第 1 回 Hi-Stat レクチャー

長期記憶時系列分析について

第 2 日 モデルの統計的推測並びに予測

**片山 直也*

一橋大学

日時：平成 16 年 2 月 20 日（金）16:30 - 18:00

会場：一橋大学第 2 研究館 2 階 217 小会議室

[1] 山本拓先生をはじめとする COE 関係者に感謝する．演者は一橋大学大学院経済学研究科の博士課程 4 年で，COE：社会科学の統計分析拠点構築の RA でもある．E-MAIL: ged0104@srv.cc.hit-u.ac.jp

2004 年 2 月 25 日作成

❑ はじめに

レクチャーにあたり，注意点をいくつか列挙する．

- 時系列解析の前提知識として，定常時系列の基礎知識を既知として話すが重要な点は触れることにする．
- 長期記憶モデルはその難解さから敬遠されがちである．そのため，さわりの部分を重点的に話す（個々の聴衆の期待する応用面に踏み込めないことをお詫びする）．
- 時間制約と演者の能力と好みから長期記憶モデルの全てを話せないため，紹介する内容には偏りがある．なるべく有効な情報を提供するように努めるが，この偏りに留意し，引用している文献等をあたっていただくことを期待する．
- 実用性を考え，（株）数理システムの販売する統計処理ソフト，S-PLUS を用いた入出力結果やグラフを紹介する．また S-PLUS の内蔵する関数の紹介もする．演者の組んだ S-PLUS のプログラム（S-code）も公開可能だが，膨大なものなのでここにはほとんど記載しない．
- 初日のレクチャーでは，長期記憶モデルを既存のほかのモデルと比較しながら様々な角度から眺めることで，モデルの存在意義を明らかにしてもらうことを目的とする．
- 2 日目のレクチャーでは統計的推論と予測の手法を紹介することで，長期記憶モデルの扱い方を理解してもらうことを目的とする．
- 引用している文献は最後に記載する．
- この資料は [8] を大いに参考・引用している．レクチャーの準備段階でこの本を読むと，この本を読んでほしいと言うだけでよいのではないかと思ったほどであった．しかしそれではつまらないので [8] に記載されている内容の部分も行間の部分を努めて説明することにした．また [8] に記載されていない，レクチャーの水準内かつ演者の興味のあるものも加えた．
- 発表では他の文献からの図表も多数用いているがここには記載していない．

2 日目は主に ARFIMA(p,d,q) モデルまたは FI(d) モデルを用いて推定，検定，予測の方法を紹介する．モデルを初日と同様に，

$$\phi(L)(1-L)^d x_t = \theta(L)\varepsilon_t \quad (\{\varepsilon_t\} \sim IID(0,\sigma^2)). \tag{1}$$

とし，MA(∞) 表現と AR(∞) をそれぞれ，

$$x_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} \pi_j x_{t-j} = \varepsilon_t, \quad \psi_0 = \pi_0 = 1, \tag{2}$$

とおく．

1　推定

1.1　平均の推定（標本平均）

ARMA モデル等の短期記憶モデルの平均の推定に標本平均を用いることが多く，最良線形推定量との効率もよいことが知られている．具体的には，スペクトル密度関数 $f(\lambda)$ が原点で発散しない定常過程より得られた標本 $\{x_t\}_{t=1}^T$ の標本平均 $\overline{x}$ は，

$$\mathrm{Var}(\sqrt{T}\overline{x}) \to 2\pi f(0), \qquad \sqrt{T}(\overline{x}-\mu) \xrightarrow{d} N(0, 2\pi f(0)), \quad \text{as } T \to \infty, \tag{3}$$

が成立する．即ち，$\overline{x} = O_p(1/\sqrt{T})$ である．

一方長期記憶モデルの 1 つ，ARFIMA モデルはスペクトル密度関数が原点で連続でないために，通常の $O_p(1/\sqrt{T})$ は成立しない．定常 ARFIMA(p,d,q) 過程では $\overline{x} = O_p(T^{d-1/2})$ となり，$d \in (0,1/2)$ のときは，収束のオーダーが $O_p(1/\sqrt{T})$ より遅い．即ち，通常の短期記憶過程の標本平均の推定量の精度より悪くなる．しかし，不偏推定量かつ一致推定量ある．また，最良線形推定量との効率もまずまずよい（[1]）．

