

Projet économétrie

Karim Hajji et Nour Mahfoudhi

Introduction :

L'analyse des séries temporelles est un outil central en économétrie pour explorer les interactions entre des variables économiques et financières dans le temps. Ce projet s'intéresse aux relations dynamiques entre deux actifs clés des marchés financiers : le prix de l'or (Gold Prices en USD) et l'indice S&P 500 (en USD). L'objectif principal est de comprendre si l'or peut être considéré comme une solution pour préserver la valeur des investissements lorsque les marchés boursiers, comme le S&P 500, connaissent une instabilité, et comment ces deux séries interagissent sur différentes périodes économiques.

Ce choix s'appuie sur des considérations pratiques et théoriques. En tant qu'étudiants en finance, ces actifs représentent des concepts fondamentaux : 1. L'or est souvent considéré comme un actif sécurisé ou "refuge", utilisé pour protéger la valeur des portefeuilles en période de crise. Les investisseurs y ont recours lorsqu'ils anticipent une baisse des marchés financiers. 2. Le S&P 500, en revanche, est un indicateur majeur de la performance économique et de la confiance des investisseurs. Sa volatilité reflète les attentes vis-à-vis de la croissance économique et des risques macroéconomiques.

Cette étude est particulièrement pertinente pour nous, étudiants en finance, pour deux raisons principales. Premièrement, en tant que futurs professionnels travaillant dans des banques, sociétés de gestion ou institutions financières, il est essentiel de comprendre les interactions entre ces actifs pour conseiller efficacement les clients, gérer des portefeuilles et anticiper les réactions des marchés face aux crises économiques. Deuxièmement, en tant que futurs investisseurs et épargnants, nous devons savoir comment préserver ou faire fructifier nos économies. Comprendre le rôle de l'or comme refuge ou celui des actions comme opportunité de croissance en fonction des conditions de marché est crucial pour prendre des décisions éclairées, en particulier en période d'instabilité économique.

Ce projet permettra d'analyser les interactions entre le prix de l'or et le S&P 500, notamment en période de crise, afin de comprendre leur comportement respectif et leurs implications pratiques pour les investisseurs et les gestionnaires de portefeuille.

Méthodologie et données :

Les données utilisées dans cette étude consistent en des observations mensuelles couvrant la période allant de janvier 2000 à décembre 2019, soit un total de 240 observations pour chaque série. Elles ont été extraites de la base de données DBnomics. Ces séries ont été choisies pour leur disponibilité sur une longue période, leur pertinence économique et leur fréquence commune, ce qui garantit une cohérence dans l'analyse.

Cette période a été retenue car elle inclut plusieurs événements économiques significatifs tels que la crise financière mondiale de 2008, suivie par une reprise progressive des marchés dans les années suivantes. Cette période permet d'analyser les dynamiques entre l'or et le S&P 500 dans des contextes variés, incluant des périodes de crise, de stabilisation et de croissance économique. L'exclusion des données post-2019 vise à éviter les distorsions liées à la crise sanitaire de la COVID-19, qui a entraîné des comportements atypiques sur les marchés financiers, et garantit une analyse basée sur des dynamiques économiques normales.

D'une part, nous appliquerons une approche univariée pour analyser chaque série séparément (prix de l'or et indice S&P 500). Nous commencerons par une exploration visuelle des séries pour identifier les tendances, les ruptures ou les périodes de volatilité. Nous calculerons les autocorrélogrammes simples (ACF) et partiels (PACF) pour détecter les dépendances temporelles. Ensuite, nous vérifierons la stationnarité de chaque série à l'aide du test de Dickey-Fuller augmenté (ADF). Si nécessaire, des transformations seront effectuées pour rendre les séries stationnaires. Une fois les séries stationnalisées, nous identifierons le modèle ARMA optimal pour l'une des deux séries, en nous appuyant sur des critères comme l'AIC ou le BIC pour sélectionner le modèle le plus adapté. Ce modèle sera ensuite utilisé pour effectuer des prévisions sur des horizons de 1 à 3 mois, en tenant compte des propriétés de la série initiale.

D'autre part, nous adopterons une approche multivariée pour examiner les interactions entre le prix de l'or et le S&P 500 via un modèle VAR. Nous déterminerons le nombre de retards optimal pour le modèle en utilisant des critères comme l'AIC et le BIC. Ensuite, nous réaliserons des tests de causalité de Granger pour explorer si les variations d'une série permettent de prédire celles de l'autre. Nous effectuerons une analyse impulsion-réponse pour mesurer l'effet d'un choc exogène sur une série (par exemple, une baisse du S&P 500) sur l'autre (le prix de l'or). Enfin, un test de cointégration de Johansen sera mené pour vérifier si une relation à long terme existe entre ces deux séries. En cas de cointégration, un modèle VECM sera estimé pour capturer les dynamiques à court et long terme entre ces variables.

Hypothèses de recherche et résultats attendus :

Impact du S&P 500 sur le prix de l'or : Nous pensons qu'une baisse du S&P 500 entraînera une hausse du prix de l'or, car l'or est souvent utilisé par les investisseurs pour protéger leur capital en période de crise. Ce résultat confirmerait que l'or agit comme un refuge face à la volatilité des marchés boursiers.

Impact du prix de l'or sur le S&P 500 : Une augmentation du prix de l'or pourrait refléter une anticipation de risques économiques par les investisseurs, ce qui pourrait peser sur la confiance dans les marchés actions représentés par le S&P 500.

Relation entre l'or et le S&P 500 : Nous nous attendons à une relation négative entre l'or et le S&P 500 en période de crise. En revanche, cette relation pourrait être neutre ou légèrement positive en période de stabilité économique, lorsque les investisseurs ne recherchent pas de refuge particulier.

Causalité entre les deux variables : Nous supposons que les variations du S&P 500 influenceront davantage les prix de l'or que l'inverse, car les investisseurs réagissent d'abord aux mouvements des marchés boursiers.

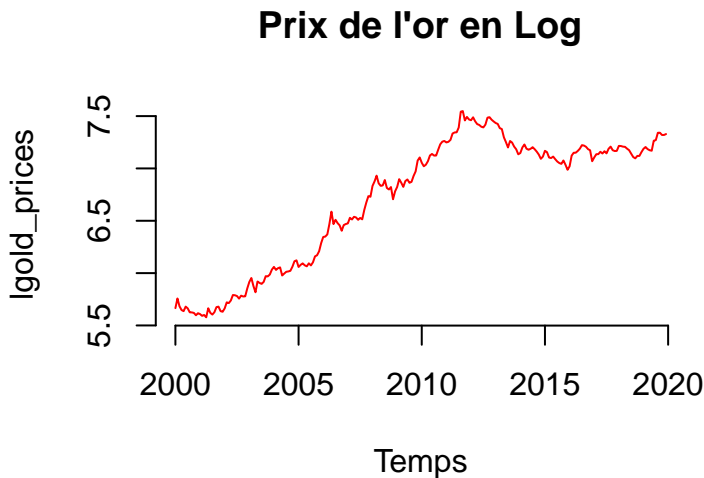
La comparaison des résultats obtenus avec ces hypothèses permettra de tester la robustesse de nos hypothèses de départ et d'affiner notre compréhension des interactions entre le prix de l'or et le S&P 500. Les interprétations des résultats éclaireront les implications pratiques pour les investisseurs et gestionnaires de fonds, en particulier dans la gestion des risques et l'allocation stratégique des portefeuilles.

Modélisation univariée

Représentations graphiques et autocorrélogrammes simples et partiels des séries

Le prix de l'or en log

```
plot(
  lgold_prices, frame = FALSE, col = 'red',
  main = "Prix de l'or en Log", xlab = 'Temps'
)
```

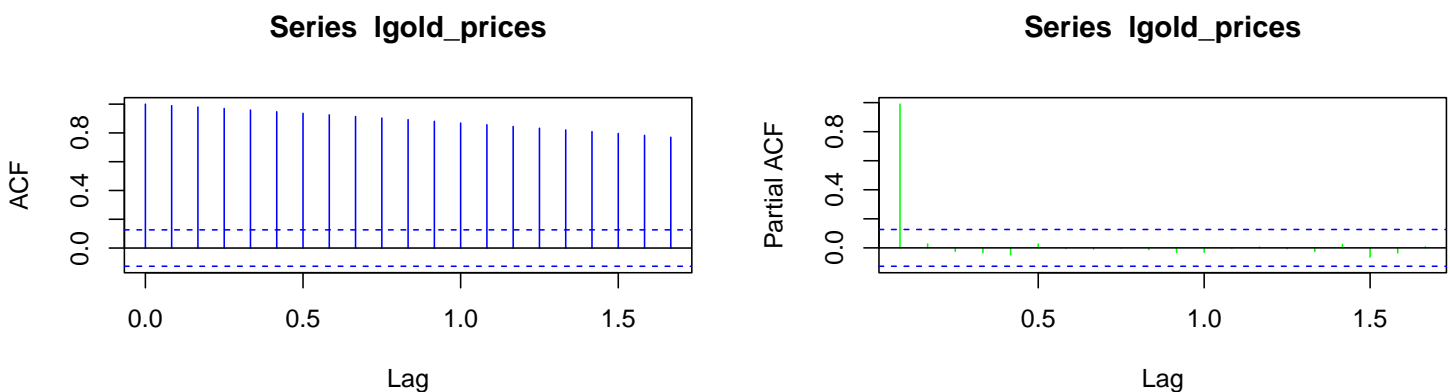


Commentaire :

L'analyse du graphique du prix de l'or en log montre une trajectoire non stationnaire, ce qui implique que les propriétés statistiques de la série, comme la moyenne et la variance ne sont pas constantes dans le temps. Une transformation de la série pourrait donc s'avérer indispensable pour stabiliser ses propriétés statistiques et permettre une analyse plus fiable des interactions entre le prix de l'or et le S&P 500.

