NUAGE DE POINTS :
-bien qu'imparfaite, on constate une légère relation linéaire croissante entre les deux variables
-un modèle linéaire estimé par les MCO pourrait donc être approprié
-on peut s'attendre à un β positif mais inférieur à 1 et à un α négatif

CORRELATION :
-nous avons un coefficient de corrélation positif et relativement élevé
-la relation linéaire entre les deux variables semble se confirmer

MCO :
-si la valeur de t-stat est inférieure à la valeur tabulé(table de student) donc n'est pas significative
-Nous obtenons un β de 1.5977, c'est à dire positif (ce que nous attendions). Il est néanmoins supérieur à 1 .
-Ainsi, une variation de 1%de la variable ersandp (variable explicative) induira une variation de 1.5977% de erford. (variable expliquée)
-La constante n'est pas négative, ce à quoi nous nous attendions à partir du nuage de points. Cela signifie que s'il n'y a aucune variation du marché, le rendement des actions de Ford sera quand même légèrement positif.
-La p-valeur associée au β est égale à 0.000 : on peut dire que la variable ersandp est statistiquement significative au seuil de 5%
-l'hypothèse de non significativité des paramètres est :
  -non-rejetée pour α. - rejetée pour β
-Significativité globale du modèle :
- le $R2$ ajusté vaut 0.287 : donc environ 30% des variation de erford sont expliquées par le modèle (ce qui est tout à fait raisonnable).

MCO « HAC » :
-on remarque que l'option "HAC" permet de mener une correction de la matrice de variance-covariance de l'erreur à la Newey-West
-cette correction semble avoir modifié les écarts-types des estimateurs (à la hausse ce qui est normal puisque l'hétéroscédasticité et l'autocorrélation ont tendance à mener à sous-estimer la variance des estimateurs).
-on pourra rappeler que ces deux problèmes sont un soucis dans la mesure où ils entrainent un estimateur erroné de la matrice de variance-covariance des erreurs, ce qui fausse l'inférence statistique (c a d les tests de significativité).

RESIDUS :
-Les résidus semble être de moyenne constante et égale à 00
-En revanche leur variance n'est pas constante
-Difficile de dire s'il existe une structure d'autocorrélation.

Breusch-Pagan / white (tuple de 4 éléments) :
-1er élément : Statistique du test LM (se référer à la table)
-2ème : p-valeur associée à la statistique de test LM (inférieure à 0.1 (white : 0,01), on rejette l'hypothèse nulle d'homoscédasticité à 10%)
-3ème : Statistique du test de Fischer (significativité globale du modèle de régression des résidus sur X)
-4ème : p-valeur associée à ce test (même interprétation puisque si la p-valeur <<0.01, le modèle est globalement significatif, et donc la relation entre les résidus et les variables explicatives est significative)
Ici nous rejettons l'hypothèse nulle d'homoscédasticité à 10%

➔VIF :
```
1 explicatives = ['ersandp','roil','d crisis','d covid']
for n,var in enumerate(l explicatives):
print('le VIF de la variable', var, ' est : ',VIF(df[l_explicatives],dropna().values,n))
```
le VIF de la variable ersandp est : 1.0591829579271606
le VIF de la variable roil est : 1.0473746769468446
le VIF de la variable d_crisis est : 1.0296646911134221
le VIF de la variable d_covid est : 1.003964217666122

Commentaires :
-Nous avons calculé les Variance Inflation Factor de toutes les variables explicatives.
-Sans avoir de règle de décision précise, nous pouvons dire qu'avec des VIF inférieur à 2 pour toutes les variables, la présence de multicolinéarité est faible.
-L'hypothèse MCO selon laquelle la matrice des variables explicatives X est de plein rang, semble remplie (hypothèse qui assure de pouvoir estimer les paramètres).

➔
```
import numpy as np
from numpy.linalg import det
D = det(df[l explicatives].corr())
# calcul de la stat : T=244 et k=4
FG = -(244 - (1/6) * (2*(4+1))+5)*np.log(D)
print(FG)
```

Commentaires :
-Nous menons ici un test de multicolinéarité de Farrar Glober
-Avec une valeur tabulée pour le chi2(1/2 k(k+1)) de 18.307 à 5%, l'hypothèse nulle d'absence de multicolinéarité est rejetée.
-Ce test confirme donc nos commentaires précédents.

-graphiques lignes pour voir l'allure des séries et, éventuellement en déduire de l'information sur la stationnarité des séries

-autocorrélogramme : permet de voir la structure autorégressive (de long terme) des séries

-autocorrélations partielles : permet de voir l'autocorrélation entre deux périodes après avoir purgé l'effet des autres périodes

-ADF : si la statistique est supérieure à la valeur tabulée nous ne rejetons pas H0 la série contient une racine unitaire. La série est non stationnaire de type stochastique : il faut différencier (ou prendre le taux de croissance pour la stationnariser. Nous devons, bien entendu vérifier que la variable ainsi transformée est elle-même stationnaire. rford est donc stationnaire et la série ford est dite intégrée d'ordre 1 puisqu'il a fallu la différencier une fois pour la stationnariser. D'où on travaillera sur la série de rendement
Si p-val est inférieure à 1% le test rejette H0 de présence d'une racine unitaire. La série

Exercice 1 :
Question1 : A l'aide de quelle(s) statistiques(s) peut-on comparer la significativité globale de 2 modèles estimés par la même méthode ?
Réponse : Pour comparer la significativité globale de deux modèles estimés par la même méthode, plusieurs statistiques peuvent être utilisées en fonction du type de modèle utilisé : test chi-2, test de fisher, le critère d'information d'Akaike (AIC) ou le critère d'information bayésien (BIC).
Question 2 : L'estimateur MCO est dit « BLUE ». Que signifie cet acronyme et à ce quoi renvoie-t-il ?
Réponse : BLUE : best linear unbiased estimator. Propriété1 « Linéarité de l'estimateur ». Propriété2 « Estimateur dans biais ». Propriété3 « ^β est de variance minimale ». Propriété totale : « ^β est un estimateur BLUE de β ». On dit de l'estimateur MCO ^ β de β est BLUE (Best Linear Unbiased Estimator).
Question 3 : Quelle propriété des variables est-il impératif d'étudier lorsque l'on analyse des séries temporelles ? Pourquoi ?
Réponse : Lors de l'analyse de séries temporelles, il est crucial d'étudier la stationnarité des variables. La stationnarité, indique que les propriétés statistiques de la série restent constantes au fil du temps. Cette propriété est essentielle pour garantir la validité des résultats d'analyse, assurer la stabilité des modèles, faciliter l'interprétation des résultats et améliorer la fiabilité des prévisions à long terme. Des méthodes graphiques et des tests statistiques, tel que le test Augmented Dickey-Fuller (ADF) est utilisé pour évaluer la stationnarité. En cas de non-stationnarité, des transformations ou ajustements peuvent être nécessaires avant d'appliquer des modèles d'analyse.

