アルキメデス・コピュラに基づく多変量 CVaR について

Molina Barreto Andres Mauricio
慶應義塾大学経済学部
e-mail : molina.mauricio@keio.jp

1 概要

Conditional Value at Risk (CVaR) は、金融分野で広く用いられるリスク指標であり、通常、単一のリスク要因に対して適用される。しかし近年、複数リスク要因間の依存関係を考慮した CVaR の多変量拡張が研究されている。コピュラは、これらの依存構造をモデル化する上で有用なツールである。本発表では、コピュラを用いた条件付き CVaR (CCVaR) を効率的に計算する手順を概説する。

2 既存の試み

まず、アルキメデス・コピュラの定義を示す。$\varphi : [0,1] \to [0,\infty]$ を連続かつ単調減少する凸関数 $\varphi(1)=0$ および $\varphi(0)=\infty$ を満たすものとする。その逆関数 $\varphi^{-1} : [0,\infty] \to [0,1]$ は $\varphi^{-1}(0)=1$ および $\varphi^{-1}(\infty)=0$ を満たすとする。
このとき、次元 $d \geq 2$ のアルキメデス・コピュラは次のように定義される。

$$C(u_1,\ldots,u_d) = \varphi^{-1}\left(\varphi(u_1)+\cdots+\varphi(u_d)\right),$$

この定義は $\varphi^{-1}(\cdot)$ が $[0,\infty)$ で次の意味で単調である場合に限り成り立つ。すなわち、

$$(-1)^k \frac{\partial^k}{\partial u^k}\varphi^{-1}(u) \geq 0, \quad \text{for } k=1,2,\ldots\ldots$$

前回の研究 [1] では、二変量ベクトル $\mathbf{X}=(X_1,X_2)$ の場合に Copula-based Conditional Value at Risk (CCVaR) を定義した。$\mathbf{X}=(X_1,X_2)$ を、同時分布関数 $H(x,y)=P(X_1 \leq x, X_2 \leq y)$ および周辺分布関数 $F_{X_j}(x)=P(X_j \leq x)$ $(j=1,2)$ を持つランダムベクトルとする。

定義 1 信頼水準 β $(0 \leq \beta < 1)$ における Copula-based Conditional Value at Risk $(\mathrm{CCVaR}_\beta(\mathbf{X}))$ は、次のように定義される。

$$\mathrm{CCVaR}_\beta(\mathbf{X}) = \frac{\iint_{\mathcal{U}_\beta}(\lambda F_{X_1}^{(-1)}(u)+(1-\lambda)F_{X_2}^{(-1)}(v))dC(u,v)}{\iint_{\mathcal{U}_\beta} dC(u,v)}, \tag{1}$$

ここで $0<\lambda<1$ であり、$\mathcal{U}_\beta$ は次の次のように定義する。

$$\mathcal{U}_\beta := \{(u,v)\,|\,C(u,v) \geq \beta\}.$$

また、コピュラ C がアルキメデス・コピュラである場合、この CCVaR を計算するための数式も提示した。

3 高次元への拡張

我々の目標は、前述の定義を次元が 2 以上のランダムベクトルに拡張することである。定義 1 を書き直せば、次のように表示できることに注意する。

$$\mathrm{CCVaR}_\beta(\mathbf{X}) = E^C\left[\lambda^t \mathbf{F}_{\mathbf{X}}^{(-1)} \mid \mathcal{U}_\beta\right],$$

ここで $\lambda^t = (\lambda, 1-\lambda)$ であり、$E^C[\cdot]$ はコピュラ C によって誘導される測度に関する期待値である。そこで、CCVaR の定義を次のように自然に拡張する。

定義 2　信頼水準 β $(0 \leq \beta < 1)$ における Copula-based Conditional Value at Risk $(\mathrm{CCVaR}_\beta(\mathbf{X}))$ は、次のように定義される。

$$\mathrm{CCVaR}_\beta(\mathbf{X}) = \frac{\int \cdots \int_{\mathcal{U}_\beta^{(d)}} (\lambda_1 F_{X_1}^{(-1)}(u_1) + \cdots + \lambda_d F_{X_d}^{(-1)}(u_d)) dC(u_1, \ldots, u_d)}{\int \cdots \int_{\mathcal{U}_\beta^{(d)}} dC(u_1, \ldots, u_d)}, \tag{2}$$

ここで $0 < \lambda_i < 1$ はポートフォリオ内の i 番目のリスク要因の重みを表し、$i = 1, \ldots, d$ に対して $\sum_{i=1}^d \lambda_i = 1$ を満たすものとする。また $\mathcal{U}_\beta^{(d)}$ は次のように定義する。

$$\mathcal{U}_\beta^{(d)} := \{(u_1, \ldots, u_d) \,|\, C(u_1, \ldots, u_d) \geq \beta\}.$$

4　数値計算結果

前述の定義 2 の数式の計算はコピュラの構造によっては難しいと考えられる。本講演では、ある種のアルキメデス・コピュラの族に対するいくつかの観察を提示し、CCVaR が数値ルーチンを通じて得られることを示す。また、いくつかの数値例と実証研究を行う。最後に、このリスク指標に関する研究のために取ることができるアプローチについて詳細は講演の折りに述べる。

参考文献

[1] Molina Barreto, A. M., & Ishimura, N. Remarks on a copula ‐ based conditional value at risk for the portfolio problem. Intelligent Systems in Accounting, Finance and Management, Vol. 30(3)(2023), pp. 150 – 170.

