Tail Dynamic Analysis of Financial Time Series
with Autoregressive Conditional Fréchet Model

Zeyu Wei

April 16, 2021

Outline

- **Intro to Extreme Value Theory**
- **Motivation**
- **Model Specification**
- **Parameter Estimation**
- **Real Data Application**
- **My work**

Two Approaches to Extreme Value

- The first method relies on deriving block maxima (minima) series generating an “Annual Maxima Series” (AMS).
- The second method relies on extracting values exceed a certain threshold (falls below a certain threshold) for some period, getting “Peak Over Threshold” (POT) values.

An Introduction to Statistical Modeling of Extreme Values
Stuart Coles, 2001.

Examples of Extreme Values

- Potential maximum daily loss across all stocks in portfolio
- Potential intra-day maximum loss of a stock
- Flood/earthquake level for civil engineering
- Extreme weather conditions
- .....

Asymptotic Models for Maxima

Let $X_1, X_2 \ldots, X_n \ldots$ be a sequence of i.i.d.-distributed random variables with distribution F, and $M_n = \max\{X_1, \ldots, X_n\}$.
In this case, we can have $Pr\{M_n < x\} = F(x)^n$

Asymptotic Models for Maxima

Let $X_1, X_2 \ldots, X_n \ldots$ be a sequence of i.i.d.-distributed random variables with distribution F, and $M_n = \max\{X_1, \ldots, X_n\}$.
In this case, we can have $Pr\{M_n < x\} = F(x)^n$
In practice, F is unknown.

Asymptotic Models for Maxima

Let $X_1, X_2 \ldots, X_n \ldots$ be a sequence of i.i.d.-distributed random variables with distribution F, and $M_n = \max\{X_1, \ldots, X_n\}$.
In this case, we can have $Pr\{M_n < x\} = F(x)^n$
In practice, F is unknown.

- Estimate $F \Rightarrow F^n$
- "Central Limit Theorem" on extreme values

Why Use Asymptotic Models for Extreme Values

- Data collection/Sample
- $\hat{F}$ unreliable, worse for $\hat{F}^n$
- The variation in the maximum behaves differently from variation in the mean

Fisher–Tippett–Gnedenko (Extremal Types) Theorem

Theorem

If there exist sequences of constants $\{a_n > 0\}$ *and* $\{b_n\}$ *such that*

$$\lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = G(x)$$

where G *is a non-degenerate distribution function, then the limit distribution* G *belongs to one of the following families:*

- $G(x) = exp\{-exp[-(\frac{x-b}{a})]\}, -\infty < x < \infty$ *(Gumbel)*
- $G(x) = exp[-(\frac{x-b}{a})^{-\alpha}], x > b;$ 0 *otherwise (Fréchet)*
- $G(x) = exp\{-[-(\frac{x-b}{a})]^{\alpha}\}, x < b;$ 1 *otherwise (Weibull)*

Fisher–Tippett–Gnedenko (Extremal Types) Theorem

Theorem

If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = G(x)$$

where G is a non-degenerate distribution function, then the limit distribution G belongs to one of the following families:

- $G(x) = exp\{-exp[-(\frac{x-b}{a})]\}, -\infty < x < \infty$ *(Gumbel)*
- $G(x) = exp[-(\frac{x-b}{a})^{-\alpha}], x > b;$ 0 *otherwise (Fréchet)*
- $G(x) = exp\{-[-(\frac{x-b}{a})]^{\alpha}\}, x < b;$ 1 *otherwise (Weibull)*

Generalized Extreme Value (GEV) Distribution:

$$G(x) = exp\Big\{-\Big[1 + \xi(\frac{x-\mu}{\sigma})\Big]^{-1/\xi}\Big\},$$

Generalized Pareto Distribution (GPD)

Theorem

Let $X_1, X_2 \ldots, X_n \ldots$ be a sequence of independent and identically-distributed random variables with distribution F, and $M_n = \max\{X_1, \ldots, X_n\}$. Denote arbitrary term in the X_i sequence by X, and suppose that the GEV distribution is satisfied that, for large n,

$$Pr\{M_n \leq z\} \approx G(z) = exp\Big\{ - \big[1 + \xi(\frac{x - \mu}{\sigma})\big]^{-1/\xi} \Big\}$$

