Financial Econometrics: Portfolio Analysis

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 11 & 18, 2014

Outline

Motivation

Roadmap

- implementing the theory of portfolio selection isn't straightforward, because there are delicate issues when 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), toward its implementation 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:

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

$$\hat{\sigma}_{ij} = \frac{1}{T-1} \sum_{t=1}^{T} (r_{t,i} - \hat{\mu}_i)(r_{t,j} - \hat{\mu}_j) \tag{2}$$

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

for a sample of returns $\{r_{t,i} : 1 \leqslant t \leqslant T, 1 \leqslant i \leqslant n\}$

The Obvious Thing to Do

- 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 to show via the delta method (van der Vaart, 1998, Chap. 3) that properties, like consistency and asymptotic distributions, of many other quantities $\Phi(\hat{\theta})$ can be straightforwardly derived from those of the plug-in estimator $\hat{\theta}$

Caveats

- the mean-variance approach is static, i.e., it's a one-shot game
- but in economic systems, most things change in the course of time
- question: do expectations about the future change as well?
- answer: if yes, then quantities like $\mathsf{E}[\, r_{t+1}\,]$ or $\sigma^2 = \mathsf{E}\big[\,(r_{t+1} - \mathsf{E}[\, r_{t+1}\,])^2\,\big]$ should account for new information arriving during the holding period $[t, t+1]$, i.e., $\mathsf{E}_t[\, r_{t+1}\,]$ or $\sigma^2_{t+1} := \mathsf{E}_t\big[\,(r_{t+1} - \mathsf{E}_t[\, r_{t+1}\,])^2\,\big]$ seem more appropriate

⇒ replace unconditional expectations $\mathsf{E}[\,.\,]$ by conditional expectations $\mathsf{E}_t[\,.\,]$

- question: but does this matter for real-world applications?
- answer: no and yes...

⇒ not (so much) for the first moment of r_t

  - 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

$$\ln P_t = \alpha + \ln P_{t-1} + \epsilon_t \qquad \text{with } \mathsf{E}[\,\epsilon_t\,] = 0 \text{ for all } t$$

Caveats

- any expectation (be it conditional or unconditional) of returns is just a constant, i.e., no exploitable dynamics

⇒ this is completely different for the second moments of r_t

- another stylized facts of financial markets is known as volatility clustering of returns
- 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" prevailing during the holding period

- question: something else to worry about?
- answer: yes, some statistical issues...

⇒ yet another stylized facts of financial markets states that returns are not normally distributed

- empirical distributions of stock returns are not symmetric, i.e., they are left-skewed
- looking at the history of returns, losses during stock market crashes are much larger (in magnitude) than the biggest gains realized in good times
- empirical distributions of returns are leptokurtic, i.e., more peaked around the mean and with fat tails

Caveats

- parametric distributions allowing for fat tails feature the potential pitfall that the variance may not exist ($\sigma^2 = \infty$) which is obviously a huge problem for the mean-variance approach
- fat tails are an artifact of extreme realizations (outliers) which the normal distributions cannot generate (in a reasonable amount of time)
- important: assuming normally distributed returns underestimates the risk of large losses which an investor faces in real markets
- it's well known from the statistical literature that the standard sample moments $\hat{\mu}$ and $\hat{\sigma}^2$ in (1) are quite unreliable $\Rightarrow$ fitted $\mathsf{N}\big(\hat{\mu}, \hat{\sigma}^2\big)$ can be substantially distorted, see Ruppert (2011, Chap. 4)
- as a remedy, robust alternatives to $\hat{\mu}$ and $\hat{\sigma}^2$ have been proposed

$$\hat{\mu}_{\text{robust}} = \text{med}(r_t)$$

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

  see Huber (1981) or Hampel et al. (1986)
- robust sample moments are either downweighting or completely eliminating the impact of outliers, but that might not be a good thing to do when caring about risk beyond variances

Caveats

- an statistically ideal approach would use robust mean and variances for modeling the distributional properties of returns around the central region and tools from extreme value theory for modeling the tails... which doesn't fit in the mean-variance framework

