

Rapport

Quels facteurs ont influencé l'évolution du volume des actions vendues aux États-Unis entre 1980 et 1995 ?

Présenté par Mahdi BEN NACIB, Ilyes BOUGHERARA, Nassim BRIKH et Yanis DIDOUCHE
L3 E2AD Groupe A2

2024-2025 Econométrie

1 - Introduction	2
Tableau MCO 1 : Totalité des variables	2
2 - Les données	4
A - Sources et descriptions de la base de données	4
B - Statistiques descriptives et représentations graphiques	5
Tableau 2: Statistiques Descriptives	5
Graphique 1 : volume en fonction de la période	5
Graphique 2 : log_volume en fonction de la période	5
Graphique 3 : log_volume en fonction du taux de trésorerie des USA	6
Graphique 4 : log_volume en fonction de l'indice des prix de S&P500	6
Graphique 5 : log_volume en fonction de log_gdp	6
3 - Modèle économique et économétrique	8
4 - Estimations et Résultats	9
Tableau 3 : MCO sous respect de H4	9
Tableau 4 : Test de White	10
Tableau 5 : Synthèse des tableaux	11
Tableau 6 : Matrice de corrélation	12
5 - Résultats et Conclusion	13
6 - Bibliographie & Sources	14

1 - Introduction

Notre problématique se concentre sur une thématique financière et macroéconomique. Nous avons choisi d'expliquer le volume d'actions vendues sur le marché financier aux États-Unis entre 1980 et 1995, une période marquée par une libéralisation financière et économique croissante. Ce contexte correspond à la fin de la répression financière et au lancement des « 3D » : déréglementation, désintermédiation et décroissement. À partir des années 1990, la forte croissance économique des États-Unis peut s'expliquer par un développement significatif et une globalisation financière libérale. Le succès de cette « nouvelle économie » a marqué une période de croissance soutenue, consolidant le statut des États-Unis en tant que première puissance économique mondiale.

Ce sujet a déjà été étudié par un certain nombre d'économistes et d'universitaires, principalement en relation avec la dynamique des marchés financiers et les liens entre les volumes de transactions et les rendements, parmi eux, Lawrence Harris, Richard Roll, Robert Shiller et Eugène Fama.

Notre méthodologie sera la suivante : nous commencerons par estimer le modèle à l'aide des moindres carrés ordinaires (MCO). Après analyse des résultats, nous testerons la présence d'hétéroscédasticité et d'autocorrélation, et appliquerons les corrections nécessaires si ces phénomènes sont détectés.

Cette démarche inclut l'analyse du respect des cinq hypothèses des MCO ainsi que l'évaluation de la pertinence de nos variables explicatives. Il est important de souligner qu'il ne s'agit pas simplement de maximiser le nombre de variables aux coefficients significatifs. En effet, même si certains coefficients individuels ne sont pas significatifs, il est possible que, collectivement, ces variables expliquent une part substantielle de la variation de la variable dépendante.

Tableau MCO 1 : Totalité des variables

Fichier Édition Tests Enregistrer Graphiques Analyse LaTeX					
Modèle 1: MCO, utilisant les observations 1980:01-1995:09 (T = 189)					
Variable dépendante: volume					
	coefficient	éc. type	t de Student	p. critique	
const	-646446	303068	-2.133	0.0343	**
period	327.134	153.681	2.129	0.0346	**
sp500	16.3276	2.32123	7.034	4.04e-11	***
tbill	-85.6901	42.5348	-2.015	0.0454	**
long	179.631	75.4171	2.382	0.0183	**
gdp	-4.35103	1.61794	-2.689	0.0078	***
cconf	12.6418	5.26377	2.402	0.0173	**
cexpect	-24.1336	6.13170	-3.936	0.0001	***
csent	25.8015	9.63712	2.677	0.0081	***
Moyenne var. dép.	3283.593	Éc. type var. dép.	1772.013		
Somme carrés résidus	59449054	Éc. type régression	574.6934		
R2	0.899294	R2 ajusté	0.894819		
F(8, 180)	200.9236	P. critique (F)	1.78e-85		
Log de vraisemblance	-1464.444	Critère d'Akaike	2946.888		
Critère de Schwarz	2976.063	Hannan-Quinn	2958.707		
rho	0.421758	Durbin-Watson	1.156039		

Voici nos résultats bruts, nous ajusterons les variables à conserver et à omettre dans la prochaine partie.

