

1. はじめに

株価データは時系列データであり、企業間の関連が非常に強い場合、株価は傾向を持つ。しかし、ランダムに企業を選択し、その相関を見ると、その振る舞いは無相関でその予測は不可能である。近年応用統計の分野ではデータマイニングの手法が注目され、時系列データへの適用が試みられるようになった。

本研究では、国内の企業の株価データを主な対象に、データ同士の関連を検証する。56企業、外為市場4通貨、TOPIXや金の価格などの主な経済指標9つの計69のデータを対象にデータの可視化を行う。分析する期間は、2002年1月1日から2009年12月31日(2921日)の記録がある期間(1963日)で始値のみを対象とする。

以下ではTOTOを取り上げ、株価の実際の動きと移動平均線による傾向の一致を示す。

2. 研究のアプローチ

時系列データの分析には時間-周波数解析、カオス解析がしばしば用いられるが、金融分野ではテクニカル分析の一種である「移動平均線」がよく用いられる。これは設定された期間の値を集計し平均したものをその日の値とするものでいわゆる加重移動平均のもっとも簡単なモデルである。移動平均線が上昇傾向ならば株価も上昇傾向、移動平均線が下降傾向ならば株価も下降傾向であることを示していて、トレンドを視覚化することができる。また、データと移動平均線の間で特定の動きや位置関係を取った場合に、「買いシグナル」または「売りシグナル」が発生し、「グランビルの法則」により次のデータの傾向を予想できる。

次に、AICを使った最適変数決定法を適用する。AICは統計モデルのよさを評価するもので、「モデルの複雑さ(使用する説明変数の数)」と「データの当てはまりの良さ(説明変数でどれだけ説明できるのか)」の2つの観点でモデルを評価する。回帰式に定数項を含む場合は以下の式で定義されている。

$$AIC = n \left(\log \left(2\pi \frac{S_e}{n} + 1 \right) \right) + 2(p+2)$$

ここで n はサンプルサイズ、 p は説明変数の数、 S_e は残差平方和である。本研究では、データより1つを目的変数、他を説明変数とした複雑な重回帰モデルに対して変数選択は減少法でモデルを評価する。

3. 数値的な分析の結果

図1はTOTOの移動平均線(短期75日・長期200日)である。縦軸は株価のデータ、横軸は時系列を表している。図から移動平均線と株価の挙動の傾向が一致していることが確認できる。また、「グランビルの法則」にしたがっていることも示されている。

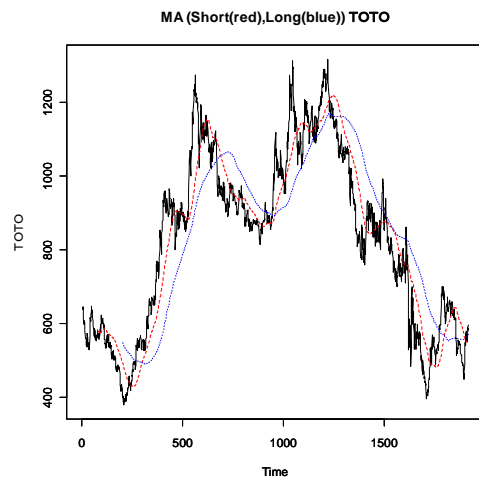


図1:TOTOの移動平均線

次にTOPIXのデータを目的変数、他の68個のデータを説明変数としたモデルを、AICを用いた最適変数決定法によって評価し、説明変数を49個とする方がよいと計算することができた。また、49個の説明変数を用いた場合でも、それぞれの相関係数が正負の両方が存在しており、回帰分析では有意でない場合もあることがわかった。

4. まとめ

本稿では、移動平均線によるデータの分析と重回帰モデルの評価を行った。重回帰モデルの評価では、変数選択にAICを用いることにより、よいモデルを決定することができた。他の変数選択を行った場合もAICで評価をして、変数選択が有意であるかの評価も可能であり、新たな変数選択法を評価する方法としての活用が考えられる。

分析に使用する企業数の増加や業種で細分化しての分析、他の変数選択法の利用が今後の課題として挙げられる。

文献

- [1] 熊谷 悦生・船尾 暢男, “Rで学ぶデータマイニング I・II”
- [2] 山田 剛史・杉澤 武俊・村井 潤一郎, “Rによるやさしい統計学”