

# 相関係数行列の行列式に基づいた多変量管理図の研究

M2011MM043 近藤佑亮

指導教員：松田眞一

## 1 はじめに

管理図は製造工程が安定した管理状態下にあるかどうかを判断するための方法であり、W.A.Shewhartによって考えられた方法である。管理すべき品質特性が複数ある場合、多変量管理図が用いられる。各品質特性の平均や分散の変化を検出するための多変量管理図の研究は多くなされているが、相関関係に着目した多変量管理図の研究はあまりなされていない。安井ら[8]は相関関係の管理に標本相関係数行列の行列式に基づく管理図(|R|管理図)を提案している。

本研究では安井ら[8]の|R|管理図と同様に標本相関係数行列の行列式に基づく管理図として、ブートストラップ|R|管理図、|R|累積和管理図を提案すると共に、その評価を行う。また先行研究で秋月[1]は相関の変化を検出する管理図として、2変量のデータに対して移動相関管理図を提案している。本研究では多変量データに対して移動相関係数行列の行列式による管理図を提案し、その評価を行う。

## 2 管理図

管理図は、工程が管理状態であるかどうかを判断するためにW.A.Shewhartによって発表された手法である。管理図は一本の中心線(CL)と上側管理限界線(UCL)、下側管理限界線(LCL)からなり、群ごとに統計量を打点していくことで、様々な判定ルール(例えば管理限界線の外に打点された)で工程が統計的管理下に置かれているかを判断する方法である。詳細はMontgomery[5]を参照のこと。

## 3 |R|管理図

今  $m$  個の群を持ち標本の大きさが  $n$  であり、 $p$  個の品質特性を持つデータベクトルを管理することを考える。第  $i$  群に属する第  $j$  番目のデータベクトルの第  $k$  要素を  $x_{ijk}$  で表すとき、データベクトルは  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijk}, \dots, x_{ijp})$  である。 $\mathbf{x}_{ij}$  は  $i, j$  に対して互いに独立で同一な多変量正規分布  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  に従うとする。安井ら[8]は相関構造を示す統計量として、標本相関係数行列  $R_i$  の行列式、すなわち、 $|R_i|$  を打点する管理図を提案した。

母相関係数行列  $P_\Sigma$  の固有値を  $\lambda_1, \dots, \lambda_p$  とするとき、母相関係数行列の行列式は単位行列  $\mathbf{I}$  を用いて、

$$|P_\Sigma| = \prod_{l=1}^p \lambda_l = \prod_{l=1}^p \lambda_l |\mathbf{I}|$$

で表される。 $\lambda_1, \dots, \lambda_p$  に対応する固有ベクトル  $\mathbf{p}_1, \dots, \mathbf{p}_p$  からなる直交行列  $P = (\mathbf{p}_1, \dots, \mathbf{p}_p)$  を用いて、 $R_i$  についても直交行列  $P$  で変換すれば、安井ら[8]は行列式  $|R_i|$  の分布は多変量正規分布  $N_p(\mathbf{0}, \mathbf{I})$  から得ら

れる標本相関係数行列  $R_I$  の行列式の分布の  $(\lambda_1 \cdots \lambda_p)$  倍になると主張している。したがって、 $|R|$  管理図における管理限界線は  $|R_I|$  の分布におけるパーセント点の  $(\lambda_1 \cdots \lambda_p)$  倍で求めることができる。

各管理項目における特性値すべてが互いに独立であるとき  $|R_i|$  が 1 に近くなり、相関が強くなるほど 0 に近くなることを考慮して打点する統計量を  $-\log |R_i|$  とする。

管理限界線を求めるのに  $-\log |R_I|$  のパーセント点が必要となるが、 $-\log |R_I|$  の分布を数学的に解くのは困難である。従って乱数を用いて元データと同じ標本の大きさをもつ 100000 群のデータから統計量を計算して、それらの経験分布から  $-\log |R_I|$  のパーセント点を求める。

