

TP Séries temporelles

Anne-Lise Caillat, Christophe Dutang

9 mars 2008

On s'intéresse dans ce TP à la série du PNB américain (son logarithme très précisément noté (X_t) dans la suite) sur la période 1954-1987 en dollars 1982. Les commandes infra affichent la courbe des PNB. On constate clairement une tendance, les PNB n'ont donc rien de stationnaire.

```
> library(tseries)
> data(USEconomic)
> GNP <- USEconomic[, 2]
> annee <- seq(1954, 1987.75, 0.25)
> plot(annee, GNP, main = "log(PNB) au cours du temps", t = "l",
+      col = "blue", xlab = "temps", ylab = "log PNB")
```

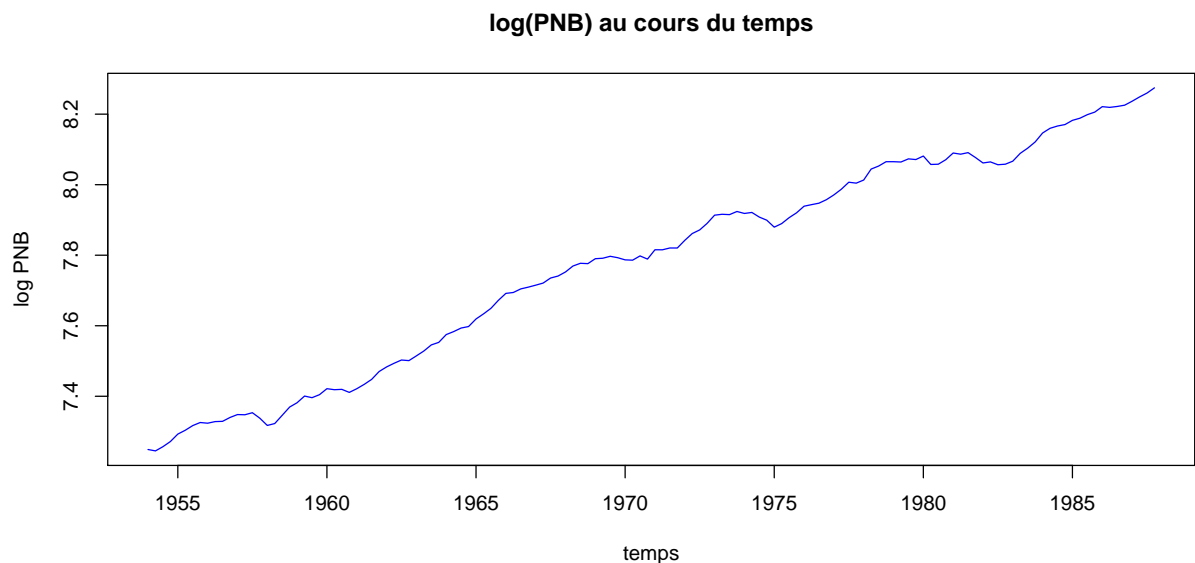


FIG. 1 – Logarithme du PNB américain

1 Stationarité de la série GNP

L'analyse de la fonction d'autocorrélation révèle la non stationarité de la série **GNP**. En effet, la fonction d'autocorrélation empirique $\hat{\rho}(h)$ n'est pas nulle pour tous les termes $h \geq 0$ (cf. figure 2).

```
> par(mfrow = c(1, 2))
> acf(GNP)
> pacf(GNP)
```

De plus on peut aussi rejeter la stationarité, car la série ne passe pas le test des points tournants*. Le test du portmanteau† rejette aussi la stationarité de la série.

```
> turnpoint.test(GNP)
```

Turning point test

t. p. statistic = -8.667377 , p-value = 0

*testant l'indépendance de l'échantillon X_1, \dots, X_T .

†testant la non-corrélation de l'échantillon.

```
> Box.test(GNP)
```

```
Box-Pierce test
```

```
data: GNP
```

```
X-squared = 129.9773, df = 1, p-value < 2.2e-16
```

