特別研究報告書

# データの過少によるリスクを考慮した資産配分について

指導教員　　　山下信雄　准教授

京都大学工学部情報学科
数理工学コース
平成 22 年 4 月入学

西崎　公彦

平成 26 年 1 月 31 日提出

# 摘要

本報告書では, 経済状況が与えられたもとで, 資産の収益率の条件つき確率分布を推定し, その分布に基づいた資産運用を考える.

資産配分問題に関する研究はこれまでに多くなされてきたが, マーコビッツ の平均・分散モデルをはじめ, それらの研究では収益率の確率分布を既知であるとして扱っていることが多い. 一方で, 収益率の確率分布を推定する手法が数多く提案されている. その一つとしてヒストリカルデータを用いるものがある. しかし, 単純なヒストリカルデータを用いた場合, 経済状況に応じた推定ができない. また, マルチファクターモデルのように経済指標の線形回帰によって計算された収益率を用いるモデルも提案されている. しかし, ある経済指標に関連したデータが少ない場合には, 単純な線形回帰では, そのデータの過小性をリスクとして組み込むことができない. そのため, リーマンショックなどのように過去と異なる状況が起きた場合には投資判断を誤ってしまう可能性がある.

本報告書ではデータの過少性をリスクとして考えるモデルを考える. そのために過去のデータを用いてガウス過程によって条件つき確率分布を推定する. さらにその分布に基づいて収益率の条件つき分散を計算する. このように計算された分散は, 一般に, 条件となる経済指標と近いデータが過去に少ないとき大きな値をとる. 逆に, データが多いときには, そのデータにおける収益率の変動にかかわらず, 小さい値をとってしまうことがある. そこで, データが多いところの分散を表すために, $\epsilon$-近傍カーネル密度推定法で導いた分散も考える. 提案モデルではこの二種類の分散の和をリスクとして扱う. さらに, ガウス過程によって得られた期待収益率と提案リスクを用いた平均分散モデルを考え, そのモデルによって資産配分を決定する資産運用のシミュレーションを行った. その結果から提案モデルの妥当性を考察する.

# 目次

# 1 序論

近年, 情報通信技術の発達により, 投資家にとってデータ解析や最適化などの数理的手法を用いた資産運用が欠かせない状況になっている. 数理的手法を用いて最適な資産配分を求めることを資産配分問題と呼ぶ [2].

資産配分問題を定式化するためにこれまで多くの研究がなされてきたが, 最も基本とされるモデルとして H. Markowitz の平均・分散モデルが挙げられる [3]. これは, 資産のリスクとリターンを表す指標として, 分散と期待収益率を用いて資産配分を行う手法である. 他にも様々な手法が提案されているが, これらは平均・分散モデルと同様に収益率の確率分布を既知としていることが多い [2]. そのため, このような確率分布が正確に推定されていないとモデルによる最適なポートフォリオは意味を成さない.

これまでに数多くの確率モデルを推定する手法が提案されている. 最もよく用いられる手法にヒストリカルデータを用いるものがある [2]. これは過去のデータを用いて, 市場の定常性を仮定した上で, 資産の期待収益率や分散をそのデータのサンプル平均, サンプル分散を用いて計算する, しかしながら単純なヒストリカルデータを用いると経済状況に応じた推定ができないという問題点がある. 経済状況に応じた確率分布, つまりある経済状況の下で具体的な確率分布が分かればよりよい資産運用ができるはずである.

本報告書では経済状況が与えられたもとでの条件つき確率分布を推定し, その分布に計算された期待収益率と分散に基づく平均分散モデルを考える.

経済状況に応じてリスクやリターンを推定するモデルにはマルチファクターモデル [4] や資産配分関数を構築するモデル [7,8] が提案されている. マルチファクターモデルは線形回帰分析を用いて資産の収益率を推定する. 一方, 資産配分関数モデルではカーネルを用いてポートフォリオを経済指標の非線形な関数として表す. この関数の構築にはヒストリカルデータと機械学習の手法が用いられている. しかしながらこれらの手法では, 投資時点の経済指標に近いデータが少なくとも, その他のデータにの補完によってリスクが推定されてしまう. そのため, 実際のリスクと異なることがある. 例えば, 過去数年にわたり円高が続いていたとき, それらのデータに基づいて計算された資産の確率分布は, 円安のときには適用できないはずである. しかし上記の手法ではなんらかのリスクが計算され, それに基づいてポートフォリオが決定されてしまう. 実際にはデータが少ない状況ではリスクが高いと考えるべきである.

ガウス過程によって推定された条件つき確率分布では, データが少ないとき分散が大きくなる傾向がある. ここで, ガウス過程とは, 確率変数が生成される過程において, その中からどの有限列を選んだとし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のことを言う. 確率過程を用いることで為替や株価の変動を数学的にモデル化することができることが

知られている

本報告書では, データの過少性をリスクとして考えるモデルを考える. まず, 過去のデータを用いてガウス過程によって条件つき確率分布を推定し, そこから条件つき収益率の分散を計算する. このように計算された分散は, 一般に, 条件となる経済指標と近いデータが過去に少ないとき大きな値をとる. 逆に, データが多いときにはそのデータの収益率の変動しか分からず小さい値をとってしまう. そこでデータの多いところでも分散の値が提案モデルに反映できるように $\epsilon$-近傍カーネル密度推定法で導いた分散も使う. 提案モデルではこの 2 種類の分散の和をリスクとして扱うことにする. さらに, ガウス過程によって得られた期待収益率と提案リスクを用いた平均分散モデルによって資産配分を決定する資産運用のシミュレーションを行い, その結果について考察する.

本報告書の構成を以下に示す. まず 2 節で準備として, ポートフォリオ理論について述べ, 条件つき確率推定, ガウス過程回帰によって得られるそれぞれのリスクとリターンについて考察する. 3 節では提案するモデルを示す. 4 節では過去の実データを用いた提案モデルのシミュレーション結果を与え, その考察を述べる.

# 2 準備

この節では本報告書の基礎となる事項を述べる.

## 2.1 ポートフォリオ最適化問題

ポートフォリオ理論とは, リスクを最小化しリターンを最大化する為には, 所持している資産をどのように配分したらよいかを決定する理論である. リスクの尺度としては収益率のばらつきを表す標準偏差が用いられ, リターンの尺度としては期待収益率を用いることが普通である.

以下では, $n$ 種の資産に対して, リターンを考慮した上で, リスクを最少化するような資産配分を考える. ただし, 取引コストについては考慮しないものとする.

また, 資産 $i$ に対する単位あたりの現在価格を $S_i$ とし, 1 期後の価格を $S_{i+1}$ とした場合, 収益率 $t_i$ を, 以下のように定義する.

$$t_i = \frac{S_{i+1} - S_i}{S_i}$$

ここで $S_{i+1}$ と $t_i$ は確率変数であることに注意する. 以上の式により, ある期間後の資産は, 期間前の資産を 1 とした場合 $1 + t_i$ と表せる. いま資産 $i$ への投資配分を $\mathrm{v}_i$ とし,

$\mathbf{v} = (\mathrm{v}_1, ..., \mathrm{v}_n)^{\mathbf{T}}$ をポートフォリオと呼ぶ. ポートフォリオ $\mathbf{v}$ は以下の式を満たすとする.

$$\Sigma \mathrm{v}_i = 1, \mathrm{v}_i \geq 0, i = 1, ..., n$$

$T_i$ を用いるとポートフォリオの収益率 $R(\mathbf{v})$ は

$$R(\mathbf{v}) = \Sigma t_i \mathrm{v}_i$$

と書け, これも確率変数となる. ポートフォリオの収益率の分散をリスク, 期待値をリターンとしたマーコビッツの資産配分問題は

$$\begin{array}{cc} \text{minimize} & \mathbf{V}(R(\mathbf{v})) \\ \text{subject to} & \mathbf{E}(R(\mathbf{v})) \geq \overline{r} \\ & \Sigma \mathrm{v}_i = 1, \mathrm{v}_i \geq 0, (i = 1, ..., n) \end{array}$$

と書ける. ここで $\mathbf{V}[\,\cdot\,]$ は分散, $\mathbf{E}[\,\cdot\,]$ は期待値を表し, また $\overline{r}$ は投資家の要望する期待収益率である.

ここでもし, $\mathbf{E}(R(\mathbf{v}))$ が $\overline{r}$ を超えるような $\mathbf{v}$ が存在しなければこの問題の解は存在しない. そのようときは

$$\begin{array}{cc} \text{minimize} & \mathbf{V}(R(\mathbf{v})) - \tau \mathbf{E}(R(\mathbf{v})) \\ \text{subject to} & \Sigma \mathrm{v}_i = 1, \mathrm{v}_i \geq 0, (i = 1, ..., n) \end{array}$$

を考えればよい. ここで $\tau$ はリスクとリターンの配分を与えるパラメータである.

分散や期待値の計算には, 資産の収益率の確率分布が既知である必要がある. 本報告書では, ある経済指標 x が与えられたあとの条件つき確率分布を推定することを考える. そこで本報告書では以下の変数を用いる.

- $\mathbf{x}^i \in \mathbb{R}^{\mathbf{M}}$:過去の入力データ
- $\mathbf{t}^i \in \mathbb{R}$:過去の出力データ
- $\mathrm{M} \in \mathbb{R}$:過去のデータ組の個数
- $\mathrm{x}^i \in \mathbb{R}^{\mathbf{N}}, (i = 1, 2, ..., M)$
- $\mathrm{t}^i \in \mathbb{R}, (i = 1, 2, ..., M)$

## 2.2 投資における条件つき確率分布推定

まず条件つき確率についての定義を与える.

**定義 2.2.1** 事象 $A = a, B = b$ に対して, $P(A = a|B = b)$ とは事象 $B = b$ が起こった上での事象 $A = a$ が起こる確率である. また, 以下の関係式を満たす.

$$p(A = a|B = b) = \frac{p(A = a, B = b)}{p(B = b)}$$

投資においては経済指標などの情報が与えられたとき, ある資産の収益率が確率分布が分かれば, その情報もないとき, つまり $p(\mathrm{t})$ のみに基づいて投資するよりもよい成果があげられるはずである. そこで本節では既存の条件つき確率分布を推定する手法を紹介する.

