MODEL RISK IN VAR CALCULATIONS

Peter Schaller, Bank Austria ∼ Creditanstalt (BA-CA) Wien, peter@ca-risc.co.at

Introductory remarks:

- Risk $\Leftarrow$ lack of information
  (We do not know the future)
- Risk depends on
  - portfolio
  - market dynamics

  and

  - information used by observer

- This has two consequences
  1. The less information we have, the higher the risk
  2. Risk measures have a subjective component:
     On the same day for the same portfolio different estimates for the same risk measure may be correct

$\Rightarrow$ Assessment of estimates to be based on a series of forecasts

VAR as risk measure

- Quantile of P/L distribution
- Drawback: not subadditive ($\Rightarrow$ not coherent)
- Still:
  - Widely used in practice
  - Enforced by regulators
- Possible reasons
  - Solely depends on P/L distribution
  - Finite for any portfolio under any distributional assumptions
  - Straightforward assessment of quality of estimates via backtesting
- Some of the ideas presented here maybe applicable to other risk measures based on the P/L distribution

Backtesting

- Back testing methods:
  1. Count number of excesses
  2. Advanced (E.g. investigate identical distribution of excesses over time)
- If an estimate fails the first test, further tests are superfluous

- Counterexample:
  - Very large estimate on 98% of days
  - Very low VAR estimate for 2% of days
  - will result in 1% of excesses
- Excluded, if we demand VAR to be function of portfolio and market history only without explicit time dependence

VAR calculation

- Calculate quantile of distribution of profits and losses
- Distribution to be estimated from historical sample
- Straightforward, if there is a large number of identically distributed historical changes of market states

However:

- Sample may be small
  - Recently issued instruments
  - Availability of data
  - Change in market dynamics !!
- Estimation from small sample induces the risk of a misestimation

Model risk

- Estimation of distribution may proceed in two steps
  1. Choose family of distributions (model specification)
  2. Select distribution within selected family (parameter estimation)
- This may be seen as inducing two types of risk
  1. Risk of misspecification of family
  2. Uncertainty in parameter estimates

- This differentiation, however, is highly artificial:
  - If there are several candidate families we might choose a more general family comprising them
  - This family will usually be higher dimensional
  - The problem of model specification is partly transformed into the problem of parameter estimation
  - Risk of misspecification is traced back to the risk from parameter misestimation
  - Indeed, uncertainty in parameter estimates will be larger for the higher dimensional family
- So, in practice, choice is not between distinct models, choice is between simple model and complex model containing the simple model

Trade off

- A simple model will not cover all features of the distribution, e.g.
  - time dependent volatility
  - fat tails
- This will result in biased (generally too small) VAR estimates
- In a more sophisticated model we will have a larger uncertainty in the estimation of the distribution
- This introduces another source of risk
- The effect will be seen in the back testing
- So, again, back testing shows an underestimation of VAR

