

Analyse de l'action Amazon.com, Inc. (AMZN) sous Python

Bilal Fatian

2024

Plan

- 1 Introduction
- 2 Description du serie
- 3 Modélisation ARIMA
- 4 Calcul de prévision

Plan

- 1 Introduction
- 2 Description du serie
- 3 Modélisation ARIMA
- 4 Calcul de prévision

Introduction

Amazon.com, Inc., initialement une librairie en ligne, s'est transformée en un géant mondial du e-commerce. Dans cette présentation, nous analyserons l'action d'Amazon, en examinant son historique boursier, les facteurs influençant son cours, et ses perspectives futures. Notre but est de comprendre les dynamiques qui façonnent la valeur de cette entreprise majeure dans l'économie numérique mondiale.

Plan

- 1 Introduction
- 2 Description du serie
- 3 Modélisation ARIMA
- 4 Calcul de prévision

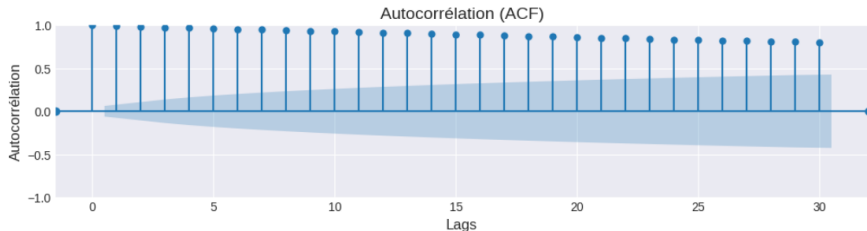
Série temporelle de AMAZON



La série originale

Le graphique montre une tendance haussière initiale suivie d'un recul marqué et d'une période de stabilisation. Une volatilité notable apparaît ensuite, reflétant une incertitude du marché. Finalement, on assiste à une reprise positive vers la fin de la période étudiée.

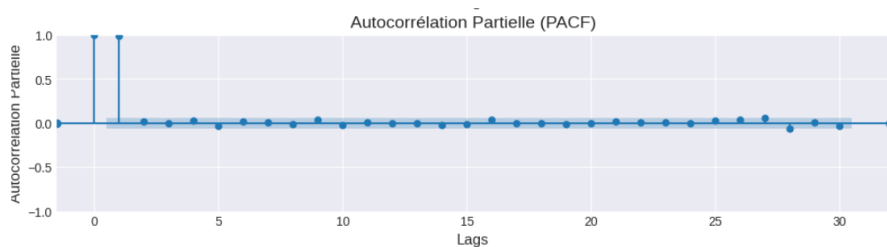
Autocorrélation



Autocorrélation

Forte autocorrélation initiale qui diminue progressivement, suggérant un fort effet de mémoire à court terme où les prix passés influencent fortement les prix futurs.

Autocorrélation partielle



Autocorrélation partielle

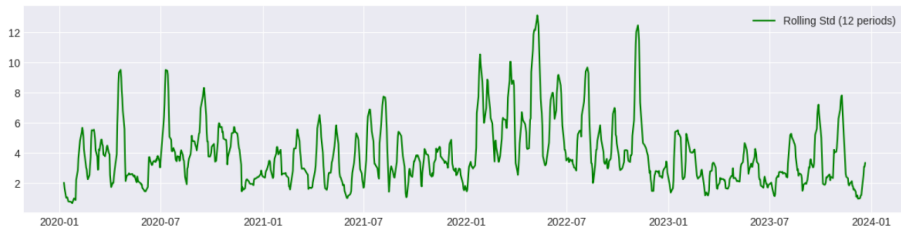
Un pic initial fort dans l'autocorrélation partielle diminue rapidement. Les décalages suivants montrent des corrélations faibles, indiquant que les prix passés n'influencent pas les prix futurs au-delà du premier décalage.

