EMH :

Definition : Un marché est dit efficient si les prix contiennent toute l'information disponible à un instant donné. Dans ce contexte, toute analyse technique (étude statistique inférentielle des prix passés) ou fondamentale (étude des informations spécifiques à l'entreprise) est futile. Selon cette théorie, chaque nouvelle information se reflète instantanement sur les prix. Ces derniers se suffisent donc à eux-mêmes et sont un parfait indicateur de la valeur intrinsèque de l'actif. C'est pourquoi tenter d'obtenir des renseignements via l'analyse de l'information liée aux faits de l'entreprise est inutile, puisque que cette information est immédiatement reflétée dans le prix. L'analyse technique est elle aussi vaine puisque les éventuelles indications fournies par les prix passés sont elles aussi déjà diffusées dans le prix actuel.

Cette idée va de pair avec la notion de marche aléatoire (random walk). La marché aléatoire est un processus stochastique dans lequel l'état de la variable aléatoire est indépendante des valeurs antérieures. En effet, si le prix contient toute l'information disponible, le prix futur ne dépend que des informations qui seront disponibles dans le futur. A l'instant T il est impossible de prévoir les informations futures, il est donc impossible de prédire les prix en T+1. De ce fait, le prix aujourd'hui ne donnent aucune indication sur les prix demain, qui résultent alors d'un processus complètement aléatoire et décorrélés des valeurs passées.

Si l'EMH est respectée, les prix suivent un processus de marche aléatoire. Vérifier que les prix soient bel et bien dictés par une marche aléatoire, permet de vérifier empiriquement la véracité de l'EMH. Hypothèse sans cesse discutée et remise en cause. Bien que son excatitude soit de plus en plus critiquée depuis les années 1980. Cette idée est née dans le désormais célèbre article de Fama : *Efficient Capital Markets: A Review of Theory and Empirical Work - Eugene F. Fama The Journal of Finance Vol. 25, No. 2*

De plus, cette hypothèse soutient l'idée qu'il est « impossible de battre le marché ». Cette expression est utilisée pour signifier qu'il est impossible de réaliser des rendements supérieurs à ceux obtenus grâce à un portfolio parfaitement diversifié*. Dans un scénario d'EMH et donc de marché aléatoire, un trader expérimenté ne dispose pas de plus d'informations quant à la future évolution des prix qu'un investisseur lambda. La solution optimale en terme de ratio rendement/risque consiste donc à investir dans un portfolio diversifié.

* parfaitement diversifié : portfolio dont les risques indiosyncratiques ont été éliminés grâce à la diversification. Le portfolio a alors la variance minimale pour un rendement donné. Ou à l'inverse, un rendement maximal pour une variace donnée (portfolio Markowitz). Voir *The Journal of Finance, Vol. 7, No. 1. (Mar., 1952), pp. 77-91. - Harry Markowitz*

Ici, nous tenterons donc de déterminer si les prix suivent un processus de marche aleatoire. Notre conclusion nous permetteras de savoir si l'EMH peut être envisagée. Nous nous intéresserons à l'indice Standards & Poors 500 (S&P 500), basé sur 500 entreprises américaines. Le S&P 500 est un des indices les plus suivis au monde, il est considéré comme représentatif du marché financier américain. C'est pourquoi nous avons choisi de travailler sur cet actif plutôt qu'un autre.

Cependant, nous ferons face à un autre problème, également formulé par Fama en 1991 dans Efficient Capital Markets: II. The Journal of Finance, 46(5), pp. 1575. Ce problème, connu sous le nom de « Joint Hypothesis Problem », implique que la vérification empirique de l'EMH est impossible au sens strict. Ceci est du au fait que même après avoir observé un actif ne se comportant pas comme il le devrait sous l'EMH, il est possible que le marché soit efficient. Ce serait alors la définition même d'efficience qui serait biaisée. La divergence entre le théorie et les observations empiriques serati attribuable à une mauvaise formulation de ce qu'est un marché efficient et de quels mécanismes sous jacents sont impliqués plutôt qu'à l'inefficience des marchés. Une des méthodes permettent d'étudier la validité empirique de l'EMH consiste à observer s'il est possible de réaliser (de manière constante) des rendements supérieurs à ceux offerts par le marché. Si tel est le cas, on peut penser qu'il est possible de tirer profit de la mauvaise évaluation des prix des actifs, autrement dit, que les prix ne reflètent pas correctement l'information disponible. Cependant pour être observés, ces rendements anormaux doivent être comparés aux rendements théoriques, et mesurés par des modèles de prix (le modèle de Black-Scholes par exemple). Dans ces circonstances, l'observation de rendements anormaux peut être attribuable à une inefficience des marchés, mais également à une mauvaise spécification du modèle de prix. Le marché serait en réalité efficient et ces rendements anormaux seraient le résultat d'une mauvaise évaluation de la valeur réelle de l'actif. En conclusion, l'observation de rendements anormaux ne suiffit pas à prouver l'échec de l'EMH, car ils peuvent résulter d'une erreur dans le modèle de prix utilisé comme référence. Cet exemple montre les deux dimensions du problème ; determiner si les actifs se comportent comme la théorie le préconnise et d'autre part, s'assurer que les critères théoriques sont correctement formulés. De ce fait le problème des hypothèses jointes nous empeche de vérifier la validité empirique de l'EMH. C'est d'ailleurs ce pourquoi la question persiste encore aujourd'hui dans la littérature.