➔Modèle de régression simple : $Y=\alpha+\beta X+\epsilon$
-Y est la variable expliquée (observée)
-X est la variable explicative (observée)
-α (la constante) et β (le coefficient de pente) sont les paramètres à estimer (inconnus)
-ϵ est le terme d'erreur (inconnu)
Dans un premier temps, on considère que X est observé sans erreur, autrement dit, X est une variable certaine. De ce fait, elle est indépendante du terme d'erreur ϵ. La variable Y est elle en revanche, une variable aléatoire (puisqu'elle dépend précisément d'ϵ).
Si les variables X et Y sont des variables qui dépendent de T observations, on écrit le modèle de régression simple comme suit : $Yt=\alpha+\beta Xt+\epsilon t$
$\forall t\in[[1;T]]$ où t peut désigner deux dimensions différentes :
-le temps (séries temporelles)
-des individus (coupes instantannées)

➔propriété du terme d'erreur : On ne peut pas mesurer ou prévoir le terme d'erreur ϵ pour chaque observation. On émet un certain nombre d'hypothèses sur ce terme....
1-**Nulité de l'erreur moyenne :**
-$\epsilon\in\mathbb{R}\Rightarrow\epsilon\leq 0$ ou $\epsilon\geq 0$, mais sans plus de chance d'avoir l'un ou l'autre
-$\Rightarrow E(\epsilon t)=0$
-Revient à dire que, en moyenne, le modèle est bien spécifié
2- **L'absence d'autocorrélation des erreurs**
-La valeur de l'erreur en t (période ou individu) ne dépend pas de celle en t'
-Il n'y a pas de corrélation entre deux erreurs à deux dates différentes ou entre deux individus
-$\Rightarrow E(\epsilon t\epsilon t')=0,\forall t\neq t'$
3- **Homoscédasticité des erreurs**
-La variance de l'erreur est constante (au cours du temps ou entre les individus)
-$\Rightarrow E(\epsilon 2t)=\sigma 2\epsilon,\forall t$
-$\sigma^2\epsilon$ représente la variance de l'erreur et ne dépend pas de t
4- **Normalité des erreurs**
-On suppose (théorème central limite) que le terme d'erreur suit une loi normale centrée et de variance constante
-Cette hypothèse se vérifie d'autant plus que le nombre d'observations est grand
-$\Rightarrow\epsilon t\sim N(0,\sigma 2\epsilon)$

➔Principe des MCO : $Yt=\alpha+\beta Xt+\epsilon t$
L'objectif des MCO est d'estimer les paramètres α et β. Une telle estimation nous donne l'équation d'une droite, appelée **droite de régression**, donnée par :
$\hat{Y}\,t=\hat{\alpha}+\hat{\beta}\,Xt$
où :
-$\hat{\alpha}$ et $\hat{\beta}$^ sont les estimateurs des paramètres α et β
-$\hat{Y}\,t$^ est la valeur estimée de Yt et donnée par le modèle (une valeur de $\hat{Y}\,t$^ correspond à un point sur l'axe des ordonnée (Yt) pour un Xt et obtenu *via* la droite de régression MCO).

➔Propriétés des estimateurs :
-Propriété 1 : la droite de régression passe par le point moyen ($\bar{X}$,$\bar{Y}$)
-Propriété 2 : la variable observée Yt et estimée Yt^sont de même moyenne : $\bar{Yt}=\bar{Yt^{\wedge}}$
-Propriété 3 : en moyenne, les résidus sont nuls (modèle en moyenne correctement spécifié) $\bar{e}=0$
-Propriété 4 : la covariance entre la variable explicative Xtet le résidus, ainsi qu'entre la variable expliquée estimée Yt^ et le résidus, est nulle (pas de corrélation avec le résidu : $Cov(Xt,et)=Cov(Yt^{\wedge},et)=0$
-Propriété 5 : les changements d'origine et d'échelle ne changent pas $\hat{\beta}$^.
-Propriété 6 : les estimateurs sont linéaires (on peut les exprimer comme des fonctions linéaires de la variable expliquée)
Très important :-**Propriété 7** : ce sont des estimateurs sans biais (en moyenne, ils sont égaux à leur vraies valeurs) $E(\hat{\alpha})=\alpha$ et $E(\hat{\beta}^{\wedge})=\beta$
-Propriété 8 : les estimateurs MCO sont convergents et de variance minimale. Autrement dit, lorsque le nombre d'observations T tend vers l'infini, la variance converge vers 0.
-les estimateurs MCO sont **BLUE** : Best Linear Unbiased Estimator

➔Estimateur MCO : Hypo du modèles de rég multinle : $Yt=\alpha+\beta 1X1,t+\beta 2X2,t+...+\beta kXk,t+\epsilon t$

-H1 : la matrice X est non aléatoire : les valeurs de X sont observées sans erreur (X et le terme d'erreur sont indépendants. $E(\epsilon t|X)=0$

-H2 : la matrice X est de plein rang : Rang(X)= k+1. Cette hypothèse revient à dire que les colonnes sont linéairement indépendantes, ou que l'on ne peut pas écrire une colonne de X comme une combinaison linéaire des autres. Nous faisons cette hypothèse pour assurer l'inversibilité de la matrice $X'X$ (dont nous avons besoin pour le calcul de β).Une hypothèse sous-jacente est que le nombre d'observations ne peut pas être inférieur au nombre de variables : $T>k+1$. Sans cela, X ne peux pas être de plein rang.

-H3 : nulité de l'espérance de l'erreur : Comme pour le modèle de régression simple, nous supposons que le terme d'erreur est nul en moyenne :$E(\epsilon)=0$. On en déduit que : $E(Y)=X\beta$ c'est à dire que le modèle est en moyenne bien spécifié.

-H4 : homoscédasticité et absence d'autocorrélation des erreurs : Soit $E(\epsilon\epsilon')$ la matrice de variance covariance du terme d'erreur, avec ϵ' la transposée du vecteur des termes d'erreur. L'hypothèse H4 s'écrit comme suit : $E(\epsilon\epsilon')=\sigma^2\epsilon I$ où I désigne la matrice identitée et $\sigma^2\epsilon$ la variance du terme d'erreur.

-H5 ; normalité des erreurs : nous supposons (en vertu du théorème central limite) que les erreurs suivent une loi normale d'espérance nulle et de variance constante :$\epsilon\sim N(0,\sigma 2\epsilon I)$ À noter que cette hypothèse ne sert pas au calcul des paramètres mais à la validation des résultats statistiques et à la construction des tests statistiques.

