人工知能学会研究会資料
SIG-FIN-021

金融時系列のための深層 t 過程回帰モデル

Deep t-Process Regression Model for Financial Time-Series Analysis

中川 慧 [1*] 角屋 貴則 [2] 内山 祐介 [2,3]
Kei Nakagawa[1] Takanori Kadoya[2] Yusuke Uchiyama[2,3]

[1] 野村アセットマネジメント株式会社
[1] Nomura Asset Management Co,Ltd.
[2] 株式会社 MAZIN
[2] MAZIN,Inc.
[3] 筑波大学 システム情報系
[3] University of Tsukuba Faculty of Engineering, Information and Systems

Abstract: 深層学習をガウス過程で構築する手法が近年提案された。これは深層学習に対するカーネル関数をガウス過程の共分散関数として使用し、深層学習モデルの完全なベイズ推論を行う。それにより通常の深層学習にはない、予測の不確実性が考慮できる等のメリットが得られる。我々は、この手法を一般に裾の厚いといわれる金融データへ適用することを目的に、ガウス過程から t 過程への拡張を行う。実証分析の結果、金融時系列においてガウス過程よりも良好な結果が得られた。

1 はじめに

ガウス過程とは、確率変数の集まりであって、どの有限個の確率変数をとっても、その結合分布がガウス分布になっているようなものをいう。そして、ガウス分布が期待値と共分散行列で特定されるのと同様に、ガウス過程は期待値関数と共分散 (カーネル) 関数という、2 つの関数によって特定される。カーネル関数について、[4] によると、単一隠れ層でユニット数が無限のニューラルネットワークをガウス過程のカーネル関数として表現できることが知られている。また、隠れ層が多層の場合のニューラルネットワーク (深層学習) に対応するカーネル関数も [1] によって導出された。以上を用いて、[2] は多層のニューラルネットに対応するカーネル関数を用いて、ガウス過程としての深層学習を構築した。深層学習をガウス過程として扱うメリットとして、大きく次の 2 点が挙げられる。1 つ目が予測の不確かさを考慮することができる点である。もう 1 つが、カーネル関数を使うことで、隠れ層に無限個のユニットがある深層学習を考えていることになるため、表現力が高く、精度向上が期待できる。実際に、[2] では MNIST や CIFAR10 といった画像データを用いて上記のメリットを確認している。以上のことから、深層学習をガウス過程として表現することは有効であると言える。一方で、画像データではなく金融市場のデータに着目すると、金融資産の価格変動は一般にガウス分布ではテール部分を当てはめることができないことが、[3] 以来広く知られている。そこで我々は、[2] の深層ガウス過程を、一般に裾が厚いといわれる金融市場データへの適用を目的に、深層 t 過程へと拡張する。t 過程は、ガウス過程より裾の厚いデータを表現でき、ガウス過程の代替として注目を集めている [7]。t 過程により、外れ値に左右されにくい推定 (回帰) が可能になる。

本稿の構成は次の通りである。次章においてまず基本となるガウス過程回帰モデルを簡単に述べる。次に、提案手法を構成する、t 過程回帰モデルおよび深層学習カーネル関数を定義する。そして、t 過程回帰モデルの導出を行い、実際の金融市場のデータを用いた実証分析によって有効性を評価する。最後に結論を延べる。

2 ガウス過程回帰モデル

確率過程 $\boldsymbol{f_{GP}}$ が、入力ベクトルから成る任意の有限集合 $\boldsymbol{X} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_n]^T \subset \mathcal{R}^d$ に対して、関数値のベクトル $\boldsymbol{f_{GP}} = [f_{GP}(\boldsymbol{x}_1), ..., f_{GP}(\boldsymbol{x}_1)]^T$ の分布がガウス分布であるとき、ガウス過程という。$\boldsymbol{f_{GP}}$ の分布は、ガウス分布であるので、その期待値を指定する平均関数 $\boldsymbol{\mu} = [\mu_1(\boldsymbol{x}_1), ..., \mu_n(\boldsymbol{x}_n)]$ および、共分散 (カーネル) 関数 $k(\boldsymbol{x}_n, \boldsymbol{x}_m)$ の値 K_{nm} を要素とする共分散行列 $\boldsymbol{K}$ を指定するによって完全に特定できる。これを、$\boldsymbol{f_{GP}} \sim N(\boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x_0}))$ と書く。直感的にはガウス

