Modélisation ARIMA
des dépenses de consommation des ménages français
de 1978 :1 à 2002 :4

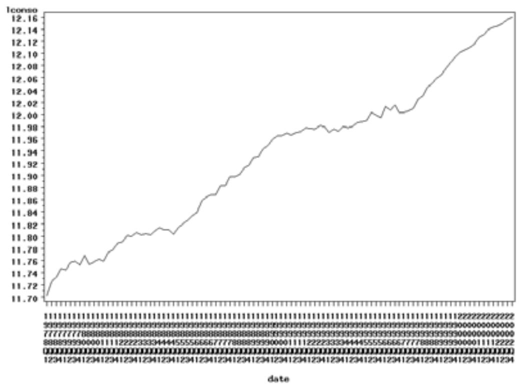
La table SAS `a.consoFra` contient la série de consommation, nommée ‘conso’ et la variable ‘date’, allant de 1978Q1 à 2002Q4, soit 100 observations.

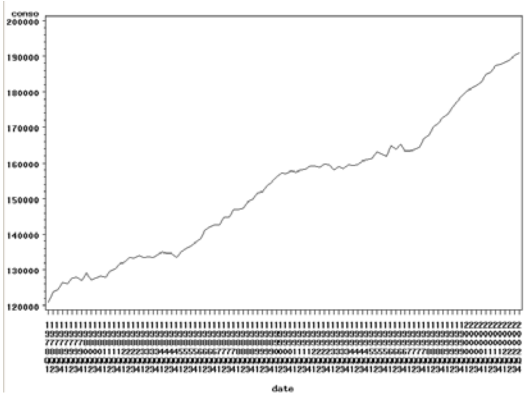
Création de la variable ‘lconso’ qui est le logarithme de la variable initiale, et de la variable ‘t’ qui correspond à une tendance :

```
data a.consoFra;
set a.consoFra;
lconso=log(conso);
t=_n_;
run;
```

Représentation graphique de la série :

```
proc gplot data=a.consoFra;
symbol i=join;
plot conso*date;
plot lconso*date;
run;
```





L’allure croissante de la série permet de conclure à la non stationnarité de la série.

Analyse des autocorrélations simples :

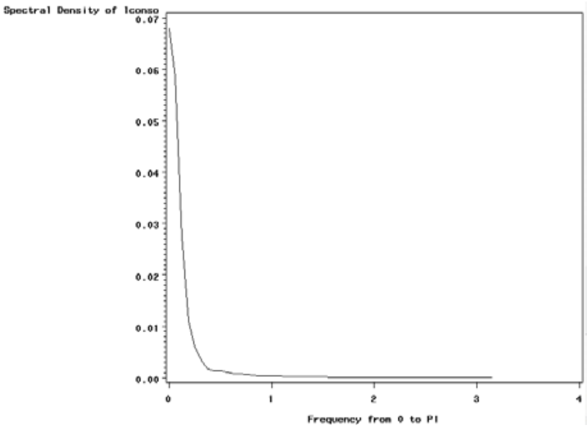
```
proc arima data=a.consoFra;
identify var=lconso;run;quit;
```

