Extreme Value Theory

And

How we can use it for financial risk management

Mandira Sarma

Centre for International Trade and Development, JNU, Delhi

16 October 2010

- Extreme value theory deals with the asymptotic behaviour of the extreme observations (maximum or minimum) of n realisations of a random variable.

- Financial risk management is all about understanding the large movements in the values of asset portfolios.

Extreme value Theory

$X \in (l, u)$ is a random variable.
$f \longrightarrow$ the density of X.
$F \longrightarrow$ the distribution function of X.
$X_1, X_2, ...X_n$ are n independent realisations of X,

and

$$Y_n = max\{X_1, X_2, ..., X_n\}$$
$$Z_n = min\{X_1, X_2, ..., X_n\}$$

EVT is the theory of the asymptotic behaviour of Y_n and Z_n as n becomes large.

The exact distribution of extremes is degenerate in the limit

Assuming $X_1, X_2, ..., X_n$ to be *iid*, the distribution function of Y_n, denoted by $P(.)$ is given by

$$\begin{aligned} P(y) &= \Pr\{Y_n \leq y\} \\ &= \Pr\{max(X_1, X_2, ...X_n) \leq y\} \\ &= \Pr\{X_1 \leq y, X_2 \leq y, ..., X_n \leq y\} \\ &= [F(y)]^n \end{aligned}$$

Similarly, the distribution function of Z_n, denoted by $K(.)$ is given by

$$K(y) = 1 - [1 - F(y)]^n$$

as $y \to l$, both $P(y)$ and $K(y)$ tend to zero, and as $y \to u$, both $G(y)$ and $K(y)$ tend to unity.

In EVT, distribution of suitably normalised extrema is studied.

Y_n and Z_n are transformed with a scale parameter $b_n (> 0)$ and a location parameter $a_n \in R$, such that the distribution of the standerdised exrema

$$Y_n^{'} = \frac{Y_n - a_n}{b_n} \quad \text{and} \quad Z_n^{'} = \frac{Z_n - a_n}{b_n}$$

is non-degenerate.

The two extremes, the maximum and the minimum are related by the following relation:

$$min\{X_1, X_2, ..., X_n\} = -max\{-X_1, -X_2, ..., -X_n\}$$

Therefore, all the results for the distribution of maxima leads to an analogous result for the distribution of minima and vice versa.

Two important theorems

1. The Fisher-Tippett theorem (1928)
2. The Pickands-Balkema-de Haan theorem (1974)

Two important distributions

1. The Generalised extreme value distribution (GEVD)
2. The Generalised Pareto distribution (GPD)

The Fisher-Tippett Theorem [Extremal Types Theorem] (1928)

If $\exists$ *constants* $b_n (> 0)$ *and* $a_n \in R$ *such that*

$$\frac{Y_n - a_n}{b_n} \xrightarrow{d} H \quad \text{as} \quad n \to \infty$$

for some non-degenerate distribution H*, then* H *must be one of the only three possible 'extreme value distributions'.*

In that case, $X \in DA(H)$.

(Similar to CLT for averages!)

The Standard Extreme Value Distributions

Type I or Gumbel Class (thin-tailed):

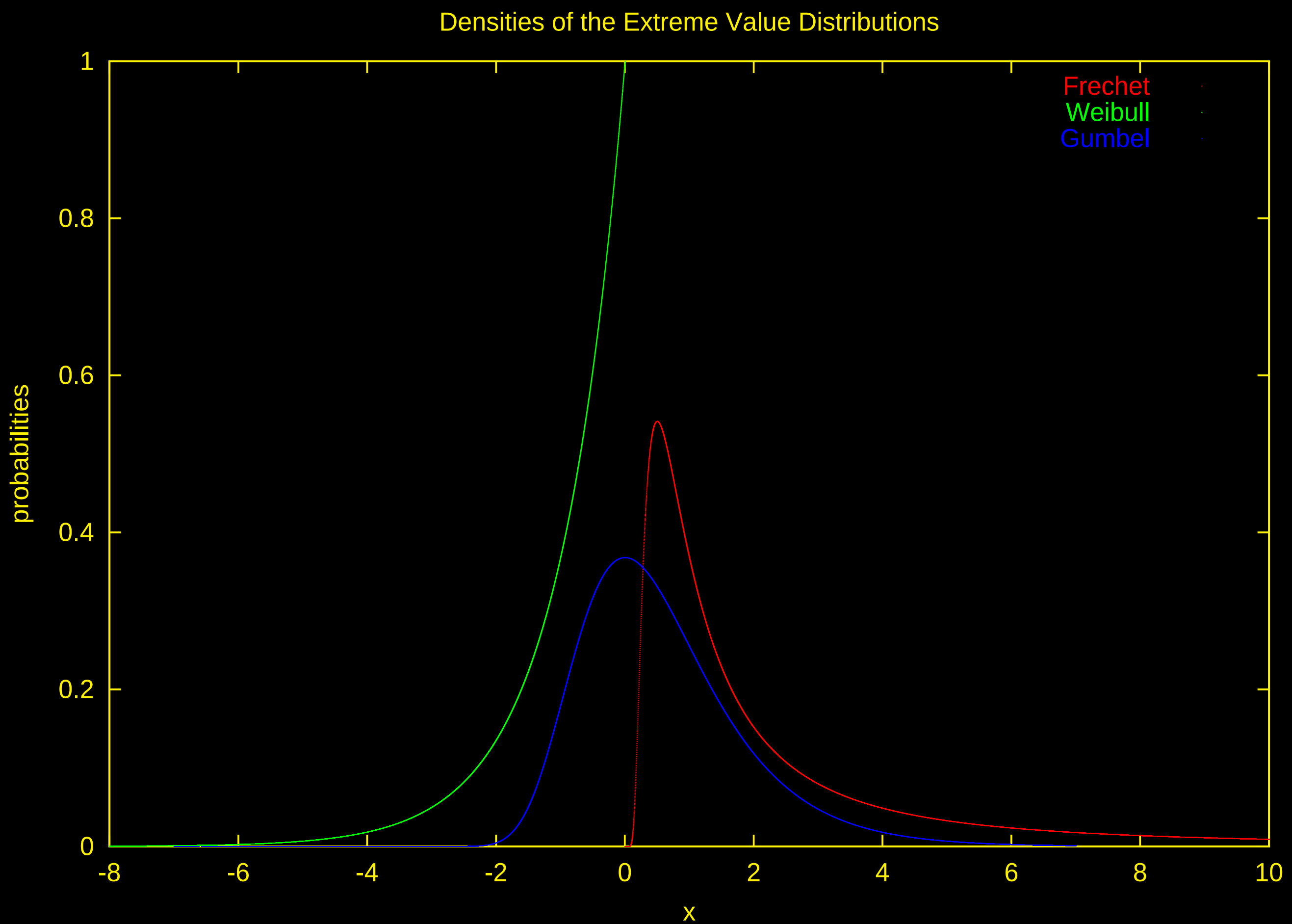
$$\bigwedge(x) = \{exp(-exp^{-x})\} \tag{1}$$