Ici, nous ne nous intéresserons pas à l'étude de rendements anormaux, en revanche nous nous concentrerons sur l'autre implication de l'EMH ; que les prix suivent un processus de marche aléatoire. Nous tenterons de détecter de la persistence dans les mouvements de prix (i.e. des autocorrelations significatives entre les prix à des dates différentes),. Le cas écheant, l'analyse des prix passés offre de l'information sur les prix futurs, ce qui rentre en contradiction directe avec l'EMH. Cependant, nous l'avons vu, cela ne constitue pas un critère de rejet de l'EMH dans sa totalité.

La marche aléatoire se présente de manière formelle sous cette forme :

$$X_t = X_{t-1} + \varepsilon_t, \, avec \, \varepsilon_t \, un \, bruit \, blanc \sim N(0,1)$$

$$X_t = X_0 + \sum_{t=1}^{t} \varepsilon_t$$

Il s'agit en fait d'un cas particulier d'un processus AR(1) :

$$X_t = \phi X_{t-1} + \varepsilon_t$$

$$-|\phi|<1 : processus \, AR \, stationaire$$
$$-|\phi|=1 : marche \, aléatoire$$
$$-|\phi|>1 : processus \, AR \, explosif$$

Nous tenterons donc de représenter notre série sous la forme d'un processus autoregressif avec un coefficient Φ strictement inférieur à 1.

Cointégration :

Par la suite nous tenterons une approche explicative. Nous essayerons de determiner la relation de long terme entre les taux d'interet de la Federal Reserve (FED) et le cours du SP500, ainsi que la relation entre l'index de volatilité du Chicago Board Option Exchange (CBOE) et le SP500.

Le lien entre taux d'interêt et économie réelle est expliqué par les modèles macroéconomiques type IS/LM. Si les taux d'interêt augmentent, il devient plus couteux d'investir et on peut s'attendre à un ralentissement de l'économie. A l'inverse, une baisse du taux d'interêt incite les ménages et les entreprises à emprunter, puisque le cout de l'emprunt diminue. Les ménages peuvent accroitre leur consommation tandis que les entreprises peuvent investir ; renouveler leur capital, s'engager dans des projets de recherche et développement, étendre leur locaux etc...Les modèles basiques suggèrent donc que l'économie et les taux d'interêts sont négativement corrélés entre eux. Ici, notre indicateur de la santé économique sera le SP500, nous ne nous intéresserons donc uniquement à une partie de la théorie qui lie taux d'interêt et économie. Notre approche sera centrée sur les entreprises constituant un échantillon représentatif de l'économie américaine. Cependant, la littérature indique également qu'une hausse de l'investissement se répercute sur les dividendes, qui décroissent avec la proportion du profit allouée à l'investissement. Le cours des actions des entreprises dont les dividendes diminuent à également tendance à diminuer. Cette relation négative entre investissement et cours de l'action provient de la théorie financière, notamment du modèle de Gordon-Shapiro (Dividend Discount Model, ou DDM). *The Review of Economics and Statistics, Vol. 41, No. 2, Part 1 (May, 1959), pp. 99-105* and *Capital Equipment Analysis: The Required Rate of Profit Myron J. Gordon and Eli Shapiro Management Science Vol. 3, No. 1 (Oct., 1956), pp. 102-110.* Les rendements d'une action proviennent de deux sources différentes ; la plus-value (éventuellement) réalisée à la revente à maturité, et les dividendes versés durant la période. Les dividendes sont déterminés par l'entreprise qui choisi d'allouer une partie des bénéfices réalisés au paiement des dividendes, le reste allant à l'investissement. Naturellement si le ratio alloué à l'investissement augmente, le ratio destiné au paiement des dividendes diminue et le rendement de l'action également, toutes choses égales par ailleurs. Une diminution des taux d'interêts, entrainant une augmentation de l'investissement peut donc avoir une incidence sur le Dividend Payout Ratio (DDR) et faire baisser le rendement de l'action. Gordon-Shapiro nous dit que le prix de l'action dépend positivement des valeurs des dividendes futures, selon notre raisonnement, le cours de l'action devrait être négativement corrélé avec les taux d'interêts.

Le CBOE met à disposition un index de volatilité implicite (implied volatility). La volatilité implicite est une mesure de la volatilité d'un actif dérivée d'un modèle de prix. Elle n'est pas identique à la volatilité historique, calculée sur les données passées. A l'inverse, la volatilité implicite est obtenue à partir du prix d'un produit dérivé, notamment d'une option. Le modèle de Black-Scholes permet d'évaluer le prix d'une option à partir de plusieurs caractéristiques, le temps restant jusqu'à maturité, le prix d'exercice de l'option, la valeur actuelle du sous-jacent, le taux sans risque et la volatilité du sous-jacent.

$$\text{Le prix d'un call} : C(S_{0,} K, r, T, \sigma) = S_0 N(d_1) - Ke^{rt} N(d_2)$$

$$avec\ N\text{ , }la\ fonction\ de\ repartition\ de\ la\ Loi\ Normale\text{ , }à\ savoir : \int_{-\infty}^{x} \left(\frac{1}{2\sqrt{\pi}} e^{\left(\frac{1}{2}x^2\right)} dx \right)$$

$$S_0 : valeur\ actuelle\ du\ sous-jacent$$

$$T : temps\ restant\ jusqu'\ à\ maturité$$

$$K : prix\ d'\ exercice\ du\ Call$$

$$r : taux\ \text{sans risque}$$

$$\sigma : volatilité\ du\ sous-jacent$$

Et ;

$$d_1 = \frac{1}{\sigma\sqrt{T}}[\ln(\frac{S_0}{K}) + (r + \frac{1}{2}\sigma^2)T]$$
$$d_2 = d_1 - \sigma\sqrt{T}$$