Graphiques de Moyenne Mobile



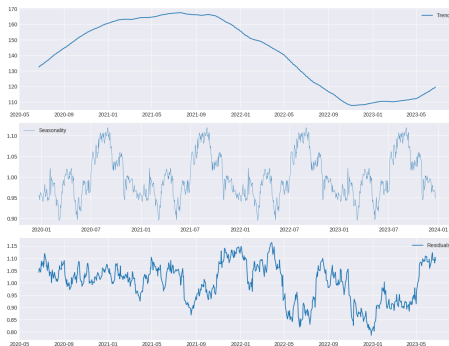
Le graphique affiche la moyenne mobile sur 12 périodes, qui souligne la tendance générale de la série temporelle. Il est visible que la moyenne mobile change avec le temps, indiquant une non-stationnarité de la tendance.

Graphiques de l'écart-type Mobile



Le troisième graphique dépeint l'écart-type mobile, et il est apparent que la volatilité change également au fil du temps, ce qui est une autre indication de non-stationnarité.

Décomposition de la série



- Le premier graphique montre une tendance non stationnaire due à une moyenne variable dans le temps.
- Le deuxième graphique indique une saisonnalité dans les données grâce à des motifs répétitifs.
- Le troisième graphique présente des résidus significatifs en amplitude.

Résultats du Test ADF

ADF Statistic: -2.1386837736273616

p-value: 0.22924672977058935

Critical Values:

1%, -3.436866962597422

Critical Values:

5%, -2.864417284603703

Critical Values:

10%, -2.5683020002450507

La série n'est pas stationnaire

Cette valeur d'ADF Statistic est trop haute par rapport aux seuils standards (1%, 5%, et 10%), ce qui signifie que la série ne semble pas être stationnaire. En d'autres termes, elle ne maintient pas un niveau constant au fil du temps. De plus, p-value de 0.2292 est plus élevée que le seuil commun de 0.05, ce qui confirme encore que la série est non-stationnaire.

Interprétation Globale

La série temporelle des prix de clôture d'Amazon n'est pas stationnaire. Elle présente des changements dans la moyenne et la volatilité au fil du temps ainsi qu'une tendance et une saisonnalité évidentes. Les résultats du test ADF confirment cette non-stationnarité.

Plan

- 1 Introduction
- 2 Description du serie
- 3 Modélisation ARIMA**
- 4 Calcul de prévision

L'ordre de différenciation $d=1$

Test Dickey et Fuller

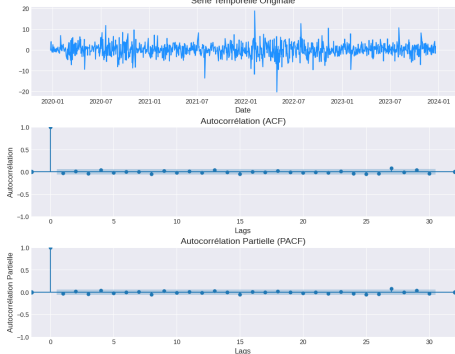
ADF Statistic for 1° differenced series: -32.551705957089794
p-value for 1° differenced series: 0.0

Étant donné que la p-value = 0, on peut rejeter l'hypothèse nulle selon laquelle la série est non stationnaire. En d'autres termes, on peut affirmer que la série est stationnaire.

L'ordre de différenciation $d=1$

l'ACF et le PACF

Série Temporelle Originale



La décomposition saisonnière



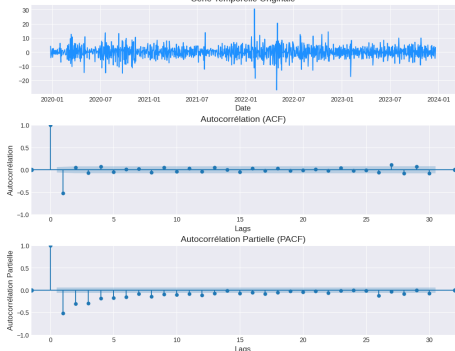
Ces représentations graphiques soutiennent la conclusion tirée de la statistique ADF que, après une première différenciation, la série est stationnaire.