また，平均部分に標本平均を与えたとき，他のパラメータの推定に与える影響は致命的にはならず，漸近理論上は他の推定量の精度に影響しない．

✎ [3] では，標本平均の他に最良線形推定量とメディアンの推定量の精度をシミュレーションにより調べている．

1.2　ARMA パラメータ，差分パラメータ，誤差の分散の推定

正規 ARFIMA(p,d,q) 過程を例に推定法を説明する．推定すべきパラメータは，誤差項の分散 σ^2 と $d,\phi_1,\phi_2,\ldots,\phi_p,\theta_1,\theta_2,\ldots,\theta_q$ である．後者のパラメタータをまとめて $\beta = (d,\phi_1,\phi_2,\ldots,\phi_p,\theta_1,\theta_2,\ldots,\theta_q)'$ とおく．いずれも非線形最適化が必要であるため，統計パッケージが必須である．そのため初期条件や収束条件なども考える必要がある．

1.2.1　最尤法

定常 ARFIMA(p,d,q) モデルである確率過程 $\{x_t\}$ が正規過程の場合は，$\boldsymbol{x}_n = (x_1,\ldots,x_n)'$ が多変量正規分布に従うので，尤度関数 $L_n(\beta,\sigma^2)$ は次で与えられる．

$$L_n(\beta,\sigma^2) = \frac{1}{(2\pi)^{n/2}} |\Gamma_n(\beta,\sigma^2)|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}_n'\Gamma_n(\beta,\sigma^2)^{-1}\boldsymbol{x}_n\right). \tag{4}$$

ここで $\Gamma_n(\beta,\sigma^2)$ は $\boldsymbol{x}_n$ の分散共分散行列である．尤度関数を最大化する β,σ^2 が**最尤推定量**である．

✎ 上記の尤度関数 $L_n(\beta,\sigma^2)$ を最大化するには $\Gamma_n(\beta,\sigma^2)$ の逆行列および行列式を求める必要がある．これらの計算は複雑になり，n が大きいときには計算コストもかかるため，近似式を代入した簡便な方法（次の擬似最尤法）が提案されている．

✎ しかし，標本数が 500 未満であれば，プログラムを入力して一晩待つ程度の時間的余裕があれば，少なくとも S-PLUS では実行することが可能である（500 は逆行列を計算する限界からである）．また 500 以上もある場合であれば，セミパラメトリックな推定法や擬似最尤法でもかなり精度のよい推定ができるだろう．

1.2.2　擬似最尤法（Whittle 推定法）

(2) の ψ_j は β に依存するので，改めて $\psi_j(\beta)$ と書き直し，$g(\lambda;\beta)$ を

$$g(\lambda;\beta) = |\sum_{j=0}^{\infty} \psi_j(\beta) e^{-ij\lambda}|$$

によって定義する．

次にピリオドグラム $I_n(\lambda)$ を

$$I_n(\lambda) = \frac{|\sum_{l=1}^{n} x_t e^{-it\lambda}|^2}{2\pi n}$$

によって定義する（これはスペクトル密度関数 $f(\lambda)$ の推定量である．

そして，

$$U_n(\beta,\sigma^2) = -\frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_j \frac{2\pi I_n(\lambda_j)}{ng(\lambda_j;\beta)} \tag{5}$$

を最小にする β,σ^2 を **Whittle 推定量**である．対数尤度関数 $\log L_n(\beta,\sigma^2)$ の定数部分を削除したとき，(5) の右辺第 1 項と第 2 項はそれぞれ対数尤度関数の $-\log|\Gamma_n(\beta,\sigma^2)|/2$ と $-\boldsymbol{x}_n'\Gamma_n(\beta,\sigma^2)^{-1}\boldsymbol{x}_n/2$ を近似する項になっている．

1.2.3　擬似最尤法（CSS 推定法）

(2) の π_j は β に依存するので，改めて $\pi_j(\beta)$ と書き直し，

$$\varepsilon_t(\beta) = \sum_{j=0}^{t-1} \pi_j(\beta) x_{t-j}$$

とおく．このとき，**CSS 関数**$S(\beta,\sigma^2)$，

$$S_n(\beta,\sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{n}\varepsilon_t(\beta)^2 \tag{6}$$

を最大化する β,σ^2 は **CSS 推定量**という．これは，

$$Q_n(\beta) = \sum_{t=1}^{n}\varepsilon_t(\beta)^2$$

を β により最小化する $\widehat{\beta}$ を求め，残差 $\varepsilon_t(\widehat{\beta}) = \sum_{j=0}^{t-1}\pi_j(\widehat{\beta})x_{t-j}$ を求め，分散の推定量 $\widehat{\sigma}^2$ を，

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t(\widehat{\beta})^2$$

とするのと同じことである．つまり，$\widehat{\beta}$ は $Q_n(\beta)$ の非線形最小二乗推定量である．また時点 0 で切断（$t \le 0$ で $x_t = 0$）したときの，対数尤度関数でもある．

