Statistics of heteroscedastic extremes: from skedasis to variations in the extreme value indices

Chen Zhou[1,2,3]

[1]De Nederlandsche Bank

[2]Erasmus University Rotterdam

[3]Tinbergen Institute

March 24, 2016

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 or frequent?
- 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.

Classic extreme value theory

- Modeling regularities in tails: X follows the distribution F

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = g(x)$$

- Consequences
  - Potential limits $g(x) = x^{-1/\gamma}$
  - In conditional probability

$$\lim_{t\to\infty} \Pr\left(\frac{X}{t} \le x | X > t\right) = 1 - x^{-1/\gamma}.$$

  - In quantile function $U = (1/(1-F))^{\leftarrow}$

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^{\gamma}$$

- Potential for application: extrapolation for high quantiles
  For some low p, even $p = p_n$ such that $np_n \to 0$

$$\frac{U(1/p)}{U(n/k)} \approx \left(\frac{k}{np}\right)^{\gamma} \Rightarrow \hat{U}(1/p) = X_{n,n-k}\left(\frac{k}{np}\right)^{\hat{\gamma}}.$$

Classic extreme value statistics

- Estimating tail properties: e.g. extreme value index
  - Idea: fitting excess ratios to Pareto distribution
  - Hill estimator: for $k \to \infty$ and $k/n \to 0$ as $n \to \infty$

$$\hat{\gamma}_H = \frac{1}{k} \sum_{i=1}^{k} \log X_{n,n-i+1} - \log X_{n,n-k}$$

- Asymptotic property
  - Requires some second order condition

$$\lim_{t \to \infty} \frac{\frac{U(tx)}{U(t)} - x^{\gamma}}{A(t)} = H(x)$$

  - The choice of k: $\lim_{n \to \infty} \sqrt{k} A(n/k) = \lambda$
  - Speed of convergence $\sqrt{k}$

$$\sqrt{k}(\hat{\gamma}_H - \gamma) \xrightarrow{d} N(bias, \gamma^2)$$

- Inference on tail events: e.g. VaR, tail probability

Beyond homoscedastic extremes

- Classic extreme value statistics assumes i.i.d. observations.
- Literature that goes beyond i.i.d.
  - Account for serial dependence
  - Nevertheless, assuming stationary distribution
- To justify "we have 'more severe' crises in certain period"
  - Must abolish "identical distribution"
  - Must keep some common properties for statistical inference
- Modeling (parametrical) distributional changes in extremes
  - Parametric models on block maxima
  - On the shift/scale of GEV
  - Some parametric approach on GPD

This talk

- Abolishing "identical distribution"
  - Consider observations $X_1, \cdots, X_n$
  - Drawn from different distributions $F_{n,1}, \cdots, F_{n,n}$
- Further assumptions
  - Some "continuity" in $F_{n,i}$ with respect to i
  - No parametric trend!
- Two recent works
  - Consider "tail comparability": Einmahl, J., de Haan, L. and Zhou, C. (2015), JRSS-B
    - Common right endpoint x^*
    - Tail comparability

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

  - Abolish "tail comparability": de Haan, L. and Zhou, C. (ongoing)

Model setup in Einmahl et al. (2015)

- 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: skedasis function $c(s)$ on $[0, 1]$
  - Uniformly for all n and all $1 \leq i \leq n$.
- Identification condition: c continuous and

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

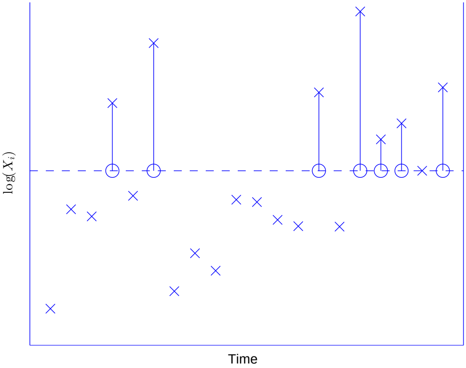
- Advantages: only assumes heteroscedasticity in extremes
- Non-parametric setup on the skedasis function
- Consequence: If $F \in \mathcal{D}_\gamma$, then all $F_{n,i}$ has the same tail index
  - Do not allow variation in extreme value index
- We will nevertheless test the model setup

