

Table des matières

Numéro 1	2
a)	2
b)	3
c)	3
d)	4
e)	5
f)	7
Annexe	8
Tableaux et figures	8
Code informatique	8

Le présent rapport se veut être l'analyse de différentes séries chronologiques. Plus précisément, ce rapport a pour but de trouver le modèle qui s'ajuste le plus exactement à nos échantillons de données. Le travail en bref sera alors de séparer nos bases de données en un échantillon *entraînement* qui servira à trouver un modèle et en un échantillon *test* qui servira à valider la précision du modèle retenu. Le travail est fait à partir de deux bases de données pour un total de cinq variables à modéliser. La première base de données, traitée au numéro 1, contient les données mensuelles du taux de change du dollar américain par rapport à la monnaie Euro de janvier 1999 à décembre 2016. La seconde base de données contient une série de statistiques de la *SAAQ*, soit le nombre d'accidents automobiles avec dommages corporels, le nombre de personnes accidentés, le nombre de demandes d'indemnités et le coût total de l'indemnisation (en dollar constants 2015) pour les années 1978 à 2015 inclusivement. Ces données seront traitées au numéro 2. Afin d'alléger la lecture du présent rapport, les figures et tableaux furent placés en annexe.

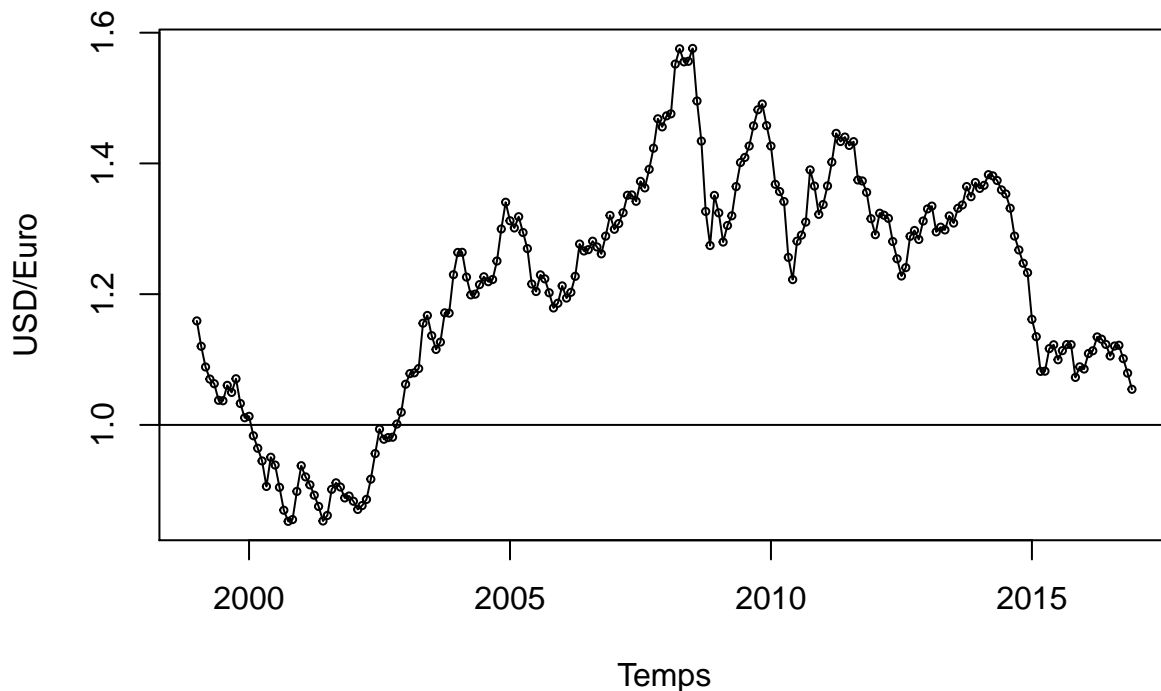
Numéro 1

a)

1

Tout d'abord, nous observons, à la *fig. 1* en annexe, la série chronologique du taux de change américain/européen depuis janvier 1999 jusqu'à décembre 2016. Les données ont été collectées de façon quotidienne pour ensuite être transformées en taux mensuelles.

