

C. Babusiaux 15/01/2008

Nous avons vu dans la partie test d'autocorrélation de ce chapitre, plus précisément dans le paragraphe portant sur le test de Durbin et Watson, la remarque particulièrement importante en pratique que l'autocorrélation est le plus souvent due à un mauvais modèle: modèle avec oubli de variables importantes, modèle à coefficients non stables, erreur de spécification du modèle. Nous en avons déduit qu'il ne faut en général pas traiter l'autocorrélation mais plutôt bien sur essayer de traiter le vrai problème (voir les quelques exemples d'illustration). Dans quelques cas on ne trouvera pas le vrai problème et on sera amené à traiter l'autocorrélation, ce sera un constat d'échec bien sur mais à l'impossible nul n'est tenu On sera donc amené à utiliser la méthode des MCG (moindres carrés généralisés).

Plutôt que de reprendre la théorie des MCG que vous trouverez dans tous les ouvrages, nous allons voir les techniques pratiques utilisées par les logiciels si on connaît la forme de l'autocorrélation.

Nous prenons l'exemple d'un modèle avec autocorrélation d'ordre 1 des erreurs

$$\begin{aligned} Y_t &= a_0 + a_1 X_t + a_2 Z_t + \epsilon_t \\ \epsilon_t &= \rho \epsilon_{t-1} + u_t \end{aligned} \tag{1}$$

avec u bruit blanc de variance σ_u^2 et indépendant des ϵ et $|\rho| < 1$. Nous avons vu que ρ est le coefficient de corrélation entre les erreurs en t et t-1.

Calcul de la variance des erreurs si on fait l'hypothèse d'homoscédasticité de ces erreurs:

$$\begin{aligned} \mathbb{V}(\epsilon_t) &= \mathbb{E}(\epsilon_t^2) = \mathbb{E}(\rho \epsilon_{t-1} + u_t)^2 = \rho^2 \mathbb{E}(\epsilon_{t-1}^2) + 2\rho \mathbb{E}(\epsilon_{t-1} u_t) + \mathbb{E}(u_t^2) \\ &= \rho^2 \mathbb{V}(\epsilon_{t-1}) + \sigma_u^2 \end{aligned}$$

l'indépendance des ϵ et des u conduit à des covariances nulles. De plus comme on fait l'hypothèse d'homoscédasticité des ϵ , en notant σ_ϵ^2 leur variance commune, on trouve:

$$\sigma_\epsilon^2 = \rho^2 \sigma_\epsilon^2 + \sigma_u^2 \implies \sigma_\epsilon^2 = \frac{\sigma_u^2}{1 - \rho^2}$$

Au passage on voit bien pourquoi on impose $|\rho| < 1$ car dans le cas contraire $|\rho| > 1$ on aurait $\rho^2 > 1$ et donc une variance négative ce qui reconnaissons-le n'est pas très raisonnable !!!

1 Cas théorique où ρ est connu

Commençons par ce cas irréaliste pour voir comment nous allons procéder.

1.1 Les MCO ne sont pas bons:

Si on se contente de faire les MCO sur le modèle de base sans tenir compte de l'autocorrélation, il est facile de voir que les estimateurs correspondants sont sans biais mais à variance trop grande. Si on écrit le modèle sous sa forme générale $\vec{Y} = X\vec{a} + \vec{\epsilon}$ on obtient le résultat des MCO

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

$$\mathbb{E}(\vec{\hat{a}}) = \vec{a} + \mathbb{E}(({}^tXX)^{-1} {}^tX\vec{\epsilon}) \quad (2)$$

On montre ainsi que si les variables explicatives sont non aléatoires alors leur covariance avec les erreurs est nulle et le résultat des MCO est toujours sans biais. Par contre la matrice de variance covariance des erreurs dépend de ρ elle est donc différente de $\sigma_\epsilon^2\mathbb{I}$, comme elle ne peut qu'être que plus importante, les MCO ont donc des variances trop grandes, et comme on le sait les estimateurs sont moins bons. Voici une façon simpliste de montrer qu'il ne faut pas utiliser les MCO dans un modèle où les erreurs sont autocorrélées.

1.2 Les MCG

Nous n'allons pas ici présenter comme on l'a déjà dit la théorie générale des MCG mais montrer en pratique comment font les logiciels pour procéder à une bonne estimation des coefficients.

Les erreurs ϵ ne sont pas bonnes, nous allons essayer de les remplacer par de "bonnes erreurs" comme le sont les erreurs u_t qui sont (si on a bien une autocorrélation d'ordre 1) des bruits blancs. (Rappelons que cela signifie $\mathbb{E}(u_t) = 0$, homoscedasticité et non autocorrélation). On voit facilement qu'il suffit de faire une transformation du modèle de base.