The purpose of the paper

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

- Under the model setup
  - Estimate the extreme value index of F, γ
  - Estimate the skedasis 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 heteroscedasticity
- Estimation of high quantile at certain time point
  - Quantify how different extreme events are in some period

The idea on estimation



- Unified threshold using a high "order statistic"
- Estimating $c(s)$ – the occurrence of POT
- Estimating γ – the magnitude of POT

Estimation

- Estimating $C(s) = \int_0^s c(u)du$
  - Threshold: $X_{n,n-k}$
    - k: as in usual extreme value statistics

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

    - It is not an order statistic (different distributions)
    - It nevertheless works as an order statistic from F
  - Count the frequency of "exceeding" in the first "s fraction"
  - Estimator: $\hat{C}(s) = \frac{1}{k} \sum_{i=1}^{[ns]} 1_{\{X_i > X_{n,n-k}\}}$
- Estimating $c(s)$
  - $C(s)$ is a distribution function with "density" $c(s)$.
  - We apply kernel density estimation to obtain c.
- Estimating γ
  - Hill estimator (as if observations are i.i.d.)

Theoretical property of the estimators

- Asymptotic normality of $\hat{C}$
  - 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 $\sqrt{k}\sup_{|u-v|\leq 1/n}|c(u) - c(v)| \to 0$
  - Theorem (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.
- Asymptotic normality of $\hat{\gamma}$
  - Usual second order condition and the condition on k
  - Under the same Skorokhod construction

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

    where N_0 follows standard normal distribution
- N_0 and $B(C(s))$ are independent

A tool for the proof: the STEP

- Sequential tail empirical process (STEP)
  - Notation $U := (1/(1-F))^{\leftarrow}$
  - Definition

$$\mathbb{F}_n(t,s) := \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{[ns]} 1_{X_i > U\left(\frac{n}{kt}\right)} - tC(s)\right).$$

  - Taking $s=1$: tail empirical process
  - Taking $t=1$: sequential process
  - The aforementioned estimators are functionals of the STEP

Theorem

There exists a standard bivariate Wiener process $W(t,s)$ on $[0,1]^2$ such that for proper weight function q, as $n \to \infty$

$$\sup_{0\leq t,s\leq 1} \frac{1}{q(t)}\left|\mathbb{F}_n(t,s) - W(t,C(s))\right| \to 0 \;\; a.s.$$

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: $T_1 := \sup_{0 \leq s \leq 1} \left| \hat{C}(s) - C_0(s) \right|$
  - Limit behavior:

$$\sqrt{k} T_1 \xrightarrow{d} \sup_{0 \leq s \leq 1} |B(C_0(s))|$$

- An alternative test
  - Test statistic: $T_2 := \int_0^1 (\hat{C}(s) - C_0(s)) dC_0(s)$
  - Limit behavior:

$$k T_2 \xrightarrow{d} \int_0^1 B^2(s) ds$$

Testing the model

- The null hypothesis: our model
  - γ is constant across the distributions
  - Skedasis 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 skedasis under the null hypothesis
  - Data violate that null, but following our model would be rejected there
  - We test constant γ in the presence of heteroscedasticity

Estimation on γ with partial sample

- Using observations in $(s_1, s_2]$
- The observations: $X_{[ns_1]+1}, \cdots, X_{[ns_2]}$
- Using a proper k: reflecting the intensity of extremes

$$k_{(s_1,s_2]} := k(\hat{C}(s_2) - \hat{C}(s_1))$$

- Estimation: using the Hill estimator $\hat{\gamma}_{(s_1,s_2]}$
- Limit behavior (under the null):

$$\sup_{s_2 - s_1 > \delta} \left| \sqrt{k} \left(\hat{\gamma}_{(s_1,s_2]} - \gamma \right) - \gamma \frac{W(C(s_2)) - W(C(s_1))}{C(s_2) - C(s_1)} \right| \to 0 \text{ a.s.}$$

- The starting point to construct test statistics

Testing constant γ

- Involving all partial samples
  - Instead of $s_2 - s_1 > \delta$, we look at $\hat{C}(s_2) - \hat{C}(s_1) > \delta$
  - Take all estimators with such subsamples
  - Test statistic: $T_3 := \sup_{\hat{C}(s_2)-\hat{C}(s_1)>\delta} \sqrt{k} \left|\hat{\gamma}_{(s_1,s_2]} - \hat{\gamma}\right|$
  - Limit behavior:

$$\sqrt{k} T_3 \xrightarrow{d} \sup_{s_2 - s_1 > \delta} \gamma \left| \frac{W(s_2) - W(s_1)}{s_2 - s_1} - W(1) \right|$$

