Econométrie approfondie

Michel BEINE

mbeine@ulb.ac.be

CADRE,Université de Lille2, France

et DULBEA, Université Libre de Bruxelles, Belgique.

http://www.users.skynet.be/fa560029

Manuel

- Le cours d'économétrie II est basé sur le livre Introductory Econometrics, a Modern Approach (second edition) de *Jeffrey M. Wooldridge*, édition Thomson South-Western.
- Sur le site web : http://www.swcollege.com/bef/wooldridge/ ... wooldridge2e/wooldridge2e.html
  il y a des données nécessaires pour les exercices et plusieurs liens intéressants.

Plan du cours

- Rappels: Chapitres 1-9 $\rightarrow$ Partie I.

Partie II: Séries temporelles.

- Chapitre 10. Introduction.
- Chapitre 11. Propriétés des MCO.
- Chapitre 12. Corrélation sérielle et hétéroscédasticité.

Plan du cours

Partie III: Techniques avancées.

- Chapitre 13. Pooling: Méthodes simples de Panel.
- Chapitre 14. Données de Panel: techniques avancées.
- Chapitre 15. Variables instrumentales et MCO en 2 étapes.
- Chapitre 16. Modèles à équations simultanées.

Infos pratiques

- Logiciel économétrique: Celui que vous voulez $\rightarrow$ Eviews, Rats, PcGive, etc.
- Examen: modalités à discuter.

Rappels Structure des données: Chapitre 1

Structure des données

Il existe 3 types de données. Chaque type de données peut appeler des techniques économétriques particulières.

- Données “Cross-section”.
- Séries temporelles.
- Données de Panel.

Données “Cross-section”

- Échantillon d’individus, ménages, firmes, ..., pris à un point du temps donné.
- Important: on peut souvent supposer que les obs. = échantillon aléatoire $\rightarrow$ simplifie l’analyse.
- Données très utilisées en économie et sciences sociales $\rightarrow$ micro appliquée: marché du travail, finances publiques, organisation industrielle, économie spatiale, démographie, économie de la santé, etc.
- Example: Wage1.wf1.

Séries temporelles

- Séries chronologiques. Ex: PNB, importations, indices de prix, etc.
- Important: Rarement indépendantes au court du temps $\longrightarrow$ complexifie l'analyse.
- Différentes fréquences: annuel, trimestriel, mensuel, hebdomadaire, journalier, intra-journalier.
- Données très utilisées en macro-économie et en finance.
- Example: PRMINWGE.wf1.

Données de panel

- Série temporelle pour chaque unité/individu.
- Important: la même unité est observée plusieurs fois au court du temps.
- Example: WAGEPAN.wf1.

Rappels Modèle de régression simple: Chapitre 2

Modèle de régression simple

- Objectif: estimer un modèle du type

$$wage = \beta_0 + \beta_1 educ + u. \quad (1)$$

- De manière générale:

$$y = \beta_0 + \beta_1 x + u. \quad (2)$$

- $\rightarrow$ (**??**) est supposé tenir sur la population d'intérêt.
- u est le terme d'erreur (aléas) = facteurs *non-observés* autres que x qui affectent y.
- u et x sont des variables aléatoires.
- Interprétation *Ceteris Paribus*: Si $\Delta u = 0$ (autres facteurs inchangés) $\rightarrow \Delta y = \beta_1 \Delta x$.

Modèle de régression simple

- Pour estimer β_0 et β_1 et garder cette interprétation *CP* il faut faire certaines hypothèses.
- $E(u) = 0$: normalisation $\rightarrow$ on ne perd rien.
- $E(u|x) = E(u)$: pour toute valeur de x, la moyenne des u correspondantes est la même.
  $\rightarrow$ implique la non-corrélation (linéaire). Ex:
  $wage = \beta_0 + \beta_1 educ + u \rightarrow E(abil|8) = E(abil|16)$.
- En combinant ces 2 hypothèses:
  $\rightarrow E(u|x) = E(u) = 0$: hyp. moyenne cond. nulle.
  $\rightarrow E(y|x) = \beta_0 + \beta_1 x$: fonction de régression de la population.

Estimation de β_0 et β_1

$(x_i, y_i) : i = 1, \ldots, n$ = échantillon aléatoire de taille n tiré de la population.

- $y_i = \beta_0 + \beta_1 x_i + u_i, \forall i$.
- $E(u) = 0 \rightarrow E(y - \beta_0 - \beta_1 x) = 0$.
- $E(u|x) = 0 \rightarrow$ Cov. nulle entre u et x
  $\rightarrow E[(y - \beta_0 - \beta_1 x)x] = 0$.
- Pour obtenir $\hat{\beta}_0$ et $\hat{\beta}_1$ on va résoudre ce système en remplaçant $E(.)$ par son équivalant empirique $1/n \sum_{i=1}^{n}(.)$.

Estimation de β_0 et β_1

$$n^{-1}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (3)$$

$$n^{-1}\sum_{i=1}^{n}[(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i] = 0 \quad (4)$$

$$\hookrightarrow \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} \quad (5)$$

$$\hookrightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2 si > 0}. \quad (6)$$

Estimation de β_0 et β_1 par la méthode des moments.

Moindres carrés ordinaires (MCO)

$min_{b_0,b_1} \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2 \rightarrow$ 2 équations “de premier ordre”:

$$-2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (7)$$

$$-2\sum_{i=1}^{n}[(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i] = 0 \quad (8)$$

$$\hookrightarrow \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} \quad (9)$$

$$\hookrightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \text{ si > 0}. \quad (10)$$

Estimation de β_0 et β_1 par la méthode des MCO donne les même $\hat{\beta}_0$ et $\hat{\beta}_1$.

Propriétés des estimateurs MCO

SLR=Simple Linear Regression.

- **SLR.1** $y = \beta_0 + \beta_1 x + u \rightarrow$ linéaire en les paramètres.
- **SLR.2** $(x_i, y_i) : i = 1, \ldots, n \rightarrow$ échantillon aléatoire de taille n tiré de la population.
- **SLR.3** $E(u|x) = 0 \rightarrow$ moyenne conditionnelle nulle. Permet de dériver les propriétés des MCO *conditionnellement* aux valeurs de x_i dans notre échantillon. Techniquement identique à supposer x_i *fixes dans des échantillons répétés* (pas très réaliste).
- **SLR.4** $\sum_{i=1}^{n}(x_i - \overline{x})^2 \neq 0 \rightarrow$ variation dans les x.
- **Théorème 2.1**: Sous les hypothèses SLR.1 - SLR.4, $E(\hat{\beta}_0) = \beta_0$ et $E(\hat{\beta}_1) = \beta_1$.

Propriétés des estimateurs MCO

- **SLR.5** $Var(u|x) = \sigma^2 \rightarrow$ homoscédasticité.
- **Théorème 2.2**: Sous les hypothèses SLR.1 - SLR.5,

$$\begin{aligned} Var(\hat{\beta}_0) &= \frac{\sigma^2 n^{-1} \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \\ Var(\hat{\beta}_1) &= \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \end{aligned}$$

- **Théorème 2.3**: Sous les hypothèses SLR.1 - SLR.5, $E(\hat{\sigma}^2) = \sigma^2$, où $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{u}_i^2$.

Rappels Modèle de régression multiple: Chapitre 3

Modèle de régression multiple

- Objectif: estimer un modèle du type

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + u. \quad (11)$$

- De manière générale:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u. \quad (12)$$

- OLS: $min_{b_0,b_1,\ldots,b_k} \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - \ldots - b_k x_{ik})^2$.

Modèle de régression multiple

Exemple: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$.
$\rightarrow$ Comment obtenir β_1 ?

- Régresser x_1 sur x_2: $\hat{x}_1 = \hat{\alpha}_0 + \hat{\alpha}_1 x_2$.
- Calculer $\hat{r}_1 = x_1 - \hat{x}_1$.
- $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2}$.
- β_1 est bien l'effet net de x_1 sur y, où net signifie après avoir tenu compte de l'effet des autres variables.

Goodness-of-Fit

Il est possible de décomposer la variabilité observée sur y, c-à-d $SST \equiv \sum_{i=1}^{n}(y_i - \overline{y})^2$, en 2 quantités:

- $SSE \equiv \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$: variabilité expliquée par le modèle;
- $SSR \equiv \sum_{i=1}^{n} \hat{u}_i^2$: variabilité **non**-expliquée par le modèle.

$\rightarrow SST = SSE + SSR$.
On peut donc définir une mesure de “qualité” de la régression (GoF):
$R^2 = SSE/SST$ compris entre 0 et 1.

Propriétés des estimateurs MCO

MLR=Multiple Linear Regression.

- **MLR.1** $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u \rightarrow$ linéaire en les paramètres.
- **MLR.2** $(x_{1i}, \ldots, x_{ki}, y_i) : i = 1, \ldots, n \rightarrow$ échantillon aléatoire de taille n tiré de la population.
- **MLR.3** $E(u|x_1, \ldots, x_k) = 0 \rightarrow$ moyenne conditionnelle nulle.
- **MLR.4** $\sum_{i=1}^{n}(x_{ji} - \overline{x})^2 \neq 0 \; \forall j = 1, \ldots, k$ et pas de relation linéaire parfaite entre les $x_j \rightarrow$ variation dans les x_j et pas de collinéarité parfaite.
- **Théorème 3.1**: Sous les hypothèses **MLR.1 - MLR.4**, $E(\hat{\beta}_j) = \beta_k \; \forall j = 0, \ldots, k$.

Propriétés des estimateurs MCO

- **MLR.5** $Var(u|x_1, \ldots, x_k) = \sigma^2 \rightarrow$ homoscédasticité.
- **Théorème 3.2**: Sous les hypothèses **MLR.1 - MLR.5**,

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{(1 - R_j^2) \sum_{i=1}^{n} (x_{ij} - \overline{x}_j)^2}$$

où R_j^2 est le R^2 de le régression de x_j sur les autres x (+ une constante).

- **Théorème 3.3**: Sous les hypothèses **MLR.1 - MLR.5** (appelées hypothèses de Gauss-Markov), $E(\hat{\sigma}^2) = \sigma^2$, où $\hat{\sigma}^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} \hat{u}_i^2$.
- **Théorème 3.4 où théorème de Gauss-Markov**: Sous les hypothèses **MLR.1 - MLR.5**, $\hat{\beta}_j$ est BLUE, $\forall i = 0, \ldots, k$.

Rappels Inférence: Chapitre 4

Hypothèse de normalité

Pour faire de l'inférence statistique, il faut ajouter une hypothèse supplémentaire:

- **MLR.6** $u \sim N(0, \sigma^2)$ et indépendant des $x_j \rightarrow$ normalité.
- Les hypothèses **MLR.1 - MLR.6** sont appelées hypothèses *classiques* du modèle linéaire (CLM) = Gauss-Markov + normalité.
- CLM $\rightarrow$ MCO à la variance minimale.
- Comment justifier cette hypothèse de normalité des résidus ?
- Si u est la somme de beaucoup de facteurs non-observés différents, on peut appeler le théorème central limite pour justifier la normalité.

Théorème Central Limite

- Soit $Y_1, Y_2, \ldots, Y_n$ un échantillon de variables aléatoires de moyenne μ et de variance σ^2.
- $\rightarrow Z_n = \frac{\overline{Y}_n - \mu}{\sigma/\sqrt{n}}$ suit asymptotiquement une distribution N(0,1).
- Si $Y \sim \chi^2(1)$, $\mu = 1$ et $\sigma^2 = 2$.
- Exemple.

