests.

The severe aspect of the problem is not that hedged behaviors inaccurately sum out, e.g., the short side's market effect has a negative risk contribution which seems to offset the long's positive. It's that they go unperceived, i.e., market risk scarcely appears anywhere in the portfolio because market is nearly orthogonal to the net thus invisible. These alignments occur because portfolios are constructed to hedge risks.

[4]Or in benchmark-relative risks, the benchmark acts as the short side.

### 7.3.2 Contribution and central idea of this paper

This paper surfaces latent risks by considering inputs across the range of uncertainty. Parameters and future portfolio weights are mathematically perturbed according to one's beliefs about their distribution. The impact of gross exposure occurs and is measured. Structural assumptions allow computation analytically, without sampling. Risk contributions, no longer derived point estimates, are accompanied with an estimate of their uncertainty.

O'Cinneide [2018] looks at perturbations from parameter uncertainty to risk contributions. There is also a paragraph on variance contributions under parameter uncertainty in the aforementioned uncertain risk parity paper. The solution presented here leverages the uncertain covariance technique used by that paper, from Shah [2015]. The author is aware of no other literature that considers both uncertainty and price movement.

## 7.4 Conventional risk contributions

Risk contributions are the portion of volatility, in either standard deviation or variance, attributed to the separate parts of the portfolio. They sum to portfolio volatility. Absent uncertainty and price movement, calculations are standard.[5]

In variance:

$$\begin{aligned} \sigma^2 &= \text{variance of portfolio} \\ v_i &= \text{part } i\text{'s contribution to portfolio variance} \\ &= \text{cov}\left[\text{part } i, \text{portfolio}\right] \\ \sum_i v_i &= \sigma^2 \end{aligned} \tag{7.4}$$

[5]For a derivation, see section 3.2.3 of Qian et al. [2007]

In standard deviation:

$$\begin{aligned} s_i &= \text{part } i\text{'s contribution to portfolio standard deviation} \\ &= v_i/\sigma \\ \sum_i s_i &= \sigma \end{aligned} \tag{7.5}$$

The contributed fraction of volatility is the same either way, $v_i/\sigma^2 = s_i/\sigma$.

### 7.4.1 Variance contributions of common partitions

Variance contributions of common portfolio partitions are stated here because they will be later modeled as uncertain. The corresponding standard deviation contribution comes from applying equation (7.5).

#### Individual securities, groups, or subportfolios

Consider the portfolio as a collection of subportfolios, $\boldsymbol{w} = \boldsymbol{p}^{(1)} + \boldsymbol{p}^{(2)} + \dots$ , e.g., by individual security, group (sector, geographic region, investment type), or separately managed holdings. Subportfolio $p$'s variance contribution is

$$\begin{aligned} v_p &= \boldsymbol{p}^{\mathrm{T}} \boldsymbol{C} \boldsymbol{w} \\ \text{where } \boldsymbol{w} &= (N \times 1) \text{ portfolio security weights} \\ \boldsymbol{C} &= (N \times N) \text{ covariance} \end{aligned} \tag{7.6}$$

If subportfolios are individual securities, the $n$'th security contributes

$$v_n = w_n \sum_m C_{n,m} w_m \tag{7.7}$$

**Factors**

If covariance is factor-modeled as

$$\boldsymbol{C} = \boldsymbol{E}\boldsymbol{F}\boldsymbol{E}^{\mathrm{T}} + \gamma \operatorname{diag}\left[\boldsymbol{\varepsilon}^2\right] \tag{7.8}$$

where $\boldsymbol{E} = (N \times K)$ factor exposures

$\boldsymbol{F} = (K \times K)$ covariance between factors

$\boldsymbol{\varepsilon}^2 = (N \times 1)$ idiosyncratic variance

$\gamma = 1$ for now. Will be used later

let $\boldsymbol{e} = (K \times 1)$ exposures of the portfolio

$$= \boldsymbol{E}^{\mathrm{T}}\boldsymbol{w} \tag{7.9}$$

$\boldsymbol{e}^p = (K \times 1)$ exposures from subportfolio $p$

$$= \boldsymbol{E}^{\mathrm{T}}\boldsymbol{p} \tag{7.10}$$

Variance contributions are

by factor

$$\text{factor } i: \quad v_i = e_i \left(\boldsymbol{F}\boldsymbol{e}\right)_i \tag{7.11}$$

$$\varepsilon^2: \quad v_{\varepsilon^2} = \gamma \sum_m w_m^2 \varepsilon_m^2 \tag{7.12}$$

$$\text{total:} \quad \sigma^2 = \sum_i v_i + v_{\varepsilon^2} \tag{7.13}$$

by factor and subportfolio

$$\text{subportfolio } p\text{, factor } i: \quad v_{p,i} = e_i^p \left(\boldsymbol{F}\boldsymbol{e}\right)_i \tag{7.14}$$

$$\text{subportfolio } p\text{, } \varepsilon^2: \quad v_{p,\varepsilon^2} = \gamma \sum_m p_m w_m \varepsilon_m^2 \tag{7.15}$$

$$\text{total:} \quad \sigma^2 = \sum_{k,i} v_{p^{(k)},i} + \sum_k v_{p^{(k)},\varepsilon^2} \tag{7.16}$$

## 7.5 Uncertain variance contributions

To add uncertainty—in $\boldsymbol{C}$ from estimation, in $\boldsymbol{w}$ from price movement—the point estimate and fixed weights are replaced by estimates of mean and variance[6].

$$\begin{aligned}
\mathrm{E}\left[\boldsymbol{C}\right] &= (N \times N) \text{ expected value of } \boldsymbol{C} \\
\mathrm{var}\left[\boldsymbol{C}\right] &= (N \times N \times N \times N) \text{ covariance between elements of } \boldsymbol{C} \\
\mathrm{E}\left[\boldsymbol{w}\right] &= (N \times 1) \text{ expected value of } \boldsymbol{w} \\
\mathrm{var}\left[\boldsymbol{w}\right] &= (N \times N) \text{ covariance between elements of } \boldsymbol{w}
\end{aligned}$$

Consider the individual security contributions of equation (7.7). With parameters and weights independent, mean and covariance are

$$\begin{aligned}
\mathrm{E}\left[v_{\text{security }n}\right] &= \mathrm{E}\left[\sum_m w_n\, C_{n,m}\, w_m\right] \\
&= \sum_m \mathrm{E}\left[C_{n,m}\right] \mathrm{E}\left[w_n w_m\right] \\
&= \sum_m \mathrm{E}\left[C_{n,m}\right] \left(\mathrm{E}\left[w_n\right] \mathrm{E}\left[w_m\right] + \mathrm{cov}\left[w_n, w_m\right]\right)
\end{aligned} \tag{7.17}$$

$$\begin{aligned}
\mathrm{cov}\left[v_n, v_m\right] &= \mathrm{cov}\left[\sum_i w_n\, C_{n,i}\, w_i, \sum_j w_m\, C_{m,j}\, w_j\right] \\
&= \sum_{i,j} \mathrm{cov}\left[w_n\, C_{n,i}\, w_i, w_m\, C_{m,j}\, w_j\right] \\
&= \sum_{i,j} \mathrm{E}\left[w_n w_i\right] \mathrm{E}\left[w_m w_j\right] \mathrm{cov}\left[C_{n,i}, C_{m,j}\right] \\
&\quad + \mathrm{cov}\left[w_n w_i, w_m w_j\right] \mathrm{E}\left[C_{n,i}\right] \mathrm{E}\left[C_{m,j}\right] \\
&\quad + \mathrm{cov}\left[w_n w_i, w_m w_j\right] \mathrm{cov}\left[C_{n,i}, C_{m,j}\right] \\
&\quad \text{since } \boldsymbol{w},\, \boldsymbol{C} \text{ independent}
\end{aligned} \tag{7.18}$$

Though equation (7.18) is straightforward, dimensionality in $\mathrm{var}\left[\boldsymbol{C}\right]$ and computational complexity can be a problem. As without uncertainty, both are reduced in the structure of a factor model.

[6] In this paper's notation, $\mathrm{var}\left[\boldsymbol{x}\right]$ means covariance of (scalar, vector, or matrix) variable $\boldsymbol{x}$ with itself and $\mathrm{cov}\left[\boldsymbol{x}, \boldsymbol{y}\right]$ is covariance between $\boldsymbol{x}$ and $\boldsymbol{y}$.

### 7.5.1 Uncertain factor-modeled covariance

Changing the parameters in the factor model of equation (7.8) from fixed to having mean and covariance yields an uncertain covariance model similar to Shah [2015] but more general[7]because no structure is imposed on $\boldsymbol{F}$.

$$\boldsymbol{C} = \boldsymbol{E}\boldsymbol{F}\boldsymbol{E}^{\mathrm{T}} + \gamma \operatorname{diag}\left[\boldsymbol{\varepsilon}^2\right] \tag{7.19}$$

Three assumptions make computations possible.

1. $\boldsymbol{E}, \boldsymbol{\varepsilon}^2, \{\boldsymbol{F}, \gamma\}$, and $\boldsymbol{w}$ are jointly independent.
2. The elements of $\boldsymbol{\varepsilon}^2$ are independent.
3. $\boldsymbol{w}$ and factor exposures (with the effect of shifting composition) of the portfolio and subportfolios are Gaussian.

The model is defined by mean and covariance of the parameters, and the portfolio by mean and covariance of its weights.

Exposures

$$\begin{aligned} \bar{\boldsymbol{E}} &= (N \times K) & \bar{\boldsymbol{E}} &= \mathrm{E}\left[\boldsymbol{E}\right] \\ \boldsymbol{X} &= (N \times K \times N \times K) & X_{m,i,n,j} &= \operatorname{cov}\left[E_{m,i}, E_{n,j}\right] \end{aligned}$$

Idiosyncratic effects

$$\begin{aligned} \bar{\boldsymbol{\varepsilon}}^2 &= (N \times 1) & \bar{\boldsymbol{\varepsilon}}^2 &= \mathrm{E}\left[\boldsymbol{\varepsilon}^2\right] \\ \boldsymbol{\omega} &= (N \times 1) & \omega_n &= \operatorname{var}\left[\varepsilon_n^2\right] \end{aligned}$$

Factor structure

$$\begin{aligned} \bar{\boldsymbol{F}} &= (K \times K) & \bar{\boldsymbol{F}} &= \mathrm{E}\left[\boldsymbol{F}\right] \\ \boldsymbol{H} &= (K \times K \times K \times K) & H_{i,j,k,l} &= \operatorname{cov}\left[F_{i,j}, F_{k,l}\right] \\ \bar{\gamma} &= (\text{scalar}) & \bar{\gamma} &= \mathrm{E}\left[\gamma\right] \\ \eta &= (\text{scalar}) & \eta &= \operatorname{var}\left[\gamma\right] \\ \boldsymbol{h} &= (K \times K) & h_{i,j} &= \operatorname{cov}\left[F_{i,j}, \gamma\right] \end{aligned}$$

[7]Appendix 7.11 converts that uncertain covariance structure to this paper's.

Portfolio weights

$$\bar{\boldsymbol{w}} = (N \times 1) \qquad \bar{\boldsymbol{w}} = \mathrm{E}\left[\boldsymbol{w}\right]$$
$$\boldsymbol{\Omega} = (N \times N) \qquad \Omega_{m,n} = \mathrm{cov}\left[w_m, w_n\right]$$

### 7.5.2 Mean and covariance of contributions

Contributions from subportfolio and factor in equations (7.14) and (7.15) are the building blocks of most quantities of interest, e.g., contributions by factor, region, or sector. Their means and covariances are derived in the following propositions: 7.1 (expectations $\mathrm{E}\left[v_{p,i}\right]$, $\mathrm{E}\left[v_{p,\varepsilon^2}\right]$) and 7.2, 7.3, 7.4 (covariance pairs $\mathrm{cov}\left[v_{p,i}\,,\, v_{q,k}\right]$, $\mathrm{cov}\left[v_{p,\varepsilon^2}, v_{q,\varepsilon^2}\right]$, $\mathrm{cov}\left[v_{p,i}, v_{q,\varepsilon^2}\right]$).

Let $P$ stand for the set of securities notionally in subportfolio $p$, even those with zero weight[8]. To avoid more notation, assume if a subportfolio holds a security, the weight is the whole portfolio's weight, i.e., $n \in P \rightarrow p_n = w_n$. Nevertheless, the formulas below allow for securities held in common so that they can be used to compute covariance with the entire portfolio.

[8] If weights are benchmark-relative, price movement in other securities shifts these away from zero.