条件つき確率分布の推定にはさまざまな手法が提案されているが, 以下では $\epsilon$-近傍カーネル密度推定を説明する. まず, $M$ 個の入出力データ $\{(\mathrm{x}^{\mathrm{i}}, \mathrm{t}^{\mathrm{i}})\}$ が与えられたとする. そのとき入力 x が与えられたときの t の条件つき分布は x の近傍にあるデータのみを用いて推定する. 今, $\epsilon$ が与えられたとき x の近傍にあるデータを

$$S(\mathrm{x}) = \{i | \|\mathrm{x}^i - \mathrm{x}\| < \epsilon\}$$

と定義する. これらのデータを用いて, 次の密度関数によって条件つき確率密度関数を与えるのが $\epsilon -$ 近傍カーネル密度推定法 である.

$$P(\mathrm{t}|\mathrm{x}) = \frac{1}{|S(\mathrm{x})|} \sum_{i \in S(\mathrm{x})} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left( \frac{\|\mathrm{t}^i - \mathrm{t}\|^2}{\sigma^2} \right) \right\}$$

ただし $\sigma$ は与えられたパラメータであり, 実際の分散よりは小さい値である. この密度関数を用いて, t の条件付期待値 $m_i(\mathrm{x})$, 条件つき分散 $\sigma(\mathrm{x})^2$ の分散は以下のように計算できる.

$$\begin{aligned}
m(\mathrm{x}) &= \frac{\sum_{i \in S(\mathrm{x})} \mathrm{t}^i}{|S(\mathrm{x})|} \\
\sigma(\mathrm{x})^2 &= \frac{1}{|S(\mathrm{x})|} \sum_{i \in S(\mathrm{x})} \int (\mathrm{t} - m(x))^2 \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{(\mathrm{t} - \mathrm{t}_i)^2}{2\sigma^2} \right) dt \\
&= \sigma^2 + \frac{1}{|S(\mathrm{x})|} \sum_{i \in S(\mathrm{x})} (\mathrm{t}_i - m(\mathrm{x}))^2 \qquad (1)
\end{aligned}$$

実際に $M = 200$, $h = 200$, x を為替の変化率,t を TOPIX の収益率として, 為替の変化率を横軸, TOPIX の収益率を縦軸とした場合の条件つき確率推定を用いた期待収益率とリスクのグラフを図 1 に表す. 赤線は平均を $m(\mathbf{x})$ としたとき $m(\mathbf{x}) \pm \sigma(\mathbf{x})$ を表している.

ここで $S(\mathrm{x}) = \emptyset$, になったり $|S(\mathrm{x})| = 1$ になるときがあり, $S(\mathrm{x}) = \emptyset$ なら $m(\mathrm{x})$ や $\sigma(\mathrm{x})^2$ は計算できない. また $|S(\mathrm{x})| = 1$ のときは,

$$\sigma(\mathrm{x})^2 = \sigma^2$$

となるので, データに依存しないで決まってしまう. つまり, データが過少であるとその分散は計算できないか小さく見積もられてしまう. 実際には x の近傍にはデータが少ないときには, そのデータに基づいて計算される. 分散は信頼できない. 資産配分においては, たとえ分散が小さくてもリスクが高いと考えるべきである.

## 2.3 ガウス過程回帰

以下ではガウス過程を用いて条件つき確率分布の推定を紹介する. 再び $M$ 個の入出力データ組 $(\mathrm{x}^i, \mathrm{t}^i)$ が与えられたとする. ここで $\mathrm{x}^i$ は時刻 $i$ の経済指標, $\mathrm{t}^i$ は時刻 $i+1$ の収益率だと考えればよい. このデータ組から収益率の予測関数 $y(\mathrm{x})$ を求める場合を考える. このとき予測関数のモデルとして

$$\mathrm{y} = \mathbf{w}^{\mathrm{T}}\phi(\mathrm{x}) \tag{2}$$

とする. ここで $\phi(\mathrm{x}) \in \mathbb{R}^{\mathbf{D}}$ は基底関数ベクトル,$\mathbf{w} \in \mathbb{R}^{\mathbf{D}}$ はパラメータベクトルである.

ここで $\mathbf{w}$ を確率変数と考え, その確率分布として

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I}_D)$$

を導入する. ここで $\alpha$ は分布の精度パラメータである. ここで入力データ $\mathrm{x}^i$ に対して出力を $\mathrm{y}^i$ とすると.(2) 式より

$$\mathrm{y} = \mathbf{w}^{\mathrm{T}}\phi(\mathrm{x}^i), i = 1, ..., M$$

となる. ここで $\mathbf{y} = (\mathrm{y}^1, ..., \mathrm{y}^M)^{\mathrm{T}}, \phi(\mathrm{x}) = (\phi_1(\mathrm{x}), ..., \phi_D(\mathrm{x}))^{\mathrm{T}}$ とすると（2）は,

$$\begin{aligned} \mathrm{y} &= \Phi\mathrm{w} \\ &= \begin{pmatrix} \phi_1(\mathrm{x}^1) & \phi_2(\mathrm{x}^1) & \dots & \phi_D(\mathrm{x}^1) \\ \phi_1(\mathrm{x}^2) & \phi_2(\mathrm{x}^2) & \dots & \phi_D(\mathrm{x}^2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathrm{x}^M) & \phi_2(\mathrm{x}^M) & \dots & \phi_D(\mathrm{x}^M) \end{pmatrix} \begin{pmatrix} \mathrm{w}_1 \\ \mathrm{w}_2 \\ \vdots \\ \mathrm{w}_D \end{pmatrix} \end{aligned}$$

となる. 一般には $\mathbf{y}$ を正確に観測することができない. 資産におけるときも x 以外の不確実要素によって実際の収益率は変わってしまうはずである.

ここで $\mathbf{w}$ がガウス分布に従うことから, ベクトル $\mathbf{y}$ もガウス分布に従う. 実際, $\mathbf{y}$ の期待値 $\mathbf{E}[\mathbf{y}]$ と共分散行列 $\mathbf{Cov}[\mathbf{y}]$ は,

$$\mathbf{E}[\mathbf{y}] = \mathbf{0} \tag{3}$$

$$\mathbf{C}ov[\mathbf{y}] = \frac{1}{\alpha}\Phi\Phi^{\mathrm{T}} = \begin{pmatrix} \frac{1}{\alpha}\phi(\mathrm{x}^1)^{\mathrm{T}}\phi(\mathrm{x}^1) & \frac{1}{\alpha}\phi(\mathrm{x}^1)^{\mathrm{T}}\phi(\mathrm{x}^2) & \dots & \frac{1}{\alpha}\phi(\mathrm{x}^1)^{\mathrm{T}}\phi(\mathrm{x}^M) \\ \frac{1}{\alpha}\phi(\mathrm{x}^2)^{\mathrm{T}}\phi(\mathrm{x}^1) & \frac{1}{\alpha}\phi(\mathrm{x}^2)^{\mathrm{T}}\phi(\mathrm{x}^2) & \dots & \frac{1}{\alpha}\phi(\mathrm{x}^2)^{\mathrm{T}}\phi(\mathrm{x}^M) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\alpha}\phi(\mathrm{x}^M)^{\mathrm{T}}\phi(\mathrm{x}^1) & \frac{1}{\alpha}\phi(\mathrm{x}^M)^{\mathrm{T}}\phi(\mathrm{x}^2) & \dots & \frac{1}{\alpha}\phi(\mathrm{x}^M)^{\mathrm{T}}\phi(\mathrm{x}^M) \end{pmatrix} \tag{4}$$

となる. さらに $k(\mathrm{x}, \mathrm{z}) = \phi(\mathrm{x})^{\mathrm{T}}\phi(\mathrm{z})$ とする. このとき関数 $k$ はカーネル関数となる. 以下では $\mathbf{K} \in \mathbb{R}^{\mathbf{M}\times\mathbf{M}}$ を

$$K_{ij} = \frac{1}{\alpha}\phi(\mathrm{x}^i)^{\mathrm{T}}\phi(\mathrm{x}^j), (i, j = 1, ..., M)$$

を要素とする行列とする.

式 (3) と (4) より, 確率密度分布は次の多変量正規分布になる.

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K})$$

いま, 実際の観測された出力データ $\mathbf{t}^M = (\mathrm{t}^1, ..., \mathrm{t}^M)^{\mathrm{T}}$ は $\mathbf{y}$ にガウスノイズ $\epsilon \in \mathbb{R}^M$ が含まれていると仮定する. すなわち,

$$\mathbf{t}^M = \mathbf{y} + \epsilon$$

とする. このときこのとき $\mathbf{t}^M$ の条件付確率分布は,

$$p(\mathbf{t}^M|\mathbf{y}) = \mathcal{N}(\mathbf{t}^M|\mathbf{0}, \beta^{-1}\mathbf{I_M})$$

となる. また, このとき $\mathbf{t}$ の条件付確率は, $\mathbf{y}, \epsilon$ がガウス分布に従うことから,

$$p(\mathbf{t}^M|\mathbf{y}) = \mathcal{N}(\mathbf{t}^M|\mathbf{0}, \beta^{-1}\mathbf{I_M})$$

となる. ただし, $\beta^{-1}$ は精度パラメータであり分散の逆数である.

いま, 求めたいのは $(\mathrm{x}, \mathrm{t})$, $(i = 1, ..., M)$ と $\mathrm{x}^{M+1}$ が与えられたときの $\mathrm{t}^{M+1}$ の確率分布, とくにその期待値と分散である. つまり, $p(\mathrm{t}^{M+1}|\mathrm{t}^1, ..., \mathrm{t}^M, \mathrm{x}^1, ..., \mathrm{x}^{M+1})$ を用いて $\mathrm{t}^{M+1}$ の期待値と分散を求めたい. それは, $p(\mathbf{t}|\mathbf{y})$ の周辺分布を求めることによって行う. まず,

$$p(\mathbf{t}^M) = \int p(\mathbf{t}^M|\mathbf{y})p(\mathbf{y})d\mathbf{y} = \mathcal{N}(\mathbf{t}^M|\mathbf{0}, \mathbf{L}_M)$$

となり. ここで $\mathbf{L}_M$ は,

$$\mathbf{L}_M = \begin{pmatrix} k(\mathrm{x}^1, \mathrm{x}^1) + \beta^{-1} & k(\mathrm{x}^1, \mathrm{x}^2) & \dots & k(\mathrm{x}^1, \mathrm{x}^M) \\ k(\mathrm{x}^2, \mathrm{x}^1) & k(\mathrm{x}^2, \mathrm{x}^2) + \beta^{-1} & \dots & k(\mathrm{x}^2, \mathrm{x}^M) \\ \vdots & \vdots & \ddots & \vdots \\ k(\mathrm{x}^M, \mathrm{x}^1) & k(\mathrm{x}^M, \mathrm{x}^2) & \dots & k(\mathrm{x}^M, \mathrm{x}^M) + \beta^{-1} \end{pmatrix}$$

である. 同様に, $\mathbf{t}^{M+1}$ の同時分布 $p(\mathbf{t}^{M+1})$

$$p(\mathbf{t}^{M+1}) = \mathcal{N}(\mathbf{t}_{M+1}|\mathbf{0}, \mathbf{L}_{M+1})$$

と書ける.

