金融市場における相関と推移

～自己回帰和分移動平均を用いて～

2007MI145 森聖也　　2007MI207 澤田康佑
指導教員: 尾崎俊治

1　はじめに

近年起きたサブプライムローン問題により, 米国住宅バブル崩壊をきっかけに, 様々な分野で資産価格の暴落が起きた. それにより, リーマン・ブラザーズも例外でなく大きな損失を抱へ, 様々な国に影響を及ぼし, リーマンショックが起きた. これらの問題によって起きた経済破綻と, 現在の NY ダウ, 日経平均, DAX，FTSE100 などの過去の時系列データに対して研究者たちはどのように取り組んでいたのか興味をもった. 将来を予測する事は, 過去のデータから何かしらの規則性を見つけ出すことにより, 将来にも反映されるであろうという期待である. そしてそのデータは定常なのか非定常なのか探る事が, 正確なモデルを選ぶ重要な岐路である. 本研究では時系列データを金融モデルにより, 比較することで各モデルの精度を検証する.

2　世界経済の破綻

2007 年頃から様々な問題により, アメリカを中心に, 様々な国で経済破綻が起きた. その主な原因が 2007 年夏頃から全体の住宅価格が下落をはじめたことから, 住宅ローンの返済の延滞率が上昇し, 住宅バブル崩壊に至った, サブプライムローン問題によるリーマンショックである. そして, ヨーロッパを中心に各国に広がり, ギリシャの経済破綻やスペインの格下げによりユーロの価値が下落しており, 2008 年 10 月 6 日から 10 日までは暗黒の一週間となり, 株価の暴落が発生し, 世界規模の恐慌に発展した. そして，日経平均株価も大暴落を起こして 6000 円台にまで下落した.

3　相関

2 つのデータ (変数), かなりの程度の規則性をもつことによって, 変化していく性質を相関という ([1] 参照).

3.1　実データの相関

現在の 4 つの株価指数の 2010 年 6 月 21 日から 7 月 20 日のデータと, 世界恐慌が起きた 2008 年 10 月 1 日から 10 月 31 日のデータの 4 つの株価指数を用いて, 相関係数の公式を元に表 1, 2 に示す. そして, どのような違いがあり, その背景には何が起きているのか検証する.

表 1　:4 指数のそれぞれの相関係数 (2010)

	日経 225	DAX	FTSE100
DAX	0.886843		
FTSE100	0.70847	0.909102	
NY ダウ	0.710756	0.88777	0.959572

表 2　:4 指数のそれぞれの相関係数 (2008)

	日経 225	DAX	FTSE100
DAX	0.937491		
FTSE100	0.945765	0.958237	
NY ダウ	0.883529	0.944864	0.913153

3.2　検証

世界恐慌の時に比べて現在は全体的に相関が小さくなっている. これは 2008 年の時に起きた世界恐慌により, 世界全体の株価が同じように下がる動きをみせたので, 2008 年の時は相関が非常に強い結果になった. 2010 年のデータでは, 日経 225 がほかの市場と比べて特に相関が弱くなっている. ともに株安とはなっているものの, ヨーロッパ圏の経済の不安定, アメリカ経済の先行きの不透明さにより市場の安定が失われていることが理由であると考えられる.

4　金融モデル

一般的には AR モデルが挙げられるが，AR モデルでは正確に時系列データを表すのは困難な場合がある．なので，様々なモデルで検証し，モデルによって正確性にどれくらい差ができて，それぞれモデルの効果や有効性を検証する．今後の指標の変化を統計学的に予測する．

4.1　AR モデル

AR モデルとは自分自身の過去を説明変数とする回帰モデルであり, AR(p) と書かれる. p 次の自己回帰構造と誤差項 u_t によって表される.

