伝達関数モデルによる時系列予測

小　村　賢　二

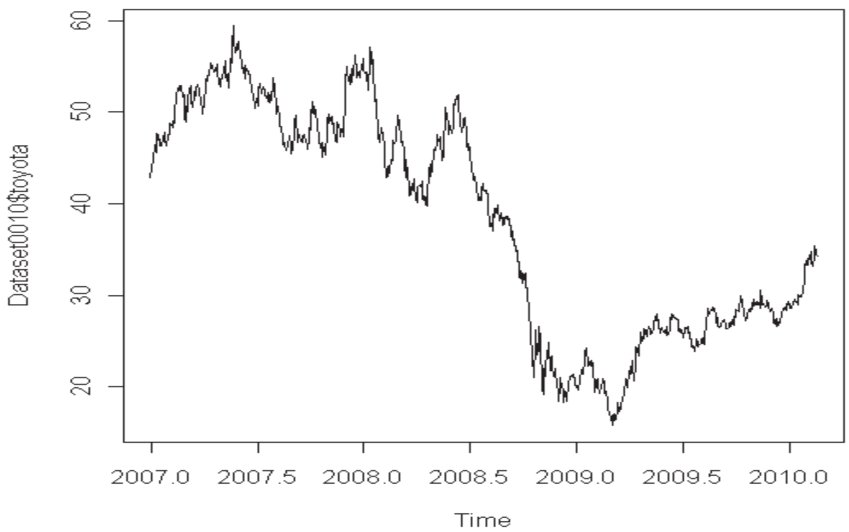
要約；時系列データ｛y_t∈T｝(2007 年１月から 2010 年２月５日まで）の予測問題を考える．この期間には図の如く 200８年９月１６日の New York でのリーマン&ブラザー証券の破綻により始まった世界的な金融不安による世界的な株価の下落が含まれている．この要因を時系列モデルの干渉モデルを使い、時系列の構造的な変化後の予測を行なう． R-2.10.1 と SAS を使い、ARIMA モデルによる予測値と干渉モデルと伝達関数(分布ラグ)モデルを使い時系列の予測値を与えた．

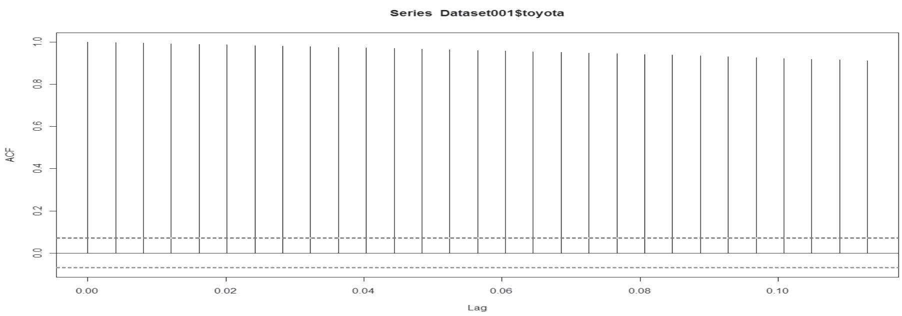
キーワード； R-2.10.1 , 伝達関数, 干渉モデル, ARIMA, AIC, クロス相関関数