Autocorrélation simple et partielle

```
par(mfrow = c(1, 2))
acf(lgold_prices, lag.max = 20, col = 'blue')
pacf(lgold_prices, lag.max = 20, col = 'green')
```



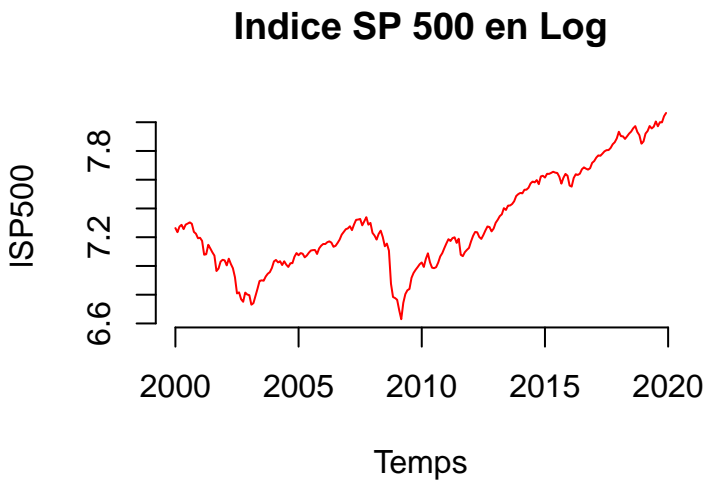
```
par(mfrow = c(1, 1))
```

Commentaire :

La décroissance très lente de l'autocorrélation simple (ACF) observée dans le graphique est un indicateur clair de la non-stationnarité de la série `log_gold_prices`. En effet, les autocorrélations restent significatives sur plusieurs lags, suggérant une forte dépendance temporelle. Du côté de l'autocorrélation partielle (PACF), seule la première observation est clairement significative, ce qui indique que les dépendances temporelles immédiates sont bien capturées par un modèle autorégressif simple. Ainsi, un modèle AR(1) avec différenciation pourrait être particulièrement adapté pour modéliser la série.

Le SP 500 en log

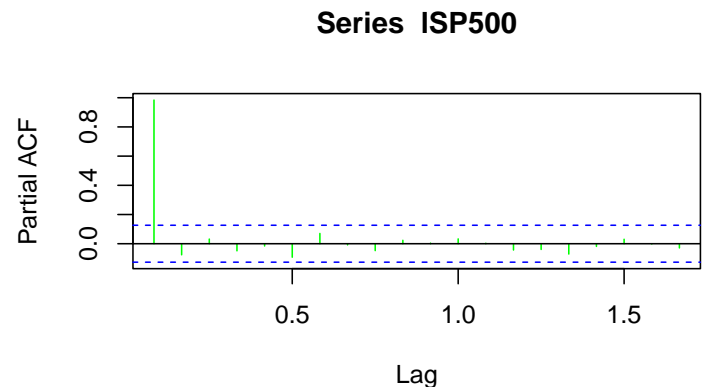
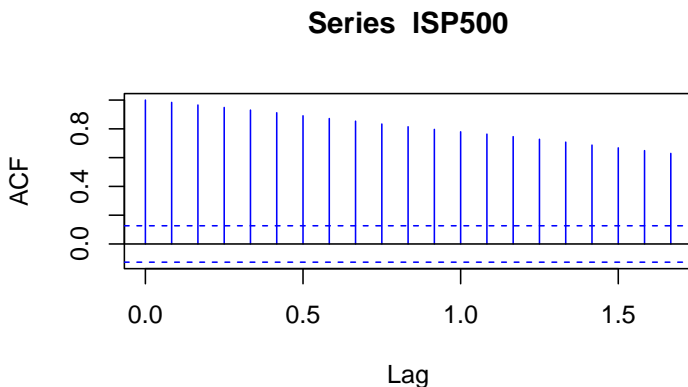
```
plot(
  lSP500, frame = FALSE, col = 'red',
  main = "Indice SP 500 en Log", xlab = 'Temps'
)
```



Commentaire : L'analyse du graphique de l'indice SP 500 en log montre également une trajectoire non stationnaire, ce qui indique que la moyenne et la variance de la série ne sont pas constantes dans le temps. Une transformation supplémentaire, comme la différenciation, pourrait être nécessaire pour stabiliser la série avant d'appliquer un modèle économétrique.

Autocorrélation simple et partielle

```
par(mfrow = c(1, 2))
acf(lSP500, lag.max = 20, col = 'blue')
pacf(lSP500, lag.max = 20, col = 'green')
```



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

Commentaire : La décroissance lente de l'autocorrélation (ACF) pour la série SP 500 en log indique une dépendance temporelle persistante, signe de non-stationnarité. De plus, l'autocorrélation partielle (PACF) montre une significativité principalement au premier lag, ce qui suggère qu'un modèle AR(1) avec différenciation pourrait être approprié pour rendre cette série stationnaire.

Evaluation de la stationnarité

Appliquons à présent la stratégie de test de racine unitaire à chacune des séries avec deux tests : Le test de Dickey Fuller augmenté et le test KPSS. Ces tests nous permettront de détecter la présence éventuelle d'une racine unitaire, et ainsi d'évaluer si nos séries sont stationnaires ou non.

Test de Dickey Fuller Augmenté (ADF)

Le test de Dickey-Fuller augmenté est défini comme suit :

$$\begin{cases} H_0 : \rho = 0 \Rightarrow X_t \text{ non stationnaire} \\ H_1 : \rho < 0 \Rightarrow X_t \text{ stationnaire} \end{cases}$$

Test de Dickey Fuller pour la série "gold_prices" Afin de mettre en oeuvre les tests de racines unitaires, on procède à une stratégie séquentielle de test : d'abord on teste à la fois la constante et la tendance dans le modèle le plus général (Modèle

3). Ensuite, nous évaluons la pertinence de ce modèle. Si ce modèle ne se révèle pas approprié, nous recommandons le test en utilisant un modèle plus restreint, jusqu'à ce que le modèle optimal soit identifié.

Le nombre de retards est fixé à 6 car nos données sont mensuelles et le nombre de retards optimal est déterminé par la minimisation du critère AIC.

Modèle 3 :

$$\Delta X_t = c + bt + \rho X_{t-1} + \sum_{j=1}^p \delta_j \Delta X_{t-j} + \varepsilon_t$$

```
test_M3_gold_prices <- ur.df(y = lgold_prices, type = 'trend', lags = 6, selectlags = 'AIC')
coefficients(test_M3_gold_prices@testreg)
```

```
##              Estimate Std. Error    t value Pr(>|t|)
## (Intercept)  5.991761e-02 5.569712e-02  1.07577577 0.28316459
## z.lag.1      -7.939599e-03 9.643082e-03 -0.82334666 0.41117124
## tt          7.645057e-06 8.664097e-05  0.08823836 0.92976468
## z.diff.lag1  9.222661e-02 6.569161e-02  1.40393292 0.16169945
## z.diff.lag2 -1.271354e-01 6.567491e-02 -1.93582964 0.05412493
```

```
print(test_M3_gold_prices@cval)
```

```
##      1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47
```

Commentaire :

D'après les résultats du test de Dickey-Fuller appliqué à notre série, avec un niveau de signification fixé à 5%, on trouve $t_{ADF} = -0.823$, tandis que la valeur critique à 5% est de -3.43. Donc $t_{ADF} > -3.43$. Comme la statistique de test est supérieure à cette valeur critique, nous ne pouvons pas rejeter l'hypothèse nulle selon laquelle la série possède une racine unitaire. On a donc une racine unitaire.

Test du coefficient b (la tendance) On teste la valeur du coefficient b avec les hypothèses suivantes :

$$\begin{cases} H_0 : b = 0 & (\text{le coefficient n'est pas significatif}) \\ H_1 : b \neq 0 & (\text{le coefficient est significatif}) \end{cases}$$

La valeur de la statistique de test est $|\hat{t}_b| = 0.088$. Comme précédemment la racine unitaire n'a pas été rejetée, nous regardons la valeur critique dans les tables de Dickey Fuller pour le modèle 3 (250 observations) qui vaut 3,12. Comme $\hat{t}_b < 3.14$ au seuil 5%, nous ne pouvons pas rejeter l'hypothèse nulle H_0 . Cela signifie que le coefficient b n'est pas significatif dans notre modèle, il vaut 0.

Considérons à présent un modèle plus contraint sans la tendance mais que la constante.

Modèle 2 :

$$\Delta X_t = c + \rho X_{t-1} + \sum_{j=1}^p \delta_j \Delta X_{t-j} + \varepsilon_t$$

```
test_M2_gold_prices <- ur.df(y = lgold_prices, type = 'drift', lags = 6, selectlags = 'AIC')
coefficients(test_M2_gold_prices@testreg)
```

```
##              Estimate Std. Error    t value Pr(>|t|)
## (Intercept)  0.055801645 0.030369472  1.837426 0.06744279
## z.lag.1      -0.007186996 0.004489332 -1.600906 0.11077643
## z.diff.lag1  0.091597330 0.065161724  1.405692 0.16117082
## z.diff.lag2 -0.127831365 0.065058232 -1.964876 0.05063769
```

```
print(test_M2_gold_prices@cval)
```

```
##      1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

Commentaire :

La statistique de test obtenue est $t_{ADF} = -1.601$, tandis que la valeur critique à 5% est de -2.88. Comme $t_{ADF} > -2.88$, nous ne pouvons pas rejeter l'hypothèse nulle selon laquelle la série possède une racine unitaire. En d'autres termes, le test indique que la série possède une racine unitaire.

Test du coefficient c (la constante)

On teste la valeur du coefficient c :

$$\begin{cases} H_0 : c = 0 \\ H_1 : c \neq 0 \end{cases}$$

La valeur de la statistique de test est $|\hat{t}_c| = 1.837$. Comme précédemment la racine unitaire n'a pas été rejetée, nous regardons la valeur critique dans la table de Dickey Fuller (250 observations) qui vaut 2.84. Comme $\hat{t}_c < 2.84$ au seuil 5%, nous ne pouvons pas rejeter l'hypothèse nulle H_0 . Cela signifie que le coefficient c n'est pas significatif dans notre modèle 2 puisque $c=0$.

Considérons un modèle plus contraint sans tendance ni constante :

Modèle 1 :

$$\Delta X_t = \rho X_{t-1} + \sum_{j=1}^p \delta_j \Delta X_{t-j} + \varepsilon_t$$

```
test_M1_gold_prices <- ur.df(y = lgold_prices, type = 'none', lags = 6, selectlags = 'AIC')
coefficients(test_M1_gold_prices@testreg)
```

```
##              Estimate   Std. Error   t value   Pr(>|t|)
## z.lag.1         0.001027493 0.0004112261  2.498608 0.01316748
## z.diff.lag1     0.096506532 0.0654423720  1.474680 0.14166619
## z.diff.lag2    -0.122861796 0.0653368937 -1.880435 0.06131254
```

```
print(test_M1_gold_prices@cval)
```

```
##      1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

Commentaire :

D'après les résultats du test de Dickey-Fuller appliqué à notre série, avec un niveau de significativité fixé à 5%, on trouve $t_{ADF} = 2.499$, tandis que la valeur critique à 5% est de -1.95. Donc $t_{ADF} > -1.95$. Comme la statistique de test est supérieure à cette valeur critique, nous ne pouvons pas rejeter l'hypothèse nulle selon laquelle la série possède une racine unitaire au seuil de 5%. On a donc la présence d'une racine unitaire puisque l'hypothèse nulle a été acceptée.

