

MODELES DYNAMIQUES

Les modèles dynamiques font intervenir des variables décalées dans le temps contrairement aux modèles statiques vus précédemment. Ils posent, comme nous le verrons au fil des chapitres, beaucoup plus de problèmes, à la fois de construction et de propriétés statistique des estimateurs. Comme vous l'avez certainement déjà constaté en économie ces modèles se retrouvent dans toute les études qu'elles soient macro ou micro-économiques ainsi que dans les modèles financiers. Dans ce chapitre nous donnerons quelques exemples montrant l'apport de ces modèles.

Un modèle dynamique fait intervenir des retards sur une ou plusieurs variables.

- Si ces variables sont uniquement exogènes on parlera de modèles à retards échelonnés comme dans la forme simple

$$y_t = c + \sum_{i=0}^r a_i x_{t-i} + \epsilon_t$$

- Si les variables retardées correspondent à l'endogène on appellera ces modèles des modèles autorégressifs

$$y_t = c + \sum_{i=1}^r a_i y_{t-i} + \epsilon_t \text{ ou } y_t = c + \sum_{i=1}^r a_i y_{t-i} + b x_t + c z_t + \epsilon_t$$

- Dans le cas général on parlera de modèles autorégressifs et à retards échelonnés

$$y_t = c + \sum_{i=1}^r a_i y_{t-i} + \sum_{j=0}^l b_j x_{t-j} + \epsilon_t$$

On remarque que j peut aller de 0 à l, mais bien sur que i ne peut aller que de 1 à r car si on prend i=0 y_t est expliqué par lui-même ce qui vous donnera $a_0 = 1$ et tous les autres coefficients nuls et un R^2 magnifique égal à 1 mais qui bien sur n'explique rien.

Autre remarque : les retards r et l ne sont pas égaux en général et tous les retards ne sont pas obligatoirement explicatifs, c'est-à-dire que des coefficients a_i ou b_j peuvent être nuls. Nous verrons que le choix de r et l n'est pas toujours simple.

REMARQUE IMPORTANTE: Nous faisons l'hypothèse que les **erreurs** des modèles présentés dans ce chapitre sont **stationnaires**. Cette hypothèse sera étudiée dans le chapitre sur l'intégration des séries.

1 Modèles dynamiques et théorie économique

Quelques exemples économiques vont être présentés ici, ils montrent le passage d'un modèle statique à un modèle dynamique.

1.1 La notion de revenu permanent:

Notons R_t^P le revenu permanent. On va le définir de la façon suivante

$$R_t^P - R_{t-1}^P = \alpha(R_t - R_{t-1}^P) \text{ où } \alpha \text{ est un coefficient fixe}$$

La consommation des ménages est supposée fonction de leur revenu disponible permanent

$$CM_t = a_0 + a_1 RD_t^P + u_t$$

Nous allons montrer que ces deux hypothèses conduisent à un modèle dynamique. Le revenu permanent est inconnu.

$$\begin{aligned} CM_t &= a_0 + a_1(R_{t-1}^P + \alpha(R_t - R_{t-1}^P)) + u_t \\ CM_t &= a_0 + a_1\alpha R_t + a_1(1 - \alpha)R_{t-1}^P + u_t \end{aligned}$$

or $CM_{t-1} = a_0 + a_1 RD_{t-1}^P + u_{t-1}$ donc $CM_{t-1} - a_0 - u_{t-1} = a_1 RD_{t-1}^P$

$$\begin{aligned} CM_t &= a_0 + a_1\alpha R_t + (1 - \alpha)(CM_{t-1} - a_0 - u_{t-1}) + u_t \\ CM_t &= \alpha a_0 + (1 - \alpha)CM_{t-1} + a_1\alpha R_t + u_t - (1 - \alpha)u_{t-1} \end{aligned}$$

On obtient donc en éliminant le revenu disponible un modèle autorégressif

1.2 La notion d'ajustement partiel:

Soit Y_t le niveau du capital d'une entreprise et Y_t^* son niveau de capital désiré qui est fonction de sa production x_t

$$Y_t^* = a + bx_t + \varepsilon_t$$

On définit l'ajustement partiel de production par l'écart réalisé, qui n'est qu'une partie de l'écart désiré.

$$\begin{aligned} Y_t - Y_{t-1} &= \lambda(Y_t^* - Y_{t-1}) \text{ avec } 0 \leq \lambda \leq 1 \\ Y_t &= Y_{t-1} - \lambda Y_{t-1} + \lambda(a + bx_t + \varepsilon_t) \\ Y_t &= (1 - \lambda)Y_{t-1} + \lambda bx_t + \lambda a + \lambda \varepsilon_t \end{aligned}$$

Soit à nouveau un modèle dynamique.

2 Les modèles dynamiques pour résoudre des problèmes de modélisation

2.1 Lorsque l'on suppose que le modèle réel n'est pas linéaire

On sait que dans ces conditions les prévisions sur ce modèle sont bonnes dans le cadre des grands échantillons en remplaçant ce modèle inconnu par un modèle autorégressif.

Soit un modèle non linéaire $Y_t = F(x_{1t}, x_{2t}, \dots, \epsilon_t)$, si la fonction est inconnue et la taille d'échantillon grande, on peut utiliser en prévision un modèle qui sera presque aussi bon (non d'un point de vue explicatif mais prévisionnel) le modèle

$$Y_t = \sum_1^r a_i Y_{t-i} + b_1 x_{1t} + b_2 x_{2t} \dots + \epsilon_t$$

Ce qui nous conduit à nouveau à un modèle autorégressif.

2.2 Lorsque l'on a oublié des variables explicatives:

Supposons que le bon modèle soit le modèle statique suivant

$$Y_t = a_0 + a_1 x_t + a_2 z_t + \epsilon_t$$

et que nous ne prenions pas en compte la variable z , parce que l'on a oublié cette variable, ou parce que l'on a pas de données sur elle

Notre modèle devient

$$Y_t = b_0 + b_1 x_t + u_t$$

Dans le cas où z_t est **lié** à z_{t-1} il suffit de prendre comme variable explicative également Y_{t-1} pour introduire dans le modèle z_{t-1} (car $Y_{t-1} = a_0 + a_1 x_{t-1} + a_2 z_{t-1} + \epsilon_{t-1}$) et donc indirectement une partie de z_t qui lui est lié. C'est un peu tordu mais souvent efficace. Le modèle devient

$$Y_t = c_0 + c_1 x_t + c_2 Y_{t-1} + v_t$$

Il est généralement bien meilleur que le modèle statique.

