

ファイナンス分析における極値分布の利用

森 棟 公 夫

この稿では大きな損失率に関する分析法を紹介する。最初に、データ系列として最大値しか得られない場合に可能な分析を解説する。たとえば収益率の日中データは得られないが、日中の最大損失率が手に入る場合に可能なリスク分析の方法である。この方法を使えば、年間最大損失率の系列を使って、40年に一度しか起きないような大きな損失率の分布を分析することが可能である。第2節では、最大値データではなく損失率の原系列が得られるとして、大きな損失率の分布を分析し、大きな損失が起きる確率を求める。第3節では、分布の形状を定めるtail indexの推定方法を説明する。

1 極値の性質

確率変数 X_1, X_2, \dots, X_n は独立同一分布(iid)に従うとする。また、確率変数はこの稿においては損失率を表現するとしよう。収益率でも分析上は同じだが、その場合は符号が逆になる。したがって損失はプラス値になり、利益が出るとマイナス値になる。極度に大きな損失は、分布関数では右裾の事象である。この様な損失率の分布関数は、任意の i について $F(x)=P(X_i \leq x)$ となるが、分布関数は共通とする。以上の準備の下で、 n 損失の内の最大損失 M_n は、よく知られた最大値の分布関数により

$$P(M_n \leq x) = F(x)^n \quad (1)$$

となる。この分布関数を用いてリスクに関する分析を行うと、中程度の損失に関しては精度の高い近似が得られるものの、大きな損失の観測値が少ないため、分布関数の右裾を正確に推定できないことが知られている。この稿では、最大値の分布を改善して、大きな損失率に対する精度の高い分析を試みる。第1節では次の様な事象の分析を目的とする。

1. 年間最大損失率がいままでの最大損失率を超えるとすれば、この損失が起きる確率はいくらか。
2. 40年に一度しか超えられないような損失率(40年損失率)はいくらか。

(1)式から n に関する極限を求めると、確率 $F(x)$ は1以下だから分布関数は0か1に収束することになり、極限で分布は退化してしまう。漸近分布がこのように退化することを避けるために、最大値を

$$Z_n = \frac{M_n - \mu_n}{\sigma_n} \quad (2)$$

と標準化する。平均と標準偏差による標準化だが、この系列は未知である。この様に標準化された系列に関して、次に述べる定理が知られている。

1-1 Fisher-Tippet定理と一般化極値分布

(2)式で定義された確率変数の漸近分布は3種類に分けることができるが、漸近分布は未知である ξ に依存する。 ξ は分布の形状を決めるが、特にその逆数 $\alpha = 1/\xi$ はtail-indexと呼ばれる。3節において、tail-indexを使った分析をさらに進める。

正規分布、対数正規分布、指数分布、ガンマ分布などのように元の分布関数 F が指数的に減衰する裾を持つならば、漸近分布はグンベル(Gumbel)属で $\xi = 0$ となり、 $-\infty < z < \infty$ に関して

$$\lim_{n \rightarrow \infty} P(Z_n < z) = H_\xi(z) = \exp\{-\exp(-z)\}, \quad (3)$$

となる。このクラスに入る分布に関しては、期待値は全ての次数について存在する。

F の裾が、パレート分布、コーシー分布、 t 分布などのように、ゆっくりと変化する関数 $L(x)$ によって

$$1 - F(x) = x^{-1/\xi} L(x) \quad (4)$$

と表現できるならば、漸近分布はフレシェ(Frechet)属で $\xi > 0$ となり、 $1 + \xi z > 0$ について

$$H_\xi(z) = \exp\{-(1 + \xi z)^{-1/\xi}\}, \quad (5)$$

となる。(domain of attractionの議論による。)元の分布に関しての期待値は、tail indexを超える次数で発散する。

F の分布が一様分布やベータ分布のように有限

範囲なら漸近分布はワイブル(Weibull)属で $\xi < 0$ となり、 $1 + \xi z < 0$ の範囲で

$$H_\xi(z) = \exp\{-(1 + \xi z)^{-1/\xi}\}, \quad (6)$$

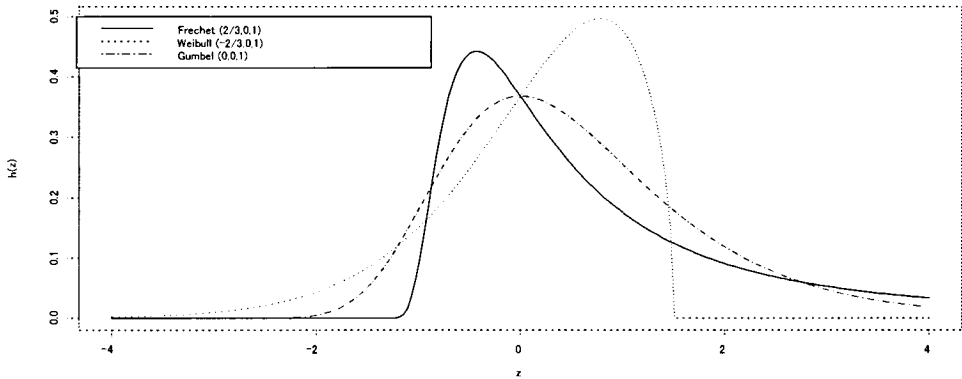
となる。元の分布についての期待値は全ての次数で存在する。

これらの分布は、まとめて一般化極値(GEV, generalized extreme value)分布と呼ばれる。このGEV分布は(2)式の標準化に依存し、次式

$$P(M_n < x) = P\left(\frac{M_n - \mu_n}{\sigma_n} < \frac{x - \mu_n}{\sigma_n}\right) \quad (7)$$