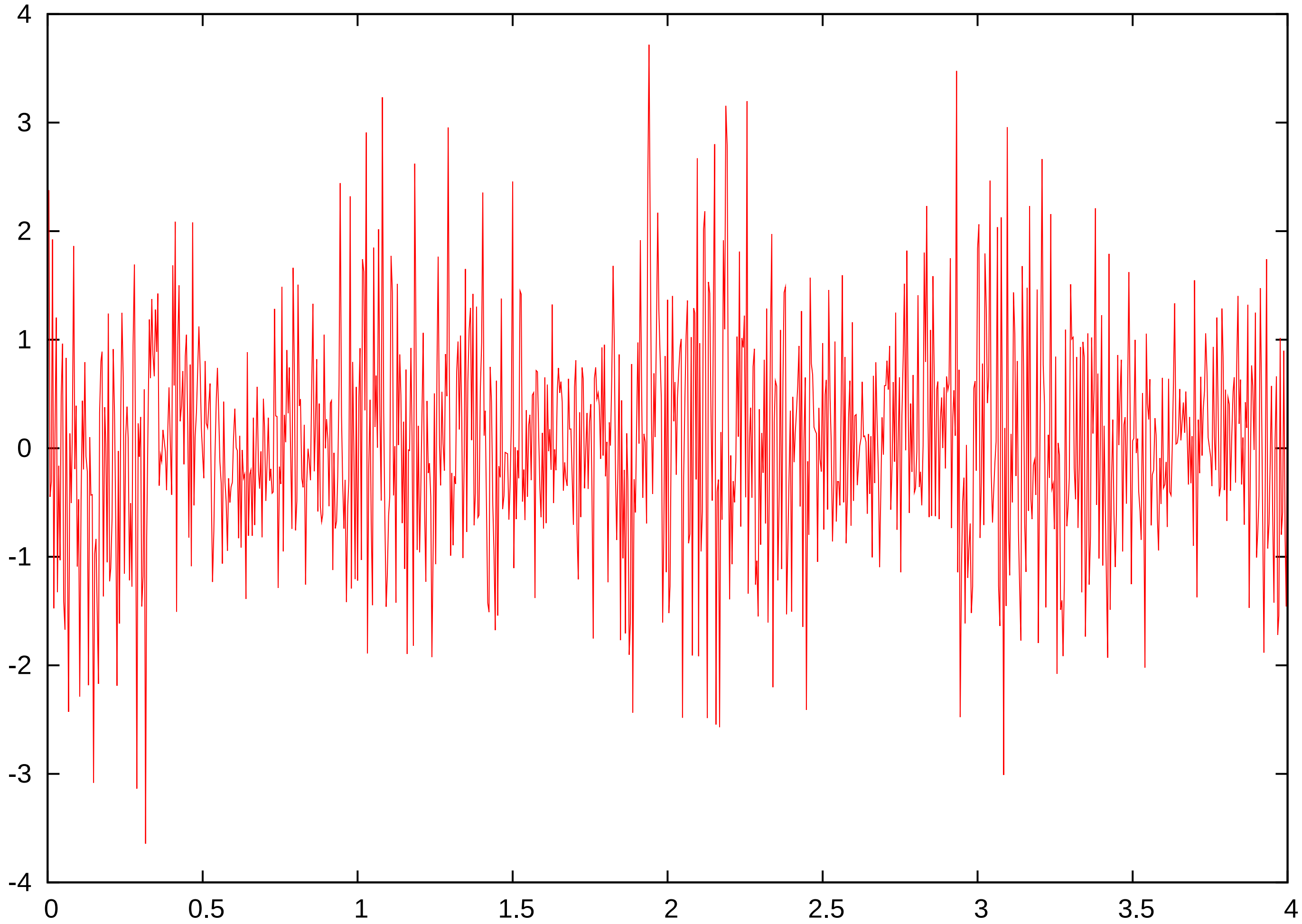
Contents:

- Guiding examples
  - Bias vs. uncertainty
  - Impact of model risk on back testing results
- Incorporating model risk into VAR
  - Classical approaches to handle model risk
  - Consistent inclusion into VAR forecast
  - Applications
- Model risk and expected shortfall
- Comparison to Bayesian approach