2.3 Lorsqu'il y a autocorrélation ou hétéroscédasticité

Parfois pour des raisons comparables à la partie ci-dessus, on constate que l'on élimine les problèmes des erreurs d'un modèle statique en le rendant dynamique. Voilà pourquoi les modèles dynamiques sont les plus courants dans les études économiques.

3 Les modèles à retards échelonnés purs

Dans le cadre des modèles statiques ($y_t = bx_t + c + \epsilon_t$) l'influence de la variable x sur y est instantanée en t . Dans le cas des modèles à retards échelonnés ($y_t = b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + c + \epsilon_t$) x_t a toujours une influence sur y en t mais les valeurs passées de x également. On dira par exemple que la consommation des ménages est fonction de leur revenu mais aussi des revenus des trimestres précédents. On va étudier séparément les deux cas classique en économie: les retards sont infinis puis les retards sont finis.

3.1 Les modèles à retards infinis

On fait l'hypothèse que l'influence sur y des retards est infinie, dans la formulation générale $r=0$ et l est ∞ . Comme on le voit on va ainsi vers de gros soucis car il faut estimer alors une infinité de coefficients b_i ce qui est impossible. Comment faire ? Il faut mettre des contraintes sur les coefficients b_i afin de diminuer le nombre de coefficients à estimer.

3.1.1 Le modèle de Koyck

Hypothèses du modèle Les économistes utilisent principalement la contrainte dite de Koyck en faisant l'hypothèse que l'influence des x_{t-i} va en décroissant au fur et à mesure que l'on s'éloigne du temps t . ($b_0 > b_1 > \dots b_i > \dots$) les coefficients tendant vers 0. Cette contrainte de Koyck s'écrit

$$b_i = b_0 \lambda^i \text{ avec } 0 \leq \lambda < 1$$

On voit donc qu'au lieu d'avoir une infinité de coefficients b_i à estimer, il suffit d'estimer b_0 et λ pour avoir tous les coefficients. Le modèle s'écrit alors

$$y_t = b_0 x_t + b_0 \lambda x_{t-1} + b_0 \lambda^2 x_{t-2} + \dots + b_0 \lambda^i x_{t-i} + \dots + c + \epsilon_t = b_0 \sum_{i=0}^{\infty} \lambda^i x_{t-i} + c + \epsilon_t$$

On montre que ce modèle est en fait un modèle autorégressif On constate ici un modèle avec contrainte sur les coefficients. Or si on écrit cette équation $t-1$

$$y_{t-1} = b_0 x_{t-1} + b_0 \lambda x_{t-2} + b_0 \lambda^2 x_{t-3} + \dots + b_0 \lambda^i x_{t-i-1} + \dots + c + \epsilon_{t-1}$$

on voit facilement qu'en multipliant la seconde équation par λ et en construisant $y_t - \lambda y_{t-1}$ on obtient

$$\begin{aligned} y_t - \lambda y_{t-1} &= b_0 x_t + c(1 - \lambda) + \epsilon_t - \lambda \epsilon_{t-1} \\ y_t &= \lambda y_{t-1} + b_0 x_t + c(1 - \lambda) + \epsilon_t - \lambda \epsilon_{t-1} \end{aligned}$$

Le modèle à retards échelonnés de Koyck est donc en fait aussi un modèle autorégressif d'ordre 1 (un seul retard sur l'endogène y) et sans retard sur x . La réciproque est vraie, un modèle autorégressif avec un retard sur l'endogène et sans retard sur l'exogène est en fait un modèle à retards échelonnés infinis de Koyck.

Ses erreurs sont en général corrélées

- Faisons l'hypothèse que les erreurs du modèle de base les ϵ_t sont homoscédastiques de variance σ_ϵ^2 et non autocorrélées donc $E(\epsilon_t \cdot \epsilon_{t-1}) = 0$ et notons $u_t = \epsilon_t - \lambda \epsilon_{t-1}$ les erreurs du modèle autorégressif. La covariance des u_t

$cov(u_t, u_{t-1}) = E(u_t \cdot u_{t-1}) = E[(\epsilon_t - \lambda \epsilon_{t-1})(\epsilon_{t-1} - \lambda \epsilon_{t-2})] = E[-\lambda \epsilon_{t-1}^2] = -\lambda \sigma_\epsilon^2 \neq 0$ les erreurs du modèle autorégressif sont corrélées

- Faisons l'hypothèse maintenant que les erreurs ϵ_t sont toujours homoscédastiques de variance σ_ϵ^2 mais maintenant autocorrélées donc $E(\epsilon_t \cdot \epsilon_{t-1}) = \rho \sigma_\epsilon^2$ alors $E(u_t \cdot u_{t-1}) = E[(\epsilon_t \epsilon_{t-1} - \lambda \epsilon_{t-1}^2 - \lambda \epsilon_t \epsilon_{t-2} + \lambda^2 \epsilon_{t-1} \epsilon_{t-2})] = \rho \sigma_\epsilon^2 - \lambda \sigma_\epsilon^2 - \lambda \rho \sigma_\epsilon^2 + \lambda^2 \rho \sigma_\epsilon^2$

Il y a donc aussi autocorrélation des erreurs u sauf dans le cas particulier où $\lambda = \rho$

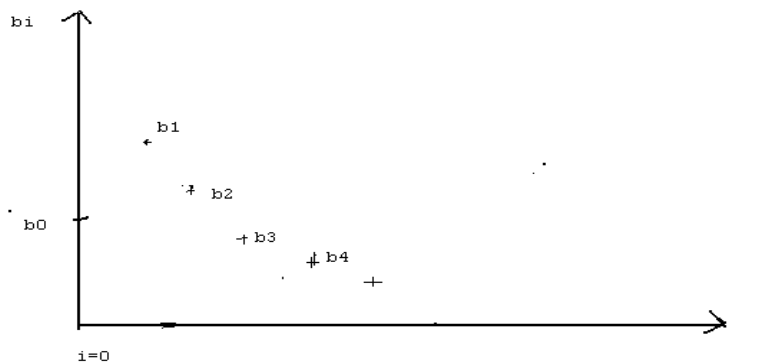
Ces résultats sont importants car nous verrons dans l'étude des modèles autorégressifs que les estimations s'effectuent différemment suivant que l'on ait autocorrélation ou non des erreurs. Nous laissons donc pour l'instant l'estimation de ces coefficients.

