Économétrie II

Ch. 4 Autocorrélation
L3 Économétrie – L3 MASS

Prof. Philippe Polomé, U. Lyon 2

Année 2015-2016

Rappel

1. $E(\varepsilon_i) = 0 \; \forall i$: **Espérance nulle**
2. $\checkmark$ $var(\varepsilon_i) = \sigma^2 \; \forall i$: **Homoscédasticité**
3. $cov(\varepsilon_t, \varepsilon_s) = 0 \; \forall t \neq s$: **Pas d'autocorrélation**
4. $E(\varepsilon_i x_i) = 0 \; \forall i$: **Exogénéité**
5. $\checkmark$ La matrice X est de plein rang : **Pas de multicolinéarité**
6. Le modèle est **correctement spécifié**
7. La variable dépendante Y est **continue**

Table des matières

Table des matières

Définition

- $\exists t, s : cov(\varepsilon_t, \varepsilon_s) \neq 0$

$$\Sigma_\varepsilon = \begin{pmatrix} \sigma^2 & \rho_{12} & \cdots & \rho_{1N} \\ & \sigma^2 & & \rho_{2N} \\ & & \ddots & \vdots \\ sym & & & \sigma^2 \end{pmatrix}$$

- Plusieurs formes sont possibles
  - Autorégressive d'ordre 1 AR(1) : $\varepsilon_t = \rho\varepsilon_{t-1} + \mu_t$
    - avec μ_t "bruit blanc"
  - À préciser (séries temporelles)
- $\Sigma_\varepsilon \neq \sigma^2 I_N$
  - La matrice de var-cov des erreurs n'est pas **diagonale**
  - Les observations ne sont **pas indépendantes** entre elles

Conséquences

- En principe, $\hat{\beta}_{MCO}$ est **non-biaisé** et **consistant**
  - Ces propriétés ne dépendent pas de Σ_ε
  - Sauf si autocorrélation causée par un problème plus grave
    - Endogénéité
    - Intégration I(1) d'une série temporelle
- Gauss-Markov ne s'applique plus $\Longrightarrow$ MCO n'est **plus efficient**
- Comme hétéroscédasticité : $\Sigma_{\hat{\beta}} = \left(X'X\right)^{-1} X'\Sigma_\varepsilon X \left(X'X\right)^{-1}$
  - Inférence basée sur $\Sigma_{\hat{\beta}} = \sigma^2 \left(X'X\right)^{-1}$ fausse : tests t, F...
  - Bootstrap faux car pas indépendance
  - $\exists$ estimateur pour $\Sigma_{\hat{\beta}}$ semblable à l'estimateur de White en cas d'hétéroscédasticité
    - Peu utilisé car l'aurocorrélation est souvent dans un contexte de séries temporelles et on veut les étudier plus en détails
    - Exception parfois avec données de panel

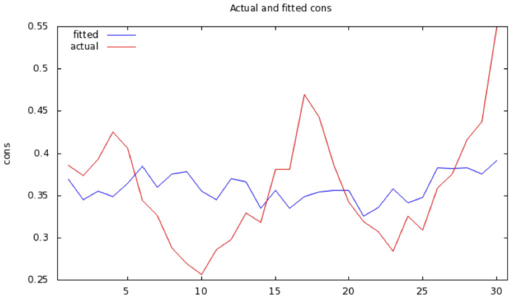
Table des matières

Autocorrélation en séries temporelles

- Forme “naturelle” : Le passé influence le présent / persistance
  - Ex. demande de monnaie
    - $\ln M1_t = \beta_0 + \beta_1 \ln PIB_t + \beta_2 \ln IPC_t + \varepsilon_t$
  - Cette influence peut être capturée par les régresseurs ou pas
- Terme d’erreur capture l’influence de variables ou de chocs **omis**
  - Variables omises : tendances et cycles (ci-dessous)
  - Chocs : Souvent l’effet dure plus d’une période
    - Cas le plus abondant et objet d’étude des séries temporelles
    - Cas AR(1) dans les sections suivantes
    - On dit souvent “autocorrélation sérielle”
  - Conséquences potentiellement graves (Endogénéité)

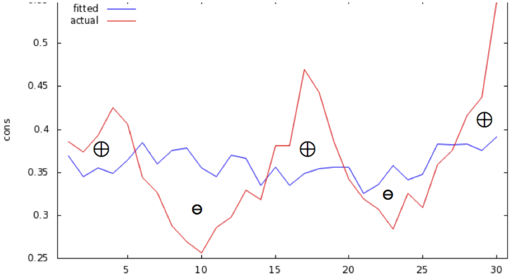
Exemple : Omission d'un terme cyclique

- Données Verbeek, moyennes / 4 semaines, Etats-Unis, 1951-53
- Expliquée : Consommation de crème glacée per capita
- Régresseurs inclus : prix et revenus moyens
- Régresseur omis : température moyenne