このとき $\mathbf{L}_{M+1}$ は $(M+1)\times(M+1)$ の共分散行列であり, 次のように分割して書ける.

$$
\mathbf{L}_{M+1} = \left(\begin{array}{cccc|c}
 & & & & k(\mathrm{x}^1,\mathrm{x}^{M+1}) \\
 & & & & k(\mathrm{x}^2,x^{M+1}) \\
 & & \mathrm{L}_M & & \vdots \\
 & & & & k(\mathrm{x}^M,\mathrm{x}^{M+1}) \\
\hline
k(\mathrm{x}^1,\mathrm{x}^{M+1}) & k(\mathrm{x}^2,\mathrm{x}^{M+1}) & \ldots & k(\mathrm{x}^M,\mathrm{x}^{M+1}) & k(\mathrm{x}^{M+1},\mathrm{x}^{M+1})+\beta^{-1}
\end{array}\right)
$$

$$
= \begin{pmatrix} \mathbf{L}_M & \mathbf{k} \\ \mathbf{k}^{\mathrm{T}} & c \end{pmatrix}
$$

以下では, $c = k(\mathrm{x}^{M+1},\mathrm{x}^{M+1}) + \beta^{-1}$, $\mathbf{k} = k(\mathrm{x}^n,\mathrm{x}^{M+1})(n=1,...,M)$ とする.

多変量ガウス分布には二つの変数集合の同時分布がガウス分布に従うならば一方の変数集合を条件としたときの他方の条件つき確率もガウス分布に従うという特徴がある. ここで, 多変量ガウス分布 $\mathcal{N}(\mathbf{t}_{M+1}|\mathbf{0},\mathbf{L}_{M+1})$ は

$$
\mathcal{N}(\mathbf{t}^{M+1}|\mathbf{0},\mathbf{L}_{M+1}) = \frac{1}{(2\pi)^{(M+1)/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{t}^{M+1})^{\mathrm{T}}\mathbf{L}_{M+1}^{-1}\mathbf{t}^{M+1}\right\} \tag{5}
$$

である. ここで $|\Sigma|$ は $\Sigma$ の行列式である.

いま変数集合を

$$
\mathbf{t^{M+1}} = \begin{pmatrix} \mathbf{t}_M \\ t^{M+1} \end{pmatrix} \tag{6}
$$

のように二つに分割した場合, 共分散行列 $\mathbf{L}_{M+1}$ も

$$
\mathbf{L}_{M+1} = \begin{pmatrix} \mathbf{L}_M & \mathbf{k} \\ \mathbf{k}^{\mathrm{T}} & c \end{pmatrix} \tag{7}
$$

となる. ここで便利のため新たに $\mathbf{L}_{M+1}^{-1}$ で定義される精度行列 $A$

$$
\mathbf{A}_{M+1} = \begin{pmatrix} \mathbf{A}_M & \mathbf{k_A} \\ \mathbf{k_A}^{\mathrm{T}} & c_A \end{pmatrix} \tag{8}
$$

を定義する.

いま, 多変量ガウス分布の指数部分の二次形式について考える. 式 (5), (6), (7) と (8) より,

$$
\begin{aligned}
(\mathbf{t}^{M+1})^{\mathrm{T}}\mathbf{L}_{M+1}^{-1}\mathbf{t}^{M+1} = &-\frac{1}{2}(\mathbf{t}^M)^{\mathrm{T}}\mathbf{A}_N\mathbf{t}^N - \frac{1}{2}(\mathbf{t}^M)^{\mathrm{T}}\mathbf{k_A}\mathrm{t}^{M+1} \\
&-\frac{1}{2}\mathrm{t}^{M+1}\mathbf{k_A}^{\mathrm{T}}\mathbf{t}^M - \frac{1}{2}c_A(\mathrm{t}^{M+1})^2
\end{aligned}
$$

この式を $\mathrm{t}^{M+1}$ の関数と見た場合, これも二次形式の形をとっており $p(\mathrm{t}^{M+1}|\mathbf{t}^M)$ もガウス分布となることが分かる. 以下では $\mathrm{t}^{M+1}$ を変数, $\mathbf{t}^M$ を定数とみなして, 多変量ガウス分布

の指数部分の係数比較をしながら $p(\mathrm{t}^{M+1}|\mathbf{t}^M)$ の条件つき平均 $m_{\mathrm{t}^{M+1}|\mathbf{t}^M}$ と条件つき分散 $\sigma^2_{\mathrm{t}^{M+1}|\mathbf{t}^M}$ を求める.

まず, $\mathrm{x}^{M+1}$ の二次の項をみると,

$$-\frac{1}{2}c_A(\mathrm{t}^{M+1})^2$$

となっていることから $p(\mathrm{t}^{M+1}|\mathbf{t}^M)$ の分散は

$$\sigma^2_{\mathrm{t}^{M+1}|\mathbf{t}^M} = c_A^{-1}$$

であることが分かる. 次に, $\mathrm{t}^{M+1}$ の線形項を見ると,

$$\mathrm{t}^{M+1}\mathbf{k_A}^{\mathrm{T}}\mathbf{t}^M$$

となっており, この $\mathrm{t}^{M+1}$ の係数は $(\sigma^2)^{-1}_{\mathrm{t}^{M+1}|\mathbf{t}^M} m_{\mathrm{t}^{M+1}|\mathbf{t}^M}$ と等しくなるので, 条件つき平均は,

$$m_{\mathrm{t}^{M+1}|\mathbf{t}^M} = (\sigma^2)^{-1}_{\mathrm{t}^{M+1}|\mathbf{t}^M}$$

となる. ここで分割された行列の逆行列に関する公式

$$\begin{pmatrix} B & C \\ D & E \end{pmatrix}^{-1} = \begin{pmatrix} (B-CE^{-1}D)^{-1} & -(B-CE^{-1}D)^{-1}CE^{-1} \\ -E^{-1}C(B-CE^{-1}D)^{-1}E^{-1} & -E^{-1}C(B-CE^{-1}D)^{-1}CE^{-1} \end{pmatrix}$$

と,$\mathbf{L}^{-1}_{M+1} = \mathbf{A}_{M+1}$ を用いれば,

$$\begin{aligned} c_A &= (c - \mathbf{k}^{\mathrm{T}}\mathbf{L}_M^{-1}\mathbf{k})^{-1} \\ \mathbf{k}_A &= -(c - \mathbf{k}^{\mathrm{T}}\mathbf{L}_M^{-1}\mathbf{k})^{-1}\mathbf{k}\mathbf{L}_M^{-1} \end{aligned}$$

を得る. これらの式より条件つき平均と条件つき分散は,

$$\begin{aligned} m_{\mathrm{t}^{M+1}|\mathbf{t}^M} &= \mathbf{k}^{\mathrm{T}}\mathbf{L}_M^{-1}\mathbf{t} \\ \sigma^2_{\mathrm{t}^{M+1}|\mathbf{t}^M} &= c - \mathbf{k}^{\mathrm{T}}\mathbf{L}_M^{-1}\mathbf{k} \end{aligned}$$

となり, 改めて条件つき平均と条件つき分散をそれぞれ $m(\mathrm{x}_{M+1})$, $\sigma^2(\mathrm{x}_{M+1})$ とおけば, 条件つき分布 $p(\mathrm{t}^{M+1}|\mathrm{t}^1,...,\mathrm{t}^M,\mathrm{x}^1,...,\mathrm{x}^{M+1})$ の期待値と分散は以下のように求まる.

$$m(\mathrm{x}_{M+1}) = \mathbf{k}^{\mathrm{T}}\mathbf{L}_M^{-1}\mathbf{t} \tag{9}$$

$$\sigma^2(\mathrm{x}_{M+1}) = c - \mathbf{k}^{\mathrm{T}}\mathbf{L}_M^{-1}\mathbf{k} \tag{10}$$

実際に $M = 200, h = 200,\mathbf{x}$ を為替の変化率,$\mathbf{t}$ を TOPIX の収益率として, 為替の変化率を横軸,TOPIX の収益率を縦軸とした場合のガウス過程による回帰を用いた期待収益率とリ

スクのグラフを図 2 に表す.

このグラフよりデータが少ない左側では分散が大きくなることが分かる. つまりデータの過少性をリスクとして捉えることができている. 条件つき確率分布を推定することで得られた条件つき期待値や条件つき分散値はデータが少ない場合ほどそれら二つの値が計算できなくなる傾向があった. また, ガウス過程回帰によって条件つき確率分布を推定することで得られた条件つき分散値は, データが少ない場合ほど分散値に対しての信頼度が低くなることがうかがえる.

一方, データ点が多く集まっている部分ほどガウス過程による条件つき分散値が小さくなっていることが見て取れる. 例えば, ドル円相場の変化率が 0 付近の TOPIX の期待収益率値を見ると条件つき分散値は小さいものの, データ点からは期待収益率値にはばらつきが見られる. これは式 (9) に見られるように $\sigma^2(\mathrm{x}^{M+1})$ の計算において収益率の情報が現れていないからである.

## 3　データの過少性をリスクとした 2 資産配分モデル

ガウス過程を用いて収益率を推定するため, ここでは TOPIX などの市場平均と安全資産である現金の 2 資産に配分するモデルを考える.TOPIX への配分を $\mathbf{v}$, 現金への配分を $1-\mathbf{v}$ とする.

TOPIX の収益率を推定するため経済指標を x とする. ガウス過程によって推定された x が与えられたときの TOPIX の期待収益率を $m(\mathrm{x})$, 分散を $\sigma_G(\mathrm{x})$ とする. さらに, $\epsilon$-近傍密度推定で推定された分散を $\sigma_\epsilon(\mathrm{x})$ とする.

本報告書では,

$$\sigma(\mathrm{x}) = \sigma_G(\mathrm{x}) + \sigma_\epsilon(\mathrm{x})$$

として分散を考える. これらを用いて x が与えられたとき以下の最適化問題の解を探し配分とする.

本報告書で数値実験を行う為に以下の平均分散モデルを用いる.

$$\begin{array}{ll} \text{minimize} & \sigma \mathrm{v}^2 - \tau m \mathrm{v} \\ \text{subject to} & 0 \leq \mathrm{v} \leq 1 \end{array}$$

ここで $\tau$ はリスクとリターンを調整するパラメータである.

## 4 数値実験

本節では, 第 3 節で考えたモデルに対して実データを用いたシミュレーションを行う. 利用するデータはドル円相場の変化率と TOPIX の収益率データとし, それぞれ 2007 年 1 月 1 日〜2013 年 12 月 31 日までのデータを用いる. 日付 d でのドル円相場 $\mathrm{k}^{\mathrm{d}}$ の変化率 $\mathrm{z}^{\mathrm{d}}$ は

$$\mathrm{z}^{\mathrm{d}} = \frac{\mathrm{k}^{\mathrm{d}} - \mathrm{k}^{\mathrm{d}-1}}{\mathrm{k}^{\mathrm{d}-1}}$$

また, 日付 d での TOPIX の株価指数 $\mathrm{f}^{\mathrm{d}}$ の収益率 $\mathrm{r}^{\mathrm{d}}$ は

$$\mathrm{r}^{\mathrm{d}} = \frac{\mathrm{f}^{\mathrm{d}} - \mathrm{f}^{\mathrm{d}-1}}{\mathrm{f}^{\mathrm{d}-1}}$$

とする.