1.2.4 上記の推定量について

上記の 3 つの推定量について次のことが言える．

- ✎ 平均が未知の場合は標本平均 $\overline{x}$ を引いた $\{x_t - \overline{x}\}$ を上述の $\{x_t\}$ に代入しても推定量の漸近的な性質は同じである（$1/\sqrt{n}$ オーダーで真のパラメータへ確率収束する）．
- ✎ 特に，FI(d) 過程の場合，3 つの推定量は $1/\sqrt{n}$ オーダーで平均 0 分散 $6/\pi^2$ の正規分布に漸近的に収束する．
- ✎ Whittle 推定法と CSS 推定法は，$d \geq 1/2$ の非定常のケースでも一致性が証明されている．
- ✎ d についてのセミパラメトリックな推定法も多数存在する（詳しくは [8] の p.154-等を参照）．強調したいのは，ARIMA モデルのモデリング法と異なり差分パラメータを決めうちする必要はなく，推定可能であるという点である．

2 S-PLUS の内臓関数

ここでは今回のレクチャーで使用したいくつかの S-PLUS の内臓関数を紹介する．S-PLUS の内臓関数には，データを生成する `arima.fracdiff.sim`，推定を行う `arima.fracdiff`，分散共分散行列を推定する `arima.fracdiff.var` がある．

`arima.fracdiff` はヘルプファイルやマニュアルを見る限り，[5] よりプログラムが作成されているが，一般には正規分布を仮定した擬似最尤推定である．

- ✎ どのような統計理論も統計パッケージへ適用されるまでおよそ 10 年（またはそれ以上）のラグがあるように思われる．これは理論上の正当化と応用・開発段階である程度時間がかかるため，仕方がないところではある．長期記憶過程についてもその類を出ない．しかし，個々のユーザーはそれをのんびり待つわけにいかない．自分でプログラムを組むのは大変だが，書籍上でプログラムを入手できたり，各統計パッケージのフォーラムやメーリングリスト，論文執筆者の HP で入手できることは少なくない．[1] では長期記憶モデルに関係する様々な S-PLUS のプログラムを公開している．
- ✎ 長期記憶モデルでは推定，検定，予測までの一連の解析手法が確立されたわけではない．例えば ARMA(p, q) の同定は自己相関，偏自己相関のグラフから決める方法が [2] では提唱されているが，このような方法がとれず，情報量基準に頼り数値計算上のみで行っているのが現状であろう．具体的にはセミパラメトリックな推定，検定手法を組み合わせる（または併用する）必要があると思われる．
- ✎ 結論から言うと，無理して `arima.fracdiff` を使う必要性はない．研究は進んでおりこのプログラムは決して最良ではない．例えば，S-PLUS は広義積分も可能なので，スペクトル密度関数を積分して分散共分散行列を計算させたり，漸化式を用いて，偏自己相関を計算させればよい．また [1] で公開しているプログラムより Whittle 推定による擬似最尤推定を行うことができる．また非線形最小二乗推定も Gauss-Newton 法を行うこともできる．さらに厳密な最尤法にはウェーブレット解析を用いる方法も開発されつつあり，いずれにせよ，近い将来よりよいものに置き換わるであろう．
- ✎ `arima.fracdiff` や `arima.fracdiff.var` の分散共分散行列の推定量は怪しいので紹介しない．分散共分散行列の推定量の計算は，検定・予測段階で必要となるが，実際には Godfrey 教授の提唱する方法が最も簡単でプログラム化されやすいのではないかと思う（演者の博士論文または COE のディスカッションペーパーを参照）．

