Projet d’économétrie

Les facteurs de ventes de Walmart

*Joudy BENKADDOUR et Ryhan CHEBREK*

*Partie 1*

*« Je déclare su l’honneur que ce mémoire a été rédigée de ma main, sans aide extérieure non autorisée, qu’il n’a pas été présenté auparavant pour évaluation et qu’il n’a jamais été publié, dans sa totalité ou en partie.*

*Toutes parties, groupes de mots ou idées, aussi limités soient-ils, y compris des tableaux, graphiques, cartes etc. qui sont empruntés ou qui font référence à d’autres sources bibliographiques sont présentés comme tels, sans exception aucune. »*

1. Estimation du modèle initiale
2. Test de significativité globale

Une image contenant table

Description générée automatiquement

*H0 : â0=â1=…=âi=…= 0*

*H1 : Il existe au moins un coefficient significativement différent de 0 : âi≠0*

Ce test permet de montrer qu’il existe des variable exogènes expliquant les ventes du Walmart.

Pour effectuer le test : Modèle => MCO => Définir variables dépendantes et explicatives => Valider

Dans notre modèle, on obtient F\* = 4.815 et la P-value est de 0.000436 .

Ainsi on rejette l’hypothèse H0 au seuil de 5% donc il existe au moins un coefficient significativement différent de 0.

Notre modèle est donc utilisable car au moins une variable peut expliquer le volume de vente du magasin.

1. Test de significativité individuelle pour toutes les variables

*H0 : âi = 0 âi significativement = 0 : La variable n’explique pas les ventes*

*H1 : âi ≠ 0 âi significativement différent de 0 : La variable explique les ventes*

Ce test permet de montrer pour chaque variable si elle explique les ventes. Nous allons le faire pour les deux premières variables c’est-à-dire le type de semaine (1 si vacances et 0 sinon) et la température.

Pour effectuer le test : Modèle => MCO => Définir variables dépendantes et explicatives => Valider

Pour notre modèle :

Pour â1 : La probabilité critique est de 7,23% donc on ne rejette pas H0 au seuil de 5%

Pour â2 : La probabilité critique est de 2,03% donc on rejette H0 au seuil de 5%

Ainsi on conclut que le type de semaine (vacance ou non) explique significativement le volume de vente tandis que la température n’explique pas significativement le modèle.

1. Test de significativité à valeur fixée (unilatéral)

*H0 : a4 = 10000*

*H1 : a4 >10000*

Ce test nous permet de savoir si l’on peut contraindre une variable à une valeur sans nuire à la qualité du modèle.

Pour effectuer ce test : Modèle => MCO => Définir variables dépendantes et explicatives => Valider => Tests => Restrictions linéaires => b[5]=10000 => Valider

Pour notre modèle : La probabilité critique vaut 33,02%.

Or Gretl calcule la p-value bilatéral donc on divise par 2 pour obtenir la valeur de la p-value unilatéral.

p-value bilatéral = 33,02% donc p-value unilatéral = 16,01%, ainsi on accepte H0 au seuil de 5%.

On en conclu que forcer le coefficient â4 = 10000 ne diminue pas drastiquement la qualité du modèle.

Une image contenant table

Description générée automatiquement

1. Test de combinaison linéaire (unilatéral)

*H0 : â5-5\*â4 = 0*

*H1 : â5-5\*â4 < 0*

Ce test nous permet de vérifier si l’on peut trouver des relations linéaires entre les variables sans nuire à la qualité du modèle.

Pour effectuer ce test : Modèle => MCO => Définir variables dépendantes et explicatives => Valider => Tests => Restrictions linéaires => b[6]-5\*b[5]=0 => Valider

Une image contenant texte

Description générée automatiquement

Ici la p-value vaut 94%.

Or Gretl calcule la p-value bilatéral donc on divise par 2 pour obtenir la valeur de la p-value unilatéral.

p-value bilatéral = 94% donc p-value unilatéral = 47%, ainsi on accepte H0 au seuil de 5%.