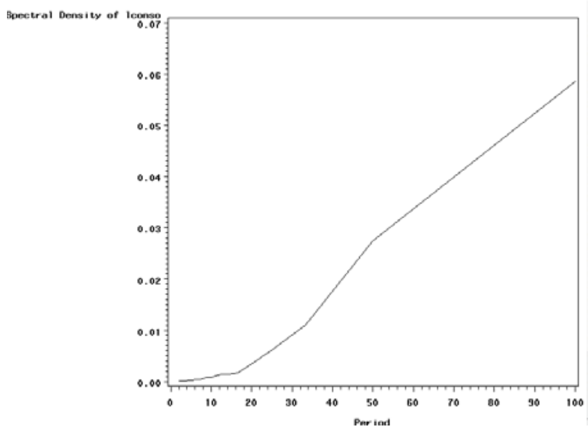
```
                              The ARIMA Procedure

                           Name of Variable = lconso

                     Mean of Working Series     11.92993
                     Standard Deviation         0.123459
                     Number of Observations          100

                                Autocorrelations

 Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

   0      0.015242       1.00000    |                    |********************|             0
   1      0.014687       0.96360    |                .   |******************* |      0.100000
   2      0.014183       0.93050    |             .      |******************* |      0.169028
   3      0.013673       0.89707    |           .        |******************  |      0.214213
   4      0.013184       0.86497    |          .         |*****************   |      0.248961
   5      0.012680       0.83192    |         .          |*****************   |      0.277390
   6      0.012182       0.79924    |        .           |****************    |      0.301308
   7      0.011698       0.76750    |       .            |***************     |      0.321811
   8      0.011184       0.73378    |      .             |***************     |      0.339622
   9      0.010722       0.70345    |      .             |**************      |      0.355123
  10      0.010217       0.67030    |     .              |************* .     |      0.368794
  11     0.0097160       0.63744    |     .              |************* .     |      0.380782
  12     0.0092107       0.60429    |    .               |************   .    |      0.391308
  13     0.0087053       0.57114    |    .               |***********    .    |      0.400531
  14     0.0082319       0.54008    |    .               |***********    .    |      0.408594
  15     0.0077707       0.50982    |   .                |**********      .   |      0.415671
  16     0.0073463       0.48197    |   .                |**********      .   |      0.421878
  17     0.0069189       0.45393    |   .                |*********       .   |      0.427349
  18     0.0065296       0.42839    |   .                |*********       .   |      0.432143
  19     0.0061323       0.40233    |   .                |********        .   |      0.436369
  20     0.0057757       0.37893    |  .                 |********         .  |      0.440063
  21     0.0054065       0.35470    |  .                 |*******          .  |      0.443314
  22     0.0050689       0.33256    |  .                 |*******          .  |      0.446143
  23     0.0047262       0.31007    |  .                 |******           .  |      0.448615
  24     0.0043928       0.28820    |  .                 |******           .  |      0.450753

                          "." marks two standard errors
```

La décroissance lente des autocorrélations simples indiquent la présence de non stationnarité (DS ou TS).

Calcul de la densité spectrale et représentation graphique :

```
PROC SPECTRA DATA=a.consoFra CENTER OUT=a.spectre P S ;
VAR lconso ;
WEIGHT PARZEN ;
RUN ;
PROC GPLOT DATA=a.spectre ;
PLOT S_01*FREQ;
PLOT S_01*PERIOD ;
RUN ;
quit;
```





Elle indique aussi la présence d'une composante tendancielle (non stationnaire).

