ANNEXE

B Guide de prise en main de Gretl

Gretl (*Gnu Regression, Econometrics and Time-series Library*) est un logiciel libre et gratuit d'économétrie, distribué sous licence GPL pour Windows, macOS et Linux (Cottrell & Lucchetti, 2023). Il réunit une interface à menus, qui permet d'estimer un modèle en quelques clics, et un langage de script, *hansl*, qui rend chaque analyse reproductible. Pour les étudiants et les chercheurs des universités africaines, c'est un outil de premier choix : il ne coûte rien, il est léger, son interface existe en français et il couvre l'essentiel des méthodes de ce manuel.

Objectifs d'apprentissage

À l'issue de cette annexe, le lecteur sera capable de :

- installer Gretl, repérer les éléments de son interface et passer d'une action de menu à la commande *hansl* équivalente ;
- importer un fichier CSV, déclarer la structure des données (coupe transversale, série chronologique, panel) et transformer des variables ;
- estimer un modèle par les moindres carrés ordinaires, lire la sortie, tester des restrictions et diagnostiquer les résidus ;
- mettre en œuvre les modèles des chapitres 6 à 15 : variables instrumentales, panel, probit, Tobit, Heckit, comptage, ARIMA, GARCH, VAR, VECM, régression quantile et différences-en-différences ;
- écrire des scripts réutilisables (boucles, fonctions) et exporter tableaux, équations et graphiques.

Cette annexe est un guide pratique : elle ne reprend pas la théorie des chapitres 5 à 15, mais indique pour chaque méthode le chapitre concerné. Les sections suivent l'ordre habituel d'une étude : installer le logiciel, importer et préparer les données, estimer, tester, puis automatiser et exporter.

Conventions de lecture

Les menus s'écrivent `Fichier > Ouvrir les données`. Chaque script *hansl* figure dans un encadré numéroté « Code » ; la sortie réelle qu'il produit est donnée juste après, suivie d'un commentaire. Pour la mise en page, les sorties sont abrégées : les messages de confirmation (« Série lprice générée ») sont omis, et une coupure est marquée `[...]`. Les nombres s'affichent avec la virgule décimale parce que Gretl suit la langue de l'interface, mais dans un script le séparateur décimal est toujours le point (`0.025`).

B.1 Installation et configuration

Le site officiel, https://gretl.sourceforge.net, propose un installateur pour Windows, un paquet pour macOS et des paquets pour les principales distributions GNU/Linux. Sous Debian ou Ubuntu, la commande `sudo apt-get install gretl gretl-data` installe le programme et le paquet `gretl-data`, qui contient les jeux de données d'exemple utilisés dans cette annexe. Les installateurs Windows et macOS les installent normalement avec le programme.

Après l'installation, deux réglages méritent un coup d'œil. Le menu `Outils > Préférences > Général...` regroupe les options globales (langue de l'interface notamment), et `Outils > Préférences > Échelle des polices...` permet d'agrandir les caractères pour un projecteur. Le menu `Fichier > Dossier de travail...` fixe le dossier où Gretl lit les fichiers désignés sans chemin complet et où il écrit ceux que créent les scripts (`store`, `tabprint`, `gnuplot --output`).

Version et reproductibilité

Les scripts de cette annexe ont été exécutés avec Gretl 2023c, en mode ligne de commande (`gretlcli -b script.inp`). Les commandes de base (`ols`, `smpl`, `series`) sont stables depuis longtemps, mais les options et les fonctions évoluent d'une version à l'autre. Devant un message « commande inconnue » ou « symbole non défini », vérifiez d'abord le numéro de version, que `gretlcli` affiche au démarrage, puis consultez l'aide d'une commande en tapant `help ols` dans la console.

Pour que chaque lecteur retrouve exactement les sorties reproduites ici, les exemples s'appuient sur des jeux de données fournis avec Gretl (tableau B.1). La série africaine du manuel (taux de change du franc congolais, annexe C du livre) s'utilise de la même façon ; un script qui lui est consacré figure à la section B.4.

Table B.1. Jeux de données livrés avec Gretl et utilisés dans cette annexe

Fichier	Contenu (source)	Obs.	Section
`data4-1`	Prix de vente et caractéristiques de 14 maisons de University City, San Diego, 1990 (Ramanathan, 2002)	14	B.4, B.6
`data7-11`	59 maisons vendues en 1984 à La Jolla et University City (Ramanathan, 2002)	59	B.6, B.12
`greene5_1`	Macroéconomie américaine trimestrielle, 1950:1 à 2000:4 (Greene, 2003)	204	B.4, B.9
`grunfeld`	10 entreprises américaines, 1935 à 1954 (Kleiber & Zeileis, 2010)	200	B.7
`mroz87`	Femmes mariées américaines, 1975 (Mroz, 1987)	753	B.7, B.8
`rac3d`	Consultations médicales d'Australiens (Cameron & Trivedi, 1998)	5 190	B.8
`b-g`	Rendements quotidiens DM/livre sterling, 1984 à 1991 (Bollerslev & Ghysels, 1996)	1 974	B.9
`denmark`	Macroéconomie danoise trimestrielle, 1974:1 à 1987:3 (Johansen, 1995)	55	B.10
`engel`	Dépense alimentaire et revenu de ménages belges (Koenker & Bassett, 1982)	235	B.11
`card_krueger94`	Restaurants du New Jersey et de Pennsylvanie (Card & Krueger, 1994)	410	B.11

B.2 L'interface graphique

B.2.1 La fenêtre principale

Au lancement, Gretl ouvre une fenêtre vide ; dès qu'un jeu de données est chargé, elle se présente comme sur la figure B.1. Ce schéma reproduit fidèlement la disposition de la fenêtre dans la version française : la barre de menus (A), le nom du jeu de données ouvert à gauche et le dossier de travail à droite (B), la liste des variables (C), la barre d'outils (D) et la barre d'état (E), qui indique la structure des données.

La barre de menus organise les fonctions par thème (tableau B.2). L'ordre des menus suit celui d'une étude : on ouvre un fichier (`Fichier`), on regarde les données (`Afficher`), on les transforme (`Ajouter`, `Échantillon`), puis on estime (`Modèle`).

La barre d'outils du bas de la fenêtre donne un accès rapide aux actions les plus fréquentes. Les infobulles, qui s'affichent au survol de chaque bouton, sont reprises dans le tableau B.3.

B.2.2 Estimer un modèle par les menus

Le menu `Modèle > Moindres Carrés Ordinaires, MCO` ouvre la boîte de dialogue « MCO ». À gauche figure la liste des variables ; à droite, le champ « Variable dépendante » et le champ « Régresseurs », où la constante `const` est déjà inscrite. Une case « Écarts-types robustes » permet de corriger l'hétéroscédasticité. Pour régresser `price` sur `sqft`, on sélectionne `price` et on l'envoie dans le champ « Variable dépendante » avec la flèche correspondante, on fait de même avec `sqft` pour les régresseurs, puis on clique sur le bouton « OK ». Les boutons « Help », « Clear » et « Cancel » ouvrent l'aide, vident les champs ou ferment la boîte.

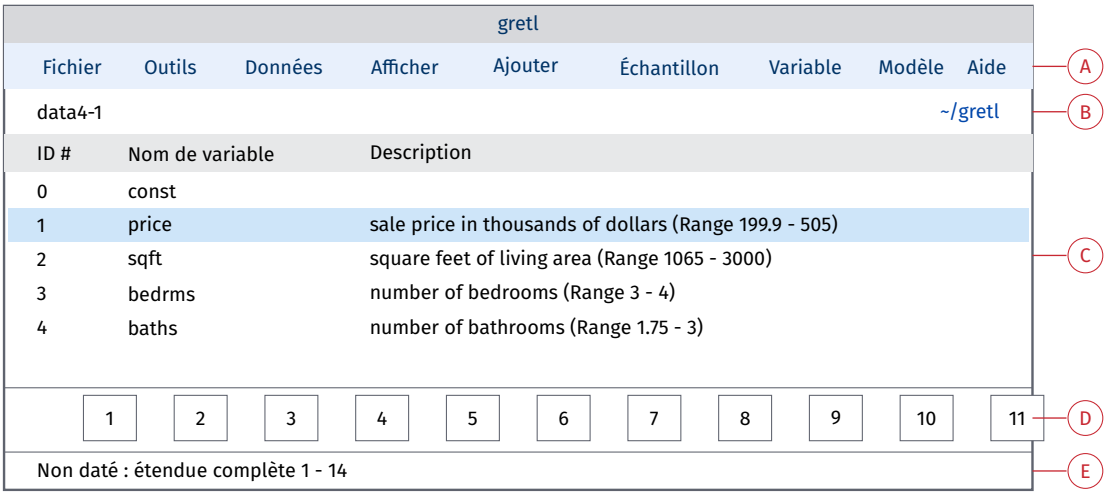


FIGURE B.1. Schéma de la fenêtre principale de GRETL (version française) après l'ouverture de `data4-1`. A : barre de menus ; B : nom du jeu de données (à gauche) et dossier de travail (à droite) ; C : liste des variables, avec les colonnes « ID # », « Nom de variable » et « Description » ; D : barre d'outils de onze boutons, numérotés de gauche à droite comme dans le tableau B.3 ; E : barre d'état, qui rappelle la structure des données.

TABLE B.2. Les menus de la fenêtre principale et les commandes *hansl* correspondantes

Menu	Contenu principal	Commandes
`Fichier`	Ouvrir, ajouter, enregistrer et exporter des données ; dossier de travail ; scripts (`Code`) ; session ; paquetages de fonctions	`open`, `store`
`Outils`	Préférences, tables statistiques, calcul de probabilités critiques, console, journal des commandes	`pvalue()`, `critical()`
`Données`	Afficher et modifier les valeurs, structure du jeu de données, listes, tri	`setobs`, `list`
`Afficher`	Statistiques descriptives, matrice de corrélation, tableau de contingence, composantes principales, graphiques	`summary`, `corr`, `gnuplot`
`Ajouter`	Logarithmes, carrés, retards, différences, indicatrices, tendance, nouvelle variable	`series`, `logs`, `lags`, `diff`
`Échantillon`	Définir, restreindre ou restaurer l'échantillon	`smpl`
`Variable`	Statistiques et graphique d'une série, tests de racine unitaire, corrélogramme, filtrage	`adf`, `kpss`, `corrgm`
`Modèle`	Tous les estimateurs : MCO, variables instrumentales, panel, modèles de choix, séries temporelles, estimation robuste	`ols`, `tsls`, `panel`, `arima`
`Aide`	Documentation du logiciel et des commandes	`help`

Table B.3. Les onze boutons de la barre d'outils, de gauche à droite

N°	Infobulle	Effet
1	Lancer le calculateur	ouvre une calculatrice
2	Nouveau code	ouvre un éditeur de script vide
3	Ouvrir la console gretl	équivaut à `Outils > Console Gretl`
4	Vue en icône de la session	affiche les objets de la session (modèles, graphiques)
5	Paquetage de functions	gère les paquetages de fonctions
6	Index des commandes	liste des commandes avec leur aide
7	Chercher les séries	recherche une variable dans le jeu de données
8	Graphique de dispersion X-Y	nuage de points de deux variables
9	Modèle MCO	ouvre la boîte de dialogue des MCO
10	Banque de données	accès aux bases de données
11	Ouvrir un jeu de données de gretl	ouvre un fichier d'exemple

Le résultat s'affiche dans une fenêtre propre au modèle, dont les menus sont `Fichier`, `Édition`, `Tests`, `Enregistrer`, `Graphiques`, `Analyse` et `LaTeX`. Ce sont eux qui servent à tester, à prévoir et à exporter (tableau B.4). Chaque entrée correspond à une commande que le script peut reproduire, d'où la règle d'or : *travailler d'abord par les menus pour explorer, puis recopier la commande dans un script pour garder une trace*.

