時系列分析のその他のトピック

見せかけの回帰とその対策

宇畑 優太（1260404）

2024-09-05

目次

準備

```
## PDF に出力する際は cairo を使用する
if (knitr::is_latex_output()) {
  knitr::opts_chunk$set(dev = "cairo_pdf")
}

#パッケージの読み込み
pacman::p_load(tidyverse,
               broom,
               coefplot,
               texreg,
               bayesplot,
               rstan,
               rstanrm,
               parallel,
               posterior,
               cmdstanr,
               patchwork,
               ggplot2,
               tidybayes,
               ggfortify,
               gridExtra,
               forecast,
               tseries,
               summarytools,
               forecast,
               lmtest,
               urca,
               prais
               )

#日本語の設定
if (.Platform$OS.type == "windows") {
  if (require(fontregisterer)) {
    my_font <- "Yu Gothic"
  } else {
    my_font <- "Japan1"
```

```r
  }
} else if (capabilities("aqua")) {
  my_font <- "HiraginoSans-W3"
} else {
  my_font <- "IPAexGothic"
}

theme_set(theme_gray(base_size = 9,
                     base_family = my_font))
```

```r
#計算の高速化
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())
```

1　ホワイトノイズへの回帰分析

全く関係ないデータ同士を回帰分析にかけると, 有意な係数は出ないはず.

正規分布に従うホワイトノイズを複数発生させて, 回帰分析を実行してみる

```r
#1 回のシミュレーションにおけるサンプルサイズ
n_sample <- 400

#シード値の設定
set.seed(1)

#データの生成
y_wn <- rnorm(n = n_sample)
x_wn <- rnorm(n = n_sample)

#回帰の実行
mod_ols_wn <- lm(y_wn ~ x_wn)

#結果の表示
summary(mod_ols_wn)
```

```
Call:
lm(formula = y_wn ~ x_wn)
```

```
Residuals:
     Min       1Q   Median       3Q      Max 
-2.91553 -0.60756 -0.06449  0.65797  2.64718 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.03993    0.04862   0.821    0.412
x_wn         0.02605    0.04500   0.579    0.563

Residual standard error: 0.9704 on 398 degrees of freedom
Multiple R-squared:  0.0008414,	Adjusted R-squared:  -0.001669 
F-statistic: 0.3352 on 1 and 398 DF,  p-value: 0.563
```

p 値をみるに, 統計的有意はない. 当然である

2 単位根のあるデータ同士の回帰分析

ホワイトノイズの累積和として，RW 過程をシミュレーションする

```r
#シード値の設定
set.seed(1)

#RW 過程
y_rw <- cumsum(rnorm(n = n_sample))
x_rw <- cumsum(rnorm(n = n_sample))

#回帰の実行
mod_ols_rw <- lm(y_rw ~ x_rw)

#結果の表示
summary(mod_ols_rw)
```

```
Call:
lm(formula = y_rw ~ x_rw)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.1820 -3.0629  0.4445  2.6711  8.3003 

Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.40661    0.29876   18.10   <2e-16 ***
x_rw        -0.28189    0.01738  -16.22   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.622 on 398 degrees of freedom
Multiple R-squared:  0.398, Adjusted R-squared:  0.3965
F-statistic: 263.1 on 1 and 398 DF,  p-value: < 2.2e-16
```