*連絡先： 野村アセットマネジメント株式会社
〒 103-0027 東京都中央区日本橋 1 丁目 12-1
E-mail: kei.nak.0315@gmail.com

過程は入力 $\boldsymbol{X}$ が似ていれば、出力 $\boldsymbol{f}_{GP}$ も似ており、似ていることの定義はカーネル関数によって与えられる。

ここで、n 個の学習データ $(\boldsymbol{x}_i, y_i), \boldsymbol{x}_i \in \mathcal{R}^d, y_i \in \mathcal{R}$ が与えられたとき、$y_i = f_{GP}(\boldsymbol{x_i}) + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$ という回帰問題を考える。$\boldsymbol{f}_{GP} = [f_{GP}(\boldsymbol{x}_1), ..., f_{GP}(\boldsymbol{x}_1)]^T$ と定義し、$\boldsymbol{f}_{GP} \sim N(\boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x_0}))$ に従うとする。$\boldsymbol{X}$ が与えられた後の $\boldsymbol{y}$ の分布は、f_{GP} と ε_i の 2 つの独立なガウス分布の畳み込みなので、再びガウス過程となり、$N(\boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x_0}) + \sigma^2 \boldsymbol{I})$ となる。次に、新たな m 個の未観測のデータ $\boldsymbol{Z} = [\boldsymbol{z}_1, ..., \boldsymbol{z}_m]^T$ が与えられたときの、予測値 $\boldsymbol{f}^*_{GP} = [f_{GP}(\boldsymbol{z}_1), ..., f_{GP}(\boldsymbol{z}_m)]^T$ を考える。観測値 $\boldsymbol{y}$ と予測値 $\boldsymbol{f}^*_{GP}$ の同時分布は、$\boldsymbol{Z}$ と $\boldsymbol{X}$ の間のカーネルを考えることで、またガウス分布になるので、

$$\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{f}^*_{GP} \end{pmatrix} \sim \tag{1}$$
$$N\left(\begin{bmatrix} \boldsymbol{\mu}(\boldsymbol{X}) \\ \boldsymbol{\mu}(\boldsymbol{Z}) \end{bmatrix}, \begin{bmatrix} \boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I} & \boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{X})^T \\ \boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{X}) & \boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{Z}) \end{bmatrix}\right)$$

と書ける。したがって、予測分布は条件付きガウス分布を考えることで、次のよう書ける。

$$p(\boldsymbol{f}^*_{GP} | \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z}) \sim N(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) \tag{2}$$

ただし、

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{X})^T (\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I})^{-1} (\boldsymbol{y} - \mu(\boldsymbol{Z})) \\ &+ \mu(\boldsymbol{Z}) \qquad (3) \\ \hat{\boldsymbol{\Sigma}} &= \boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{Z}) \\ &- \boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{X})^T (\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I})^{-1} \boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{X}) \qquad (4) \end{aligned}$$

である。ガウス過程回帰は線形回帰モデルを含んでおり、線形回帰の拡張となっている。線形回帰とは異なり、カーネル法を用いることでより柔軟な関数近似が可能となり、外挿に強いなどの利点がある。