Inférence

- **Théorème 4.1**: Sous les hypothèses **MLR.1 - MLR.6**, $\hat{\beta}_j \sim N(\beta_j, Var(\hat{\beta}_j))$, $\forall i = 0, \ldots, k$, où $Var(\hat{\beta}_j) = \frac{\sigma^2}{(1-R_j^2)\sum_{i=1}^{n}(x_{ij}-\overline{x}_j)^2}$.
- Exemple.
- **Théorème 4.2**: Sous les hypothèses **MLR.1 - MLR.6**, $\frac{\hat{\beta}_j-\beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$, $\forall i = 0, \ldots, k$, où $se(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)}$ et σ^2 est remplacé par $\hat{\sigma}^2$.
- On peut donc tester $H_0 : \beta_j = a_j$ contre $H_1 : \beta_j <, \neq, > a_j$.

P-valeur

- P-valeur: “Quelle est le plus petit niveau de significativité auquel H_0 serait rejeté”.
- Comment la calculer? Prendre la t-stat et regarder à quel pourcentile elle correspond dans la distribution de Student appropriée.
- $\rightarrow$ Rejeter H_0 si la P-Valeur < au seuil fixé (ex: 5%).

Tests multiples de restrictions linéaires

- Si $H_0 : \beta_j = 0$ pour $j \in 0, 1, \ldots, k$. Ex: $H_0 : \beta_1 = \beta_2 = 0$.
- et $H_1 : H_0$ n'est pas vrai.
- $\rightarrow 1)$ Estimer le modèle non contraint.
- $\rightarrow 2)$ Estimer le modèle contraint.
- $\rightarrow 3)$ Calculer la statistique $F \equiv \frac{(SSR_c - SSR_{nc})/q}{SSR_{nc}/(n-k-1)}$, où SSR dénote la somme des carrés des résidus et les indices c et nc signifient respectivement contraint et non-contraint.
- Sous H_0, $F \sim F_{q,n-k-1}$, où $F =$Fisher.
- Rejet H_0, si $F > c_\alpha$, où $c = F^{\alpha}_{q,n-k-1}$.

Exercices récapitulatifs

- Exercice 4.12.
- Exercice 4.17.
- Exercice 4.19.

Rappels propriétés asymptotiques des MCO: Chapitre 5

Propriétés exactes des MCO

- Pour prouver le caractère non-biaisé et BLUE des MCO nous avons imposé des hypothèses assez fortes (**MLR.1-MLR.5**).
- Idem pour effectuer de l'inférence statistique (**MLR.6**: normalité).
- Ces propriétés sont vraies quel que soit la taille de l'échantillon ($\forall n$).
- $\rightarrow$ On parle dès lors de propriété exacte, d'échantillon fini, ou encore de petit échantillon.

Propriétés asymptotiques des MCO

- Dans certains cas, le rejet de certaines hypothèses ne signifie pas que les MCO sont invalides.
- Ex: non-normalité de u.
- $\rightarrow$ En effet, les MCO peuvent être encore valides en grand échantillon *sous des hypothèses plus faibles*.
- $\rightarrow$ Étudier les propriétés statistiques pour n grand = étudier les propriétés asymptotiques.
- Nouveau concept: CONVERGENCE.

Propriétés asymptotiques des MCO

- Pour rappel. **Théorème 3.1**: Sous les hypothèses **MLR.1 - MLR.4**, $E(\hat{\beta}_j) = \beta_k \; \forall j = 0, \ldots, k$.
- Pour rappel. **MLR.3** $E(u|x_1, \ldots, x_k) = 0 \rightarrow$ moyenne conditionnelle nulle.
- On a vu que $E(u|x_1) = 0$ implique $Cov(u, x_1) = 0$.
- **Définition**: *plim = Limite en probabilité = valeur vers laquelle un estimateur converge lorsque la taille de l'échantillon tend vers l'infini.* Voir page 741.
- Hors, il est possible de montrer que:

$$\begin{aligned} plim\ \hat{\beta}_1 &= \beta_1 + \frac{Cov(x_1, u)}{Var(x_1)} \\ &= \beta_1, \text{ si } Cov(u, x_1) = 0. \end{aligned}$$

Propriétés asymptotiques des MCO

- $\rightarrow$ Il est possible de relâcher **MLR.3** pour prouver tout de même que les MCO sont convergents.
- **MLR.3'** $E(u) = 0$ et $Cov(x_j, u) = 0 \; \forall j = 1, \ldots, k \rightarrow$ moyenne nulle et correlation nulle.
- $\rightarrow$ Sous les hypothèses **MLR.1 - MLR.2 - MLR.3' - MLR.4**, $\hat{\beta}_j$ est un estimateur convergent de β_j, $\forall j = 1, \ldots, k$.

Normalité asymptotique

- La normalité ne joue aucun rôle dans le caractère non-biasés des MCO ni dans leur caractère BLUE.
- Par contre, pour effectuer de l'inférence statistique nous avons supposé que $u \sim N(0, \sigma^2)$ et donc que la distribution de $y|x_1, \ldots, x_k$ est normale.
- $\rightarrow$ distribution symétrique autour de sa moyenne.
- $\rightarrow$ peut prendre des valeurs sur $\Re$.
- $\rightarrow$ plus de 95% des observations sont comprises entre 2 écart-types.

Normalité asymptotique

- Devons nous abandonner les t-stats si u n'est pas normalement distribué ?
- Non: on fait appel au théorème central limite pour conclure que les MCO satisfont la normalité asymptotique.
- Si n grand, $\hat{\beta}_j$ est *approximativement* $N(\beta_j, Var(\hat{\beta}_j))$.
- Illustration via une Simulation de Monte-Carlo: Prg, Output n=2000,Output n=30.

Rappels Hétéroscédasticité: Chapitre 8

Conséquences de l'hétéroscédasticité

- Une fois de plus l'hétéroscédasticité n'a pas d'influence sur le caractère non-biaisé ou convergent des MCO.
- Par contre la formule traditionnelle de $Var(\hat{\beta}_j)$ donne un estimateur biaisé de la variance des estimateurs si l'hypothèse d'homoscédasticité est violée.
- $\rightarrow$ Inférence incorrecte si on utilise cette formule.
- $\rightarrow$ MCO ne sont plus "BLUE".
- *Solution 1*: Corriger cette formule $\rightarrow$ Écart-types robustes à l'hétéroscédasticité (White, 1980).
- *Solution 2*: Déterminer la source de cette hétéroscédasticité et la prendre en compte dans l'estimation $\rightarrow$ Moindres Carrés Pondérés.

Régression avec des séries temporelles: Chapitre 10

Séries temporelles

Les séries chronologiques diffèrent des séries "cross-section" pour plusieurs raisons:

- $\rightarrow$ l'ordre (temporel) importe;
- $\rightarrow$ le passé influence souvent le futur;
- $\rightarrow$ la notion d'échantillon aléatoire est plus discutable car on n'a qu'une seule réalisation (sauf si on pense que des conditions initiales $\neq$ auraient donné une réalisation $\neq$).
- Example: Phillips.wf1.
- Question: Peut-on toujours utiliser les MCO avec ces séries ?

Séries temporelles

Il y a deux types principaux de modèles de séries temporelles simples:

- $\rightarrow$ les modèles statiques du type:

$$y_t = \beta_0 + \beta_1 z_t + u_t, t = 1, \ldots, n.$$

- $\rightarrow$ les modèles dynamiques du type:

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \ldots + \delta_q z_{t-q} + u_t, t = 1, \ldots, n.$$

Modèles statiques

$y_t = \beta_0 + \beta_1 z_t + u_t, t = 1, \ldots, n.$

- $\Delta y_t = \beta_1 \Delta z_t$ si $\Delta u_t = 0 \rightarrow$ Effet immédiat.
- Example 1. Courbe de Phillips (annuel):
  $inf_t = \beta_0 + \beta_1 unem_t + u_t$.
- Example 2. Consommation de tabac (mensuel):
  $ConsTabac_t = \mu_0 + \mu_1 DepPubl_t + u_t$.

Modèles dynamiques

FDL(2): $y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t, t = 1, \ldots, n.$

- Supposons que $z = c$, $\forall$ observation avant t.
- En $t \rightarrow z = c + 1$.
- En $t + 1, t + 2, \ldots, n \rightarrow z = c$.
- Pour simplifier supposons que $u_t = 0$.

Modèles dynamiques: Δ temporaire

$$
\begin{aligned}
y_{t-1} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c \\
y_t &= \alpha_0 + \delta_0 (c+1) + \delta_1 c + \delta_2 c \\
y_{t+1} &= \alpha_0 + \delta_0 c + \delta_1 (c+1) + \delta_2 c \\
y_{t+2} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 (c+1) \\
y_{t+3} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c
\end{aligned}
$$

Multiplicateurs d'impact

Après un changement temporaire d'une unité de z (c-à-d $\Delta z = 1$)

$\longrightarrow y_t - y_{t-1} = \delta_0 =$ effet immédiat sur y.

$\longrightarrow y_{t+1} - y_{t-1} = \delta_1 =$ effet dans 1 période sur y.

$\longrightarrow y_{t+2} - y_{t-1} = \delta_2 =$ effet dans 2 périodes sur y.

Modèles dynamiques: Δ permanente

$$
\begin{aligned}
y_{t-1} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c \\
y_t &= \alpha_0 + \delta_0 (c+1) + \delta_1 c + \delta_2 c \\
y_{t+1} &= \alpha_0 + \delta_0 (c+1) + \delta_1 (c+1) + \delta_2 c \\
y_{t+2} &= \alpha_0 + \delta_0 (c+1) + \delta_1 (c+1) + \delta_2 (c+1)
\end{aligned}
$$

Multiplicateur de long-terme

Après un changement permanent d'une unité de z (c-à-d $\Delta z = 1$)

$\rightarrow y_t - y_{t-1} = \delta_0 =$ effet immédiat sur y.

$\rightarrow y_{t+1} - y_{t-1} = \delta_0 + \delta_1 =$ effet dans 1 période sur y.

$\rightarrow y_{t+2} - y_{t-1} = \delta_0 + \delta_1 + \delta_2 =$ effet dans 2 périodes sur y.

$\rightarrow \sum_{i=0}^{q} \delta_i =$ multiplicateur de long-terme.

Propriétés des MCO

Nous allons montrer que les MCO sont encore valables avec des séries temporelles et faire le parallèle avec les séries “cross-section”. **Hypothèses de base:**

- **MLR.1** $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u \rightarrow$ linéaire en les paramètres.

  $\Downarrow$

  TS.1 $y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t \rightarrow$ linéaire en les paramètres. Ex: $x_{t3} = salaire_{t-2}$.

- **MLR.2** $(x_{1i}, \ldots, x_{ki}, y_i) : i = 1, \ldots, n \rightarrow$ échantillon aléatoire de taille n tiré de la population.

  $\Downarrow$

  Pas nécessaire.

