

移動相関管理図の構成とその有用性の研究

M2010MM001 秋月良康

指導教員：松田眞一

1 はじめに

品質管理などに欠かせない管理図は、多くの種類が存在するが、現在移動相関だけを使った管理図はない。私は、株価などの日々変動するが、企業同士で関係性のあるデータに対して、相関だけに注目して、相関の変化から異常を発見することができないかと思った。そこで本研究では、移動相関のみを使った管理図を構成し、その有用性や、どのようなデータに対して使用できるかを目的とする。

2 データについて

使用した実データはYAHOO!ファイナンスの時系列データ [6] から、トヨタ、日産、三菱、マツダ、ホンダ、スバル、スズキ、ダイハツの自動車メーカー 8 社の株価、カプコン、バンダイナムコ、スクエアエニックスのゲーム会社 3 社の株価、円ドル・円ユーロ・円ポンドの為替であり、気象庁の過去データ [2] から、名古屋、東京、大阪、博多、広島、札幌、静岡、仙台、鹿児島 の 9 都市の平均気温である。使用したデータの範囲は 2007 年 1 月初めから 2010 年 12 月末までで、1 日ごとのデータを使用した。

3 管理図とは

管理図は、工程における異常を検出するため、シュハート (W.A.Shewhart) により 1926 年に発表された手法である。管理図は、1 本の中心線と、上部管理限界線 (UCL:Upper Control Limit) と下部管理限界線 (LCL:Lower Control Limit) からなり、データを群ごとに打点 (プロット) したものである。(Peter [4] 参照) 本研究では計量値の管理図を考える。また、管理図にはいくつかの判定ルールがあるが詳細は日本工業標準調査会 [3] 参照のこと。

4 移動平均管理図とは

移動平均管理図とは連続する n 個の観測値の平均を用いて、工程水準を評価する管理図である。普通、移動平均管理図と書かれている場合、大抵が単純移動平均を使用した管理図を指している。平均の計算に当たっては、 $n+1$ 個の中で一番古いデータを、現時点でのデータに置き換える。この場合の移動平均管理図では重み付けされない残存効果が n 点続くという短所を持つ。移動平均にはこのほかに、個々に重み付けを行った加重移動平均や、指数関数的に重みを減少させる指数移動平均などがある。(日本工業標準調査会 [3] 参照)

5 移動相関とは

移動相関とは、移動平均を作成するのと同じように、データの初めから、 n 組のデータの標本相関係数を求め、データを 1 つずらして、データの最後まで標本相関係数を求めることである。

6 移動相関管理図

私の考えた移動相関管理図は上記で書いた移動相関の始めのデータから一定数のデータの平均を中央線とし、この平均をフィッシャーの Z 変換にかけて $3\sqrt{\frac{1}{n-3}}$ を加えたものをフィッシャーの Z 変換の逆変換した値を UCL とし、引いて逆変換した値を LCL として引いたものである。

7 類似の研究

加納ら [1] では、SPC チャートに変わる運転監視方式として、主成分分析を用いる方法があるが、この方法では変数間の関係 (運転状態) が変化しても、運転データが管理限界内に存在する限り、その変化を検出することができないという点に対して、運転状態がデータの分布に反映されることに着目し、データの分布の変化を監視する方法を提案している。この論文と私の論文の違いは、私の研究では相関の関係に注目しているが、この研究では位置の動きを監視しているという点である。

8 自動車メーカー株価における移動相関管理図の結果

ウィンドウ幅を 200 日として移動相関管理図 (図 1) を作成したが大部分で管理限界線を超えているのがわかる。このままでは管理図として成り立たないので前日との差分をとり、その差分を用いた差分移動相関管理図 (図 2、図 3、図 4) を作成した。図 2 では一か所だけ長期にわたって管理限界線を超えているところがある、ここは他のメーカー同士でやった場合でも同じように飛び出している。この時期はアメリカでサブプライム問題やリーマンショックが起こったため、これが原因と考えられる。図 3 では 4ヶ所飛び出ている。最初の 2ヶ所はサブプライム問題とリーマンショックが分かれたのではないかと考えられる。次の下に飛び出ている部分はトヨタのリコール問題と考えられるが、その後の上に飛び出ている部分の異常は原因を発見することが出来なかった。図 4 ではより多くの部分で飛び出しており上で述べた以外の異常の原因を発見することはできなかった。