**Proposition 7.1.** $\mathrm{E}\left[v_{p,i}\right]$ *and* $\mathrm{E}\left[v_{p,\varepsilon^2}\right]$

$$\mathrm{E}\left[v_{p,i}\right] = \overbrace{\bar{e}_i^p \left(\bar{\boldsymbol{F}}\bar{\boldsymbol{e}}\right)_i}^{mean} \quad + \quad \overbrace{\sum_j \bar{F}_{i,j} \operatorname{cov}\left[e_i^p, e_j\right]}^{uncertainty} \tag{7.20}$$

$$= \overbrace{\bar{e}_i^p \left(\bar{\boldsymbol{F}}\bar{\boldsymbol{e}}\right)_i}^{mean} \quad + \quad \sum_j \bar{F}_{i,j} \sum_{n \in P, m} \overbrace{X_{n,i,m,j}\bar{w}_n\bar{w}_m}^{E\ uncertainty} + \overbrace{\bar{E}_{n,i}\bar{E}_{m,j}\Omega_{n,m}}^{w\ uncertainty} + \overbrace{X_{n,i,m,j}\Omega_{n,m}}^{E,w\ unc.} \tag{7.21}$$

$$\mathrm{E}\left[v_{p,\varepsilon^2}\right] = \bar{\gamma} \sum_{n \in P} \left[ \underbrace{\bar{w}_n^2}_{mean} + \underbrace{\Omega_{n,n}}_{w\ unc.} \right] \bar{\varepsilon}_n^2 \tag{7.22}$$

$$\textit{where } \bar{\boldsymbol{e}}^p = \mathrm{E}\left[\boldsymbol{e}^p\right] = \bar{\boldsymbol{E}}^T \bar{\boldsymbol{p}} \tag{7.23}$$

$$\bar{\boldsymbol{e}} = \bar{\boldsymbol{E}}^T \bar{\boldsymbol{w}} \tag{7.24}$$

$\mathrm{E}\left[v_{p,i}\right]$ *computes in* $O(K)$ *given* $\bar{\boldsymbol{e}}^p$, $\bar{\boldsymbol{e}}$, $\operatorname{cov}\left[\boldsymbol{e}^p, \boldsymbol{e}\right]$. $\mathrm{E}\left[v_{p,\varepsilon^2}\right]$ *computes in* $O(N)$, $\bar{\boldsymbol{e}}^p$ *and* $\bar{\boldsymbol{e}}$ *in* $O(KN)$, *and* $\operatorname{cov}\left[\boldsymbol{e}^p, \boldsymbol{e}\right]$ *in* $O(K^2N^2)$.

*Proof.*

$$\begin{aligned}
\mathrm{E}\left[v_{p,i}\right] &= \mathrm{E}\left[e_i^p \left(\boldsymbol{F}\boldsymbol{e}\right)_i\right] && (7.25)\\
&= \mathrm{E}\left[e_i^p \sum_j F_{i,j} e_j\right] \\
&= \sum_j \mathrm{E}\left[F_{i,j}\right] \mathrm{E}\left[e_i^p e_j\right] \quad \text{since } \boldsymbol{e}, \boldsymbol{F} \text{ ind} \\
&= \sum_j \bar{F}_{i,j} \left(\mathrm{E}\left[e_i^p\right] \mathrm{E}\left[e_j\right] + \mathrm{cov}\left[e_i^p, e_j\right]\right) \\
&= \overbrace{\bar{e}_i^p \left(\bar{\boldsymbol{F}}\bar{\boldsymbol{e}}\right)_i}^{\text{mean}} \quad + \quad \overbrace{\sum_j \bar{F}_{i,j} \,\mathrm{cov}\left[e_i^p, e_j\right]}^{\text{uncertainty}} && (7.26)
\end{aligned}$$

$$\begin{aligned}
\mathrm{cov}\left[e_i^p, e_j^q\right] &= \mathrm{cov}\left[\sum_{n\in P} E_{n,i}\, w_n, \sum_{m\in Q} E_{m,j}\, w_m\right] \\
&= \sum_{n\in P, m\in Q} \mathrm{cov}\left[E_{n,i}\, w_n, E_{m,j}\, w_m\right] \\
&= \sum_{n\in P, m\in Q} \overbrace{X_{n,i,m,j} \bar{w}_n \bar{w}_m}^{E \text{ uncertainty}} + \overbrace{\bar{E}_{n,i} \bar{E}_{m,j} \Omega_{n,m}}^{w \text{ uncertainty}} + \overbrace{X_{n,i,m,j} \Omega_{n,m}}^{E,w \text{ unc.}} \\
&\qquad \text{since } \boldsymbol{E}, \boldsymbol{w} \text{ ind} && (7.27)
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\left[v_{p,\varepsilon^2}\right] &= \mathrm{E}\left[\gamma \sum_{n\in P} w_n^2 \varepsilon_n^2\right] && (7.28)\\
&= \mathrm{E}\left[\gamma\right] \sum_{n\in P} \mathrm{E}\left[w_n^2\right] \mathrm{E}\left[\varepsilon_n^2\right] \quad \text{since } \gamma, \boldsymbol{w}, \boldsymbol{\varepsilon}^2 \text{ ind} && (7.29)\\
&= \bar{\gamma} \sum_{n\in P} \left[\underbrace{\bar{w}_n^2}_{\text{mean}} + \underbrace{\Omega_{n,n}}_{w \text{ unc.}}\right] \bar{\varepsilon}_m^2 && (7.30)
\end{aligned}$$

□

*Notice that equations (7.21) and (7.22) separate expectation into sources—mean and different uncertainties.*

**Proposition 7.2.** $\mathrm{cov}\left[v_{p,i}\,,\, v_{q,k}\right]$

*If portfolio and subportfolio factor exposures* $\boldsymbol{e}$, $\boldsymbol{e}^p$, *and* $\boldsymbol{e}^q$ *are Gaussian,*

$$\begin{aligned}\operatorname{cov}\left[v_{p,i}\,,\, v_{q,k}\right] = \sum_{j,l} H_{i,j,k,l}\left(\bar{e}_i^p\bar{e}_j + \operatorname{cov}\left[e_i^p, e_j\right]\right)\left(\bar{e}_k^q\bar{e}_l + \operatorname{cov}\left[e_k^q, e_l\right]\right)\\ + \left(\bar{F}_{i,j}\bar{F}_{k,l} + H_{i,j,k,l}\right)\Bigg[\\ \left(\bar{e}_i^p\bar{e}_k^q + \operatorname{cov}\left[e_i^p, e_k^q\right]\right)\left(\bar{e}_j\bar{e}_l + \operatorname{cov}\left[e_j, e_l\right]\right)\\ + \left(\bar{e}_i^p\bar{e}_l + \operatorname{cov}\left[e_i^p, e_l\right]\right)\left(\bar{e}_k^q\bar{e}_j + \operatorname{cov}\left[e_k^q, e_j\right]\right)\\ - 2\bar{e}_i^p\bar{e}_k^q\bar{e}_j\bar{e}_l\Bigg]\end{aligned} \tag{7.31}$$

*and is computed in* $O(K^2)$ *given* $\bar{\boldsymbol{e}}, \bar{\boldsymbol{e}}^p, \bar{\boldsymbol{e}}^q, \operatorname{var}\left[\boldsymbol{e}\right], \operatorname{cov}\left[\boldsymbol{e}, \boldsymbol{e}^p\right], \operatorname{cov}\left[\boldsymbol{e}, \boldsymbol{e}^q\right]$, *and* $\operatorname{cov}\left[\boldsymbol{e}^p, \boldsymbol{e}^q\right]$.

*Proof.*

$$\begin{aligned}\operatorname{cov}\left[v_{p,i}\,,\, v_{q,k}\right] &= \operatorname{cov}\left[e_i^p \sum_j F_{i,j}e_j,\ e_k^q \sum_l F_{k,l}e_l\right]\\ &= \sum_{j,l} \operatorname{cov}\left[F_{i,j}e_i^p e_j,\ F_{k,l}e_k^q e_l\right]\\ &= \sum_{j,l} \operatorname{cov}\left[F_{i,j}, F_{k,l}\right]\operatorname{E}\left[e_i^p e_j\right]\operatorname{E}\left[e_k^q e_l\right]\\ &\qquad + \left(\bar{F}_{i,j}\bar{F}_{k,l} + \operatorname{cov}\left[F_{i,j}, F_{k,l}\right]\right)\operatorname{cov}\left[e_i^p e_j, e_k^q e_l\right]\\ &\qquad \text{since } \boldsymbol{F}, \boldsymbol{e} \text{ ind}\end{aligned} \tag{7.32}$$

$$\begin{aligned}&= \sum_{j,l} H_{i,j,k,l}\left(\bar{e}_i^p\bar{e}_j + \operatorname{cov}\left[e_i^p, e_j\right]\right)\left(\bar{e}_k^q\bar{e}_l + \operatorname{cov}\left[e_k^q, e_l\right]\right)\\ &\qquad + \left(\bar{F}_{i,j}\bar{F}_{k,l} + H_{i,j,k,l}\right)\underbrace{\operatorname{cov}\left[e_i^p e_j, e_k^q e_l\right]}_{①}\end{aligned} \tag{7.33}$$

$$\begin{aligned}① &= \left(\bar{e}_i^p\bar{e}_k^q + \operatorname{cov}\left[e_i^p, e_k^q\right]\right)\left(\bar{e}_j\bar{e}_l + \operatorname{cov}\left[e_j, e_l\right]\right)\\ &\quad + \left(\bar{e}_i^p\bar{e}_l + \operatorname{cov}\left[e_i^p, e_l\right]\right)\left(\bar{e}_k^q\bar{e}_j + \operatorname{cov}\left[e_k^q, e_j\right]\right) - 2\bar{e}_i^p\bar{e}_k^q\bar{e}_j\bar{e}_l\\ &\quad \text{since } \boldsymbol{e}, \boldsymbol{e}^p, \text{ and } \boldsymbol{e}^q \text{ Gaussian}\end{aligned} \tag{7.34}$$

□

**Proposition 7.3.** $\operatorname{cov}\left[v_{p,\varepsilon^2}, v_{q,\varepsilon^2}\right]$

*If* $\boldsymbol{w}$ *is Gaussian,*

$$\begin{aligned}\operatorname{cov}\left[v_{p,\varepsilon^2}, v_{q,\varepsilon^2}\right] &= \eta \operatorname{E}\left[v_{p,\varepsilon^2}\right] \operatorname{E}\left[v_{q,\varepsilon^2}\right] / \bar{\gamma}^2 \\ &\quad + \left(\bar{\gamma}^2 + \eta\right) \sum_{n\in P, m\in Q} \left(4\bar{w}_n \bar{w}_m \Omega_{n,m} + 2\Omega_{n,m}^2\right) \bar{\varepsilon}_n^2 \bar{\varepsilon}_m^2 \\ &\quad + \left(\bar{\gamma}^2 + \eta\right) \sum_{n\in P\cap Q} \left(\bar{w}_n^4 + 6\bar{w}_n^2 \Omega_{n,n} + 3\Omega_{n,n}^2\right) \omega_n \end{aligned} \tag{7.35}$$

*and is computed in* $O(N^2)$ *given* $\operatorname{E}\left[v_{p,\varepsilon^2}\right]$ *and* $\operatorname{E}\left[v_{q,\varepsilon^2}\right]$.

*Proof.*

$$\operatorname{cov}\left[v_{p,\varepsilon^2}, v_{q,\varepsilon^2}\right] = \operatorname{cov}\left[\gamma \sum_{n\in P} w_n^2 \varepsilon_n^2, \gamma \sum_{m\in Q} w_m^2 \varepsilon_m^2\right] \tag{7.36}$$

$$\begin{aligned} &= \operatorname{cov}\left[\gamma, \gamma\right] \operatorname{E}\left[\sum_{n\in P} w_n^2 \varepsilon_n^2\right] \operatorname{E}\left[\sum_{m\in Q} w_m^2 \varepsilon_m^2\right] \\ &\quad + \left(\operatorname{E}^2\left[\gamma\right] + \operatorname{cov}\left[\gamma, \gamma\right]\right) \operatorname{cov}\left[\sum_{n\in P} w_n^2 \varepsilon_n^2, \sum_{m\in Q} w_m^2 \varepsilon_m^2\right] \\ &\quad \text{since } \gamma, \{\boldsymbol{w}, \boldsymbol{\varepsilon}^2\} \text{ ind} \end{aligned} \tag{7.37}$$

$$\begin{aligned} &= \eta \frac{\operatorname{E}\left[v_{p,\varepsilon^2}\right]}{\bar{\gamma}} \frac{\operatorname{E}\left[v_{q,\varepsilon^2}\right]}{\bar{\gamma}} \\ &\quad + \left(\bar{\gamma}^2 + \eta\right) \underbrace{\operatorname{cov}\left[\sum_{n\in P} w_n^2 \varepsilon_n^2, \sum_{m\in Q} w_m^2 \varepsilon_m^2\right]}_{①} \end{aligned} \tag{7.38}$$