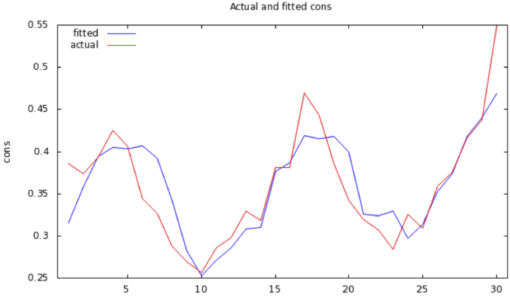
Omission d'un terme cyclique (2)

- La température suit des cycles annuels (été–hiver)
  - Les ventes de crème glacée aussi !
- Si la température n'est pas incluse comme régresseur, les résidus suivront les cycles été–hiver
  - Ces résidus auront tendance à apparaître groupés par signe = autocorrélation



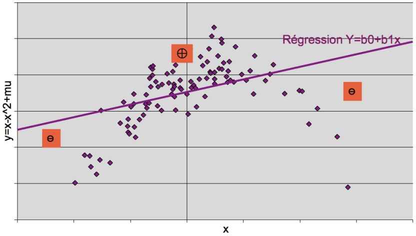
Omission d'un terme cyclique : Correction de l'autocorrélation

- Si on a les données correspondantes au régresseur manquant, on ajoute ce régresseur manquant
  - Si on a pas les données ...



Régresseur quadratique manquant Regr manquant.ods

- Monte-Carlo : $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ avec $\beta_1 > 0$ et $\beta_2 < 0$
  - Mais on régresse $Y = \beta_0 + \beta_1 x + \mu$
    - On “ajuste” une droite dans un nuage de points “courbe”
  - Résidus $\ominus$ pour des valeurs petites et grandes de X et $\oplus$ pour des valeurs moyennes de X
  - Il faut ordonner selon X pour le voir
    - Pas facilement détecté par logiciel, sauf si X croit avec le temps

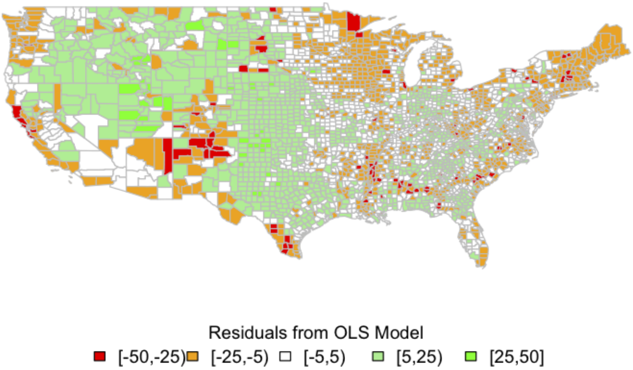


- Conséquences potentiellement graves (Endogénéité)

Autocorrélation en coupe transversale

- Généralement : autocorrélation si données **structurées**
  - Coupe transversale, échantillonnage aléatoire ⇒ pas d'autocorrélation
  - Corrélation spatiale

Résidus régression % County votes pour Bush contre revenu par habitant



Source Y.M. Zukhov, 2010

Table des matières

Séries temporelles

- En dehors du cas spatial, l'autocorrélation n'est considérée que pour les séries chronologiques
  - Variables macro-économiques : PIB, inflation, chômage...
  - Suivi mensuel d'un individu / population : emploi, salaire, consommation...
  - Cotation d'une action : annuelle, mensuelle, journalière, à la minute...
  - Taux de change
  - Flux de passagers dans une compagnie aérienne
  - Ventes / achats
  - Rendements : plante, entreprise, titre...
- La suite de ce chapitre leur est consacrée

Séries temporelles vs. coupes transversales

- Les observations des séries chronologiques sont ordonnées
  - l'ordre des observations d'une coupe transversale n'a pas d'importance
- Les observations de séries chronologiques sont issues d'un **processus stochastique** (aléatoire)
  - pas d'un échantillon aléatoire (coupes transversales)
- Les modèles chronologiques sont généralement indexés par t :

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_k x_{kt} + \varepsilon_t$$

Modèles à retards distribués

- Le modèle $y_t = \beta_0 + \delta_0 x_t + \varepsilon_t$ est dit **statique**
  - Courbe de Phillips classique $inflation_t = \beta_0 + \beta_1 chomage_t + \varepsilon_t$
- Modèles à **retards distribués** (du régresseur) finis
  - un ou plusieurs x affectent y avec un ou plusieurs retards (lags)
  - $gf_t = \beta_0 + \delta_0 te_t + \delta_1 te_{t-1} + \delta_2 te_{t-2} + \varepsilon_t$
    - gf “general (average) fertility”
    - te “tax exemption”
    - “à l’ordre 2”
  - δ_0 = impact immédiat (= de court terme) de x sur y
  - L’ensemble $\delta_0, \delta_1, \ldots, \delta_q$ décrit la relation de long terme entre x et y