The SP500 Index

- SP500 is a value-weighted index comprising the stocks of the 500 biggest U.S. companies; see p. 12
- it's considered to be one of the most important overall stock market indices for the U.S. economy
- this index can be interpreted as a portfolio of constituent stocks
- weights are computed as

$$w_i = \frac{n_i P_i}{\sum_{j=1}^{500} n_j P_j}$$

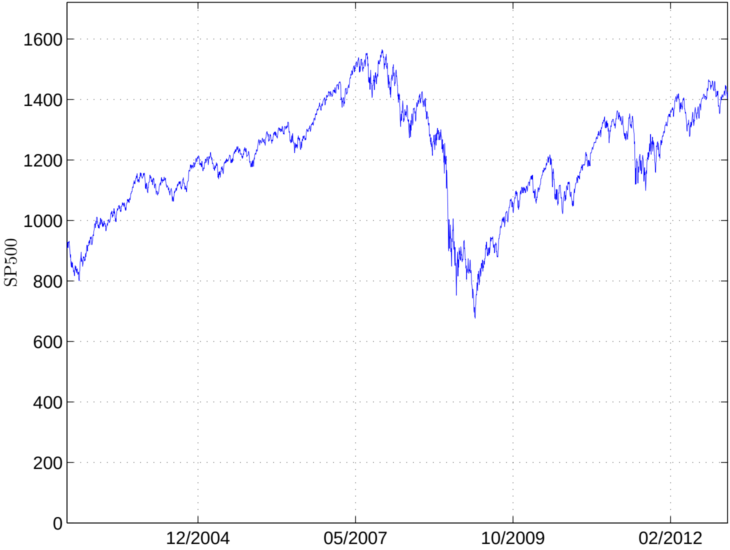
where n_i is stock i's number of common shares outstanding and P_i is the share price of stock i

- value of the SP500 index is computed as

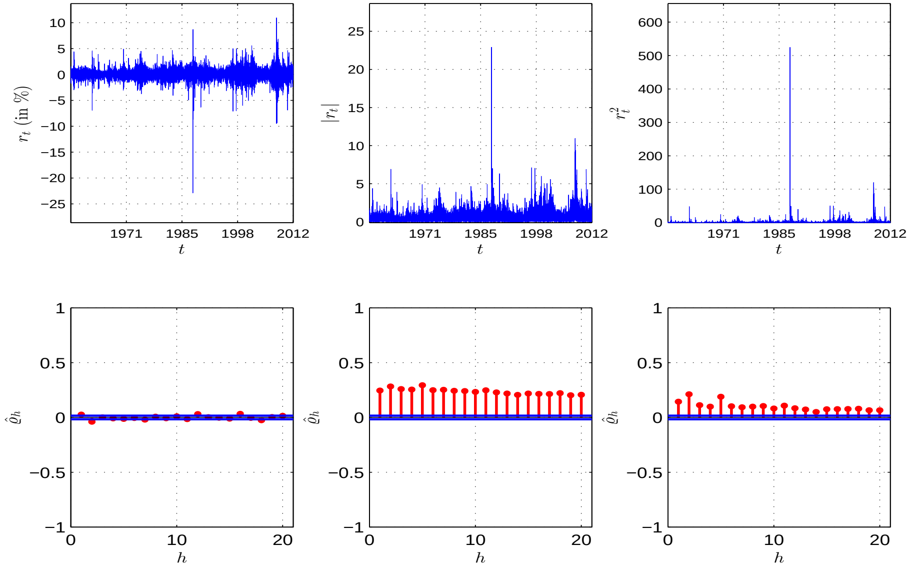
$$\text{SP500} = \sum_{j=1}^{500} w_j P_j$$

- the aforementioned stylized facts of financial markets are still present in this weighted average of individual stocks; see pp. 13–14

