状態変化を伴うモデルの推定[†]

1. TAR モデルを推定する。

1.1 パッケージ “tseriesChaos” と “tsDyn” をインスツールする。

TAR モデルを推定するために R のパッケージ tseriesChaos と tsDyn 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)」→「tseriesChaos」→「OK」とクリックする(tsDyn についても同様)。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(tseriesChaos)
```

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 tseriesChaos を使用できる様になる(tsDyn についても同様)。

次に使用するデータを読み込む。今回は R にあらかじめ用意してある lynx というデータを使う。これは北西カナダのマッケンジー川地区で捕獲されたオオヤマネコの数の年次データである。このデータを読み込むには

```
> str(lynx)
```

とタイプする。また

```
> lynx
```

および

```
> summary(lynx)
```

とすればデータそのもの、およびデータの標準的な記述統計量が出力される。データをプロットするには

```
> plot(lynx)
```

とタイプする。このデータの常用対数(底が 10 の対数)をとったものを

```
> x = log10(lynx)
```

とする。以降ではこの変換後のデータ x を分析する

[†]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のでご了承ください。

1.2. `setar()`関数による TAR モデルの推定

上記のデータに対して、まず通常の AR(m)モデル推定してみよう。通常の AR(m)モデルも `tsDyn` パッケージによって推定できる。`liner()`関数を用いて推定する。

```
 > ar2 = linear(x, m=2)
```

推定結果は

```
> ar2

Non linear autoregressive model

AR model
Coefficients:
     const       phi.1      phi.2
 1.0576005  1.3842377 -0.7477757
```

となる(もしくは `summary(ar2)` とすれば推定結果についてより詳しい情報が得られる)。

次に TAR モデルを推定しよう。以下の TAR モデルを推定する。

$$x_t = \begin{cases} c_1 + \phi_{11}x_{t-1} + \phi_{12}x_{t-2} + \ldots + \phi_{1L}x_{t-L} + \varepsilon_t, & x_{t-\delta} \le \tau, \\ c_2 + \phi_{21}x_{t-1} + \phi_{22}x_{t-2} + \ldots + \phi_{2H}x_{t-H} + \varepsilon_t & x_{t-\delta} > \tau, \end{cases} \qquad (1)$$

$$E(\varepsilon_t) = 0,\ \mathrm{var}(\varepsilon_t) = \sigma_\varepsilon^2$$

ここで 整数 δ は 1 以上の整数とする(状態変数が x_t 自体の過去の値である事、レジームごとに AR の次数が異なっている事、および ε_t の分散はどちらのレジームでも同じになっている事に注意。このように状態変数が x_t 自体の過去の値になっているモデルの事を **Self exciting TAR (SETAR) モデル**と呼ぶ)。まず $L = H = m$ のケース(つまりどちらのレジームでも AR の次数は m)および δ =1 の場合を考える。この場合に上記のモデルにおいて $m = 3$ のモデルを推定するには

```
> set1 = setar(x,m=3)
```

とタイプする。推定結果は(青字)

```
> summary(set1)

Non linear autoregressive model

SETAR model ( 2 regimes)
Coefficients:
Low regime:
     phiL.1     phiL.2      phiL.3    const L
 1.1175812 -0.1165550 -0.1532651  0.5886583

High regime:
     phiH.1     phiH.2      phiH.3    const H
 1.4029086 -0.6538670 -0.2364283  1.4051764

Threshold:
 -Variable: Z(t) = + (1) X(t)+ (0)X(t-1)+ (0)X(t-2)
```

```
-Value: 2.588
Proportion of points in low regime: 30.63%      High regime: 69.37%

Residuals:
       Min        1Q    Median        3Q       Max
-0.566419 -0.105803 0.013666 0.135648 0.426856

Fit:
residuals variance = 0.03842,  AIC = -354, MAPE = 5.683%

Coefficient(s):