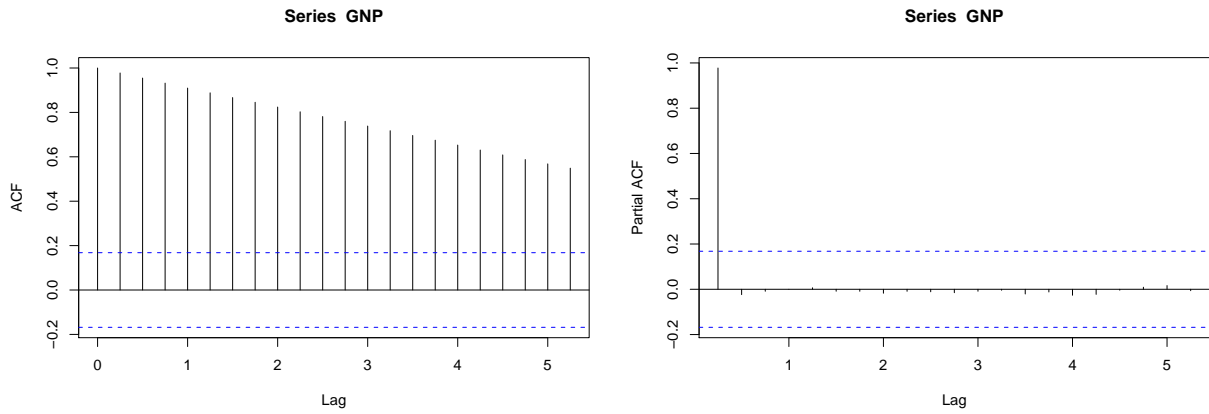


FIG. 2 – Fonctions d'autocorrélation empiriques

2 Etude de la série diff(GNP)

Nous allons maintenant travailler sur la série $(Y_t = X_t - X_{t-1})_{2 \leq t \leq T}$. Sur le tracé de Y_t , on constate que la série ne semble plus avoir une quelconque tendance (comportement très ératique).

```
> diffGNP <- diff(GNP)
> T <- length(annee)
> plot(annee[2:T], diffGNP, main = "diffGNP", t = "l", col = "blue",
+      xlab = "temps", ylab = "diff GNP")
```

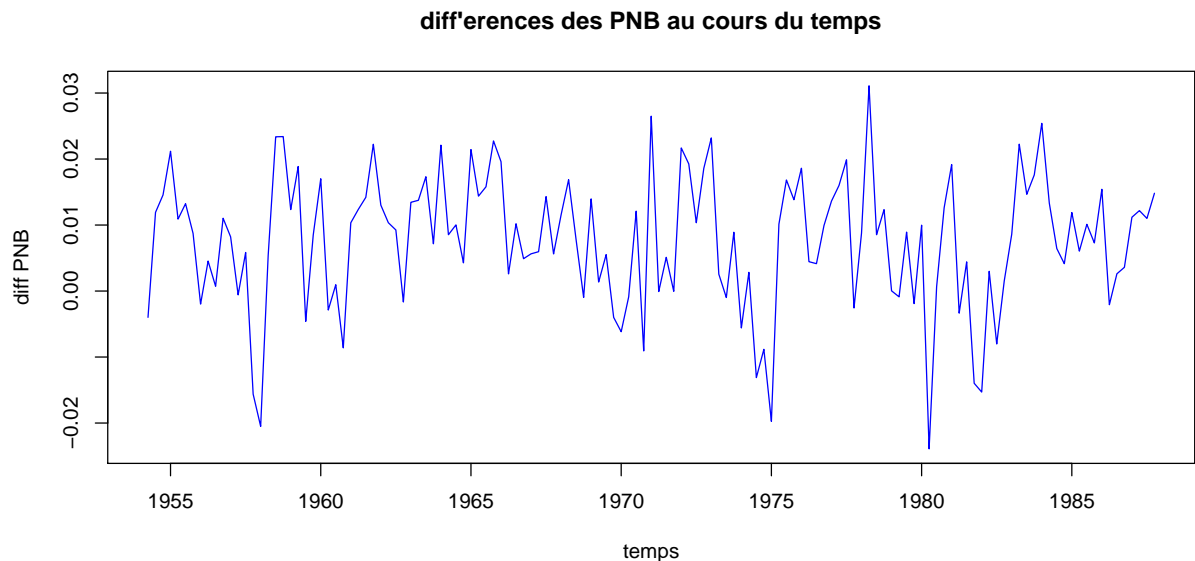


FIG. 3 – Série des différences

On peut aussi tracer la fonction d'autocorrélation empirique. Comme les fonctions d'autocorrélations (normale et partielle) décroissent très vite (cf. figure 4), on peut considérer que la série Y_t des différences est stationnaire.

```
> par(mfrow = c(1, 2))
> acf(diffGNP, lag.max = 30)
> pacf(diffGNP, lag.max = 30)
```

De plus, la série Y_t n'est pas centrée malgré une faible moyenne empirique de 0,007596586. Ainsi, nous avons besoin d'un terme constant dans la modélisation de Y . En effet, lorsqu'on teste l'hypothèse " \bar{Y}_T " avec le test de Student, on rejette l'hypothèse nulle.

```
> t.test(diffGNP)
```

One Sample t-test

```
data: diffGNP
t = 8.6739, df = 134, p-value = 1.223e-14
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.005864407 0.009328764
sample estimates:
```