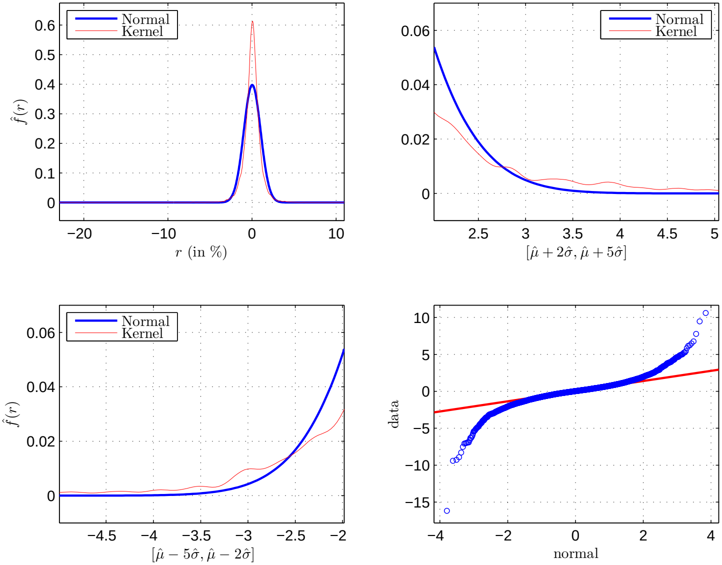
The SP500 Index



Random-Walk Property and Volatility Clustering in SP500 Returns



Non-Normality and Fat Tails in SP500 Returns



An Example: The SP500 Portfolio

Flashback: Relevant Quantities

- 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

- sample means vector of returns:

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

$$= \frac{1}{T}\begin{bmatrix} \boldsymbol{r}_1^{\mathsf{T}} \\ \vdots \\ \boldsymbol{r}_n^{\mathsf{T}} \end{bmatrix}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \frac{\boldsymbol{R}^{\mathsf{T}}\mathbf{1}_T}{T} \qquad (3)$$

Flashback: Relevant Quantities

- sample covariance matrix of returns:

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

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

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

Flashback: Relevant Quantities

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

$$=: \frac{1}{T-1}\begin{bmatrix} \tilde{\boldsymbol{r}}_1^\mathsf{T} \\ \vdots \\ \tilde{\boldsymbol{r}}_n^\mathsf{T} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{r}}_1 & \cdots & \tilde{\boldsymbol{r}}_n \end{bmatrix} =: \frac{\tilde{\boldsymbol{R}}^\mathsf{T}\tilde{\boldsymbol{R}}}{T-1} \tag{4}$$

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

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

is the ith unit vector

Flashback: Relevant Quantities

- 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

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

- 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

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

- estimator of the minimum variance portfolio:

$$\hat{\boldsymbol{x}}_{MVS} = \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{R}} \left(\hat{\boldsymbol{R}}^{\mathsf{T}} \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{R}} \right)^{-1} \boldsymbol{\mu} \tag{5}$$

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

- 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

- 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

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

which is useful when picking the tangential portfolio without short-selling

The Stock Data

- downloading each and every stock contained in the SP500 index from Yahoo! Finance manually 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”)

⇒ 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

⇒ holding period is one day

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

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

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

The Risk-Free Rate

- 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 p. 27

⇒ "*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 the unique point $(0, r_f)$ in the (σ, μ)-diagram ⇒ we need one single value for r_f
- instead, use 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