Table B.4. Équivalences entre les menus de la fenêtre d'un modèle et les commandes

Menu de la fenêtre du modèle	Commande ou accesseur
`Tests > Omettre les variables`	`omit x1 x2`
`Tests > Restrictions linéaires`	`restrict ... end restrict`
`Tests > Test du RESET de Ramsey`	`reset`
`Tests > Hétéroscédasticité` (dont White)	`modtest --breusch-pagan`, `modtest --white`
`Tests > Normalité des résidus`	`modtest --normality`
`Enregistrer > Résidus`	`series e = $uhat`
`Enregistrer > Série ajustée`	`series yh = $yhat`
`Analyse > Prévisions...`	`fcast`
`Analyse > Colinéarité`	`vif`
`Graphiques > Corrélogramme des résidus`	`corrgm e 12`
`LaTeX`	`tabprint`, `eqnprint`

B.2.3 Console, éditeur de code et journal

Trois outils complètent les menus. La *console* (`Outils > Console Gretl`) exécute une commande à la fois, comme dans une calculatrice : on y tape `help ols` ou `summary price`. L'*éditeur de code* (`Fichier > Code > Nouveau code`) permet d'écrire un script entier, de l'enregistrer au format `.inp` et de l'exécuter d'un coup. Enfin, le *journal des commandes* (`Outils > Journal des commandes`) conserve toutes les commandes générées par vos clics.

C'est le moyen le plus rapide d'apprendre la syntaxe : on exécute une action par menu, puis on lit la commande correspondante dans le journal.

B.2.4 Un flux de travail type

La figure B.2 résume la démarche suivie dans le reste de l'annexe. Chaque étape a son menu et sa commande ; le script `.inp` qui les enchaîne est la mémoire de l'étude.

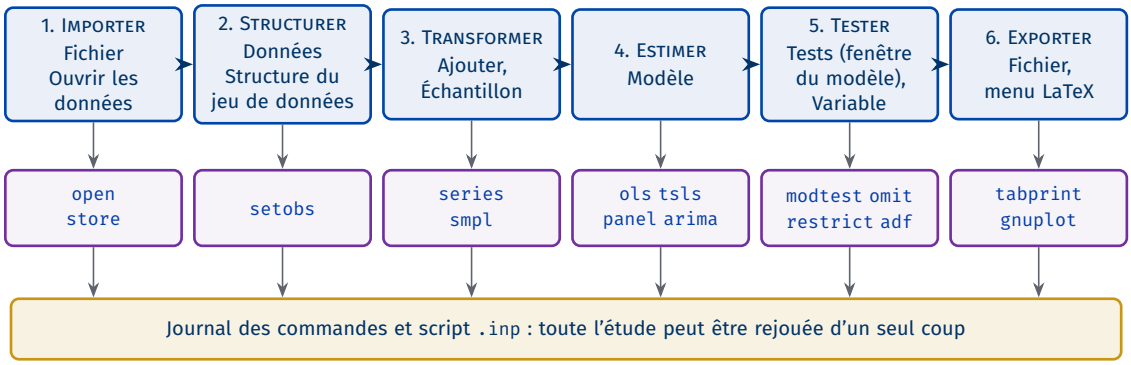


Figure B.2. Flux de travail dans Gretl. Les rectangles bleus sont les six étapes d'une étude avec le menu à utiliser ; les rectangles violets donnent les commandes *hansl* correspondantes ; le bandeau du bas rappelle que toutes les étapes se retrouvent dans le journal des commandes et dans le script.

B.3 Le langage hansl en bref

Une commande *hansl* tient sur une ligne. Elle commence par le nom de la commande, suivi des variables ou des objets concernés, puis des options introduites par deux tirets. La commande `ols price const sqft --robust` se lit donc ainsi : `ols` (la méthode), `price` (la variable dépendante), `const sqft` (les régresseurs) et `--robust` (l'option qui demande des écarts-types robustes). Le dièse `#` ouvre un commentaire qui s'étend jusqu'à la fin de la ligne. Gretl distingue les majuscules des minuscules : `Price` et `price` sont deux noms différents.

Cinq règles de syntaxe à retenir

- La constante s'écrit `const` ; elle doit figurer explicitement dans la liste des régresseurs, sinon la régression est sans constante et le R^2 n'est plus comparable.
- Le décalage d'une série se note entre parenthèses : `y(-1)` est la valeur retardée d'une période.
- Un accesseur commence par `$` et renvoie un résultat de la dernière estimation : `$coeff`, `$stderr`, `$rsq`, `$uhat`.
- L'option `--quiet` supprime l'affichage d'une commande, ce qui est utile quand on veut seulement récupérer ses résultats.
- `help commande` affiche la syntaxe et les options d'une commande.

Le langage manipule cinq types d'objets : des scalaires, des séries (une valeur par observation), des matrices, des listes de séries et des chaînes de caractères. Le code B.1 les illustre sur les 14 maisons de `data4-1`.

Code B.1. Les types d'objets de hansl

```
open data4-1 --quiet
scalar n = $nobs                      # scalaire : un nombre
series ppm = price / sqft             # série : une valeur par observation
matrix A = {price, sqft}              # matrice 14 x 2
list X = const sqft bedrms baths      # liste : un groupe de séries
string ville = "San Diego"            # chaîne de caractères
bundle res = _(n = n, ville = ville) # bundle : objets nommés
printf "%s : %d maisons, prix moyen %.1f\n", res.ville, res.n, mean(price)
summary ppm --simple
```

```
San Diego : 14 maisons, prix moyen 317,5

Statistiques descriptives, utilisant les observations 1 - 14
pour la variable « ppm » (14 observations valides)

  Moyenne            0,16842
  Minimum            0,13833
  Maximum            0,19519
  Écart-type        0,018663
  Obs. manq.               0
```

Le scalaire `n` compte les observations ; la série `ppm` est le prix au pied carré de chaque maison, en milliers de dollars ; la matrice `A` range deux séries côte à côte ; la liste `X` regroupe des séries pour les passer ensemble à une commande ; la chaîne `ville` contient du texte ; le *bundle* `res` rassemble des objets de natures différentes sous un seul nom, ce qui est pratique pour renvoyer plusieurs résultats d'une fonction. Le prix moyen est de 317,5 milliers de dollars, soit environ 0,168 millier de dollars (168 dollars) le pied carré, avec une variation modeste d'une maison à l'autre (écart-type 0,019).

Les matrices permettent de retrouver à la main le résultat d'une régression. Le code B.2 calcule l'estimateur des MCO $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ et la somme des carrés des résidus $\hat{\mathbf{e}}'\hat{\mathbf{e}}$.

Code B.2. Les MCO calculés à la main avec des matrices

```
open data4-1 --quiet
matrix y = {price}              # vecteur des prix (14 x 1)
matrix X = {const, sqft}        # matrice des régresseurs (14 x 2)
matrix b = inv(X'X) * X'y       # b = (X'X)^(-1) X'y
matrix e = y - X*b              # résidus
printf "b0 = %.4f, b1 = %.5f, SCR = %.2f\n", b[1], b[2], e'e
```

```
b0 = 52,3509, b1 = 0,13875, SCR = 18273,57
```

On obtient $\hat{\beta}_0 = 52{,}35$, $\hat{\beta}_1 = 0{,}1388$ et une somme des carrés résidus de 18 273,57, exactement les valeurs de la commande `ols` de la section B.6 (code B.11). Cette vérification rassure sur ce que fait la commande `ols` et prépare l'écriture d'estimateurs sur mesure.

B.4 Importer, structurer et transformer les données

B.4.1 Ouvrir un jeu de données

Gretl lit son format natif `.gdt` ainsi que les fichiers CSV, Excel, Stata, SPSS et d'autres formats. Dans l'interface, on passe par `Fichier > Ouvrir les données > fichier utilisateur...` pour un fichier personnel et par `Fichier > Ouvrir les données > fichier d'exemple...` pour les jeux livrés avec le logiciel. Dans un script, la commande `open` devine le format à partir de l'extension (code B.3).

Code B.3. Ouvrir un jeu de données et décrire deux séries

```
open data4-1 --quiet       # jeu d'exemple : maisons de San Diego
varlist                    # liste des séries disponibles
summary price sqft         # statistiques descriptives
```

```
Liste des 5 variables :
  0) const     1) price     2) sqft      3) bedrms    4) baths

                  Moyenne        Médiane        Minimum        Maximum
price              317,49         291,50         199,90         505,00
sqft               1910,9         1835,0         1065,0         3000,0

                Ec. type           C.V.      Asymétrie   Ex. kurtosis
price              88,498        0,27874        0,65346       -0,52983
sqft               577,76        0,30234        0,48526       -0,67212

                      IQ     Obs. manq.
price              154,50              0
sqft               832,75              0
```

`varlist` dresse la liste des cinq séries (la constante porte le numéro 0). `summary` calcule la moyenne, la médiane, les extrêmes, l'écart-type, le coefficient de variation, l'asymétrie, le kurtosis en excès et l'écart interquartile. Le prix moyen des maisons est de 317,5 milliers de dollars pour une médiane de 291,5 : la moyenne dépasse la médiane, signe d'une distribution étirée vers les prix élevés, ce que confirme l'asymétrie positive (0,65).

B.4.2 Enregistrer et relire un fichier CSV

Le CSV est le format d'échange le plus simple entre Gretl, Excel, R et Python. À l'ouverture d'un fichier CSV, Gretl détecte le séparateur et la présence d'une ligne de noms de variables ; si les noms sont mal reconnus, vérifiez que le fichier est enregistré en UTF-8 et que le séparateur décimal correspond à celui de vos paramètres régionaux. Le code B.4 écrit puis relit un fichier à la française (colonnes séparées par un point-virgule, virgule décimale).

Code B.4. Enregistrer et relire un fichier CSV

```
open data4-1 --quiet
set csv_delim semicolon              # colonnes séparées par « ; »
store "maisons.csv" --decimal-comma  # CSV à la française
open "maisons.csv" --quiet           # relecture : format détecté seul
print price sqft --byobs
```

```
maisons.csv écrit

          price         sqft

 1        199,9         1065
 2        228,0         1254
 3        235,0         1300
 4        285,0         1577
 5        239,0         1600
 6        293,0         1750
 7        285,0         1800
 8        365,0         1870
 9        295,0         1935
10        290,0         1948
11        385,0         2254
12        505,0         2600
13        425,0         2800
14        415,0         3000
```

La commande `store` écrit dans le dossier de travail, et l'extension `.csv` suffit à choisir le format (il n'existe pas d'option `--csv`). L'option `--decimal-comma` produit la virgule décimale, et `set csv_delim` choisit le séparateur de colonnes. À la relecture, Gretl reconnaît seul ce format : les valeurs imprimées sont celles de `data4-1`, que l'on retrouve dans le tableau.

B.4.3 Déclarer la structure des données

Avant toute estimation sur séries temporelles ou sur panel, il faut dire à Gretl comment les lignes sont ordonnées. Dans l'interface, c'est le menu `Données > Structure du jeu de données...`, qui guide pas à pas ; dans un script, c'est la commande `setobs`. La figure B.3 aide à choisir la bonne forme.

Le code B.5 déclare une série trimestrielle ; on exporte d'abord le PIB et le chômage américains vers un CSV sans colonne de dates, ce qui reproduit le cas courant où le fichier arrive sans calendrier.