Propriétés des MCO

- **MLR.3** $E(u|x_1, \ldots, x_k) = 0 \rightarrow$ moyenne conditionnelle nulle.

$$\Downarrow$$

TS.2 $E(u_t|\mathbf{X}) = 0 \; \forall t \rightarrow$ moyenne conditionnelle nulle étant donné les variables explicatives $\forall t$
= exogénéité stricte.
$\rightarrow$ implique exogénéité contemporaine (hypothèse plus faible) : $E(u_t|\mathbf{x}_t) = 0$).
Pour des séries "cross-section", pas nécessaire de vérifier $E(u_t|\mathbf{X}) = 0 \; \forall t$, c-à-d $Corr(u_i, x_j)$ pour $i \neq j$ car **MLR.2**.
! Ne dit rien sur $Corr(u_t, u_s)$ ou $Corr(y_t, y_s)$ pour $t \neq s$!

Causes du rejet de TS.2

Modèle statique simple: $y_t = \beta_0 + \beta_1 z_t + u_t$.
TS.2 $\rightarrow Corr(u_t, z_t) = 0$ et $Corr(u_t, z_s) = 0 \; \forall t \neq s$.

- Variables omises $z_{t-1} \rightarrow Corr(u_t, z_{t-1}) \neq 0$.
- Erreurs de mesure dans des variables exogènes.
- $Corr(u_t, z_{t+1}) \neq 0$.
  Ex: $mrdrte_t = \beta_0 + \beta_1 polpc_t + u_t$ où $mrdrte =$ taux de criminalité dans une ville et $polpc =$ nombre de policiers par personne.
  Quand $\Delta^+ mrdrte_t$ non expliquée $\rightarrow \Delta^+ u_t$.
  Si dans ce cas la ville engage plus de policiers $(\Delta^+ polpc_t) \rightarrow corr(polpc_{t+1}, u_t) > 0$.

Propriétés des MCO

- **MLR.4** $\sum_{i=1}^{n}(x_{ji} - \overline{x})^2 \neq 0 \; \forall j = 1, \ldots, k$ et pas de relation linéaire parfaite entre les x_j $\rightarrow$variation dans les x_j et pas de collinéarité parfaite.

$$\Downarrow$$

TS.3 Pas de variable explicative constante et pas de collinéarité parfaite.

- **Théorème 10.1**: Sous les hypothèses **TS.1 - TS.3**, $E(\hat{\beta}_j) = \beta_j \; \forall j = 0, \ldots, k$ (conditionnellement à **X**).

Propriétés des MCO

- **MLR.5** $Var(u|x_1, \ldots, x_k) = \sigma^2 \rightarrow$ homoscédasticité.

  $\Downarrow$

  TS.4 $Var(u_t|\mathbf{X}) = Var(u) = \sigma^2 \rightarrow \forall t = 1, \ldots, n$ homoscédasticité.
  Exemple de violation: NASDAQ, données journalières (+ de 4000 observations).

- **TS.5** $Corr(u_t, u_s|\mathbf{X}) = 0, \forall\, t \neq s \rightarrow$ pas de corrélation sérielle.
  Exemple de violation: $Corr(u_t, u_{t-1}) = 0.5$.

- **Théorème 10.2**: Sous les hypothèses **TS.1 - TS.5** (appelées hypothèses de Gauss-Markov des séries temporelles), $Var(\hat{\beta}_j) = \frac{\sigma^2}{(1-R_j^2)\sum_{t=1}^{n}(x_{tj}-\overline{x}_j)^2}$, où R_j^2 est le R^2 de la régression de x_j sur les autres x (+ une cst).

Propriétés des MCO

- **Théorème 10.3**: Sous les hypothèses **TS.1 - TS.5**, $E(\hat{\sigma}^2) = \sigma^2$, où $\hat{\sigma}^2 = \frac{1}{n-k-1} \sum_{t=1}^{n} \hat{u}_t^2$.
- **Théorème 10.4 où théorème de Gauss-Markov**: Sous les hypothèses **TS.1 - TS.5**, $\hat{\beta}_j$ est BLUE, $\forall i = 0, \ldots, k$ (conditionnellement à **X**).
- $\rightarrow$ même propriétés qu'avec des séries "cross-section".
- Pour effectuer de l'inférence statistique il faut également supposer:
  MLR.6 $u \sim N(0, \sigma^2)$ et indépendant des $x_j \rightarrow$ normalité.

  $\Downarrow$

  TS.6 $u \sim N(0, \sigma^2)$ et indépendant de **X** $\rightarrow$ normalité.
- L'inférence classique est d'application.

Exemple 10.2 et exercice 10.7

- En utilisant la base de données INTDEF.wf1, estimez le modèle de tx d'intérêt suivant:
  $i3_t = \alpha_0 + \alpha_1 inf_t + \alpha_2 def_t + u_t,$
  où $i3$ est le tx d'intérêt obligataire à 3 mois, inf est le tx d'inflation annuel basé sur l'indice des prix à la consommation et def est le déficit budgétaire en % du PNB.
- En Octobre 1979, la FED a changé sa politique pour modifier la masse monétaire en utilisant directement l'instrument du tx d'intérêt.
  $\rightarrow$ Comment tenir compte de cette information ?
  $\rightarrow$ Est-ce que cela change les résultats ?

Trend linéaire

- Beaucoup de séries ont un trend. Ex: Output par heure aux USA (47-87).
- Modèle du type: $y_t = \alpha_0 + \alpha_1 t + e_t, t = 1, \ldots, n$.
- $\rightarrow$ Quand $\Delta e_t = 0$, $\Delta y_t = y_t - y_{t-1} = \alpha_1$.
- $\rightarrow$ Ceteris paribus, α_1 mesure la Δy_t due au passage du temps.
- $\rightarrow E(y_t) = \alpha_0 + \alpha_1 t$. Croissant ou décroissant en fonction de α_1.
- $\rightarrow Var(y_t) = Var(e_t)$. Le trend n'a pas d'effet sur la variance de y_t.
- $\rightarrow$ On appelle ce type de trend: trend linéaire.

Trend exponentiel

- Beaucoup de séries sont mieux représentées par des trend exponentiels.
- Ex: Série mensuelle représentant le nombre de demandeurs d'emploi dans le canton d'Anderson dans l'Indiana sur la période 1980.1 à 1988.11. Figures: Chômeurs et log(Chômeurs).
- Appelle plutôt une modélisation du type $log(y_t) = \beta_0 + \beta_1 t + e_t, \, t = 1, \ldots, n$.
- $\rightarrow$ comme $\Delta log(y_t) \approx \frac{y_t - y_{t-1}}{y_{t-1}}$ pour $\Delta log(y_t)$ petit, $\Delta log(y_t) = \beta_1$ est approximativement le taux de croissance moyen par période de y_t.

Trend quadratique

- Autre modèle: $y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + e_t$, $t = 1, \ldots, n$.
- $\rightarrow \frac{\Delta y_t}{\Delta t} = \alpha_1 + 2\alpha_2 t$. Si $\alpha_1 > 0$ et $\alpha_2 < 0$, le trend a une forme en bosse.

Exemple

- Données annuelles (1947-1988) sur l'investissement immobilier ($invpc$) et sur un indice des prix immobilier ($price$, $1982 = 1$). Graphique des données.
- Modèle simple sans trend.
- L'élasticité Investissement-Indice de prix est très élevée: 1.241 et statistiquement > 0.
- Modèle simple avec trend.
- L'élasticité Investissement-Indice de prix est négative mais non significative. Par contre le trend est significatif et implique environ une hausse de 1% par an de $invpc$. De plus le R^2 est passé de 0.208 à 0.341.
- $\rightarrow$ relation fallacieuse expliquée par un trend commun, c-à-d l'existence d'une variable omise qui explique l'évolution conjointe des deux séries.

Autre interprétation

- Modèle estimé: $\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{t1} + \hat{\beta}_2 x_{t2} + \hat{\beta}_3 t$.
- Il est facile de montrer que l'on peut obtenir également $\hat{\beta}_1$ et $\hat{\beta}_2$ de cette manière.
- Régresser par MCO:
  y_t sur cst et trend; $\rightarrow$ sauver les résidus: y'_t.
  x_{t1} sur cst et trend; $\rightarrow$ sauver les résidus: x'_{t1}.
  x_{t2} sur cst et trend; $\rightarrow$ sauver les résidus: x'_{t2}.
- Régresser par MCO y'_t sur x'_{t1} et x'_{t2} (avec ou sans constante).
- Revenons à l'exemple précédant. Graphiques log(invpc) et log(price).
- Important: le $R^2 = 0.008 \rightarrow$ sur les variables "détrendées", $log(price)$ n'explique rien de l'évolution de $log(invpc)$.

Séries temporelles (plus) avancées: Chapitre 11

Rappels

- Nous avons étudié les propriétés des MCO en échantillon fini.
- Nous avons obtenu les même propriétés avec des séries temporelles qu'avec des séries "cross-section" mais au prix d'hypothèses très fortes.
- Nous allons tenter de relâcher certaines hypothèses comme celle d'exogénéité stricte, hypothèse violée si par exemple **X** comprend y_{t-i} avec $i > 0$.
- On ne pourra plus étudier le comportement en échantillon fini des MCO mais bien le comportement asymptotique (n grand).
- On va devoir introduire pour cela 2 nouveaux concepts: la stationnarité et la dépendance faible.

Stationnarité

- Définition formelle: *Le processus* $\{x_t : t = 1, 2, \ldots\}$ *est strictement stationnaire si pour chaque collection d'indices de temps* $1 \leq t_1 < t_2 < \ldots < t_m$*, la distribution jointe de* $(x_{t_1}, x_{t_2}, \ldots, x_{t_m})$ *est la même que la distribution jointe de* $(x_{t_1+h}, x_{t_2+h}, \ldots, x_{t_m+h})$*,* $\forall$ *entier* $h \geq 1$*.*
- Les x_t peuvent être corrélés mais la corrélation doit être constante.
- Exemple de processus non-stationnaire:
  $y_t = \alpha_0 + \alpha_1 t + e_t$, $t = 1, \ldots, n$ et $e_t \sim iid(0, 1)$.
  $\rightarrow E(y_t) = \alpha_0 + \alpha_1 t$.
- Par contre, $y_t - \alpha_0 - \alpha_1 t$ est stationnaire.
- Il est très difficile de tester l'hypothèse de stationnarité.

Processus Covariance-stationnaire

- Définition: *Un processus stochastique* $\{x_t : t = 1, 2, \ldots\}$ *avec second moment fini* $[E(x_t^2) < \infty]$ *est* *covariance-stationnaire* *si:*
  i) $E(x_t)$ *est constante;*
  ii) $Var(x_t)$ *est constante;*
  iii) $\forall\ t, h \geq 1, Cov(x_t, x_{t+h})$ *dépend seulement de* h *et non de* t*.*
- $\rightarrow$ implique uniquement les 2 premiers moments.
- Si un processus stationnaire a $E(x_t^2) < \infty \Rightarrow$ il est covariance-stationnaire.
- L'inverse n'est pas vrai.
- Exemples : Rejet et non rejet de l'hypothèses de covariance-stationnarité.

Dépendance faible

- Définition 1: *Un processus stationnaire* $\{x_t : t = 1, 2, \ldots\}$ *est faiblement dépendant si* x_t *et* x_{t+h} *sont* $\pm$ *indépendant au fur et à mesure que* $h \to \infty$*.*
- Définition 2: *Un processus covariance-stationnaire* $\{x_t : t = 1, 2, \ldots\}$ *est faiblement dépendant si* $Corr(x_t, x_{t+h}) \to 0$ *suffisamment rapidement si* $h \to \infty$*.*
- Cette hypothèse remplace celle d'échantillon aléatoire permettant au théorème centrale limite et à la loi des grands nombres de s'appliquer aux séries temporelles.