Example I: time dependent volatility

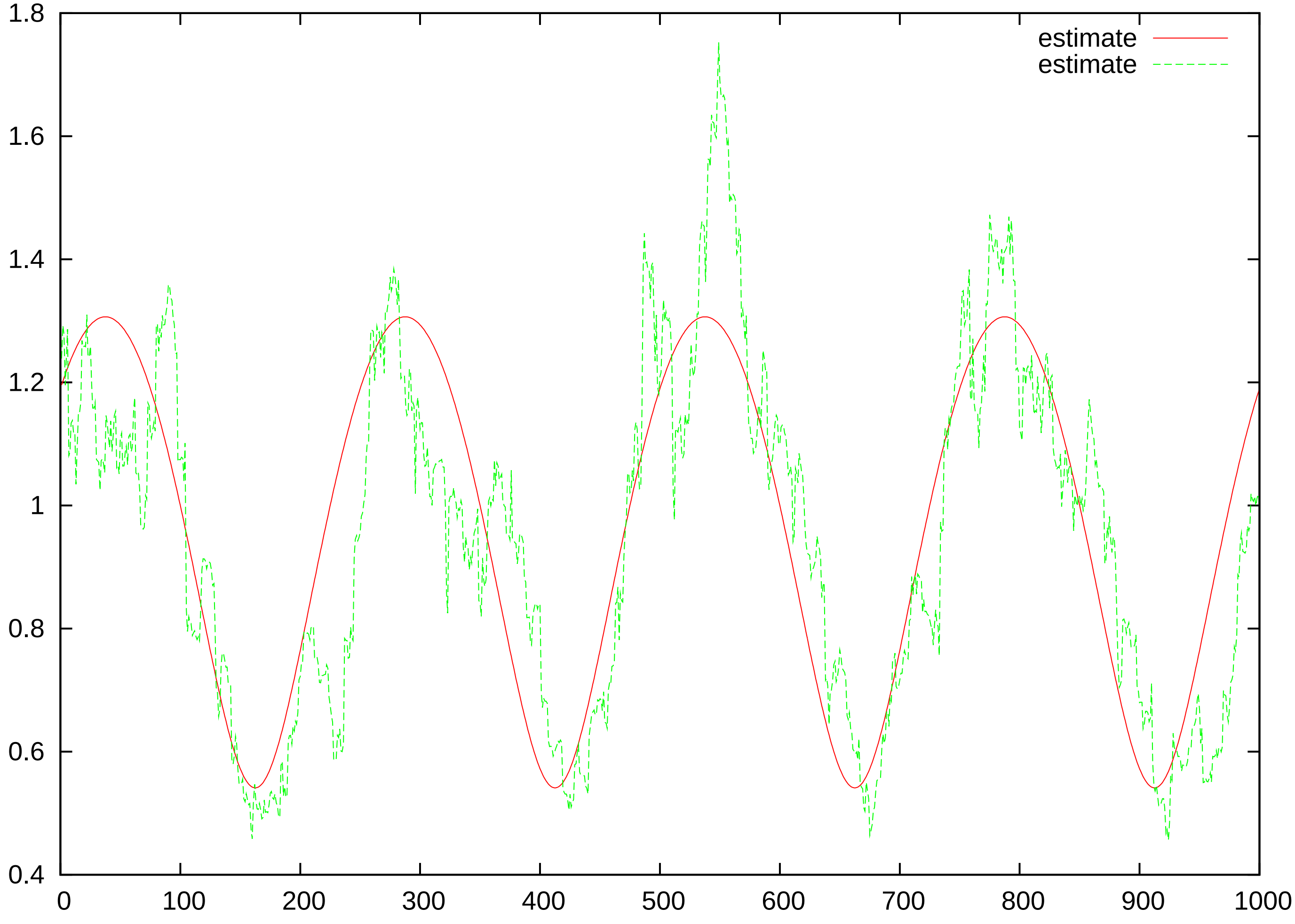
- Daily returns are normally distributed, time dependent volatility
- Volatility varies between 0.55 and 1.3
- average volatility is 1
- e.g.: $\sigma^2 = 1 + 0.7 * \sin(2\pi t)$

Time series of normally distributed returns with varying volatility (4 years)



- With normal distribution assumption and a long term average of the volatility ($\sigma = 1$) we get a $\text{VAR}_{0.99}$ of 2.33
- Back testing will show 1.4% of excess values rahter than 1%
- Note: Excesses not identically distributed over time
- Way out: Calculate volatility from most recent 25 returns to get time dependent volatility
- Again we will find some 1.4% of excesses
- Note: Excesses now (almost) identically distributed over time

Volatility estimate from 25 returns



- Estimating time dependent volatility:
  - Long lookback period leads to systematic error (bias)
  - Short lookback period leads to stochastic error (uncertainty)
- Both seen in back testing results