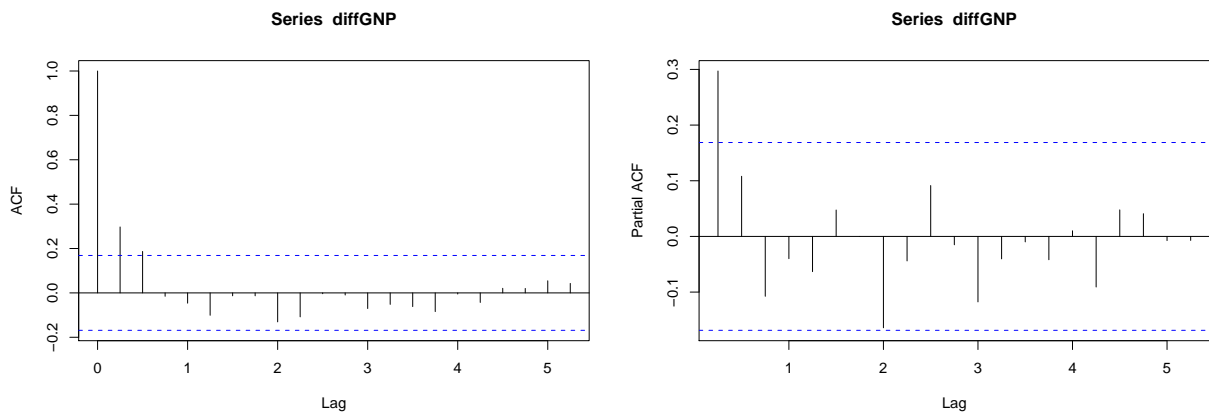


FIG. 4 – Fonctions d'autocorrélation empiriques

```
mean of x
0.007596586
```

Pour la série X_t des PNB, cela signifie qu'elle est du type marche aléatoire puisque $Y_t = X_t - X_{t-1}$ est stationnaire.

Au vu des autocorrélogrammes, un modèle ARMA(0,2) semble être le plus adapté. L'estimation par maximum de vraisemblance se fait avec la fonction `arima` de R.

```
> arima(diffGNP, c(0, 0, 2))
```

Call:

```
arima(x = diffGNP, order = c(0, 0, 2))
```

Coefficients:

	ma1	ma2	intercept
	0.2681	0.1976	0.0076
s.e.	0.0851	0.0790	0.0012

```
sigma^2 estimated as 9.178e-05: log likelihood = 435.86, aic = -863.73
```

On peut aussi choisir un modèle ARMA(7,2) ou ARMA(0,1). En effet, la fonction d'autocorrélation $\hat{\rho}(h)$ est quasiment en dehors de l'intervalle de confiance pour $h = 7$ et la fonction d'autocorrélation partielle $\hat{\alpha}(h)$ est quasiment dans l'intervalle de confiance pour $h = 2$.

```
> arima(diffGNP, c(7, 0, 2))
```

Call:

```
arima(x = diffGNP, order = c(7, 0, 2))
```

Coefficients:

	ar1	ar2	ar3	ar4	ar5	ar6	ar7	ma1	ma2
	0.0470	-0.7223	0.2004	0.0760	-0.1751	0.0187	-0.0205	0.2314	1.0000
s.e.	0.0899	0.0873	0.1052	0.1067	0.1051	0.0895	0.0891	0.0243	0.0415
	intercept								
	0.0076								
s.e.	0.0011								

```
sigma^2 estimated as 8.42e-05: log likelihood = 439.62, aic = -857.23
```

```
> arima(diffGNP, c(0, 0, 1))
```

Call:

```
arima(x = diffGNP, order = c(0, 0, 1))
```

Coefficients:

```
      ma1  intercept
      0.2278    0.0076
s.e.  0.0707    0.0010
```

```
sigma^2 estimated as 9.578e-05:  log likelihood = 433.03,  aic = -860.05
```

Pour comparer les modèles entre eux, on cherche à minimiser le critère AIC (Aikake's Information Criterion). Le modèle, dont le critère AIC est le plus bas (-863.73), est le modèle ARMA(0,2). Il semble donc être le mieux adapté aux données. Cependant, on peut aussi comparer ces 3 modèles en testant la stationarité de leurs résidus.

```
> Box.test((result02 <- arima(diffGNP, c(0, 0, 2)))$residual)
```

Box-Pierce test

```
data: (result02 <- arima(diffGNP, c(0, 0, 2)))$residual
X-squared = 0.0043, df = 1, p-value = 0.9474
```

```
> Box.test((result72 <- arima(diffGNP, c(7, 0, 2)))$residual)
```

Box-Pierce test

```
data: (result72 <- arima(diffGNP, c(7, 0, 2)))$residual
X-squared = 2e-04, df = 1, p-value = 0.9895
```

```
> Box.test((result01 <- arima(diffGNP, c(0, 0, 1)))$residual)
```