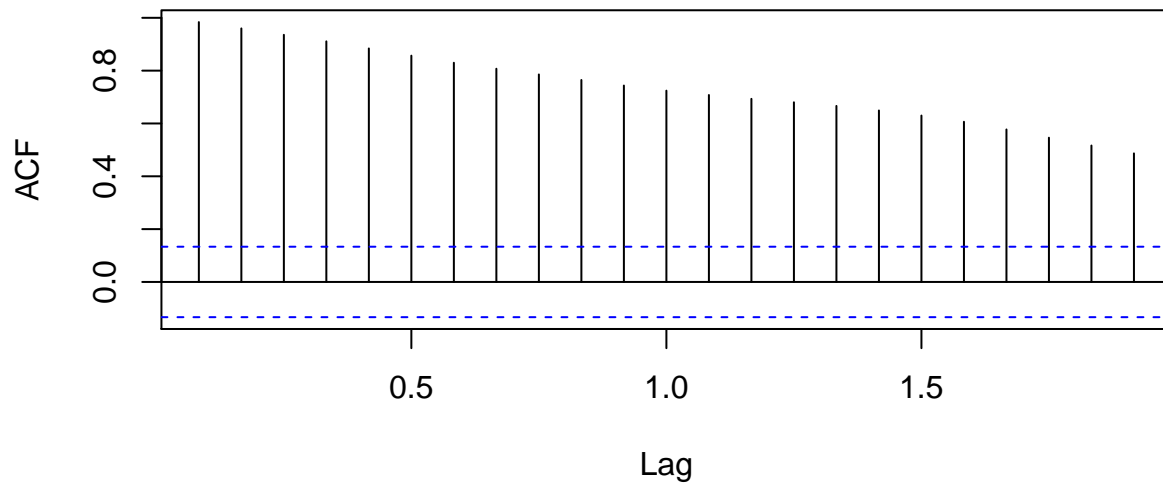
Figure 1: taux de change US/Euro



b)

Grâce à une première analyse, on remarque sur le tableau de la fonction d'autocorrélation échantillonnale (ACF) qu'il y a une forte présence d'autocorrélation et que celle-ci diminue lentement plus le lag augmente (voir figure 2). On peut donc déduire que le processus est non-stationnaire. Il n'est pas nécessaire de poursuivre notre analyse avec le graphique de la fonction d'autocorrélation partielle (PACF) puisque nous devons effectuer une différentiation sur le modèle pour créer l'effet de stationnarité.

Figure 2: fonction d'autocorrélation

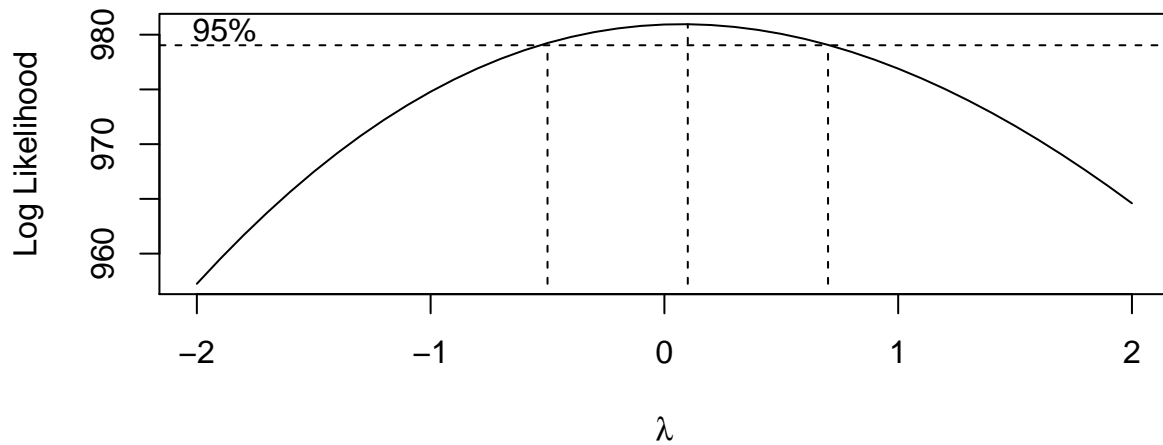


c)