$$y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + u_t \quad t = 1, 2, \cdot\cdot, T, p \leqq t \tag{1}$$

y_t は時系列データ，ϕ_iはパラメータである．そして, u_tは
(1) 平均が 0，分散σ_2であり，互いに無相関な系列
(2) 各 t について u_t と $y_{t-k} (k = 1, 2, \cdot\cdot)$ は独立という仮定をみたすものとする
仮定を満たす確率変数列をホワイトノイズという.
また，自己相関は自己共分散を分散で割った量なので

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}$$

であるから，自己相関は

$$\begin{cases} \rho(1) = \phi_1 + \phi_2\rho(1) + \cdot\cdot + \phi_p\rho(p-1) \\ \rho(2) = \phi_1\rho(1) + \phi_2 + \cdot\cdot + \phi_p\rho(p-2) \\ \qquad\qquad \vdots \\ \rho(p) = \phi_1\rho(p-1) + \phi_2\rho(p-2) + \cdot\cdot + \phi_p \end{cases}$$

となる.

定常性を持つための条件は，特性方程式

$$\phi(y) = 1 - (\phi y + \phi_2 y^2 + \cdot\cdot + \phi_p y^p) = 0$$

の解の絶対値が 1 より大きいときである ([6] 参照).

4.2　MA モデル

MA モデルは $y(t)$ を今期から過去にさかのぼった誤差項についての移動加重和として表され，MA(q) と書かれる．

$$y_t = \mu + u_t - \sum_{i=1}^{q} \theta_i u_{t-i} \tag{2}$$

ここで y_t は時系列データ，θ_iはパラメータである．そして，u_tは平均が 0 で，分散 σ^2のホワイトノイズである．そして，平均，分散，自己相関は

$E[y_t] = \mu, Var(y_t) = \sigma^2 \sum_{i=1}^{q} \theta_i^2$

$$\rho(k) = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \cdot\cdot + \theta_{q-k}\theta_q}{1+\theta_1^2+\cdot\cdot+\theta_q^2} & (k = 1, 2, \cdot\cdot, q) \\ 0 & (k > q) \end{cases}$$

どれも t に依存しないので，MA モデルは常に定常である ([6] 参照).

4.3　ARMA モデル

ARMA モデルは回帰モデルと時系列モデルの一種である．考察の対象が同一変数の過去と現在の自己相関を求めることで導く事のできる，自己回帰モデルと移動平均モデルからなるモデルである．また AR モデル，MA モデルを単独で使用したときと比較すると，より少ないパラメータで定常過程の性質を表すことができる特徴がある．

t を時点を表す整数とすると，y_t を $E(y_t)$=0 である弱定常過程

(確率過程 $Y(t)$ が，$E(Y^2(t)) < \infty$を満たすと性質として，

(1)$E(Y(t)) = m(t$ に無関係に一定値である)

(2) 任意においた 2 時点 s と t に対して $Y(s)$ と $Y(t)$ の共分散 $E((Y(s) - m)(Y(t) - m))$ が $t - s$ で決めることができる．

この性質をもつとき $Y(t)$ を弱定常過程と呼ぶ．)

u_t を $E(u_t)$=0, $V(u_t)$=σ^2,E$(u_t u_s) = 0 (t \neq s)$

を満たすホワイトノイズとする．L を時間を戻す作用をもつ演算子である，$L^i y_t = y_{t-i}$, $L^i u_t = u_{t-i} (i = 1, 2, \cdot\cdot)$ として,$\phi(L)$,$\theta(L)$ を

$\phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i$,$\theta(L) = 1 + \sum_{i=1}^{q} \theta_i L^i$

で定義される多項式演算子とする．ここで，多項式$\phi(z) = 0$,$\theta(z) = 0$ に共通根はない．そして,$\phi_i (i = 1, 2, \cdot\cdot, p)$, $\theta_i (i = 1, 2, \cdot\cdot, q)$ はパラメータになる．そして，弱定常過程 y_tの確率変動が$\phi(L) y_t = \theta(L) u_t$これは

$$y_t = c + \phi_1 y_{t-1} + \cdot\cdot + \phi_p y_{t-p} + u_t - \theta_1 u_{t-1} - \cdot\cdot - \theta_q u_{t-q} \tag{3}$$

このモデルを次数 (p,q) の自己回帰移動平均モデルといい，ARMA(p,q) モデルになる ([2][6] 参照).