Type II or Fréchet Class (fat-tailed):

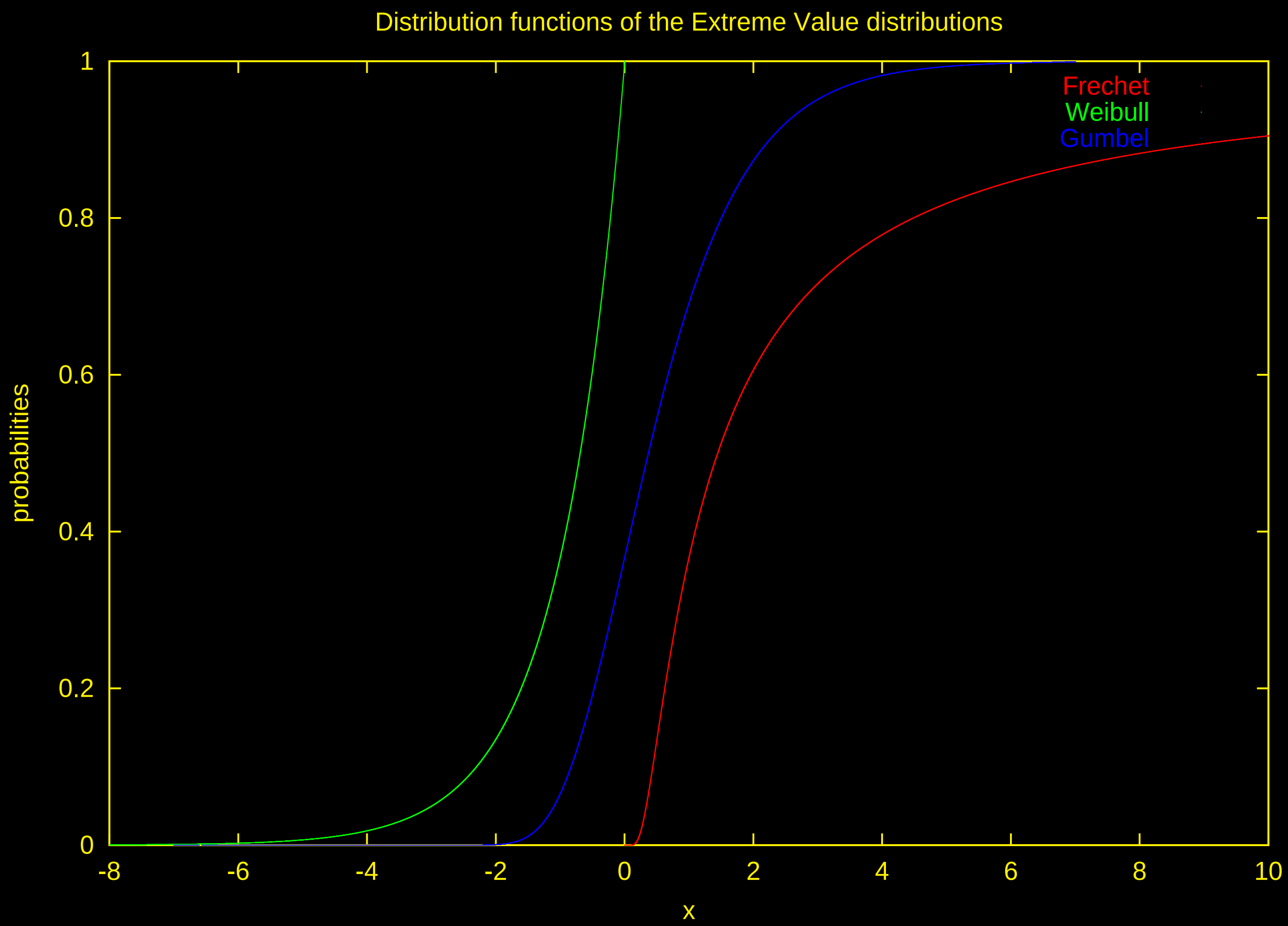
$$\Phi_\alpha(x) = \begin{cases} 0\ ; & x \leq 0 \\ exp(-x^{-\alpha}); & x > 0,\ \ \alpha > 0 \end{cases} \tag{2}$$

Type III or Weibull Class (no tail):

$$\Psi_\alpha(x) = \begin{cases} exp(-(-x)^\alpha)\ ; & x \leq 0,\ \ \alpha > 0 \\ 1; & x > 0 \end{cases} \tag{3}$$