        Estimate  Std. Error  t value  Pr(>|t|)
const L   0.58866    0.29640   1.9860 0.0496118 *
phiL.1    1.11758    0.16442   6.7970 6.482e-10 ***
phiL.2   -0.11655    0.20023  -0.5821 0.5617302
phiL.3   -0.15327    0.11730  -1.3066 0.1941901
const H   1.40518    0.23618   5.9495 3.513e-08 ***
phiH.1    1.40291    0.11249  12.4719 < 2.2e-16 ***
phiH.2   -0.65387    0.18814  -3.4754 0.0007403 ***
phiH.3   -0.23643    0.13197  -1.7915 0.0760691 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold
Variable: Z(t) = + (1) X(t) + (0) X(t-1)+ (0) X(t-2)

Value: 2.588
```

となる。これより推定されたモデルは(推定結果を小数点以下適当な桁数で四捨五入して)

$$x_t = \begin{cases} 0.59 + 1.12x_{t-1} - 0.12\,x_{t-2} - 0.15x_{t-3} + \varepsilon_t, & x_{t-1} \le 2.59, \\ 1.41 + 1.4\,x_{t-1} - 0.65\,x_{t-2} - 0.24\,x_{t-3} + \varepsilon_t & x_{t-1} > 2.59, \end{cases}$$
$$E(\varepsilon_t) = 0,\ \ \mathrm{var}(\varepsilon_t) = 0.038$$

となる。

　また、2 つのレジームで異なった次数の AR モデルを推定する事もできる。例えば `set1` の推定結果において、ローレジームでは次数 2 と 3 の AR 係数が有意ではなく、またハイレジームでは次数 3 の AR 係数が有意ではない。よって AR の次数としてローレジームでは 1 (L = 1)、ハイレジームでは 2 (H = 2)としたモデルを推定しなおしてみよう。以下のコマンドで推定する(ここでは状態変数は 1 期前の値とする)。

```
> set2 = setar(x,mL=1,mH=2)
```

ここで `mL` が `L` の値、`mH` が `H` の値である。推定結果 (青字) は

```
> summary(set2)

Non linear autoregressive model

SETAR model ( 2 regimes)
Coefficients:
Low regime:
```

```
    phiL.1   const L
0.9863628 0.1930968

High regime:
     phiH.1     phiH.2     const H
 1.5476983 -0.9562741  1.1808695

Threshold:
-Variable: Z(t) = + (1) X(t)+ (0)X(t-1)
-Value: 2.558
Proportion of points in low regime: 27.68%      High regime: 72.32%

Residuals:
      Min        1Q    Median        3Q       Max
-0.689146 -0.127983  0.029727  0.142420  0.474525

Fit:
residuals variance = 0.0443,  AIC = -343, MAPE = 6.156%

Coefficient(s):

         Estimate  Std. Error  t value    Pr(>|t|)
const L  0.193097    0.292719   0.6596     0.5109
phiL.1   0.986363    0.132005   7.4722  2.093e-11 ***
const H  1.180869    0.213553   5.5296  2.220e-07 ***
phiH.1   1.547698    0.088773  17.4343  < 2.2e-16 ***
phiH.2  -0.956274    0.073336 -13.0396  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold
Variable: Z(t) = + (1) X(t) + (0) X(t-1)

Value: 2.558
```

となり、推定されたモデルは

$$x_t = \begin{cases} 0.19 + 0.99\, x_{t-1} + \varepsilon_t, & x_{t-1} \le 2.59, \\ 1.18 + 1.55\, x_{t-1} - 0.96 x_{t-2} + \varepsilon_t & x_{t-1} > 2.59, \end{cases}$$
$$E(\varepsilon_t) = 0,\ \ \mathrm{var}(\varepsilon_t) = 0.044$$

となる。

上記のモデルにおいて 状態変数は x_{t-1} (x_t の 1 期前の値)であったが、この値を x_t の 2 期以上前の過去の値($x_{t-\delta}$, $2 \le \delta$)とすることができる。この時、引数に "thDelay = δ - 1" を加える(**δ から 1 を引いた値であることに注意)**。例えば上記のモデルにおいて状態変数を x_{t-2} とするのであれば($L = 1, H = 2$ とする)、

