時系列分析 1
時系列分析の基礎

担当:　長倉　大輔
(ながくらだいすけ)

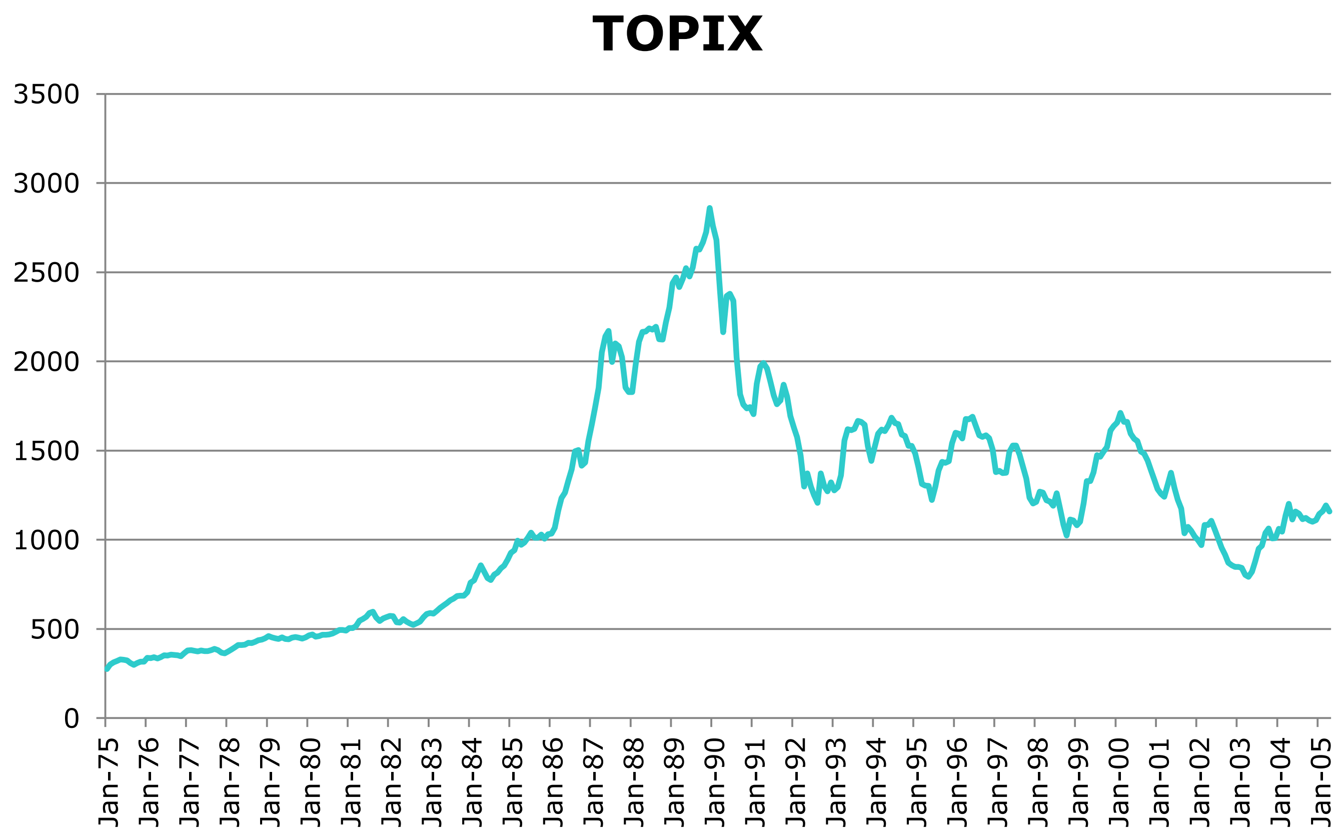
時系列分析の基礎

- 時系列データとは？

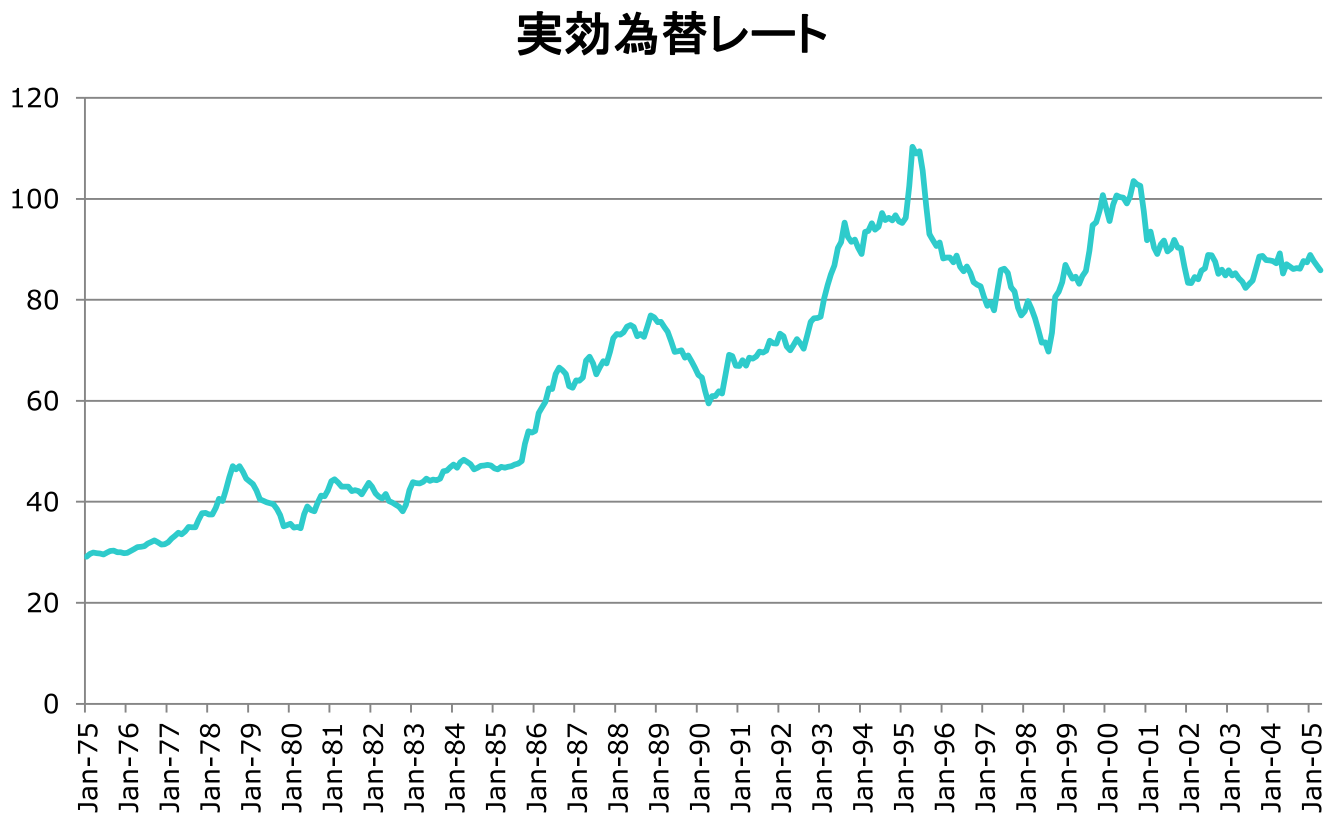
時系列データとは、時間の推移とともに観測されるデータの事です。例えば、為替レート、株価、国内総生産(GDP)、インフレ率などのようなデータです。観測される**順番**に意味のある事が大きな特徴です。

- 時系列分析とは？

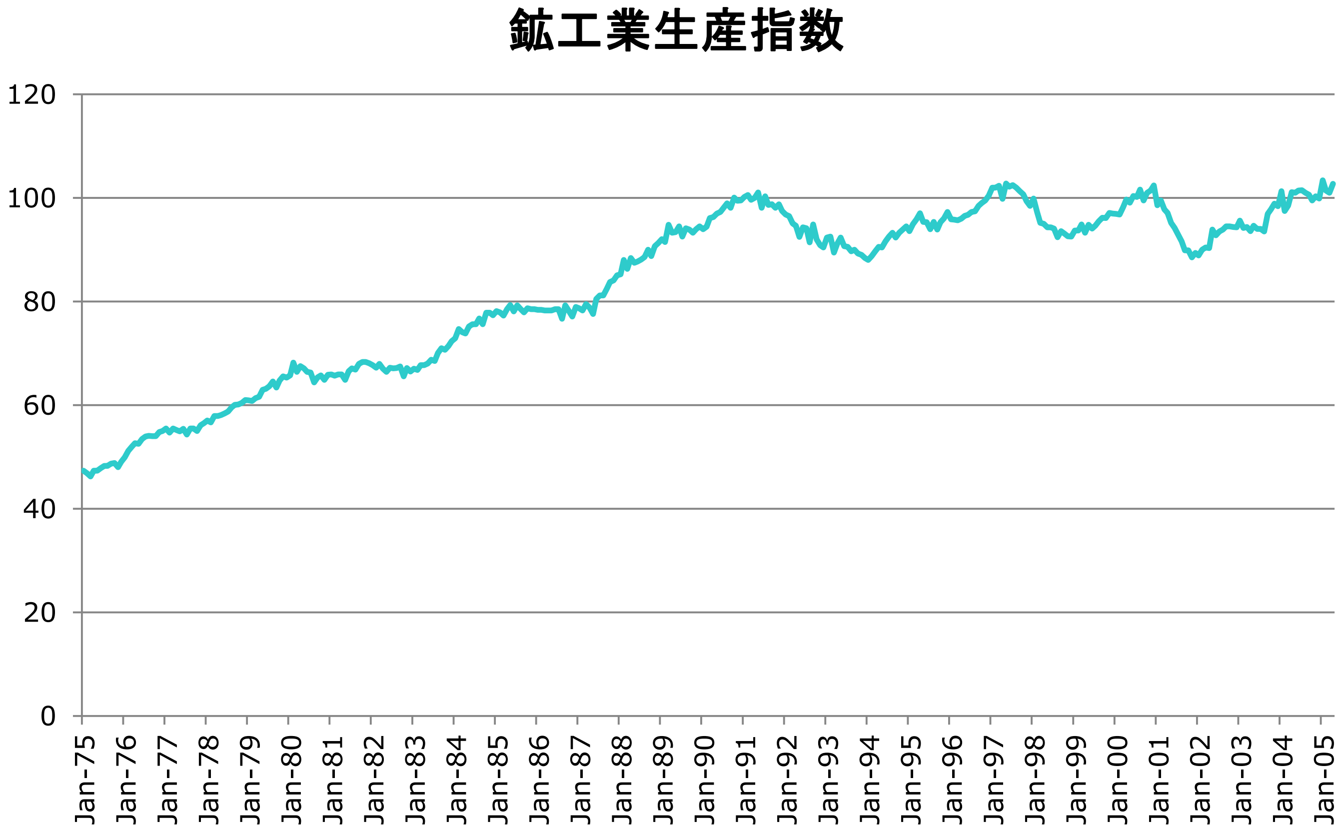
時系列分析とは時系列データを分析するための手法です。



TOPIX
3500
3000
2500
2000
1500
1000
500
0
Jan-75
Jan-76
Jan-77
Jan-78
Jan-79
Jan-80
Jan-81
Jan-82
Jan-83
Jan-84
Jan-85
Jan-86
Jan-87
Jan-88
Jan-89
Jan-90
Jan-91
Jan-92
Jan-93
Jan-94
Jan-95
Jan-96
Jan-97
Jan-98
Jan-99
Jan-00
Jan-01
Jan-02
Jan-03
Jan-04
Jan-05