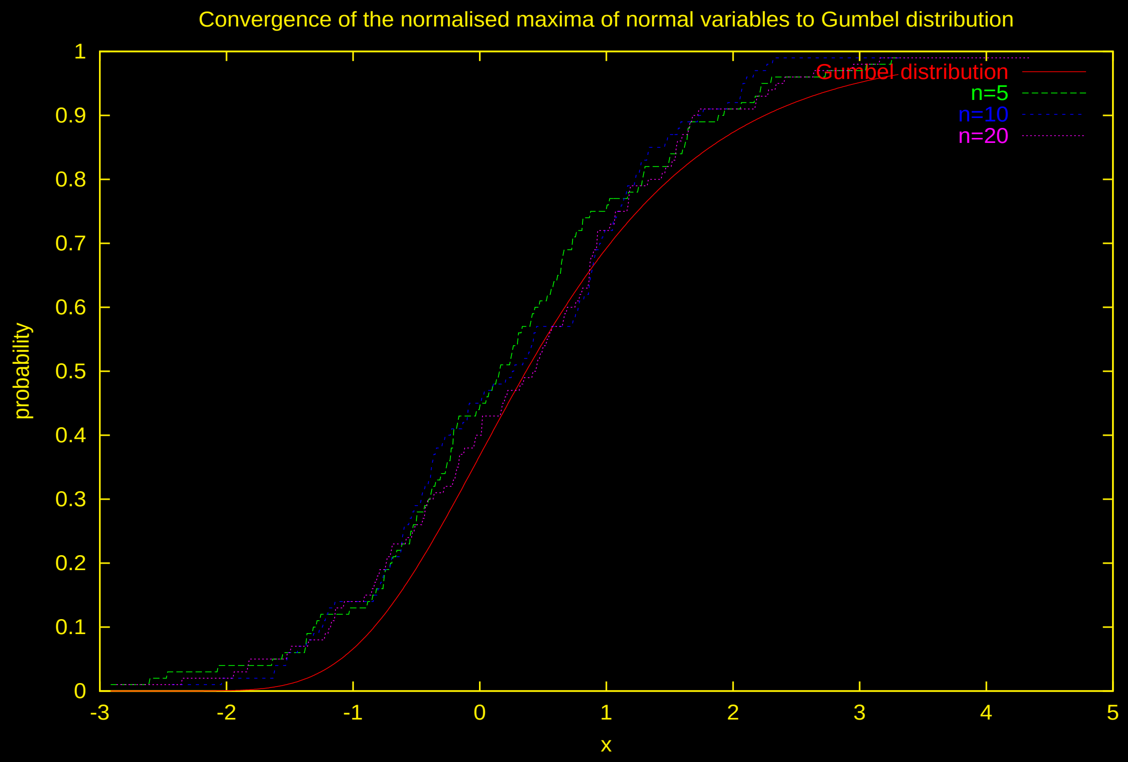


For the Fréchet and the Weibull distributions, $\alpha = 1$ is chosen.



For the Fréchet and the Weibull distributions, $\alpha = 1$ is chosen.

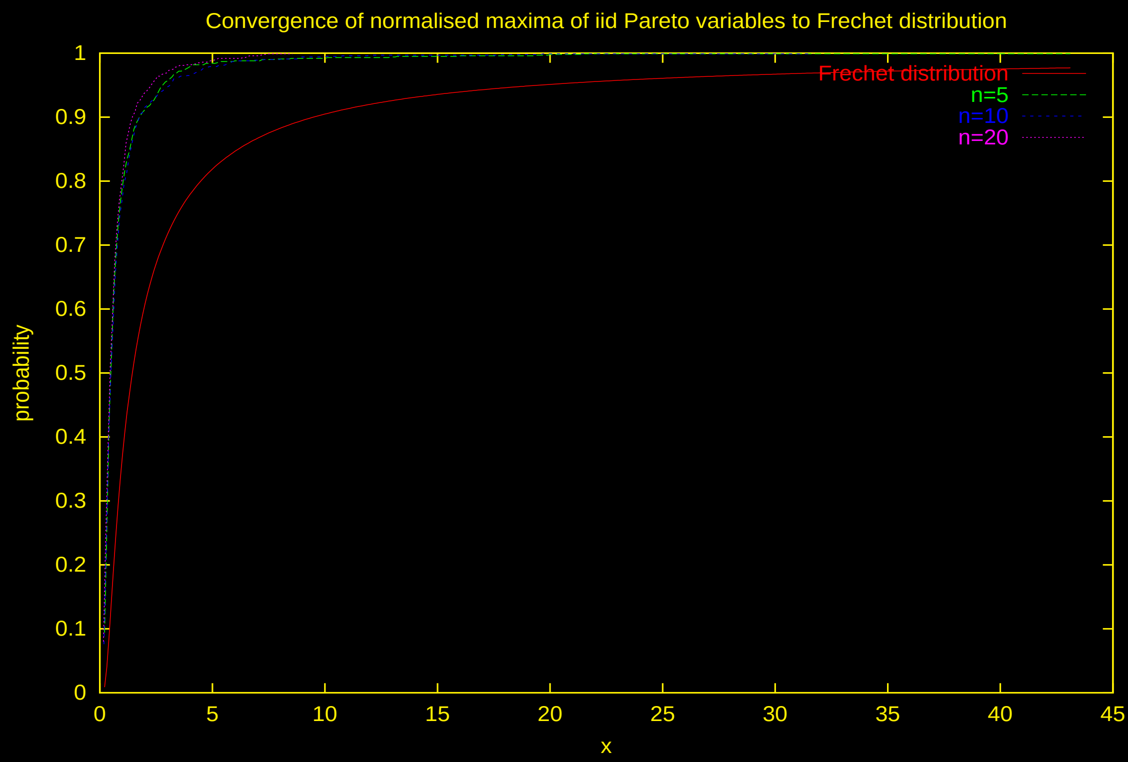
- Normal, Exponential, lognormal (and other monotone transformation of the normal distribution) $\in$ DA(Gumbel)
- Pareto, Cauchy, students-t, fat-tailed distributions $\in$ DA(Fréchet)
- Uniform, beta $\in$ DA(Weibull)
- Poisson, Geometric $\notin$ any domain of attraction



If Y_n is the maximum of n *iid* standard normal variables, then the distribution of $U_n = \frac{Y_n - a_n}{b_n}$ converges to the Gumbel distribution, as n increases, where

