PIVOTAL QUANTILE ESTIMATES IN VAR CALCULATIONS

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

Contents

References

- P.Schaller: Uncertainty of parameter estimates in VAR calculations; Working paper, Bank Austria, Vienna, 2002; SSRN abstract_id 308082.
- G.Pflug, P.Schaller: Pivotal quantile estimates in VAR calculations; in preparation.

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
  - Uncertainty in parameter estimates will be larger for the higher dimensional family
  - Eventually, problem of model specification is partly transformed into problem of parameter estimation
- In practice, choice is often not between distinct models, choice is between simple model and complex model containing the simple model

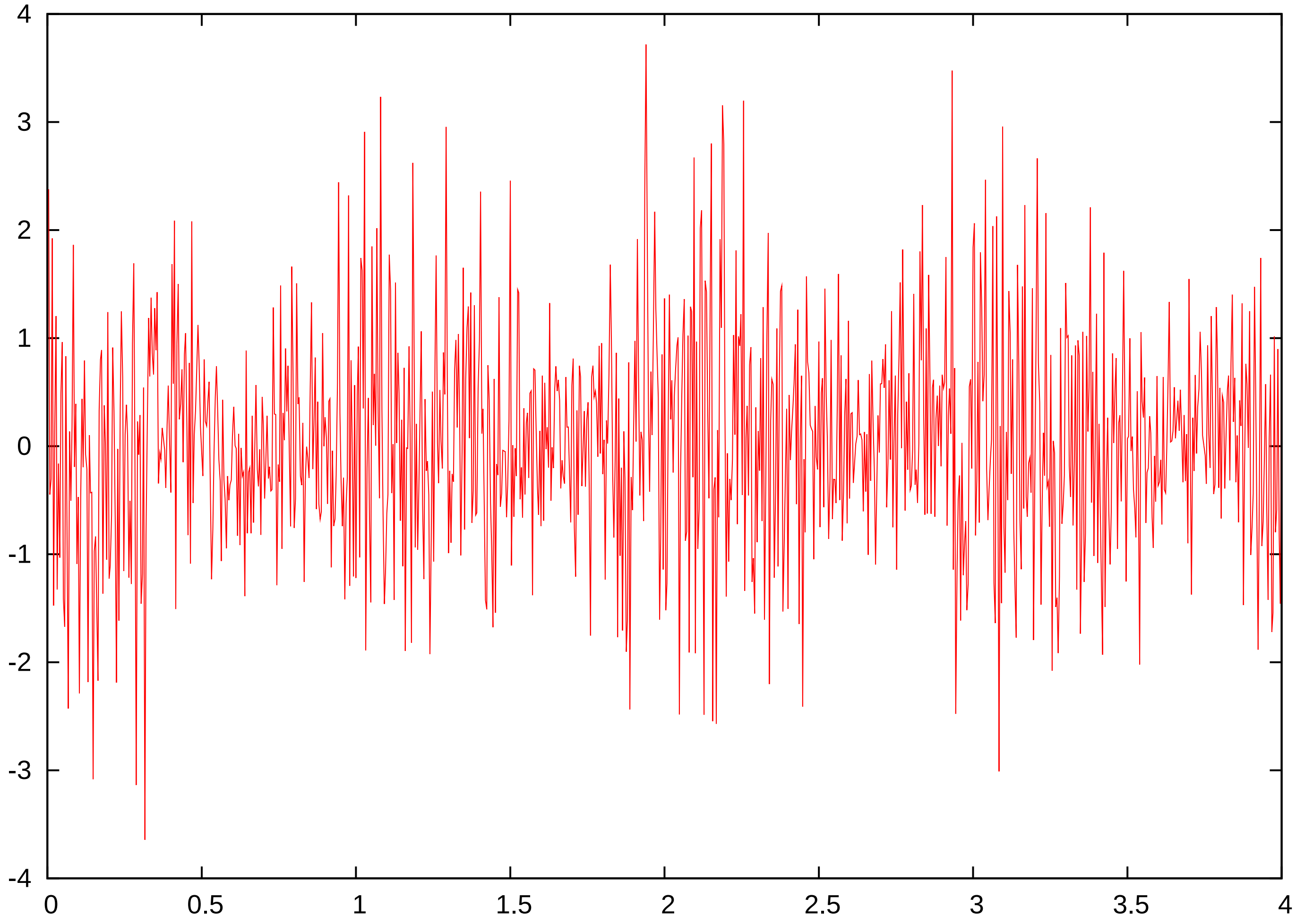
Trade off

- 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 exposes us to model risk

Example I: time dependent volatility

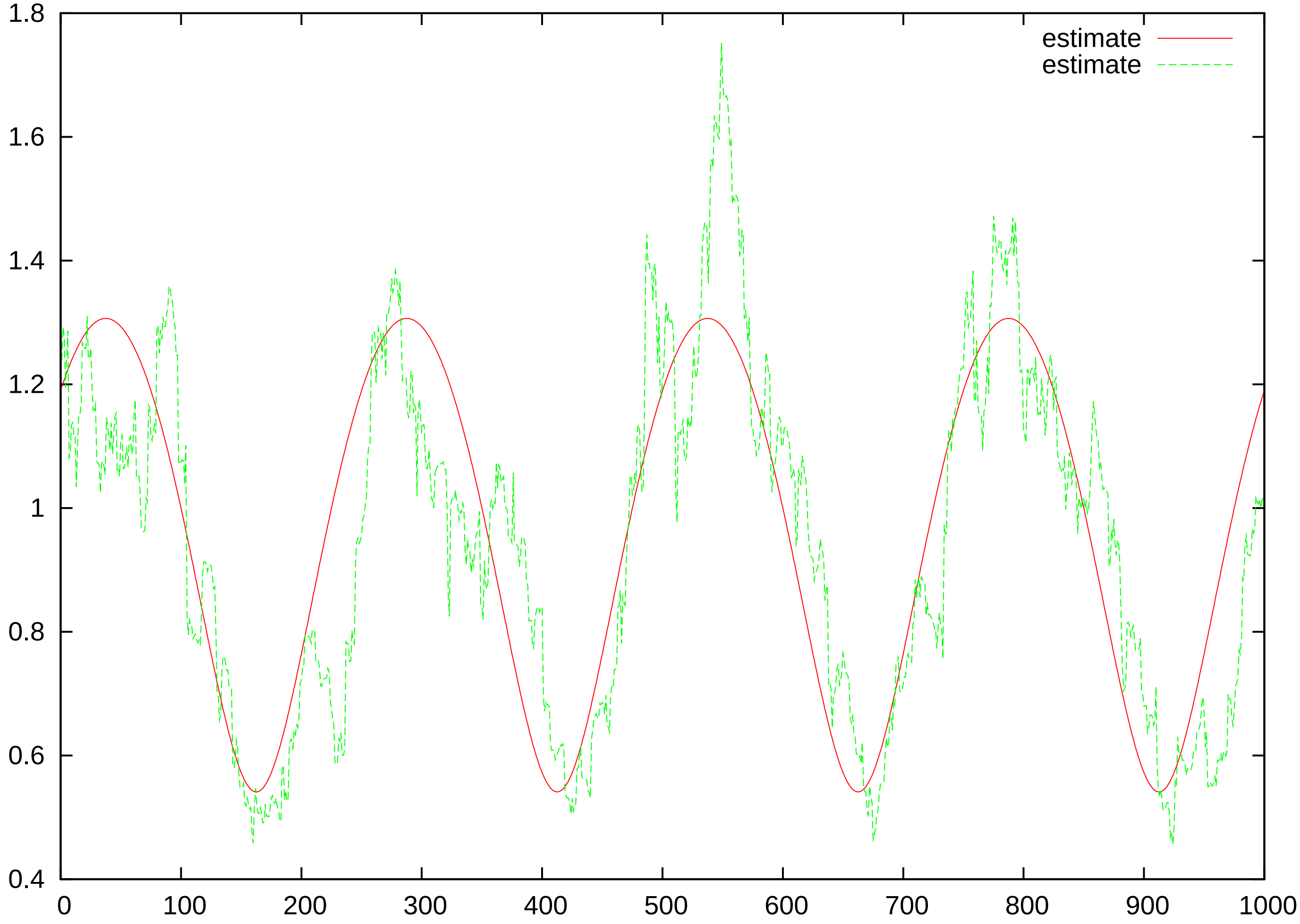
- Daily returns are normally distributed, time dependent volatility
- Volatility varies between 0.55 and 1.3
- average volatility is 1
- e.g.: $\sigma^2 = 1 + 0.7 * \sin(2\pi t)$