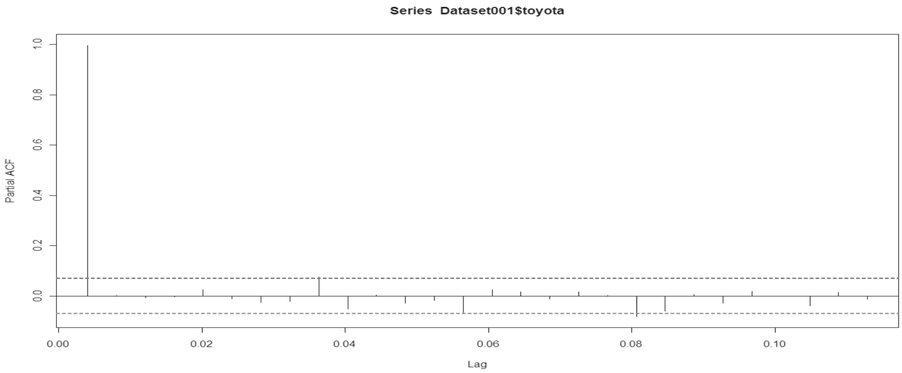
§―１　時系列のプロットと R－2.10.1 による予測

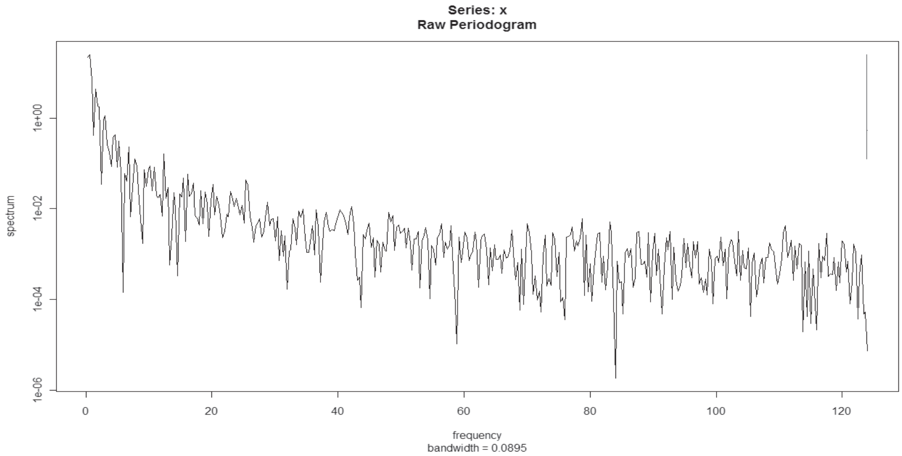
R―2.10.1 の package　Rcmdr-Plulag.In の TS-model において Run Multiple ARMA と Run Multiple GARCH の機能が新たに追加され多変量解析の機能がアップされた．
時系列データ｛y_t∈T｝をＲ－MySQL によって Yahoo のデータベース(NYSE)から取得する．
時系列データ　｛y_t∈T｝の特徴を把握するためにプロットし、定常性を確認する．非定常過程であれば、１次の階差∇=(1-B)、または２次階差∇^2=(1-B)^2 を取り、自己相関関数またはコレログラムやＤＦ検定（単位根）で確認する．まず干渉モデルのステップ関数を含まないARIMA(12, 1, 0), ARIMA(12, 1, 1)とARIMA(24, 1, 0)によって予測値を比較する．

時系列$\{y_t \in T\}$の自己共分散関数と偏自己相関関数のプロットのグラフは以下である.









summary(arma(y))

> print(forecast(fit, h=6))

以下はR-2.10.1による　ARIMA(12, 1, 0)モデルの予測値 (h=6) である.

予測標本	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2010.14516129032	34.18068	33.05311	35.30826	32.45621	35.90516
2010.14919354839	34.32548	32.73863	35.91232	31.89861	36.75235
2010.15322580645	34.60547	32.66404	36.54691	31.63630	37.57464
2010.15725806452	34.57247	32.32683	36.81811	31.13806	38.00688
2010.16129032258	34.45219	31.96378	36.94061	30.64649	38.25789