$$a_n = \sqrt{2 \ln n} - \frac{\ln 4\pi + \ln \ln n}{2(2 \ln n)^{1/2}}$$

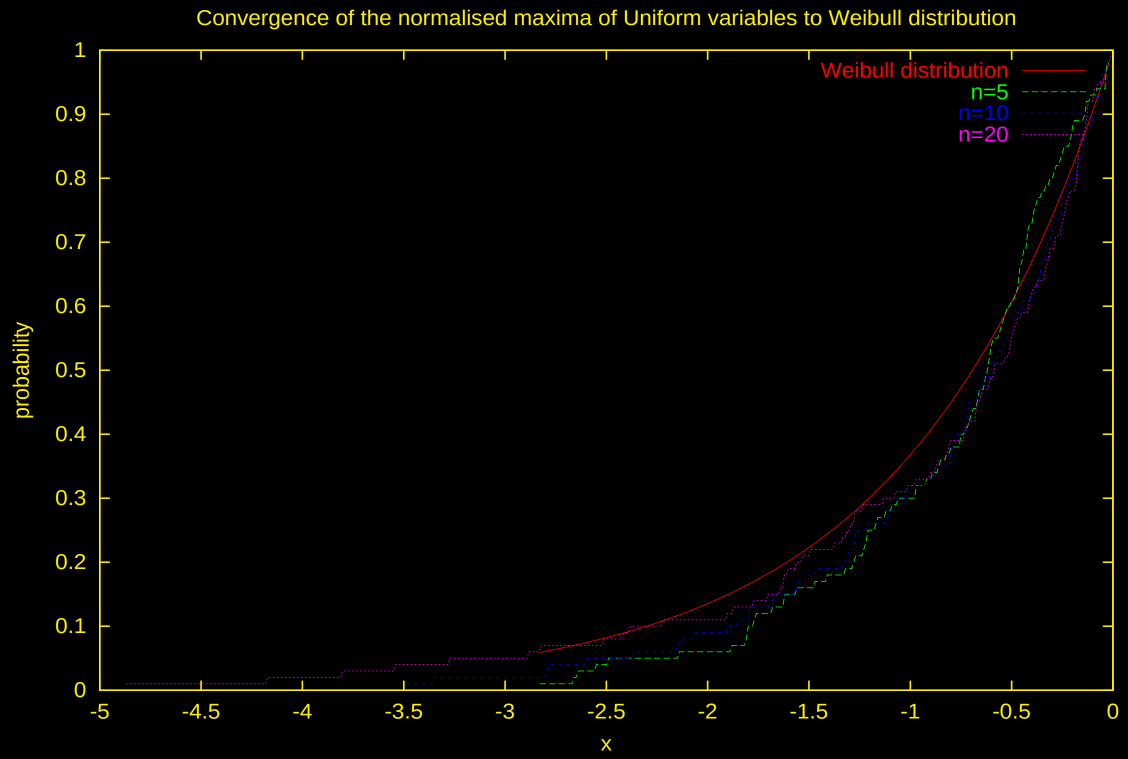
$$b_n = (2 \ln n)^{-1/2}$$



If Y_n is the maximum of n *iid* Pareto variables with distribution function $F(x) = 1 - Kx^{-a}$; where $a > 0$, $K > 0$, $x \geq K^{1/a}$, then the distribution of $U_n = \frac{Y_n - a_n}{b_n}$ converges to the Fréchet distribution, as n increases, where

$$a_n = 0 \text{ and } b_n = (Kn)^{1/a}$$

Here $K = 2$, $a = 1.5$ for the Pareto distribution and $\alpha = 1$ for the Fréchet distribution are assumed.



If Y_n is the maximum of n *iid* uniform variables, then the distribution of $U_n = \frac{Y_n - a_n}{b_n}$ converges to the Weibull distribution, as n increases, where

$$a_n = 1$$
$$b_n = n^{-1}$$

The Generalised Extreme Value Distribution

The three families of extreme value distributions can be nested into a single parametric representation (Jeskinson and Von Mises)

$$H_\xi(x) = exp\{-(1+\xi x)^{-\frac{1}{\xi}} \tag{4}$$

where

$$1 + \xi x > 0$$

$$\begin{aligned} x &> -\frac{1}{\xi} \text{ if } \xi > 0 \\ x &< \frac{1}{\xi} \text{ if } \xi < 0 \\ x &\in R \text{ if } \xi = 0 \end{aligned}$$

The parameter ξ, called the tail index, models the distribution tails.

- $\xi > 0 \longrightarrow$ Fréchet distribution
- $\xi < 0 \longrightarrow$ Weibull distribution
- $\xi = 0 \longrightarrow$ the Gumbel distribution.

Extremal Types theorem for stationary time series (Leadbetter et al. 1983)

Suppose that $\tilde{X}_1, \tilde{X}_2, ...\tilde{X}_n$ be a stationary time series with a marginal (unconditional) distribution function F. Also, let $\tilde{Y}_n = max(\tilde{X}_1, \tilde{X}_2, ...\tilde{X}_n)$.

Denote by $X_1, X_2, ..., X_n$ as associated *iid* series with the same marginal distribution F and let $Y_n = max(X_1, X_2, ..., X_n)$. Then,

$$\lim_{n\to\infty} \Pr\{(Y_n - a_n)/b_n \leq x\} = H(x),$$

for a non-degenerate $H(x)$ if and only if

$$\lim_{n\to\infty} \Pr\{(\tilde{Y}_n - a_n)/b_n \leq x\} = H^{\theta}(x),$$

where $H^{\theta}(x)$ is also non-degenerate, and $0 \leq \theta \leq 1$ is known as the "extremal index" of the stationary process.

Extremal index

- The extremal index models the relationship between the dependence structure and the extremal behaviour of a stationary process. It can be interpreted as the reciprocal of the mean cluster size.
- $\theta = 1$ for independent processes. The stronger the dependence, the lower the value of θ.
- Various methods of estimating θ are elaborated in text books on EVT (Chap 8, Embrechts et al.).

Pickands-Balkema-de Haan Theorem (1974)

Suppose that $X_1, X_2, ...X_n$ are n independent realisations of a random variable X with a distribution function $F(x)$. Let x_0 be the finite or infinite right endpoint of the distribution F. The distribution function of the excesses over certain high threshold u by

$$\Phi_u(x) = \Pr\{X - u \leq x | X > u\} = \frac{F(x+u) - F(u)}{1 - F(u)}$$

for $0 \leq x < x_0 - u$.

If $F \in DA(H_\xi)$ then $\exists$ a positive measurable function $\sigma(u)$ such that

$$\lim_{u \to x_0} \sup_{0 \leq x < x_0 - u} |\Phi_u(x) - G_{\xi,\sigma(u)}(x)| = 0$$

and vice versa, where $G_{\xi,\sigma(u)}(x)$ denote the Generalised Pareto distribution.

The Generalised Pareto Distribution (GPD)

$$G_{\xi,\sigma}(x) = \begin{cases} 1-(1+\xi x/\sigma)^{-1/\xi} & if \xi \neq 0 \\ 1-exp(-x/\sigma) & if \xi = 0 \end{cases} \quad (5)$$

where $\sigma > 0$, and the support of x is $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\sigma/\xi$ when $\xi < 0$.

- $\xi > 0 \rightarrow G_{\xi,\sigma}$ is a reparameterised version of the ordinary Pareto distribution.
- $\xi = 0 \rightarrow G_{\xi,\sigma}$ is exponential
- $\xi < 0 \rightarrow G_{\xi,\sigma}$ is Pareto type II distribution

EVT provides a robust framework for measuring financial risk

- EVT does not require *a priori* assumption about the return distribution while the conventional parametric approaches need the assumption of normal distribution.

- EVT based methods inherently incorporate separate estimation of both the tails, thereby allowing for modelling possible asymmetry.

Two broad categories of EVT models

1. The Block Maxima (BM) model (A GEV approach)

   Models for the largest (smallest) observations collected from non-overlapping blocks (samples) of size n from the data.

   - Estimation of stress loss (McNeil 1999)
   - Estimation of VaR (Longin 2000)

2. The Peaks-Over-Threshold (POT) model (A GPD approach)

   Models for large observations exceeding certain threshold.

