Financial Econometrics: Portfolio Analysis

II.2 Portfolio Selection: The Practice

Serkan Yener[1]

[1]Institut für Statistik
Ludwig-Maximilians-Universität München
Akademiestr. 1/I
80799 München
serkan.yener@stat.uni-muenchen.de

Nov 26 & Dec 03 & 10, 2013

Outline

Roadmap

- when implementing the theory of portfolio selection, however, there are issues doing it the obvious way
- the first set of issues is somewhat inherent to the transition from the idealized framework, originally put forward by Markowitz (1952), to the choice to be made by a real-world investor, i.e., there are some implicit "degrees of freedom" and caveats
- if this set of issues is settled by making the appropriate assumptions, another problem arises when trying to estimate optimal portfolios
- however, it turns out that the severity of this problem can only be illustrated after introducing simulation techniques
- these techniques help to understand and alleviate the so-called estimation risk of portfolio selection
- lastly, knowing the source of the problem allows us to look into the vast arsenal of statistical methods for practical alternatives

The Obvious Thing to Do

- flashback: all quantities of interest depend on two parameters only: $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$
- use unbiased estimators for their components:

$$\hat{\mu}_i = \frac{1}{T} \sum_{t=1}^{T} r_{t,i}$$

$$\hat{\sigma}_{ij} = \frac{1}{T-1} \sum_{t=1}^{T} (r_{t,i} - \hat{\mu}_i)(r_{t,j} - \hat{\mu}_j)$$

$$\hat{\sigma}_i^2 = \hat{\sigma}_{ii}$$

  for a sample of returns $\{r_{t,i} : 1 \leqslant t \leqslant T, 1 \leqslant i \leqslant n\}$
- from a theoretical perspective, a more convenient assumption is that these estimators should be consistent (which they are), $\operatorname{plim}_{T\to\infty} \hat{\theta}_T = \theta$, in order show via the delta method (van der Vaart, 1998, Chap. 3) that properties, like consistency and asymptotic distributions, of most quantities $\Phi(\hat{\theta})$ are straightforwardly derived from those of the plug-in estimator $\hat{\theta}$

Caveats I

- the mean-variance approach is static, i.e., there is no room for changes whatsoever
- but in economic systems, most things change in the course of time

$\Rightarrow$ do expectations about the future change as well?

- if yes, then quantities like $\mathrm{E}[\, r_{t+1} \,]$ or $\mathrm{E}\big[\, (r_{t+1} - \mathrm{E}[\, r_{t+1} \,])^2 \,\big]$ should account for a change in the portfolio holding period $[t, t+1]$, i.e., $\mathrm{E}_t[\, r_{t+1} \,]$ or $\mathrm{E}_t\big[\, (r_{t+1} - \mathrm{E}_t[\, r_{t+1} \,])^2 \,\big]$

$\Rightarrow$ replace unconditional expectations $\mathrm{E}[\, . \,]$ by conditional expectations $\mathrm{E}_t[\, . \,]$

- question: but does this matter for real-world applications?
- answer: no and yes..
- one of the stylized facts of financial markets states that expected (excess) returns are unpredictable
- put differently, the EMH postulates that the random walk with drift is the best model for (log-)stock prices $\Rightarrow$ any expectation (be it conditional or unconditional) of its returns is just a constant

Caveats II

- this is completely different for second moments, i.e., another stylized facts of financial markets is known as volatility clustering
- put differently, variances (and covariances) are time-varying in a way which exhibits a memory
- there are gains to be made if we can predict the “volatility regime” in the holding period
- yet another stylized facts of financial markets states that returns are not normally distributed, i.e., empirical distributions of returns are leptokurtic (more peaked around the mean and with fat tails)
- parametric distributions allowing for fat tails have the problem that the variance may not exist, i.e., $\sigma^2 = \infty$ which is obviously a problem for the mean-variance approach
- outliers (extreme realizations) $\Rightarrow$ fat tails
- it is well known that, in the face of outliers, alertstandard sample moments like $\hat{\mu}$ and $\hat{\sigma}^2$ quite unreliable $\Rightarrow$ fitted $\mathsf{N}\big(\hat{\mu}, \hat{\sigma}^2\big)$ can be substantially distorted; see Ruppert (2011, Chap. 4)

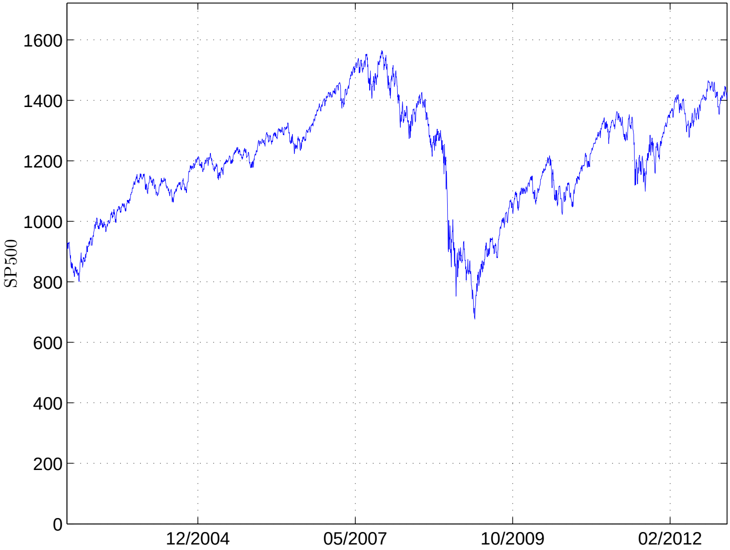
Caveats III

- as a remedy, robust alternatives to $\hat{\mu}$ and $\hat{\sigma}^2$ have been proposed in the literature

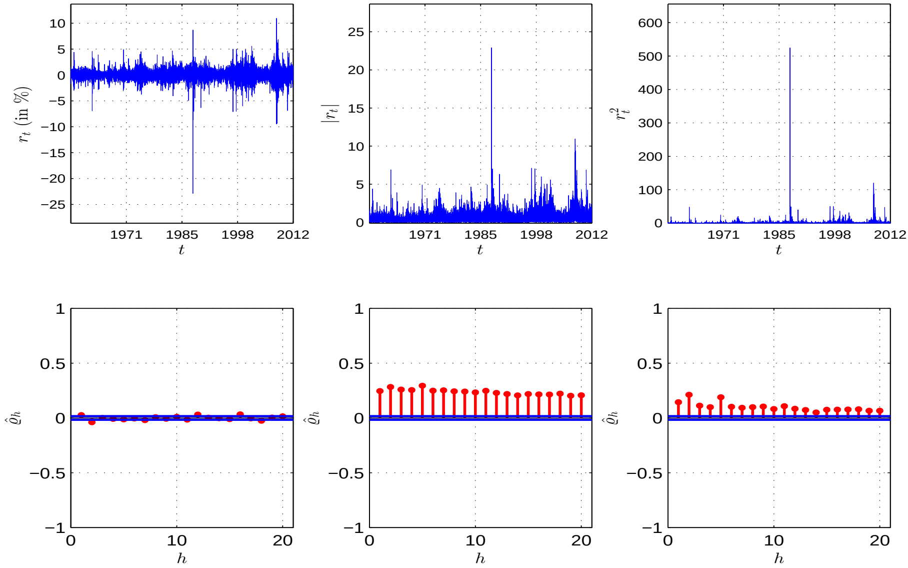
$$\hat{\mu}_{rob} = \text{med}(r_t)$$
$$\hat{\sigma}_{rob} = 1.4826 \cdot \text{med}\Big\{ \Big| r_t - \text{med}(r_t) \Big| \Big\}$$

see Huber (1981) or Hampel et al. (1986)

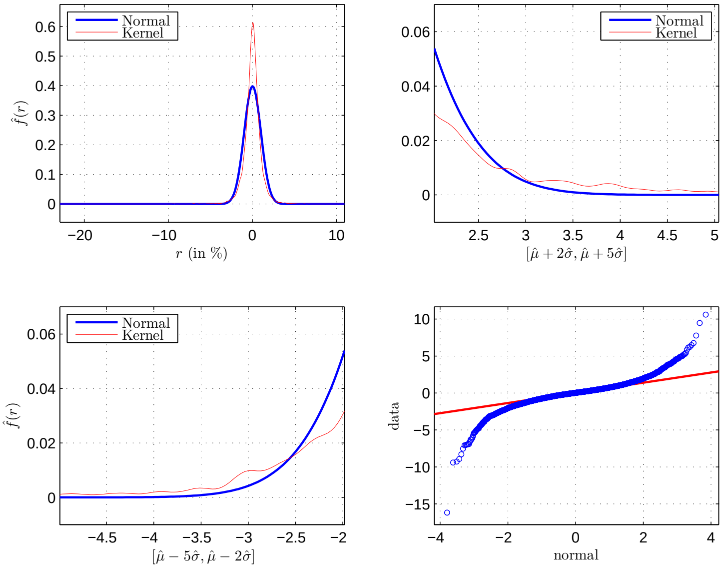
The SP500 Index



Random-Walk Property and Volatility Clustering in SP500 Returns



Non-Normality and Fat Tails in SP500 Returns



Flashback: Relevant Quantities I

- for all stocks $i = 1, \ldots, n$, use log-returns:

$$r_{t,i} = \ln\left(\frac{P_{t,i}}{P_{t-1,i}}\right) \cdot 100\%$$

- returns matrix:

$$\underbrace{\boldsymbol{R}}_{T\times n} = \begin{bmatrix} \boldsymbol{r}_1 & \boldsymbol{r}_2 & \cdots & \boldsymbol{r}_n \end{bmatrix} = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,n} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{T,1} & r_{T,2} & \cdots & r_{T,n} \end{bmatrix}$$