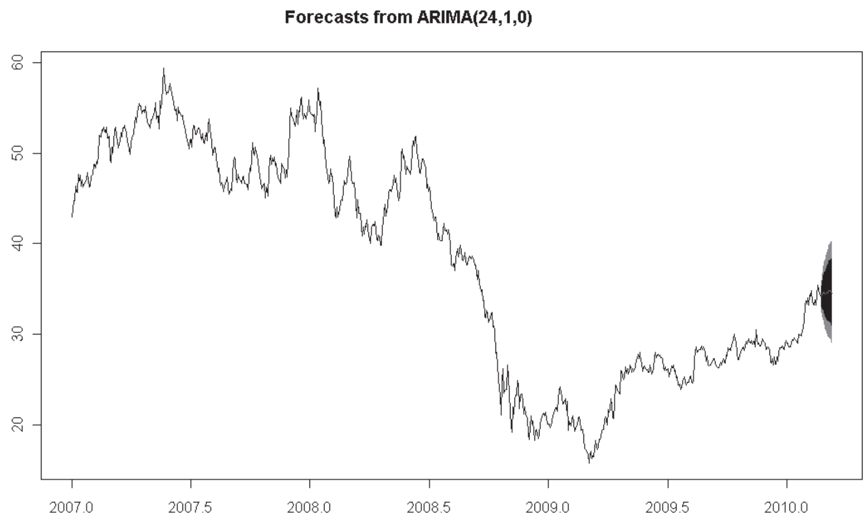
ここで　2010.14516129032 は 780 番目の時系列、以下同様に 2010.14919354839 は 781 番目の時系列である．80%と 95%の予測信頼区間である．予測値の精度は標本誤差測定基準として、MSE(平均２乗誤差)、MPE(平均パーセント誤差),MAPE(平均絶対パーセント誤差)、RMSE

等があり、値の小さい方の予測値が選択される. 赤池情報量基準 AIC は $\ln(\hat{\sigma}_a^2) + r\dfrac{2}{n} + cons.$

である.　$\hat{\sigma}_a^2$ は white noise の分散の最尤推定値.　パラメーターの個数は r=p+q+1　である.

BIC は次の量である.　n;標本数

$$BIC = \ln(\hat{\sigma}_a^2) + r\frac{\ln(n)}{n}.$$



> ArimaModel.1

Series: Dataset0010$y

ARIMA(12,1,1)

Call: arima(x = Dataset0010$y, order = c(12, 1, 1), seasonal = list(order = c(0, 0, 0), period = 248), include.mean = 1)

Coefficients:

	ar1	ar2	ar3	ar4	ar5	ar6	ar7	ar8	ar9	ar10	ar11	ar12	ma1
	0.4348	0.0038	0.0016	-0.0519	0.0332	0.0220	-0.0048	-0.0731	0.0748	-0.0138	0.0142	0.0289	-0.4416
s.e.	0.4465	0.0391	0.0390	0.0391	0.0455	0.0392	0.0410	0.0393	0.0507	0.0439	0.0395	0.0396	0.4458

sigma^2 estimated as 0.7874:　log likelihood = -1011.02

AIC = 2050.04　AICc = 2050.59　BIC = 2115.23

> predar3(ArimaModel.1, fore1=6)

	予測値	lower	upper
1	34.32866	32.58940	36.06791
2	34.38549	31.93415	36.83683
3	34.44285	31.44319	37.44251
4	34.37877	30.91488	37.84266
5	34.27694	30.44296	38.11093
6	34.33219	30.15298	38.51140

Series: y

ARIMA(12, 1, 0) の出力結果は以下である。

Call: arima(y, order = c(12, 1, 0))

Coefficients:

	ar1	ar2	ar3	ar4	ar5	ar6	ar7	ar8	ar9	ar10	ar11	ar12
	-0.0072	0.0007	0.0079	-0.0532	0.0076	0.0249	0.0098	-0.0690	0.0421	0.0066	0.0125	0.0232
s.e.	0.0358	0.0358	0.0358	0.0359	0.0358	0.0358	0.0359	0.0359	0.0360	0.0361	0.0360	0.0361

sigma^2 estimated as 0.7883: log likelihood = -1012.72

AIC = 2051.44 AICc = 2051.92 BIC = 2112

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE
-0.01101189	0.88727380	0.67034878	-0.06630615	1.94192744	0.99689155

```
y <- ts(y$Close, start=c(2007, 1), frequency=248)
fit <- arima(y, order=c(24, 1 ,0))   #ARIMA(24, 1, 0)
plot(forecast(fit, h=12))
summary(fit)
```