$$
\begin{aligned}
① &= \operatorname{cov}\left[\sum_{n\in P} w_n^2\varepsilon_n^2, \sum_{m\in Q} w_m^2\varepsilon_m^2\right] && (7.39)\\
&= \sum_{n\in P, m\in Q} \operatorname{cov}\left[w_n^2\varepsilon_n^2, w_m^2\varepsilon_m^2\right] && (7.40)\\
&= \sum_{n\in P, m\in Q} \operatorname{cov}\left[w_n^2, w_m^2\right]\bar{\varepsilon}_n^2\bar{\varepsilon}_m^2 \\
&\qquad + \left(\operatorname{E}\left[w_n^2\right]\operatorname{E}\left[w_m^2\right] + \operatorname{cov}\left[w_n^2, w_m^2\right]\right)\operatorname{cov}\left[\varepsilon_n^2, \varepsilon_m^2\right] && (7.41)\\
&\qquad \text{since } \boldsymbol{w}, \boldsymbol{\varepsilon}^2 \text{ ind} \\
&= \sum_{n\in P, m\in Q} \operatorname{cov}\left[w_n^2, w_m^2\right]\bar{\varepsilon}_n^2\bar{\varepsilon}_m^2 \\
&\quad + \sum_{n\in P\cap Q} \left(\operatorname{E}^2\left[w_n^2\right] + \operatorname{var}\left[w_n^2\right]\right)\operatorname{var}\left[\varepsilon_n^2\right] \quad \text{since elements of } \boldsymbol{\varepsilon}^2 \text{ ind} && (7.42)\\
&= \sum_{n\in P, m\in Q} \left(4\bar{w}_n\bar{w}_m\Omega_{n,m} + 2\Omega_{n,m}^2\right)\bar{\varepsilon}_n^2\bar{\varepsilon}_m^2 \\
&\quad + \sum_{n\in P\cap Q} \left(\operatorname{E}^2\left[w_n^2\right] + 4\bar{w}_n^2\Omega_{n,n} + 2\Omega_{n,n}^2\right)\omega_n \quad \text{since } \boldsymbol{w} \text{ Gaussian} && (7.43)\\
&= \sum_{n\in P, m\in Q} \left(4\bar{w}_n\bar{w}_m\Omega_{n,m} + 2\Omega_{n,m}^2\right)\bar{\varepsilon}_n^2\bar{\varepsilon}_m^2 \\
&\quad + \sum_{n\in P\cap Q} \left(\left[\bar{w}_n^2 + \Omega_{n,n}\right]^2 + 4\bar{w}_n^2\Omega_{n,n} + 2\Omega_{n,n}^2\right)\omega_n && (7.44)\\
&= \sum_{n\in P, m\in Q} \left(4\bar{w}_n\bar{w}_m\Omega_{n,m} + 2\Omega_{n,m}^2\right)\bar{\varepsilon}_n^2\bar{\varepsilon}_m^2 \\
&\quad + \sum_{n\in P\cap Q} \left(\bar{w}_n^4 + 6\bar{w}_n^2\Omega_{n,n} + 3\Omega_{n,n}^2\right)\omega_n && (7.45)
\end{aligned}
$$

□

**Proposition 7.4.** $\operatorname{cov}\left[v_{p,i}, v_{q,\varepsilon^2}\right]$

*If* $\boldsymbol{w}$ *is Gaussian,*

$$\operatorname{cov}\left[v_{p,i}, v_{q,\varepsilon^2}\right] = c + \sum_j \left(\bar{F}_{i,j}\,\bar{\gamma} + h_{i,j}\right) \operatorname{cov}\left[e_i^p e_j, \sum_{n\in Q} w_n^2 \varepsilon_n^2\right] \tag{7.46}$$

$$\approx c$$

$$\text{where } c = \frac{\operatorname{E}\left[v_{q,\varepsilon^2}\right]}{\bar{\gamma}} \sum_j h_{i,j}\left(\bar{e}_i^p \bar{e}_j + \operatorname{cov}\left[e_i^p, e_j\right]\right) \tag{7.47}$$

$$\text{and } \operatorname{cov}\left[e_i^p e_j, \sum_{n\in Q} w_n^2 \varepsilon_n^2\right] = 2 \sum_{m\in P, n\in Q, s} \left(\bar{E}_{m,i}\bar{E}_{s,j} + X_{m,i,s,j}\right) \bar{\varepsilon}_n^2$$

$$\left[\bar{w}_m \bar{w}_n \Omega_{s,n} + \bar{w}_s \bar{w}_n \Omega_{m,n} + \Omega_{m,n}\Omega_{s,n}\right] \tag{7.48}$$

*Computation is* $O(K)$ *for approximate covariance* $c$ *and* $O(KN^3)$ *for exact* $\operatorname{cov}\left[v_{p,i}, v_{q,\varepsilon^2}\right]$ *given* $\bar{\boldsymbol{e}}^p, \bar{\boldsymbol{e}}, \operatorname{cov}\left[\boldsymbol{e}^p, \boldsymbol{e}\right]$*, and* $\operatorname{E}\left[v_{q,\varepsilon^2}\right]$.

*The approximation ignores price movement's linking factor exposures and idiosyncratic risk. The effect is secondary and, in a model and one built from estimates, seems of needless precision. The result is what would be if idiosyncratic risk arose from an independent set of portfolio weights following the same distribution, i.e., if* $\sigma^2 = \boldsymbol{w}^T \boldsymbol{E}\boldsymbol{F}\boldsymbol{E}^T \boldsymbol{w} + \gamma \boldsymbol{z}^T \operatorname{diag}\left[\boldsymbol{\varepsilon}^2\right] \boldsymbol{z}$ *where* $\boldsymbol{z}$ *ind of* $\boldsymbol{w}$*,* $\operatorname{E}\left[\boldsymbol{z}\right] = \operatorname{E}\left[\boldsymbol{w}\right]$*,* $\operatorname{var}\left[\boldsymbol{z}\right] = \operatorname{var}\left[\boldsymbol{w}\right]$*. Thus, the matrix of contributions' approximated covariance is positive semidefinite.*

*Proof.*

$$\operatorname{cov}\left[v_{p,i}, v_{q,\varepsilon^2}\right] = \operatorname{cov}\left[e_i^p \sum_j F_{i,j} e_j, \gamma \sum_{n\in Q} w_n^2 \varepsilon_n^2\right] \tag{7.49}$$

$$= \sum_{n\in Q,j} \operatorname{cov}\left[e_i^p F_{i,j} e_j, \gamma w_n^2 \varepsilon_n^2\right] \tag{7.50}$$

$$= \sum_j \operatorname{cov}\left[F_{i,j}, \gamma\right] \operatorname{E}\left[e_i^p e_j\right] \operatorname{E}\left[\sum_{n\in Q} w_n^2 \varepsilon_n^2\right] + \left(\bar{F}_{i,j}\,\bar{\gamma} + \operatorname{cov}\left[F_{i,j}, \gamma\right]\right) \operatorname{cov}\left[e_i^p e_j, \sum_{n\in Q} w_n^2 \varepsilon_n^2\right] \tag{7.51}$$

since $\{\boldsymbol{F}, \gamma\}$ ind of $\{\boldsymbol{e}, \boldsymbol{w}, \boldsymbol{\varepsilon}^2\}$

$$= \sum_j h_{i,j} \left(\bar{e}_i^p \bar{e}_j + \operatorname{cov}\left[e_i^p, e_j\right]\right) \frac{\operatorname{E}\left[v_{q,\varepsilon^2}\right]}{\bar{\gamma}} + \left(\bar{F}_{i,j}\,\bar{\gamma} + h_{i,j}\right) \underbrace{\operatorname{cov}\left[e_i^p e_j, \sum_{n\in Q} w_n^2 \varepsilon_n^2\right]}_{①} \tag{7.52}$$

$$① = \operatorname{cov}\left[e_i^p e_j, \sum_{n\in Q} w_n^2 \varepsilon_n^2\right]$$

$$= \operatorname{cov}\left[\sum_{m\in P,s} E_{m,i} w_m E_{s,j} w_s, \sum_{n\in Q} w_n^2 \varepsilon_n^2\right]$$

$$= \sum_{m\in P,n\in Q,s} \operatorname{cov}\left[w_m w_s E_{m,i} E_{s,j},\ w_n^2 \varepsilon_n^2\right]$$

$$= \sum_{m\in P,n\in Q,s} \operatorname{E}\left[E_{m,i} E_{s,j}\right] \operatorname{E}\left[\varepsilon_n^2\right] \operatorname{cov}\left[w_m w_s, w_n^2\right] \tag{7.53}$$

since $\boldsymbol{E}$, $\boldsymbol{w}$, $\boldsymbol{\varepsilon}^2$ ind

$$= \sum_{m\in P,n\in Q,s} \left(\bar{E}_{m,i} \bar{E}_{s,j} + X_{m,i,s,j}\right) \bar{\varepsilon}_n^2 \underbrace{\operatorname{cov}\left[w_m w_s, w_n^2\right]}_{②}$$

$$② = 2\left(\bar{w}_m \bar{w}_n \Omega_{s,n} + \bar{w}_s \bar{w}_n \Omega_{m,n} + \Omega_{m,n} \Omega_{s,n}\right) \tag{7.54}$$

since $\boldsymbol{w}$ Gaussian

□

### 7.5.3 Computational complexity

If the portfolio is divided into $M$ subportfolios, the mean of the $M \times (K+1)$ parts computes in $O(MK^2N^2)$ broken out in table 7.1. For variance, generally only the

Table 7.1: Computational complexity of reporting contributions' mean

| Quantity | Complexity |
|---|---|
| $\bar{\boldsymbol{e}}, \bar{\boldsymbol{e}}^p$ | $M \times O(KN)$ |
| $\mathrm{cov}\left[\boldsymbol{e}^p, \boldsymbol{e}\right]$ | $M \times O(K^2N^2)$ |
| $\mathrm{E}\left[\boldsymbol{v}_p\right]$ | $M \times O(K^2)$ |
| $\mathrm{E}\left[v_{p,\varepsilon^2}\right]$ | $M \times O(N)$ |
| | $O(MK^2N^2)$ |

range around a reported number is of interest; covariance with others is irrelevant with the exception of total portfolio variance, needed to convert numbers to standard deviation. Together, these compute in $O(MK^2N^2 + MK^4)$ broken out in table 7.2.

Table 7.2: Computational complexity of reporting contributions' variance

| Quantity | Complexity |
|---|---|
| $\mathrm{var}\left[\boldsymbol{e}^p\right], \mathrm{var}\left[\boldsymbol{e}\right]$ | $M \times O(K^2N^2)$ |
| $\mathrm{var}\left[\boldsymbol{v}_p\right], \mathrm{var}\left[\boldsymbol{v}_{portfolio}\right]$ | $M \times O(K^4)$ |
| $\mathrm{cov}\left[\boldsymbol{v}_p, \boldsymbol{v}_{portfolio}\right]$ | $M \times O(K^4)$ |
| $\mathrm{var}\left[v_{p,\varepsilon^2}\right], \mathrm{var}\left[v_{portfolio,\varepsilon^2}\right]$ | $M \times O(N^2)$ |
| $\mathrm{cov}\left[v_{p,\varepsilon^2}, v_{portfolio,\varepsilon^2}\right]$ | $M \times O(N^2)$ |
| approx $\mathrm{cov}\left[v_p, v_{p,\varepsilon^2}\right], \mathrm{cov}\left[v_{portfolio}, v_{portfolio,\varepsilon^2}\right]$ | $M \times O(K^2)$ |
| approx $\mathrm{cov}\left[v_p, v_{portfolio,\varepsilon^2}\right], \mathrm{cov}\left[v_{portfolio}, v_{p,\varepsilon^2}\right]$ | $M \times O(K^2)$ |
| | $O(MK^2N^2 + MK^4)$ |

## 7.6 Uncertain standard deviation contributions

Volatility is more tangible in standard deviation than variance, e.g., a portfolio standard deviation of 15% vs. variance of $225\%^2$. Can uncertain standard deviation

contributions be computed? Equation (7.5), $s_i = v_i/\sigma$, converts variance contributions to standard deviation. However, nonlinearity makes the mean and covariance of $\boldsymbol{s}$, unlike $\boldsymbol{v}$, unavailable.

The usual estimate of center would be $\mathrm{E}[\boldsymbol{s}] = \mathrm{E}\left[\boldsymbol{v}/\sqrt{\sigma^2}\right]$. Instead, a notional center of

$$\mathrm{ctr}[\boldsymbol{s}] = \frac{\mathrm{E}[\boldsymbol{v}]}{\sqrt{\mathrm{E}[\sigma^2]}} \tag{7.55}$$

agrees in the absence of uncertainty and properly sums to a notional center of portfolio volatility, $\sum_i \mathrm{ctr}[s_i] = \sum_i \mathrm{E}[v_i]/\sqrt{\mathrm{E}[\sigma^2]} = \sqrt{\mathrm{E}[\sigma^2]}$.

Approximate covariance comes from linearizing sensitivity to variance contributions and taking the variance of the linearization.

