大阪府×大阪府立大学　共催セミナー

実践！時系列データ分析

2021年12月1日

大阪府立大学　立花実

本日の内容

１． 株式リターンの統計的な特徴

２． 株式リターンを時系列モデルに当てはめて推定

3． “R”の使い方

（その前に、より広い視点からの話）

データ分析に必要な知識・スキル

① 分析手法に関する知識

② 統計ソフトを用いて①を活用できるスキル

「① 分析手法に関する知識」について

～経済データを分析する場合～

- 必須　　　確率論・統計学の基礎　＆　計量経済学の基礎

- 発展的な分野

　因果推論、構造推定、時系列分析、ベイズ、機械学習　など

（発展分野についてはすべてを習得する必要はない。分析目的や
扱うデータに応じて必要な知識をその都度、学べばよい。）

「② 統計ソフトを用いて①を活用できるスキル」について

～経済データの分析によく使われる統計ソフト～

- 操作が簡単

  gretl、EViews（ただし有料）など

- 柔軟なプログラミングが可能（より高度な分析ができる）

  R、Python、MATLAB（ただし有料）など

- それらの中間

  Stata（ただし有料。経済学の実証研究者がよく利用）など

Rの利点

① 無料

② 様々な“パッケージ”が利用できる

③ 柔軟なプログラミングが可能（より高度な分析ができる）

④ インターネットや書籍による解説が豊富

Rを習得するコツ（私見）

① 分からないことがあったら、とにかくインターネットで検索。

② ただし、初めは（あるいはどこかの時点で）本を買ってRの基本的な操作方法を体系的に学んだ方がよいかもしれない。

（私が参考にした本）

・Garrett Grolemund (著), 大橋 真也 (監訳), 長尾 高弘（訳）
『RStudioではじめるRプログラミング入門』オライリー・ジャパン

・Kun Ren (著), 湯谷 啓明・松村 杏子・市川 太祐 (訳)
『Rプログラミング本格入門: 達人データサイエンティストへの道』共立出版

③ パッケージにある関数の使い方が分からない場合は、CRAN(*)にリファレンス・マニュアルや解説があるのでそれを参考にするとよい（ただし英語）。

（＊）CRAN (Comprehensive R Archive Network)：
R本体や各種パッケージをダウンロードするためのWebサイト

例えば、本セミナーで使用する“rugarch”というパッケージに含まれている関数の使い方を知りたい時には、検索エンジンで「rugarch CRAN」などと検索し、rugarchのCRANページに移動する。（次スライドへ続く）

rugarchのCRANページ

rugarch: Univariate GARCH Models

ARFIMA, in-mean, external regressors and various GARCH flavors, with methods for fit, forecast, simulation, inference and plotting.

Version:	1.4-4
Depends:	R (≥ 3.5.0), methods, parallel
Imports:	Rsolnp, nloptr, ks, numDeriv, spd, xts, zoo, chron, SkewHyperbolic, Rcpp, graphics, stats, grDevices, utils
LinkingTo:	Rcpp (≥ 0.10.6), RcppArmadillo (≥ 0.2.34)
Published:	2020-07-16
Author:	Alexios Ghalanos [aut, cre], Tobias Kley [ctb]
Maintainer:	Alexios Ghalanos <alexios at 4dscape.com>
License:	GPL-3
Copyright:	see file COPYRIGHTS
URL:	http://www.unstarched.net, https://bitbucket.org/alexiosg
NeedsCompilation:	yes
Citation:	rugarch citation info
Materials:	README ChangeLog
In views:	Finance, TimeSeries
CRAN checks:	rugarch results

Documentation:

Reference manual:	rugarch.pdf
Vignettes:	Introduction to the rugarch package

上がリファレンス・マニュアル、下が解説書。

ただし解説書がないパッケージも多い。

Downloads:

Package source:	rugarch_1.4-4.tar.gz

ちなみにR上でも関数の検索ができる

Console画面に“?関数名”と入れて実行すると、右下の画面にヘルプの結果が表示される。

（例）“rugarch”というパッケージの中の“ugarchfit”という関数の使い方を知りたい場合

Console画面の一番下に

```
> ?ugarchfit
```

と入れてEnterを押すと、

右下にこの画面が表示される。

ugarchfit-methods {rugarch}

function: Univariate GARCH Fitting

Description

Method for fitting a variety of univariate GARCH models.