Series: y

ARIMA(24,1,0)

Call: arima(y, order = c(24, 1, 0))

Coefficients:

ar1 ar2 ar3 ar4 ar5 ar6 ar7 ar8 ar9 ar10 ar11 ar12

ar13 ar14 ar15 ar16 ar17 ar18 ar19 ar20 ar21 ar22 r23 ar24

-0.0098 0.0017 0.0090 -0.0500 0.0175 0.0202 0.0019 -0.0695 0.0414 0.0031 0.0142 0.0283

0.0621 -0.0243 -0.0275 0.0005 -0.0179 0.0021 0.0732 0.0593 0.0132 0.0288 -0.0379 0.0036

s.e. 0.0358 0.0358 0.0358 0.0359 0.0358 0.0357 0.0358 0.0358 0.0359 0.0360 0.0360

0.0360 0.0360 0.0360 0.0360 0.0360 0.0360 0.0360 0.0360 0.0361 0.0361 0.0360 0.0361 0.0361

sigma^2 estimated as 0.7741: log likelihood = -1005.84

AIC = 2061.67 AICc = 2063.4 BIC = 2178.12

ARIMAモデルによる予測値は過去の時系列の加重ウエイトの和として表現されるから、モデルの推定方法として、ユール・ウォーカー(YW)及び最尤法(ML)による予測値より最適である．過去のリーマンショックによる影響もモデルに含めることができる．ARIMAモデル $\varphi(B)y_t=\theta(B)a_t$ の予測値は加重ウエイト平均、$\bar{y}_{t+l-1}(\pi)=\sum_{j=1}^{\infty}\pi_j y_{t+l-j}$ である．ここで $\sum_{i=1}\pi_i=1$.

§—２　SASによる干渉モデルと伝達関数(分布ラグ)モデル

以下干渉モデルと伝達関数(分布ラグ)モデルを使い解析すると、干渉モデルよりもモデルの適合度, AICの基準やモデルの識別や推定方法を考えると分布ラグ予測値が選択される.

変数名 = y

系列の階差	1
系列の平均	−0.01112
標準偏差	0.892817
オブザベーション数	778
階差で取り除かれたオブザベーション	1

自己相関係数

ラグ	共分散	相関係数	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1	標準誤差
0	0.797122	1.00000	| |********************|	0
1	−0.0072504	−.00910	| .|. |	0.035852
2	−0.0006273	−.00079	| .|. |	0.035855
3	0.0028787	0.00361	| .|. |	0.035855
4	−0.039295	−.04930	| *|. |	0.035855
5	0.0070781	0.00888	| .|. |	0.035942
6	0.021238	0.02664	| .|* |	0.035945
7	0.0051234	0.00643	| .|. |	0.035970
8	−0.054994	−.06899	| *|. |	0.035972
9	0.034299	0.04303	| .|* |	0.036142
10	0.0014749	0.00185	| .|. |	0.036207
11	0.011900	0.01493	| .|. |	0.036207
12	0.025488	0.03197	| .|* |	0.036215
13	0.044722	0.05610	| .|* |	0.036252
14	−0.023943	−.03004	| *|. |	0.036363
15	−0.021547	−.02703	| *|. |	0.036395

16	0.0013358	0.00168	| .|. |	0.036421
17	−0.023463	−.02943	| *|. |	0.036421
18	0.0069524	0.00872	| .|. |	0.036451
19	0.061636	0.07732	| .|** |	0.036454
20	0.039065	0.04901	| .|* |	0.036664
21	−0.014750	−.01850	| .|. |	0.036748
22	0.030866	0.03872	| .|* |	0.036760
23	−0.033722	−.04231	| *|. |	0.036813
24	−0.0008028	−.00101	| .|. |	0.036875

