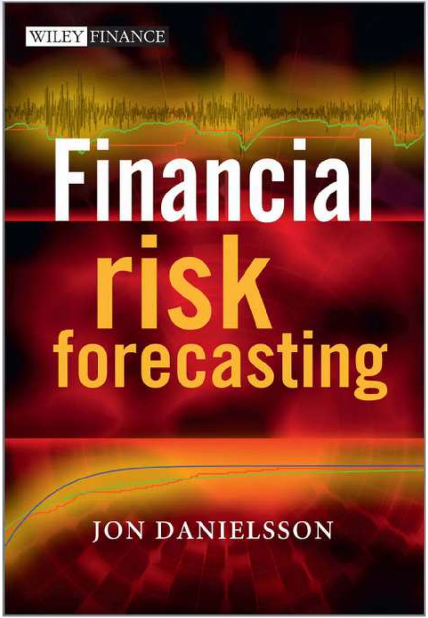
Financial Risk Forecasting
Chapter 5
Implementing Risk Forecasts

Jon Danielsson ©2025
London School of Economics

To accompany
Financial Risk Forecasting
FinancialRiskForecasting.com

Published by Wiley 2011

Version 10.0, August 2025


WILEY FINANCE
Financial risk forecasting
JON DANIELSSON

Introduction

The Focus of This Chapter

- Techniques for implementing risk forecasting, especially
- Historical simulation
- Risk measures and parametric methods
- Expected returns
- VaR with time-dependent volatility

From Theory to Practice — Risk Measurement

- In previous chapters, we explored:
  - Properties of financial data
  - Volatility modelling using methods such as GARCH
  - Theoretical definitions of Value-at-Risk (VaR) and Expected Shortfall (ES)
- This chapter focuses on applying those concepts to real data
- We move from theoretic models to data-driven methods of estimating financial risk

Some Concepts

- So far we have defined a sample to have a size T
- When it comes to what follows, it is useful to have a separate indication for number of observations used in the estimation, called *estimation window*, or W_E
- The methods below fall into two main categories

nonparametric No distribution of data assumed, no estimation, and hence no distributional parameters (like historical simulation)

parametric Assume a distribution and estimate its parameters

Notation new to this Chapter

W_E	Estimation window

In Sample, Out-of-Sample and Forecasting

- If we use information from $t = 1$ to $t = T$ to produce analysis about time T, we are doing *in-sample* analysis
- But here we want to *forecast* risk
- That means using information from $t = 1$ to $t = T$ to produce a forecast of what might happen later, perhaps at $t = T + 1$

Reminder of learning outcomes

1. Understand the difference between parametric and nonparametric methods
2. The properties, strength and weaknesses of historical simulation
3. Derive VaR for simple and compound returns
4. Implementation of risk forecasts for time-dependent models

Historical Simulation (HS)

Historical Simulation — A Nonparametric Approach to Risk

- Historical simulation is a nonparametric method for estimating VaR and ES
- Nonparametric means we do not assume any functional form for the return distribution
- We use the empirical distribution — the observed returns themselves — as the basis for risk estimates
- This avoids distributional assumptions but may be sensitive to rapidly changing and outdated data

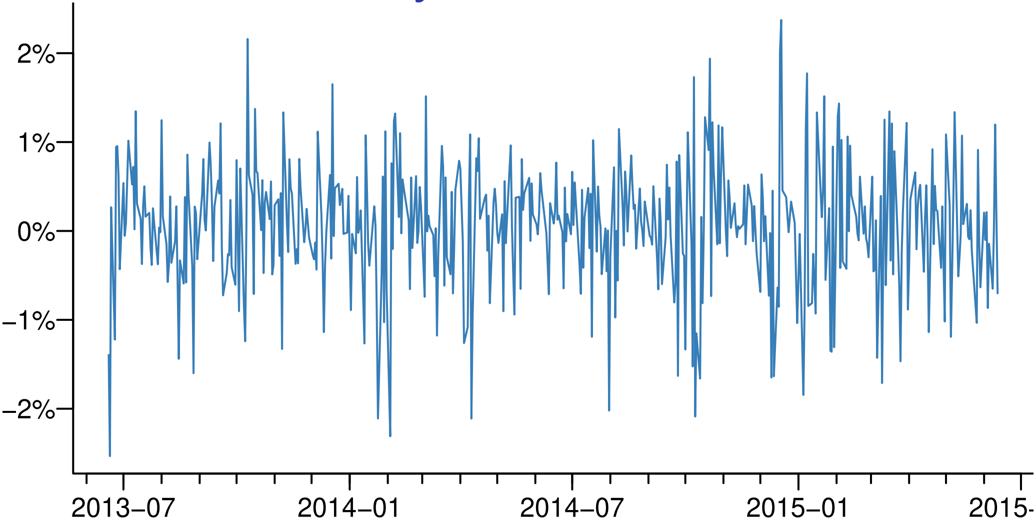
Historical Simulation: One Asset

- Assumes that one of the observations in the estimation window will be the next days return, therefore
- Assume history repeats itself
- VaR is one of the observations in the estimation window, multiplied by the *monetary* value of the asset holdings, the portfolio value ϑ_t

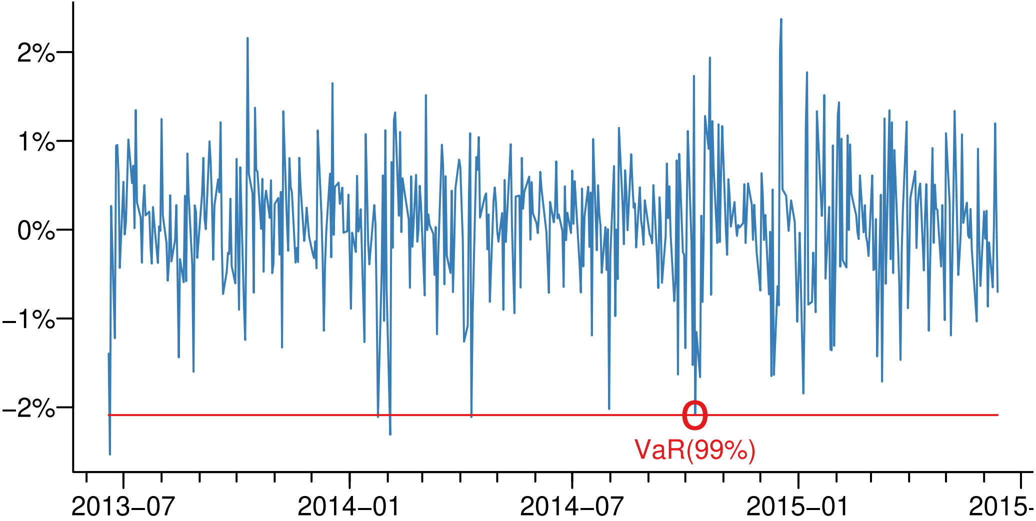
$$\vartheta_t = \text{ number of stocks owned} \times p_t$$

- VaR_t is the negative of the $(W_E \times \rho)^{\text{th}}$ smallest return, times ϑ_t
- Best seen by example