Pour discuter de cette problématique, nous allons d'abord présenter nos données, ensuite nous en déduirons notre meilleur modèle économique et économétrique afin d'obtenir des résultats ainsi que des estimations pertinentes.

Pour finir nous analyserons nos résultats et conclurons.

2 - Les données

A - Sources et descriptions de la base de données

Notre base de données provenant de Gretl dont les sources ont été compilées par Brian Wampler d'après les sources : United States Statistical Abstract, Standard and poor quarterly reports, citibase data series, et l'université d'Alabama.

Nous avons une proposition de 8 variables explicatives qu'il va falloir trier en fonction de leur pertinence afin d'estimer notre meilleur modèle.

Nous avons une série temporelle avec un pas temporel d'1 mois, étalée sur 15 ans et 9 mois soit un ensemble d'observations de 189 et les variables explicatives :

- 1) **volume** : Le volume d'actions rapporté par Wall Street (quantitative discrète)
- 2) **SP500** : L'indice des prix des actions ordinaires (quantitative continue)
- 3) **tbill** : Le taux de trésorerie des USA sur 3 mois en pourcentage (quantitative continue)
- 4) **long** : Le rendement des obligations de trésorerie sur 10 ans (quantitative continue)
- 5) **gdp** : Le PIB (quantitative continue) (gdp)
- 6) **cconf** : L'indice de confiance des consommateurs (quantitative continue)
- 7) **cexpect** : L'indice des prévisions des consommateurs (quantitative continue)
- 8) **csent** : L'indice du sentiment des consommateurs (quantitative continue)

Nous avons décidé de retirer la 8ème variable explicative (**csent**) de notre modèle, par manque de pertinence et une potentielle corrélation avec la 6ème variable (**c_expect**)

Nous observons la possibilité de mettre "*period*" en variable explicative.

Nous avons jugé intéressant de mettre la période en variable explicative car le volume d'actions vendues peut fortement dépendre de la période en fonction du contexte financier et macroéconomique, comme par exemple le krach boursier de 1978 qui se situe dans notre période étudiée.

B - Statistiques descriptives et représentations graphiques

Tableau 2: Statistiques Descriptives

	Moyenne	Médiane	Minimum	Maximum
period	1987.4	1987.1	1980.0	1995.1
sp500	280.44	269.05	102.97	578.77
tbill	7.3996	7.0700	2.8400	16.295
long	9.3641	8.9200	5.9000	14.140
cconf	87.667	90.400	47.300	120.70
cexpect	93.203	95.500	50.000	124.30
l_volume	7.9331	8.0790	6.5132	9.0865
l_gdp	7.3573	7.3853	6.8096	7.7968

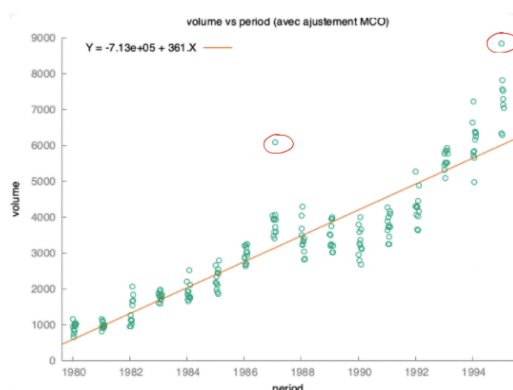
Au vu des statistiques descriptives du tableau 2, on peut observer par rationalité que tous nos résultats sont cohérents.

Nous avons mis les variables *volume* et *gdp* car nous avons estimé qu'elles seraient plus pertinentes pour les interprétations, avoir une semi élasticité est plus intéressant pour commenter une évolution de la variable *volume* en pourcentage. Pour la variable *gdp*, le log nous permet d'atténuer les valeurs extrêmes et d'obtenir une stabilité des données.

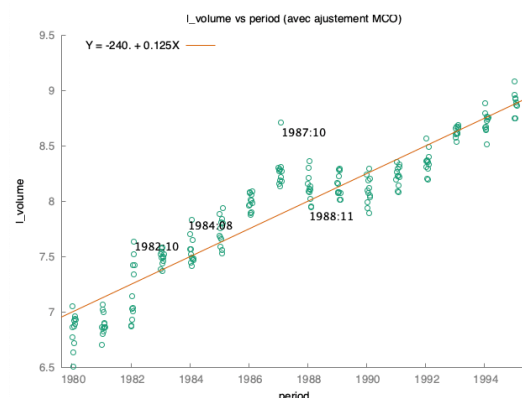
Et pour les deux variables (*volume* et *gdp*) cela nous permet d'obtenir une élasticité, simplifiant considérablement l'interprétation.