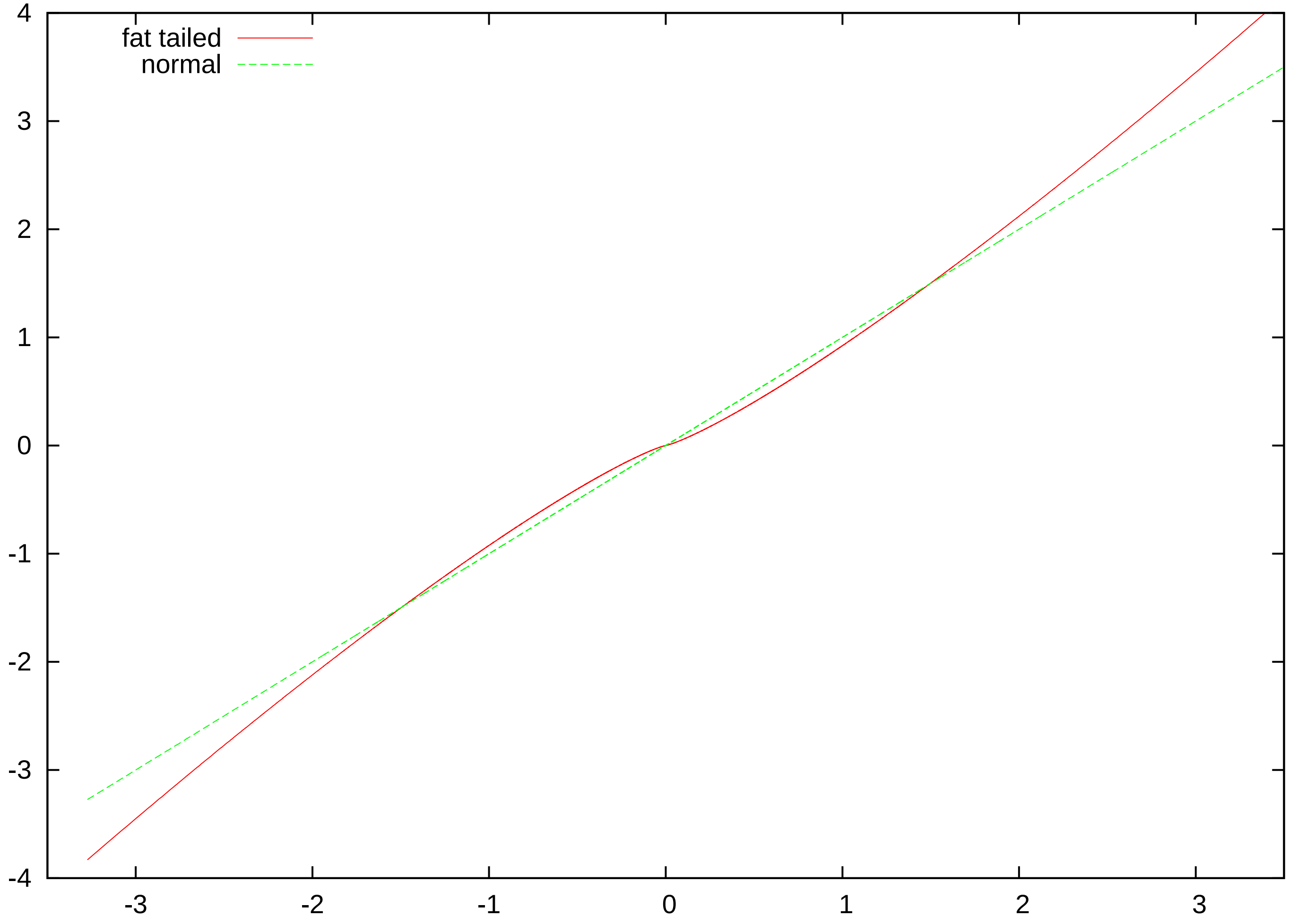
- Assume returns normally distributed with $\sigma = 1$
- Volatility estimated from n-day lookback period
- 99% quantile calculated from estimated volatility under normal distribution assumption
- the following table show, how average number of excess values depend on n

n	excesses
10	2.1%
20	1.5%
50	1.2%
100	1.1%

Example II: Fat tailed distribution

- Model fat tailed returns as function of normally distributed variable: e.g.: $x = a * \text{sign}(y) * |y|^b$, y normally distributed
- parameter b determines tail behavior:
  - normal for $b = 1$
  - fat tailed for $b > 1$
- volatility depends on scaling parameter a

Fat tailed distributions for b=1.25:



- Modeling as normal distribution:
  - Assume perfect volatility estimate
  - 1.5% excesses of estimated $\mathrm{VAR}_{0.99}$
- Modeling as fat tailed distribution
  - Two parameters have to be estimated
  - With a lookback period of 50 days we obtain 1.5% of excesses
- Compare normal distribution:
  50 days of lookback period $\Rightarrow$ 1.2% of excesses
- The result does not depend on the actual value of b

- Interpretation: With the complexity of the model the uncertainty of the parameter estimates increases
- Again there is a trade off between
  - bias in the simple model
  - uncertainty in the complex model

The general situation

- Distribution $P(\vec{\alpha})$ member of family $\mathbf{P}$ of distributions labeled by some parameters $\vec{\alpha}$
- For estimation of $\vec{\alpha}$ a (possibly small) sample $< \vec{x} >$ of independent draws from $P(\vec{\alpha})$ available

Estimation of parameters:

- Choose estimator $\hat{\alpha}(\vec{x})$
- Calculate $\hat{\alpha}$ value for given sample
- Identify this value with $\vec{\alpha}$

However:

- $\hat{\alpha}$ is itself a random number
- A value of $\vec{\alpha}$ different from the observed value could have produced sample

Classical approaches:

Statistical testing

- Use distribution of $\hat{\alpha}$ to formulate conditions on a reasonable choice of $\vec{\alpha}$
- A range of values of $\vec{\alpha}$ will match
- Satisfactory, if admissible range of values is small

Bayesian approach:

- Assume prior distribution for $\vec{\alpha}$
  $\Rightarrow$ Conditional distribution of $\vec{\alpha}$ depending on observed value of $\hat{\alpha}$
  $\Rightarrow$ Stochastic mixture of distributions from family $\mathbf{P}$
- Calculate VAR estimate from the latter
- Some features
  - Assumes, that VAR is quantile of some distribution $P(\vec{\alpha}) \in \mathbf{P}$
    Effectively calculates VAR from stochastic mixture of distributions
  - In this way includes risk of misestimation of $\vec{\alpha}$ into VAR
  - However, depends on choice of prior distribution
  - In general, will not lead to a VAR figure behaving well in the back testing

Method

- In http://papers.ssrn.com/sol3/papers.cfm?abstract_id=308082 method was presented, which
  - incorporates the uncertainty in the parameter estimates
  - does not depend on the assumption of a prior distribution
  - behaves well in the back testing
- We will shortly review it

Starting point: Given is

- A family $\mathbf{P}$ probability distributions parameterized by a set of parameters $\vec{\alpha}$
- A finite sample $< x_1, ...x_n >$ of independent draws from a particular member $P(\vec{\alpha}) \in \mathbf{P}$.
- A priory nothing is known about $\vec{\alpha}$
- VAR estimate should produce correct back testing results

Back testing:

- V_q (VAR for confidence level q): is a function of $< x_1, ...x_n >$
- Repeat experiment k times $\rightarrow$ k samples $< x_1^a, ...x_n^a >$
- $\rightarrow$ k quantile estimates $V_q^a = V_q(x_1^a, ..., x_n^a)$
- No explicit time dependence (V_q^a dep. on a via sample only)
- Compare V_q^a with next draw x_{n+1}^a
- x_{n+1}^a should exceed V_q^a in q percent of the cases.

Note:

- Different functions of sample may be correct quantile estimates

Different point of view

- Effectively we have a $n + 1$-dimensional sample of i.i.d. variables
- $P(\alpha) \in \mathbf{P}$ induces multivariate distr. $P_{mult}(\alpha)$ of samples
- Assume function $\Phi(x_1, ..., x_n; x_{n+1})$ such that
  - distribution of Φ does not depend on $\vec{\alpha}$
  - $\Phi_0(x_{n+1}) := \Phi(x_1^0, ..., x_n^0; x_{n+1})$ is strictly monotonic in x_{n+1}
- Given q and historical sample $< x_1^0, ..., x_n^0 >$
  - calculate q-quantile for distribution of Φ
  - calculate corresponding value of x_{n+1} from inverse of Φ_0

Eventually we have

$$V_q(x_1^{(0)}, ..., x_n^{(0)}) = {\Phi_0}^{-1}(Q_q^{\Phi})$$

Result:

- Obviously the above construction will produce a VAR estimate behaving correctly under the back testing described above

Remarks:

- Different choices of Φ lead to different (albeit correct) VAR estimates
- Distribution of Φ depends neither on historical sample nor on $\vec{\alpha}$
  $\Rightarrow$ Determination of Q_q^Φ has to be done once only
  $\Rightarrow$ Possible even if it needs expensive simulation
- Though inspired by a problem from financial risk management, the method may be well applicable in other fields.