**Proposition 7.5.** *Approximate* $\mathrm{var}[\boldsymbol{s}]$

*If contributions to variance* $\boldsymbol{v}$ *have mean* $\mathrm{E}[\boldsymbol{v}]$ *and covariance* $\mathrm{var}[\boldsymbol{v}]$*, contributions to standard deviation* $\boldsymbol{s}$ *have approximate covariance*

$$\begin{aligned}\mathrm{var}[\boldsymbol{s}] \approx \frac{\mathrm{var}[\boldsymbol{v}]}{\mathrm{E}[\sigma^2]} - \frac{\mathrm{E}[\boldsymbol{v}]\,\mathrm{cov}[\sigma^2, \boldsymbol{v}] + \mathrm{cov}[\boldsymbol{v}, \sigma^2]\,\mathrm{E}[\boldsymbol{v}]^T}{2\left(\mathrm{E}[\sigma^2]\right)^2} \\ + \frac{\mathrm{var}[\sigma^2]}{4\mathrm{E}[\sigma^2]} \frac{\mathrm{E}[\boldsymbol{v}]\,\mathrm{E}[\boldsymbol{v}]^T}{\left(\mathrm{E}[\sigma^2]\right)^2}\end{aligned} \tag{7.56}$$

$$\textit{So} \quad \mathrm{var}[s_k] \approx \frac{\mathrm{var}[v_k]}{\mathrm{E}[\sigma^2]} - \frac{\mathrm{cov}[v_k, \sigma^2]\,\mathrm{E}[v_k]}{\mathrm{E}^2[\sigma^2]} + \frac{\mathrm{var}[\sigma^2]\,\mathrm{E}^2[v_k]}{4\mathrm{E}^3[\sigma^2]} \tag{7.57}$$

$$\mathrm{var}[\sigma] \approx \frac{\mathrm{var}[\sigma^2]}{4\,\mathrm{E}[\sigma^2]} \tag{7.58}$$

$$\textit{where} \quad \mathrm{E}\left[\sigma^2\right] = \mathbf{1}^T \mathrm{E}[\boldsymbol{v}] \tag{7.59}$$

$$\mathrm{cov}\left[\boldsymbol{v}, \sigma^2\right] = \mathrm{var}[\boldsymbol{v}]\,\mathbf{1} \tag{7.60}$$

$$\mathrm{var}\left[\sigma^2\right] = \mathbf{1}^T \mathrm{var}[\boldsymbol{v}]\,\mathbf{1} \tag{7.61}$$

*Proof.*

First note that portfolio variance is the sum of the variance contributions, $\sigma^2 = \mathbf{1}^{\mathrm{T}}\boldsymbol{v}$. So, $\mathrm{E}[\sigma^2] = \mathbf{1}^{\mathrm{T}}\mathrm{E}[\boldsymbol{v}]$, $\mathrm{var}[\sigma^2] = \mathbf{1}^{\mathrm{T}}\mathrm{var}[\boldsymbol{v}]\,\mathbf{1}$, and $\mathrm{cov}[\boldsymbol{v}, \sigma^2] = \mathrm{var}[\boldsymbol{v}]\,\mathbf{1}$.

For the covariance of $\boldsymbol{s}$, linearize around the mean of $\boldsymbol{v}$ and $\sigma^2$.

$$\boldsymbol{s} = \boldsymbol{v}/\sqrt{\sigma^2} \tag{7.62}$$

$$s_i \approx \text{constant} + \sum_j \frac{\partial s_i}{\partial v_j}\left(v_j - \mathrm{E}\left[v_j\right]\right) + \frac{\partial s_i}{\partial \sigma^2}\left(\sigma^2 - \mathrm{E}\left[\sigma^2\right]\right) \tag{7.63}$$

$$\frac{\partial s_i}{\partial v_j} = \begin{cases} \frac{1}{\sqrt{\mathrm{E}[\sigma^2]}} & j = i \\ 0 & j \neq i \end{cases} \tag{7.64}$$

$$\frac{\partial s_i}{\partial \sigma^2} = -\frac{\mathrm{E}\left[v_i\right]}{2\left(\mathrm{E}\left[\sigma^2\right]\right)^{3/2}} \tag{7.65}$$

Approximate the covariance as the covariance of the linearization.

$$\begin{aligned} \boldsymbol{J} &= \text{Jacobian of } \begin{bmatrix} \boldsymbol{s} \\ \sigma \end{bmatrix} \text{ to } \boldsymbol{v}, \sigma^2 \\ &= \frac{1}{\sqrt{\mathrm{E}\left[\sigma^2\right]}} \begin{pmatrix} 1 & 0 & \cdots & 0 & -\frac{\mathrm{E}[v_1]}{2\mathrm{E}[\sigma^2]} \\ 0 & 1 & \ddots & \vdots & -\frac{\mathrm{E}[v_2]}{2\mathrm{E}[\sigma^2]} \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & -\frac{\mathrm{E}[v_K]}{2\mathrm{E}[\sigma^2]} \\ 0 & \cdots & \cdots & 0 & 1/2 \end{pmatrix} \end{aligned} \tag{7.66}$$

$$= \frac{1}{\sqrt{\mathrm{E}\left[\sigma^2\right]}} \begin{pmatrix} \boldsymbol{I} & -\frac{1}{2\mathrm{E}[\sigma^2]}\mathrm{E}\left[\boldsymbol{v}\right] \\ \boldsymbol{0} & 1/2 \end{pmatrix} \tag{7.67}$$

$$\operatorname{var}\begin{bmatrix} \boldsymbol{s} \\ \sigma \end{bmatrix} \approx \boldsymbol{J} \operatorname{var}\begin{bmatrix} \boldsymbol{v} \\ \sigma^2 \end{bmatrix} \boldsymbol{J}^{\mathrm{T}} \tag{7.68}$$

$$= \boldsymbol{J} \begin{pmatrix} \operatorname{var}\left[\boldsymbol{v}\right] & \operatorname{cov}\left[\boldsymbol{v}, \sigma^2\right] \\ \operatorname{cov}\left[\sigma^2, \boldsymbol{v}\right] & \operatorname{var}\left[\sigma^2\right] \end{pmatrix} \boldsymbol{J}^{\mathrm{T}} \tag{7.69}$$

Note: this correctly gives $\operatorname{var}\left[\sum_i s_i\right] = \operatorname{var}\left[\sigma\right]$.

$$\text{So var}\left[\boldsymbol{s}\right] \approx \frac{\text{var}\left[\boldsymbol{v}\right]}{\text{E}\left[\sigma^2\right]} - \frac{\text{E}\left[\boldsymbol{v}\right]\text{cov}\left[\sigma^2,\boldsymbol{v}\right] + \text{cov}\left[\boldsymbol{v},\sigma^2\right]\text{E}\left[\boldsymbol{v}\right]^{\text{T}}}{2\left(\text{E}\left[\sigma^2\right]\right)^2} + \frac{\text{var}\left[\sigma^2\right]}{4\text{E}\left[\sigma^2\right]}\frac{\text{E}\left[\boldsymbol{v}\right]\text{E}\left[\boldsymbol{v}\right]^{\text{T}}}{\left(\text{E}\left[\sigma^2\right]\right)^2} \tag{7.70}$$

$$\text{cov}\left[\boldsymbol{s},\sigma\right] \approx \frac{\text{cov}\left[\boldsymbol{v},\sigma^2\right]}{2\,\text{E}\left[\sigma^2\right]} - \frac{\text{var}\left[\sigma^2\right]}{4\text{E}\left[\sigma^2\right]}\frac{\text{E}\left[\boldsymbol{v}\right]}{\text{E}\left[\sigma^2\right]} \tag{7.71}$$

$$\text{var}\left[\sigma\right] \approx \frac{\text{var}\left[\sigma^2\right]}{4\,\text{E}\left[\sigma^2\right]} \tag{7.72}$$

□

## 7.7 Modeling weight dynamics

The perturbation on portfolio weights is specified by their covariance over the horizon of interest. Not normally at hand, it can be derived from a model of return covariance, initial holdings, and if a benchmark is involved, benchmark weights and whether it is fixed weight or floating.

The return model enlisted is likely of shorter horizon than the risk model. For example, one might be interested in how the annual risk of the portfolio can vary until the next rebalance period. In that case, the return model's horizon is ideally the time until rebalancing.

**Proposition 7.6.** *Future weight covariance* $\boldsymbol{\Omega}$

$$\boldsymbol{\Omega} \approx \boldsymbol{J}\,\text{var}\left[\boldsymbol{r}\right]\,\boldsymbol{J}^T \tag{7.73}$$

*where*

$$\boldsymbol{J} = \begin{cases} -\left(\bar{\boldsymbol{w}}+\boldsymbol{b}\right)\left(\bar{\boldsymbol{w}}+\boldsymbol{b}\right)^T + \text{diag}\left[\bar{\boldsymbol{w}}+\boldsymbol{b}\right] & \textit{fixed benchmark} \\ -\left(\bar{\boldsymbol{w}}+\boldsymbol{b}\right)\left(\bar{\boldsymbol{w}}+\boldsymbol{b}\right)^T + \text{diag}\left[\bar{\boldsymbol{w}}\right] + \boldsymbol{b}\boldsymbol{b}^T & \textit{floating benchmark} \end{cases} \tag{7.74}$$

$\bar{\boldsymbol{w}} = (n \times 1)$ *current benchmark-relative weights*

$\boldsymbol{b} = (n \times 1)$ *benchmark weights*

$\boldsymbol{r} = (n \times 1)$ *future returns*

*If there isn't a benchmark,* $\boldsymbol{b} = \mathbf{0}$*; otherwise,* $\boldsymbol{b}$ *sums to 1. In both cases,* $\bar{\boldsymbol{w}} + \boldsymbol{b}$ *sums to 1.*

*Proof.*

Let $\boldsymbol{w} = (n \times 1)$ future benchmark-relative weights.

With a fixed weight benchmark:

$$w_i = \frac{(1 + r_i)(\bar{w}_i + b_i)}{\sum_k (1 + r_k)(\bar{w}_k + b_k)} - b_i \tag{7.75}$$

$$= \alpha_i - b_i \tag{7.76}$$

With a floating benchmark:

$$w_i = \frac{(1 + r_i)(\bar{w}_i + b_i)}{\sum_k (1 + r_k)(\bar{w}_k + b_k)} - \frac{(1 + r_i) b_i}{\sum_k (1 + r_k) b_k} \tag{7.77}$$

$$= \alpha_i - \beta_i \tag{7.78}$$

$$\text{where } \alpha_i = \frac{(1 + r_i)(\bar{w}_i + b_i)}{\sum_k (1 + r_k)(\bar{w}_k + b_k)} \tag{7.79}$$

$$\beta_i = \frac{(1 + r_i) b_i}{\sum_k (1 + r_k) b_k} \tag{7.80}$$

The approximate weight covariance is the covariance of linearization around the mean return.

$$\begin{aligned} \boldsymbol{\Omega} &= \operatorname{var}[\boldsymbol{w}] \\ &\approx \boldsymbol{J} \operatorname{var}[\boldsymbol{r}] \, \boldsymbol{J}^{\mathrm{T}} \end{aligned} \tag{7.81}$$

$$\text{where } \boldsymbol{J} = \text{Jacobian of } \boldsymbol{w} \text{ to } \boldsymbol{r} \text{ evaluated at } \boldsymbol{r} = \bar{\boldsymbol{r}}$$

$$J_{i,j} = \frac{\partial w_i}{\partial r_j} = \begin{cases} \frac{\partial \alpha_i}{\partial r_j} & \text{fixed benchmark} \\ \frac{\partial \alpha_i}{\partial r_j} - \frac{\partial \beta_i}{\partial r_j} & \text{floating benchmark} \end{cases} \tag{7.82}$$

$$\frac{\partial \alpha_i}{\partial r_j} = -\left(\bar{w}_j + b_j\right) \frac{\left(1 + r_i\right)\left(\bar{w}_i + b_i\right)}{\left[\sum_k \left(1 + r_k\right)\left(\bar{w}_k + b_k\right)\right]^2} + \ 1_{\{i=j\}} \frac{\left(\bar{w}_i + b_i\right)}{\left[\sum_k \left(1 + r_k\right)\left(\bar{w}_k + b_k\right)\right]^2} \quad (7.83)$$

$$= -\left(\bar{w}_j + b_j\right)\left(\bar{w}_i + b_i\right) + \ 1_{\{i=j\}}\left(\bar{w}_i + b_i\right) \quad (7.84)$$

evaluated at $\boldsymbol{r} = \bar{\boldsymbol{r}} = \mathbf{0}$[9] since $\sum_k \bar{w}_k + b_k = 1$

Likewise,

$$\frac{\partial \beta_i}{\partial r_j} = -b_j b_i + \ 1_{\{i=j\}} b_i \quad (7.85)$$

$$\boldsymbol{J} = \begin{cases} -\left(\bar{\boldsymbol{w}} + \boldsymbol{b}\right)\left(\bar{\boldsymbol{w}} + \boldsymbol{b}\right)^{\mathrm{T}} + \ \mathrm{diag}\left[\bar{\boldsymbol{w}} + \boldsymbol{b}\right] & \text{fixed benchmark} \\ -\left(\bar{\boldsymbol{w}} + \boldsymbol{b}\right)\left(\bar{\boldsymbol{w}} + \boldsymbol{b}\right)^{\mathrm{T}} + \ \mathrm{diag}\left[\bar{\boldsymbol{w}} + \boldsymbol{b}\right] + \boldsymbol{b}\boldsymbol{b}^{\mathrm{T}} - \ \mathrm{diag}\left[\boldsymbol{b}\right] & \text{floating benchmark} \end{cases} \quad (7.86)$$

□

## 7.8 An example

Data are 106 periods of weekly Wed–Wed returns from Wed Dec 20, 2017 through Tue Dec 31, 2019. The first 80 periods, 12/20/2017–7/3/2019, are the training set. The final 26 periods, 7/3/2019–12/31/2019, are the testing set. The universe consists of the 474 securities in the S&P 500 as of Dec 20, 2017 with complete series. Each security belongs to one of eleven sectors. ETFs—IWV-iShares Russell 3000, IWB-iShares Russell 1000, IWM-iShares Russell 2000, IWF-iShares Russell 1000 Growth, IWD-iShares Russell 1000 Value—are used in the risk factors.