Ici, on peut voir l'atténuation des valeurs passant de 600 à environ 7 à l'ordonnée à l'origine et des valeurs extrêmes telles que les deux entourées en rouge.

Graphique 1 : volume en fonction de la période



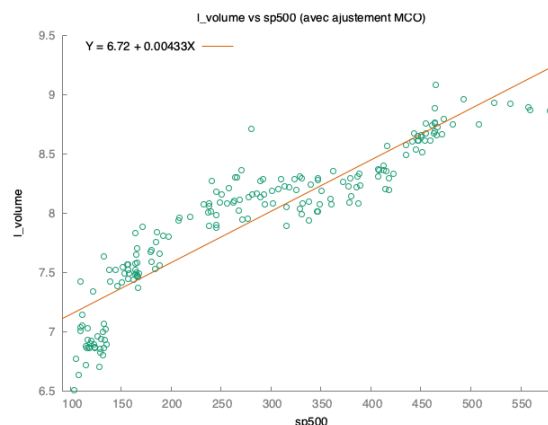
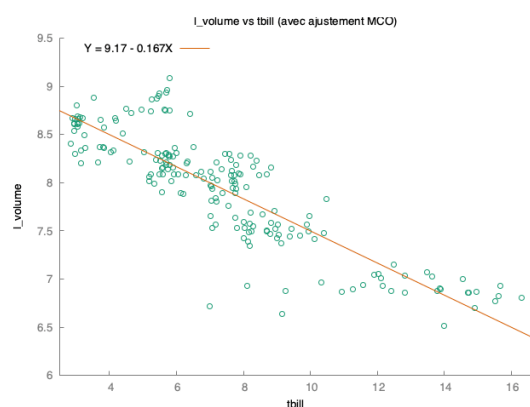
Graphique 2 : log_volume en fonction de la période



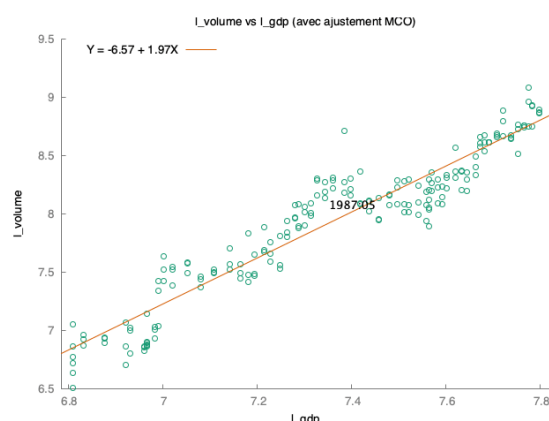
Graphique 3 : \log_volume en fonction du taux de trésorerie des USA



Graphique 4 : \log_volume en fonction de l'indice des prix de S&P500



Graphique 5 : \log_volume en fonction de \log_gdp



Le graphique 1 nous permet de constater une régression linéaire représentant l'évolution du volume des actions vendues au fil des années ce qui permet de confirmer nos intuitions d'une forte croissance au fil des années.

Le graphique 2 représente le même graphique que le 1 mais avec la variable *volume* en logarithme ce qui nous permet d'atténuer les valeurs extrêmes tel que les deux entourés en rouge et d'obtenir une stabilité numérique.

Le graphique 3 nous permet d'identifier une relation linéaire négative entre le \log_volume et le taux de trésorerie ce qui peut s'expliquer par le fait que si le taux de trésorerie des USA est élevé cela entraîne une réduction de l'incitation à spéculer.

Le graphique 4 illustre une relation linéaire positive entre le *log_volume* et *l'indice des prix de S&P500*, car si cet indice des prix augmente alors les ventes des actions vont augmenter également:

- les vendeurs voulant retirer leurs bénéfices.
- les acheteurs misant sur une continuité de la hausse.

Le graphique 5 traduit lui aussi une relation linéaire positive car il représente le *log_volume* en fonction du *log_gdp*, soit une hausse du PIB des USA entraîne une hausse des ventes d'actions.

L'ensemble des graphiques nous permet de déduire que notre modèle est une régression linéaire.

3 - Modèle économique et économétrique

Facteurs explicatifs du volume d'actions vendues aux USA.

