

時系列データにおけるボラティリティ推定について

Volatility Estimation in Time Series Data

谷山 徹[†]

Toru Taniyama

1. はじめに

1.1 研究背景

ボラティリティとは金融工学の用語であり、時系列データの変動の激しさを指標である。この指標は、金融商品のリスクの度合いとして捉えるのが一般的である。つまり、ボラティリティが高い商品はリスクが高く、逆に低い商品はリスクも低いと言える。よって、ボラティリティを推定することは、リスク管理の観点で重要である。先行研究には、ARCH モデルや SV モデル、さらにはそれらを発展させた GARCH モデル、GPRSV モデル [1] などがある。本研究で提案する手法は、GPRSV モデルに潜在変数を増やし記述力を高めたものと位置付けることができる。

1.2 アプローチ

本研究では、ボラティリティのモデル化に GPDM[2] を用いる。

2. 収益率

収益率 (リターン) r_t は前期から現在への増減率を表し、 y_t を現在の値、 y_{t-1} を前期の値 (例えば株価の終値) とすると、

$$r_t = \log \frac{y_t}{y_{t-1}}, \quad (1)$$

で計算できる。本研究では、収益率の時系列データを扱っていくこととする。

3. GPDM

GPDM とは Gaussian Process Dynamical Models の略である。名前の通り、ガウス過程と動的モデルを合わせたモデルになっている。

3.1 モデル

状態方程式と出力方程式は以下のように表せる。

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}; \mathbf{A}) + \mathbf{n}_{x,t} \quad (2)$$

$$\mathbf{y}_t = g(\mathbf{x}_t; \mathbf{B}) + \mathbf{n}_{y,t} \quad (3)$$

$n_{x,t}, n_{y,t}$ はそれぞれ独立に等方的なガウス分布に従う確率変数だとする。関数 f, g は、

$$f(\mathbf{x}_{t-1}; \mathbf{A}) = \mathbf{s}_{t-1}^T \mathbf{A}, \quad (4)$$

[†] 関西学院大学大学院, Graduate School of Science and Technology, Kwansei Gakuin University

$$g(\mathbf{x}_t; \mathbf{B}) = \mathbf{r}_t^T \mathbf{B}. \quad (5)$$

ただし、

$$S_{t,i} = \phi_i(\mathbf{x}_t) \quad (6)$$

$$R_{t,j} = \psi_j(\mathbf{x}_t) \quad (7)$$

であり、 $\mathbf{s}_{t-1}, \mathbf{r}_t$ は行を抜き出したベクトルである。関数値をまとめて表記すると、

$$\mathbf{F} = \mathbf{S}\mathbf{A}, \quad (8)$$

$$\mathbf{G} = \mathbf{R}\mathbf{B} \quad (9)$$

と書ける。ここで $\mathbf{A}$ と $\mathbf{B}$ に平均 $\mathbf{0}$ の等方的なガウス分布を仮定すると、 $\mathbf{F}, \mathbf{G}$ はガウス分布の線形和となり、 $\mathbf{F}, \mathbf{G}$ 自体もガウス分布に従う。関数値の同時確率がガウス分布に従うので、ガウス過程の表記を使い、

$$f \sim \mathcal{GP}(\mathbf{0}, \mathbf{C}_x), \quad (10)$$

$$g \sim \mathcal{GP}(\mathbf{0}, \mathbf{C}_y) \quad (11)$$

と表せる。共分散行列 $\mathbf{C}_x, \mathbf{C}_y$ はグラム行列である。変数依存は、 $\mathbf{C}_x$ が $\mathbf{x}_{1:N-1}$ とカーネルのパラメータで決まり、 $\mathbf{C}_y$ が $\mathbf{x}_{1:N}$ とカーネルのパラメータで決まる。

3.2 状態変数の同時分布

$\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ の同時分布を考える。 $\mathbf{X}$ は一次マルコフ連鎖を成しているるので、超パラメータへの依存性を省略すると、以下のように記述できる。

$$p(\mathbf{X}) = p(\mathbf{x}_1) \prod_{t=2}^N p(\mathbf{x}_t | \mathbf{x}_{t-1}) \quad (12)$$

$$= p(\mathbf{x}_1) p(\mathbf{x}_{2:N}) \quad (13)$$

$p(\mathbf{x}_1)$ は、パラメータをもったガウス分布を仮定する。 $\mathbf{n}_{x,t}$ の分散パラメータを α とすると、 $p(\mathbf{x}_{2:N})$ は $\mathbf{F}$ が平均で、分散が $\alpha \mathbf{I}$ で与えられるガウス分布に従うので、 $\mathbf{F}$ を積分消去して、

$$\begin{aligned} p(\mathbf{x}_{2:N}) &= \int \mathcal{N}(\mathbf{x}_{2:N} | \mathbf{F}, \alpha \mathbf{I}) \mathcal{N}(\mathbf{F} | \mathbf{0}, \mathbf{C}_x) d\mathbf{F} \quad (14) \\ &= \mathcal{N}(\mathbf{x}_{2:N} | \mathbf{0}, \mathbf{K}_x) \quad (15) \end{aligned}$$

を得る。ここで、グラム行列 $\mathbf{K}_x$ は、 $\mathbf{x}_{1:N-1}$ に依存するため、 $p(\mathbf{x}_{2:N})$ 自体はガウス分布に従わないことに注意する。