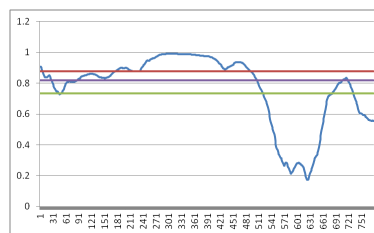


図 1 トヨタと日産の移動相関管理図

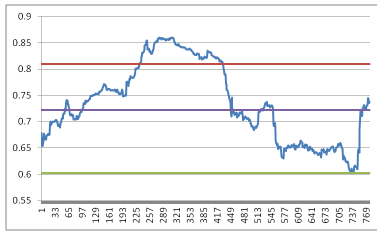


図 2 トヨタと日産のウィンドウ幅 200 日の差分移動相関管理図

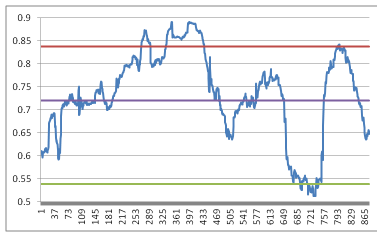


図 3 トヨタと日産の 100 日間での差分移動相関管理図

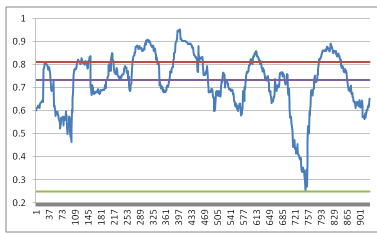


図 4 トヨタと日産の 50 日間での差分移動相関管理図

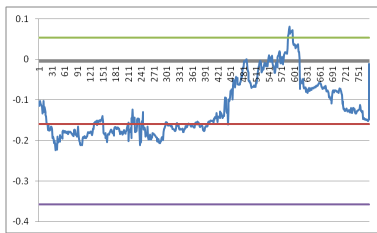


図 5 円ドルと円ユーロの 200 日間での差分移動相関管理図

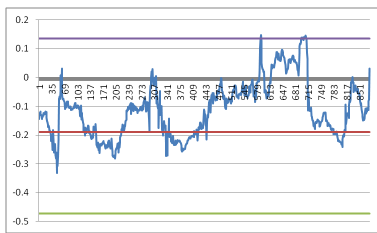


図 6 円ドルと円ユーロの 100 日間での差分移動相関管理図

9 為替データにおける移動相関管理図の結果

為替の場合も同じようにまずウィンドウ幅 200 日の移動相関管理図を作ったが、ほとんどのところで管理限界線を超えている。よって自動車メーカーと同様に差分移動相関管理図 (図 5、図 6) を作った。図 5 では 1 箇所飛び出ているところがある。これはトヨタのリコール問題が発生した時期で、それによって円が世界的に下がったことによって発生した異常と推測できる。次に図 6 では、2ヶ所飛び出している、これはトヨタのリコール問題の 2009 年 11 月と 2010 年 1 月の 2 回行われた時と考えることが出来る。50 日では、図 6 の間にもう一か所飛び出している、この理由を見つけることは出来なかった。このことから、為替ではウィンドウ幅 100 日を使うのがいいと考えられる。