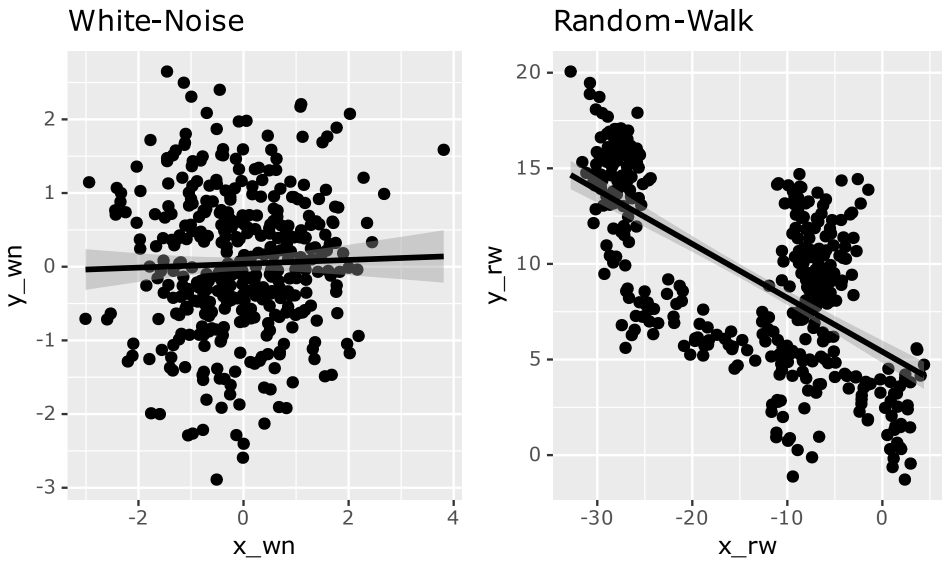
係数が, 統計的有意である.

全く関係のない RW 過程を回帰分析にかけると, 統計的有意になる. この現象を「見せかけの回帰」と呼ぶ.

図示してみよう

```
#WN

#データの整形
df_wn <- data.frame(x_wn = x_wn, y_wn = y_wn)

#図示
p_wn <- ggplot(df_wn,
               aes(x = x_wn,
                   y = y_wn))+
  geom_point() +
  geom_smooth(method = "lm", color = 1) +
  ggtitle("White-Noise")


#RW
#データの整形
df_rw <- data.frame(x_wn = x_rw, y_wn = y_rw)

#図示
p_rw <- ggplot(df_rw,
               aes(x = x_rw,
                   y = y_rw))+
  geom_point() +
  geom_smooth(method = "lm", color = 1) +
  ggtitle("Random-Walk")
```

```
#表示
grid.arrange(p_wn, p_rw, ncol = 2)
```



3 AR 定常過程への回帰分析

単位根ではなく,定常 AR 過程に従うシミュレーション

```
set.seed(1)

#定常 AR 過程に従うデータ
y_ar <- arima.sim(
  n = n_sample,
  model = list(order = c(1,0,0), ar = c(0.8))
)

x_ar <- arima.sim(
  n = n_sample,
  model = list(order = c(1,0,0), ar = c(0.8))
)

#回帰の実行
mod_old_ar <- lm(y_ar ~ x_ar)

#結果
summary(mod_old_ar)
```

```
Call:
lm(formula = y_ar ~ x_ar)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.2996 -0.9923  0.1009  0.9839  4.0034 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)  0.19888    0.07660   2.596  0.00978 **
x_ar         0.08634    0.04955   1.742  0.08223 . 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.511 on 398 degrees of freedom
Multiple R-squared:  0.007569,	Adjusted R-squared:  0.005076 
F-statistic: 3.036 on 1 and 398 DF,  p-value: 0.08223
```

有意な回帰係数が得られてしまった. これも「見せかけの回帰」である

4 残差の自己相関と見せかけの回帰

詳細は note

5 DW 検定

詳細は note

```
#DW 統計量
resid_old <- mod_ols_rw$residuals

dw <- sum(diff(resid_old)^2) / sum((resid_old)^2)

dw
```

```
[1] 0.08021259
```

2 ではないので, 自己相関がありそう

検定もできる

```
#WN
dwtest(mod_ols_wn)
```

```
	Durbin-Watson test

data:  mod_ols_wn
DW = 2.0935, p-value = 0.8261
alternative hypothesis: true autocorrelation is greater than 0
```

```
#RW
dwtest(mod_ols_rw)
```

```
	Durbin-Watson test

data:  mod_ols_rw
DW = 0.080213, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

```
#AR(1)過程
dwtest(mod_old_ar)
```

```
	Durbin-Watson test

data:  mod_old_ar
DW = 0.48088, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