An uncertain covariance model is constructed from the training set as follows:

1. Three factors are from ETF returns: market = $r_{3000}$, size = $r_{1000} - r_{2000}$, and growth-value = $r_{1000G} - r_{1000V}$. Inferred on security returns stripped of these,

[9] Omitted here for clarity, considering drift ($\bar{\boldsymbol{r}} \neq \mathbf{0}$) adds scalars.

four principal components are added for a total of 7 factors. Their mean covariance $\bar{\boldsymbol{F}}$ is the sample covariance of the series.

2. Exposures $\boldsymbol{E}$ and idiosyncratic variances $\boldsymbol{\varepsilon}^2$ are first point-estimated as the coefficients and variance of the residual in OLS regression of security returns against factors. Using the universe median and variance of these as a prior—exposures Gaussian, idiosyncratic variance inverse gamma—they are re-estimated with mean and variance via Bayesian regression, yielding $\bar{\boldsymbol{E}}, \boldsymbol{X}, \bar{\boldsymbol{\varepsilon}}^2, \boldsymbol{\omega}$. ($\boldsymbol{X}$ assumes no exposure covariance between securities.)

3. For simplicity, $\boldsymbol{H}$, the variance of factor covariance $\boldsymbol{F}$, is taken as the standard sampling error.[10] Note that this underestimates since the real uncertainty isn't from sample size but changes in state. For example, $\text{VIX}^2$, a good proxy for $F_{1,1}$, observed weekly over the same training period has variance $275\times$ $H_{1,1,1,1}$.

4. A measure of the idiosyncratic variance scalar $\gamma$ is the cross-sectional median of each period's squared OLS regression residuals, divided by the mean across time for normalization. $\gamma$'s variance $\eta$ is the variance of this series; its mean $\bar{\gamma}$ is 1.

5. $\boldsymbol{h}$, the covariance between $\boldsymbol{F}$ and $\gamma$, is for simplicity assumed to be zero.

6. To report in annualized variance, $\boldsymbol{F}$ and $\bar{\gamma}$ are multiplied by 52, and $\boldsymbol{H}$, $\eta$, and $\boldsymbol{h}$ by $52^2$.

Portfolio weight covariance $\boldsymbol{\Omega}$ is computed by applying the method in section 7.7 to the horizon-scaled sample covariance of the training set.

### 7.8.1 Portfolio 1: Diversified long, cash benchmark

The 472 securities are equal-weighted (0.21%) and analyzed with a 26 week horizon and no benchmark. Subportfolios are holdings in each sector.

Table 7.3 shows the mean and standard deviation of factor exposures contributed by sector. Standard deviations are small because the portfolio is diversified and exposure uncertainty was modeled as uncorrelated across securities.

[10] For the usual unbiased variance estimate of multivariate Gaussian $(a, b, c, d)$ from $N$ points, $\operatorname{cov}\left[\hat{\sigma}_{ab}, \hat{\sigma}_{cd}\right] = \left(\sigma_{a,c}\,\sigma_{b,d} + \sigma_{a,d}\,\sigma_{b,c}\right)/(N-1)$.

Table 7.3: Diversified long—mean and SD of factor exposures contributed by sector

| | | | Factor | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | N | wt % | mkt | size | gv | pc1 | pc2 | pc3 | pc4 |
| All | 474 | 100 | 1.059 | -0.098 | -0.333 | 0.326 | -0.301 | 0.190 | 0.080 |
| ± | | | 0.023 | 0.015 | 0.023 | 0.050 | 0.040 | 0.029 | 0.033 |
| Comm Svcs | 5 | 1.1 | 0.009 | -0.002 | -0.005 | 0.003 | -0.009 | 0.002 | 0.002 |
| | | | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| Cons Discr | 76 | 16.0 | 0.174 | -0.036 | -0.038 | 0.125 | -0.087 | 0.042 | 0.065 |
| | | | 0.010 | 0.006 | 0.007 | 0.012 | 0.011 | 0.010 | 0.012 |
| Cons Staples | 32 | 6.8 | 0.045 | 0.005 | -0.032 | -0.032 | -0.061 | 0.033 | 0.010 |
| | | | 0.004 | 0.003 | 0.005 | 0.006 | 0.007 | 0.006 | 0.005 |
| Energy | 28 | 5.9 | 0.080 | -0.023 | -0.048 | 0.007 | 0.070 | 0.006 | 0.106 |
| | | | 0.011 | 0.005 | 0.007 | 0.005 | 0.011 | 0.005 | 0.016 |
| Financials | 65 | 13.7 | 0.166 | -0.007 | -0.106 | 0.110 | -0.028 | -0.073 | -0.058 |
| | | | 0.011 | 0.004 | 0.009 | 0.010 | 0.005 | 0.008 | 0.006 |
| Healthcare | 57 | 12.0 | 0.129 | -0.019 | -0.020 | -0.004 | -0.057 | -0.013 | 0.029 |
| | | | 0.009 | 0.004 | 0.006 | 0.006 | 0.008 | 0.007 | 0.008 |
| Industrials | 66 | 13.9 | 0.175 | -0.001 | -0.061 | 0.077 | 0.018 | 0.026 | -0.056 |
| | | | 0.009 | 0.004 | 0.006 | 0.009 | 0.007 | 0.007 | 0.008 |
| Tech | 64 | 13.5 | 0.164 | -0.005 | 0.048 | 0.108 | -0.016 | 0.035 | -0.023 |
| | | | 0.013 | 0.004 | 0.007 | 0.013 | 0.008 | 0.009 | 0.008 |
| Materials | 22 | 4.6 | 0.059 | -0.003 | -0.023 | 0.037 | -0.003 | 0.011 | -0.006 |
| | | | 0.004 | 0.002 | 0.003 | 0.004 | 0.004 | 0.004 | 0.004 |
| Real Estate | 32 | 6.8 | 0.037 | -0.013 | -0.020 | -0.034 | -0.085 | 0.064 | 0.004 |
| | | | 0.004 | 0.003 | 0.003 | 0.005 | 0.010 | 0.008 | 0.004 |
| Utilities | 27 | 5.7 | 0.021 | 0.007 | -0.030 | -0.072 | -0.044 | 0.056 | 0.008 |
| | | | 0.003 | 0.002 | 0.005 | 0.011 | 0.008 | 0.009 | 0.004 |

Table 7.4 shows risk decomposed by factor conventionally and with uncertainty. Conventional uses the point estimate model left after discarding uncertainty parameters:

$$C_{conventional} = \bar{\boldsymbol{E}}\bar{\boldsymbol{F}}\bar{\boldsymbol{E}}^{\mathrm{T}} + \bar{\gamma}\ \operatorname{diag}\left(\bar{\boldsymbol{\varepsilon}}^2\right) \tag{7.87}$$

Notice that uncertain contributions come with an approximate standard deviation, e.g., portfolio variance of $14.47 \pm 1.20$ vs. the conventional's point estimate of 14.45. The table's final 4 columns split the uncertain's center into sources—means and different uncertainties as described by equations (7.21) and (7.22).

Table 7.5 shows ex-ante modeled and realized risk decomposed by sector. Realized

Table 7.4: Diversified long—conventional and uncertain risk contributions by factor

| | | Uncertain | | | | | |
|---|---|---|---|---|---|---|---|
| | Conventional | center | ± | means | $\boldsymbol{w}$ unc | $\boldsymbol{E}$ unc | $\boldsymbol{w}, \boldsymbol{E}$ unc |
| All | 14.45 | 14.47 | 1.20 | 14.44 | 0.03 | 0 | 0 |
| mkt | 14.68 | 14.67 | 2.39 | 14.66 | 0.01 | 0 | 0 |
| size | 0.24 | 0.24 | 0.10 | 0.24 | 0.00 | 0 | 0 |
| gv | -1.11 | -1.11 | 0.34 | -1.11 | 0.00 | 0 | 0 |
| pc1 | 0.30 | 0.31 | 0.26 | 0.30 | 0.01 | 0 | 0 |
| pc2 | 0.20 | 0.21 | 0.20 | 0.20 | 0.00 | 0 | 0 |
| pc3 | 0.07 | 0.07 | 0.11 | 0.07 | 0.00 | 0 | 0 |
| pc4 | 0.01 | 0.01 | 0.05 | 0.01 | 0.00 | 0 | 0 |
| $\varepsilon^2$ | 0.06 | 0.07 | 0.02 | 0.06 | 0.00 | | |

contributions are measured with two sets of portfolio weights: $\boldsymbol{w}_0$-as constructed, $\boldsymbol{w}_T$-at the horizon (26 weeks in the future) adjusted for price movement. Realized covariance is the sample covariance of returns over the test set (the same 26 weeks).

Table 7.5: Diversified long—ex-ante modeled and realized risk contributions by sector

| | | Uncertain | | Realized | |
|---|---|---|---|---|---|
| | Conventional | center | ± | $\boldsymbol{w}_0$ | $\boldsymbol{w}_T$ |
| All | 14.45 | 14.47 | 1.20 | 13.84 | 13.78 |
| Comm Svcs | 0.13 | 0.13 | 0.02 | 0.13 | 0.12 |
| Cons Discr | 2.59 | 2.60 | 0.17 | 2.76 | 2.77 |
| Cons Staples | 0.54 | 0.54 | 0.07 | 0.30 | 0.30 |
| Energy | 0.99 | 0.99 | 0.16 | 1.61 | 1.41 |
| Financials | 2.06 | 2.06 | 0.17 | 2.43 | 2.45 |
| Healthcare | 1.81 | 1.81 | 0.13 | 1.30 | 1.31 |
| Industrials | 2.29 | 2.29 | 0.14 | 2.44 | 2.44 |
| Tech | 2.58 | 2.58 | 0.23 | 1.95 | 2.08 |
| Materials | 0.79 | 0.79 | 0.06 | 0.70 | 0.69 |
| Real Estate | 0.53 | 0.53 | 0.09 | 0.13 | 0.13 |
| Utilities | 0.16 | 0.16 | 0.08 | 0.09 | 0.09 |

### 7.8.2 Portfolio 2: 44 stock long-short, cash benchmark

A 44 stock long-short portfolio is constructed from the the top 4 securities alphabetically by name in each sector. The first two are long; the remaining two are short.

Weights come from minimizing variance as measured by the conventional factor model of equation (7.87) with longs constrained positive and summing to 1 and shorts negative and summing to -1. The horizon is again 26 weeks for modeled future weight covariance $\boldsymbol{\Omega}$ and realized weights $\boldsymbol{w}_T$.

Table 7.6 shows factor exposures by side. Tables 7.7, 7.8, and 7.9 are as the preceding tables 7.3, 7.4, and 7.5.

Table 7.6: 44 stock long-short—mean and SD of factor exposures contributed by side

| | | | Factor | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | N | wt % | mkt | size | gv | pc1 | pc2 | pc3 | pc4 |
| All | 44 | 0 | 0.008 | 0.016 | 0.003 | 0.014 | 0.061 | -0.005 | 0.022 |
| ± | | | 0.062 | 0.068 | 0.079 | 0.093 | 0.099 | 0.106 | 0.110 |
| Long | 22 | 100 | 0.851 | 0.031 | -0.286 | -0.105 | -0.570 | 0.351 | -0.088 |
| | | | 0.127 | 0.048 | 0.066 | 0.085 | 0.096 | 0.087 | 0.080 |
| Short | 22 | -100 | -0.843 | -0.015 | 0.289 | 0.119 | 0.631 | -0.356 | 0.110 |
| | | | 0.146 | 0.051 | 0.070 | 0.112 | 0.107 | 0.088 | 0.087 |

In this less diversified portfolio, uncertainty has a measurable effect on contributions' center—reported in the last four columns of table 7.8—accounting for 14% ($\frac{0.37+0.34+0.02}{5.22}$) of portfolio volatility.

## 7.9 Summary

Risk contribution information enters throughout investment management. Because what's measured is behavior in the direction of net risk, the portfolio's actual risks, viewed statically, can be hidden by hedging alignments. Looking instead with parameters and portfolio weights modeled as random evaluates the system across the distribution of what's possible. Latent fragility surfaces, and contributions go from a point estimate to having a center and range, offering robust and realistic assessments of risk.