[2] AM. モリナバレート, 石村, 直之, 2023, Value at Risk の拡張とその応用: 企業研究所, 3–17 p.

Sig-Wasserstein GAN を利用した資産価格過程のシミュレーション

白井 佑介 [1]
[1] 立命館大学大学院理工学研究科基礎理工学専攻 M2
e-mail : howaitojun91@gmail.com

1　概要

本報告では，Hao Ni et al [1] による Signature を用いた資産価格過程を生成する GAN（Sig-Wasserstein GAN）の改良版として、訓練データとして，ボラティリティ過程の Signature を付加するという手法を提案し、いくつかの数値実験結果について報告する．実際の市場では直接観測できないボラティリティ過程に対しては高頻度統計的手法を用いた推定量を用いる必要があるが，本研究では Malliavin-Mancino による推定法 ([2],[3]) を使って推定する．また，モデルデータとマーケットデータの 2 種類をトレーニングデータとして使用し，それぞれの結果について報告を行う．本研究は，北野力暉氏（立命館大学）との共同研究である．

参考文献

[1] Hao Ni, Lukasz Szpruch, Marc Sabate-Vidales Baoren Xiao, Magnus Wiese, Shujian Liao, Sig-Wasserstein GANS for Time Series Generation, in ICAIF ’21: Proceedings of the Second ACM International Conference on AI in Finance Article No.: 28, Pages 1-8 (2022)

[2] Malliavin, P.,Mancino, M.E., ourier series method for measure- ment of multivariate volatilities. Finance and Stochastics 4,(2002), 49–61.

[3] Malliavin, P., Mancino, M.E., A Fourier transform method for nonparametric estimation of multivariate volatility. Annals of statisics No.4 (2009).

Ross Recovery の無限次元的拡張としての熱核法

西川 翔 [1]
[1] 立命館大学大学院理工学研究科基礎理工学専攻 M2
e-mail : sho05295290@gmail.com

1 概要

S.Ross [1] は、ある仮定を置くことで実確率の下での状態価格 (state price) が復元可能 (Recover) であることを示した。これは、連続時間で考えたときには、金融派生商品価格から実確率での「情報」がある種の過程の下で復元できる ∼ calibration できるという可能性を示唆したものととらえることができる。実際，P.Carr と J.Yu[2] は状態価格が変数分離型であるという仮定からいわゆる market price of risk が calibration 可能であることを示した。本講演ではその「状態価格にたいする仮定」を無限次元のものと理解することで，Akahori et al [3] の「熱核モデル」が得られることを示し，それを用いたモデル同定の方法とその実装について報告する．本報告は赤堀次郎氏，松永耕英氏（立命館大)，Hoang Vu 氏 (カリフォルニア大学サンタバーバラ校) との共同研究に基づく。

参考文献

[1] S. Ross, The recovery theorem, The Journal of Finance 70 (2015) 615–648.

[2] P. Carr and J-M. YU, Risk, Return, and Ross Recovery, The Journal of Derivatives 20 (2012), 38–59.

[3] J. Akahori, Y. Hishida, J. Teichmann, T. Tsuchiya, A heat kernel approach to interest rate models, JJIAM 31 (2014), 419–439

混合 LSTM ニューラルネットワークを用いた日経 225 の予測

YANG Yanbo[1], 安田 和弘 [2]
[1] 法政大学大学院総合理工学インスティテュート，[2] 法政大学理工学部経営システム工学科
e-mail：k_yasuda@hosei.ac.jp

