Statistics of heteroscedastic extremes

John Einmahl[1] Laurens de Haan[2,3] Chen Zhou[4,3]

[1]Tilburg University

[2]University of Lisbon

[3]Erasmus University Rotterdam

[4]De Nederlandsche Bank

May 22, 2013

Motivation

"We are going through a financial crisis more severe and unpredictable than any in our lifetimes."

– Henry M. Paulson, Nov 18, 2008

- Is that true?
  - Are financial crises nowadays more severe than those in the past?
- Challenge to statistics
  - Analyze tail events
  - Account for potential distributional changes
- Do extreme value statistics work here?
  - Yes: tools for tails
  - No: usually assuming i.i.d.

Homoscedastic extremes

- Classic extreme value statistics
  - Observations are assumed to be i.i.d.
  - Estimating tail properties: e.g. extreme value index
  - Inference on tail events: e.g. VaR, tail probability
- Beyond i.i.d.
  - Account for serial dependence
  - Nevertheless, assuming stationary distribution
- So far, distributional changes are assumed away
- To justify "we have 'more severe' crises in certain period"
  - Must abolish "identical distribution"
    - Given crisis magnitude, more frequent
    - Given crisis frequency, more severe
  - Must keep some common properties for statistical inference

Model setup

- Consider observations $X_1, \cdots, X_n$
- Drawn from different distributions $F_{n,1}, \cdots, F_{n,n}$
- Common right endpoint x^*
- Tail comparability

$$\lim_{x \to x^*} \frac{1 - F_{n,i}(x)}{1 - F(x)} = c\left(\frac{i}{n}\right)$$

  - Comparable tail: common distribution function $F \in \mathcal{D}_\gamma$
  - Heteroscedastic extremes: scaling function $c(s)$ on $[0, 1]$
  - Uniformly for all n and all $1 \leq i \leq n$.
- Identification condition: c continuous and

$$\int_0^1 c(s)ds = 1$$

Model evaluation

- Advantages
  - The model allows non-identical distribution
  - The model only assumes heteroscedasticity in extremes
    - Typical extreme value setup
  - Non-parametric setup on the scaling function
    - Flexible for further modeling the trend c
- Consequence: If $F \in \mathcal{D}_\gamma$, then all $F_{n,i}$ has the same tail index
  - Do not allow variation in extreme value index
- Such a feature is not necessarily bad
  - Otherwise, no common parameter to estimate
- We will nevertheless test the model setup

The purpose of the paper

General purpose: provide a set of tools on extreme value statistics with non-identically distributed observations

- Assume the model
  - Estimate the extreme value index of F, γ
  - Estimate the scaling function $c(s)$
  - Testing hypothesis $c(s) = c_0(s)$ for a given c_0
    - Rejecting the null that $c(s) = 1$ confirms the statement that "in some period, extreme events are more severe than other".
- Testing the model
  - Testing the null hypothesis of constant γ
  - In the presence of scale changes
- Estimation of high quantile at certain time point
  - Quantify how different extreme events are in some period

Estimating the scaling function

- We first estimate $C(s) = \int_0^s c(u)du$
- A Peak-Over-Threshold (POT) idea
  - Threshold: $X_{n,n-k}$ ($(k+1)-$th highest observation among all)
  - It is not an order statistic (different distributions)
  - It nevertheless works as an order statistics from F
  - Then count the frequency of "exceeding" in the first "s fraction"
- Estimator: $\hat{C}(s) = \frac{1}{k}\sum_{i=1}^{[ns]} 1_{\{X_i > X_{n,n-k}\}}$
- Choice of k
  - As in usual extreme value statistics

$$\lim_{n\to+\infty} k(n) = +\infty, \quad \lim_{n\to+\infty} \frac{k}{n} = 0$$

  - Extra conditions for proving asymptotic normality

Theoretical property of estimators

- Conditions
  - Quantifying speed of convergence: $\frac{\frac{1-F_{n,i}(x)}{1-F(x)} - c\left(\frac{i}{n}\right)}{A_1(x)} = O(1)$
  - Extra conditions on k:
    $\sqrt{k}A_1(n/k) \to 0$ and $\sqrt{k}\sup_{|u-v|\leq 1/n}|c(u) - c(v)| \to 0$
- Conclusion (under a Skorokhod construction)

$$\sup_{0\leq s\leq 1}\left|\sqrt{k}(\hat{C}(s) - C(s)) - B(C(s))\right| \to 0 \text{ a.s.}$$

  - $B(s)$ is a standard Brownian bridge.
- A consistent estimator of the $c(s)$ function
  - A uniform kernel approach
  - A bandwidth $h > 0$ such that $h \to 0$ and $kh \to \infty$ as $n \to \infty$
  - Estimator: $\hat{c}(s) = \frac{\hat{C}(\min(s+h,1)) - \hat{C}(\max(s-h,0))}{\min(s+h,1) - \max(s-h,0)}$