3.1.2 Une généralisation de ce modèle de Koyck

Nous avons précisé tout à l'heure qu'un modèle

$$y_t = \lambda y_{t-1} + b_0 x_t + c(1 - \lambda) + \epsilon_t - \lambda \epsilon_{t-1} \text{ est équivalent à } y_t = b_0 \sum_{i=0}^{\infty} \lambda^i x_{t-i} + c + \epsilon_t$$

Prenons maintenant un modèle à retards échelonnés infinis laissant libre le premier coefficient et commençant en t-1 seulement les coefficients de Koyck. Cela arrive en Economie lorsque l'on pense que la décroissance des coefficients ne commence qu'en t-1 et que le coefficient en t a par exemple une importance moindre.



Le modèle s'écrit

$$y_t = b_0 x_t + b_1 \sum_{i=0}^{\infty} \lambda^i x_{t-1-i} + c + \epsilon_t$$

On montre facilement en écrivant y_{t-1} et en retranchant λy_{t-1} à y_t , qu'il est équivalent au modèle autorégressif et à retard échelonné :

$$y_t = \lambda y_{t-1} + b_0 x_t + (b_1 - \lambda b_0) x_{t-1} + c(1 - \lambda) + \epsilon_t - \lambda \epsilon_{t-1}$$

On peut continuer la généralisation en laissant libres b_0 et $b_1 \dots$

3.2 les modèles à retards échelonnés finis

$$y_t = c + \sum_{j=0}^p b_j x_{t-j} + \epsilon_t$$

où p est une constante. Sur ces modèles on peut faire les MCO si les erreurs sont IID car p est fini. Ils présentent cependant deux difficultés d'estimation

- tout d'abord en général les retards sur la variable x sont corrélés en eux et comme on le sait cela conduit à des coefficients mal estimés et des variances très grandes. Contre

cela on ne peut pas grand chose, mais il est bon de se souvenir que l'on ne peut faire confiance à la valeur des coefficients estimés, seul l'ensemble $\sum_{j=0}^p \widehat{b}_j x_{t-j}$ compte. Il existe aussi des méthodes de traitement de la colinéarité comme la régression Ridge qui aident à une meilleure estimation de ces coefficients.

- ensuite il faut connaître la valeur de p. Pour cela on va utiliser les critères classiques de sélection de modèles.

3.2.1 Valeur de p

On va rappeler ici les principaux critères de sélection de modèles. La plupart de ces critères sont basés sur un ordre dans les variables introduites. Ici l'ordre est le temps.

Nous allons voir que tous ces critères reposent sur le fait que la **taille d'échantillon doit être la même pour tous les modèles**.

On dispose de K variables explicatives et on cherche parmi elles celles qui donneront le meilleur modèle au sens du critère choisi.

- Critère du $R_k^2 = 1 - SCR_k / \sum (y_t - \bar{y})^2$ maximiser le R^2 revient à minimiser la somme des carrés des résidus SCR_k si la taille d'échantillon reste la même. Nous savons si le nombre k de variables explicatives du modèle augmente alors son SCR va diminuer, autrement dit plus on ajoute de variables, plus le modèle est bon au sens de ce critère, ce qui bien sur conduit à ajouter des variables non significatives. Nous savons que ce critère n'est pas bon. On le note tout de même.

$$\min SCR_k \sim \min \text{Log}(SCR_k) \sim \min \text{Log}(SCR_k/n)$$

n étant les même pour tous les modèles. Suivant ce critère, le meilleur modèle est pour k=K

- Critère du R_k^2 corrigé noté aussi $\overline{R}_k^2 = 1 - \frac{SCR_k/(n-k)}{\sum (y_t - \bar{y})^2 / (n-1)}$ maximiser ce \overline{R}_k^2 est identique à minimiser $SCR_k/(n-k) = s_k^2$ la variance résiduelle car n étant fixé $\sum (y_t - \bar{y})^2 / (n-1)$ ne change pas si k varie. Afin de pouvoir comparer les différents critères on va passer au $\text{LOG}(s_k^2)$ en se souvenant que n est fixé.

$$\text{Log}(s_k^2) = \text{Log}(SCR_k) - \text{Log}(n-k) = \text{Log}\left(\frac{SCR_k}{n}\right) + \text{Log}\left(\frac{n}{n-k}\right) = \text{Log}\left(\frac{SCR_k}{n}\right) + \text{Log}\left(1/\left(1 - \frac{k}{n}\right)\right)$$

or si n est grand $\text{Log}(1/(1 - \frac{k}{n})) \sim k/n$ et

$$\min s_k^2 \sim \min \text{Log}(s_k^2) \sim \min[(\text{Log}(SCR_k/n) + k/n)]$$

Si on compare au critère précédent on constate que l'on a ajouté à $\text{Log}(SCR_k)$ la valeur k/n. Le critère introduit le nombre de variables explicatives k qui varie (alors que n ne bouge pas); si on augmente k alors SCR_k va diminuer mais k va augmenter, il faut regarder lequel l'emporte sur l'autre. Tout se passe comme si on "payait" k/n pour ajouter une variable en plus, cette valeur se nomme pénalité. Avec le critère du $\text{Log}(SCR_k/n)$ la pénalité est nulle, cela ne coute rien d'ajouter des variables alors inutile de se gêner on va toutes les mettre même si elles n'ont aucun intérêt. Par contre ici on va être plus parcimonieux car une pénalité existe.

Ces deux critères sont ceux de base, voyons maintenant les critères autres proposés par différents économètres.

- Le critère d'Akaike proposé en 1974 il consiste à minimiser AIC_k

$$\text{minimiser } AIC_k = \text{Log}(SCR_k/n) + 2k/n$$

On voit ici que la pénalité $2k/n$ est le double de celle du critère du $\overline{R_k^2}$ on sera donc encore plus regardant à ajouter des variables il faudra que l'intérêt statistique soit important.

- Le critère de Tsai en 1989 qui consiste à minimiser $AICC_k$

$$\text{minimiser } AICC_k = \text{Log}(SCR_k/n) + \frac{n+k}{n-k-2}$$

La pénalité est encore plus forte que l'AIC, elle est de $\frac{n+k}{n-k-2}$ mais tend vers le critère AIC si n très grand.