```
> set3=setar(x,mL=1,mH=2,thDelay=1)
```

とすればよい。推定結果は

```
> summary(set3)

Non linear autoregressive model

SETAR model ( 2 regimes)
Coefficients:
```

```
Low regime:
    phiL.1     const L
1.08403352  0.01640729

High regime:
  phiH.1     phiH.2   const H
1.492201  -1.158738  2.069329

Threshold:
-Variable: Z(t) = + (0) X(t)+ (1)X(t-1)
-Value: 2.867
Proportion of points in low regime: 49.11%      High regime: 50.89%

Residuals:
     Min        1Q    Median        3Q       Max
-0.725279  -0.117072  0.023711  0.115959  0.511283

Fit:
residuals variance = 0.04379,  AIC = -345, MAPE = 5.959%

Coefficient(s):

         Estimate  Std. Error  t value  Pr(>|t|)
const L  0.016407   0.168480   0.0974   0.9226
phiL.1   1.084034   0.066237  16.3660  < 2.2e-16 ***
const H  2.069329   0.402274   5.1441  1.196e-06 ***
phiH.1   1.492201   0.086225  17.3060  < 2.2e-16 ***
phiH.2  -1.158738   0.122828  -9.4339   8.375e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold
Variable: Z(t) = + (0) X(t) + (1) X(t-1)

Value: 2.867
```

となる。

状態数を 3 つにするする場合、すなわち

$$x_t = \begin{cases} c_1 + \phi_{11}x_{t-1} + \phi_{12}x_{t-2} + \ldots + \phi_{1L}x_{t-L} + \varepsilon_t, & x_{t-\delta} \le \tau_1, \\ c_2 + \phi_{21}x_{t-1} + \phi_{22}x_{t-2} + \ldots + \phi_{2M}x_{t-M} + \varepsilon_t & \tau_1 < x_{t-\delta} \le \tau_2, \\ c_3 + \phi_{31}x_{t-1} + \phi_{32}x_{t-2} + \ldots + \phi_{3H}x_{t-H} + \varepsilon_t & \tau_2 < x_{t-\delta}, \end{cases}$$

$$E(\varepsilon_t) = 0,\ \mathrm{var}(\varepsilon_t) = \sigma_\varepsilon^2$$

というモデルを推定するには (L=1, M =2, H=3, δ =2 の時)

```
> set4=setar(x,mL=1,mM=2,mH=3,thDelay=1,nthresh=2)
```

とする。ここで `nthresh` は number of thresholds (閾値 τ_i の数)で、これを `nthresh=2` とすることによって状態数が 3 つになる。推定結果は

```
> summary(set4)

Non linear autoregressive model

SETAR model ( 3 regimes)
Coefficients:
```

```
Low regime:
   const.L     phiL.1
0.07826235 1.09951710

Mid regime:
    const.M      phiM.1      phiM.2
-0.56938414  1.25437711  0.01871932

High regime:
     const.H       phiH.1       phiH.2       phiH.3
 2.064934905  1.496642082 -1.168974266  0.007485407

Threshold:
-Variable: Z(t) = + (0) X(t)+ (1)X(t-1)
-Value: 2.446 2.867
Proportion of points in low regime: 21.62%      Middle regime: 27.03%
High regime: 51.35%

Residuals:
       Min         1Q     Median         3Q        Max
-0.5595518 -0.1340602 -0.0015146  0.1348984  0.5075138