- A "block POT" approach
  - Take m blocks as $0 = s_0 < s_1 < \cdots < s_m = 1$
  - Equal intensity in each block: $\hat{C}(s_j) - \hat{C}(s_{j-1}) = 1/m$ for $j = 1, \cdot, m$
  - Test statistic: $T_4 := \frac{1}{m} \sum_{j=1}^{m} \left(\frac{\hat{\gamma}_{(s_{j-1},s_j]}}{\hat{\gamma}} - 1 \right)^2$
  - Limit behavior:

$$kT_4 \xrightarrow{d} \chi^2(m-1)$$

VaR prediction

- We predict high quantiles at the "next" time point
- Assumptions
  - $c(s)$ is defined on $[0, 1+\varepsilon]$ for $\varepsilon > 0$
  - All conditions hold also with $i = n+1$
- Estimator

$$\widehat{U_{n,n+1}(1/p)} = X_{n,n-k}\left(\frac{k\hat{c}(1)}{np}\right)^{\hat{\gamma}_H}.$$

  - Need to estimate $\hat{c}(1)$
  - Use a boundary kernel:

$$\hat{c}(1) = \frac{1}{kh}\sum_{i=1}^{n} 1_{\left\{X_i^{(n)} > X_{n,n-k}\right\}} G_b\left(\frac{1-\frac{i}{n}}{h}\right),$$

  where

$$G_b(x) = \frac{\int_0^1 u^2 G(u)du - x\int_0^1 uG(u)du}{\frac{1}{2}\int_0^1 u^2 G(u)du - \left(\int_0^1 uG(u)du\right)^2} G(x);$$

Asymptotic normality for the predicted quantile

- Bandwidth choice
  - $kh \to \infty$
  - $hk^{1/5} \to \lambda \in [0, \infty)$
  - $\sqrt{h} \log(k/(np)) \to \beta \in [0, \infty)$
- Theorem

$$\sqrt{kh} \left(\frac{\widehat{U_{n,n+1}\left(\frac{1}{p}\right)}}{U_{n,n+1}\left(\frac{1}{p}\right)} - 1 \right) \xrightarrow{d} N\left(bias, variance\right)$$

$$bias = \lambda^{5/2} \frac{\gamma c''(1)}{2c(1)} \int_0^1 x^2 G_b(x) dx$$

$$variance = \gamma^2 \left(\frac{\int_0^1 G_b^2(x) dx}{c(1)} + \beta^2 \right)$$

Simulations

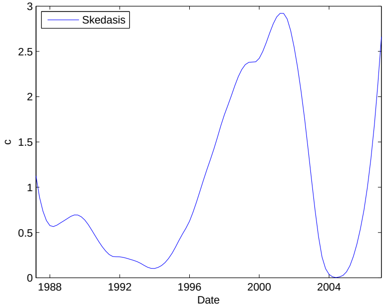
- Simulated observations
  - DGP 1: i.i.d. standard Fréchet $c(s) = 1$
  - DGP 2: $c(s) = 0.5 + s$
  - DGP 3: $c(s) = 2s + 0.5$, for $s \in [0, 0.5]$, $c(s) = -2s + 2.5$ for $s \in (0.5, 1]$
  - DGP 4: $c(s) = 0.8$, for $s \in [0, 0.4] \cup [0.6, 1]$, $c(s) = 20s - 7.2$ for $s \in (0.4, 0.5]$, $c(s) = -20s + 12.8$ for $s \in (0.5, 0.6)$.
- Sample size $n = 5,000$ (similar to that in application)
- Number of samples 1000
- Report: rejections under $1\%, 5\%, 10\%$ confidence level

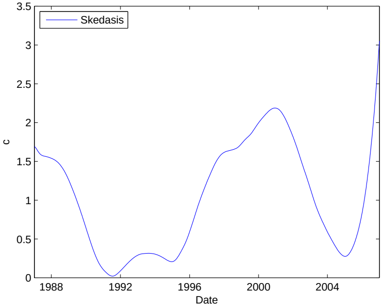
α	1%		5%		10%	
Test	T_1	T_2	T_1	T_2	T_1	T_2
DGP 1	8	12	44	47	95	98
DGP 2	990	998	998	999	1000	1000
DGP 3	455	570	838	921	941	987
DGP 4	663	521	930	903	979	978