Flashback: Relevant Quantities II

- sample means vector of returns:

$$\underbrace{\hat{\boldsymbol{\mu}}}_{n\times 1} = \begin{pmatrix} \hat{\mu}_1 \\ \vdots \\ \hat{\mu}_n \end{pmatrix} = \begin{pmatrix} T^{-1}\sum_{t=1}^{T} r_{t,1} \\ \vdots \\ T^{-1}\sum_{t=1}^{T} r_{t,n} \end{pmatrix} = \frac{1}{T} \underbrace{\begin{bmatrix} r_{1,1} & \cdots & r_{T,1} \\ \vdots & \vdots & \vdots \\ r_{1,n} & \cdots & r_{T,n} \end{bmatrix}}_{n\times T} \underbrace{\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}_{T\times 1}$$

$$= \frac{1}{T} \begin{bmatrix} \boldsymbol{r}_1^{\mathsf{T}} \\ \vdots \\ \boldsymbol{r}_n^{\mathsf{T}} \end{bmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \frac{\boldsymbol{R}^{\mathsf{T}}\mathbf{1}_T}{T} \tag{1}$$

- sample covariance matrix of returns:

Flashback: Relevant Quantities III

$$\underbrace{\hat{\mathbf{\Sigma}}}_{n\times n} = \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1n} \\ \hat{\sigma}_{12} & \hat{\sigma}_2^2 & \cdots & \hat{\sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{1n} & \hat{\sigma}_{2n} & \cdots & \hat{\sigma}_{1n} \end{bmatrix}$$

$$= \frac{1}{T-1} \begin{bmatrix} \sum_{t=1}^T (r_{t,1} - \hat{\mu}_1)^2 & \cdots & \sum_{t=1}^T (r_{t,1} - \hat{\mu}_1)(r_{t,n} - \hat{\mu}_n) \\ \vdots & \ddots & \vdots \\ \sum_{t=1}^T (r_{t,1} - \hat{\mu}_1)(r_{t,n} - \hat{\mu}_n) & \cdots & \sum_{t=1}^T (r_{t,n} - \hat{\mu}_n)^2 \end{bmatrix}$$

$$= \frac{1}{T-1} \underbrace{\begin{bmatrix} (r_{1,1} - \hat{\mu}_1) & \cdots & (r_{T,1} - \hat{\mu}_1) \\ \vdots & \vdots & \vdots \\ (r_{1,n} - \hat{\mu}_n) & \cdots & (r_{T,n} - \hat{\mu}_n) \end{bmatrix}}_{n\times T} \underbrace{\begin{bmatrix} (r_{1,1} - \hat{\mu}_1) & \cdots & (r_{1,n} - \hat{\mu}_n) \\ \vdots & \vdots & \vdots \\ (r_{T,1} - \hat{\mu}_1) & \cdots & (r_{T,n} - \hat{\mu}_n) \end{bmatrix}}_{T\times n}$$

Flashback: Relevant Quantities IV

$$\underbrace{\hat{\boldsymbol{\Sigma}}}_{n\times n} = \frac{1}{T-1}\begin{bmatrix} (\boldsymbol{r}_1 - \hat{\mu}_1 \mathbf{1}_T)^\top \\ \vdots \\ (\boldsymbol{r}_n - \hat{\mu}_n \mathbf{1}_T)^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{r}_1 - \hat{\mu}_1 \mathbf{1}_T & \cdots & \boldsymbol{r}_n - \hat{\mu}_n \mathbf{1}_T \end{bmatrix}$$

$$=: \frac{1}{T-1}\begin{bmatrix} \tilde{\boldsymbol{r}}_1^\top \\ \vdots \\ \tilde{\boldsymbol{r}}_n^\top \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{r}}_1 & \cdots & \tilde{\boldsymbol{r}}_n \end{bmatrix} =: \frac{\tilde{\boldsymbol{R}}^\top \tilde{\boldsymbol{R}}}{T-1} \tag{2}$$

with mean-adjusted returns $\tilde{\boldsymbol{r}}_i = \boldsymbol{r}_i - \hat{\mu}_i \mathbf{1}_T = \boldsymbol{r}_i - (\boldsymbol{e}_i^\top \hat{\boldsymbol{\mu}})\mathbf{1}_T$ where

$$\underbrace{\boldsymbol{e}_i^\top}_{1\times n} := (\, 0 \ \cdots \ 0 \ \underbrace{1}_{\substack{i\text{th} \\ \text{position}}} \ 0 \ \cdots \ 0 \,)$$

is the *i*th unit vector

Flashback: Relevant Quantities V

- estimator of the fundamental matrix:

$$\hat{\mathbf{\Omega}} = \begin{bmatrix} \mathbf{1}_n^\top \hat{\mathbf{\Sigma}}^{-1} \mathbf{1}_n & \mathbf{1}_n^\top \hat{\mathbf{\Sigma}}^{-1} \hat{\boldsymbol{\mu}} \\ \hat{\boldsymbol{\mu}}^\top \hat{\mathbf{\Sigma}}^{-1} \mathbf{1}_n & \hat{\boldsymbol{\mu}}^\top \hat{\mathbf{\Sigma}}^{-1} \hat{\boldsymbol{\mu}} \end{bmatrix} =: \begin{bmatrix} \hat{a} & \hat{b} \\ \hat{b} & \hat{c} \end{bmatrix}$$

  with determinant

$$\hat{d} := \det(\hat{\mathbf{\Omega}}) = \hat{a}\hat{c} - \hat{b}^2$$

- estimator of the global minimum variance portfolio:

$$\hat{\boldsymbol{x}}_{GMVP} = \frac{\hat{\mathbf{\Sigma}}^{-1} \mathbf{1}_n}{\hat{a}}$$

$$\hat{\mu}_{GMVP} = \frac{\hat{b}}{\hat{a}}$$

$$\hat{\sigma}^2_{GMVP} = \frac{1}{\hat{a}}$$

Flashback: Relevant Quantities VI

- estimator of the minimum variance set:

$$\hat{\sigma}^2_{MVS} = \frac{\hat{a}\mu^2 - 2\hat{b}\mu + \hat{c}}{\hat{d}}$$

  - with short-selling: define a grid of points $\mu \in \{\mu_{\text{min}}, \ldots, \mu_{\text{max}}\}$ where μ_{min} and μ_{max} can be chosen arbitrarily
  - without short-selling: define a grid of points $\mu \in \{\mu_{\text{min}}, \ldots, \mu_{\text{max}}\}$ where $\mu_{\text{min}} := \min_{\{1\leqslant i\leqslant n\}} \hat{\mu}_i$ and $\mu_{\text{max}} := \max_{\{1\leqslant i\leqslant n\}} \hat{\mu}_i$

- estimator of the minimum variance portfolio:

$$\hat{\boldsymbol{x}}_{MVS} = \hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{R}}\left(\hat{\boldsymbol{R}}^{\mathsf{T}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{R}}\right)^{-1}\boldsymbol{\mu} \qquad (3)$$

where

$$\hat{\boldsymbol{R}} = \begin{bmatrix} \mathbf{1}_n & \hat{\boldsymbol{\mu}} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\mu} = \begin{pmatrix} 1 \\ \mu \end{pmatrix}$$

Flashback: Relevant Quantities VII

- estimator of the tangential (market) portfolio:

$$\hat{\boldsymbol{x}}_m = \frac{\hat{\boldsymbol{\Sigma}}^{-1}\tilde{\hat{\boldsymbol{\mu}}}}{\hat{b} - \hat{a}r_f}$$

$$\hat{\mu}_m = \frac{\hat{c} - \hat{b}r_f}{\hat{b} - \hat{a}r_f}$$

$$\hat{\sigma}_m^2 = \frac{\hat{a}r_f^2 - 2\hat{b}r_f + \hat{c}}{(\hat{b} - \hat{a}r_f)^2}$$

with the estimator for expected excess returns

$$\tilde{\hat{\boldsymbol{\mu}}} := \begin{pmatrix} \hat{\mu}_1 - r_f \\ \hat{\mu}_2 - r_f \\ \vdots \\ \hat{\mu}_n - r_f \end{pmatrix} = \hat{\boldsymbol{\mu}} - r_f \mathbf{1}_n$$