Modèles à retards distribués

- Modèle d'ordre 2 $y_t = \beta_0 + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \varepsilon_t$
- Choc transitoire (1 t) Δsur x constant intervenant en t
  - $y_t = \beta_0 + \delta_0 (x + \Delta) + \delta_1 x + \delta_2 x + \varepsilon_t$
  - $y_{t+1} = \beta_0 + \delta_0 x + \delta_1 (x + \Delta) + \delta_2 x + \varepsilon_{t+1}$
  - $y_{t+2} = \beta_0 + \delta_0 x + \delta_1 x + \delta_2 (x + \Delta) + \varepsilon_{t+2}$
- Choc permanent (à partir de t) Δsur x constant
  - $y_t = \beta_0 + \delta_0 (x + \Delta) + \delta_1 x + \delta_2 x + \varepsilon_t$
  - $y_{t+1} = \beta_0 + \delta_0 (x + \Delta) + \delta_1 (x + \Delta) + \delta_2 x + \varepsilon_{t+1}$
  - $y_{t+2} = \beta_0 + \delta_0 (x + \Delta) + \delta_1 (x + \Delta) + \delta_2 (x + \Delta) + \varepsilon_{t+2}$

Table des matières

Modèles à tendance

- Le modèle $y_t = \beta_0 + \delta_0 t + \varepsilon_t$ est une **tendance**
  - y “suit” le temps avec un bruit stochastique ε
- Multiples spécifications de tendance
  - Linéaire $y_t = \beta_0 + \beta_1 z_t + \beta_2 \mathbf{t} + \varepsilon_t$, $t = 1, 2 \ldots T$
  - Exponentielle $\ln y_t = \beta_0 + \beta_1 z_t + \beta_2 \mathbf{t} + \varepsilon_t$
    - Arrive lorsque y a le même taux de croissance t après t (ci-dessous)
  - Quadratique $y_t = \beta_0 + \beta_1 z_t + \beta_2 \mathbf{t} + \beta_3 \mathbf{t^2} + \varepsilon_t$

Note sur log

- $\Delta \log(y_t) = \log(y_t) - \log(y_{t-1}) \approx \frac{y_t - y_{t-1}}{y_{y-1}}$
  - Le **différentiel** de log (ln en fait) est approximativement égal au taux de croissance
  - Pour des taux plutôt faibles
- Une tendance exponentielle sans régresseur est alors $\ln y_t = \beta_0 + \beta_2 \mathbf{t} + \varepsilon_t$
  - $\Delta ln(y_t) = \beta_2 + \Delta\varepsilon_t$: tx de croissance constant + une erreur à espérance nulle

Régression spurieuse

- Différentes variables économiques chronologiques ont souvent une **tendance** temporelle
- *Régresser une tendance sur une tendance semble souvent bon*
  - R^2 , t élevés
  - Des facteurs **inobservés** par l'économètre peuvent causer les tendances
    - $\equiv$ problème des cigognes
  - Ces facteurs inobservés peuvent être **contrôlés** en modélisant la tendance temporelle
    - p.e. introduire un régresseur $t = 1, \ldots, T$ peut ramener la significativité des autres régresseurs à leur juste niveau
  - **Régression fallacieuse** (spurieuse/spurious)

Exemple tableur Trend Coint RndWalk.odt

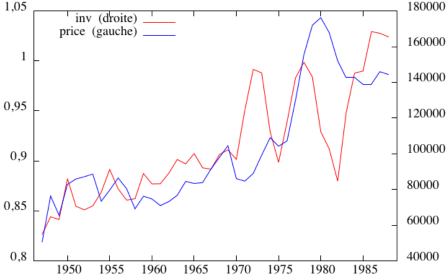
- Générer y=a+bt+µ
  - =2+3*C3+10*SQRT(-2*LN(RAND()))*SIN(2*PI()*RAND())
- Générer x=ct+v :
  - =-10*C3+10*SQRT(-2*LN(RAND()))*SIN(2*PI()*RAND())
  - indépendant de y car RAND() génére indépendamment
- Régresser y sur t et une constante (plage L)
  - =LINEST(A3 :A53 ;C3 :C53 ;1 ;1)
- Régresser y sur x et une constante (plage M)
  - =LINEST(A3 :A53 ;H3 :H53 ;1 ;1)
  - Calculer t-stat : x semble significatif
- Insérer tendance dans régression y sur x
  - x n'est plus significatif
- Régresser detrended y sur detrended x

Exemple : investissement immobilier et prix

- Fichier Gretl hseinv.gdt dans données Wooldridge
- Série 1947-88 “housing investment and housing price index”

$$\ln \widehat{(invpc)} = -.55 + 1.24 \ln (price)$$

- L’élasticité de l’investissement individuel (pc) p/r prix est significativement différente de zéro, mais pas de un.
  - Un changement de prix est répercuté complèment sur l’investissement
  - Mais les deux séries suivent une tendance



Exemple : investissement immobilier et prix

- En rajoutant une tendance

$$\widehat{\ln(invpc)} = -.91 - .38\ln(price) + .0098t$$

- Le prix cesse d'être significatif
  - Mais l'investissement (réel) croit d'environ 1% l'an
  - Possiblement, cela reflète l'effet de regresseurs omis
- Le résultat antérieur était spurieux
- Si y et x ont des trends opposés
  - Introduire une tendance peut **accroitre** la significativité de x
  - La t-stat sur une tendance n'est pas nécessairement correcte (section sur I(1))