La stratégie séquentielle s'arrête ici et on conclut à un processus intégré **I(1)**. Notre processus associé à la série `gold_prices` n'est donc pas stationnaire. Il correspond à un processus DS (Differency Stationary). Pour le stationnariser il faudra donc le différencier une fois.

Test de Dicky Fuller pour la série "SP500"

Modèle 3 :

```
test_M3_SP500 <- ur.df(y = lSP500, type = 'trend', lags = 6, selectlags = 'AIC')
coefficients(test_M3_SP500@testreg)
```

```
##              Estimate   Std. Error   t value   Pr(>|t|)
## (Intercept)  0.1982970698 8.181694e-02  2.42366768 1.615832e-02
## z.lag.1      -0.0300298618 1.203569e-02 -2.49506687 1.331505e-02
## tt          0.0001886092 6.048621e-05  3.11821751 2.058063e-03
## z.diff.lag1  0.2634104740 6.446997e-02  4.08578527 6.119891e-05
## z.diff.lag2 -0.0902333548 6.601168e-02 -1.36693023 1.730175e-01
## z.diff.lag3  0.1176332965 6.627896e-02  1.77482111 7.728551e-02
## z.diff.lag4  0.0029076973 6.640666e-02  0.04378623 9.651138e-01
## z.diff.lag5  0.1777295259 6.628395e-02  2.68133560 7.878635e-03
## z.diff.lag6 -0.1889434416 6.500189e-02 -2.90673755 4.018940e-03
```

```
print(test_M3_SP500@cval)
```

```
##      1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47
```

Commentaire :

D'après les résultats du test de Dickey-Fuller appliqué à notre série, avec un niveau de significativité fixé à 5%, la statistique de test obtenue est $t_{ADF} = -2.495$, tandis que la valeur critique à 5% est de -3.43. Comme $t_{ADF} > -3.43$, nous ne pouvons pas rejeter l'hypothèse nulle selon laquelle la série possède une racine unitaire.

Test du coefficient b (la tendance)

On teste la valeur du coefficient b avec les hypothèses suivantes :

$$\begin{cases} H_0 : b = 0 & (\text{le coefficient n'est pas significatif}) \\ H_1 : b \neq 0 & (\text{le coefficient est significatif}) \end{cases}$$

La valeur calculée de la statistique de test est $|\hat{t}_b| = 3.118$. Comme précédemment, la racine unitaire n'a pas été rejetée, nous regardons la valeur critique dans les tables de Dickey-Fuller pour le modèle 3, qui vaut 3.12. Comme $\hat{t}_b < 3.12$ au seuil de 5%, nous ne pouvons pas rejeter l'hypothèse nulle H_0 . Cela signifie que le coefficient b n'est pas significatif dans notre modèle et vaut 0.

Considérons à présent un modèle plus contraint.

Modèle 2 :

```
test_M2_SP500 <- ur.df(y = 1SP500, type = 'drift', lags = 6, selectlags = 'AIC')
coefficients(test_M2_SP500@testreg)
```

```
##              Estimate Std. Error    t value    Pr(>|t|)
## (Intercept) -0.0021507635 0.051585350 -0.04169330 9.667802e-01
## z.lag.1      0.0006183784 0.007079832  0.08734366 9.304760e-01
## z.diff.lag1  0.2748940148 0.065600554  4.19042215 4.001189e-05
## z.diff.lag2 -0.0929469103 0.067273302 -1.38163147 1.684552e-01
## z.diff.lag3  0.1214772249 0.067539879  1.79859997 7.342223e-02
## z.diff.lag4  0.0010865306 0.067679095  0.01605415 9.872054e-01
## z.diff.lag5  0.1728621066 0.067537919  2.55948227 1.113844e-02
## z.diff.lag6 -0.1982787167 0.066179673 -2.99606672 3.041328e-03
```

```
print(test_M2_SP500@cval)
```

```
##      1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

Commentaire :

La statistique de test obtenue est $t_{ADF} = 0.087$, tandis que la valeur critique à 5% est de -2.88. Comme $t_{ADF} > -2.88$, nous ne pouvons pas rejeter l'hypothèse nulle selon laquelle la série possède une racine unitaire. En d'autres termes, le test indique que la série possède une racine unitaire dans le modèle 2.

Test du coefficient c (la constante) On teste la valeur du coefficient c :

$$\begin{cases} H_0 : c = 0 \\ H_1 : c \neq 0 \end{cases}$$

La valeur de la statistique de test est $|\hat{t}_c| = 0.042$. Comme précédemment, la racine unitaire n'a pas été rejetée, nous regardons la valeur critique sur la table de Dickey Fuller, qui vaut 2,86. Comme $\hat{t}_c < 2.84$ au seuil de 5%, nous ne pouvons pas rejeter l'hypothèse nulle H_0 donc c vaut 0.

Considérons un modèle plus contraint sans tendance ni constante :

Modèle 1 :

```
test_M1_SP500 <- ur.df(y = 1SP500, type = 'none', lags = 6, selectlags = 'AIC')
coefficients(test_M1_SP500@testreg)
```

```
##              Estimate Std. Error    t value    Pr(>|t|)
## z.lag.1      0.0003234995 0.0003197917  1.01159431 3.128139e-01
## z.diff.lag1  0.2751209138 0.0652298891  4.21771243 3.572322e-05
## z.diff.lag2 -0.0926879160 0.0668377851 -1.38675924 1.668815e-01
## z.diff.lag3  0.1217554309 0.0670608543  1.81559618 7.075762e-02
## z.diff.lag4  0.0013839828 0.0671532148  0.02060933 9.835755e-01
## z.diff.lag5  0.1731592719 0.0670122895  2.58399277 1.039582e-02
## z.diff.lag6 -0.1979063112 0.0654291588 -3.02474180 2.776260e-03
```

```
print(test_M1_SP500@cval)
```

```
##      1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

Commentaire :

D'après les résultats du test de Dickey-Fuller appliqué à notre série, avec un niveau de significativité fixé à 5%, on trouve $t_{ADF} = 1.012$, tandis que la valeur critique à 5% est de -1.95. Donc $t_{ADF} > -1.95$. Comme la statistique de test est supérieure à cette valeur critique, nous ne pouvons pas rejeter l'hypothèse nulle selon laquelle la série possède une racine unitaire au seuil de 5%. On a donc la présence d'une racine unitaire puisque l'hypothèse nulle a été acceptée.

La stratégie séquentielle s'arrête ici (au modèle 1) et on conclut à un processus intégré **I(1)**. Notre processus associé à la série SP 500 n'est donc pas stationnaire. Il correspond à un processus DS (Differency Stationary). Pour le stationnariser il faudra de même le différencier.

Effectuons à présent un test KPSS sur les modèles :

Test de stationnarité de KPSS

Le test KPSS ne peut être fait que dans le modèle M2 qui est le modèle le plus proche du M1 retenu précédemment dans les stratégies séquentielles. Dans ce test, l'hypothèse nulle est que la série est stationnaire autour d'une constante.

Le test KPSS est défini comme suit :

$$\begin{cases} H_0 : X_t \text{ stationnaire} \\ H_1 : X_t \text{ non stationnaire} \end{cases}$$

```
## Test KPSS pour la série "log_gold_prices"
kpssM2_gold_prices <- ur.kpss(y=lgold_prices,type="mu",lags="long")
show(kpssM2_gold_prices)
```

```
##
## #####
## # KPSS Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: 1.4528
print(kpssM2_gold_prices@cval)

##              10pct  5pct  2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Commentaire :

D'après les résultats du test de KPSS appliqué au modèle 2, avec un niveau de significativité fixé à 5%, on trouve $LM_{\text{obs}} = 1.4528$, tandis que la valeur critique à 5% est de 0.463. Donc $LM_{\text{obs}} > 0.463$. Comme la statistique de test est supérieure à cette valeur critique, rejet de l'hypothèse nulle de stationnarité au seuil 5%. Le processus est I(1).

Test KPSS pour la série "log_SP500"

```
kpssM2_SP500 <- ur.kpss(y=lSP500,type="mu",lags="long")
show(kpssM2_SP500)
```

```
##
## #####
## # KPSS Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: 1.2429
print(kpssM2_SP500@cval)

##              10pct  5pct  2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Commentaire :

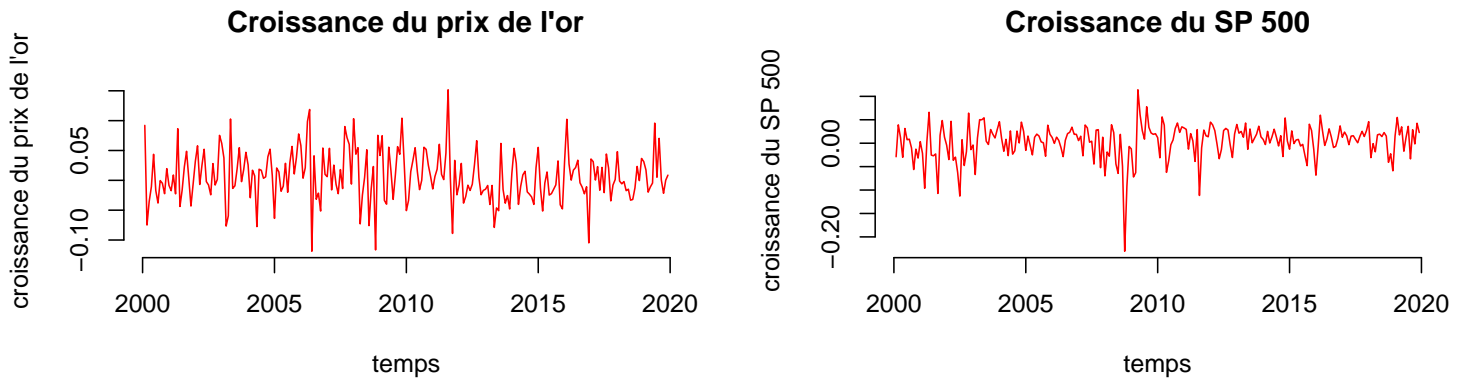
D'après les résultats du test de KPSS appliqué au modèle 2, avec un niveau de significativité fixé à 5%, on trouve $LM_{\text{obs}} = 1.2429$, tandis que la valeur critique à 5% est de 0.463. Donc $LM_{\text{obs}} > 0.463$. Comme la statistique de test est supérieure à cette valeur critique, rejet de l'hypothèse nulle de stationnarité au seuil 5%. Le processus est I(1).