また, 経済指標 $\mathbf{u}^{\mathrm{d}} \in \mathbb{R}^4$ として, ドル円相場の変化率と過去 3 日分の収益率を用いて以下のように表す,

$$\mathbf{u}^{\mathrm{d}} = \begin{pmatrix} \mathrm{z}^{\mathrm{d}} \\ \mathrm{r}^{\mathrm{d}-1} \\ \mathrm{r}^{\mathrm{d}-2} \\ \mathrm{r}^{\mathrm{d}-3} \end{pmatrix}$$

そして, 平均分散モデルの期待収益率の尺度として, (9) を用い, 分散の尺度として $\epsilon-$ 近傍カーネル密度推定法 によって計算された (1) 式の条件つき分散と, ガウス過程回帰によって計算された (10) 式の条件つき分布 $p(\mathrm{t}^{M+1}|\mathrm{t}^1, ..., \mathrm{t}^M, \mathrm{x}^1, ..., \mathrm{x}^{M+1})$ の分散値の和を利用するものとする.

また, $m(\mathrm{x})$, $\sigma_G(\mathrm{x})$ を求める際に, $\theta_1$ をパラメータとした以下のカーネルを用いた.

$$k(\mathrm{x}^n, \mathrm{x}^m) = \exp\left\{-\frac{\theta_1}{2}\|\mathrm{x}^n - \mathrm{x}^m\|^2\right\}$$

本報告書の, 数値実験においては $\theta_1 = 100$ とした. また, $\beta$ は精度パラメータと呼ばれるパラメータであり, 分散の逆数をとったものである. $\beta$ の値を小さくとればリスクをより考慮した場合を想定できる.

本実験では 2007 年 4 月 13 日から 200 日分を初期の訓練データとし, 201 日目の条件つき期待収益率と条件つき分散をもとに資産配分を決定する. それ以降は 2007 年 1 月 2 日から 200 日分を新たに訓練データとし再び 201 日目の資産配分を決定するといったように訓練データを一日ずつずらしながら最終的に 2013 年 12 月 30 日まで実験を進める.

実験結果は図 1 から図 26 にまとめ, 付録 A に掲載する.

図 3 から図 5 では $\tau = 10$, $\beta = 20$ として資産運用のシミュレーション結果, $\sigma(\mathrm{x})$ の変化, $\sigma_G(\mathrm{x})$ の変化の順に示した.

図 6 から図 8 では $\tau = 10$, $\beta = 200$, 図 9 から図 11 では $\tau = 50$, $\beta = 20$, 図 12 から図 14 では $\tau = 50$, $\beta = 200$, 図 15 から図 17 では $\tau = 100$, $\beta = 20$, 図 18 から図 20 では $\tau = 100$, $\beta = 200$, 図 21 から図 23 では $\tau = 200$, $\beta = 20$, 図 24 から図 26 では $\tau = 200$, $\beta = 200$, として同様に示す.

シミュレーションの結果, $\beta$ の値を小さくとれば $\sigma_G(\mathrm{x})$ の値が大きくなり, リスクを大きく考慮するようになるため, 投資は少なくなり, $\tau$ の値が大きくなると投資が増えることがわかる．さらに, TOPIX の資産価値が大きく下がったり上がったりしている点, すなわち $\sigma(\mathrm{x})$, $\sigma_G(\mathrm{x})$ に対してリスクの値が一時的に上がっている 3 点は, 左からリーマンショック, 東北大震災, アベノミクスの時の動向を示しており, それらの点では 経済状況が大きく変わってしまっている. よってそれらの点の直後では, データの過少性による リスクを考慮して投資をしないという選択を取っていることが分かる.

# 5 まとめと今後の課題

本報告書ではデータの過少性を考慮したモデルを提案した. データの過少性を表すためにガウス過程によって条件つきの分散を導出した. またシミュレーションによって, 実際の運用経緯を調べた. リーマンショック直後など, 対応するデータが少ない状況では, それをリスクとして捉えることができた.

今後の課題として, 経済指標の選び方が挙げられる. 本実験においては経済指標としてドル円相場や TOPIX の過去の収益率を用いたが, より相関のある経済指標を見つけることができれば実験結果よりも良い結果が得られると思われる. また, 本報告書では分散の値を単純に $\sigma_G(\mathbf{x})$ と, $\sigma_\epsilon(\mathbf{x})$ の和でとったが, さらに良い結果が出るような分散のとり方を考えることも検討すべきであり, $\sigma_G(\mathbf{x})$ を計算するために用いたカーネル関数のパラメータ $\theta_1$ のとり方も考える必要がある. 最後に, 実際の場では取引コストを考慮した場合を考える必要がありこれも課題である.

# 謝辞

本研究に取り組むにあたって多大なる助言をいただき，本報告書の作成にあたっては細部に至るまで多くの御指導を賜った山下信雄准教授に深く感謝の意を表します．また, 福田助教ならびに最適化数理研究室の皆様に厚く御礼申し上げます.

## 参考文献


[1] Bishop, C.M., パターン認識と機械学習 (上), (下), シュプリンガー・ジャパン, 2007.

[2] 今野浩, 理財工学 I -平均・分散モデルとその拡張-, 日科技連出版社, 2004.

[3] Markowitz, H., Portfolio Selection : Efficient Diversification of Investments, John Wiley & Sons, New York, 1959.

[4] Perold, A., "Large Scale Portforio Optimization", Management Science, Vol.30, pp.1143-1160, 1984.

[5] 小針 宏, 確率統計入門, 岩波書店, 1973.

[6] 松井 亮大, "取引コストを考慮した資産配分関数の構築", 特別研究報告書 数理工学コース, 京都大学, 2012.

[7] 吉田雅基, 山下信雄, "カーネル法を用いた資産配分関数の構築", Technical Report 2009-013, 数理工学専攻, 京都大学, 2009.

[8] Sugiyama, M., Takeuchi, I., Suzuki, T., Kanamori, T, Hachiya, H., Okanohara, D., "Least-Squares Conditional Density Estimation", IEICE Transactions 93-D(3), pp.583-594, 2010.


## 付録 A

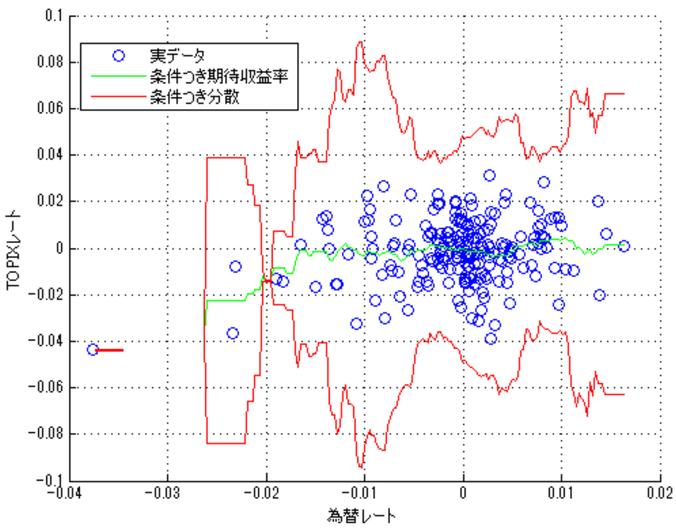


図 1　為替を経済指標とした TOPIX の $\epsilon -$ 近傍カーネル密度推定法 を用いて得られた条件つき期待収益率と条件つき分散

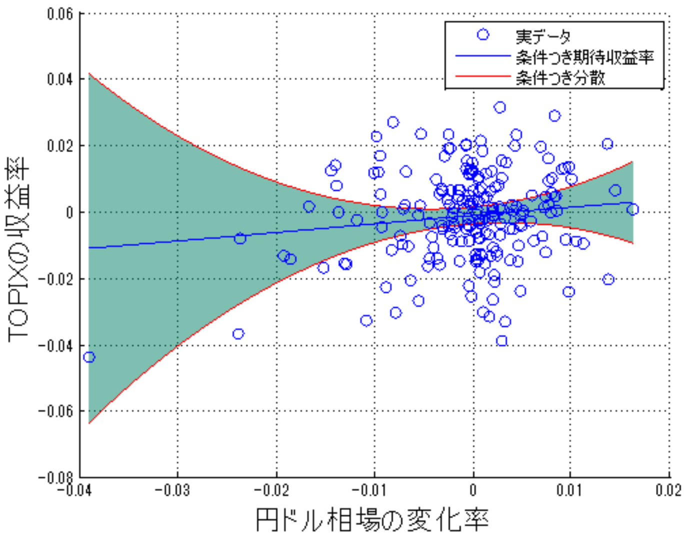


図 2　為替を経済指標とした TOPIX のガウス過程を用いて得られた条件つき期待収益率と条件つき分散


図 3　初期資産を 1 とした場合の日ごとの資産変化 ($\tau = 10, \beta = 20$)

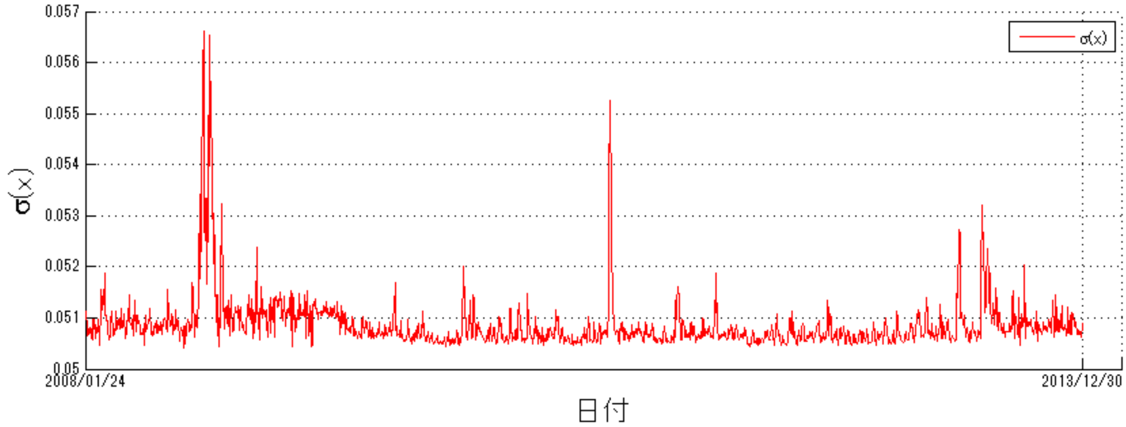


図 4　日ごとの $\sigma(\mathrm{x})$ の変化 ($\tau = 10, \beta = 20$)