<u>Caractérisation de la non-stationnarité :</u>

1) On régresse la série sur une tendance déterministe et on récupère le résidu, pour analyser (visuellement grâce aux autocorrélations) sa stationnarité :

```
proc reg data=a.consoFra;
model lconso=t;
output out=a.resTS r=res p=fit;
run;quit;
proc arima data=a.resTS;
identify var=res;
run;
```

```
                              The ARIMA Procedure

                            Name of Variable = res

                     Mean of Working Series      -25E-16
                     Standard Deviation         0.018078
                     Number of Observations          100

                                Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0    0.00032680        1.00000    |                    |********************|              0
  1    0.00030119        0.92165    |                .   |******************  |       0.100000
  2    0.00028733        0.87924    |             .      |******************  |       0.164282
  3    0.00025950        0.79409    |            .       |****************    |       0.206034
  4    0.00023139        0.70806    |           .        |**************      |       0.234652
  5    0.00020392        0.62401    |          .         |************        |       0.255124
  6    0.00016836        0.51519    |         .          |**********.         |       0.269956
  7    0.00014078        0.43079    |         .          |********* .         |       0.279615
  8    0.00010589        0.32403    |         .          |******    .         |       0.286175
  9    0.00008240        0.25215    |        .           |*****      .        |       0.289820
 10    0.00005114        0.15649    |        .           |***        .        |       0.292006
 11    0.00002387        0.07305    |        .           |*          .        |       0.292843
 12    -8.2523E-6        -.02525    |        .          *|           .        |       0.293026
 13    -0.0000349        -.10683    |        .         **|           .        |       0.293047
 14    -0.0000616        -.18846    |        .       ****|           .        |       0.293436
 15    -0.0000896        -.27411    |        .      *****|           .        |       0.294644
 16    -0.0001139        -.34846    |        .    *******|           .        |       0.297184
 17    -0.0001431        -.43779    |        .  *********|           .        |       0.301242
 18    -0.0001613        -.49370    |        . **********|           .        |       0.307538
 19    -0.0001842        -.56351    |       . ***********|            .       |       0.315364
 20    -0.0001957        -.59894    |       .************|            .       |       0.325277
 21    -0.0002099        -.64217    |       *************|            .       |       0.336125
 22    -0.0002142        -.65548    |      .*************|             .      |       0.348177
 23    -0.0002148        -.65730    |      .*************|             .      |       0.360306
 24    -0.0002122        -.64930    |     . *************|              .     |       0.372104

                          "." marks two standard errors
```

Le fait de retirer une tendance semble avoir créé une autocorrélation très forte.

2) On stationnarise par différenciation et on analyse (visuellement) sa stationnarité :

```
proc arima data=a.consoFra;
identify var=lconso(1);run;
```

```
                              The ARIMA Procedure

                          Name of Variable = lconso

              Period(s) of Differencing                          1
              Mean of Working Series                      0.004618
              Standard Deviation                          0.006644
              Number of Observations                            99
              Observation(s) eliminated by differencing          1

                                Autocorrelations

 Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

   0    0.00004414        1.00000    |                    |********************|             0
   1    -8.5637E-6        -.19402    |                ****|    .               |      0.100504
   2    0.00001488        0.33717    |               .    |*******             |      0.104218
   3    2.00275E-6        0.04537    |              .     |*    .              |      0.114709
   4    -2.0553E-6        -.04656    |              .    *|     .              |      0.114890
   5    9.22861E-6        0.20908    |              .     |****.              |      0.115080
   6    -7.5031E-6        -.16999    |              . ***|     .              |      0.118855
   7    5.30176E-6        0.12011    |              .     |**   .              |      0.121286
   8    -8.0188E-6        -.18167    |              .****|     .              |      0.122482
   9    4.72761E-6        0.10711    |              .     |**   .              |      0.125174
  10    -4.1891E-6        -.09491    |              .   **|     .              |      0.126097
  11    4.74577E-6        0.10752    |              .     |**   .              |      0.126816
  12     -6.115E-6        -.13854    |              . ***|     .              |      0.127733
  13    2.67743E-6        0.06066    |              .     |*    .              |      0.129242
  14    2.38289E-6        0.05399    |              .     |*    .              |      0.129530
  15    -9.5254E-7        -.02158    |              .     |     .              |      0.129757
  16     4.8777E-6        0.11051    |              .     |**   .              |      0.129793
  17    -8.8354E-6        -.20017    |              .****|     .              |      0.130740
  18    4.03438E-6        0.09140    |              .     |**   .              |      0.133800
  19    -8.8598E-6        -.20072    |              .****|     .              |      0.134429
  20    1.42341E-6        0.03225    |              .     |*    .              |      0.137423
  21    -8.0893E-6        -.18327    |              .****|     .              |      0.137499
  22    -4.9181E-6        -.11142    |             .    **|      .             |      0.139945
  23    -3.1577E-6        -.07154    |             .     *|      .             |      0.140838
  24    -7.3613E-6        -.16678    |             .  ***|      .             |      0.141205

                          "." marks two standard errors
```

La série différenciée semble stationnaire mais sa moyenne n'est pas forcément nulle. On peut alors penser que la série de consommation (en log) est un processus I(1) avec dérive. Pour confirmer cette intuition, on mène les tests de Dickey et Fuller (appelé DF par la suite).
Il faut pour cela choisir au préalable le nombre de retards à introduire dans la régression du test de DF pour blanchir les résidus.