実効為替レート
120
100
80
60
40
20
0
Jan-75
Jan-76
Jan-77
Jan-78
Jan-79
Jan-80
Jan-81
Jan-82
Jan-83
Jan-84
Jan-85
Jan-86
Jan-87
Jan-88
Jan-89
Jan-90
Jan-91
Jan-92
Jan-93
Jan-94
Jan-95
Jan-96
Jan-97
Jan-98
Jan-99
Jan-00
Jan-01
Jan-02
Jan-03
Jan-04
Jan-05


鉱工業生産指数
120
100
80
60
40
20
0
Jan-75
Jan-76
Jan-77
Jan-78
Jan-79
Jan-80
Jan-81
Jan-82
Jan-83
Jan-84
Jan-85
Jan-86
Jan-87
Jan-88
Jan-89
Jan-90
Jan-91
Jan-92
Jan-93
Jan-94
Jan-95
Jan-96
Jan-97
Jan-98
Jan-99
Jan-00
Jan-01
Jan-02
Jan-03
Jan-04
Jan-05

時系列分析とは？

- 時系列分析の応用

時系列分析の応用には分析の目的にもよりますが、おおざっぱに言って

1. データの特徴を捉える。
2. 経済理論などの検証
3. 時系列変数の予測

などが挙げられます。

時系列分析の基礎

- 時系列データの表し方

観測点を t で表し、その時点における観測値を y_t などと表します。時点 $t=1,\ldots,T$ までのデータの集合

$$\{\, y_1, y_2, \ldots, y_T \,\}$$

は $\{y_t\}_{t=1}^{T}$ などと表されます。

時系列分析の基礎

- 時系列データの種類

時系列データそのものは**原系列**とよばれます。原系列データに様々な変換を施して、新たな時系列データを作る事ができます。

- よく使われる変換

原系列への変換としてよく使われる変換に**対数変換**と**階差をとる**という 2 つの変換があります。対数変換とは $\log y_t$ のように y_t の対数をとる事、階差をとるとは $y_t - y_{t-1}$ のように 1 時点離れたデータとの差を取る事です。

時系列分析の基礎

- 対数差分による変化率

(通常の変化率の定義)

$$\frac{y_t - y_{t-1}}{y_{t-1}}$$

(対数差分による変化率)

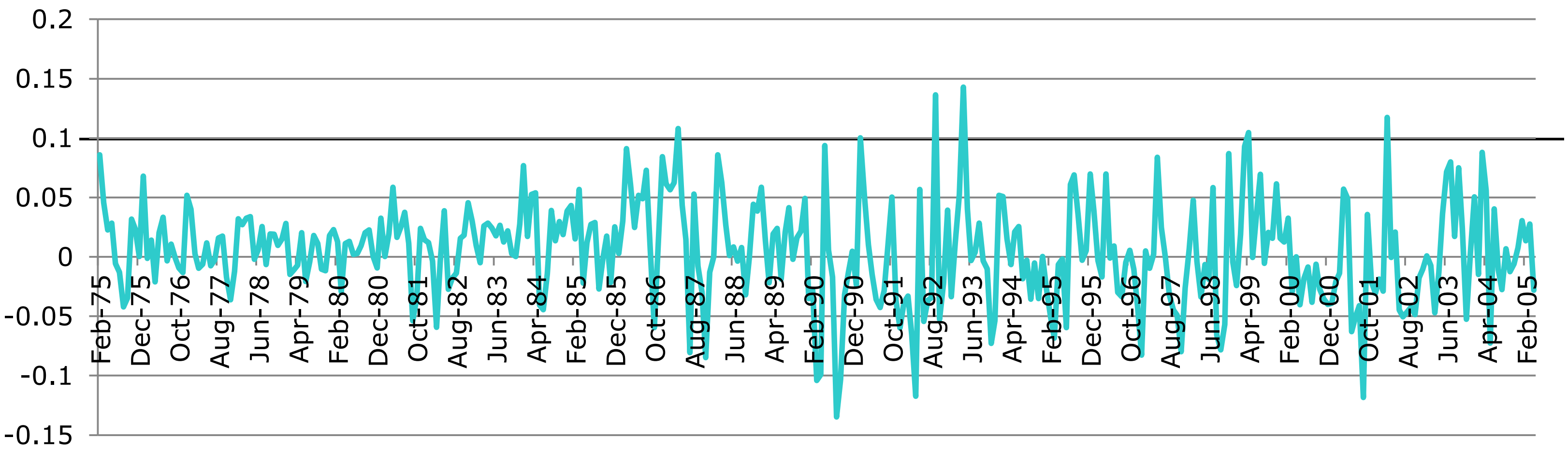
$$\log y_t - \log y_{t-1}$$

変化率が小さい時

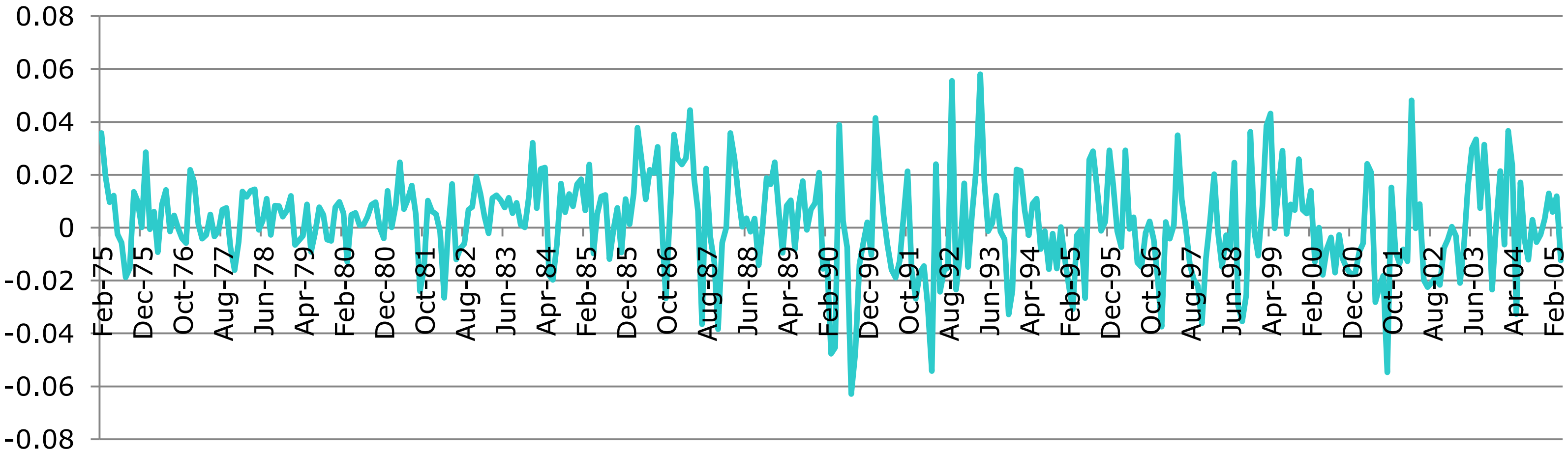
$$\log y_t - \log y_{t-1} = \log \frac{y_t}{y_{t-1}} = \log\left(1 + \frac{y_t - y_{t-1}}{y_{t-1}}\right) \approx \frac{y_t - y_{t-1}}{y_{t-1}}$$

が成り立ちます。

TOPIXの変化率



TOPIX (対数差分による)変化率



時系列分析の基礎

- 基本統計量

データの性質を要約する統計量を基本統計量といいます。時系列データに関する基本統計量として代表的なものには

1. **期待値**
2. **分散**
3. **標準偏差(ボラティィリティ)**
4. **自己共分散**
5. **自己相関係数**

などがあります。

時系列分析の基礎

- 期待値と分散

y_t の期待値は $E(y_t)$ 、分散は $\mathrm{var}(y_t)$ と表されます。

それぞれ、通常の期待値や分散と同じ定義
(つまり分散は $\mathrm{var}(y_t) = E[(y_t - \mu_t)^2]$ と定義されます。
ここで $\mu_t = E(y_t)$ です)。解釈の仕方も同じ。

注)　μ_t に下付き文字 t がついている事に注意。
一般には y_t と y_{t-1} の期待値、分散は**異なります**。

時系列分析の基礎

- 自己共分散

時点 t における k 次の自己共分散 γ_{kt} は次のように定義されます。

$$\gamma_{kt} = \mathrm{cov}(y_t, y_{t-k}) = E[(y_t - \mu_t)(y_{t-k} - \mu_{t-k})]$$

ここで $\mu_t = E(y_t)$, $\mu_{t-k} = E(y_{t-k})$ です。
つまり、**時点 t より k 時点前のy_t との共分散**です。

また $k = 0$ の時(つまり 0 次の自己共分散) γ_{0t} は定義により y_t の分散と等しくなります。

時系列分析の基礎

- 自己相関係数

時点 t における k 次の自己相関係数、ρ_{kt}、は**時点 t より k 時点前のデータとの相関係数**、すなわち

$$\rho_{kt} = \frac{\mathrm{cov}(y_t, y_{t-k})}{\sqrt{\mathrm{var}(y_t)\,\mathrm{var}(y_{t-k})}} = \frac{\gamma_{kt}}{\sqrt{\gamma_{0t}\gamma_{0,t-k}}}$$