Construction of Φ

- Assume $\mathbf{P}$ generated by the action of some Lie group G on $\mathbb{R}$, i.e.:
  - Fix distribution P_0
  - X P_g-distributed for $X = g\,Y$ with Y P_0-distributed and $g \in G$
- Assume that only identity acts trivial on P_0
- Assume some estimator $\hat{g}(\vec{x})$ for the group element g corresponding to the distribution the sample $\vec{x}$ was taken from
- Let the estimator be G-homogeneous: $\hat{g}(g(\vec{x})) = g\,\hat{g}(x)$
- Let $\Phi = \hat{g}^{-1}(x_1, ..., x_n)\, x_{n+1}$
- Distribution on Φ does not depend on distribution of $< x_1, ...x_{n+1} >$

Proof:

- $\hat{g}(x)$ solves the equation $\hat{g}(g^{-1}(x)) = id$ w.r.t. g
- $\hat{g}(< y_1, ...y_n >) = id$ generates $(n+1-d)$–dimensional surface in $\mathbb{R}^{n+1}$ $(d = \dim(G))$
- Action of G forms d-dim G-invariant orbits in $\mathbb{R}^{n+1}$
- These orbits are invariant under group transformations
- $\hat{g}^{-1}\, x_{n+1}$ is $(n{+}1)$-th coordinate of intersection point between this surface and orbit through $< x_1, ...x_{n+1} >$
- Change of distribution induced by G transformation
- G-invariance of Φ immediately follows from G-invariance of orbits
- q.e.d.

G-homogeneous estimators

- Assume r.v. X $g_1 P_0$ distributed
- Consider $Y = g_2 X$ as different variable on same probability space
- Estimate for probability space should not depend on parametrization of event space
- From this point of view homogeneity of $\hat{g}$ appears as natural condition

G-homogeneous estimators, examples

- Most likelihood estimator
- Construction used in the cited paper
  - Denote by f a $\mathbb{R}^d$-valued functional on $\mathbf{P}$ ($d = \dim(g)$) with $f(Y) = 0 \;\Leftrightarrow\; Y$ P_0-distributed
  - Denote by $\hat{f}(x_1, ..., x_n)$ an estimator of f for the sample $< x_1, ...x_n >$ of size n
  - $\hat{f}(\hat{g}^{-1}(\vec{x}))$ defines homogeneous estimator $\hat{g}$ for g

$d = 1$

- Consider scale transformation $X \to \alpha X$
  ($G = (\mathbb{R}_+, \times)$)
- Generates family of distributions characterized by scale parameter α
- Any reasonable estimator $\tilde{\alpha}$ for α will be homogeneous
  ($\hat{\alpha}(\lambda\vec{x}) = \lambda\hat{\alpha}(x)$)
- Choose $\Phi = x_{n+1}/\hat{\alpha}(x_1, ..., x_n)$
- Result:

$$p_\Phi = E_{P_0^n}[\hat{\alpha} p_0(\hat{\alpha}\Phi)]$$

  with p_0 ... density of P_0
  and P_0^n ... dist. of n independent draws from P_0

Example: Normal distribution

- Standard deviation σ as scale parameter
- As an estimator choose weighted sum $\hat{\sigma} = \sqrt{\sum w_i x_i^2}$ with $\sum w_i$=1
- Sample may be infinite, but recent returns have higher weights than past returns. This has a similar effect as a finite sample.
- Result (N denotes normalization constant:)

$$p(\Phi) = N \prod_{i=1}^{n} \frac{1}{\sqrt{1 + w_i \Phi^2}} E[\sqrt{\mu(x_i)}]$$

with

$$\mu(x_i) = \sum_{i=1}^{n} \frac{w_i x_i^2}{1 + w_i \Phi^2}$$

and $E[.]$ denoting the expectation value w.r.t. standard normal dist.