Exemple: MA(1)

- $MA(1) : x_t = e_t + \alpha_1 e_{t-1}$, où $e_t \sim i.i.d.(0, \sigma_e^2)$.
- $\rightarrow Corr(x_t, x_{t+1}) = \frac{\alpha_1}{1+\alpha_1^2}$.
- $\rightarrow Corr(x_t, x_{t+h}) = 0 \; \forall h > 1$.
- $MA(1)$ est faiblement dépendent.
- Exemple de processus $MA(1)$.

Exemple: AR(1)

- $AR(1) : y_t = \rho_1 y_{t-1} + e_t$, où $e_t \sim i.i.d.(0, \sigma_e^2)$.
- Si $|\rho_1| < 1 \rightarrow Corr(y_t, y_{t+h}) = \rho_1^h \; \forall h \geq 1$.
- $AR(1)$ est faiblement dépendent si $|\rho_1| < 1$ car dans ce cas, $\rho_1^h \rightarrow 0$ si $h \rightarrow \infty$.
- Exemple de processus $AR(1)$ stable.
- Voir les détails mathématiques dans le bouquin.

Propriétés asymptotiques des MCO

Nous allons montrer que les MCO sont encore valables avec des hypothèses plus faibles que celles postulées au Chapitre 10.

Hypothèses de base:

- **TS.1** $y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t \rightarrow$ linéaire en les paramètres.

$$\Downarrow$$

TS.1' $y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t \rightarrow$ linéaire en les paramètres et $(\mathbf{x}_t, y_t) : t = 1, 2, \ldots$ est stationnaire et faiblement dépendant.

Propriétés asymptotiques des MCO

- **TS.2** $E(u_t|\mathbf{X}) = 0 \; \forall t \rightarrow$ moyenne conditionnelle nulle étant donné les variables explicatives $\forall t$
  = exogénéité stricte.

  $\Downarrow$

  TS.2' $E(u_t|\mathbf{x}_t) = 0 \rightarrow$ moyenne conditionnelle nulle étant donné les variables explicatives contemporaines
  = exogénéité contemporaine
  $\rightarrow$ implique que $E(u_t) = 0$ et $Cov(x_{tj}, u_t) = 0$, $\forall j = 1, \ldots, k$.

Propriétés asymptotiques des MCO

- **TS.3** Pas de variable explicative constante et pas de collinéarité parfaite.

$$\Downarrow$$

TS.3' Idem.

- **Théorème 11.1**: Sous les hypothèses **TS.1', TS.2'** et **TS.3'**, $plim\hat{\beta}_j = \beta_j \; \forall j = 0, \ldots, k$.
- AR(1): $y_t = \beta_0 + \beta_1 y_{t-1} + u_t$, où $E(u_t|y_{t-1}, y_{t-2}, \ldots) = 0$.
  $\rightarrow E(y_t|y_{t-1}, y_{t-2}, \ldots) = E(y_t|y_{t-1}) = \beta_0 + \beta_1 y_{t-1}$.
  $\rightarrow$ seulement y_{t-1} affecte la valeur espérée de y_t.
  $\rightarrow$ la valeur espérée de y_t est une fonction linéaire de y_{t-1}.

Estimation du modèle AR(1)

- Comme $\mathbf{x}_t$ contient seulement y_{t-1}, $E(u_t|y_{t-1}, y_{t-2}, \ldots) = 0$ implique que l'hypothèse **TS.2'** tient.
- Par contre **TS.2** ne tient pas. En effet cette hypothèse implique que $\forall t$, u_t est non corrélé avec $x_1, \ldots, x_t, \ldots, x_n$. Ceci est faux pour un processus AR(1) car $Cov(x_{t+1}, u_t) = Cov(y_t, u_t) = Cov(\beta_0 + \beta_1 y_{t-1} + u_t, u_t) = \beta_1^2 Cov(y_{t-1}, u_t) + Cov(u_t, u_t) = 0 + \sigma^2 > 0$, alors que $Cov(y_{t-1}, u_t) = 0 \rightarrow$ exogénéité contemporaine. Pour que l'hypothèse de dépendance faible tienne il faut que $|\beta_1| < 1$. $\rightarrow$ Théorème 11.1: Convergence des MCO.
- Illustration: Modèle classique et Modèle AR(1).

Propriétés asymptotiques des MCO

- **TS.4** $Var(u_t|\mathbf{X}) = Var(u) = \sigma^2, \forall t = 1, \ldots, n \rightarrow$ homoscédasticité.

$$\Downarrow$$

TS.4' $Var(u_t|\mathbf{x}_t) = \sigma^2, \forall t = 1, \ldots, n \rightarrow$ homoscédasticité contemporaine.

- **TS.5** $Corr(u_t, u_s|\mathbf{X}) = 0, \forall\ t \neq s \rightarrow$ pas de corrélation sérielle.

$$\Downarrow$$

TS.5' $E(u_t u_s|\mathbf{x}_t, \mathbf{x}_s) = 0, \forall\ t \neq s \rightarrow$ pas de corrélation sérielle.

En pratique on omet souvent le conditionnement sur $\mathbf{x}_t$ et $\mathbf{x}_s$ et on vérifie que u_t et u_s sont non-corrélés.

Propriétés asymptotiques des MCO

- **Théorème 11.2**: Sous les hypothèses **TS.1' - TS.5'**, les estimateurs MCO sont asymptotiquement normalement distribués. De plus, les écart-types standards sont asymptotiquement valides ainsi que les t, F et LM tests.
- Illustration: Modèle classique et Modèle AR(1).
- Exemple: Tester l'hypothèse d'efficience des marchés. y_t est le rendement hebdomadaire en % (le mercredi) du New York Stock Exchange composite index. Une forme faible de l'hypothèse d'efficience des marché stipule que l'information publique disponible en $t-1$ ne doit pas permettre pour prédire y_t
  $\rightarrow E(y_t|y_{t-1}, y_{t-2}, \ldots) = E(y_t)$.
- $\rightarrow$ Estimer un $AR(q)$ et tester H_0 : coefficient $AR(q)$ sont nuls.

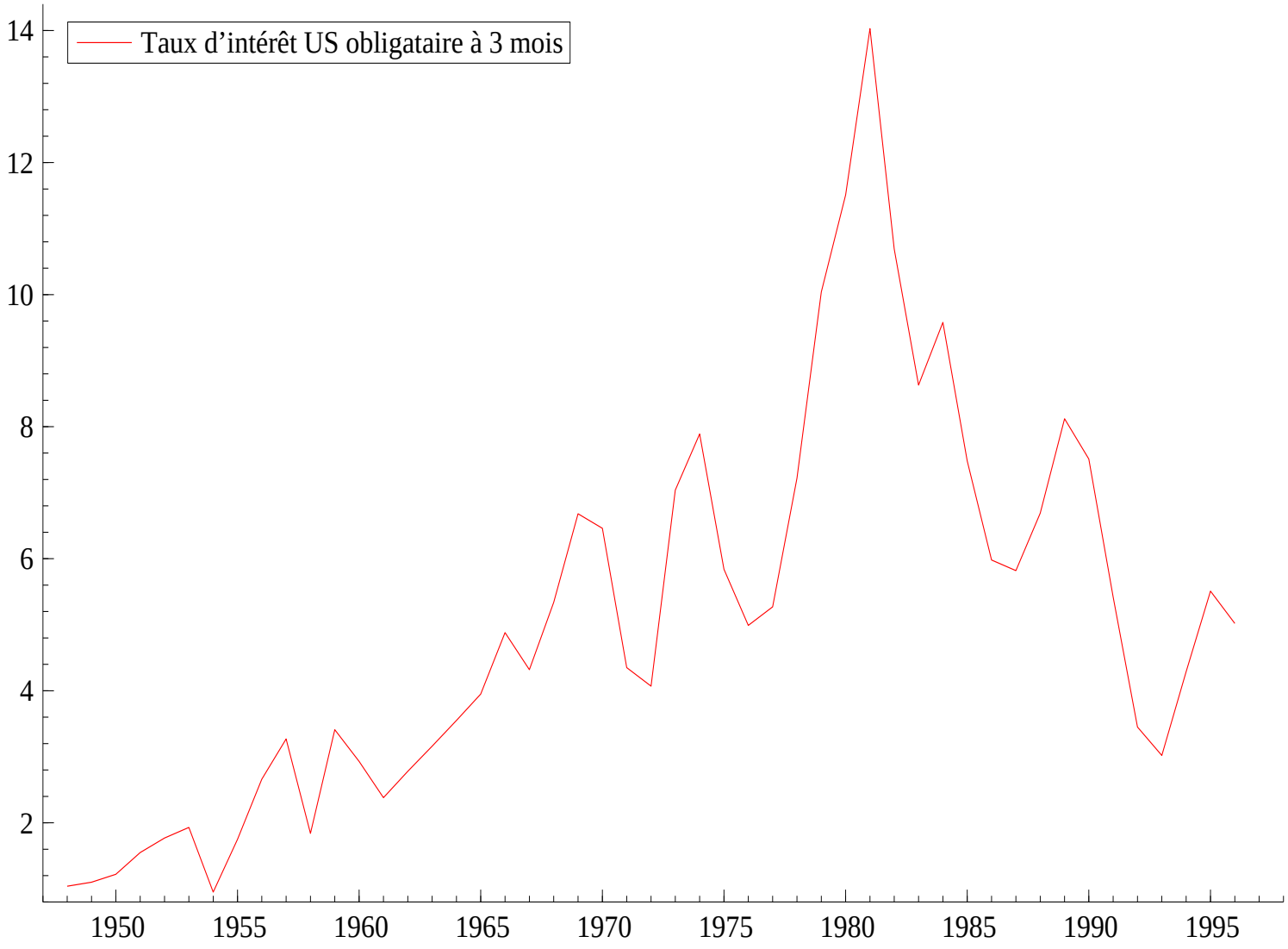
Séries temporelles très persistantes

- Beaucoup de séries chronologiques sont très persistantes, ce qui peut mener au rejet de l'hypothèse de dépendance faible évoquée plus haut.
- L'exemple le plus courant est celui d'une marché aléatoire.
- $\rightarrow y_t = y_{t-1} + e_t, t = 1, 2, \ldots,$ où $e_t \sim iid(0, \sigma_e^2)$.
- $\rightarrow AR(1)$ avec $\rho_1 = 1$.
- $\rightarrow y_t = e_t + e_{t-1} + \ldots + e_1 + y_0$.
- $\rightarrow E(y_t) = E(e_t) + E(e_{t-1}) + \ldots + E(e_1) + E(y_0)$.
- $\rightarrow E(y_t) = E(y_0), \forall t \geq 1$.
- $\rightarrow Var(y_t) =$
  $Var(e_t) + Var(e_{t-1}) + \ldots + Var(e_1) + Var(y_0) = t\sigma_e^2$.

Séries temporelles très persistantes

- Persistance ?
- $\rightarrow y_{t+h} = e_{t+h} + e_{t+h-1} + \ldots + e_{t+1} + y_t$.
- $\rightarrow E(y_{t+h}|y_t) = y_t, \forall h \geq 1$.
- Alors que pour un $AR(1) : E(y_{t+h}|y_t) = \rho_1^h y_t$.
- $\rightarrow Corr(y_t, y_{t+h}) = \sqrt{t/(t+h)} \approx 1$ pour t grand.
- $\rightarrow$ une marche aléatoire n'est pas covariance-stationnaire.
- $\rightarrow$ une marche aléatoire n'est pas faiblement dépendante.
- Illustration.