L'ordre de différenciation $d=2$

Maintenant on assure qu'on ne doit pas choisir la différenciation d'ordre 2

l'ACF et le PACF

Série Temporelle Originale

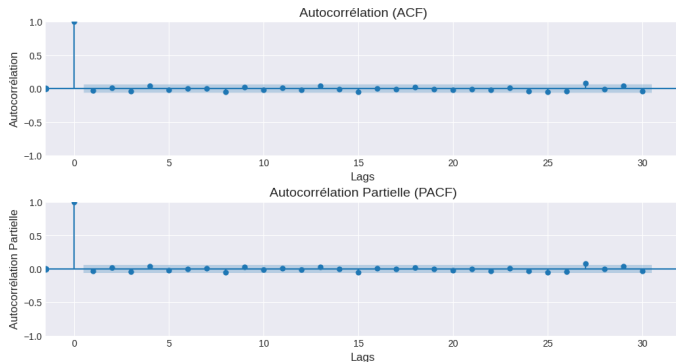


La décomposition saisonnière



Ces observations renforcent l'idée que $d=1$ est plus approprié que $d=2$. La différenciation doit être utilisée avec parcimonie, en différenciant seulement jusqu'à ce que la série devienne stationnaire. Aller au-delà de cela ne fait qu'augmenter la variabilité et réduire la qualité des informations que la série peut fournir pour les prévisions futures.

L'ordre du terme ARIMA(p,1,q)



Sur le graphique ACF et PACF, tous les lags tombent rapidement en dessous du seuil de signification. Cela suggère que l'ordre du terme AR et MA sont respectivement $p=0$ et $q=0$

Choix des modèles ARIMA à l'aide des critères d'information AIC et BIC

Le modèle optimisé selon l'AIC : ARIMA(2, 1, 2)

Best Model Summary (AIC):

SARIMAX Results

```

=====
Dep. Variable:      Close      No. Observations:      1007
Model:              ARIMA(2, 1, 2)  Log Likelihood      -2558.259
Date:              Fri, 12 Jan 2024  AIC              5126.519
Time:              02:37:06         BIC              5151.087
Sample:            0              HQIC             5135.854
Covariance Type:    - 1007
                    opg
=====

```

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-1.6914	0.011	-159.798	0.000	-1.712	-1.671
ar.L2	-0.9693	0.010	-94.869	0.000	-0.989	-0.949
ma.L1	1.7076	0.008	206.369	0.000	1.691	1.724
ma.L2	0.9979	0.010	104.279	0.000	0.979	1.017
sigma2	9.4509	0.251	37.719	0.000	8.960	9.942

```

=====
Ljung-Box (L1) (Q):      0.76  Jarque-Bera (JB):      800.20
Prob(Q):                0.38  Prob(JB):           0.00
Heteroskedasticity (H):  0.80  Skew:              -0.16
Prob(H) (two-sided):    0.04  Kurtosis:          7.36
=====

```

Le modèle optimisé selon l'AIC indiquant un ajustement plus fin aux données d'entraînement qui pourrait potentiellement capturer une plus grande complexité dans la structure de la série temporelle.

Le modèle optimisé selon le BIC : ARIMA(0, 1, 0)

Best Model Summary (BIC):

SARIMAX Results

```

=====
Dep. Variable:      Close      No. Observations:      1007
Model:              ARIMA(0, 1, 0)  Log Likelihood      -2566.161
Date:              Fri, 12 Jan 2024  AIC              5134.322
Time:              02:37:06         BIC              5139.236
Sample:            0              HQIC             5136.189
Covariance Type:    - 1007
                    opg
=====

```

	coef	std err	z	P> z	[0.025	0.975]
sigma2	9.6199	0.246	39.086	0.000	9.137	10.102

```

=====
Ljung-Box (L1) (Q):      0.77  Jarque-Bera (JB):      703.39
Prob(Q):                0.38  Prob(JB):           0.00
Heteroskedasticity (H):  0.78  Skew:              -0.13
Prob(H) (two-sided):    0.02  Kurtosis:          7.09
=====

```

Le modèle préféré selon le BIC est essentiellement une marche aléatoire sans composantes autorégressives ou de moyenne mobile supplémentaires. Ce modèle est favorisé pour sa simplicité et sa parcimonie, ce qui peut souvent conduire à de meilleures performances de prévision en raison d'une moindre propension au surajustement.