On en conclu que forcer le coefficient â5 = 5\*â4 ne diminue pas drastiquement la qualité du modèle.

1. Test par analyse de la variance
2. Test égalité conjointe de 2 coefficients

*H0 : â1 = 90000 et â2 = -2000*

*H1 : 3 possibilités*

Ce test nous permet de savoir si l’on va pouvoir conjointement fixer deux variables sans nuire à la qualité du modèle.

Il existe 3 causes de rejets possible de H0 : â1 ≠ 90000 ou â2 ≠ -2000 ou â1 ≠ 90000 et â2 ≠ -2000 .

Pour effectuer ce test : Modèle => MCO => Définir variables dépendantes et explicatives => Valider => Tests => Restrictions linéaires => b[2]= 90000 et b[3]= -2000 => Valider

Une image contenant texte

Description générée automatiquement

Ainsi on a p-value = 98,48% , alors on accepte H0 au seuil de 5%.

Cela signifie que l’on peut garder â1 = 90000 et â2 = -2000 sans trop dégrader le modèle.

1. Test d’ajout de variable

**H0 : SCR = SCR1**

**H1 : SCR ≠ SCR1**

Ce test nous permet de savoir si l’on peut ajouter ou supprimer des variables sans nuire à la qualité du modèle.

On a décidé d’omettre les variables : inflation, température, type de semaine qui sont les coefficients avec les plus grandes probabilités critiques donc les variables qui expliquent le mieux les ventes.

Pour effectuer ce test : Modèle => MCO => Définir variables dépendantes et explicatives => Valider => Tests => Omettre les variables => Sélectionner inflation, température et type de semaine => Valider

Une image contenant texte

Description générée automatiquement

On a donc une p-value très faible, on rejette l’hypothèse H0 au seuil de 5%.

Assez logiquement, si l’on enlève les variables explicatives les plus pertinentes, le modèle se détériore drastiquement la qualité du modèle.

1. Test de Chow

*2 août: USA - Après des semaines de bataille, le Congrès autorise un relèvement du plafond de la dette (plus de 14.500 mds de dollars), évitant au pays un défaut de paiement. Avec un taux de chômage de 9,2% et une croissance faible (1,3%), les Etats-Unis se voient privés pour la première fois de leur "AAA" par l'agence de notation Standard and Poor's.*

Ce test nous permet de savoir le modèle reste stable lorsqu’on scinde l’échantillon.

Nous avons décider de scinder notre modèle en 2 parties : la première avant le 2 août 2011 et la deuxième partie après cette date. Cette date nous semble pertinente car il nous permet de voir si les décisions du congrès ont eu des conséquences macroéconomiques sur les ventes et l’influence des coefficients des variables explicatives.

*H0 : SCR – (SCR1 – SCR2) = 0*

*H1 : SCR – (SCR1 – SCR2) ≠ 0*

Pour effectuer ce test: Modèle => MCO => Définir variables dépendantes et explicatives => Valider => Tests => Test de Chow => Scindé à 82 => Valider

Une image contenant table

Description générée automatiquement

On obtient une p-value de 35,48%, ainsi on accepte H0 au seuil de 5%, cela signifie que les coefficient de chaque variable peuvent être considérés égaux avant et après la date du 11 août 2011. Le modèle est stable avant et après ce 11 août.

Ainsi, la décision du congrès n’a pas eu de conséquences directes sur les ventes du Walmart et sur l’importance des coefficients des variables explicatives.

1. Test de contrainte sur les coefficients

*H0 : â1=â5 et â4=15000*

*H1 : Plusieurs possibilités*

Ce test permet de savoir si l’on peut contraindre des coefficients sans nuire au modèle.

Il y a possiblement 3 raisons de rejeter H0 : â1 significativement différent de â5 ou â4 significativement différent de 15000 ou bien les 2 à la fois.

Pour effectuer ce test : Modèle => MCO => Définir variables dépendantes et explicatives => Valider => Tests => Omettre les variables => Sélectionner inflation, température et type de semaine => Valider

Une image contenant texte, personne, capture d’écran

Description générée automatiquement

Ainsi, notre test nous renvoie une p-value = 91,73%, on accepte donc H0 au seuil de 5%

Cela signifie que nos hypothèses ne détériore pas significativement le modèle.

