

金融リスク管理手法の対比較型バックテスト

川崎 能典 モデリング研究系 教授

1 問題設定

金融資産の市場リスクを管理するための統計モデルの適否は、例えばリスク尺度を Value-at-Risk (すなわち分位点、以後 VaR) と決めた上でその信頼水準を例えば 99% と決め、ボラティリティの外挿予測を繰り返した時に、経験超過比率が期待比率と合致しているかどうかの検定 (Kupiec 検定) や、超過にクラスター性がないこと、すなわち独立性の検定 (Christoffersen 検定) によって論じられる。しかし、複数のリスク管理手法が検定をパスしたら、どの方法が最も優れたリスク管理手法なのだろうか。

本報告では、Kaibuchi, Kawasaki and Stupfler (2022) で提案された GARCH-UGH 法, McNeil and Frey (2000, *J Empir Financ*) の GARCH-EVT 法, GARCH-UGH 法で GARCH パートを捨象した UGH 法の 3 者について、対比較型バックテスト (comparative backtesting) の枠組みで性能比較を行う。

2 VaR の誘導性とスコア関数

複数のリスク管理手法の優劣を論じるためには、例えば分位点の予測値 (r) と実現値 (X , 観測されれば x) との距離を測る関数 (スコア関数) が必要である。リスク尺度が満たすことが望ましい性質は、様々な立場から論じられているが、そのひとつに誘導性 (elicitability) がある。あるリスク尺度 η に対し少なくともひとつ対応するスコア関数 $S(r, X)$ が存在し、 $r = \eta(X)$ と取れば、スコア関数の期待値 $E[S(r, X)]$ を最小化できるとき、リスク尺度 η は誘導性を持つ (elicitable) と言う。

VaR 推定値を $\hat{q}_{\tau,1}, \hat{q}_{\tau,2}, \dots, \hat{q}_{\tau,N}$ と書き、損失率の実現値を x_1, x_2, \dots, x_N と書く。このときスコア関数が存在すれば、例えば N 時点で $S_{\text{VaR}}(\hat{q}_{\tau,N}, x_N)$ を実現値として計算できる。評価期間を通じて、スコア関数の値が有意に小さい方が良い予測方式と判断される。

ところで、VaR に関しては、 \mathbb{R} 上の増加関数 G を使って

$$S_{\text{VaR}}(q, x) = (1 - \tau - \mathbb{1}\{x > q\})G(q) + \mathbb{1}\{x > q\}G(x)$$

と書けるスコア関数は全て、確率 $\tau \in (0, 1)$ のもとでの VaR q_{τ} と整合的であることが知られている (Nolde and Ziegel 2017, *Ann Appl Stat*)。つまり、VaR に対応するスコア関数はクラスしか決まらず任意性がある。恣意的だが、ひとつの選択は $G(q) = q$, $G(x) = x$ と取ることである。これを 1 次同次スコア関数と呼ぶ。更に、 $G(q) = \log q$, $G(x) = \log x$ (ただし $q, x > 0$) と取った時を 0 次同次スコア関数と呼ぶ。

3 Diebold-Mariano 検定

予測の比較にしばしば用いられるのが Diebold-Mariano 検定である。

$$DM = \frac{\sqrt{N} \times \bar{d}}{\hat{\sigma}_N}, \quad \bar{d} = \frac{1}{N} \sum_{t=1}^N (S_{\text{VaR}}^1(\hat{q}_{\tau,t}, x_t) - S_{\text{VaR}}^2(\hat{q}_{\tau,t}, x_t))$$

ここで S_{VaR}^2 はベンチマークモデルのスコア関数、 S_{VaR}^1 は対抗馬のスコア関数とする。対抗馬の方が優れている (S^1 の方が小さい) と期待するならば、スコア関数の差は系統的に負になるはずである。統計量 DM は、 \bar{d} の変動性を考慮してもスコア関数の差が有意かを調べる t 統計量に他ならない。 ($\hat{\sigma}_N$ の推定には所謂 HAC 推定量を用いる。)

4 対比較型 VaR バックテスト

ここで我々は、「対抗馬は最悪でもベンチマークの性能と同等」という仮説 (これを H_0^- と書く) と、「対抗馬は良く行ってベンチマークの性能と同等」という仮説 (これを H_0^+ と書く) を考える。 H_0^- が棄却されず、 H_0^+ が棄却されるのであれば、対抗馬の方が優れていると結論できる。図 1, 図 2 で青く塗られたセルは、対比較の結果、行方向のモデルが列方向のモデルよりバックテスト結果が優れていたことを示している。なお、列