Fit:
residuals variance = 0.03881,  AIC = -348, MAPE = 5.711%

. . . . . (以下略)
```

となる。

また `setar()` 関数において状態変数を外生変数として与えてやる事もできる。外生変数として `y` が与えられているのであれば `"thVar = y"` とすることにより状態変数を `y` とすることができる。例えば外生変数を時点とした以下のモデル:

$$x_t = \begin{cases} c_1 + \phi_{11}x_{t-1} + \phi_{12}x_{t-2} + \ldots + \phi_{1L}x_{t-L} + \varepsilon_t, & t-1 \le \tau, \\ c_2 + \phi_{21}x_{t-1} + \phi_{22}x_{t-2} + \ldots + \phi_{2H}x_{t-H} + \varepsilon_t & \tau < t-1, \end{cases}$$

$$E(\varepsilon_t) = 0, \ \mathrm{var}(\varepsilon_t) = \sigma_\varepsilon^2$$

(つまり $t = \tau + 2$ で構造変化が起こっている)を推定するには、まず外生変数として時点 time を作成する(x の観測数と同じだけの時点が必要。ここでは x の観測数は 114 なので):

```
> time=1:114
```

(`a:b` で `a` から `b` までの整数を並べたベクトルが作成される)。これを用いて (L=3, H=2 として)

```
> set5=setar(x,mH=2,mL=3,thVar=time)
```

と入力する。結果は

```
> summary(set5)

Non linear autoregressive model

SETAR model ( 2 regimes)
Coefficients:
Low regime:
   const.L     phiL.1     phiL.2     phiL.3
 1.2984306  1.2479485 -0.4796353 -0.2270403
```

```
High regime:
   const.H     phiH.1     phiH.2 
 1.1022246  1.2981732 -0.6566474 

Threshold:
-Variable: external-Value: 75
Proportion of points in low regime: 67.57%       High regime: 32.43% 

Residuals:
      Min        1Q    Median        3Q       Max 
-0.637445 -0.136761  0.019679  0.137486  0.559789 

Fit:
residuals variance = 0.047,  AIC = -333, MAPE = 6.329%
```

．．．．．．（以下略）

となる。構造変化点の推定値は $t=77$ 時点から状態 2 になっているので 77 時点であることがわかる。

setar() 関数ではモデルから定数項を除いたり、トレンドを入れたり、トレンドと定数項の両方を入れたりすることもできる。例えば、(1)式において $L=H=3$、$\delta=1$ のモデルで定数項がないモデルを推定するには

```
> set=setar(x,m=3,include=c("none"))
```

とする。トレンド項だけを入れる場合は `include=c("trend")`、トレンド項と定数項の両方を入れる場合には `include=c("both")` とする。

1.3. lstar() 関数による LSTAR モデルの推定

以下の LSTAR モデルを推定する。

$$
\begin{aligned}
x_t &= (c_1+\phi_{11}x_{t-1}+\phi_{12}x_{t-2}+\ldots+\phi_{1m}x_{t-m})(1-G(x_{t-\delta};\gamma,a)) \\
&\quad +(c_2+\phi_{11}x_{t-1}+\phi_{12}x_{t-2}+\ldots+\phi_{1m}x_{t-m})G(x_{t-\delta};\gamma,a)+\varepsilon_t\,, \\
&\quad E(\varepsilon_t)=0,\ \operatorname{var}(\varepsilon_t)=\sigma_\varepsilon^2\,, \\
&\quad G(x_t;\gamma,a)=\frac{1}{1+\exp(-\gamma(x_{t-\delta}-a))},\ \gamma>0
\end{aligned}
$$

これは lstar() 関数によって推定できる(LSTAR モデルにおいて全てのレジームで AR の次数は同じにならないといけないことに注意)。例えば $m=2$ 、$\delta=1$ のモデルを推定するには

```
> ls1=lstar(x,m=2)
```

とタイプする。推定結果は

```
> summary(ls1)

Non linear autoregressive model

LSTAR model
Coefficients:
Low regime:
     const1     phi1.1      phi1.2 
 0.4145994  1.2337266 -0.3270309
```

```
High regime:
     const2     phi2.1      phi2.2
 0.7981470  0.3103511 -0.6352231

Smoothing parameter: gamma = 100

Threshold
Variable: Z(t) = + (1) X(t) + (0) X(t-1)