Usage

```
ugarchfit(spec, data, out.sample = 0, solver = "solnp", solver.control = list(),
fit.control = list(stationarity = 1, fixed.se = 0, scale = 0, rec.init = 'all',
trunclag = 1000),
numderiv.control = list(grad.eps=1e-4, grad.d=0.0001,
grad.zero.tol=sqrt(.Machine$double.eps/7e-7), hess.eps=1e-4, hess.d=0.1,
hess.zero.tol=sqrt(.Machine$double.eps/7e-7), r=4, v=2),...)
```

Arguments

`data`	A univariate data object. Can be a numeric vector, matrix, data.frame, zoo, xts, timeSeries, ts or irts object.
`spec`	A univariate GARCH spec object of class uGARCHspec.
`out.sample`	A positive integer indicating the number of periods before the last to keep for out of sample forecasting (see details).
`solver`	One of either “nlminb”, “solnp”, “lbfgs”, “gosolnp”, “nloptr” or “hybrid” (see notes).
`solver.control`	Control arguments list passed to optimizer.

- 以降はRを動かしながら解説を進める。

- 本PPT資料の以後のスライドは、あくまで参照用としてのみ使う。

- なお、時系列分析の手法をしっかり学びたい人には以下の図書がお薦め。

  沖本 竜義（著）『経済・ファイナンスデータの計量時系列分析』朝倉書店

データの対数差分をとるとそのデータの「変化率の近似」になる。

$$\ln(y_t) - \ln(y_{t-1}) \approx \frac{y_t - y_{t-1}}{y_{t-1}}$$

証明

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

$\ln(1+z) \approx z$（ただしzが0に近い場合）という公式を適用

（標本）平均　$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$

（標本）分散　$s^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2$

（標本）標準偏差　$s = \sqrt{s^2} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2}$

（標本）歪度（わいど）

$$Skewness = \frac{1}{T}\sum_{t=1}^{T}\frac{(y_t - \bar{y})^3}{s^3}$$

（標本）尖度（せんど）

$$Kurtosis = \frac{1}{T}\sum_{t=1}^{T}\frac{(y_t - \bar{y})^4}{s^4}$$

ジャック＝ベラ検定

・帰無仮説 H_0：標本データは正規分布に従う

・検定統計量：$JB = \frac{T}{6}\left(Skewness^2 + \frac{1}{4}(Kurtosis - 3)^2\right)$

帰無仮説が正しければJBは自由度2のカイ2乗分布に従う

k次の（標本）自己相関

$$\rho_k = \frac{\text{k次の（標本）自己共分散}}{\text{（標本）分散}} = \frac{(1/T)\ \sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{(1/T)\ \sum_{t=1}^{T}(y_t - \bar{y})^2}$$

- $-1 \leq \rho_k \leq 1$
- $\rho_k \neq 0$のとき「k次の自己相関がある」という

リュング＝ボックス検定

・帰無仮説 H_0：$\rho_1 = \rho_2 = \cdots = \rho_m = 0$

・検定統計量： $Q(m) = T(T+2)\sum_{k=1}^{m} \frac{\rho_k^2}{T-k}$

帰無仮説が正しければ$Q(m)$は自由度mのカイ2乗分布に従う

（弱）定常性の定義

任意の t と k について以下の（1）と（2）が成立するとき（弱）定常という。

（1）$E(Y_t) = \mu$ ⟵ 期待値が時間を通じて一定

（2）$Cov(Y_t, Y_{t-k}) = E[(Y_t - \mu)(Y_{t-k} - \mu)] = \gamma_k$

分散（k=0のとき）と自己共分散が時間を通じて一定。
ただし、自己共分散は時間差 k には依存する。

- AR(p)モデル (Autoregressive model；自己回帰モデル)

