R による時系列分析 3†

1. 単位根検定をする。

1.1 パッケージ “fUnitRoots” をインスツールする。

単位根検定をするために R のパッケージ fUnitRoots をインスツールする。パッケージとは通常の R には含まれていない、追加的な R のコマンドの集まりのようなものである。R には追加的に 600 以上のパッケージが用意されており、それぞれ分析の目的に応じて標準の R にパッケージを追加していくことになる。

インターネットに接続してあるパソコンで R を起動させ、「パッケージ」→「パッケージのインストール...」→「Japan(Tokyo)」→「fUnitRoots」→「OK」とクリックする。すると（いろいろとインストールの途中経過が表示されて）パッケージのインストールが自動的に終わる。（上記の作業は次回以降はやる必要はないが、以下の作業は R を起動するたびに毎回やる必要がある）。次にインストールしたパッケージを使うためにコマンドウィンドウ (R Console) に

```
> library(fUnitRoots)
```

と入力すると(library()関数はインスツールしたパッケージを読み込むための関数)、再びコマンドウィンドウ上にいろいろと表示されパッケージ fUnitRoots を使用できる様になる。

次に使用するデータを読み込む。今回は以前使用した tsdata.txt のデータをもう一度使用する。データはホームページにアップしてある(tsdata.txt ファイル)。

tsdata.txt を任意のファイルに移動し、「ファイル」→「ディレクトリの変更」によってデータ tsdata.txt のあるディレクトリに移動。データを読み込み最初の 5 行だけ表示してみる。

```
> tsdata=read.table("tsdata.txt",header=T)
> head(tsdata,5)
date  topix exrate indprod
1 Jan-75 276.09  29.13   47.33
2 Feb-75 299.81  29.70   46.86
3 Mar-75 313.50  29.98   46.24
4 Apr-75 320.57  29.80   47.33
5 May-75 329.65  29.79   47.33
```

データは左から、年月、TOPIX、実効為替レート、鉱工業生産指数を表している。ここでは TOPIX の収益率に対して単位根検定を行う。TOPIX を時系列データに直し、その収益率を（対数階差×100 で）計算する。

```
> topixrate = diff(log(tsdata$topix))*100
```

†この資料は私のゼミおよび講義で R の使用法を説明するために作成した資料です。ホームページ上で公開しており、自由に参照して頂いて構いません。ただし、内容について、一応検証してありますが、間違いがあるかもしれません。間違いがあった場合でもそれによって生じるいかなる損害、不利益について責任は負いかねますますのでご了承下さい。

1.2. `unitrootTest()` 関数によって単位根検定を行う

まず単位根検定には帰無仮説と対立仮説の組み合わせとして、3 つのケースが考えられることを思い出そう(スライド参照)。例えば、DF 検定に関しては

ケース①

帰無仮説: ドリフトの**ない**単位根過程

$$y_t = y_{t-1} + \varepsilon_t.$$

対立仮説: 定数項の**ない**定常な AR(1) 過程

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad |\rho| < 1.$$

ケース②

帰無仮説: ドリフトの**ない**単位根過程

$$y_t = y_{t-1} + \varepsilon_t.$$

対立仮説: 定数項のある定常な AR(1)過程

$$y_t = c + \rho y_{t-1} + \varepsilon_t, \quad |\rho| < 1.$$

ケース③

帰無仮説: ドリフトの**ある**単位根過程

$$y_t = \delta + y_{t-1} + \varepsilon_t.$$

対立仮説: トレンド定常過程

$$y_t = c + \delta t + \rho y_{t-1} + \varepsilon_t, \quad |\rho| < 1.$$

となる。それぞれの場合で帰無仮説のもとでの検定統計量の分布が異なる。どの場合も、対立仮説の回帰モデルを推定し、ρ が 1 より小さいかどうかを片側検定する。

では、今回用いる TOPIX の収益率の場合はどのケースを考えるのが適当であろうか? これは収益率をプロットして判断するとよい。収益率を

```
> plot(topixrate,type="l")
```

によってプロットしたものをみると、明らかにトレンドは存在しない。よってケース③は除外される。また、対立仮説として、必ずしも定数項があるかないか(つまり定常平均が 0 かどうか)は判断できないので、対立仮説のもとで定数項が 0 でない可能性も考慮するとケース②が妥当である事がわかる。