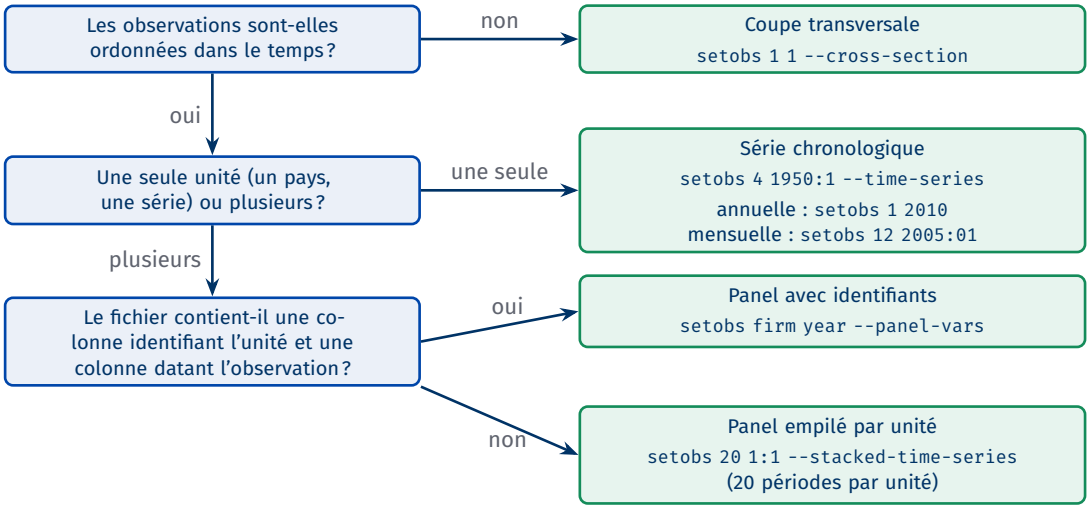


FIGURE B.3. Choisir la forme de `setobs`. Les rectangles bleus sont des questions sur le fichier, les rectangles verts la commande à exécuter. Pour un panel empilé, le premier argument est le nombre de périodes par unité (`--stacked-time-series`) ou le nombre d'unités (`--stacked-cross-section`).

Code B.5. Déclarer une série trimestrielle

```
open greene5_1 --quiet
store "pib_us.csv" realgdp unemp --omit-obs  # CSV sans colonne de dates
open "pib_us.csv" --quiet            # série non datée : 204 lignes
setobs 4 1950:1 --time-series        # trimestrielle, début en 1950:1
smpl 1990:1 1990:4
print realgdp --byobs
```

```
pib_us.csv écrit
Plage des données complète: 1950:1 - 2000:4 (n = 204)

Plage des données complète: 1950:1 - 2000:4 (n = 204)
Échantillon actuel: 1990:1 - 1990:4 (n = 4)

            realgdp

1990:1       6716,3
1990:2       6731,7
1990:3       6719,4
1990:4       6664,2
```

À la relecture, le fichier est traité comme des données sans date, avec 204 lignes. La commande `setobs 4 1950:1 --time-series` indique une fréquence trimestrielle (4) et un début au premier trimestre 1950, ce qui fixe l'étendue 1950:1 à 2000:4. La commande `smpl` permet ensuite de désigner les quatre trimestres de 1990 : le PIB réel passe de 6 716,3 à 6 664,2 milliards de dollars, soit un recul au dernier trimestre.

Pour un panel, le cas le plus simple est celui d'un fichier « long » avec une colonne pour l'unité et une colonne pour la date (code B.6).

Code B.6. Déclarer un panel à partir de variables d'identification

```
open grunfeld --quiet
store "grunfeld.csv" firm year invest value kstock --omit-obs
open "grunfeld.csv" --quiet          # format long : une ligne par couple
setobs firm year --panel-vars        # identifiants individu et date
```

```
grunfeld.csv écrit
Plage des données complète: 1:01 - 10:20 (n = 200)
```

L'option `--panel-vars` indique les deux colonnes qui identifient l'unité (`firm`) et la date (`year`). Gretl trie alors les données et affiche l'étendue `1:01 - 10:20` : 10 entreprises observées sur 20 années, soit 200 observations.

Notation des dates dans Gretl

Pour des données trimestrielles, la date s'écrit `AAAA:T` (`1950:1` désigne le premier trimestre de 1950) ; pour des données mensuelles, `AAAA:MM` (`2005:01`) ; pour des données annuelles, l'année seule. Dans un panel, une observation se note `unité:période`, par exemple `3:07` pour la septième période de la troisième unité.

La série mensuelle du taux de change du franc congolais (fichier BCC-01 de l'annexe C du livre) se manipule de la même manière. Le code B.7 la charge, calcule la dépréciation mensuelle et teste la racine unitaire. Ce fichier n'est pas livré avec Gretl : le script n'a donc été exécuté que pour vérifier sa syntaxe sur un fichier de même structure, et aucune sortie n'est reproduite.

Code B.7. Script pour la série BCC-01 de l'Annexe E (non exécuté ici)

```
open "bcc_taux_change_cdf_usd_2005_2023.csv" --quiet
series depr = 100 * ldiff(cdf_usd)   # dépréciation mensuelle (%)
adf 4 cdf_usd --ct                   # racine unitaire dans le niveau
```

La colonne `date` au format `AAAA-MM` est reconnue comme calendrier mensuel et n'apparaît plus comme variable. La fonction `ldiff` donne la différence des logarithmes, et le facteur 100 exprime la dépréciation mensuelle en pourcentage. Pour l'estimation d'un modèle de panel, voir la section B.7, qui s'appuie sur un jeu livré avec Gretl.

B.4.4 Créer et transformer des variables

Les transformations s'obtiennent par le menu `Ajouter` (logarithme, différences, retards, indicatrices) ou par la commande `series`. Le code B.8 applique les plus courantes au PIB réel américain.

Code B.8. Transformations usuelles de variables

```
open greene5_1 --quiet
series lpib   = log(realgdp)         # logarithme népérien
series croiss = 100 * diff(lpib)     # croissance trimestrielle (%)
series croiss2 = 100 * ldiff(realgdp)  # même résultat, en une commande
series lpib_1 = lpib(-1)             # retard d'ordre 1
series post80 = ($obsmajor >= 1980)  # indicatrice : 1 à partir de 1980
series cycle  = hpfilt(lpib)         # cycle de Hodrick-Prescott
series tendance = lpib - cycle       # tendance
smpl 1979:3 1980:2
print lpib croiss croiss2 lpib_1 post80 --byobs
```

```
Plage des données complète: 1950:1 - 2000:4 (n = 204)
Échantillon actuel: 1979:3 - 1980:2 (n = 4)

               lpib       croiss      croiss2       lpib_1       post80

1979:3     8,502323     0,708934     0,708934     8,495234            0
1979:4     8,505647     0,332361     0,332361     8,502323            0
1980:1     8,508939     0,329243     0,329243     8,505647            1
1980:2     8,488341    -2,059828    -2,059828     8,508939            1
```

La croissance trimestrielle `croiss` s'obtient de deux façons équivalentes, par `diff` sur le logarithme ou par `ldiff` sur la série brute ; les colonnes sont identiques. Au deuxième trimestre de 1980, l'économie américaine se contracte de 2,06 %, ce qui illustre le caractère « volatil » du taux de croissance par rapport au niveau du PIB. L'indicatrice `post80` vaut 0 jusqu'en 1979:4 et 1 à partir de 1980:1 (`$obsmajor` est la partie principale de la date, ici l'année). Enfin, `hpfilt` renvoie la composante cyclique du filtre de Hodrick-Prescott (avec le paramètre de lissage 1 600, valeur usuelle pour des données trimestrielles) et la tendance s'en déduit par différence.

B.4.5 Restreindre l'échantillon

La commande `smpl` (menu `Échantillon > Définir l'échantillon...` ou `Échantillon > Restreindre, basé sur le critère...`) limite l'analyse à une période ou à un sous-groupe. Il faut toujours restaurer l'échantillon complet avant de passer à autre chose (code B.9).

Code B.9. Définir et restaurer l'échantillon

```
open greene5_1 --quiet
smpl 1980:1 2000:4          # sous-période
smpl --full                 # retour à l'échantillon complet
smpl unemp > 7 --restrict   # trimestres où le chômage dépasse 7 %
summary unemp --simple
smpl --full
```

```
Plage des données complète: 1950:1 - 2000:4 (n = 204)
Échantillon actuel: 1980:1 - 2000:4 (n = 84)

Plage des données complète: 1950:1 - 2000:4 (n = 204)

Échantillon complet : 204 observations
Échantillon actuel : 42 observations

Statistiques descriptives, utilisant les observations 1 - 42
pour la variable « unemp » (42 observations valides)

  Moyenne             7,9595
  Minimum             7,1000
  Maximum             10,700
  Écart-type         0,96905
  Obs. manq.               0

Plage des données complète: 1950:1 - 2000:4 (n = 204)
```

La première commande retient 84 trimestres (1980 à 2000), la troisième seulement ceux où le chômage dépasse 7 % : 42 trimestres, avec un taux moyen de 7,96 %. Un oubli de `smpl --full` est une erreur classique : les estimations suivantes porteraient sans bruit sur le sous-échantillon.

B.5 Statistiques descriptives et graphiques

Les entrées `Statistiques descriptives` et `Matrice de corrélation` du menu `Afficher` correspondent aux commandes `summary` et `corr`. Pour un graphique, le même menu propose `Graphique des variables spécifiées`, et le menu `Variable` propose `Graphique de séries temporelles` : chacune ouvre une fenêtre dont le menu permet d'enregistrer l'image. Dans un script, la commande `gnuplot` accepte l'option `--output` pour écrire directement un fichier (code B.10).

Code B.10. Corrélations et nuage de points avec droite ajustée

```
open data4-1 --quiet
corr price sqft bedrms baths
gnuplot price sqft --fit=linear --output="prix_surface.pdf"
```

```
Coefficients de corrélation, utilisant les observations 1 - 14
Two-tailed critical values for n = 14: 5% 0,5324, 1% 0,6614

         price          sqft        bedrms         baths
        1,0000        0,9058        0,3156        0,6696  price
                      1,0000        0,4647        0,7873  sqft
                                    1,0000        0,5323  bedrms
                                                  1,0000  baths

~/gretl/prix_surface.pdf écrit
```

Le prix est très corrélé à la surface (0,906) et nettement moins au nombre de chambres (0,316) ; la surface est elle-même corrélée au nombre de salles de bains (0,787), ce qui annonce un problème de colinéarité dans la régression multiple. Avec 14 observations, un coefficient doit dépasser 0,532 en valeur absolue pour être significatif à 5 % (valeur critique affichée par Gretl). Le fichier `prix_surface.pdf` contient le nuage des 14 maisons, la droite des moindres carrés et, dans la légende, son équation $\hat{y} = 52{,}4 + 0{,}139\,x$: un graphique de cette nature est reproduit, avec les mêmes données réelles, à la figure B.6.

B.6 Régression par les moindres carrés ordinaires

La régression par les moindres carrés ordinaires (chapitre 5) est la commande de base : `ols` suivie de la variable dépendante puis des régresseurs, la constante étant écrite `const`.

B.6.1 La régression simple

Code B.11. Régression simple

```
open data4-1 --quiet
ols price const sqft       # prix (milliers de $) sur la surface
```

```
Modèle 1: MCO, utilisant les observations 1-14
Variable dépendante: price

             coefficient    éc. type    t de Student   p. critique
  ------------------------------------------------------------------
  const       52,3509      37,2855         1,404        0,1857
  sqft         0,138750     0,0187329      7,407        8,20e-06   ***

Moyenne var. dép.     317,4929   Éc. type var. dép.    88,49816
Somme carrés résidus  18273,57   Éc. type régression   39,02304
R2                    0,820522   R2 ajusté             0,805565
F(1, 12)              54,86051   P. critique (F)       8,20e-06
Log de vraisemblance -70,08421   Critère d'Akaike      144,1684
Critère de Schwarz    145,4465   Hannan-Quinn          144,0501
```

Chaque pied carré supplémentaire de surface habitable augmente le prix de vente prévu de 0,1388 millier de dollars, soit 138,75 dollars (statistique t de 7,41, probabilité critique inférieure à 10^{-5}) ; la constante (52,35) n'est pas significativement différente de zéro (probabilité critique 0,186). Le R^2 de 0,8205 indique que la surface explique 82 % de la variance des prix, et la statistique $F(1, 12) = 54{,}86$ rejette la nullité de la pente.