- Le critère de Schwartz en 1978 qui consiste à minimiser le BIC_k

$$\text{minimiser } BIC_k = \text{Log}(SCR_k/n) + \frac{k}{n} \text{Log}(n)$$

On constate que la pénalité est supérieure à celle de AIC dès que $n > 7$

- Le critère d'Hannan-Quinn en 1980 qui consiste à minimiser ϕ_k

$$\text{minimiser } \phi_k = \text{Log}(SCR_k/n) + \frac{kc}{n} \text{Log}(\text{Log}(n)) \text{ c constante } \geq 2$$

En général on prend $c=2$

$$\text{minimiser } \phi_k = \text{Log}(SCR_k/n) + \frac{2k}{n} \text{Log}(\text{Log}(n))$$

- Pour finir on donnera le critère de Mallow noté C_p de Mallow (il a noté p les k variables du modèle). Ce critère est totalement différent dans son principe. Il n'a pas besoin d'ordre pour les variables.

minimiser $C_p = SCR_p/s_K^2 + 2p/n$ où s_K^2 est le s^2 du modèle avec toutes les variables

3.2.2 Conclusion

Maintenant que ces critères sont énoncés, deux problèmes se posent

- Quelle valeur donner à K c'est-à-dire en fait quelle longueur maximale va-t-on donner aux retards ? Il est évident que cette longueur doit dépendre de la taille de l'échantillon. Il est évident également que cet échantillon ne doit être trop petit.

Schwert en 1989 a conclu à l'aide de modèles simulés que le nombre maximum de retards doit être de $4(n/100)^{1/4}$ résultat que l'on utilise en général pour des données trimestrielles, ou $12(n/100)^{1/4}$ que l'on utilise pour des données mensuelles.

- Ensuite parmi tous ces critères lequel prendre? Tout dépend de la taille de l'échantillon. En général si l'échantillon est très petit (<50) on utilise le critère du \overline{R}_k^2 (identique à minimiser s^2), AIC ou AICc. Si $50 < n < 150$ on utilise AIC, AICc ou ϕ et dans le cas des grands échantillons ϕ ou plus souvent BIC.

3.3 Exemple de modèle à retards finis

L'exemple se trouve dans le fichier echelon.prg et les données dans echelon.rat. Ces données au nombre de 124 sont des chroniques mensuelles. On explique y_t par la variable x_t et par des retards sur x . Quel retard maximum va-t-on prendre pour que tous les modèles aient la même taille d'échantillon? On prend le critère de Schwert pour des données mensuelles $12(n/100)^{1/4} \simeq 12.9$, le retard max étant de 12 on prendra l'échantillon de 13 à 124 soit de 112.

retards sur x	AIC	AICC	BIC	HQ	s	DW	n
0	22.824	19.161	22.836	18.160	8617.17831	1.081	112
1	21.745	18.101	21.774	17.109	5047.90098	2.240	112
2	21.492	17.868	21.542	16.885	4469.75894	2.585	112
3	21.408	17.804	21.480	16.828	4304.93623	2.639	112
4	21.406	17.822	21.502	16.854	4321.38799	2.641	112
5	21.404	17.841	21.526	16.880	4337.64686	2.637	112
6	21.401	17.859	21.550	16.905	4352.60914	2.621	112
7	21.376	17.855	21.552	16.907	4317.93368	2.617	112
8	21.373	17.875	21.579	16.932	4333.86082	2.616	112
9	21.373	17.897	21.609	16.960	4355.23085	2.617	112
10	21.368	17.915	21.634	16.982	4364.82973	2.610	112
11	21.359	17.929	21.656	17.000	4366.90493	2.650	112
12	21.335	17.929	21.665	17.004	4337.33698	2.656	112

On remarque que la taille de l'échantillon utilisé est bien toujours de 112. Le critère AIC ne donne rien car le min est pour une valeur supérieure à 12, il prend beaucoup trop de retards. Les autres critères trouvent tous un minimum pour 3 retards sur x . Ces critères sont les seuls intéressants car $n > 100$.

Le modèle retenu est donc

```
Linear Regression - Estimation by Least Squares
Dependent Variable Y
Usable Observations      112      Degrees of Freedom  107
Centered R**2      0.992644      R Bar **2      0.992369
Uncentered R**2    0.999737      T x R**2      111.970
Mean of Dependent Variable      254556.26335
Std Error of Dependent Variable  49280.30005
Standard Error of Estimate      4304.93623
Sum of Squared Residuals      1982974926.5
Regression F(4,107)      3609.6883
Significance Level of F      0.00000000
Log Likelihood      -1093.52557
Durbin-Watson Statistic      2.639186
```

Variable	Coeff	Std Error	T-Stat	Signif
1. Constant	11688.330702	2096.765708	5.57446	0.00000019
2. X	0.511436	0.129400	3.95237	0.00013914
3. X{1}	1.080753	0.215688	5.01074	0.00000216
4. X{2}	0.204729	0.215358	0.95065	0.34392682
5. X{3}	0.393505	0.128154	3.07055	0.00270872

On remarque que le deuxième retard n'est pas significatif, cela est en général du aux problèmes de colinéarité engendrés par les variables retardées.

3.4 Les hypothèses des polynômes d'Almon

3.4.1 Théorie

Quand les critères de choix de retards sur la variable x conduisent à prendre un grand nombre de retards p, on constate comme on l'a déjà indiqué que les variables retardées sont très corrélées, cela entraîne (voir le chapitre sur la colinéarité) des coefficients mal estimés et des variances grandes des estimateurs donc des "t" de Student petits alors que les retards sont intéressants. Les économistes n'aiment pas beaucoup les mauvais t de Student. Pour cela dans certains cas économiques ils préfèrent ajouter des contraintes aux coefficients des retards. La forme la plus connue de ces contraintes est la méthode polynomiale d'Almon qui a imposé aux coefficients indice i de suivre un polynome en i de degré allant de 1 à 4 à choisir suivant l'exemple (ALmon a construit ses polynômes dans le cadre de modèles d'investissements).

$$\text{Modèle} : Y_t = c + \sum_{i=0}^p b_i x_{t-i} + \varepsilon_t$$

$$\text{Contraintes sur les } b_i : b_i = \alpha_0 + \sum_{j=1}^r \alpha_j i^j$$

Par exemple pour r=2 :

$$b_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2$$

$$Y_t = c + \sum_{i=0}^p (\alpha_0 + \alpha_1 i + \alpha_2 i^2) x_{t-i} + \varepsilon_t$$

$$Y_t = c + \alpha_0 \sum_{i=0}^p x_{t-i} + \alpha_1 \sum_{i=0}^p i x_{t-i} + \alpha_2 \sum_{i=0}^p i^2 x_{t-i} + \varepsilon_t$$

$$Y_t = c + \alpha_0 Z1_t + \alpha_1 Z2_t + \alpha_2 Z3_t + \varepsilon_t$$

Au lieu des p+1 variables corrélées on n'a plus que les 3 variables Z1, Z2 et Z3. Connaissant les estimations des α_i , on revient aux estimations des b_i .

3.4.2 Exemple

Prenons un modèle où la variable y est expliquée par x et des retards sur x.

