6.1 モデル評価の基準

6.1.1 モデルの評価指標

いろいろなリスク(=VaR)の算出方法(=モデル)があったけど、結局どのモデルがよかったのだろう？となる。
ちゃんとモデルのリスク値が正確に現実のリスクをとらえられていたかを後から**評価**する必要がある。

バックテスト

観察期間が終了した後で、その期間の実際の値動きとその時々に計算したリスク値を見比べ、どのくらいの割合で想定したリスクから外れてしまう現象が生じたのかを確かめる方法。一番基本的な概念。（図6.1）

実際に図を書いてみて、モデルの評価指標をどうやって設定しようか？
→三つ紹介。

6.1.2 超過回数による検証

観察期間Nに対して、毎日の値動きx_tがその日のVaRを下回った回数をカウントする。モデルが99%信頼区間のVaRを算出するものであれば、その回数はざっくり100回に1回であってほしい。

N回中下回った回数をz_1とすると、$z_1 \sim Bin(N, p)$。モデルが正しければ$p = 0.01$のハズ。

ということで二項片側検定。

$$H_0 : p = 0.01$$
$$H_1 : p > 0.01$$

でもこれ独立なサンプルではないから、検定使ってるけどちょっと微妙...
やっぱり分布の形で評価しないとね！

6.1.3 確率プロットを利用した収益率分布の検証

観察期間に得られた収益率の系列$\{x_t\}_{t=1}^n$と、それぞれの時刻にモデルが予測した収益率の確率分布の系列$\{\mathbb{P}_t\}_{t=1}^n$があるとする。

この確率分布の系列$\{\mathbb{P}_t\}_{t=1}^n$はちゃんと観測された収益率の確率分布になってるのか？を調べる。

確率分布$\mathbb{P}_t$の累積分布関数を$F_t(x)$とする。時刻tの収益率X_t(確率変数)に対して、**もし本当に$X_t \sim \mathbb{P}_t$ならば、$F_t(X_t) \sim U(0,1)$のハズ**である。

これが正しければn個の観測に対して作成した系列$\{F_t(x_t)\}_{t=1}^n$は一様分布からのサンプリング結果となる。

小さい順に並び変えれば、大体k番目に小さい値$\dot{F}_k(\dot{x}_k)$は$(k-0.5)/n$くらいの値になる。（まあ感覚的にはわかるけど、一様分布のなんの値なんだこれ）

というわけで$\dot{F}_k(\dot{x}_k)$と$(k-0.5)/n$をそれぞれ縦軸と横軸にしてプロットしてみたのが**確率プロット**。正しく確率分布が推定できていれば45°線にのる。この45°線とプロットの乖離がモデルの評価指標になる。

蛇足

ちなみに確率プロットは本来確率分布同士の類似度を図る方法なので、別に一様分布にしなくても、分位点を並べるだけでもおｋ。
コラムにあるローレンツ曲線は、累積分布関数を比較しましょ、という話。密度関数を比較するより、離散⇔連続の分布間の評価がしやすい。（個人的な印象）
KS検定とかも分布関数で評価する統計量だし。

補足

証明：
確率変数Xの従う確率分布の累積分布関数を$F_X(x)$とする。また$Y = F_X(X)$とする。$Y \sim U(0,1)$を示す。

Yの分布関数$F_Y(y)$について、

$$
\begin{aligned}
F_Y(y) &= P[Y < y] \\
&= P[F_X(X) < y] \\
&= P[X < F_X^{-1}(y)] \\
&= \int_{-\infty}^{F_X^{-1}(y)} f_X(x)dx \\
&= F_X(F_X^{-1}(y)) - F_X(-\infty) \\
&= y
\end{aligned}
$$

よって$Y \sim U(0,1)$。

6.1.4 BIS規制対応のための評価

BIS規制で使用を許可されているリスクモデルもいくつかある。銀行としてはなるべく申請するリスクを小さく抑えたい（リスクに見合った資産を保有しなきゃならなくなり、それは避けたい。なるべくお金を融資に回したい）。

ということで観察期間nの間に算出されたVaRの平均値をそれぞれのモデルで導出して、一番小さいモデルを使おう、ってなった。

数学的な意味はない。