En t on a $Y_t = a_0 + a_1X_t + a_2Z_t + \epsilon_t$

En t-1 on a $Y_{t-1} = a_0 + a_1X_{t-1} + a_2Z_{t-1} + \epsilon_{t-1}$

On en déduit

$$\begin{aligned} Y_t - \rho Y_{t-1} &= a_0(1 - \rho) + a_1(X_t - \rho X_{t-1}) + a_2(Z_t - \rho Z_{t-1}) + \epsilon_t - \rho \epsilon_{t-1} \\ &= a_0(1 - \rho) + a_1(X_t - \rho X_{t-1}) + a_2(Z_t - \rho Z_{t-1}) + u_t \end{aligned}$$

On constate ainsi que ce nouveau modèle possède la bonne erreur u on peut donc appliquer les MCO. Ces MCO donneront de bons estimateurs de $a_0(1 - \rho)$ donc de a_0 car ρ est connu, de a_1 et de a_2 . On a utilisé un modèle transformé, où toutes les variables sont les variables en t moins ρ * (les variables en t-1).

Pour résumer on estime le modèle transformé par les MCO, cela nous donne les \hat{a}_i qui sont de bons estimateurs. Ces estimateurs seront donc ceux du modèle de base $Y_t = a_0 + a_1X_t + a_2Z_t + \epsilon_t$. $\implies \hat{Y}_t = \hat{a}_0 + \hat{a}_1X_t + \hat{a}_2Z_t$. Comme ces estimateurs ne sont pas obtenus par les MCO du modèle du base on dit qu'ils sont obtenus par les MCG dans le modèle de base ce qui correspond aux MCO du modèle transformé. Si vous avez étudié la théorie des MCG vous constatez bien ce même résultat (sauf pour t=1).

2 Cas où ρ est inconnu

La partie précédente était une présentation pratique d'une forme des MCO. On pratique le coefficient ρ est inconnu, comme il est présent dans le modèle transformé il est impossible de faire les MCO sur ce modèle. Nous allons définir les MCQG (moindres carrés quasi-généralisés).

2.1 Les MCQG

Comme dans tous les problèmes statistiques lorsqu'un coefficient est une constante inconnue on va essayer de l'estimer. Avant de donner les techniques qui vont permettre d'estimer ρ voyons la méthode MCQG. Supposons donc pour l'instant que ρ est estimé par $\hat{\rho}$, la méthode des MCQG consiste simplement à faire les MCO sur le modèle transformé en remplaçant la valeur théorique ρ par son estimateur. Le modèle transformé devient

$$Y_t - \hat{\rho}Y_{t-1} = a_0(1 - \hat{\rho}) + a_1(X_t - \hat{\rho}X_{t-1}) + a_2(Z_t - \hat{\rho}Z_{t-1}) + \epsilon_t - \hat{\rho}\epsilon_{t-1} \quad (3)$$

Une nouveauté apparaît ici, les variables explicatives $X_t - \hat{\rho}X_{t-1}$ et $Z_t - \hat{\rho}Z_{t-1}$ sont maintenant des variables aléatoires car elles contiennent le coefficient aléatoire $\hat{\rho}$. Plus haut, de l'équation (2) on avait déduit que les variables explicatives étant non aléatoires les MCO sont sans biais, ici ce n'est plus le cas, les variables explicatives étant aléatoires nous n'avons plus $\mathbb{E}(({}^tX X)^{-1} {}^tX \vec{\epsilon}) = \vec{0}$ par conséquent l'estimateur des MCO dans ce modèle transformé (3) ne sont plus sans biais. Le théorème central limite indique que si l'estimateur de ρ est un bon estimateur alors les MCO sur ce modèle transformé sont asymptotiquement sans biais. En conséquence, pour faire correctement les MCO sur le modèle transformé il faut

- que la taille de l'échantillon soit grande
- que l'on estime correctement ρ

Dans ces conditions on estime les coefficients du modèle transformé par les MCO, ces coefficients deviendront les estimateurs dans le modèle de base ce qui s'appelle les MCQG dans ce modèle de base (équation 1).

Il ne reste plus qu'à savoir comment estimer correctement ρ . Il existe dans les logiciels deux techniques principalement utilisées, la méthode de COCHRANE-ORCUTT et la méthode d'HILDRETH-LU

2.2 La méthode de COCHRANE-ORCUTT

C'est une méthode itérative c'est-à-dire que l'on va faire une série d'estimations qui vont améliorer à chaque fois l'estimation et on s'arrêtera lorsque l'on ne gagne presque plus rien, lorsque deux estimations consécutives sont pratiquement identiques.

- phase 0: faire les MCO sur le modèle de base (équation 1) $Y_t = a_0 + a_1X_t + a_2Z_t + \epsilon_t$ on sait que l'estimation des coefficients n'est pas très bonne. On récupère les résidus

e_t^0 de ce modèle et on cherche à estimer ρ dans $\epsilon_t = \rho\epsilon_{t-1} + u_t$, pour cela on remplace les erreurs par les résidus.

$$e_t^0 = \rho e_{t-1}^0 + u_t^0$$

Les MCO donnent $\hat{\rho}_0$. On peut aussi si le logiciel nous donne la statistique de Durbin et Watson $DW \simeq 2(1 - \hat{\rho}_0) \implies \hat{\rho}_0 \simeq 1 - DW/2$.