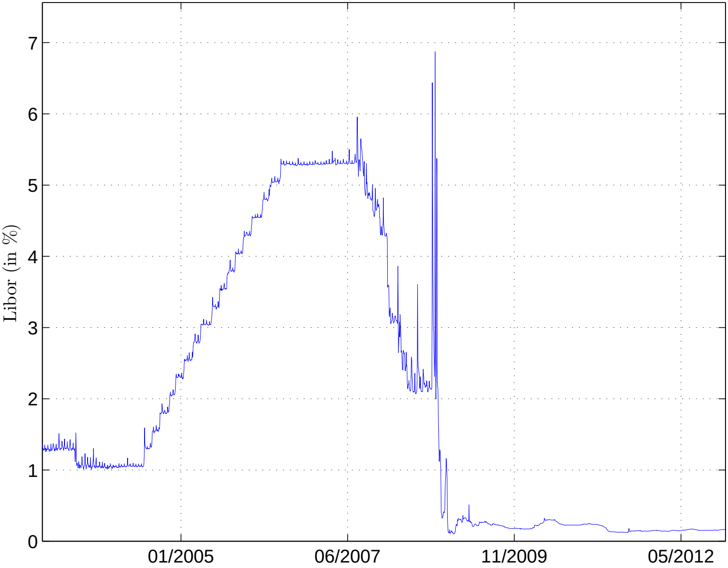
The Risk-Free Rate

⇒ use the following proxy as risk-free interest rate:

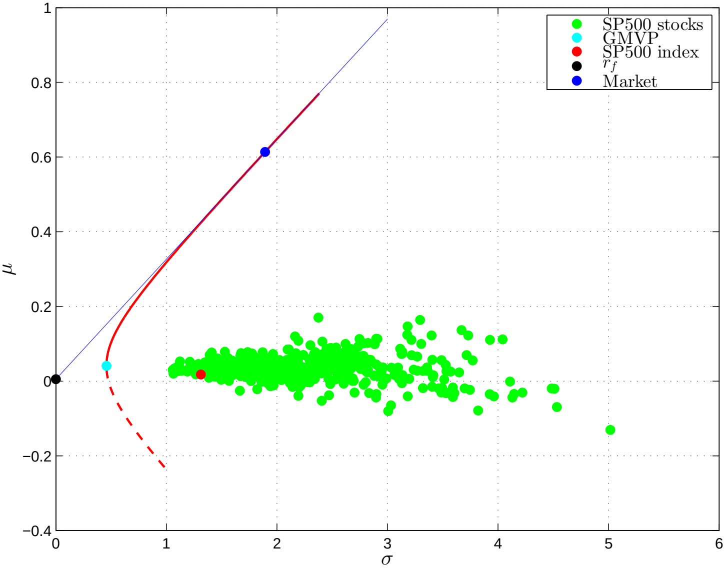
$$r_f := \frac{1}{T}\sum_{t=1}^{T}\frac{Libor_t}{365\text{ days}} = 0.00524\%$$

- 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 $r_f = 0$
  - 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

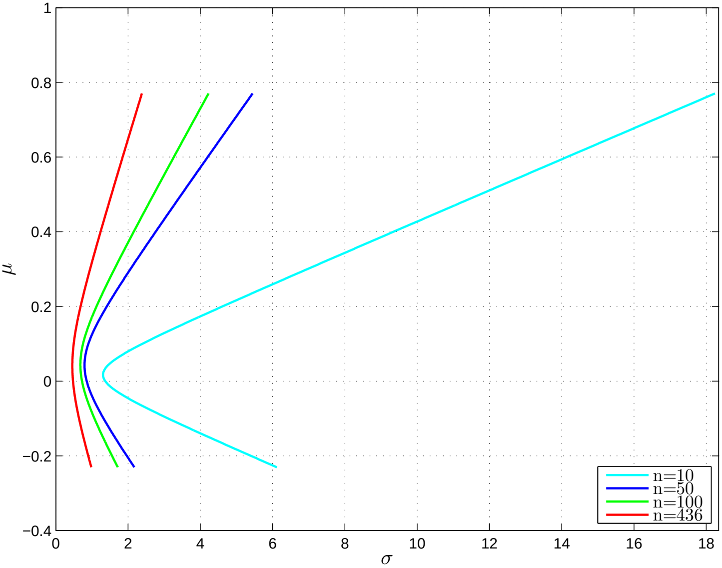
The Overnight Libor



Mean-Variance Analysis (with Short-Selling)



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



No Short-Selling

- in the theory part to the mean-variance analysis, we didn't cover the general n-asset case