On commence par rechercher le nombre de retards p suivant les différents critères:

```
*****
retards   akaike   aicc     bic      s        D.W     stat Q   ns de Q  n
sur x
*****
0          5.649     6.666     5.693   16.72159  0.365   282.44   0.00000  129
1          5.005     6.023     5.071   12.07081  0.912   126.32   0.00000  129
2          4.304     5.323     4.392    8.47007   1.148   69.58    0.00000  129
```

3	3.692	4.713	3.803	6.21446	1.622	47.42	0.00129	129
4	3.164	4.187	3.297	4.75519	1.853	34.56	0.04301	129
5	3.063	4.088	3.218	4.50412	2.034	28.94	0.14645	129
6	2.999	4.026	3.176	4.34676	2.192	33.28	0.05808	129
7	2.947	3.977	3.147	4.22066	2.175	36.09	0.02968	129
8	2.961	3.994	3.182	4.23386	2.176	35.62	0.03331	129

Pour tous les critères, le minimum du critère est obtenu pour p=7

$$Y_t = c + \sum_{i=0}^7 b_i x_{t-i} + \varepsilon_t$$

```
Linear Regression - Estimation by Least Squares
Dependent Variable Y
Usable Observations    129      Degrees of Freedom    120
Centered R**2          0.996536    R Bar **2            0.996305
Uncentered R**2        0.999865    T x R**2             128.983
Mean of Dependent Variable    343.02494949
Std Error of Dependent Variable 69.43076504
Standard Error of Estimate    4.22065952
Sum of Squared Residuals      2137.6760163
Regression F(8,120)          4314.7542
Significance Level of F      0.00000000
Log Likelihood              -364.13728
Durbin-Watson Statistic      2.175095
```

Variable	Coeff	Std Error	T-Stat	Signif
1. Constant	48.09965571	1.66903887	28.81877	0.00000000
2. X	0.43760837	0.12684423	3.44997	0.00077443
3. X{1}	0.24043509	0.21276952	1.13003	0.26071894
4. X{2}	0.70698013	0.21575097	3.27683	0.00137305
5. X{3}	0.20597968	0.22554440	0.91326	0.36293935
6. X{4}	0.82307758	0.23373870	3.52136	0.00060791
7. X{5}	0.07503054	0.23416282	0.32042	0.74920695
8. X{6}	-0.07902655	0.23280886	-0.33945	0.73486546
9. X{7}	0.40635545	0.14072486	2.88759	0.00460621

Les problèmes de colinéarité entraînent de mauvais t de Student et des coefficients mal estimés. Cela n'empêche pas ce modèle d'être bon en prévision, seulement il n'est pas "beau" car beaucoup de retards semblent (à tort) non significatifs. On va utiliser la méthode d'Almon. Rats redonne directement les coefficients de base et non ceux des ZI.

Avec un polynome de degré 3

```
Linear Regression - Estimation by Least Squares
Dependent Variable Y
Usable Observations    129      Degrees of Freedom    124
Centered R**2          0.996344    R Bar **2            0.996227
Uncentered R**2        0.999857    T x R**2             128.982
Mean of Dependent Variable    343.02494949
Std Error of Dependent Variable 69.43076504
Standard Error of Estimate    4.26500753
Sum of Squared Residuals      2255.5958644
Regression F(4,124)          8449.3597
Significance Level of F      0.00000000
Log Likelihood              -367.60060
Durbin-Watson Statistic      2.249365
```

Variable	Coeff	Std Error	T-Stat	Signif
1. Constant	48.296528752	1.684196129	28.67631	0.00000000

2.	X	0.293823085	0.081540606	3.60340	0.00045311
3.	X{1}	0.500733391	0.041878489	11.95682	0.00000000
4.	X{2}	0.543252986	0.061624967	8.81547	0.00000000
5.	X{3}	0.476756196	0.036706362	12.98838	0.00000000
6.	X{4}	0.356617350	0.039127662	9.11420	0.00000000
7.	X{5}	0.238210775	0.063580663	3.74659	0.00027335
8.	X{6}	0.176910801	0.041603179	4.25234	0.00004126
9.	X{7}	0.228091754	0.088161401	2.58721	0.01082937

Avec un polynôme de degré 2

***methode d almon avec 7 retards et un polyn\U{f4}me de degre 2

```
Linear Regression - Estimation by Least Squares
Dependent Variable Y
Usable Observations    129      Degrees of Freedom    125
Centered R**2          0.996276      R Bar **2            0.996187
Uncentered R**2        0.999855      T x R**2             128.981
Mean of Dependent Variable    343.02494949
Std Error of Dependent Variable 69.43076504
Standard Error of Estimate    4.28724532
Sum of Squared Residuals      2297.5590531
Regression F(3,125)          11148.4843
Significance Level of F      0.00000000
Log Likelihood              -368.78954
Durbin-Watson Statistic      2.222200
```

Variable	Coeff	Std Error	T-Stat	Signif	

1.	Constant	48.138826939	1.689757263	28.48860	0.00000000
2.	X	0.400339241	0.041818797	9.57319	0.00000000
3.	X{1}	0.440854394	0.014200241	31.04556	0.00000000
4.	X{2}	0.454188044	0.019046419	23.84638	0.00000000
5.	X{3}	0.440340193	0.027938901	15.76083	0.00000000
6.	X{4}	0.399310840	0.027360503	14.59443	0.00000000
7.	X{5}	0.331099985	0.017475749	18.94625	0.00000000
8.	X{6}	0.235707628	0.015319145	15.38647	0.00000000
9.	X{7}	0.113133770	0.045444355	2.48950	0.01410660

Avec un polynôme de degré 1

***methode d almon avec 7 retards et un polyn\U{f4}me de degre 1

```
Linear Regression - Estimation by Least Squares
Dependent Variable Y
Usable Observations    129      Degrees of Freedom    126
Centered R**2          0.996099      R Bar **2            0.996037
Uncentered R**2        0.999848      T x R**2             128.980
Mean of Dependent Variable    343.02494949
Std Error of Dependent Variable 69.43076504
Standard Error of Estimate    4.37064165
Sum of Squared Residuals      2406.9160615
Regression F(2,126)          16087.7790
Significance Level of F      0.00000000
Log Likelihood              -371.78872
Durbin-Watson Statistic      2.115515
```

Variable	Coeff	Std Error	T-Stat	Signif	

1.	Constant	47.405329964	1.695128857	27.96562	0.00000000
2.	X	0.491535727	0.019097893	25.73769	0.00000000
3.	X{1}	0.451914205	0.013718649	32.94160	0.00000000
4.	X{2}	0.412292683	0.008391308	49.13330	0.00000000
5.	X{3}	0.372671161	0.003371607	110.53221	0.00000000
6.	X{4}	0.333049639	0.003326013	100.13479	0.00000000
7.	X{5}	0.293428117	0.008336543	35.19782	0.00000000
8.	X{6}	0.253806595	0.013662887	18.57635	0.00000000
9.	X{7}	0.214185073	0.019041847	11.24812	0.00000000

Nous prendrons pour ces trois celui qui a le S le plus petit, soit le modèle avec le polynôme d'Almon de degré 3. On constate que les t Student des coefficients sont bien significatifs. Rappelons que ce modèle a pour seul intérêt de donner des bons t de Student, le modèle sans Almon est aussi bon en prévision.