Value: 2.568

Residuals:
      Min        1Q    Median        3Q       Max
-0.595535 -0.132499  0.040486  0.129661  0.467416

Fit:
residuals variance = 0.04036,  AIC = -350, MAPE = 5.94%

Coefficient(s):
          Estimate  Std. Error  t value  Pr(>|z|)
const1    0.414599    0.289072   1.4342 0.1515036
phi1.1    1.233727    0.146648   8.4128 < 2.2e-16 ***
phi1.2   -0.327031    0.098753  -3.3116 0.0009277 ***
const2    0.798147    0.346389   2.3042 0.0212117 *
phi2.1    0.310351    0.167131   1.8569 0.0633204 .
phi2.2   -0.635223    0.121885  -5.2117 1.872e-07 ***
gamma   100.000113   77.576158   1.2891 0.1973782
th        2.567993   0.014846 172.9720 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Non-linearity test of full-order LSTAR model against full-order AR
model
 F = 8.529 ; p-value = 0.00036569

Threshold
Variable: Z(t) = + (1) X(t) + (0) X(t-1)
```

となる。これより推定されたモデルは

$$
\begin{aligned}
x_t &= (0.41 + 1.23x_{t-1} - 0.33x_{t-2})(1 - G(x_t; 100, 2.57)) \\
&\quad + (0.8 + 0.31x_{t-1} - 0.64x_{t-2})G(x_t; 100, 2.57) + \varepsilon_t\,, \\
&\qquad E(\varepsilon_t) = 0,\ \ \mathrm{var}(\varepsilon_t) = 0.04, \\
&\qquad G(x_t; \gamma, a) = \frac{1}{1 + \exp(-100(x_t - 2.57))},\ \gamma > 0
\end{aligned}
$$

となる。

LSTAR モデルの推定においても SETAR モデルの時と同様 "thVar" や "thDelay" の値を設定する事ができる。ただし、これらの値の設定がうまくいかないと LSTAR モデルにおいて推定がうまくいかなくなる場合があることに留意する必要がある。

2. 予測値の計算

setar(), lstar() 関数によって推定した結果を用いて予測を行うことができる。予測には predict()関数を用いる。ここでは、x の 114 個の観測値のうち最初の 104 個を用いてモデルの推定を行い、推定したモデルを用いて(最初の 104 個に基づいて)10 期先予測を行い、実際の 10 個との比較を行ってみよう。まず、window()関数を用いて、x の最初の 104 個のデータを取り出す。x1 と名前を付けよう。ここで window()は時系列オブジェクト(時系列属性を持つデータ)から特定のデータを抜き出す関数である(時系列オブジェクトの作り方については他の資料を参照)。

```
> x1=window(x,end=1924)
```

次に最後の 10 個を取り出す。

```
> x2=window(x,start=1925)
```

それぞれの関数で推定

```
> set.x1=setar(x1,m=2,thDelay=1)
> ls.x1=lstar(x1,m=2,thDelay=1)
```

次に推定したモデルを用いて 10 期先予測を行う。予測の仕方としていくつかあるが、ここではまずモンテカルロ法(MC 法)を用いた予測をしてみる(これは誤差項が正規分布と仮定して行う)

```
> set.pred=predict(set.x1,n.ahead=10,type=c("MC"),nboot=10000,ci=0.95,
+ boot1Zero=FALSE)
```

ここで ci=α とすると予測の α% および $(1-\alpha)$ %分位点を計算する。nboot は MC 法に用いるシミュレーションの回数である。結果を確認する。

```
> set.pred
$pred
Time Series:
Start = 1925
End = 1934
Frequency = 1
 [1] 3.556083 3.424928 3.076340 2.735621 2.582736 2.633636 2.787521 2.942092
 [9]  3.048769  3.093558