以下はヘルプファイルにある例である．

まず 1 では

$$(1 - 0.2L)(1 - L)^{0.3} x_t = (1 - 0.4L)\varepsilon_t,$$

のデータを 3000 を生成し，

2 では初期値を NA として ARFIMA$(1, d, 1)$ 過程に当てはめ,
3 では推定量の分散共分散行列を求めている.

```
#1# generate a fractionally-differenced ARIMA(1,d,1) model given initial values
ts.sim <- arima.fracdiff.sim(model = list(d = .3, ar = .2, ma = .4), n = 3000)
#2# estimate the parameters in an ARIMA(1,d,1) model for the simulated series
fd.out <- arima.fracdiff(ts.sim, model = list(ar = NA, ma = NA))
#3# modify the covariance estimate by changing the finite-difference interval
arima.fracdiff.var(ts.sim, fd.out, h = .0001)

# arima.fracdiff は d が 1/2 より大きい"d.range"を指定すると（おそらく）計算できない.
# d が 1/2 より小さい範囲であればきちんと出力される. また FI(d) 過程でなおかつ
# 標本の数が 100 未満であればデフォルトで exact な最尤推定量を導くが,
# そうでないケースでは"M"を標本数にする必要がある. 一般の ARIMA(p,d,q) では
# exact な最尤推定ではなく, 擬似最尤推定量である.
#
# arima.fracdiff の作者のページ>
# http://www.stat.washington.edu/raftery/Research/TS/ts_software.html

########## d=-0.40 ###########
# arima.fracdiff の d が 0 より小さいケースのテスト

# FI(-0.40) 過程を生成する自作のプログラムで 200 個の FI(-0.40) を生成
> FIdm040 <- FIdm040.gen(500)[301:500]
# arima.fracdiff による exact な最尤推定量
> arima.fracdiff(FIdm040, model = list(ar = NULL, ma = NULL), d.range=c(-1,.5), M=200)$model$d
[1] -0.4677091
# arima.fracdiff による擬似最尤推定量
> arima.fracdiff(FIdm040, model = list(ar = NULL, ma = NULL), d.range=c(-1,.5))$model$d
[1] -0.4732519
>
>
########## d=1.40 ###########
# 今度は d=1.40 のケースをやってみる.
# FI(1.40) 過程を生成する自作のプログラムで 200 個の FI(1.40) を生成
> FId140 <- FId140.gen(500)[301:500]
> FId140diff <- diff(FId140) # 差分をとる
# arima.fracdiff による exact な最尤推定量
> arima.fracdiff(FId140diff, model = list(ar = NULL, ma = NULL), d.range=c(-1,.5), M=200)$model$d
[1] 0.4228736
# arima.fracdiff による擬似最尤推定量
> arima.fracdiff(FId140diff, model = list(ar = NULL, ma = NULL), d.range=c(-1,.5))$model$d
[1] 0.4256767
# 自作プログラムによるの非線形最小二乗推定量
> CSSE.FId.nocat(1.40,FId140)
[1] 1.177682
# 自作プログラムによる（差分を取ったデータの）非線形最小二乗推定量
> CSSE.FId.nocat(0.40,FId140diff)
[1] 0.4639064
```

自作プログラムの $d = 1.40$ の精度が悪いのは，単にデータの生成の仕方が悪かったためである*1.
以下はユニットルート過程（I(1) 過程）を生成し，パラメータを推定したものである．推定量の精度は（理論上，シミュレーション上で最良ではないにせよ）非線形最小二乗推定量でも十分使える.

```
> FId100.gen <-function(n){cumsum(rnorm(5*n+100))[(5*n+100 - n + 1):(5*n+100)]}
> FId100 <- FId100.gen(200)
> FId100diff <- diff(FId100)
> CSSE.FId.nocat(1.00,FId100)
[1] 1.018946
```

*1 FI(1.40) のデータを生成する関数 `FId140.gen` は，[1] などにあるコレスキー分解を用いてデータを生成している．生成したデータの近似が良くなるためにはデータ数が 100 程度が限界のプログラムとなっているのでこのような結果となったと思われる.

```
> CSSE.FId.nocat(0.00,FId100diff)
[1] 0.04586729
> arima.fracdiff(FId100diff, model = list(ar = NULL, ma = NULL), d.range=c(-1,.5))$model$d
[1] 0.04425555
> arima.fracdiff(FId100diff, model = list(ar = NULL, ma = NULL), d.range=c(-1,.5), M=200)$model$d
[1] 0.04405239
```

3 検定

初日に検定問題

$$\begin{aligned} &(\text{検定問題 } 1) \qquad H_0:\ \text{真の過程は } I(0) \quad \text{vs} \quad H_1:\ \text{真の過程は } I(1), \\ &(\text{検定問題 } 2) \qquad H_0':\ \text{真の過程は } I(1) \quad \text{vs} \quad H_1':\ \text{真の過程は } I(0) \end{aligned} \tag{7}$$

を挙げた．従来の $d = 1$ あるいは $d = 0$ の 2 分法による単位根検定では，より適合度の高い差分パラメータ $d \in (0, 1)$ のときには，それが抜け落ちる可能性があることを指摘した．この検定を原型列を差分を取ることにより，

$$(\text{検定問題 } 3) \qquad H_0:\ d = 0 \quad \text{vs} \quad H_1:\ d \neq 0 \tag{8}$$

に拡張することが可能である．検定問題としては，前節で紹介したパラメトリックなモデルをあてはめ通常の尤度比検定，LM（ラグランジュ未定乗数）検定やワルド検定を行うことが考えられる（例えば [7]）．しかしモデルの特定化を誤った場合，推定量は一致性を持たず，この検定方法は信頼性を失う．

そこで，ここでは検定問題 3 について，セミパラメトリックモデルでの LM 検定とパラメトリックモデルでの LM 検定の 2 種類を紹介する．

3.1 セミパラメトリックモデルでの LM 検定

スペクトル密度関数 $f(\lambda)$ について

$$f(\lambda) = C\lambda^{-2d}, \quad (\lambda \text{ は原点の近傍，} d < 1/2,\ C \text{ は定数}) \tag{9}$$