10 気象データの結果

日本は四季があるのでウィンドウ幅を 90 日とするのが妥当であると判断して移動相関管理図にかけたが、多くのところで限界線を超えており、使用不可であった。よって差分を取ったが、季節ごとの変動がうかがえるため使うことができないのもう一度差分をとり 2 階差分移動相関管理図を求めたが、緩和されたものの、まだ季節ごとの変動があるため、ウィンドウ幅 90 日は使用不可と判断した。そのため季節ごとの変動を回避するためウィンドウ幅を 365 日に設定した結果が図 7 と図 8 で、図 7 の管理限界線を超えている場所はそれぞれ、エルニーニョ現象とナニーニャ現象が活発だったところである。また、図 7 と図 8 をみると二つの地域が近いほど相関が高くなっているのがわかる。

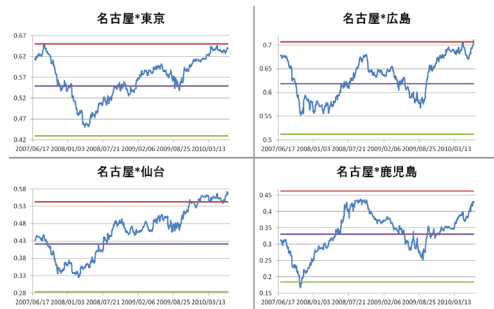


図 7 ウィンドウ幅 365 日で管理限界線を超える

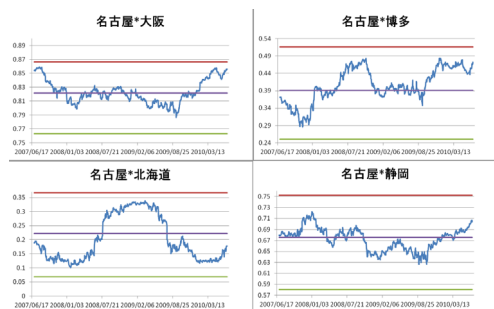


図 8 ウィンドウ幅 365 日で管理限界線を超えない

11 実データにおけるまとめ

全体を通してまとめると、どのデータでも移動相関管理図は使うことができない。このことから蓄積性のあるデータは差分を取らないと移動相関管理図を使用できないと考えられる。株価や為替ではウィンドウ幅が200のときは世界規模レベルの異常時で管理限界線を超えることがほとんどで、ウィンドウ幅が100のときはいくつか理由を発見することができないものはあるものの、ほとんどが理由付けすることができるが、ウィンドウ幅が50の場合だと理由付けが出来ない物が多々あり、使用が難しいと考えられる。また、季節変動など非常に強い影響力のある周期性の物を移動相関管理図で使用する場合、それにそったウィンドウ幅で移動相関管理図を構成しないと周期性につられて移動相関管理図が変動してしまうので、注意が必要である。

12 シミュレーション用のプログラム

1. r2norm: 指定した相関をもつ2つの正規乱数の生成
2. countmc: 指定したシード値までの移動相関管理図を構成し、管理限界線を超えた数とそのシード値を返還
3. mcc: 移動相関管理図を構成し、管理限界線を超えた場合その数とその場所を返還
4. nodifference: 差分のデータを差分前のデータに変換

13 株価データ等の再現

蓄積性のあるデータについては差分を取らないと移動相関管理図を構成することが出来ないかどうか、差分の状態が無相関で変化のないデータを、差分前のデータにもどし、ウィンドウ幅200、管理限界線作成の幅を365として、movingcorrelにかけて移動相関管理図を構成した結果が図9で無相関であるにもかかわらず、大きく変化している。

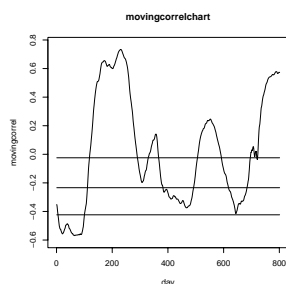


図9 蓄積性のあるデータの移動相関管理図

また、引数を変更したり、相関のあるデータなど、いろいろ移動相関管理図を構成してみたところ、管理限界線を超える位置は変化するが、どれも様に大部分で管理限界線を超えていた。このことから、蓄積性のあるデータについては差分を取らないと移動相関管理図を使用することはできないといえる。