➔ Estimateur de la variance des erreurs
On ne connait pas les erreurs dans un modèle. Pour estimer la variance des erreurs $\sigma 2\epsilon$, il convient d'utiliser les résidus e : $\hat{\sigma}\,2\epsilon=e'e/(T-k-1)$ $\hat{\sigma}\,2\epsilon$ est un estimateur sans biais de la variance des erreur $\sigma 2\epsilon$

➔test sur un coefficient de régression : test de student : À partir de la statistique t définie à la section précédente, nous pouvons déduire un intervalle de confiance à $100(1-p)\%$ pour βi:
$\hat{\beta}_i \pm t_{n\backslash 2}\hat{\sigma}_\epsilon\sqrt{a_{i+1,i+1}}$
On peut donc tester si β i^ s'écarte plus ou moins une certaine valeur $\beta 0$: -$H0$, l'hypothèse nulle : β i=$\beta 0$ - $H1$, l'hypothèse alternative : β i≠$\beta 0$
-en pratique, on utilise ce test comme un test de significativité, c'est à dire qu'on prend $\beta 0$=0: -$H0$, l'hypothèse nulle : β i=0. -$H1$, l'hypothèse alternative : β i≠0

➔test sur plusieurs coefficients : test de Fisher : Le test de Student nous permet de tester la significativité statistique d'une variable seule. Nous procédons maintenant à la description du test de Fisher, qui permet de tester la significativité de plusieurs coefficients simultanément. Idée générale du test : poser une contrainte sur plusieurs coefficients en même temps, $R\beta=r$. Où R est une matrice de plein rang et de taille(q,k+1)avec q le nbr de contraintes et r le vecteur des contraintes de taille(q,1)

➔Equation de l'analyse de la variance : $Y=X\beta+\epsilon$. nous avons obtenus, après estimation, une expression pour le résidu : $e=y-x\hat{\beta}$. où l'on note X et Y en minuscule pour signifier qu'elles sont centrées (on leur retire leur moyenne).

➔coefficient de détermination (R^2) : mesure la part de la variance totale expliquée par les variables explicatives et permet de juger de la qualité de l'ajustement du modèle. R²=SCE/SCT= 1-SCR/SCT où 0<=R²<=1. ➔coefficient de détermination corrigé/ajusté (R^{-2}) : le coefficient de détermination classique augment avec le nombre variables explicatives intégrées dans le modèle(puisque l'introduction de nouvelles variables ne peut pas faire diminuer la SCR) on peut donc le corriger de ce biais : $R^{-2}=1-((T-1)/(T-k-1))*(1-R^2)$

➔Critère d'information : critère d'information d'Akaike(AIC) : plus ce critère est faible ,plus le modèle a un pouvoir explicatif important. Critère d'information de Schwarz(SCI) : ce critère est aussi un critère qu'il faut minimiser.

➔Prévision : estimation d'un modèles : - permet d'expliquer ce qui est observé : pouvoir explicatif (en échantillon). -permet d'émettre des prévision : pouvoir prédictif du modèle (hors échantillon : on estime le modèle sur une partie de l'échantillon, puis on regarde l'erreur de prévision sur la partie restante de l'échantillon).

➔Propriété des estimateurs en présence d'autocorrélation et/ou hétéroscédasticité : modèle : $Y=X\beta+\epsilon$ où X est non aléatoire de plein rang. L'estimateur des MCO donne :$\hat{\beta}=(X'X)-1X'Y$; en présence d'hétéroscédasticité ou d'autocorrélation, la variance de l'estimateur MCO n'est plus minimale. Les MCO ne sont alors plus *"Best"*.
➔Hétéroscédasticité des erreurs : On rappelle que l'hétéroscédasticité renvoie au fait que les termes situés sur la diagonale de la matrice de variance-covariance des erreurs sont non-égaux entre eux. En série temporelle : la variance des erreurs est non constante dans le temps.

➔D'où peut venir l'hétéroscédasticité ?
-hétérogénéité de l'échantillon (pays différents ou époques différentes)
-oubli d'une variable explicative dans le modèle
-asymétrie dans la distribution des certaines variables
-une mauvaise transformation des variables (par exemple : prendre le modèle linaire alors que le bon modèle aurait été le modèle log-linéaire).

➔Tests d'hétéroscédasticité : *1-test graphique : (pour détecter*
-il existe divers tests permettant de la détecter
-une première intuition graphique est toujours utile :
- estimation du modèle comme s'il n'y avait pas d'hétéroscédasticité
- graphique (nuage de points) de la variable expliquée estimée (^Y) en fonction des résidus au carré
- s'il existe une relation apparente entre les deux séries, nous pouvons avoir présomption de présence d'hétéroscédasticité

-il est aussi possible de faire le nuage de points entre les résidus et une variable explicative : de même si une relation apparaît, nous avons présomption d'hétéroscédasticité.
-une autre analyse graphique consiste à faire le graphique de l'évolution des résidus selon les observations : la variance doit apparaître constante.

2-quelques tests d'hétéro connus : Le test de Goldfeld-Quandt, développé en 1965, sert à détecter l'hétéroscédasticité due à l'une des variables explicatives dans une régression linéaire. Le processus implique le classement des valeurs de la variable suspecte par ordre croissant, la suppression des valeurs centrales, l'estimation par les MCO sur les deux sous-échantillons résultants, le calcul des sommes des carrés des résidus (SCR), et enfin la comparaison des SCR. Une différence significative suggère la présence d'hétéroscédasticité dans le modèle. Test de Glesjer (1969) : a pour but de détecter la présence ET la forme de l'hétéroscédasticité.

3-test de breusch-pagan(1979) : -test général couvrant un grand nombre de cas d'hétéroscédasticité
- valable que pour des échantillons assez grands (beaucoup d'observation) : test asymptotique
- n'implique pas de savoir quelle variable explicative est à l'origine de l'hétéroscédasticité
- test simple : ne se base que sur les résidus issus de la régression MCO du modèle

On considère le modèle de régression multiple suivant :
$Yt=\alpha+\beta 1X1,t+\beta 2X2,t+...+\beta kXk,t+\epsilon t$

- $BP<\chi 2p$, l'hypothèse nulle d'homoscédasticité n'est pas rejetée
- $BP>\chi 2p$l'hypothèse nulle d'homoscédasticité est rejetée

4-test WHITE :- ne repose pas sur l'hypothèse de normalité des erreurs
-ne s'intéresse pas à la forme de l'hétéroscédasticité.
- si $TR^2<\chi^2_{2k}$, on ne rejette pas l'hypothèse nulle d'homoscédasticité
- si $TR^2>\chi^2_{2k}$, on rejette l'hypothèse nulle d'homoscédasticité

➔Procédure d'estimation en présence d'hétéro : Problème : cette variance est utilisée dans le calcul des test de significativité. Solution : -estimation du modèle par les MCG, il faut connaitre la matrice de variance covariance du terme d'erreur.
-Correction des MCO par l'utilisation d'un estimateur de la matrice de var-cov.

1-estimateur de la matrice de var-cov de white :
$\hat{\Omega}^W\epsilon=(T/(T-k-1))*(X'X)^{-1}(\sum e^2tXtX't)(X'X)^{-1}$

2-estimateur de la matrice de var-cov de newey west : -La proposition de correction de White repose sur l'hypothèse de non-autocorrélation des résidus.
-La proposition faite par Newey et West (1987) non

➔Autocorrélation des erreurs : nous travaillerons sur l'autocorrélation en série temporelle : le terme d'erreur à un instant, dépend du même terme d'erreur à un autre instant/ en coupe instantanée : autocorrélation spatiale.