と定義されます。

自己相関係数は $k=0$ の時に 1、$k>0$ の時に $|\rho_{kt}| \leq 1$ が成り立ちます。

定常性

- 定常性

時系列分析において最も重要な概念が**定常性**です。

ある時系列データが**定常**であるとは次の事を意味します。

全ての t について

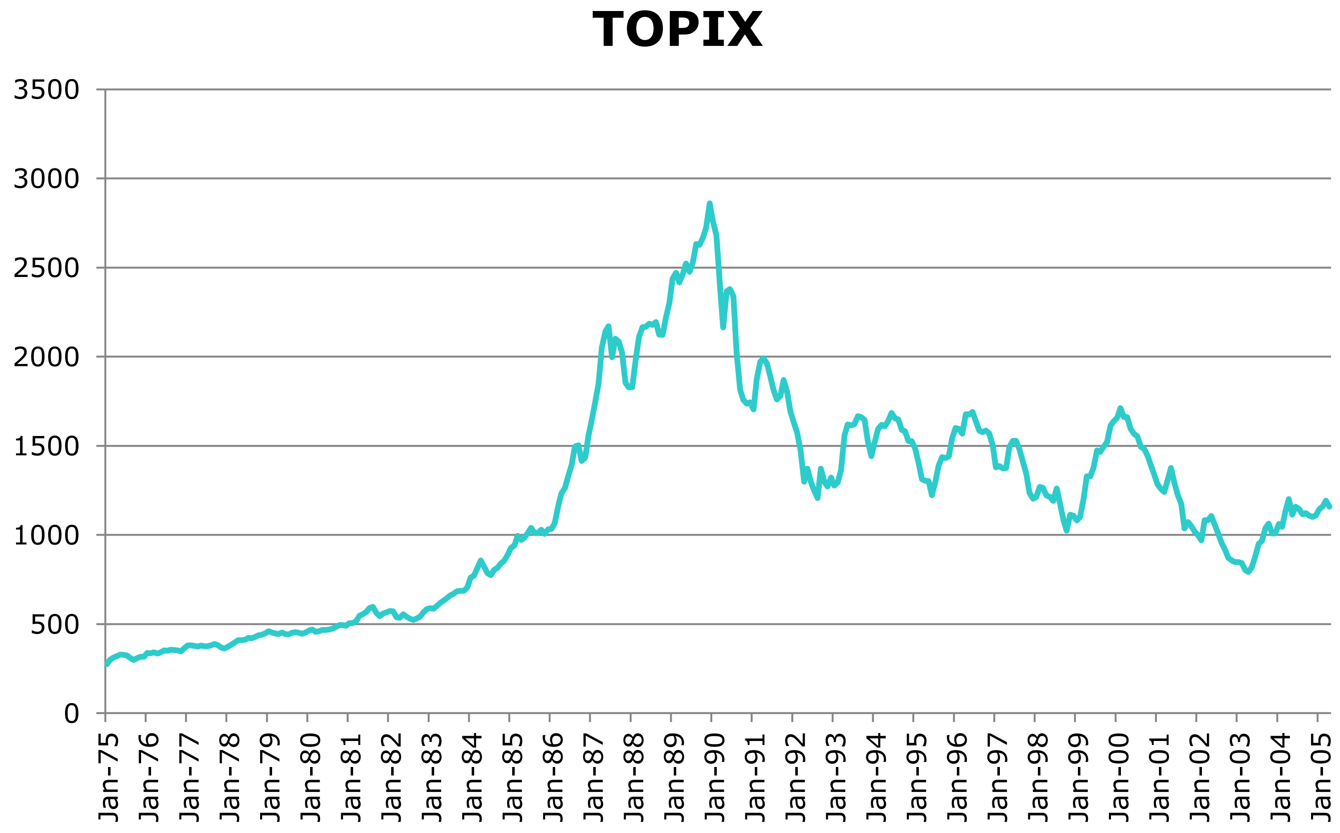
$$E(y_t) = \mu \quad \text{および} \quad \gamma_{kt} = \gamma_k \text{、}$$

ここで γ_k とは k にのみ依存した値(つまり k の関数)
を表しています。

定常性

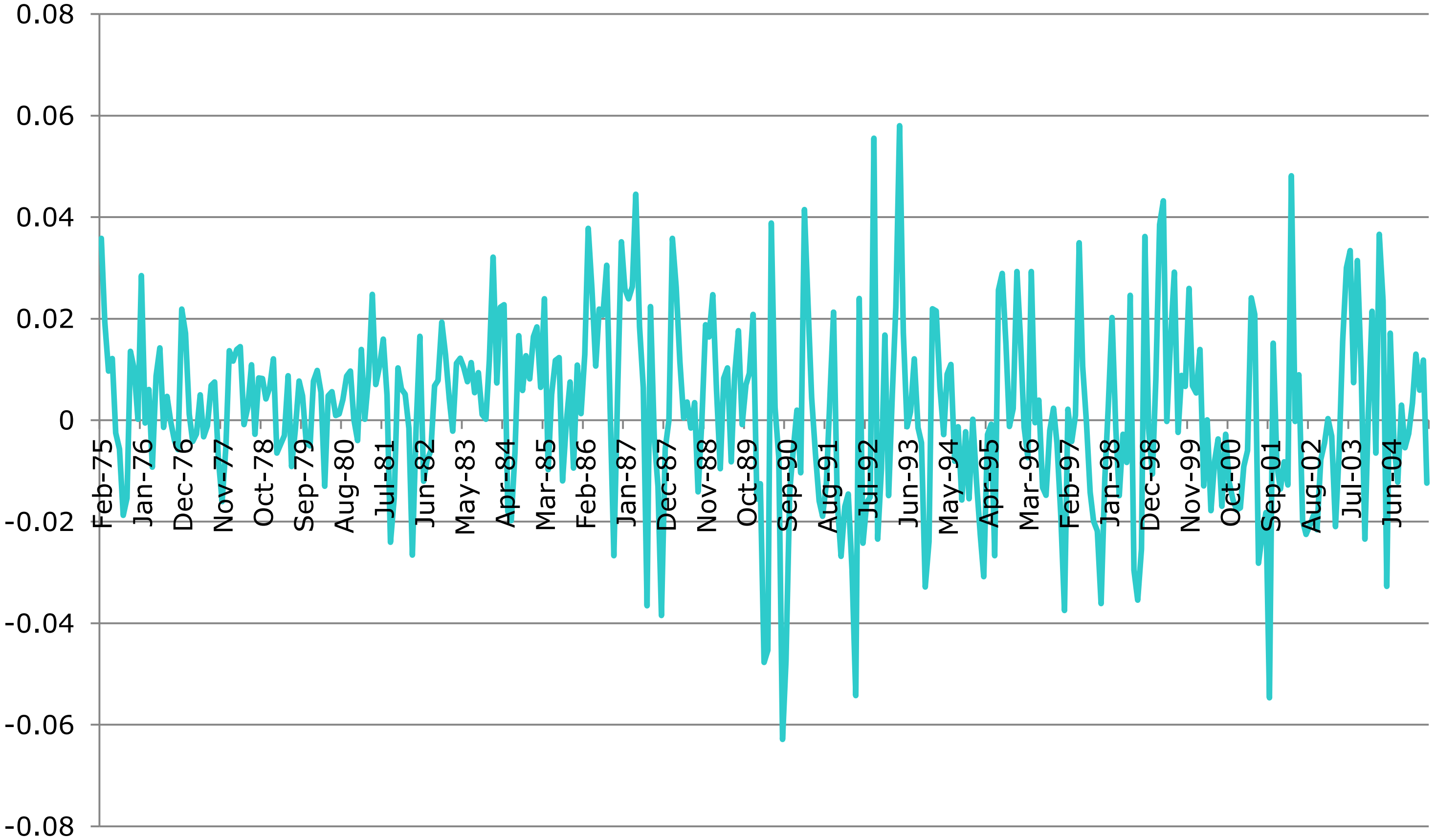
- 定常性の意味

1. 一般には y_t と y_s $(t \neq s)$の期待値や自己共分散は異なる。
2. しかし、これらを推定するためのデータ y_t は1つの時系列に付き 1 つしか観測されない(例えば2008年の11月6日のTOPIXの終値は一度しか観測されない)。
3. y_t の1つの観測値より、y_tがたくさん観測された時に平均的にとるであろう値である、期待値を精度良く推定する事はほぼ不可能。
4. 定常性の仮定をおく事により期待値の推定に(同じ期待値を持つ) y_t, y_{t-1} , … を用いる事ができるようになる。
5. ほとんどの場合にデータが定常性の仮定を満たすようにデータを変換する事ができる (例: 対数階差)。

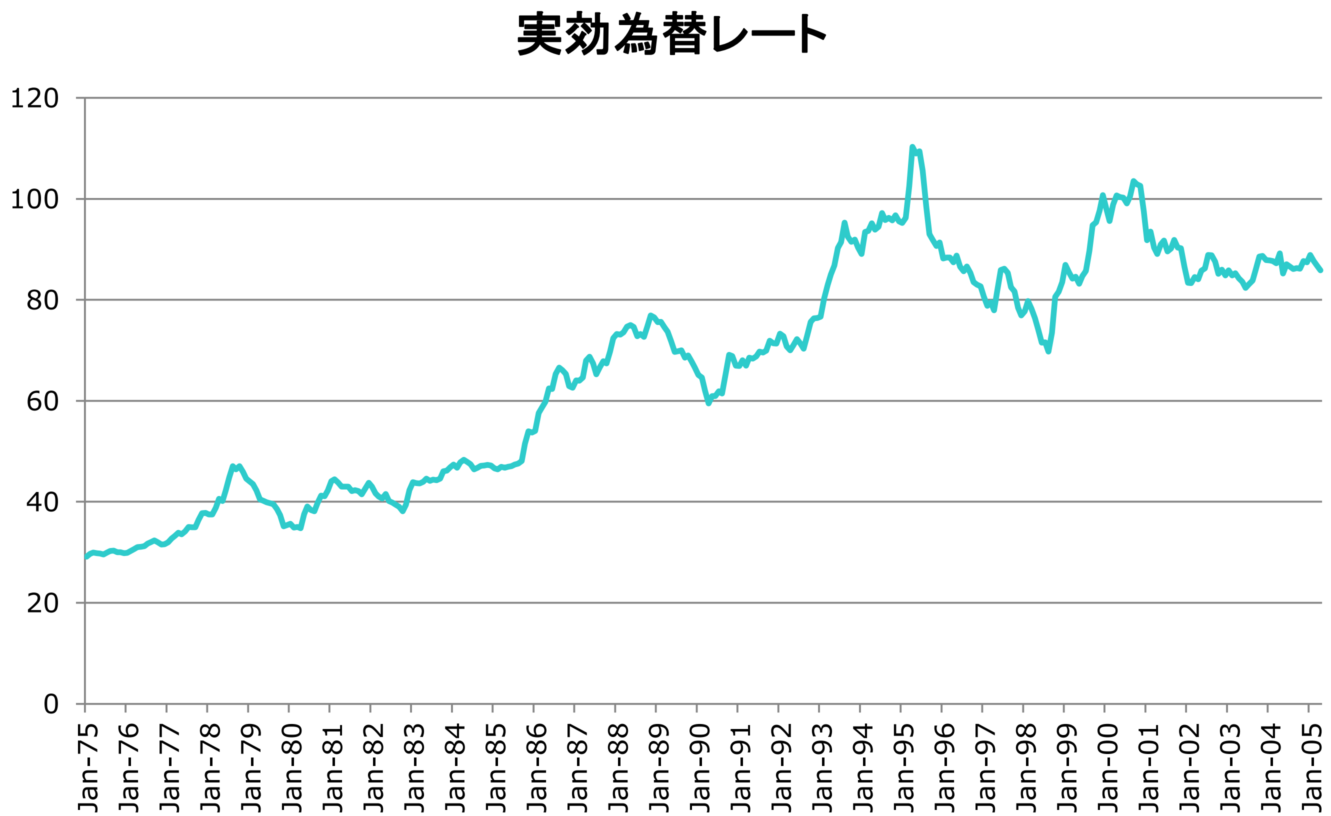


TOPIX
3500
3000
2500
2000
1500
1000
500
0
Jan-75
Jan-76
Jan-77
Jan-78
Jan-79
Jan-80
Jan-81
Jan-82
Jan-83
Jan-84
Jan-85
Jan-86
Jan-87
Jan-88
Jan-89
Jan-90
Jan-91
Jan-92
Jan-93
Jan-94
Jan-95
Jan-96
Jan-97
Jan-98
Jan-99
Jan-00
Jan-01
Jan-02
Jan-03
Jan-04
Jan-05