   - Estimation of VaR (Danielsson and de Vries 1997; McNeil and Frey 1999)
   - Estimation of Expected shortfall (McNeil and Frey 1999; Longin 2000)

The BM approach

Estimation of Stress Loss (McNeil 1999)

●$X_1, X_2, ..X_T$ are daily (negative) logarithmic returns.
● Divide the data into k non-overlapping blocks of same size n.
●$Y_n^j = min\{(X_1^j, X_2^j, ..., X_n^j\}$, the minimum of the n observations in the block j.
● Use MLE to fit the GEVD H_ξ to the block minima $Y_n^1, Y_n^2, ...Y_n^k$.
● The p^{th}quantile of the fitted distribution is known as the p^{th} "stress loss".

For example, if $n = 25$ (a month), then $H_{\hat{\xi}}^{-1}(0.05)$ gives the magnitude of the daily loss level which can be expected to reach once in 20 months.

Estimation of VaR (Longin 2000)

• Use MLE to estimate the parameters a_n, b_n and ξ of the asymptotic distribution of the minimal returns. A Goodness-of-fit test can be carried out to test for the statistical validity of the estimates.

• If the data is non-iid and stationary (which is generally the case) then the extremal index θ needs to be estimated as well.

• If p^* is the probability of the minimal daily returns (over n days) exceeding certain threshold and p the corresponding probability for the underlying returns, then,

$$\begin{aligned} p^* &= \Pr\{min(X_1, X_2, ...X_n) \leq z\} \\ &= \Pr\{X_1 \leq z, X_2 \leq z, ..., X_n \leq z\} \\ &= p^n \end{aligned} \tag{6}$$

• In case of non-iid stationary series, (6) takes the form

$$p^* = (p^n)^\theta$$

Estimation of VaR (contd.)

• The VaR formula, expressed in terms of the the distribution of the minimal returns, can be obtained as

$$\begin{aligned} p^* &= 1 - H(VaR) \\ &= exp\left[- \left(1 + \xi \left(\frac{VaR - b_n}{a_n} \right) \right)^{1/\xi} \right] \end{aligned}$$

leading to

$$VaR = -b_n + \frac{a_n}{\xi} \left[1 - (-\ln(p^*))^{\xi} \right]$$

The POT approach

The tail of the return distribution over certain threshold is estimated. Two different approaches exist.

- The semi-parametric models based on the Hill estimator and its relatives (Danielsson and de Vries 1997).
- Fully parametric models based on the Generalised Pareto Distribution (McNeil and Frey 1999).
- The Hill estimator based method is applicable only for fat-tailed distributions.
- The GPD version provides simple parameteric formulae for measuring extreme risk which can be supplemented with estimates of standard error using MLE.

The GPD approach

Using the Pickands-Balkema-de Haan theorem and the GPD approximation for the distribution of the excesses over a threshold u, we get

$$F(x) \;=\; (1 - F(u))G_{\xi,\sigma}(x-u) + F(u)$$

which gives the tail estimator

$$\hat{F}(x) \;=\; 1 - \frac{N_u}{N}\left(1 + \hat{\xi}\frac{x-u}{\hat{\sigma}}\right)^{-1/\xi}$$

For a given probability $p > F(u)$, the VaR is estimated by inverting the tail estimation formula

$$\hat{V}aR_p = u + \hat{\sigma}/\hat{\xi}\left(\left(N/N_u(1-p)\right)^{-\xi} - 1\right)$$

Expected shortfall

$$ES_p = VaR_p + E\left[X - VaR_p | X > VaR_p\right]$$

A nice stability property of the excess distribution above threshold u is that if a higher threshold is taken then the distribution of the excess above the higher threshold is also GPD with the same shape parameter but a different scale parameter.

$$F_{VaR_p}(y) = G_{\xi, \sigma + \xi(VaR_p - u)}(y)$$

Dynamic VaR (McNeil 1999; McNeil and Frey 1999)

Let $\{X_t\}$ is a strictly stationary time series whose dynamics are given by

$$X_t = \mu_t + \sigma_t Z_t \tag{7}$$

where μ_t is the mean process and σ_t the volatility dynamics of X_t, and,

$$Z_t \ \sim f_Z(z)$$

where $f_Z(z)$ is white noise.

The p^{th} quantile of the distribution of X_t at time t can be obtained by using that of Z_t, as,

$$x_p^t = \mu_t + \sigma_t z_p \tag{8}$$

where z_p is the p^{th} quantile on the distribution of Z_t, which, by assumption, is *iid*.

Dynamic VaR and ES (contd.)

1. Fit a time series model to the return series without making any assumption about $f_Z(z)$ and using a pseudo-maximum likelihood (PML) approach. Estimate μ_t and σ_t from the fitted model and extract the residuals Z_ts.
2. Consider the residuals to be the realisations of a strict white noise process and use extreme value theory (EVT) to model the tail of $F_Z(z)$. Use this EVT model to estimate z_p.
3. VaR_p^t and ES_p^t for the observed returns are

$$
\begin{aligned}
VaR_p^t &= \mu_t + \sigma_t VaR(z)_p \\
ES_p^t &= \mu_t + \sigma_t ES(z)_p
\end{aligned}
$$

PML

Gourieroux et al. (1984) establish that a distribution belonging to linear and quadratic exponential family (eg. the normal distribution) can generate consistent and asymptotically normally distributed estimators for the first two moments of the true distribution, regardless of the exact form of the true underlying distribution.

A POT analysis of the Nifty returns: An illustration(Apologies for showing you an old work)

- Nifty data:

  Estimation window: 3 July 1990 – 7 May 1996 (1250 daily observations)

  Forecast period: 8 May 1996 – 15 March 2002 (1446 daily forecasts).
- Time series model: An AR(1)-GARCH(1,1) model.
- Estimation of 95% and 99% VaR for a long and a short position in Nifty portfolio.