Flashback: Relevant Quantities VIII

- alternatively, once we have computed the (σ, μ)-combinations on the upper branch of the MVS, i.e., the efficient frontier, the position of the market portfolio in the (σ, μ)-plane is simply that point on the efficient frontier which is associated to the maximal value of

$$\frac{\mu_{MVS}}{\sigma_{MVS}} - \frac{r_f}{\sigma_{MVS}} \qquad \text{for } \mu_{MVS} \geqslant \mu_{GMVP}$$

The Stock Data

- downloading each and every stock contained in the SP500 index from Yahoo! Finance is cumbersome
- fortunately, there is a nice Excel macro which does the job (you have to change the security settings such that the spreadsheet is "trusted")

$\Rightarrow$ all stock prices should be adjusted to dividend payments and stock splits

- time series on the SP500 index itself can be download directly from Yahoo! Finance
- the sampling frequency of all time series is daily

$\Rightarrow$ holding period is one day

- the sampling period for all time series is 01/02/2003 — 12/31/2012

$\Rightarrow$ these 10 years of data correspond to $T = 2516$ observations

- deleting all stocks with missing observations (trading halts, etc.), there are $n = 436$ stocks available for our portfolio

The Risk-Free Rate I

- with respect to the risk-free rate r_f, the overnight London Interbank Offered Rate (LIBOR) was downloaded from the database of the Federal Reserve Bank of St. Louis; see page 22

⇒ "*The overnight US Dollar (USD) LIBOR interest rate is the average interest rate at which a selection of banks in London are prepared to lend to one another in American dollars with a maturity of 1 day.*"

⇒ "*Libor is the most widely used "benchmark" or reference rate for short term interest rates.*"

- however, a Libor time series with thousands of observations is not useful because it cannot be plotted as $(0, r_f)$ in the (σ, μ)-diagram
- instead, use the sample overnight Libors to estimate an average overnight Libor over the sampling period
- flashback: since simple interest rates are stated as annualized rates, we have to adjust to the daily holding period by assuming that interest is earned linearly

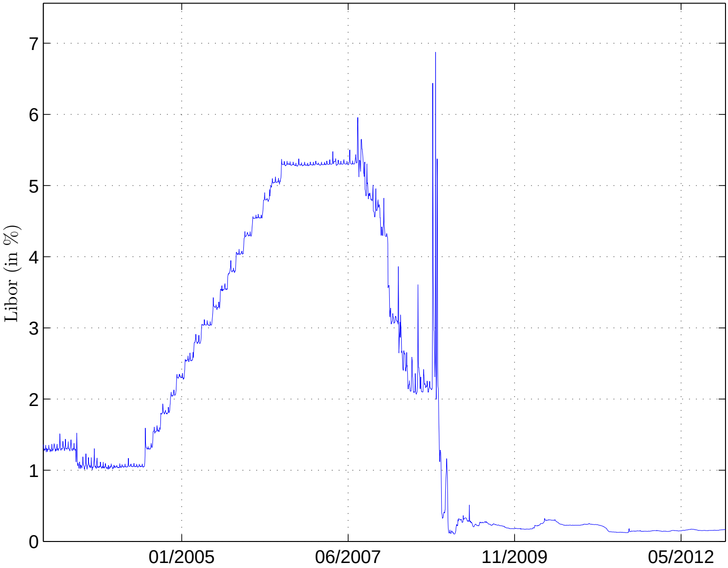
The Risk-Free Rate II

$\Rightarrow$ use the following proxy as risk-free interest rate:

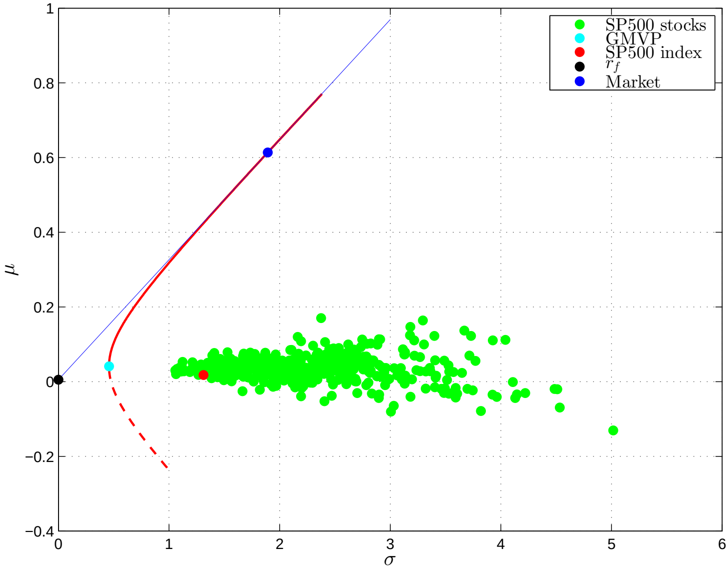
$$r_f := \frac{\frac{1}{T}\sum_{t=1}^{T} Libor_t}{365 \text{ days}} = 0.00524\%$$

- however, the choice of r_f is no essential issue
- in his original paper (with a monthly sampling frequency), Sharpe (1966) used an annual risk-free rate of 3% derived from secondary-market yields on 10-year US government bonds; see also Ruppert (2011, p. 306)
- in the financial industry, Sharpe ratios are usually published by setting $r_f = 0$
- idea:
  - Sharpe ratios should only reflect the pure risk-return trade-off μ_p/σ_p of corresponding portfolios
  - investors should decide by themselves which risk-free rate suits their purposes best

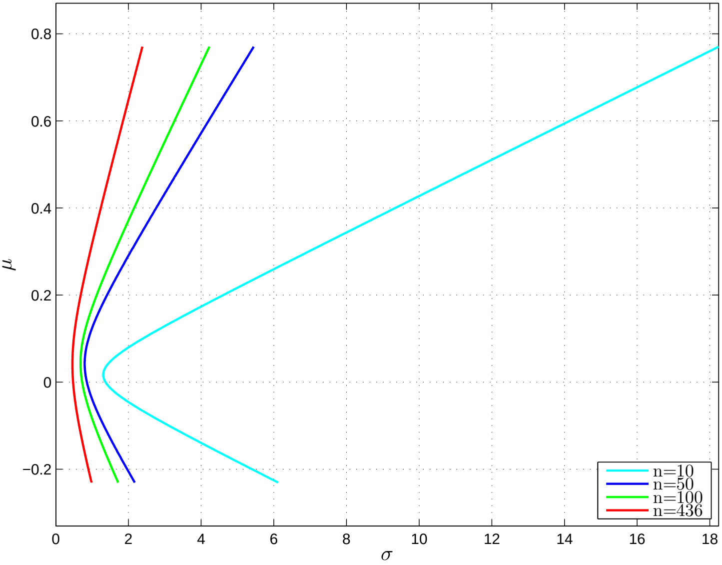
The Overnight Libor



Mean-Variance Analysis (with Short-Selling)



MVS for Most Liquid n Stocks (with Short-Selling)



No Short-Selling I

- in the theory part to the mean-variance analysis, we did not cover the general n-asset case

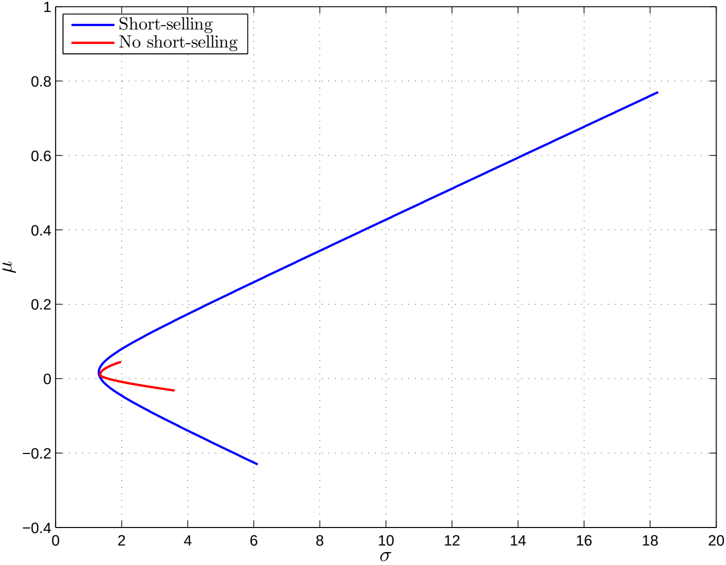
$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x} \\ \text{s.t.} \quad & \mathbf{1}_n^\top \boldsymbol{x} = 1 \\ & \boldsymbol{\mu}^\top \boldsymbol{x} = \bar{\mu} \\ & \boldsymbol{x} \geqslant \mathbf{0}_n \end{aligned}$$

  when short-selling is not allowed because it is tough to tackle in an analytical manner
- allowing for corner solutions in optimal portfolio weights requires a more general approach, known as Karush-Kuhn-Tucker conditions, which nests the Lagrange multiplier approach; see Sundaram (1999, Chap. 6) or Boyd and Vandenberghe (2004, Section 5.5.3)
- unfortunately, the Karush-Kuhn-Tucker conditions (though analytical) are not particularly helpful in delivering solutions but rather tell us whether a given point is a solution or not