Étant donné que les données sont positives, Il est possible d'utiliser Box-Cox afin de transformer notre processus. En ce faisant, la variance diminuera. On rappelle que la famille des fonctions de puissance est définie de la façon suivante:

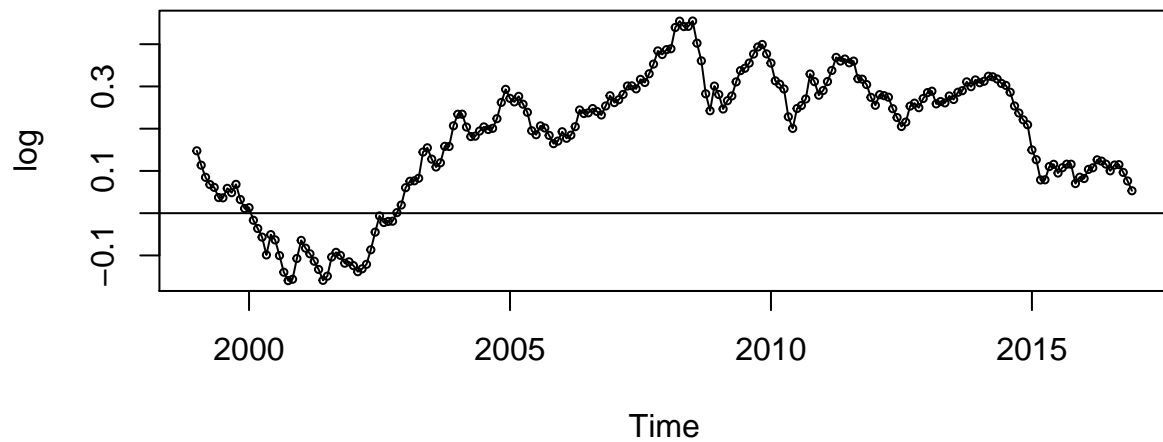
$$g(x) = \frac{x^\lambda - 1}{\lambda} \times 1_{\{\lambda \neq 0\}} + \ln(x) \times 1_{\{\lambda = 0\}}$$

λ est donc déterminé en maximisant la fonction de log-vraisemblance de nos données que voici:



On constate que $\lambda = 0.1$ semble être l'estimé MLE situé au centre de l'intervalle de confiance 95%, soit $] - 0.5, 0.7[$. Puisque $\lambda = 0$ est dans notre IC, cette valeur du paramètre peut également être à considérer. Donc, on utilise la transformé logarithmique pour notre modèle (voir figure 3).