RW と AR 過程に自己相関が見られる

6 シミュレーションによる見せかけの回帰

以下のコードで p 値だけ取り出せる

```
summary(mod_ols_wn)$coefficient["x_wn", "Pr(>|t|)"]
```

```
[1] 0.5629569
```

この作業を何回も繰り返してみよう

```r
n_sim <- 200       #シミュレーションの回数

n_sample <- 400     #サンプルサイズ

p_wn <- numeric(n_sim)
p_rw <- numeric(n_sim)

set.seed(1)

for(i in 1:n_sim){
  #自己相関のないシミュデータ
  y_wn <- rnorm(n = n_sample)
  x_wn <- rnorm(n = n_sample)

  #線形回帰分析
  mod_wn <- lm(y_wn ~ x_wn)

  p_wn[i] <- summary(mod_wn)$coefficient["x_wn", "Pr(>|t|)"]

  #ランダムウォークにするシミュ
  y_rw <- cumsum(rnorm(n = n_sample))
  x_rw <- cumsum(rnorm(n = n_sample))

  #線形回帰分析
  mod_rw <- lm(y_rw ~ x_rw)

  p_rw[i] <- summary(mod_rw)$coefficient["x_rw", "Pr(>|t|)"]
}

#WN
sum(p_wn < 0.05) / n_sim
```

```
[1] 0.055
```

```r
#RW
sum(p_rw < 0.05) / n_sim
```

```
[1] 0.85
```

理論値は 0.05 であることを考慮すると，0.85 はかなり大きい．

7　見せかけの回帰を防ぐ方法

詳細は note

8

9　単位根検定

```r
#RW への ADF 検定
summary(ur.df(y_rw, type = "none"))
```

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max
-2.69362 -0.76935 -0.01875  0.61984  3.10863

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.007932   0.008430  -0.941    0.347
z.diff.lag -0.037636   0.050540  -0.745    0.457

Residual standard error: 1 on 396 degrees of freedom
Multiple R-squared:  0.004097,  Adjusted R-squared:  -0.0009326
F-statistic: 0.8146 on 2 and 396 DF,  p-value: 0.4436


Value of test-statistic is: -0.9409
```

```
Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

```r
summary(ur.df(x_rw, type = "none"))
```

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4968 -0.6984  0.0219  0.8117  3.0420

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.0008817  0.0030911  -0.285    0.776
z.diff.lag -0.0148436  0.0501355  -0.296    0.767

Residual standard error: 1.04 on 396 degrees of freedom
Multiple R-squared:  0.0004466, Adjusted R-squared:  -0.004602
F-statistic: 0.08847 on 2 and 396 DF,  p-value: 0.9153


Value of test-statistic is: -0.2852

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

1.96 以下である

単位根を持つという帰無仮説を棄却できなかった.

```r
#定常 AR(1) 過程への ADF 検定

summary(ur.df(y_ar,type = "none"))
```

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max
-2.94892 -0.57068 -0.04344  0.70965  2.59868

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.22706    0.03419  -6.641 1.03e-10 ***
z.diff.lag -0.01848    0.05022  -0.368    0.713
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9782 on 396 degrees of freedom
Multiple R-squared:  0.1158,    Adjusted R-squared:  0.1113
F-statistic: 25.92 on 2 and 396 DF,  p-value: 2.633e-11


Value of test-statistic is: -6.6414

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

```r
summary(ur.df(x_ar,type = "none"))
```

```
################################################
# Augmented Dickey-Fuller Test Unit Root Test #
################################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0561 -0.8481 -0.0768  0.6517  3.8238

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.30771    0.03843  -8.008 1.32e-14 ***
z.diff.lag  0.03521    0.05003   0.704    0.482
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.089 on 396 degrees of freedom
Multiple R-squared:  0.1516,    Adjusted R-squared:  0.1473
F-statistic: 35.37 on 2 and 396 DF,  p-value: 7.348e-15


Value of test-statistic is: -8.0076

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

1.96 異常であるので, 単位根がないという帰無仮説を棄却する. 単位根はない.

10 一般化最小二乗法：GLS

詳細は note

11　R による Prains-Winsten 法

まずは残差を求めよう