14 移動相関管理図の異常判定の構想

移動相関管理図は、移動相関を用いているためデータ同士の関係性が高いので、JIS規格の8つのルールのうち、ルール2~8は使用不可と考えられる。また、ルール1も、一つ飛びだしたらではなく、いくらかの様子見期間をつけるべきと考えて、ランダムシード値100、それぞれデータ数1000個で、移動相関管理図を構成し管理限界線を超える場合をみることによって様子見期間を設定する。管理限界線設定の幅365、ウィンドウ幅を50、100、200としてcountmcにかけた結果、超えた回数はそれぞれ39、32、23である。この結果から分かったランダムシード値のデータをmccにかけた結果を用いて、様子見期間を20として、管理限界線を超えた日数で調べた。この結果からウィンドウ幅が200の場合は20の様子見は妥当と考えられるが100と50は大きすぎるので、100は10で、50は5で行った。それぞれの様子見の数と様子見の範囲を加えた場合の個数と、除いた場合の個数が、表1である。この結果をまとめると様子見の範囲はウィンドウ幅の10分の1で設定するとよさそうである。異常を示している部分の数は総データ数 $(1000 - 365) * 100 = 63500$ 個の中のそれぞれ37、44、57と非常に少なく 3σ で管理限界線を引いた場合の確率より少ないので十分有用である。飛び出している期間は、ウィンドウ幅が短くなるほど、減少することが分かった。これは、管理限界線を 3σ で引いているためウィンドウ幅が短くなると管理限界線の幅が広がることに原因があると考えられる。また、様子見の範囲以上で超えている場合の個数はどれもほとんど変わらないが、様子見の範囲以下も加えた場合は、ウィンドウ幅が短くなるほど個数が増えている。このことからウィンドウ幅が短くなるほど、短期の飛び出しが増えることが分かった。

表1 結果

ウィンドウ幅	200	100	50
様子見の範囲	20	10	5
様子見の範囲以上の個数	20	23	22
範囲以下も加えた個数	34	44	57

15 移動相関管理図の異常検出力の検証

ランダムシード値は4、ウィンドウ幅 r 、50、100、200、元の相関 c 、0と0.5、相関の変化量 d 、0の場合、0.3と0.1、0.5の場合、 ± 0.2 、変化量の期間の長さ w 、50と20とし、元の相関400個、変化量の相関 w 個、元の相関400個でデータを作成し、それを使って異常検出力の検証を行った結果以下のようなことが分かった。図11のように、 c が0.5で、 d が0.3のときは、どのウィンドウ幅でも異常を検出することが出来ていない。このことと図12から、元のデータが中心線より上にある時に、下のほうの異常値が発生しても検出しにくいといえる。つまり、元のデータにある程度影響を受けるといことが分かった。このことをふまえて見ると、 r が50と100の場合は、 w が20では異常を検出出来ないが、50の場合は異常を

検出できる (図 12、図 13)。 r が 200 の場合は、 w が 20 のときも 50 のときも異常を検出できるとわかった。ただし、どのウィンドウ幅でも元のデータに影響を受けるのでそこに注意をする必要がある。