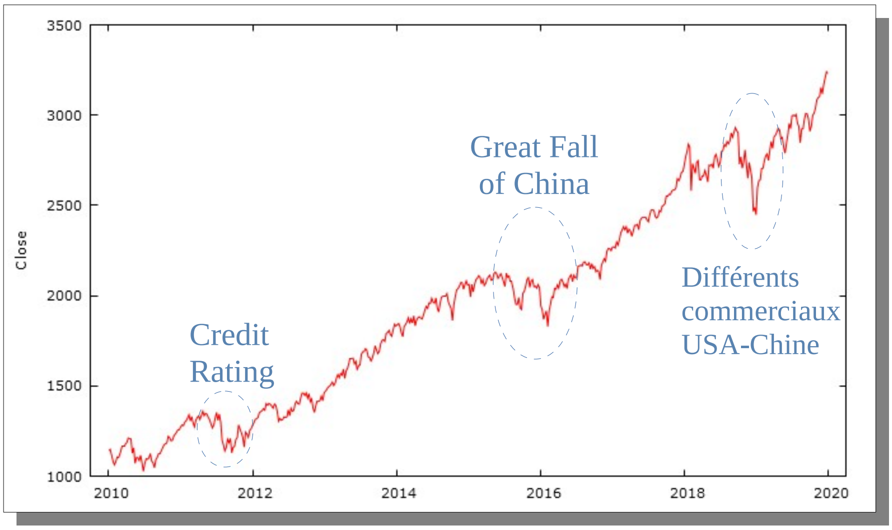
Cette volatilité peut être estimée de plusieurs façon. En revanche, en observant à posteriori le prix d'une action, il est possible d'utiliser la formule de Black-Scholes pour en tirer une estimation de la volatilité du sous-jacent. Le prix de l'action étant décidé sur les marchés financiers, la volatilité dérivée de la formule de Black-Scholes donne une mesure de la volatilité selon le marché. C'est pourquoi elle est dite « implicite », elle permet de constater le sentiment du marché sur la volatilité d'un actif, en utilisant le prix d'une option ayant cet actif pour sous-jacent.

Le VIX, l'index de volatilité fourni par le CBOE, mesure la volatilité implicite du SP500 en utilisant la formule de Black-Scholes avec une moyenne des prix des puts et des calls sur le SP500. Puisqu'elle est mesuré sur des option avec une maturité de 30 nours, la volatilité implicite est une mesure de la volatilité éspérée. Le VIX est courement surnomé le « fear index » (index de la peur). Ce surnom lui vient du fait qu'une correlation négative a été observée entre la volatilité d'un actif et son cours. A savoir, quand la volatilité augmente, le cours diminue et inversement. Une augmentation du VIX pourrait donc signifier une baisse du cours du SP500. Nous tenterons de retrouver cette correlation négative entre SP500 et VIX dans nos données.

1.1 – Analyse descriptives des données

Les données utilisées proviennent de Yahoo ! Finance. Nous travaillerons sur la période 2010 à 2020, et ce dans le but d'éviter d'inclure les valeurs aberrantes de 2008 et 2009 dans nos calculs. Valeurs qui pourraient biaiser les résultats des estimations. Le graphique1.1 présente la série sur la période sélectionné. Nous utilisons les prix hebdomadaires à la fermeture (Close).

Graphique 1.1



On constate que la série présente une tendance à la hausse très visible, elle est très probablement non-stationnaire. Nous remarquons 3 périodes de chutes significatives; en 2011, 2015-2016 et 2018. Nous pouvons attribuer la baisse de 2011 au 8 aout 2011, quand la notation de la dette souveraine des Etats-Unis est passée de AAA à AA+. L' agence de notation Standards & Poors (qui developpe aussi le SP500) a, pour la première fois de l'histoire, revu à la baisse la note de la dette américaine, causant une chute sur les marchés financiers. Ce jours le SP500 a perdu 6,66 %, le Dow Jones 5,55 % et le NASDAQ 6,90 %. Les chocs de 2015 et 2016 peuvent être du à l'explosion de la bulle spéculative en Chine. De nombreuse actions surévaluées ont été soudainement vendues, causant un krach boursier durant l'été 2015. Le 7 juillet, après avoir perdu 30 % en 3 semaine, la bourse de Shanghai a interdit les transactions sur la majorité des titres. Quant à l'année 2018, la baisse de 11 % du SP500 sur le mois de décembre s'explique par les tensions entre les Etats-Unis et la Chine quant aux accords commerciaux en vigueur entre les deux pays.

Le tableau 1.1 reprend les principales statistiques descriptives de notre série :