Tests de validation sur les résidus estimés du modèle ARIMA

Tests de validation sur les résidus estimés du modèle qui minimise le BIC : ARIMA(0,1,0)

Test de Ljung-Box:
lb_stat lb_pvalue
10 8.605849 0.569873

H0 : Les résidus du modèle ne présentent pas d'autocorrélation.
p-value > 0.05 , alors nous ne rejetons pas l'hypothèse nulle.

Test de Breusch-Pagan:
Lagrange Multiplier p-value: 0.6122664976834833

Test de White:
Test Statistic p-value: 0.05784949536021828

H0 : la variance des résidus est constante (homoscédasticité).
dans les deux test : p-value > 0.05 , alors nous ne rejetons pas l'hypothèse nulle.

Test de Jarque-Bera:
JB Statistic: 703.3929406263005 p-value: 1.8203866342687895e-153
H0 : Les résidus suivent une distribution normale.
p-value < 0.05 , alors nous rejetons l'hypothèse nulle.

Test de Ljung-Box:
lb_stat lb_pvalue
10 3.484421 0.967625

H0 : Les résidus du modèle ne présentent pas d'autocorrélation.
p-value > 0.05 , alors nous ne rejetons pas l'hypothèse nulle.

Test de Breusch-Pagan:
Lagrange Multiplier p-value: 0.6081730375249801

Test de White:
Test Statistic p-value: 0.06544097877195035

H0 : la variance des résidus est constante (homoscédasticité).
dans les deux test : p-value > 0.05 , alors nous ne rejetons pas l'hypothèse nulle.

Test de Jarque-Bera:
JB Statistic: 796.3928512198348 p-value: 1.162758824062317e-173
H0 : Les résidus suivent une distribution normale.
p-value < 0.05 , alors nous rejetons l'hypothèse nulle.

Les tests sur les résidus pour les deux modèles indique l'absence d'autocorrélation significative dans les résidus, permettant de les considérer comme du bruit blanc. ils montrent qu'ils sont homoscédastiques, mais leur non-normalité indique une possible limite du modèle pour certaines analyses ou prédictions.

Problématique de la Non-Normalité des Résidus !!!

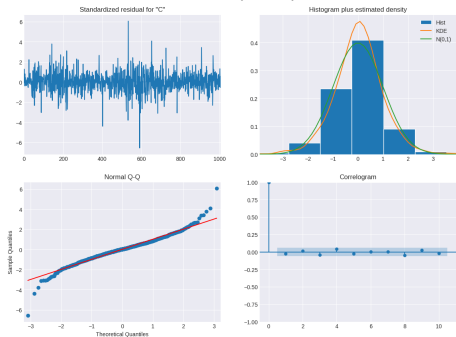
Est ce qu'il est possible d'obtenir un modèle ARIMA dont les résidus suivent une distribution normale pour la serie de l'action Amazon.com, Inc ?

Problématique de la Non-Normalité des Résidus !!!

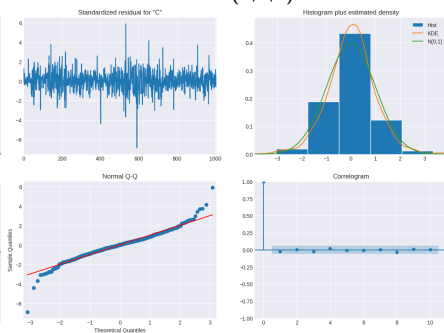
Pour la série temporelle d'Amazon, tous les modèles ARIMA testés, avec des combinaisons de p , d , et q (où p et q varient entre 0 et 3 et d est 1 ou 2), aboutissent à des résidus non-normaux, même en analysant chaque année séparément. Cela suggère que la structure complexe des données financières, l'impact des facteurs externes sur les prix des actions, et les limites des modèles ARIMA à saisir pleinement ces motifs complexes, rendent difficile l'obtention de résidus normaux.