の右辺の近似である。 ξ 、 μ 、 σ はいずれも未知とする。イメージをつかむために ξ の値を $-2/3$ 、 0 、 $2/3$ として3種類のGEV密度関数を作図すると図1ようになる。図では μ は 0 、 σ は 1 と設定してある。

図1 GEV密度関数

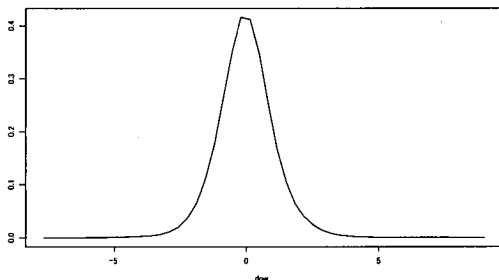


2 日次収益率を用いた推定

推定は非常に簡単で、極値分布から密度関数を一次導関数として求め、極値密度関数のパラメータを推定する。この手法の漸近的な妥当性はともかく、推定がどのような結論をもたらすか数値例で確認しよう。最初に日次収益率が得られる場合について、日次収益率から月間最大値を求め、月間最大値の分布に関する推定を行う。利用する系列はダウ平均株価で、期間は1990年1月1日から2004年9月30日である。

系列の分布を理解するために1990年1月以後のダウ平均株価の連続日次収益率を計算し、日次収益率の符号を逆にして、経験密度をカーネル法で求める。最小値は-7.15、最大値は6.15であり分布は

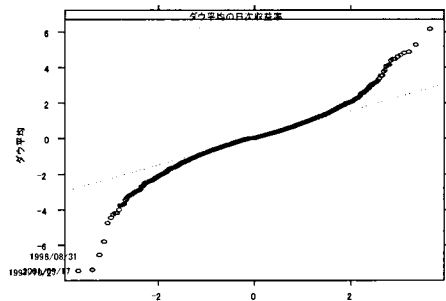
図2 損失率の分布



原点对称に近いことが分かる。

正規QQプロットを描くが、収益率の分布は標準正規より裾が厚くなっている。したがって、極値分布はグンベルではなくフレシエ型になり、 $\xi > 0$ であると予想できる。もちろん分布範囲に切断は無いので、ワイブル分布は適切でない。

図3 ダウ平均株価損失率のQQプロット

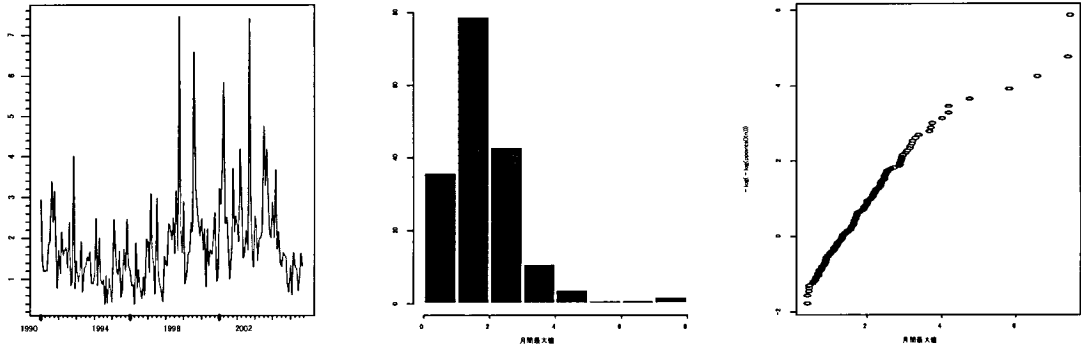


次に、日次収益率から月間最大損失を求め、最大値の分布の性質を検討する。図4では最大値系列をまず描き、中央でヒストグラムを作成する。この最大値に関するヒストグラムをグンベルやフレシエ分布と比較すれば、後者に似ていることが確認できよう。QQプロットでは基準となる分布型はグンベルで、もし観測値の標本分布がグンベル

に乗れば、QQプロットは直線になり、フレシエ分布が採用できないことになる。この例ではプロットは直線上にあるとは考えられないので、分布

はグンベルでなくフレシエであることが再確認できる。

図4 月間最大値の性質



以上の準備により、フレシエ型分布を推定する。結果は、 $\xi = 0.156(0.06)$, $\sigma = 0.687(0.05)$, $\mu = 1.33(0.06)$ となった。括弧内は標準誤差である。

次に推定結果に関する二種類の診断分析を行う。診断では M_i を第 i 月の月間最大値として、係数推定値を用いて W_i を

$$W_i = (1 + \frac{\xi}{\hat{\sigma}} \frac{M_i - \hat{\mu}}{\hat{\sigma}})^{-1/\xi} \quad (8)$$

と変換する。一種の残差であるが、この変換値の時系列がトレンドを持つか否か、またQQプロットが直線に乗るか否かの判定を行う。 M_i が iid であれば、変換値の時系列にトレンドは存在しない。次にこの変換統計量の分布を導こう。