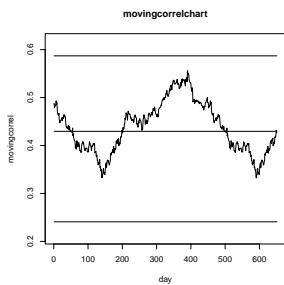


図 10 $d = 0.3$ 、 $w = 50$

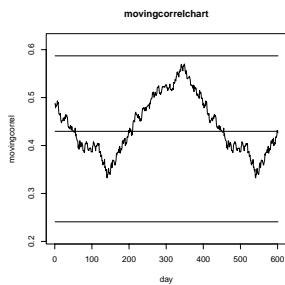


図 11 元相関のみ

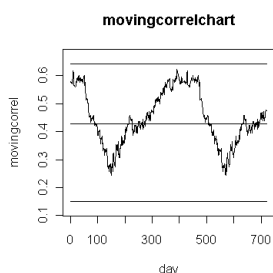


図 12 $r = 100$ 、 $c = 0.5$ 、

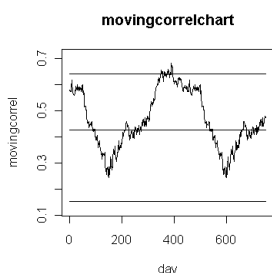


図 13 $r = 100$ 、 $c = 0.5$ 、
 $d = 0.3$ 、 $w = 20$ $d = 0.3$ 、 $w = 50$

16 シミュレーション結果のまとめ

全体を通してわかったことは、

1. 蓄積性のあるデータについては、差分移動相関管理図を使用する必要がある。
2. 様子見の範囲はウィンドウ幅の10分の1で設定する。
3. 管理限界線は 3σ で妥当である。
4. 飛び出している期間は、ウィンドウ幅が短くなるほど減少する。
5. ウィンドウ幅が短いほど、短期の飛び出しが増加。
6. ウィンドウ幅 50 と 100 は、変化後の幅が 20 では異常を検出することが出来ないが、変化後の幅が 50 の場合は異常を検出することができる。
7. ウィンドウ幅 200 の場合は変化後の幅が 20 のときも 50 のときも異常を検出することができる。
8. どのウィンドウ幅でも元のデータに影響を受けるのでそこに注意をする必要がある。

である。

17 シミュレーション結果を元に実データを再検証

シミュレーション結果をもとに、実データを見直すと、図 3 は理由の説明できなかった異常の部分に 10 日未満な

ので、異常ではない。よって、図 3 は全ての異常が説明できた。図 4 ではいくつかの異常は様子見の範囲内で減ったが、まだ説明できない異常がいくつか存在した。よって実データではウィンドウ幅は 100 日が良いと考えられる。

18 まとめ

これまでのことから、次のことがわかった。

1. 蓄積性のあるデータについては、移動相関管理図ではなく、差分移動相関管理図を使用する必要がある。
2. 季節性などの強い周期性があるデータを使用する場合は、周期性に合わせたウィンドウ幅を取らないと周期性に影響を受けるので注意が必要である。
3. 管理図の異常判定方法はウィンドウ幅の10分の1の期間様子見をして、飛び出しているならそこからさかのぼって管理限界線作成の幅分を取って引きなおす。
4. 管理限界線は 3σ で妥当である。
5. ウィンドウの幅は周期性を持たない場合は 100 がよい。
6. ウィンドウ幅が小さいほど異常の検出力は下がる。
7. 元データの移動相関の変動が激しい場合は、異常の検出力が下がる。

よって、移動相関管理図は、データを多く集めることが出来るもので、移動相関が大きく変動しないものに対しては、有用であると証明できた。移動相関管理図を使うのに適したものは、株価や為替などであると考えられる。

19 おわりに

この研究を通して、移動相関管理図は有用なものであるとわかった。しかし、様子見の期間があるのは管理するものによっては致命的な場合もあるので、異常の検出力を上げるために、移動平均管理図のように、加重移動相関管理図や、指数移動相関管理図を新たに作り出す必要があるかもしれないと思った。

参考文献

- [1] 加納 学・大野 弘・長谷部 伸治・橋本 伊織:時系列データの非類似度に基づく運転監視, 化学工学論文集, 第 25 巻, 第 6 号, pp.1004-1009, 1992
- [2] 気象庁データベース: <http://finance.yahoo.co.jp/>
- [3] 日本工業標準調査会: <http://www.jisc.go.jp/index.html>
- [4] Peter, W.M. john: Statistical Methods in Engineering and Quality Assurance, 1990.
- [5] 竹澤邦夫:統計解析フリーソフト R の備忘録頁, <http://cse.naro.affrc.go.jp/takezawa/r-tips/r.html>
- [6] YAHOO!ファイナンス: <http://finance.yahoo.co.jp/>