方向の 5% から 25% は極値理論を援用する際に順序統計量を大きい方から全体の何パーセント使って裾指数を推定するかを示している。赤いセルは、 H_0^- を棄却し、 H_0^+ が棄却されなかったケースで、ベンチマークモデルの方が優れている。黄色いセルは、それ以外のケースで、明確な結論が出なかったケースである。

円ポンド為替レート (2000 年 1 月 2 日 ~ 2010 年 12 月 14 日) の損失率に対し、推定ウィンドウを 1000 (営業日にして概ね 4 年) とし、1 日先 Value-at-Risk 予測を行う。推定が終わったらサイズ 1000 のウィンドウを 1 時点シフトし、モデルは再推定する。こうして 3000 日分の外挿予測を逐次的に行う。予測値の系列がモデル別に 3 本得られ、そこから得られるスコア関数の値の列に対し DM 検定を実行する。0 次同次スコア関数の設定で行った結果が図 1, 1 次同次スコア関数の設定で行った結果が図 2 である。GARCH-UGH vs. GARCH-EVT では、全 30 ケース中 GARCH-UGH の 17 勝 10 敗 3 引き分けである。用いる順序統計量の数が少なく、また信頼水準を甘めに取った時に GARCH-EVT に軍配が上がるのは、構造上自然である。