➔tests d'autocorrélation : *1-détection graphique de la présence d'autocorrélation :* de la même manière que pour l'hétéroscédasticité, il est possible de repérer la présence d'autocorrélation dans les erreurs en nous intéressant à la structure des résidus, de manière graphique

*2-test de durbin-watson :*Permet de tester l'autocorrélation à l'ordre 1 (en série temporelle, cela l'autocorrélation à l'ordre 1 renvoie à une autocorrélation du terme d'erreur entre 2 périodes). Test basé sur une hypothèse de distribution des résidus

3-test de Breusch-godfrey : permet de déceler d'autocorrélation à des ordres supérieurs à 1. Le test se base sur le processus autorégressif d'ordre p (AR(p)). $BG<\chi 2p$, l'hypothèse nulle d'absence d'autocorrélation n'est pas rejetée. $BG>\chi 2p$, l'hypothèse nulle d'absence d'autocorrélation est rejetée
➔procédure d'estimation en présence d'autocorrélation :
-Même remarques que pour l'hétéroscédasticité :
- autocorrélation ⇒ variance de l'estimateur non minimale
- variance non minimale ⇒ tests de significativité biaisé
- modèle non interprétable.

-Il est donc important de corriger de l'autocorrélation
-Pour effectuer cette correction : même problème que pour l'homoscédasticité concernant les MCG.
-On préférera donc procéder à une correction du modèle en utilisant un estimateur de la matrice de var-cova des erreurs.
-Nous avons vu que la correction de Newey-West permettait est aussi adaptée à la correction de modèles en présence d'autocorrélation : c'est donc la correction que nous utiliserons.

➔Problème lié aux variables explicatives : variables aléatoires : -Hypothèse, X est non aléatoire : $E(\epsilon t|X)=0$
renvoie à l'hypothèse d'exogénéité des variables explicatives : l'espérance conditionnelle de l'erreur étant nulle, les variables explicatives ne sont pas déterminées par le modèle et lui sont donc exogène.
-Lorsque cette hypothèse est non vérifée : même si la taille de l'échantillon croît indéfiniment, les estimateurs MCO ne se rapprochent plus de la vraie valeur.
-Problème : trouver un estimateur convergent
-L'idée :
- variables explicatives ne sont pas réellement "exogènes"
- solutions : les remplacer par des variables qui le sont
- enjeux : trouver des variables exogènes, représentative de ce que l'on souhaite mesurer

-Pour détecter le problème : étude de la corrélation entre les variables explicatives et le résidu de la régression
-origine du problème : mauvaise mesure des variables ou manque d'une variable explicative importante dans le modèle ou la variable expliquée est aussi variable explicative du modèle.

➔si la corrélation est forte : -Passer le modèle en dynamique (avec des retards). Vérifier si une variable explicative importante ne manque pas. Etudier le sens de causalité de la relation.

➔Problème lié aux variables explicatives : mutlicolinéarité :
Nous avons étudié l'hypothèse selon laquelle X est une matrice non aléatoire. Nous nous intéressons maintenant, plus en détails à l'hypothèse qui établit que X est de plein rang.

➔Détection multicolinéarité:VIF(variance inflation factor :
-Cette statistique est appelée facteur d'influence de la variance (Variance Inflation Factor).
-Des valeurs élevées pour cette statistique pour une variable explicative donnée est le signe de la présence éventuelle d'une situation de multicolinéarité presque parfaite l'impliquant.
-Il n'existe cependant pas une règle établissant un seuil pour cette statistique au-delà duquel on peut prendre une décision. La statistique est néanmoins utile car donnant des indications sur les paramètres $\hat{\beta}$j pouvant être estimés avec peu de précision.
-Parfois considéré que : $VIF<2$ est parfait, et qu'à partir de 10, la multicolinéarité est trop forte.

➔test de FG : corrélation entre variables explicatives :
-Test basé sur la matrice des coefficients de corrélation entre les variables explicatives.
-Si les variables sont linéairement corrélées, le déterminant de cette matrice est nul
-La statistique de test est données par :
$FG=-(T-1-1/6(2(k+1)+5))\times log(D)$
où D est le déterminant de la matrice de corrélation des variables explicatives.
-Sous l'hypothèse nulle d'othogonalité (non-multicolinéarité) des variables explicatives :
$FG\sim\chi^2_{1/2k(k+1)}$
-La règle de décision est données par :
- Si $FG<\chi^2_{1/2k(k+1)}$, l'hypothèse nulle n'est pas rejetée (pas de multicolinéarité)
- Si $FG>\chi^2_{1/2k(k+1)}$, l'hypothèse nulle est rejetée (présomption de multicolinéarité)

➔Solutions à la multicolinéarité : - *Augmenter le taille de l'échantillon* : augmenter le nombre d'observations pour diminuer la part des observations corrélées entre elles. Plus T est supérieur à k, moins la multicolinéarité a de chance d'être forte.
La régression ridge : l'une des colonnes de X est égale à une combinaison linéaire des autres. En ajoutant une constante sur la diagonale de X, cette multicolinéarité disparaît.

➔Test de stabilité des coefficients : -En robustesse de l'estimation menée : tester la stabilité des coefficients (en série temporelle principalement).
-S'assurer que les coefficients ne varient pas selon le contexte
-Etudier l'impact d'un changement structurel.
-Parmis d'autres approches, nous étudions ici le test de Chow.

➔3 types de structure de données : -Données en coupe instantannée (plusieurs individus à un instant donné)
-Données en série temporelle (plusieurs périodes pour un individu donné)
-Données de panel (plusieurs individus sur plusieurs périodes)

➔Pourquoi introduire les retards ? -une variable que l'on cherche à expliquer a de grandes chances de dépendre des valeurs passées des variables explicatives : modèles à retards échelonnés
-la variable expliquée peut dépendre de ses propres retards : modèles autorégressifs

➔Dans quel cadre utiliser les retards ?- analyse financière, prévision autorégressive, prévision échelonnée, persistance ou effet structurel d'un choc.

➔Nombre de retard et estimation : -Regarder la significativité des coefficients.- Estimation de plusieurs modèles avec plusieurs retards : on retient le modèle qui a les meilleurs critères d'information (minimiser AIC, BIC, SIC) et le plus grand $R2$ ajusté.

➔Modèles à retards échelonnés : PGL :
$Yt=\mu+\delta 0Xt+\delta 1Xt-1+\delta 2Xt-2+...+\delta hXt-h+\epsilon t$

➔Modèles à retards échelonnés autorégressifs : Dans de nombreux cas (séries financières par exemple) : la variable endogène dépend de ses propres retards. On parle de processus autorégressifs : $Yt=\mu+\phi_1Y_{t-1}+\phi 2Y_{t-2}+...+\phi pY_{t-p}+\epsilon t$
Le cas le plus connu étant celui de la marche aléatoire : la valeur d'aujourd'hui est la même que celle d'hier à un bruit prêt (finance ou météo). $Yt=\mu+\phi_1Y_{t-1}+\epsilon t$

➔Caractéristiques des séries temporelles :
Saisonnalité : caractéristique des séries temporelles où les données décrivent des variations régulières et prévisibles.
On parle aussi de périodicité : tout schéma se répétant avec un intervalle de temps régulier. (Climat, ventes, finance).