TOPIX 変化率

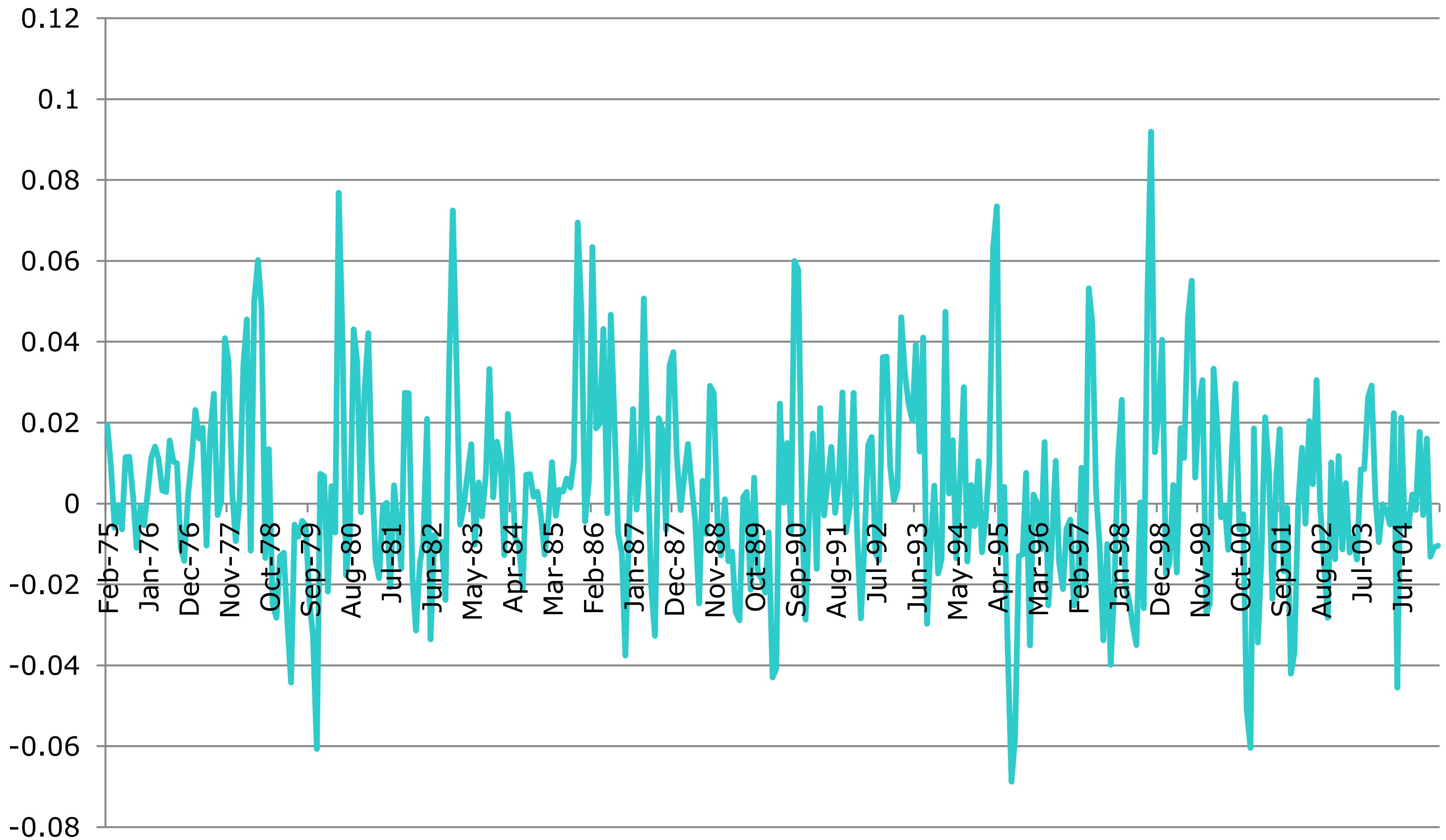


0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
Feb-75
Jan-76
Dec-76
Nov-77
Oct-78
Sep-79
Aug-80
Jul-81
Jun-82
May-83
Apr-84
Mar-85
Feb-86
Jan-87
Dec-87
Nov-88
Oct-89
Sep-90
Aug-91
Jul-92
Jun-93
May-94
Apr-95
Mar-96
Feb-97
Jan-98
Dec-98
Nov-99
Oct-00
Sep-01
Aug-02
Jul-03
Jun-04

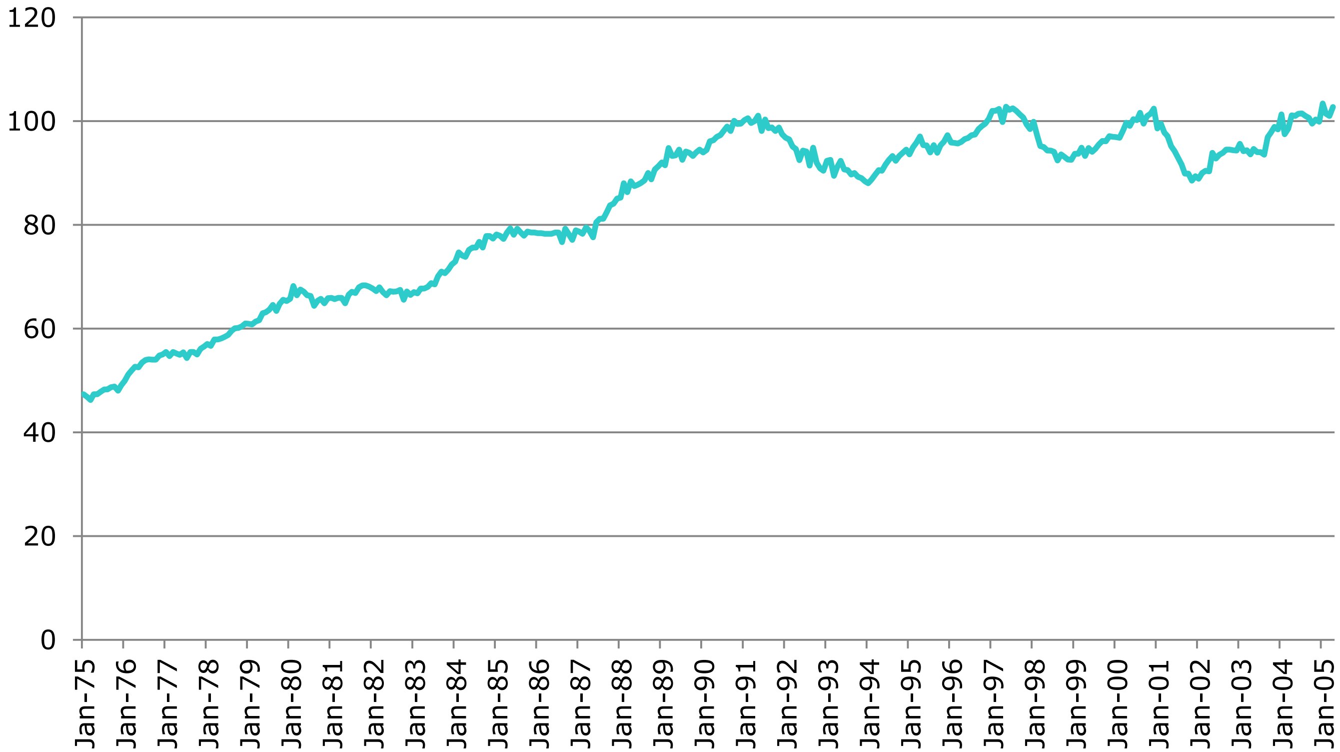


実効為替レート
120
100
80
60
40
20
0
Jan-75
Jan-76
Jan-77
Jan-78
Jan-79
Jan-80
Jan-81
Jan-82
Jan-83
Jan-84
Jan-85
Jan-86
Jan-87
Jan-88
Jan-89
Jan-90
Jan-91
Jan-92
Jan-93
Jan-94
Jan-95
Jan-96
Jan-97
Jan-98
Jan-99
Jan-00
Jan-01
Jan-02
Jan-03
Jan-04
Jan-05