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

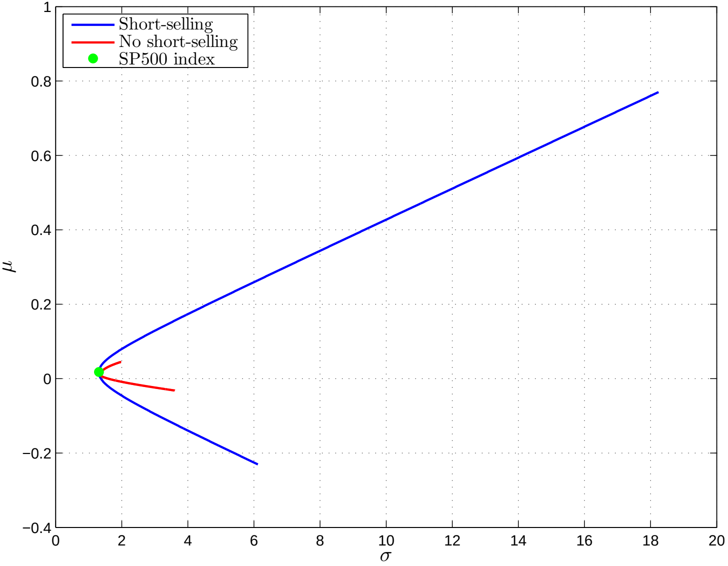
  when short-selling is not allowed because it is tough to tackle analytically

- 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

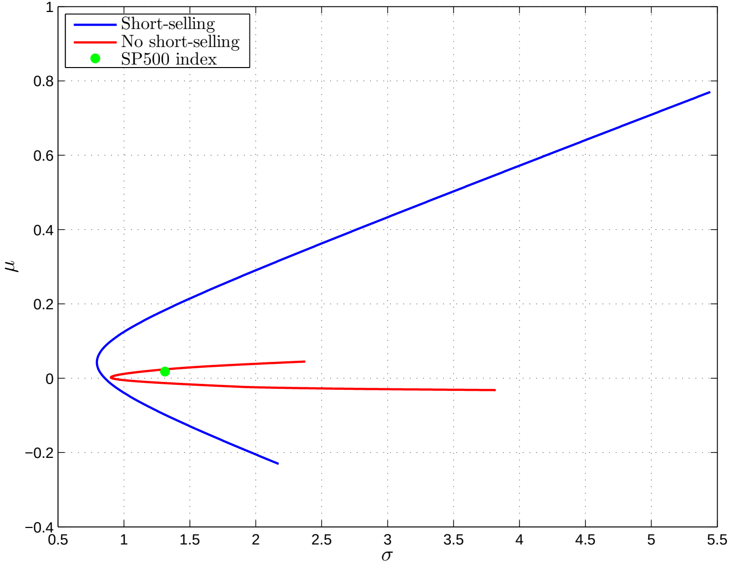
No Short-Selling

- 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

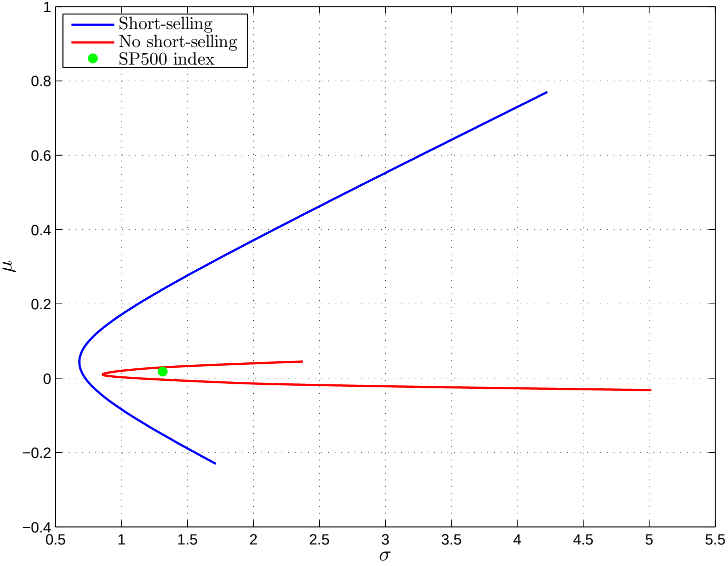
MVS Comparison for Most Liquid 10 Stocks