Then for large enough u (threshold), the distribution function of $(X - u)$, conditional on $X > u$, is approximately

$$H(y) = 1 - (1 + \frac{\xi y}{\tilde{\sigma}})^{-1/\xi}$$

where $\tilde{\sigma} = \sigma + \xi(u - \mu)$, defined on $\{y : y > 0$ and $(1 + \xi y/\tilde{\sigma}) > 0\}$

Maxima-GEV vs. POT-GPD

Maxima-GEV

- Direct modeling of maxima, especially for high-frequency time series
- Comparable performance with POT-GPD for large sample size

POT-GPD

- More data used, more efficient
- Sensitive to threshold

Autoregressive Conditional Fréchet (AcF) Model

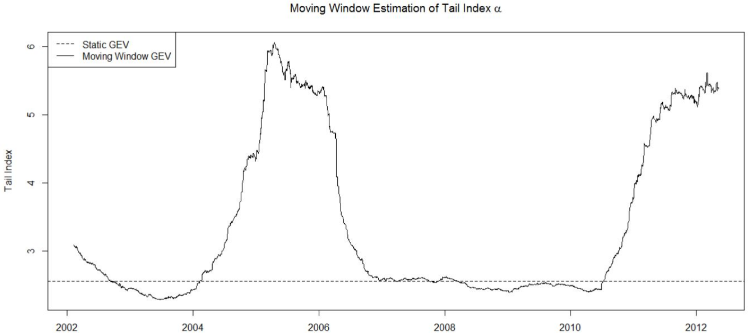
Modeling Maxima with Autoregressive Conditional Fréchet Model
Zifeng Zhao, Zhengjun Zhang, Rong Chen, 2018
Journal of Econometrics

Motivation

The behavior of the underlying time series may change through time, and the static approach cannot capture the dynamics of series.

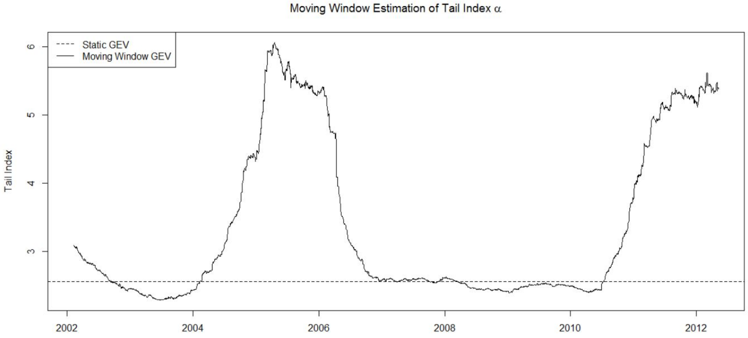
Motivation

The behavior of the underlying time series may change through time, and the static approach cannot capture the dynamics of series.
Tail index of financial markets varies through time:



Motivation

The behavior of the underlying time series may change through time, and the static approach cannot capture the dynamics of series.
Tail index of financial markets varies through time:



Recent studies on dynamic POT-GPD models, but little on dynamic GEV models.

Autoregressive Conditional Fréchet (AcF) Model

Let $\{X_{it}\}_{i=1}^{p}$ be a set of time series, and let $\{Q_t\}$ be univariate time series of maxima that $Q_t = max_{1 \leq i \leq p} X_{it}$.

Autoregressive Conditional Fréchet (AcF) Model

Let $\{X_{it}\}_{i=1}^{p}$ be a set of time series, and let $\{Q_t\}$ be univariate time series of maxima that $Q_t = max_{1 \leq i \leq p} X_{it}$.
Model Q_t conditionally with Fréchet distribution with parameters $(\mu_t, \sigma_t, \alpha_t) \in \mathcal{F}_{t-1} = \sigma(Q_{t-1}, Q_{t-2}, \ldots)$ where $\alpha_t = 1/\xi_t$, and the parametrization:

$$Q_t = \mu_t + \sigma_t Y_t^{1/\alpha_t}$$

where $\{Y_t\}$ is a sequence of *i.i.d.* unit Fréchet random variables.