Table 7.7: 44 stock long-short—mean and SD of factor exposures contributed by sector

| | | | Factor | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | N | wt % | mkt | size | gv | pc1 | pc2 | pc3 | pc4 |
| All | 44 | 0.0 | 0.008 | 0.016 | 0.003 | 0.014 | 0.061 | -0.005 | 0.022 |
| ± | | | 0.062 | 0.068 | 0.079 | 0.093 | 0.099 | 0.106 | 0.110 |
| Comm Svcs | 4 | 3.0 | 0.021 | 0.001 | -0.059 | -0.015 | -0.038 | -0.029 | 0.006 |
| | | | 0.017 | 0.015 | 0.020 | 0.022 | 0.024 | 0.025 | 0.025 |
| Cons Discr | 4 | -1.2 | -0.020 | -0.011 | 0.055 | -0.041 | -0.032 | -0.026 | 0.018 |
| | | | 0.020 | 0.017 | 0.024 | 0.027 | 0.026 | 0.027 | 0.028 |
| Cons Staples | 4 | 3.9 | 0.031 | -0.020 | -0.031 | 0.001 | -0.007 | 0.047 | 0.058 |
| | | | 0.015 | 0.017 | 0.021 | 0.022 | 0.025 | 0.030 | 0.028 |
| Energy | 4 | -2.0 | -0.017 | 0.013 | 0.011 | -0.000 | 0.027 | 0.016 | -0.039 |
| | | | 0.014 | 0.012 | 0.014 | 0.013 | 0.018 | 0.017 | 0.028 |
| Financials | 4 | -6.4 | -0.065 | -0.011 | 0.005 | -0.046 | 0.082 | -0.069 | 0.007 |
| | | | 0.026 | 0.025 | 0.030 | 0.031 | 0.037 | 0.039 | 0.041 |
| Healthcare | 4 | 1.2 | -0.015 | 0.025 | 0.014 | -0.028 | -0.026 | -0.019 | -0.048 |
| | | | 0.021 | 0.018 | 0.019 | 0.023 | 0.025 | 0.027 | 0.030 |
| Industrials | 4 | 1.5 | 0.027 | 0.017 | -0.003 | 0.011 | 0.054 | 0.024 | 0.026 |
| | | | 0.017 | 0.013 | 0.015 | 0.020 | 0.022 | 0.021 | 0.023 |
| Tech | 4 | 6.0 | 0.070 | -0.007 | -0.016 | 0.011 | -0.036 | 0.005 | -0.060 |
| | | | 0.028 | 0.024 | 0.030 | 0.031 | 0.034 | 0.035 | 0.039 |
| Materials | 4 | -3.8 | -0.046 | -0.009 | 0.015 | 0.030 | 0.009 | 0.027 | 0.028 |
| | | | 0.022 | 0.021 | 0.024 | 0.027 | 0.029 | 0.032 | 0.034 |
| Real Estate | 4 | -1.1 | -0.018 | 0.021 | 0.016 | 0.023 | -0.016 | 0.026 | -0.009 |
| | | | 0.018 | 0.027 | 0.030 | 0.033 | 0.041 | 0.042 | 0.041 |
| Utilities | 4 | -1.2 | 0.041 | -0.003 | -0.004 | 0.067 | 0.045 | -0.005 | 0.036 |
| | | | 0.021 | 0.030 | 0.033 | 0.041 | 0.042 | 0.044 | 0.045 |

Table 7.8: 44 stock long-short—conventional and uncertain risk contributions by factor

| | Conventional | Uncertain center | ± | means | $\boldsymbol{w}$ unc | $\boldsymbol{E}$ unc | $\boldsymbol{w}, \boldsymbol{E}$ unc |
|---|---|---|---|---|---|---|---|
| All | 4.85 | 5.22 | 1.03 | 4.50 | 0.37 | 0.34 | 0.02 |
| mkt | 0.00 | 0.13 | 0.20 | 0.00 | 0.08 | 0.05 | 0 |
| size | 0.00 | 0.05 | 0.08 | 0.00 | 0.00 | 0.04 | 0 |
| gv | 0.00 | 0.06 | 0.11 | 0.00 | 0.01 | 0.05 | 0 |
| pc1 | 0.00 | 0.07 | 0.10 | 0.00 | 0.01 | 0.05 | 0 |
| pc2 | 0.02 | 0.08 | 0.12 | 0.02 | 0.01 | 0.05 | 0 |
| pc3 | 0.00 | 0.06 | 0.08 | 0.00 | 0.00 | 0.05 | 0 |
| pc4 | 0.00 | 0.06 | 0.09 | 0.00 | 0.01 | 0.05 | 0 |
| $\varepsilon^2$ | 4.81 | 4.71 | 1.12 | 4.46 | 0.25 | | |

Table 7.9: 44 stock long-short—ex-ante modeled and realized risk contributions by sector

| | | Uncertain | | Realized | |
|---|---|---|---|---|---|
| | Conventional | center | ± | $\boldsymbol{w}_0$ | $\boldsymbol{w}_T$ |
| All | 4.85 | 5.22 | 1.03 | 4.47 | 4.30 |
| Comm Svcs | 0.26 | 0.29 | 0.11 | 0.23 | 0.16 |
| Cons Discr | 0.32 | 0.35 | 0.14 | 0.38 | 0.54 |
| Cons Staples | 0.34 | 0.36 | 0.11 | 0.21 | 0.27 |
| Energy | 0.14 | 0.17 | 0.10 | 1.11 | 0.59 |
| Financials | 0.64 | 0.68 | 0.22 | -0.16 | -0.28 |
| Healthcare | 0.32 | 0.37 | 0.15 | 0.36 | 0.32 |
| Industrials | 0.23 | 0.25 | 0.11 | 0.02 | 0.03 |
| Tech | 0.55 | 0.61 | 0.29 | 0.19 | 0.48 |
| Materials | 0.45 | 0.49 | 0.17 | 1.25 | 1.18 |
| Real Estate | 0.71 | 0.74 | 0.18 | 0.34 | 0.44 |
| Utilities | 0.87 | 0.90 | 0.27 | 0.54 | 0.56 |

## 7.10 Mathematical results

**Lemma 7.1.**

$$\textit{For} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

$$\begin{aligned} \mathrm{E}[x_1 x_2 x_3 x_4] = {} & \mu_1\mu_2\mu_3\mu_4 \\ & + \mu_1\mu_2\sigma_{3,4} + \mu_3\mu_4\sigma_{1,2} + \sigma_{1,2}\sigma_{3,4} \\ & + \mu_1\mu_3\sigma_{2,4} + \mu_2\mu_4\sigma_{1,3} + \sigma_{1,3}\sigma_{2,4} \\ & + \mu_1\mu_4\sigma_{2,3} + \mu_2\mu_3\sigma_{1,4} + \sigma_{1,4}\sigma_{2,3} \end{aligned}$$

*Proof.*

$$\begin{aligned}
\mathrm{E}\left[x_1x_2x_3x_4\right] &= \mathrm{E}\left[\prod_{i=1}^{4}\left[\mu_i + (x_i - \mu_i)\right]\right] \\
&= \mu_1\mu_2\mu_3\mu_4 \\
&\quad + \mu_1\mu_2\sigma_{3,4} + \mu_3\mu_4\sigma_{1,2} \\
&\quad + \mu_1\mu_3\sigma_{2,4} + \mu_2\mu_4\sigma_{1,3} \\
&\quad + \mu_1\mu_4\sigma_{2,3} + \mu_2\mu_3\sigma_{1,4} \\
&\quad + \mathrm{E}\left[\prod_{i=1}^{4}(x_i - \mu_i)\right] \\
&\quad \text{since odd moments are } 0
\end{aligned}$$

$$\mathrm{E}\left[\prod_{i=1}^{4}(x_i - \mu_i)\right] = \sigma_{1,2}\sigma_{3,4} + \sigma_{1,3}\sigma_{2,4} + \sigma_{1,4}\sigma_{2,3}$$

by Isserlis' theorem Isserlis [1918] since 0 mean Gaussians

$$\mathrm{cov}\left[x_1x_2, x_3x_4\right] = \mathrm{E}\left[x_1x_2x_3x_4\right] - \underbrace{\mathrm{E}\left[x_1x_2\right]}_{(\mu_1\mu_2+\sigma_{12})}\underbrace{\mathrm{E}\left[x_3x_4\right]}_{(\mu_3\mu_4+\sigma_{34})}$$

□

**Corollary 7.1.** *Covariance of squared Gaussians*

*For* $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right), \mathrm{cov}\left[x_1^2, x_2^2\right] = 4\mu_1\mu_2\sigma_{1,2} + 2\sigma_{1,2}^2$

**Lemma 7.2.** *For* $\{x_1, x_2\}, \{y_1, y_2\}$ *jointly independent,*

$$\begin{aligned}
\mathrm{cov}\left[x_1y_1, x_2y_2\right] &= \mathrm{E}\left[x_1\right]\mathrm{E}\left[x_2\right]\mathrm{cov}\left[y_1, y_2\right] + \mathrm{cov}\left[x_1, x_2\right]\mathrm{E}\left[y_1\right]\mathrm{E}\left[y_2\right] \\
&\quad + \mathrm{cov}\left[x_1, x_2\right]\mathrm{cov}\left[y_1, y_2\right] \qquad (7.91) \\
&= \left(\mathrm{E}\left[x_1\right]\mathrm{E}\left[x_2\right] + \mathrm{cov}\left[x_1, x_2\right]\right)\mathrm{cov}\left[y_1, y_2\right] \\
&\quad + \mathrm{cov}\left[x_1, x_2\right]\mathrm{E}\left[y_1\right]\mathrm{E}\left[y_2\right] \qquad (7.92)
\end{aligned}$$

*Proof.*

$$\begin{aligned}
\operatorname{cov}[x_1y_1, x_2y_2] &= \operatorname{E}[x_1y_1x_2y_2] - \operatorname{E}[x_1y_1]\operatorname{E}[x_2y_2] \\
&= \operatorname{E}[x_1x_2]\operatorname{E}[y_1y_2] - \operatorname{E}[x_1]\operatorname{E}[y_1]\operatorname{E}[x_2]\operatorname{E}[y_2] \\
&= \Big(\operatorname{E}[x_1]\operatorname{E}[x_2] + \operatorname{cov}[x_1, x_2]\Big)\Big(\operatorname{E}[y_1]\operatorname{E}[y_2] + \operatorname{cov}[y_1, y_2]\Big) \\
&\quad - \operatorname{E}[x_1]\operatorname{E}[y_1]\operatorname{E}[x_2]\operatorname{E}[y_2] \\
&= \operatorname{E}[x_1]\operatorname{E}[x_2]\operatorname{cov}[y_1, y_2] + \operatorname{cov}[x_1, x_2]\operatorname{E}[y_1]\operatorname{E}[y_2] \\
&\quad + \operatorname{cov}[x_1, x_2]\operatorname{cov}[y_1, y_2]
\end{aligned}$$

□

## 7.11 Mean and variance of factor covariance from uncertain covariance model

The covariance model in this paper is a generalization of the form in Shah [2015], which imposes the following structure on $\boldsymbol{F}$:

$$\boldsymbol{C} = \boldsymbol{E}\left(\overbrace{\boldsymbol{M}\operatorname{diag}[\boldsymbol{\theta}]\boldsymbol{M}^{\mathrm{T}}}^{\boldsymbol{F}}\right)\boldsymbol{E}^{\mathrm{T}} + \gamma\operatorname{diag}\left[\boldsymbol{\epsilon}^2\right] \tag{7.93}$$

where $\boldsymbol{M} = (K \times K)$ fixed

$\boldsymbol{\theta} = (K \times 1)$ random and dependent on $\gamma$

The values $\operatorname{E}[\boldsymbol{F}]$, $\operatorname{var}[\boldsymbol{F}]$, and $\operatorname{cov}[\boldsymbol{F}, \gamma]$ used in this paper are easily computed from that form.

$$\operatorname{E}[\boldsymbol{F}] = \boldsymbol{M}\operatorname{diag}[\operatorname{E}[\boldsymbol{\theta}]]\boldsymbol{M}^{\mathrm{T}} \tag{7.94}$$

$$\operatorname{cov}[\boldsymbol{F}, \gamma] = \boldsymbol{M}\operatorname{diag}[\operatorname{cov}[\boldsymbol{\theta}, \gamma]]\boldsymbol{M}^{\mathrm{T}} \tag{7.95}$$

$$\operatorname{cov}[F_{i,j}, F_{k,l}] = \sum_{s,t} M_{i,s}M_{j,s}M_{k,t}M_{l,t}\operatorname{cov}[\theta_s, \theta_t] \tag{7.96}$$

*Proof.*

$$
\begin{aligned}
\boldsymbol{F} &= \boldsymbol{M}\,\mathrm{diag}\left[\boldsymbol{\theta}\right]\boldsymbol{M}^{\mathrm{T}} \\
\mathrm{E}\left[\boldsymbol{F}\right] &= \mathrm{E}\left[\boldsymbol{M}\,\mathrm{diag}\left[\boldsymbol{\theta}\right]\boldsymbol{M}^{\mathrm{T}}\right] \\
&= \boldsymbol{M}\,\mathrm{diag}\left[\mathrm{E}\left[\boldsymbol{\theta}\right]\right]\boldsymbol{M}^{\mathrm{T}} \\
\mathrm{cov}\left[\boldsymbol{F},\gamma\right] &= \mathrm{cov}\left[\boldsymbol{M}\,\mathrm{diag}\left[\boldsymbol{\theta}\right]\boldsymbol{M}^{\mathrm{T}},\gamma\right] \\
&= \boldsymbol{M}\,\mathrm{diag}\left[\mathrm{cov}\left[\boldsymbol{\theta},\gamma\right]\right]\boldsymbol{M}^{\mathrm{T}} \\
F_{i,j} &= \sum_{k} M_{i,k}M_{j,k}\theta_{k} \\
\mathrm{cov}\left[F_{i,j},F_{k,l}\right] &= \mathrm{cov}\left[\sum_{s} M_{i,s}M_{j,s}\theta_{s},\sum_{t} M_{k,t}M_{l,t}\theta_{t}\right] \\
&= \sum_{s,t} M_{i,s}M_{j,s}M_{k,t}M_{l,t}\,\mathrm{cov}\left[\theta_{s},\theta_{t}\right]
\end{aligned}
$$

□

# CHAPTER 8

# CONCLUSION

Motivated largely by the problems caused by estimation error in investment portfolio optimization, this dissertation developed techniques to improve covariance forecasting, first by incorporating side information, then by projecting uncertainty in the values of parameters of factor-modeled covariance to uncertainty in covariance between securities.