Box-Pierce test

```
data: (result01 <- arima(diffGNP, c(0, 0, 1)))$residual
X-squared = 0.2829, df = 1, p-value = 0.5948
```

On constate que le modèle ARMA(7,2) semble meilleur, puisque la p-valeur du test de Box Pierce est légèrement plus élevée. Ainsi, on ne peut pas rejeter le fait que les résidus sont non corrélés. Le tracé des autocorrélogrammes (cf. figure 5) confirme que les modèles ARMA(7,2) et ARMA(0,2) collent bien aux données.

```
> par(mfrow = c(1, 3))
```

```
> acf(result02$residual, main = "modele ARMA(0,2)")
```

```
> acf(result72$residual, main = "modele ARMA(7,2)")
```

```
> acf(result01$residual, main = "modele ARMA(0,1)")
```

Nous allons maintenant tester les performances des 3 modèles. Pour ce faire, nous enlevons les 10 derniers points de notre série Y_t , pour ensuite comparer nos 10 valeurs prédites avec nos 10 valeurs réelles.

```
> n <- 10
```

```
> index <- 1:(T - n - 1)
```

```
> res02 <- predict(arima(diffGNP[index], c(0, 0, 2)), n)
```

```
> res72 <- predict(arima(diffGNP[index], c(7, 0, 2)), n)
```

```
> res01 <- predict(arima(diffGNP[index], c(0, 0, 1)), n)
```

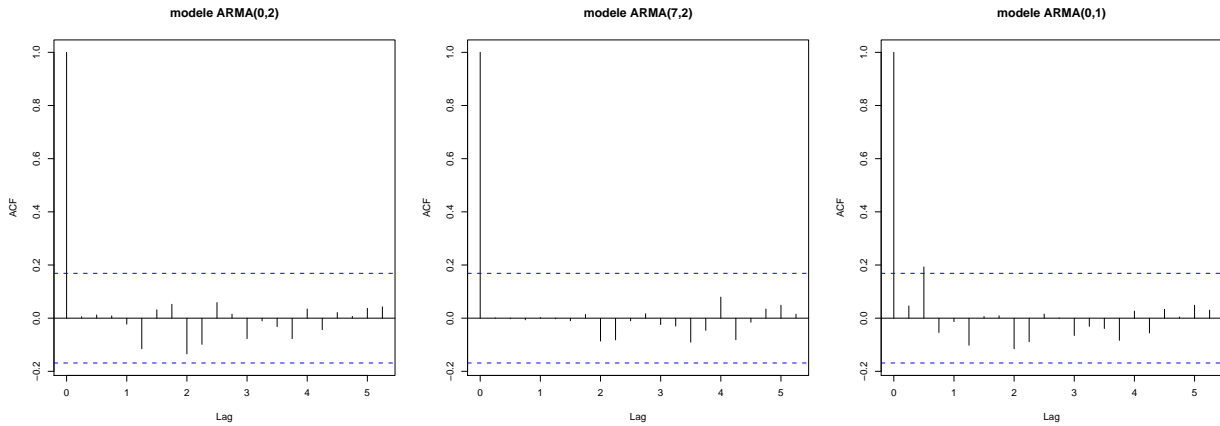


FIG. 5 – Autocorrélogrammes sur les résidus

```

> par(mfrow = c(1, 3))
> plot(annee[(T - 4 * n):T], diffGNP[(T - 4 * n):T - 1], main = "prevision ARMA(0,2)",
+      t = "l", col = "blue", xlab = "temps", ylab = "diff PNB")
> lines(annee[(T - n):T], c(diffGNP[T - n - 1], res02$pred))
> lines(annee[(T - n):T], c(diffGNP[T - n - 1], res02$pred) + c(0,
+      res02$se) * 1.96, lty = 2)
> lines(annee[(T - n):T], c(diffGNP[T - n - 1], res02$pred) - c(0,
+      res02$se) * 1.96, lty = 2)

> plot(annee[(T - 4 * n):T], diffGNP[(T - 4 * n):T - 1], main = "prevision ARMA(7,2)",
+      t = "l", col = "blue", xlab = "temps", ylab = "diff PNB")
> lines(annee[(T - n):T], c(diffGNP[T - n - 1], res72$pred), col = "red")
> lines(annee[(T - n):T], c(diffGNP[T - n - 1], res72$pred) + c(0,
+      res72$se) * 1.96, lty = 2, col = "red")
> lines(annee[(T - n):T], c(diffGNP[T - n - 1], res72$pred) - c(0,
+      res72$se) * 1.96, lty = 2, col = "red")

> plot(annee[(T - 4 * n):T], diffGNP[(T - 4 * n):T - 1], main = "prevision ARMA(0,1)",
+      t = "l", col = "blue", xlab = "temps", ylab = "diff PNB")
> lines(annee[(T - n):T], c(diffGNP[T - n - 1], res01$pred), col = "green")
> lines(annee[(T - n):T], c(diffGNP[T - n - 1], res01$pred) + c(0,
+      res01$se) * 1.96, lty = 2, col = "green")
> lines(annee[(T - n):T], c(diffGNP[T - n - 1], res01$pred) - c(0,
+      res01$se) * 1.96, lty = 2, col = "green")

```

Au vu des graphes (cf. figure 6), on en déduit que les prédictions du modèle ARMA(0,1) sont plus mauvaises que celles des deux autres modèles.