Au vu de la représentation des autocorrélations partielles de la série différenciée, on peut penser à un AR(2). On peut alors mener le test de DF pour p=2.

```
                        Partial Autocorrelations

 Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

   1       -0.19402    |                ****|    .               |
   2        0.31124    |               .    |******              |
   3        0.17352    |               .    |***.                |
   4       -0.13897    |               .***|    .               |
   5        0.13004    |               .    |***.                |
   6       -0.07661    |               .  **|    .               |
   7       -0.02800    |               .   *|    .               |
   8       -0.14164    |               .***|    .               |
   9        0.09850    |               .    |**  .               |
  10       -0.02563    |               .   *|    .               |
  11        0.12164    |               .    |**  .               |
  12       -0.16314    |               .***|    .               |
  13        0.04697    |               .    |*   .               |
  14        0.09050    |               .    |**  .               |
  15        0.05535    |               .    |*   .               |
  16       -0.04770    |               .   *|    .               |
```

```
proc arima data=a.consoFra;
identify var=lconso stationarity=(adf=(2)) ;
run;
```

Augmented Dickey-Fuller Unit Root Tests

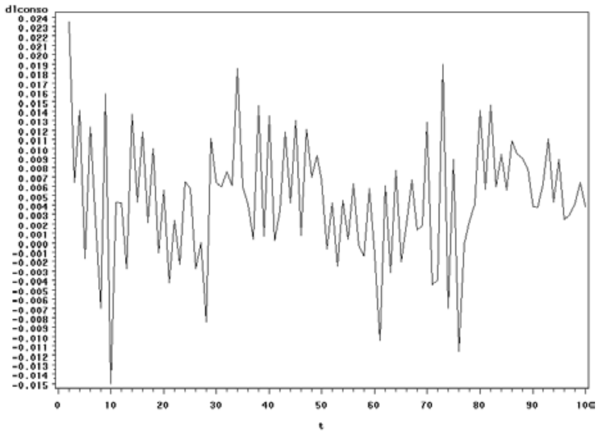
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	2	0.0353	0.6889	4.07	0.9999		
Single Mean	2	0.3030	0.9692	0.52	0.9865	8.33	0.0010
Trend	2	-7.5634	0.6041	-1.78	0.7097	1.89	0.8000

La p-value associée au test de DF dans le modèle M3 (Trend) (Pr<Tau=0.7097) conduit à accepter l'hypothèse nulle de racine unitaire. On teste alors conjointement la racine unitaire et la nullité du coefficient de la tendance par le test F3 afin de s'assurer que le modèle M3 était pertinent pour tester la racine unitaire. La pvalue de ce test (Pr>F=0.8) conduit à accepter l'hypothèse nulle de nullité jointe. Il faut donc refaire le test de DF dans le modèle M2 (Single Mean) : la pvalue du test DF (Pr<Tau=0.9865) conduit à accepter l'hypothèse nulle de racine unitaire. On teste alors conjointement la racine unitaire et la nullité du coefficient de la constante par le test F2 afin de s'assurer que le modèle M2 était pertinent pour tester la racine unitaire. La pvalue de ce test (Pr>F=0.001) conduit à rejeter l'hypothèse nulle de nullité jointe. Ainsi, on peut conclure que la consommation est I(1) avec dérive, c'est-à-dire qu'il y a une racine unitaire avec une constante significative pour sa différence première (soit son taux de croissance), ce qui signifie aussi que la consommation est non stationnaire de type stochastique (DS) et donc que son taux de croissance est stationnaire.