とする．前節で紹介した Whittle 推定量は，$\lambda_j \in (-\pi, \pi)$ の範囲にあるピリオドグラム I_j を全て利用していたが，ここでは原点まわりの低周波部分の情報のみのある中心化関数 $R(d)$ で考える．

このとき検定統計量は，

$$LM_s = m\left(\left.\frac{\partial R(d)}{\partial d}\right|_{d=0} \right)^2 \Bigg/ \left(\left.\frac{\partial^2 R(d)}{\partial d^2}\right|_{d=0} \right) \tag{10}$$

により定義される．これを近似を用いて

$$LM_s' = m(\widehat{C}_1/\widehat{C}_0)^2 \quad where\ \widehat{C}_k = \frac{1}{m}\sum_{j=1}^{m}\left(\log j - \frac{1}{m}\sum_{i=1}^{m}\log i\right)^k I_j,\ k = 0, 1;\ \lambda_j = \frac{2\pi j}{n} \tag{11}$$

で定義される LM' は m を n とともに大きくすると自由度 1 の χ^2 分布にしたがう．これより χ^2 検定を行うことができる．

✎ 定常 ARFIMA モデルでは，スペクトル密度関数が，

$$f(\lambda) \sim C\lambda^{-2d}, \quad (\text{as } \lambda \to 0),$$

である．スペクトル密度関数が原点近傍でこのような振る舞いをすることでセミパラメトリックモデルを定義することも多い．

3.2 パラメトリックモデルでの LM 検定

ARFIMA(p,d,q) モデルにおいて，最尤推定（または CSS 推定）の中心化関数，$L_n(\beta,\sigma^2)$（$S_n(\beta,\sigma^2)$）を用いるときの検定問題 3 の検定統計量を説明する．ここでは中心化関数に CSS 関数を用いた例を説明する．

中心化関数を $S_n(\beta,\sigma^2)$ とするとき，ARMA パラメータを γ とし，$S_n(d,\gamma,\sigma^2)$ と書き換える．

$$LM_p = \frac{1}{\sqrt{n}}\left(\left.\frac{\partial S_n(d,\gamma,\sigma^2)}{\partial d}\right|_{d=0,\,\gamma=\widehat{\gamma},\,\sigma^2=\widehat{\sigma}^2}\right)\Bigg/\left(\left.\frac{\partial^2 S_n(d,\gamma,\sigma^2)}{\partial d^2}\right|_{d=0,\,\gamma=\widehat{\gamma},\,\sigma^2=\widehat{\sigma}^2}\right)^{1/2} \tag{12}$$

これを書き直して，近似的に残差自己相関を用いた LM_p 検定統計量の表現を得る．

$$LM_p = \sqrt{n}\sum_{j=1}^{n-1}\frac{1}{j}\widehat{\rho}(j)\Big/\omega, \tag{13}$$

$$where \quad \widehat{\rho}(j) = \frac{\sum_{t=1}^{n-j}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t+j}}{\sum_{t=1}^{n}\widehat{\varepsilon}_t^2}, \quad \widehat{\varepsilon}_t = \varepsilon_t(\widehat{\beta}) = \sum_{j=0}^{t-1}\pi_j(\widehat{\beta})x_{t-j} \tag{14}$$

ここで ω は帰無仮説の下で，d の CSS 推定量 $\widehat{d}$ の漸近標準偏差である．即ち，

$$\sqrt{n}(\widehat{d}-d) \xrightarrow{d} N(0,\omega^{-2}), \quad \text{as } n\to\infty.$$

ω の一致推定量は [7] により提案されている．LM_p は帰無仮説の下で標準正規分布に漸近的に近似できるので，標準正規分布を用いた両側検定を行うことができる．さらに対立仮説が，

$$\text{(検定問題 3')} \qquad H_0:\ d=0 \quad \text{vs} \quad H_1:\ d>0 \quad (\text{または } d<0) \tag{15}$$

の片側検定でも標準正規分布の片側の裾の評価で検定できる．

$p=q=0$，即ち，FI(d) 過程であることを仮定し，検定問題 (3) を行う場合は，最も有効な検定統計量は，[7] によるスコア検定統計量，

$$LM_p' = \sqrt{n}\sum_{j=1}^{n-1}\frac{1}{j}\widehat{\rho}(j)\Big/\sqrt{\frac{\pi^2}{6}}, \quad where \quad \widehat{\rho}(j) = \frac{\sum_{t=1}^{n-j}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t+j}}{\sum_{t=1}^{n}\widehat{\varepsilon}_t^2}, \quad \widehat{\varepsilon}_t = x_t, \tag{16}$$

であろう．これは今の設定では推定量を用いていない．LM_p と同様に検定問題 (3) において，帰無仮説の下，漸近的に標準正規分布に従うことから，正規分布の両側検定を行えばよい．また前述のように片側検定も可能である．