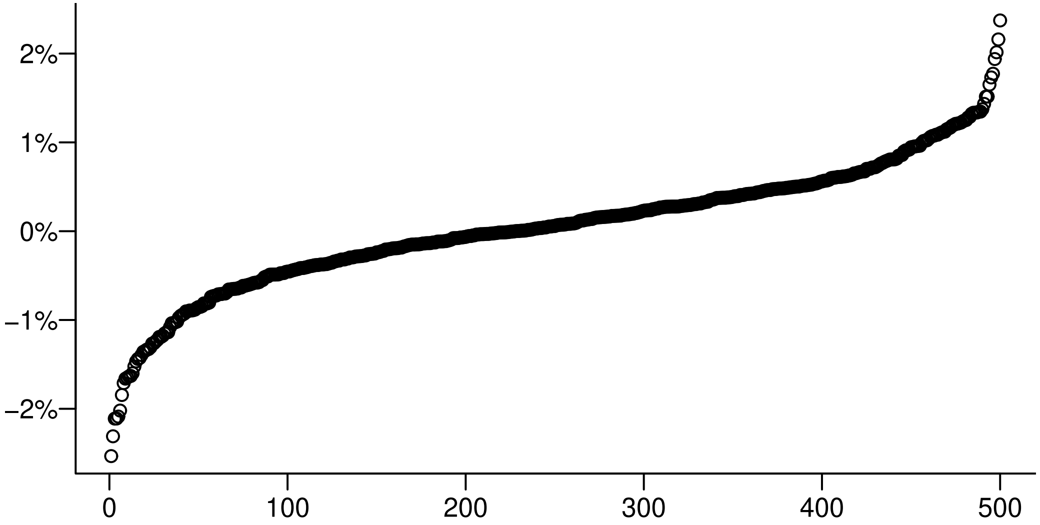
500 Days of the S&P-500



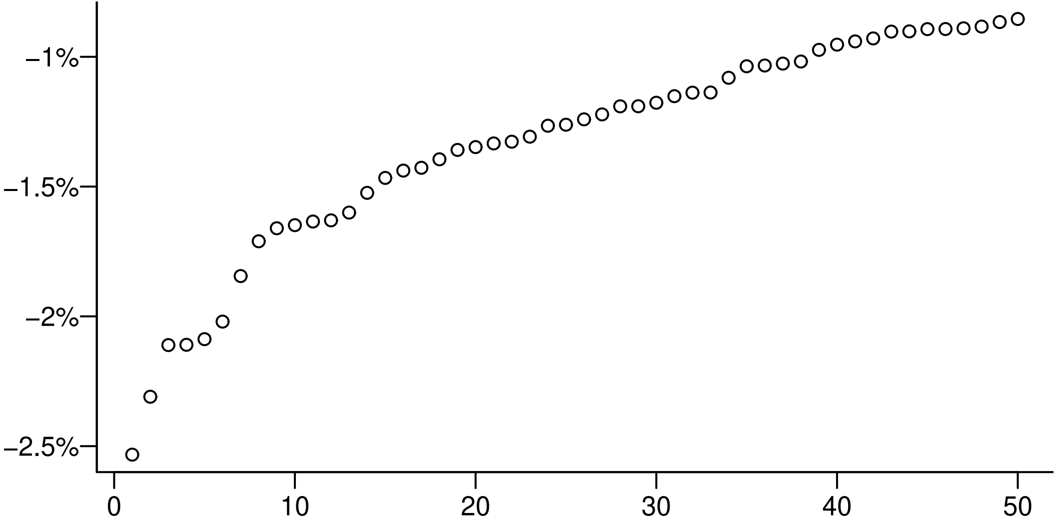
500 Days of the S&P-500



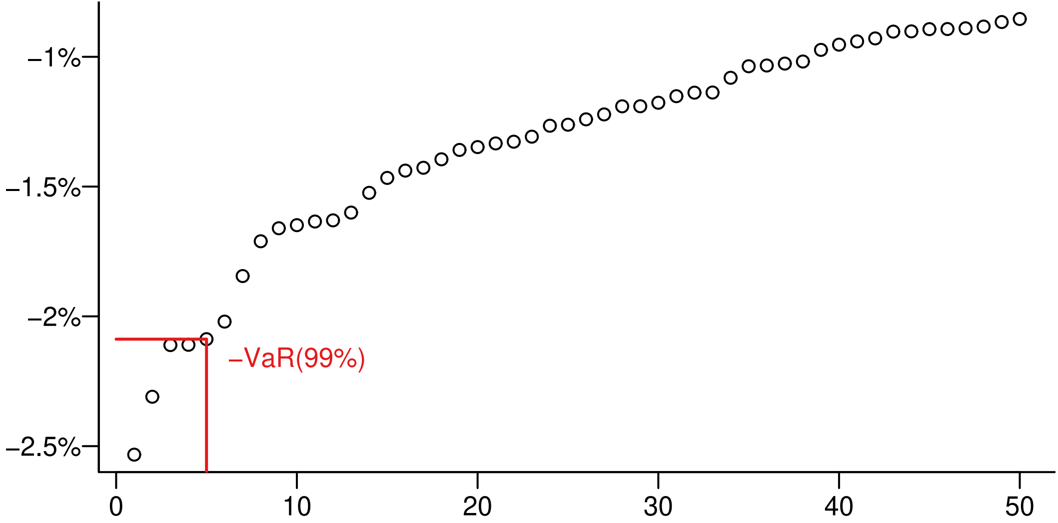
Sorted Returns



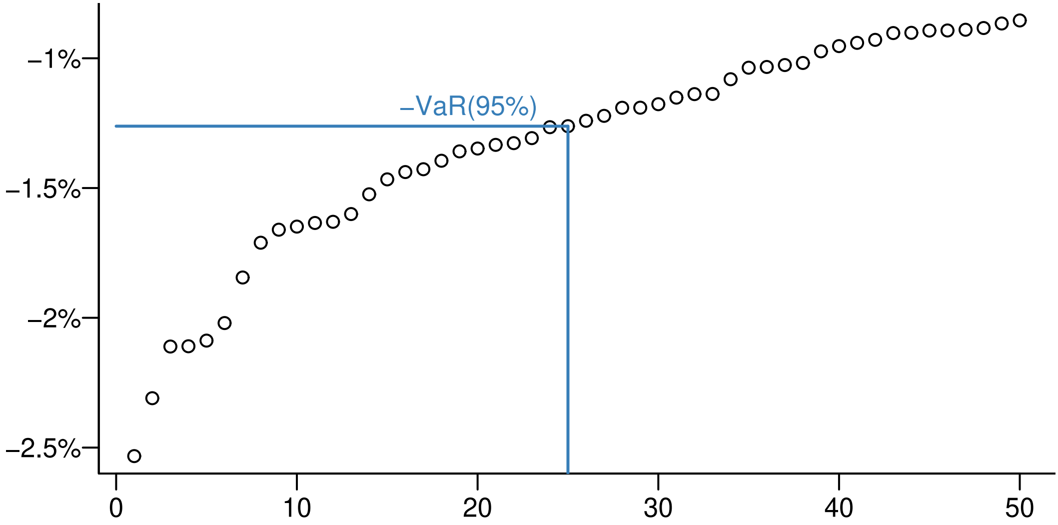
Zoom In



Zoom In



Zoom In



Procedure

- Decide on a probability, ρ, for example, 1%
- Have a sample of returns, $y = \{y_1, \ldots, y_{W_E}\}$, with length W_E, for example, 1000
- Sort the y from the smallest to the largest, call that ys
- At the daily frequency simple and compound returns are almost the same, so here use compound
- Take the $(W_E \times \rho)^{\text{th}} = (1000 \times 0.01)^{\text{th}}$ smallest value of ys, call that $ys_{(W_E \times \rho)} = ys_{10}$
- If you own one stock, and $p_{t-1} = 1$, then VaR is the 10^{th} smallest return, that is,

$$\text{VaR}_t = -ys_{10}$$

- Otherwise have to multiply that by the number of stocks you own and their $t-1$ price

$$\text{VaR}_t = -ys_{10} \times p_{t-1} \times \text{number of stocks}$$

Multiple Assets

- Take a matrix of historical portfolio returns
- y is a $W_E \times K$ matrix of returns
- w is a $K \times 1$ vector of portfolio weights
- We then get the time series vector of portfolio returns by

$$y_{\text{portfolio}} = yw$$

- And then you can simply treat the portfolio as if it were a single asset and apply HS

Expected Shortfall Estimation

- The expected losses conditional on VaR being violated
- Estimated by HS by taking the mean of all observations less than or equal to $-\text{VaR}$
- Continuing as with the single asset asset case from above

$$\text{ES} = -\frac{1}{10}\sum_{i=1}^{10} ys_i$$

Importance of Sample Size

- The most extreme observations fluctuate a lot more than observations that are less extreme
- You will see that in the Monte Carlo simulations a few slides down
- Therefore, the bigger the sample the more precise the estimation of HS should be
- The downside is that old data may not be all that representative
- And if there is a structural break in the data (like in 2008) the VaR forecasts take longer to adjust to structural changes in risk
- As a general rule
- Minimum recommended sample size

$$\frac{3}{\rho}$$

Issues

- No model assumptions needed
- In the absence of structural breaks HS tends to perform well
- It captures non-linear dependence directly
- But performs badly when data has structural breaks
- This can be seen in the discussion about backtesting in Chapter 8

Estimation

- Use stock data
- Use the last 1,000 days for the estimation ($W_E = 1,000$)
- $p = 0.01$
- $w = c(0.4, 0.6)'$
- Portfolio value is $1,000

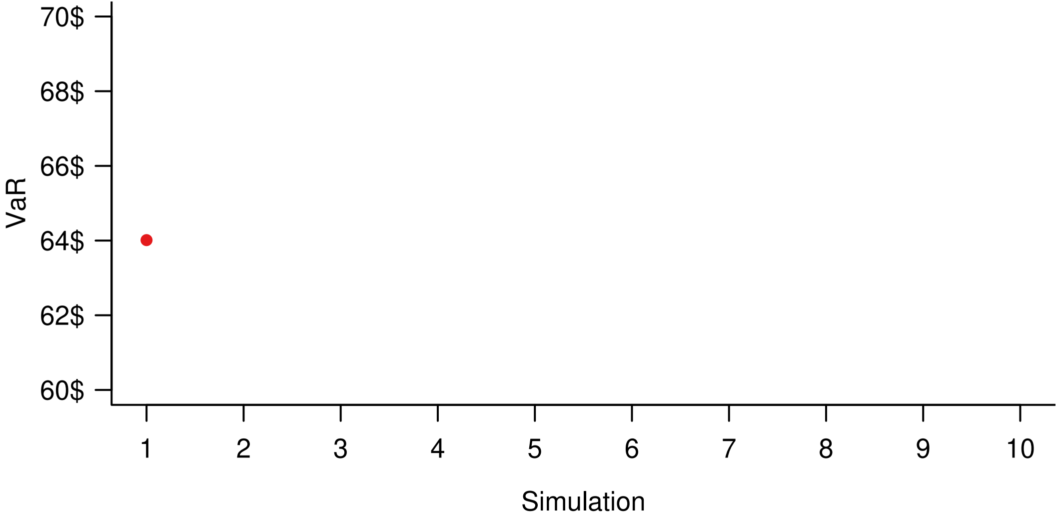
R Estimation

```
y1 # vector of asset 1
y2 # vector of asset 2
portfolio = 1000
w = c(0.4,0.6)
rho = 0.01;
WE = 1000
y1 = tail(y1,WE)
y2 = tail(y2,WE)
VaR = -sort(y1)[rho*WE] * portfolio
42.19389
y = cbind(y1,y2) %*% w
VaRPortfolio = -sort(y)[p*WE] * portfolio
38.88712
```