一般に確率変数 X が分布関数 $F(x)$ を持つならば、確率変数 $F(X)$ の分布関数は $(0,1)$ 間の一様分布で

$$P(F(X) < x) = x, (0 < x < 1) \quad (9)$$

となる。 W_i はフレシエ分布関数の $-\log$ に一致するから、 W_i の分布関数は、

$$P(-\log(F(X)) < z) = P(F(X) > \exp(-z)) = 1 - \exp(-z) \quad (10)$$

より、指数分布となる。これを確認するのが第2図である。

図5の左図は各 W_i が iid 確率変数であるか否かを調べるために描かれているが、この図から特に際だったトレンドは見られない。図中の曲線は、各 W_i の近辺の値を計算に入れたカーネル平均値の軌跡で、時系列的な性質があればこの曲線にトレンドが表れるはずである。右図は指数分布と W_i のQQプロットで、推定された係数を用いた残差の変換 W_i が指数分布に従うなら、散布図は直線に乗る。この例のフィットは裾の5個ほどを除いて全体としてはよいと言える。

図5 変換 w のプロットとQQプロット

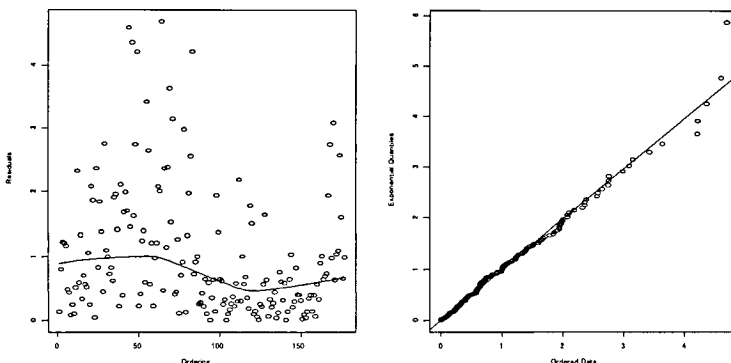
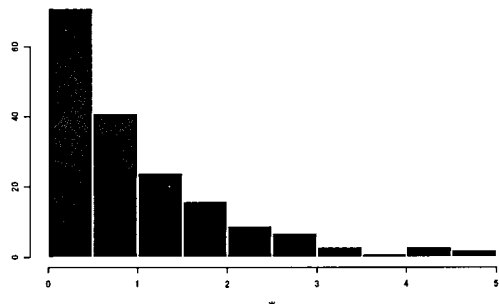


図6は変換値のヒストグラムだが、裾を除いて綺麗な指数密度関数になっていることが確認できる。

図6 W値のヒストグラム



以上の分析を用いて翌月の最大損失が、過去の最大値を超える確率を求めてみる。最初にデータ系列の中での最大月間損失を求めると、7.455となるから、推定結果を用いてフレシェ分布上でこの値を超える確率を求めると、0.00374、つまり0.374%となった。同様に、最小損失率minを求めると0.388で、翌月の最大損失がこの値以下になる確率を求めると0.916%であった。中央値medianは1.645、この値以下になる確率は52.6%であった。月間最大値ではなく、例えば年間最大値についての分析を行えば、この確率は当然高くなるが、現在のデータでは、年次観測個数は15なので、推定結果は信頼できない。

2-1 損失率分布の分位点

このように、極限分布を用いれば特定の座標値を超える確率を容易に計算することができる。同様に、パーセント点により、所与の小さな確率で

しか生じない月間最大損失率水準などを計算することができる。例えば月間最大値の系列を使い、一年に一度の割合でしか生じないような月間最大損失率水準はいくらかといった計算は

$$P(M_{20} > m) = \frac{1}{12} \quad (11)$$

を満たすmを求めればよい。mは収益水準 (return level,あるいは損失水準) と呼ばれる。一般的に(1/k)に対しては、mは(1-(1/k))パーセント点で、

$$\exp\{-(1+\xi z)^{-1/\xi}\} = 1 - \frac{1}{k} \quad (12)$$

を解けば、

$$z = \frac{m - \hat{\mu}}{\hat{\sigma}} = -\frac{1}{\xi} \{1 - (-\log(1 - \frac{1}{k}))\}^{-\xi} \quad (13)$$

となり、この式を満たすmを求める。現在の月間最大損失のデータに関して、5年に一度しか起きないような大きな損失率水準を求めると、60月に一回起きるような損失は分位点では(59/60)となり

$$P(M_{20} < m) = \frac{59}{60} \quad (14)$$

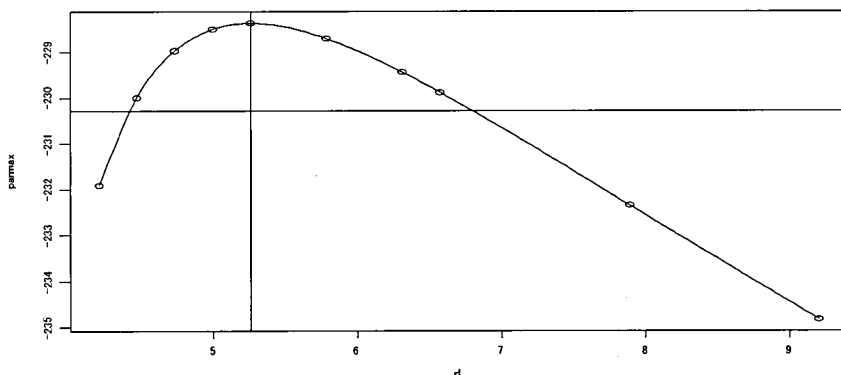
を満たすmを求める。集約尤度関数(プロファイル)を使うが、損失水準はプロファイルの頂点で、この例では5.258となり、95%信頼区間はx軸の平行線とプロファイルの交点で決まる。