"." は 2 標準誤差を示します

ホワイトノイズの自己相関検証

ラグ	カイ 2 乗	自由度	Pr > ChiSq	自己相関係数					
6	2.60	6	0.8571	−0.009	−0.001	0.004	−0.049	0.009	0.027
12	8.83	12	0.7170	0.006	−0.069	0.043	0.002	0.015	0.032
18	13.38	18	0.7684	0.056	−0.030	−0.027	0.002	−0.029	0.009
24	23.00	24	0.5196	0.077	0.049	−0.019	0.039	−0.042	−0.001

変数 s の差分を取りました。

y と s の相関

系列の階差	2
入力変数の分散 =	0.002532
オブザベーション数	777
階差で取り除かれたオブザベーション	2

相互相関

ラグ	共分散	相関係数	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
−24	0.00022919	0.00507	| .|. |
−23	−0.0003104	−.00686	| .|. |
−22	0.00023959	0.00530	| .|. |
−21	−0.0001075	−.00238	| .|. |
−20	−0.0012271	−.02713	| *|. |
−19	−0.0006478	−.01432	| .|. |
−18	−0.0010875	−.02404	| .|. |
−17	−0.0020148	−.04454	| *|. |
−16	0.00027283	0.00603	| .|. |
−15	−0.0007878	−.01742	| .|. |

```
-14   -0.0020510   -.04534   |          *|.          |
-13   -0.0008271   -.01828   |          .|.          |
-12   -0.0005590   -.01236   |          .|.          |
-11   -0.0012902   -.02852   |          *|.          |
-10   -0.0015513   -.03429   |          *|.          |
 -9   -0.0021814   -.04822   |          *|.          |
 -8   -0.0024152   -.05339   |          *|.          |
 -7   0.00084522   0.01868   |          .|.          |
 -6   0.00070447   0.01557   |          .|.          |
 -5   -0.0016250   -.03592   |          *|.          |
 -4   -0.0006373   -.01409   |          .|.          |
 -3    0.0011733   0.02594   |          .|*          |
 -2   0.00085311   0.01886   |          .|.          |
 -1   -0.0018610   -.04114   |          *|.          |
  0   -0.0019504   -.04311   |          *|.          |
  1   -0.0002997   -.00662   |          .|.          |
  2   -0.0037616   -.08315   |         **|.          |
  3   -0.0031900   -.07052   |          *|.          |
  4   -0.0018407   -.04069   |          *|.          |
  5   -0.0034241   -.07569   |         **|.          |
  6   -0.0032500   -.07184   |          *|.          |
  7   -0.0022612   -.04999   |          *|.          |
  8   -0.0036622   -.08096   |         **|.          |
  9    0.0018165   0.04015   |          .|*          |
 10    0.0067095   0.14832   |          .|***        |
 11   -0.0018387   -.04064   |          *|.          |
 12   -0.0031246   -.06907   |          *|.          |
 13   0.00066003   0.01459   |          .|.          |
 14    0.0037240   0.08232   |          .|**         |
 15    0.0018860   0.04169   |          .|*          |
 16   -0.0041013   -.09066   |         **|.          |
 17   -0.0049530   -.10949   |         **|.          |
 18   -0.0045410   -.10038   |         **|.          |
 19   -0.0028898   -.06388   |          *|.          |
 20    0.0028378   0.06273   |          .|*          |
 21    0.0025548   0.05647   |          .|*          |
```

22	0.00089852	0.01986	| .|. |
23	0.0027645	0.06111	| .|* |
24	0.00080705	0.01784	| .|. |

"." は 2 標準誤差を示します

予備推定

初期自己回帰パラメータの推定値

	推定値
1	-0.00603
2	0.00156
3	0.00292
4	-0.05157
5	0.01009
6	0.02615
7	0.00709
8	-0.06999
9	0.04273
10	0.00524
11	0.01581
12	0.02410

ARIMA 推定最適化の要約

推定法	Conditional Least Squares
パラメータ推定値	14
停止基準	Maximum Relative Change in Estimates
反復ストップ値	0.001
基準値	0.000757
代替基準	Relative Change in Objective Function
代替基準値	2.119E-9
勾配の最大絶対値	0.014795
最終反復からの R2 乗変換	0.000038
目的変数	Sum of Squared Residuals
目的変数値	599.8167
Marquardt の Lambda 係数	1E-6
数値的導関数 摂動デルタ	0.001
反復	13

分散 推定値	0.79236
標準誤差 推定値	0.890146
AIC	2022.433
SBC	2087.5
残差の数	771

＊AIC と SBC には行列式の対数を含みません.