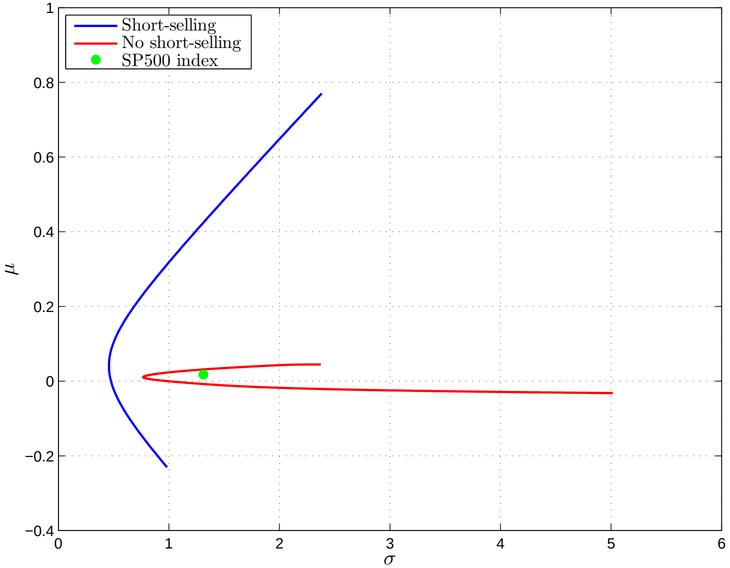
MVS Comparison for Most Liquid 50 Stocks



MVS Comparison for Most Liquid 100 Stocks



MVS Comparison for Most Liquid 436 Stocks



Statistical Inference via Resampling Techniques

Monte Carlo (MC) Simulations

- consider the classical linear regression model

$$\underbrace{\boldsymbol{y}}_{T\times 1} = \underbrace{\boldsymbol{X}}_{T\times k}\underbrace{\boldsymbol{\beta}}_{k\times 1} + \underbrace{\boldsymbol{u}}_{T\times 1} \tag{6}$$

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

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

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

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\mathsf{T}\boldsymbol{X})^{-1}\boldsymbol{X}^\mathsf{T}\boldsymbol{y} \tag{7}$$

exists and follows the probability law

$$\hat{\boldsymbol{\beta}} \overset{d}{=} \mathsf{N}\Big(\,\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^\mathsf{T}\boldsymbol{X})^{-1}\,\Big)$$

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

Monte Carlo (MC) Simulations

- 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 only have a small sample size T, and even when T is large, we usually don't know if convergence has proceeded far enough to provide a good approximation

Monte Carlo (MC) Simulations

- put differently, we have to distinguish between the exact, finite-sample distribution of $\hat{\boldsymbol{\beta}}$ (holds for any fixed $T < \infty$) and the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ (holds for $T \to \infty$ only)

⇒ in cases where we suspect the asymptotic approximation of the (unknown) finite-sample distribution to be inaccurate, we need alternative ways of approximation:

1. proposal is purely theoretical and refers to so-called Edgeworth and saddlepoint approximations which are refinements to the second-order approximation of asymptotic statistics allowing for more flexibility to accommodate features of the finite-sample distribution; see Hall (1992), Rothenberg (1984), and Ullah (2004)
2. proposal is to use MC simulations to approximate the finite-sample distribution

- while the first approach is more general (explicitly laying out all required assumptions), it is much more complicated and harder to accomplish than MC simulations

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

⇒ MC simulations are case-by-case studies

How and Why MC Works: An Example

- suppose that we have no clue about the finite-sample and asymptotic properties of $\hat{\boldsymbol{\beta}}$
- 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 $\hat{\boldsymbol{\beta}}$ to some real-world data
- however, this is not helping because we neither know if the underlying model (6) is appropriate nor, if it were true, what the true parameter $\boldsymbol{\beta}$ looks like
- a better way of analyzing the properties of $\hat{\boldsymbol{\beta}}$ is to simulate an artificial data set from model (6) which is completely known
- but before we can start any simulation the free parameters in model (6) 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