3 t 過程回帰モデル

ガウス過程と同様に、確率過程 $\boldsymbol{f}_{TP}$ が、入力の任意の有限集合 $\boldsymbol{X}$ に対して、関数値のベクトル $\boldsymbol{f}_{TP}$ の分布が多変量 t 分布であるとき、t 過程という。$\boldsymbol{f}_{TP}$ の分布は、t 分布であるので、自由度 $\nu \in R^+ \backslash [0, 2]$ と平均関数 $\boldsymbol{\mu}$ および、共分散 (カーネル) 関数を要素とする共分散行列 $\boldsymbol{K}$ を指定するによって完全に特定できる。これを、$\boldsymbol{f}_{TP} \sim T(\nu, \boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x_0}))$ と書く。ここで、$T(\nu, \boldsymbol{\mu}, \boldsymbol{K})$ は多変量 t-分布であり、確率変数 $\boldsymbol{y} \in R^n$ に対して、次の密度関数を持つ。

$$\begin{aligned} T(\nu, \boldsymbol{\mu}, \boldsymbol{K}) &= \frac{\Gamma(\frac{\nu+n}{2})}{((\nu-2)\pi)^{\frac{n}{2}} \Gamma(\frac{\nu}{2})} (\det \boldsymbol{K})^{-\frac{1}{2}} \\ &\times \left(1 + \frac{(\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{K}^{-1} (\boldsymbol{y} - \boldsymbol{\mu})}{\nu - 2}\right)^{-\frac{\nu+n}{2}} \end{aligned} \tag{5}$$

ただし、$\Gamma(\cdot)$ はガンマ関数である。容易にわかるように[1]、多変量 t 分布は $\nu \to \infty$ としたものがガウス分布となり、ガウス分布のある種の一般化となっている。

ここで、n 個の学習データ $(\boldsymbol{x}_i, y_i), \boldsymbol{x}_i \in \mathcal{R}^d, y_i \in \mathcal{R}$ が与えられたとき、$y_i = f_{TP}(\boldsymbol{x_i}) + \varepsilon_i, \varepsilon_i \sim T(\nu, 0, \sigma^2)$ という回帰問題を考える。$\boldsymbol{f}_{TP} = [f_{TP}(\boldsymbol{x}_1), ..., f_{TP}(\boldsymbol{x}_1)]^T$ と定義し、$\boldsymbol{f}_{TP} \sim T(\boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x_0}))$ に従うとする。$\boldsymbol{X}$ が与えられた後の $\boldsymbol{y}$ の分布は、f_{TP} と ε_i の 2 つの独立な t 分布の畳み込みなので、再び t 過程となり、$T(\nu, \boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x_0}) + \sigma \boldsymbol{I})$ となる。次に、新たな m 個の未観測のデータ $\boldsymbol{Z} = [\boldsymbol{z}_1, ..., \boldsymbol{z}_m]^T$ が与えられたときの、予測値 $\boldsymbol{f}^*_{TP}$ を考える。観測値 $\boldsymbol{y}$ と予測値 $\boldsymbol{f}^*_{TP}$ の同時分布は、$\boldsymbol{Z}$ と $\boldsymbol{X}$ の間のカーネルを考えることで、また t 分布になるので、

$$\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{f}^*_{GP} \end{pmatrix} \sim \tag{6}$$
$$T\left(\nu, \begin{bmatrix} \boldsymbol{\mu}(\boldsymbol{X}) \\ \boldsymbol{\mu}(\boldsymbol{Z}) \end{bmatrix}, \begin{bmatrix} \boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I} & \boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{X})^T \\ \boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{X}) & \boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{Z}) \end{bmatrix}\right)$$

と書ける。したがって、予測分布は 6 章の条件付き t 分布を考えることで、次のよう書ける。

$$p(\boldsymbol{f}^*_{TP} | \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z}) \sim T(\hat{\nu}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) \tag{7}$$

ただし、

$$\begin{aligned} \hat{\nu} &= \nu + n, \qquad (8) \\ \hat{\boldsymbol{\mu}} &= \boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{X})^T (\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I})^{-1} (\boldsymbol{y} - \mu(\boldsymbol{Z})) \\ &+ \mu(\boldsymbol{Z}) \qquad (9) \\ \hat{\Sigma} &= \frac{\nu - \beta - 2}{\nu - n - 2} \times \\ & [\boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{Z}) \qquad (10) \\ & - \boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{X})^T (\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I})^{-1} \boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{X})] \end{aligned}$$