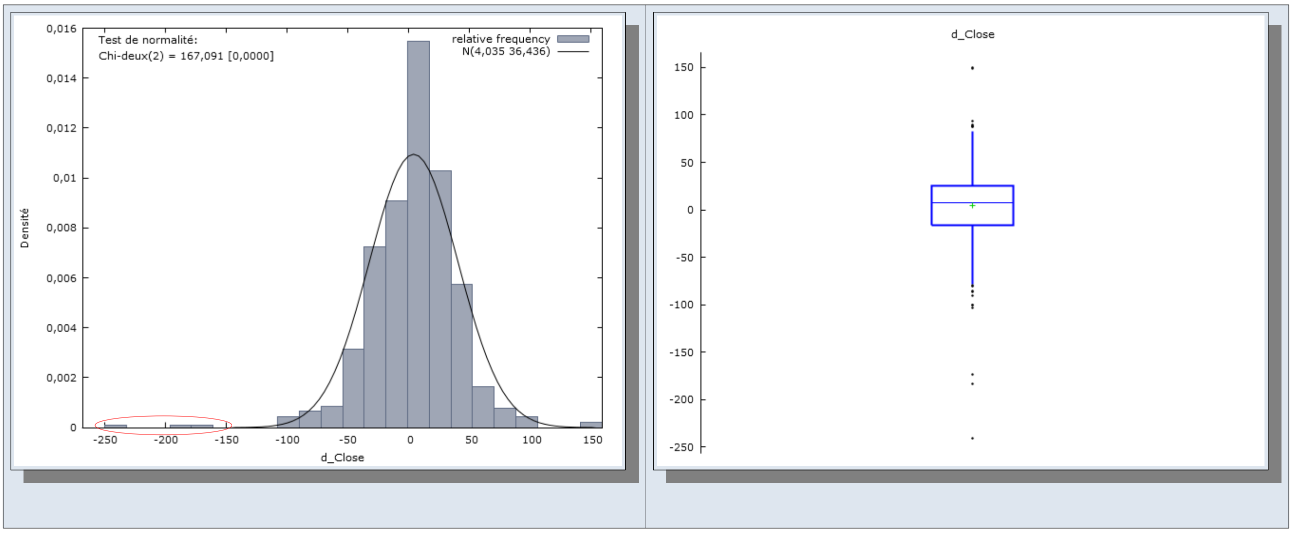
Tableau 1.1

Période	Minimum	Maximum	Moyenne	Ecart-type
2010-2020	1027,4	3239,9	1962,6	590,49

La série débute le 01/01/2010 avec une valeur de 1141.69, et atteint son minimum le 25/06/2010. La série se termine le 20/12/2019 avec une valeur de 3239,9, la plus haute valeur de la série sur la période. La moyenne se situe à 1962,6 et l'écart à la moyenne est de 590,49. Les graphiques 1.2 et 1.3 représentent la distribution de la série différenciée Z_t, c'est à dire :

$$Z_t = Y_t - Y_{(t-1)}, \text{avec } Y_t \text{ la série brute}$$

Graphique 1.2	**Graphique 1.3**



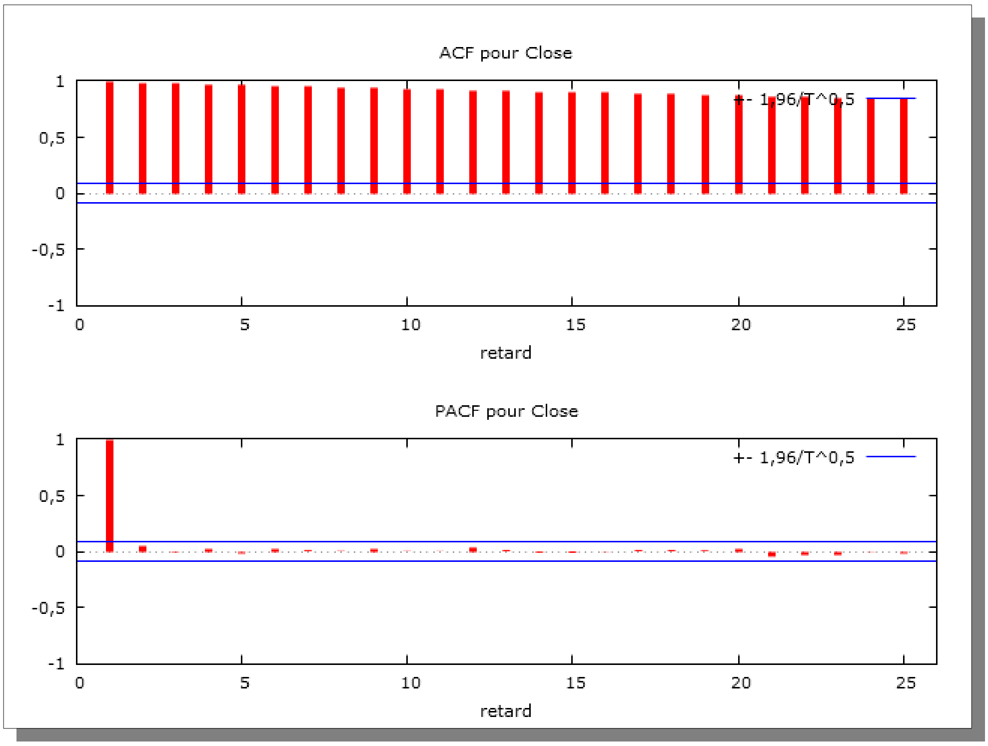
Nous constatons que la série n'est pas normalement distribuée. Elle est légèrement leptokurtique, c'est à dire moins applatie que la fonction de densité de la loi Normale. De ce fait, la série présente des valeurs extrêmes plus fréquement qu'une série distribuée normalement. Et ce, principalement pour les valeurs négatives. Cette observation est cohérente avec les travaux de Manderlbrot et Fama dans les années 1960 qui ont remis en cause la normalité des rendements en finance. Le Kurtosis est de 6,4596, largement supérieur à celui de la loi Normale (0). La série est également légèrement asymétrique sur la gauche (coefficient de skewness égal à -0,9898), ce qui confirme que l'occurence de valeurs négatives est plus probable que l'occurence de valeurs positives. On remarque la présence de ces valeurs extrêmes sur le boxplot de droite, où les outliers sont mis en évidence, et se situent surtout dans la partie basse du graphique.