1. Intégration d’une variable indicatrice

…

1. Elimination de variable

Nous allons éliminer les variables des moins significatives aux plus significatives donc des probabilités critiques les plus élevés au moins élevés.

Une image contenant table

Description générée automatiquement

On commence d’abord par supprimer la variable « Prix de gasoil »

Une image contenant table

Description générée automatiquement

On remarque les probabilités critiques des autres variables ont été modifiés tout comme leur coefficient. Cependant on a toujours des variables peu significative (p-value > 0.05).

La plus petite p-value correspond au taux de chômage, on supprime donc cette variable.

Une image contenant table

Description générée automatiquement

On a maintenant quasiment que des variables significatives même si la p-value du type de semaine est très légèrement supérieur à 5%.

Nous allons quand même la supprimer.

Une image contenant table

Description générée automatiquement

Il nous reste maintenant que 2 variables qui sont belle et bien significatives avec des p-value bien inférieur à 5%.

On peut en conclure que l’on peut expliquer les ventes avec la température et le taux d’inflation. Avec les coefficients, on peut donc avancer qu’une hausse des températures cause une baisse des ventes tandis qu’une hausse de l’inflation conduit à une hausse des ventes.

La relation entre température et vente du magasin est un peu biaisé car les périodes de fêtes où l’on consomme le plus sont en hiver.

Partie 2

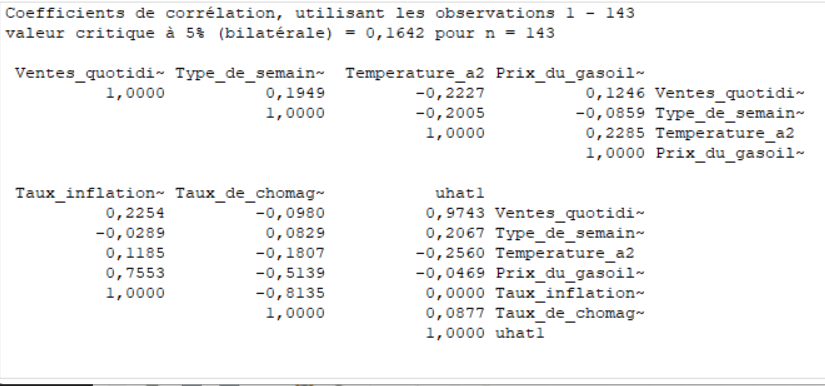
1. Régression par étage

Pour commencer, on utilise les coefficients de corrélation entre les variables exogènes et la variable endogène et on retient la variable exogène avec la plus forte corrélation en valeur absolue avec les ventes du magasin qui se trouve être ici le taux d’inflation.