Équation à estimer gardée :

$$\log(volume_i) = \beta_0 + \beta_1 sp500_i + \beta_2 tbill_i + \beta_3 long_i + \beta_4 cconf_i + \beta_5 cexpect_i + \beta_6 \log(gdp_i) + \beta_7 period_i + u_i$$

Les intuitions sur les signes sont les suivants :

tbill (taux de trésorerie) : Signe : Négatif.

Interprétation : Des taux de trésorerie plus élevés rendent les investissements en actions moins attractifs, diminuant donc le volume des transactions boursières.

log(gdp) (PIB) : Signe : Positif.

Interprétation : Une croissance économique continue stimule la confiance des investisseurs et le volume des échanges.

cconf (indice de confiance des consommateurs) : Signe : Négatif.

Interprétation : Une confiance plus élevée des consommateurs les incite à ne pas vendre leurs actions pour attendre que ces dernières prennent de la valeur.

SP500 (indice de prix des actions) : Signe : Positif.

Interprétation : Plus les actions prennent de la valeur et plus les détenteurs d'actions choisiront d'attendre pour les vendre à prix plus élevé.

long (rendement des obligations) : Signe : Négatif.

Interprétation : Un rendement élevé des obligations peut détourner les investisseurs des actions vers des obligations, réduisant donc le volume des transactions en actions.

cexpect (indice des prévisions des consommateurs) : Signe : Négatif.

Interprétation : Si le marché est très volatil, cela aura pour incidence de réduire les prévisions de vente des consommateurs car les déstabilisant.

Nous verrons par la suite que seuls les signes intuités des paramètres de *cexcept* et *cconf* diffèrent des signes obtenus.

Forme fonctionnelle : Modèle log-niveau pour la plupart des variables explicatives (*semi-élasticité*) exceptée le PIB en log (*élasticité*) avec relation linéaire.

4 - Estimations et Résultats

Équation estimée :

$$\widehat{\log(\text{volume}_i)} = -169,3 + 0,00025\text{sp500}_i - 0,061\text{tbill}_i - 0,010\text{long}_i + 0,0054\text{cconf}_i + 0,00022\text{cexpect}_i - 0,19\log(\text{gdp}_i) + 0,089\text{period}_i$$

(59,3763) (.0005) (.01) (.018) (.0001) (.0013) (.388) (.0312)

écart-type

T = 189 mois, 15 ans et 9 mois

Tableau 3 : MCO sous respect de H4

Modèle 8: MCO, utilisant les observations 1980:01-1995:09 (T = 189)
Variable dépendante: l_volume

	coefficient	éc. type	t de Student	p. critique		
const	-169.339	59.3763	-2.852	0.0049	***	✓
period	0.0898984	0.0311920	2.882	0.0044	***	✓
sp500	0.000255066	0.000544093	0.4688	0.6398		
tbill	-0.0619737	0.0100012	-6.197	3.79e-09	***	✓
long	-0.0107191	0.0179938	-0.5957	0.5521		
cconf	0.00543359	0.00109331	4.970	1.55e-06	***	✓
cexpect	0.000225949	0.00131371	0.1720	0.8636		
l_gdp	-0.191046	0.387923	-0.4925	0.6230		
Moyenne var. dép.	7.933056	Éc. type var. dép.	0.604488			
Somme carrés résidus	3.672056	Éc. type régression	0.142435			
R2	0.946546	R2 ajusté	0.944479			
F(7, 181)	457.8738	P. critique (F)	1.6e-111			
Log de vraisemblance	104.2447	Critère d'Akaike	-192.4893			
Critère de Schwarz	-166.5553	Hannan-Quinn	-181.9828			
rho	0.367295	Durbin-Watson	1.264931			

Significativité des paramètres :

- Les paramètres des variables : constante, *period*, *tbill* et *cconf* sont **significatifs** au seuil de 1%.
- Les paramètres des variables : *sp500*, *long*, *cexpect* et *l_gdp* sont **non significatifs**.

Test de Fisher : $H_0 : \beta_i = 0$ contre $H_1 : \beta_i \neq 0$ avec $i \in \{1, 2, \dots, k\}$

F(calc) = 457,87

F(7,181) = 2,05

F(calc) > F(7,181), on rejette H0 donc le modèle est globalement significatif.

La constante: $\hat{\beta}_0$ est égale à -169.339, c'est le volume des actions vendues lorsque tous les autres facteurs sont nuls.