<u>Modélisation ARMA de la partie stationnaire, c'est-à-dire de (1-L)lconso :</u>

Création de la variable et représentation graphique :

```
data a.dlconso;set a.consoFra;
dlconso=dif(lconso);
run;
proc gplot data=a.dlconso;
plot dlconso*t ;
run;
```



Identification du modèle ARMA(p,q):

```
proc arima data=a.consoFra;
identify var=lconso(1) nlag=12;
run;
```

L'option `nlag=12` permet de ne calculer que 12 autocorrélations simples, inverses et partielles pour faciliter l'examen graphique.

```
                         Name of Variable = lconso

           Period(s) of Differencing                            1
           Mean of Working Series                        0.004618
           Standard Deviation                            0.006644
           Number of Observations                              99
           Observation(s) eliminated by differencing            1

                                  Autocorrelations

 Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

   0    0.00004414        1.00000    |                    |********************|             0
   1    -8.5637E-6        -.19402    |                ****|   .                |      0.100504
   2    0.00001488        0.33717    |                  . |*******             |      0.104218
   3    2.00275E-6        0.04537    |                 .  |*    .              |      0.114709
   4    -2.0553E-6        -.04656    |                 . *|     .              |      0.114890
   5    9.22861E-6        0.20908    |                 .  |****.              |      0.115080
   6    -7.5031E-6        -.16999    |                 .***|     .              |      0.118855
   7    5.30176E-6        0.12011    |                 .  |**   .              |      0.121286
   8    -8.0188E-6        -.18167    |                .****|     .              |      0.122482
   9    4.72761E-6        0.10711    |                 .  |**   .              |      0.125174
  10    -4.1891E-6        -.09491    |                 . **|     .              |      0.126097
  11    4.74577E-6        0.10752    |                 .  |**   .              |      0.126816
  12     -6.115E-6        -.13854    |                 .***|     .              |      0.127733

                          "." marks two standard errors

                            Inverse Autocorrelations

                 Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

                   1        0.04447    |                  . |*   .               |
                   2       -0.28821    |               ******|    .               |
                   3       -0.11577    |                 . **|    .               |
                   4        0.10282    |                 .  |**  .               |
                   5       -0.14583    |                 .***|    .               |
                   6        0.04856    |                 .  |*   .               |
                   7        0.03743    |                 .  |*   .               |
                   8        0.15383    |                 .  |***.               |
                   9       -0.05552    |                 .  *|    .               |
                  10       -0.08668    |                 . **|    .               |
                  11       -0.06775    |                 .  *|    .               |
                  12        0.12992    |                 .  |***.               |
```

```
                         Partial Autocorrelations

 Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

   1       -0.19402     |               ****|   .               |
   2        0.31124     |               .   |******             |
   3        0.17352     |               .   |***.               |
   4       -0.13897     |               .***|   .               |
   5        0.13004     |               .   |***.               |
   6       -0.07661     |               . **|   .               |
   7       -0.02800     |               .  *|   .               |
   8       -0.14164     |               .***|   .               |
   9        0.09850     |               .   |** .               |
  10       -0.02563     |               .  *|   .               |
  11        0.12164     |               .   |** .               |
  12       -0.16314     |               .***|   .               |
```

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	--------Autocorrelations--------					
6	23.76	6	0.0006	-0.194	0.337	0.045	-0.047	0.209	-0.170
12	34.76	12	0.0005	0.120	-0.182	0.107	-0.095	0.108	-0.139

On peut hésiter entre un AR(2) et un MA(2).
Afin d'obtenir le critère BIC pour différents modèles ARMA estimés, il suffit de mettre l'option `minic.`

```
proc arima data=a.consoFra;
identify var=lconso(1) minic;run;
```