```
#定常 AR(1) 過程に従うデータを OLS でモデル化する

mod_ols_ar <- lm(y_ar ~ x_ar)

#残差
resid_ols_ar <- mod_ols_ar$residual
```

残差に対して OLS 推定を行い, 残差の自己相関の推定値を求める

```
mod_resid <- lm(resid_ols_ar[-1] ~ resid_ols_ar[-n_sample]-1)

ro <- as.numeric(mod_resid$coefficients)

ro
```

```
[1] 0.7585326
```

初期時点のデータの変換を行う

```
y_trans_1 <- sqrt(1 - ro^2) * y_ar[1]
x_trans_1 <- sqrt(1 - ro^2) * x_ar[1]
psi_trans_1 <- sqrt(1 - ro^2)
```

2 時点以降

```
y_trans_2 <- y_ar[-1] - ro*y_ar[-n_sample]
x_trans_2 <- x_ar[-1] - ro*x_ar[-n_sample]
psi_trans_2 <- rep(1 - ro, n_sample -1 )
```

二つを結合する

```
y_trans_all <- c(y_trans_1,y_trans_2)
x_trans_all <- c(x_trans_1,y_trans_2)
psi_trans_all <- c(psi_trans_1,psi_trans_2)
```

あとは普通に OLS 推定量を求めるだけ

```
mod_gls_hand <- lm(y_trans_all ~ psi_trans_all + x_trans_all -1)

summary(mod_gls_hand)
```

```
Call:
lm(formula = y_trans_all ~ psi_trans_all + x_trans_all - 1)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.00775 -0.00374 -0.00285 -0.00159  0.40275 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
psi_trans_all 0.011590   0.004202   2.758  0.00607 ** 
x_trans_all   0.998326   0.001047 953.383  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02044 on 398 degrees of freedom
Multiple R-squared:  0.9996,	Adjusted R-squared:  0.9996 
F-statistic: 4.551e+05 on 2 and 398 DF,  p-value: < 2.2e-16
```

統計的有意は得られなかった. 見せかけの回帰を回避できた

12 差分系列への回帰分析

単位根があるデータへの回帰分析 (非定常過程に従う)

差分系列に回帰するのが簡単である

```r
mod_lm_diff <- lm(diff(y_rw) ~ diff(x_rw))

summary(mod_lm_diff)
```

```
Call:
lm(formula = diff(y_rw) ~ diff(x_rw))

Residuals:
     Min       1Q   Median       3Q      Max 
-2.63491 -0.73367  0.04338  0.65956  3.14882 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) -0.02864    0.05012  -0.571    0.568
diff(x_rw)  -0.01742    0.04818  -0.362    0.718

Residual standard error: 1.001 on 397 degrees of freedom
Multiple R-squared:  0.0003292, Adjusted R-squared:  -0.002189
F-statistic: 0.1307 on 1 and 397 DF,  p-value: 0.7179
```

13 共和分

共和分の定義は note

変数を y_t と x_t のみとし, 各々は単位根を有する [I(1)] が, 線形結合が定常過程 [I(0)] になる場合を考える.

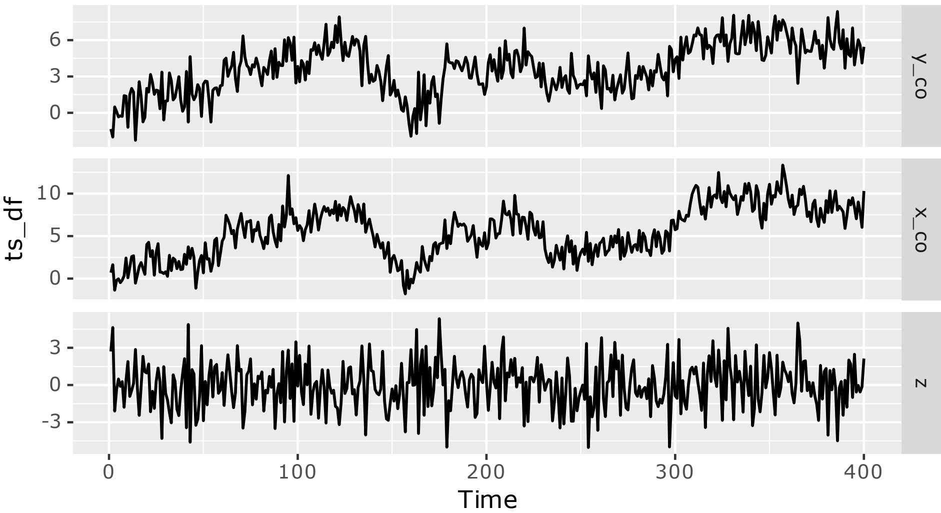
```
set.seed(1)