Autoregressive Conditional Fréchet (AcF) Model

Let $\{X_{it}\}_{i=1}^{p}$ be a set of time series, and let $\{Q_t\}$ be univariate time series of maxima that $Q_t = max_{1 \leq i \leq p} X_{it}$.
Model Q_t conditionally with Fréchet distribution with parameters $(\mu_t, \sigma_t, \alpha_t) \in \mathcal{F}_{t-1} = \sigma(Q_{t-1}, Q_{t-2}, \ldots)$ where $\alpha_t = 1/\xi_t$, and the parametrization:

$$Q_t = \mu_t + \sigma_t Y_t^{1/\alpha_t}$$

where $\{Y_t\}$ is a sequence of *i.i.d.* unit Fréchet random variables.

- The scale parameter σ_t should not be taken exactly as volatility of Q_t
  The conditional variance of Q_t depends on both on σ_t and α_t
  However, σ_t can be closely related to the volatility process of the underlying time series $\{X_{it}\}_{i=1}^{p}$

Model Specification

- μ_t is set to be constant
- $\log \sigma_t = \beta_0 + \beta_1 \log \sigma_{t-1} + \eta_1(Q_{t-1})$
- $\log \alpha_t = \gamma_0 + \gamma_1 \log \alpha_{t-1} + \eta_2(Q_{t-1})$
- $\beta_1, \gamma_1 \geq 0$
- $\eta_1(.)$ is assumed to be continuous increasing function of Q_{t-1}, and $\eta_2(.)$ is assumed to be continuous decreasing function of Q_{t-1} (clustering of extreme events)

Model Specification

- μ_t is set to be constant
- $\log \sigma_t = \beta_0 + \beta_1 \log \sigma_{t-1} + \eta_1(Q_{t-1})$
- $\log \alpha_t = \gamma_0 + \gamma_1 \log \alpha_{t-1} + \eta_2(Q_{t-1})$
- $\beta_1, \gamma_1 \geq 0$
- $\eta_1(.)$ is assumed to be continuous increasing function of Q_{t-1}, and $\eta_2(.)$ is assumed to be continuous decreasing function of Q_{t-1} (clustering of extreme events)

Further let $\eta_1(.)$, $\eta_2(.)$ to be exponential functions of the form $a_0 exp(-a_1 x)$

- Simplified version of the widely used logistic function $\frac{L}{1+a_0 exp(-a_1 x)}$
- Monotonicity, differentiability
- Implies upper bound for $\{\sigma_t\}, \{\alpha_t\}$, does not affect flexibility but facilitates numerical and techical tractability

Model Specification (Cont.)

The specific model is:

$$Q_t = \mu + \sigma_t Y_t^{1/\alpha_t}$$

$$\log \sigma_t = \beta_0 + \beta_1 \log \sigma_{t-1} - \beta_2 \exp(-\beta_3 Q_{t-1})$$

$$\log \alpha_t = \gamma_0 + \gamma_1 \log \alpha_{t-1} + \gamma_2 \exp(-\gamma_3 Q_{t-1})$$

where $\{Y_t\}$ is a sequence of *i.i.d.* unit Fréchet random variables, $0 \leq \beta_1 \neq \gamma_1 < 1, \beta_2, \beta_3, \gamma_2, \gamma_3 > 0$.

Model Specification (Cont.)

The specific model is:

$$Q_t = \mu + \sigma_t Y_t^{1/\alpha_t}$$

$$\log \sigma_t = \beta_0 + \beta_1 \log \sigma_{t-1} - \beta_2 \exp(-\beta_3 Q_{t-1})$$

$$\log \alpha_t = \gamma_0 + \gamma_1 \log \alpha_{t-1} + \gamma_2 \exp(-\gamma_3 Q_{t-1})$$

where $\{Y_t\}$ is a sequence of *i.i.d.* unit Fréchet random variables, $0 \leq \beta_1 \neq \gamma_1 < 1, \beta_2, \beta_3, \gamma_2, \gamma_3 > 0$.