Exemple de marche aléatoire



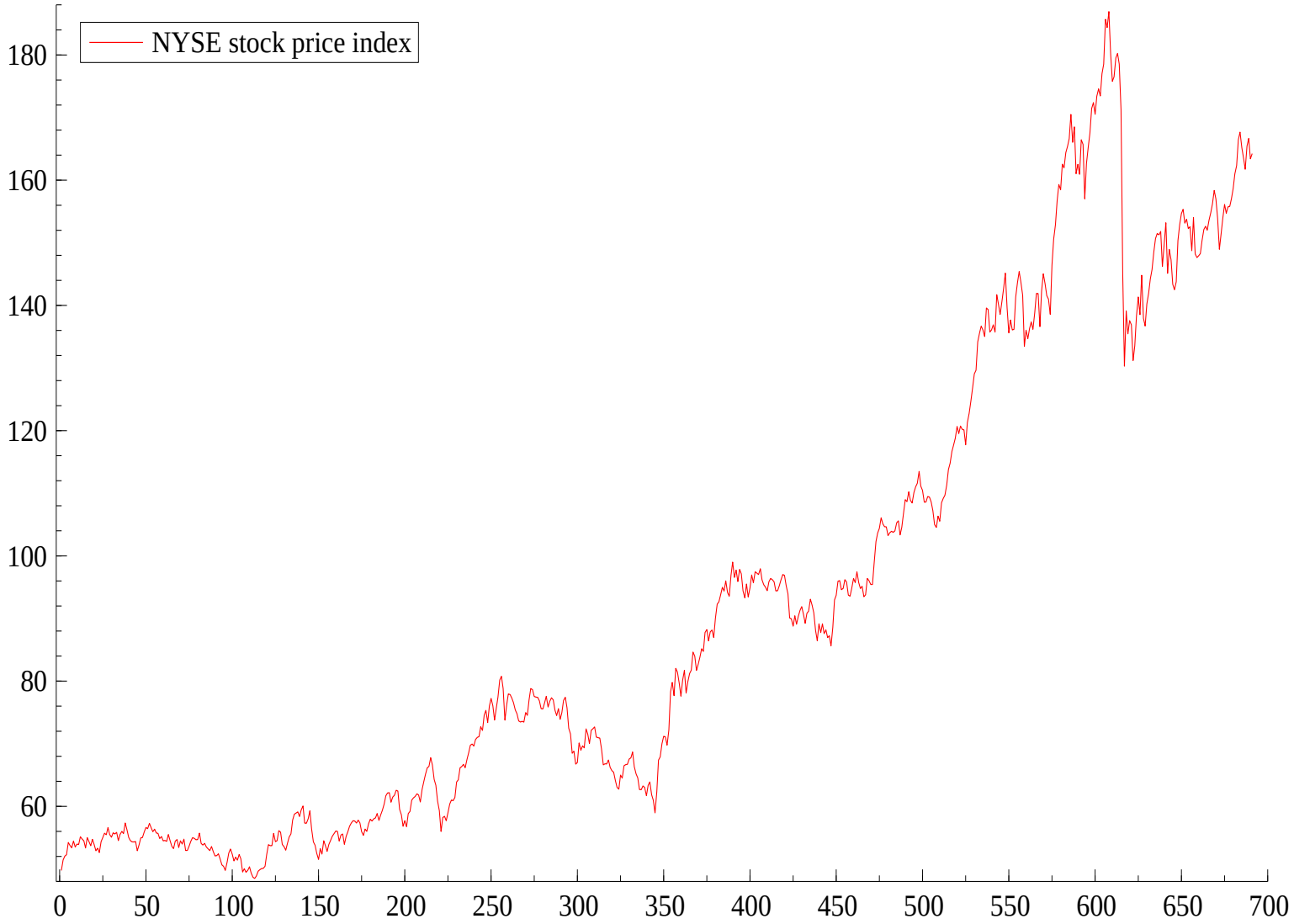
Implication économique

- Si y est faiblement dépendant, une Δ de politique économique affectant y (ex: PNB) aujourd'hui aura peu d'effet dans quelques années.
- Par contre, si y suit une marche aléatoire (ou plus généralement a une racine unitaire, c-à-d $\rho_1 = 1$), une Δ de politique économique affectant y aujourd'hui aura un effet important dans plusieurs années.
- Tester la présence d'une racine unitaire a donc une implication économique importante.

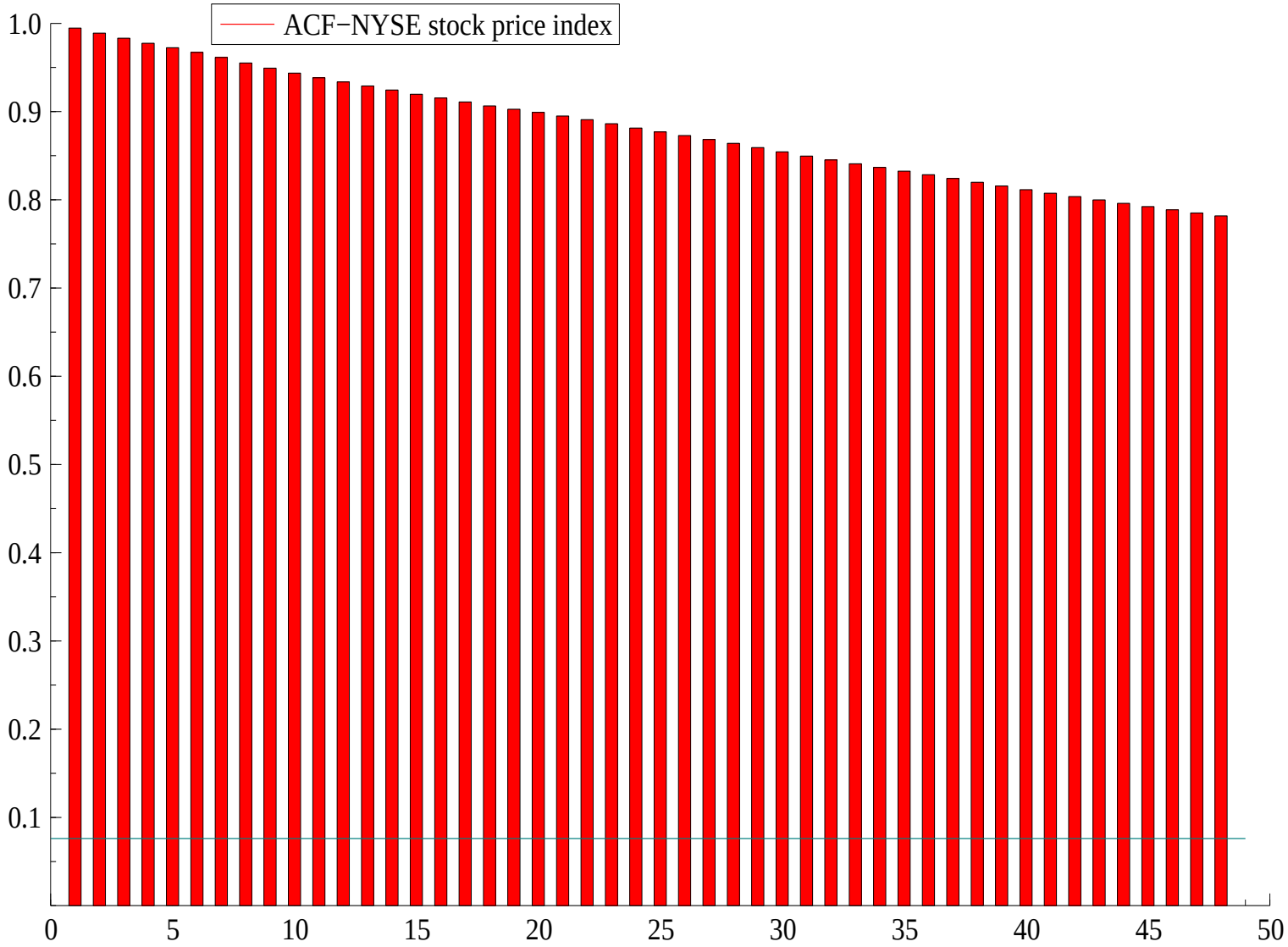
Marche aléatoire avec tendance (drift)

- $y_t = \alpha_0 + y_{t-1} + e_t : AR(1)$ avec constante.
- $\rightarrow y_t = \alpha_0 t + e_t + e_{t-1} + \ldots + e_1 + y_0$.
- $\rightarrow E(y_t) = \alpha_0 t$ si $E(y_0) = 0$.
- $\rightarrow E(y_{t+h}|y_t) = \alpha_0 h + y_t, \forall h \geq 1$.
- $\rightarrow Var(y_t) = t\sigma_e^2$.
- Exemple de marche aléatoire avec tendance.
- Comment détecter une RU ?