残差の自己相関検証

ラグ	カイ 2 乗	自由度	Pr > ChiSq	自己相関係数					
6	.	0	.	−0.001	−0.000	0.001	−0.002	0.006	−0.001
12	.	0	.	−0.004	−0.001	0.002	−0.004	0.002	0.004
18	4.40	6	0.6225	0.047	−0.032	−0.037	−0.014	−0.025	−0.001
24	11.29	12	0.5039	0.066	0.045	−0.018	0.019	−0.038	−0.011
30	18.08	18	0.4506	0.036	−0.032	−0.008	−0.052	0.056	−0.018
36	27.00	24	0.3043	0.049	−0.019	−0.053	−0.062	−0.020	0.035
42	28.92	30	0.5219	−0.028	−0.022	0.019	0.003	0.013	0.024
48	29.87	36	0.7545	0.004	0.010	0.002	−0.022	−0.013	0.020

偏自己相関係数

```
ラグ   相関係数    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
  1   -0.00087   |                   .|.                   |
  2   -0.00002   |                   .|.                   |
  3    0.00091   |                   .|.                   |
  4   -0.00152   |                   .|.                   |
  5    0.00588   |                   .|.                   |
  6   -0.00053   |                   .|.                   |
  7   -0.00423   |                   .|.                   |
  8   -0.00076   |                   .|.                   |
  9    0.00193   |                   .|.                   |
 10   -0.00419   |                   .|.                   |
 11    0.00186   |                   .|.                   |
 12    0.00441   |                   .|.                   |
 13    0.04734   |                   .|*                   |
 14   -0.03239   |                   *|.                   |
 15   -0.03741   |                   *|.                   |
 16   -0.01417   |                   .|.                   |
 17   -0.02473   |                   .|.                   |
```

18	-0.00181	| .|. |
19	0.06716	| .|* |
20	0.04651	| .|* |
21	-0.01786	| .|. |
22	0.01880	| .|. |
23	-0.03826	| *|. |
24	-0.01353	| .|. |

ARIMA プロシジャ

時系列変数 y のモデル

系列の階差　1

モデル内に平均項はありません.

自己回帰モデルの因子

Factor 1:　1 + 0.01601 B**(1) + 0.01264 B**(2) + 0.00874 B**(3) + 0.05919 B**(4) + 0.00009 B**(5) - 0.02043 B**(6) - 0.00136 B**(7) + 0.07091 B**(8) - 0.04018 B**(9) - 0.00099 B**(10) - 0.01012 B**(11) - 0.01596 B**(12)

入力番号 1

入力変数	s
シフト	1
系列の階差	2
総合的な回帰の因子	-0.84978

分母の因子

Factor 1:　1 - 0.83553 B**(2)

変数の予測値: y

Obs	予測値	標準誤差	95% 信頼限界	
780	34.3181	0.8901	32.5734	36.0628
781	34.3544	1.2488	31.9067	36.8020
782	34.4005	1.5191	31.4232	37.3778
783	34.3305	1.7444	30.9116	37.7494
784	34.2481	1.9212	30.4826	38.0137
785	34.3209	2.0838	30.2368	38.4051

ここで、干渉モデルの独立変数、2008 年9月16日以後をステップ(step)反応関数で、入力変数の値を S(ω)=1 とする.　$S_t(\omega) = 1$　if t ≧ 2008-9-16 . それ以外は $S_t(\omega) = 0$.
以下は伝達関数(分布ラグ)モデルとして、時系列の構造解析を行なう. 伝達関数のシステムの入力変数 S(ω)の条件付最小2乗法による推定は以下である.