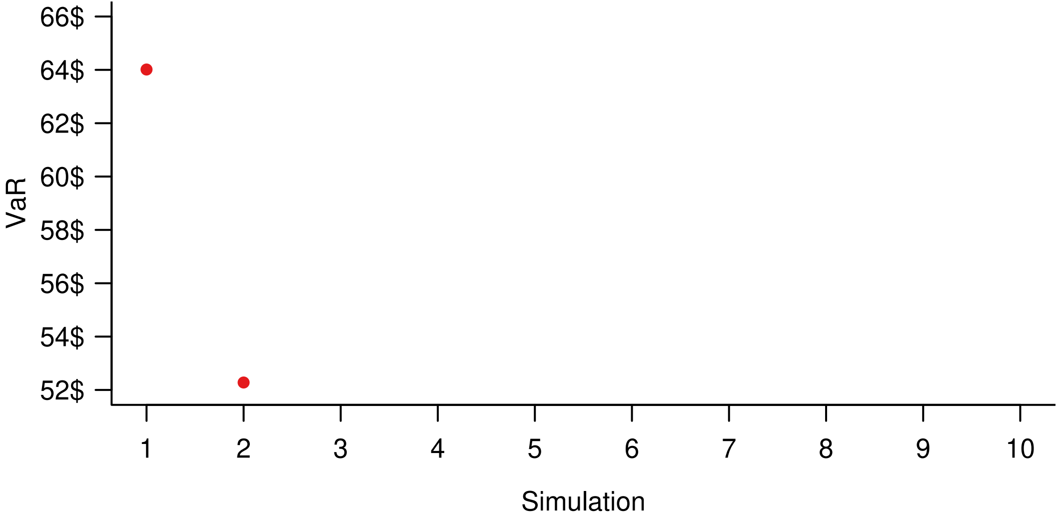
Simulation Examination

- We can compare the impact of the sample size by using Monte Carlo simulations
- In the following, the true distribution is Student-t (3), so you can easily work out the true VaR
- You can use the method below to create a confidence bound for a VaR estimated by historical simulation
- This setup lets us assess the variability of the estimated VaR and construct confidence bounds

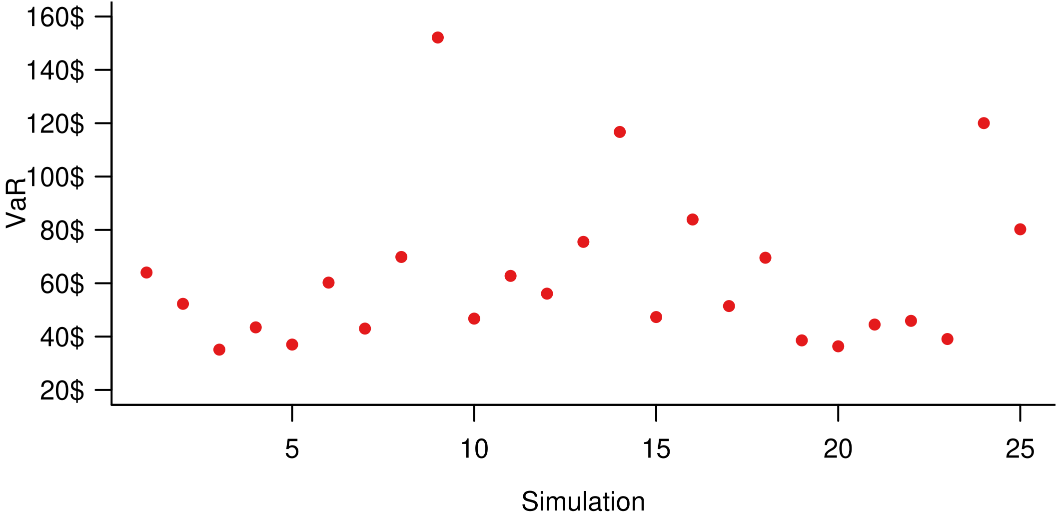
Simulated HS: $T = 100$.



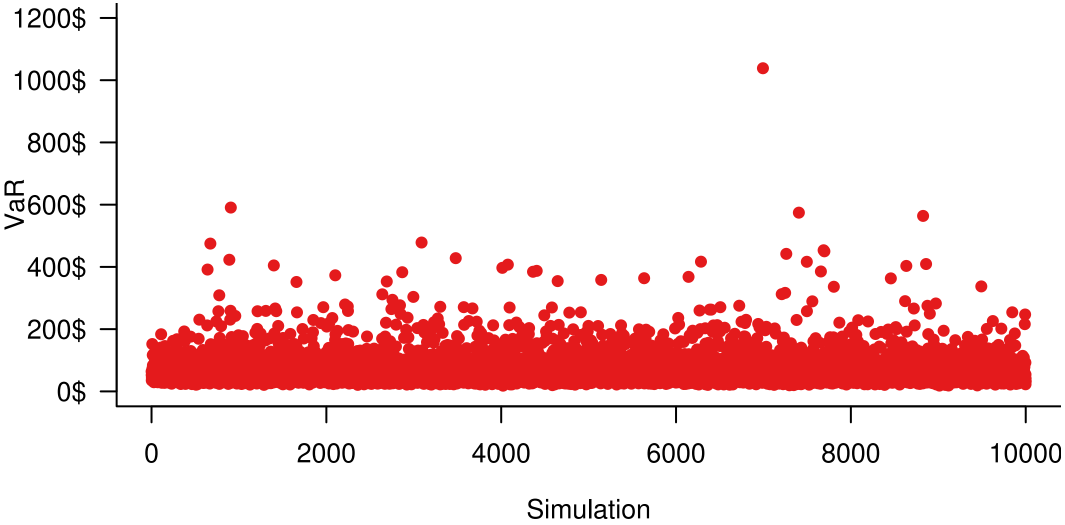
Simulated HS: $T = 100$.



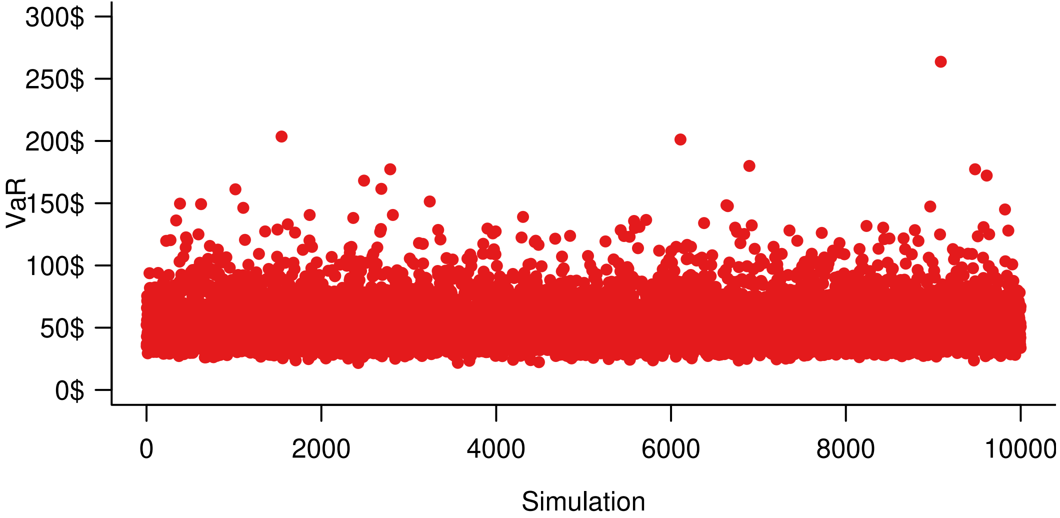
Simulated HS: $T = 100$.



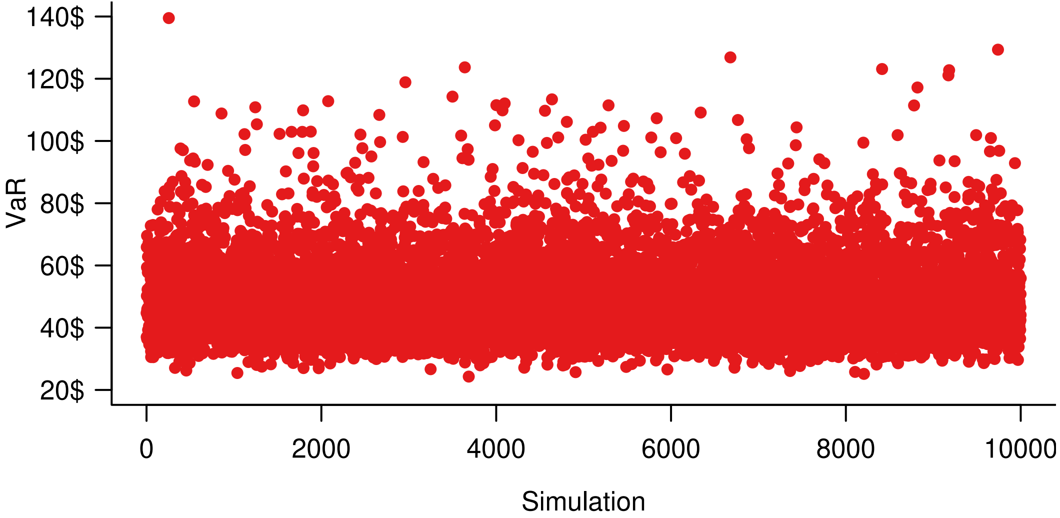
Simulated HS: $T = 100$.