Conclusion

Les résultats des tests ADF nous amènent à conclure que le le prix de l'or (en log) des Etats-Unis et le SP 500 (en log) sont des séries DS sans tendance ni constante. On dit aussi qu'elles sont des processus DS sans dérive. Ainsi, on doit les différencier pour les stationnariser car ce sont des processus DS.

Stationnarisation de nos deux séries"

```
dlgold_prices <- diff(lgold_prices)
par(mfrow = c(1, 2))
plot(dlgold_prices, main="Croissance du prix de l'or",frame=F,xlab="temps",ylab="croissance du prix de l'or",col="green")
d1SP500 <- diff(lSP500)
plot(d1SP500, main="Croissance du SP 500",frame=F,xlab="temps",ylab="croissance du SP 500",col="red")
```

Estimation de la série gold_prices, tests de validation et prévisions

L'analyse de séries temporelles est un outil central en économétrie pour comprendre et modéliser les fluctuations observées dans les données économiques et financières. Les modèles ARMA (AutoRegressive Moving Average) sont largement utilisés pour capturer les dynamiques temporelles, permettant d'expliquer les variations passées et de produire des prévisions sur des horizons courts.

Dans ce projet, nous décidons de chercher le modèle ARMA optimal d'une des deux séries : l'indice S&P 500 (dlSP500)

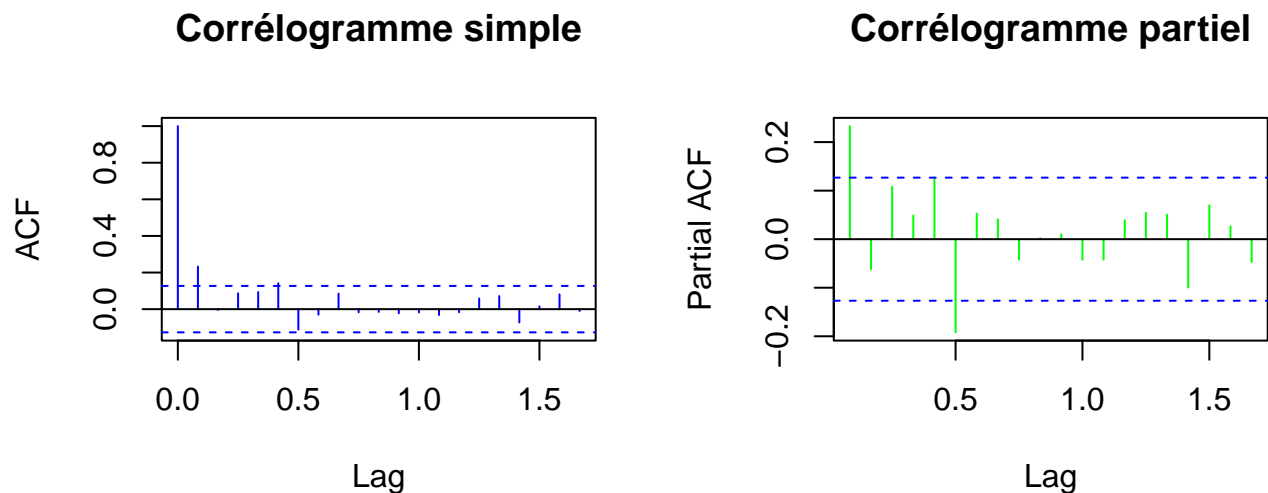
L'objectif de cette étude est de déterminer si le modèle ARMA peut efficacement capturer les variations observées dans cette série temporelle. Nous effectuerons des tests statistiques, tels que les tests de stationnarité et les tests sur les résidus, pour valider la qualité du modèle. Enfin, nous examinerons les performances prédictives (calcul des prévisions) de ce modèle sur un horizon de prévision de 1 à 3 mois.

Choix et estimation du modèle

Après que la série S&P 500 a été stationnarisée, cherchons le modèle ARMA satisfaisant les conditions de validité. La série initiale correspondant à l'indice correspond à un processus intégré d'ordre 1 DS. Ainsi, on va modéliser la variable en différence, qui correspond au processus filtré stationnaire associé.

Choix du modèle

```
par(mfrow = c(1, 2))
acf(dlSP500, lag.max=20, main="Corrélogramme simple",col="blue")
pacf(dlSP500, lag.max=20, main="Corrélogramme partiel",col="green")
```



En examinant attentivement les graphiques ACF et PACF, il est évident que le modèle le plus approprié pour décrire les dynamiques de ces données est un processus moyenne mobile, MA(1), puisque :

-> Sur l'ACF, le premier retard montre une corrélation significative, tandis que les corrélations suivantes diminuent rapidement et oscillent autour de zéro. Ce comportement est caractéristique d'un modèle MA(1) (Moving Average d'ordre 1).

-> Sur le PACF, les corrélations partielles décroissent également rapidement après le premier retard et oscillent ensuite autour de zéro. Cela confirme que le modèle le plus probable est un MA(1).

Ainsi, les caractéristiques des corrélogrammes indiquent que le modèle identifié est un MA(1). Si des tests supplémentaires ou des critères d'information (AIC, BIC) confirment ce choix, cela validerait le modèle.

Balayage des critères d'information

```
mat_aic <- matrix(0,5,5)
mat_bic <- matrix(0,5,5)
for (p in 0:4) {
  for (q in 0:4) {
    ARMA_pq <- Arima(y=dlSP500,order=c(p,0,q),method='ML')
    mat_aic[p+1,q+1] <- ARMA_pq$aic
    mat_bic[p+1,q+1] <- ARMA_pq$bic
  }
}
rownames(mat_aic) <- c('p=0','p=1','p=2','p=3','p=4')
colnames(mat_aic) <- c('q=0','q=1','q=2','q=3','q=4')
rownames(mat_bic) <- c('p=0','p=1','p=2','p=3','p=4')
colnames(mat_bic) <- c('q=0','q=1','q=2','q=3','q=4')

print(mat_aic)
```

```
##           q=0          q=1          q=2          q=3          q=4
## p=0 -897.3666 -910.3966 -909.0028 -909.7522 -907.8618
## p=1 -908.7004 -910.9878 -909.9701 -908.8619 -906.8893
## p=2 -907.6393 -910.2071 -909.7316 -907.7709 -905.9905
## p=3 -908.4836 -907.4526 -907.7617 -906.8368 -905.2468
## p=4 -907.0725 -905.5695 -905.9715 -908.7856 -914.6057
```

```
print(mat_bic)
```

```
##           q=0          q=1          q=2          q=3          q=4
## p=0 -890.4137 -899.9672 -895.0970 -892.3698 -887.0031
## p=1 -898.2710 -897.0819 -892.5878 -888.0031 -882.5541
## p=2 -893.7334 -892.8248 -888.8728 -883.4356 -878.1788
## p=3 -891.1013 -886.5938 -883.4264 -879.0251 -873.9587
## p=4 -886.2137 -881.2343 -878.1598 -877.4975 -879.8410
```

Commentaire :

On identifie un modèle MA(1) à partir du critère BIC, et nous choisissons de retenir celui qui minimise ce critère, tout en ayant un AIC parmi les plus faibles. Ce choix s'accorde également avec l'analyse des corrélogrammes, qui confirme que le modèle MA(1) est approprié.

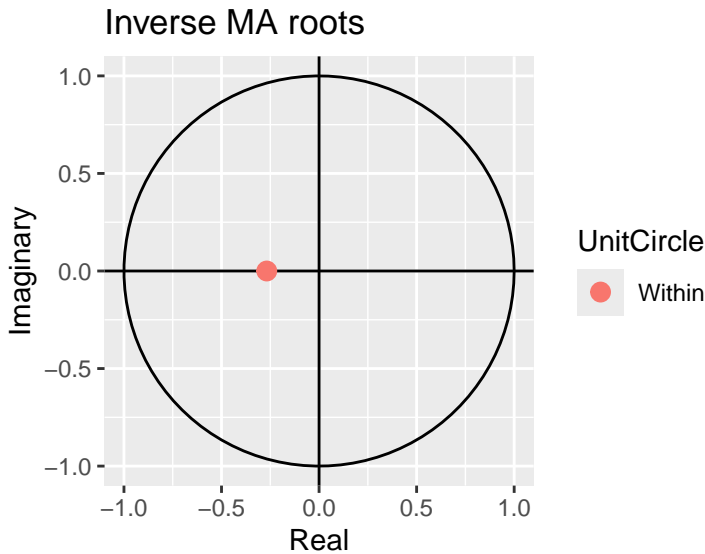
Estimation du modèle MA(1)

```
MA1 <- Arima(y = dlSP500, order = c(0, 0, 1))
summary(MA1)
```

```
## Series: dlSP500
## ARIMA(0,0,1) with non-zero mean
##
## Coefficients:
##           ma1      mean
##           0.2681  0.0033
## s.e.      0.0675  0.0029
##
## sigma^2 = 0.001276: log likelihood = 458.2
## AIC=-910.4 AICc=-910.29 BIC=-899.97
##
## Training set error measures:
##           ME          RMSE          MAE          MPE          MAPE          MASE
## Training set 4.168224e-05 0.03557037 0.02558477 72.61989 143.0148 0.6636679
##           ACF1
## Training set -0.01296435
```

Tests de validation du modèle

```
autoplot(MA1)
```



La commande autoplot reporte l'inverse des racines du polynome $\Phi(L)$. Il se situe bien à l'intérieur du disque unité complexe. Donc cela vérifie que le processus estimé est stationnaire.