$|R|$  管理図の信頼水準  $100 \times (1 - \alpha)\%$  の管理限界線の計算は以下の手順で行う。

- 母相関係数行列の推定値  $\hat{P}_\Sigma = \frac{1}{m} \sum_{i=1}^m R_i$  を計算し、その行列式  $|\hat{P}_\Sigma|$  を計算する。
- $\theta_I = -\log |R_I|$  の経験分布を乱数を用いて求める。
- 上側管理限界線(UCL)、下側管理限界線(LCL)はそれぞれ、

$$UCL = -\log |\hat{P}_\Sigma| + \theta_{I,(1-\alpha/2)}$$

$$LCL = -\log |\hat{P}_\Sigma| + \theta_{I,(\alpha/2)}$$

となる。ここで  $\theta_{I,(1-\alpha/2)}$  および  $\theta_{I,(\alpha/2)}$  はそれぞれ  $\theta_I$  の  $1 - \alpha/2$  パーセント点、 $\alpha/2$  パーセント点である。

## 4 ブートストラップ法

ブートストラップ法はEfron[3]によって提案された数値計算法である。理論に基づく解析を、計算機を用いた反復計算で置き換えたものであり、母集団に特定の分布形を仮定せずに統計量の標準誤差や信頼区間を求めることができる。ここで説明のために観測値  $X$  を用いた工程管理を考える。 $X$  の分布の形は未知であるが、工程を特徴づけるパラメータ  $\theta$  を推定したいとする。 $\theta$  としては平均、標準偏差などが考えられる。母集団から  $n$  個の標本を抽出し、それらを  $X_1, \dots, X_n$  とする。また関心のあるパラメータ  $\theta$  の推定値を  $\hat{\theta}$  とする。ブートストラップ法では観測された標本  $X_1, \dots, X_n$  から無作為復元抽出を行い、 $B$  個のブートストラップ標本  $\mathbf{X}_b^* = (X_1^*, \dots, X_n^*)$ ,  $b = 1, \dots, B$  を作成する。この  $B$  個のブートストラップ標本に対して、それぞれブートストラップ推定値  $\hat{\theta}^*$  を計算する。このようにして計算された  $B$  個のブートストラップ推定値  $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$  は推定される  $\hat{\theta}$  の標本分布を与える。

## 5 ブートストラップ|R|管理図

工程が管理状態にあるとき観測値がモデル

$$x_{ijk} = \mu_k + \varepsilon_{ijk}$$

に従うと仮定する。ここで  $\mu_k, k = 1, \dots, p$  は各品質特性  $k$  の母平均である。また  $\varepsilon_{ijk}$  は誤差項である。そして, Seppala et al.[6] のバランス・サブグループ・ブートストラップ法を用いて, ブートストラップ |R| 管理図の管理限界線を構成する。その手順は以下のようになる。

1. 元の標本の残差

$$\begin{aligned} \mathbf{e}_1 &= \mathbf{e}_{11} = (e_{111}, \dots, e_{112}, \dots, e_{11p}) \\ \mathbf{e}_2 &= \mathbf{e}_{12} = (e_{121}, \dots, e_{122}, \dots, e_{12p}) \\ &\vdots \\ \mathbf{e}_N &= \mathbf{e}_{mn} = (e_{mn1}, \dots, e_{mn2}, \dots, e_{mnp}) \end{aligned}$$

を計算する。ここで  $N = mn$  であり,  $e_{ijk} = x_{ijk} - \bar{x}_{i.k}$  である。 $\bar{x}_{i.k}$  は群ごとの特性それぞれの標本平均である。

2.  $B = Mm \geq 10000$  となるような整数  $M$  を決める。 $(M$  は計算時間を考慮して決める。)
3.  $N$  個の残差ベクトルを  $M$  回複製する。

$$\underbrace{\mathbf{e}_1, \dots, \mathbf{e}_1}_{M \text{ 回}}, \underbrace{\mathbf{e}_2, \dots, \mathbf{e}_2}_{M \text{ 回}}, \dots, \underbrace{\mathbf{e}_N, \dots, \mathbf{e}_N}_{M \text{ 回}}$$

4.  $MN$  個の残差ベクトルを無作為に並び替えて, それぞれの標本の大きさが  $n$  となる  $B$  個の残差ベクトルの群を作る。

$$\underbrace{\mathbf{e}_{11}^*, \dots, \mathbf{e}_{1n}^*}_{\mathbf{e}_1^*}, \underbrace{\mathbf{e}_{21}^*, \dots, \mathbf{e}_{2n}^*}_{\mathbf{e}_2^*}, \dots, \underbrace{\mathbf{e}_{B1}^*, \dots, \mathbf{e}_{Bn}^*}_{\mathbf{e}_B^*}$$