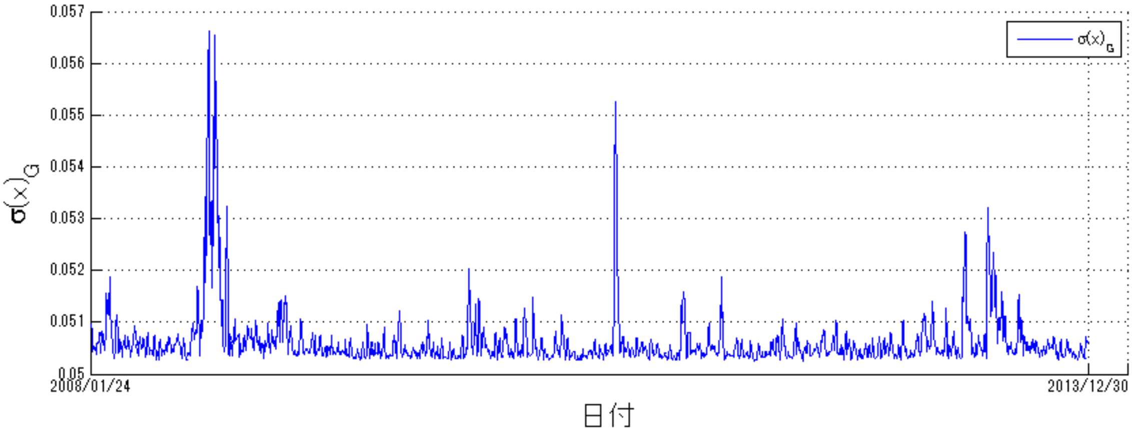


図 5　日ごとの $\sigma_G(\mathrm{x})$ の変化 ($\tau = 10, \beta = 20$)

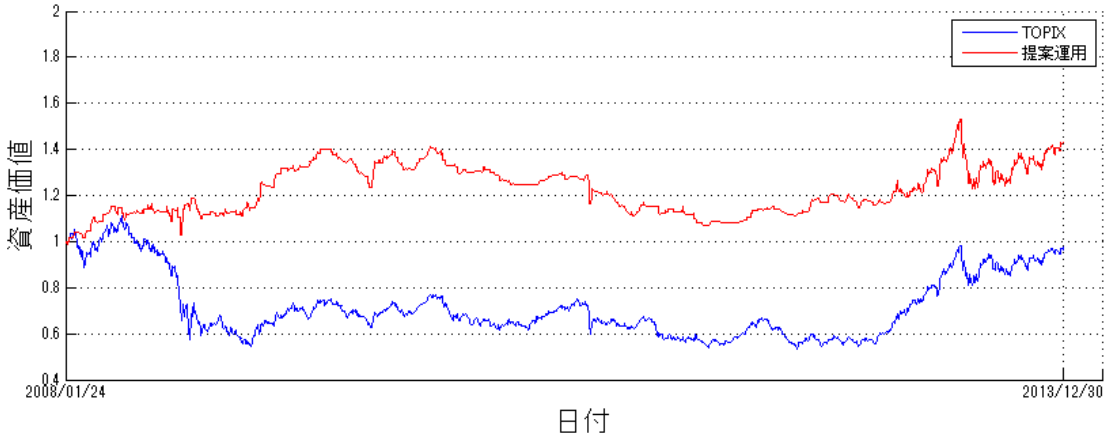


図 6　初期資産を 1 とした場合の日ごとの資産変化 ($\tau = 10, \beta = 200$)

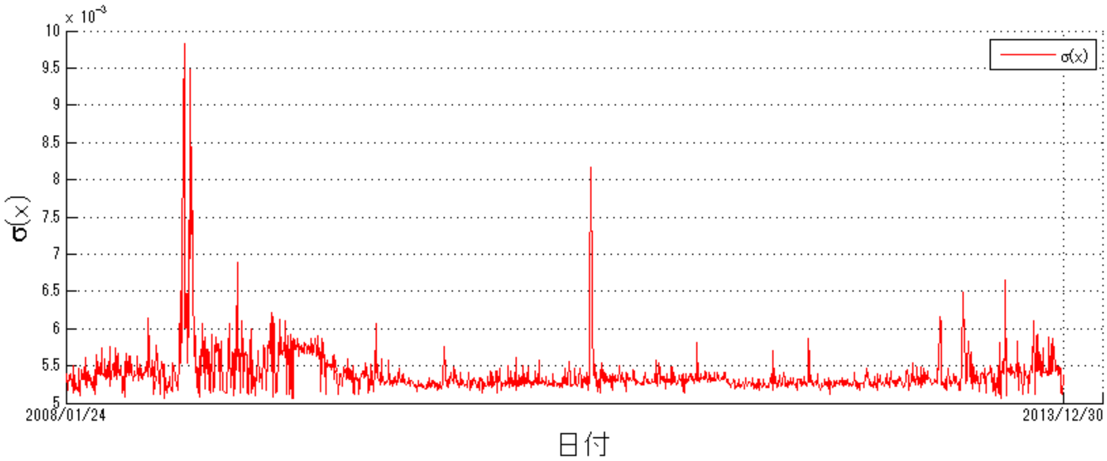


図 7　日ごとの $\sigma(\mathrm{x})$ の変化 ($\tau = 10, \beta = 200$)

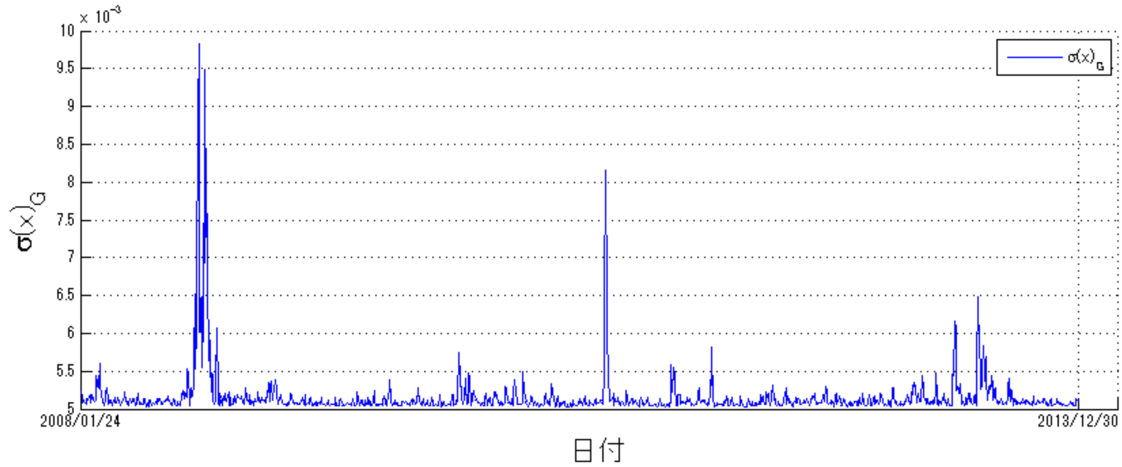


図 8　日ごとの $\sigma_G(\mathrm{x})$ の変化 ($\tau = 10, \beta = 200$)

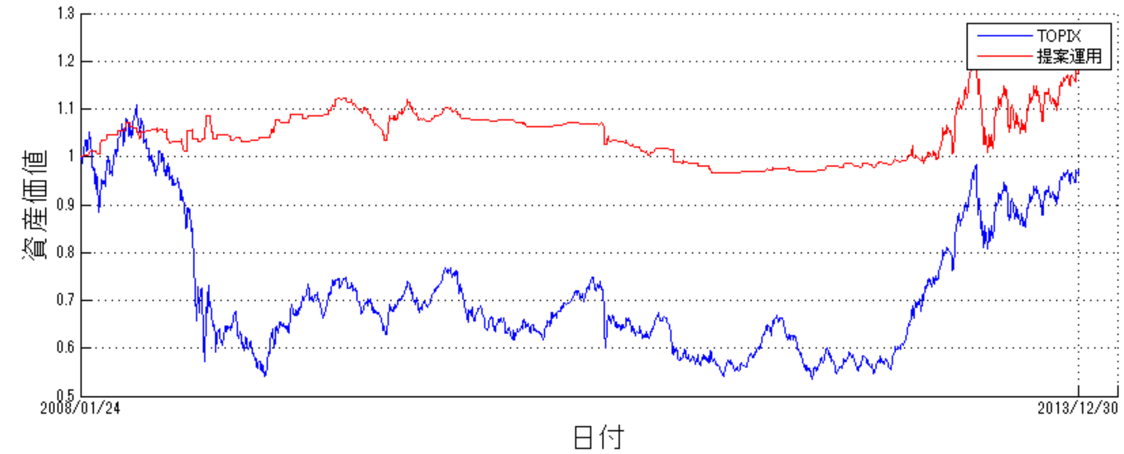


図 9　初期資産を 1 とした場合の日ごとの資産変化 ($\tau = 50, \beta = 20$)

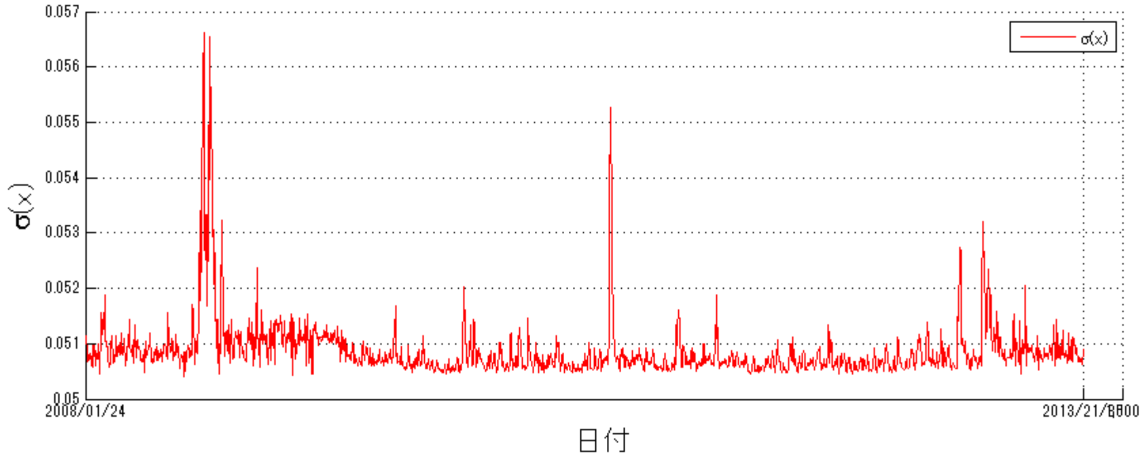


図 10　日ごとの $\sigma$(x) の変化 ($\tau = 50, \beta = 20$)