まず DF テストを行う。結果に `utresult` という名前を付けるとして、以下のコマンドを入力する。

```
> utresult=unitrootTest(topixrate, type="c", lags=0)
> utresult@test
```

(ここで `type` は `"c"`, `"nc"`, `"ct"` の 3 つがあり、それぞれケース②、ケース①、ケース③に相当する。`lags` は ADF 検定における `lag` の数、`lags=0` が DF 検定に相当する)

すると

```
$data.name
[1] "topixrate"

$regression

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

Residuals:
     Min       1Q   Median       3Q      Max 
-14.2057  -2.5007  -0.0439   2.0718  13.6110 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.24877    0.20676   1.203     0.23    
y.lag.1     -0.69166    0.04991 -13.859   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 3.915 on 360 degrees of freedom
Multiple R-squared: 0.3479,     Adjusted R-squared: 0.3461 
F-statistic: 192.1 on 1 and 360 DF,  p-value: < 2.2e-16

$statistic
       DF 
-13.85921 

$p.value
           t            n 
1.177472e-26  5.092225e-02 

$parameter
Lag Order 
        0 
```

と表示される。最初の部分は回帰モデルの最少二乗法による推定結果を示している。

推定結果を見ると `lm(formula = y.diff ~ y.lag.1 + 1)` とあり、また `(Intercept)` と `y.lag1` の係数だけが推定されている事がわかる。これは

$$\Delta y_t = c + \beta y_{t-1} + \varepsilon_t$$

という回帰モデルを推定したという事である。被説明変数が y_t の階差になっていることに注意しよう。この場合、

$$y_t - y_{t-1} = c + \beta y_{t-1} + \varepsilon_t$$
$$\Leftrightarrow \; y_t = c + \beta y_{t-1} + y_{t-1} + \varepsilon_t$$
$$\Leftrightarrow \; y_t = c + (1+\beta) y_{t-1} + \varepsilon_t$$

であるから、$\beta = 0$ の時に単位根過程になる。よってこの回帰式では $\beta = 0$ を検定する事になる。この場合の検定についての説明はスライドの $\hat{\rho}_T - 1$ の部分を全て、この回帰モデルの β の OLS 推定量 $\hat{\beta}_{ols}$ に置き換えたものになる。例えば、スライドの DF ρ 検定統計量 λ_ρ は

$$\lambda_\rho = T(\hat{\rho}_T - 1) = T\hat{\beta}_{ols}$$

によって計算される。ここでは標本数 $T = 363$ (`length(topixrate)` で確認できる)なので、λ_ρ の値は

$$\lambda_\rho = 363 \times (-0.69166) \approx -251.07$$

となる。ちなみに後述する DF t 検定統計量と異なり、DF ρ 検定統計量は自動的に計算されないので、このように自分で計算しなければならない。

同様にスライドの DF t 検定統計量 τ_t の値は β の OLS 推定量の t 値となる。下の方に

```
$statistic
        DF
 -13.85921
```

とあるのが DF t 検定統計量の値である。また、その下の **`$p.value`** は DF t 検定統計量の P 値である(`t` の下の数字の方。`n` の下の数値は何の値か不明)。これを用いて検定してもよい。スライドに載っている臨界値 (%点) より、DF ρ 検定でも DF t 検定でも単位根過程という帰無仮説有意水準 1%で棄却されることがわかる。

次に ADF 検定を行おう。これは先ほどの `unitrootTest()` において `lags` の数を変更するだけである。例えば対立仮説として AR(3)モデルを考えるのであれば `lags = 2` とすればよい (`lags` は「対立仮説の AR の次数 -1 」になる)。この場合、

```
> utresult2= unitrootTest(topixrate, type="c", lags=2)
```

とし

```
> utresult2@test
```

とすれば、ADF t 検定統計量、τ_t、の値が出力されている。

また ADF t 検定統計量の値の出力だけでよいのであれば

```
>unitrootTest(topixrate, type="c", lags=2)
```

とすれば

```
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 2
  STATISTIC:
    DF: -9.5888
  P VALUE:
    t: 2.517e-16
    n: 0.1465
```

と出力される。ADF t 検定統計量だけで検定するのであればこれだけでも十分である。

練習問題

1. 実効為替レートの変化率に対して、対立仮説を AR(3) モデルとした ADF 検定 を行いなさい。