4.4　ARIMA モデル

自己回帰和分移動平均モデル（ARIMA）は，自己回帰移動平均モデル（ARMA) という時系列データに基づく統計モデルを階差をとった時系列に当てはめる統計モデルである．そして d を自然数とすると，次数 d の階差演算子である $(1 - L)^d$ を

$(1 - L)^d = \sum_{i=0}^{d} C_i (-1)^i L^i (d = 1, 2, \cdot\cdot)$ と定義する．

ここで非定常過程 x_tの d 階差である $y_t = (1 - L)^d x_t$が弱定常になるときに, x_tは階差次数である d の和文モデルに従うことになり，このモデルを $I(d)$ モデルと略記する . $y_t = (1 - L)^d x_t$が特にARMA(p, q) モデルに従い,

$$\phi(L)(1 - L)^d x_t = \theta(L) u_t \tag{4}$$

と表すことができるとき，このモデルを次数 (p,d,q) の自己回帰和分移動平均モデルと呼び，ARIMA モデルと略記します．また，ARIMA$(p,0,q)$ モデルが ARMA(p,q) と同じものになり，ARIMA(0,1,0) モデルはランダムウォークになる ([2][4][7] 参照).

4.5　ARCH モデル

経済の動きとして，ボラティリティーのショックには高い持続性があり，ボラティリティーに小さい変動が起きるとその後も同じように小さな変動が続きやすく，大きな変動が起きるとその後も同じように大きな変動が続きやすい性質をしているので，不均一な分散の時系列データが多く存在してしまう．この現象はボラティリティクラスタリングと呼ばれている．この現象がおきると，AR モデルでは正確にモデル化できないことが多い．そこで時系列データかつ，分散までもモデル化することにより不均一な時系列の構造を解決できるのが ARCH モデルである．このモデルは $y_t = \mu + u_t$ と表す．ここで μ は条件付平均であり，u_t は誤差項である．そして，分散 σ_t^2 をモデル化する．

$$\sigma_t^2 = \gamma + \sum_{i=1}^{p} \phi_i u_{t-i}^2 \tag{5}$$

これが ARCH モデルであり，ARCH(p) と記す．

また ARCH モデルが定常になるのは，$E[u_t^2] < \infty$ ならば，定常となる．定常性の条件のもとで分散は

$$E[u_t^2] = \frac{\gamma}{1 - \phi_1 - \phi_2 - \cdot\cdot - \phi_p}$$

となり，定常性が保障されるためには

$$\phi_1 + \phi_2 + \cdot\cdot + \phi_p < 1$$

というパラメータ制約を満たさなければならない．また (5) 式のσ_t^2が常に否負になるように

$$\gamma \geq 0, \phi_1 \geq 0, \cdot\cdot, \phi_p \geq 0$$

のような非負制約も必要である ([3][5] 参照).

4.6　ARCH モデルの推定

定式化すると，尤度を解析的に求めることができるのでパラメータを最尤法で推定できる．この対数尤度は

$$L(\theta) = -\frac{T-p}{2} \log 2\pi - \frac{1}{2} \sum_{t=p+1}^{T} \log \sigma t^2 -$$

$\frac{1}{2}\sum_{t=p+1}^{T}\frac{(y_t-\mu)^2}{\sigma_t^2}$

で表される ([3][5] 参照).