Purger une tendance (detrending)

- Au lieu d'introduire une tendance linéaire : purger les données de la tendance (**detrend**)

1. Régression de chaque variable du modèle sur une tendance
2. Utilisation des résidus de chaque équation comme nouvelles variables

- Par exemple $y_t = \beta_0 + \beta_1 z_t + \beta_2 t + \varepsilon_t$

1. Création de variables purgées de la tendance
   $y_t = \gamma_0 + \gamma_1 t + \zeta_t \dashrightarrow y_t^d = \hat{\zeta}_t$
   $z_t = \theta_0 + \theta_1 t + \xi_t \dashrightarrow z_t^d = \hat{\xi}_t$
2. Régression sur les variables purgées $y_t^d = \lambda z_t^d + \nu_t$
   - Plus utile de mettre une constante car $E\left(y_t^d\right) = E\left(z_t^d\right) = 0$ par construction

Purger une tendance (2)

- Purger la tendance est une méthode en 2 étapes qui introduit une **erreur de mesure** sur la deuxième étape
  - Les variables purgées de la tendance sont construites à partir de **paramètres estimés**
  - $\hat{\xi}_t$ est une mesure avec erreur de z_t^d
- L'introduction d'une tendance et l'utilisation de variables purgées de la tendance sont deux approches **équivalentes**
- Remarque sur le R^2
  - Les régressions en séries temporelles ont souvent un R^2 élevé du seul fait de la tendance
    - Mais ne correspond pas au pouvoir explicatif réel du modèle estimé
  - Le R^2 de la régression avec variables purgées de la tendance reflète de manière plus juste le pouvoir explicatif du modèle économique

Table des matières

Stationnarité

- Un processus stochastique est **stationnaire**
  - Lorsque sa distribution ne change pas dans le temps
    - paramètres compris
  - La stationnarité est semblable à "identiquement distribués"
- Les tendances ne sont pas stationnaires
  - car leur espérance change avec le temps
- Un processus stochastique est **covariance-stationnaire**
  - Si son espérance et sa variance sont constantes dans le temps
  - Et si la covariance entre 2 périodes ne dépend que du nombre de périodes entre elles

Intégration

- Un processus stationnaire est
  - **Faiblement dépendant** ou
  - **Intégré d'ordre zéro I(0)** si
  - x_t et x_{t+h} sont "presque indépendant" quand $h \to \infty$
- Une définition semblable existe pour un proc. non stationnaire
- Une série covariance-stationnaire est I(0) si la corrélation entre x_t et $x_{t+h} \to 0$ quand $h \to \infty$
- I(0) implique que la loi des grands nombres et le théorème central limite s'appliquent
  - La dépendance faible est semblable à "indépendamment distribués"
  - Stationnaire + I(0) remplace l'hypothèse d'échantillon aléatoire simple (iid)
  - I(0) est une condition suffisante pour pouvoir utiliser une série en régression

MA(1) : Processus de moyenne mobile d'ordre 1

- MA(1) $x_t = e_t + \alpha e_{t-1},\ t = 1, 2, \ldots$
  - $\{e_t : t = 0, 1, \ldots\}$ est une séquence i.i.d avec moyenne zéro et variance σ_e^2
    - **Bruit blanc**
- Un MA(1) est I(0)
  - Les termes adjacents dans une séquence sont corrélés
  - Dès qu'il y 2 périodes ou plus entre 2 termes d'un MA(1), la corrélation est zéro car e_t est i.i.d.
  - Comme e_t est i.i.d., le MA(1) est stationnaire

Table des matières

AR(1) : $\varepsilon_t = \rho \varepsilon_{t-1} + \mu_t$ processus autorégressif d'ordre 1

- AR(1) est dit **stable** lorsque $|\rho| < 1$
- $\mu_t \sim iid\left(0, \sigma_\mu^2\right)$ **bruit blanc**
  - Espérance 0, variance constante et covariance 0
- On peut écrire $\varepsilon_t = \mu_t + \rho \mu_{t-1} + \rho^2 \mu_{t-2} + \ldots$
  - D'où $var\left(\varepsilon_t\right) = \sigma_\varepsilon^2 = \sigma_\mu^2 + \rho^2 \sigma_\mu^2 + \rho^4 \sigma_\mu^2 + \ldots = \dfrac{\sigma_\mu^2}{1-\rho^2}$
  - Et $cov\left(\varepsilon_t, \varepsilon_{t-1}\right) = cov\left(\rho \varepsilon_{t-1} + \mu_t, \varepsilon_{t-1}\right) = \rho \sigma_\varepsilon^2 = \dfrac{\rho \sigma_\mu^2}{1-\rho^2}$
- Par substitutions répétées dans AR(1)

$$\varepsilon_t = \rho \varepsilon_{t-1} + \mu_t = \rho\left(\rho \varepsilon_{t-2} + \mu_{t-1}\right) + \mu_t = \rho^2 \varepsilon_{t-2} + \rho \mu_{t-1} + \mu_t = \ldots = \rho^s \varepsilon_{t-s} + \sum_{i=0}^{s-1} \rho^i \mu_{t-i}$$