- 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 (6) is fixed, i.e., deterministic $\Rightarrow$ this allows us to simplify the notation by using $\mathsf{E}[\,.\,]$ instead of $\mathsf{E}[\,.|\boldsymbol{X}\,]$
- in 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 $\text{Unif}(0, 1)$
  - the third column of $\boldsymbol{X}$ corresponds to a $(T \times 1)$ vector of iid realizations drawn from the standard normal distribution $\mathsf{N}(\,0, 1\,)$

How and Why MC Works: An Example

- thus, the randomness of $\hat{\boldsymbol{\beta}}$ can be traced back to $\boldsymbol{u}$ by substituting (6) in (7):

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

- 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 generate 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 $\mathsf{N}\big(\,\mathbf{0}_{T\times 1}, \sigma^2\boldsymbol{I}_T\,\big)$

⇒ 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

- suppose we have drawn m samples $\boldsymbol{u}^{(1)}, \ldots, \boldsymbol{u}^{(m)}$ from $\mathsf{N}\left(\mathbf{0}_{T\times 1}, \sigma^2 \boldsymbol{I}_T\right)$ and subsequently computed $\boldsymbol{y}^{(j)}$ and $\hat{\boldsymbol{\beta}}^{(j)}$ via (6) and (7), respectively, with $j = 1, \ldots, m$ such that

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

can be averaged over all m equations:

$$\begin{aligned}\frac{1}{m}\sum_{j=1}^{m}\hat{\boldsymbol{\beta}}^{(j)} &= \frac{1}{m}\sum_{j=1}^{m}\boldsymbol{\beta} + \frac{1}{m}\sum_{j=1}^{m}(\boldsymbol{X}^\mathsf{T}\boldsymbol{X})^{-1}\boldsymbol{X}^\mathsf{T}\boldsymbol{u}^{(j)} \\ &= \boldsymbol{\beta} + (\boldsymbol{X}^\mathsf{T}\boldsymbol{X})^{-1}\boldsymbol{X}^\mathsf{T}\left(\frac{1}{m}\sum_{j=1}^{m}\boldsymbol{u}^{(j)}\right) \qquad (8)\end{aligned}$$

How and Why MC Works: An Example

- notice that (8) is the sample counterpart of

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

which corresponds to

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

because $\mathsf{E}[\,\boldsymbol{u}\,] = \mathbf{0}_{T\times 1}$, by construction

- the only requirement for this result to hold is that

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

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

⇒ this is the basic idea of MC simulations!

- more generally, 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

- 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

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

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

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

$$= \widehat{\mathrm{Bias}}\left[\hat{\theta}\right]^2 + \widehat{\mathrm{StD}}\left[\hat{\theta}\right]^2$$

How and Why MC Works: An Example

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

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

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

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

- 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

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

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

How and Why MC Works: An Example

- 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

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

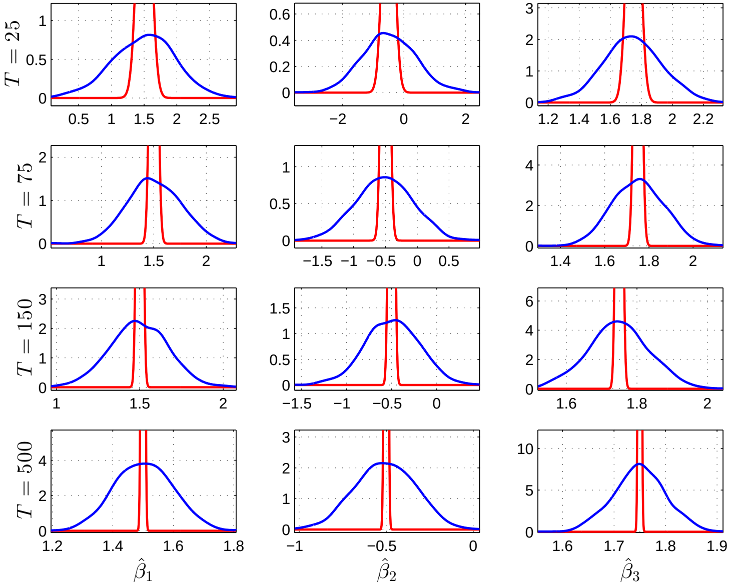
or

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

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., $\forall\, t = 1, \dots, T$