Lire un tableau de résultats des MCO

Le tableau donne, pour chaque variable, le coefficient estimé, son écart-type (« éc. type »), la statistique t de Student et la probabilité critique bilatérale ; les étoiles marquent les seuils de 10 % (*), 5 % (**) et 1 % (***). Au-dessous figurent la moyenne et l'écart-type de la variable dépendante, la somme des carrés des résidus, l'écart-type de la régression $\hat{\sigma}$, le R^2, le R^2 ajusté $\bar{R}^2 = 1 - (1 - R^2)(n-1)/(n-k)$, la statistique F de significativité globale avec sa probabilité critique, la log-vraisemblance et les critères d'information d'Akaike (AIC $= -2\ell + 2k$), de Schwarz (BIC) et de Hannan-Quinn. Pour des séries temporelles, Gretl ajoute le coefficient d'autocorrélation d'ordre 1 des résidus (« rho ») et la statistique de Durbin-Watson.

B.6.2 La régression multiple

Code B.12. Régression multiple

```
open data4-1 --quiet
ols price const sqft bedrms baths
```

```
Modèle 1: MCO, utilisant les observations 1-14
Variable dépendante: price

             coefficient    éc. type    t de Student   p. critique
  ------------------------------------------------------------------
  const      129,062       88,3033         1,462         0,1746
  sqft         0,154800     0,0319404      4,847         0,0007    ***
  bedrms     -21,5875      27,0293        -0,7987        0,4430
  baths      -12,1928      43,2500        -0,2819        0,7838

Moyenne var. dép.     317,4929   Éc. type var. dép.    88,49816
Somme carrés résidus  16700,07   Éc. type régression   40,86572
R2                    0,835976   R2 ajusté             0,786769
F(3, 10)              16,98894   P. critique (F)       0,000299
Log de vraisemblance -69,45391   Critère d'Akaike      146,9078
Critère de Schwarz    149,4641   Hannan-Quinn          146,6712
```

La surface reste le seul déterminant significatif (0,1548, $t = 4{,}85$). Une fois la surface contrôlée, le nombre de chambres (−21,6) et de salles de bains (−12,2) n'a pas d'effet significatif (probabilités critiques 0,44 et 0,78), et les signes négatifs reflètent surtout la colinéarité avec la surface (corrélation de 0,79 entre `sqft` et `baths`). Le R^2 augmente mécaniquement (0,836),

mais le R^2 ajusté baisse de 0,806 à 0,787 : les deux variables ajoutées ne méritent pas de rester dans le modèle.

B.6.3 Récupérer les résultats d'une estimation

Après chaque estimation, Gretl garde les résultats dans des accesseurs que l'on peut réutiliser : `$coeff` (coefficients), `$stderr` (écarts-types), `$nobs`, `$df`, `$ncoeff`, `$rsq`, `$sigma`, `$Fstat`, `$uhat` (résidus), `$yhat` (valeurs ajustées), `$lnl`, `$aic` et `$bic`. Le code B.13 s'en sert pour recalculer le test F global et un intervalle de confiance.

Code B.13. Accéder aux résultats d'une estimation

```
open data4-1 --quiet
ols price const sqft bedrms baths --quiet
scalar fstat = $Fstat                 # F de significativité globale
scalar pF = pvalue(F, $ncoeff - 1, $df, fstat)
matrix b  = $coeff                    # coefficients
matrix se = $stderr                   # écarts types
scalar tc = critical(t, $df, 0.025)   # quantile de Student à 5 %
printf "n = %d, R2 = %.4f, sigma = %.3f\n", $nobs, $rsq, $sigma
printf "F(%d, %d) = %.3f (p = %.4f)\n", $ncoeff - 1, $df, fstat, pF
printf "IC 95 %% de sqft : [%.4f ; %.4f]\n", b[2]-tc*se[2], b[2]+tc*se[2]
```

```
n = 14, R2 = 0,8360, sigma = 40,866
F(3, 10) = 16,989 (p = 0,0003)
IC 95 % de sqft : [0,0836 ; 0,2260]
```

On retrouve $F(3, 10) = 16{,}989$ et sa probabilité critique 0,0003, calculée par `pvalue(F, ...)`. L'intervalle de confiance à 95 % de la pente de `sqft` est $\hat{\beta} \pm t_{0{,}975;10}\,\widehat{\text{se}}$, soit [0,0836 ; 0,2260] : il est plus large que dans la régression simple, ce qui traduit la perte de précision due à la colinéarité.

Pas d'accesseur pour le R^2 ajusté

Contrairement à ce que laissent croire certains manuels, Gretl ne fournit pas d'accesseur `$adjrsq` : le message « Le symbole '$adjrsq' n'est pas défini » apparaît. Le R^2 ajusté s'obtient par `1 - (1 - $rsq) * ($nobs - 1) / $df`. De même, `$pvalue` renvoie la probabilité critique du dernier test exécuté, pas celle du F global : on la calcule avec `pvalue(F, $ncoeff - 1, $df, $Fstat)`.

B.6.4 Un modèle log-linéaire avec indicatrices

Quand la variable dépendante est en logarithme, chaque coefficient se lit comme une semi-élasticité. Le code B.14 explique le logarithme du prix des 59 maisons de `data7-11` par la surface, l'âge et trois indicatrices (quartier de La Jolla, piscine, vue).

Code B.14. Modèle log-linéaire avec indicatrices

```
open data7-11 --quiet            # 59 maisons de La Jolla et University City
series lprice = log(price)       # log du prix : semi-élasticités
ols lprice const sqft age lajolla pool view
```

```
Modèle 1: MCO, utilisant les observations 1-59
Variable dépendante: lprice

            coefficient    éc. type    t de Student  p. critique
  ------------------------------------------------------------------
  const      4,36572      0,117554       37,14        1,27e-39   ***
  sqft       0,000314719  4,84937e-05     6,490       2,98e-08   ***
  age       -0,000489849  0,00280356     -0,1747      0,8620
  lajolla    0,536909     0,0615239       8,727       7,87e-12   ***
  pool       0,110912     0,0661499       1,677       0,0995     *
  view       0,110954     0,0567962       1,954       0,0560     *

Moyenne var. dép.     5,331199   Éc. type var. dép.    0,439191
Somme carrés résidus  2,101096   Éc. type régression   0,199106
R2                    0,812193   R2 ajusté             0,794476
F(5, 53)              45,84098   P. critique (F)       4,62e-18
Log de vraisemblance  14,66744   Critère d'Akaike     -17,33487
Critère de Schwarz   -4,869646   Hannan-Quinn         -12,46895
```

Cent pieds carrés supplémentaires augmentent le prix d'environ 3,1 % ($100 \times 0{,}000315 \times 100$). Une maison de La Jolla se vend en moyenne $e^{0{,}537} - 1 \approx 71\,\%$ plus cher qu'une maison comparable de University City, avec une statistique t de 8,73. La piscine et la vue ajoutent chacune environ 11,7 % ($e^{0{,}111} - 1$) mais ne sont significatives qu'au seuil de 10 %, tandis que l'âge n'a aucun effet mesurable (probabilité critique 0,86). Le modèle explique 81 % de la variance du log du prix ($\bar{R}^2 = 0{,}794$).

B.6.5 Diagnostics et écarts-types robustes

Les tests de diagnostic se lancent depuis le menu `Tests` de la fenêtre du modèle ou par script (code B.15). On range les régresseurs dans une liste `X` pour ne pas les recopier.

Code B.15. Tests de diagnostic et écarts types robustes

```
open data7-11 --quiet
series lprice = log(price)
list X = const sqft age lajolla pool view
ols lprice X --quiet
modtest --breusch-pagan --quiet   # H0 : variance des erreurs constante
modtest --white --quiet           # H0 : homoscédasticité (forme générale)
modtest --normality --quiet       # H0 : erreurs normales (Doornik-Hansen)
reset --quiet                     # H0 : forme fonctionnelle correcte
vif                               # facteurs d'inflation de la variance
ols lprice X --robust             # écarts types robustes (White)
```

```
Test de Breusch-Pagan pour l'hétéroscédasticité

Statistique de test: LM = 33,455614,
avec p. critique = P(Khi-deux(5) > 33,455614) = 0,000003

Test de White pour l'hétéroscédasticité

Statistique de test: TR^2 = 39,429931,
avec p. critique = P(Khi-deux(17) > 39,429931) = 0,001556

Test de l'hypothèse nulle de normalité de la distribution:
Khi-deux(2) = 17,143 avec p. critique 0,00019

Test RESET pour la spécification (carrés et cubes)
Statistique de test: F = 0,377277,
avec p. critique = P(F(2,51) > 0,377277) = 0,688

Facteurs d'inflation de variance
Valeur minimale possible = 1.0
Des valeurs > 10.0 peuvent indiquer un problème de colinéarité

        sqft    1,311
         age    1,409
     lajolla    1,408
        pool    1,119
        view    1,197
[...]
Modèle 1: MCO, utilisant les observations 1-59
Variable dépendante: lprice
Écarts-types robustes (hétéroscédasticité), variante HC0

            coefficient    éc. type    t de Student  p. critique
  ------------------------------------------------------------
  const      4,36572      0,125310       34,84        3,30e-38   ***
  sqft       0,000314719  6,08626e-05     5,171       3,63e-06   ***
  age       -0,000489849  0,00288420     -0,1698      0,8658
  lajolla    0,536909     0,0605151       8,872       4,65e-12   ***
  pool       0,110912     0,0743429       1,492       0,1417
  view       0,110954     0,0482967       2,297       0,0256     **
[...]
F(5, 53)              127,5032   P. critique (F)       2,79e-28
```

Les tests de Breusch-Pagan (LM = 33,46, $p < 0{,}0001$) et de White ($TR^2 = 39{,}43$ avec 17 degrés de liberté, $p = 0{,}0016$) rejettent l'homoscédasticité des erreurs (Breusch & Pagan, 1979). Le test de normalité rejette aussi la loi normale des résidus ($\chi^2(2) = 17{,}14$, $p = 0{,}0002$). En revanche, le test RESET ne détecte aucune erreur de forme fonctionnelle ($p = 0{,}69$) et les facteurs d'inflation de la variance, tous inférieurs à 1,5, écartent la colinéarité. Face à l'hétéroscédasticité, on conserve les coefficients mais on corrige les écarts-types : avec `--robust`, l'écart-type de `view` passe de 0,0568 à 0,0483, ce qui rend la vue significative à 5 % ($p = 0{,}026$), tandis que celui de `pool` augmente (la piscine n'est plus significative, $p = 0{,}14$). Le F affiché devient alors une statistique de Wald robuste (127,5). Par défaut, l'option `--robust` donne la variante HC0 de White ; la commande `set hc_version 1` sélectionne la variante HC1.

B.6.6 Tester des restrictions linéaires

Le test d'exclusion de plusieurs variables se fait avec `omit`, et un ensemble de restrictions linéaires quelconques avec `restrict` (code B.16). Dans une restriction, le membre de droite doit être un nombre : on écrit `b[pool] - b[view] = 0` et non `b[pool] = b[view]`.