Donc $cov\left(\varepsilon_t, \varepsilon_{t-s}\right) = \dfrac{\rho^s \sigma_\mu^2}{1-\rho^2} = \rho^s \sigma_\varepsilon^2$

Matrice var-cov des erreurs AR(1)

$$
\begin{aligned}
\Sigma_\varepsilon &= \sigma_\varepsilon^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ & 1 & \rho & \cdots & \rho^{T-2} \\ & & \ddots & & \vdots \\ & & & 1 & \rho \\ sym & & & & 1 \end{pmatrix} \\
&= \sigma_\varepsilon^2 I_T + \sigma_\varepsilon^2 \begin{pmatrix} 0 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ & 0 & \rho & \cdots & \rho^{T-2} \\ & & \ddots & & \vdots \\ & & & 0 & \rho \\ sym & & & & 0 \end{pmatrix} \\
&= \sigma_\varepsilon^2 I_T + \sigma_\varepsilon^2 \Phi
\end{aligned}
$$

AR(1) stable est I(0)

- Stationnaire car μ_t i.i.d.
- Cov $\rightarrow$ 0 quand écart entre périodes $\rightarrow \infty$

Matrice var-cov des coefficients MCO avec erreurs AR(1)

- $y = X\beta + \varepsilon$ avec $\varepsilon_t = \rho\varepsilon_{t-1} + \mu_t$
- $$\begin{aligned}\Sigma_{\hat{\beta}} &= \left(X'X\right)^{-1} X'\Sigma_\varepsilon X \left(X'X\right)^{-1} \\ &= \left(X'X\right)^{-1} X' \left[\sigma_\varepsilon^2 I_T + \sigma_\varepsilon^2 \Phi\right] X \left(X'X\right)^{-1} \\ &= \sigma_\varepsilon^2 \left(X'X\right)^{-1} + \sigma_\varepsilon^2 \left(X'X\right)^{-1} X'\Phi X \left(X'X\right)^{-1}\end{aligned}$$
- On ne peut pas dire si elle est + grande que $\Sigma_{\hat{\beta}} = \sigma_\varepsilon^2 \left(X'X\right)^{-1}$
  - On ne peut pas dire si les t-stats seront sur- ou sous-évaluées

Table des matières

Examen visuel des résidus

- Voir exemples antérieurs
  - Autocorrélation $\oplus$
    - Les résidus tendent à être groupés par signe
    - Ils changent de signe trop peu souvent par rapport à des résidus non-autocorrélés
  - Autocorrélation $\ominus$ c'est l'inverse
    - Les résidus alternent plus d'une fois sur deux
- Le plus souvent, l'analyse graphique est **délicate** à interpréter

Test de Durbin-Watson

- Test d'autocorrélation de type AR(1) : $\varepsilon_t = \rho \varepsilon_{t-1} + \mu_t$
  - $|\rho| < 1$
  - $\mu_t \sim iid\left(0, \sigma_\mu^2\right)$
- On veut tester $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$
  - Stat du test : $DW = \dfrac{\sum_{t=2}^{T} \left(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1}\right)^2}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2}$ où $\hat{\varepsilon} = y - X\hat{\beta}$

Analyse de DW

- Si on estime $\hat{\varepsilon} = \alpha\hat{\varepsilon}_{-1} + v$ avec $\hat{\varepsilon}_{-1} =$ vecteur des $\hat{\varepsilon}_{t-1}$
  - $\hat{\alpha} = \left(\hat{\varepsilon}_{-1}'\hat{\varepsilon}_{-1}\right)^{-1}\hat{\varepsilon}_{-1}'\hat{\varepsilon} = \dfrac{\sum_{t=2}^{T}\hat{\varepsilon}_{t-1}\hat{\varepsilon}_t}{\sum_{t=2}^{T}\hat{\varepsilon}_{t-1}^2} = \hat{\rho}$
    - $\hat{\rho}$ est un estimateur non-biaisé de ρ
  - $DW = \dfrac{\sum_{t=2}^{T}(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{T}\hat{\varepsilon}_t^2} = \dfrac{\sum_{t=2}^{T}\left(\hat{\varepsilon}_t^2 + \hat{\varepsilon}_{t-1}^2\right) - 2\sum_{t=2}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}}{\sum_{t=1}^{T}\hat{\varepsilon}_t^2} \approx 2 - 2\hat{\rho}$

Règle de décision

- Comme $\rho \in [-1, 1]$, $DW \in [0, 4]$ avec
  - Autocorrélation $\ominus$: ρ proche de -1 $\Leftrightarrow$ DW proche de 4
  - Autocorrélation $\oplus$: ρ proche de 1 $\Leftrightarrow$ DW proche de 0
  - Pas d'autocorrélation :ρ proche de 0 $\Leftrightarrow$ DW proche de 2
    - C'est la 2º "règle du 2" en économétrie (la 1º était le t-stat)
- Durbin et Watson (1950) ont tabulé les valeurs critiques de DW au seuil de 5%
  - Taille de l'échantillon T
  - Nombre de variables explicatives k
  - La table donne 2 valeurs d_L et d_U (Low et Up)