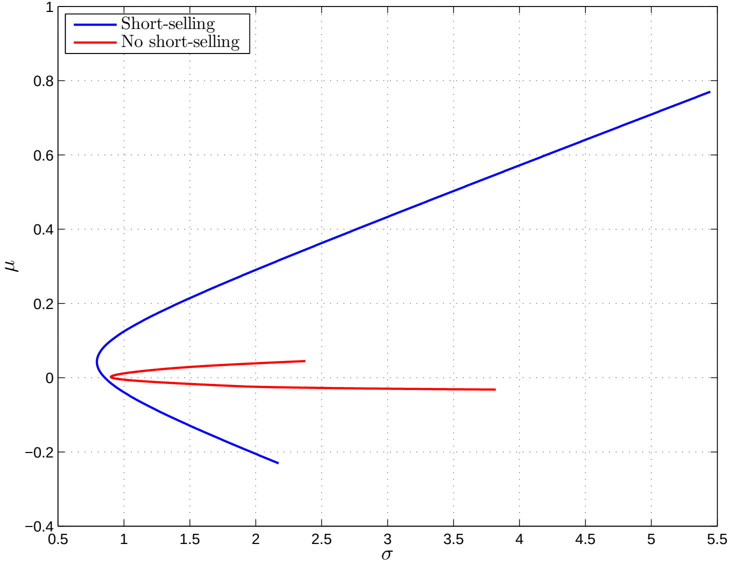
No Short-Selling II

- for obtaining numerical solutions, one has to resort to numerical methods known as quadratic programming
- most statistical software package include a function or routine to perform quadratic programming

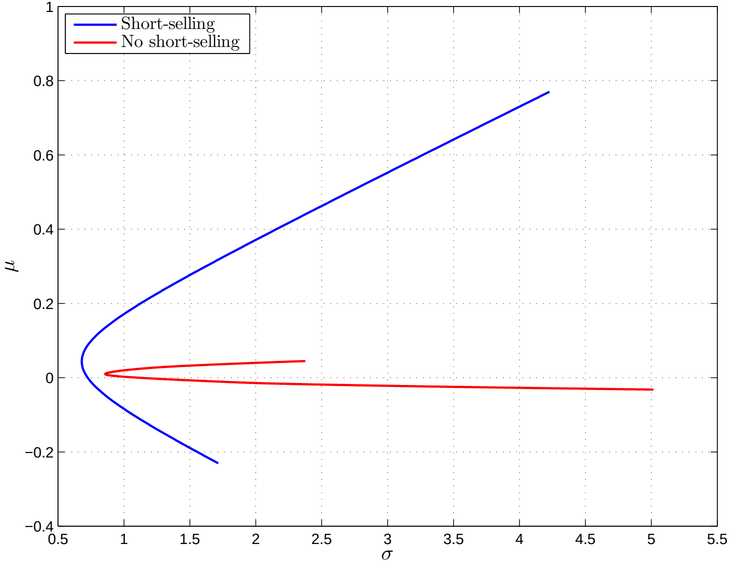
MVS Comparison for Most Liquid 10 Stocks



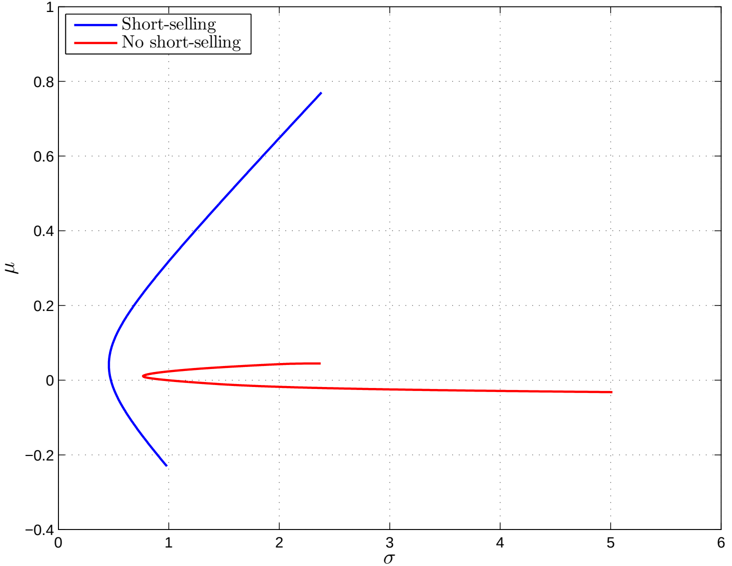
MVS Comparison for Most Liquid 50 Stocks



MVS Comparison for Most Liquid 100 Stocks



MVS Comparison for Most Liquid 436 Stocks



Monte Carlo (MC) Simulations I

- consider the classical linear regression model

$$\underbrace{\boldsymbol{y}}_{T\times 1} = \underbrace{\boldsymbol{X}}_{T\times k}\underbrace{\boldsymbol{\beta}}_{k\times 1} + \underbrace{\boldsymbol{u}}_{T\times 1} \tag{4}$$

where we make a host of assumptions, i.e.,

  - $\mathrm{rk}(\boldsymbol{X}) = k < T$
  - $\boldsymbol{u} \stackrel{d}{=} \mathrm{N}\big(\boldsymbol{0}_{T\times 1}, \sigma^2 \boldsymbol{I}_T\big)$,

in order to guarantee that the ordinary least-squares (OLS) estimator

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y} \tag{5}$$

exists and follows the probability law

$$\hat{\boldsymbol{\beta}} \stackrel{d}{=} \mathrm{N}\Big(\boldsymbol{\beta}, \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\Big)$$

which allows us to derive confidence intervals, t- and F-tests, etc.

Monte Carlo (MC) Simulations II

- unfortunately, these assumptions turn out to be unrealistically stringent casting doubts on the unbiasedness and efficiency of $\hat{\boldsymbol{\beta}}$ and on the validity of its confidence intervals and tests
- for relaxing some of these assumptions, one often resorts to asymptotic statistics and its limit laws, i.e., the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT), to establish consistency and asymptotic normality
- one particularly important result is that the normality of $\hat{\boldsymbol{\beta}}$ (asymptotically) carries over even when $\boldsymbol{u}$ is non-normally distributed (given its variance exists)
- but there is a price to be paid for this generality: limit laws hold exactly for $T \to \infty$ and approximately when the convergence has proceeded far enough
- in some instances, we have only a small sample size T, and even when T is large, we usually do not know if the convergence has proceeded far enough to provide a good approximation

$\Rightarrow$ we have to distinguish between the exact, finite-sample distribution of $\hat{\boldsymbol{\beta}}$ (which depends upon and holds for a fixed $T < \infty$ only) and the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ ("$T = \infty$")

Monte Carlo (MC) Simulations III

- thus, in cases where we suspect the asymptotic approximation to the (unknown) finite-sample distribution to be inaccurate, we need alternative ways of approximation:
  - the first proposal in the literature is purely theoretical and refers to so-called Edgeworth and saddlepoint approximations which are refinements of the asymptotic distribution allowing for more flexibility of the approximation to accommodate features of the finite-sample distribution; see Hall (1992), Rothenberg (1984), and Ullah (2004)
  - the second proposal is to use MC simulations to approximate the finite-sample distribution
- while the first approach is more general (laying out all required assumptions explicitly), it is much more complicated and harder to compute than MC simulations
- the popularity of MC simulations stems from its simplicity and wide applicability which only requires that we are able to simulate a model (or data-generating process) a larger number of times

Monte Carlo (MC) Simulations IV

- of course, the drawback of MC simulations (compared to the first approach) is that they only hold for the simulated models and that they cannot be generalized

$\Rightarrow$ MC simulations are case-by-case studies

How and Why MC Works: An Example I

- suppose that we have no clue about the finite-sample and asymptotic properties of $\hat{\beta}$
- in what follows, we will motivate and illustrate the usage of MC simulations to shed some light on these properties
- a first try would be to apply $\hat{\beta}$ to some real-world data
- however, this is not helping because we neither know whether the underlying model (4) is appropriate to be fitted to the data nor, if it were true, whether the true parameter β is unknown which prevents us from a comparison
- a better way to analyze the properties of $\hat{\beta}$ is by simulating an artificial data set from model (4) which is completely known
- but before we can start any simulation the free parameters in model (4) have to be fixed:
  - the number of exogenous regressor, say, $k = 3$
  - the sample size, say, $T = 25$