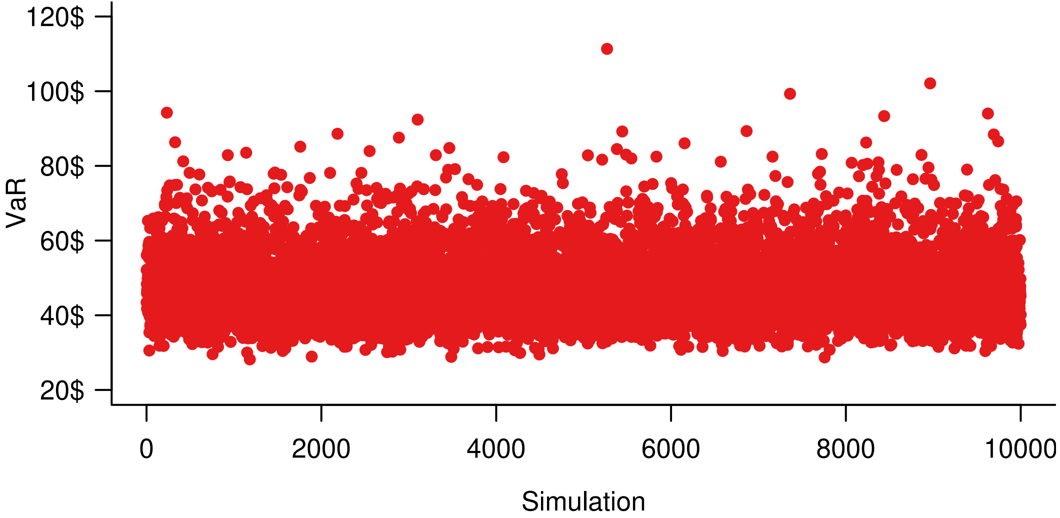
$T = 200$ observations



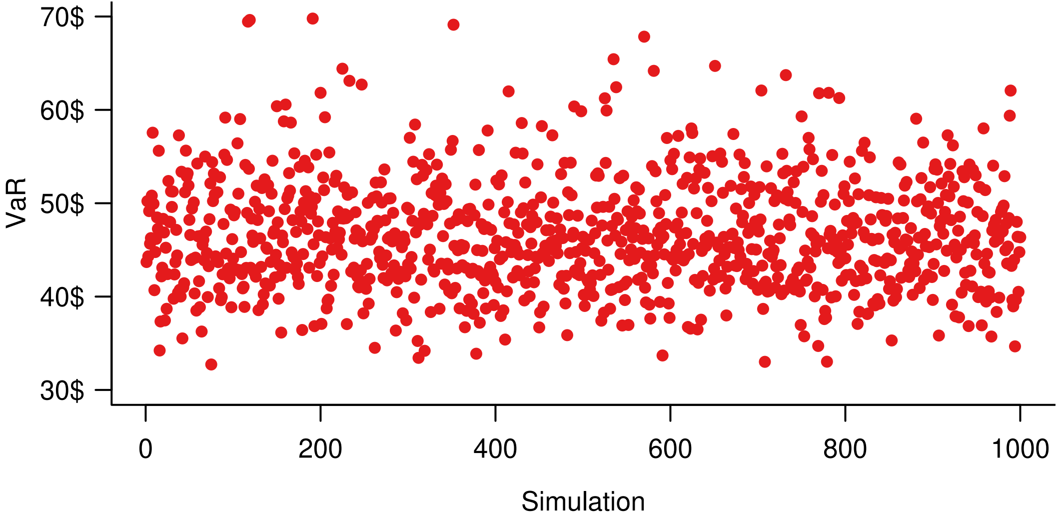
$T=300$ observations



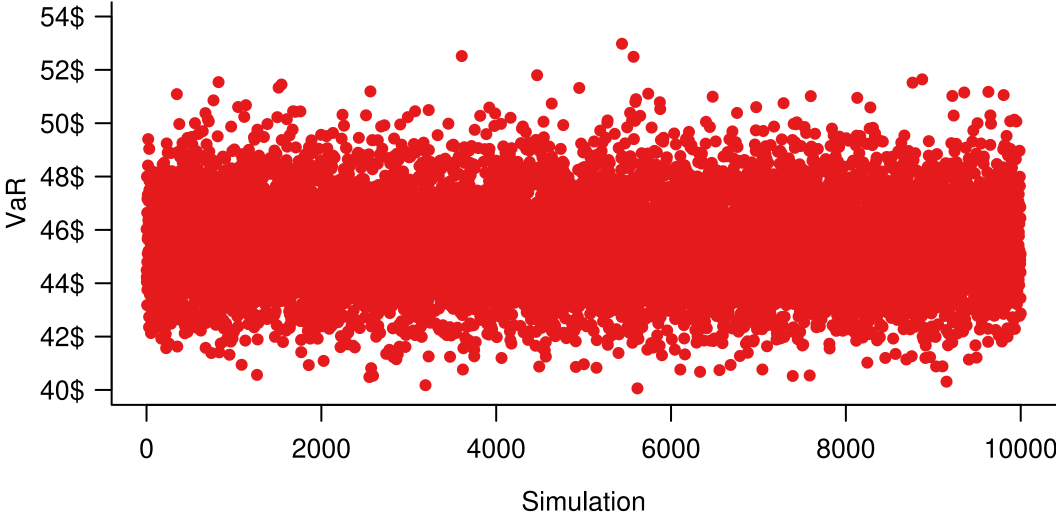
T=500 observations



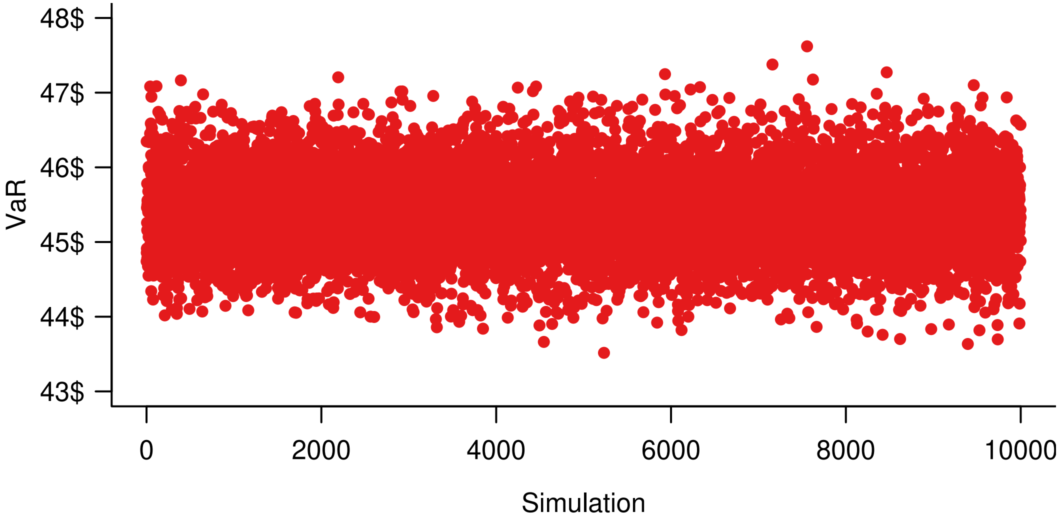
T=1,000 observations



T=10,000 observations



T=100,000 observations



Historical Simulation — Key Takeaways

- Historical simulation is a simple and intuitive nonparametric method for estimating VaR and ES
- It relies on the empirical distribution of past returns — no model assumptions are needed
- Accuracy depends heavily on the sample size and the relevance of past data
- Performs well when market conditions are stable but struggles with structural breaks
- Monte Carlo studies help assess estimation uncertainty and guide practical implementation

Parametric Methods

Risk Measures — From Nonparametric to Parametric

- Historical simulation estimates risk directly from past data without assuming a model
- We now turn to parametric methods — where we assume a functional form for the return distribution
- This allows for analytical expressions for risk measures like VaR and ES
- Parametric methods can be more efficient and flexible, but also introduce particular types model risk (discussed later)

Setup — Parametric Risk Measurement

- We now abstract from time and focus on calculating VaR and ES on day t
- This is conditional on assuming a return distribution on day t
- Parametric methods require us to specify a density $f(\cdot)$ and a corresponding distribution function $F()$
- Common choices include the normal and the Student-t distribution

Setup

- The normal distribution is often denoted by $\phi()$ for the density and $\Phi()$ for the cumulative distribution
- Each distribution has associated parameters θ

  Normal $\theta = (\sigma, \mu)'$

  Student-t $\theta = (\sigma, \mu, \nu)'$

- We assume in this Section that the mean return is zero

Analytical Derivation of Risk Measures

- We now derive explicit formulas for VaR under common parametric distributions
- These derivations assume known distribution functions — typically normal or Student-t
- The goal is to express risk measures in terms of quantiles and distribution parameters
- This helps link statistical assumptions directly to financial risk forecasts

Recall VaR

- The definition of VaR is

$$\rho = \mathbb{P}[q_t \leq -\text{VaR}_t(\rho)]$$
$$= \int_{-\infty}^{-\text{VaR}_t(\rho)} f(x)dx$$

- Profit and loss when we own one stock

$$q_t = p_t - p_{t-1}$$

VaR for Simple Returns and Holding One Stock

The definition of simple returns is

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}}$$

The definition of VaR (we own one stock)

$$\begin{aligned}
\rho =& \mathbb{P}[q_t \leq -\text{VaR}_t(\rho)] \\
=& \mathbb{P}(p_t - p_{t-1} \leq -\text{VaR}_t(\rho)) \\
=& \mathbb{P}(p_{t-1} r_t \leq -\text{VaR}_t(\rho)) \\
=& \mathbb{P}\left(\frac{r_t}{\sigma} \leq -\frac{\text{VaR}_t(\rho)}{p_{t-1}\sigma}\right)
\end{aligned}$$

VaR for Simple Returns and Holding One Stock (Cont.)

- So

$$\begin{aligned}\rho =& \mathbb{P}[q_t \leq -\mathrm{VaR}_t(\rho)] \\ =& \mathbb{P}\left(\frac{r_t}{\sigma} \leq -\frac{\mathrm{VaR}_t(\rho)}{p_{t-1}\sigma}\right)\end{aligned}$$

- We now want to solve for VaR
- Recall from Chapter 1 that a probability of an outcome can be expressed as a value from the CDF

VaR for Simple Returns and Holding One Stock (Cont.)