5. ブートストラップ標本  $X_{bjk}^*, b = 1, \dots, B$  は  $\mu_k$  の推定量  $\bar{x}_{..k} = \frac{1}{m} \sum_{i=1}^m \bar{x}_{i.k}$  を用いて

$$X_{bjk}^* = \bar{x}_{..k} + a e_{bjk}^*$$

となる。ここで  $a = \sqrt{n/(n-1)}$ ,  $n > 1$  である。

6. ブートストラップ標本  $\mathbf{X}_{b1}^*, \dots, \mathbf{X}_{bn}^*$  からブートストラップ統計量  $T_b^* = -\log |R_b^*|$  を計算する。ここで  $R_b^*$  はブートストラップ標本の標本相関係数行列である。
7. パーセンタイル法により管理限界線を構成する。ブートストラップ |R| 管理図における信頼水準  $100 \times (1 - \alpha)\%$  の上側管理限界線 (UCL) はブートストラップ統計量  $T_b^*, b = 1, \dots, B$  の  $1 - \alpha/2$  パーセント点  $T_{b,(1-\alpha/2)}^*$  であり, 下側管理限界線 (LCL) は  $\alpha/2$  パーセント点  $T_{b,(\alpha/2)}^*$  となる。また中心線 (CL) は元のデータの相関係数行列の行列式を対数変換したものの平均である。

$$\begin{aligned} UCL &= T_{b,(1-\alpha/2)}^* \\ CL &= \frac{1}{m} \sum_{i=1}^m -\log |R_i| \\ LCL &= T_{b,(\alpha/2)}^* \end{aligned}$$

## 6 移動相関係数行列の行列式による管理図

秋月 [1] は 2 変量のデータに対して移動相関を管理する移動相関管理図を提案した。本研究ではこれを多変量に拡張したものとして移動相関係数行列の行列式による管理図を提案する。その構成法は以下の手順で行う。

1. 群の数が  $m$ , 標本の大きさが 1, 特性の数が  $p$  であるデータ  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip}), i = 1, 2, \dots, m$  が得られたとする。(時系列データであるならば, 前日との差分を取る等を行い, データが過去に依存しないようにする)
2. あるブートストラップ反復係数  $B$  を与えて, 標本ベクトル  $\mathbf{x}_i, i = 1, 2, \dots, m$  に対して無作為復元抽出を行い, ブートストラップ標本  $\mathbf{x}_b^*, b = 1, 2, \dots, B$  を作成する。
3. ブートストラップ標本  $\mathbf{x}_b^*, b = 1, 2, \dots, B$  を用いて, ウィンドウ幅  $w$  の移動相関係数行列を計算する。

$$R_j^* = \begin{pmatrix} r_{1,1}^{*(j)} & r_{1,2}^{*(j)} & \dots & r_{1,p}^{*(j)} \\ & r_{2,2}^{*(j)} & & \vdots \\ & & \ddots & r_{p-1,p}^{*(j)} \\ & & & r_{p,p}^{*(j)} \end{pmatrix}$$

ただし,  $r_{k,k'}^{*(j)}, k = 1, 2, \dots, p, k' = 1, 2, \dots, p, j = w, w+1, \dots, B$  は, ブートストラップ標本から計算される第  $j$  時点での第  $k, k'$  成分間の移動相関であり,  $(x_{bk}, x_{bk'}^*), b = j-w+1, j-w+2, \dots, j$  の標本相関係数である。

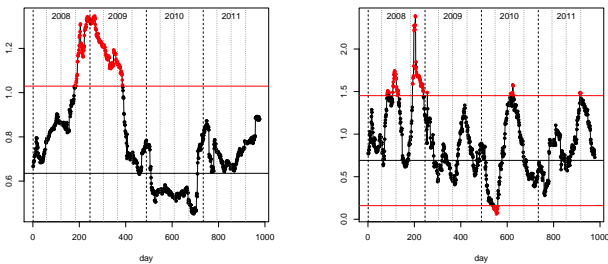
4.  $R_j^*$  の行列式  $|R_j^*|$  を計算し, それを対数変換したものをブートストラップ統計量  $T_j^* = -\log |R_j^*|, j = w, w+1, \dots, B$  とする。
5. 移動相関係数行列の行列式による管理図における信頼水準  $100 \times (1 - \alpha)\%$  の上側管理限界線 (UCL) はブートストラップ統計量  $T_j^*$  の  $1 - \alpha/2$  パーセント点  $T_{j,(1-\alpha/2)}^*$  であり, 下側管理限界線 (LCL) は  $\alpha/2$  パーセント点  $T_{j,(\alpha/2)}^*$  となる。中心線 (CL) は元のデータの移動相関係数行列の行列式を対数変換したものの平均である。