How and Why MC Works: An Example II

- set

$$\boldsymbol{\beta} = \begin{pmatrix} 1.5 \\ -0.5 \\ 1.75 \end{pmatrix} \quad \text{and} \quad \sigma^2 = 1.25$$

- lastly, in order to isolate the source of randomness, we assume that the regressor matrix $\boldsymbol{X}$ in (4) is fixed, i.e., deterministic $\Rightarrow$ this allows us to simplify the notation by using $\mathrm{E}[\,.\,]$ instead of $\mathrm{E}[\,.|\boldsymbol{X}\,]$
- in the subsequent simulations, we use the same regressor matrix $\boldsymbol{X}$ (for same T) generated by the following scheme:
  - the first column of $\boldsymbol{X}$ corresponds to a $(T \times 1)$ ones-vector $\mathbf{1}_{T\times 1}$
  - the second column of $\boldsymbol{X}$ corresponds to a $(T \times 1)$ vector of iid realizations drawn from the uniform distribution $\mathrm{Unif}(0,1)$
  - the third column of $\boldsymbol{X}$ corresponds to a $(T \times 1)$ vector of iid realizations drawn from the standard normal distribution $\mathrm{N}(\,0,1\,)$

How and Why MC Works: An Example III

- thus, the randomness of $\hat{\boldsymbol{\beta}}$ can be traced back to $\boldsymbol{u}$ by substituting (4) in (5):

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{u}$$
$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{u}$$

- to measure the discrepancy between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$, we use the usual Euclidean norm, i.e.,

$$||\boldsymbol{u}|| = \sqrt{u_1^2 + \cdots + u_T^2}$$

- for example, if we generated a $\boldsymbol{u}^{(j)}$ with a large $||\boldsymbol{u}^{(j)}||$, then the discrepancy $||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}||$ tends to be large as well
- this happens when $\boldsymbol{u}^{(j)}$ represents an extreme realization from $\mathrm{N}\left(\boldsymbol{0}_{T\times 1}, \sigma^2 \boldsymbol{I}_T \right)$

⇒ it turns out that the influence of these "bad" simulated data can be averaged out by repeated simulations (or resampling)

How and Why MC Works: An Example IV

- suppose we have drawn *m* samples $\boldsymbol{u}^{(1)}, \ldots, \boldsymbol{u}^{(m)}$ from $\mathrm{N}\big(\,\mathbf{0}_{T\times 1}, \sigma^2 \boldsymbol{I}_T\,\big)$ and subsequently computed $\boldsymbol{y}^{(j)}$ and $\hat{\boldsymbol{\beta}}^{(j)}$ with $j = 1, \ldots, m$ such that

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{(1)} &= \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{u}^{(1)} \\ &\vdots \\ \hat{\boldsymbol{\beta}}^{(m)} &= \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{u}^{(m)}\end{aligned}$$

can be averaged over all m equations:

$$\begin{aligned}\frac{1}{m}\sum_{j=1}^{m} \hat{\boldsymbol{\beta}}^{(j)} &= \frac{1}{m}\sum_{j=1}^{m} \boldsymbol{\beta} + \frac{1}{m}\sum_{j=1}^{m} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{u}^{(j)} \\ &= \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \left(\frac{1}{m}\sum_{j=1}^{m} \boldsymbol{u}^{(j)} \right) \qquad (6)\end{aligned}$$

How and Why MC Works: An Example V

- notice that (6) is the sample counterpart of

$$\mathrm{E}\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta} + (\boldsymbol{X}^\mathsf{T}\boldsymbol{X})^{-1}\boldsymbol{X}^\mathsf{T}\mathrm{E}[\,\boldsymbol{u}\,]$$

  which corresponds to

$$\mathrm{E}\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$$

  because $\mathrm{E}[\,\boldsymbol{u}\,] = \mathbf{0}_{T\times 1}$, by construction
- the only requirement for this result to hold is that

$$\frac{1}{m}\sum_{j=1}^{m}\boldsymbol{u}^{(j)} \xrightarrow{p} \mathrm{E}[\,\boldsymbol{u}\,]$$

  which, by the weak LLN, holds whenever $\mathrm{E}[\,\boldsymbol{u}\,] < \infty$

$\Rightarrow$ this is the basic idea of MC simulations!

- assume we have generated an MC sample of estimates $\left\{\hat{\theta}^{(j)}\right\}_{j=1}^{m}$ for a scalar parameter θ

How and Why MC Works: An Example VI

- a measure of unbiasedness of $\hat{\theta}$ is the MC sample mean

$$\bar{\hat{\theta}} := \frac{1}{m} \sum_{j=1}^{m} \hat{\theta}^{(j)}$$

- a measure of dispersion of $\hat{\theta}$ about its sample mean is the MC sample standard deviation

$$\widehat{\text{StD}}\left[\hat{\theta}\right] := \sqrt{\frac{1}{m} \sum_{j=1}^{m} \left(\hat{\theta}^{(j)} - \bar{\hat{\theta}}\right)^2}$$

- a combined measure of bias and estimation error of $\hat{\theta}$ is the MC sample mean-squared error

$$\begin{aligned} \widehat{\text{MSE}}\left[\hat{\theta}\right] &:= \left(\bar{\hat{\theta}} - \theta\right)^2 + \frac{1}{m} \sum_{j=1}^{m} \left(\hat{\theta}^{(j)} - \bar{\hat{\theta}}\right)^2 \\ &= \widehat{\text{Bias}}\left[\hat{\theta}\right]^2 + \widehat{\text{StD}}\left[\hat{\theta}\right]^2 \end{aligned}$$

How and Why MC Works: An Example VII

- thus, for $m \to \infty$, we obtain

$$\bar{\hat{\theta}} \xrightarrow{P} \mathsf{E}\left[\hat{\theta}\right]$$

$$\widehat{\mathsf{StD}}\left[\hat{\theta}\right] \xrightarrow{P} \mathsf{StD}\left[\hat{\theta}\right]$$

$$\widehat{\mathsf{MSE}}\left[\hat{\theta}\right] \xrightarrow{P} \mathsf{MSE}\left[\hat{\theta}\right]$$

- although $m \to \infty$ is impossible for numerical simulations on a computer, we can nevertheless expect that these convergence results to hold pretty well for large m, say $m = 1000$
- as an indication of weak consistency, we need to find evidence that

$$\widehat{\mathsf{MSE}}\left[\hat{\theta}\right] \to 0$$

as $T \to \infty$ because "$\xrightarrow{m.s.}$" $\Rightarrow$ "$\xrightarrow{P}$"

How and Why MC Works: An Example VIII

- as an indication of convergence in distribution, we need to find evidence that the nonparametric kernel density estimate of a MC sample for $\hat{\theta}$ converges to the asymptotic distribution of $\hat{\theta}$ as $T \to \infty$
- digression: the CLT for $\hat{\boldsymbol{\beta}}$ states that

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathrm{N}\Big(\mathbf{0}_{k\times 1}, \sigma^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \Big)$$

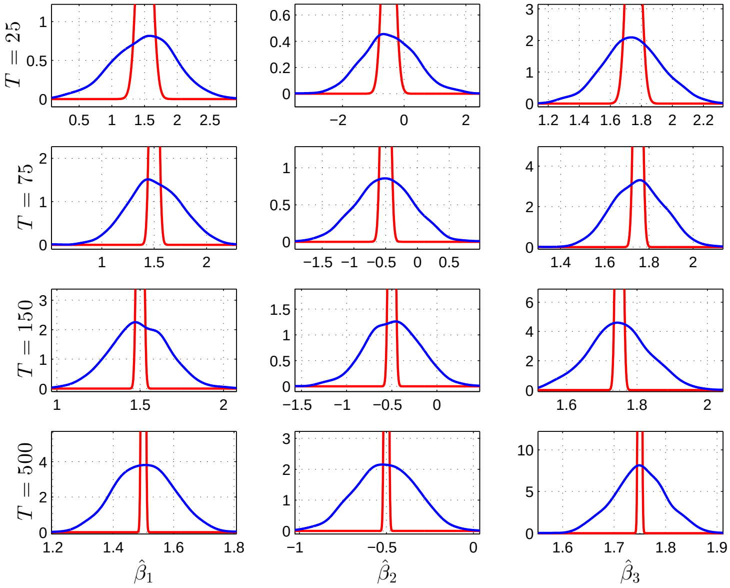
or

$$\hat{\boldsymbol{\beta}} \overset{asy}{=} \mathrm{N}\Big(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}/T \Big)$$