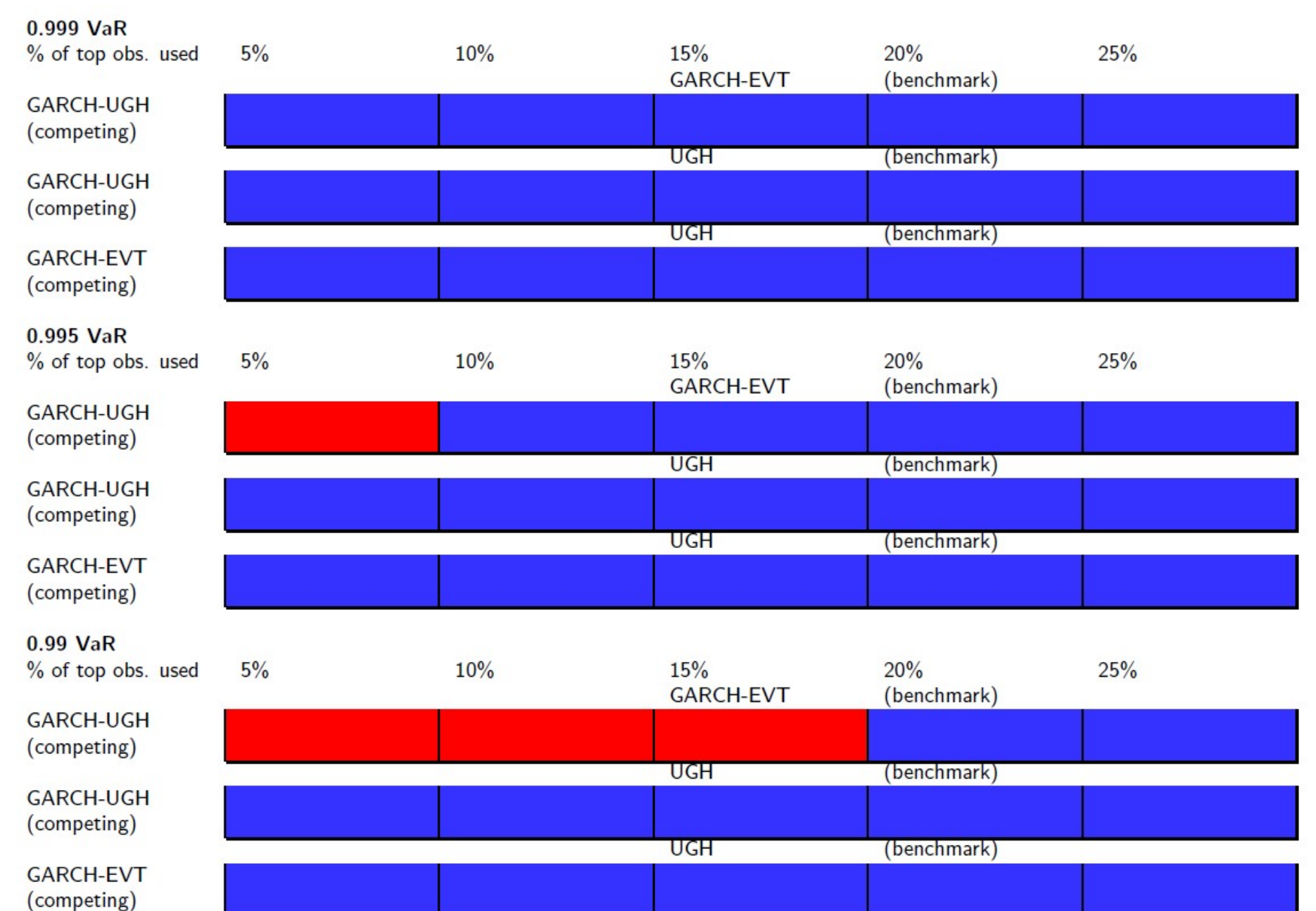


図 1: 0 次同次スコア関数のもとでの DM 検定

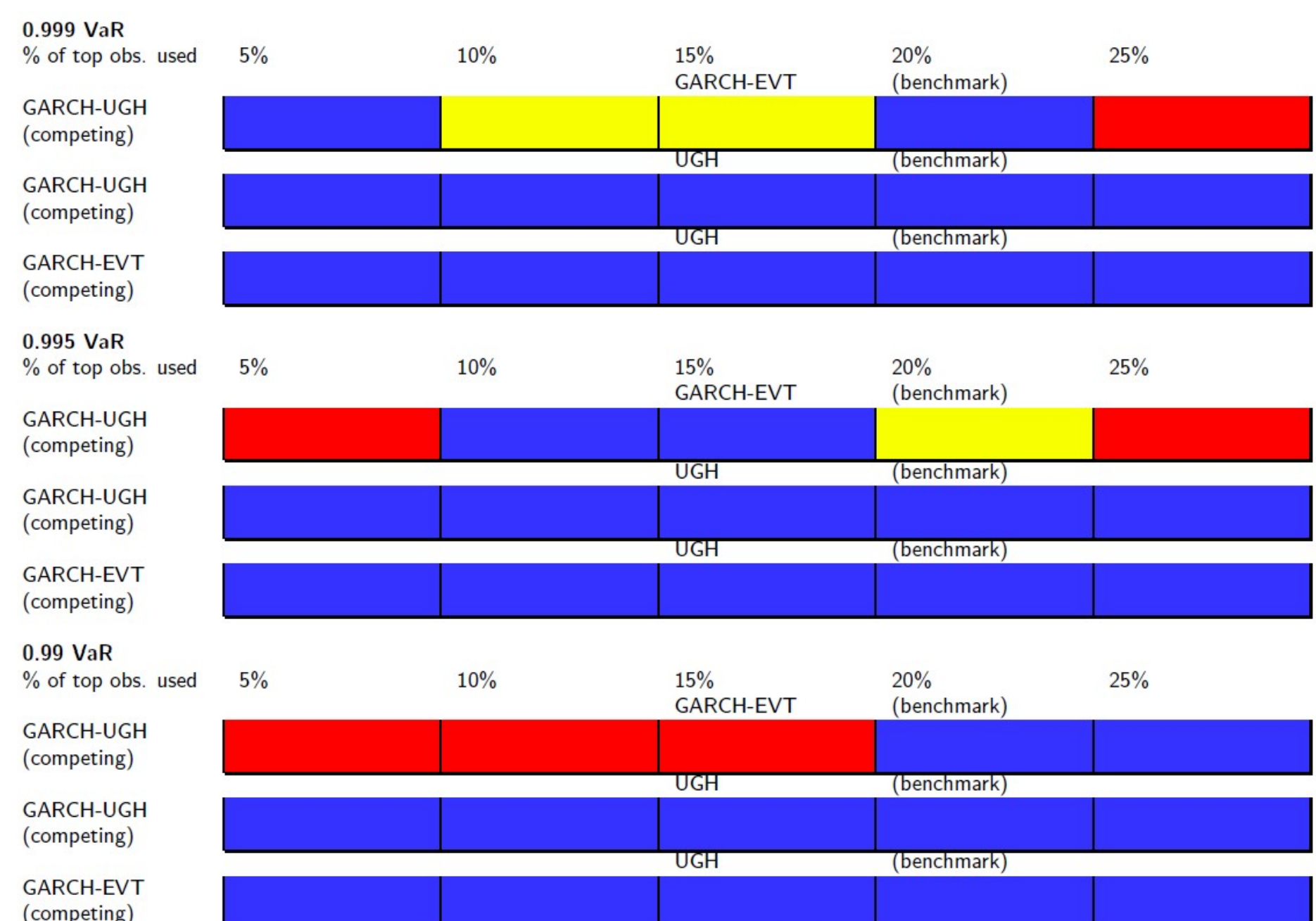


図 2: 1 次同次スコア関数のもとでの DM 検定

参考文献

Kaibuchi, H., Kawasaki, Y. and Stupfler, G. (2022), GARCH-UGH: A bias-reduced approach for dynamic extreme Value-at-Risk estimation in financial time series, *Quantitative Finance*, 22(7), 1277–1294.