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

- MA(q)モデル (Moving average model；移動平均モデル)

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

- ARMA(p,q)モデル

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

- GARCHモデル(Generalized autoregressive conditional heteroskedasticity model)

GARCHモデルを定義する前に「条件付き期待値」のモデルを定義する

$$y_t = \mu_t + u_t = \mu_t + \sqrt{h_t}\, v_t$$

（μ_tにはARMAモデルなどが入る）

h_tはy_tの条件付き分散になり、これを以下のようにGARCHモデルとして定式化する

GARCH(m,r)モデル： $h_t = \omega + \alpha_1 u_{t-1}^2 + \cdots + \alpha_m u_{t-m}^2 + \beta_1 h_{t-1} + \cdots + \beta_r h_{t-r}$

特にGARCH(1,1)モデルがよく使われる

$$h_t = \omega + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1}$$

“レバレッジ効果” を捉えるためにGJR-GARCH(1,1)モデルもよく使われる

$$h_t = \omega + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} + \gamma_1 I(u_{t-1} < 0) u_{t-1}^2$$

（ $I(u_{t-1} < 0)$ は、$u_{t-1} < 0$の時には1、それ以外の時には0をとる定義関数）

- DCCモデル(Dynamic conditional correlation model)

DCCモデルを定義する前に「多変量バージョンの条件付き期待値」のモデルを定義する。

$$\boldsymbol{y_t} = \boldsymbol{\mu_t} + \boldsymbol{u_t} = \boldsymbol{\mu_t} + \boldsymbol{H_t^{1/2}}\ \boldsymbol{v_t}$$

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{bmatrix} = \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \\ \vdots \\ \mu_{nt} \end{bmatrix} + \begin{bmatrix} h_{11,t} & h_{12,t} & \cdots & h_{1n,t} \\ h_{21,t} & h_{22,t} & \cdots & h_{2n,t} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1,t} & h_{n2,t} & \cdots & h_{nn,t} \end{bmatrix}^{1/2} \begin{bmatrix} v_{1t} \\ v_{2t} \\ \vdots \\ v_{nt} \end{bmatrix}$$

（$\boldsymbol{H_t}$は$\boldsymbol{y_t}$の条件付き分散共分散行列になる）

Engle型のDCCモデル

$$\boldsymbol{H_t} = \boldsymbol{D_t R_t D_t} = [diag(h_{11,t}, \cdots, h_{nn,t})]^{\boldsymbol{1/2}}\ \boldsymbol{R_t}\ [diag(h_{11,t}, \cdots, h_{nn,t})]^{\boldsymbol{1/2}}$$

（$diag()$はカッコ内の要素を対角成分に持つ対角行列、$\boldsymbol{R_t}$は相関行列）

$$\boldsymbol{R_t} = [diag(q_{11,t}, \cdots, q_{nn,t})]^{\boldsymbol{-1/2}}\ \boldsymbol{Q_t}\ [diag(q_{11,t}, \cdots, q_{nn,t})]^{\boldsymbol{-1/2}}$$

$$\boldsymbol{Q_t} = (1 - a - b)\boldsymbol{\bar{Q}} + a\boldsymbol{\varepsilon_{t-1}\varepsilon'_{t-1}} + b\boldsymbol{Q_{t-1}}$$

（$\boldsymbol{\varepsilon_t}$ は標準化残差であり、$\boldsymbol{\varepsilon_t} = \boldsymbol{D_t^{-1} u_t}$と定義）

（他にもTse & Tsui型のDCCモデルもある）