Les intervalles de confiance (à 95%) sur les valeurs prédites supposent que les résidus sont gaussiens (ce qui semblent être vrai pour le modèle ARMA(7,2) et ARMA(0,2), puisque on ne peut pas rejeter l'hypothèse de normalité).

```

> shapiro.test(arima(diffGNP, c(0, 0, 2))$residual)

```

Shapiro-Wilk normality test

```

data: arima(diffGNP, c(0, 0, 2))$residual
W = 0.9885, p-value = 0.323

```

```
> shapiro.test(arima(diffGNP, c(7, 0, 2))$residual)
```

Shapiro-Wilk normality test

```
data: arima(diffGNP, c(7, 0, 2))$residual  
W = 0.9893, p-value = 0.3885
```

```
> shapiro.test(arima(diffGNP, c(0, 0, 1))$residual)
```

Shapiro-Wilk normality test

```
data: arima(diffGNP, c(0, 0, 1))$residual  
W = 0.9776, p-value = 0.02492
```

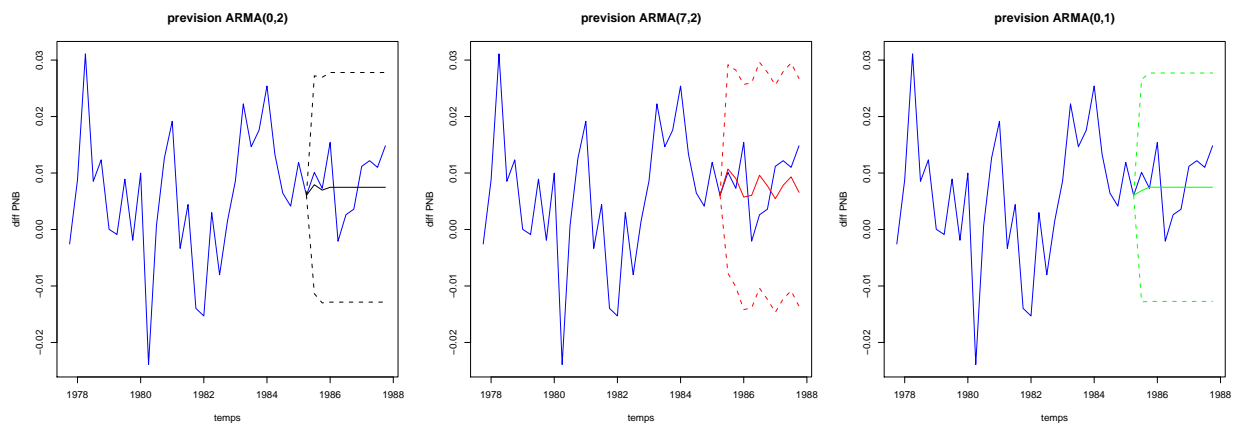


FIG. 6 – Prédictions

En conclusion de l'analyse de la série différenciée Y_t , on est amené à préférer le modèle ARMA(0,2), car il minimise le critère AIC sans perdre la stationarité des résidus.

3 Modèle ARIMA pour la série GNP

La série Y_t des différences de la série GNP nous semblait centrée, mais si on trace l'histogramme à classe de même effectif, on constate qu'elle n'est pas du tout symétrique.

```
> histo_eff(diffGNP)
```