Tracés Résiduels pour le Modèle ARIMA(0,1,0) et ARIMA(2,1,2)

ARIMA(0,1,0)



ARIMA(2,1,2)



- Résultats similaires pour ARIMA(0,1,0) et ARIMA(2,1,2)
- Pas d'hétéroscédasticité évidente
- Légère déviation de la distribution normale dans les résidus
- Queues plus lourdes que normales et absence de distribution parfaitement normale dans les résidus
- Pas d'autocorrélation significative entre les résidus à différents décalages

Conclusion de l'Analyse des Résidus

Les résultats des graphiques viennent confirmer les observations issues des tests précédents, assurant que :

- Absence de biais : Les résidus oscillent autour de zéro, indiquant peu de biais systématique.
- Autocorrélation : Aucune autocorrélation significative dans les résidus, le modèle capture bien la corrélation temporelle.
- Stabilité de la variance : Tests confirment une variance stable des résidus dans le temps.
- Normalité : les résidus ne suivent pas une distribution normale.

La Sélection du modèle ARIMA(0,1,0)

L'analyse des résidus montre que les modèles ARIMA(2,1,2) et ARIMA(0,1,0) sont similaires. Le modèle ARIMA(0,1,0), en plus de minimiser le BIC, favorise le principe de parcimonie, c'est-à-dire qu'il offre une qualité d'ajustement comparable à celui d'ARIMA(2,1,2), tout en étant plus simple. De plus, les graphiques ACF et PACF suggèrent que (0,1,0) est l'option optimale.

le modèle ARIMA optimal (Out-of-Time Cross validation)

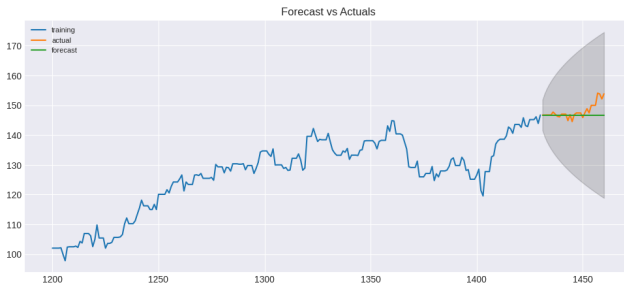
La meilleure combinaison de paramètres basée sur la moyenne des RMSE est : $(0, 1, 0)$ avec
RMSE moyen de : 26.282508349223306
MSE moyen de : 690.7702451269928
MAE moyen de : 20.007969747158725
MAPE moyen de : 16.816136739641898

Après plusieurs tests, le modèle ARIMA(0,1,0) a été retenu avec un RMSE de 26.28 pour les prix de clôture d'Amazon. Ce modèle, basé sur la différenciation, est le plus précis parmi ceux examinés. Les fluctuations passées ne sont pas de bons indicateurs pour les prévisions futures au-delà de la tendance générale, suggérant l'imprévisibilité des mouvements futurs, typique d'un marché efficient.

Plan

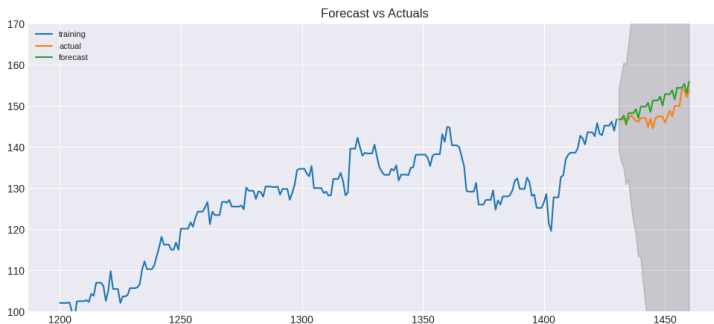
- 1 Introduction
- 2 Description du serie
- 3 Modélisation ARIMA
- 4 Calcul de prévision**

Calcul de prévisions hors échantillon du 22/11/2023 au 22/12/2023 avec ARIMA(0,1,0)