Detecting heteroscedasticity in extremes

- Testing the null $c(s) = c_0(s)$ or $C(s) = C_0(s)$
  - Example: $c_0(s) = 1$ or $C_0(s) = s$: no trend
  - Economic interpretation
- A Kolmogorov-Smirnov type test
  - Test statistic: $T := \sup_{0 \le s \le 1} \left| \hat{C}(s) - C_0(s) \right|$
  - Limit behavior:

$$\sqrt{k}\, T \xrightarrow{d} \sup_{0 \le s \le 1} |B(C_0(s))|$$

- Other potential test
  - The null $C(s_0) = s_0$
  - The crisis frequency before s_0 is the same as that after s_0
  - Break point (A Chow-type test)

Extreme value statistics under heteroscedastic extremes

- So far, we did not use the domain of attraction condition
  - The estimation of trend does not depend on F
- Next, we assume that $F \in \mathcal{D}_\gamma$ for $\gamma > 0$.
  - γ is the common EVI shared by all $F_{n,i}$
- Estimation of γ: the Hill estimator
- Under the same Skorokhod construction as in estimating $C(s)$,

$$\sqrt{k}(\hat{\gamma}_H - \gamma) \to \gamma N_0 \text{ a.s.},$$

  where N_0 follows standard normal distribution
- N_0 and $B(C(s))$ are independent
  - The estimation of the tail property and the scaling function are independent

Testing the model

- The null hypothesis: our model
  - γ is constant across the distributions
  - Scales may vary across observations
- The alternative: γ variation
- Comparing with other tests in literature
  - Quintos et al. (2001) tested constant γ, by taking the null hypothesis that observations are i.i.d.
    - They require constant scale under the null hypothesis
  - Data violate that null, but following our model would be rejected there
  - We test constant γ in the presence of scale variation

Constructing the test

- A recursive idea
  - Estimate γ based on the first $[ns]$ observations: $\hat{\gamma}_s$, $s \in [\delta, 1]$
  - The choice of k matches the scaling effect $k_s = kC(s)$
  - Use the estimated $C(s)$: $k_s = k\hat{C}(s)$
  - Compare that with $\hat{\gamma}_1$ when using all observations
- Formally,

$$\sqrt{k}(\hat{\gamma}_s - \gamma) \to \gamma \frac{\tilde{W}(C(s))}{C(s)},$$

  where $\tilde{W}$ is a standard Brownian motion.
- Test statistics: $T_\delta := \sup_{\delta \le s \le 1} |\hat{\gamma}_s - \hat{\gamma}_1|$.

$$\sqrt{k}\, T_\delta \xrightarrow{d} \gamma \sup_{\delta \le s \le 1} \left| \frac{\tilde{W}(C(s))}{C(s)} - \tilde{W}(1) \right|.$$

  - Limit process: use estimated $\hat{C}(s)$ and simulation

Potential problems and solutions

- The recursive test can identify γ change
  - Any increase in γ after $[n\delta]$-th observation
- It cannot identify
  - Changes in the first δ fraction
  - Decreasing γ
- Two solutions:
  - Combining with a reverse recursive test
  - A circle idea
- Choice of δ
  - We choose $\delta = 1/2$ to guarantee the number of observations
  - It is not a problem after adopting the circle idea

Simulations

- Simulated observations
  - Model 1: i.i.d. standard Fréchet
  - Model 2: multiplied with $c(s) = 0.2 + 1.6 * s$
  - Model 3: adjusted to the power $d(s) = 0.7 + 0.6 * s$
- Sample size $n = 5,000$ (similar to that in application)
- Number of samples $m = 100$
- Report: the number rejections under 5%(10%) confidence level

	Testing constant γ	Testing constant scale
Model 1	4 (8)	4 (10)
Model 2	3 (12)	100 (100)
Model 3	50 (62)	–