The latter result made it possible to approach several investment problems explicitly modeling uncertainty — mean-variance portfolio optimization, risk parity portfolio construction, reporting risks under uncertainty and price movement. Algorithmic trading would be the next, and surely there are others.

At the same time, applications of the uncertain covariance model in chapter 5 aren't restricted to quantitative finance. Covariance appears everywhere. Mean and variance are the first two terms of a Taylor series, and forecasting higher moments (particularly for the purpose of making decisions) is hard. There are surely many non-finance problems that could benefit from a tractable, predictive distribution of covariance.

# BIBLIOGRAPHY


Karim M Abadir, Walter Distaso, and Filip Žikeš. Design-free estimation of variance matrices. *Journal of Econometrics*, 181(2):165–180, 2014. ISSN 0304-4076. doi: 10.1016/j.jeconom.2014.03.010.

Evan W Anderson and Ai-Ru Cheng. Robust Bayesian portfolio choices. *The Review of Financial Studies*, 29(5):1330–1375, 2016.

Werner Antweiler and Murray Z Frank. Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, 59(3):1259–1294, 2004. doi: 10.1111/j.1540-6261.2004.00662.x.

Adam Atkins, Mahesan Niranjan, and Enrico Gerding. Financial news predicts stock market volatility better than close price. *The Journal of Finance and Data Science*, 4(2):120–137, 2018.

Doron Avramov and Guofu Zhou. Bayesian portfolio analysis. *Annual Review of Financial Economics*, 2(1):25–47, 2010. doi: 10.1146/annurev-financial-120209-133947.

Xi Bai, Katya Scheinberg, and Reha Tutuncu. Least-squares approach to risk parity in portfolio selection. *Quantitative Finance*, 16(3):357–376, 2016. doi: 10.1080/14697688.2015.1031815.

Ole E Barndorff-Nielsen and Neil Shephard. Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):167–241, 2001.

Christopher B Barry. Portfolio analysis under uncertain means, variances, and covariances. *The Journal of Finance*, 29(2):515–522, 1974.

David Bauder, Taras Bodnar, Nestor Parolya, and Wolfgang Schmid. Bayesian mean–variance analysis: optimal portfolio selection under parameter uncertainty. *Quantitative Finance*, 21(2):221–242, 2021.

Luc Bauwens, Sébastien Laurent, and Jeroen VK Rombouts. Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109, 2006.

Stan Beckers. Variances of security price returns based on high, low, and closing prices. *The Journal of Business*, pages 97–112, 1983.

Simone Bernardi, Markus Leippold, and Harald Lohre. Second-order risk of alternative risk parity strategies. *Journal of Risk*, 21(3):1–25, 2019. doi: 10.21314/jor.2018.401.

Dimitris Bertsimas, David B Brown, and Constantine Caramanis. Theory and applications of robust optimization. *SIAM review*, 53(3):464–501, 2011. doi: 10.1137/080734510.

Fischer Black and Robert Litterman. Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43, 1992. doi: 10.2469/faj.v48.n5.28.

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

Tim Bollerslev. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The Review of Economics and Statistics*, pages 498–505, 1990.

Tim Bollerslev, Robert F. Engle, and Daniel B. Nelson. ARCH models. In *Handbook of Econometrics*, volume 4 of *Handbook of Econometrics*, chapter 49, pages 2959–3038. Elsevier, 1994. doi: 10.1016/S1573-4412(05)80018-2.

Jean-Philippe Bouchaud. Radical complexity. *arXiv preprint arXiv:2103.09692*, 2021.

Michael W Brandt and Christopher S Jones. Volatility forecasting with range-based EGARCH models. *Journal of Business & Economic Statistics*, 24(4):470–486, 2006.

L Breiman. Optimal gambling systems for favorable games. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California, 1961.

Mark Broadie. Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, 45(1):21–58, 1993. doi: 10.1007/bf02282040.

Joël Bun, Jean-Philippe Bouchaud, and Marc Potters. Cleaning large correlation matrices: tools from random matrix theory. *Physics Reports*, 666:1–109, 2017.

Fabiano Cavadini, Alessandro Sbuelz, and Fabio Trojani. A simplified way of incorporating model risk, estimation risk and robustness in mean variance portfolio management. *Working paper, Tilburg University*, 2001.

Sebastián Ceria and Robert A Stubbs. Incorporating estimation errors into portfolio selection: Robust portfolio construction. *Journal of Asset Management*, 7(2):109–127, 2006.

Gary Chamberlain. A characterization of the distributions that imply meanvariance utility functions. *Journal of Economic Theory*, 29(1):185–201, 1983. ISSN 0022-0531. doi: 10.1016/0022-0531(83)90129-1.

Tun-Jen Chang, Sang-Chin Yang, and Kuang-Jung Chang. Portfolio optimization problems in different risk measures using genetic algorithm. *Expert Systems with Applications*, 36(7):10529–10537, 2009. ISSN 0957-4174. doi: 10.1016/j.eswa.2009.02.062.

James Chilson, Raymond Ng, Alan Wagner, and Ruben Zamar. Parallel computation of high-dimensional robust correlation and covariance matrices. *Algorithmica*, 45 (3):403–431, 2006.

Ray Yeutien Chou. Forecasting financial volatilities with extreme values: the conditional autoregressive range (carr) model. *Journal of Money, Credit and Banking*, pages 561–582, 2005.

Ray Yeutien Chou, Chun-Chou Wu, and Nathan Liu. Forecasting time-varying covariance with a range-based dynamic conditional correlation model. *Review of Quantitative Finance and Accounting*, 33(4):327–345, 2009.

Ray Yeutien Chou, Hengchih Chou, and Nathan Liu. Range volatility models and their applications in finance. In *Handbook of Quantitative Finance and Risk Management*, pages 1273–1281. Springer, 2010.

Giorgio Costa. *Advances in Risk Parity Portfolio Optimization*. PhD thesis, Mechanical and Industrial Engineering, University of Toronto, 2021. https://hdl.handle.net/1807/106376.

Giorgio Costa and Roy Kwon. A robust framework for risk parity portfolios. *Journal of Asset Management*, 21(5):447–466, 2020a.

Giorgio Costa and Roy H Kwon. Risk parity portfolio optimization under a Markov regime-switching framework. *Quantitative Finance*, 19(3):453–471, 2019.

Giorgio Costa and Roy H Kwon. Generalized risk parity portfolio optimization: an admm approach. *Journal of Global Optimization*, 78(1):207–238, 2020b.

Peter Cotton. *Microprediction: Building an Open AI Network*. MIT Press, 2022a.

Peter Cotton. Schur complementary portfolios – a unification of machine learning and optimization-based allocation, Nov 2022b. URL https://medium.com/geekculture/schur-complementary-portfolios-fix-hierarchical-risk-parity-28b0efa1f35f.

Aditi Dandapani, Paul Jusselin, and Mathieu Rosenbaum. From quadratic Hawkes processes to super-Heston rough volatility models with Zumbach effect. *Quantitative Finance*, pages 1–13, 2021.

Michael J Daniels and Robert E Kass. Shrinkage estimators for covariance matrices. *Biometrics*, 57(4):1173–1184, 2001.

Sanjiv R Das and Mike Y Chen. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science*, 53(9):1375–1388, 2007.

Mark Davis and Sébastien Lleo. Black–litterman in continuous time: the case for filtering. *Quantitative Finance Letters*, 1(1):30–35, 2013.

Marielle de Jong. Portfolio optimisation in an uncertain world. *Journal of Asset Management*, 19(4):216–221, 2018. doi: 10.1057/s41260-017-0066-3.

Beatriz Vaz de Melo Mendes and Ricardo Pereira Câmara Leal. Robust multivariate modeling in finance. *International Journal of Managerial Finance*, 2005.

Gianluca De Nard, Olivier Ledoit, and Michael Wolf. Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics*, 01 2019. ISSN 1479-8409. doi: 10.1093/jjfinec/nby033.

Gianluca De Nard, Robert F Engle, Olivier Ledoit, and Michael Wolf. Large dynamic covariance matrices: enhancements based on intraday data. *University of Zurich, Department of Economics, Working Paper*, (356), 2021.

Marcos Lopez De Prado. Building diversified portfolios that outperform out of sample. *The Journal of Portfolio Management*, 42(4):59–69, 2016.

Victor DeMiguel and Francisco J Nogales. Portfolio selection with robust estimation. *Operations Research*, 57(3):560–577, 2009. doi: 10.1287/opre.1080.0566.

Dan diBartolomeo and Sandy Warrick. Making covariance-based portfolio risk models sensitive to the rate at which markets reflect new information. In John Knight and Stephen Satchell, editors, *Linear Factor Models in Finance*, Quantitative Finance, chapter 12, pages 249 – 261. Butterworth-Heinemann, Oxford, 2005. ISBN 978-0-7506-6006-8. doi: 10.1016/B978-075066006-8.50013-6.

Thomas Dimpfl and Stephan Jank. Can internet search queries help to predict stock market volatility? *European Financial Management*, 22(2):171–192, 2016.

Robert Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.

Robert Engle. Risk and volatility: Econometric models and financial practice. *American Economic Review*, 94(3):405–420, 2004.

Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007, 1982. doi: 10.2307/1912773.

Robert F. Engle and Kenneth F. Kroner. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150, 1995. ISSN 02664666, 14694360. URL http://www.jstor.org/stable/3532933.

Robert F Engle, Victor K Ng, and Michael Rothschild. Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics*, 45(1-2):213–237, 1990.

Robert F Engle, Olivier Ledoit, and Michael Wolf. Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2):363–375, 2019. doi: 10.1080/07350015.2017.1345683.

Frank J Fabozzi, Petter N Kolm, Dessislava A Pachamanova, and Sergio M Focardi. *Robust Portfolio Optimization and Management*. John Wiley & Sons, 2007.

Frank J Fabozzi, Dashan Huang, and Guofu Zhou. Robust portfolios: contributions from operations research and finance. *Annals of Operations Research*, 176(1):191–220, 2010.

Jianqing Fan, Yingying Fan, and Jinchi Lv. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1):186–197, 2008.

Jianqing Fan, Yuan Liao, and Martina Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 75(4), 2013. doi: 10.1111/rssb.12016.

William Feller. The asymptotic distribution of the range of sums of independent random variables. *The Annals of Mathematical Statistics*, 22(3):427–432, 1951.

Peter A Frost and James E Savarino. An empirical bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, 21(03):293–305, 1986. doi: 10.2307/2331043.

Lorenzo Garlappi, Raman Uppal, and Tan Wang. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies*, 20(1):41–81, 2007.

Nicolae Gârleanu and Lasse Heje Pedersen. Dynamic trading with predictable returns and transaction costs. *The Journal of Finance*, 68(6):2309–2340, 2013.

Mark B Garman and Michael J Klass. On the estimation of security price volatilities from historical data. *The Journal of Business*, pages 67–78, 1980.

Lisa R Goldberg, Alex Papanicolaou, and Alexander Shkolnik. The dispersion bias. *Available at SSRN 3071328*, 2018. doi: 10.2139/ssrn.3071328.

Donald Goldfarb and Garud Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38, 2003. doi: 10.1287/moor.28.1.1.14260.

Bjarni V Halldórsson and Reha H Tütüncü. An interior-point method for a class of saddle-point problems. *Journal of Optimization Theory and Applications*, 116(3): 559–590, 2003. doi: 10.1023/a:1023065319772.

Lars Peter Hansen and Thomas J Sargent. *Robustness*. Princeton University Press, 2011.

Peter R Hansen and Asger Lunde. A forecast comparison of volatility models: does anything beat a GARCH(1, 1)? *Journal of Applied Econometrics*, 20(7):873–889, 2005.

Peter Reinhard Hansen, Zhuo Huang, and Howard Howan Shek. Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906, 2012.

Peter Huber. *Robust statistics*. John Wiley & Sons, 1981. doi: 10.1002/0471725250.

Leon Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2): 134–139, 1918.

J David Jobson and Bob Korkie. Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371):544–554, 1980. doi: 10.2307/2287643.

Iain M Johnstone. High dimensional statistical inference and random matrices. In *Proceedings of the International Congress of Mathematicians Madrid, August 22–30, 2006*, pages 307–333, 2007.

Philippe Jorion. Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(03):279–292, 1986. doi: 10.2307/2331042.

Philippe Jorion. Risk management lessons from Long-Term Capital Management. *European financial management*, 6(3):277–300, 2000. doi: 10.1111/1468-036X.00125.