➔Stationnarité au second ordre : Modélisation en série temporelle : étude obligatoire de la stationnarité !!
-Nous nous intéresserons qu'à la stationnarité du second ordre / stationnarité faible : le plus courant
-Définissions de la stationnarité au second ordre d'un processus Yt :
- Le moment d'ordre 2 est fini et constant au cours du temps (homoscédasticité) : $E(Y2t)<\infty$, $\forall t$
- La moyenne est constante : $E(Yt)=m$, $\forall t$
- La covariance entre t et $t+h$ ne dépend que de h: $Cov(Yt,Y_{t+h})=\gamma_h$, $\forall t$

➔Fonction d'autocovariance, d'autocorrélation : La fonction d'autocovariance est définie, pour un processus Yt de variance finie, par :
$\gamma_h=Cov(Y_t,Y_{t+h})=E((Y_t-E(Y_t))(Y_{t+h}-E(Y_{t+h}))$
c'est la covariance d'une même variable entre deux dates.
De là, nous pouvons définir la fonction d'autocorrélation :
$\rho_h=\gamma_h/\gamma_0$
elle mesure les liaisons temporelles d'une série. On se sert de la fonction d'autocorrélation pour calculer les corrélogrammes.

➔Processus TS et Processus DS : Nous cherchons à caractériser un aspect essentiel des séries temporelles : la stationnarité
-Deux façons de caractériser la non-stationnarité en moyenne :
Processus TS (Trend Stationnary) : non stationnarité déterministe
Processus DS (Difference Stationnary) : non stationnarité stochastique
IMPORTANT : pourquoi faut-il que les séries soient stationnaires ??
Pas de stationnarité ⇒propriétés asymptotiques non valables (estimateurs à forte variance)
Pas de stationnarité ⇒⇒ possibilité de **régressions fallacieuses**

➔Processus TS : Dans le cas simple : tendance polynomiale d'ordre 1 : $Yt=\gamma+t\beta+\epsilon t$

➔Processus DS : Le processus DS est un processus non stationnaire que l'on peut stationnariser en appliquant une différence.
Le processus DS peut s'écrire (dans son cas le plus simple) :
$Yt=\gamma+Yt-1+\epsilon t$
Ainsi en appliquant une différence simple on a :n$Yt-Yt-1=\gamma+\epsilon t$
D'où, si ϵest stationnaire (un bruit blanc par exemple), la différence première est stationnaire.
Ainsi : espérance et variance d'un processus DS dépendent du temps t.
la non stationnarité d'un processus DS : déterministe (par l'espérance)
et stochastique : variance suit une tendance linéaire

➔Décomposition tendance-saisons ; traitement de la saisonnalité ; Il existe de nombreuses méthodes pour prendre en compte la présence de saisonnalité dans une série :
-souvent les séries sont déjà désaisonnalisées
-le très fameux filtre Hodrick Prescott : particulièrement utilisé en macro pour différencier la tendance du cycle

-méthodes basées sur la décomposition tendance/cycle, tendance/saisons

➔Détection de la non-stationnarité : Les tests pemettant de savoir si les séries sont stationnaires ou non sont appelés tests de racine unitaire $\frac{\phi_{ols}}{\sqrt{1/T}}$

-Il en existe un grand nombre. Nous présentons ici le plus répandu : Dickey-Fuller (DF) et Dickey-Fuller Augmenté (ADF)

-Dans les deux cas :

- $H0$: la série est non-stationnaire, c a d qu'elle contient au moins une racine unitaire
- $H1$: la série est stationnaire, c a d qu'elle ne comporte pas de racine unitaire

➔Test de Dickey-Fuller : L'idée du test : tester les différentes structures possibles de la série à laquelle nous nous intéressons en fonction des processus TS et DS. Précisément, trois modèles nous intéressent : Modèle (3) : $\Delta Yt=\alpha+\beta t+\phi Yt-1+\epsilon t$

Modèle (2) : $\Delta Yt=\alpha+\phi Yt-1+\epsilon t$. Modèle (1) : $\Delta Yt=\phi Yt-1+\epsilon t$

Soit : Modèle (3) : avec constante et tendance déterministe

Modèle (2) : avec constante, sans tendance déterministe

Modèle (1) : sans constante ni tendance déterministe

-On regarde ici si ϕ est significativement différent de 00 : nous utilisons des tables spécifiques pour cela (voir test ADF).

-si ϕ est significativement égale à 00 : on ne rejette pas $H0$ et la série est considérée comme non-stationnaire

-sinon, on considère que la série est stationnaire

-Pour savoir dans quel modèle nous devons-nous placer : on teste la significativité de la tendance et de la constante dans les modèles qui les contiennent (idem, nous devrons faire appelles à des tables spécifiques).

-Les modèles du test simple sont trop restrictif quant à la structure autorégressive des séries... d'où ADF

➔Comment interpréter les résultats du test ADF : Valeur du test DF

-P-valeur associée au test (hypothèse nulle : la série contient au moins une racine unitaire, donc si la p-value > 0.05, on ne rejette pas l'hypothèse nulle et la série contient au moins une racine unitaire)

-Un dictionnaire est donné ensuite avec les valeurs tabulées du test en fonction du nombre d'observations

-Le dernier élément est un objet contenant les résultats du modèle

-Nous devons différencier notre série et tester à nouveau la stationnarité : si la série différenciée est stationnaire, on dit qu'elle est intégrée d'ordre 1 et on peut travailler sur sa différence.

➔Modèles de la classe AR(I)MA : -On parle de processus a-théoriques : non fondé théoriquement. -Modélisation : valeurs passées de la série et des valeurs présentes et passées du terme d'erreur/bruit.

➔Modèles MA(q) : Un processus MA (Moving average) est un processus stationnaire vérifiant une relation du type:

$Yt=\epsilon t-\theta 1\epsilon t-1-...-\theta q\epsilon t-q$ avec $\theta i\in\mathbb{R}$ et $\epsilon t\ BB(0,\sigma^2\epsilon)$

pouvant être réécrit : $Yt=\Theta(L)\epsilon t$

avec $\Theta(L)=1-\theta 1L-...-\theta qL^q$ et Lq l'opérateur retard à l'ordre q

-La fonction d'autocorr. théorique permet l'identification de l'ordre q du processus **MA**.

-Elle s'annule pour tout $h>q$.

-Etant donné que nous disposons d'un échantillon, nous utiliserons la fonction d'autocorélation empirique pour vérifier cette propriété.

-Si le vrai proc générateur de donnée est **MA** alors nous devrions observer que les autocorr a un ordre $h>q$ s'annulent à cet horizon (utiliser test de significativité).

➔Modèle AR(p) : Un processus AR (Auto Regressif) est un processus stationnaire vérifiant une relation du type :

$Yt-\phi 1Yt-1-...-\phi pYt-p=\epsilon t$

-La fonction d'autoc partielle permet l'identification du nombre de retards p à inclure dans un processus $AR(p)$.