MC Results: Classical Linear Regression Model

		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
$T = 25$	$\bar{\hat{\beta}}_i$	1.4924	-0.4743	1.7429
	$\widehat{\text{StD}}$	0.4726	0.8544	0.1880
	$\widehat{\text{MSE}}$	0.2234	0.7307	0.0354
$T = 75$	$\bar{\hat{\beta}}_i$	1.5040	-0.5164	1.7510
	$\widehat{\text{StD}}$	0.2558	0.4457	0.1167
	$\widehat{\text{MSE}}$	0.0655	0.1990	0.0136
$T = 150$	$\bar{\hat{\beta}}_i$	1.4948	-0.4928	1.7482
	$\widehat{\text{StD}}$	0.1730	0.3000	0.0861
	$\widehat{\text{MSE}}$	0.0300	0.0901	0.0074
$T = 500$	$\bar{\hat{\beta}}_i$	1.5026	-0.5066	1.7490
	$\widehat{\text{StD}}$	0.0951	0.1705	0.0505
	$\widehat{\text{MSE}}$	0.0091	0.0291	0.0025

Density Comparison (CLT=red, MC=blue)



MC Simulation: Non-Normal Error Terms

- now, let us analyze the effects of non-normal $\boldsymbol{u}$
- assume that the error terms are iid draws from a non-central t-distribution, i.e., $\forall\, t = 1, \ldots, T$

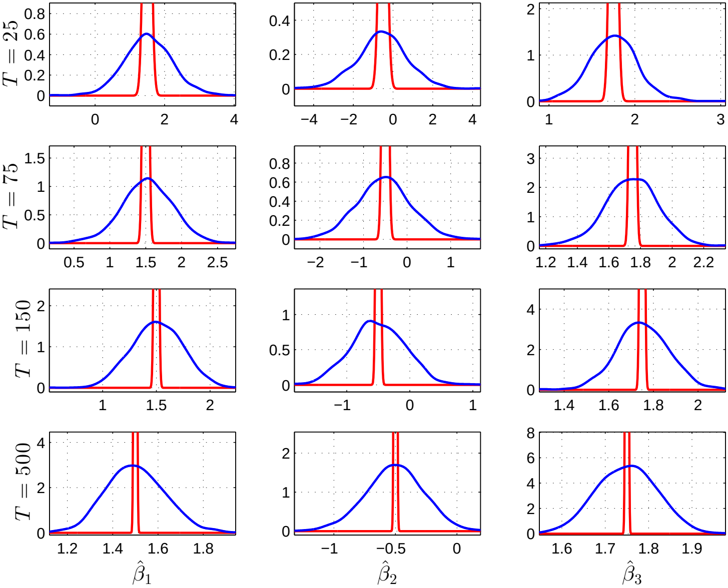
$$u_t \overset{\text{iid}}{=} t_{\nu,\delta}$$

- ν is the parameter of the degrees of freedom
  - the smaller ν the more leptokurtic $t_{\nu,\delta}$
  - the variance of $t_{\nu,\delta}$ even does not exist for $\nu < 2$
  - $t_{\nu,\delta}$ converges to a normal distribution for $\nu \to \infty$
- δ is the non-centrality parameter, i.e., the distribution is centered at the origin or $\mathrm{E}[\, t_{\nu,\delta}\,] = 0$ when $\delta = 0$
- consider two cases:
  1) $\nu = 4$ and $\delta = 0$
  2) $\nu = 4$ and $\delta = 1$
- question: what do you conclude?

MC Results: $u \overset{iid}{=} t_{\nu,\delta}$ with $\nu = 4$ and $\delta = 0$

		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
$T = 25$	$\bar{\hat{\beta}}_i$	1.4850	-0.4920	1.7690
	$\widehat{\text{StD}}$	0.6860	1.2406	0.2737
	$\widehat{\text{MSE}}$	0.4708	1.5392	0.0753
$T = 75$	$\bar{\hat{\beta}}_i$	1.4925	-0.4970	1.7477
	$\widehat{\text{StD}}$	0.3446	0.5952	0.1679
	$\widehat{\text{MSE}}$	0.1188	0.3543	0.0282
$T = 150$	$\bar{\hat{\beta}}_i$	1.4995	-0.5111	1.7481
	$\widehat{\text{StD}}$	0.2483	0.4302	0.1184
	$\widehat{\text{MSE}}$	0.0616	0.1852	0.0140
$T = 500$	$\bar{\hat{\beta}}_i$	1.4928	-0.4874	1.7470
	$\widehat{\text{StD}}$	0.1367	0.2417	0.0713
	$\widehat{\text{MSE}}$	0.0188	0.0586	0.0051

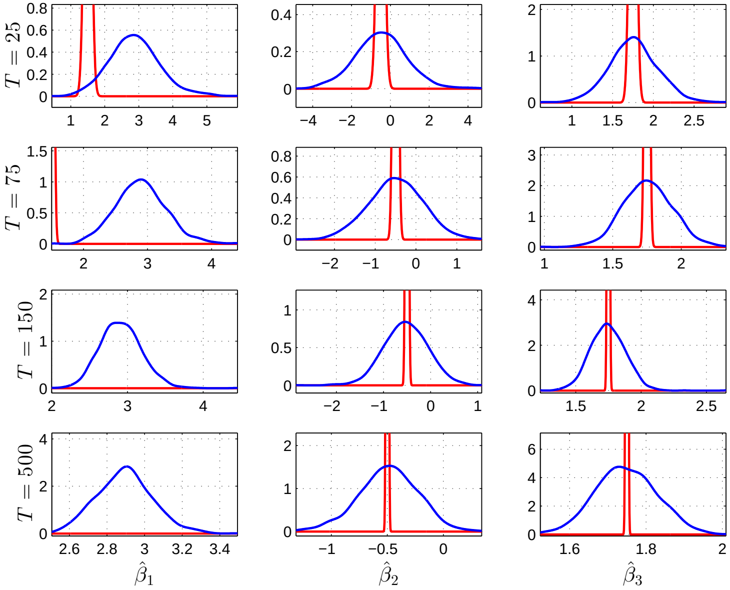
Density Comparison (CLT=red, MC=blue)



MC Results: $u \overset{\text{iid}}{=} t_{\nu,\delta}$ with $\nu = 4$ and $\delta = 1$

		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
$T = 25$	$\bar{\hat{\beta}}_i$	2.8980	-0.4878	1.7632
	$\widehat{\text{StD}}$	0.7289	1.3123	0.2913
	$\widehat{\text{MSE}}$	2.4858	1.7224	0.0851
$T = 75$	$\bar{\hat{\beta}}_i$	2.9131	-0.5117	1.7668
	$\widehat{\text{StD}}$	0.4150	0.7232	0.1832
	$\widehat{\text{MSE}}$	2.1692	0.5231	0.0338
$T = 150$	$\bar{\hat{\beta}}_i$	2.9091	-0.5048	1.7469
	$\widehat{\text{StD}}$	0.2771	0.4625	0.1340
	$\widehat{\text{MSE}}$	2.0622	0.2140	0.0180
$T = 500$	$\bar{\hat{\beta}}_i$	2.9030	-0.5022	1.7554
	$\widehat{\text{StD}}$	0.1531	0.2607	0.0768
	$\widehat{\text{MSE}}$	1.9918	0.0680	0.0059

Density Comparison (CLT=red, MC=blue)



MC Simulation: Estimation Error of MV Approach I

- now, we illustrate the usage of MC techniques to get an idea of the estimation error when the plug-in estimator is applied to the MV approach
- at this stage, we stick to all assumptions underlying the MV approach, i.e., returns are drawn from a multivariate normal distribution with parameters $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$:

$$\underbrace{\boldsymbol{r}_t}_{n\times 1} \stackrel{\text{iid}}{=} \mathrm{N}(\,\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\,)$$

  where $\boldsymbol{r}_t^{\mathsf{T}} := (\, r_{t,1} \;\cdots\; r_{t,n}\,)$
- again, since the true parameters, $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$, of the data-generating process are unknown, potential deficiencies of the plug-in estimator cannot be evaluated

MC Simulation: Estimation Error of MV Approach II

- the solution is to set

$$\boldsymbol{\mu}_0 := \hat{\boldsymbol{\mu}}$$
$$\boldsymbol{\Sigma}_0 := \hat{\boldsymbol{\Sigma}} \,,$$

  where $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are computed from real data, and to simulate m iid samples $\boldsymbol{r}_t^{(1)}, \dots, \boldsymbol{r}_t^{(m)}$ from $\mathrm{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$
- since any basic (pseudo-)normal random number generator can produce arbitrary arrays of iid realizations of standard normal variables, the only challenge here is to introduce the dependence structure between stock returns in $\boldsymbol{r}_t$ implied by $\boldsymbol{\Sigma}_0$

$\Rightarrow$ this works pretty much as in the scalar case, except that we need a multivariate version of the square root of a variance