Modélisation

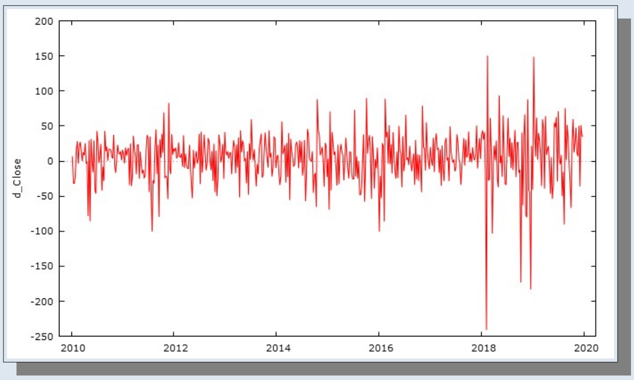
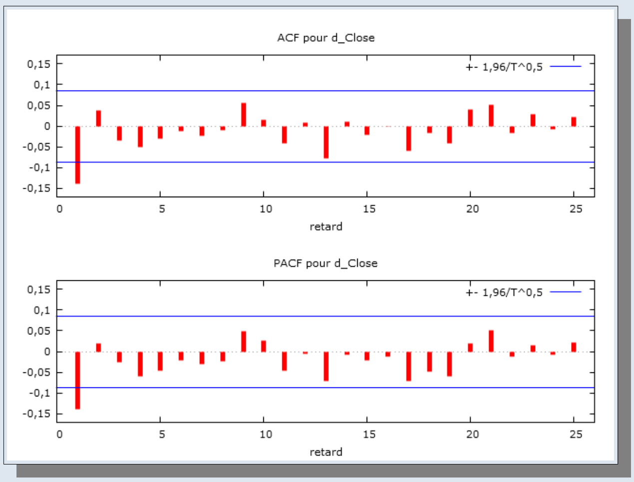
Présentation Methodo ARIMA

Nous commencons par vérifier la stationnarité de notre variable brute Y_t. Comme nous l'avons vu précedement, le graphique 1.1 il y a de forte chance que la série ne soit pas stationaire en tendance. Nous construisons le corrélogramme de la série, qui apparaît sur le graphique 2.1

Graphique 2.1



Le correlogramme indique des autocorrelations significatives jusqu'à 25 semaines. Ces autocorrelations sont exponentiellement décroissantes, ce qui indique clairement un cas de non-stationnarité. Précisement ; l'ésperance mathématique de notre série dépend du temps. Nous pouvons utiliser des tests statistiques de racine unitaire (Augmented Dickey-Fuller, Kwiatkowski-Phillips-Schmidt-Shin), mais le graphique et le corrélogramme suffisent à déterminer que la série n'est pas stationnaire elle doit donc être intégrée au moins une fois. Nous différencions notre série brute Y_t une première fois pour obtenir une série intégrée d'ordre 1. Puis nous vérifions la stationnarité de cette série.

Graphique 2.2	**Graphique 2.3**
	

Le graphique 2.2 montre la série différenciée une fois. On constate que la nouvelle série évolue autour de 0 et que la variance semble relativement constante sur la période. La série est légèrement plus volatile entre 2018 et 2020, à cause de la guerre commerciale Chine-Etats-

Unis comme suggéré plus haut. Le graphique semble indiquer que la série est stationnaire en tendance. Le corrélogramme du graphique 2.3 confirme l'hypothèse de stationnarité ; en effet on peut remarquer une autocorrelation singnificative au premier retard, puis des autocorrelations non-significatives pour les retards supérieurs à 1. Nous utilisons les tests de racine unitaire pour vérifier notre analyse graphique. Les résultats sont reportés dans le tableaux 2.1 et 2.2.

Tableau 2.1	**Tableau 2.2**
Test de Dickey-Fuller augmenté pour d_Close testing down from 18 lags, criterion AIC taille de l'échantillon 518 hypothèse nulle de racine unitaire : a = 1 test sans constante avec 0 retards de (1-L)d_Close modèle: (1-L)y = (a-1)*y(-1) + e valeur estimée de (a - 1): -1,12741 statistique de test: tau_nc(1) = -25,8207 p. critique 1,037e-038 Coeff. d'autocorrélation du 1er ordre pour e: 0,004	Test KPSS pour d_Close T = 519 Paramètre du délai de troncation = 6 Statistique de test = 0,117956 10% 5% 1% Valeurs critiques: 0,348 0,462 0,742 P. critique > .10

L'hypothèse nulle du test ADF considère la présence d'une racine unitaire. La p-value est inférieure à 0,05, on peut donc rejeter l'hypothèse nulle de non-stationnarité. Le test KPSS ne permet pas de rejeter l'hypothèse nulle de stationnarité à 5 % ni 10 %. Les deux tests confirment l'analyse graphique, la série intégrée d'ordre 1 (*d* = 1) est stationnaire, nous avons donc :

$$Z_t = Y_t - Y_{(t-1)} = \nabla^1 Y_t = (1 - B^1) Y_t$$

Nous essayons de trouver un modèle approprié pour modeliser Z_t. Nous utilisons le corrélogramme du graphique 2.3 pour determiner les ordres *p* et *q* du processus ARIMA(p,1,q) que nous appliquerons à notre série. La fonction d'autocorrelation (ACF pour Autocorrelation Function) met en évidence une valeur significative pour le premier retard puis se coupe brusquement, ce qui suggère un modèle MA(1). La fonction d'autocorrelation partielle (PACF pour Partial Autocorrelation Function) montre le même schéma ; une autocorrelation significative au premier retard puis une coupure brusque au-delà de 1 retard. La PACF suggère un modèle AR(1). Puisque nous ne pouvons pas trancher visuellement, ous allons nous orienter vers 3 modèles ; AR(1), MA(1) et une combinaison, ARIMA(1,1,1). Nous choisirons le plus approprié après avoir estimé ces 3 modèles sur notre échantillon.