-En effet la fonction d'autocor partielle s'annule à partir du rang p+1.

-Comme pour le proc MA, il faut confirmer statistiquement la significativité des coefficients de corr partielles.

-La stat de test pour cette confirmation (pour de grands échantillons) est donnée par :

➔Modèles ARMA(p,q) : Un processus Yt, est un processus ARMA(p,q) s'il vérifie :

$Yt-\phi 1Yt-1-...-\phi pYt-p=\epsilon t-\theta 1\epsilon t-1-...-\theta q\epsilon t-q$

où $\phi i,\ \theta j\epsilon R$ and $\epsilon t\ BB(0,\sigma 2\epsilon)$

➔La méthodologie de BOX et jenkins : 4 étapes :

-identification des modèles, estimations du ou des modèles identifiés, validation des modèles sélectionnées : au final un modèle est retenu, prévision a un horizon t+h avec intervalle de prévision.

1-Identification : L'objectif de cette étape est de trouver les valeurs des paramètres p et q correspondant respectivement a la partie AR et MA du processus ARMA.

On trouve ces valeurs a l'aide des valeurs des fonctions d'autocorrelation et d'autocorrélation partielle empiriques. Voir les fonctions décrites plus haut. Si, par exemple, l'étude de ces fonctions nous amènes a trouver p = 1 et q = 2, notre ensemble de modèles de départ sera : ARMA(1, 2), ARMA(1, 1), AR(1), MA(1), MA(2)

2-estimation : Cette étape consiste a estimer les coefficients associés aux termes autorégréssifs et de moyenne mobile.

Pour l'estimation, il est possible d'appliquer la méthode des MCO si les erreurs sont non autocorrélées, sinon la méthode du maximum de vraissemblance ou la méthode des moindre carrés non linéaires.

3-validation : Vérifier si tous les coefficients d'un modèle sont significatifs (tests habituels). Si ce n'est pas le cas, renouvelez l'estimation sans le/les coefficients non significatifs.

-Tester les bonnes propriétés des résidus:

Non autocorrélation, homoscédasticité

-Il faut donc écarter les modèles à coeffs non significatifs ou ne respectant pas les bonnes propriétés résiduelles.

S'il demeurent plusieurs modèles à cet instant, on peut utiliser plusieurs méthodes pour sélectionner le "meilleur".

Les critères standards, Les critères d'informations

-Les critères standards, que l'on cherche à minimiser :

L'erreur absolue moyenne, la racine de l'erreur quadratique moyenne, l'erreur absolue moyenne en pourcentage.

-les critères d'information, que l'on cherche à minimiser :

AIC=$log\hat{\sigma}^2\epsilon+2(p+q)/T$

SIC=$log\hat{\sigma}^2\epsilon+(p+q)*(logT)/T$

➔Modèles de la classe ARCH : -Les modèles de types ARCHs ont été développés suite à des insuffisances des modèles linéaires. On distingue les processus à non linéarité en moyenne et les processus à non linéarité en variance. -Ce document se concentre sur les processus non linéaires en variance : modèles de la classe ARCH.

-Jusqu'à présent, lorsque nous souhaitions modéliser des séries temporelles, nous utilisions des modèles ARMA (modèles linéaires). 2 inconvénients majeurs aux processus ARMA : -"Les modèles ARMA ne permettent pas de prendre en compte les phénomènes d'asymétrie (caractérisant les cycles économiques, les coûts d'ajustements, les rigidités...), ni les ruptures de forte amplitude." -"Les modèles ARMA ne prennent en compte que les moments d'ordre 2 au travers de la fonction d'autocovariance, ce qui implique un exploitation incomplète de l'information contenue dans la série Xt. Les modèles non linéaires prennent en considération les moments d'ordre supérieur à 2 et permettent donc d'affiner l'étude de la dynamique de la série."

-Le concept clé de cette classe de modèle est la **variance conditionelle** : cad une variance conditionnée par le passé.

-Dans les modèles **GARCH** classiques, la variance conditionnelle est exprimée comme une fonction linéaire des carrés des valeurs passées de la série.

➔Faits stylisés : -La modélisation à l'aide des modèles de la classe ARMA n'est pas suffisante pour l'étude des séries financières.

-Les modèles ARMA peuvent être complètements caractérisés par leurs fonctions d'autocorrélations : seuls les deux premiers moments sont inclus. Si des régularités apparaissent lorsque l'on considère des moments d'ordre supérieurs, alors les modèles de la classe ARMA sont insuffisants.

-L'existence de régularités statistiques (faits stylisés), dans les séries financières notamment, rend difficile leurs modélisations.

-L'observation de ces régularités peut dépendre de la nature de la série et de sa fréquence. Les faits stylisés présentés ci-dessous sont valables pour la vaste majorité des séries de rendements boursiers en fréquence journalière.

1-non stationnarité des séries de prix. : Les séries de prix sont généralement proche de marches aléatoires. Les rendements, à l'inverse, sont la plupart du temps compatible avec les hypothèses de stationnarités au second ordre. Sur une fenêtre relativement longue, on peut considérer une magnitude constante.

2-Absence d'autocorrélation des rendements : L'étude des fonctions d'autocorrélation des rendements montrent qu'elles décroissent très rapidement (rapidement non significativement différentes de zéro). Si on se satisfait de cette observation, on dira que la série est proche d'un bruit blanc. Si on étudie les autocorrélations de séries à fréquence intra-journalière, on pourra observer, à l'inverse, des autocorélations significative (market microstructure effects).

3-forte autocorrélation des rendements au carré : Les rendements aux carrés ou en terme absolus sont fortement autocorrélés. Les rendements ne sont plus considérés comme une série de bruit blanc fort (r_t et r_{t+h} sont dépendants...)

4-Clusteres de volatilité : des périodes de turbulences (haute volatilité) sont suivies de périodes de calme (faibles volatilité). Ces périodes apparaissent souvent mais de façon non périodique. Les clusters de volatilité contredisent l'hypothèse d'homo (localement : les tests de stationnarités usuelles ne captent que la variance sur longue période, c'est pour cela que l'on conclu le plus souvent à la stationnarité des rendements.

5-Distributions aux queues épaisses : Si l'on trace la distribution empirique des rendements (ou la densité associée), on remarque qu'elle s'éloigne d'une distribution gaussienne (rejet H0 de Jarque-Berra par exemple). Elles sont le plus souvent *leptokurtiques* : retour vers zéro lent et présence d'une pic plus élevé à zéro (mesuré le plus souvent avec le kurtosis : $E[r^4]/V[r]^2$, avec des valeurs souvent supérieures à 3). Si on baisse la fréquence d'observation (intra -> daily -> weekly ...), cette distribution converge vers une distribution Gaussienne.

6-Effets de levier : Un effet asymétrique est observé. Une valeur passé négative ou positive du rendement impacte différentement la volatilité comptemporaine: Les rendements négatifs (correspondant à des baisses de prix) ont tendance à accroître la volatilité dans une plus large mesure que les rendements positifs (hausses de prix) de même ampleur.