$se
Time Series:
Start = 1925
End = 1934
Frequency = 1
           5%      95%
1925 3.222738 3.884888
1926 2.795722 4.049314
1927 2.337862 3.866821
1928 1.991964 3.506214
1929 1.741605 3.409401
1930 1.670183 3.566844
1931 1.784149 3.696169
1932 1.997395 3.791213
1933 2.153271 3.862638
1934 2.177127 3.912146
```

青字が予測値、緑が 5%と 95%の予測の分位点である(これは 90%予測区間と考えることができる)。予測値と実際の値をプロットしてみる。まず実際の値をプロットする

```
> plot(x2,ylim=c(1.5,4.5),lwd=3)
```

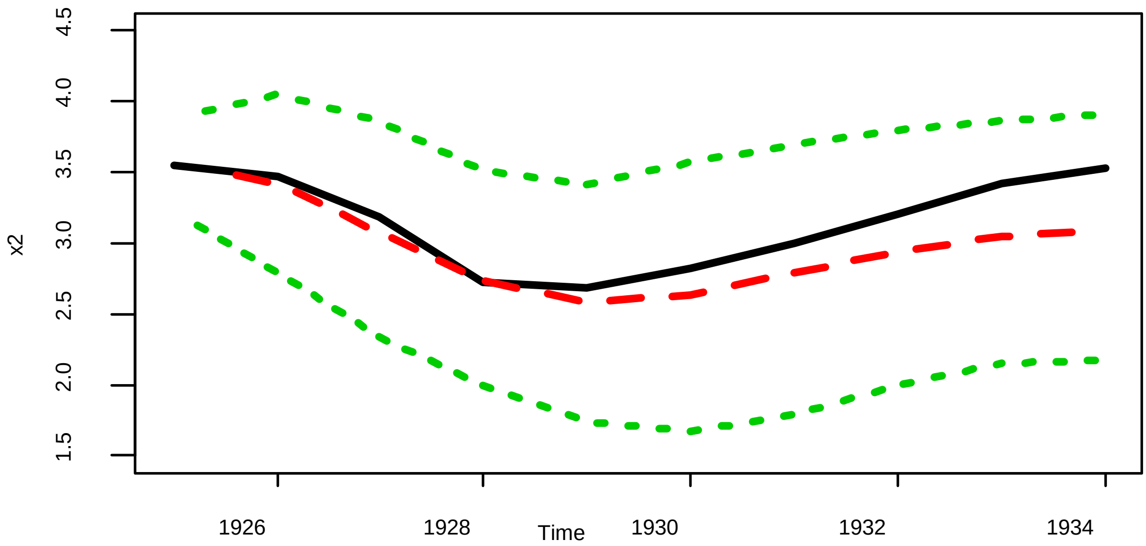
次に予測値を赤の点線で書き込んでみる。