$$\begin{aligned} UCL &= T_{j,(1-\alpha/2)}^* \\ CL &= \frac{1}{m-w+1} \sum_{j=1}^{m-w+1} -\log |R_j| \\ LCL &= T_{j,(\alpha/2)}^* \end{aligned}$$

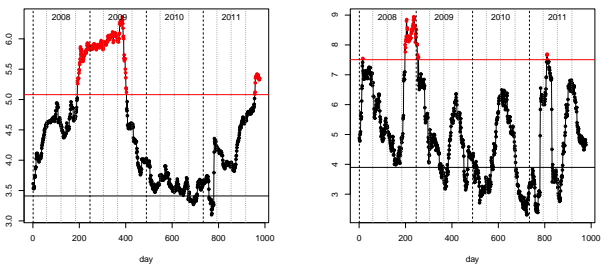
実データによる移動相関係数行列の行列式による管理図の適用例として, YAHOO!ファイナンス [7] の自動車メーカーの株価を用いる。用いたデータはトヨタ, 日産, 三菱, マツダ, ホンダ, 富士, スズキ, ダイハツの 8 社である。使用したデータの範囲は 2008 年 1 月初めから 2011 年 12 月末までで, 1 日ごとの終値を用いる。管理限界線の作成には 2007 年のデータを用いる。

2 社 (トヨタ, 日産) による結果を図 1a, 図 1b に, 8 社の結果を図 1c, 図 1d に示す。全ての図で 2008 年後半に

管理外に飛び出している。これはアメリカでのサブプライム問題、リーマン・ショックが原因と考えられる。図 1b の 2010 年初めの飛び出しはトヨタのリコール問題と考えられ、その次の飛び出しは急激な円高が原因と考えられる。図 1b、図 1c の 2011 年後半の飛び出しは欧州債務危機等から生じた世界同時株安が原因と考えられる。また、図 1d の 2011 年前半の飛び出しは東日本大震災が原因と思われる。



(a) 2 社データ (ウィンドウ幅 200) (b) 2 社データ (ウィンドウ幅 50)



(c) 8 社データ (ウィンドウ幅 200) (d) 8 社データ (ウィンドウ幅 50)

図 1: 移動相関係数行列の行列式による管理図

## 7 |R| 累積和管理図

管理図の一種として累積和統計量を管理対象とする累積和管理図がある。累積和管理図は比較的小さな変化に鋭敏であることが知られている。そこで相関係数行列の行列式 |R| に対して累積和統計量をとることが考えられるが、一般的な累積和管理図は正規分布の仮定を置いているため、そのままでは使用することができない。そのため Chatterjee and Qiu[2] の Distribution-Free 累積和管理図を適用することで、|R| 累積和管理図を構成する。

累積和管理図における統計量は

$$C_i^+ = \max(0, \theta_i - k + C_{i-1}^+), C_0^+ = 0$$

$$C_i^- = \max(0, -k - \theta_i + C_{i-1}^-), C_0^- = 0$$

となる。 $C^+$  は正の方向の変化を、 $C^-$  は負の方向の変化を検出するための統計量であり、それぞれ上側累積和統計量、下側累積和統計量と呼ばれる。 $\theta_i$  は通常の累積和管理図では標本値や標本平均が用いられるが、|R| 累積和管理図では相関係数行列の行列式を用いる。以降の説明では  $C^+$  のみを対象とするが、 $C^-$  についても同様と考