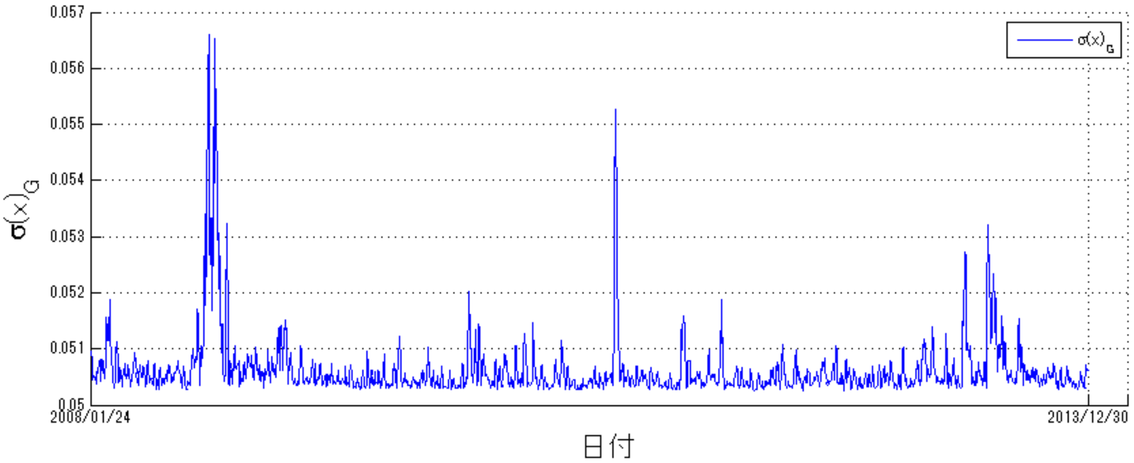


図 11　日ごとの $\sigma_G$(x) の変化 ($\tau = 50, \beta = 20$)

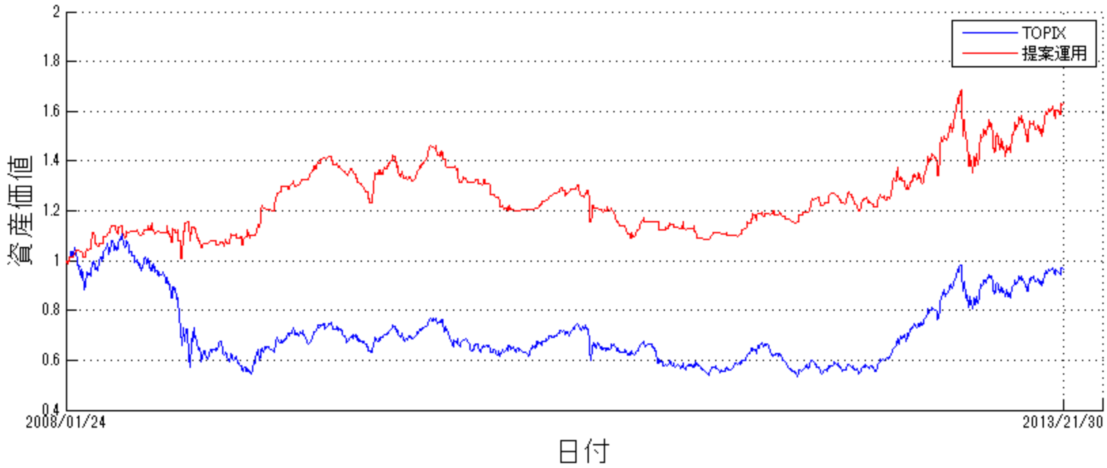


図 12　初期資産を 1 とした場合の日ごとの資産変化 ($\tau = 50, \beta = 200$)

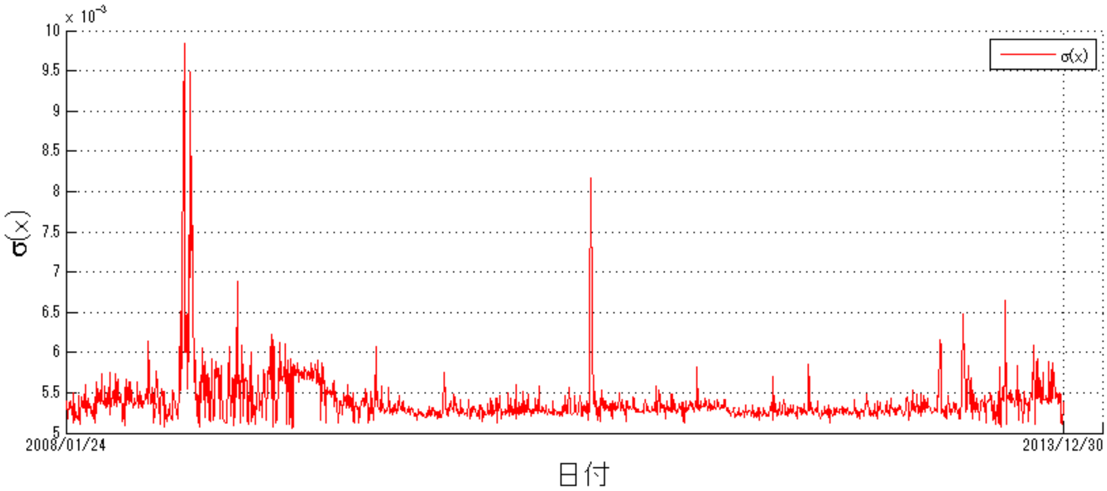


図 13　日ごとの $\sigma(\mathrm{x})$ の変化 ($\tau = 50, \beta = 200$)

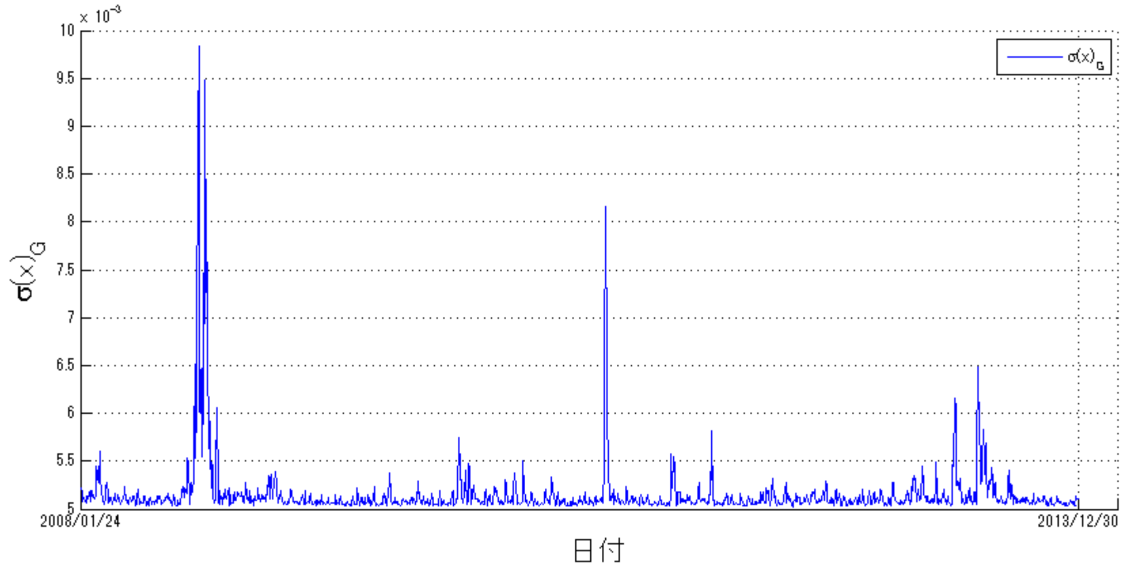


図 14　日ごとの $\sigma_G(\mathrm{x})$ の変化 ($\tau = 50, \beta = 200$)

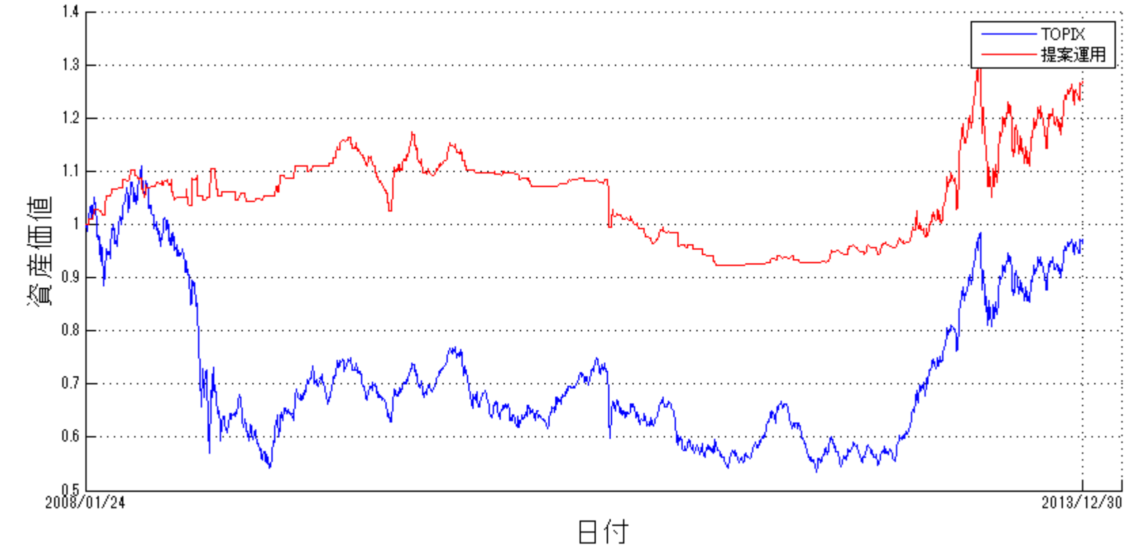


図 15　初期資産を 1 とした場合の日ごとの資産変化 ($\tau = 100, \beta = 20$)

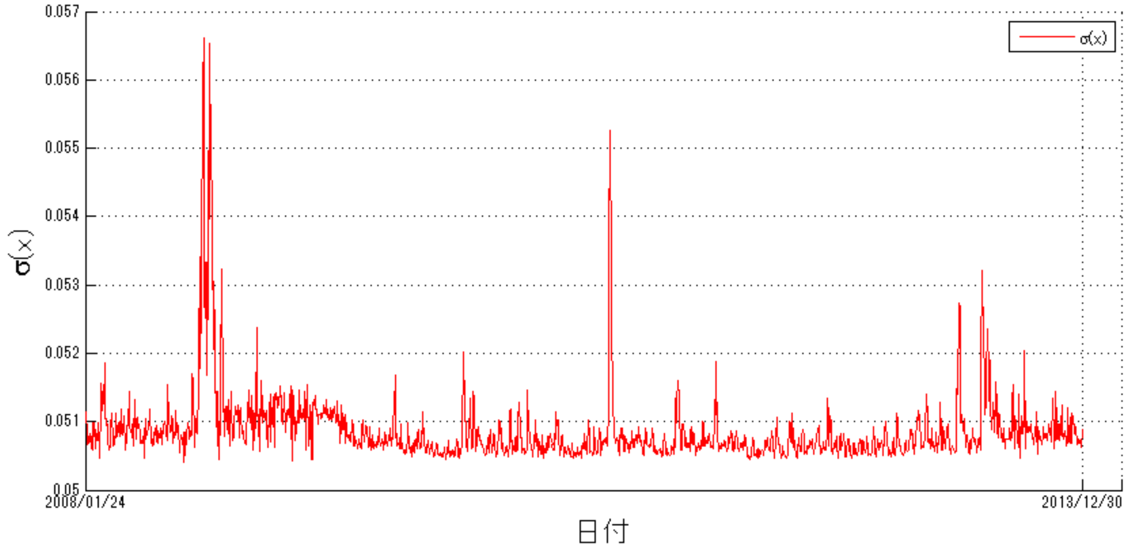


図 16　日ごとの $\sigma(\mathrm{x})$ の変化 ($\tau = 100, \beta = 20$)