ARIMA(1,1,1)	coefficient erreur std. z p. critique --- const 3,96939 1,41182 2,812 0,0049 *** phi_1 −0,274180 0,300514 −0,9124 0,3616 theta_1 0,136795 0,309782 0,4416 0,6588 Éc. type var. dép. 36,44606 R2 ajusté 0,018166 Log de vraisemblance −2596,917 Critère d'Akaike 5201,835
AR(1)	coefficient erreur std. z p. critique --- const 3,96785 1,38801 2,859 0,0043 *** phi_1 −0,140331 0,0434506 −3,230 0,0012 ***

	Éc. type var. dép. 36,44606 R2 ajusté 0,019738 Log de vraisemblance −2597,003 Critère d'Akaike 5200,006
MA(1)	coefficient erreur std. z p. critique --- const 3,96692 1,36857 2,899 0,0037 *** theta_1 −0,135739 0,0428412 −3,168 0,0015 *** Éc. type var. dép. 36,44606 R2 ajusté 0,018975 Log de vraisemblance −2597,207 Critère d'Akaike 5200,415

Les résultats des différents modèles sont repris dans le tableau 2.3 :

Tableau 2.3

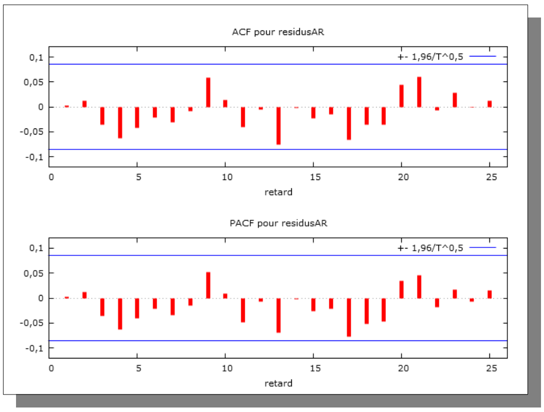
Modèle	Coefficients	Ecart-type	R2 Ajusté	AIC
ARIMA(1,1,1)	$\Phi_1 = -0{,}2748$ p-value = 0,36 $\Theta_1 = 0{,}136795$ p-value = 0,65	36,446	0,0182	5201,84
AR(1)	$\Phi_1 = -0{,}140331$ p-value = **0,0012**	36,446	**0,0197**	**5200,01**
MA(1)	$\Theta_1 = -0{,}135739$ p-value = 0,0015	36,446	0,01898	5200,42

Ces 3 modèles fournissent des résultats très similaires, cependant les coefficients éstimés pour le modèle ARIMA(1,1,1) ne sont pas significatifs[1], avec des p-values de 0,36 pour le coefficient associé à la partie AR et 0,65 pour la partie MA. De plus, ce modèle a le R2 ajusté le moins élévé et l'AIC le plus élévé. Tous les éléments semblent suggérer d'exclure ce modèle. A première vue le modèle AR(1) est le meilleur de ces 3 modèles avec le R2 le plus élevé (0,0197), bien que très faible en valeur absolue. AR(1) a égalament le plus faible AIC, et c'est son coefficient qui est le plus significatif avec une p-value inférieure à 0,01, on peut conclure que Φ_1 est significatif au seuil de 1 %. Le coefficient Θ_1 est lui aussi significatif au seuil de 1 %. Il est en revanche impossible de discrimer les modèles en se basant sur l'écart-type, identique dans les trois cas.

Rien dans le corrélogramme ne semble indiquer un modèle AR(p), p ≠ 1, ou MA(q), q≠ 1. Nous séléctionnons donc les modèles AR(1) et MA(1), Equivalents en terme de qualité d'estimation. Pour les départager nous comparons les résidus des modèles.

Modèle AR(1)	
Corrélogramme des résidus	**Test Ljung-Box**

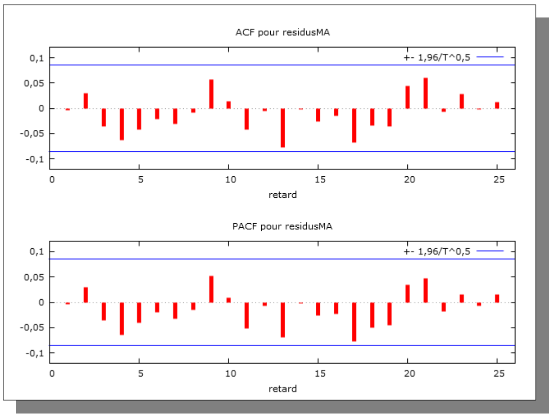
1 Hypothèse H_0 du z-test : $\phi_i = 0$; Hypothèse alternative H_a : $\phi_i \neq 0$