MC Simulation: Estimation Error of MV Approach III

- Cholesky decomposition: any symmetric, positive definite matrix $\boldsymbol{\Sigma}_0$ can be uniquely decomposed as

$$\boldsymbol{\Sigma}_0 = \boldsymbol{C}^\mathsf{T} \boldsymbol{C}$$

where $\boldsymbol{C}$ is an upper triangular matrix which can be interpreted as the square-root matrix of $\boldsymbol{\Sigma}_0$

- then, $\boldsymbol{r}_t \overset{\text{iid}}{=} \mathrm{N}(\,\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\,)$ can be computed via

$$\underbrace{\boldsymbol{r}_t}_{n\times 1} = \underbrace{\boldsymbol{\mu}_0}_{n\times 1} + \underbrace{\boldsymbol{C}^\mathsf{T}}_{n\times n} \underbrace{\boldsymbol{z}}_{n\times 1} \tag{7}$$

where we only have to simulate $\boldsymbol{z} \overset{\text{iid}}{=} \mathrm{N}(\,\boldsymbol{0}_{n\times 1}, \boldsymbol{I}_n\,)$

MC Simulation: Estimation Error of MV Approach IV

- to see that (7) does the job, note that $\boldsymbol{r}_t$ is multivariate normally distributed, because it is a linear combination of normally distributed random variables in $\boldsymbol{z}$, with mean

$$\mathrm{E}[\,\boldsymbol{r}_t\,] = \mathrm{E}\Big[\,\boldsymbol{\mu}_0 + \boldsymbol{C}^{\mathsf{T}}\boldsymbol{z}\,\Big] = \mathrm{E}[\,\boldsymbol{\mu}_0\,] + \mathrm{E}\Big[\,\boldsymbol{C}^{\mathsf{T}}\boldsymbol{z}\,\Big] = \boldsymbol{\mu}_0 + \boldsymbol{C}^{\mathsf{T}}\underbrace{\mathrm{E}[\,\boldsymbol{z}\,]}_{=\boldsymbol{0}_{n\times 1}} = \boldsymbol{\mu}_0$$

  and covariance matrix

$$\begin{aligned}\mathrm{E}\Big[\,(\boldsymbol{r}_t - \boldsymbol{\mu}_0)(\boldsymbol{r}_t - \boldsymbol{\mu}_0)^{\mathsf{T}}\,\Big] &= \mathrm{E}\Big[\,(\boldsymbol{\mu}_0 + \boldsymbol{C}^{\mathsf{T}}\boldsymbol{z} - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_0 + \boldsymbol{C}^{\mathsf{T}}\boldsymbol{z} - \boldsymbol{\mu}_0)^{\mathsf{T}}\,\Big] \\ &= \mathrm{E}\Big[\,\boldsymbol{C}^{\mathsf{T}}\boldsymbol{z}(\boldsymbol{C}^{\mathsf{T}}\boldsymbol{z})^{\mathsf{T}}\,\Big] = \mathrm{E}\Big[\,\boldsymbol{C}^{\mathsf{T}}\boldsymbol{z}\boldsymbol{z}^{\mathsf{T}}\boldsymbol{C}\,\Big] \\ &= \boldsymbol{C}^{\mathsf{T}}\underbrace{\mathrm{E}\Big[\,\boldsymbol{z}\boldsymbol{z}^{\mathsf{T}}\,\Big]}_{=\boldsymbol{I}_n}\boldsymbol{C} = \boldsymbol{C}^{\mathsf{T}}\boldsymbol{C} = \boldsymbol{\Sigma}_0\end{aligned}$$

- although (7) tells us how to simulate along the cross-sectional dimension of the data, we also need to simulate its time series observations

MC Simulation: Estimation Error of MV Approach V

- to this end, transpose (7),

$$\underbrace{\boldsymbol{r}_t^{\top}}_{1\times n} = \underbrace{\boldsymbol{\mu}_0^{\top}}_{1\times n} + \underbrace{\boldsymbol{z}^{\top}}_{1\times n}\underbrace{\boldsymbol{C}}_{n\times n},$$

  simulate this T-times, and stack all T time series observations

$$\underbrace{\boldsymbol{R}}_{T\times n} = \underbrace{\boldsymbol{M}}_{T\times n} + \underbrace{\boldsymbol{Z}}_{T\times n}\underbrace{\boldsymbol{C}}_{n\times n} \tag{8}$$

  where $\boldsymbol{Z}$ is a $(T \times n)$ matrix of iid realizations drawn from a standard normal distribution

$$\boldsymbol{M} = \left.\begin{bmatrix} \boldsymbol{\mu}_0^{\top} \\ \vdots \\ \boldsymbol{\mu}_0^{\top} \end{bmatrix}\quad\right\} T \text{ times}$$

- simulate (8) in order to generate m samples $\boldsymbol{R}^{(1)}, \ldots, \boldsymbol{R}^{(m)}$ which are used in (1) and (2) to compute $\hat{\boldsymbol{\mu}}^{(j)}$ and $\hat{\boldsymbol{\Sigma}}^{(j)}$ for $j = 1, \ldots, m$

MC Simulation: Estimation Error of MV Approach VI

- in the following MC simulations, we analyze two important quantities
  - the first one is functional: the MVS (or the efficient frontier)
  - the second one is scalar: the market portfolio's Sharpe ratio sr_m

MC Simulation: Parameter Comparison for $n = 10$ and $m = 250$

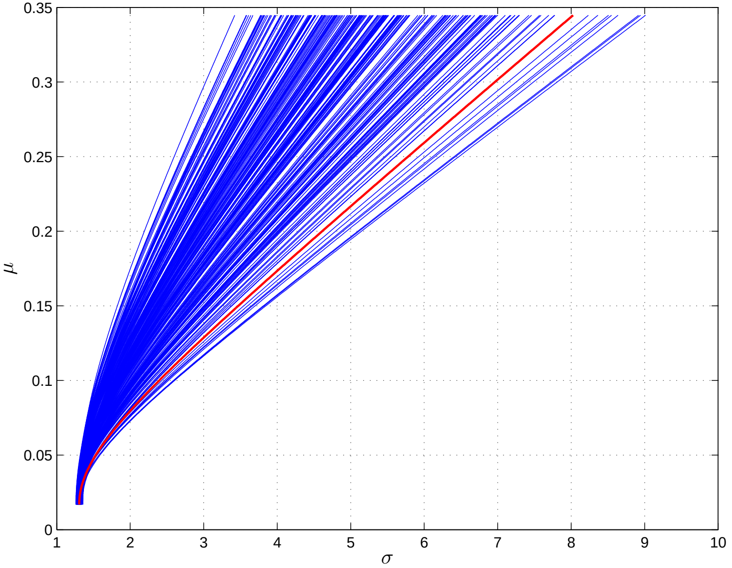
$$\mu_0^{\mathsf{T}} = \begin{pmatrix} -0.0321 & 0.0160 & 0.0167 & 0.0056 & 0.0173 & 0.0328 & 0.0108 & 0.0447 & 0.0064 & 0.0246 \end{pmatrix}$$

$$\overline{\hat{\mu}^{(j)}}^{\mathsf{T}} = \begin{pmatrix} -0.0368 & 0.0139 & 0.0113 & 0.0027 & 0.0141 & 0.0283 & 0.0070 & 0.0412 & 0.0047 & 0.0216 \end{pmatrix}$$