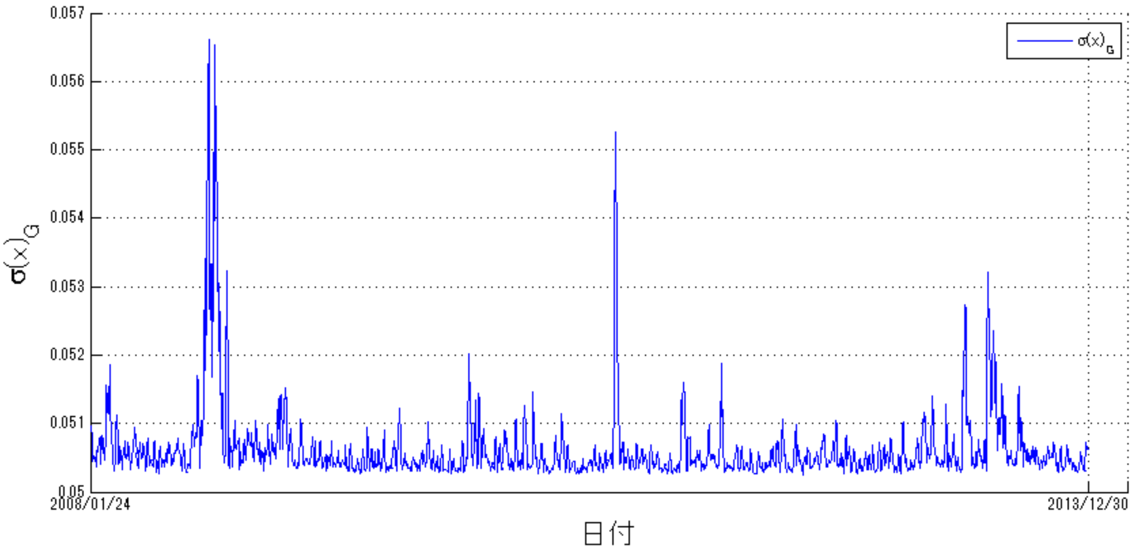


図 17　日ごとの $\sigma_G(\mathrm{x})$ の変化 ($\tau = 100, \beta = 20$)

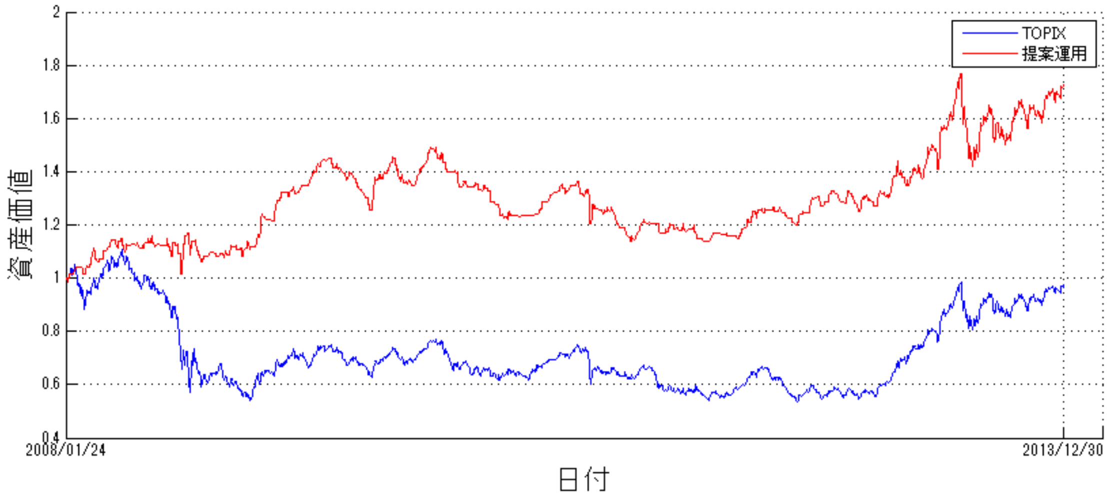


図 18　初期資産を 1 とした場合の日ごとの資産変化 $(\tau = 100, \beta = 200)$

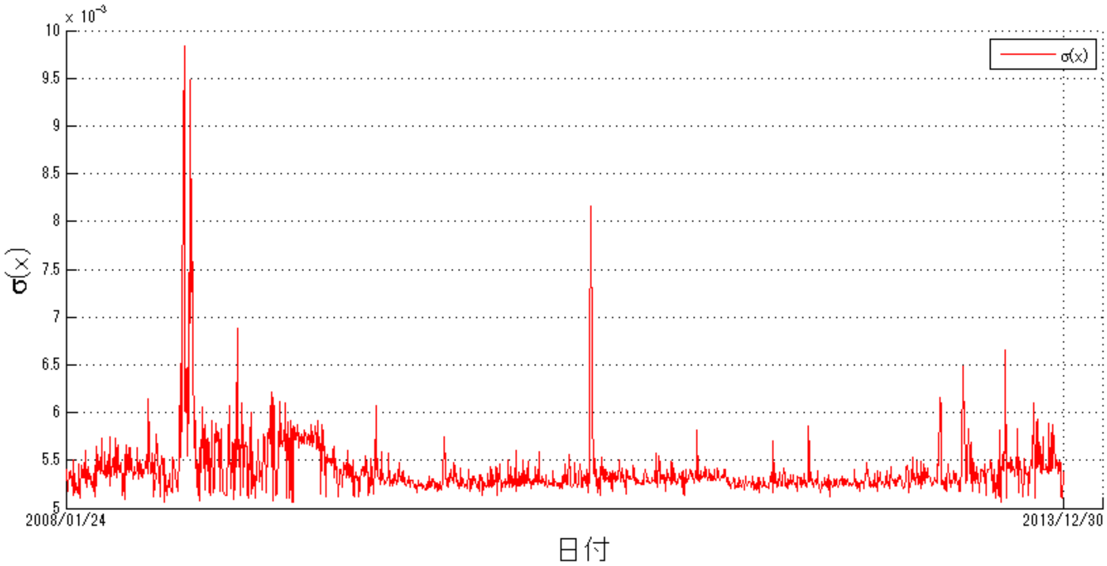


図 19　日ごとの $\sigma(\mathrm{x})$ の変化 $(\tau = 100, \beta = 200)$

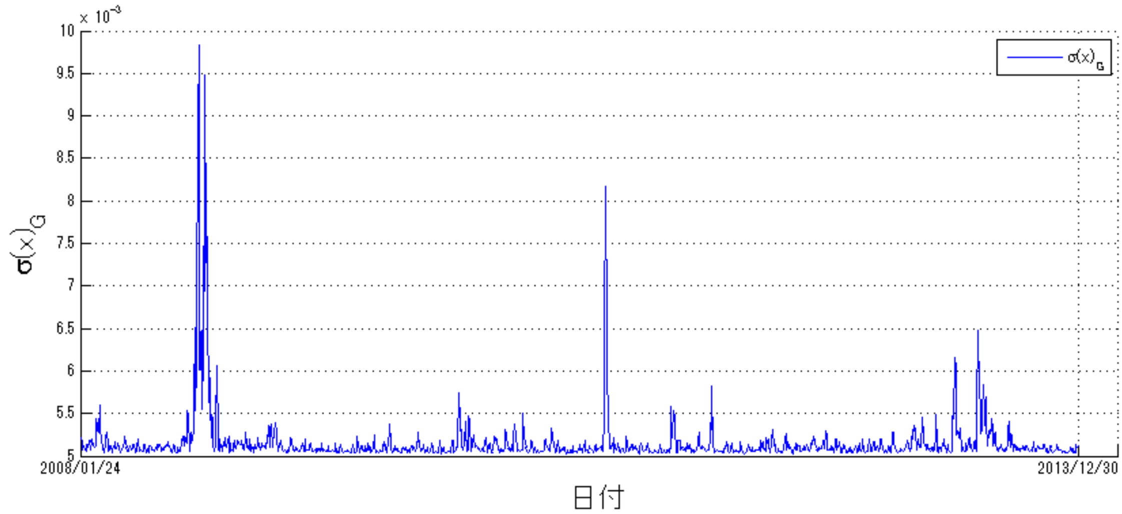


図 20　日ごとの $\sigma_G(\mathrm{x})$ の変化 $(\tau = 100, \beta = 200)$

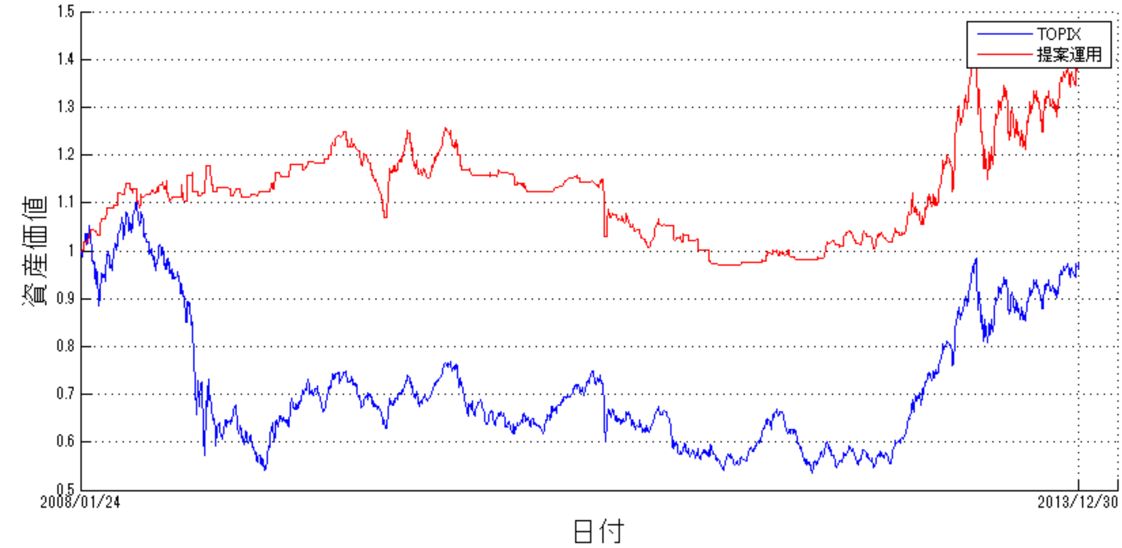


図 21　初期資産を 1 とした場合の日ごとの資産変化 ($\tau = 200, \beta = 20$)