Figure 3: logarithme du taux

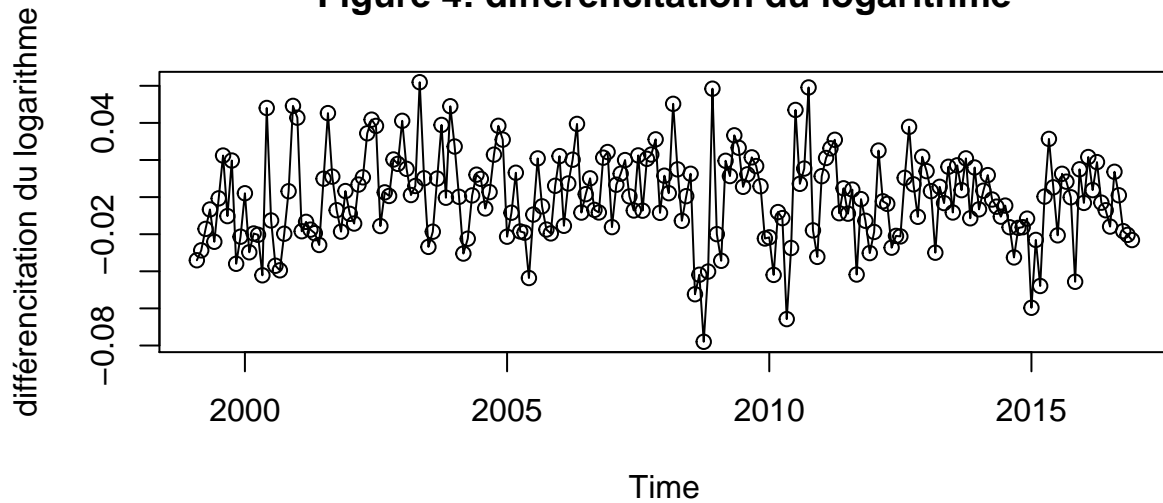


Cependant, en analysant le graphique de la fonction d'autocorrélation du logarithme de notre série, on remarque qu'il n'y a pas de variation notable. Le modèle est toujours non stationnaire. PLUGGER LA THÉORIE DE YAN SUR LE MFE PI LES SHIT DE RENDEMENT.

d)

En revanche, si on effectue une première différentiation du log de notre série, on obtient la série de la figure 4 qui semble plus stable.

Figure 4: différenciation du logarithme



On vérifie alors si une première différenciation amène la stationnarité à notre processus en utilisant le test augmenté de Dickey-Fuller (ADF). En se concentrant sur la partie autorégressive, on obtient un processus AR(1) associé à notre modèle de la première différenciation du logarithme. Par la suite, on applique le test ADF à 95% et on obtient comme estimé a/σ_a , $t_{DF} = -9.5181$, avec une p-value de 0.01. Le test semble corroborer que la série chronologique est stationnaire, i.e qu'une première différenciation doit s'appliquer.

e)

On cherche maintenant le modèle de notre série stationnaire. En observant le graphique ACF et PACF de la figure 5 et 6. Il est à noter que le lag est exprimé en année, ce qui veut dire que 0.1 équivaut à 1/10 d'année et que 0.5 équivaut à 6 mois.