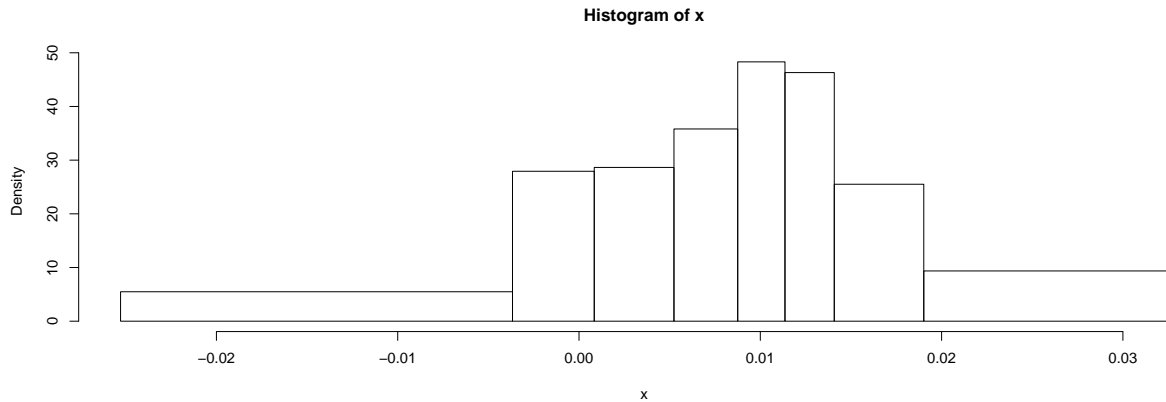


FIG. 7 – Histogramme de Y_t

Ainsi, on va différencier une deuxième fois la série GNP, on pose $Z_t = \Delta^2 X_t (= \Delta Y_t)$. On constate que la série est plus centrée et surtout plus symétrique.

```
> d2GNP <- diff(diffGNP)
> par(mfrow = c(1, 2))
> plot(annee[3:T], d2GNP, main = "Z au cours du temps", t = "l",
+      col = "blue", xlab = "temps", ylab = "Z")
> par(mfrow = c(1, 2))
> histo_eff(d2GNP)
```

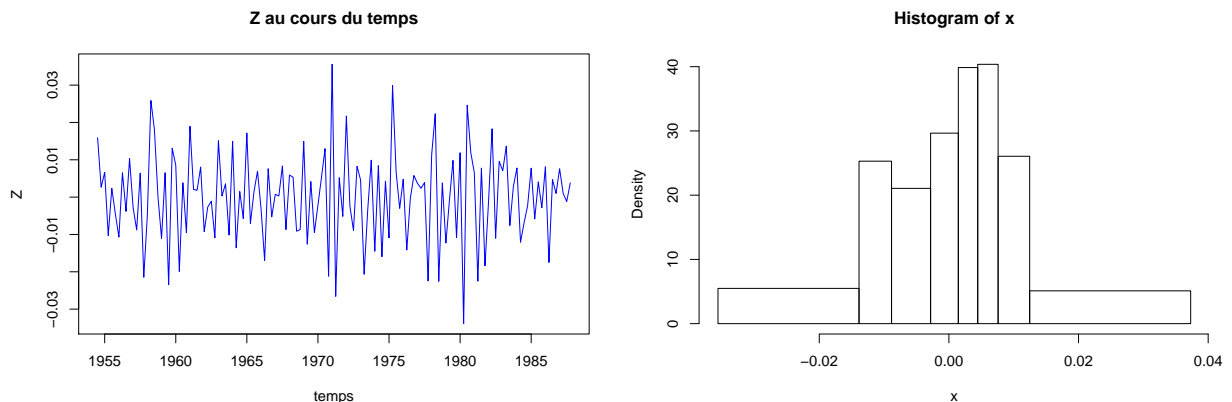


FIG. 8 – Diagnostique graphique de Z_t

Le test de Student révèle qu'on ne peut pas rejeter l'hypothèse que $\bar{Z}_T = 0$.

```

> t.test(d2GNP)

One Sample t-test

data:  d2GNP
t = 0.1348, df = 133, p-value = 0.893
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.001918837  0.002199523
sample estimates:
 mean of x
0.0001403431

```

Comme pour les modèles ARMA, nous allons maintenant trouver les coefficients p et q du modèles ARIMA, puis tester la stationarité des résidus et la performance de nos prédictions. Le tracé des autocorrélogrammes révèle que $q = 1$ et $p = 4$ ou 8.

```

> par(mfrow = c(1, 2))
> acf(d2GNP, lag.max = 30)
> pacf(d2GNP, lag.max = 30)