Test de significativité sur les coefficients :

```
coeftest(MA1)
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ma1      0.2681123  0.0675275  3.9704 7.175e-05 ***
## intercept 0.0033090  0.0029159  1.1348  0.2565
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

On remarque que le coefficients θ_1 est significatif au seuil de 5%, tandis que la constance ne l'ai pas. Ainsi, on garde le coefficient ma1 et on retire la constante du modèle.

Rajoutons un terme AR(1) et un terme MA(1) pour vérifier qu'ils ne sont pas significatifs

```
ARMA11 <- Arima(y=dlSP500,order=c(1,0,1),include.constant=FALSE)
coeftest(ARMA11)
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1 -0.46647    0.15804 -2.9515  0.003162 **
## ma1  0.71500    0.12580  5.6838  1.317e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
MA2 <- Arima(y = dlSP500, order = c(0, 0, 2),include.constant=FALSE)
coeftest(MA2)
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ma1  0.271112  0.064617  4.1957 2.721e-05 ***
## ma2 -0.048429  0.068621 -0.7057  0.4803
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Commentaire : Le terme ar1 est significatif au seuil de 5% et le terme ma2 n'est pas significatif au seuil de 5%. Ainsi, dans ce qui suit nous allons retenir le modèle ARMA(1,1).

Test d'absence d'autocorrélation sur les résidus : Test de Ljung Box

Les hypothèses du test sont les suivantes :

$$\begin{cases} H_0 : \rho_1 = \rho_2 = \dots = \rho_H = 0 \\ H_1 : \exists j \in \{1, 2, \dots, H\} \text{ tel que } \rho_j \neq 0 \end{cases}$$

```
checkresiduals(ARMA11,lags=10, plot=FALSE)
```

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,1) with zero mean
## Q* = 22.645, df = 22, p-value = 0.422
##
## Model df: 2.    Total lags used: 24
```

On remarque que $p\text{ value} = 42,2\% > 5\%$. On ne rejette pas l'hypothèse nulle du test. Donc pas d'AR des résidus. Les résidus ne sont pas autocorrélés. Les autocorrélations ne sont pas significatives individuellement au seuil de 5%.

Test d'homoscedasticité sur les résidus : Test d'effets ARCH de Engle

Les hypothèses du test sont les suivantes :

$$\begin{cases} H_0 : \alpha_1 = \dots = \alpha_p = 0 \\ H_1 : \overline{H_0} \end{cases}$$

```
e <- ARMA11$residuals
ArchTest(e,lags=6)
```

```
##
##  ARCH LM-test; Null hypothesis: no ARCH effects
##
## data:  e
## Chi-squared = 18.721, df = 6, p-value = 0.004662
```

$$LM_{\text{obs}} = 18.721 > \chi_{6,0.95}^2 = 12.592 \text{ ou } p\text{-value} = 0.4662\% < 5\%$$

Donc on rejette l'hypothèse nulle du test. Ainsi, il y a donc des effets ARCH d'ordre 6 sur les résidus. Donc les résidus ne sont pas homoscedastiques. Nous allons tenter de les corriger dans la suite de notre modèle.

Test de normalité des résidus : Test de Jarque Bera

Les hypothèses du test sont les suivantes :

$$\begin{cases} H_0 : \text{Skewness} = 0 \text{ et Kurtosis} = 3 \Rightarrow \epsilon_t \text{ suit une loi normale} \\ H_1 : \overline{H_0} \end{cases}$$

```
jarque.bera.test(ARMA11$residuals)
```

```
##
##  Jarque Bera Test
##
## data:  ARMA11$residuals
## X-squared = 371.87, df = 2, p-value < 2.2e-16
```

$$JB_{\text{obs}} = 371.87 > \chi_{2,0.95}^2 = 5.99 \text{ ou } p\text{-value} \approx 0 < 5\%.$$

Donc les résidus ne sont pas normalement distribués.

Essayons de corriger la non-normalité des résidus en ajoutant des variables indicatrices. Deux approches seront testées.

Dans un premier temps, nous introduirons une variable indicatrice pour capturer les "outliers", définis comme les résidus situés dans les extrêmes de leur distribution. Cette méthode vise à corriger les problèmes potentiels d'hétéroscédasticité et de non-normalité.

Dans un second temps, nous tenterons de créer une indicatrice pour prendre en compte la crise financière de 2008.

```
#{r}
dates <- base1$period
#Méthode 1 : Identification des résidus aux extrémités
lower <- quantile(e, 0.05)
upper <- quantile(e, 0.95)
dum_outliers <- ifelse(e < lower | e > upper, 1, 0)
ARMA11bis <- Arima(y = dlSP500, order = c(1, 0, 1), xreg = dum_outliers)
# Vérification des coefficients
coeftest(ARMA11bis)
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1      -0.4886186  0.1665852 -2.9331  0.003355 **
```

```
## ma1      0.7072330  0.1354270  5.2222 1.768e-07 ***
## intercept 0.0046932  0.0027110  1.7312 0.083420 .
## xreg     -0.0138552  0.0074722 -1.8542 0.063707 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

ebis <- ARMA11bis$residuals
# Test de normalité (Jarque-Bera)
jarque.bera.test(ebis)

##
## Jarque Bera Test
##
## data:  ebis
## X-squared = 271.64, df = 2, p-value < 2.2e-16
# Test ARCH (détection d'hétéroscédasticité)
ArchTest(ebis, lags = 6)

##
## ARCH LM-test; Null hypothesis: no ARCH effects
##
## data:  ebis
## Chi-squared = 26.668, df = 6, p-value = 0.0001671
#Méthode 2 : Identification d'une période à exlure : (crise financière de 2008)
dum_bulle <- ifelse((dates >= "2007-10" & dates <= "2009-06"),1,0)
dum_bulle <- dum_bulle[-1]
ARMA11ter <- Arima(y = dlSP500, order = c(1, 0, 1), xreg = dum_bulle)
# Vérification des coefficients
coefTest(ARMA11ter)

##
## z test of coefficients:
##
##          Estimate Std. Error z value Pr(>|z|)
## ar1      -0.5231906  0.1478156 -3.5395 0.0004009 ***
## ma1      0.7346585  0.1165446  6.3037 2.907e-10 ***
## intercept 0.0056743  0.0026796  2.1176 0.0342090 *
## xreg     -0.0269330  0.0090141 -2.9879 0.0028093 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

eter <- ARMA11ter$residuals
# Test de normalité (Jarque-Bera)
jarque.bera.test(eter)

##
## Jarque Bera Test
##
## data:  eter
## X-squared = 261.84, df = 2, p-value < 2.2e-16
# Test ARCH (détection d'hétéroscédasticité)
ArchTest(eter, lags = 6)

##
## ARCH LM-test; Null hypothesis: no ARCH effects
##
## data:  eter
## Chi-squared = 29.433, df = 6, p-value = 5.037e-05
```

Malgré notre approche avec des variables indicatrices (qui s'avèrent significatives) en utilisant deux méthodes, nous ne parvenons pas à régler la non-normalité de nos résidus et les effets d'ARCH. Cela est typique des séries financières. Les résidus des séries financières ne suivent pas une distribution normale en raison de la présence de queues épaisses (fat tails), qui rendent les événements extrêmes plus probables. Ces résidus sont souvent asymétriques, reflétant des rendements non équilibrés entre gains et pertes. Enfin, les marchés sont affectés par des chocs soudains et irréguliers, éloignant les résidus de la normalité. Cela affectera notamment nos prévisions, car la non-normalité des résidus est une des hypothèses. Malgré cela nous allons passer à la prévision du modèle.

Prévision sur les horizons de 1 à 3 des prix du S&P500 (Modèle sans indicatrice)

Nous allons dans un premier temps prévoir le processus différencié (dlSP500) sur la base du modèle estimé précédemment, c'est à dire un ARMA(1,1).

Une fois les prévisions sur ce processus sont effectuées, nous pourrons dans un second temps construire les prévisions du modèle initial du S&P 500.

Prévisions du processus différencié

```
forecast_ARMA11 <- forecast(ARMA11, h=3, level=95)
forecast_ARMA11
```

##	Point Forecast	Lo 95	Hi 95
## Jan 2020	-0.007281905	-0.07712512	0.06256131
## Feb 2020	0.003396770	-0.06857114	0.07536468
## Mar 2020	-0.001584482	-0.07400645	0.07083749

À présent, calculons les prévisions à la main pour vérifier nos résultats : - Calcul de la prévision pour $h=1$. Nous avons le modèle suivant :

$$dlSP500_t = \phi \cdot dlSP500_{t-1} + \theta \cdot \varepsilon_{t-1} + \varepsilon_t$$

Avec les coefficients estimés par R :

$$\phi = -0.4665, \quad \theta = 0.7150$$

La prévision pour $h = 1$ est donnée par :

$$\hat{Y}_T(1) = E(dlSP500_{T+1} | I_T) = \phi \cdot dlSP500_T + \theta \cdot E(\varepsilon_T | I_T) + E(\varepsilon_{T+1} | I_T)$$

Sachant que $E(\varepsilon_{T+1} | I_T) = 0$ (l'erreur est centrée) que $E(\varepsilon_T | I_T) = \varepsilon_T$, nous avons :

$$\hat{Y}_T(1) = \phi \cdot dlSP500_T + \theta \cdot \varepsilon_T$$

Regardons les valeurs de $dlSP500_T$ en Décembre 2019 :

```
tail(dlSP500)
```

##	Jul	Aug	Sep	Oct	Nov
## 2019	0.036002054	-0.033468326	0.028798860	-0.001503661	0.042587101
##	Dec				
## 2019	0.022795533				

On trouve les valeurs suivantes :

$$dlSP500 \text{ (Décembre 2019)} = 0.022795533$$

Affichage des premières valeurs des résidus