RETARD	ACF	PACF	Q	[p. crit.]
1	0,0020	0,0020	0,0020	[0,964]
2	0,0120	0,0120	0,0775	[0,962]
3	-0,0358	-0,0359	0,7500	[0,861]
4	-0,0633	-0,0634	2,8553	[0,582]
5	-0,0419	-0,0411	3,7768	[0,582]
6	-0,0209	-0,0208	4,0062	[0,676]
7	-0,0308	-0,0347	4,5079	[0,720]
8	-0,0085	-0,0153	4,5457	[0,805]
9	0,0578	0,0521	6,3190	[0,708]
10	0,0143	0,0083	6,4281	[0,778]
11	-0,0412	-0,0495	7,3305	[0,772]
12	-0,0060	-0,0072	7,3496	[0,834]
13	-0,0756 *	-0,0700	10,4024	[0,661]
14	-0,0029	-0,0026	10,4068	[0,732]
15	-0,0237	-0,0266	10,7093	[0,773]
16	-0,0152	-0,0216	10,8333	[0,820]
17	-0,0663	-0,0771 *	13,1980	[0,723]
18	-0,0353	-0,0513	13,8689	[0,738]
19	-0,0353	-0,0472	14,5430	[0,751]
20	0,0446	0,0343	15,6185	[0,740]
21	0,0598	0,0463	17,5574	[0,677]
22	-0,0068	-0,0184	17,5827	[0,730]
23	0,0273	0,0165	17,9888	[0,758]
24	-0,0008	-0,0069	17,9892	[0,804]
25	0,0118	0,0148	18,0654	[0,840]

Modèle MA(1)

Corrélogramme des residus



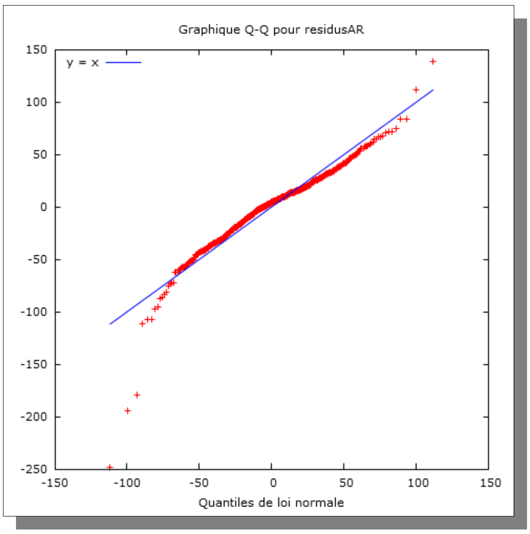
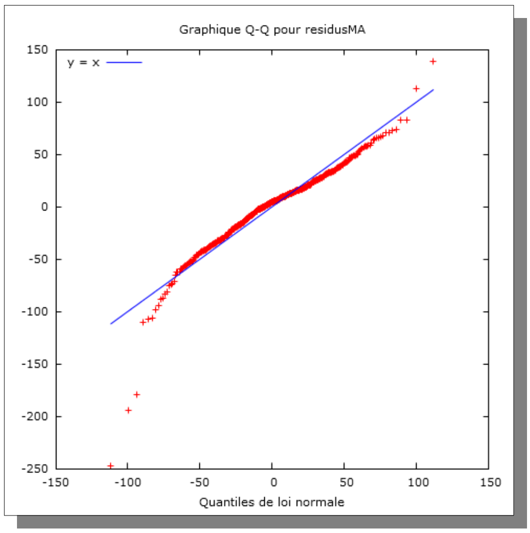
Test Ljung-Box

RETARD	ACF	PACF	Q	[p. crit.]
1	-0,0035	-0,0035	0,0063	[0,937]
2	0,0299	0,0299	0,4753	[0,788]
3	-0,0365	-0,0364	1,1745	[0,759]
4	-0,0627	-0,0640	3,2416	[0,518]
5	-0,0428	-0,0413	4,2031	[0,521]
6	-0,0219	-0,0200	4,4556	[0,615]
7	-0,0304	-0,0331	4,9452	[0,667]
8	-0,0088	-0,0153	4,9860	[0,759]
9	0,0563	0,0516	6,6683	[0,672]
10	0,0139	0,0088	6,7710	[0,747]
11	-0,0416	-0,0518	7,6935	[0,740]
12	-0,0052	-0,0073	7,7078	[0,808]
13	-0,0766 *	-0,0695	10,8466	[0,624]
14	-0,0027	-0,0030	10,8507	[0,698]
15	-0,0264	-0,0268	11,2247	[0,737]
16	-0,0155	-0,0224	11,3545	[0,787]
17	-0,0672	-0,0773 *	13,7864	[0,682]
18	-0,0343	-0,0504	14,4206	[0,701]
19	-0,0354	-0,0452	15,0993	[0,716]
20	0,0437	0,0346	16,1364	[0,708]
21	0,0595	0,0478	18,0567	[0,645]
22	-0,0064	-0,0191	18,0787	[0,701]
23	0,0287	0,0158	18,5269	[0,728]
24	-0,0018	-0,0071	18,5286	[0,777]
25	0,0125	0,0148	18,6147	[0,815]

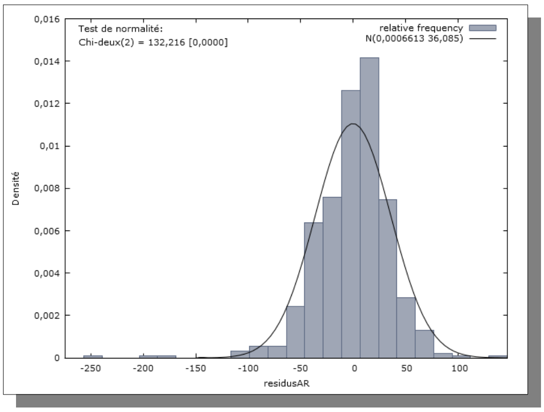
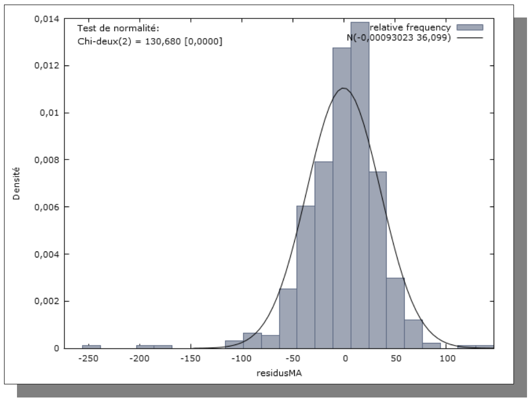
Pour les deux modèles, les résidus sont non-significatifs à 5 %. En revanche les 13e et 17e valeurs sont significatives à 10 %, dans les deux modèles. Visuellement, les résidus semblent donc stationnaires, nous verifions cette approche graphique à l'aide des tests ADF et KPSS.