えてよい。ここで新たな統計量  $T_i$  を導入する。その定義は次のようになる。

$$T_i = \begin{cases} 0, & C_i^+ = 0 \\ j, & C_i^+ \neq 0, \dots, C_{i-j+1}^+ \neq 0, C_{i-j}^+ = 0 \end{cases}$$

$$j = 1, 2, \dots, i$$

つまり、 $T_i$  は累積和統計量  $C_i^+$  が最後に 0 であった所から何点経過したかを表しており、 $T_i$  を sprint length と呼ぶ。Distribution-Free 累積和管理図での管理限界線は  $T_i$  に応じて決められる。 $T_i = j, j = 1, 2, \dots, j_{max}$  である  $C_i^+$  に対しては管理限界線  $h_j, j = 1, 2, \dots, j_{max}$  が用いられる。 $T_i$  が  $j_{max}$  を超える  $C_i^+$  に対しては管理限界線  $h^*$  を用いる。ここで  $j_{max}$  は事前に決めておく定数であるが、Chatterjee and Qiu[2] は多くても 50 程で良いとしている。

|R| 累積和管理図の構成法の手順を簡単に説明する。

1. 目標とする平均的な  $T_i$ , ASL(average sprint length) を決める。
2. ブートストラップ法を用いて  $C_b, b = 1, 2, \dots, B$  を計算し、 $T_b$  の平均が ASL と十分近くなるように  $k$  を反復法により調整する。
3. 求められた  $k$  を用いて、目標とする ARL(average sprint length) に十分近くなるように管理限界線  $\{h_j, j = 1, 2, \dots, j_{max}; h^*\}$  をブートストラップ法と反復計算により求める。

|R| 累積和管理図の例を図 2 に示す。図の縦軸の 0 より上に  $C^+$ 、下に  $C^-$  が打点されている。また図の曲線が管理限界線を表している。|R| 累積和管理図では  $C^+$  と  $C^-$  の  $k$  は異なる値をとるため  $C^+$  と  $C^-$  が共に 0 より大きい箇所がある。

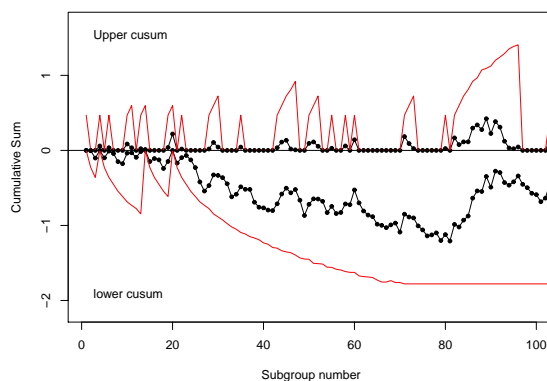


図 2: |R| 累積和管理図

## 8 シミュレーション

ブートストラップ |R| 管理図、|R| 累積和管理図が有意水準を守っているかを検証する。特性  $p = 2, 6$  の 2 種類の多変量正規分布  $N(\mu, \Sigma)$  に従うデータを用いて、ブートストラップ |R| 管理図では有意水準 0.0027, |R| 累積和

管理図では有意水準水準 0.0027 に対応する  $ARL = 370$  を満たす管理限界線を作成する (有意水準  $\alpha$  と  $ARL$  には  $\alpha = 1/ARL$  の関係がある)。また 2 変量データに対しては検出力も評価する。6 変量正規分布のパラメータ  $\mu, \Sigma$  は

$$\mu = \begin{pmatrix} 1 \\ 2 \\ 2 \\ 3 \\ 1 \\ 3 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0.41 & 0.64 & 0.79 & 0.58 & 0.35 \\ & 1 & 0.64 & 0.43 & 0.19 & 0.48 \\ & & 1 & 0.54 & 0.42 & 0.47 \\ & & & 1 & 0.49 & 0.37 \\ & & & & 1 & 0.30 \\ & & & & & 1 \end{pmatrix}$$

を用いる。2 変量正規分布では標準状態の相関係数を  $\rho_0 = 0.6$  とする。また検出力の評価では相関係数を  $\rho_1 = 0.4, \rho_2 = 0.5, \rho_3 = 0.7, \rho_4 = 0.8$  の 4 種類に変化させることで検出力を確かめる。ここで検出力とは変化が生じた後、平均何点目で検出できるかを表す。