5　次数選択

AR モデル，MA モデル，ARMA モデルを用いた時系列検証を行う場合，時系列データから次数の p と q を定める必要があり，同定と呼ぶ．次数 p と q が決まればパラメータを最小二乗法により，推定することができる．そこで最も有名なものは，AIC 法である．AIC 法は時系列データとの適合性と，モデルの複雑さとのバランスをうまくとるために使用される．
公式は以下のようになる．

$$-2\log L+2k \tag{6}$$

L は最大尤度を示し，k は自由パラメータを示している．
公式を最小とする次数を選択する．
ここで，誤差項が独立であり，確率分布が正規分布のとき
$\mathrm{AIC}=\sum_{i=0}^{n}\log(2\pi\sigma_i^2)+2$
$=\sum_{i=0}^{n}\log\sigma_i^2+2k+n\log 2\pi$
$=\sum_{i=0}^{n}\log\sigma_i^2+2k$
n は標本数であり,σ_iは標準誤差である．
誤差項の標準誤差が等しいときには
AIC$=n\log(2\pi\sigma^2)+2k=n\log\sigma^2+2k+n\log 2\pi$
AIC$=n\log\sigma^2+2k$
と単純化することができる ([4][5] 参照).

6　最小二乗法による推定

AR モデル，MA モデル，ARMA モデルのパラメータを推定する方法である．最小二乗法による係数の推定値は以下の式になっている．

6.1　自己回帰モデル AR(p) の推定

AR(p) のモデル
$y_t=\phi_0+\sum_{i=1}^{p}\phi_i y_{t-i}+u_t$
この式を，残差の二乗和である下式を最小にする係数 $\phi_0,\phi_1,\cdot\cdot,\phi_p$を推定値
$u_t=\sum_{t=p+1}^{n}(y_t-(\phi_0+\phi_1 y_{t-1}+\phi_2 y_{t-2}+\cdot\cdot+\phi_p y_{t-p}))^2$
係数の推定値は次のように与えられる ([6] 参照).

$$\begin{pmatrix}\hat{\phi}_0\\ \hat{\phi}_1\\ \vdots\\ \hat{\phi}_p\end{pmatrix}=(X^TX)^{-1}X^T\begin{pmatrix}y_{p+1}\\ y_{p+2}\\ \vdots\\ y_n\end{pmatrix}$$

$$\text{ここで},X=\begin{pmatrix}1 & y_p & \cdot\cdot & y_1\\ 1 & y_{p+1} & \cdot\cdot & y_2\\ \vdots & \vdots & & \vdots\\ 1 & y_{n-1} & \cdot\cdot & y_{n-p}\end{pmatrix}$$

6.2　移動平均モデル MA(q) の推定

MA(q) のモデル
$y_t=\mu+u_t-\sum_{i=1}^{q}\theta_i u_{t-i}$
この式は,AR(p) と異なっていて最小二乗法を使って推定することはできないため，時系列データ $y_{t\,t=1}^{\;n}$と，時刻が 0 以前の誤差項を
$e_0=e_{-1}=e_{-2}=\cdot\cdot=e_{1-q}=0$
と仮定して，残差を
$e_1=y_1-\theta_0+\theta_1 e_0+\theta_2 e_{-1}+\cdot\cdot+\theta_q e_{1-q}$
$e_2=y_2-\theta_0+\theta_1 e_1+\theta_2 e_0+\cdot\cdot+\theta_q e_{2-q}$
$\vdots$
$e_n=y_n-\theta_0+\theta_1 e_{n-1}+\theta_2 e_{n-2}+\cdot\cdot+\theta_q e_{n-q}$

として $\sum_{t=1}^{n}e_t^2$を最小化する$\theta_0,\theta_1,\cdot\cdot,\theta_q$を推定量$\hat{\theta}_0,\hat{\theta}_1,\cdot\cdot,\hat{\theta}_q$と考える．ただし，非線形関数になるため，明示的に解を求めることはできず，ニュートン法を使って計算を行う必要がある ([6] 参照).