4 Modèles autorégressifs

Dans ces modèles des retards sur la variable endogène décalée se retrouvent comme variables explicatives. Nous avons vu un exemple dans le cas de modèle à retards échelonnés infinis. Economiquement on retrouve très souvent des retards sur l'endogène, ils indiquent que l'explication de Y est fonction de ses valeurs passées (par exemple, le comportement des ménages variant peu d'un trimestre à l'autre l'épargne en t est très liée à celui en t-1 d'où un modèle autorégressif pour l'épargne).

4.1 Quels sont les problèmes liés aux modèles autorégressifs?

Dans le modèle, par exemple, $y_t = a + bx_t + \epsilon_t$ nous avons jusqu'ici supposé que la variable x était non aléatoire, donc sous sa forme générale la matrice X des variables explicatives est non aléatoire. Par contre les erreurs ϵ sont aléatoires ce qui entraîne bien sur y aléatoire. Nous avons déduit de cela dans le chapitre sur les propriétés statistiques des MCO que L'estimateur des MCO était sans biais.

4.1.1 La matrice X est maintenant aléatoire:

Maintenant dans un modèle $y_t = a + bx_t + cy_{t-1} + \epsilon_t$ la variable y_t est bien sur aléatoire mais sa valeur y_{t-1} est pour les mêmes raisons aussi aléatoire. Nous nous trouvons donc dans un cas très différent du modèle précédent car nous avons comme variable explicative une variable aléatoire. La matrice X des MCO devient aléatoire. (notons que la taille d'échantillon est n-1)

4.1.2 Alors les MCO sont biaisés

La matrice X est formée dans cet exemple du vecteur unité du vecteur de la variable x et du vecteur engendré par les y_{t-1}

$$X_{(n-1, n-1)} = \begin{bmatrix} 1 & x_2 & y_1 \\ \cdot & \cdot & \cdot \\ 1 & x_t & y_{t-1} \\ \cdot & \cdot & \cdot \\ 1 & x_n & y_{n-1} \end{bmatrix} \quad \vec{Y} = \begin{pmatrix} y_2 \\ \cdot \\ y_t \\ \cdot \\ y_n \end{pmatrix} \quad \text{et} \quad \vec{\epsilon} = \begin{pmatrix} \epsilon_2 \\ \cdot \\ \epsilon_t \\ \cdot \\ \epsilon_n \end{pmatrix}$$

le résultat des MCO est

$$\vec{\hat{a}} = ({}^tXX)^{-1t}XY = ({}^tXX)^{-1t}X(X\vec{a} + \vec{\epsilon}) = \vec{a} + ({}^tXX)^{-1t}X\vec{\epsilon}$$

donc

$$E(\vec{\hat{a}}) = \vec{a} + E[({}^tXX)^{-1t}X\vec{\epsilon}] = \vec{a} + E\left[\left(\frac{{}^tXX}{(n-1)}\right)^{-1} \frac{{}^tX\vec{\epsilon}}{(n-1)}\right]$$

Dans notre exemple

$$\frac{{}^t X \vec{\epsilon}}{(n-1)} = \begin{bmatrix} \sum \epsilon_t / (n-1) \\ \sum x_t \epsilon_t / (n-1) \\ \sum y_{t-1} \epsilon_t / (n-1) \end{bmatrix} \xrightarrow{n \rightarrow \infty} \begin{bmatrix} E(\epsilon) \\ Cov(x, \epsilon) \\ Cov(y_{-1}, \epsilon) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ Cov(y_{-1}, \epsilon) \end{bmatrix}$$

car l'espérance des erreurs est nulle et la variable x n'étant pas aléatoire (ou dans certains cas aléatoire mais indépendante des erreurs) a donc une covariance nulle avec les erreurs. Deux cas peuvent se présenter:

- Si la covariance entre les erreurs et les variables explicatives existe c'est-à-dire que les erreurs en t sont liées à la variable y en t-1 soit

$$\frac{{}^t X \vec{\epsilon}}{n-1} \neq \vec{0} \text{ si } n \rightarrow \infty$$

dans ces conditions on voit que $E[(\frac{{}^t X X}{n-1})^{-1} \frac{{}^t X \vec{\epsilon}}{n-1}] \neq \vec{0}$ en faisant l'hypothèse $(\frac{{}^t X X}{n-1})^{-1} \rightarrow$ une matrice finie et alors $E(\vec{\hat{a}}) \neq \vec{a}$ et l'estimateur des MCO est asymptotiquement biaisé, donc les MCO ne sont pas utilisables même si l'échantillon est grand.

Dans quel cas a-t-on ce résultat ? On sait que y_t est lié par construction à ϵ_t , de même y_{t-1} est lié par construction à ϵ_{t-1} . Si les erreurs sont autocorrélées alors ϵ_t est lié à ϵ_{t-1} et donc à y_{t-1} et alors dans l'équation $y_t = a + bx_t + cy_{t-1} + \epsilon_t$ une variable explicative ici y_{t-1} est liée à l'erreur ϵ_t .

- par contre si la covariance est nulle, les erreurs en t ne sont pas liées à la variable y en t-1 et alors $\frac{{}^t X \vec{\epsilon}}{n-1} \rightarrow \vec{0}$ si $n \rightarrow \infty$ et $E(\vec{\hat{a}}) = \vec{a}$ les MCO sont asymptotiquement sans biais et donc utilisables si n est grand.
- Remarque: dans les cas où $(\frac{{}^t X X}{n-1})^{-1} \rightarrow$ la matrice nulle si $n \rightarrow \infty$ alors on voit que dans les deux cas le biais est asymptotiquement nul et les MCO sont applicables, ce cas se retrouvera dans les problèmes de cointégration.