Modèle	P-value ADF	P-value KPSS	Conclusion
AR(1)	4,644e-040 < 0,05	> 0,10	Résidus stationnaires
MA(1)	4,603e-040 < 0,05	> 0,10	Résidus stationnaires

L'hypothèse de non stationnarité est rejetée pour le test ADF, pour les deux modèles. Et on ne peut pas rejeter l'hypothèse d'indépendence des résidus pour le test KPSS. Les deux modèles ont des résidus stationnaires, nous vérifions qu'ils soient i.i.d normaux $(0, \sigma^2)$[2].

Modèle	AR(1)	MA(1)
Tests	Test de Doornik-Hansen = 132,216, avec p. critique 1,94846e-029 Shapiro-Wilk W = 0,926008, avec p. critique 2,62076e-015 test de Lilliefors = 0,076148, avec p. critique ~= 0 test de Bera-Jarque = 1140,56, avec p. critique 2,13972e-248	Test de Doornik-Hansen = 130,68, avec p. critique 4,20037e-029 Shapiro-Wilk W = 0,926026, avec p. critique 2,63243e-015 test de Lilliefors = 0,0800678, avec p. critique ~= 0 test de Bera-Jarque = 1124,93, avec p. critique 5,29163e-245
QQ Plot		

2 Les test de Doornik-Hansen, Shapiro-Wilk, Lilliefors et Jarque-Bera ont pour hypothèse nulle, H_0 : les residus sont distribués normalement

Histogramme		

On s’aperçoit sur la QQ plot que les résidus sont leptokurtiques, avec plus de valeurs autour de la moyenne et dans les queues (surtout la queue basse), que dans la distribution normale. Ceci nous est confirmé par l’histogramme qui montre le même schéma, avec une distribution de fréquences plus pointue dans les résidus de nos modèles que dans une loi Normale ainsi que les valeurs négatives extrêmes bien plus fréquentes. Par ailleurs les quatres tests de Normalité de Gretl suggèrent une non-Normalité des résidus. Il faut donc être parcimonieux avec l’interpretation de nos modèles. En revanche, pour définitivement éliminer l’un de deux, nous utiliserons les résultats du test de Ljung-Box avec un retard maximal de 25 semaines. La p-value du test sur les résidus du modèle AR(1) est légèrement supérieure à celle du modèle MA(1) ; 0,84 contre 0,815. L’hypothèse nulle d’absence d’auto-corrélation d’ordre 1 à 25 ne peut pas être rejetée, pour aucun des deux modèles. Nous choisirons le modèle avec la p-value la plus élévée (dont les résidus sont le moins autocorrélés). Nous séléctionnons donc le modèle AR(1) pour modéliser notre série différenciée.

Prévision

Notre modèle final est une modèle AR(1) :

$$Z_t = \delta + \phi_1 Z_{(t-1)} + \varepsilon_t , \varepsilon_t \sim N(0,\sigma)\, i.i.d$$
$$\hat{Z}_t = (-0{,}140331) Z_{(t-1)} + 3{,}96785$$

Z_t est notre série différenciée une fois (intégrée d’ordre 1). Pour faire la prévision à une période , nous devons rebasculer vers notre série brute :

$$(1 - B^1 \hat{\phi}_1) * Z_t = \varepsilon_t + \delta$$
$$Z_t = Z_t - B^1 \hat{\phi}_1 Z_t = Z_t - \hat{\phi}_1 Z_{(t-1)} + \delta$$
$$Z_T(1) = E[Z_{(T+1)} | H_T] - \hat{\phi}_1 Z_T + \delta$$

De plus ;

$$Z_t = \nabla^1 Y_t = Y_t - Y_{(t-1)}$$
$$\Leftrightarrow Y_{(T+1)} = Y_T + Z_{(t+1)} = Y_T + Z_T(1)$$

$$Y_{(T+1)} = 3205{,}370 + 3{,}976262 - 0{,}140331 * 36{,}8000 = 3204{,}182081$$

Parallèlement ;

$$P(3205{,}37-71{,}296 \leqslant Y_{(T+1)} \leqslant 3205{,}37+70{,}017)=95\,\%$$

Autrement dit, l'intervalle de confiance à 95 % pour la valeur $Y_{(T+1)}$ est :

$$Y_{(T+1)} \in [3134{,}074\,;3275{,}387]$$

La valeur théorique de la série à la date T+1 (12/20/2019) est 3239,90[3]. La valeur prédite par notre modèle AR(1) est 3204,18. Nous avons donc un résidus estimé à 35,72, légèrement en dessous de notre écart-type des résidus (36,0498). La valeur du SP500 au 12/20/2019 n'est donc pas aberrante, notre estimation n'est pas véritablement précise. Nous avions observé un faible R2 ajusté et un écart -type important, il est donc naturel de constater une mauvaise prévision quant à l'estimation même à h = 1.

Cointégration :

3 L'observation a été retirée de l'echantillon avant l'estimation.