Si l'indice des prix des actions ordinaires augmente d'une unité, le volume augmentera de 0.0255%.

Si le taux de trésorerie des USA augmente d'une unité, le volume diminue de 6.2%.

Si le rendement des obligations de trésorerie augmente d'une unité, le volume diminue de 1.07%.

Si l'indice de confiance des consommateurs augmente d'une unité, le volume augmentera de 0.54%.

Si l'indice des prévisions des consommateurs augmente d'une unité, le volume augmentera de 0.0225%.

Si le PIB augmente d'un % , le volume des actions vendues diminue de 0.19%.

La qualité d'ajustement est égal à $R^2 = 0,946$ cela signifie que 94,6% de la part de la variance est expliquée par le modèle, cette valeur semble trop élevée pour un modèle réaliste, nous soupçonnons donc une potentielle présence d'autocorrélation, et d'hétéroscédasticité donc un non respect de l'hypothèse H2.

Tableau 4 : Test de White

Test de White pour l'hétéroscédasticité
MCO, utilisant les observations 1980:01-1995:09 (T = 189)
Variable dépendante: uhat^2

	coefficient	éc. type	t de Student	p. critique
const	38068.7	84343.5	0.4514	0.6524
period	-41.0906	89.0079	-0.4617	0.6450
sp500	-0.0218980	1.19332	-0.01835	0.9854
tbill	-15.3903	19.8990	-0.7734	0.4405
long	35.3660	35.0974	1.008	0.3152
cconf	0.415928	2.53855	0.1638	0.8701
cexpect	1.06700	2.91165	0.3665	0.7145
l_gdp	707.295	1166.00	0.6066	0.5450
sq_period	0.0110826	0.0234907	0.4718	0.6377
X2_X3	-6.39637e-07	0.000625474	-0.001023	0.9992
X2_X4	0.00823968	0.0105325	0.7823	0.4352
X2_X5	-0.0192374	0.0185641	-1.036	0.3017
X2_X6	-0.000234982	0.00133637	-0.1758	0.8607
X2_X7	-0.000555511	0.00153121	-0.3628	0.7173
X2_X8	-0.378119	0.619310	-0.6105	0.5424
sq_sp500	-5.27003e-06	5.69769e-06	-0.9249	0.3565
X3_X4	0.000131445	0.000190403	0.6904	0.4910
X3_X5	-0.000227087	0.000369999	-0.6138	0.5403
X3_X6	1.76773e-06	2.55820e-05	0.06910	0.9450
X3_X7	1.14738e-05	2.97213e-05	0.3860	0.7000
X3_X8	0.00359055	0.00757347	0.4741	0.6361
sq_tbill	0.000638895	0.00238730	0.2676	0.7894
X4_X5	0.00337207	0.00794040	0.4247	0.6717
X4_X6	-0.000306411	0.000445644	-0.6876	0.4928
X4_X7	0.000396749	0.000501706	0.7908	0.4303
X4_X8	-0.146776	0.161624	-0.9081	0.3652
sq_long	-0.00259379	0.00751087	-0.3453	0.7303
X5_X6	0.000614676	0.000817687	0.7517	0.4534
X5_X7	-0.000506857	0.000938884	-0.5399	0.5901
X5_X8	0.402231	0.281727	1.428	0.1554
sq_cconf	1.78015e-05	4.26097e-05	0.4178	0.6767
X6_X7	-1.36960e-05	7.96822e-05	-0.1719	0.8638
X6_X8	0.00612590	0.0184725	0.3316	0.7406
sq_cexpect	1.09504e-05	4.12380e-05	0.2655	0.7909
X7_X8	0.00472857	0.0209028	0.2262	0.8213
sq_l_gdp	2.69195	4.59503	0.5858	0.5588

R2 non-ajusté = 0.337543

Statistique de test: $TR^2 = 63.795666$,
avec p. critique = $P(Khi\text{-deux}(35) > 63.795666) = 0.002083$

Détection d'hétéroscédasticité : Test de White :

Hypothèses : $H_0 : \sigma_i^2 = \sigma_u^2 \forall i$ vs. $H_1 : \sigma_i^2 \neq \sigma_u^2$

Statistique de test :
$$Q_W = NR^2 \approx \chi^2 \left(\frac{(k+1)(k+2)}{2} - 1 \right)$$