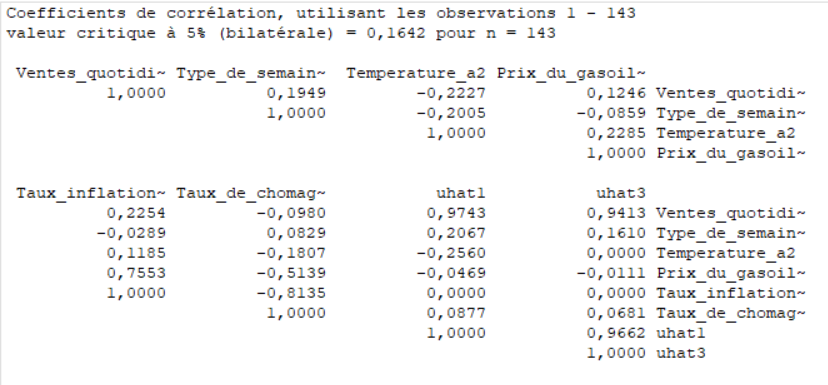
Une image contenant texte

Description générée automatiquement

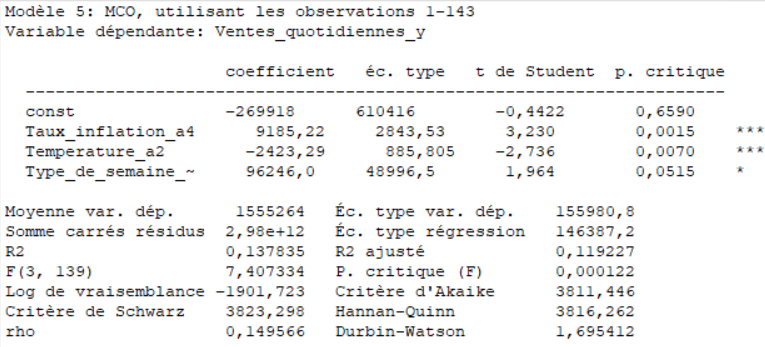
Ensuite, on effectue un MCO entre les ventes et le taux d’inflation et on retient le résidu pour ensuite regarder les corrélations entre ce résidu et les autres variables exogènes.



Ici la variable à retenir est la température. On fait de nouveau une MCO avec la température et l’inflation puis on retient le résidu pour avoir la corrélation de ce dernier avec les autres variables exogènes.



La corrélation la plus forte avec le résidu est la variable type de semaine, cependant la MCO avec cette variable indique qu’elle n’est pas significative ainsi le modèle à retenir est le modèle précèdent avec les deux variables taux d’inflation et température.



On obtient au final le même résultat qu’avec l’élimination de variable.

Une image contenant texte

Description générée automatiquement

1. Autocorrélation des erreurs
   1. Examen visuel du graphique des résidus



Il semble y avoir une légère autocorrélation négative des résidus, ce qui reste à confirmer avec les prochains tests.

* 1. Test de Durbin-Watson

Avant d’effectuer le test, on vérifie les 4 hypothèses :

* Le nombre d’observation est supérieur à 15 (ici 143)
* Le modèle doit comporter un terme constant ou bien les données sont centrés sur leur moyenne (nous avons bien un terme constant)
* Un modèle en série temporelle ou bien si en coupe instantanée, les données doivent être triés (nous avons une série temporelle)
* La variable explicative ne doit pas figuré en tant que variable explicative décalé : yt = b\*yt-1 (c’estbien le cas)

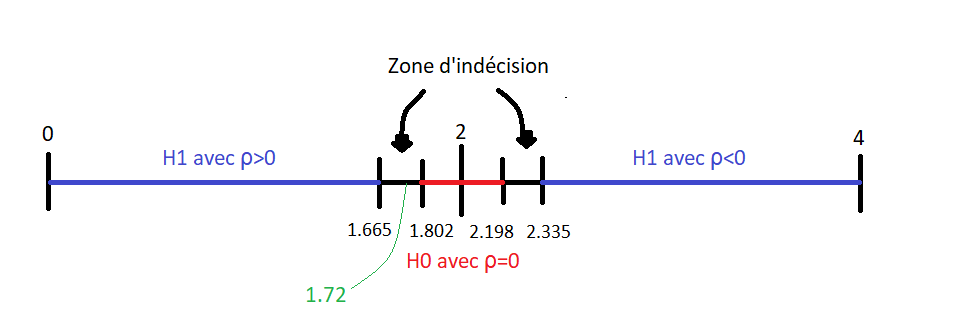
On pose : εt = εt-1 + vt avec -1<ρ <1 et vt suit une loi normale centré à écart type constant.

*H0 : ρ = 0*

*H1 : ρ ≠ 0*

On obtient sur Gretl que le MCO renvoie 1,72 pour Durbin-Watson, on recherche une table sur internet avec n = 143 et k = 5 (le nombre de variable explicatives) on trouve d1 = 1,665 et d2 = 1,802.

On calcule donc 4-d1 = 2,335 et 4-d2= 2,198.



On se retrouve donc dans la zone d’indécision avec H0 : ρ > 0 et H1 : ρ < 0 or comme on se retrouve plus proche de d1, on préfère donc l’hypothèse selon laquelle ρ > 0 donc H1 .