	推定値	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	−0.0040637	0.03151	−0.13	0.8974	0	y	0
NUM1	−1.64594	0.86529	−1.90	0.0575		s	0
NUM1,1	−0.11406	0.86529	−0.13	0.8952		s	0
NUM1,2	0.36594	0.86529	0.42	0.6725		s	0
NUM1,3	2.57594	0.86529	2.98	0.0030		s	0
NUM1,4	−0.07406	0.86529	−0.09	0.9318		s	0
NUM1,5	1.52594	0.86529	1.76	0.0782		s	0
NUM1,6	1.15594	0.86529	1.34	0.1820		s	0
NUM1,7	1.39594	0.86529	1.61	0.1071		s	0
NUM1,8	0.38594	0.86529	0.45	0.6557		s	0
NUM1,9	2.48594	0.86529	2.87	0.0042		s	0
NUM1,10	−3.87406	0.86529	−4.48	<.0001		s	0
NUM1,11	−1.31406	0.86529	−1.52	0.1293		s	0
NUM1,12	2.76594	0.86529	3.20	0.0014		s	0

Constant Estimate −0.00406

Variance Estimate 0.747727

Std Error Estimate 0.864712

AIC 1964.995

SBC 2029.971

Number of Residuals 766

＊AIC and SBC do not include log determinant.

分布ラグモデルとしてモデルを推定し、モデルの切片と入力変数Sである.

Model for variable y

Estimated Intercept −0.00406

Period(s) of Differencing 1

Input Number 1

Input Variable s

Period(s) of Differencing 1

numerator Factors

Factor 1:　−1.6459 + 0.11406 B**(1) − 0.36594 B**(2) − 2.57594 B**(3) + 0.07406 B**(4) − 1.52594 B**(5) − 1.15594 B**(6) − 1.39594 B**(7) − 0.38594 B**(8) − 2.48594 B**(9) + 3.87406 B**(10) + 1.31406 B**(11) − 2.76594 B**(12)

変数yの予測値 y,

	予測値	標準誤差	95%	信頼区間
780	34.2559	0.8647	32.5611	35.9507
781	34.2519	1.2229	31.8551	36.6487
782	34.2478	1.4977	31.3123	37.1833
783	34.2437	1.7294	30.8541	37.6334
784	34.2397	1.9336	30.4500	38.0294
785	34.2356	2.1181	30.0842	38.3870

結論;ARIMA モデルと伝達関数(分布ラグ)モデルは時系列の構造変化に対応でき、予測の使いやすさとモデルの識別やモデル選択の基準 AIC において、干渉モデルよりも最適である. R においてODBC(Open DataBase Connectivity)のパッケージは CRAN から求められる. またRでよく使われ RMySQL も同様に求められる. R－2.10.1 と SAS とのデータのやり取りにはパッケージ foreign がインストールされている必要がある. ここでR－2.10.1 のパッケージすべて(2200 個)をパソコンに入れると反応のスピードは少し遅くなることを述べておきたい.

参考文献

Box, G. E.P., G. M. Jenkins , and Reinsel ,G., (2008), *Time Series Analysis, Forecasting and Control.* Wily , Fourth Edition NJ.

Brockwell , P. J. and Davis, R . A., (2002) . *Introduction to Time Series and Forecasting*, 2nd ,Springer ,New York.

Douglas C. Montgomery, Cheryl L .Jennings and Murat, Kulahci .,(2008), *Introduction to Time Series Analysis and Forecasting*. John Wily& Sons ,Inc , New Jersey.

K , Komura ., (2001). 予測のためのデータ科学 . 晃洋書房

John , M. Chambers., (2008), *Software for Data Analysis Programming with R ,* Springer .CA.

Phil , Spector .,(2008) ., *Data Manipulation with R.* Springer. New York.

Robert ,A . Muenchen. , (2009) , *R for SAS and SPSS Users*, Springer ,New York.

CRAN., (2009). , *R-2.10.1 Reference manuals*