3.3 出力変数の同時分布

続いて、 $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ の同時分布を考える。 $\mathbf{n}_{y,t}$ の分散パラメータを β とすると、 $p(\mathbf{Y}|\mathbf{X})$ は $\mathbf{G}$ が平均で、分散が $\beta\mathbf{I}$ で与えられるガウス分布に従うので、 $\mathbf{G}$ を積分消去して、

$$p(\mathbf{Y}|\mathbf{X}) = \mathcal{N}(\mathbf{Y}|\mathbf{0}, \mathbf{K}_y) \quad (16)$$

を得る。グラム行列 $\mathbf{K}_y$ は、 $\mathbf{X}$ に依存するため、 $p(\mathbf{Y}|\mathbf{X})$ はガウス分布に従う。

4. 提案モデル

収益率 r_t の平均を 0 と仮定すると、 $r_t = \sigma_t \epsilon_t$ となる。ここで、 σ_t はボラティリティであり、 ϵ_t は平均 0、分散 1 に従う分布である。現時点では、簡単のためガウス分布を仮定する。ボラティリティのモデル化に GPDM を利用する。GPDM の出力変数にボラティリティをあてはめる。潜在変数 $\mathbf{z}_t$ は、ボラティリティの状態を説明する変数と解釈できる。パラメータを省いたグラフィカルモデルは、図 1 である。

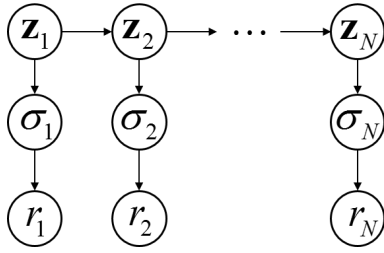


図 1: 提案モデルのグラフィカルモデル。 r が収益率、 σ がボラティリティ、 $\mathbf{z}$ が潜在変数を表す。与えられる変数は、収益率のみである。

確率変数の方程式は以下ようになる。

$$\mathbf{z}_t = f(\mathbf{z}_{t-1}) + \nu_t, \quad f \sim \mathcal{GP}(\mathbf{0}, \mathbf{C}_z) \quad (17)$$

$$\sigma_t = g(\mathbf{z}_t) + \eta_t, \quad g \sim \mathcal{GP}(\mathbf{0}, \mathbf{C}_\sigma) \quad (18)$$

$$r_t = \sigma_t \epsilon_t \quad (19)$$

ν_t, η_t は、それぞれ独立に平均 0 の等方的なガウス分布に従うとする。

4.1 同時確率

ボラティリティと潜在変数がちょうど GPDM の関係になるため、それらの同時確率は、パラメータを省略すると、以下で表現される。

$$p(\mathbf{Z}) = p(\mathbf{z}_1) \mathcal{N}(\mathbf{z}_{2:N}|\mathbf{0}, \mathbf{K}_z) \quad (20)$$

$$p(\sigma|\mathbf{Z}) = \mathcal{N}(\sigma|\mathbf{0}, \mathbf{K}_\sigma) \quad (21)$$

$$p(\mathbf{r}|\sigma) = \mathcal{N}(\mathbf{0}, \Sigma) \quad (22)$$

$\mathbf{K}_z$ のパラメータを α 、 $\mathbf{K}_\sigma$ のパラメータを β とすると、モデル全体の同時分布は、

$$p(\mathbf{r}, \sigma, \mathbf{Z}|\alpha, \beta) = p(\mathbf{r}|\sigma) p(\sigma|\mathbf{Z}, \beta) p(\mathbf{Z}|\alpha) \quad (23)$$

となる。

4.2 推定

パラメータ $\sigma, \mathbf{Z}$ と超パラメータ α, β の決定に MAP 推定を用いる。そのために、超パラメータの事前分布を以下のように与える。

$$p(\alpha) \propto \prod \frac{1}{\alpha_i} \quad (24)$$

$$p(\beta) \propto \prod \frac{1}{\beta_i} \quad (25)$$

これらより、事後確率は、

$$p(\sigma, \mathbf{Z}, \alpha, \beta|\mathbf{r}) \propto p(\mathbf{r}, \sigma, \mathbf{Z}, \alpha, \beta) \quad (26)$$

となり、これを最大化するようにパラメータと超パラメータを決定する。

5. モデル比較

提案モデルで推定したボラティリティと既存のモデルで推定したボラティリティの比較を行うため、データから計算できる指標を導入する。

5.1 Realized Volatility

これは、日次データだけでなく日中のデータ (例えば 5 分ごとの収益率) も利用する。ボラティリティの推定量である。 t 日の日中の収益率を $r_{t,i}$ とし、その数を n とすると、以下のように表せる。

$$RV_t = \sum_{i=1}^n r_{t,i}^2 \quad (27)$$

6. 今後の課題

まずは、提案モデルの学習を行うことが課題である。続いて、既存手法の学習をし、提案手法と比較、さらには予測モデルへの発展を課題としたい。

参考文献

- [1] Jianan, H. and X. Zhang (2015). Financial time series volatility analysis using Gaussian process state-space models. *Signal and Information Processing (GlobalSIP), 2015 IEEE Global Conference on*, 358-362
- [2] Wang, J. M., D. J. Fleet, and A. Hertzmann (2005). Gaussian Process Dynamical Models. *NIPS 18*, MIT Press