Estimation of AR(1)-GARCH(1,1) model

Parameter	Estimates	SE	Confidence bounds
The mean equation:			
Constant	-0.036	0.040	(-0.114, 0.043)
AR(1)	0.225	0.029	(0.167,0.282)
The variance equation:			
Constant	0.039	0.015	(0.010, 0.067)
ARCH(1)	0.101	0.016	(0.069, 0.132)
GARCH(1,1)	0.803	0.015	(0.863, 0.923)

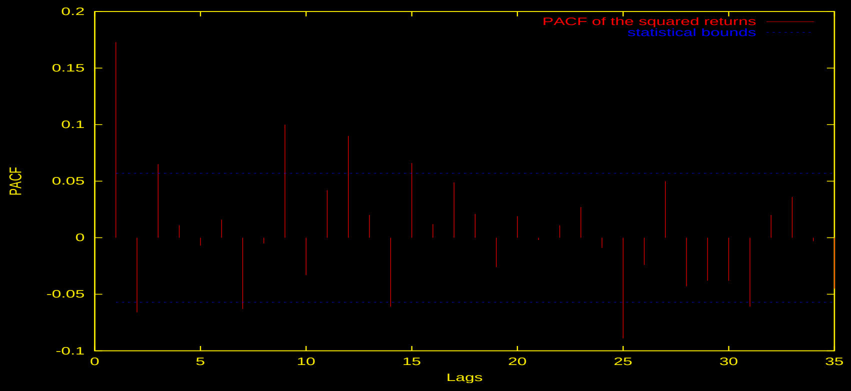
Descriptive statistics: Returns and standardised residuals

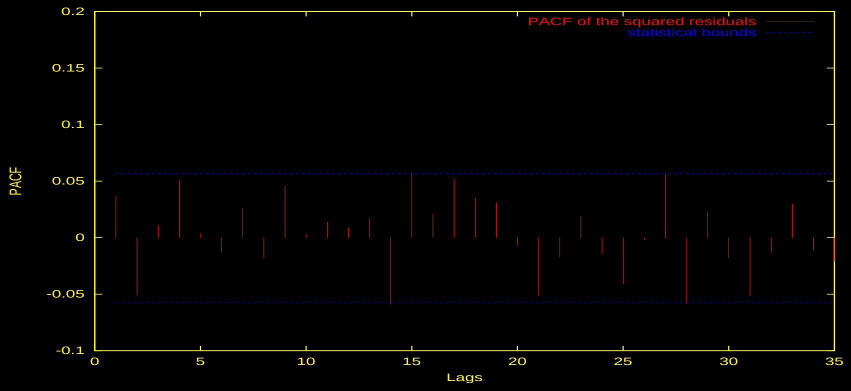
	Mean	Variance	Skewness	Kurtosis
Returns	0.1094	4.1675	0.0758	8.6126
SR	0.0268	1.0033	0.2223	4.3623

Tests for skewness, kurtosis and auto correlation

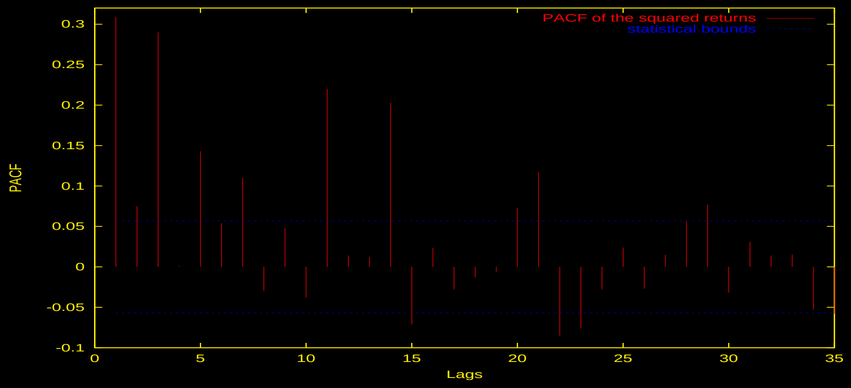
	statistic	p-value
Panel A: The returns series (r_t)		
skewness	15.7839	0.000
kurtosis	292.3220	0.000
H.C. Ljung-Box	57.2202	0.01
Panel B: The residual series (z_t)		
skewness	46.2049	0.000
kurtosis	70.78355	0.000
H.C. Ljung-Box	47.7039	0.074

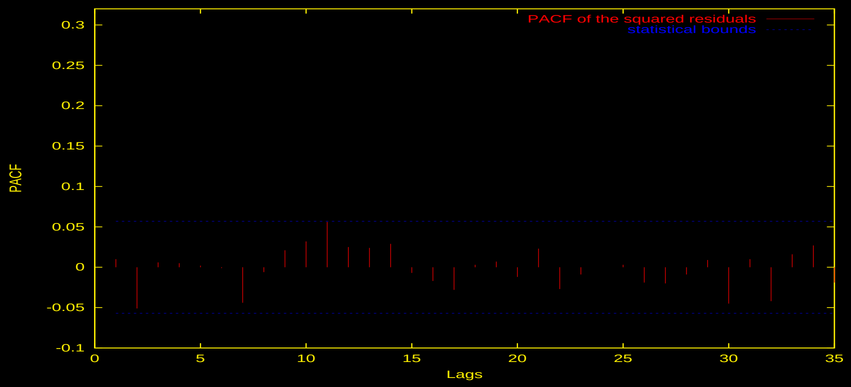
Correlograms of the returns and the standard residuals





Correlograms of the squared returns and the squared standard residuals





Modeling Peaks-over-threshold

- How to choose the threshold?

  - Mean-excess plot (McNeil and Frey 1999; McNeil 1996).
  - Arbitrary threshold level (Gavin 2000).
  - $1.65 \times \sigma$ (Neftci 2000).

Maximum likelihood estimates of the GPD parameters

	N_u	u	F_u	$\hat{\xi}$	$\hat{\sigma}$
left tail	50	-1.6493	0.9599	0.2027	0.4099
				(0.2095)	(0.1030)
Right tail	66	1.6494	0.9471	-0.0064	0.6460
				(0.2736)	(0.1950)