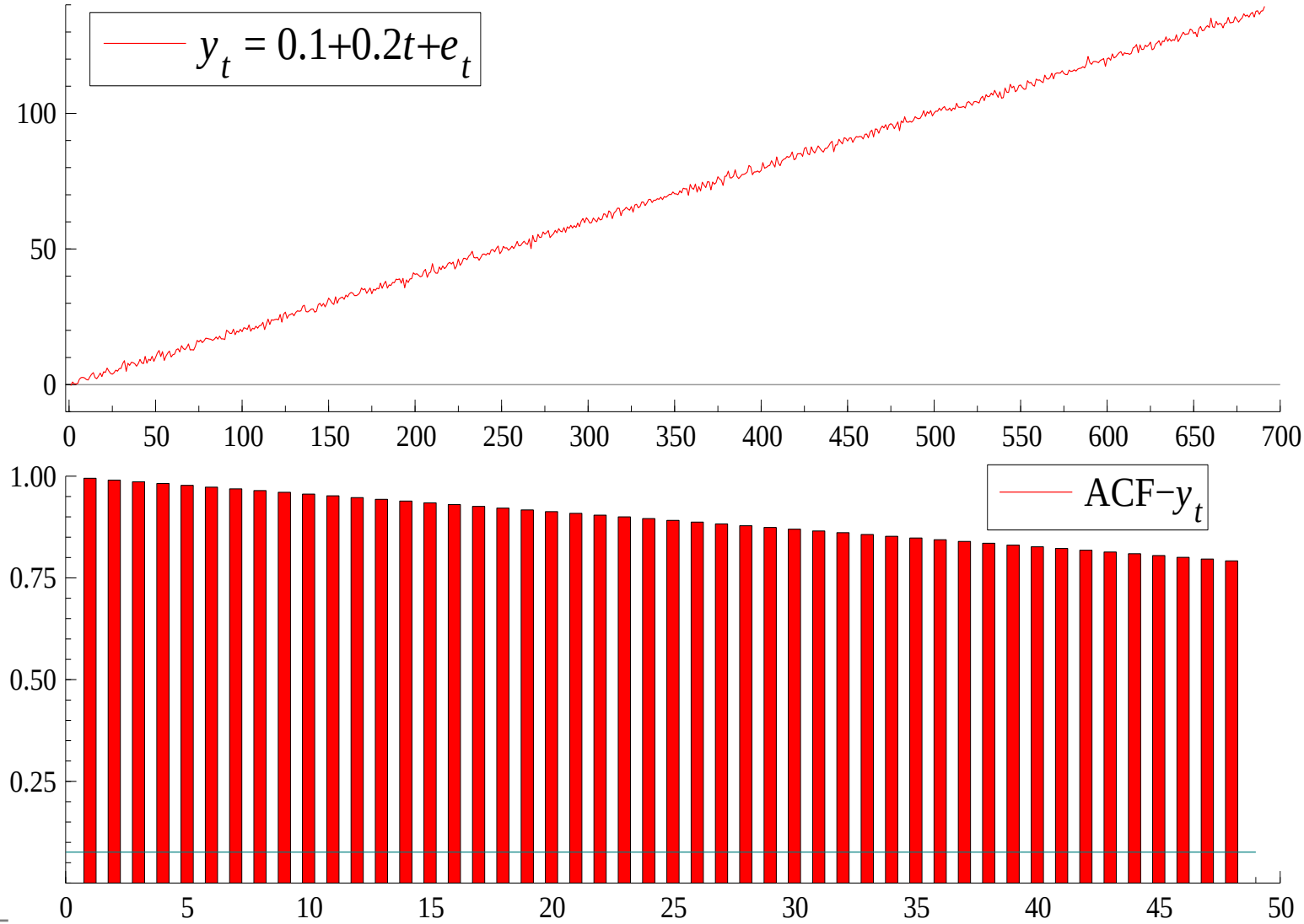
Inspection Visuelle



Correlogramme



Difficile de distinguer RU et trend



Tester la présence d'une RU

- Modèle $AR(1) : y_t = \alpha + \rho y_{t-1} + e_t$.
- y_0 est supposé observé et $E(e_t|y_{t-1}, y_{t-2}, \ldots, y_0) = 0$.
- RU si $\rho = 1 \rightarrow H_0 : \rho = 1$ et $H_1 : \rho < 1$.
- Notons que: $y_t - y_{t-1} = \alpha + (\rho - 1)y_{t-1} + e_t$.
- $\Leftrightarrow \Delta y_t = \alpha + \theta y_{t-1} + e_t$.
- RU si $H_0 : \theta = 0$ et $H_1 : \theta < 0$.
- Problème sous H_0, y_{t-1} n'est pas faiblement dépendant et donc les t-stats n'ont pas une distribution standard (théorème centrale limite pas applicable, même pour n grand).
- $\rightarrow$ Il faut utiliser des valeurs critiques tabulées par Dickey Fuller (1979).

Tester la présence d'une RU

- Valeurs critiques asymptotiques pour un test de RU avec constante mais sans trend:

α	1%	2.5%	5%	10%
Valeur critique	-3.43	-3.12	-2.86	-2.57

- Exemple: Taux d'intérêt US obligataire à 3 mois.
- Notez que à 5% la valeur critique traditionnelle pour n grand est -1.65 contre -2.86.
- Limite: dans le test DF usuel, on suppose sous H_0 que Δy_t n'a pas d'autocorrélation.

Test ADF

- Le test ADF (Augmenté) généralise le test de DF: $\Delta y_t = \alpha + \theta y_{t-1} + \gamma_1 \Delta y_{t-1} + \ldots + \gamma_p \Delta y_{t-p} + e_t$.
- Les valeurs critiques sont les mêmes que le test de DF.
- Les t-stats de $\gamma_1, \ldots, \gamma_p$ sont standards asymptotiquement.
- Le F-test $\gamma_1 = \ldots = \gamma_p = 0$ est valide asymptotiquement.
- Inclure le bon nombre de lags est important car cela pourrait affecter la fiabilité du test de RU. En effet, les valeurs critiques sont calculées en supposant que toute la dynamique de Δy_t a été modélisée.
- Exemple: Taux d'intérêt US obligataire à 3 mois.

Test ADF avec trend

- Nous avons évoqué le fait qu'il est parfois difficile de distinguer RU et tendance.
- $\rightarrow$ une série peut donc être stationnaire autour d'un trend, c-à-d $y_t - \hat{\delta} t$ est faiblement dépendant.
- Il est donc important d'inclure un trend dans le test ADF si celui-ci est apparent (visuellement).
- Le test ADF (Augmenté) généralise le test de DF: $\Delta y_t = \alpha + \delta t + \theta y_{t-1} + \gamma_1 \Delta y_{t-1} + \ldots + \gamma_p \Delta y_{t-p} + e_t$.
- Sous $H_0 : \theta = 0$ et sous $H_1 : y_t$ est stationnaire autour d'un trend.

Test ADF avec trend

- Les valeurs critiques sont différentes.
  Valeurs critiques asymptotiques pour un test de RU avec constante et avec trend:

α	1%	2.5%	5%	10%
Valeur critique	-3.96	-3.66	-3.41	-3.12

- Le t-stat relatif au trend n'a pas une distribution standard.
- Exemple: Taux d'intérêt US obligataire à 3 mois.
- Il y a beaucoup d'autres tests de RU qui ont tous leurs avantages et inconvénients.

Que faire si y_t a une RU ?

- Si $y_t = \alpha + y_{t-1} + e_t \rightarrow \Delta y_t \equiv y_t - y_{t-1} = \alpha + e_t$.
- Si par contre Δy_t est stationnaire, y_t est dit Intégré d'ordre 1 ou $I(1)$ et Δy_t est dit Intégré d'ordre 0 ou $I(0)$.
- Si l'application le permet, on peut donc estimer par MCO un modèle sur Δy_t.
- Dans beaucoup d'applications économiques, on définit des rendements en %: $r_t = 100\Delta log(y_t) \simeq 100\frac{y_t - y_{t-1}}{y_{t-1}}$.
- Si différentier les séries n'est pas possible de part la nature de l'application, on peut avoir recourt à des techniques plus avancées connues sous le nom Cointégration (voir section 18.4), permettant de modéliser y_t et non r_t.
- Qu'est-ce qu'une régression fallacieuse (ou spurious regression) ?

Corrélation sérielle et hétéroskedasticité: Chapitre 12

Biais des MCO et autocorrélation

- Supposons que $y_t = \beta_0 + \beta_1 x_t + u_t$.
- Supposons que le terme d'erreur suit un $AR(1) : u_t = \rho u_{t-1} + e_t$ avec $|\rho < 1|$ et $e_t \sim iid(0, \sigma_e^2)$.
- Chapitre 10: **TS.1-TS.3** $\rightarrow$ MCO non-biaisés pour autant que x soit strictement exogène.
- Chapitre 11: **TS.1'-TS.3'** $\rightarrow$ MCO consistant pour autant que y_t soit faiblement dépendant et $E(u_t|x_t) = 0$.
- Rien n'est dit sur le fait que u_t ne puisse pas suivre un $AR(1)$.
- Par contre les MCO ne sont plus BLUE car violation de **TS.5** et **TS.5'**.

Efficience-Inférence des MCO

- Estimation du modèle $y_t = \beta_0 + \beta_1 x_t + u_t$ par MCO, où u_t supposé $iid(0, \sigma^2)$ et pour simplifier $\overline{x} = 0$.
- Rappelons que u_t suit en réalité un $AR(1)$.
- $\hat{\beta}_1 = \beta_1 + SST_x^{-1} \sum_{i=1}^{n} x_t u_t$, où $SST_x = \sum_{i=1}^{n} x_t^2$.
-

$$Var(\hat{\beta}_1) = SST_x^{-2} Var(\sum_{i=1}^{n} x_t u_t)$$

$$= SST_x^{-2} \left(\sum_{i=1}^{n} x_t^2 Var(u_t) + 2 \sum_{t=1}^{n-1} \sum_{j=1}^{n-t} x_t x_{t+j} E(u_t u_{t+j}) \right)$$

$$= \frac{\sigma^2}{SST_x} + 2 \frac{\sigma^2}{SST_x^2} \sum_{t=1}^{n-1} \sum_{j=1}^{n-t} \rho^j x_t x_{t+j}$$

Efficience-Inférence des MCO

- La formule traditionnelle ignore le second terme.
- Si $\rho > 0 \rightarrow$ on sous-estime $Var(\hat{\beta}_1)$.
- Si $\rho < 0 \rightarrow$ on sur-estime $Var(\hat{\beta}_1)$.
- Question: Supposons que $u_t = e_t + \alpha e_{t-1}$. Montrez que la formule traditionnelle pour calculer $Var(\hat{\beta}_1)$ est incorrecte si $\alpha \neq 0$.

Tester la présence d'autocorrélation

- Dans le modèle général:
  $y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t \rightarrow$ comment tester la présence de corrélation sérielle ?
- Deux types de tests:
- 1. Tests basés sur l'hypothèse que X est strictement exogène: t-test ou DW.
- 2. Tests basés sur l'hypothèse que X n'est pas strictement exogène.

Si X est strictement exogène

- Supposons que $y_t = \beta_0 + \beta_1 x_t + u_t$.
- Nous voulons tester si le terme d'erreur suit un $AR(1) : u_t = \rho u_{t-1} + e_t$ avec $|\rho < 1|$.
- $H_0 : \rho = 0$.
- Développons tout d'abord un test valide quand n est grand et quand X est strictement exogène.
- Hypothèses supplémentaires disant en quelque sorte que e_t est *iid* :
  $E(e_t|u_{t-1}, u_{t-2}, \ldots) = 0$ et $Var(e_t|u_{t-1}) = Var(e_t) = \sigma_e^2$.
- Si les u_t étaient observés on pourrait utiliser le **Théorème 11.2** pour valider l'utilisation asymptotique des t-tests dans la régression $u_t = \rho u_{t-1} + e_t$.
- Que ce passe-t-il si on remplace u_t par $\hat{u}_t$? A noter que $\hat{u}_t$ dépend de $\hat{\beta}_0$ et de $\hat{\beta}_1$.

Marche à suivre pour le t-test

- Estimer $y_t = \beta_0 + \beta_1 x_t + u_t$ et obtenir $\hat{u}_t$.
- Estimer $\hat{u}_t = \rho \hat{u}_{t-1} + e_t$.
- Utiliser un t-test pour tester $H_0 : \rho = 0$ contre $H_1 : \rho <\neq> 0$.

Test de Durbin-Watson

- Le test de Durbin-Watson est une autre test pour tester la corrélation sérielle d'ordre 1 sous l'hypothèse que X est strictement exogène.
- La statistique $DW = \frac{\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n}\hat{u}_t^2}$, où $\hat{u}_t$ est le résidu d'une régression du type $y_t = \beta_0 + \beta_1 x_{1t} + \ldots + beta_k x_{kt} + u_t$.
- Il est facile de montrer que $DW \approx 2(1 - \hat{\rho})$.
- Durbin et Watson (1951) ont dérivé la distribution de DW conditionnellement à X sous l'hypothèse que les hypothèses du modèle linéaire classique sont vérifiées (incluant la normalité).
- La distribution de DW dépend de n, k, et du fait qu'on a inclut une constante ou pas.

Test de Durbin-Watson

- Notons que
  $\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2 = \sum_{t=2}^{n}\hat{u}_t^2 + \sum_{t=2}^{n}\hat{u}_{t-1}^2 - 2\sum_{t=2}^{n}\hat{u}_t\hat{u}_{t-1}$.
- Si $\hat{\rho} = 1 \rightarrow DW = 0$.
- Si $\hat{\rho} = -1 \rightarrow DW = 4$.
- Si $\hat{\rho} = 0 \rightarrow DW = 2$.
- $H_0 : \rho = 0$ contre généralement $H_1 : \rho > 0$.
- Durbin et Watson (1951) reportent des valeurs critiques inférieures d_L et supérieures d_U.
- Si $d_L \leq DW \leq d_U$, on ne rejette pas H_0.
- La plupart des logiciels économétriques reportent DW mais pas les valeurs critiques.