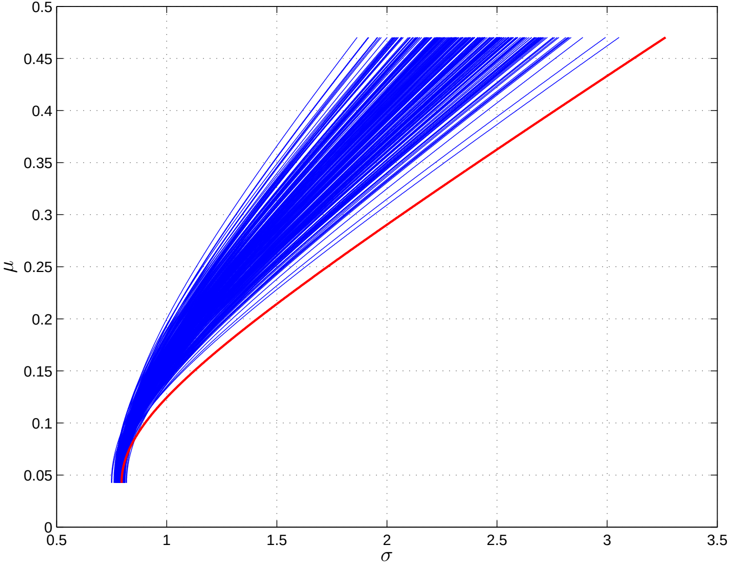
$$\Sigma_0 = \begin{bmatrix}
12.9920 & 3.2504 & 4.7579 & 4.5281 & 3.0903 & 7.9347 & 2.6418 & 2.8773 & 2.4008 & 8.7937 \\
3.2504 & 4.2127 & 2.5211 & 2.0216 & 2.5481 & 2.8176 & 2.0916 & 2.4087 & 1.3831 & 2.5522 \\
4.7579 & 2.5211 & 9.3836 & 2.7482 & 2.5830 & 3.5506 & 2.1450 & 2.2811 & 1.5076 & 3.8168 \\
4.5281 & 2.0216 & 2.7482 & 3.8903 & 2.0742 & 3.5659 & 1.6626 & 1.9062 & 1.4897 & 3.6964 \\
3.0903 & 2.5481 & 2.5830 & 2.0742 & 4.0095 & 2.7472 & 2.1082 & 2.3400 & 1.3486 & 2.6207 \\
7.9347 & 2.8176 & 3.5506 & 3.5659 & 2.7472 & 7.5972 & 2.3025 & 2.5591 & 1.9467 & 6.6633 \\
2.6418 & 2.0916 & 2.1450 & 1.6626 & 2.1082 & 2.3025 & 2.9597 & 2.0033 & 1.1844 & 2.2235 \\
2.8773 & 2.4087 & 2.2811 & 1.9062 & 2.3400 & 2.5591 & 2.0033 & 3.9275 & 1.2618 & 2.2610 \\
2.4008 & 1.3831 & 1.5076 & 1.4897 & 1.3486 & 1.9467 & 1.1844 & 1.2618 & 2.3822 & 1.9222 \\
8.7937 & 2.5522 & 3.8168 & 3.6964 & 2.6207 & 6.6633 & 2.2235 & 2.2610 & 1.9222 & 8.6770
\end{bmatrix}$$

$$\overline{\hat{\Sigma}^{(j)}} = \begin{bmatrix}
13.0130 & 3.2357 & 4.7750 & 4.5363 & 3.0857 & 7.9431 & 2.6373 & 2.8752 & 2.4014 & 8.8199 \\
3.2357 & 4.2089 & 2.5218 & 2.0110 & 2.5481 & 2.8030 & 2.0907 & 2.4065 & 1.3860 & 2.5507 \\
4.7750 & 2.5218 & 9.3972 & 2.7531 & 2.5874 & 3.5612 & 2.1443 & 2.2866 & 1.5147 & 3.8309 \\
4.5363 & 2.0110 & 2.7531 & 3.8885 & 2.0724 & 3.5647 & 1.6645 & 1.9035 & 1.4926 & 3.7058 \\
3.0857 & 2.5481 & 2.5874 & 2.0724 & 4.0119 & 2.7451 & 2.1129 & 2.3399 & 1.3499 & 2.6258 \\
7.9431 & 2.8030 & 3.5612 & 3.5647 & 2.7451 & 7.5936 & 2.2955 & 2.5500 & 1.9462 & 6.6754 \\
2.6373 & 2.0907 & 2.1443 & 1.6645 & 2.1129 & 2.2955 & 2.9637 & 2.0052 & 1.1857 & 2.2231 \\
2.8752 & 2.4065 & 2.2866 & 1.9035 & 2.3399 & 2.5500 & 2.0052 & 3.9231 & 1.2660 & 2.2648 \\
2.4014 & 1.3860 & 1.5147 & 1.4926 & 1.3499 & 1.9462 & 1.1857 & 1.2660 & 2.3817 & 1.9268 \\
8.8199 & 2.5507 & 3.8309 & 3.7058 & 2.6258 & 6.6754 & 2.2231 & 2.2648 & 1.9268 & 8.7035
\end{bmatrix}$$

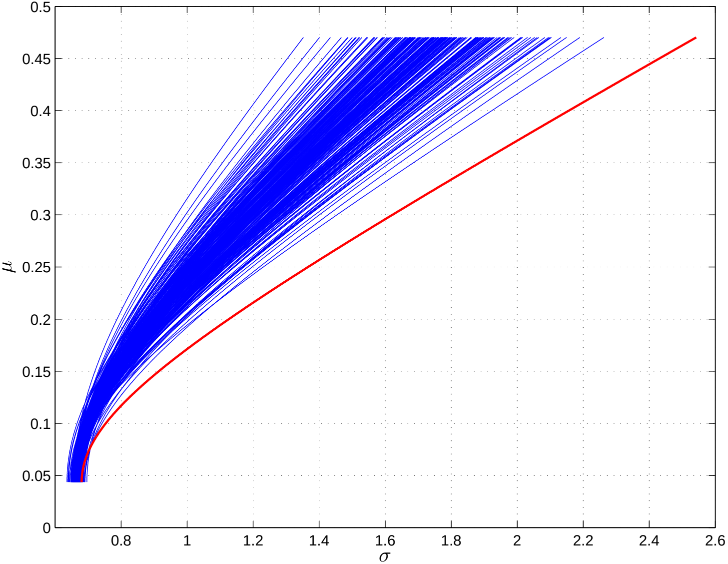
MC Simulation of Efficient Frontier ($n = 10$, $m = 250$)



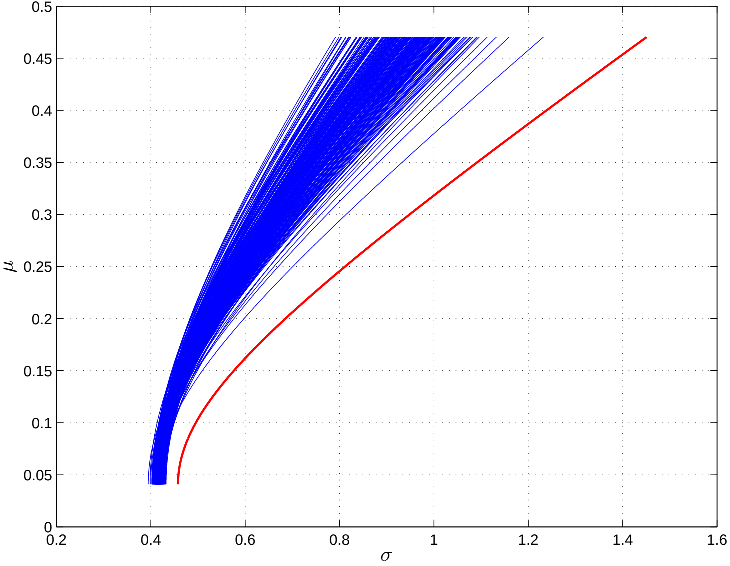
MC Simulation of Efficient Frontier ($n = 50$, $m = 250$)



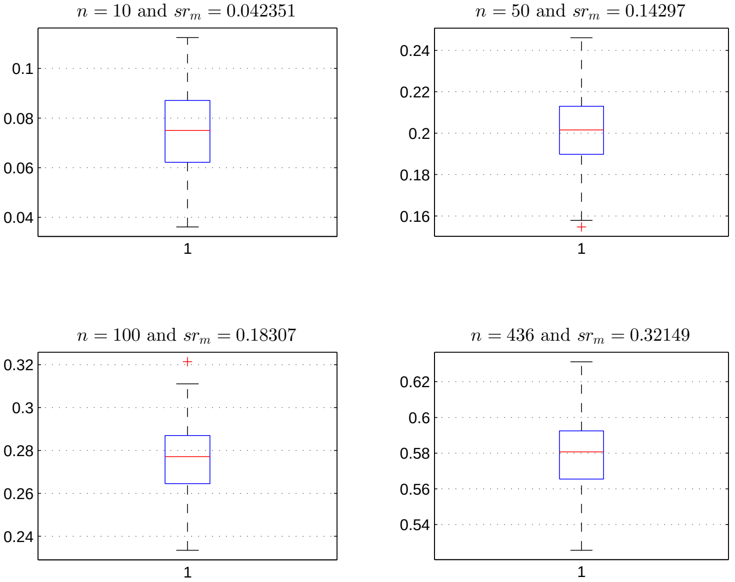
MC Simulation of Efficient Frontier ($n = 100$, $m = 250$)



MC Simulation of Efficient Frontier ($n = 436$, $m = 250$)



MC Simulation of Sharpe Ratios ($m = 250$)



Bibliography I

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley.

Huber, P. J. (1981). *Robust Statistics*. Wiley.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, **7** 77–91.

Rothenberg, T. J. (1984). Approximating the distribution of econometric estimators and test statistics. In *Handbook of Econometrics* (Z. Griliches and M. D. Intriligator, eds.), vol. 2. North-Holland, 881–935.

Ruppert, D. (2011). *Statistics and Data Analysis for Financial Engineering*. Springer.

Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business*, **39** 119–138.

Sundaram, R. (1999). *A First Course in Optimization Theory*. reprinted ed. Cambridge University Press.

Ullah, A. (2004). *Finite Sample Econometrics*. Oxford University Press.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.