```
residuals <- ARMA11$residuals
tail(residuals)
```

##	Jul	Aug	Sep	Oct	Nov	Dec
## 2019	0.02647081	-0.03560111	0.03864168	-0.01569864	0.05311018	0.00468738

En remplaçant les valeurs de $dlSP500_T$ dans la formule, on obtient la prévision pour Janvier 2020 :

$$\hat{Y}_{jan2020} = \hat{Y}_T(1) = -0.4665 \cdot 0.022795533 + 0.7150 \cdot 0.00468738 = -0.00728$$

La prévision pour $h = 2$ est donnée par :

$$\hat{Y}_T(2) = \phi \cdot \hat{Y}_T(1)$$

En remplaçant les valeurs :

$$\hat{Y}_{fev2020} = \hat{Y}_T(2) = -0.4665 \cdot -0.00728 = 0.003396$$

La prévision pour $h = 3$ est donnée par :

$$\hat{Y}_T(3) = \phi \cdot \hat{Y}_T(2)$$

En remplaçant les valeurs :

$$\hat{Y}_{mars2020} = \hat{Y}_T(3) = -0.4665 \cdot 0.003396 = -0.001584$$

Les prévisions du processus filtré calculées avec Forecast sont donc compatibles avec les prévisions du processus filtré calculées à la main.

Prévisions du processus différencié

Une fois les prévisions du processus filtré calculées, on peut construire sur cette base les prévisions du processus initial. Revenons à notre modèle non stationnaire, qui correspond à un processus DS, $X_T \sim I(1)$:

Le modèle DS non filtré est donné par :

$$(1 - L)X_t = Y_t \quad \Leftrightarrow \quad X_t = Y_t + X_{t-1}$$

Cela implique les relations suivantes pour les prévisions :

$$\hat{X}_T(1) = \hat{Y}_T(1) + X_T$$

$$\hat{X}_T(2) = \hat{Y}_T(2) + \hat{X}_T(1)$$

$$\hat{X}_T(3) = \hat{Y}_T(3) + \hat{X}_T(2)$$

Les prévisions filtrées sont :

$$\hat{Y}_T(1) = -0.00728, \quad \hat{Y}_T(2) = 0.003396, \quad \hat{Y}_T(3) = -0.001584$$

`tail(1SP500)`

```
##          Jul          Aug          Sep          Oct          Nov          Dec
## 2019 8.005071 7.971603 8.000402 7.998898 8.041485 8.064281
```

La valeur initiale non filtrée (Décembre 2019) est :

$$X_T = 8.064281$$

Calcul des prévisions non filtrées :

Pour $h = 1$:

$$\hat{X}_T(1) = \hat{Y}_T(1) + X_T$$

En remplaçant les valeurs :

$$\hat{X}_T(1) = -0.00728 + 8.064281 = 8.057001$$

Pour $h = 2$:

$$\hat{X}_T(2) = \hat{Y}_T(2) + \hat{X}_T(1)$$

En remplaçant les valeurs :

$$\hat{X}_T(2) = 0.003396 + 8.057001 = 8.060397$$

Pour $h = 3$:

$$\hat{X}_T(3) = \hat{Y}_T(3) + \hat{X}_T(2)$$

En remplaçant les valeurs :

$$\hat{X}_T(3) = -0.001584 + 8.060397 = 8.058813$$

Mois	Réels (X)	Prévus (X^\wedge)	Erreur ($ X - X^\wedge $)
Jan 2020	8,095	8,057	0,038
Fév 2020	8,096	8,060	0,036
Mar 2020	7,883	8,058	0,175

Sur les trois premiers mois de 2020, notre modèle ARMA produit des prévisions relativement proches des valeurs réelles pour janvier et février, avec des erreurs faibles. Cependant, en mars 2020, l'erreur augmente de manière significative. Cela pourrait s'expliquer par la survenue de la crise du COVID-19, un événement exogène et imprévisible à ce moment-là pour notre modèle. Pour mieux capturer l'effet de tels événements exogènes, nous pourrions envisager d'intégrer dans le modèle une indicatrice pour la crise Covid.

Modélisation multivariée

Alors que nous nous aventurons plus profondément dans les intrications de la finance et de l'économie, notre projet prend une tournure résolument multivariée. Nous vous invitons à explorer avec nous les interactions complexes et souvent subtiles entre deux piliers des marchés financiers : le prix de l'or et l'indice S&P 500.

Avant de poursuivre, clarifions un point essentiel : lorsque nous parlons des séries temporelles du prix de l'or et du S&P 500, nous faisons référence à des données ajustées pour garantir leur stationnarité, un processus crucial pour rendre leur moyenne et leur variance constantes dans le temps. Ce voyage analytique nous est facilité par l'élégante structure d'un modèle vectoriel autorégressif (VAR).

Imaginez ces deux séries temporelles comme des acteurs dans une danse économique complexe. Chaque mouvement, chaque variation est à la fois une réaction et une anticipation de l'autre. À travers le prisme du modèle VAR, nous allons décrypter cette relation dynamique, en examinant non seulement comment chaque série influence son propre avenir à travers ses retards, mais aussi comment elle prédit et est influencée par sa partenaire.

La sélection méticuleuse de l'ordre du modèle VAR, guidée par des critères statistiques éprouvés, ouvrira la voie à une analyse causale de Granger. Cette approche nous permettra de répondre à des questions essentielles telles que : « Le prix de l'or réagit-il en tant qu'actif refuge face aux variations du S&P 500 ? » ou encore, « Le S&P 500 reflète-t-il les anticipations des investisseurs liées à la volatilité de l'or ? »

Ensuite, grâce aux fonctions de réponse impulsionnelle, nous observerons les répercussions d'un choc inattendu dans une série sur l'autre, dessinant ainsi une cartographie des effets en cascade entre ces deux actifs. Et comme dans toute histoire financière, nous chercherons à comprendre les parts d'incertitude en analysant la décomposition de la variance de l'erreur de

prévision. Cela nous révélera la proportion de l'incertitude future du prix de l'or ou du S&P 500 pouvant être attribuée à des chocs passés de l'autre actif.

Rejoignez-nous dans cette enquête économétrique qui, nous l'espérons, enrichira non seulement notre compréhension mais aussi notre appréciation des relations dynamiques qui régissent les marchés financiers. À travers cette étude, nous espérons offrir des éclairages pratiques pour les gestionnaires de portefeuille, les investisseurs et tous ceux qui cherchent à naviguer dans les eaux complexes de l'économie mondiale.

Estimation du VAR

Dans cette partie, on ne considère que des processus stationnaires. L'estimation du VAR se fait équation par équation par la méthode des MCO ou par le maximum de vraisemblance.

Choix du nombre de retards avec les critères d'information

```
endogen <- ts.intersect(dlgold_prices, dlSP500) #concaténation sur une plage temporelle commune
pselec <- VARselect(y=endogen, lag.max=6, type='none')
pselec$criteria #affichage des valeurs des critères de sélection
```

```
##          1          2          3          4          5
## AIC(n) -1.300063e+01 -1.298378e+01 -1.296758e+01 -1.295433e+01 -1.295350e+01
## HQ(n)  -1.297674e+01 -1.293600e+01 -1.289590e+01 -1.285877e+01 -1.283405e+01
## SC(n)  -1.294139e+01 -1.286529e+01 -1.278984e+01 -1.271735e+01 -1.265727e+01
## FPE(n)  2.258904e-06  2.297305e-06  2.334874e-06  2.366076e-06  2.368172e-06
##          6
## AIC(n) -1.297062e+01
## HQ(n)  -1.282728e+01
## SC(n)  -1.261515e+01
## FPE(n)  2.328152e-06
```

pselec\$selection #nb de retards optimal selon les critères de sélection, on choisit l'ordre qui minimise les c

```
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      1      1      1      1
```

On remarque que : - le critère AIC est minimum pour un retard de 1, - le critère HQ est minimum pour un retard de 1, - le critère SC est minimum pour un retard de 1, - le critère FPE est minimum pour un retard de 1. Ainsi, on estime le modèle VAR avec p=1 retard.

Estimation du modèle VAR(1)

```
VAR1 <- VAR(y=endogen, lag.max=1, type='none')
coef(VAR1)
```

```
## $dlgold_prices
##           Estimate Std. Error  t value  Pr(>|t|)
## dlgold_prices.l1  0.1055809  0.06375130  1.656137  0.09902308
## dlSP500.l1        0.1600531  0.07345438  2.178946  0.03032548
##
## $dlSP500
##           Estimate Std. Error  t value  Pr(>|t|)
## dlgold_prices.l1 -0.002800748  0.05501162 -0.05091194  0.9594387528
## dlSP500.l1       0.239449080  0.06338450  3.77772279  0.0002004188
```

Vérifions que le VAR est bien stationnaire :

```
roots(VAR1, modulus=TRUE)
```

```
## [1] 0.2360123 0.1090177
```

Le module des inverses des racines du polynôme caractéristique vaut 0.236 et 0.109. Ces valeurs sont strictement inférieures à 1 ce qui signifie que notre modèle VAR est bien stationnaire.

Vérifions à présent que les résidus ne sont pas autocorrélés :

```
serial.test(VAR1, lags.pt=10, type="PT.adjusted")
```

```
##
## Portmanteau Test (adjusted)
##
## data: Residuals of VAR object VAR1
## Chi-squared = 47.346, df = 36, p-value = 0.0977
```

On remarque que la p-value = 0.0977 = 9.77% > 5%. Ainsi on ne rejette pas l'hypothèse nulle. Donc les résidus ne sont pas autocorrélés.

Commentaire :

Les résultats du modèle VAR pour les séries stationnarisées du prix de l'or différencié (dlgold_prices) et des variations du S&P 500 (dlSP500) révèlent des différences marquantes dans l'impact des retards propres à chaque série.

Dans l'équation de dlgold_prices, le coefficient du retard propre (dlgold_prices.l1) est relativement faible (0.10558) et marginalement significatif avec une p-value de 0.0990. Cela indique une légère persistance dans les variations du prix de l'or, mais cet effet reste modéré. En revanche, le coefficient de (dlSP500.l1) dans la même équation est plus élevé (0.16005) et statistiquement significatif (p-value = 0.0303). Cela suggère que les variations passées du S&P 500 ont un impact notable sur les variations du prix de l'or, reflétant une interdépendance entre les deux marchés.

Dans l'équation de dlSP500, le coefficient du retard propre (dlSP500.l1) est significatif avec une valeur élevée (0.239449) et une p-value très faible (0.0002), indiquant une forte persistance dans les variations du S&P 500. À l'inverse, le coefficient de (dlgold_prices.l1) est non significatif (p-value = 0.9594), ce qui montre que les variations passées du prix de l'or n'ont pas d'effet direct sur les variations actuelles du S&P 500.

La matrice de corrélation des résidus indique une faible corrélation négative (-0.1238) entre les deux séries, suggérant que les chocs spécifiques aux marchés de l'or et du S&P 500 sont principalement indépendants l'un de l'autre.

En somme, cette analyse montre que, bien que les variations du S&P 500 influencent significativement les variations du prix de l'or, l'inverse n'est pas vrai. De plus, chaque série conserve une dynamique interne distincte, avec une persistance plus marquée dans le S&P 500 que dans le prix de l'or. Ces résultats reflètent une interaction asymétrique entre les deux marchés financiers étudiés.

Tests de causalité au sens de Granger

Causalité du prix de l'or sur l'indice S&P 500 :

$$\text{Hypothèses du test d'absence de causalité : } \begin{cases} H_0 : a_{21} = 0 & \Rightarrow \text{dlgold_prices ne cause pas dlSP500,} \\ H_1 : a_{21} \neq 0 & \Rightarrow \text{dlgold_prices cause dlSP500.} \end{cases}$$