- For constant weight over sample of size n we obtain StudentT distribution with n degrees of freedom
  (Note that $\hat{\sigma}$ is square root of χ^2 distr. variable
- For general choice of weights:
  - Expand $\sqrt{\mu}$ into Taylor series at $\mu_0 = E[\mu]$
  - Allows approximation of result in terms of moments of normal distr. to arbitrary order in $\mu - \mu_0$
- Popular:
  - EWMA: $w_i = \lambda^{n-i} / \sum \lambda^{n-i}$
  - GARCH(1,1): $w_i = p/n + (1-p)\,\lambda^{n-i} / \sum \lambda^{n-i}$

Note on GARCH(1,1)

- volatility estimate for GARCH(1,1) may be written as wighted average of long term estimate and EWMA estimate:
  $\hat{\sigma}^2_{GARCH} = p\,\sigma_0^2 + (1-p)\,\hat{\sigma}^2_{EWMA}(\lambda)$
- σ_0 is the long term average of the volatility
  weight p, and decay factor λ depend on parameters α, β, and γ of GARCH process

d=2 example

- Characterization of $\mathbf{P}$
  - P_0 ... standard normal distribution
  - Variable from $P(a,b) \in \mathbf{P}$ is generated by transformation $x = g(a,b) \cdot y := a\,\mathrm{sgn}(y)\,|y|^b\,, \quad a,b > 0$
- Straightforward to prove that this transformations form a group
- Note: $P(a,b)$ fat tailed if $b > 1$
- Standard normal distr. may e.g. be characterized by variance and kurtosis
- Standard estimators for these quantities may be used (e.g. empirical values of the sample)
- Distr. of Φ may be generated by simulation (Once only even in the case of daily estimates!!)

Coherent extension of VAR (CVAR)

- In contrast to quantile the conditional mean of the events beyond the quantile is coherent (i.p. sub additive) risk measure
- Can we calculate this quantity from Φ (E.g. By multiplying volatility estimate with conditional mean of Φ in case of one parameter family of distributions)?

Gedanken experiment

- For normally distr. losses choose size of historical sample $n = 1$

$$\Downarrow$$

  - absolute value of most recent return is estimate for std. dev.
  - Φ is StudentT distr. with one degree of freedom
- 75% quantile of the latter equals 1
  $\Rightarrow$ Abs. value of most recent loss is VAR for confidence level of 75%

Compare

- Cond. mean of StudentT distr. with one deg. of freedom is infinite
- Naive back testing would produce a finite result for the CVAR:

$$CVAR_{back-testing} = \frac{1}{0.25}\hat{E}_{normal-distr}[(x_t)\theta(x_t - |x_{t-1}|)]$$

However

- This back testing assumes constant size of portfolio
- Assume
  - Family of distributions related by scale transformations
  - Portfolio with constant VAR limit l: Whenever VAR estimate deviates from l portfolio will be resized by a factor l/VAR
- Apply back testing with this regularly resized portfolio
- Product of cond. mean of Φ and limit l is CVAR result compatible with back testing

Comparison with Bayesian approach

Setting

- Consider one parameter family of distributions:
  - P_1 ... arbitrary distr. with standard deviation of 1
  - P_α ... distr. generated from P_1 by transformation $x \rightarrow \alpha \cdot x$
- Choose homogeneous estimator ϕ for stand. dev.: $\Phi = x_{n+1}/\phi \qquad (*)$
- Is there a prior distr. for stand. dev. such that Bayesian approach generates correct result?

Note:

- In the Bayesian approach VAR is calculated from a stochastic mixture of distributions
- In view of (*) distr. of Φ may be interpreted as stochastic mixture of P_σ distributions where σ has distr. of quantity $1/\phi$ (calculated with α=1)

Result:

- After some calculations using
  - Homogeneity of ϕ
  - $p_{az}(x) = \frac{1}{a} p_z(x/a)$
  - $p_{1/z}(x) = \frac{1}{x^2} p_z(1/x)$

  we find:

- Bayesian approach gives same result as our method, if density of prior distribution for std. dev. is chosen according to $p_{prior}(\sigma) = 1/\sigma$