- Denote the distribution of standardised returns (where the area under the PDF is 1), r_t/σ, by

$$F_r(\cdot)$$

- We need this because a CDF $F()$ for an asset will likely be in respect to a volatility that is low compared to a standardised (like 0.01 for a normal where 1 is for standardised normal)
- The inverse distribution is $F_r^{-1}(\rho)$

SO

- We have from above

$$\begin{aligned}\rho =& \mathbb{P}[q_t \leq -\text{VaR}_t(\rho)] \\ =& \mathbb{P}\left(\frac{r_t}{\sigma} \leq -\frac{\text{VaR}_t(\rho)}{p_{t-1}\sigma}\right)\end{aligned}$$

- And the standardised distribution for r_t/σ is

$$F_r(\cdot)$$

Solving

- Solving we get

$$\text{VaR}_t(\rho) = -\sigma \times F_r^{-1}(\rho) \times p_{t-1}$$

- Because
- $F_r^{-1}(\rho)$ is the standardised quantile associated with probability ρ
- Need to rescale it to the distribution we have, hence the $-\sigma\times$
- And finally take into account our position: $\times p_{t-1}$

VaR for Continuously Compounded Returns

- VaR has so far been derived from simple returns

$$R_t = \frac{p_t - p_{t-1}}{p_{t-1}}$$

- We now switch to log-returns

$$y_t = \log p_t - \log p_{t-1}$$

- The link is $R_t = e^{y_t} - 1$ — for small y_t the two measures are almost identical
- Log-returns aggregate additively over time and give cleaner parametric formulas
- Our next step is to obtain the VaR quantile from the distribution of y_t and, if needed, translate it back to the simple-return scale

VaR for Continuously Compounded Returns

- Definition

$$y_t = \log p_t - \log p_{t-1}$$

$$\begin{aligned}\rho &= \mathbb{P}\left(p_t - p_{t-1} \leq -\text{VaR}_t(\rho)\right) \\ &= \mathbb{P}\left(p_{t-1}\left(e^{y_t} - 1\right) \leq -\text{VaR}_t(\rho)\right) \\ &= \mathbb{P}\left(\frac{y_t}{\sigma} \leq \log\left(-\frac{\text{VaR}_t(\rho)}{p_{t-1}} + 1\right)\frac{1}{\sigma}\right)\end{aligned}$$

when

$$\frac{-\text{VaR}_t(\rho)}{p_{t-1}} \leq 1$$

VaR for Continuously Compounded Returns (Cont.)

- Denote the distribution of standardised log returns (y_t/σ) by

$$F_y(\cdot)$$

- The inverse distribution is $F_y^{-1}(\rho)$
- So in the same as for the simple returns we get

$$\text{VaR}_t(\rho) = -\left(\exp(F_y^{-1}(\rho)\sigma) - 1\right)p_{t-1}$$

- Not very convenient but (next slide)

By Approximation

- How different is the log return version

$$e^{F_y^{-1}(\rho)\sigma} - 1$$

- From the simple return version

$$\sigma F_r^{-1}(\rho)?$$

- It depends on the sampling frequency and volatility
- Recall the discussion in the first section of Chapter 1
- Where simple returns are approximately same as compound returns at the daily frequency, therefore

$$F_r \approx F_y$$

- For daily returns the volatility is low (perhaps 0.01)

So

- The two are not that different in practice

$$e^{F_y^{-1}(\rho)\sigma} - 1 \approx \sigma F_r^{-1}(\rho)?$$

- With typical vol, 0.01, and 1% probability

$$\exp(-2.3 * 0.01) - 1 = -0.02274 \approx -0.023 = 0.01 \times -2.3$$

We Therefore Get

- Therefore for small $F_y^{-1}(\rho)\sigma$ the VaR for holding one unit of asset is:

$$\text{VaR}_t(\rho) \approx -\sigma F_y^{-1}(\rho) p_{t-1}$$

- Meaning the VaR for continuously compounded returns is approximately the same as the VaR using simple returns
- And $F_y^{-1}(\rho)\sigma$ is much easier to work with
- Besides, σ is only accurately estimated to 1 or at best 2 digits
- So little or nothing is lost by the approximation

VaR for Portfolios — Beyond a Single Asset

- So far we have derived VaR for a single asset using its own return variance
- Real portfolios contain several assets whose risks interact through correlation
- We capture these interactions with the weight vector w and the covariance matrix Σ
- Portfolio risk is summarised by $\sigma^2_{\text{portfolio}}$ which feeds directly into the parametric VaR formula

When There Is More Than One Asset

- In the two assets case

$$\sigma^2_{\text{portfolio}} = \begin{pmatrix} w_1 & w_2 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

- And generally in the K asset case

$$\sigma^2_{\text{portfolio}} = w'\Sigma w$$

- Then, as before,

$$\text{VaR}_t(\rho) = -\sigma_{\text{portfolio}} F^{-1}(\rho)\vartheta$$

VaR When Returns Are Normally Distributed

- $\vartheta = 1, \sigma = 1, p = 0.05$

$$\text{VaR} = -\Phi^{-1}(0.05) = 1.64$$

- If $\sigma \neq 1$, then the VaR is:

$$\text{VaR} = \sigma 1.64$$

- If portfolio value is not equal to 1:

$$\text{VaR} = \sigma 1.64 \vartheta$$

VaR Under the Student-t Distribution

- Advantage of Student-t VaR over normal is fat tails
- ν indicates how fat tails are
- When $\nu = \infty$ the Student-t becomes normal

VaR When Returns Are Student-t

$$\vartheta = 1000, \sigma = 0.01, p = 0.05$$

In R: `qt(0.05,5)`

ν	F^{-1}	VaR
∞	-1.645	$\$15.45 = \lvert -1.645\rvert \times 1000 \times 0.01$
5	-2.015	\$20.15
3	-2.353	\$23.53

- There is one important complication that arises because by convention tGARCH uses the standardised distribution. See discussion in Chapter 2

Expected Shortfall Under Normality

- VaR under normality $\text{VaR}_t(\rho) = -\vartheta\sigma\Phi^{-1}(\rho)$
- ES is the expected loss beyond VaR
- Result derivation in Appendix

$$
\begin{aligned}
\text{ES} &= \int_{-\infty}^{-\text{VaR}_t(\rho)} x f_{\text{VaR}}(x) dx \\
&= \vartheta\sigma\frac{\phi(\Phi^{-1}(\rho))}{\rho}
\end{aligned}
$$

- The ES / VaR ratio is fixed by ρ; e.g. $\rho = 1\% \Rightarrow \text{ES} \approx 2.66$ VaR

R Probabilities

```
-qt(0.05,1000) * 1000 * 0.01
16.46379
-qt(0.05,5) * 1000 * 0.01
20.15048
-qt(0.05,3) * 1000 * 0.01
23.53363
for(df in c(1000,5,3))
cat(round(-qt(0.05,df) * 1000 * 0.01,1),"\n")
16.5
20.2
23.5
```

VaR and ES in R

```
sigma=0.01
portfolio=1000
p=0.01
-sigma * portfolio * qnorm(p)
23.26348
sigma * portfolio * dnorm(qnorm(p))/p
26.65214
```

Non-Zero Returns

Expected Returns

- Returns have (at least) a mean and variance
- Is it reasonable to assume $\mu_t = 0$?
- Given that statistical uncertainty is more than 10% in most VaR calculations, VaR calculation is only significant to one digit
- Mean is smaller than that
- For S&P-500: $\mu = 0.019\%, \sigma = 1.15\%$

VaR With Mean

- The definition of VaR is

$$p = \mathbb{P}[q \leq -\text{VaR}(\rho)]$$
$$= \int_{-\infty}^{-\text{VaR}(\rho)} f(x)dx$$

- Profit and loss when we own one stock

$$\text{E}(q_t) = \text{E}(p_t) - \text{E}(p_{t-1})$$

VaR With Mean (Cont.)

- If mean is not zero, $\mathrm{E}(q) \neq 0$, then the definition of VaR becomes

$$\mathbb{P}[q \leq -\mathrm{VaR}(\rho)] = \mathbb{P}[q - \mu \leq -\mathrm{VaR}(\rho) - \mu] = p$$

  but now the expected return, μ needs to enter the VaR equation

$$\mathrm{VaR}(\rho) = -\sigma F^{-1}(\rho) - \mu$$

- Supposing F is the distribution of the de-meaned returns
- Note how the mean has a minus in front of it
- $-\sigma F^{-1}(\rho) > 0$ and $-\mu < 0$
- A positive mean pulls the distribution to the right, making the VaR smaller

Time Aggregation of VaR With Mean

- If the returns are IID, then both mean and variance aggregate at the same rate
- Mean and variance over T days is equal to T times mean and variance over one day
- Which implies that the volatility aggregates at the square root of time
- The T-period VaR is therefore:

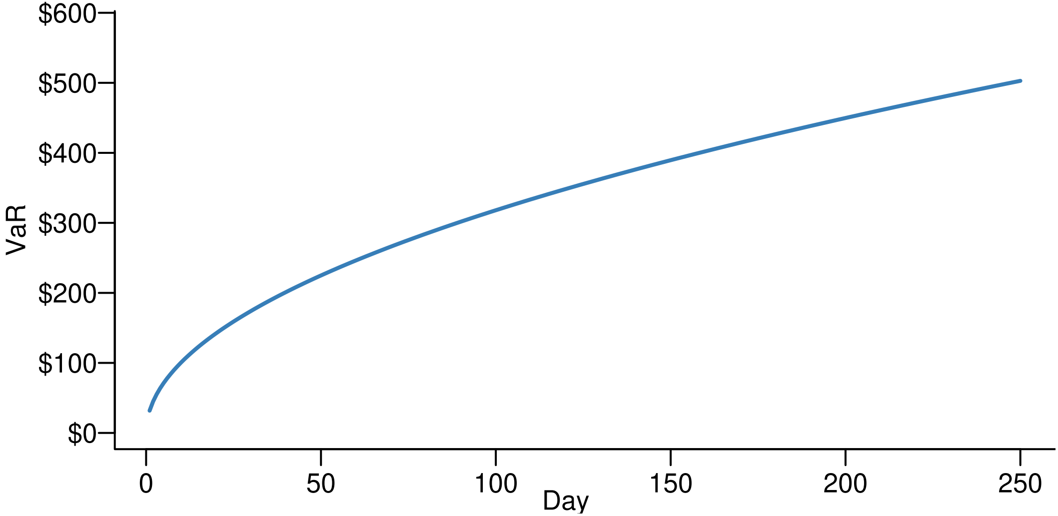
$$\begin{aligned}\text{VaR}(T\text{day}) &= -\sigma(T\text{day})F^{-1}(p) - \mu(T\text{day}) \\ &= -\sqrt{T}\sigma(1\text{day})F^{-1}(p) - T\mu(1\text{day})\end{aligned}$$

- And $\sqrt{T} < T$

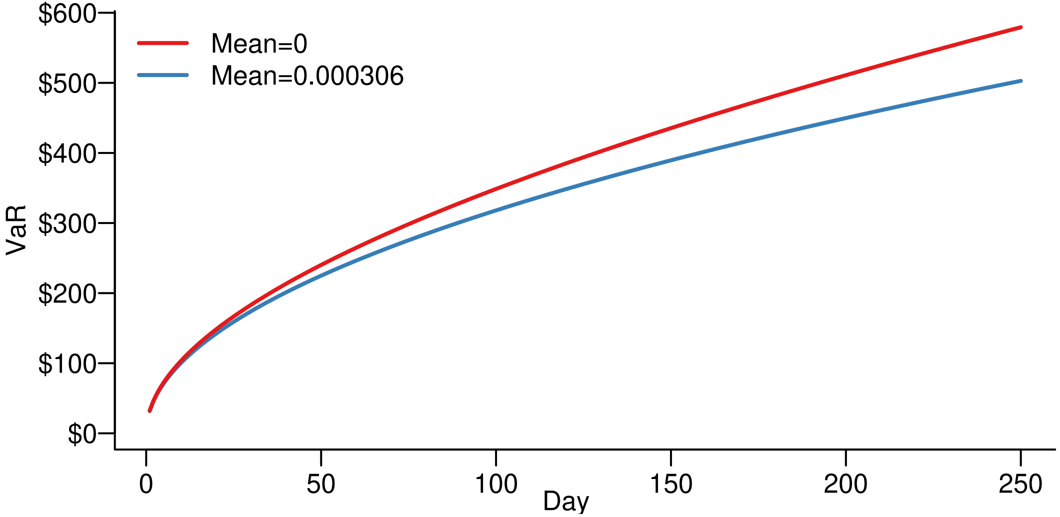
Therefore

- The assumption $\mu = 0$ is relatively harmless as the error is small at the daily level
- Daily S&P-500: $\mu = 0.019\%, \sigma = 1.15\%$
- Annual: $\mu = 4.87\%, \sigma = 18.2\%$

VaR When Returns Do Not Have Zero Mean



VaR When Returns Do Not Have Zero Mean



Issues in Including the Mean

- It is much more difficult to estimate the mean than the variance
- And unless necessary, should be avoided
- And for VaR over one day or 10 days is not necessary in most cases
- Therefore, it is common practice to remove the mean (de-mean) from returns before estimating a volatility model
- But, as a practical matter, the error made by just ignoring it, will usually be tiny

What if We Want the Mean?

- Of course, finance is all about the mean
- And we may want to forecast it along with volatility
- The idea being that the accuracy of the mean forecast is higher if we allow for time varying volatility
- So start with a standard GARCH model

$$\sigma_t^2 = \omega + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2$$

- Where

$$y_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$$

Specifying the Mean

- There are many ways one can forecast the mean, here is a simple way; ARMA(1,1)

$$\mu_t = a + by_{t-1} + c\epsilon_{t-1}$$

- Which we saw in Chapter 2

VaR With Time-Dependent Volatility

VaR with Time-Dependent Volatility — Next Steps

- The last Section assumed a fixed volatility σ when computing next-day VaR
- Empirical data show volatility clusters and varies through time
- We now model the conditional variance σ^2_{t+1} with GARCH-type processes
- VaR becomes time dependent $\text{VaR}_{t+1}(\rho)$

VaR With Time-Dependent Volatility

- When using conditional volatility models, like EWMA and GARCH
  - Estimate the model
  - Take the last in-sample volatility
  - And use that along with the last return observation
  - And the model parameters
  - To calculate one-day-ahead VaR forecast

Normal GARCH

- GARCH(1,1) is:

$$\hat{\sigma}_t^2 = \omega + \alpha y_{t-1}^2 + \beta \hat{\sigma}_{t-1}^2$$

- Software packages usually return the last volatility of the in-sample volatility ($\hat{\sigma}_t$), and the parameters
- One then needs to update the volatility by using $\hat{\sigma}_t$, y_t and the model parameters

$$\hat{\sigma}_{t+1}^2 = \hat{\omega} + \hat{\alpha} y_t^2 + \hat{\beta} \hat{\sigma}_t^2$$

- So the VaR at $t+1$ is

$$\widehat{\text{VaR}}_{t+1} = -\hat{\sigma}_{t+1} \Phi^{-1}(p) \vartheta_t$$

tGARCH

- The only difference with the tGARCH is that instead of using the inverse normal
- We use the inverse Student-t, (F), where the degrees of freedom is the parameter ν
- So the VaR at $t+1$ is
$$\widehat{\text{VaR}}_{t+1} = -\hat{\sigma}_{t+1} F_{\nu}^{-1}(\rho)\vartheta_t$$
- Except *standardisation* – see next Slide

Standardised t

- The variance of a Student-t random variable is

$$\frac{\nu}{\nu - 2}$$

- By convention, the tGARCH uses the *standardised* t density, that is, the variance is scaled so it becomes *one*
- What that means for risk calculations is that for software that uses the standardisation (like `rugarch`) we have to use the standardised t
- Because otherwise the risk estimate is too low
- So the VaR at $t+1$ is

$$\widehat{\text{VaR}}_{t+1} = -\hat{\sigma}_{t+1} F_{\nu}^{-1}(\rho) \vartheta_t \sqrt{\left(\frac{\nu - 2}{\nu}\right)}$$

Stressed VaR (and ES)

Why Stressed VaR — Motivation

- Ordinary VaR based on recent data can fall sharply in calm periods then rise in crises
- This pattern understates risk just before stressful events
- If we suspect the sample period is unusually tranquil, the VaR may be too low
- Regulations (Basel III) therefore require a Stressed VaR computed from a severe historical window
- The aim is to hold capital that would have been adequate through an earlier market turmoil
- We discuss regulations later in the course

Identifying a Stress Window

- Select a fixed length window of W_S trading days
- Slide the window through history and compute VaR (or ES) for each position
- Two approaches
  1. Choose the window giving the highest VaR
  2. Refer to a list of crisis periods, e.g. 2007 — 2009, March 2020 or April 2025
- The Stressed VaR today is then calculated with returns from that window
- Often indicated by S-VaR and S-ES

Stressed VaR — Takeaways

- Captures tail risk observed in past crises and curbs procyclic capital swings
- Sensitive to choice of stress window length and market segment
- May overstate risk for portfolios now quite different from those in the stress period
- Common practice is to hold capital as the maximum of current and stressed VaR
- Robust implementation needs regular review of the designated stress window

Model Risk of Risk Models

Model Risk — Why Forecast Choice Matters

- VaR and ES depend on the model used for volatility dynamics and return tails
- Competing models fitted to the same data can yield forecasts that differ by a factor of two or more
- Under- or over-estimating risk has direct consequences for trading limits and regulatory capital
- We quantify this dispersion with a risk-ratio framework and examine how it varies through time

Model Risk of Risk Forecast models

"Every model is wrong – Some models are useful"

The risk of loss, or other undesirable outcomes like financial crises arising from using risk models to make financial decisions

- Infinite number of candidate models
- Infinite number of different risk forecasts for the same event
- Infinite number of different decisions, many ex ante equally plausible
- Hard to discriminate

Model Risk of Risk Models

- Based on a research paper called "Model Risk of Risk Models" (2016) with Kevin James (PRA), Marcela Valenzuela (University of Chile) and, Ilknur Zer (Federal Reserve), published in the *Journal of Financial Stability*
- It can be downloaded here papers.ssrn.com/sol3/papers.cfm?abstract_id=2425689

Risk Ratios: Our Proposed Model Risk Methodology

- Consider the problem of forecasting risk for day $t+1$ using information available on day t
- Suppose we have N candidate models to forecast the risk, each providing different forecasts

$$\left\{\text{Risk}_{t+1}^{n}\right\}_{n=1}^{N}$$

- We then define *model risk as the ratio the highest to the lowest risk forecasts*

$$\text{Risk ratio}_{t+1} = \frac{\max\left\{\text{Risk}_{t+1}^{n}\right\}_{n=1}^{N}}{\min\left\{\text{Risk}_{t+1}^{n}\right\}_{n=1}^{N}}$$