```
TC1 <- causality(VAR1, cause='dlgold_prices')
TC1$Granger
```

```
##
## Granger causality H0: dlgold_prices do not Granger-cause dlSP500
##
## data: VAR object VAR1
## F-Test = 0.002592, df1 = 1, df2 = 472, p-value = 0.9594
```

On remarque que p-value = 95.94% > 5%. On ne rejette donc pas l'hypothèse nulle au seuil 5% → Le prix de l'or ne cause pas l'indice S&P 500.

Causalité de l'indice S&P 500 sur le prix de l'or :

$$\text{Hypothèses du test d'absence de causalité : } \begin{cases} H_0 : a_{12} = 0 & \Rightarrow \text{dlSP500 ne cause pas dlgold_prices,} \\ H_1 : a_{12} \neq 0 & \Rightarrow \text{dlSP500 cause dlgold_prices.} \end{cases}$$

```
TC2 <- causality(VAR1, cause='dlSP500')
TC2$Granger
```

```
##
## Granger causality H0: dlSP500 do not Granger-cause dlgold_prices
##
## data: VAR object VAR1
## F-Test = 4.7478, df1 = 1, df2 = 472, p-value = 0.02983
```

On remarque que p-value = 2.933% < 5%. On rejette donc l'hypothèse nulle. → L'indice S&P 500 cause le prix de l'or.

Conclusion générale : Les tests de causalité de Granger révèlent une relation dynamique asymétrique entre les variations logarithmiques de l'indice S&P 500 (dlSP500) et les variations logarithmiques des prix de l'or (dlgold_prices). Les résultats indiquent que les fluctuations des prix de l'or n'ont pas d'influence significative sur les variations du S&P 500, confirmant l'absence de causalité dans ce sens. En revanche, les fluctuations du S&P 500 exercent une influence significative sur les variations des prix de l'or, suggérant que les mouvements de l'indice S&P 500 jouent un rôle dans la détermination des variations des prix de l'or. Cela met en évidence une relation unidirectionnelle, où le S&P 500 impacte le marché de l'or, mais pas l'inverse.

Fonctions impulsion-réponse

Les fonctions de réponse à l'impulsion, représentées dans les graphiques, permettent d'analyser l'impact d'un choc standardisé dans une série temporelle sur elle-même ainsi que sur l'autre série au fil du temps. Voici l'analyse pour chaque graphique en prenant en compte les deux variables, dlgold_prices et dlSP500, qui représentent respectivement les variations différenciées des prix de l'or et des variations du S&P 500 avant l'événement étudié. Ces réponses permettent de capturer les interactions dynamiques entre les deux marchés financiers et d'identifier la manière dont un choc affecte chaque variable individuellement et mutuellement au fil du temps.

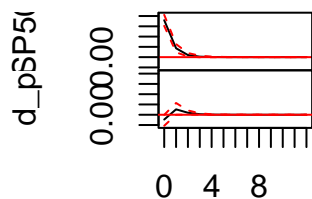
Justification du choix de l'ordre des variables dans la décomposition de Cholesky : D'après le cours, l'ordre des variables dans la décomposition de Choleski doit refléter leur degré d'indépendance causale, avec la variable la plus indépendante en premier et la plus dépendante en dernier. Dans notre cas, les tests de causalité ont montré que l'indice S&P 500 (dlSP500) n'est pas causé par les variations des prix de l'or (dlgold_prices), tandis que les prix de l'or sont influencés par le S&P 500. Par conséquent, dlSP500 est positionné en premier dans la décomposition, car il agit comme une variable exogène, tandis que dlgold_prices est placé en second, reflétant sa dépendance aux variations du S&P 500. Ce choix est également justifié économiquement, car le marché des actions influence souvent les marchés des actifs refuges comme l'or, mais l'inverse est moins fréquent.

```
endogen <- ts.intersect(dlSP500, dlgold_prices)
VAR1 <- VAR(y=endogen, type='none', lag.max=1)
```

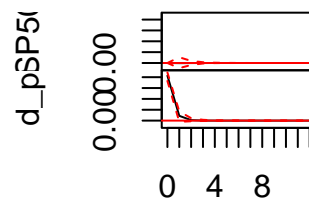
Calcul classique des fonctions de réponse (itération sur le VAR estimé)

```
irf <- irf(VAR1, n.ahead = 12, ortho = TRUE, ci=0.95)
plot(irf)
```

Orthogonal Impulse Response from dlSP500 Orthogonal Impulse Response from dlgold_prices



95 % Bootstrap CI, 100 runs



95 % Bootstrap CI, 100 runs

Remarque : Les fonctions de réponse convergent vers 0, signe que le VAR est bien stationnaire.

Analyse des fonctions de réponse avec la méthode classique (méthode du VAR)

Analyse du premier graphique (Orthogonal Impulse Response from dlSP500) :

1. Réaction de dlSP500 à un choc dans dlSP500 : Le graphique montre qu'un choc dans dlSP500 entraîne une réponse initiale positive, qui décroît progressivement pour atteindre zéro à moyen terme. Cela signifie que les variations du S&P 500 sont autocorrélées à court terme mais n'ont pas de persistance significative à long terme. La dynamique positive initiale reflète l'impact immédiat d'un choc, probablement lié à la structure du marché du S&P 500, mais cet effet disparaît rapidement.
2. Réaction de dlgold_prices à un choc dans dlSP500 : Un choc dans dlSP500 provoque une réponse initialement négative sur dlgold_prices. Cela indique que les variations du S&P 500 peuvent temporairement réduire les prix de l'or, ce qui pourrait refléter une dynamique de fuite des investisseurs des actifs refuges comme l'or vers des actifs plus risqués (comme les actions du S&P 500). Cependant, cette influence négative est temporaire, car la réponse retourne rapidement à zéro, indiquant que l'effet du S&P 500 sur l'or est transitoire et limité.

Analyse du deuxième graphique (Orthogonal Impulse Response from dlgold_prices) :

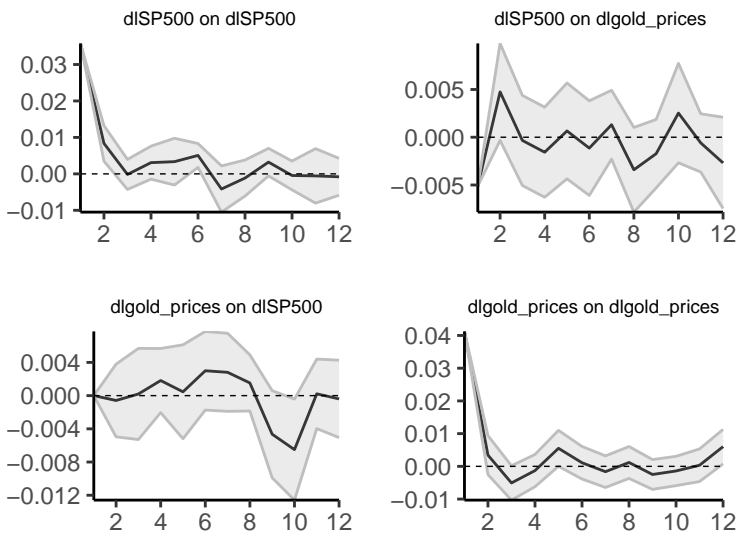
1. Réaction de dlSP500 à un choc dans dlgold_prices : La réponse de dlSP500 à un choc dans dlgold_prices est négligeable, avec des intervalles de confiance qui incluent constamment la ligne zéro. Cela suggère qu'un choc dans les prix de l'or n'a pas d'effet direct ni significatif sur les variations du S&P 500. Cette indépendance pourrait être liée à la nature différente des deux marchés (l'or étant souvent considéré comme un actif refuge, tandis que le S&P 500 représente un marché d'actions plus risqué).
2. Réaction de dlgold_prices à un choc dans dlgold_prices : Un choc dans dlgold_prices génère une réponse initialement positive et significative, qui diminue rapidement avant de s'annuler à long terme. Cela reflète une autocorrélation à court terme dans les variations des prix de l'or, mais avec une dynamique de retour à la moyenne. Cette réponse montre que les ajustements du marché de l'or sont rapides et que l'impact des chocs ne persiste pas.

Conclusion générale : Les tests de causalité ont montré que le S&P 500 cause les prix de l'or, une conclusion appuyée par les graphiques des fonctions de réponse. La réponse initiale négative des prix de l'or à un choc dans le S&P 500 indique qu'une hausse du S&P 500 peut entraîner une baisse temporaire des prix de l'or. Cette dynamique est probablement liée à des mouvements d'investissements entre actifs risqués (comme les actions du S&P 500) et refuges (comme l'or). Cependant, cet effet est transitoire, et la réponse des prix de l'or revient à zéro à moyen et long terme. Cette absence d'effet persistant soulève une contradiction apparente entre les résultats des tests de Granger et les IRF. Alors que les tests de Granger détectent une causalité dynamique globale sur l'ensemble des données, les IRF montrent que l'impact des chocs isolés est temporaire. Cette différence peut s'expliquer par les horizons temporels différents analysés par chaque méthode : les tests de Granger évaluent les relations sur une fenêtre temporelle étendue, tandis que les IRF se concentrent sur des réponses immédiates et localisées. Les fonctions de réponse révèlent également une asymétrie dans les relations entre ces deux marchés. Les variations différenciées des prix de l'or n'ont pas d'effet significatif sur le S&P 500, confirmant l'indépendance relative du

marché des actions par rapport à celui de l'or. En revanche, les variations du S&P 500 influencent temporairement les prix de l'or, soulignant une influence unidirectionnelle et limitée.

Calcul des fonctions de réponse avec des projections locales