7- Saisonnalité.

➔Modèle ARCH : Le processus ARCH(q) modélise la volatilité conditionnelle d'une série temporelle Yt avec une moyenne μ. L'erreur ϵt suit une distribution normale conditionnelle avec une variance $\sigma^2 t$, où $\sigma^2 t$ dépend du temps t selon l'équation

$\Sigma^2 t=\alpha 0+\sum\alpha i\epsilon^2_{t-i}$. Cette structure permet de capturer la dépendance temporelle de la volatilité. La variance conditionnelle peut aussi être exprimée comme $\sigma^2 t=\alpha 0+\alpha(L)\epsilon^2 t$, où $\alpha(L)=\sum L^i\alpha i$. Les paramètres $\alpha 0$ et αi sont des coefficients positifs. Ce modèle est utilisé pour caractériser l'hétéroscédasticité conditionnelle dans les séries temporelles financières.

➔Modèles alternatifs : les modèles ARCH et GARCH reproduisent bien une caractéristique importante observée sur les séries financière : la succession de période de calme et de crise. Ces modèles ont aussi l'avantage de permettre une étude étendue et simple des propriété statistiques et probabilistes des séries financière.

Cependant, certaines limites apparaissent : autocorrélations positives des carrés pour tous les lags (pourquoi pas négatives ?). Cette contrainte entraine également des difficultés pour l'inférence sur ces coefficients. La formulation standard ARCH/GARCH ne permet pas d'inclure des mesures de volatilité provenant d'autres variables (variables macro par exemple). Par construction, la variance conditionnelle ne dépend que du carré des variables (ϵ et σ) passées : les innovations positives et négatives passées ont le même effet sur la volatilité actuelle l'augmentation de la volatilité due à une baisse des prix est généralement plus forte que celle résultant d'une hausse des prix de même ampleur. Modèle "classiques" alternatifs : EGARCH, TGARCH, GARCH-M, etc...

➔Vecteur AutoRégressif VAR : $Xt=A+\sum BiXt-i+\epsilon t$

-Définitions et propriétés : Les séries doivent être stationnaires : si les séries ne sont pas stationnaires, on teste l'existence de cointégration. S'il n'y a pas de relation de cointégration, on peut alors stationnariser les séries pour appliquer le modèle VAR.

Fonction d'autocorrélation partielle : il est possible d'obtenir les fonctions d'autocovariance ou encore d'autocorrélation partielle. On les utilise pour déterminer l'ordre des processus AR. Dans le cadre des VAR, elles sont complexes à calculer et on aura recours à des critères d'informations pour sélectionner un ordre le plus approprié.

➔Tests de spécification : *1-Selection du nombre de retards :* critères d'informations (AIC, SIC, HQ) : l'idée est de faire tourner plusieurs modèles avec différents ordres de retard, et de retenir celui qui minimise les critères d'informations (eux-mêmes basés sur des estimateurs de la matrice des résidus). *Note 1* : les critères d'information SIC et HQIC favorisent un estimateur convergent de p (le "vrai" nombre de retards), AIC conduit vers un estimateur efficace de p (i.e. sans biais et de variance faible) et le FPE (Final Prediction Error) tente de minimiser l'erreur de prévision.

Note 2 : On peut également avoir recours au test du rapport de maximum de vraisemblance pour acter le nombre de retards choisis.

L'idée est de tester la significativité des paramètres associés au retards d'ordre p+1.

2-Causalité : on ne s'intéresse ici qu'à la causalité au sens de Granger (1969). On dit que X cause Y si l'erreur de prévision de Y est moindre lorsque l'on connait conjontement le passé de Y et X, que lorsque l'on ne connait que le passé de Y.

-Tester la non-causalité : on peut, à partir de l'expression ci-dessus, mener un test de causalité. $\xi=T\times C_{X\to Y}\sim\chi^2_{rT-rp}$

où T est le nombre de périodes, r le nombre de contraintes, p le

$$C_{X\to Y}=\frac{det(V_\epsilon(Y_t|Y_{t-1}))}{det(V_\epsilon(Y_t|Y_{t-1},X_{t-1}))}$$

Règle de décision :

$\xi>\chi^2_{rT-rp}$, on accepte l'hypothèse nulle d'absence de causalité

$\xi\leq\chi^2_{rT-rp}$, on rejette l'hypothèse nulle d'absence de causalité

-A quoi ca sert ? Il peut être intéressant d'étudier la causalité pour justifier le recours à un modèle VAR. En effet, une hypothèse sous-jacente de cette spécification est que les variables s'entre-déterminent. Par conséquent, dans un modèle bivarié VAR(p) par exemple, on pourra regarder la causalité dans les deux sens (de x_1 vers x_2, et vice versa) pour appuyer la pertinence de ce modèle.

3-Exogénéité : concept relativement complexe et pas de manière simple de mener des tests d'exogénéité sous python pour le moment. L'idée est que le bloc de variables ($X_{1,t-1},X_{1,t-2},...,X_{1,t-p}$) est exogène au bloc ($X_{2,t-1},X_{2,t-2},...,X_{2,t-p}$) si le fait de rajouter $X1,t$ n'améliore pas significativement la détermination des variables $X2,t$.

➔Note : on peut estimer un VAR uniquement sur des séries stationnaires, sans saisonnalité, sans tendance, par MCO ou maximum de vraisemblance.

➔Interprétation : Nous avons principalement deux outils d'analyse des VAR : les fonctions de réponses impulsionnelles (IRF) et l'analyse de la décomposition de variance. **Fonctions de réponses impulsionnelles** : l'idée ici est de s'intéresser à l'effet d'un choc sur l'erreur d'une des équations (choc exogène) sur la variable endogène de cette équation, mais également sur les autres variables, *via* la structure dynamique du système VAR. L'idée des IRFs peut être résumée comme suit : On se sert de la réécriture d'un modèle VAR sous sa forme VMA (Vector Moving Average) *via* le théorème de Wold. Cela permet d'isoler l'effet des chocs/des erreurs. Il est également possible de passer par l'écriture du VAR canonique sous sa forme structurelle : techniquement, une matrice de passage P est appliquée à l'ensemble du modèle. Cette matrice de passage est déduite de la relation entre les erreurs (ϵ) et les chocs passés structurant les variables contemporaines (ω). On utilise donc un estimateur de cette matrice de passage pour obtenir un estimateur de chocs interprétables économiquement.