Model Choice

MA Moving average

EWMA Eponentially weighted moving average

GARCH Normal innovations

t–GARCH Student-t innovations

HS Historical simulation

EVT Extreme value theory

- All models re-estimated every day

We can, and have, tried the new fancy model.
Each new model will weakly increase the *RR*

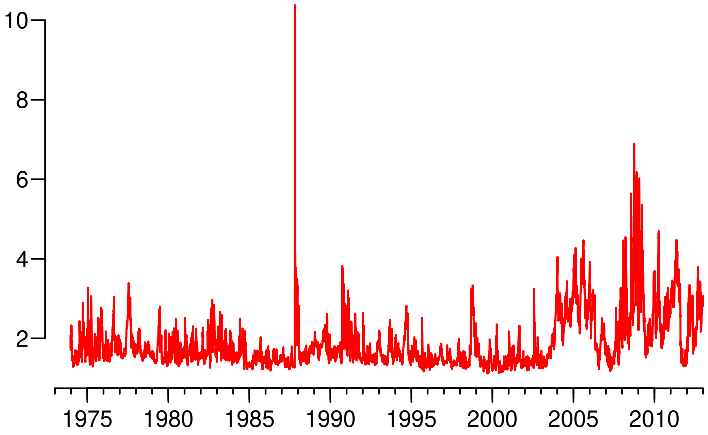
Risk Measures and Data

- Current Basel: VaR 99%
- Proposed Basel III: ES 97.5%, overlapping estimation windows
- Large financials traded on the NYSE, AMEX, and NASDAQ
  - Banking, insurance, real estate, and trading sectors
- January 1970 to December 2012.
- Sampling frequencies daily
- Sample size shown here, 1,000 days

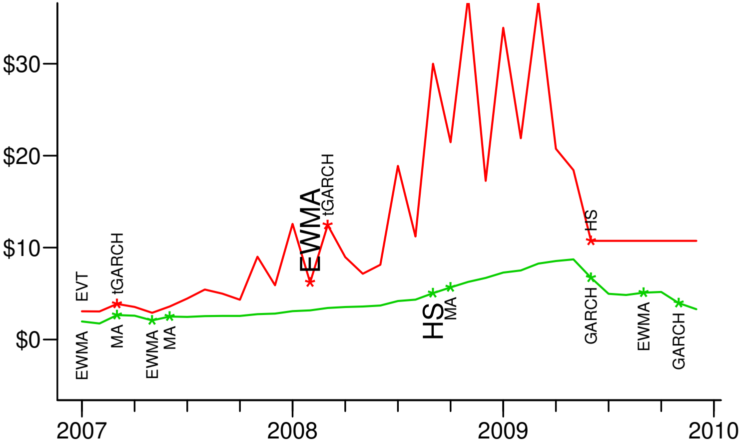
Sample Results: JPM January 3, 2007, $100 Portfolio

Model	VaR
HS	$ 3.22
MA	$ 2.91
EWMA	$ 1.96
GARCH	$ 2.13
tGARCH	$ 2.74
EVT	$ 3.22
Model risk	1.64

JPM: Model Risk (Risk Ratios)



Zooming in (End of Quarter): VaR



Each Model Has Its Own Merits

There is no clear conclusion of which one is better

Particular models perform best in a particular environments and purposes

- React quickly to news – volatility-based, like GARCH
- Easy to compute – EWMA, MA, HS
- Very small samples – EWMA, MA
- Low volatility of risk – HS
- Tails – EVT, tGARCH

Model Risk — Key Takeaways

- Forecast dispersion across models can rival the size of the market risk itself
- Risk ratios well above one flag material model uncertainty that should feed into capital buffers
- No single model dominates across all regimes — combining or adaptively switching models mitigates error
- Backtesting (later discussion) reveals bias but cross-model comparison captures dispersion
- Robust governance — documentation, independent validation and periodic recalibration — is essential

“Why Risk Is So Hard to Measure” (2016)

Jon Danielsson and Chen Zhou, 2016,

Evaluating Risk Forecast Performance — Why It Matters

- Previous Sections developed VaR and ES via historical, parametric and stressed methods
- A critical issue is how well these forecasts perform in finite samples
- Sampling noise and model error can mask genuine changes in market risk
- We draw on Danielsson and Zhou (2016) to quantify this uncertainty with Monte Carlo simulations and CRSP stock data
- The goal is to measure the dispersion of VaR and ES estimates and assess the reliability of their confidence bounds

How Accurate Are Market Risk Measurements?

- Consider the various market risk forecasting methods
- A lot of papers exist on the asymptotic properties of various methods
- Or comparing method A to method B
- We could not find any paper on how the various methods work in small samples
- That is, in practice
- Calculate the *empirical* 99% confidence bounds

Objective

- What is the relationship between ES and VaR?
  - VaR(99%) and ES(97.5%) because of Basel
- What are the small sample properties of these risk measures?
- Risk measures compared by Monte Carlo simulations
  - 10^7 simulations (yes, we need that many)
  - Because we are estimating quantiles of quantiles
- And theoretic analysis
- And with CRSP data
- Across sample sizes and tail thicknesses

Methodology — Monte Carlo and Empirical Design

- Generate 10^7 artificial return samples for each length $T \in \{100, 250, 500, 1000, \ldots\}$
- Data-generating processes — iid Normal$(0, 1)$ and Student-t$(\nu{=}3)$ scaled to unit variance
- For every sample compute 1% VaR and ES via historical simulation
- Build the sampling distribution of each estimator and extract 95% and 99% confidence bounds from empirical quantiles
- Empirical check — roll a window over CRSP US stock returns and repeat the VaR / ES estimation
- Compare Monte Carlo bounds with the dispersion observed in real data to assess finite-sample reliability

Estimation Method

- Historical simulation
- For most of the data generating processes it is appropriate
- And we don't want to get into the business of "which method is best for which DGP"
- Instead, a widely used method in practice

Accuracy of VaR and ES

- We know the asymptotic properties
- But what happens when the sample size becomes smaller and smaller?
- Across various tail thicknesses

CRSP Stocks 99% Risk

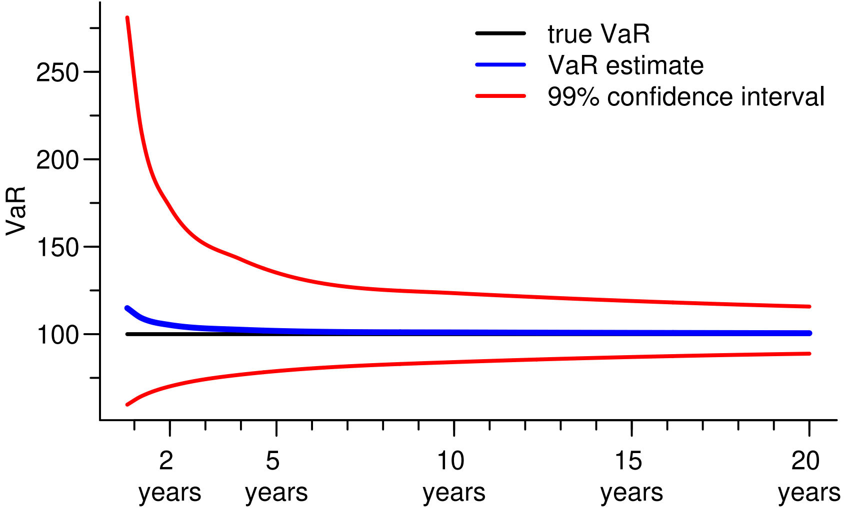
Expect VaR and ES to be one

	VaR	ES
Sample size	99% confidence bound	
300 days	[0.65,1.49]	[0.63,1.28]
4 years	[0.74,1.35]	[0.69,1.35]
20 years	[0.84,1.20]	[0.80,1.24]

Might a random number generator be cheaper and as reliable?

Finite Sample Properties of VaR With Monte Carlo

$\alpha = 3$



Performance Study — Key Takeaways

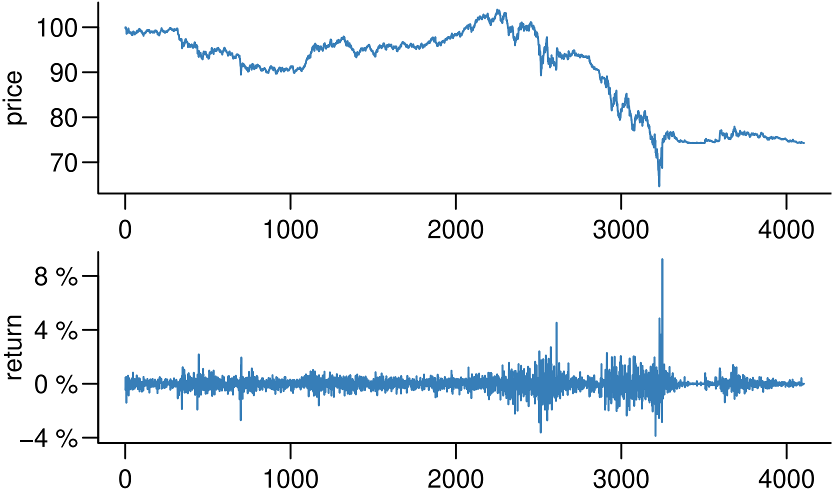
- Small samples inflate the estimation error of both VaR and ES — the dispersion can exceed the risk itself for $T \leq 250$
- Student-t tails widen confidence bounds dramatically compared with the normal DGP
- Monte Carlo bounds line up with the spread seen in CRSP stocks, confirming that sampling noise dominates short windows
- ES is more efficient than VaR asymptotically but requires larger samples to outperform in practice
- Regulatory backtesting should account for finite-sample uncertainty when flagging VaR / ES exceptions