1 概要

本講演では，機械学習の Long Short Term Memory (LSTM) ニューラルネットワークを基とした混合モデルを用いて日経 225 の翌日の終値の予測を試みる．時系列データを予測する機械学習の手法として LSTM が有効である報告が多数成されている．また，その後の発展として LSTM に他の統計的手法等を組み合わせた混合モデルも多々提案されている．本講演では混合モデルとして，まず LSTM に Empirical Mode Decomposition (EMD) を合わせたモデル (EMD-LSTM) やそれにノイズを加える形で改良した Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) を合わせたモデル (CEEMDAN-LSTM) を考える．EMD は，1998 年に Huang et al. [1] によって提案された時系列データの分解方法で，基底関数を用いずに行われるという特徴がある．そのため非定常な時系列データの分解に有効と言われている．EMD-LSTM や CEEMDAN-LSTM などを用いた時系列データの予測は近年，金融データに限らず広く用いられるようになっている．本研究では予測対象となっているデータ以外の複数のデータも用いて予測を試みる．複数のデータから主要なデータを選択するために主成分分析 （PCA） を用いて，用いるデータの選別をする．EMD と LSTM に PCA まで組み合わせたモデルになると，先行研究として Srijiranon et al. [2] があるが限られる．本講演では，これらのモデルを用いて日経 225 の翌日の終値の予測をし，モデルによる予測精度の比較を行う．

2 モデル

まず，EMD について簡単に述べておく．EMD は，元のデータを複数の Intrinsic Mode Function (IMF) と残余信号に分解する分解方法である．IMF は，次の 2 条件を満たす関数である．
鷲沢，田中 [3] より引用：

- 信号の極値の数と零交差の数が等しいか差が 1 である．すなわち，零以上の極小値や零以下の極大値を持たない．
- 任意の時刻において，極大値を結ぶ包絡線を極小値を結ぶ包絡線の平均値が零である．

$x(t)$ を元データ，$\mathrm{IMF}_i(t)$ を i 番目の IMF（$i = 1, 2, ..., n$），$r(t)$ を残余信号とすると，EMD は，

$$x(t) = \sum_{i=1}^{n} \mathrm{IMF}_i(t) + r(t)$$

といった分解になる．IMF の詳しい計算方法等は [3] などを参照してください．この分解の特徴として，IMF_i や r などがすべてデータだけから決定することがあげられる．CEEMDAN は，適宜，データに正規ノイズを加えて EMD を行う方法である（詳しくは，Torres et al. [4] などを参照）．

次に，本講演で用いるモデルの概要を述べる．

1) 予測対象のデータの他に予測に役立つと考えられるいくつかのデータを用いる．用いるデータ

表 1. 各モデルで求められた各指標の値

モデル	MAE	RMSE	MAPE	R^2
CP-LSTM	25.85764	33.68400	0.00589	0.99446
CEEMDAN-LSTM	29.60383	38.86866	0.00674	0.99258
LSTM	42.55073	55.47839	0.00972	0.98544
Gradient Boosting Regression	44.24533	58.34926	0.01014	0.98343

すべてに対して PCA を用いて主成分および各寄与率を求め，累積寄与率を求める．累積寄与率がある水準以上になるまでの主成分数を採用する．ここでは，予測対象のデータとの相互情報量が大きいデータ順に採用することとする．

2) 予測対象のデータに対して CEEMDAN により各 IMF_i および残余信号を求める．
3) 2) で求めた各 IMF_i および残余信号と 1) で採用されたデータを合わせて LSTM を用いて，各 IMF_i および残余信号の予測値を求める．
4) 3) で求めた各 IMF_i および残余信号の予測値を足し合わせることで予測対象データの予測値とする．

3　数値実験および結果

予測対象は日経 225 の翌日の終値で，用いるデータ期間は 2013 年 1 月 4 日から 2024 年 4 月 30 日の日次データである．その他のデータとして，同じ期間の始値，終値，高値，安値，取引量，値幅率，価格変化額，価格変化率を用いている．ここでは，予測の指標として，平均絶対誤差 （MAE)，二乗平均平方誤差 (RMSE)，平均絶対パーセント誤差 (MAPE)，決定係数 (R^2) の 4 種類を用いる．

PCA を用いた結果，累積寄与率が 99% を超えるのが第 4 主成分であったため，ここでは相互情報量を求め，大きい 4 つのデータとして始値，終値，高値，低値を採用する．また，終値に対して CEEMDAN で分解をし，前節のモデルで得られた予測値を用いて求められた各指標の値が表 1 の CP-LSTM の行である．また，他のいくつかのモデルで求めた結果も合わせて表 1 に記載している．

参考文献

[1] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung and H. H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, in: Proc. of the Royal Society of London. Series A: Mathematical physical and engineering science, Vol. 454(1971), pp. 903-995, 1998.

[2] K. Srijiranon, Y. Lertratanakham and T. Tanantong, A hybrid framework using PCA, EMD and LSTM methods for stock market price prediction with sentiment analysis, Applied Sciences, 12, (2022), 10823.

[3] 鷲澤嘉一，田中聡久, 経験的モード分解: チュートリアル, in: 第 22 回信号処理シンポジウム論文集, pp. 135-140, 2007.

[4] M. E. Torres, M. A. Colominas, G. Schlotthauer and P. Flandrin, A complete ensemble empirical mode decomposition with adaptive noise, in: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4144?4147, 2011.