FIGURE – Règle de Durbin-Watson

Valeur de dw				
Autoc. positive	?	Pas d'autoc.	?	Autoc. négative

0 dL dU 2 4-dU 4-dL 4

- Si $DW < d_L$: Autocorrélation $\oplus$
- Si $DW > 4 - d_L$: Autocorrélation $\ominus$
- Si $DW \in [d_U, 4 - d_U]$: pas d'autocorrélation
- Entre d_L et d_U et entre $4 - d_U$ et $4 - d_L$ on ne peut conclure

FIGURE – Table de Durbin-Watson à 5%

k = nbr de régresseurs (constante exclue)
n = nbr observations (au minimum 15)

	k = 1		k = 2		k = 3		k = 4	
n	dL	dU	dL	dU	dL	dU	dL	dU
15	1.08	1.36	.95	1.54	.82	1.75	.69	1.97
16	1.10	1.37	.98	1.54	.86	1.73	.74	1.93
17	1.13	1.38	1.02	1.54	.90	1.71	.78	1.90
18	1.16	1.39	1.05	1.53	.93	1.69	.82	1.87
19	1.18	1.40	1.08	1.53	.97	1.68	.86	1.85
20	1.20	1.41	1.10	1.54	1.00	1.68	.90	1.83

Conditions d'utilisation de Durbin Watson

- DW ne permet de tester que l'autocorrélation d'ordre 1
  - Mais souvent c'est la principale
- Intercept obligatoire
- Nombre d'observations ≥ 15
- Pas y_{t-1} dans les variables explicatives
  - AR(1) : $\varepsilon_t = \rho \varepsilon_{t-1} + \mu_t$
  - Si $y_t = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \varepsilon_t$
  - Alors
    $y_t = \beta_0 + \beta_1 z_t + \beta_2 [\beta_0 + \beta_1 z_{t-1} + \beta_2 y_{t-2} + \varepsilon_{t-1}] + [\rho \varepsilon_{t-1} + \mu_t]$
    - Donc : corrélation erreur-régresseur : Ch. endogénéité
    - On cumule les problèmes (on ne traite pas)

Tests alternatifs

- DW est un vieux test
  - Un de ses intérêts principaux est de minimiser les calculs
    - On réutilise les résidus de MCO pour un calcul “simple”
- Test de résidus AR(1) : estimer $\hat{\varepsilon} = \alpha\hat{\varepsilon}_{-1} + v$
  - Si $\hat{\alpha}$ est significatif, (t-test) il y a AR(1), $\oplus$ ou $\ominus$ selon le signe
  - Mais si ρ proche de 1, ce test n'est plus valable
- Test de résidus AR(q) : estimer $\hat{\varepsilon} = \alpha\hat{\varepsilon}_{-q} + v$
  - t-test sur $\hat{\alpha}$
- Test d'autocorrélation plus générale : $\hat{\varepsilon} = \sum_{i=1}^{q} \alpha_i \hat{\varepsilon}_{-i} + v$
  - F-test $H_0 : \hat{\alpha}_1 = \hat{\alpha}_2 = \ldots = \hat{\alpha}_q = 0$

Table des matières

Application : Capital Asset Pricing Model

- Modèle d'évaluation des actifs financiers (MEDAF)
- Suppose que tout investisseur compose son portefeuille d'actifs en équilibre entre sa rentabilité et son risque
  - Investisseur au sens large, p.e. entreprise multiproduit
  - Risque = variance de la rentabilité (return)
- D'où chaque investisseur détient un portefeuille dit **mean variance efficient**
  - qui donne le return maximum pour un niveau de risque donné : le risque maximum accepté par l'investisseur
- Si
  1. Tous les investisseurs ont les mêmes croyances sur les risques et les returns
  2. Il n'y a pas de coût de transaction

  Alors, le portefeuille de marché (la somme de tous les portefeuilles individuels) doit **aussi** être mean variance efficient

Formalisation

- Si le marché est mean variance efficient,
  - alors le **return espéré d'un actif individuel est une proportion du return espéré du marché** :

$$E(r_{jt} - r_f) = \beta_j E(r_{mt} - r_f)$$

- où r_{jt} est le return (risqué) sur l'actif j dans la période t
- r_{mt} est le return du marché en t
- r_f est le return sans risque (p.e. bon d'état)
  - La différence espérée $r_{jt} - r_f$ est une rémunération du risque
- β_j = facteur de proportionnalité (inconnu)
  - indique comment des fluctuations du marché affectent le return de l'actif j
  - β_j = volatilité de l'actif / celle du marché : $\beta_j = \dfrac{cov(r_{jt}, r_{mt})}{var(r_{mt})}$

CAPM comme modèle économétrique

- On n'observe pas les espérances E mais seulement les cours réalisés
- Hypothèse d'espérances rationnelles
  - en moyenne les agents ne se trompent pas

Définitions

Return **inespéré de l'actif j** $\mu_{jt} = r_{jt} - E(r_{jt})$
Return inespéré du **marché** $\mu_{mt} = r_{mt} - E(r_{mt})$