```

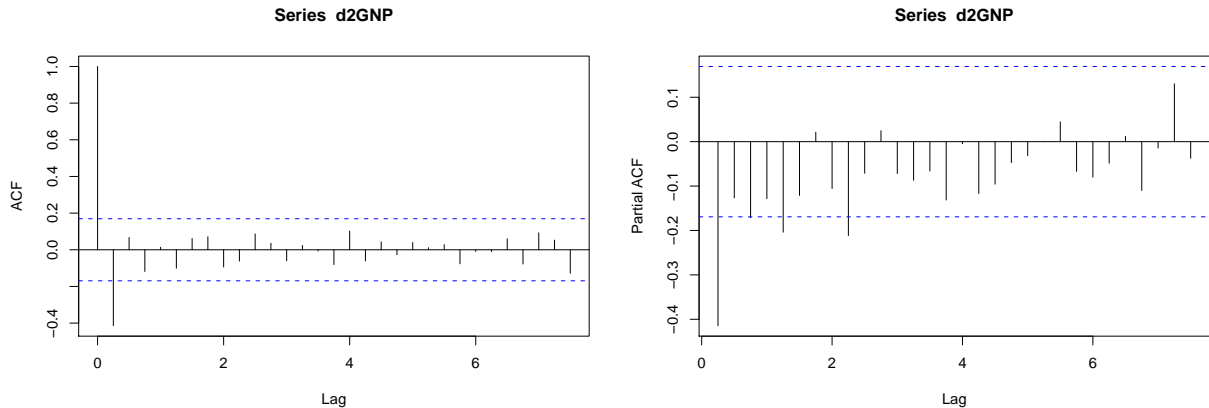


FIG. 9 – Autocorrélogrammes de Z_t

Nous allons donc tester deux modèles ARIMA(4,2,1) et ARIMA(8,2,1). On trouve comme coefficients :

```

> arima(GNP, c(4, 2, 1), inc = FALSE)

Call:
arima(x = GNP, order = c(4, 2, 1), include.mean = FALSE)

Coefficients:
      ar1      ar2      ar3      ar4      ma1
 0.2817  0.1464 -0.0933 -0.0357 -1.0000
s.e.  0.0868  0.0894  0.0891  0.0867  0.0233

sigma^2 estimated as 9.19e-05:  log likelihood = 430.43,  aic = -848.87

> arima(GNP, c(8, 2, 1), inc = FALSE)

```

```
Call:
arima(x = GNP, order = c(8, 2, 1), include.mean = FALSE)
```

```
Coefficients:
      ar1      ar2      ar3      ar4      ar5      ar6      ar7      ar8
 0.2786  0.1511 -0.0925 -0.0320 -0.0846  0.0730  0.0469 -0.1529
s.e.  0.0856  0.0886  0.0888  0.0895  0.0888  0.0886  0.0880  0.0853
      ma1
-1.0000
s.e.  0.0266
```

```
sigma^2 estimated as 8.887e-05:  log likelihood = 432.43,  aic = -844.87
```

Selon le critère AIC, le modèle ARIMA(4,2,1) semble mieux coller aux données. Testons la stationnarité et la normalité des résidus.

```
> Box.test((res421 <- arima(GNP, c(4, 2, 1), inc = FALSE))$residual)
```

```
Box-Pierce test
```

```
data: (res421 <- arima(GNP, c(4, 2, 1), inc = FALSE))$residual
X-squared = 0.0185, df = 1, p-value = 0.8917
```

```
> Box.test((res821 <- arima(GNP, c(8, 2, 1), inc = FALSE))$residual)
```

```
Box-Pierce test
```

```
data: (res821 <- arima(GNP, c(8, 2, 1), inc = FALSE))$residual
X-squared = 0.0233, df = 1, p-value = 0.8786
```

```
> shapiro.test(res421$residual)
```

```
Shapiro-Wilk normality test
```

```
data: res421$residual
W = 0.9904, p-value = 0.4781
```

```
> shapiro.test(res821$residual)
```

```
Shapiro-Wilk normality test
```

```
data: res821$residual
W = 0.9885, p-value = 0.3188
```

On en déduit qu'on ne peut pas rejeter la stationnarité et la normalité des résidus pour les modèles ARIMA(4,2,1) et ARIMA(8,2,1). De plus, le modèle ARIMA(4,2,1) apparaît meilleur.

Enfin le tracé des fonctions d'autocorrélogrammes (cf. figure 10) ne montre pas de corrélation pour les résidus. Au final, le modèle ARIMA(4,2,1) semble meilleur que l'ARIMA(8,2,1) sur ces données.

```
> par(mfrow = c(1, 2))
> acf(res421$residual)
> acf(res821$residual)
```

Enfin testons la performance des prédictions.