Les IRF ne sont pas interprétables lorsque les **résidus des différentes équations sont corrélés entre eux**. Par conséquent, l'application de P est faite de sorte que les résidus transformés ($\omega=P$) soient orthogonalisés (non-corrélés). En supposant que les chocs ω sont non-corrélés entre eux et de variance unitaire, on peut alors recourir à la décomposition de Cholesky : la matrice P est alors la matrice de passage de la diagonalisation de la matrice de variance covariance des erreurs, $\Sigma\epsilon$ (ou matrice dont les vecteurs sont les vecteurs propres de $\Sigma\epsilon$). Aucune condition économique n'est posée sur l'ordre des variables, seulement qu'elles soient **ordonnées de la plus exogène vers la plus endogène** (cf causalité pour trouver l'ordre). Afin d'appliquer une approche par **VAR structurel**, il est nécessaire de poser des ***a priori* économiques** sur l'ordre d'influence des variables. C'est selon cet ordonnément que l'on classe les variables dans le système avant d'estimer P.

➔Décomposition de variance : l'analyse des IRF est généralement complétée par une étude de la décomposition de variance des erreurs de prévisions. Ici, l'objectif est de calculer la contribution de chacun des chocs à la variance de l'erreur. On écrit la variance de l'erreur de prévision en fonction de la variance de l'erreur attachée à chacune des variables du système ; puis on rapporte cette quantité à la variance totale.

➔Cointégration : La régression entre deux séries non stationnaires entraine l'invalidité de certaines lois asymptotiques. Modélisations VAR et ARMA ne fonctionnent pas sur les séries non-stationnaires

Stationnarisation *via* la différence : -Masque les propriétés de long terme des séries. -Lorsque deux séries non-stationnaires sont liées par une relation de long terme, on va parler de **cointégration** que l'on modélise au travers de modèles à **correction d'erreur**. Ce type de relation n'est pas pris en compte par les modèles classiques.

-Définition : À court terme, les deux variables peuvent diverger, mais elles gardent une relation commune de long terme : elles sont linéaires autour d'une tendance. $zt=Xt-\alpha Yt$

Définitions: Econométrie : est la science de la mesure de l'économie. C'est donc une branche des sciences économiques qui fait appel à 4 disciplines : - La théorie économique - La statistique - Les mathématiques - L'informatique **Articulation multidisciplinaire :** 1- On dégage un problème et on établit une théorie économique permettant de l'expliquer de façon simplifiée (modèle théorique) 2- On extrait des données sur ce problème. 3- Après une étude statistique des données, on fait appel à un modèle mathématique approprié permettant de valider ou non la théorie économique. 4- Pour mener cette analyse, il convient de faire appel à certains outils informatiques. **Python :** est un langage informatique, donc un grand dictionnaire de commandes que l'on combine entre elle pour former des « phrases » permettant de produire un résultat. **Exemple macroéconomique: la loi d'Okun** La relation entre le chômage et la croissance économique : Lorsque la croissance est faible, les entreprises produisent peu et anticipent une faible production à venir. Il y a alors peu d'embauche et le taux de chômage tend à augmenter. Inversement, lorsque le chômage augmente il est probable que les entreprises produisent moins et donc que la croissance baisse. **Théorie économique :** on pourrait penser qu'il existe une relation négative entre les variations du taux de chômage et celles de la croissance : $\Delta Ut=a+b\times Gt+\epsilon t$ avec U : le taux de chômage, G : le taux de croissance du PIB et a, b, c sont les paramètres à estimer.(Selon la théorie économique, nous devrions à partir de données observées obtenir un b négatif. **Len() :** permet de compter le nombre de caractères, aussi permet de compter le nombre d'éléments dans une liste , tuple, dicos. Ne fonctionne pas directement pour les nombres, il faut d'abord les convertirs en String grace à « str() » \ **Type() :** utile pour connaitre ou se rappeler du type d'une variable **% :** est un opérateur sert à obtenir le reste d'une division Les **Boucles : La boucle FOR: permet** d'itérer / de parcourir à travers une variable qui le permet (des 'itérables') : str, list, tuple, dict. **La boucle WHILE :** permet de répéter une instruction plusieurs fois, tant qu'une condition est vraie. Les boucles vont nous permettre d'introduire la fonction range() qui permet de parcourir les entiers avec un certain pas. **Les principaux packages utilisés :** Numpy : pour faire des calculs Pandas : pour manipuler les bases de données (très important) StatsModels : pour les modèles d'économétrie ScikitLearn : pour le machine learning Matplotlib et Seaborn : pour les graphiques : Arch : pour mener des estimations de modèles ARCH. **La formalisation** -Version en différence : $\Delta ut=a+bgt+ct$ où gt est la croissance du PIB -Capture la corrélation contemporaine entre la croissance du PIB et les variations du taux de chômage. -La valeur du coefficient β attendue est négative : une augmentation du taux de croissance rapide serait associée à une chute du taux de chômage. Une croissance molle ou négative serait, à l'inverse associée avec un taux de chômage s'accroit. -La relation - a b donne le taux de croissance associé à un taux de chômage stable. Ou autrement dit, quel rythme de croissance est nécessaire pour maintenir un certain taux de chômage. **Estimation des coefficients** -On procède au calcul de : X'X -Puis (X'X)-1 -Il ne nous reste plus qu'à multiplier par X'Y pour obtenir β^: β^= (X'X**), Estimateur de la variance des erreurs**

$$\hat{\sigma}^2_\epsilon=\frac{e'e}{T-k-1}$$

Analyse de la variance Equation de l'analyse de la variance y'y = β^'x'y + e'e Où : -La somme des carrés totale (SCT) est donnée par : y'y = $\sum$t y²t = $\sum$t (Yt - $\bar{Y}$)² -La somme des carrés expliquée (SCE) est donnée par : β^'x'y -La somme des carrés des résidus (SCR) est donnée par : e'e = $\sum$te²t .On peut retrouver la mêle équation pour le cas de variables non centrées Coefficient de détermination et coefficient de détermination corrigé : Le coefficient de détermination (R^2) : Mesure la part de la variance totale expliquée par les variables explicatives et permet de juger de la qualité de l'ajustement du modèle. R^2 = SCE /SCT = 1- SCR/ SCT Prévision Estimation d'un modèle Permet d'explique ce qui est observé : pouvoir explicatif (en échantillon) Permet d'émettre des prévisions : pouvoir prédictif du modèle (hors échantillon)

Rappels sur le modèle de régression multiple :

- Modèle de régression multiple

Y= Xβ+ε

- Matrices :

  - Le vecteur colonne de la variable expliquée : **Y ∈ $M_{n,1}$ (IR)**
  - La matrice des variables explicatives, dont la première colonne est composée de 11 pour la constante : **X ∈ $M_{n,k+1}$ (IR)**
  - Le vecteur colonne des paramètres à estimer : **β ∈ $M_{k+1,1}$ (IR)**
  - Le vecteur colonne des termes d'erreur : **ε $M_{n,1}$ (IR)**

- Hypothèses :
  - X est non aléatoire : **R (ε,) X)=0**
  - X est de plein rang : **Rang(X)=k+1**
  - L'erreur est nulle en moyenne : **E(ε)=0**
  - Non autocorrélation et homoscédasticité des erreurs : **E(εε')=σ²I**

Normalité des erreurs : $\epsilon\sim N(0,\sigma^2\ \epsilon I)$ • Résidus : e=Y−Xβ^ • Estimateurs MCO : $\beta\hat{}=(X'X)-1X'$