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
- On the average this will lead to 1.4% of excess returns rather than 1%
- Note: Excesses not uniformly 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) uniformly distributed over time

Volatility estimate from 25 returns

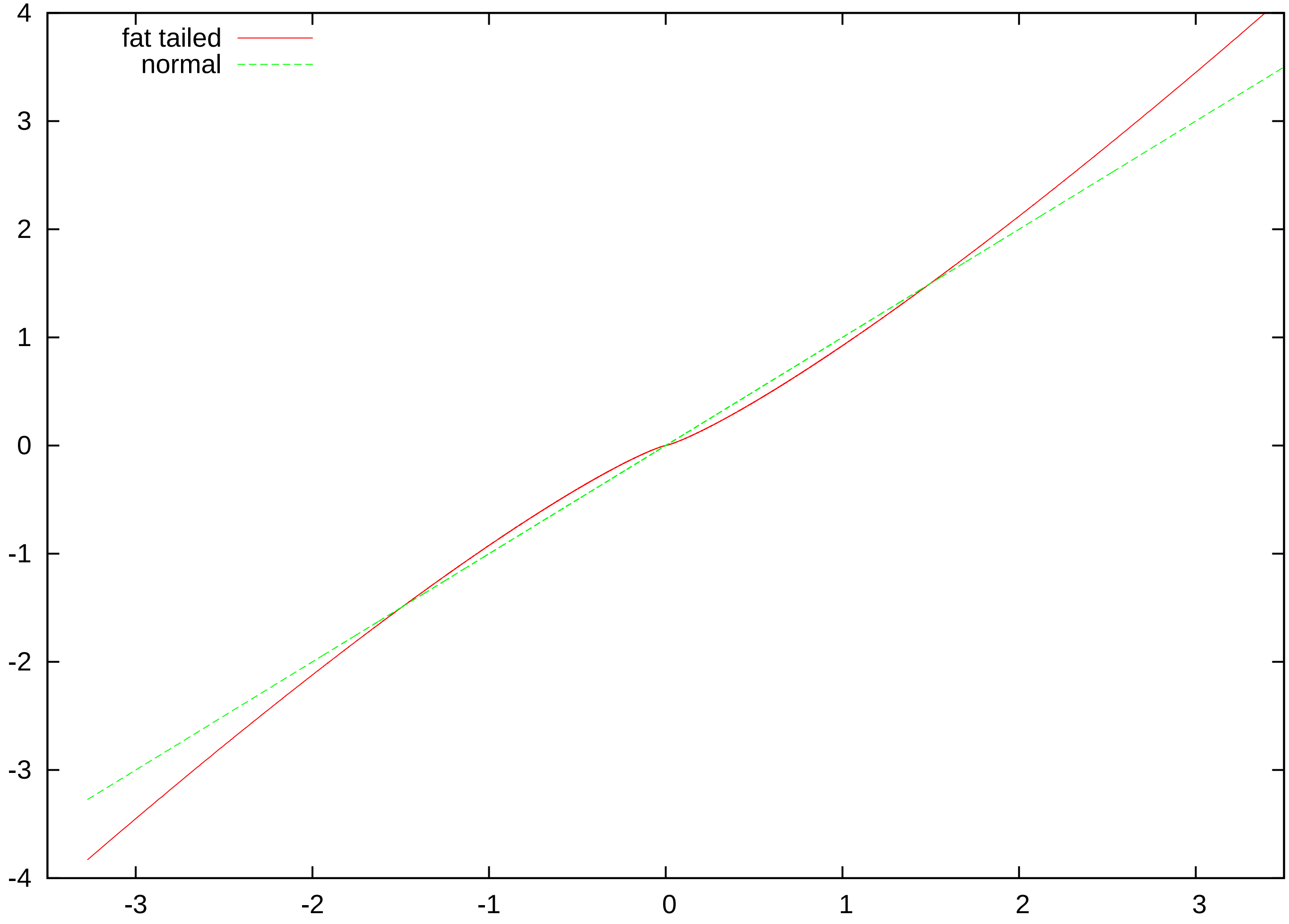


- Estimating time dependent volatility:
  - Long lookback period leads to systematic error (bias)
  - Short lookback period leads to stochastic error (uncertainty)
- Both seen in back testing of the VAR estimate:
  Probability of excess return is higher than expected from VAR confidence level

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
- The result for the two parameter model does not depend on the actual value of b:
  - The model would also generate 1.5% of excesses for b=1 (corresp. to norm.dist.)
  - Compare to normal distribution assumption:
    50 days of lookback period $\Rightarrow$ 1.2% of excesses for norm.dist. returns

- 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

Example III: Oprisk Capital

- 99.9% quantile (VAR with 99.9% confidence level) of yearly aggregate losses to be calculated
- Typical observation period: 5 years
- Sample may be increased by external data
- Still, direct estimation of the quantile is not possible
- Bootstrapping
  - Split yearly loss into series of independent loss events
  - Estimate distribution of size of events (severities)
  - Estimate frequency
  - Calculate distribution of yearly losses by convolution

Remarks:

- Sampling is always subject to lower threshold
- Frequencies are (approximately) Poisson distributed by definition
- Severities will be fat tailed (E.g. Pareto tails with exponent close to one)

Synthetic example

- Assume severity distribution is Pareto: $F(x) = 1 - x^{-1/\chi}$
  $x \in \{1, ..., .\infty\}$ ... ratio between severity and sampling threshold