Code B.16. Tests de restrictions linéaires

```
open data7-11 --quiet
series lprice = log(price)
list X = const sqft age lajolla pool view
ols lprice X --quiet
omit age pool view --quiet        # F : H0 age = pool = view = 0
ols lprice X --quiet
restrict --quiet                  # H0 : pool = view et age = 0
  b[pool] - b[view] = 0
  b[age] = 0
end restrict
```

```
  Hypothèse nulle : les paramètres de la régression sont nuls pour les variables
    age, pool, view
  Statistique de test: F(3, 53) = 3,20126, p. critique 0,030543

Ensemble de restrictions
 1: b[pool] - b[view] = 0
 2: b[age] = 0

Statistique de test: F(2, 53) = 0,0153182, avec p. critique = 0,984803
```

Le premier test rejette, au seuil de 5 %, la nullité conjointe de l'âge, de la piscine et de la vue ($F(3, 53) = 3{,}20$, $p = 0{,}031$), bien que l'âge soit insignifiant seul : la piscine et la vue ont un pouvoir explicatif commun. Le second test ne rejette pas l'hypothèse jointe « la piscine et la vue ont le même effet et l'âge n'en a pas » ($F(2, 53) = 0{,}015$, $p = 0{,}98$), ce qui autorise une version plus simple du modèle.

B.7 Variables instrumentales et données de panel

B.7.1 Variables instrumentales

Quand un régresseur est corrélé à l'erreur, les MCO sont biaisés et on recourt aux doubles moindres carrés (chapitre 6). Dans l'interface, le menu est `Modèle > Variables instrumentales > Doubles Moindres Carrés, 2MC` ; dans un script, la commande `tsls` sépare par un point-virgule la liste des régresseurs et celle des instruments. Le code B.17 estime le rendement de l'éducation des femmes salariées de `mroz87` en instrumentant l'éducation `WE` par celle des parents (`WMED`, `WFED`).

Code B.17. Variables instrumentales : rendement de l'éducation

```
open mroz87 --quiet
smpl WW > 0 --restrict                # 428 femmes salariées
series lwage   = log(WW)              # log du salaire horaire
series exper   = AX                   # expérience
series expersq = AX^2
ols lwage const WE exper expersq --quiet   # MCO : WE supposée exogène
scalar b_mco = $coeff(WE)
tsls lwage const WE exper expersq ; const exper expersq WMED WFED
printf "Éducation : MCO = %.4f ; 2MC = %.4f\n", b_mco, $coeff(WE)
```

```
Modèle 1: 2MC, utilisant les observations 1-428
Variable dépendante: lwage
Instrumenté: WE
Instruments: const exper expersq WMED WFED

            coefficient    éc. type    t de Student  p. critique
  ------------------------------------------------------------
  const      0,0481003    0,400328        0,1202       0,9044
  WE         0,0613966    0,0314367       1,953        0,0515    *
  exper      0,0441704    0,0134325       3,288        0,0011    ***
  expersq   -0,000898970  0,000401686    -2,238        0,0257    **

Moyenne var. dép.     1,190173   Éc. type var. dép.    0,723198
Somme carrés résidus  193,0200   Éc. type régression   0,674712
R2                    0,145660   R2 ajusté             0,139615
F(3, 424)             8,140709   P. critique (F)       0,000028

Test de Hausman -
  Hypothèse nulle: Les estimateurs MCO sont non biaisés
  Statistique du test asymptotique: Khi-deux(1) = 2,8256
  avec p. critique = 0,0927721

Test de Sargan de sur-identification -
  Hypothèse nulle: tous les instruments sont valides
  Statistique de test: LM = 0,378071
  avec p. critique = P(Khi-deux(1) > 0,378071) = 0,538637

Test d'instruments faibles -
  Statistique F, première étape (2, 423) = 55,4003
[...]
Éducation : MCO = 0,1075 ; 2MC = 0,0614
```

Le rendement estimé de l'éducation chute de 10,75 % par année d'études (MCO) à 6,14 % (2MC), avec un écart-type de 0,031 : le coefficient 2MC n'est significatif qu'à 10 % ($p = 0{,}052$). Le test de Hausman, qui compare MCO et 2MC, rejette l'exogénéité de l'éducation au seuil de 10 % seulement ($p = 0{,}093$). Il y a deux instruments pour un régresseur endogène : le test de Sargan de sur-identification ne rejette pas la validité des instruments ($p = 0{,}54$), et la statistique F de première étape (55,4, très supérieure au seuil habituel de 10) écarte le risque d'instruments faibles. Gretl affiche ces trois diagnostics sans qu'on les demande. Le R^2 d'une estimation par variables instrumentales n'a pas d'interprétation habituelle.

B.7.2 Données de panel

Les estimateurs de panel (chapitre 6) sont dans `Modèle > Panel > Effets fixes ou aléatoires`, ou dans la commande `panel`. Le code B.18 estime la fonction d'investissement des 10 entreprises de Grunfeld, déjà déclarées comme panel dans le fichier.

Code B.18. Panel : effets fixes, effets aléatoires et test de Hausman

```
open grunfeld --quiet                 # 10 entreprises, 1935-1954
panel invest const value kstock --fixed-effects
panel invest const value kstock --random-effects --quiet
matrix b = $coeff           # coefficients des effets aléatoires
printf "EA : value = %.4f, kstock = %.4f\n", b[2], b[3]
matrix h = $hausman         # (statistique, ddl, p-valeur)
printf "Hausman : chi2(%d) = %.3f, p = %.4f\n", h[2], h[1], h[3]
```

```
Modèle 1: Effets fixes, utilisant 200 observations
10 processus individuels inclus
Dimension temporelle = 20
Variable dépendante: invest

             coefficient    éc. type    t de Student   p. critique
  ------------------------------------------------------------------
  const      -58,7439      12,4537         -4,717       4,66e-06   ***
  value        0,110124     0,0118567       9,288       3,92e-17   ***
  kstock       0,310065     0,0173545      17,87        2,22e-42   ***
[...]
Test de différence des constantes individuelles -
  Hypothèse nulle: Constante commune pour tous les individus
  Statistique de test: F(9, 188) = 49,1766
  avec p. critique = P(F(9, 188) > 49,1766) = 8,70012e-45

EA : value = 0,1098, kstock = 0,3081
Hausman : chi2(2) = 2,186, p = 0,3352
```

Le modèle à effets fixes donne un effet de 0,110 pour la valeur boursière (`value`) et de 0,310 pour le stock de capital (`kstock`), tous deux très significatifs. Le test de différence des constantes individuelles ($F(9, 188) = 49{,}18$) rejette la constante commune : les entreprises diffèrent. Les effets aléatoires (ligne « EA » de la sortie) donnent des valeurs presque identiques (0,110 et 0,308), et le test de Hausman ne rejette pas la cohérence de cet estimateur ($\chi^2(2) = 2{,}19$, $p = 0{,}34$) : on peut retenir les effets aléatoires, plus efficaces. Gretl n'a pas de commande `hausman` : le test s'affiche tout seul après `panel --random-effects`, et l'accesseur `$hausman` contient la statistique, les degrés de liberté et la probabilité critique.

B.8 Variables qualitatives, limitées et de comptage

Les modèles de cette section (chapitre 8) se trouvent dans `Modèle > Variable dépendante censurée`. Ils s'estiment par maximum de vraisemblance. Le code B.19 et les suivants reprennent l'étude de Mroz (1987) sur la participation des femmes mariées au marché du travail.

B.8.1 Probit et effets marginaux

Code B.19. Probit de la participation et effet marginal moyen

```
open mroz87 --quiet
series nwifeinc = (FAMINC - WW*WHRS) / 1000   # autres revenus du ménage
series exper    = AX
series expersq  = AX^2
list X = const nwifeinc WE exper expersq WA KL6 K618
probit LFP X                                  # participation à l'emploi
series xb = lincomb(X, $coeff)                # indice linéaire X'b
scalar ame_kl6 = mean(dnorm(xb)) * $coeff(KL6)
printf "Effet marginal moyen de KL6 : %.4f\n", ame_kl6
```

```
Modèle 1: Probit, utilisant les observations 1-753
Variable dépendante: LFP
Écarts-types basés sur la matrice hessienne

             coefficient    éc. type        z         pente
  ------------------------------------------------------------
  const       0,270077     0,508593       0,5310
  nwifeinc   -0,0120237    0,00483984    -2,484    -0,00469623
  WE          0,130905     0,0252542      5,183     0,0511287
  exper       0,123348     0,0187164      6,590     0,0481771
  expersq    -0,00188708   0,000599986   -3,145    -0,000737055
  WA         -0,0528527    0,00847724    -6,235    -0,0206432
  KL6        -0,868329     0,118522      -7,326    -0,339151
  K618        0,0360050    0,0434768      0,8281    0,0140628

Moyenne var. dép.     0,568393   Éc. type var. dép.    0,495630
R2 de McFadden        0,220581   R2 ajusté             0,205043
Log de vraisemblance -401,3022   Critère d'Akaike      818,6044
Critère de Schwarz    855,5969   Hannan-Quinn          832,8557

Nombre de cas 'correctement prédis' = 553 (73,4%)
f(beta'x) à la valeur moyenne des variables indépendantes = 0,391
Test du ratio de vraisemblance: Khi-deux(7) = 227,142 [0,0000]
[...]
Effet marginal moyen de KL6 : -0,2612
```

Le jeune enfant est le déterminant le plus fort : un enfant de moins de six ans diminue l'indice de participation de 0,868. La colonne « pente » donne l'effet marginal *à la moyenne* des régresseurs : la probabilité de travailler recule de 34 points de pourcentage par enfant de moins de six ans, et augmente de 5,1 points par année d'études. L'effet marginal moyen sur

l'échantillon, calculé à la main avec la densité normale, est un peu plus faible (−0,261). Le modèle classe correctement 73,4 % des femmes, pour un R^2 de McFadden de 0,22. Pour un logit, on remplace `probit` par `logit` ; si la variable dépendante prend plus de deux valeurs ordonnées, la même commande `logit` estime un logit ordonné (il n'y a pas de commande `ologit`). Les commandes `mfx` et `lrtest`, parfois citées, n'existent pas dans Gretl.

B.8.2 Données censurées : le modèle Tobit

Code B.20. Modèle Tobit des heures de travail

```
open mroz87 --quiet
series nwifeinc = (FAMINC - WW*WHRS) / 1000
series exper    = AX
series expersq  = AX^2
list X = const nwifeinc WE exper expersq WA KL6 K618
tobit WHRS X               # heures de travail censurées à 0
```

```
Modèle 1: Tobit, utilisant les observations 1-753
Variable dépendante: WHRS
Écarts-types basés sur la matrice hessienne

             coefficient    éc. type       z      p. critique
  ---------------------------------------------------------------
  const       965,305      446,436       2,162     0,0306     **
  nwifeinc     -8,81424      4,45910    -1,977     0,0481     **
  WE           80,6456      21,5832      3,736     0,0002     ***
  exper       131,564       17,2794      7,614     2,66e-14   ***
  expersq      -1,86416      0,537662   -3,467     0,0005     ***
  WA          -54,4050       7,41850    -7,334     2,24e-13   ***
  KL6        -894,022      111,878      -7,991     1,34e-15   ***
  K618        -16,2180      38,6414     -0,4197    0,6747

Khi-deux(7)           253,8619   p. critique           4,17e-51
Log de vraisemblance -3819,095   Critère d'Akaike      7656,189
Critère de Schwarz    7697,806   Hannan-Quinn          7672,222

sigma = 1122,02 (41,5791)
Observations censurées à gauche: 325
Observations censurées à droite: 0
```

Les 325 femmes qui ne travaillent pas ont des heures censurées à 0, d'où le recours au Tobit. Les coefficients mesurent l'effet sur les heures *désirées* (la variable latente) et non sur les heures observées : une année d'études supplémentaire accroît les heures désirées de 80,6 heures par an, un enfant de moins de six ans les réduit de 894 heures, et l'écart-type de l'erreur est estimé à $\hat{\sigma} = 1\,122$. L'effet sur les heures observées est plus faible, car il combine la hausse de la probabilité de travailler et celle des heures pour celles qui travaillent.