Il y a une légère autocorrélation des résidus comme pressenti avec le graphique de résidu, ainsi on subodore qu’il existe une corrélation des erreurs.

Il convient désormais de corriger cette autocorrélation. Il existe deux principales causes de l’autocorrélation des erreurs, la première l’oublie d’une variable explicative et la seconde la mauvaise forme d’une variable (log, niveau…).

* 1. Correction

1. Cochrane-Orcutt

Pour cela on commence avec la méthode Cochrane-Orcutt qui est une méthode itérative. Modèle => Série temporelles univariés => Erreur AR => AR(1) => Cochrane-Orcutt

Une image contenant table

Description générée automatiquement

On estime ρ=0,15 d’après le test.

1. Hildreth-Lu

Il existe une deuxième méthode de correction Hildreth-Lu (méthode de balayage) :

Modèle => Série temporelles univariés => Erreur AR => AR(1) => Hildreth-Lu



Une image contenant texte, reçu

Description générée automatiquement

On obtient exactement les mêmes résultats avec les deux méthodes de correction.

1. Hétéroscédasticité
2. Test de Goldfelt-Quandt

Pour commencer, on va prendre comme variable exogène, les variables les plus significatives définis aux questions précédentes donc : Type de semaine, la température et le taux d'inflation. Ce test permet de savoir s’il existe une hétéroscédasticité structurelle du modèle.

*H0 : SCR1 = SCR2*

*H1 : SCR1 ≠ SCR2*

Scindons l’horizon temporel en trois partie avec n1 = 47 de 1 à 47 et n2=47 de 96 à 143 et le reste de 48 à 95 pour calculer SCR sur la première et la dernière partie afin de bien déterminer le test.

Une image contenant texte

Description générée automatiquement

SCR1 = 1,18\*1012

Une image contenant texte

Description générée automatiquement

SCR2 = 1,02\*1012 et donc SCR2 < SCR1, ce qui nous permet d’écrire correctementle F\*.

F\* = (SCR1/ddl1)/(SCR2/ddl2)= SCR1/SCR2 = 1,156 car nous avons pris soin d’avoir n1=n2.

Or le Flu à 5% = 1,62 d’après Excel.

Ainsi on accepte H0 à 5%, il n’y a pas d’hétéroscédasticité structurelle dans notre modèle.

1. Test de White

On garde toujours les trois variables les plus significatives pour ce test permettant de savoir s’il existe une hétéroscédasticité dû à une variable explicative.

**H0 : a1=a2=a3=… (aucune relation entre résidu et variable explicative)**

**H1 : Il existe au moins une valeur différente de 0 et donc une relation entre le résidu et au moins une variable explicative.**

Une image contenant texte

Description générée automatiquement

On regarde la probabilité critique la plus basse soit 39,28% pour le produit x2\*x3, on conserve donc H0 au seuil de 5% ainsi il n’y a pas d’hétéroscédasticité dû aux variables explicatives.

Mais nous allons tout de même corriger le modèle pour notre bon plaisir.

La plus petite probabilité critique  et donc plus forte relation avec le carré des résidus est le produit croisée x2\*x3. On crée une variable POND = 1/abs(X2\_X3).



On utilise maintenant le MCP avec le facteur POND et on obtient :

Une image contenant table

Description générée automatiquement

Le MCP a supprimé la variable type de semaine pour cause de colinéarité parfaite, ce qui nous permet d’avoir de nouveaux coefficients corrigés pour la température et le taux d’inflation.

Pour conclure, ce projet nous a permis d’effectuer plusieurs tests afin de savoir si l’on peut simplifier le modèle sans pour autant le dégrader, si le modèle est stable ou encore s’il y a un problème d’hétéroscédasticité ou d’autocorrélation. Tout cela afin d’avoir le modèle le plus précis possible et donc de pouvoir dans notre cas exprimer le volume de vente en fonction de toutes nos variables exogènes.

FIN.