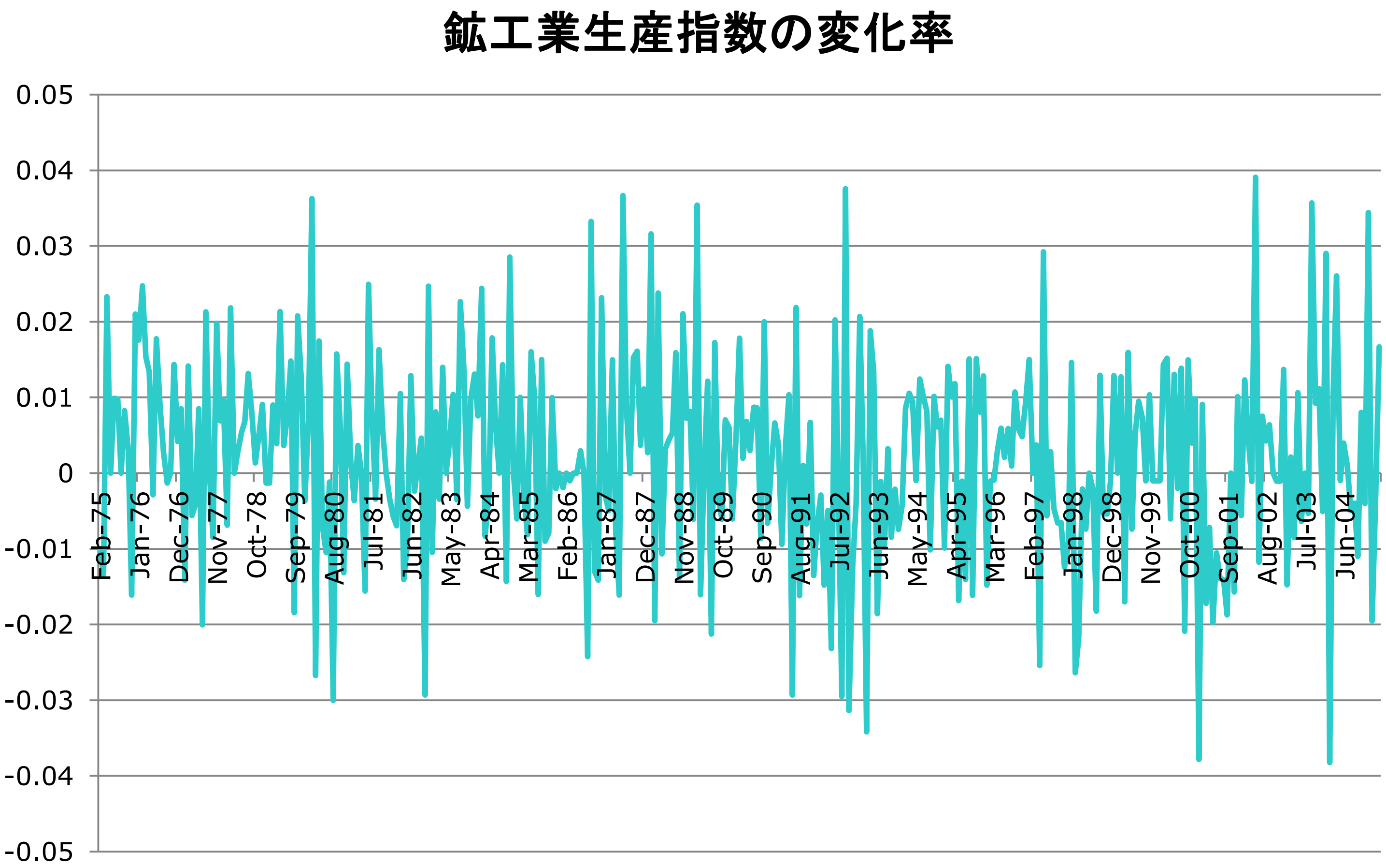
為替レートの変化率


0.12
0.1
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
Feb-75
Jan-76
Dec-76
Nov-77
Oct-78
Sep-79
Aug-80
Jul-81
Jun-82
May-83
Apr-84
Mar-85
Feb-86
Jan-87
Dec-87
Nov-88
Oct-89
Sep-90
Aug-91
Jul-92
Jun-93
May-94
Apr-95
Mar-96
Feb-97
Jan-98
Dec-98
Nov-99
Oct-00
Sep-01
Aug-02
Jul-03
Jun-04

鉱工業生産指数



120
100
80
60
40
20
0
Jan-75
Jan-76
Jan-77
Jan-78
Jan-79
Jan-80
Jan-81
Jan-82
Jan-83
Jan-84
Jan-85
Jan-86
Jan-87
Jan-88
Jan-89
Jan-90
Jan-91
Jan-92
Jan-93
Jan-94
Jan-95
Jan-96
Jan-97
Jan-98
Jan-99
Jan-00
Jan-01
Jan-02
Jan-03
Jan-04
Jan-05


鉱工業生産指数の変化率
0.05
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
-0.05
Feb-75
Jan-76
Dec-76
Nov-77
Oct-78
Sep-79
Aug-80
Jul-81
Jun-82
May-83
Apr-84
Mar-85
Feb-86
Jan-87
Dec-87
Nov-88
Oct-89
Sep-90
Aug-91
Jul-92
Jun-93
May-94
Apr-95
Mar-96
Feb-97
Jan-98
Dec-98
Nov-99
Oct-00
Sep-01
Aug-02
Jul-03
Jun-04

定常性

■　強定常性

先ほどの定常性は厳密に言うと、**弱定常性**といわれるものです。定常性にはもうひとつ**強定常性**とよばれるものがあります。これは以下のように定義されます。

(強定常性)
全ての t と k について $(y_t, y_{t+1}, \ldots, y_{t+k})$ の同時分布が等しい。

通常の分析においては弱定常性で十分ですが、場合によっては強定常性を仮定する事があります。

定常性

例題1

弱定常過程において、

$$\gamma_k = \gamma_{-k}$$

を確認して下さい。

推定

- 平均、分散、自己共分散、自己相関の推定

(標本平均)

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

(標本自己共分散)

$$\hat{\gamma}_k = \frac{1}{T}\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})$$

推定

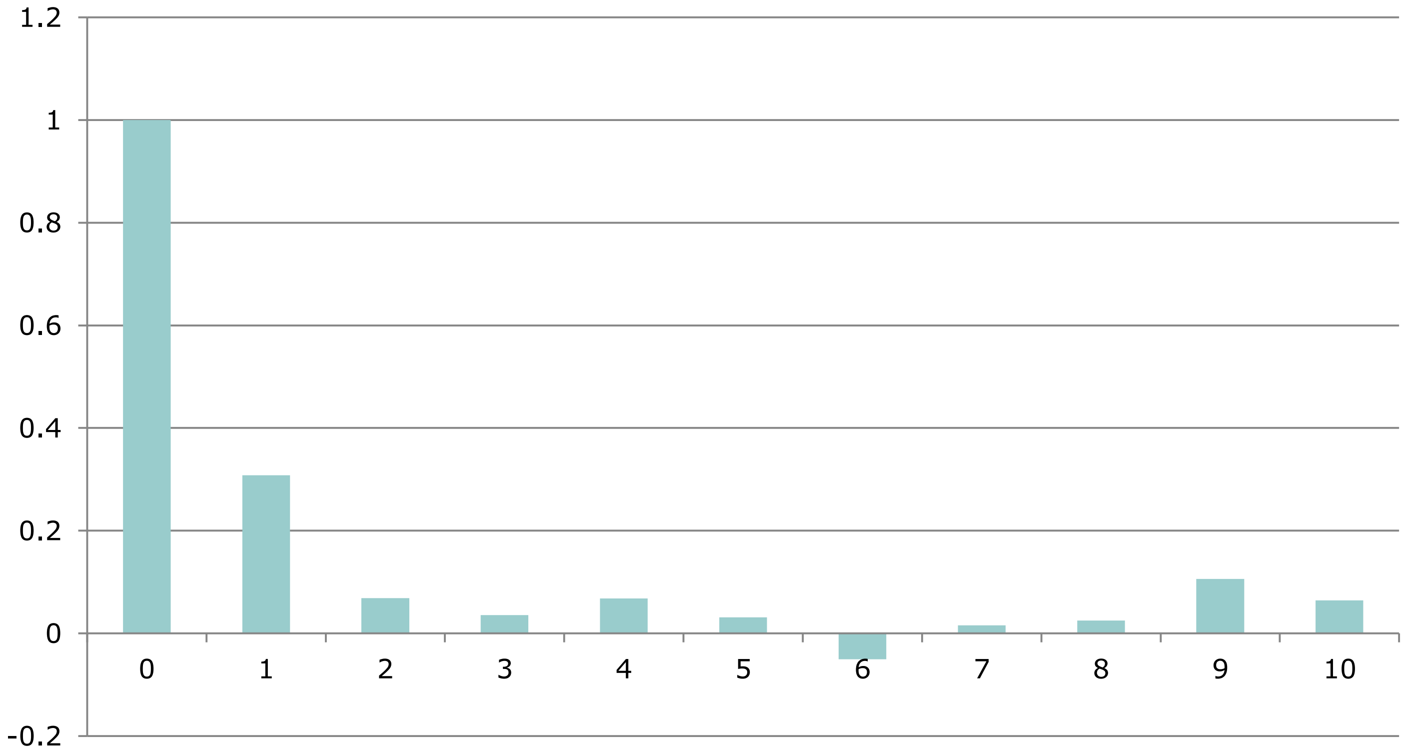
(標本自己相関)

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$$

- (標本)コレログラム

縦軸に(標本)自己相関、横軸に次数 k をとって図を描いたものを(標本)コレログラムといいます。

TOPIX 変化率のコレログラム



定常過程の例

- ホワイトノイズ

ε_t が全ての t において

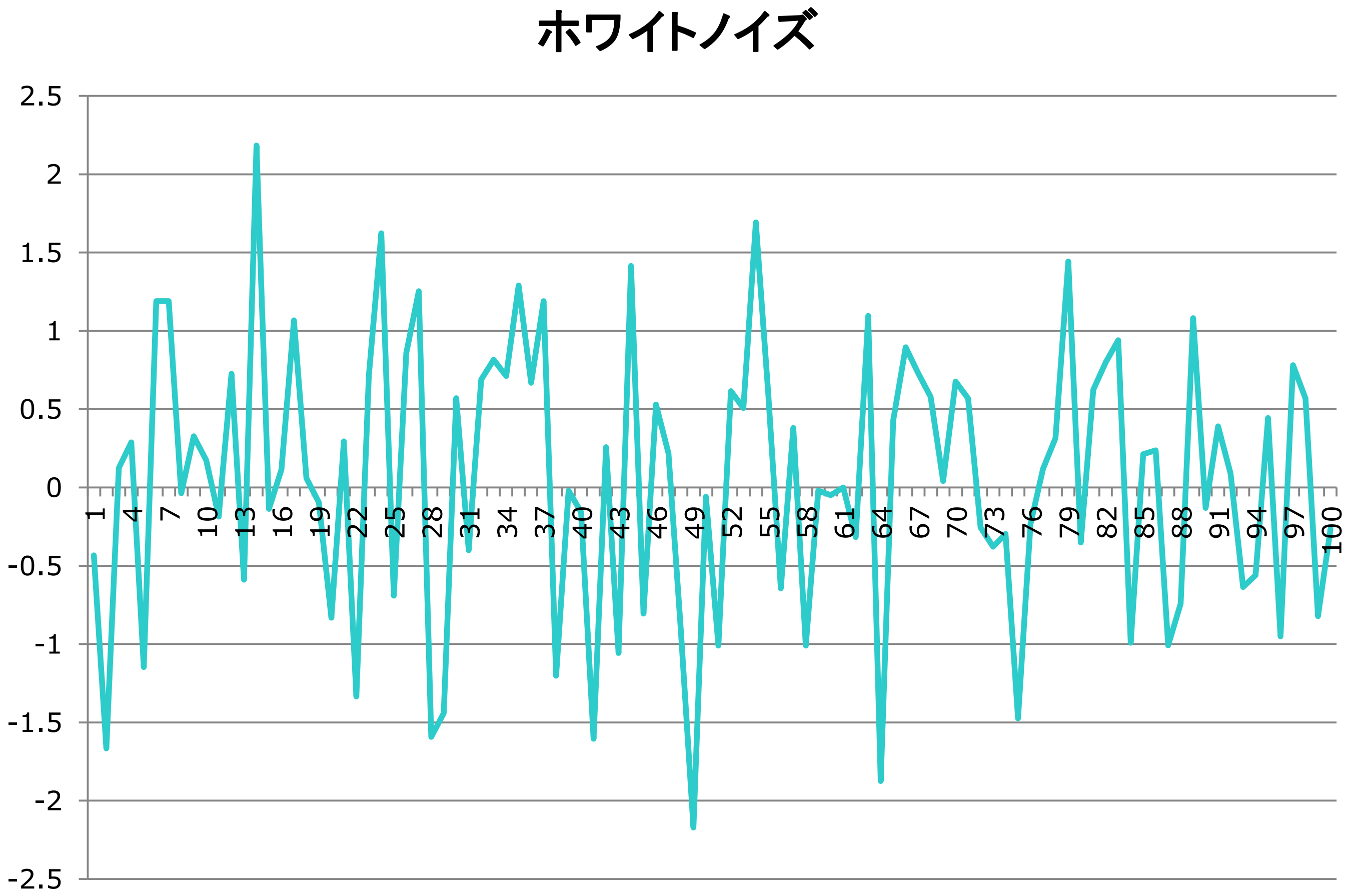
$$(1)\quad E(\varepsilon_t) = 0$$

$$(2)\quad \gamma_k = E(\varepsilon_t \varepsilon_{t-k}) = \begin{cases} \sigma^2, & k = 0 \\ 0\ , & k \neq 0 \end{cases}$$