- Alors modèle de régression sans intercept
  - $r_{jt} - r_f = \beta_j (r_{mt} - r_f) + \varepsilon_{jt}$
    - avec $\varepsilon_{jt} = \mu_{jt} - \beta_j \mu_{mt}$
  - Pourrait être hétéroscédastique et/ou autocorrélé
  - Sous espérances rationnelles $E(\varepsilon_{jt}) = 0$
  - On peut montrer que $(r_{mt} - r_f)$ n'est pas corrélé à ε_{jt}

Estimation

- Gretl file **capm.gdt**, données Verbeek non-préchargées, disponibles en ligne sur www.econ.kuleuven.ac.be/GME/
  - 1988 :1 à 1996 :2
- Returns de 3 grandes entreprises belges sur la bourse de Bruxelles
  - Pétrofina, Générale de banque, CBR
- Return du marché = Belgian All Share index
- Actif sans risque : bons du trésor à 3 mois

Résultats pour Générale de Banque

- On regarde Générale de Banque : définir les rendements nets
  - Volatilité Générale (coefficient estimé) : $\hat{\beta}_{gén} = .72$
- CAPM implique que d'autres régresseurs ne devraient pas être pertinents y-compris une constante
  - Constante pas significative : conforme à la théorie
  - Effet Janvier : il existe quelques indications que janvier serait un mois dans lequel les returns seraient plus élevés
    - pas dans ces données-ci

Résultats pour Générale de Banque

- Menu modèle → MCO
  - Remarque : boutton “retards” = retards distribués des régresseurs
  - DW est en sortie standard : 1.80
    - avec 98 données & 3 régresseurs, on voit tout de suite qu’on est dans la zone “pas d’autocorrélation”
    - Peut paraitre bizarre, mais la régression porte sur des différences de séries
- post-estimation menu “tests”
  - p-critique DW = 15% environ
  - Autocorrélation AR(x)
    - Sélectionner x = ordre d’autocorrélation
    - Plusieurs tests
    - Sortie MCO modifiée
  - Tests (fin d’output) : rejette Hét
- R^2 “moyen” 47%, mais peu de variables explicatives
  - significativité globale très élevée

Résultats pour CBR

- DW < valeur critique [via tools / statistical tables] : autocorrél $\oplus$
  - C'est une erreur ?
  - Voir aussi DW p-value dans postestimation
- On va voir à présent ce qu'on peut faire

Table des matières

Moindres Carrés Généralisés

- Pour autant que l'autocorrélation ne révèle pas un problème plus grave
  - Régresseur manquant / endogénéité
  - Série temporelle I(1)
- On a vu qu'en présence d'autocorrélation
  - MCO sans biais (si exogénéité stricte, cfr ch. suivant)
  - MCO ne sont plus de variance minimale
- On cherche un estimateur qui soit de variance minimale
- Proposition
  - Soit le modèle $Y = X\beta + \varepsilon$ avec $\Sigma_\varepsilon = \sigma^2 \Psi \neq \sigma^2 I$
  - Alors l'estimateur $\hat{\beta}_{MCG} = \left[X'\Psi^{-1}X\right]^{-1} X'\Psi^{-1}Y$ est efficient

Preuve

- Toute matrice de var-cov est définie positive, donc Ψ aussi
- Pour toute matrice définie positive, on peut définir une matrice carrée non-singulière non nécessairement unique P t.q. $P'P = \Psi^{-1}$
- On transforme le modèle : $PY = PX\beta + P\varepsilon \dashrightarrow Y^* = X^*\beta + \varepsilon^*$
  - $\Sigma_{\varepsilon^*} = E\left(\varepsilon^*\varepsilon^{*'}\right) = E\left(P\varepsilon\varepsilon' P'\right) = PE\left(\varepsilon\varepsilon'\right)P' = \sigma^2 P\Psi P'$
  - $P\Psi P' = P\left(P'P\right)^{-1}P' = PP^{-1}\left(P'\right)^{-1}P' = I$
  - Donc $\Sigma_{\varepsilon^*} = \sigma^2 I$
    - les données transformées ne sont plus autocorrélées
  - Et $E(\varepsilon^*) = 0$
  - Donc MCO appliqué à ces données est efficient

Propriétés des MCG

- L'estimateur des MCG
  - $\hat{\beta}_{MCG} = \left(X^{*'}X^{*}\right)^{-1} X^{*'}Y^{*} = \left(X'\Psi^{-1}X\right)^{-1} X'\Psi^{-1}Y$
  - est BLUE
  - suit asymptotiquement une loi normale centrée en β et de variance

$$\Sigma_{\hat{\beta}_{MCG}} = \sigma^2 \left(X^{*'}X^{*}\right)^{-1} = \sigma^2 \left(X'\Psi^{-1}X\right)^{-1}$$

- Les MCO sont un cas particulier des MCG pour $\Psi = I$

Appliquer les MCG

- Dans la pratique on ne connaît pas Ψ
  - Il faut l'estimer dans une première étape.
- Les tests d'hypothèses des MCO peuvent être directement appliqués au modèle transformé.
- Le R^2 n'est plus comparable à celui des MCO
  - $\exists$ plusieurs alternatives (non développées ici)

MCG faisables

- Pour appliquer les MCG, on suit 3 étapes
  - Estimation de Ψ : $\hat{\Psi}$
  - Transformer les données
  - Appliquer MCO aux données transformées
- L'ensemble s'appelle MCG faisables
  - Parfois MC "quasi-généralisés"
- Rem. La première étape introduit des paramètres estimés (avec erreur) dans la deuxième étape.