図 22　日ごとの $\sigma$(x) の変化 ($\tau = 200, \beta = 20$)

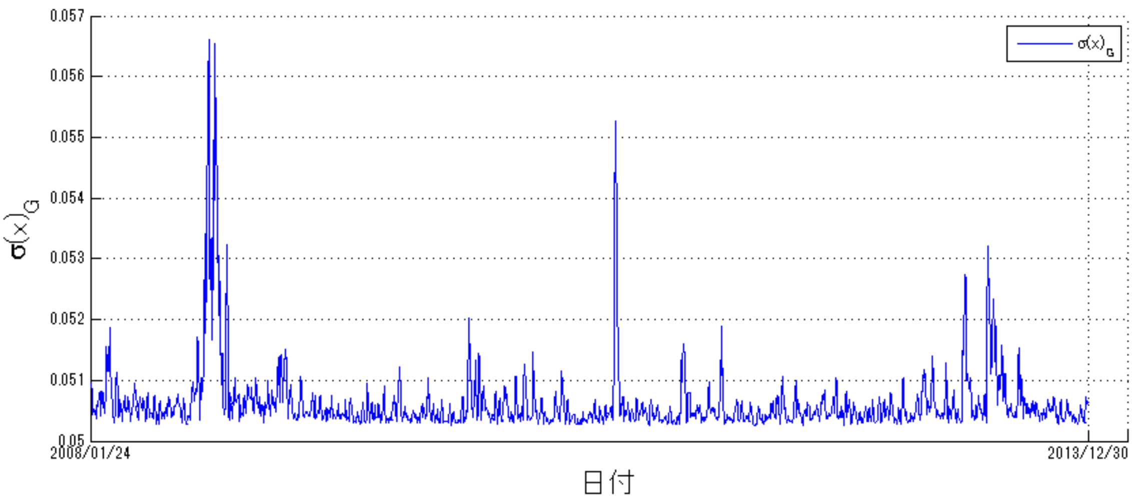


図 23　日ごとの $\sigma_G$(x) の変化 ($\tau = 200, \beta = 20$)

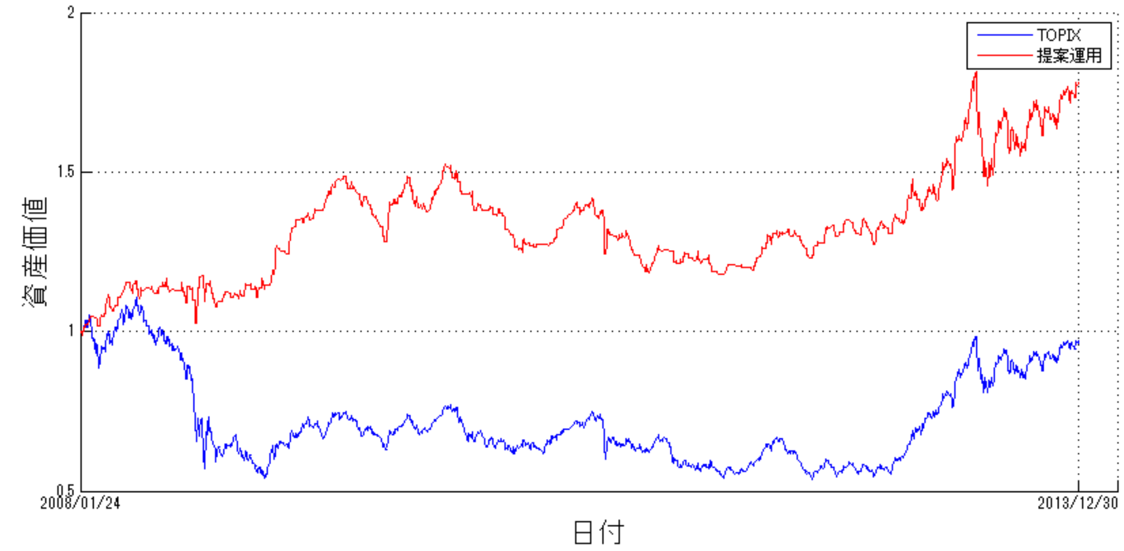


図 24　初期資産を 1 とした場合の日ごとの資産変化 ($\tau = 200, \beta = 200$)

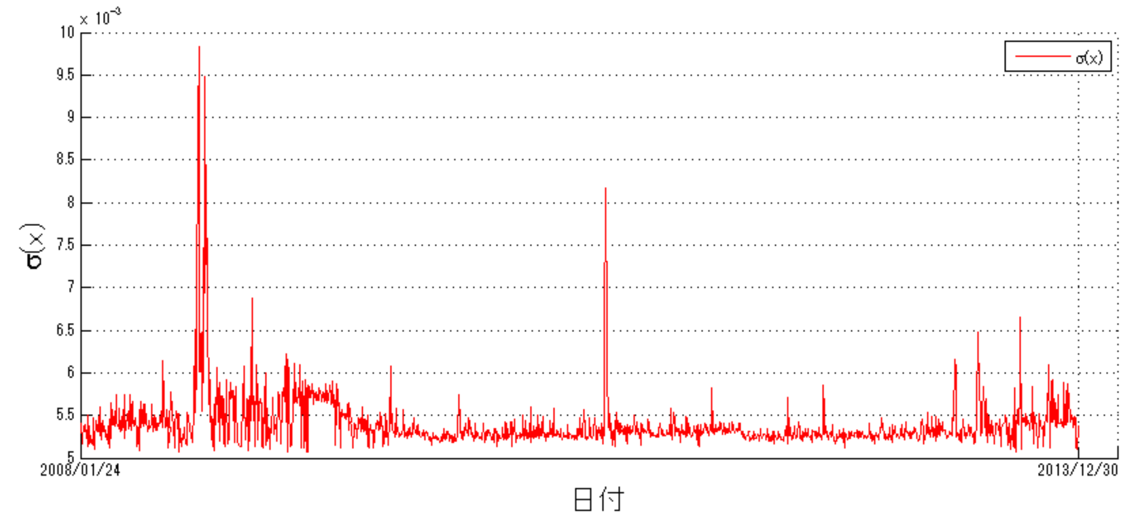


図 25　日ごとの $\sigma(\mathrm{x})$ の変化 ($\tau = 200, \beta = 200$)

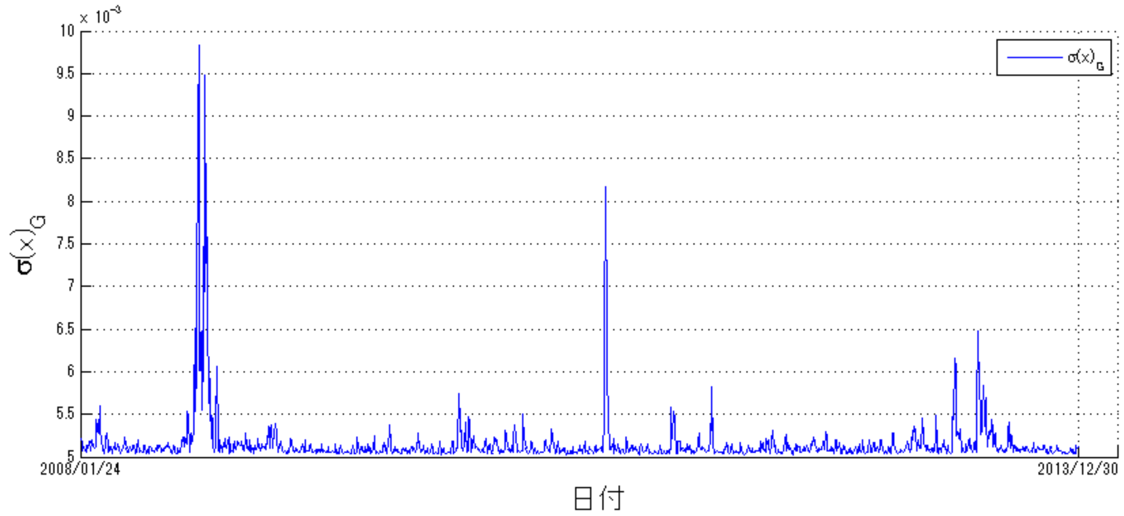


図 26　日ごとの $\sigma_G(\mathrm{x})$ の変化 ($\tau = 200, \beta = 200$)

特別研究報告書

# データの過少によるリスクを考慮した資産配分について

指導教員　　　山下信雄　准教授

京都大学工学部情報学科
数理工学コース
平成 22 年 4 月入学

西崎　公彦

平成 26 年 1 月 31 日提出

データの過少によるリスクを考慮した資産配分について

西崎　公彦

平成25年度

# データの過少によるリスクを考慮した資産配分について

西崎　公彦

**摘要**

本報告書では, 経済状況が与えられたもとで, 資産の収益率の条件つき確率分布を推定し, その分布に基づいた資産運用を考える.

資産配分問題に関する研究はこれまでに多くなされてきたが, マーコビッツ の平均・分散モデルをはじめ, それらの研究では収益率の確率分布を既知であるとして扱っていることが多い. 一方で, 収益率の確率分布を推定する手法が数多く提案されている. その一つとしてヒストリカルデータを用いるものがある. しかし, 単純なヒストリカルデータを用いた場合, 経済状況に応じた推定ができない. また, マルチファクターモデルのように経済指標の線形回帰によって計算された収益率を用いるモデルも提案されている. しかし, ある経済指標に関連したデータが少ない場合には, 単純な線形回帰では, そのデータの過小性をリスクとして組み込むことができない. そのため, リーマンショックなどのように過去と異なる状況が起きた場合には投資判断を誤ってしまう可能性がある.

本報告書ではデータの過少性をリスクとして考えるモデルを考える. そのために過去のデータを用いてガウス過程によって条件つき確率分布を推定する. さらにその分布に基づいて収益率の条件つき分散を計算する. このように計算された分散は, 一般に, 条件となる経済指標と近いデータが過去に少ないとき大きな値をとる. 逆に, データが多いときには, そのデータにおける収益率の変動にかかわらず, 小さい値をとってしまうことがある. そこで, データが多いところの分散を表すために, $\epsilon$-近傍カーネル密度推定法で導いた分散も考える. 提案モデルではこの二種類の分散の和をリスクとして扱う. さらに, ガウス過程によって得られた期待収益率と提案リスクを用いた平均分散モデルを考え, そのモデルによって資産配分を決定する資産運用のシミュレーションを行った. その結果から提案モデルの妥当性を考察する.