“What The Swiss FX Shock Says About Risk Models ”

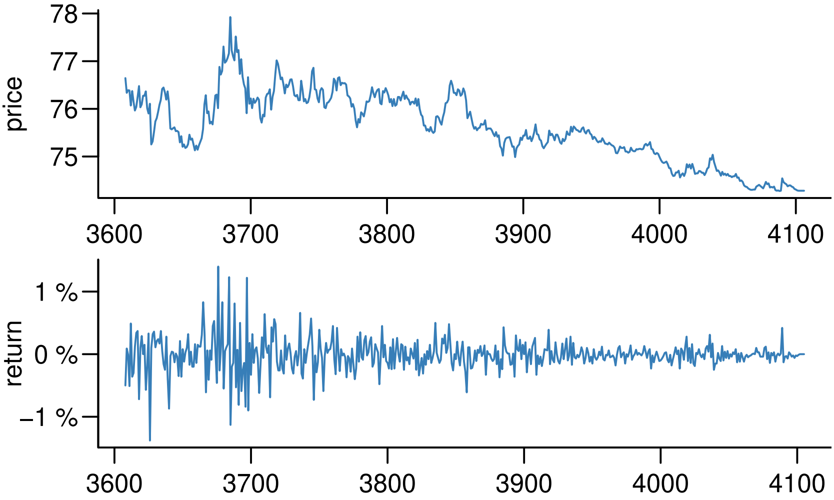
(2016)

Jon Danielsson (2016)
VoxEU.org

Some Actual Price Series



Some Actual Price Series (Zoom In)



VaR and ES

Estimated with "reputable" models generally accepted by authorities and industry

MA Moving average

EWMA Exponentially weighted moving average

GARCH Normal innovations

t–GARCH Student-t innovations

HS Historical simulation

EVT Extreme value theory

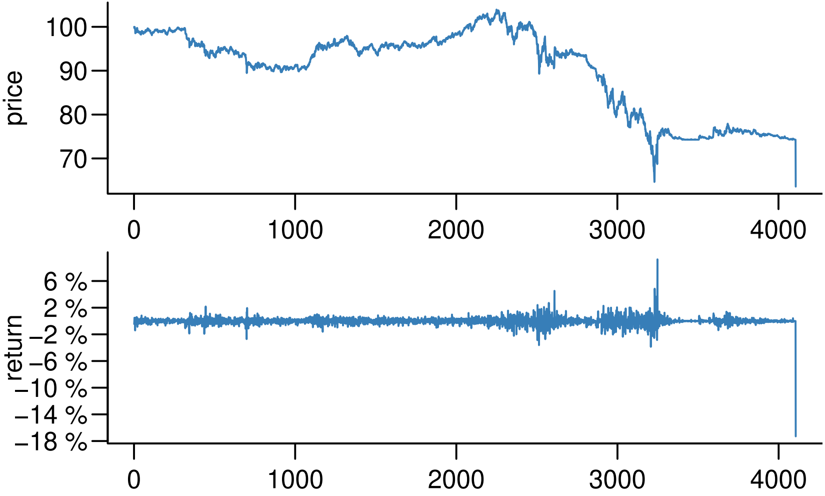
- While other models may be discussed, these six cover the vast amount of use cases
- Estimation period, 1,000 days
- Other assumptions give qualitatively similar results

Risk for the Next Day $(T+1)$

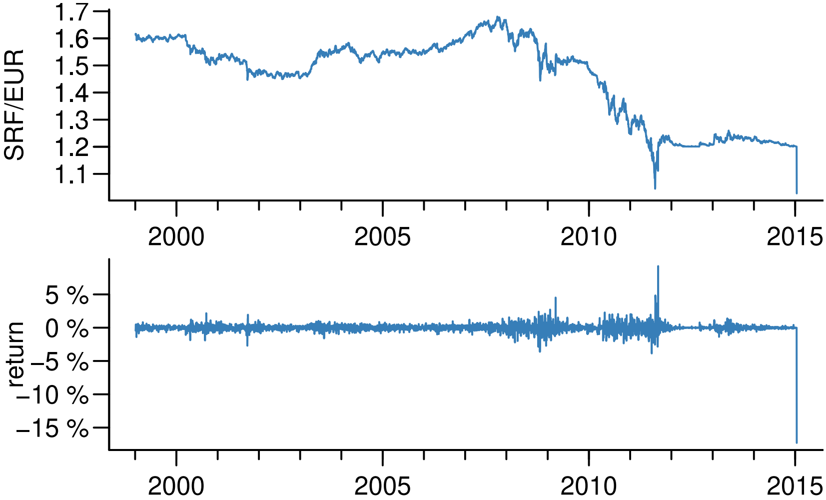
Portfolio value is 1,000

Model	VaR	ES
HS	*14.04*	20.33
MA	11.42	13.09
EWMA	*1.59*	1.82
GARCH	1.71	1.96
tGARCH	2.10	2.89
EVT	13.90	24.41
Model risk	*8.85* $= {}^{14.04}/{}_{1.59}$	$13.43 = {}^{24.41}/{}_{1.82}$

Let's Add One More Day



How many francs a euro buys

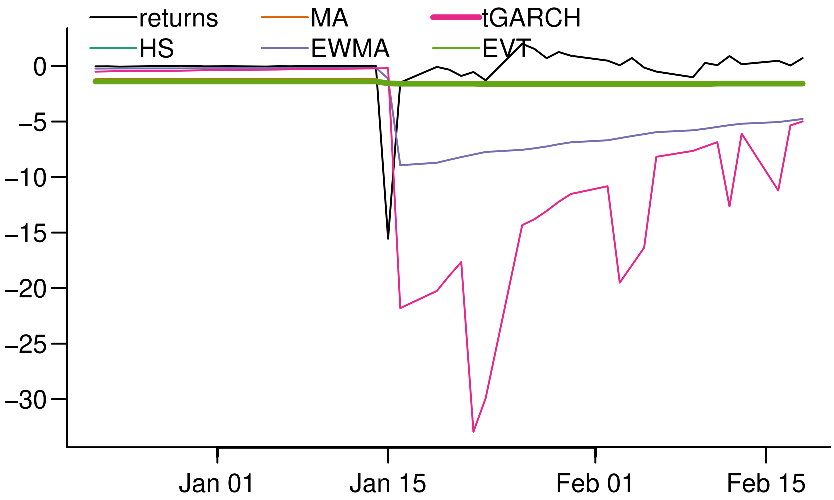


How Frequently Do the Swiss Appreciate by 15.5%?

measured in once every X years

Model	frequency	
EWMA	never	
GARCH	never	
MA	2.7×10^{217}	age of the universe is about 1.4×10^{10}
tGARCH	1.4×10^{7}	age of the earth is about 4.5×10^{9}
EVT	109	

Even More Interesting After the Event

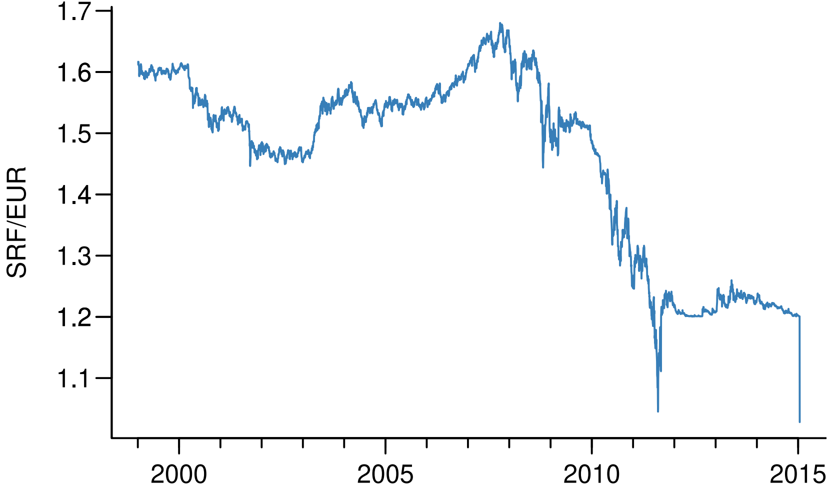


So

- Depending on model, risk, may or may not, not move
- Some models signal very high risk when we know nothing else will happen

But Is the Event All That Extraordinary?

just eyeballing it seems not that much



Should We Care?

- I am often told that nobody manages FX risk with such methods
- But still they are a part of Basel III market risk
- Some countries use them for pension fund regulations
- The accuracy, or lack thereof, is representative for many other situations and methodologies

Could We Forecast Risk Better?

- If one considers who owns the Swiss National Bank
- And some factors, perhaps
  - SNB dividend payments
  - Money supply
  - Reserves
  - Government bonds outstanding
- Yes, we can do much much better than the models used here

The Swiss Takeaway

- Before the shock: “1-in-a million year event”
- After the shock: “It’ll happen again tomorrow”
- Both wrong. But both from the same model
- It’s not just wrong — it’s confidently wrong

Appendix

Derivation of ES under Normality

$$\begin{aligned}
\text{ES} &= \frac{1}{\rho}\int_{-\infty}^{-\text{VaR}_t(\rho)} x\, f(x)\, dx \\
&= \frac{\vartheta\sigma}{\rho}\int_{-\infty}^{-\Phi^{-1}(\rho)} z\,\phi(z)\, dz && \text{with } z=\frac{x}{\vartheta\sigma} \\
&= \frac{\vartheta\sigma}{\rho}\,\phi\big(\Phi^{-1}(\rho)\big)
\end{aligned}$$

◂ Back to main ES normal slides