3.1 パラメータ推定

GPR と同様に、カーネル内のハイパーパラメータと自由度を含むすべてのパラメータは、次の尤度関数を用いた最尤法によって推定することができる。入力 X、

[1] カーネル部分 $(1 + \frac{(\boldsymbol{y}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{y}-\boldsymbol{\mu})}{\nu-2})^{-\frac{\nu+n}{2}}$ に着目すると、指数関数の定義 $e^x = \lim_{n\to\infty}(1 + \frac{x}{n})^n$ から、$e^{\frac{-(\boldsymbol{y}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{y}-\boldsymbol{\mu})}{2}}$ となり、ガウス分布に収束することがわかる。

ハイパーパラメータ θ としたとき尤度関数は次のように書ける。

$$
\begin{aligned}
L &= -\log p(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\theta},v) \\
&= \frac{\nu+n}{2}\log\left(1+\frac{\delta}{\nu-2}\right)+\frac{1}{2}\log\det\boldsymbol{K_\theta} \\
&\quad +\frac{n}{2}\log\left((\nu-2)\pi\right) \\
&\quad +\log\Gamma\left(\frac{\nu}{2}\right)-\log\Gamma\left(\frac{\nu+n}{2}\right) \qquad (11)
\end{aligned}
$$

ここで $\delta=(\boldsymbol{y}-\boldsymbol{\mu})^T\boldsymbol{K_\theta}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})$ および $K_\theta=K(\boldsymbol{X},\boldsymbol{X})$ である。

4 深層学習カーネル

[2] は、$\sigma_b^2, \sigma_w^2 \in \mathcal{R}$ を与えたとき、次の式がカーネル関数として、無限のユニットを持つ全結合の深層学習と同様の表現力を持つことを示した。

$$
\begin{aligned}
&K^l(\boldsymbol{x},\boldsymbol{x}')=\sigma_b^2 \qquad (12)\\
&+\sigma_w^2F_\phi(K^{l-1}(\boldsymbol{x},\boldsymbol{x}),K^{l-1}(\boldsymbol{x},\boldsymbol{x}'),K^{l-1}(\boldsymbol{x}',\boldsymbol{x}'))
\end{aligned}
$$

ここで、F は l 層目の非線形変換 ϕ から決定的に求められる。具体的に ReLU を非線形関数として選んだ場合の再帰式は次のように書ける。この深層学習カーネルを計算し、あとは t 過程回帰の予測式にこれを代入し計算すれば、深層 t 過程回帰による予測が得られる。

$$
\begin{aligned}
K^l(\boldsymbol{x},\boldsymbol{x}')&=\sigma_b^2+\frac{\sigma_w^2}{2\pi}\sqrt{K^{l-1}(\boldsymbol{x},\boldsymbol{x})K^{l-1}(\boldsymbol{x}',\boldsymbol{x}')}\\
&\times\sin\theta^{l-1}+(\pi-\theta^{l-1})\cos\theta^{l-1}, \qquad (13)
\end{aligned}
$$

ただし、

$$
\theta^l=\cos^{-1}\left(\frac{K^l(\boldsymbol{x},\boldsymbol{x}')}{\sqrt{K^l(\boldsymbol{x},\boldsymbol{x})K^l(\boldsymbol{x}',\boldsymbol{x}')}}\right) \qquad (14)
$$