- On the average 200 losses per year above threshold
- 5 years of observation $\Rightarrow$ Sample size N=1000
- Relevant external data may increase sample size to $N = 10000$
- Estimate χ via MLE ($\chi = \overline{log(x)}$)
- stdev. of estimator $\sigma_\chi = \chi/\sqrt{N}$

- Single loss approximation
  - For fat tailed distribution loss in bad years is dominated by single huge loss
  - ⇒ For calculation of high quantiles distribution of aggregated losses can be approximated by distribution of annual loss maxima
- Result for $\chi = 1$
  - VAR=200000 (in units of lower threshold)
  - With an error of ± 2 stddev. for χ estimate will lead to result fluctuating between 92400 and 432600 (internal data only) res. 156600 and 255100 (with external data)
  - Accuracy of single loss approximation:
    FFT result for $\chi = 1$ is 202500

- Use lower sampling threshold to increase sample size
  - Problematic in view of the large quotient between result and sampling threshold
  - Complete sampling may be difficult to achieve for low threshold
  - In practice, the opposite is done (Peak over Threshold method)
- To be on the safe side would be costly!

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

Naive argument:

- With some probability we will underestimate quantile

  ⇒ Probability that next year's loss will exceed quantile estimate is higher than 1-q

- With some probability the we will overestimate quantile

  ⇒ Probability that next year's loss will exceed quantile estimate is lower than 1-q

- Effects might average out and overall probability that next year's loss is above the estimate might be 1-q

- The estimate could then be interpreted as VAR with a confidence level of q

- Unfortunately it does not work out, as seen in the examples

Question

- Can we find estimate such that probability of next year's loss to be above estimate is precisely 1-q ?
  (q ... confidence level of VAR estimate)