を満たす時、ε_t を**ホワイトノイズ**(白色雑音) といいます。
ε_t が分散 σ^2 のホワイトノイズである事を

$$\varepsilon_t \sim \mathrm{W.N.}(\sigma^2)$$

のように書きます。



ホワイトノイズ
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
1
4
7
10
13
16
19
22
25
28
31
34
37
40
43
46
49
52
55
58
61
64
67
70
73
76
79
82
85
88
91
94
97
100

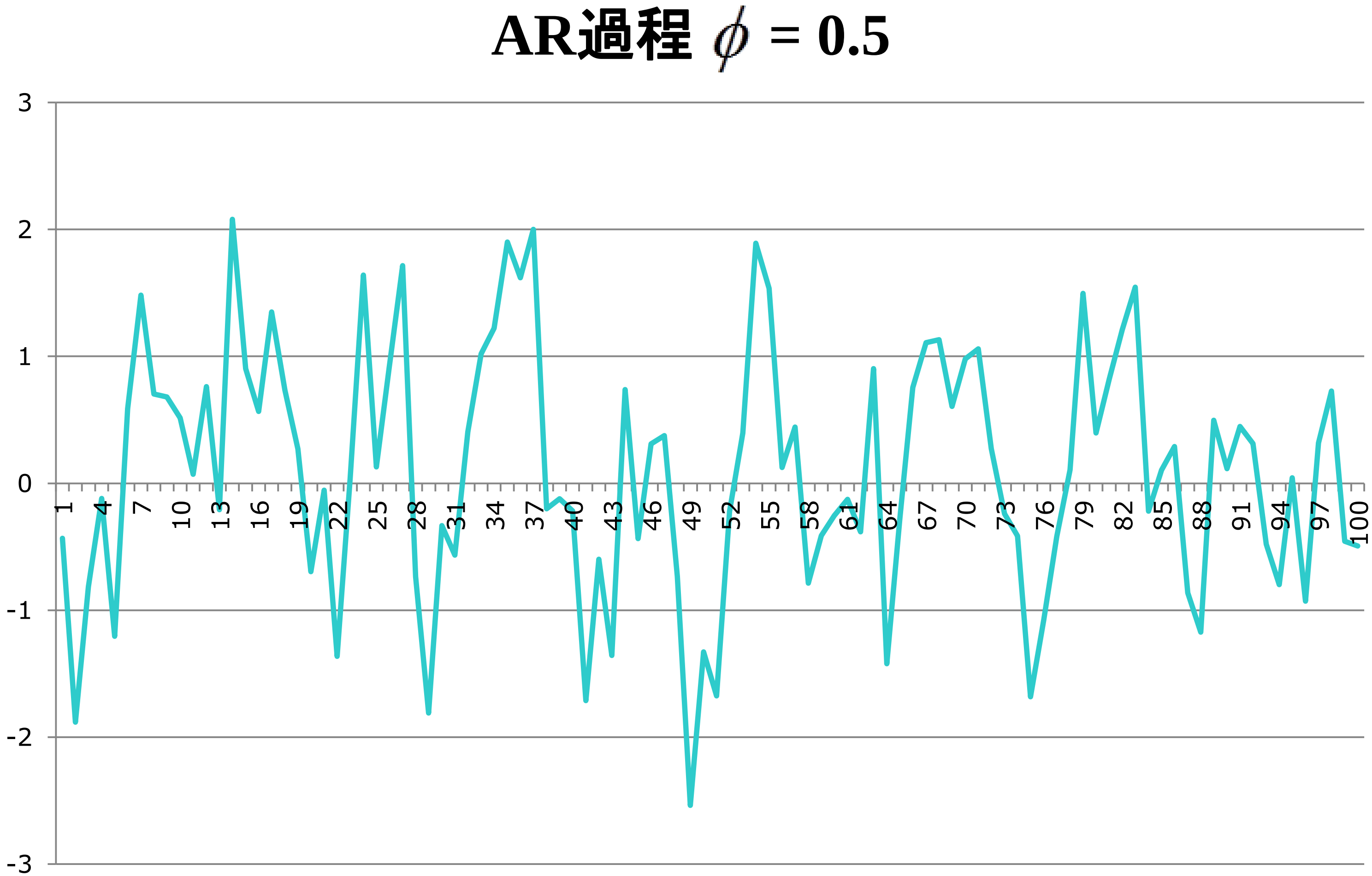
定常過程の例

- 自己回帰過程　(autoregressive (AR) process)

$$y_t = c + \phi y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{W.N.}(\sigma^2)$$