ブートストラップ |R| 管理図において検証に用いる群の数を 100000, 管理限界線の作成で用いる群の数  $m = 30$ , 標本の大きさ  $n = 20, 50$ , ブートストラップ反復係数  $M = 1000$  として, 上記の設定のもとでシミュレーションを行う。また |R| 累積和管理図において, 管理限界線の作成で用いる群の数  $m = 30$ , 標本の大きさ  $n = 20, 50$ , 目標とする ASL を 5 とし, シミュレーションを行う。シミュレーション回数は共に計算時間を考量して 100 回とする。

シミュレーションの結果を表 1, 2 に示す。表の括弧内の数字は標準偏差を表す。表 1a, 2a の結果よりブートストラップ |R| 管理図, |R| 累積和管理図ともにそれぞれの目標とする有意水準,  $ARL$  より少し甘い値となっている。また, どちらの管理図も第 1 種の過誤の点で大きな差はない。一方, 検出力は表 1b, 表 2b より, ブートストラップ |R| 管理図は  $n = 50$  の時で相関係数の変化が  $\pm 0.2$  である  $\rho_1, \rho_4$  では早く検出できているが, 相関係数の変化が  $\pm 0.1$  である  $\rho_2, \rho_3$  では検出力が悪い。一方, |R| 累積和管理図は非常に早く検出できている。

表 1: ブートストラップ |R| 管理図

(a) 第 1 種の過誤

$n$	2 変量データ	6 変量データ
20	0.0032(0.0010)	0.0032(0.0012)
50	0.0031(0.0009)	0.0034(0.0010)

(b) 検出力 (2 変量データ)

$n$	$\rho_1$	$\rho_2$	$\rho_3$	$\rho_4$
20	59.8(30.1)	164.0(74.9)	118.1(65.2)	13.9(6.0)
50	8.7( 2.9)	52.1(25.0)	34.4(16.8)	2.6(0.5)

## 9 まとめ

移動相関係数行列の行列式による管理図が一部の株価データに対してのみでしか検証を行っていないが, 有用

表 2: |R| 累積和管理図

(a)  $ARL$

$n$	2 変量データ	6 変量データ
20	299.8(114.4)	313.4(145.7)
50	296.7(111.9)	303.4(118.9)

(b) 検出力 (2 変量データ)

$n$	$\rho_2$	$\rho_3$
20	40.4(31.0)	31.3(20.4)
50	12.1( 4.9)	9.2( 3.2)

なものであることがわかった。また, シミュレーションによりブートストラップ |R| 管理図および, |R| 累積和管理図の管理限界線が第 1 種の過誤を守っていることを示すことができた。|R| 累積和管理図はブートストラップ |R| 管理図より検出力の点で非常に優れていた。しかし, |R| 累積和管理図は累積和統計量の計算や複数の管理限界線があることなど, 実際に使用する際には非常に複雑な手法であることが欠点であると言える。

## 10 おわりに

この研究を通して, 各管理図が有用なものであるとわかった。今後の課題としては, シミュレーションで用いるデータの吟味である。今回多変量データとして多変量正規分布を用いたが, そのパラメータについての議論は不十分であるため, しっかりと吟味をする必要がある。

## 参考文献

- [1] 秋月良康: 移動相関管理図の構成とその有用性の研究, 南山大学数理情報研究科修士論文, 2012.
- [2] Chatterjee, S. and Qiu, P. : Distribution-Free Cumulative Sum Control Charts Using Bootstrap-Based Control Limits. *The Annals of Applied Statistics*, **3**, 349–369, 2009.
- [3] Efron, B. : Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7**, 1–26, 1979.
- [4] Hawkins, D. M. : Cumulative Sum Control Charting: An Underutilized SPC Tool. *Quality Engineering*, **5**, 463–477, 1993.
- [5] Montgomery, D. C. : *Introduction to Statistical Quality Control*. John Wiley & Sons, 2001.
- [6] Seppala, T. and Moskowicz, H., Plante, R. and Tang, J. : Statistical Process Control via the Subgroup Bootstrap. *Journal of the Quality Technology*, **27**, 139–153, 1995.
- [7] YAHOO!ファイナンス : <http://finance.yahoo.co.jp/>, 2012.
- [8] 安井清一, 尾島善一, 仙石直人 : 相関構造の変化を検出するための多変量管理図. 日本品質管理学会第 80 回研究発表会研究発表要旨集, 259–262, 2006.