Figure 5: ACF de la différence du logarithme

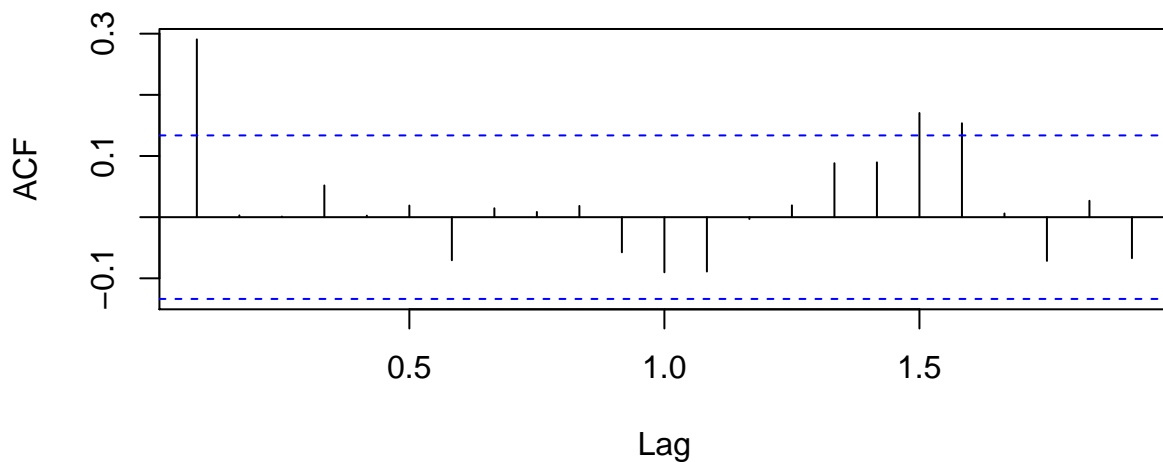
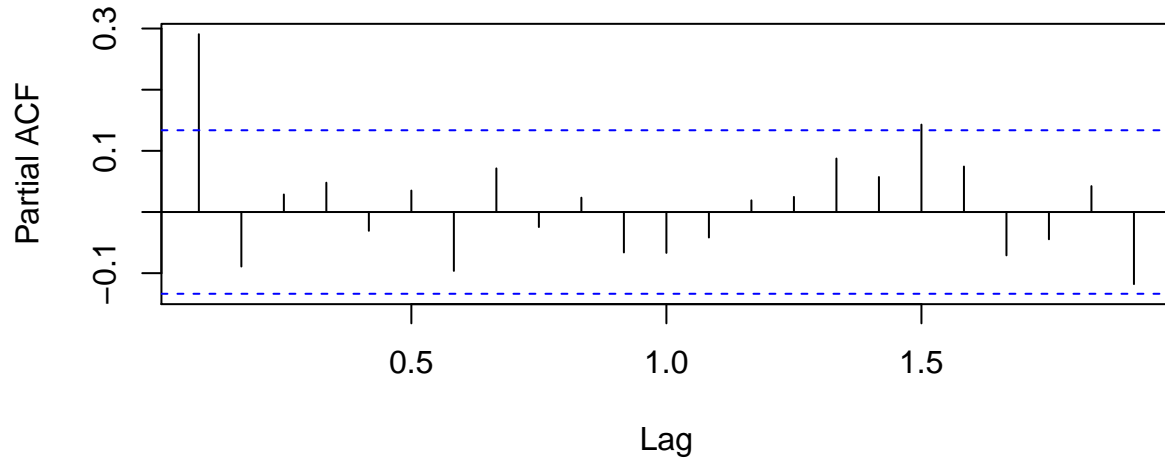


Figure 6: PACF de la différence du logarithme

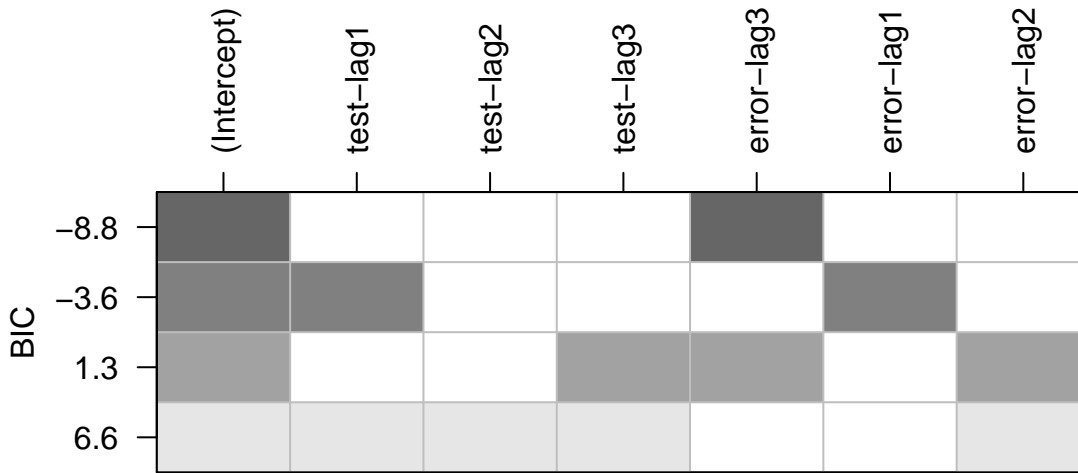


Il semblerait que l'on soit en présence d'un ARIMA(1,1,0) en se fiant au tableau ACF. Pour ce qui est du graphique PACF, on observe une ARIMA(0,1,1). Nous confirmons le tout avec la fonction d'autocorrélation étendue (EACF) ci-dessous.