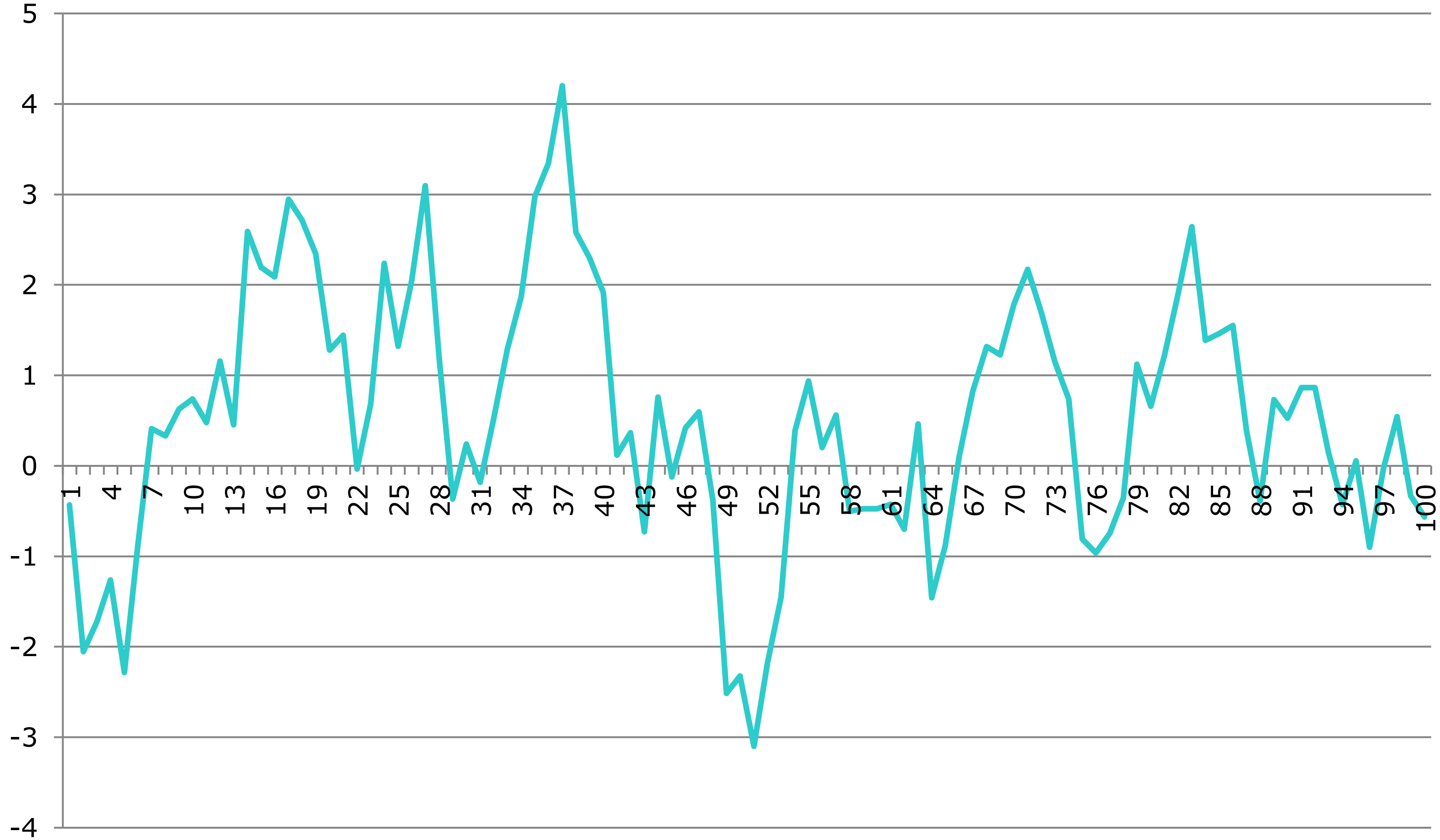
のようにy_tが決定される時、y_tを自己回帰過程（AR過程)といいます。

ϕ が $|\phi| < 1$ の時にAR過程は定常となります。

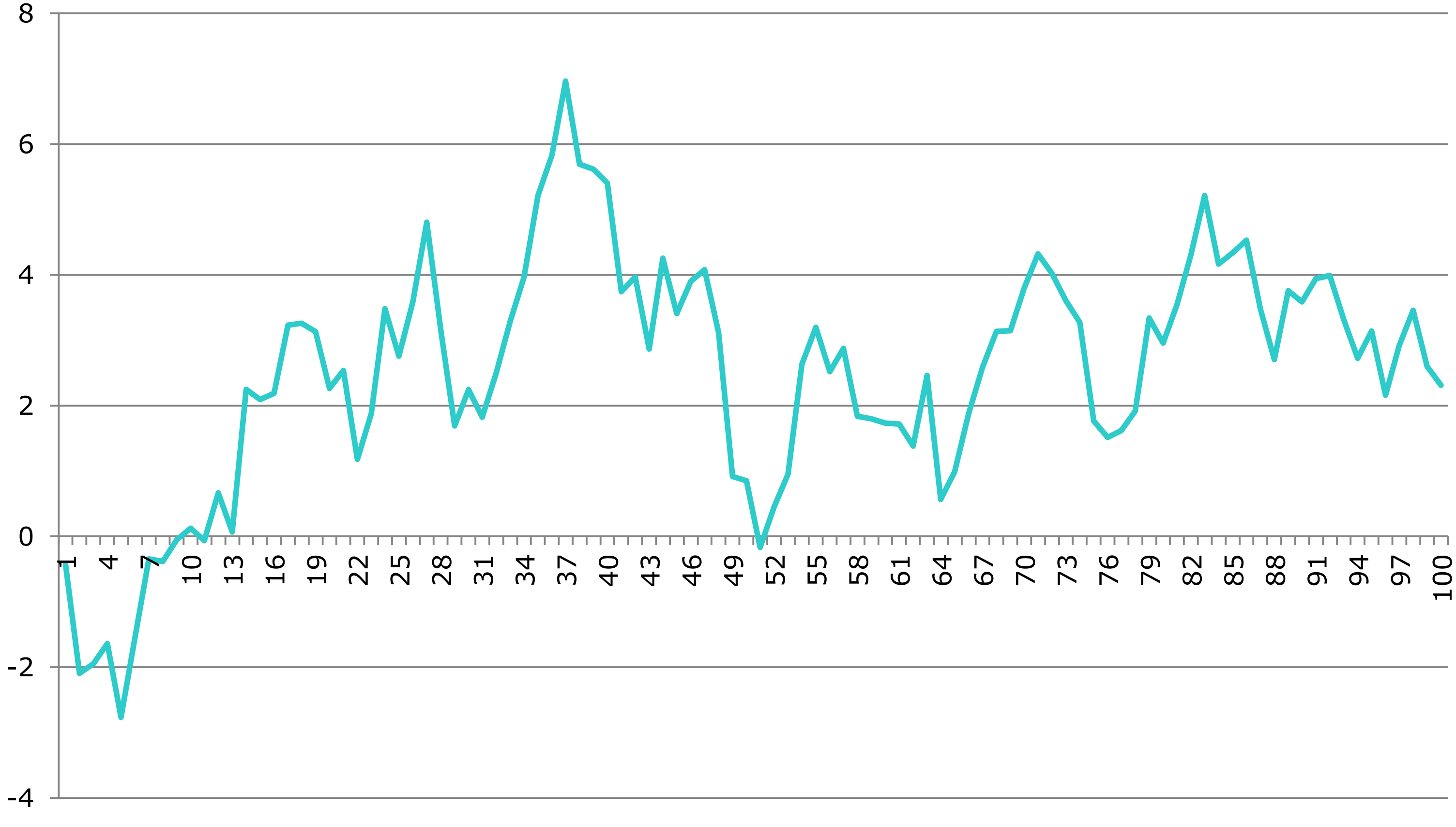


AR過程 ϕ = 0.5
3
2
1
0
-1
-2
-3
1
4
7
10
13
16
19
22
25
28
31
34
37
40
43
46
49
52
55
58
61
64
67
70
73
76
79
82
85
88
91
94
97
100

AR過程 $\phi = 0.9$


5
4
3
2
1
0
-1
-2
-3
-4
1
4
7
10
13
16
19
22
25
28
31
34
37
40
43
46
49
52
55
58
61
64
67
70
73
76
79
82
85
88
91
94
97
100

AR過程　$\phi = 0.99$



8
6
4
2
0
-2
-4
1
4
7
10
13
16
19
22
25
28
31
34
37
40
43
46
49
52
55
58
61
64
67
70
73
76
79
82
85
88
91
94
97
100

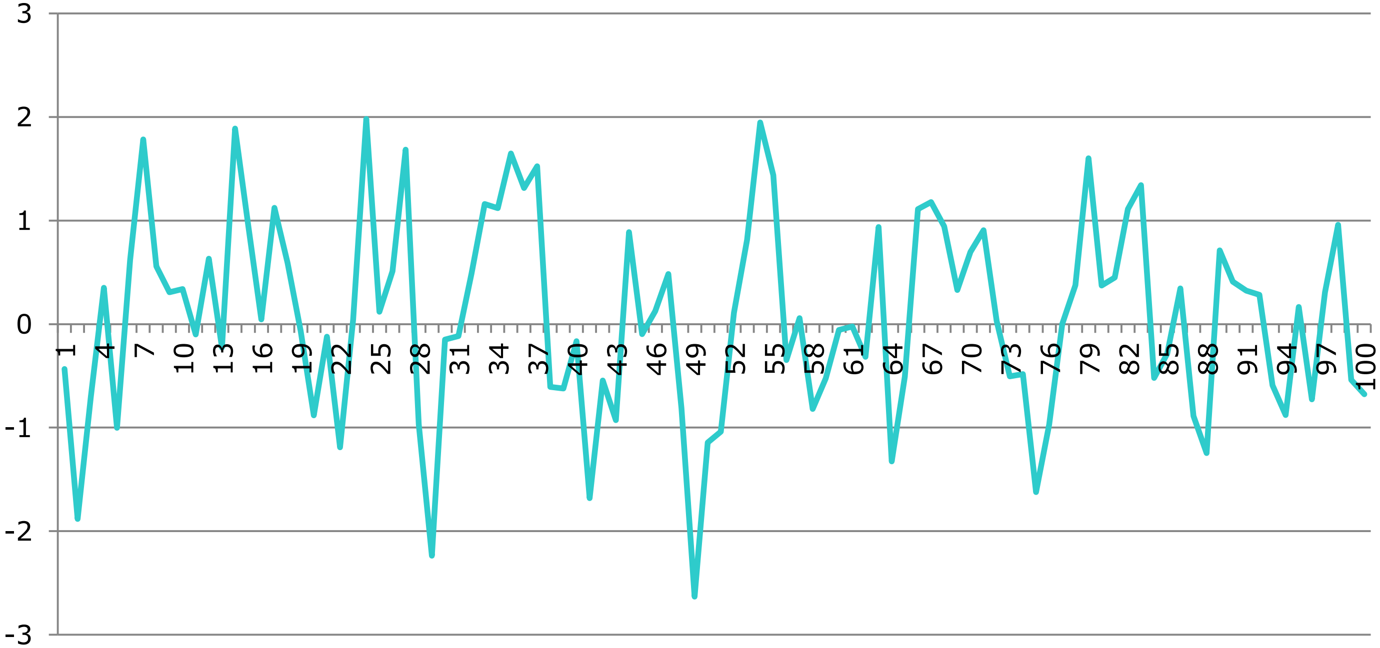
定常過程の例

- 移動平均過程
  (moving average (MA) process)

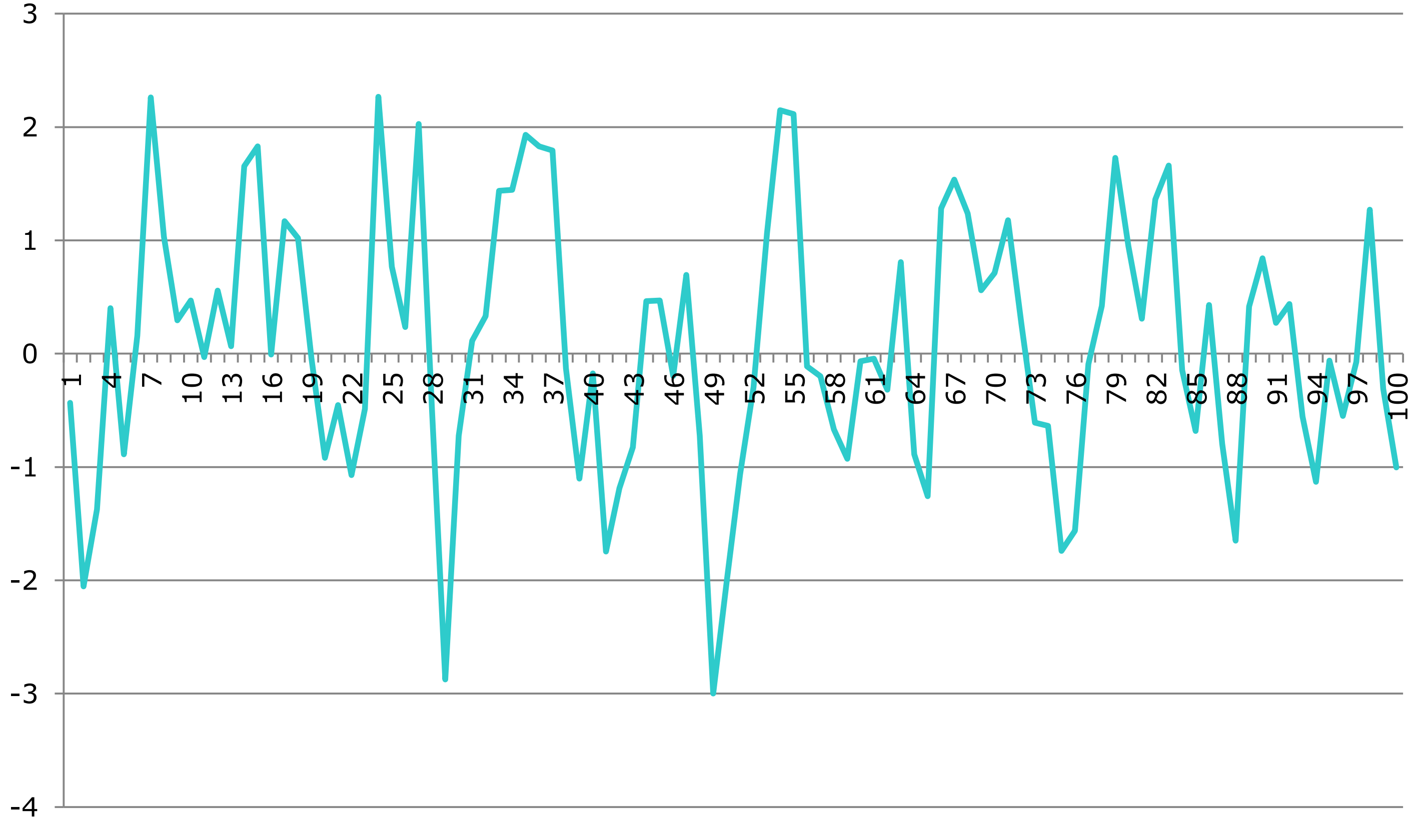
$$y_t = c + \theta \varepsilon_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{W.N.}(\sigma^2)$$

のようにy_tが決定される時、y_tを移動平均過程（MA過程)といいます。

MA過程(θ = 0.5)



MA過程(θ = 0.9)



MA過程(θ = 0.99)

定常過程

例題2

$$y_t = \mu + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{W.N.}(\sigma^2)$$

が弱定常過程である事を確認して下さい。

定常過程

例題3

$$y_t = \theta \varepsilon_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{W.N.}(\sigma^2)$$

が弱定常過程である事を確認して下さい。

自己相関の検定

- 自己相関の推定

定常性を仮定すると、自己共分散、自己相関は

$$\hat{\gamma}_k = \frac{1}{T}\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})$$