日々の損失率の分布を使うと、一月に20取引日あるとして

$$F(x)^{20} = \frac{59}{60} \quad (15)$$

だから、平均すれば日々の収益(損失)率分布F(x)の0.99916という分位点、つまり99.916%点に対応している。現在の系列は3848日分の損失率で

図7 プロファイルと信頼区間

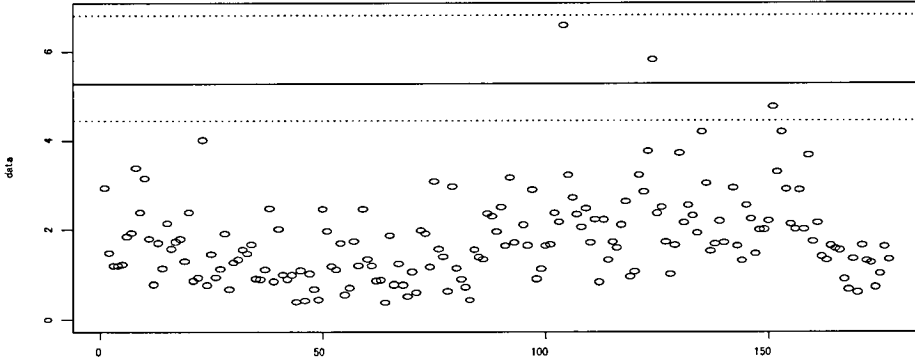


含み、各観測値が均等な確率を持つとすると、99.916%点は3845番目、大きな損失最大値から見て4番目の値になり、4.81%である。漸近分布から求まる結果5.26とはかなり違っている。

図8では全ての月間最大損失率を点で示し、5年

に一度起こりうる損失率を実線で示す。95%信頼区間の下限は4.443、上限は6.800だが、これは点線でしめしてある。信頼区間に入る観測値は3個あるが、観測期間177からして直感に合う結果になっている。

図8 月間最大損失率と5年水準



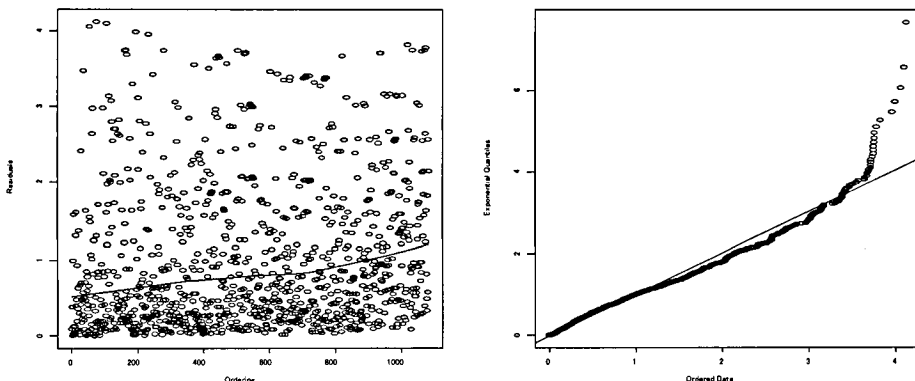
2-2 日中最大損失率データの分析

ダウ平均株価の例では日次データから月間最大値系列を計算して極値分布を使った分析を行った。原系列が最大値の観測値だけから構成されている場合でも、以上の分析を応用できる。ここで、TDKの始値tdko.d,高値dtdkh.d,安値tdkl.d,終値tdkc.dを用い日中の最大損失率を $\log(\text{tdkl.d}/\text{tdko.d})$ とし、符号を逆にして分析を行う。収益率のデータは無いが、日々の最大損失率の観測値は使えるという想定である。ただし始値と低値が同額である日の比率が高いため、損失率が0である観測値は除外し、プラス値だけの分析を行う。先の例では図2を描き、元の損失率の分布形状を検討することができたが、この計算では損失率が得られな

いため分布の形状に関する情報を得ることができないのである。観測期間は2000年1月4日以後2004年9月30日までとする。ダウ平均株価と同様の分析を繰り返すが、推定結果は ξ が0.222 (0.032)、 σ が0.00934 (0.00028)、 μ が0.0120 (0.00034)となった。括弧内は標準誤差で、係数は全て有意になる。最後に変換した残差の図を示すが、左のトレンド図および右のQQプロットから判断して、フィットが良いとは言えない。

分布の右裾で推定が上手くいかないようである。先の例と同様に過去データ中の最大値を求め、その値における確率を推定する。日次の最大損失率は0.173であり、この値を超える確率は0.0008となったが、過剰推定の可能性がある。最後に100日

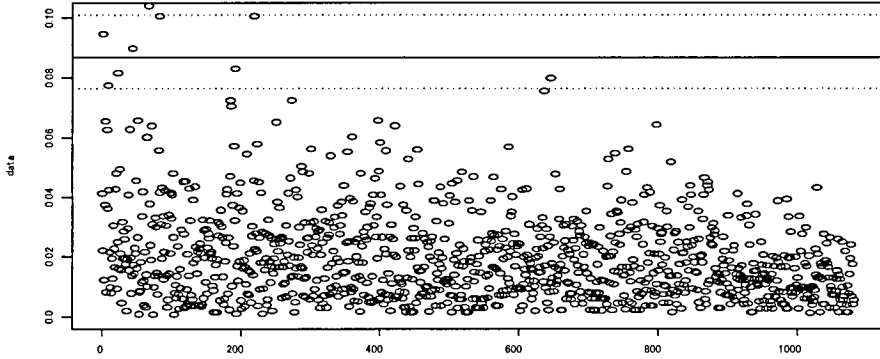
図9 残差の診断プロット



に一度起きるような損失率を求めると、0.0868と求まった。またこの値の95%信頼区間は上限が0.101、下限が0.0763だった。1082個の観測値の中

で5%はほぼ50個だが、図10の信頼区間にはほぼ8個の観測値しか含まれていない。

図10 100日水準



3 分布の裾に関する分析

利用可能なデータが最大値しかない場合は、前節の分析は有効であろう。しかし、もし最大値だけでなく収益率に関する詳細なデータが利用できるのであれば、大きな損失率の分布に関するさらに詳細な分析が可能である。この節では損失率の境界値 u を指定し、この境界値を超える損失率の分布を求める。このような分析は、POT(peaks over thresholds)と呼ばれる。もちろんthresholdは境界値 u を意味する。