この検定では，Godfrey 教授の手法を用いることも有効である（例えば [4]）．

$$\boldsymbol{X} = -\begin{array}{c}\left(\begin{array}{c|c|c} \displaystyle\sum_{l=1}^{i-1}\frac{\widehat{\varepsilon}_{i-l}}{l} & \dfrac{\widehat{\varepsilon}_{i-j}}{\widehat{\phi}(L)} & \dfrac{\widehat{\varepsilon}_{i-j}}{\widehat{\theta}(L)}\end{array}\right) n, \\ \begin{array}{ccc} 1 & \quad p & \quad q \end{array}\end{array} \tag{17}$$

この行列 $\boldsymbol{X}$ は $n\times(1+p+q)$ 行列で，帰無の下 $(\varepsilon_1(\beta),\varepsilon_2(\beta),\ldots,\varepsilon_n(\beta))$ を β について微分し，（$d=0$ として）ARMA 推定量を代入（ハット，$\widehat{\ }$）したラグ多項式を用いて表している．注意として，$\widehat{\varepsilon}_t = 0$ for $t\le 0$ である．このとき

$$\widehat{\boldsymbol{I}}^{-1} = n\widehat{\sigma}^2\big(\boldsymbol{X}'\boldsymbol{X}\big)^{-1} \quad where \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\widehat{\varepsilon}_t^2. \tag{18}$$

は $\widehat{\beta}$ の漸近分散の一致推定量となる．また $\widehat{\boldsymbol{I}}^{-1}$ の $(1,1)$ 成分は ω の一致推定量である．

さらに

$$nR^2 = n\,\widehat{\boldsymbol{\varepsilon}}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\varepsilon}}\Big/\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}, \quad where \quad \widehat{\boldsymbol{\varepsilon}} = (\widehat{\varepsilon}_1, \widehat{\varepsilon}_2, \ldots, \widehat{\varepsilon}_n)' \tag{19}$$

は帰無の下で漸近的に自由度 1 の χ^2 分布に従うので，自由度 1 の χ^2 分布の右の裾から棄却するかどうか判断する．

✎ 残差自己相関を用いた検定として，いわゆる**（修正）かばん検定**の実行も ARMA モデルのケースと同様可能である．

4　予測

ここでは一般の長期記憶モデルの予測の方法（セミパラメトリックな予測）と ARFIMA(p,d,q) モデルを仮定して，パラメータを推定したときの予測量を構成する方法（パラメトリックな予測）を紹介する．

4.1　セミパラメトリックな予測

✎ 有限標本における最良線形予測量を構成する議論は時系列に関係するどの本にも掲載されているものである．この最良線形予測量を最小二乗法で推定する，という方法は定常過程で MA(∞) 表現可能であれば有効である．

定常長期過程 $\{x_t\}$ を次のように定義する．

$$x_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \tag{20}$$

ここで $\{\varepsilon_t\} \sim NID\,(0, \sigma^2)$, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, そして自己共分散 $\gamma_x(j)$ は $j \to \infty$ で

$$\gamma_x(j) \sim \beta_x(j)\, j^{2d-1} \quad (\text{where } d < 0.5 \text{ and } |\beta_x(j)| < M \text{ for some constant } M > 0). \tag{21}$$

が成立するものとする．$\{x_t\}_{t=1}^{p}$ を標本とするとき，x_{p+h} を予測する最良線形予測量 $\hat{x}_{p+h}^{(B)}$ は次で与えられる．

$$\hat{x}_{p+h}^{(B)} = [x^{(p)}]'\Phi = \Phi_1 x_1 + \cdots + \Phi_p x_p \tag{22}$$

ここで，$\Phi_1, \Phi_2, \ldots, \Phi_p$ は

$$\mathrm{E}[x_{p+h} - (\Phi_1 x_1 + \cdots + \Phi_p x_p)]^2 \quad \text{を } \Phi_1, \Phi_2, \ldots, \Phi_p \text{ について最小化する} \tag{23}$$

ことで与えられる．この解は次のようになる．

$$\begin{aligned} x^{(p)} &= (x_p, \ldots, x_1)', \\ \Phi &= (\Phi_p, \ldots, \Phi_1)' = \Sigma_p^{-1}\gamma_h^{(p)}, \\ \Sigma_p &= E[x^{(p)}[x^{(p)}]'] = [(\gamma_x(i-j))]_{i,j=1}^{p}, \\ \gamma_h^{(p)} &= E[x_{p+h}x^{(p)}] = (\gamma_x(h), \ldots, \gamma_x(p+h-1))', \end{aligned} \tag{24}$$