B.8.3 Sélection d'échantillon : le modèle de Heckman

Le salaire n'est observé que pour les femmes qui travaillent. L'estimateur de Heckman en deux étapes (Heckman, 1979) corrige ce biais de sélection : la commande `heckit` prend l'équation de salaire, puis, après le point-virgule, la variable de sélection et ses déterminants.

Code B.21. Modèle de sélection de Heckman en deux étapes

```
open mroz87 --quiet
series nwifeinc = (FAMINC - WW*WHRS) / 1000
series exper    = AX
series expersq  = AX^2
series lwage    = log(WW)        # manquant pour les non-salariées
list Z = const nwifeinc WE exper expersq WA KL6 K618
heckit lwage const WE exper expersq ; LFP Z --two-step
```

```
Modèle 1: Heckit en deux étapes, utilisant les observations 1-753
Variable dépendante: lwage
Variable de sélection: LFP

             coefficient     éc. type        z      p. critique
  ---------------------------------------------------------------
  const      -0,578103      0,305006      -1,895     0,0580     *
  WE          0,109066      0,0155230      7,026     2,12e-12   ***
  exper       0,0438873     0,0162611      2,699     0,0070     ***
  expersq    -0,000859114   0,000438916   -1,957     0,0503     *
  lambda      0,0322619     0,133625       0,2414    0,8092

                     Équation de sélection

  const       0,270077      0,508593       0,5310    0,5954
  nwifeinc   -0,0120237     0,00483984    -2,484     0,0130     **
  WE          0,130905      0,0252542      5,183     2,18e-07   ***
  exper       0,123348      0,0187164      6,590     4,39e-11   ***
  expersq    -0,00188708    0,000599986   -3,145     0,0017     ***
  WA         -0,0528527     0,00847724    -6,235     4,53e-10   ***
  KL6        -0,868329      0,118522      -7,326     2,37e-13   ***
  K618        0,0360050     0,0434768      0,8281    0,4076

Moyenne var. dép.    1,190173   Éc. type var. dép.   0,723198
sigma                0,663629   rho                  0,048614

Observations totales: 753
Observations censurées: 325 (43,2%)
```

L'équation de sélection exclut le revenu du ménage hors salaire, l'âge et le nombre d'enfants, qui jouent le rôle de variables d'exclusion nécessaires à l'identification. Le coefficient de la variable `lambda` (rapport de Mills inverse) est petit et non significatif (0,032, $p = 0{,}81$) : on ne détecte pas de biais de sélection. Le rendement de l'éducation, estimé à 10,9 %, reste donc proche de celui des MCO sur les seules femmes salariées (10,75 %).

B.8.4 Données de comptage

Pour une variable qui compte des événements (visites chez le médecin, accidents), on estime un modèle de Poisson, puis un binomial négatif si la variance dépasse la moyenne (chapitre 8).

Code B.22. Modèle binomial négatif pour des données de comptage

```
open rac3d --quiet           # consultations médicales, Australie
list X = const SEX AGE AGESQ INCOME LEVYPLUS FREEPOOR FREEREPA
list X += ILLNESS ACTDAYS HSCORE CHCOND1 CHCOND2
summary DVISITS --simple     # moyenne 0,30 ; écart type 0,80
poisson DVISITS X --quiet
scalar aic_p = $aic
negbin DVISITS X
printf "AIC : Poisson = %.1f ; binomiale négative = %.1f\n", aic_p, $aic
```

```
Statistiques descriptives, utilisant les observations 1 - 5190
pour la variable « DVISITS » (5190 observations valides)

  Moyenne            0,30173
  Minimum             0,0000
  Maximum             9,0000
  Écart-type         0,79813
  Obs. manq.               0

Modèle 1: Loi binomiale négative, utilisant les observations 1-5190
Variable dépendante: DVISITS
Écarts-types basés sur la matrice hessienne

             coefficient    éc. type       z      p. critique
  ---------------------------------------------------------------
  const      -2,19001      0,233580     -9,376     6,86e-21   ***
  SEX         0,216644     0,0693873     3,122     0,0018     ***
  AGE        -0,216159     1,28102      -0,1687    0,8660
  AGESQ       0,609159     1,40618       0,4332    0,6649
  INCOME     -0,142202     0,108190     -1,314     0,1887
  LEVYPLUS    0,118064     0,0855379     1,380     0,1675
  FREEPOOR   -0,496611     0,206890     -2,400     0,0164     **
  FREEREPA    0,144982     0,116949      1,240     0,2151
  ILLNESS     0,214341     0,0242276     8,847     8,99e-19   ***
  ACTDAYS     0,143754     0,00781400   18,40      1,39e-75   ***
  HSCORE      0,0380601    0,0137992     2,758     0,0058     ***
  CHCOND1     0,0993550    0,0787023     1,262     0,2068
  CHCOND2     0,190327     0,104409      1,823     0,0683     *
  alpha       1,07704      0,103012     10,46      1,38e-25   ***

Log de vraisemblance -3198,744   Critère d'Akaike      6425,488
Critère de Schwarz    6517,251   Hannan-Quinn          6457,588

AIC : Poisson = 6737,1 ; binomiale négative = 6425,5
```

Le nombre moyen de consultations est de 0,30 pour un écart-type de 0,80 : la variance (0,64) dépasse largement la moyenne, ce qui signale une surdispersion que le modèle de Poisson ne peut représenter. Le paramètre de dispersion α du binomial négatif est estimé à 1,08 et il est très significatif, et le critère d'Akaike diminue fortement (6 425 contre 6 737 pour Poisson). Chaque jour d'activité réduite augmente le nombre attendu de consultations d'environ 15 % ($e^{0{,}144} - 1$), et chaque point de l'indice de maladie d'environ 24 % ($e^{0{,}214} - 1$).

B.9 Séries temporelles univariées

Cette section (chapitre 7) utilise la croissance du PIB réel américain construite à la section B.4. Les menus sont `Variable > Corrélogramme`, `Variable > Tests de racine unitaire` et `Modèle > Série temporelle univariée`.

B.9.1 Corrélogramme

Code B.23. Corrélogramme de la croissance du PIB

```
open greene5_1 --quiet           # PIB réel américain, 1950-2000
series lpib   = log(realgdp)
series croiss = 100 * diff(lpib) # croissance trimestrielle (%)
corrgm croiss 12 --plot=none     # ACF, PACF et statistique de Ljung-Box
```

```
 RETARD   ACF      PACF         Q  [p. critique]

   1   0,3497  ***   0,3497 ***     25,1935  [0,000]
   2   0,1577  **    0,0403         30,3416  [0,000]
   3   0,0085       -0,0668         30,3567  [0,000]
   4  -0,0773       -0,0765         31,6072  [0,000]
   5  -0,1513  **   -0,1072         36,4206  [0,000]
   6  -0,0650        0,0388         37,3133  [0,000]
   7  -0,0606       -0,0306         38,0928  [0,000]
   8  -0,0582       -0,0453         38,8153  [0,000]
   9   0,0591        0,0951         39,5645  [0,000]
  10   0,0757        0,0245         40,8006  [0,000]
  11  -0,0047       -0,0715         40,8054  [0,000]
  12  -0,1260  *    -0,1496 **      44,2619  [0,000]
```

La commande `corrgm` affiche l'autocorrélation (ACF), l'autocorrélation partielle (PACF) et la statistique de Ljung-Box Q pour chaque retard. Le premier coefficient est de 0,35 et il est très significatif, les suivants s'éteignent vite, et la PACF s'annule dès le deuxième retard : cette signature (figure B.4) est celle d'un processus autorégressif d'ordre 1. La statistique Q rejette sans ambiguïté l'absence d'autocorrélation (44,3 avec 12 retards).

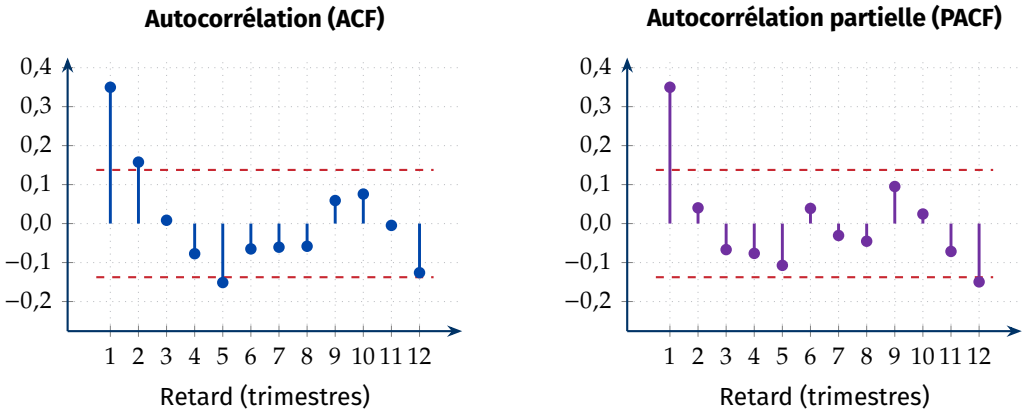


FIGURE B.4. Corrélogramme de la croissance trimestrielle du PIB réel américain, de 1950 à 2000 (sortie du code B.23). À gauche, l'ACF ; à droite, la PACF. Les pointillés rouges délimitent l'intervalle $\pm 1{,}96/\sqrt{T}$ (avec $T = 203$) : un coefficient hors de cet intervalle est significativement différent de zéro à 5 %. Seul le premier coefficient de la PACF est nettement significatif.

B.9.2 Racine unitaire et stationnarité

Code B.24. Tests de racine unitaire et de stationnarité

```
open greene5_1 --quiet
series lpib   = log(realgdp)
series croiss = 100 * diff(lpib)
adf 8 lpib --ct --test-down=BIC    # H0 : racine unitaire
kpss -1 lpib --trend               # H0 : stationnarité (tendance)
adf 8 croiss --c --test-down=BIC
kpss -1 croiss
```

```
Test de Dickey-Fuller augmenté pour lpib
test à reculons à partir de 8 retards, suivant le critère BIC
taille de l'échantillon 202
hypothèse nulle de racine unitaire : a = 1

  avec constante et tendance temporelle
  avec un retard de (1-L)lpib
  modèle: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
  valeur estimée de (a - 1): -0,0483387
  statistique de test: tau_ct(1) = -2,89175
  p. critique asymptotique 0,1651

Test KPSS pour lpib (avec tendance)

T = 204
Paramètre du retard de troncation = 4
Statistique de test = 0,602516

                      10%      5%      1%
Valeurs critiques: 0,120   0,148   0,217
P. critique < .01

Test de Dickey-Fuller augmenté pour croiss
test à reculons à partir de 8 retards, suivant le critère BIC
taille de l'échantillon 202
hypothèse nulle de racine unitaire : a = 1

  test avec constante
  avec 0 retards de (1-L)croiss
  modèle: (1-L)y = b0 + (a-1)*y(-1) + e
  valeur estimée de (a - 1): -0,650011
  statistique de test: tau_c(1) = -9,93349
  p. critique asymptotique 5,423e-19

Test KPSS pour croiss

T = 203
Paramètre du retard de troncation = 4
Statistique de test = 0,135843

                      10%      5%      1%
Valeurs critiques: 0,348   0,462   0,739
P. critique > .10
```

Les deux tests ont des hypothèses nulles opposées : le test de Dickey-Fuller augmenté (`adf`) teste la racine unitaire, le test KPSS (Kwiatkowski et al., 1992) teste la stationnarité. Pour le logarithme du PIB, avec constante et tendance, l'ADF ne rejette pas la racine unitaire ($\tau = -2{,}89$, $p = 0{,}165$) et le KPSS rejette la stationnarité autour d'une tendance (0,603, supérieur à la valeur critique de 0,217 à 1 %) : les deux tests concordent, la série est intégrée d'ordre 1. Pour la croissance, l'ADF rejette massivement la racine unitaire ($\tau = -9{,}93$) et le KPSS ne rejette pas la stationnarité (0,136) : la croissance est stationnaire. L'option `--test-down=BIC` choisit le nombre de retards de l'ADF par le critère de Schwarz à partir du maximum indiqué (8) ; `kpss -1` sélectionne automatiquement la fenêtre.