3-1 平均超過関数の漸近分布

境界値 u を所与、 $y > 0$ とすると、 u を超える値の条件付き分布関数は

$$F_u(y) = P(X - u < y | X > u)$$

$$= \frac{P(u < X < y + u)}{P(u < X)} = \frac{F(y + u) - F(u)}{1 - F(u)} \quad (16)$$

となる。分布関数 $F(\cdot)$ は未知だから、再び漸近分布を求めると、 $F_u(y)$ は $\xi > 0$ なら一般化パレート分布 (GPD, generalized Pareto distribution) で、 $y > 0$ について、

$$G_{\xi, \beta}(y) = \frac{-(1 + \xi \frac{y + u - \mu}{\sigma})^{-1/\xi} + (1 + \xi \frac{u - \mu}{\sigma})^{-1/\xi}}{(1 + \xi \frac{u - \mu}{\sigma})^{-1/\xi}}$$

$$= 1 - (1 + \frac{\xi y}{\beta(u)})^{-1/\xi} \quad (17)$$

となる。ただし、分布関数をフレシェ型で

$$F(y + u) = \exp\left\{-\left(1 + \xi \frac{y + u - \mu}{\sigma}\right)^{-1/\xi}\right\}$$

などと近似し、 $\exp(-x)$ は $(1-x)$ で 1 次近似する。 $\xi = 0$ なら指数分布で、 $y > 0$ について

$$G_{\xi, \beta}(y) = 1 - \exp\left(-\frac{y}{\beta(u)}\right), \quad (18)$$

$\xi < 0$ なら GPD の第 2 型で、 $0 < y - \beta(u) / \xi$ について

$$G_{\xi, \beta}(y) = 1 - \left(1 + \frac{\xi y}{\beta(u)}\right)^{-1/\xi} \quad (19)$$

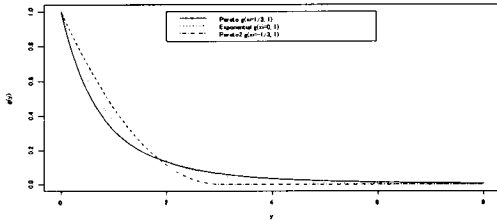
となる。ただし、 ξ は分布型を決めるので形状パラメータ、 $\beta(u)$ は正値をとる関数でスケールパラメータと呼ばれる。 ξ と $\beta(u)$ は共に未知である。

この結果は Fisher-Tippett 定理に基づいており、GPD と GEV 分布には密接な関係がある。特に、 ξ は共通である。収益率の分析で通常扱われるのは $\xi > 0$ のケースだが、GEV ではフレシェ型分布だった。GPD では、tail index ($\alpha = 1/\xi$) 以上の次数では収益率のモーメントが存在しない。これは、ボラティリティが発散することを意味している。

境界値が u_0 、スケールパラメータが $\beta(u_0)$ である GPD が与えられているなら、境界値が $u > u_0$ と変われば、スケールパラメータは $\beta(u_0) + \xi(u - u_0)$ となり線形関係を維持して変化する。したがって、 $\beta(u_0)$ と ξ が求まっていれば、新しいスケールパラメータは容易に計算できる。

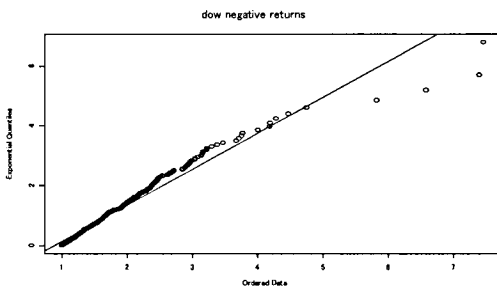
$\beta(u_0)$ は1として、 ξ が1/3、0、-1/3の3ケースについてGPDの密度関数を描いてみる。GPDの第2型では、分布は+3が上限である。

図11 一般化パレート密度関数



境界を越える損失率（-収益率）の分布がどの分布型であるかを判定する基準として、指数分布を比較基準としたQQプロットを使う。月間最大値に関して、参照分布としてゲンベルの二重指数分布を使ったのと同じ考えである。データは再びダウの収益率を用いる。境界値を1とし、指数分布と比較するためのQQプロットを描くが、このプロットは、収益率分布が指数分布からはずれていることを示しており、 ξ は0でないことが判定できる。この結果は境界値には依存しない。

図12 損失率QQプロット



3-2 スケールパラメター

境界値 u_0 を超える超過損失 $X - u_0$ の分布が、1より小さい ξ とスケールパラメター $\beta(u_0)$ のGPDで近似できるとする。GPDの性質を用いて計算すると、超過損失の期待値は理論的に

$$E(X - u_0 | X > u_0) = -\int_0^{\infty} \frac{y}{\beta(u)} \left(1 + \frac{\xi y}{\beta(u)}\right)^{-1/\xi} dy = \frac{\beta(u_0)}{1 - \xi} \quad (20)$$