4.1.3 Conclusion

Remarquons bien que l'on raisonne toujours dans le cas asymptotique ce qui signifie que pour construire un modèle autorégressif il faut un échantillon de taille importante.

- Si les erreurs d'un modèle autorégressif ne sont pas autocorrélées alors les MCO sont asymptotiquement sans biais et on peut utiliser ce résultat.
- Si les erreurs d'un modèle autorégressif sont autocorrélées alors les MCO sont asymptotiquement biaisés et cette méthode n'est plus utilisable.

4.1.4 Le test de DURBIN

Comme nous venons de le voir il est très important de savoir s'il y a ou non autocorrélation d'ordre 1 $\epsilon_t = \rho\epsilon_{t-1} + u_t$ des erreurs dans un modèle autorégressif. Voir le chapitre "tests d'autocorrélation"

H0 : il y a non autocorrélation $\iff \rho = 0$

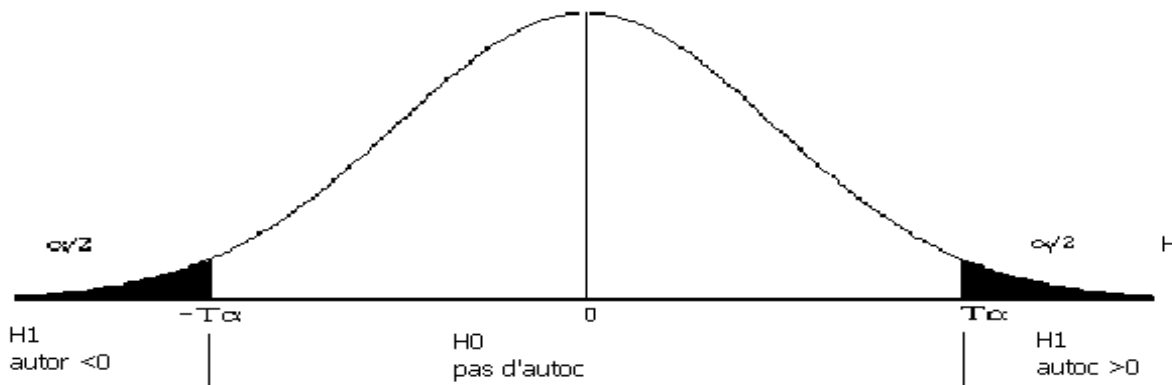
H1 : il y a autocorrélation $\iff \rho \neq 0$ avec toujours $|\rho| < 1$

Dans les cas classique on utilise la statistique de DW. Ici cette statistique est biaisée sous H1 car les MCO étant asymptotiquement biaisés, les résidus sont aussi biaisés et donc également DW qui utilise ces résidus. De plus on montre que ce biais tend à privilégier la décision H0, Durbin a construit un nouveau test qui réduit ce biais sous H1. Nous en avons parlé dans le chapitre " tests d'autocorrélation" avec des exemples. On se contente ici de rappeler les résultats. Durbin a construit la statistique notée H

$$H = \frac{\hat{\rho}}{\sqrt{\frac{1}{n} - \widehat{Var}(\hat{a}_1)}} \simeq \frac{1 - \frac{DW}{2}}{\sqrt{\frac{1}{n} - \widehat{Var}(\hat{a}_1)}}$$

$\hat{\rho}$ est l'estimation de ρ , comme on a vu que $DW \simeq 2(1 - \hat{\rho})$ on en déduit que $\hat{\rho} \simeq 1 - DW/2$; n est la taille de l'échantillon et $\widehat{Var}(\hat{a}_1)$ est la variance estimée du **premier retard sur l'endogène**.

Durbin a montré que sous l'hypothèse H0 de non autocorrélation des erreurs cette variable H suit asymptotiquement une loi Normale centrée réduite. Le fait que ce test soit asymptotique n'est pas un problème car il est indispensable d'avoir des échantillons importants pour travailler sur des modèles autorégressifs.



Pour effectuer ce test il faut bien sur que $\frac{1}{n} - \widehat{Var}(\hat{a}_1) > 0$

Dans le cas où $\widehat{Var}(\hat{a}_1) > \frac{1}{n}$ on remplace le test de Durbin par le test suivant très proche du test de Goldfeld et Breusch. On effectue les MCO sur le modèle autorégressif et on récupère les résidus e_t , puis on construit l'équation:

$$e_t = \rho e_{t-1} + \text{toutes les variables explicatives du modèle de base} \\ \text{y compris les retards sur l'endogène sauf la constante}$$

et on effectue simplement le test de Student (ici avec la loi Normale car on est toujours dans le cadre des grands échantillons) sur le coefficient $\hat{\rho}$ du retard sur le résidu ou si on

veut être plus précis le test du nR^2 (non centré car on n'a pas de terme constant) qui suit sous l'hypothèse H_0 un χ_1^2 .

4.1.5 Le test de Ljung-Box

Les hypothèses du test sont les mêmes, on teste:

H0 $\rho_1 = \rho_2 = \dots = \rho_r = 0$ pas d'autocorrélation des erreurs d'ordre 1 à r

H1 l'un au moins des $\rho_i \neq 0$ il y a autocorrélation des erreurs d'ordre entre 1 et r

Pour l'effectuer on récupère les résidus e_t du modèle de base et on construit

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \dots + \rho_r e_{t-r} + u_t$$

Les MCO sur ce modèle donnent des estimations $\hat{\rho}_i$ des ρ_i . Ljung et Box ont montré que sous l'hypothèse H0 la variable Q

$$Q = n(n+2) \sum_1^r \frac{\hat{\rho}_i^2}{n-i}$$

suit sous l'hypothèse H0 une loi du χ^2 à (r-am) degrés de liberté où am est le nombre de retards de l'endogène et des variables explicatives(AR et MA)

Ce test est le plus utilisé des tests d'autocorrélation d'ordre supérieur à 1. Le plus souvent les logiciels ne le fournissent pas directement, il faut les calculer à part.