Table de Durbin-Watson pour $H_1 : \rho > 0$

Table E.8 *Critical Values d_L and d_U of the Durbin-Watson Statistic D (Critical Values Are One-Sided)*[a]

	α = .05										α = .01									
	P = 1		P = 2		P = 3		P = 4		P = 5		P = 1		P = 2		P = 3		P = 4		P = 5	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	.95	1.54	.82	1.75	.69	1.97	.56	2.21	.81	1.07	.70	1.25	.59	1.46	.49	1.70	.39	1.96
16	1.10	1.37	.98	1.54	.86	1.73	.74	1.93	.62	2.15	.84	1.09	.74	1.25	.63	1.44	.53	1.66	.44	1.90
17	1.13	1.38	1.02	1.54	.90	1.71	.78	1.90	.67	2.10	.87	1.10	.77	1.25	.67	1.43	.57	1.63	.48	1.85
18	1.16	1.39	1.05	1.53	.93	1.69	.82	1.87	.71	2.06	.90	1.12	.80	1.26	.71	1.42	.61	1.60	.52	1.80
19	1.18	1.40	1.08	1.53	.97	1.68	.86	1.85	.75	2.02	.93	1.13	.83	1.26	.74	1.41	.65	1.58	.56	1.77
20	1.20	1.41	1.10	1.54	1.00	1.68	.90	1.83	.79	1.99	.95	1.15	.86	1.27	.77	1.41	.68	1.57	.60	1.74
21	1.22	1.42	1.13	1.54	1.03	1.67	.93	1.81	.83	1.96	.97	1.16	.89	1.27	.80	1.41	.72	1.55	.63	1.71
22	1.24	1.43	1.15	1.54	1.05	1.66	.96	1.80	.86	1.94	1.00	1.17	.91	1.28	.83	1.40	.75	1.54	.66	1.69
23	1.26	1.44	1.17	1.54	1.08	1.66	.99	1.79	.90	1.92	1.02	1.19	.94	1.29	.86	1.40	.77	1.53	.70	1.67
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	.93	1.90	1.04	1.20	.96	1.30	.88	1.41	.80	1.53	.72	1.66
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	.95	1.89	1.05	1.21	.98	1.30	.90	1.41	.83	1.52	.75	1.65
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	.98	1.88	1.07	1.22	1.00	1.31	.93	1.41	.85	1.52	.78	1.64
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86	1.09	1.23	1.02	1.32	.95	1.41	.88	1.51	.81*	1.63
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85	1.10	1.24	1.04	1.32	.97	1.41	.90	1.51	.83	1.62
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84	1.12	1.25	1.05	1.33	.99	1.42	.92	1.51	.85	1.61
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83	1.13	1.26	1.07	1.34	1.01	1.42	.94	1.51	.88	1.61
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83	1.15	1.27	1.08	1.34	1.02	1.42	.96	1.51	.90	1.60
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82	1.16	1.28	1.10	1.35	1.04	1.43	.98	1.51	.92	1.60
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	.94	1.59
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	.95	1.59
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	.97	1.59
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	.99	1.59
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

[a] *n = number of observations; P = number of independent variables.*

Source: This table is reproduced from Biometrika, *41 (1951): 173 and 175, with the permission of the* Biometrika *Trustees.*

Si X n'est pas strictement exogène

- Durbin (1970 propose un autre test valable si X n'est pas strictement exogène, par exemple si le modèle contient y_{t-1} comme variable explicative.
- $i)$ Estimer le modèle $y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t$ et obtenir $\hat{u}_t$, $\forall t = 1, 2, \ldots, n$.
- $ii)$ Estimer le modèle $\hat{u}_t = \gamma_0 + \gamma_1 x_{t1} + \ldots + \gamma_k x_{tk} + \rho \hat{u}_{t-1}$, $\forall t = 2, 3, \ldots, n$ et obtenir $\hat{\rho}$ ainsi que $t_{\hat{\rho}}$.
- $iii)$ Utiliser $t_{\hat{\rho}}$ de la manière usuelle pour tester $H_0 : \rho = 0$ contre $H_1 : \rho <\neq> 0$.
- $\rightarrow$ On régresse $\hat{u}_t$ sur x_t et $\hat{u}_{t-1}$ et donc on permet à chaque x_{tj} d'être corrélé avec u_{t-1}.
- $\rightarrow$ $t_{\hat{\rho}}$ a approximativement une distribution en t si n est grand.

Tester un *AR(q)*

- $i)$ Estimer le modèle $y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t$ et obtenir $\hat{u}_t$, $\forall t = 1, 2, \ldots, n$.
- $ii)$ Estimer le modèle
  $\hat{u}_t = \gamma_0 + \gamma_1 x_{t1} + \ldots + \gamma_k x_{tk} + \rho_1 \hat{u}_{t-1} + \ldots + \rho_q \hat{u}_{t-q}$,
  $\forall t = q+1, q+2, \ldots, n$.
- $iii)$ Effectuer le F-test suivant $H_0 : \rho_1 = \ldots = \rho_q = 0$ contre H_1 : un des $\rho_j <\neq> 0$, $\forall j = 1, \ldots, q$.
- Si les x_t sont supposés strictement exogènes, ont peut les omettre dans les étapes $i)$ et $ii)$.
- A noter que ces tests supposent $Var(u_t | x_t, u_{t-1}, \ldots, u_{t-q}) = \sigma^2$. Il existe une version de ces tests robuste à l'hétéroskedasticité comme on le verra plus loin.

Tester un *AR(q)*

- Une alternative a F-test est d'utiliser un Lagrange Multiplier (LM) Test: $LM = (n-q)R^2_{\hat{u}}$,
  où $R^2_{\hat{u}}$ est le R^2 de la régression
  $\hat{u}_t = \gamma_0 + \gamma_1 x_{t1} + \ldots + \gamma_k x_{tk} + \rho_1 \hat{u}_{t-1} + \ldots + \rho_q \hat{u}_{t-q}$,
  $\forall t = q+1, q+2, \ldots, n$.
- Sous $H_0, LM \sim \chi^2_q$ asymptotiquement.
- Ce test est connu sous le nom de test de **Breusch-Godfrey**.
- Il est test est disponible en Eviews (avec le F-test).
- Exemple: Taux d'intérêt US obligataire à 3 mois.
  $ci3_t = \alpha_0 + \alpha_1 ci3_{t-1}$.

Correction pour corrélation sérielle

- Si on détecte de la corrélation sérielle, on peut modifier le modèle initial pour tenter d'obtenir un modèle dynamiquement complet (ex: $AR(1) \rightarrow AR(2)$).
- Dans certains cas nous ne sommes pas intéressé par modéliser cette dynamique $\rightarrow$ l'intérêt réside plutôt dans les autres variables incluses dans le modèle.
- Mais l'inférence est compromise $\rightarrow$ Que faire ?
- $\rightarrow$ Calculer des écart-types robustes à n'importe quelle forme de corrélation sérielle.

Écart-types robustes

- Considérons le modèle $y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t$, $t = 1, \ldots, n$.
- Comment obtenir un écart-type pour $\hat{\beta}_1$ robuste à la corrélation sérielle ?
- $x_{t1} = \delta_0 + \delta_2 x_{t2} \ldots + \delta_k x_{tk} + r_t$, où $E(r_t) = 0$ et $Corr(r_t, x_{tj}) = 0, \forall j \geq 2$.
- Il est possible de montrer que $Avar(\hat{\beta}_1) = \frac{Var\left(\sum_{i=1}^{n} r_t u_t\right)}{\left(\sum_{i=1}^{n} E(r_t^2)\right)^2}$, où $Avar$ dénote la variance asymptotique.
- Sous l'hypothèse **TS.5'**, $a_t \equiv r_t u_t$ est non corrélé sériellement et donc la formule traditionnelle de $Var(\hat{\beta}_1)$ est valide. Par contre si **TS.5'** ne tient pas, $Avar(\hat{\beta}_1)$ doit tenir compte de la corrélation entre a_t et a_s $\forall t \neq s$.

Écart-types robustes

- Newey et West (1987) et Wooldridge (1989) ont montré que $Avar(\hat{\beta}_1)$ peut être estimé de la manière suivante.
- $i)$ Estimer par MCO $y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t$, $t = 1, \ldots, n$. $se(\hat{\beta}_1)$ dénote l'écart-type de $\hat{\beta}_1$ et $\hat{\sigma}$ l'écart-type de $\hat{u}_t$.
- $ii)$ Estimer la régression auxiliaire: $x_{t1} = \delta_0 + \delta_2 x_{t2} \ldots + \delta_k x_{tk} + r_t$.
- $iii)$ Calculer $\hat{a}_t = \hat{r}_t \hat{u}_t$, $\forall t = 1, \ldots, n$.
- $iv)$ Pour une valuer $g > 0$ donnée, calculer: $\hat{v} = \sum_{t=1}^{n} \hat{a}_t^2 + 2 \sum_{h=1}^{g} [1 - h/(g+1)] \left(\sum_{t=h+1}^{n} \hat{a}_t \hat{a}_{t-h} \right)$. $\rightarrow g$ contrôle la "quantité" de corrélation sérielle que nous permettons.

Écart-types robustes

- Ex: $g = 1$, $\hat{v} = \sum_{t=1}^{n} \hat{a}_t^2 + \sum_{t=2}^{n} \hat{a}_t \hat{a}_{t-1}$.
- $v)$ L'écart-type robuste à la corrélation sérielle de $\hat{\beta}_1$ est: $se^*(\hat{\beta}_1) = [se(\hat{\beta}_1)/\hat{\sigma}^2]\sqrt{\hat{v}}$.
- On peut montrer que cet estimateur est aussi robuste à toute forme d'hétéroskedasticité $\rightarrow$ cas plus général de ce qui est exposé au Chapitre 8.
- Comment choisir g ?
- La théorie nous dit que g doit croître avec n.

Choix de g

- Certains travaux ont suggéré pour:
  - des données annuelles $\rightarrow g = 1, 2$;
  - des données trimestrielles $\rightarrow g = 4, 8$;
  - des données mensuelles $\rightarrow g = 12, 24$.
- Newey et West (1987) recommandent de prendre la partie entière de $4(n/100)^{2/9} \rightarrow$ implémenté en Eviews.
- Exemple: Taux d'intérêt US obligataire à 3 mois.
  $ci3_t = \alpha_0 + \alpha_1 ci3_{t-1}$.

Hétéroscédasticité

- Pour beaucoup de séries temporelles, l'hypothèse **TS.4** ou **TS.4'** d'homoscédasticité est violée.
- Exemple: Rendements journaliers du NYSE
- $\rightarrow$ n'affecte pas le caractère non-biaisé ou convergent des MCO mais invalide l'inférence statistique traditionnelle.
- Il existe deux manières d'aborder le problème lié à l'hétéroscédasticité.
- $i)$ Corriger les écart-types pour effectuer de l'inférence correctement.
- $ii)$ Modéliser la dynamique présente dans la variance.