となり、これを平均超過関数と呼ぶが、この期待値はスケールパラメターと ξ で決まる事が分か

る。既に述べたようにスケールパラメターは境界値に関して線形という性質があるので、 u_0 より大である任意の境界値 u に関して

$$E(X - u | X > u) = \frac{\beta(u_0) + \xi(u - u_0)}{1 - \xi} \quad (21)$$

となり、分子のスケールパラメターだけが変化する。この式を u_0 を超える超過分 y で表現すると、 $u = u_0 + y$ だから

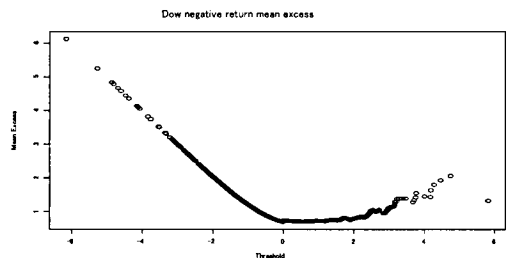
$$e(u_0 + y) = E(X - (u_0 + y) | X > u_0 + y) = \frac{\beta(u_0) + \xi y}{1 - \xi} \quad (22)$$

となる。平均超過関数の、超過分 y に関する線形関係が明らかであろう。この結果は、 u_0 を見つければ超過関数は y の一次関数になるということであるから、境界値 u_0 を見つける方法をも示している。つまり、一般的に、任意に選んだ境界値 u を超える損失 $x_{(i)}$ の平均は、境界値を超える損失の観測個数が n_u として

$$e_n(u) = \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) \quad (23)$$

によって推定できる。したがって、全ての u に関してこの超過関数を推定し、超過関数が線形になる値を求めれば u_0 の推定値となる。ダウ損失率について、平均超過関数を求めると図13のようになる。綺麗な直線にはならないが、2に近い値でプラスの直線が始まると見て良いだろう。ここでは損失率を扱っているので負の範囲は境界値として意味は持たない。プラス収益率の分布であれば、直線による近似が正確であると予想できる。

図13 ダウの平均超過関数



3-3 分布関数の裾の推定

境界値 u が十分に大きければ、境界値を超える

損失の分布はGPDを使って近似できるが、近似式を求めるために条件付き確率を再現すると

$$F_u(y) = P(X-u < y | X > u) \\ = \frac{F(y+u) - F(u)}{1 - F(u)} \approx G_{\xi, \beta}(y) \quad (24)$$

である。yはuを超える超過分だから(y+u)は損失率、(y+u)をxと表記すれば裾の確率は

$$F(x) \approx G_{\xi, \beta}(y)(1 - F(u)) + F(u) \quad (25)$$

となる。さらにF(u)は境界値までの累積確率だから観測確率で推定すると、kを境界値を超える観測個数として、(n-k)/nである。ここで、 ξ が0で無い場合の $G_{\xi, \beta}(y)$ を代入し、かつF(u)を観測確率で置き換えると、境界値を超える領域における裾の確率F(x)を1から引いた値は

$$1 - \hat{F}(x) = \frac{k}{n} \left(1 + \xi \frac{x-u}{\hat{\beta}(u)}\right)^{-1/\xi} \quad (26)$$

と推定できる。

この確率は全てのx値について標本確率によっても推定できるから、標本確率と(26)式の差異を見ることにより、推定の精度が検討できる。なお、損失率は収益率の符号を逆にした値だから、 $1 - F(x)$ は、収益率で考えたときの分布関数の左裾になっていることに注意しよう。

(16)式を使ってGPDパラメターの最尤推定を行う場合は、データとしてはuを超える超過損失 $x_{(i)}$ を用いる。漸近分布の密度関数の対数値は(17)式の一次微分より、 ξ が0でなければ

$$-\log \beta(u) - \left(1 + \frac{1}{\xi}\right) \log \left(1 + \frac{\xi(x_{(i)} - u)}{\beta(u)}\right), \quad (27)$$

となり、対数尤度関数は(27)式の $x_{(i)}$ に関する和になる。 $y_{(i)} = x_{(i)} - u$ と表記すれば、 $y_{(i)}$ の範囲は(17)式で示されている。 ξ が0なら

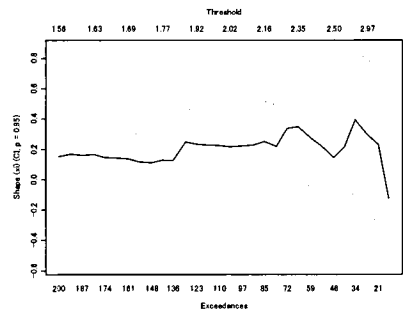
$$-\log \beta(u) - \frac{x_{(i)} - u}{\beta(u)} \quad (28)$$

である。

uを2と指定して、最尤推定を行うと、uを超える観測個数は109となった。u以下の観測値の確率は、3848分の3739だから0.972となる。係数推定値は ξ が0.24(0.12)、 β が0.59(0.09)となった。括弧内は標準誤差であり両係数は有意になり、また ξ は予想通り正值であった。

ξ の値の安定性を確認する。図では境界値を1.56(境界を越える観測値は200個)から徐々に上げていき、 ξ の値を逐次推定した結果を示す。既に求めた109個の場合の推定結果も途中で示されるが、uを2以下にした場合の ξ の推定結果が0.2に少し満たない値で安定していることが分かる。境界値が2を大きく超えると観測個数が減るために推定値は不安定になる。図中、下軸は境界を越える観測個数、上軸はuの値で、uが1.5の場合には観測個数は218である。