Mais avant, posons-nous la question de la valeur de r, RATS propose de regarder au maximum pour **rmax = partie entière du MIN(n/4, $2\sqrt{n}$)**

4.2 Exemple

Modèle dynamique avec 1 retard sur l'endogène et un retard sur la variable explicative z:

```
Linear Regression - Estimation by Least Squares
Dependent Variable Y
Usable Observations    132      Degrees of Freedom    128
Total Observations    137      Skipped/Missing      5
Centered R**2         0.993583    R Bar **2           0.993433
Uncentered R**2      0.999736    T x R**2            131.965
Mean of Dependent Variable 340.40074765
Std Error of Dependent Variable 70.77342918
Standard Error of Estimate 5.73546741
Sum of Squared Residuals 4210.6350579
Regression F(3,128)    6606.2806
Significance Level of F 0.00000000
Log Likelihood        -415.82930
Durbin-Watson Statistic 3.048487
```

Variable	Coeff	Std Error	T-Stat	Signif

1. Constant	15.395468026	2.833212264	5.43393	0.00000027
2. Y{1}	0.684482351	0.032934885	20.78290	0.00000000
3. Z	0.244444774	0.162123378	1.50777	0.13407760
4. Z{1}	0.642927516	0.206242440	3.11734	0.00225397

4.2.1 Etude de ce modèle 1

Pour savoir si les MCO sont ou non sans biais il faut étudier la statistique H de Durbin

$$H \simeq \frac{1 - \frac{DW}{2}}{\sqrt{\frac{1}{n} - \widehat{Var}(\widehat{a}_1)}} = \frac{1 - 3.04848/2}{\sqrt{\frac{1}{132} - (0.0329)^2}} = -6.507$$

N'oublions pas que la variance utilisée est celle du premier retard sur l'endogène Y.

La valeur de H est très éloignée de l'intervalle -1.96 1.96 on en déduit une forte autocorrélation d'ordre 1 .Le modèle est autorégressif donc les MCO sont biaisés et comme nous venons de voir qu'il y a autocorrélation d'ordre 1 des erreurs, alors les MCO sont asymptotiquement biaisés , donc inutilisables même si l'échantillon est grand.

La valeur de la statistique de Ljung-Box calculée par Rats est

statistique de Ljung-Box Q(20-2)= 96.8885. Significance Level 0.000

Elle teste donc 22 retards et on enlève deux degrés de liberté correspondant au retard sur Y et au retard sur X. Le niveau de significativité étant presque nul, on en déduit une forte autocorrélation entre 1 et 22 retards sur les erreurs.

4.2.2 recherche du bon nombre de retards sur l'endogène et l'exogène

Nous n'avons donné qu'un retard sur chaque, ce n'est certainement pas assez. Pour avoir une idée du nombre de retards à mettre, nous prenons les critères de sélection qui sont identiques à ceux indiqués pour les modèles à retards échelonnés. Voici les résultats fournis par RATS

CRITERE AIC AVEC DES RETARDS SUR L ENDOGENE ET 1 VARIABLE(S) EXOGENE(S)

avec le retard max de schwert = int(4*(n/100)**.25)= 4

AIC	2.9468	retards sur Y	sur Z
		3	4
AIC	2.9615	retards sur Y	sur Z
		4	4
AICc	3.9768	retards sur Y	sur Z
		3	4
AICc	3.9920	retards sur Y	sur Z
		2	4
BIC	3.1422	retards sur Y	sur Z
		2	4
BIC	3.1463	retards sur Y	sur Z
		3	4
HQ	3.0279	retards sur Y	sur Z
		3	4

HQ 3.0369 retards sur Y 2 sur Z 4

Le programme donne les deux meilleurs résultats pour chaque critère (le plus petit étant comme vous le voyez à sa valeur étant le premier pour chaque critère).

Le critère du AIC par exemple donne avec sa valeur minimale de 2.9468 donne 3 retards sur Y et 4 sur Z. Avec la valeur suivante un peu plus grande 2.9615 il donne 4 et 4.

Le critère du AICc donne pour sa plus petite valeur 3 et 4 retards, comme le critère HQ. Seul le BIC donne 2 et 4 retards.

On va donc donner les résultats pour les critères AIC

4.2.3 Etude du meilleur AIC, AICc et HQ

```

Linear Regression - Estimation by Least Squares
Dependent Variable Y
Usable Observations 130 Degrees of Freedom 121
Centered R**2 0.996596 R Bar **2 0.996371
Uncentered R**2 0.999865 T x R**2 129.982
Mean of Dependent Variable 342.15712418
Std Error of Dependent Variable 69.86535336
Standard Error of Estimate 4.20889776
Sum of Squared Residuals 2143.4932579
Regression F(8,121) 4427.9812
Significance Level of F 0.00000000
Log Likelihood -366.63476
Durbin-Watson Statistic 2.010341
  
```

Variable	Coeff	Std Error	T-Stat	Signif
1. Constant	37.36241406	3.79621683	9.84201	0.00000000
2. Y{1}	-0.08004454	0.08603579	-0.93036	0.35403544
3. Y{2}	0.18652407	0.07113852	2.62198	0.00986480
4. Y{3}	0.11816447	0.06008657	1.96657	0.05152198
5. Z	0.40982297	0.12665079	3.23585	0.00156446
6. Z{1}	0.33644108	0.21560079	1.56048	0.12125783
7. Z{2}	0.63041356	0.21917662	2.87628	0.00475689
8. Z{3}	0.11801800	0.22498927	0.52455	0.60085611
9. Z{4}	0.68907802	0.19728122	3.49287	0.00066809

Valeur de H de Durbin:

$$H \simeq \frac{1 - \frac{DW}{2}}{\sqrt{\frac{1}{n} - Var(\hat{a}_1)}} = \frac{1 - 2.01/2}{\sqrt{\frac{1}{130} - (0.086)^2}} = -0.303$$

Cette valeur de H étant comprise dans l'intervalle -1.96 1.96 alors les MCO sont biaisés mais asymptotiquement sans biais et comme n est grand, nous gardons les MCO

Valeur de Q:

Q(20-7)= 21.8716. Significance Level 0.05738390

On accepte la non autocorrélation des erreurs entre 1 et 20 retards (remarque: le niveau de significativité est très proche de 5% donc ce résultat n'est pas très bon).

5 Modèles autorégressifs asymptotiquement biaisés

On repère ces modèles, comme nous venons de le voir, au fait qu'ils présentent une autocorrélation d'ordre 1 (statistique H de Durbin). Dans ces condition les MCO ne sont plus acceptables.

Reportons-nous au chapitre autocorrélation dans les modèles statiques, les remarques sont les mêmes.

Avant d'effectuer tout traitement de l'autocorrélation des erreurs, il faut se demander si l'autocorrélation n'est pas due à une cause liée à la construction même du modèle:

- Un oubli de variable
- un oublié de retard dans les variables endogène ou exogènes
- une mauvaise spécification du modèle : modèle en Log par exemple
- des points aberrants non traités ...

Si on veut cependant traiter le modèle, la technique la plus simple est de faire les MCG (voir le cours sur le traitement de l'autocorrélation) mais en utilisant la méthode d'Hildreth-Lu et non celle de Cochrane qui conduit trop souvent à un optimum local.