6.3　自己回帰移動平均モデル ARMA(p,q) の推定

ARMA(p,q) は MA(q) と同様に最小二乗法は適用することは難しく，時系列データ $y_{t\,t=1}^{\;n}$と，時刻が 0 以前の誤差項を
$e_p=e_{p-1}=e_{p-2}=\cdot\cdot=e_{p-q+2}=e_{p-q+1}=0$
と仮定して，残差を
$e_{p+1}=y_{p+1}-(c+\phi_1 y_p+\cdot\cdot+\phi_p y_1)+(\theta_1 e_p+\theta_2 e_{p-1}+\cdot\cdot+\theta_q e_{p-q+1})$
$e_{p+2}=y_{p+2}-(c+\phi_1 y_{p+1}+\cdot\cdot+\phi_p y_2)+(\theta_1 e_{p+1}+\theta_2 e_p+\cdot\cdot+\theta_q e_{p-q+2})$
$\vdots$
$e_n=y_n-(c+\phi_1 y_{n-1}+\cdot\cdot+\phi_p y_{n-p})+(\theta_1 e_{n-1}+\theta_2 e_{n-1}+\cdot\cdot+\theta_q e_{n-q})$

として $\sum_{t=p+1}^{n}e_t^2$を最小化する $c,\theta_1,\cdot\cdot,\theta_p,\phi_1,\cdot\cdot,\phi_q$を推定量$\hat{c},\hat{\theta}_1,\cdot\cdot,\hat{\theta}_p,\hat{\phi}_1,\cdot\cdot,\hat{\phi}_q$と考える．ただし，非線形関数になるため，明示的に解を求めることはできず，ニュートン法を使って計算を行う必要がある ([6] 参照).

7　実データを用いた検証

2010/6/20〜2010/7/20 の日経 225，DAX，FTSE100，NY ダウの指数からそれぞれに対する最適な次数，モデルの検証を行う．

7.1　次数の検証

まず，AR，MA，ARMA，ARIMA それぞれ最適な次数について考える．まず，AIC 規準を実際に計算してみて最小となる次数はそれぞれ表 3 である．次に各次数からモデルの推定し予測をする．そして, 実際の値との残差の二乗和を比較していく．ここで残差の二乗和が小さいほどモデルの当てはまりが良い．残差の二乗和を最小となる次数は表 4 である．表 3 と表 4 でともに一致する次

数は 5 つと少なく，今回のデータは AIC 法で次数を決定することは難しいことがわかる．残差の二乗和を最小とする次数は AIC 規準が最小となる次数から並べたとき前半 5 割には属する．