図14 境界値uと形状パラメター



最尤推定の結果についての診断を、図を使って行う。最初の図は、uを超える観測値に関して推定されたGPDと累積相対頻度だが、推定されたGPDは累積相対頻度に密着している。2図は、x軸y軸ともに対数座標だが、(26)式の推定結果を元にして求めた裾確率の推定値と、相対頻度のプロットで、フィットはよい。超過分の観測値を

$$res = \frac{1}{\xi} \log \left(1 + \frac{\xi(x-u)}{\beta}\right) \quad (29)$$

と変換する。3図は推定した残差値の時系列プロットであるが、残差にトレンドは見られず乱数の性質を示していると考えられる。この変換については、GPDにより

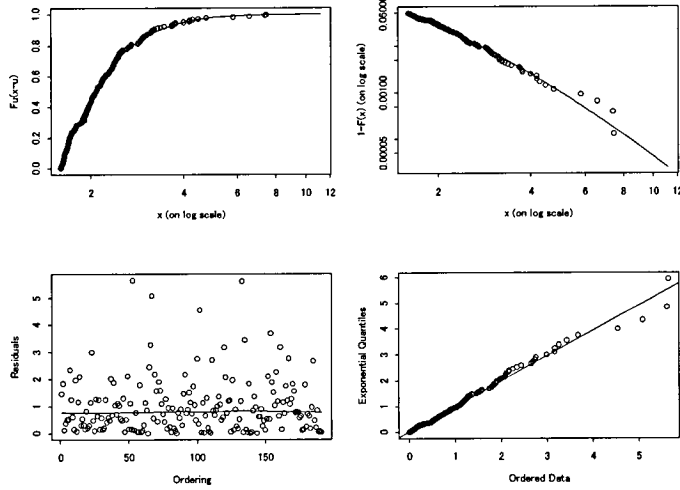
$$res = -\log(1 - G),$$

である事が分かる。既に分布関数の分布は(0,1)一様分布であるという性質を使ったが、この性質によりresの分布は

$$P\{res < z\} = P(G < 1 - e^{-z}) = 1 - e^{-z} \quad (30)$$

となり、指数分布であることが分かる。したがって、resの推定値の累積相対分布と指数分布のQQプロットを求めれば、GPDのフィットを確かめることができるが、4図はこのQQプロットである。

図15 GPD推定の診断図



4 バリュアットリスク

最大値データの分析は前節で行ったが、詳細なデータが利用できる場合は、収益率の分布を基として、バリュアットリスク (VaR) や期待損失 (expected shortfall, ES) が計算できる。損失率の分布関数を $F(x)$ とすれば、1% の VaR は 99% 点

$$\text{VaR}_{99} = F^{-1}(0.99) \quad (31)$$

で、この値を超える損失率は 1% の確率でしか起きないことを示す。収益率の分布が正規分布で近似できるとすれば、標準正規分布の 99% 点を q_{99} とし、1% の VaR は

$$\text{VaR}_{99} = \mu + \sigma \cdot q_{99} \quad (32)$$

となる。平均と標準偏差は収益率データから推定する。

ES は、VaR を超える損失の平均値である。例えば 1% の ES は

$$\text{ES}_{99} = E(X|X > \text{VaR}_{99}) \quad (33)$$

と定義され、1% の VaR を超える損失の条件付き期待値になっている。再び正規分布を仮定するならば、この条件付き期待値は $\phi(z)$ を標準正規密度関数、 $\Phi(z)$ を標準正規分布関数、 $z = (\text{VaR}_{99} - \mu) / \sigma$ とし

$$\text{ES}_{99} = \mu + \sigma \frac{\phi(z)}{1 - \Phi(z)} \quad (34)$$

と計算できる。平均と標準偏差を収益率データから推定するだけで、1% の VaR と ES を計算することができ、非常に簡単である。

ダウに関して、この方法で VaR と ES を計算すると、95% の VaR と ES は 1.62、2.05 となり、99% の

VaR と ES は 2.31 と 2.65 であった。

VaR などは分布の裾に関する情報であるから、正確な値を得るためには、正規分布による近似法ではなく極値に関する近似法を使うべきであろう。境界値 u に対して、 $F(u)$ よりも 99% が大きい値ならば、VaR も裾分布関数の近似式 (26) を使って計算されるべきであろう。 ξ が正ならば任意の高確率 q について

$$q = \hat{F}(\text{VaR}_q) = 1 - \frac{k}{n} \left(1 + \xi \frac{\text{VaR}_q - u}{\hat{\beta}(u)} \right)^{-1/\xi} \quad (35)$$

と近似できるから、100(1-q)% の VaR は係数を推定量に置き換えて

$$\text{VaR}_q = u + \frac{\beta(u)}{\xi} \left(\left(\frac{n}{k} (1-q) \right)^{-\xi} - 1 \right) \quad (36)$$

と推定できる。この VaR について ES は

$$\text{ES}_q = E(X|X > \text{VaR}_q) \quad (37)$$

だから、

$$\text{ES}_q = \text{VaR}_q + E(X - \text{VaR}_q | X > \text{VaR}_q) \quad (38)$$

となり、平均超過関数 (21) 式を使えば、

$$\begin{aligned} E(X - \text{VaR}_q | X > \text{VaR}_q) &= \frac{\beta(u) + \xi(\text{VaR}_q - u)}{1 - \xi} \end{aligned} \quad (39)$$