- phase 1: on utilise cette estimation de ρ dans le modèle transformé (3) modèle sur lequel on peut asymptotiquement avoir de bons résultats des MCO

$$Y_t - \hat{\rho}_0 Y_{t-1} = a_0(1 - \hat{\rho}_0) + a_1(X_t - \hat{\rho}_0 X_{t-1}) + a_2(Z_t - \hat{\rho}_0 Z_{t-1}) + \epsilon_t - \hat{\rho}_0 \epsilon_{t-1}$$

- phase 2: de ces nouveaux MCO on déduit les estimations \hat{a}_0, \hat{a}_1 et \hat{a}_2 et on en déduit de nouveaux résidus du modèle de base

$$e_t^1 = Y_t - \hat{a}_0 - \hat{a}_1 X_t + \hat{a}_2 Z_t$$

qui sont meilleurs que les e_t^0 car les coefficients a_i sont mieux estimés

- phase 3: on estime à nouveau ρ dans le modèle

$$e_t^1 = \rho e_{t-1}^1 + u_t^1$$

Les MCO donnent $\hat{\rho}_1$

On reprend alors la phase 1, puis 2 et 3 pour obtenir une estimation $\hat{\rho}_2$ meilleure que $\hat{\rho}_1$ et on continue ainsi jusqu'à ce que l'écart entre deux estimations consécutives soit inférieur à une valeur très petite fixée par le logiciel. On obtient ainsi une "bonne" estimation de ρ qui conduit à une "bonne" estimation des coefficients du modèle, tout cela bien sûr comme on va le voir dans quelques exemples à la condition que l'erreur u soit une erreur non autocorrélée.

le problème est la possibilité d'obtenir non l'optimum mais seulement un optimum local, ce qui est assez rare en pratique.

2.3 La méthode de balayage de Hildreth-Lu

dans le modèle transformé

$$Y_t - \rho Y_{t-1} = a_0(1 - \rho) + a_1(X_t - \rho X_{t-1}) + a_2(Z_t - \rho Z_{t-1}) + \epsilon_t - \rho \epsilon_{t-1}$$

On essaie "toutes les valeurs de ρ comprises entre -1 et +1 et on prend le meilleur résultat c'est-à-dire celui qui a donné la plus petite SCR ou le plus petit s .

En pratique on fait un premier balayage entre -1 et +1 en prenant un pas de 0,1 et on repère où se trouve la meilleure somme des carrés des résidus ou le plus petit s . Puis on recommence autour de cette valeur de ρ avec un pas de 0,01 et ainsi de suite. généralement dans votre logiciel vous avez une fonction qui fait ce test simple.

Prenons l'exemple suivant : $Y = a_0 + a_1 X_1 + a_2 X_2 + \epsilon$ Ce modèle a $n=50$, $DW = 0.6$, $s=2.5$ et $R^2 = 0.96$

On commence avec un pas de 0.1

$\rho = 0.1$	$s = 2.27$	$DW = 0.8$
$\rho = 0.2$	$s = 2.13$	$DW = 0.9$
$\rho = 0.3$	$s = 2.02$	$DW = 1.07$
$\rho = 0.4$	$s = 1.93$	$DW = 1.25$
$\rho = 0.5$	$s = 1.87$	$DW = 1.44$
$\rho = 0.6$	$s = 1.84$	$DW = 1.68$
$\rho = 0.7$	$s = 1.86$	$DW = 1.78$
$\rho = 0.8$	$s = 1.90$	$DW = 1.89$

On constate que s est minimum pour $\rho = 0.6$ pour cette valeur le $DW=1.68$ est dans la zone d'acceptation de H_0 (les bornes du test sont 1.46 et 1.63)

Pour avoir plus de précision on recommence entre 0.5 et 0.7 avec un pas de 0.01

2.4 Rappel important

Nous avons vu dans la partie sur les tests d'autocorrélation que l'autocorrélation était souvent due à d'autres problèmes comme des variables oubliés, un modèle instable ...

Je ne vous recommande pas d'utiliser cette méthode avant d'avoir vu si les problèmes ne peuvent être réglés. En fait cette méthode n'est pas utilisée en pratique, on préfère passer directement aux modèles dynamiques. En effet vous constatez que le modèle transformé peut aussi s'écrire:

$$Y_t = \rho Y_{t-1} + a_0(1 - \rho) + a_1 X_t - a_1 \rho X_{t-1} + a_2 Z_t - a_2 \rho Z_{t-1} + \epsilon_t - \rho \epsilon_{t-1}$$

C'est-à-dire sous forme d'un modèle autorégression (variable explicative Y_{t-1}) et à retards échelonnés (des retards sur les variables explicatives) mais avec des contraintes sur les coefficients. En effet le coefficient de X_{t-1} n'est pas quelconque mais le produit du coefficient de Y_{t-1} par le coefficient de X_t . On préfère construire de simple modèles dynamiques en oubliant ces contraintes afin de faciliter la pratique. D'où maintenant le passage au chapitre suivant sur les modèles dynamiques.