表 3 :AIC 規準が最小となる次数

	AR	MA	ARMA	ARIMA
日経 225	2	2	2 , 1	3, 3, 1
DAX	3	3	3 , 1	1, 3,1
FTSE100	3	3	3 , 1	1, 3, 1
NY ダウ	1	3	3 , 2	0, 3, 2

表 4 :残差の二乗和が最小となる次数

	AR	MA	ARMA	ARIMA
日経 225	3	2	3 , 1	2,1,3
DAX	2	1	1 , 3	0,3,1
FTSE100	1	3	1 , 1	3,2,3
NY ダウ	1	3	1 , 1	0,3,2

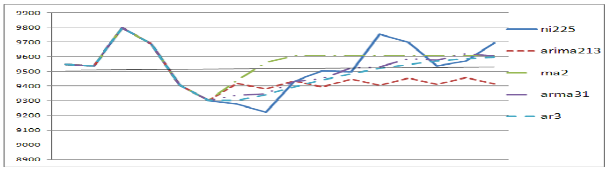

図 1 :日経 225

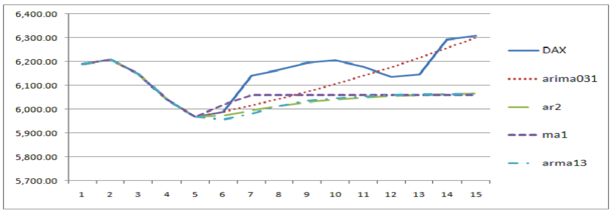

図 2 :DAX

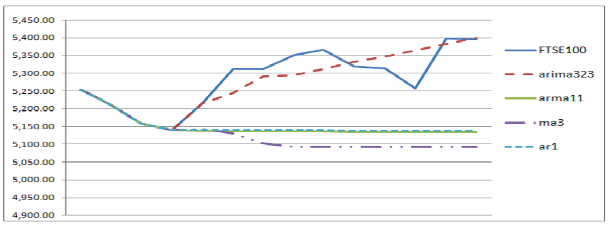

図 3 :FTSE100

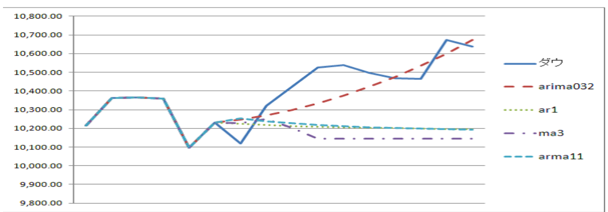

図 4 :NY ダウ

7.2 モデルの検証

すべてのモデルの内，残差の二乗和を最小とするモデルは日経 225 で ARMA(3, 1)，DAX で ARIMA(0, 3, 1)，FTSE100 で ARIMA(3, 2, 3)，NY ダウで ARIMA(0, 3, 2) である．モデルごとの残差の二乗和の平均を見てみる．ここで ARIMA は次数によっての残差の二乗和が大きく違うことと残差の二乗和を最小とする次数が AIC 規準を最小とする次数の順に並べたときに前半 5 割に属することから前半 5 割の平均とする．日経 225 は MA ＞ ARIMA ＞ AR ＞ ARMA ，DAX は ARMA ＞ AR ＞ MA ＞ ARIMA，FTSE100 は MA ＞ ARMA ＞ AR ＞ ARIMA，NY ダウは MA ＞ AR ＞ ARIMA ＞ ARMA の順になった．日経 225 は ARMA モデル，DAX は ARIMA モデル，FTSE100 は ARIMA モデルが最適なモデルであることがわかる．NY ダウについて，ARIMA モデルは全体的にばらつきがあるのに対して，ARMA モデルは平均的に適していると考える．

8 終わりに

今回はリーマンショックが起き，世界恐慌が起きた時の時系列データと現在の時系列データを用いたので，市場の変動が大きくなり，分析，予測が困難ではあったが様々なモデルを活用することにより，正確性を近づける事ができた．そして市場も 4 つの，日経 225，DAX，FTSE100，NY ダウにしぼり詳しく検証したので，このときの日本は安定していた事や，ヨーロッパ圏の経済の不安定，アメリカ経済の先行きの不透明さにより市場の安定が失われていることなどもわかった．そして，時系列データによって，モデルや次数選択の規準も様々であることがわかった．時系列データより正確に検証していくためには，各モデルの理解，応用をできるようにする事で選択肢が増え，より正確に検証できると思う．実際に時系列データは決して同じような動きは見せず，その国によって経済の不安定などで流れはつねに変化していく．だから，時系列データから予測することはとても難しい事だが，よりモデルの理解を深め，実用的な予測ができるように繋げていきたい．

参考文献

[1] 白旗慎吾：統計解析入門．共立出版株式会社，1992.

[2] J. ダービン，S.J. クープマン：状態空間モデリングによる時系列分析入門．シーエーピー出版株式会社，2004.

[3] 渡部敏明：ボラティリティ変動モデル．株式会社朝倉書店，2000.

[4] 廣松毅, 浪花貞夫：経済時系列分析．株式会社朝倉書店，1990.

[5] 小暮厚之：ファイナンスへの計量分析．株式会社朝倉書店，1996.

[6] 日笠克己：モデリング．株式会社サンワ，2009.

[7] 朝倉邦造：金融工学辞典．株式会社朝倉書店，2004.