となる。これを前式に代入して

$$\text{ES}_q = \text{VaR}_q + \frac{\beta(u) + \xi(\text{VaR}_q - u)}{1 - \xi}$$

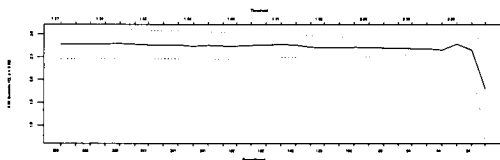
$$= \frac{\text{VaR}_q}{1-\xi} + \frac{\beta(u) - \xi u}{1-\xi} \quad (40)$$

が求まる。

ダウの例を続けると、95%のVaRとESは1.69、2.37、同じく、99%のVaRとESは2.70と3.70だった。したがって5%の確率で収益率は-1.69以下になり、-1.69以下になった場合の平均は-2.37である。さらに1%の確率で収益率は-2.37以下になり、その場合の平均は-3.7という低い値になる。ξが0.24であったから、正規分布を用いた値よりも、かなり低い収益率を示す。

最後に、境界値uを変化させ、99%VaRの安定性を検討してみる。結果は図16に示されるが、境界値が2.5以下ならばVaRはほとんどuに依存していないことが分かる。境界値は上軸に与えられている。

図16 境界値とVaR



4-1 tail-indexのヒル推定量

既に述べた分位点の推定法はGPDに基づいているが、ヒルの方法は順序統計量

$$X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(n)} \quad (41)$$

を用いて形状パラメータを直接推定するもので、特定の分布型に依存しないノンパラメトリックな推定法である。ここでk番目の統計量を取り上げると、ヒルの推定量は $X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(k)}$ を用いて、

$$\hat{\xi}(k) = \frac{1}{k} \sum_{i=1}^k (\log X_{(i)} - \log X_{(k)}) \quad (42)$$

と定義される。 $\hat{\xi}(k)$ は正值である。tail-indexの推定量は、 $\hat{\alpha} = 1/\hat{\xi}$ となる。GPDの分析では、kは境界値uを超える観測個数であったが、このkも同様の働きをする。理論的には、 $k \rightarrow \infty$ かつ $k/n \rightarrow 0$ 、さらに確率変数の極値の分布がGEVに収束するという条件の下で、 $\hat{\xi}(k)$ はξの一致推定量になる。漸近分布は正規分布、漸近分散は ξ^2/k で、t統計量の計算に用いることができる。同様に、tail-indexについても、漸近分散は α^2/k と求まる。

実際は、Tは固定しているから、特定範囲のkについて $\hat{\xi}(k)$ を全て計算して、推定値の安定性を確かめる。

分布の裾は、(4)式と同じく裾部分 $1-F(x)$ が $x^{-\alpha}$ L(x)と近似でき、かつ $x > X_{(k+1)}$ として

$$1 - \hat{F}(x) = \left(\frac{x}{X_{(k+1)}} \right)^{-\hat{\alpha}(k)} \frac{k}{T} \quad (43)$$

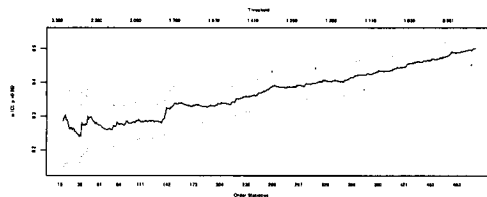
と推定される。分位点qの推定は、 $q > 1 - k/T$ として、

$$\hat{x} = X_{(k+1)} \left(\frac{1-q}{k/T} \right)^{-\hat{\xi}} \quad (44)$$

となる。99%のVaRなら、 $q=0.99$ を代入する。

ダウの損失率についてξ(k)を求めてみよう。kの範囲は下限は15上限を500とする。

図17 形状パラメータのヒル推定



95% VeR	xi	順序	境界値
2.12	0.29	120	1.94
2.12	0.29	119	1.94
2.12	0.29	118	1.95
2.12	0.29	117	1.95
2.12	0.29	116	1.95
2.12	0.28	115	1.96
2.12	0.29	114	1.97
2.12	0.29	113	1.97
2.12	0.29	112	1.97
2.12	0.29	111	1.98
2.12	0.28	110	1.99
2.12	0.28	109	2.00
2.12	0.28	108	2.00
2.13	0.28	107	2.00
2.12	0.28	106	2.01
2.12	0.29	105	2.01
2.11	0.29	104	2.02
2.12	0.29	103	2.02
2.12	0.29	102	2.03
2.12	0.29	101	2.03
2.13	0.28	100	2.04

ξ(k)と信頼区間が図示されているが、kが140位を超えるとパラメータ値は線形に増加していて安

定しているという印象は得られない。GPDを用いた推定値は0.24だったが、 k が100と140の間では似た値になっている。形状パラメーターはGPDの結果より大きな値になる。しかし境界値 u はほとんど変わらない事が分かる。(この推定での境界値とは、境界となる順序統計量 $X(k)$ である。)形状パラメーターの値をまとめて、表にする。

参考文献

Hill, B. M. (1975), "a simple general approach to inference about the tail of a distribution," *The Annals of Statistics*, 3, 1163-1174.

Embrechts, P., C. Klöppelberg, and T. Mikosch (1997), *Modelling Extremal Events*, Springer-Verlag, Berlin.

Tsay, R. S. (2001), *Analyses of Financial Time Series*, John Wiley & Sons, New York