- Can include q_1 autoregressive terms and q_2 lagged terms of Q_t
- Increase complexity and instability, but not necessarily improve model performance

Stationarity and Ergodicity

$\{\sigma_t, \alpha_t\}$ form a homogeneous Markov chain in $\mathbb{R}^2$

Theorem

For an AcF with $0 \leq \beta_1 \neq \gamma_1 < 1, \beta_2, \beta_3, \gamma_2, \gamma_3 > 0, \beta_0, \gamma_0, \mu \in \mathbb{R}$, *the latent process* $\{\sigma_t, \alpha_t\}$ *is stationary and geometrically ergodic.*

Since $\{Q_t\}$ is a coupled process of $\{\sigma_t, \alpha_t\}$, it is also stationary and ergodic.

Parameter Estimation

Denote all the parameters in the model by
$\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \gamma_0, \gamma_1, \gamma_2, \gamma_3, \mu)$
Denote the space by
$\Theta_s = \{\theta | \beta_0, \gamma_0, \mu \in \mathbb{R}, \quad \beta_1, \gamma_1 \in [0, 1), \quad \beta_2, \beta_3, \gamma_2, \gamma_3 > 0\}$

Parameter Estimation

Denote all the parameters in the model by
$\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \gamma_0, \gamma_1, \gamma_2, \gamma_3, \mu)$
Denote the space by
$\Theta_s = \{\theta | \beta_0, \gamma_0, \mu \in \mathbb{R}, \quad \beta_1, \gamma_1 \in [0, 1), \quad \beta_2, \beta_3, \gamma_2, \gamma_3 > 0\}$

The conditional p.d.f. of Q_t given $(\mu, \sigma_t, \alpha_t)$ is

$$f_t(\theta) = f(Q_t | \mu, \sigma_t, \alpha_t) = \alpha_t \sigma_t^{\alpha_t} (Q_t - \mu)^{-(\alpha_t + 1)} \exp\{-\sigma_t^{\alpha_t} (Q_t - \mu)^{-(\alpha_t + 1)}\}$$

Parameter Estimation

Denote all the parameters in the model by
$\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \gamma_0, \gamma_1, \gamma_2, \gamma_3, \mu)$
Denote the space by
$\Theta_s = \{\theta | \beta_0, \gamma_0, \mu \in \mathbb{R}, \quad \beta_1, \gamma_1 \in [0,1), \quad \beta_2, \beta_3, \gamma_2, \gamma_3 > 0\}$

The conditional p.d.f. of Q_t given $(\mu, \sigma_t, \alpha_t)$ is

$$f_t(\theta) = f(Q_t|\mu, \sigma_t, \alpha_t) = \alpha_t \sigma_t^{\alpha_t} (Q_t - \mu)^{-(\alpha_t+1)} \exp\{-\sigma_t^{\alpha_t} (Q_t - \mu)^{-(\alpha_t+1)}\}$$

Hence, with conditional independence, the log-likelihood function with observations $\{Q_t\}_{t=1}^n$ is

$$\begin{aligned} L_n(\theta) &= \frac{1}{n} \sum_{t=1}^{n} l_t(\theta) \\ &= \frac{1}{n} \sum_{t=1}^{n} [\log \alpha_t + \alpha_t \log \sigma_t - (\alpha_t + 1) \log(Q_t - \mu) - \sigma_t^{\alpha_t} (Q_t - \mu)^{-(\alpha_t+1)}] \end{aligned}$$

$\{\sigma_t, \alpha_t\}_{t=1}^n$ can be obtained recursively.

Parameter Estimation

- True value of (σ_1, α_1) unknown

Parameter Estimation

- True value of (σ_1, α_1) unknown

Doesn't matter!

With $0 \leq \beta_1, \gamma_1 < 1$, the influence of initial value decay exponentially. The asymptotic distribution does not depend on (σ_1, α_1). The consistency and asymptotic normality of MLE do not depend on (σ_1, α_1).

Use static GEV estimation for initial value (burn-in)

Consistency and Asymptotic Normality of MLE

Let $\tilde{L}_n(\theta)$ be the log-likelihood based on arbitrary initial values $(\tilde{\sigma}_1, \tilde{\alpha}_1)$

Theorem (Consistency)

Assume the parameter space Θ is a compact set of Θ_s. Suppose the observations $\{Q_t\}_{t=1}^n$ are generated by a stationary and ergodic AcF with true parameter θ_0 and θ_0 is in the interior of Θ, then there exists a sequence $\hat{\theta}_n$ of local maximizer of $\tilde{L}_n(\theta)$ s.t. $\hat{\theta}_n \to_p \theta_0$ and $||\hat{\theta}_n - \theta_0|| \leq \tau_n$, where $\tau_n = O_p(n^{-r}), 0 < r < \frac{1}{2}$. Hence $\hat{\theta}_n$ is consistent.