Figures in parenthesis indicate standard error

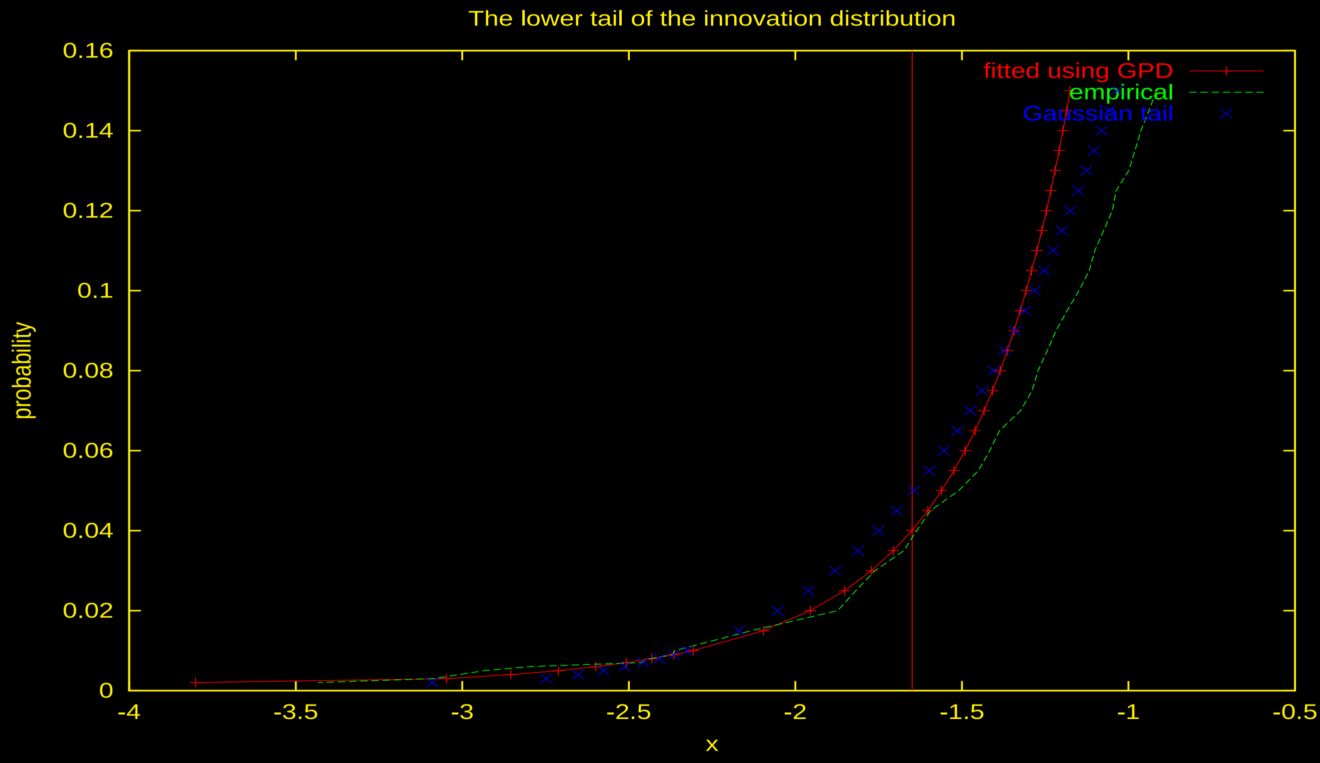
Quantiles on the lower tail of the i.i.d. residuals

p	EVT	Empirical	Normal
0.06	-1.4907	-1.4164	-1.5548
0.05	-1.5608	-1.5092	-1.6449
0.04	-1.6503	-1.6358	-1.7507
0.03	-1.7717	-1.7618	-1.8808
0.02	-1.9555	-1.8128	-2.0537
0.01	-2.3067	-2.3615	-2.3263
0.009	-2.3646	-2.3704	-2.3656
0.008	-2.4307	-2.4348	-2.4082
0.007	-2.5076	-2.4637	-2.4573
0.006	-2.5991	-2.7965	-2.5121
0.005	-2.7109	-2.9345	-2.5758
0.004	-2.8526	-3.0110	-2.6521
0.003	-3.0474	-3.0858	-2.7478
0.002	-3.3404	-3.2280	-2.8782
0.001	-3.8005	-3.4330	-3.0902

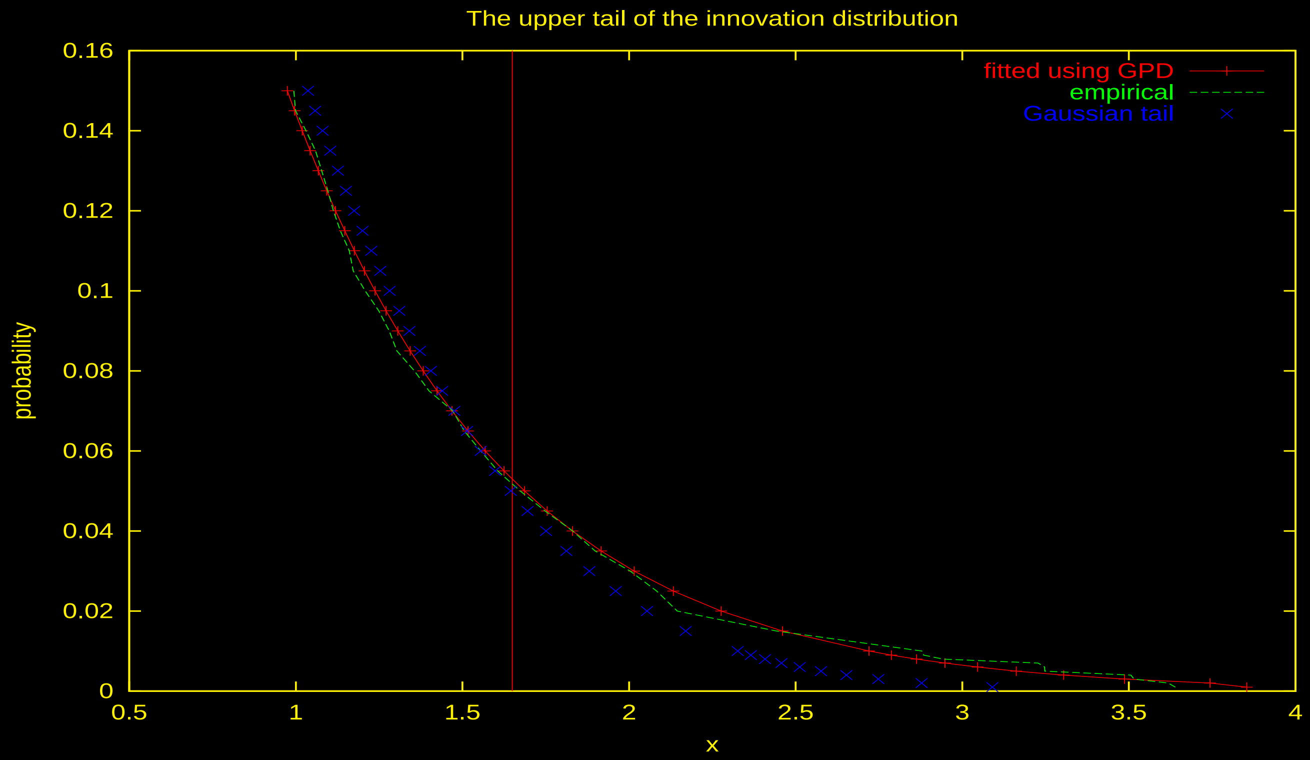
Quantiles on the upper tail of the i.i.d. residuals

p	EVT	Empirical	Normal
0.94	1.5684	1.5550	1.5548
0.95	1.6862	1.6728	1.6449
0.96	1.8302	1.8310	1.7507
0.97	2.0155	2.0034	1.8808
0.98	2.2761	2.1447	2.0537
0.99	2.7202	2.8785	2.3263
0.991	2.7875	2.8839	2.3656
0.992	2.8627	2.9422	2.4082
0.993	2.9479	3.2272	2.4373
0.994	3.0461	3.2469	2.5121
0.995	3.1622	3.2476	2.5758
0.996	3.3041	3.5057	2.6521
0.997	3.4867	3.5153	2.7478
0.998	3.7436	3.6180	2.8782
0.999	3.8535	3.6394	3.0902

Lower tail of the Nifty innovations



Upper tail of the Nifty innovations



Testing the discrepancy between estimated and empirical tail quantiles

$F(x)$, $G(x)$ and $\phi(x)$ denote the empirical, the estimated and the normal distribution functions.