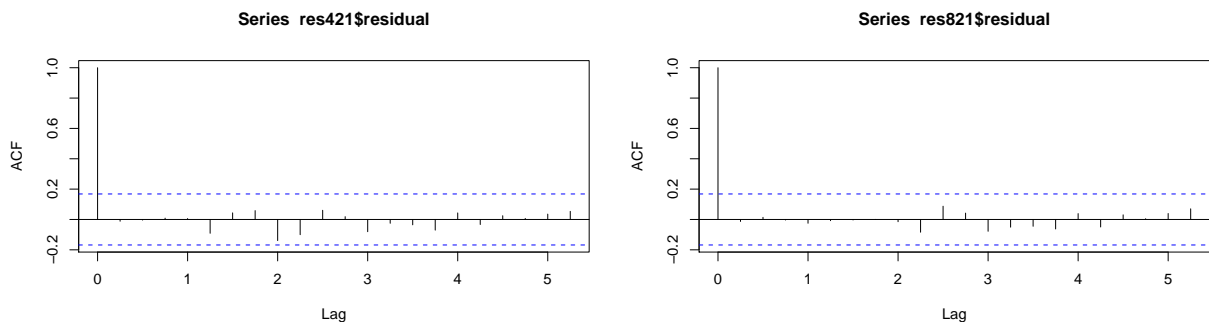


FIG. 10 – Autocorrélogrammes sur les résidus

```

> index <- 1:(T - n)
> res421 <- predict(arima(GNP[index], c(4, 2, 1), inc = FALSE),
+   n)
> res821 <- predict(arima(GNP[index], c(8, 2, 1), inc = FALSE),
+   n)
> par(mfrow = c(1, 2))
> plot(annee[(T - 4 * n):T], GNP[(T - 4 * n):T], main = "prevision ARIMA(4,2,1)",
+   t = "l", col = "blue", xlab = "temps", ylab = "diff PNB")
> lines(annee[(T - n):T], c(GNP[T - n], res421$pred))
> lines(annee[(T - n):T], c(GNP[T - n], res421$pred) + c(0, res421$se) *
+   1.96, lty = 2)
> lines(annee[(T - n):T], c(GNP[T - n], res421$pred) - c(0, res421$se) *
+   1.96, lty = 2)

> plot(annee[(T - 4 * n):T], GNP[(T - 4 * n):T], main = "prevision ARMA(8,2,1)",
+   t = "l", col = "blue", xlab = "temps", ylab = "diff PNB")
> lines(annee[(T - n):T], c(GNP[T - n], res821$pred), col = "red")
> lines(annee[(T - n):T], c(GNP[T - n], res821$pred) + c(0, res821$se) *
+   1.96, lty = 2, col = "red")
> lines(annee[(T - n):T], c(GNP[T - n], res821$pred) - c(0, res821$se) *
+   1.96, lty = 2, col = "red")

```

Au vu des prévisions, on conclut que le modèle ARIMA(4,2,1) est mieux adapté aux données.

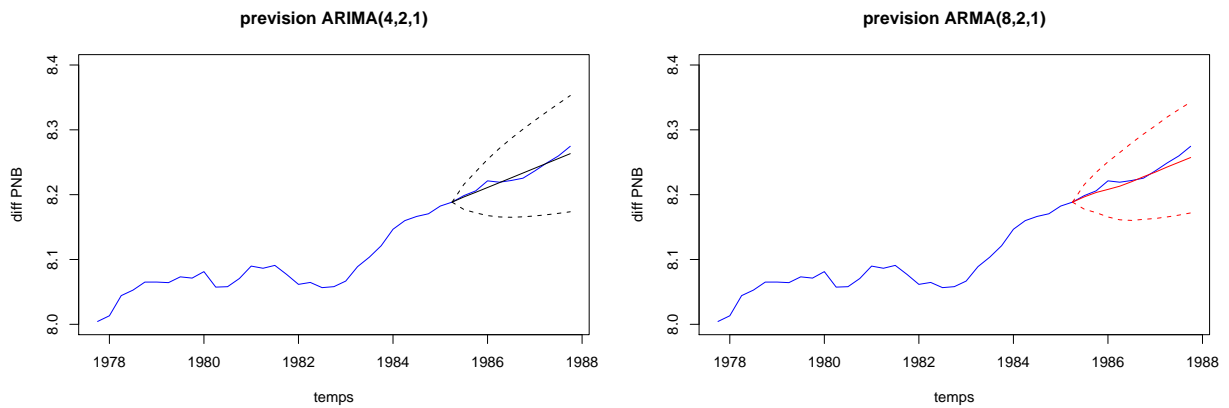


FIG. 11 – Prédications

Le tableau infra récapitule l'ensemble des résultats des différents modèles présentés dans ce compte rendu.

Critères	ARMA(0,2)	ARIMA(4,2,1)	ARMA(0,1)	ARMA(7,2)	ARIMA(8,2,1)
AIC	-863,73	-848,87	-860,05	-857,23	-844,87
Vraisemblance	435,86	430,43	433,03	439,62	432,43
p-valeur Box Pierce	0,9474	0,8917	0,5948	0,9895	0,8786
p-valeur Pts Tourn.	0,1320	0,0477	0,055	0,2731	0,0169
p-valeur Shapiro	0,323	0,4781	0,2492	0,3885	0,3188

TAB. 1 – Bilan : critère AIC, log vraisemblance et tests sur les résidus

En conclusion, le modèle ARMA(0,2) sur la série différenciée des GNP est meilleur compte tenu des critères choisis que le modèle ARIMA(4,2,1) sur la série GNP.