Règle de décision : Si $Q_W > \chi_{th}^2$, les perturbations sont hétéroscédastiques
Si $Q_W < \chi_{th}^2$, les perturbations sont homoscedastiques

Application : $Q_W = T * R^2 = 0.337543 * 189 \approx 63.79$

$\chi^2(35) = 43,773$

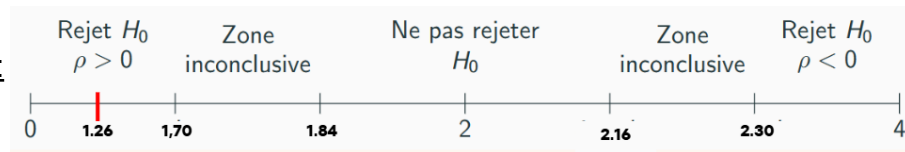
$Q_W > \chi^2(35)$ donc les perturbations sont hétéroscédastiques.

Détection d'autocorrélation : Test de Durbin-Watson :

Hypothèses : $H_0 : \rho = 0$ et $H_1 : \rho \neq 0$

Statistique de test :
$$DW = \frac{\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^T \hat{u}_t^2} \approx 2(1 - \hat{\rho})$$

Règle de décision :



Conclusion :

$0 < DW < 1.70$, on rejette H_0 , il y a une autocorrélation des perturbations.

On détecte de l'autocorrélation et de l'hétéroscédasticité, nous pouvons donc décider de les corriger avec les MCQG, nous obtenons le choix entre plusieurs modèles et nous pouvons dresser ce tableau :

Tableau 5 : Synthèse des tableaux

	Variables	Estimations des paramètres	Ecart-types estimés	DW
Moindres Carrés Ordinaires	const	-169.339	59.3763	1.264931
	period	0.0898984	0.0311920	
	sp500	0.000255066	0.000544093	
	tbill	-0.0619737	0.0100012	
	long	-0.0107191	0.0179938	
	cconf	0.00543359	0.00109331	
	cexpect	0.000225949	0.00131371	
	l_gdp	-0.191046	0.387923	
Prais-Winsten	const	-186.202	73.9800	2.271726
	period	0.0986538	0.0388944	
	sp500	8.54494e-05	0.000696672	
	tbill	-0.0570126	0.0131000	
	long	-0.0162986	0.0239160	
	cconf	0.00552989	0.00152567	
	cexpect	-0.000538268	0.00175033	
	l_gdp	-0.247080	0.501658	
Cochrane-Orcutt	const	-185.473	75.1381	2.266703
	period	0.0982567	0.0395416	
	sp500	8.59305e-05	0.000698415	
	tbill	-0.0570517	0.0131429	
	long	-0.0160860	0.0242245	
	cconf	0.00550823	0.00157249	
	cexpect	-0.000515343	0.00179318	
	l_gdp	-0.239220	0.520035	
Hildreth et Lu	const	-185.473	75.1382	2.266706
	period	0.0982567	0.0395416	
	sp500	8.59299e-05	0.000698416	
	tbill	-0.0570517	0.0131429	
	long	-0.0160861	0.0242245	
	cconf	0.00550823	0.00157249	
	cexpect	-0.000515348	0.00179318	
	l_gdp	-0.239220	0.520036	

Les 3 résultats des modèles des MCQG sont similaires puisqu'ils ont une statistique de Durbin-Watson proche de 2, nous allons donc privilégier le modèle Prais-Winsten car il conserve la première observation contrairement au test Cochrane-Orcutt.

Remise en question de H3 : La matrice **X** est non-aléatoire. (corrélacion entre une variable explicative et le terme d'erreur), cela aura pour incidence de biaiser l'estimateur des MCO et de le rendre non convergent.

Rappel : Afin de vérifier la validité de H3, il convient de tester les variables pour une éventuelle exogénéité : une variable est considérée comme telle si on peut la considérer comme une donnée sans perte d'information, cependant en économie les variables individuelles sont souvent corrélées avec les informations explicatives observées.

Il est également important de rappeler que la matrice nous sert à nous faire une idée de l'intensité des liens inter-variables, afin de nous mettre sur la piste de l'endogénéité.

Pour la détecter factuellement il faut établir des variables instrumentales, les tester (test de Sargan) et réaliser un test d'Hausman, l'endogénéité sera alors corrigée via les Doubles Moindres Carrés.