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x o o o o o o o o o o o o o
## 1 x x o o o o o o o o o o o o
## 2 x o o o o o o o o o o o o o
## 3 x x x x o o o o o o o o o
## 4 x x o x o o o o o o o o o
## 5 x o x x o o o o o o o o o
## 6 x x x x o o o o o o o o o
## 7 x x x x x x o o o o o o o
```

Ce graphique porte à croire que l'on est en présence d'un modèle ARIMA(0,1,1), i.e. IMA(1). On continue notre analyse en appliquant le critère d'information Bayésienne (BIC).

```
## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax,
## force.in = force.in, : 2 linear dependencies found
## Reordering variables and trying again:
```



Nous concluons à partir de ce graphique que le meilleur modèle à considérer est l'ARIMA(0,1,3).

Pour résumé, après avoir analysé les différents graphiques de corrélation et les différents tests, nous considérerons les 3 modèles suivants basés sur le logarithme:

- ARIMA(1,1,0);
- ARIMA(0,1,1);
- ARIMA(0,1,3).

f)

Par la suite, il faut estimer les paramètres de chacun des modèles énumérés en (e) par la méthode du maximum de vraisemblance (MLE).

Pour l'ARIMA(0,1,1), le paramètre θ_1 de notre MA(1) est de 0.3118, ce qui nous donne le modèle suivant:

$$\log Y_t = \log Y_{t-1} + e_t + 0.3118e_{t-1}$$

Pour l'ARIMA(1,1,0), le paramètre ϕ_1 de notre AR(1) est de 0.2933, ce qui nous donne le modèle suivante:

$$\log Y_t = 1,2933 \log Y_{t-1} - 0.2933 \log Y_{t-2} + e_t$$

Pour l'ARIMA(0,1,3), le paramètre ϕ_1 de notre AR(1) est de 0.2933, ce qui nous donne le modèle suivante:
voir chapitre 5 pour écrire le modèle

Annexe

Tableaux et figures

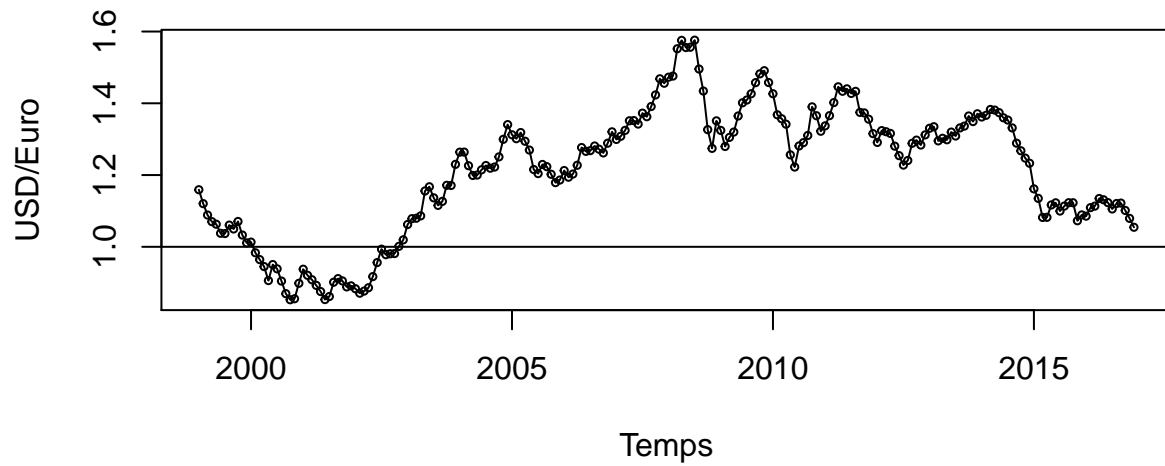


Fig 1. Taux de change US/Euro

Code informatique

```
# Test augmenté de Dickey-Fuller (ADF)
ar(log(ttaux))
adf.test(log(ttaux), k=2)

ttaux011 <- arima(log(ttaux), order=c(0,1,1), method='ML')
```