B.9.3 Modèle ARIMA et prévision

Code B.25. Estimation ARIMA et prévisions hors échantillon

```
open greene5_1 --quiet
series lpib = log(realgdp)
smpl 1950:1 1999:4              # estimation jusqu'en 1999:4
arima 1 1 0 ; lpib              # ARIMA(1,1,0) avec constante
fcast 2000:1 2000:4 lpib_prev   # prévisions hors échantillon
smpl 2000:1 2000:4
print lpib lpib_prev --byobs
```

```
Modèle 1: ARIMA, utilisant les observations 1950:2-1999:4 (T = 199)
Estimation par AS 197 (MV exact)
Variable dépendante: (1-L) lpib
Écarts-types basés sur la matrice hessienne

             coefficient    éc. type      z     p. critique
  ---------------------------------------------------------
  const      0,00876555    0,00103755   8,448    2,96e-17   ***
  phi_1      0,362255      0,0668974    5,415    6,13e-08   ***

Moyenne var. dép.     0,008674   Éc. type var. dép.    0,010051
Moyenne innovations  -0,000045   Éc. type innovations  0,009358
Log de vraisemblance  647,1950   Critère d'Akaike     -1288,390
Critère de Schwarz   -1278,510   Hannan-Quinn         -1284,391

Échantillon actuel: 2000:1 - 2000:4 (n = 4)

               lpib    lpib_prev

2000:1     9,116304     9,123312
2000:2     9,130149     9,133540
2000:3     9,133470     9,142835
2000:4     9,138189     9,151792
```

Le modèle ARIMA(1,1,0) du logarithme du PIB suppose que la croissance suit un processus AR(1) : la constante de 0,0088 est la croissance moyenne trimestrielle (environ 3,5 % par an) et le coefficient autorégressif vaut 0,362 ($z = 5{,}4$), valeur proche de la première autocorrélation observée. Les quatre prévisions de 2000 surestiment un peu le PIB réel : l'erreur de prévision (observation moins prévision du logarithme) est de −0,007, −0,003, −0,009 et −0,014, soit un écart de 0,3 à 1,4 %, croissant avec l'horizon, comme attendu. Avec `fcast`, la prévision est enregistrée dans une nouvelle série (ici `lpib_prev`).

B.9.4 Volatilité : le modèle GARCH

Les rendements financiers ont des périodes agitées et des périodes calmes : la variance dépend du passé (chapitre 7). Le code B.26 estime un GARCH(1,1) pour les rendements quotidiens du taux deutschemark/livre sterling.

Code B.26. Modèle GARCH(1,1) des rendements quotidiens

```
open b-g --quiet                # rendements DM/livre sterling, 1974 jours
garch 1 1 ; Y                   # GARCH(1,1) : garch p q ; variable
matrix b = $coeff
printf "Persistance alpha(1) + beta(1) = %.4f\n", b[3] + b[4]
series sigma = sqrt($h)         # volatilité conditionnelle estimée
summary sigma --simple
```

```
Modèle 1: GARCH, utilisant les observations 1-1974
Variable dépendante: Y
Écarts-types basés sur la matrice hessienne

             coefficient    éc. type       z      p. critique
  ---------------------------------------------------------
  const      -0,00619040   0,00846212   -0,7315    0,4644

  alpha(0)    0,0107614    0,00285271    3,772     0,0002     ***
  alpha(1)    0,153134     0,0265228     5,774     7,76e-09   ***
  beta(1)     0,805974     0,0335527    24,02      1,67e-127  ***

Moyenne var. dép.    -0,016427   Éc. type var. dép.    0,470244
Log de vraisemblance -1106,608   Critère d'Akaike      2223,216
Critère de Schwarz    2251,155   Hannan-Quinn          2233,481

Variance non conditionnelle de l'erreur = 0,263165
Test du ratio de vraisemblance pour les termes (G)ARCH:
  Khi-deux(2) = 408,977 [1,55508e-89]

Persistance alpha(1) + beta(1) = 0,9591

Statistiques descriptives, utilisant les observations 1 - 1974
pour la variable « sigma » (1974 observations valides)

  Moyenne            0,44951
  Minimum            0,24155
  Maximum             1,3610
  Écart-type         0,16774
  Obs. manq.               0
```

La variance conditionnelle suit $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$ avec $\hat{\alpha}_0 = 0{,}0108$, $\hat{\alpha}_1 = 0{,}153$ et $\hat{\beta}_1 = 0{,}806$, tous très significatifs. La persistance $\hat{\alpha}_1 + \hat{\beta}_1 = 0{,}959$, proche de 1, signifie que les chocs de volatilité se dissipent lentement (une demi-vie d'environ 17 jours). Le test du rapport de vraisemblance ($\chi^2(2) = 409$) rejette l'absence d'effets (G)ARCH. La volatilité conditionnelle estimée varie de 0,24 % à 1,36 % par jour, pour une moyenne de 0,45 %. Dans Gretl, les coefficients portent les noms `alpha(0)`, `alpha(1)` et `beta(1)` : on les atteint par leur position dans la matrice `$coeff`, et non par un nom comme `$coeff(alpha_1)`.

B.10 VAR, cointégration et réponses impulsionnelles

Cette section (chapitre 7) étudie une relation de demande de monnaie sur données danoises : le logarithme de la monnaie réelle (`LRM`), celui du revenu réel (`LRY`), le taux des obligations (`IBO`) et le taux des dépôts bancaires (`IDE`). Les menus se trouvent dans `Modèle > Séries temporelles multivariées`.

B.10.1 Choisir l'ordre et estimer un VAR

Code B.27. Choix de l'ordre et estimation d'un VAR

```
open denmark --quiet                # monnaie, revenu, taux d'intérêt
var 6 LRM LRY IBO IDE --lagselect  # critères d'information, p = 1 à 6
var 2 LRM LRY IBO IDE --quiet      # VAR(2) : tests de Fisher par équation
```

```
Système VAR, ordre des retards max 6

Les astérisques indiquent les meilleures (donc les plus faibles) valeurs
des critères d'information suivants, AIC = critère d'Akaike,
BIC = critère bayésien de Schwartz et HQC = critère d'Hannan-Quinn.

retards   log-vrais.    p(LR)       AIC          BIC          HQC

   1     594,38410           -23,444249   -22,672078*  -23,151288*
   2     615,43480  0,00038  -23,650400*  -22,260491   -23,123070
   3     624,97613  0,26440  -23,386781   -21,379135   -22,625083
   4     636,03766  0,13926  -23,185210   -20,559827   -22,189144
   5     658,36014  0,00016  -23,443271   -20,200150   -22,212836
   6     669,88472  0,11243  -23,260601   -19,399743   -21,795797

Système VAR, ordre des retards 2
Estimation MCO, observations 1974:3-1987:3 (T = 53)
Log de vraisemblance = 653,3993
Déterminant de la matrice de covariance = 2,3008439e-16
AIC = -23,2981
BIC = -21,9598
HQC = -22,7834
Test du Portmanteau: LB(13) = 191,533, ddl = 176 [0,2004]

Équation de LRM:
Tests de Fisher d'absence de restriction:

Tous les retards de LRM            F(2, 44) =   22,617 [0,0000]
Tous les retards de LRY            F(2, 44) =   1,1524 [0,3252]
Tous les retards de IBO            F(2, 44) =   8,7514 [0,0006]
Tous les retards de IDE            F(2, 44) =   1,5937 [0,2147]
Toutes les variables, retard 2     F(4, 44) =   1,3124 [0,2802]
[...]
Pour le système dans l'ensemble:

  Hypothèse nulle: l'ordre des retards max est 1
  Hypothèse alternative: l'ordre des retards max est 2
  Test du ratio de vraisemblance: Khi-deux(16) = 46,2573 [0,0001]
```

L'option `--lagselect` compare les ordres 1 à 6 : le critère d'Akaike retient 2, ceux de Schwarz et de Hannan-Quinn retiennent 1. On choisit 2, car le test du rapport de vraisemblance rejette l'ordre 1 au profit de l'ordre 2 ($\chi^2(16) = 46{,}3$, $p = 0{,}0001$) et le test portmanteau ne détecte pas d'autocorrélation des résidus ($p = 0{,}20$). Pour chaque équation, Gretl fournit des tests de Fisher d'absence de restriction sur les retards de chaque variable : ce sont les tests de causalité au sens de Granger. Dans l'équation de la monnaie, les retards du taux des obligations sont très significatifs ($F(2, 44) = 8{,}75$, $p = 0{,}0006$), pas ceux du revenu ($p = 0{,}33$) ni du taux des dépôts ($p = 0{,}21$).

B.10.2 Cointégration et modèle à correction d'erreurs

Code B.28. Test de Johansen et modèle à correction d'erreurs

```
open denmark --quiet
johansen 2 LRM LRY IBO IDE --quiet   # tests de la trace et de Lmax
vecm 2 1 LRM LRY IBO IDE --quiet     # VECM : 2 retards, rang 1
```

```
Test de Johansen:
Nombre d'équations = 4
Ordre des retards = 2
Période d'estimation: 1974:3 - 1987:3 (T = 53)
Cas 3 : constante sans restriction

Log de vraisemblance = 803,807 (constante incluse: 653,399)

Rang Valeur propre Test de la Trace  p. critique  Test Lmax  p. critique
   0    0,44821     48,804 [0,0389]     31,514 [0,0120]
   1    0,17421     17,290 [0,6274]     10,145 [0,7345]
   2    0,11690     7,1449 [0,5673]     6,5889 [0,5467]
   3   0,010436    0,55602 [0,4559]    0,55602 [0,4559]

Corrigé pour la taille de l'échantillon  (ddl = 44)
Rang Test de la Trace p. critique
   0     48,804 [0,0692]
   1     17,290 [0,6699]
   2     7,1449 [0,5875]
   3    0,55602 [0,4696]
```

```
Système VECM, ordre des retards 2
Estimation Maximum de vraisemblance, observations 1974:3-1987:3 (T = 53)
Rang de cointégration = 1
Cas 3 : constante sans restriction

beta (vecteurs cointégrants, écarts-types entre parenthèses)

LRM       1,0000
         (0,0000)
LRY      -0,97565
         (0,14438)
IBO        5,4086
         (0,55145)
IDE       -4,1624
          (1,1362)

alpha (vecteurs d'ajustement)

LRM      -0,28147
LRY      0,037469
IBO    -0,0039022
IDE      0,019960
```

Le test de la trace rejette l'absence de relation de cointégration (48,8, $p = 0{,}039$) mais pas l'existence d'une seule (17,3, $p = 0{,}63$) ; le test de la valeur propre maximale (Lmax) conclut de même. Le rang retenu est donc $r = 1$. Avec la correction pour petit échantillon, la probabilité critique du premier test passe à 0,069 : le rejet n'est plus acquis qu'à 10 % (Johansen, 1995). La relation de long terme normalisée sur la monnaie est $\text{LRM} - 0{,}976\,\text{LRY} + 5{,}41\,\text{IBO} - 4{,}16\,\text{IDE} \approx 0$: l'élasticité de la demande de monnaie au revenu est proche de 1, le taux des obligations (coût d'opportunité de la monnaie) a un effet négatif et le taux des dépôts (rémunération) un effet positif. La vitesse d'ajustement de la monnaie, $\hat{\alpha} = -0{,}281$, indique qu'environ 28 % de l'écart à l'équilibre est corrigé chaque trimestre ; les trois autres variables s'ajustent à peine.