Application

- Data: S&P500 daily returns (1988–2012) 6,302 obs
  - Testing constant γ: T_3 and T_4, strong rejection
  - Not possible to apply the theory
- Sub-sample: 1988-2007 (5,043 obs)
  - Testing constant γ: $p = 0.98(T_3)$ and $p = 0.76(T_4)$
  - Testing constant $c(s)$: T_1 and T_2, strong rejection
  - Next, we plot the estimated $c(s)$
- Robustness check: weekly returns (1,043 obs)

The skedasis function over time





The ongoing work: abolishing the "tail comparability"

- Recall the notation $F_{n,i}$ as the distribution function of X_i
  - A series of distribution functions $F_s(x) := F(s,x)$: $F_s \in \mathcal{D}_{\gamma(s)}$
  - $F_{n,i} = F_{\frac{i}{n}}$ for $i = 1, 2, \cdots, n$
  - Note that $\gamma(s)$ is now varying across s!
- The goal: estimate $\gamma(s)$ with observations $X_1, \cdots, X_n$
- Second order condition: Denote $U_s = (1/(1-F_s))^{\leftarrow}$, then

$$\lim_{t\to\infty} \frac{\frac{U_s(tx)}{U_s(t)} - x^{\gamma(s)}}{A_s(t)} = x^{\gamma(s)} \frac{x^{\rho(s)} - 1}{\rho(s)},$$

holds uniformly for all $s \in [0,1]$ and $x > 1$.

  - $\rho(s)$: continuous negative function
  - $A_s(t) := A(s,t)$ continuous with respect to s

Further assumptions on continuity and smoothness

- Intermediate sequence and band width: $h \to 0$, $kh \to \infty$.
- Notation: $\overline{\gamma} = \sup_{0 \le s \le 1} \gamma(s)$ and $\underline{\gamma} = \inf_{0 \le s \le 1} \gamma(s)$
- The quantile functions varies slowly:

$$\sqrt{k} \sup_{|s_1 - s_2| \le h} \left| \frac{U_{s_1}\left(\frac{n}{k}\right)}{U_{s_2}\left(\frac{n}{k}\right)} - 1 \right| \to 0.$$

- The function $\gamma(s)$ varies slowly: for some $\varepsilon > 0$,

$$k^{1/2+\overline{\gamma}+\varepsilon} \sup_{|s_1 - s_2| \le h} |\gamma(s_1) - \gamma(s_2)| \to 0.$$

- No asymptotic bias in our asymptotic theory: for some $\varepsilon > 0$,

$$k^{1/2+\overline{\gamma}+\varepsilon} \sup_{0 \le s \le 1} \left| A_s\left(\frac{n}{k}\right) \right| \to 0.$$

Asymptotic theories: local versus global

- Local estimation
  - Local estimator for $\gamma(s)$: Hill estimator in a h-neighborhood
    - Top $[2kh]$ order statistics among $[2nh]$ local observations
  - Local asymptotic theory

$$\sqrt{2kh}\left(\widehat{\gamma(s)} - \gamma(s)\right) \xrightarrow{d} N(0, (\gamma(s))^2).$$

- Global estimation
  - The goal: $\Gamma(s) = \int_0^s \gamma(u)du$
  - Estimator:

$$\widehat{\Gamma(s)} = 2h \sum_{s_t \leq s} \widehat{\gamma(s_t)}.$$

    - The series $s_t = (2t-1)h$ for $t = 1, 2, \cdots$.
  - Asymptotic theory

$$\sqrt{k}\left(\widehat{\Gamma(s)} - \Gamma(s)\right) \xrightarrow{d} \int_0^s \gamma(u)dW(u).$$

Conclusion

- We can handle extreme value statistics when observations are drawn from different distributions
- We can identify whether heteroscedastic extremes are due to the variation of γ or skedasis
- If the skedasis varies, we can quantify that variation
- If the γ varies, we can also estimate the variation in γ.
- Handle the γ constant case: the Sequential Tail Empirical Process (STEP)
  - A useful tool that can be applied to other estimators
  - *It was the first STEP towards non-stationarity.*
- *Now we have made the second step!*