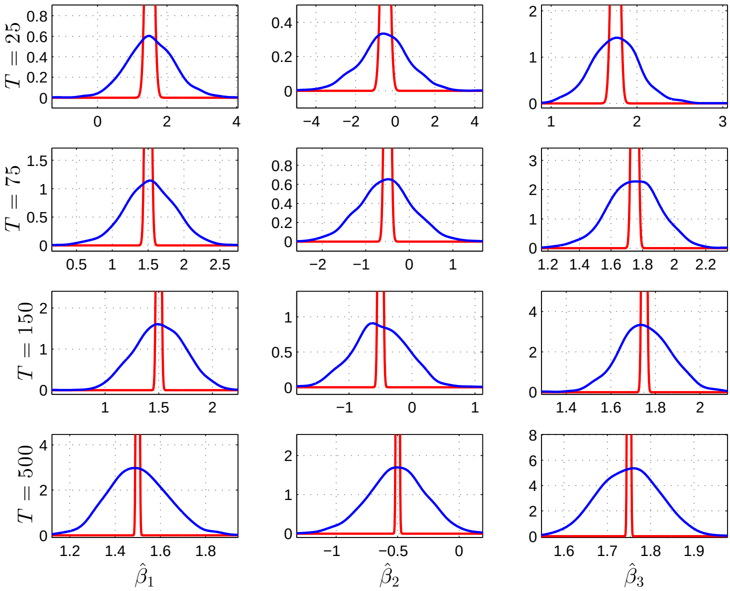
$$u_t \overset{\text{iid}}{=} t_{\nu,\delta}$$

- 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 $\mathsf{E}[\, t_{\nu,\delta}\,] = 0$ 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: $u \overset{\text{iid}}{=} t_{\nu,\delta}$ 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

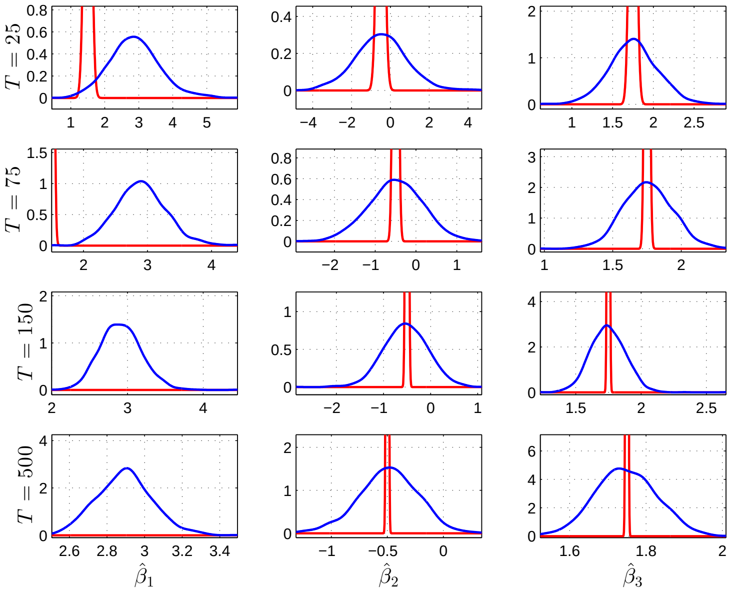
Density Comparison (CLT=red, MC=blue)



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)



Further Estimation Methods

Bibliography

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. *The Journal of Business*, **39** 119–138.

Bibliography

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.