rw <- cumsum(rnorm(n = n_sample))          #単位根がある

x_co <- 0.6 * rw + rnorm(n = n_sample)     #単位根がある

y_co <- 0.4 * rw + rnorm(n = n_sample)     #単位根がある
```

なぜ線形結合で単位根が消えるのか

$$z_t = x_t - \frac{0.6}{0.4} y_t$$

上記のような線形結合で RW が消える.

```
#データをまとめる

df <- data.frame(
  y_co = y_co,
  x_co = x_co,
  z = x_co - (0.6/0.4)*y_co
)

#ts 型に変形
ts_df <- ts(df)

#図示
autoplot(ts_df, facets = T)
```



14 共和分検定

単位根を持つデータに対して OLS 推定で回帰直線を求める. 残差を求める

残差に対して単位根検定を行う. 単位根がないなら, 共和分がある.

なぜなら, 回帰式は一種の線型結合だから. 共和分の関係にあれば, 単位根は消えるはず.

```
#データの整形
data_mat <- matrix(nrow = n_sample,ncol = 2)

data_mat[,1] <- y_co

data_mat[,2] <- x_co

#共和分検定

summary(ca.po(data_mat, demean = "none"))
```

```
########################################
# Phillips and Ouliaris Unit Root Test #
########################################

Test of type Pu
detrending of series none
```

```
Call:
lm(formula = z[, 1] ~ z[, -1] - 1)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.4650 -0.6586  0.0970  0.9177  3.4208 

Coefficients:
        Estimate Std. Error t value Pr(>|t|)    
z[, -1]  0.64249    0.00997   64.44   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.251 on 399 degrees of freedom
Multiple R-squared:  0.9123,    Adjusted R-squared:  0.9121 
F-statistic:  4153 on 1 and 399 DF,  p-value: < 2.2e-16


Value of test-statistic is: 236.8936 

Critical values of Pu are:
                  10pct    5pct    1pct
critical values 20.3933 25.9711 38.3413
```

共和分がないという帰無仮説が棄却された. 対立仮説を採用する. 共和分関係はある.

共和分関係にあるデータに差分系列への回帰分析を行うと, その関係は見えなくなる

```r
#共和分にあるデータに,差分を取ってから回帰
y_co_diff <- diff(y_co)
x_co_diff <- diff(x_co)

mod_lm_dif_cointegrate <- lm(y_co_diff ~ x_co_diff)
summary(mod_lm_dif_cointegrate)
```

```
Call:
lm(formula = y_co_diff ~ x_co_diff)

Residuals:
    Min      1Q  Median      3Q     Max 
```

```
-4.2689 -0.9975 -0.0603  0.9983  5.4392

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.01614    0.07767   0.208    0.836
x_co_diff    0.03659    0.04712   0.777    0.438

Residual standard error: 1.551 on 397 degrees of freedom
Multiple R-squared:  0.001517,  Adjusted R-squared:  -0.0009984
F-statistic: 0.603 on 1 and 397 DF,  p-value: 0.4379
```

統計的有意になってないことがわかる.差分をとればいいってもんじゃない