Test 1:

$$H_0 : F(x) = G(x)$$

against the alternative hypothesis

$$H_1 : F(x) \neq G(x)$$

Two-sided Kolmogorov-Smirnov test to test this hypothesis.

Test 2:

$$H_0' : F(x) = \phi(x)$$

and the alternative hypothesis is

$$H_1' : F(x) > \phi(x)$$

A one-sided Kolmogorov-Smirnov test for this.

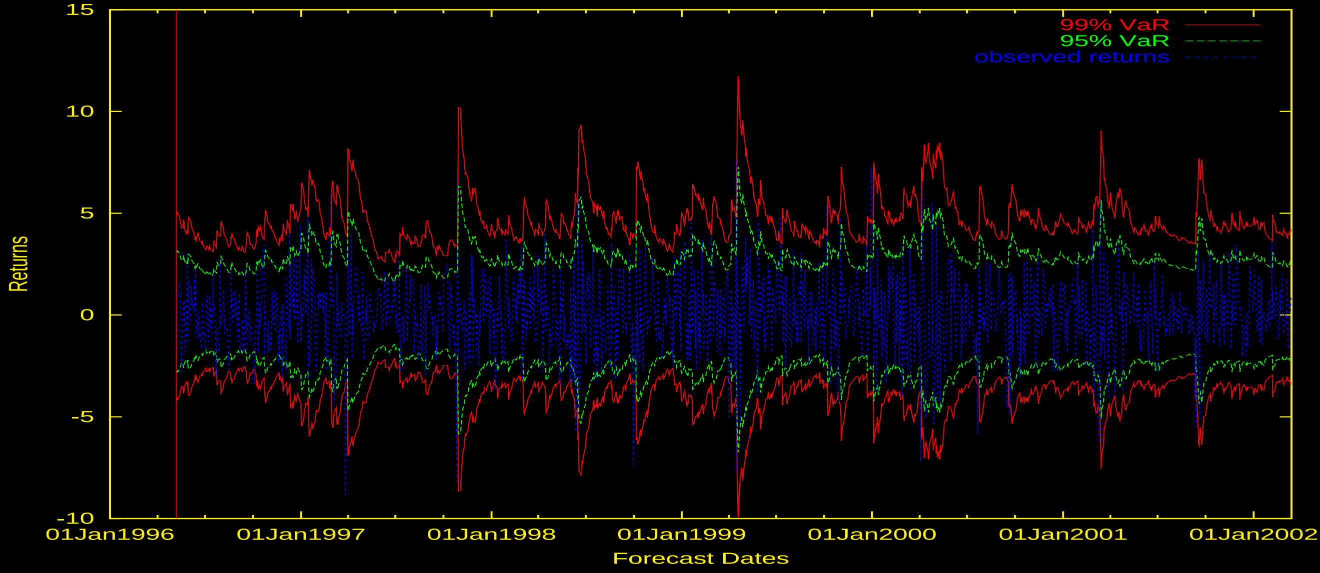
Results of the Kolmogorov-Smirnov tests of discrepancy

	Upper tail	Lower tail
D	0.2794	0.1574
D^+	0.8536^*	0.3589

Critical value of D at 0.05 level of significance = 0.467

Critical value of D^+ at 0.05 level of significance = 0.400

95% and 99% VaR measures estimated with the POT model



Statistical precision of the VaR measures

- A “good” VaR model should generate the pre-specified failure probability p, conditionally at each point of time (conditional efficiency) [Christoffersen, 1998].

- Given r_t and the *ex-ante* VaR forecasts, the following indicator variable may be defined

$$I_t = \begin{cases} 1 \text{ if} & r_t < v_t \\ 0 & \text{otherwise} \end{cases}$$

- The VaR forecasts are said to be efficient if they display “correct conditional coverage”, i.e., if

$$E\left[I_{t|t-1}\right] = p \, \forall \, t$$

- This is equivalent to saying that the $\{I_t\}$ series is *iid* with mean p.

Tests for correct conditional coverage

1. Christoffersen's test (1998)

2. Regression-based tests (Christoffersen and Diebold 2000; Clements and Taylor 2000)

The regression test

Define the regression

$$
\begin{aligned}
I_t &= \alpha_0 + \sum_{s=1}^{S} \alpha_s I_{t-s} + \sum_{s=1}^{S-1} \mu_s D_{s,t} + \epsilon_t \\
t &= S+1, S+2, ..., T \\
D_{s,t} & \quad \text{are explanatory variables.}
\end{aligned}
$$

Conditional efficiency of the I_t process can be tested by testing the joint hypothesis:

$$H : \Phi = 0, \alpha_0 = p \tag{9}$$

where

$$\Phi = [\alpha_1, ...\alpha_S, \mu_1, ..., \mu_S]'$$

•The hypothesis (9) can be tested by using an F-statistic in the usual OLS framework.

Results of the test of “conditional coverage” for the POT VaR

	$\hat{p}$ (p-value)	F-stat (p-value)
Panel A: 95% VaR estimation		
Short Nifty	0.0500 (0.4990)	1.0633 (0.3878)
Long Nifty	0.0452 (0.6778)	0.6155 (0.8016)
99% VaR estimation		
Short Nifty	0.0056 (0.8359)	0.7701 (0.6579)
Long Nifty	0.0087 (0.6397)	0.29651 (0.98208)

Figures in parentheses indicate p-values

Summary

- The innovation distribution of Nifty returns is asymmetric.
- There is significant “tail–thickness” on the right tail of the Nifty innovations while the left tail behaves like the left tail of the standard normal distribution.
- The VaR forecasts generated by the POT approach displays the property of “correct conditional coverage”.

Thank you