および

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$$

によって推定する事ができます。

自己相関の検定

- 自己相関の検定

通常の統計分析同様、推定の後は**検定**をすることを考えます。

ここでは自己相関 ρ_k が 0 かどうかの検定を考えましょう。

帰無仮説は H_0: $\rho_k = 0$、対立仮説は H_1: $\rho_k \neq 0$ です。

自己相関の検定

- 自己相関の検定

$\hat{\rho}_k$ を用いて検定します。

そのためには帰無仮説のもとでの $\hat{\rho}_k$ の**漸近分布**を知る必要があります。

y_t が i.i.d.であれば、$\hat{\rho}_k$ の分布は漸近的に

$$\hat{\rho}_k \sim N(0, 1/T)$$

となる事が知られています。

自己相関の検定

- 自己相関の検定

よって有意水準5%で検定するのであれば、$\hat{\rho}_k$ が

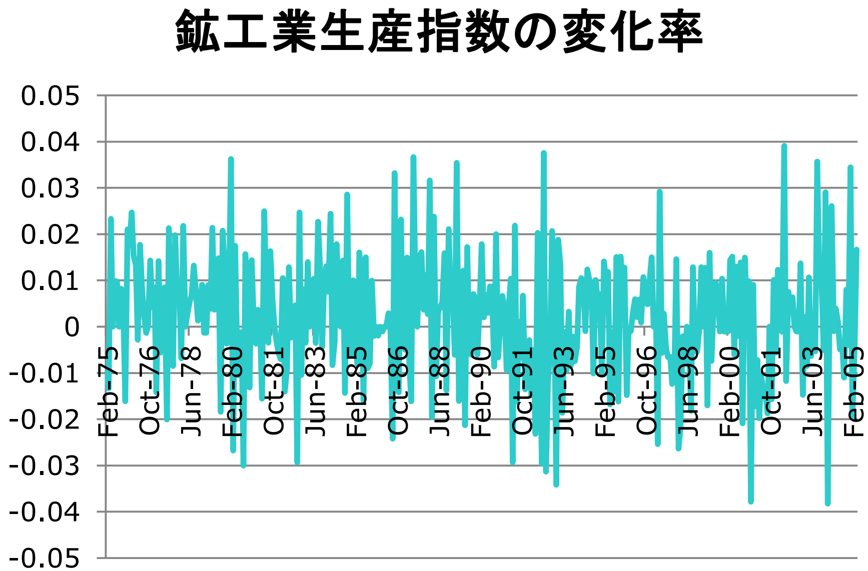
$$|\hat{\rho}_k| > \frac{1.96}{\sqrt{T}}$$

を満たす時に帰無仮説 $H_0\!: \rho_k = 0$ を**棄却**します
(つまり ρ_k は 0 ではないと結論する！)。

自己相関の検定

- 自己相関の検定の例

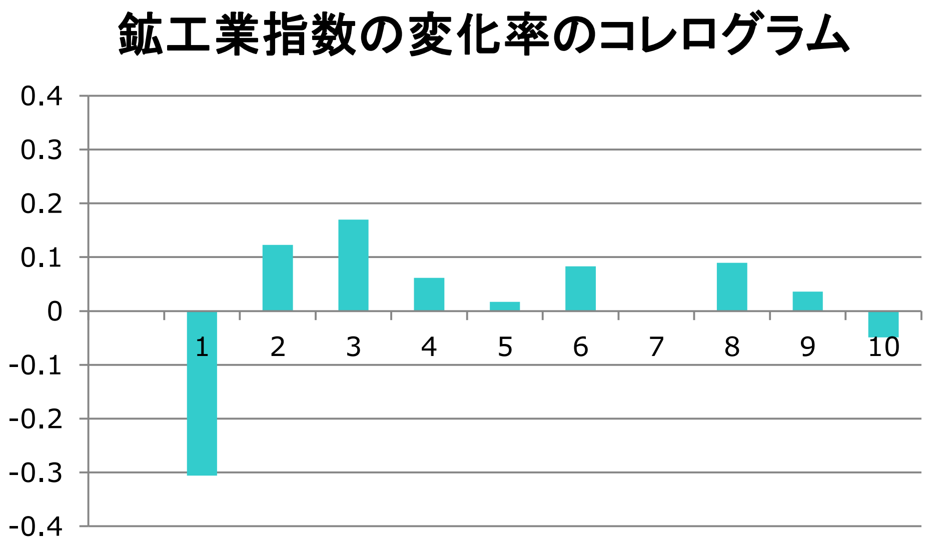
鉱工業指数の変化率に対して、自己相関の有無を検定してみましょう。



自己相関の検定

- 自己相関の検定の例

まず、変化率の自己相関係数を計算します。
以下はそのコレログラムです。



自己相関の検定

- 自己相関の検定の例

ρ_k の具体的な値は以下のようになります。

1	2	3	4	5
-0.305	0.123	0.170	0.062	0.017
6	**7**	**8**	**9**	**10**
0.083	-0.001	0.090	0.036	-0.049

標本数は $T = 363$ であるので ρ_k の絶対値が

$$1.96/\sqrt{T} = 1.96/\sqrt{363} \approx 0.103$$

より大きければ有意水準 5 %で $\rho_k = 0$ を棄却。

自己相関の検定

- 自己相関の検定の例

ρ_kの具体的な値は以下のようになります。

1	2	3	4	5
-0.305	0.123	0.170	0.062	0.017
6	**7**	**8**	**9**	**10**
0.083	-0.001	0.090	0.036	-0.049

標本数は $T = 363$ であるので ρ_k の絶対値が

$$1.96/\sqrt{T} = 1.96/\sqrt{363} \approx 0.103$$

より大きければ有意水準 5 %で $\rho_k = 0$ を棄却。

自己相関の検定

- 自己相関の検定の例

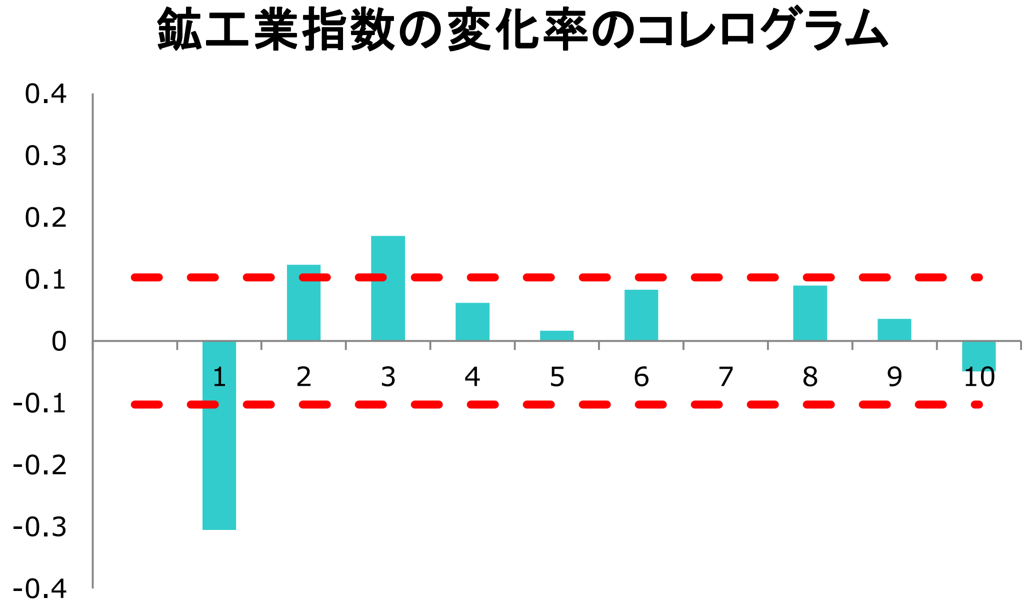
よって、最初の3つの自己相関（3次までの自己相関）は有意(に 0 と異なる)。

4 次から10次までの自己相関は有意ではない。
→本当の値は 0 かもしれない。

自己相関の検定

- 自己相関の検定の例

以下のようにコレログラムに臨界値を示すとわかりやすい。



自己相関の検定

- 自己相関の検定のまとめ

以上の流れをまとめておきましょう。

(1) 自己相関を推定する。
(2) コレログラムを書く。
(3) 自己相関を検定する。
(4) コレログラムに臨界値を示す線を書きこむ。
(5) 結果を解釈する。

自己相関の検定

- かばん検定

先ほどは個々の自己相関が 0 かどうかの検定を考えたが、まとめて m 次までの自己相関がすべて 0 かどうかを検定するという方法があります。

このような検定を**かばん検定**と呼びます。

自己相関の検定

- かばん検定の注意点

かばん検定の帰無仮説は

「m 次までの全ての自己相関が 0 である」

というものです。よって対立仮説は

「m 次までの自己相関のうち少なくとも 1 つは 0でない」

となる事に注意が必要です。

自己相関の検定

- Ljung – Box 統計量

代表的なかばん検定統計量に **Ljung – Box 統計量**があります。その検定統計量は

$$Q(m) = T(T+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{T-k}$$

によって定義されます。

自己相関の検定

- Ljung – Box 統計量

$Q(m)$ は 帰無仮説

$$H_0: \rho_1 = \rho_2 = \rho_3 = \ldots = \rho_m = 0,$$

のもとで（より正確には y_t が i.i.d.であれば）

$$Q(m) \sim \chi^2(m)$$

となります。よって $Q(m)$ と $\chi^2(m)$ の95%点を比べて、$Q(m)$ の方が大きい場合に帰無仮説を**棄却します**。

自己相関の検定

- Ljung – Box 統計量の例

鉱工業指数の変化率のデータの Ljung – Box 統計量を計算してみましょう。

（m = 10 までの Ljung – Box 統計量）

m	1	2	3	4	5	6	7	8	9	10
$Q(m)$	34.20	39.76	50.40	51.80	51.91	54.46	54.46	57.46	57.94	58.84
P値	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

自己相関の検定

- Ljung – Box 統計量の例

鉱工業指数の変化率のデータの Ljung – Box 統計量を計算してみましょう。

（m = 10 までの Ljung – Box 統計量）

m	1	2	3	4	5	6	7	8	9	10
$Q(m)$	**34.20**	**39.76**	**50.40**	**51.80**	**51.91**	**54.46**	**54.46**	**57.46**	**57.94**	**58.84**
P値	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**

m = 10 までのすべての P 値が 0。
⇒　少なくとも 10次までの自己相関のどれか1つが 0ではない。

自己相関の検定

- Ljung – Box 統計量の問題点

Ljung – Box 統計量のひとつの問題点は m の選択が難しい事にあります。

m が小さすぎる　⇒　高次の相関を見逃す可能性

m が大きすぎる　⇒　検出力の低下の可能性

実際には、複数の m に対して $Q(m)$を計算して総合的に判断するのが一般的。