Pivotal quantile estimate

- Definition: A quantity $Q_q(X_1, \ldots, X_n)$ is denoted as pivotal quantile estimate, if

$$Prob\{X_{n+1} \leq Q_q(X_1, \ldots, X_n)\} = q \quad \forall \alpha$$

- Example:
  - Consider family of all continuous probability distributions on $\mathbb{R}$.
  - Let $Y_1, \ldots, Y_n$ be the order statistics of a sample of i.i.d. variables from some member of this family. Then
  - Y_k is a pivotal quantile estimate for $q = k/(n+1)$.
- In the following we will consider families of distributions allowing a pivotal quantile estimate for all levels of q

- Lemma: The following statements (a) and (b) are equivalent:
  (a) A family of distributions (P_α) allows for a pivotal quantile estimate $Q_q(X_1, \ldots, X_n)$ for all $q \in (0, 1)$.
  (b) A pivotal function (i.e. a function whose distribution does not depend on α) $V(X_1, \ldots, X_{n+1})$ exists, such that the distribution of V is continuous and V is strictly monotonic in X_{n+1}

- Proof:
  - $(a) \Rightarrow (b)$: The inverse of $Q_q(X_1, \ldots, X_n)$ with respect to q applied to X_{n+1} is uniformly distributed for all q
  - $(b) \Rightarrow (a)$: Denote by Q^V the quantile function for the distribution of V: $Prob\{V \leq Q^V(q)\} = q$.
    Then the inverse of V w.r.t X_{n+1} applied to $Q^V(q)$ is a pivotal quantile estimate.

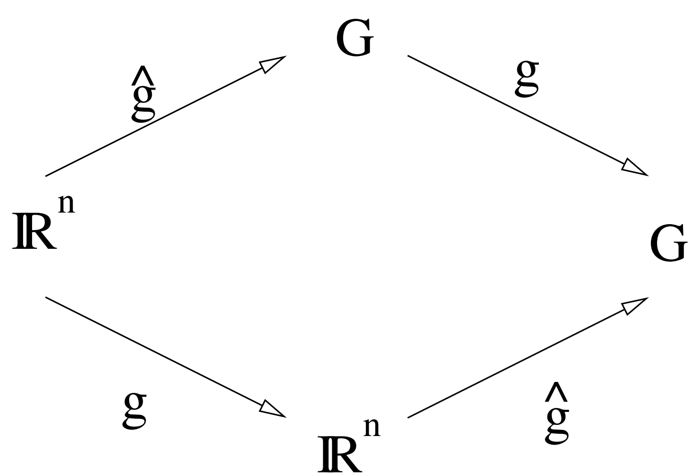
Structure models

- Let G be a group of monotonic bijective transformations on the real line and let P be some probability measure on $\mathbb{R}$
- By P^g we denote the transformed measure

$$P^g(A) = P(g^{-1}(A)).$$

- A mapping $x^{(n)} \mapsto \hat{g}_{x^{(n)}}$, which maps $\mathbb{R}^n$ into G is called G-equivariant, if for all $g \in G$ and all vectors $x^{(n)}$

$$\hat{g}_{g(x^{(n)})} = g \circ \hat{g}_{x^{(n)}}.$$



- Consider a structure model $(P^g)_{g \in G}$ is given.
  - Let $X_1, \ldots, X_n, X_{n+1}$ be an i.i.d. sequence from P^g for some unknown g.
  - Let $\hat{g}_{x^{(n)}}$ be G-equivariant.

  $\Rightarrow$ $V = \hat{g}_{X^{(n)}}^{-1}(X_{n+1})$ is pivotal.

  $\Rightarrow$ If V has a continuous distribution function F, a pivotal quantile estimate is given by

$$Q_q(X^{(n)}) := \hat{g}_{X^{(n)}}(F^{-1}(q)).$$

Construction of equivariant maps

- For all $x^{(n)} \in R^n$, let

$$\mathcal{O}(x^{(n)}) = \{y^{(n)} : \exists g \in G \text{ such that } x^{(n)} = g(y^{(n)})\}$$

  be the orbit of $x^{(n)}$.

  – For $x^{(n)}$ and $y^{(n)}$ on the same orbit, there is a g with $y^{(n)} = g(x^{(n)})$.