```
irf_LP <- lp_lin(endog_data = as.data.frame(endogen),
  lags_endog_lin = NaN,
  lags_criterion = 'BIC',
  max_lags = 6,
  trend = 0,
  shock_type = 0,
  confint = 1.96,
  hor = 12,
  adjust_se=TRUE)
plot(irf_LP)
```



Analyse des fonctions de réponse avec la méthode des projections locales

Les graphiques présentés sont des estimations de la fonction de réponse impulsionnelle (IRF) pour deux séries temporelles différenciées : dLSP500 (variations différenciées du S&P 500) et dLgold_prices (variations différenciées des prix de l'or).

Réponse de dLSP500 à un choc dans dLSP500 : Ce graphique montre la réaction de dLSP500 à un choc en lui-même. La réponse initiale est positive et forte, mais elle diminue rapidement au fil du temps pour osciller autour de zéro avant de s'éteindre. Cela indique une forte autocorrélation à court terme dans les variations différenciées du S&P 500, mais cet effet est transitoire et la série revient à sa moyenne à long terme.

Réponse de dLgold_prices à un choc dans dLSP500 : Un choc dans dLSP500 a un impact initial positif sur dLgold_prices, ce qui est cohérent avec les résultats des tests de causalité de Granger indiquant que dLSP500 cause dLgold_prices. Cependant, cet effet décroît progressivement pour devenir statistiquement non significatif à long terme. Cela suggère que les variations différenciées du S&P 500 influencent temporairement les variations différenciées des prix de l'or, probablement en raison de facteurs liés à l'interconnexion des marchés financiers, mais cet impact ne persiste pas sur le long terme.

Réponse de dLSP500 à un choc dans dLgold_prices : Ce graphique montre que la réaction de dLSP500 à un choc dans dLgold_prices est initialement proche de zéro et reste non significative tout au long de l'horizon temporel. Cela indique que les variations différenciées des prix de l'or n'ont pas d'effet notable sur les variations différenciées du S&P 500, confirmant l'indépendance dynamique entre ces deux séries.

Réponse de dLgold_prices à un choc dans dLgold_prices : La réponse de dLgold_prices à un choc en lui-même est initialement forte et positive, mais décroît rapidement pour osciller autour de zéro. Cela suggère que les variations différenciées des prix de l'or présentent une autocorrélation notable à court terme, mais avec une dynamique de retour à la moyenne qui limite l'impact des chocs à long terme.

Conclusion générale : Les fonctions de réponse impulsionnelle, combinées aux résultats des tests de causalité de Granger, montrent une relation asymétrique entre les deux séries. Les variations différenciées du S&P 500 (dLSP500) influencent significativement les variations différenciées des prix de l'or (dLgold_prices) à court terme, comme en témoigne une réponse initiale positive. Cependant, l'effet est transitoire et s'atténue rapidement. En revanche, il n'y a pas de preuve que les variations différenciées des prix de l'or influencent les variations différenciées du S&P 500. Ces résultats renforcent l'idée que le S&P 500 agit comme un moteur de dynamique sur les marchés financiers, tandis que le marché de l'or répond davantage à des impulsions externes sans affecter directement les indices boursiers.

Comparaison des conclusions : Les deux méthodes (classique et projections locales) donnent des conclusions similaires : - Une forte autocorrélation transitoire dans dLSP500 et dLgold_prices. - dLSP500 influence dLgold_prices de manière limitée et transitoire, mais pas l'inverse. - Les chocs dans dLgold_prices ou dLSP500 n'ont pas d'effet durable ou significatif sur l'autre série.

Cointégration de Johansen

Dans le cadre de notre analyse économétrique, nous avons déterminé que les séries étudiées, à savoir les prix de l'or (lgold_prices) et l'indice S&P 500 (lSP500), sont toutes deux intégrées d'ordre un (I(1)). Cela signifie qu'elles ne sont pas stationnaires dans leurs niveaux, mais qu'elles deviennent stationnaires après différenciation première, ce qui donne naissance aux séries différenciées (dlgold_prices et dlSP500), ce que nous avons montré précédemment. Cette caractéristique commune ouvre la voie à l'application potentielle des techniques de cointégration, qui reposent sur l'existence d'une relation de long terme entre des séries temporelles intégrées.

La cointégration est une méthode particulièrement utile lorsque des séries I(1) présentent une tendance commune sur le long terme, malgré des fluctuations à court terme. Cette tendance partagée permet de modéliser les séries à travers des modèles à correction d'erreur, qui intègrent à la fois les dynamiques de court terme et la convergence vers un équilibre de long terme. Dans notre cas, l'existence de cette relation pourrait être vérifiée à travers des tests formels de cointégration, tels que le test de Johansen ou celui d'Engle-Granger.

Si une relation de cointégration est établie entre lgold_prices et lSP500, cela indiquerait l'existence d'une tendance commune sous-jacente, malgré des variations distinctes à court terme. Cette relation permettrait d'utiliser des modèles à correction d'erreur, capables de capter les ajustements de court terme tout en respectant une convergence vers un équilibre de long terme. Dans notre cas, une relation de cointégration pourrait refléter une interaction structurelle entre les marchés boursiers et les prix de l'or, par exemple via des arbitrages entre actifs risqués (actions) et actifs refuges (or). En revanche, si aucune cointégration n'est détectée, cela signifierait que ces deux marchés évoluent de manière largement indépendante sur le long terme, bien qu'ils puissent être liés par des chocs transitoires ou des corrélations de court terme.

Ainsi, l'identification ou l'absence de cointégration guidera la suite de l'analyse : si une relation de cointégration est confirmée, des modèles à correction d'erreur seront nécessaires pour modéliser la dynamique des séries. À défaut, des modèles VAR différenciés pourront être employés pour capturer les interactions entre lgold_prices et dlSP500. Cette distinction souligne l'importance d'une vérification préalable de la cointégration avant de choisir une méthodologie appropriée.

Choix du retard p dans le VAR en niveau :

```
endogen_co <- ts.intersect(lSP500 , lgold_prices)
pselec <- VARselect(endogen_co, lag.max = 6, type = "none")
pselec$selection # On prend 2 retards
```

```
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      2      2      1      2
pniv <- 2
```

Sur la base des critères AIC, HQ et FPE, on retient 2 retards dans le VAR en niveau, soit 1 retard dans le VECM.

Test de cointégration de Johansen :

Débutons la stratégie séquentielle :

Hypothèses du test de cointégration : $\begin{cases} H_0 : r = 0 & \Rightarrow \text{Pas de relation de cointégration entre les séries,} \\ H_1 : r > 0 & \Rightarrow \text{Existence d'une ou plusieurs relations de cointégration.} \end{cases}$

```
test_coint <- ca.jo(endogen_co, ecdet = 'none', spec = 'transitory', type = 'trace', K = 2)
summary(test_coint)
```

```
##
## #####
## # Johansen-Procedure #
## #####
##
## Test type: trace statistic , with linear trend
##
## Eigenvalues (lambda):
## [1] 0.019093324 0.008077327
##
## Values of teststatistic and critical values of test:
##
##          test 10pct  5pct  1pct
## r <= 1 |  1.93   6.50   8.18 11.65
## r = 0  |  6.52 15.66 17.95 23.52
##
## Eigenvectors, normalised to first column:
## (These are the cointegration relations)
##
##                lSP500.l1 lgold_prices.l1
## lSP500.l1         1.000000      1.000000
## lgold_prices.l1 -5.238593      -0.2494737
##
## Weights W:
## (This is the loading matrix)
```

```
##
##               lSP500.l1 lgold_prices.l1
## lSP500.d      -0.001372977   -0.006085228
## lgold_prices.d 0.001211673   -0.009888465
```

On remarque que $TR^* = 6.52$ et $VC = 17.95$. Donc comme $TR^* < VC$, on ne rejette pas l'hypothèse nulle au seuil de 5%. Ainsi $r=0$ (le rang de cointégration est nul) Comme nous ne pouvons pas rejeter l'hypothèse nulle, cela indique qu'il n'y a pas de relation de cointégration entre les séries. La stratégie séquentielle s'arrête ici.

Conclusion : D'après le test de Johansen, il n'existe aucune relation de cointégration entre les séries analysées (lSP500 et lgold_prices). Les séries évoluent donc indépendamment sur le long terme.

Validation du modèle VAR :

Nos séries lSP500 et lgold_prices ont été confirmées comme I(1) (intégrées d'ordre 1). Le test de cointégration de Johansen a montré qu'il n'existe aucune relation de cointégration.. Cela signifie qu'il n'y a pas de relation de long terme entre lSP500 et lgold_prices, comme spécifié dans nos tests précédents. Dans ce cas, le modèle VAR en différences premières est approprié. Ce modèle analyse les relations dynamiques entre les variations des séries entre dlSP500 et dlgold_prices.

CONCLUSION :

À l'issue de notre exploration multivariée, nous avons pu approfondir la compréhension des dynamiques entre l'indice S&P 500 et les prix de l'or à travers diverses méthodes économétriques.

Les tests de causalité de Granger ont révélé que l'indice S&P 500 cause les variations des prix de l'or, indiquant une relation unidirectionnelle où les fluctuations du marché boursier américain influencent temporairement le marché de l'or. Toutefois, l'absence de relation de causalité inverse suggère une indépendance relative des variations des prix de l'or sur les mouvements futurs du S&P 500.

Les fonctions de réponse impulsionnelle ont confirmé l'existence d'une influence dynamique mais transitoire entre les deux séries. En particulier, un choc dans le S&P 500 provoque une réaction initiale dans les prix de l'or, bien que cet effet disparaisse à long terme. Cette dynamique peut être attribuée à une réallocation des capitaux entre actifs risqués (actions) et refuges (or), mais sans impact persistant.

En ce qui concerne la cointégration, le test de Johansen n'a pas mis en évidence de relation de long terme entre les deux séries. Cela justifie l'utilisation d'un modèle VAR en différences premières, qui se concentre sur les relations dynamiques à court termes.

En conclusion, cette étude multivariée met en lumière des interactions dynamiques et asymétriques entre l'indice S&P 500 et les prix de l'or. Bien qu'aucune relation de long terme n'ait été établie, les résultats indiquent une influence significative à court terme du S&P 500 sur l'or, ce qui ouvre la voie à de futures recherches intégrant des approches non linéaires ou des variables additionnelles pour mieux capturer la complexité des marchés financiers et des matières premières.

Les hypothèses initiales sont partiellement vérifiées. Si l'influence unidirectionnelle du S&P 500 sur le prix de l'or est bien établie, l'effet refuge supposé de l'or en réponse à une baisse du S&P 500 n'est pas systématique ni persistant. L'absence d'impact significatif du prix de l'or sur le S&P 500 réfute une anticipation de risques traduite par le prix de l'or. Enfin, la relation globale entre les deux variables est dynamique mais non toujours négative, et semble fortement dépendre du contexte économique. Ces résultats éclairent les limites des hypothèses initiales et soulignent l'importance de facteurs contextuels dans les relations entre l'or et le marché boursier.