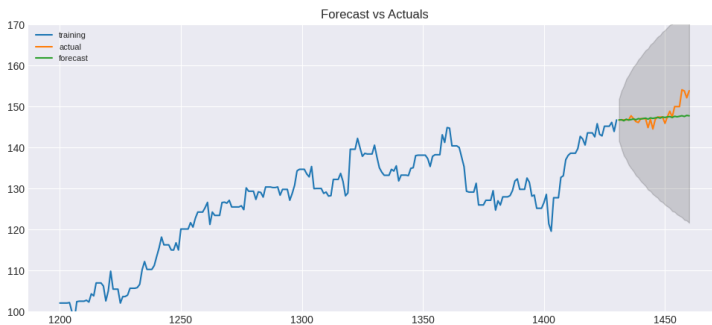
Les prévisions du modèle ARIMA(0,1,0) pour les prix de clôture d'Amazon divergent des données réelles, malgré sa relative précision par rapport aux autres modèles testés. Ceci suggère que ce modèle ne parvient pas à saisir la complexité des mouvements de prix sur le marché, ce qui signifie que les données passées ne sont pas une base solide pour prédire les prix futurs.

Calcul de prévisions hors échantillon du 22/11/2023 au 22/12/2023 avec SARIMA order=(0, 1, 0) et seasonal order=(0, 1, 0, 5)



Ce modèle suit effectivement la tendance de la série temporelle. Cela signifie que le modèle est capable de prédire la direction générale que prend la série temporelle, en partie grâce à la différenciation qui aide à stabiliser la moyenne sur le temps. La composante saisonnière avec une période de 5 permet au modèle de prendre en compte et d'ajuster les prévisions en fonction des schémas qui se répètent tous les 5 intervalles de temps.

Calcul de prévisions hors échantillon du 22/11/2023 au 22/12/2023 avec SARIMA $\text{order}=(2, 1, 2)$ et $\text{seasonal order}=(2, 1, 2, 5)$



Le modèle SARIMAX avec $\text{order}=(2, 1, 2)$ et $\text{seasonal_order}=(2, 1, 2, 5)$ ne capture pas la tendance de la série temporelle comme souhaité. Par conséquent, il semble judicieux de choisir le modèle SARIMAX plus simple avec $\text{order}=(0, 1, 0)$ et $\text{seasonal_order}=(0, 1, 0, 5)$, qui a démontré une meilleure capacité à suivre la tendance de la série.

Auto Arima Forecast

```

Best model: ARIMA(0,1,1)(0,0,0)[0] intercept
Total fit time: 16.637 seconds

=====
SARIMAX Results
=====
Dep. Variable:          y      No. Observations:      1460
Model:                SARIMAX(0, 1, 1)      Log Likelihood      -3447.788
Date:                Thu, 11 Jan 2024      AIC      6901.576
Time:                21:38:42      BIC      6917.432
Sample:              12-23-2019      HQIC      6907.491
                  - 12-21-2023
Covariance Type:      opg
=====
              coef      std err      z      P>|z|      [0.025      0.975]
-----
intercept      0.0440      0.063      0.694      0.488      -0.080      0.168
ma.L1          -0.0585      0.020     -2.960      0.003      -0.097     -0.020
sigma2          6.6085      0.128     51.530      0.000      6.357      6.860
=====
Ljung-Box (L1) (Q):              0.00      Jarque-Bera (JB):              2872.52
Prob(Q):                        1.00      Prob(JB):                  0.00
Heteroskedasticity (H):          0.78      Skew:                      -0.13
Prob(H) (two-sided):            0.01      Kurtosis:                  9.87
=====

```

La fonction `auto.arima` a choisi le modèle $ARIMA(0,1,1)$ pour prédire les prix de clôture d'Amazon. Ce modèle se base principalement sur les variations récentes plutôt que sur les tendances à long terme.

Code et Référence

Pour accéder au code Jupyter contenant tous les résultats et graphiques mentionnés précédemment, veuillez visiter le lien suivant :

```
https://colab.research.google.com/drive/1W66S74l\_  
VApKex4R9CnNKe0F9WzRfWvI?usp=sharing
```