Marcin Kacperczyk and Paul Damien. Asset allocation under distribution uncertainty. *McCombs Research Paper Series No. IROM-01*, 11, 2011.

Petko S Kalev, Wai-Man Liu, Peter K Pham, and Elvis Jarnecic. Public information arrival and volatility of intraday stock returns. *Journal of Banking & Finance*, 28 (6):1441–1467, 2004.

Michalis Kapsos, Nicos Christofides, and Berc Rustem. Robust risk budgeting. *Annals of Operations Research*, 266(1):199–221, 2018.

Ioannis Karatzas and Steven Shreve. *Brownian Motion and Stochastic Calculus*, volume 113. Springer, 1991.

J. Kelly. A new interpretation of information rate. *IRE Transactions on Information Theory*, 2(3):185–189, 1956. doi: 10.1109/TIT.1956.1056803.

Jafar A Khan, Stefan Van Aelst, and Ruben H Zamar. Robust linear model selection based on least angle regression. *Journal of the American Statistical Association*, 102(480):1289–1299, 2007.

Hyuksoo Kim and Saejoon Kim. Reduction of estimation error impact in the risk parity strategies. *Quantitative Finance*, pages 1–14, 2021.

Seung-Jean Kim and Stephen Boyd. Robust efficient frontier analysis with a separable uncertainty model. Technical report, Technical report, Department of Electrical Engineering, Stanford University, 2007.

Roger W Klein and Vijay S Bawa. The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, 3(3):215–231, 1976.

Petter Kolm and Gordon Ritter. On the Bayesian interpretation of Black-Litterman. *European Journal of Operational Research*, 258(2):564–572, 2017.

Petter Kolm and Gordon Ritter. Factor investing with Black-Litterman-Bayes: Incorporating factor views and priors in portfolio construction. *The Journal of Portfolio Management*, 47(2):113–126, 2020.

Petter Kolm, Gordon Ritter, and Joseph Simonian. Black-Litterman and beyond: The Bayesian paradigm in investment management. *The Journal of Portfolio Management*, 47(5):91–113, 2021.

Hiroshi Konno and Hiroaki Yamazaki. Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. *Management Science*, 37(5):519–531, 1991.

Jaya Krishnakumar and Elvezio Ronchetti. Robust estimators for simultaneous equations models. *Journal of Econometrics*, 78(1):295–314, 1997.

Naoto Kunitomo. Improving the Parkinson method of estimating security price volatilities. *The Journal of Business*, pages 295–302, 1992.

Tze Leung Lai, Haipeng Xing, and Zehao Chen. Mean–variance portfolio optimization when means and covariances are unknown. *The Annals of Applied Statistics*, 5(2A): 798–823, 2011.

Clifford Lam. Nonparametric eigenvalue-regularized precision or covariance matrix estimator. *The Annals of Statistics*, 44(3):928–953, 2016. doi: 10.1214/15-AOS1393.

Nathan Lassance, Victor DeMiguel, and Frédéric Vrins. Optimal portfolio diversification via independent component analysis. *Operations Research*, 70(1):55–72, 2022. doi: 10.1287/opre.2021.2140.

GJ Lauprete, AM Samarov, and RE Welsch. Robust portfolio optimization. In *Developments in Robust Statistics*, pages 235–245. Springer, 2003.

Olivier Ledoit and Michael 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. doi: 10.1016/S0927-5398(03)00007-0.

Olivier Ledoit and Michael Wolf. Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004. doi: 10.3905/jpm.2004.110.

Olivier Ledoit and Michael Wolf. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2):1024–1060, 2012.

Olivier Ledoit and Michael Wolf. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *The Review of Financial Studies*, 30(12):4349–4388, 06 2017. ISSN 0893-9454. doi: 10.1093/rfs/hhx052.

Pui Lam Leung and Foon Yip Ng. Improved estimation of a covariance matrix in an elliptically contoured matrix distribution. *Journal of Multivariate Analysis*, 88(1): 131–137, 2004.

Harald Lohre, Heiko Opfer, and Gabor Orszag. Diversifying risk parity. *Journal of Risk*, 16(5):53–79, 2014. doi: 10.21314/jor.2014.284.

Zhaosong Lu. Robust portfolio selection based on a joint ellipsoidal uncertainty set. *Optimization Methods & Software*, 26(1):89–104, 2011. doi: 10.1080/10556780903334682.

Sébastien Maillard, Thierry Roncalli, and Jérôme Teïletche. The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management*, 36(4): 60–70, 2010. doi: 10.3905/jpm.2010.36.4.060.

Benoit Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419, 1963.

Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952. doi: 10.1111/j.1540-6261.1952.tb01525.x.

Harry Markowitz. Mean–variance approximations to expected utility. *European Journal of Operational Research*, 234(2):346–355, 2014.

Harry M Markowitz. *Portfolio Selection*. Yale University Press, jul 1959. ISBN 9780300013726.

Ricardo A Maronna and Ruben H Zamar. Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics*, 44(4):307–317, 2002.

Jose Menchero and Ben Davis. Risk contribution is exposure times volatility times correlation: Decomposing risk using the x-sigma-rho formula. *The Journal of Portfolio Management*, 37(2):97–106, 2011. ISSN 0095-4918. doi: 10.3905/jpm.2011.37.2.097.

Robert C Merton. Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413, 1971. ISSN 0022-0531. doi: 10.1016/0022-0531(71)90038-X.

Attilio Meucci. Fully flexible views: Theory and practice. *Risk*, 21(10):97–102, 2008.

Attilio Meucci. Review of statistical arbitrage, cointegration, and multivariate Ornstein-Uhlenbeck. *SSRN*, 2009. doi: 10.2139/ssrn.1404905.

Attilio Meucci. 'P' versus 'Q': Differences and commonalities between the two areas of quantitative finance. *GARP Risk Professional*, pages 47–50, 2 2011. Available at SSRN: https://ssrn.com/abstract=1717163.

Richard O. Michaud. The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45(1):31–42, 1989. doi: 10.2469/faj.v45.n1.31.

Richard O. Michaud. *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Oxford University Press, 1998.

Luiza Miranyan. Incorporating forward-looking market data into linear multifactor fundamental models. *Journal of Risk*, 14(4):3–34, 2012. doi: 10.21314/JOR.2012.245.

Leela Mitra, Gautam Mitra, and Dan Dibartolomeo. Equity portfolio risk estimation using market information and sentiment. *Quantitative Finance*, 9(8):887–895, 2009. doi: 10.1080/14697680903448361.

Colm A. O'Cinneide. Risk contributions: duality and sensitivity. *Quantitative Finance*, 18(12):2023–2033, 2018. doi: 10.1080/14697688.2017.1423175.

Michael Parkinson. The extreme value method for estimating the variance of the rate of return. *The Journal of Business*, 53(1):61–65, 1980. doi: 10.1086/296071.

Markus Pelger. Deep learning statistical arbitrage. NYU Courant, M.S. in Mathematics in Finance Seminars, Feb 2023.

Cdric Perret-Gentil and Maria-Pia Victoria-Feser. Robust mean-variance portfolio selection. fame research paper 140. *International Center for Financial Asset Management and Engineering, Geneva*, 2004.

Vasiliki Plerou, Parameswaran Gopikrishnan, Bernd Rosenow, Luis A Nunes Amaral, Thomas Guhr, and H Eugene Stanley. Random matrix approach to cross correlations in financial data. *Physical Review E*, 65(6):066126, 2002.

Marc Potters and Jean-Philippe Bouchaud. *A First Course in Random Matrix Theory: For Physicists, Engineers and Data Scientists*. Cambridge University Press, 2020.

Marc Potters, Jean-Philippe Bouchaud, and Laurent Laloux. Financial applications of random matrix theory: Old laces and new pieces. *Acta Physica Polonica B*, 36 (9):2767, 2005.

Edward Qian. Risk parity portfolios: Efficient portfolios through true diversification. *Panagora Asset Management*, 2005.

Edward Qian. Risk parity and diversification. *The Journal of Investing*, 20(1):119–127, 2011. ISSN 1068-0896. doi: 10.3905/joi.2011.20.1.119.

Edward E Qian, Ronald H Hua, and Eric H Sorensen. *Quantitative equity portfolio management: modern techniques and applications*. CRC Press, 2007. doi: 10.1201/9781420010794.

Calum Robertson, Shlomo Geva, and Rodney C Wolff. Can the content of public news be used to forecast abnormal stock market behaviour? In *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, pages 637–642. IEEE, 2007.

R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41, 2000. ISSN 1465-1211. doi: 10.21314/JOR.2000.038.

R Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002.

David M Rocke. Robustness properties of S-estimators of multivariate location and shape in high dimension. *The Annals of Statistics*, pages 1327–1345, 1996.

L Christopher G Rogers and Stephen E Satchell. Estimating variance from high, low and closing prices. *The Annals of Applied Probability*, pages 504–512, 1991.

Leonard CG Rogers and Fanyin Zhou. Estimating correlation from high, low, opening and closing prices. *The Annals of Applied Probability*, 18(2):813–823, 2008.

Thierry Roncalli and Guillaume Weisang. Risk parity portfolios with risk factors. *Quantitative Finance*, 16(3):377–388, 2016. doi: 10.1080/14697688.2015.1046907.

Peter J Rousseeuw. Multivariate estimation with high breakdown point. In W Grossman, G Pflug, I Vincze, and W Wertz, editors, *Mathematical Statistics and Applications*, volume 8, pages 283–297. Reidel Publishing Company, 1985.

Peter J Rousseeuw and Katrien Van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3):212–223, 1999.

Berc Rustem, Robin G Becker, and Wolfgang Marty. Robust min–max portfolio strategies for rival forecast and risk scenarios. *Journal of Economic Dynamics and Control*, 24(11-12):1591–1621, 2000.

Juliane Schäfer and Korbinian Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1), 2005.

Bernd Scherer. Portfolio resampling: Review and critique. *Financial Analysts Journal*, 58(6):98–109, 2002. doi: 10.2469/faj.v58.n6.2489.

Maria Grazia Scutellà and Raffaella Recchia. Robust portfolio asset allocation and risk measures. *Annals of Operations Research*, 204(1):145–169, 2013. doi: 10.1007/s10479-012-1266-3.

Anish Shah. Short-term risk from long-term models. Presented at Northfield Information Services Research Conference in Key Largo FL, 10 2007. URL: http://northinfo.com/documents/286.pdf.

Anish Shah. Mitigating estimation error in optimization. Presented at Northfield Information Services seminar in London, 11 2009. URL: https://www.northinfo.com/documents/361.pdf.

Anish Shah. Uncertain risk parity. *Journal of Investment Strategies*, 10(1):29–41, 2021a. doi: 10.21314/JOIS.2021.009.

Anish Shah. Risk decomposition under uncertainty and price movement. CFA Society VBA Netherlands, Feb 2021b.

Anish Shah. Risk under uncertainty. Qwafaxnew NYC, Mar 2021c.

Anish Shah. Risk under uncertainty and price movement. London Quant Group, Mar 2022.

Anish R Shah. Short-term risk and adapting covariance models to current market conditions. *SSRN*, 2014. doi: 10.2139/ssrn.2501071.

Anish R Shah. Uncertain covariance models and uncertainty-penalized portfolio optimization. *SSRN*, 2015. doi: 10.2139/ssrn.2616109.

Anish R Shah. Portfolio risks under estimation uncertainty and price movement. *SSRN*, 2020. doi: 10.2139/ssrn.3774239.

Annastiina Silvennoinen and Timo Teräsvirta. *Multivariate GARCH Models*, pages 201–229. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. ISBN 978-3-540-71297-8. doi: 10.1007/978-3-540-71297-8_9.

Charles Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 197–206, 1956.

Nassim Nicholas Taleb. *Antifragile: Things that gain from disorder*, page 425. Random House Incorporated, 2012. ISBN 9780812979688. appendix II, table 12.

Edward O. Thorp. Portfolio choice and the Kelly criterion. *Proceedings of the Business and Economics Section of the American Statistical Association*, pages 215–224, 1971.

Ruey S Tsay. *Analysis of Financial Time Series*. John Wiley & Sons, 2010.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Ioannis D Vrontos, Petros Dellaportas, and Dimitris N Politis. A full-factor multivariate GARCH model. *The Econometrics Journal*, 6(2):312–334, 2003.

Roy E Welsch and Xinfeng Zhou. Application of robust statistics to asset allocation models. *REVSTAT–Statistical Journal*, 5(1):97–114, 2007.

Dennis Yang and Qiang Zhang. Drift-independent volatility estimation based on high, low, open, and close prices. *The Journal of Business*, 73(3):477–492, 2000.

Liusha Yang, Romain Couillet, and Matthew R McKay. A robust statistics approach to minimum variance portfolio optimization. *IEEE Transactions on Signal Processing*, 63(24):6684–6697, 2015. doi: 10.1109/TSP.2015.2474298.

Chao Zhang, Yihuang Zhang, Mihai Cucuringu, and Zhongmin Qian. Volatility forecasting with machine learning and intraday commonality. *Journal of Financial Econometrics*, 03 2023. ISSN 1479-8409. doi: 10.1093/jjfinec/nbad005.