```
> lines(set.pred$pred,lty=2,col=2,lwd=3)
```

次に 90%予測区間を緑のより細かい点線で書き込んでみる。

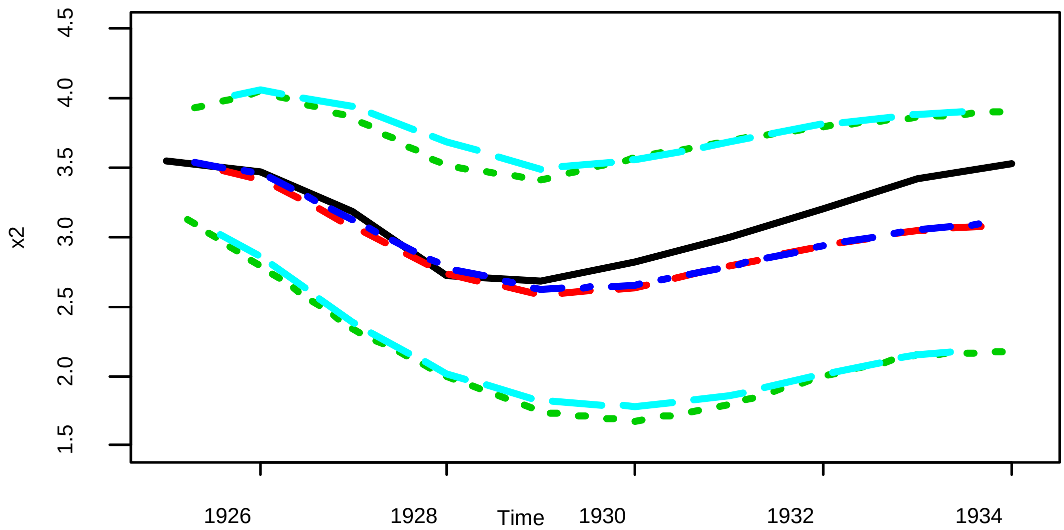
```
> lines(set.pred$se[,1],lty=3,col=3,lwd=3)
> lines(set.pred$se[,2],lty=3,col=3,lwd=3)
```

以下の図のようになる。



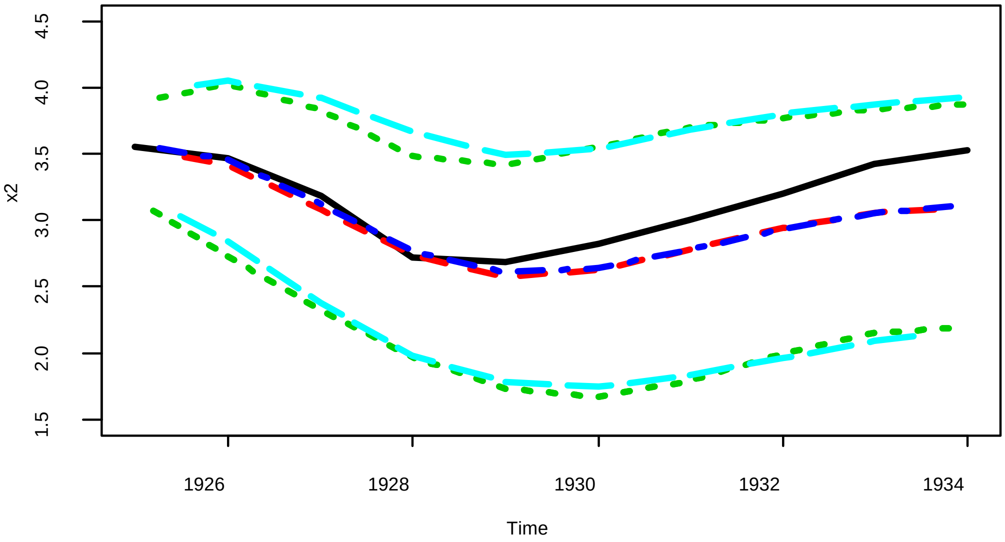
次に `lstar()` 関数の推定結果(LSTAR モデル)を用いて同様の予測を行い、予測値を青線で、90%予測区間を水色で書き込んでみる。

```
> ls.pred=predict(ls.x1,n.ahead=10,type=c("MC"),nboot=10000,ci=0.95,
+ boot1Zero=FALSE)
> lines(ls.pred$pred,lty=4,col=4,lwd=3)
> lines(ls.pred$se[,1],lty=5,col=5,lwd=3)
> lines(ls.pred$se[,2],lty=5,col=5,lwd=3)
```



どちらのモデルで予測してもあまり変わらないことがわかる。同様のことを今度は Bootstrap 法を用いて行ってみる(これは正規分布の仮定は必要ない)。これは `predict()` の中で

`type=c("MC")` を `type=c("bootstrap")` に変えるだけで他は同じでよい。以下の図のようになる(コードは略す)。



どちらの方法でやってもあまり変わらないことがわかる。

練習問題

tsdata.txt にある Topix の値についてその収益率に対して SETAR モデルと(L=H=1)と LSTAR モデル(m=1)を推定しなさい。

tsdata.txt にある topix の値についてその収益率に対して様々な SETAR モデル、および時点を外生変数とした SETAR モデルを推定しなさい。