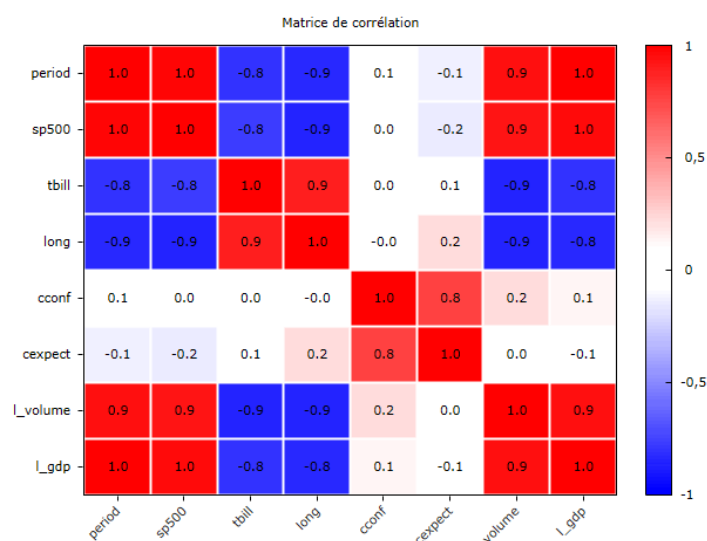


Tableau 6 : Matrice de corrélation

Ici, on remarque que la plupart des variables sont corrélées entre elles avec un coefficient $\approx \pm 1$.

5 - Résultats et Conclusion

Deux coefficients n'ont pas les signes attendus : β_4 et β_5 qui sont les paramètres associés aux variables **cconf** et **cexpect**. Les comportements humains sont les variables les plus incertaines en économie car elles peuvent varier en fonction de beaucoup de facteurs, ils sont d'ailleurs la cause de beaucoup de crises économiques et financières (Grande Dépression, la bulle internet, la crise des subprimes...). Nous avons pour intuitions que **cconf** ait un coefficient négatif car plus les détenteurs d'actions ont confiance que le marché va prendre de la valeur, plus ils seront incités à attendre pour les vendre; nos intuitions peuvent toutefois être fausses car plus les consommateurs ont confiance en l'évolution du marché, plus ils seront incité à acheter, l'analyse dépend d'un point de vu socio-économique.

Pour conclure sur la partie des signes des coefficients, nous estimons H4 respectée.

Notre estimateur des MCO n'était pas un estimateur de Gauss-Markov car les perturbations étaient hétéroscédastiques et autocorrélées grâce aux tests de White et Durbin-Watson.

C'est la raison pour laquelle nous avons estimé notre modèle par les Moindre Carrés Quasi Généralisés (Prais-Winsten & Cochrane-Orcutt).

Toutefois, même après correction, nous remarquons un coefficient de détermination toujours anormalement élevé, la matrice des corrélations nous a donc mené sur la piste du non-respect de H3 à savoir l'endogénéité.

S'il y a endogénéité, l'estimateur des MCQG n'est pas un estimateur de Gauss-Markov qui doit valider les 5 hypothèses :

$H_1 : E[u] = 0, \forall t$ l'espérance mathématique de l'erreur est nulle

$H_2 : V[u] = E(uu^T) = \sigma_u^2 I_N$ la variance de l'erreur est constante

H_3 : matrice X est non-aléatoire

H_4 : le modèle est correctement spécifié

H_5 : la matrice X est de plein rang : $k + 1 < T$

Pour poursuivre le travail, nous pourrions faire un test de Hausman pour détecter l'endogénéité, si ce test en détecte nous pourrions poursuivre avec une correction via les Doubles Moindres Carrés pour corriger l'endogénéité.

Pour conclure, nous pourrions répondre à notre problématique initiale en disant que le taux de trésorerie, l'indice de confiance des consommateurs et le PIB ont joué des rôles majeurs sur cette période de l'histoire quant à l'évolution et à la démocratisation des échanges d'actions sur les marchés financiers.

6 - Bibliographie & Sources

La base de données provient de Gretl, dans le fichier "data9-11" du fichier de données Ramanathan.

Les sources ont été compilées par Brian Wempler d'après les sources : United States Statistical Abstract, Standard and poor quarterly reports, citibase data series, et l'université d'Alabama.

https://www.census.gov/library/publications/time-series/statistical_abstracts.html

<https://estima.com/webhelp/topics/citibase.html>

<https://catalog.ua.edu/undergraduate/arts-sciences/mathematics/data-science-bs/>