Proposition (Asymptotic Uniqueness)

Denote $V_n = \{\theta \in \Theta | \mu \leq cQ_{n,1} + (1-c)\mu_0\}$ where $Q_{n,1} = min_{1 \leq t \leq n} Q_t$, under the conditions in previous Theorem, for any fixed $0 < c < 1$, there exists a sequence of $\hat{\theta}_n = \arg\max_{\theta \in V_n} \tilde{L}_n(\theta)$ such that $\hat{\theta}_n \to_p \theta_0$ and $||\hat{\theta}_n - \theta_0|| \leq \tau_n$, where $\tau_n = O_p(n^{-r}), 0 < r < \frac{1}{2}$, and
$\mathbb{P}(\hat{\theta}_n \text{is the unique global maximizer of } \tilde{L}_n(\theta) \text{ over } V_n) \to 1.$

This proposition gives a partial answer to the asymptotic uniqueness of MLE.V_n can be made arbitararily close to Θ_n by the fact that $Q_{n,1} \downarrow \mu_0$ a.s. and by setting c close to 1.

Consistency and Asymptotic Normality of MLE

Theorem (Asymptotic Normality)

Under the consistency theorem, we have $\sqrt{n}(\hat{\theta}_n - \theta_0) \to^d N(0, M_0^{-1})$*, where* $\hat{\theta}_n$ *is the estimator in previous theorem and* M_0 *is the Fisher Information matrix evaluated at* θ_0*. Further, the sample variance of plug-in estimated score functions* $\{\frac{\partial}{\partial\theta} l_t(\hat{\theta}_n)\}_{t=1}^n$ *is a consistent estimator of* M_0*.*

The consistency and asymptotic normality theorem show that there always exists a sequence $\hat{\theta}_n$, which is a local maximizer of $\tilde{L}_n(\theta)$, such that $\hat{\theta}_n$ is consistent and asymptotically normal regardless of the initial values $(\tilde{\sigma}_1, \tilde{\alpha}_1)$.

Simulations

- Under factor model setting

$$X_{it} = f_i(Z_{1t}, Z_{2t}, \ldots, Z_{dt}) + \sigma_{it}\epsilon_{it}$$

- With some mild conditions, the conditional distribution of maxima $Q_t = \max_{1 \le i \le p} X_{it}$ can be well approximated by a Fréchet distribution.

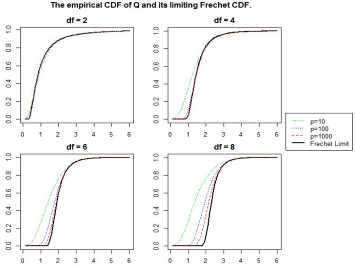
Simulations

- Under factor model setting

$X_{it} = f_i(Z_{1t}, Z_{2t}, \ldots, Z_{dt}) + \sigma_{it}\epsilon_{it}$

- With some mild conditions, the conditional distribution of maxima $Q_t = \max_{1 \le i \le p} X_{it}$ can be well approximated by a Fréchet distribution.
- Convergence of maxima in factor model

$X_i = \beta_i Z + \sigma_i \epsilon_i, i = 1, \ldots, p,$
$Z \sim N(0,1)$, $\beta_i \sim^{i.i.d.} U(-2,2)$, $\sigma_i \sim^{i.i.d.} \frac{1}{2}U(0.5, 1.5) + \frac{1}{2}U(0.75, 1.25)$, ϵ_i i.i.d. t-distributions with degrees of freedom ν. Repeated 1000 times.