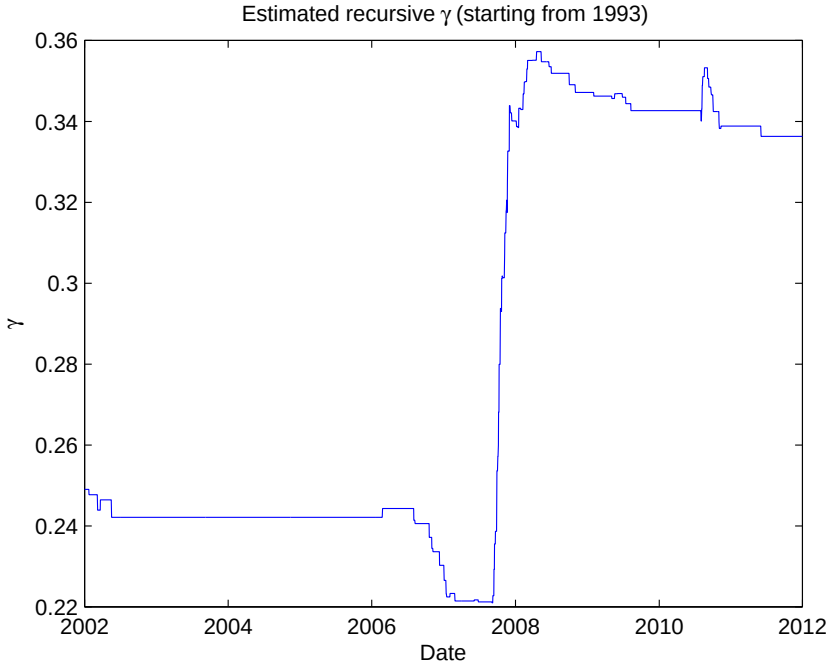
Application

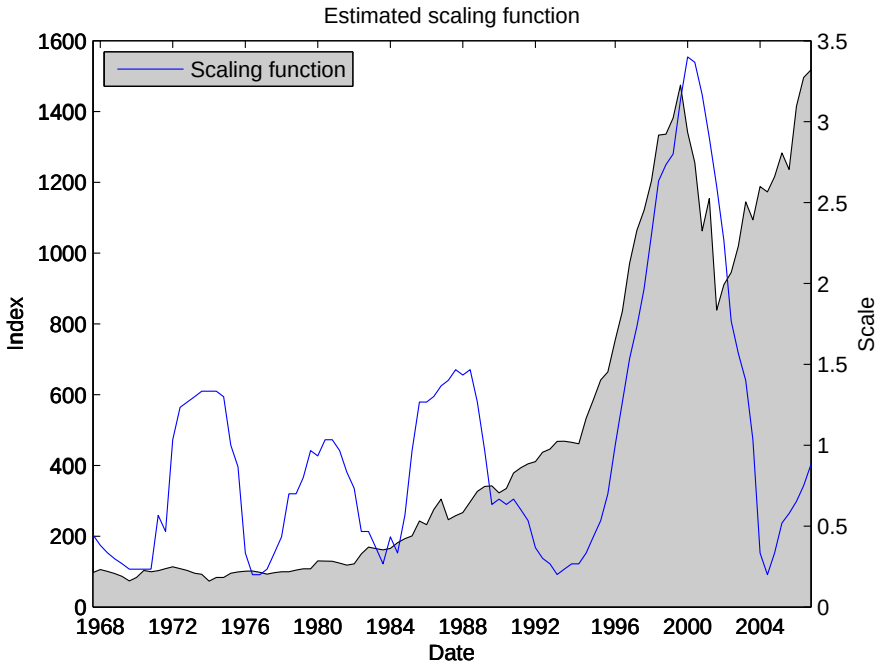
- Data: S&P500 daily returns (1968–2012, 11,327 obs)
- Sub-samples
  - Sample 1: 1968-1987 (5,025 obs)
  - Sample 2: 1988-2007 (5,043 obs)
  - Sample 3: 1993-2012 (5,037 obs)

	k	Testing constant γ	Testing constant scale
Full Sample	320	0.11	0
Sample 1	130	0.09	0
Sample 2	130	0.16	0
Sample 1&2	300	0.91	0
Sample 3	130	0.002	–

- Next, we plot
  - The estimated $\hat{\gamma}_s$ in Sample 3
  - The estimated $c(s)$ in Sample 1 & 2

The scale function over time





A by-product: VaR at each time point

- Different distributions lead to different high quantiles
- We estimate high quantiles at each time point
  - Economic significance of the different scaling
    - Given crisis frequency, the difference in crisis magnitude
  - Forecasting high quantile in the next time period
    - Exploiting the continuity of the c function
- A two-step approach
  - Estimate the quantile of the common distribution F

$$\widehat{VaR}(p) = X_{n,n-k} \cdot \left(\frac{k}{np}\right)^{\hat{\gamma}_H}.$$

  - Use the relation that

$$VaR_i(p) = VaR\left(p/c\left(\frac{i}{n}\right)\right) \sim VaR(p)c\left(\frac{i}{n}\right)^{\gamma}$$

  - The consistent estimator

$$\widehat{VaR}_i(p) = X_{n,n-k} \cdot \left(\frac{k\hat{c}\left(\frac{i}{n}\right)}{np}\right)^{\hat{\gamma}_H}.$$

Conclusion

- We can handle extreme value statistics when observations are draw from different distributions
- We identify whether heteroscedastic extremes are due to the variation of γ or scale
- If the changes are in the scale, we can quantify that variation
  - in terms of frequency, given magnitude
  - in terms of magnitude, given frequency
- Omitted in the presentation
  - The proofs: based on the Sequential Tail Empirical Process (STEP).
  - A useful tool to that can be applied to other problems in extreme value statistics when having heteroscedastic extremes.

It is just a first STEP towards non-stationarity!