  – Orbits are either disjoint or identical.

- Let $r(x^{(n)})$ be a maximalinvariant selection
  (i.e. $r(x^{(n)}) \in \mathcal{O}(x^{(n)})$, r constant in each orbit

- Let $\hat{g}$ be defined through the relation

$$\hat{g}_{x^{(n)}} r(x^{(n)}) = x^{(n)}.$$

$\Rightarrow$ $\hat{g}(x^{(n)})$ is G-equivariant

Example: MLE

- The most likelihood estimator is equivariant.
- r is given by samples with the following property:
  The maximum of the likelihood function is located at P.

Example: Location-scale families.

- $g_{a,b}(x) = a + bx \ (b > 0)$
- A location estimate $\hat{\mu}(X^{(n)})$ is location/scale equivariant, if for all a and all $b > 0$

$$\hat{\mu}(a + bX^{(n)}) = a + b\hat{\mu}(X^{(n)})$$

- A scale estimate is equivariant, if $\hat{\sigma}(a + bX^{(n)}) = b\,\hat{\sigma}(X^{(n)})$.
- $(X^{N+1} - \hat{\mu})/\hat{\sigma}$ is pivotal
- Transformations with $a = 0$ form subgroup

Results I: Normal distribution with time dependent volatility

- Standard deviation 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.
- Popular schemes like EWMA, GARCH(1,1) may be treated in this way.
- $V = x^{n+1}/\hat{\sigma}$ is pivotal
- Pivotal quantile estimate given by the product of $\hat{\sigma}$ and the quantile of V

- Probability density of V given by

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

with

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

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 $\sqrt{\nu}$ into Taylor series at $\nu_0 = E[\nu]$
  - Allows approximation of result in terms of moments of normal distr. to arbitrary order in $\nu - \nu_0$

Results II: fat tails

- 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
- Standard normal distr. may e.g. be characterized by variance and kurtosis:
  - With standard estimators $\hat{V}$, $\hat{K}$ for these quantities (e.g. empirical values of the sample):
  - Maximalinvariant selection given by $\hat{V} = 1$ and $\hat{K} = 3$
  - Solve $\hat{V}(g^{-1}(\hat{a},\hat{b})\,x^{(n)}) = 1$ and $\hat{K}(g^{-1}(\hat{a},\hat{b})\,X^{(n)}) = 3$ w.r.t. $\hat{a}$, $\hat{b}$
  - Pivotal function given by $V = (g^{-1}(\hat{a},\hat{b})\,)X_{n+1}$

Note:

- As an alternative MLE for a, b could be used as $\hat{a}$, $\hat{b}$
- Distr. of V may be generated by simulation (Once only even in the case of daily estimates!!), as it does not depend on actual values of a,b

Results III: Oprisk VAR

- Choose $F_0 = 1 - 1/x$
- Transformation $x \to x^{\chi_0}$ will generate Pareto distribution with parameter χ_0
- In single loss approximation for Oprisk VAR target quantity is $x_a = \max(x_1, .., x_f)$, of Pareto distributed variables, where f is the annual frequency of losses
- Under change of transformations it will transform in the same way as severity x
- We choose MLE estimator $\hat{\chi} = \sum_{i=1,\ldots,N} \log(x_i)/N$ from historical severities x_i
- distribution of $V = x_a^{1/\hat{\chi}}$ is invariant under change of transformation

Conservative estimate of VAR

- Compute distribution of V (e.g. by simulation)
- Determine its 99.9% quantile Q
- Need to be done after each change of sample size/frequency
- Estimate $\hat{\chi}$ from available historical data
- $Q^{\hat{\chi}}$ is then to be taken as VAR estimate
- If distributional assumptions are correct, it will be exceeded with a probability of 0.1%
- Some additional term may be necessary to account for the error in the single loss approximation

Numerical result

- 200 losses/year, sample size 1000
- Simulation of the distribution of V with 10 Mio runs leads to 213000 ± 2000 as 99.9% quantile of V
- Note, that this result does not depend on the value of χ