そして予測の平均二乗誤差 $\sigma_p^2(h)$ は次で与えられる．

$$\sigma_p^2(h) \equiv E[x_{p+h} - \hat{x}_{p+h}^{(B)}]^2 = \gamma_x(0) + [\gamma_h^{(p)}]'\Sigma_p^{-1}\gamma_h^{(p)} \tag{25}$$

ここで Σ_p に逆行列が存在する仮定を追加した（実は存在しなくても一般化逆行列で代替できる．ここまでは [1] の section 8.7 を参照)．

(23) より最良線形予測量は h 期先のデータを L^2 ノルムの意味で最小化しているといえる．紹介する最小二乗推定量を用いる方法はユークリッドノルムを用いて h 期先のデータを最小化しようとしているといえよう．

上で定義した $n+p+h-1$ 個の標本 $(x_{-n-h+2},\ldots,x_0,x_1,\ldots,x_p)$ が得られたとする．ここで n は十分大とする． x_{p+h} を $(x_1,\ldots,x_p)$ で予測するために，Φ を推定する必要が在る．そのために，モデルを次のように書き換える:

$$y = X\Phi + \xi \tag{26}$$

ここで y は p-vector で $y=(x_{-n+p+1},\ldots,x_p)'$, X は $n\times p$ 行列で (i,j) 成分は $x_{p-h-n+1+i-j}$ である．ξ は n-vector で

$$\xi = (\xi_{-n+p+1},\ldots,\xi_p)',\quad \xi_t \equiv \Phi(L)x_t,\quad \Phi(L) = 1-\Phi_p L^h - \cdots - \Phi_1 L^{p+h-1} \tag{27}$$

即ち ξ_t は人工的な AR(p) モデルの誤差項である．回帰を行い最小二乗推定量 $\hat\Phi$ は,

$$\hat\Phi = [X'X]^{-1}X'y \tag{28}$$

で得られる．これを $\hat x^{(B)}_{p+h}$ の Φ に代入して（分散共分散の構造が未知のケースの）予測量をえる．

✎ (28) の最小二乗推定は**ユール＝ウォーカー法**により代替できるため計算コストはさほどかからない．

✎ (28) を用いる予測量は p を固定したとき，$n\to\infty$ で予測の平均二乗誤差は最良線形予測量のそれ (25) に一致する．

✎ この予測法の大きな欠点は非定常のケースへ拡張が難しいことである．

✎ 情報量基準を用いて最良の p を探すことが必要となり，R.J.Bhansali 教授が積極的に行っている．Bhansali 教授の HP より最新論文がダウンロード可能である[*2]．この推定量を用いる予測方法の理論検証は慶応大学の柴田里程教授がパイオニアとなった．

✎ また矢島美寛教授も統計グループの東大輪講で 1991 年に発表されている．さらに矢島美寛教授は J.Hidalgo 教授との共著 [6] で周波数領域のアプローチで予測を行っている．

✎ おそらく季節性を持つ時系列では Φ のいくつかの成分は 0 としたほうが良いであろう．0 制約のための検定手法の開発は未解決である．

✎ 観測する標本の大きさ n の時系列の分散共分散の構造が既知であれば，$p=n$ とした予測量 $\hat x^{(B)}_{n+h}$ を構成すれば，それが最良の線形予測量である．しかしそういうケースは稀なのでこのような手法を用いている．p を闇雲に大きくできない（実際には $p\le 10$ 程度）のは，最小二乗推定量 $\hat\Phi$ を推定するときの有効な標本数が p を大きくすると減り，$\hat\Phi$ の Φ に対する精度が落ちるためである．即ち $\hat\Phi$ に使う標本数と，予測量を構成する p との trade-off がある．

4.2 パラメトリックな予測

✎ 無限標本における最良線形予測量を構成する議論は時系列に関係するどの本にも掲載されているものである．最良線形予測量の表現に推定量を代入して代替する，という方法は，（いくらかの仮定のもとで）AR(∞) かつ MA(∞) 表現可能であれば有効である．

定常かつ反転可能な系列:

$$\sum_{j=0}^{\infty}\pi_j y_{t-j} = \varepsilon_t, \qquad \text{and} \qquad y_t = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}, \qquad t=0,\pm1,\ldots, \tag{29}$$

$\sum_{j=0}^{\infty}\pi_j^2$, $\sum_{j=0}^{\infty}\psi_j^2<\infty$, そして $\pi_0=\psi_0=1$ をおく．このとき，$y_n,y_{n-1},\ldots$ を用いる y_{n+h} の最良線形予測量 $y_n(h)$ は次で表される．

$$y_n(h) = \sum_{j=1}^{\infty}c_j(h)y_{n+1-j} = \sum_{j=h}^{\infty}\psi_j\varepsilon_{n+h-j},\quad c_j(h) = -\sum_{i=0}^{h-1}\psi_i\pi_{j+h-i-1}\quad \text{for } j\ge 1, \tag{30}$$