Application des MCGF aux erreurs AR(1)

- AR(1) : $\varepsilon_t = \rho \varepsilon_{t-1} + \mu_t$
  - $|\rho| < 1$
  - $\mu_t \sim iid\left(0, \sigma_\mu^2\right)$
- $\Sigma_\varepsilon = \sigma_\varepsilon^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ & 1 & \rho & \cdots & \rho^{T-2} \\ & & \ddots & & \vdots \\ & & & 1 & \rho \\ sym & & & & 1 \end{pmatrix} = \sigma_\varepsilon^2 \Psi$

Application des MCGF aux erreurs AR(1)

- Applique MCO $\dashrightarrow \hat{\varepsilon}$
- On estime $\hat{\varepsilon} = \rho \hat{\varepsilon}_{-1} + v$ avec $\hat{\varepsilon}_{-1} =$ vecteur des $\hat{\varepsilon}_{t-1}$ comme indiqué dans Durbin-Watson $\dashrightarrow \hat{\rho}$
- L'estimation de Ψ est alors

$$\hat{\Psi} = \begin{pmatrix} 1 & \hat{\rho} & \hat{\rho}^2 & \cdots & \hat{\rho}^{T-1} \\ & 1 & \hat{\rho} & \cdots & \hat{\rho}^{T-2} \\ & & \ddots & & \vdots \\ & & & 1 & \hat{\rho} \\ sym & & & & 1 \end{pmatrix}$$

- L'estimation MCGF est $\hat{\beta}_{MCGF} = \left(X' \hat{\Psi}^{-1} X\right)^{-1} X' \hat{\Psi}^{-1} Y$

Procédure itérative de Cochrane-Orcutt

- On itère MCGF :
  - Avec les résidus MCGF $\hat{\varepsilon}_{MCGF} = Y - X\hat{\beta}_{MCGF}$
    - on calcule une nouvelle estimation de ρ : $\hat{\rho}_1$
  - Avec $\hat{\rho}_1$ on calcule une nouvelle estimation MCGF des coefficients
  - On itère jusqu'à ce que les coefficients estimés ne changent plus
- En principe, pas d'avantage à itérer, l'estimateur a toujours les mêmes propriétés
  - Mais beaucoup de logiciels proposent la procédure Cochrane-Orcutt
- Prais-Winsten est une procédure équivalente à Cochrane-Orcutt

Procédure de balayage de Hildreth-Lu

- Première étape : Statistique de Durbin-Watson
  - Autocorrélation positive $\hat{\rho}_0 > 0$ ou négative $\hat{\rho}_0 < 0$
- Deuxième étape : Régression pour l'intervalle des valeurs possibles de ρ
  - Si $\hat{\rho}_0 > 0$ alors on "balaie" (on en essaie plusieurs) sur des valeurs de ρ sur [0,1]
  - Si $\hat{\rho}_0 < 0$ alors on balaie sur des valeurs de ρ sur [-1,0]
- On retient la valeur de ρ qui minimise la somme des carrés des résidus en balayant l'intervalle avec un pas correspondant au degré de précision désiré
- Gretl : Cochrane-Orcutt, Prais-Winsten, Hildreth-Lu accessibles via "modèle" "séries temporelles"

Propriétés de MCGF

- Si l'estimateur de ρ est consistant
  - l'estimation en 2 étapes MCGF est **asymptotiquement** équivalente à utiliser le vrai paramètre ρ
- Les propriétés à échantillon fini de l'estimateur MCGF sont inconnues dans le cas général.
  - N'est plus «sans biais»
  - Les statistiques de tests ne sont que des approximations du fait de l'approximation de ρ par $\hat{\rho}$
- Différentes études (Monte-Carlo) montrent que MCGF est **souvent plus efficient** que MCO
  - mais, si le problème d'autocorrélation n'est pas trop grave, MCO peut être plus efficient que MCGF sur **petits** échantillons
- Si erreur sérieuse sur la forme de l'autocorrélation
  - $\hat{\rho}$ n'est pas un estimateur consistant de ρ (mauvaise spécification)
  - Peut induire inconsistance de MCGF alors que MCO consistant

Application CAPM : Résultats pour CBR

- On a vu : DW < valeur critique [via tools / statistical tables] : autocorrél $\oplus$
- MCGF dans Gretl : Cochrane-Orcutt, Prais-Winsten, Hildreth-Lu
  - accessibles via menu modèle $\rightarrow$ séries temporelles $\rightarrow$ AR(1)
  - menu modèle $\rightarrow$ séries temporelles $\rightarrow$ modèle autorégressif = Cochrane-Orcutt avec des AR(x)