Simulations

- Estimation for conditional VaR of maxima
- For $0 < q < 1$, the conditional Value at Risk for time t, $cVaR_t^q$ is defined as the $1 - q$ extreme quantile of Q_t, given all past information $\mathcal{F}_{t-1}$

$X_{it} = 0.009(\beta_i Z_t + \sigma_i \epsilon_{it})$
$Z_t \sim N(0,1)$, $\beta_i \sim^{i.i.d.} U(-2,2)$, $\sigma_i \sim^{i.i.d.} \frac{1}{2}U(0.5,1.5) + \frac{1}{2}U(0.75,1.25)$, ϵ_i i.i.d. t-distributions with degrees of freedom ν_t
$\log \nu_t = \gamma_0 + \gamma_1 \log \nu_{t-1} + \gamma_2 \exp(-\gamma_3 Q_{t-1})$,
$(\gamma_0, \gamma_1, \gamma_2, \gamma_3) = (-0.1, 0.9, 0.3, 5)$
Train AcF on $T_1 = 1000, 2000, 5000$, calculate next $T_2 = 100$ Q_t values.
Repeat the experiment 500 times for each combination

T_1	$\bar{q}$ $(q^0 = 0.1)$	$\bar{q}$ $(q^0 = 0.05)$	$\bar{q}$ $(q^0 = 0.01)$	mean cor.	median cor.
1000	0.095	0.049	0.012	0.871	0.928
2000	0.096	0.049	0.012	0.909	0.952
5000	0.097	0.051	0.012	0.947	0.973

Simulations

- Robust to mild dependence among ϵ_{it}

For each day, 100 ϵ_{it} are generated from 10 different multivariate t-distributions of size 10, with moderate pairwise-correlations (≤ 0.5) within each multivariate t-distribution.

T_1	$\bar{q}$ ($q^0 = 0.1$)	$\bar{q}$ ($q^0 = 0.05$)	$\bar{q}$ ($q^0 = 0.01$)	mean cor.	median cor.
1000	0.094	0.048	0.012	0.862	0.921
2000	0.097	0.047	0.011	0.876	0.936
5000	0.096	0.048	0.011	0.918	0.960

- Robust to heterogeneous ϵ_{it}

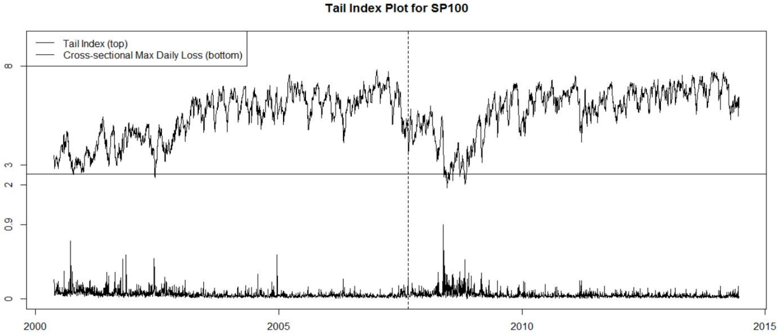
For each day, 100 ϵ_{it} are generated independently from 100 t-distributions of degree of freedom $c_i\nu_t$, c_i generated independently from $U(0.8, 1)$

T_1	$\bar{q}$ ($q^0 = 0.1$)	$\bar{q}$ ($q^0 = 0.05$)	$\bar{q}$ ($q^0 = 0.01$)	mean cor.	median cor.
1000	0.098	0.051	0.013	0.864	0.922
2000	0.097	0.050	0.012	0.905	0.953
5000	0.098	0.051	0.012	0.956	0.974

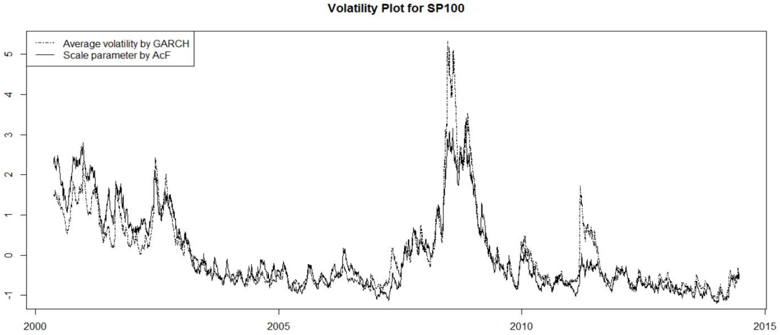
Real Data Applications

Negative daily log-returns of S&P100 Index components from January 1, 2000 to December 31, 2014 (3773 obs).