[*2] http://www.liv.ac.uk/maths/SOR/HOME/RJ_Bhansali.html

([2] の (A5.2.3) を参照)．さらに予測の平均二乗誤差 $\sigma_y^2(h)$ は次で与えられる．

$$\sigma_y^2(h) \equiv \mathrm{E}\left[y_{n+h} - y_n(h)\right]^2 = \mathrm{E}\left[\sum_{j=0}^{h-1} \psi_j \varepsilon_{n+h-j}\right]^2 = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2. \tag{31}$$

定常 ARFIMA モデルで予測をする場合，最も困るのは AR(∞) 表現でも MA(∞) 表現でも有限次のラグ多項式が使えないことである．我々は有限標本しか持ち得ないので，どうのように予測量を構成すればよいのかわからない．

演者は，博士論文で h より十分大きい n 個のデータ $\{x_t\}_{t=1}^n$ が得られたとき，n 個のデータを使って予測量を構成するケースを考えた．(1) の ARFIMA(p,d,q) 過程で時点 0 で切断（$\varepsilon_t = x_t = 0,\ t \le 0$）されているものとする．また正規性を仮定する．時点 0 で切断されているときの最良線形予測量 $x_n(h)$ の表現は単に (30) の時点 0 で切断しただけの表現となる（また y_t を x_t で置き換える）．

$$x_n(h) = \sum_{j=1}^{n} c_j(h) x_{n+1-j} = \sum_{j=h}^{n+h-1} \psi_j \varepsilon_{n+h-j}, \quad c_j(h) = -\sum_{i=0}^{h-1} \psi_i \pi_{j+h-i-1} \quad \text{for } j \ge 1. \tag{32}$$

また，この予測の平均二乗誤差は (31) と同じく

$$\sigma_x^2(h) \equiv \mathrm{E}\left[x_{n+h} - x_n(h)\right]^2 = \mathrm{E}\left[\sum_{j=0}^{h-1} \psi_j \varepsilon_{n+h-j}\right]^2 = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2. \tag{33}$$

となる．

(32) の $c_j(h)$ は d と ARMA パラメータより構成されているので，2 節で用いたパラメータベクトル β を用いて，$c_j(h,\beta)$ と書き換える．また π_j と ψ_j も同じ理由から $\pi_j(\beta)$ と $\psi_j(\beta)$ と書く．また 1.2.3 節の CSS 推定法による β の推定量を $\widehat{\beta}$ とおくとき，最良線形予測量表現に基づく CSS 推定量を与えた $\widehat{x}_n(h)$ は次で与えられる．

$$\widehat{x}_n(h) = \sum_{j=1}^{n} c_j(h,\widehat{\beta}) x_{n+1-j} = \sum_{j=h}^{n+h-1} \psi_j(\widehat{\beta}) \widehat{\varepsilon}_{n+h-j} \tag{34}$$

ここで $\{\widehat{\varepsilon}_t\}_{t=1}^n$ は残差系列で $\widehat{\varepsilon}_t = \sum_{j=0}^{t-1} \pi_j(\widehat{\beta}) x_{t-j}$ で定義される．

✎ 演者の博士論文では，$\widehat{x}_n(h)$ の予測の平均二乗誤差歯、$n \to \infty$ で (33) の $\sigma_x^2(h)$ へ一致することを示している．また非定常性，季節性，非正規分布を持つモデルで結果を拡張している．

5　2 日目の終わりに

参考にした文献を挙げる．

参考文献

[1] Beran, J.: *Statistics for Long-Memory Processes*. Chapman and Hall, London, 1994.

[2] Box, G.E.P. and Jenkins, G.M.: *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1976.

[3] Chung, C.-F. and Baillie, R.T.: Small sample bias in conditional sum-of-squares estimators of fractionally integrated ARMA models. *Empirical Economics* 18 (1993), 791-806.

[4] Godfrey, L.G.: *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches*. Cambridge: Cambridge University Press, (1988).

[5] Haslett, J. and Raftery, A.E.: . Space-time modelling with longmemory dependence: Assessing Ireland's wind power resource (with Discussion). *J. of the Royal Stat. Soc., series C-Applied Statistics*, (1989), 38:1?50.

[6] Hidalgo, J. and Yajima, Y.: Prediction and signal extraction of strongly dependent processes in the frequency domain. *Econometric Theory* 18 (2002), 584-624.

[7] Tanaka, K.: The nonstationary fractional unit root. *Econometric Theory* 15 (1999), 549-582.

[8] 矢島美寛: 長期記憶を持つ時系列モデル. 経済時系列の統計ーその数理的基礎，統計科学のフロンティア第8巻，岩波書店，(2003), 103-202.