5 実証分析

ここでは、提案手法である深層 t 過程回帰モデルの有効性を確認するために、実際の金融市場のデータを用いた実証分析を行う。使用するデータとして、SP500、TOPIX、DAX 指数の月次データを使用する。各指数の配当利回り、PBR、PER、ROE、12-1 か月モメンタムを説明変数とし、翌月のリターンを非説明変数とする。データは 2018/6 末から直近 12 年分 (144 サンプル) を使用し、うち推定に 7 割 (100 サンプル) の、テストに 3 割のデータを使用する。全期間における各指数の統計量は表 5 の通りである。よく知られているように各指数ともに正規性を満たさないため、本研究のような裾の厚い分布を当てはめる余地は十分にあると言える。

	TPX	SPX	DAX
年率リターン [%]	2.37	7.38	8.15
年率リスク [%]	18.02	14.24	18.32
歪度	-0.47	-0.82	-0.56
尖度	5.22	6.20	5.69
Jarque-Beta 統計量	12.94	41.75	23.67
p-値 [%]	0.1550	0.0000	0.0007

表 1: 各指数における統計量

具体的な推定手順は下記の通りである。

1. 深層学習カーネルのハイパーパラメータ σ_w^2 と σ_b^2 の初期値を 0.1 とする
2. 観測ノイズのパラメータ σ を 0.005 で与える
3. カーネルをフィッティングし、テストデータを予測
4. ガウス過程、t 過程をそれぞれ最尤推定する

分析にあたっては、先行研究にて提案されている深層学習カーネルを用いたガウス過程 (GP) と、本研究の提案手法である深層学習カーネルを用いた t 過程 (TP) の予測誤差を RMSE と MAE それぞれで評価する。結果のサマリーが表 5 である。上段が Total、中段が予測誤差が上位 50%、下段が予測誤差が下位 50%の RMSE と MAE を記載している。また、太字は GP または TP うち精度の良い方を示している。まずは、表 5 の上段より、全ての指数において、回帰モデルとして TP は GP よりも RMSE、MAE の両方で良いことが確認できる。次に、表 5 の中段、下段は予測誤差が大きいケースと小さいケースで誤差にどの程度の差があるかを評価したものである。GP、TP ともに予測誤差が大きい場合、すなわち予測の不確実性が高い場合には、予測精度が悪くなっていることがわかる。これは、[2] と整合的な結果であり、通常の深層学習を含めた識別モデルの場合には得られない分布を仮定することによって得られる情報である。TP は予測誤差の大小に関わらず、GP よりも精度が改善している。以上から、提案手法は金融市場において回帰モデルとして [2] のモデルより有効であると言える。

6 条件付き多変量 t 分布

[7] において、多変量 t 分布の条件付き分布の導出が行われているが、カーネル部分のみであり、自由度のパラメータに不完全な部分が残る。そこで、本稿ではカーネル部分以外も含めた完全な多変量 t 分布の条件付き分布の導出を行う。

指数	TPX		SPX		DAX	
Total	GP	TP	GP	TP	GP	TP
RMSE	0.3511	**0.2158**	1.0357	**0.5076**	0.2934	**0.1134**
MAE	0.2914	**0.1622**	0.8910	**0.4282**	0.2489	**0.0808**

予測誤差 (大)	GP	TP	GP	TP	GP	TP
RMSE	0.3561	**0.2623**	1.3646	**0.6856**	0.3285	**0.1390**
MAE	0.2937	**0.2079**	1.3431	**0.6674**	0.2759	**0.1014**

予測誤差 (小)	GP	TP	GP	TP	GP	TP
RMSE	0.3458	**0.1524**	0.5650	**0.2343**	0.2553	**0.0800**
MAE	0.2892	**0.1142**	0.4594	**0.1999**	0.2231	**0.0602**

表 2: 各指数における RMSE と MAE のサマリー:上段が Total、中段が予測誤差が上位 50%、下段が予測誤差が下位 50%。太字は GP または TP のうち精度の良い方を示している。

6.1 ブロック行列に関する補題

まずは証明にあたり必要な補題を証明なしに準備する。これらの証明については、例えば [6] 中にある。ブロック行列の逆行列および行列式の補題を以下のように準備する。