Écart-types robustes à la White (1980)

- White (1980) propose une méthode permettant de rendre les écart-types robustes à toute forme d'hétéroscédasticité.
- $\rightarrow$ offre une solution intéressante, pour autant que l'intérêt ne se porte que sur la modélisation de la moyenne conditionnelle.
- Eviews, ainsi que beaucoup d'autres logiciels économétriques, offre cette option.
- Il est possible de montrer que la formule de White (1980) est un cas particulier de la formule de Newey et West (1987) qui permet de tenir compte également d'une possible autocorrélation des résidus.
- Exemple: AR(1) sur NYSE

Tester la présence d'hétéroscédasticité

- Avant de modéliser la dynamique présente dans la variance, il est judicieux de tester la présence d'hétéroscédasticité afin d'avoir une meilleure idée de la spécification à adopter.
- Pour appliquer les tests présentés ci-dessous il faut supposer que les résidus u_t sont non corrélé sériellement$\rightarrow$ tester avant.
- Test de **Breusch-Pagan**:
  $u_t^2 = \delta_0 + \delta_1 x_{t1} + \ldots + \delta_k x_{tk} + v_t$; $H_0 : \delta_1 = \ldots + \delta_k = 0$.
- Pour utiliser un F-test, il faut que les écart-types des MCO soient valables et donct que v_t satisfasse **TS.4** ou **TS.4'** et **TS.5** ou **TS.5'**.
- Exemple:
  AR(1) sur NYSE: Estimer $\hat{u}_t^2 = \alpha_0 + \alpha_1 return_{t-1} + e_t$

Modèles ARCH

- Considérons un modèle simple: $y_t = \beta_0 + \beta_1 z_t + u_t$.
- Une caractéristique largement admise des séries financières à fréquence élevée est que la variance n'est pas constante au court du temps et qu'il existe des grappes de volatilité.
- $\rightarrow$ Si la variance en t est grande, elle le sera probablement demain et les jours qui suivent.
- $\rightarrow$ Si la variance en t est petite, elle le sera probablement demain et les jours qui suivent.
- Engle (1982) a proposé un modèle appelé ARCH: Autoregressive Conditional Heteroskedasticity.

Modèles ARCH

- La caractéristique du modèle ARCH(1) est que: $E(u_t^2|u_{t-1}, u_{t-2}, \ldots) = E(u_t^2|u_{t-1}) = \alpha_0 + \alpha_1 u_{t-1}^2$, alors que $E(u_t^2|u_{t-1}, u_{t-2}, \ldots) = 0$.
- Les u_t sont non corrélés sériellement alors que les u_t^2 le sont.
- Conditions de positivité: $E(u_t^2|u_{t-1}) > 0$, $\forall t \rightarrow \alpha_0 > 0$ et $\alpha_1 \geq 0$.
- Si $\alpha_1 = 0 \rightarrow$ homoscédasticité.
- $\rightarrow$ on peut tester la présence d'effets ARCH en estimant ce modèle (sur $\hat{u}_t$), voir une version plus étendue (plus de retards).
- On peut tester $\alpha_1, \ldots, \alpha_q = 0$ en utilisant un $LM-$test ou $F-$test.

[ps,default,slideColor,colorBG]prosper
[latin1]inputenc pstricks,pst-node,pst-text,pst-3d amsmath

Pooling de données “cross-section” ou méthodes simples de données de panel: Chapitre 13

Pooling

- Une série “cross-section” constitue bien souvent un ensemble de données relatives à des unités (individus, firmes, etc.) interrogées à un moment donné.
- Dans certains cas, l’enquête est répétée plusieurs fois donnant lieu à des échantillons différents, représentatifs de la population.
- La technique du pooling suppose que les différents échantillons sont chaque fois tirés aléatoirement de la population.
- $\rightarrow$ On n’observe pas nécessairement les mêmes unités.
- $\rightarrow$ On dispose de plusieurs échantillons indépendants.
- $\rightarrow$ Par conséquent, $Corr(u_t, u_s) = 0$, $\forall t \neq s$ et donc on peut donc empiler les enquêtes et effectuer une analyse MCO traditionnelle.

Panel

- Par contre, lorsqu'on observe la même unité au court du temps, on parle de données de panel ou longitudinales.
- Par conséquent, on ne peut pas supposer que les observations sont indépendantes.
- $\longrightarrow$ Un facteur non-observé (comme le QI) qui affecte le salaire d'un individu en 1995 va également affecter son salaire en 2000.
- $\longrightarrow$ Requiert des techniques particulières pour traiter ce problème.
- $\longrightarrow$ Empiler les échantillons et utiliser les MCO donne des estimateurs biaisés.

Technique de pooling

- Beaucoup d'enquêtes auprès des ménages sont répétées au court du temps.
- Vu que le taux d'attrition est souvent assez grand, on veille à interroger de nouvelles personnes pour accroître l'échantillon.
- $\rightarrow$ On a alors des échantillons indépendants.
- Pourquoi effectuer plusieurs enquêtes ?
- $\rightarrow$ Pour avoir plus d'observations et donc plus de précision dans l'estimation des paramètres et des écart-types.

Technique de pooling

- $\rightarrow$ Pour effectuer des tests nécessitant l'utilisation d'observations à différents moments du temps. Exemple: Pour tester l'efficacité d'une politique économique il faut des observations avant et après la mise en oeuvre de la politique.
- Etude de cas: La base de données FERTIL1.wf1 concerne l'étude de Sanders (1994) sur la fertilité aux USA.
- Fréquence: une enquête tous les 2 ans de 1972 à 1984 $\rightarrow$ 7 vagues.
- La question posée est Comment évolue la fertilité au court du temps ? $\rightarrow$ après avoir contrôlé pour des facteurs tels que éducation âge, race, région (à 16 ans), environnement (à 16 ans).

Variable dépendante: KIDS

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.742457	3.051767	-2.537040	0.0113
EDUC	-0.128427	0.018349	-6.999272	0.0000
AGE	0.532135	0.138386	3.845283	0.0001
AGESQ	-0.005804	0.001564	-3.710324	0.0002
BLACK	1.075658	0.173536	6.198484	0.0000
EAST	0.217324	0.132788	1.636626	0.1020
NORTHCEN	0.363114	0.120897	3.003501	0.0027
WEST	0.197603	0.166913	1.183867	0.2367
FARM	-0.052557	0.147190	-0.357072	0.7211
OTHRURAL	-0.162854	0.175442	-0.928248	0.3535
TOWN	0.084353	0.124531	0.677367	0.4983
SMCITY	0.211879	0.160296	1.321799	0.1865
Y74	0.268183	0.172716	1.552737	0.1208
Y76	-0.097379	0.179046	-0.543881	0.5866
Y78	-0.068666	0.181684	-0.377945	0.7055
Y80	-0.071305	0.182771	-0.390136	0.6965
Y82	-0.522484	0.172436	-3.030016	0.0025
Y84	-0.545166	0.174516	-3.123871	0.0018

R-squared	0.129512	Mean dependent var	2.743136
Adjusted R-squared	0.116192	S.D. dependent var	1.653899
S.E. of regression	1.554847	Akaike info criterion	3.736447
Sum squared resid	2685.898	Schwarz criterion	3.816627
Log likelihood	-2091.224	F-statistic	9.723282
Durbin-Watson stat	2.010694	Prob(F-statistic)	0.000000

Données de Panel à 2 périodes

- Supposons maintenant que nous disposons des données individuelles pour lesquelles un individu est observé en $t = 1$ et $t = 2$.
- Exemple: la base de donnée CRIME2 comprend des données sur les taux de criminalité et de chômage pour 46 villes américaines en 1982 et 1987.
- Estimation pour l'année 1987 : Résultat.
- Comment expliquer ce résultat ? $\leftarrow$ Si chômage augmente, le crime diminue.
- Variables omises observables ? On pourrait contrôler pour des facteurs tels que (la distribution de) l'âge, éducation, etc.

Données de Panel à 2 périodes

- Variables omises non-observables ? On peut imaginer que certains de ces facteurs sont constants au court du temps et certains varient au court du temps.
- $y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + a_i + u_{it}, t = 1, 2.$
- $i \rightarrow$ unité et $t \rightarrow temps$.
- $d2_t = 0$ si $t = 1$ et $d2_t = 1$ si $t = 2 \rightarrow$ intercept différent pour $t = 1$ ou 2.
- a_i capture tous les facteur non-observé affectant $y_{it} \rightarrow$ effet non-observé ou fixe
- Ce modèle s'appelle donc modèle à effet non-observé ou à effet fixe.
- u_{it} est le terme d'erreur idiosyncratique ou variant dans le temps.

Exemple

- $crmrte_{it} = \beta_0 + \delta_0 d87_t + \beta_1 unem_{it} + a_i + u_{it}, t = 1, 2.$
- a_i reprend tous les autres facteurs affectant le taux de criminalité qui ne varient pas entre $t = 1$ et 2.
- $\rightarrow$ caractéristiques géographiques (localisations aux USA).
- $\rightarrow$ caractéristiques démographiques (age, race, éducation, etc.): a vérifier.
- $\rightarrow$ certaines villes peuvent avoir leurs propres comptabiliser les crimes.
- Comment estimer β_1 ?
- Pooling ?

Pooling

- $y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + v_{it}, t = 1, 2.$
- $\rightarrow v_{it} = a_i + u_{it}$
- Pour estimer ce modèle par MCO il faut supposer que a_i est non corrélé avec x_{it} sinon v_{it} (*erreur composé*) serait corrélé avec x_{it}.
- Si ce n'est pas le cas les MCO donne lieu à un biais d'hétérogénéité non observée.
- $\rightarrow$ Biais dû à l'omission d'une variable constante au court du temps.
- Exemple: CRIME. $\rightarrow$ 92 observations: 46 villes et 2 périodes.

Panel

- Si a_i est corrélé avec x_{it}, la technique du pooling est donc inappropriée.
- Solution simple: estimer un modèle en différence première.
- $$\begin{aligned} y_{i1} &= \beta_0 + \beta_1 x_{i1} + a_i + u_{i1} \ (t = 1) \\ y_{i2} &= (\beta_0 + \delta_0) + \beta_1 x_{i2} + a_i + u_{i2} \ (t = 2) \end{aligned}$$
- Par conséquent,

$$\begin{aligned} (y_{i2} - y_{i1}) &= \delta_0 + \beta_1 \Delta(x_{i2} - x_{i1}) + (u_{i2} - u_{i1}) \\ \Delta y_i &= \delta_0 + \beta_1 \Delta x_i + \Delta u_i \end{aligned}$$

Panel

- $\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i$ peut être estimé par MCO si les hypothèses traditionnelles sont validées.
- En particulier, il faut que $Corr(\Delta u_i, \Delta x_i) = 0$, ce qui est naturellement vrai si $Corr(u_{it}, x_{it}) = 0$, $\forall t = 1, 2$.
- Dans notre exemple $Corr(\Delta u_i, \Delta unem_i) = 0$?
- Faux si par exemple si l'effort d'application de la loi ($\in u_{it}$) augmente plus dans des villes où le taux de chômage diminue $Corr(u_{it}, unem_{it}) < 0 \rightarrow$ biais des MCO.
- Important: on ne peut estimer par cette méthode l'effet de variables constantes au court du temps (z_i). Exemple: Distance par rapport à la capitale $\rightarrow \Delta z_i = 0$.
- Exemple: CRIME.