Minimum Information Criterion

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	-10.3489	-10.3044	-10.3574	-10.3307	-10.2863	-10.3114
AR 1	-10.3232	-10.3077	-10.3573	-10.3701	-10.332	-10.3462
AR 2	-10.415	-10.3815	-10.3355	-10.3536	-10.3207	-10.3274
AR 3	-10.4025	-10.394	-10.3577	-10.3139	-10.286	-10.2922
AR 4	-10.3618	-10.3613	-10.3237	-10.2778	-10.2407	-10.2467
AR 5	-10.3565	-10.3288	-10.2865	-10.2635	-10.2322	-10.2028

Error series model: AR(10)
Minimum Table Value: BIC(2,0) = -10.415

Le critère BIC nous conduit à choisir plutôt un AR(2).

L'estimation de l'AR(2) est obtenue par la commande :

```
estimate p=2;run;
```

Le modèle AR(2) est alors estimé sur la variable définie dans la commande `identify` qui a été exécuté en dernier (ici lconso(1)).

The ARIMA Procedure

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.0046922	0.0007645	6.14	<.0001	0
AR1,1	-0.13328	0.09697	-1.37	0.1725	1
AR1,2	0.31212	0.09701	3.22	0.0018	2

Constant Estimate	0.003853
Variance Estimate	0.00004
Std Error Estimate	0.006289
AIC	-719.747
SBC	-711.962
Number of Residuals	99

* AIC and SBC do not include log determinant.

Correlations of Parameter Estimates

Parameter	MU	AR1,1	AR1,2
MU	1.000	0.001	0.012
AR1,1	0.001	1.000	0.193
AR1,2	0.012	0.193	1.000

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	--------Autocorrelations--------					
6	7.32	4	0.1199	-0.054	0.075	0.084	-0.093	0.187	-0.101
12	11.54	10	0.3170	0.004	-0.112	0.027	0.016	0.053	-0.145
18	16.29	16	0.4329	0.039	0.085	0.041	0.067	-0.155	0.025
24	25.11	22	0.2920	-0.105	0.009	-0.143	-0.121	-0.045	-0.140

```
Model for variable lconso

Estimated Mean              0.004692
Period(s) of Differencing          1

Autoregressive Factors

Factor 1:  1 + 0.13328 B**(1) - 0.31212 B**(2)
```

Si on avait voulu estimer un MA(2) :

```
estimate q=2;run;
```

Pour estimer un ARMA(1,3):

```
estimate p=1 q=3;run;
```

Par défaut, la procédure ARIMA utilise les moindres carrés conditionnels pour estimer les paramètres. Si on veut utiliser la méthode du maximum de vraisemblance :

```
estimate p=2 method=ml;run;
```

ou la méthode des moindres carrés non conditionnels:

```
estimate p=2 method=uls;run;
```

Rajouter l'option `plot` permet d'avoir la représentation des autocorrélations simples, inverses et partielles des résidus du modèle estimé.

Pour obtenir les résidus estimés (pour faire ensuite des tests de validation) et l'ajustement de la série par le modèle que l'on vient d'estimer, il faut passer à la commande forecast :

```
forecast out=a.res lead=0;run;
```

L'option `lead=n` spécifie l'horizon auquel on construit la prévision. Ici `lead=0` signifie qu'on calcule l'ajustement et non la prévision.

Dans la table `a.res`, il y a 6 variables : `lconso` (série intiale), `forecast` (ajustement de la série si on a mis `lead=0`), `std` (l'écart type de la prévision), `L95` (limite inférieur de l'intervalle de confiance à 95%), `U95` (limite supérieure de l'intervalle de confiance à 95%) et `residual` (les résidus estimés).

Une autre option de la commande `forecast` est `back=n` qui spécifie le nombre `n` d'observations avant la fin de l'échantillon à partir duquel les prévisions doivent commencer. Par défaut `back=0`.