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12}^T & \Lambda_{22} \end{bmatrix} = \Lambda \qquad (15)$$

ここで、

$$\Lambda_{11} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T)^{-1}, \qquad (16)$$
$$\Lambda_{22} = (\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}, \qquad (17)$$
$$\Lambda_{12} = \Lambda_{11}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^T\Lambda_{11}\Sigma_{12})^{-1} \qquad (18)$$

であり、行列式については以下の関係が成り立つ。

$$\det|\Sigma| = \frac{\det|\Sigma_{22}|}{\det|\Lambda_{11}|} \qquad (19)$$

6.2 多変量 t 分布の分割

ガウス過程の証明 [5] と同様に次元 n、平均 μ、共分散行列 Σ、自由度 ν の多変量 t 分布 X を X_1, X_2 に分割する。そして、X_2 が所与のもとでの X_1 の条件付き分布の導出を行う。X_1, X_2 のそれぞれの次元は n_1, n_2、平均ベクトル μ_1, μ_2、共分散行列を $\Sigma_{11}, \Sigma_{12}, \Sigma_{22}$、自由度を ν_1, ν_2 とする。次元および自由度については、$n_1 + n_2 = n$ と $\nu_1 + \nu_2 = \nu$ の関係が成立する。すなわち、

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{bmatrix} \qquad (20)$$

である。以上を用いて記号の簡便化のためマハラノビス距離を次のように定義する。

$$\Delta = (X-\mu)^T\Lambda(X-\mu) = \Delta_1^* + \Delta_2, \qquad (21)$$
$$\Delta_1^* = (X_1 - \mu_{1|2})^T\Lambda_{11}(X_1 - \mu_{1|2}), \qquad (22)$$
$$\mu_{1|2} = \mu_1 - \Lambda_{11}^{-1}\Lambda_{12}(X_2 - \mu_2), \qquad (23)$$
$$\Delta_2 = (X_2 - \mu_2)^T\Lambda_{22}(X_2 - \mu_2) \qquad (24)$$

さて、X_2 が所与のもとでの X_1 の条件付き分布は、多変量 t 分布の定義から、

$$\begin{aligned} p(X_1|X_2) &= \frac{p(X_1, X_2)}{p(X_2)} \\ &= \frac{\Gamma(\frac{\nu+n}{2})(\det\Sigma)^{-\frac{1}{2}}}{((\nu-2)\pi)^{\frac{n}{2}}\Gamma(\frac{\nu}{2})}\left(1 + \frac{\Delta}{\nu-2}\right)^{-\frac{\nu+n}{2}} \\ &\times \frac{((\nu-2)\pi)^{\frac{n_2}{2}}\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+n_2}{2})(\det\Sigma_{22})^{-\frac{1}{2}}}\left(1 + \frac{\Delta_2}{\nu-2}\right)^{\frac{\nu+n_2}{2}} \end{aligned} \qquad (25)$$

と書ける。まずは $p(X_1, X_2)$ と $p(X_2)$ の規格化定数に着目すると、

$$= \frac{\Gamma(\frac{\nu+n}{2})(\det\Sigma)^{-\frac{1}{2}}}{((\nu-2)\pi)^{\frac{n}{2}}\Gamma(\frac{\nu}{2})}\frac{((\nu-2)\pi)^{\frac{n_2}{2}}\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+n_2}{2})(\det\Sigma_{22})^{-\frac{1}{2}}} \qquad (26)$$

$$= \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu+n_2}{2})}\frac{(\det\Lambda_{11})^{\frac{1}{2}}}{((\nu-2)\pi)^{\frac{n_1}{2}}} \qquad (27)$$