S&P100	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3	μ
Mean	-0.068	0.890	0.328	5.33	-0.050	0.961	-0.051	6.68	-0.069
S.D.	0.014	0.013	0.063	1.27	0.006	0.004	0.0072	1.01	0.006



Real Data Applications



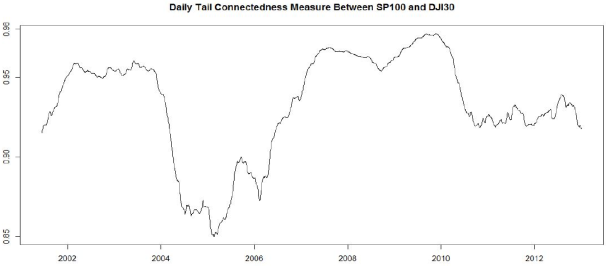
Implications for Tail-connectedness

- Let the estimated tail indexes for the S&P100 data be $\{\hat{\alpha}_t^S\}$
- Similarly applied AcF to DJI30 data, and denote the etimated tail indexes be $\{\hat{\alpha}_t^D\}$

Overall correlation between $\{\hat{\alpha}_t^S\}$ $\{\hat{\alpha}_t^D\}$ is 0.93

Let $\hat{\boldsymbol{\alpha}}_t = (\hat{\alpha}_t^S, \hat{\alpha}_t^D\ )$, let $\hat{\Sigma}$ be the sample covariance matrix of $\hat{\boldsymbol{\alpha}}_t$ based on sample $\hat{\boldsymbol{\alpha}}_{t=1}^T$.

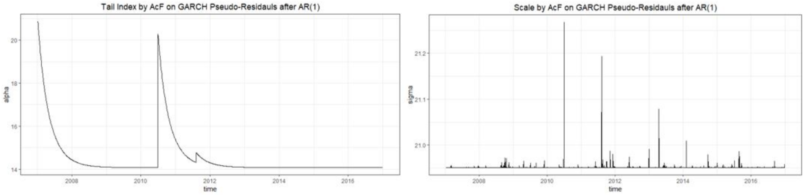
Use the ratio between the maximum eigenvalue of $\hat{\Sigma}$ and the sum of all eigenvalues as a measure of tail-connectedness. Rolling window of 36-months (about 756) days.



Post tail dynamic analysis

Apply AR(1) + GARCH(1,1) to each S&P500 component stock returns

- On daily maximal pseudo-residuals
- Test normality: Normal aggregates to Gumbel(μ, α), and exponential of Gumbel is Fréchet with location 0, scale $\exp(\mu)$, and tail $1/\alpha$.
- Apply AcF to pseudo-residuals w/o manipulation

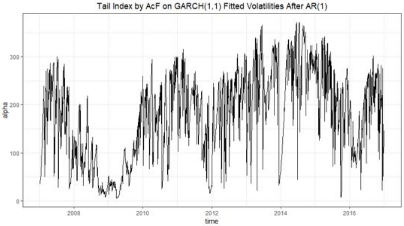


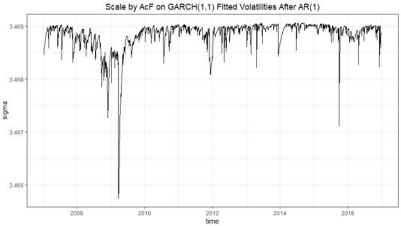
γ_1 significant while β_1 not. The aggregated maximal pseudo-residuals have phenomena of volatility clustering and clustering of extremal values, and hence the pseudo-residuals of GARCH(1,1) are not as random and independent as normally believed.

Post tail dynamic analysis

Apply AR(1) + GARCH(1,1) to each S&P500 component stock returns

- On daily maximal fitted-volatilities





The "market-wise" maximal volatilities of stock returns have its own dynamics.

Other Related Works/Further Research

- Construct pseudo-series (ghost data)
  Get the max pseudo-residual and max fitted volatility at the given time, multiply back for a pseudo-volatility. Apply AcF.
  Magnified trends
- Fama French industry classification
- Other η functions for different types of data
- Extension to light-tailed data
- ......

Thanks