と変形できる。次に $\times$ の後ろのカーネル部分に着目すると、

$$\left(1+\frac{\Delta}{\nu-2}\right)^{-\frac{\nu+n}{2}}\left(1+\frac{\Delta_2}{\nu-2}\right)^{\frac{\nu+n_2}{2}}$$
$$=\left(1+\frac{\Delta_2}{\nu-2}\right)^{-\frac{\nu+n}{2}}\left(1+\frac{\Delta_1}{\nu-2+\Delta_2}\right)^{-\frac{\nu+n}{2}}$$
$$\times\left(1+\frac{\Delta_2}{\nu-2}\right)^{\frac{\nu+n_2}{2}} \tag{28}$$
$$=\left(1+\frac{\Delta_2}{\nu-2}\right)^{-\frac{\nu+n_1}{2}}\left(1+\frac{\Delta_1}{\nu-2+\Delta_2}\right)^{-\frac{\nu+n}{2}} \tag{29}$$

と変形できる。ここで両者を合わせて、さらに整理すると、

$$\frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\frac{\nu+n_2}{2}\right)}\frac{(\det\Lambda_{11})^{\frac{1}{2}}}{((\nu-2)\pi)^{\frac{n_1}{2}}}$$
$$\times\left(1+\frac{\Delta_2}{\nu-2}\right)^{-\frac{\nu+n_1}{2}}\left(1+\frac{\Delta_1}{\nu-2+\Delta_2}\right)^{-\frac{\nu+n}{2}} \tag{30}$$
$$=\frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\frac{\nu+n_2}{2}\right)}\frac{(\det\Lambda_{11})^{\frac{1}{2}}}{(((\nu-2)+\Delta_2)\pi)^{\frac{n_1}{2}}}$$
$$\times\left(1+\frac{\Delta_1}{\nu-2+\Delta_2}\right)^{-\frac{\nu+n}{2}} \tag{31}$$

となる。これは多変量 Student t-分布の密度関数の形から、$\alpha=\nu+n_2$、$\beta=\nu+\Delta_2$ とすると、自由度 $\nu+n_2$, 平均 $\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(X_2-\mu_2)$, 共分散 $\frac{\beta}{\alpha}\Lambda_{11}=\frac{\nu+(x_2-\mu_2)^T\Sigma_{22}^{-1}(x_2-\mu_2)}{\nu+n_2}(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T)$ の多変量 t 分布であることがわかる。

7　まとめ

本稿では、一般に裾の厚い金融市場のデータへの応用を踏まえ、深層 t 過程回帰モデルの提案を行った。本稿の貢献は次の通り。

- t 過程回帰モデルに必要となる多変量条件付き t 分布の完全な導出
- 実証分析により深層 t 過程回帰モデルの有効性の確認
- 深層 t 過程の予測の不確実性と予測精度の考察

今後の展望として、理論的にはガウス過程、t 過程を包含するより広いクラスである加法過程による回帰モデルなどへの拡張の余地などを検討すること、また実証的には、伝統的に回帰モデルによって構築されるマルチファクターモデルに対して適用することで、手法の安定性を検証することが挙げられる。

参考文献

[1] Youngmin Cho and Lawrence K Saul. Kernel methods for deep learning. In *Advances in neural information processing systems*, pp. 342–350, 2009.

[2] Jaehoon Lee, Jascha Sohl-dickstein, Jeffrey Pennington, Roman Novak, Sam Schoenholz, and Yasaman Bahri. Deep neural networks as gaussian processes. In *International Conference on Learning Representations*, 2018.

[3] Benoit B Mandelbrot. The variation of certain speculative prices. In *Fractals and scaling in finance*, pp. 371–418. Springer, 1997.

[4] Radford M Neal. Priors for infinite networks. In *Bayesian Learning for Neural Networks*, pp. 29–53. Springer, 1996.

[5] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Advanced lectures on machine learning*, pp. 63–71. Springer, 2004.

[6] James R Schott. *Matrix analysis for statistics.* John Wiley & Sons, 2016.

[7] Amar Shah, Andrew Wilson, and Zoubin Ghahramani. Student-t processes as alternatives to gaussian processes. In *Artificial Intelligence and Statistics*, pp. 877–885, 2014.