B.10.3 Réponses impulsionnelles et décomposition de la variance

Les fonctions `irf` et `fevd` renvoient des matrices : la première est la réponse d'une variable à un choc d'un écart-type sur une autre, la seconde la part de chaque choc dans la variance de l'erreur de prévision. Les arguments sont les numéros des variables dans la commande `var`, 0 désignant « toutes ».

Code B.29. Réponse impulsionnelle et décomposition de la variance

```
open denmark --quiet
set horizon 12                  # horizon : 12 trimestres
var 2 LRM LRY IBO IDE --silent
matrix rep = irf(1, 3)          # réponse de LRM à un choc sur IBO
matrix vd  = fevd(1, 0)         # décomposition de la variance de LRM
print rep
printf "Part de IBO dans la variance de LRM (h = 12) : %.3f\n", vd[12, 3]
```

```
rep (12 x 1)

      0,0000
   -0,011006
   -0,019310
   -0,025536
   -0,029118
   -0,031110
   -0,032031
   -0,032321
   -0,032164
   -0,031671
   -0,030906
   -0,029931

Part de IBO dans la variance de LRM (h = 12) : 0,598
```

La figure B.5 trace cette réponse. Un choc d'un écart-type sur le taux des obligations n'a aucun effet immédiat sur la monnaie réelle, par construction : `LRM` est classée avant `IBO` dans l'ordre de la décomposition de Cholesky. La monnaie recule ensuite pendant sept trimestres jusqu'à −3,2 %, puis remonte lentement. À douze trimestres, le taux des obligations explique 60 % de la variance de l'erreur de prévision de la monnaie réelle.

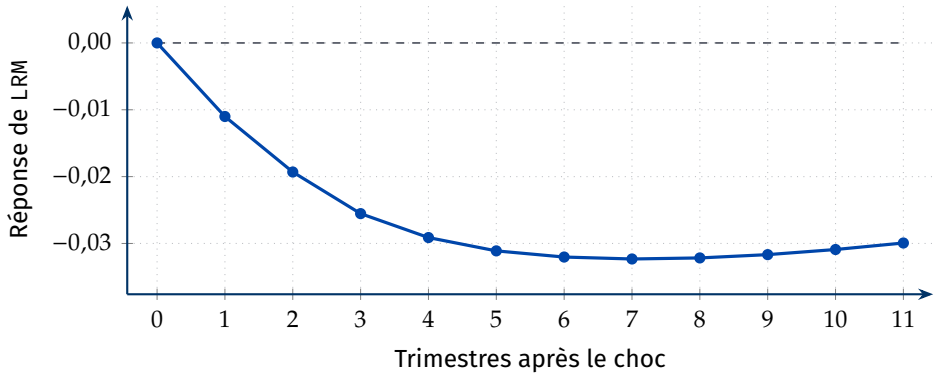


FIGURE B.5. Réponse de la monnaie réelle (`LRM`, en logarithme) à un choc d'un écart-type sur le taux des obligations (`IBO`) dans le VAR(2) du code B.29 (matrice `rep`, 12 trimestres). La réponse est nulle à la date du choc, puis négative et persistante.

B.11 Régression quantile et différences-en-différences

B.11.1 La régression quantile

La régression quantile (Koenker & Bassett, 1978) estime l'effet d'une variable en différents points de la distribution (chapitre 8). Dans `Modèle > Estimation robuste > Régression de quantile`, ou par la commande `quantreg`, on peut donner un seul quantile ou une matrice de quantiles. Le code B.30 reprend les données d'Engel des 235 ménages belges de Koenker et Bassett (1982).

Code B.30. Régression quantile sur cinq quantiles

```
open engel --quiet          # 235 ménages belges
ols foodexp const income --quiet
matrix tau = {0.1, 0.25, 0.5, 0.75, 0.9}
quantreg tau foodexp const income   # un appel pour tous les quantiles
```

```
Modèle 1: Estimation du quantile, utilisant les observations 1-235
Variable dépendante: foodexp
Écarts-types asymptotiques sous l'hypothèse d'erreurs IID

             tau    coefficient     éc. type t de Student
  ------------------------------------------------------
  const     0,100       110,142      17,8638      6,16562
            0,250       95,4835      15,8619      6,01968
            0,500       81,4822      13,2391      6,15468
            0,750       62,3966      9,85947      6,32860
            0,900       67,3509      20,5674      3,27464

  income    0,100      0,401766    0,0160831      24,9807
            0,250      0,474103    0,0142807      33,1989
            0,500      0,560181    0,0119193      46,9977
            0,750      0,644014   0,00887662      72,5518
            0,900      0,686299    0,0185171      37,0630
```

La propension marginale à dépenser pour l'alimentation augmente avec le quantile : 0,40 franc belge de dépense alimentaire par franc de revenu supplémentaire pour les ménages du premier décile, 0,56 pour la médiane et 0,69 pour le dernier décile ; les MCO, qui ne retiennent que la moyenne, donnent 0,49. La figure B.6 montre l'éventail des droites, qui reflète l'hétéroscédasticité : la dispersion des dépenses alimentaires croît avec le revenu.

B.11.2 Différences-en-différences

L'étude de Card et Krueger (1994) compare l'emploi dans les restaurants de restauration rapide du New Jersey, où le salaire minimum a augmenté en avril 1992, et de Pennsylvanie, où il est resté inchangé (chapitre 15). Un restaurant est caractérisé par l'indicatrice `STATE` (1 pour le New Jersey). L'estimateur des différences-en-différences s'obtient en régressant la *variation* de l'emploi sur cette indicatrice : le coefficient de `STATE` est la différence entre les variations des deux groupes (code B.31).

Code B.31. Différences-en-différences sur les données de Card et Krueger

```
open card_krueger94 --quiet       # fast-foods du NJ et de Pennsylvanie
series fte_av = EMPFT  + NMGRS  + 0.5*EMPPT     # emploi en ETP, avant
series fte_ap = EMPFT2 + NMGRS2 + 0.5*EMPPT2    # emploi en ETP, après
series dfte = fte_ap - fte_av                   # variation par restaurant
summary fte_av fte_ap --by=STATE --simple       # STATE = 1 : New Jersey
ols dfte const STATE                            # DiD : coefficient de STATE
```

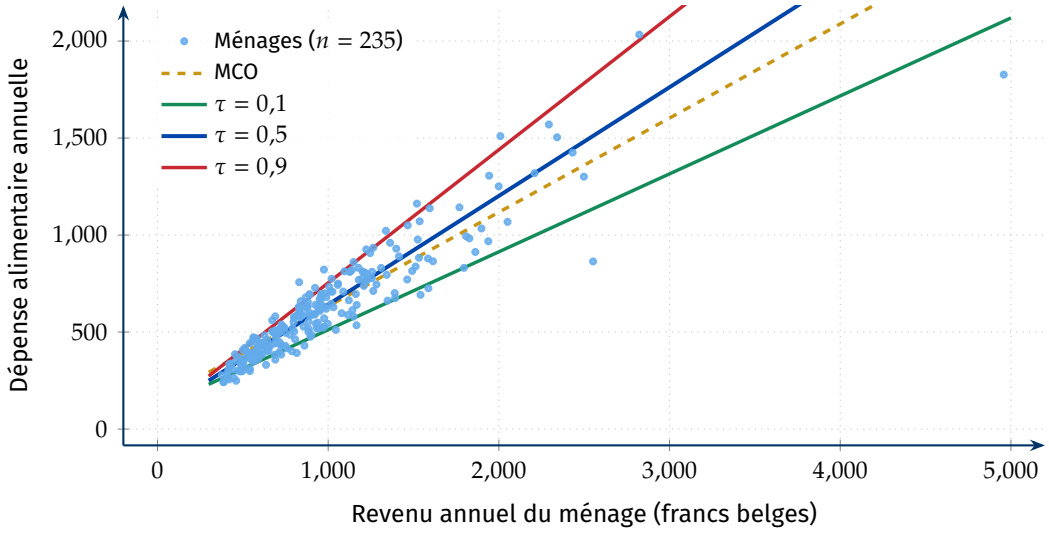


FIGURE B.6. Données d'Engel : dépense alimentaire en fonction du revenu pour 235 ménages belges (les 235 points sont les données réelles du fichier `engel`). La droite en tirets est la droite des MCO ($147{,}5 + 0{,}485\,x$) ; les droites pleines sont les régressions quantiles aux ordres 0,1 (vert), 0,5 (bleu) et 0,9 (rouge) du code B.30.

```
STATE = 0 (n = 77):

              Moyenne    Médiane       E.T.        Min        Max
fte_av          23,33      20,50      11,86      7,500      70,50
fte_ap          21,17      20,00      8,277      0,000      43,50

STATE = 1 (n = 321):

              Moyenne    Médiane       E.T.        Min        Max
fte_av          20,44      19,25      9,106      5,000      85,00
fte_ap          21,03      20,50      9,293      0,000      60,50

Modèle 1: MCO, utilisant les observations 1-410 (n = 384)
Suppression d'observations manquantes ou incomplètes: 26
Variable dépendante: dfte

             coefficient   éc. type   t de Student   p. critique
  ------------------------------------------------------------
  const       -2,28333     1,03551       -2,205        0,0280    **
  STATE        2,75000     1,15436        2,382        0,0177    **
```

En Pennsylvanie, l'emploi moyen passe de 23,33 à 21,17 équivalents temps plein (ETP) par restaurant, soit une baisse de 2,16 ; dans le New Jersey, il passe de 20,44 à 21,03, soit une hausse de 0,59. L'estimateur des différences-en-différences est donc $0{,}59 - (-2{,}16) = 2{,}75$ ETP (écart-type 1,15, $p = 0{,}018$), identique au coefficient de `STATE` : le salaire minimum n'a pas réduit l'emploi, ce qui contredit la prédiction de base du modèle concurrentiel. Les 384 observations utilisées sont celles des restaurants dont l'emploi est connu aux deux dates.

La régression de discontinuité (chapitre 14) se met en œuvre de la même façon, avec `smpl` pour retenir une fenêtre autour du seuil et `ols` avec une indicatrice de traitement et des

pentes différentes de part et d'autre. Gretl ne fournit pas de jeu de données adapté : aucun script n'est donné ici.

B.12 Scripts, boucles et fonctions

Un script est un fichier texte d'extension `.inp`. On l'exécute depuis l'éditeur de code, par la commande `run script.inp` dans la console ou en ligne de commande par `gretlcli -b script.inp`. Un bon script commence par des commentaires (objet, jeu de données, date), puis ouvre les données, déclare leur structure, enchaîne transformations, estimations et tests, et se termine par l'exportation des résultats.

B.12.1 Répéter une opération : la boucle

Code B.32. Boucle sur une liste de séries

```
open greene5_1 --quiet
list niveaux = realgdp realcons realinvs realgovt
loop foreach i niveaux
    series l_$i = log($i)
    adf 4 l_$i --ct --quiet          # constante et tendance
    printf "%-10s tau = %7.3f   p = %.3f\n", "l_$i", $test, $pvalue
endloop
```

```
l_realgdp  tau =  -2,504   p = 0,326
l_realcons tau =  -1,984   p = 0,609
l_realinvs tau =  -3,730   p = 0,020
l_realgovt tau =  -4,525   p = 0,001
```

La boucle `loop foreach i niveaux` parcourt les noms de la liste, et `$i` est remplacé par le nom courant. Avec constante et tendance, le test ADF ne rejette pas la racine unitaire pour le PIB réel ($p = 0{,}33$) ni la consommation ($p = 0{,}61$) mais la rejette pour l'investissement ($p = 0{,}02$) et les dépenses publiques ($p = 0{,}001$). Ces résultats dépendent du nombre de retards et de la spécification (ici 4 retards avec tendance) : il ne faut pas les lire mécaniquement, mais les confronter à ceux d'autres spécifications.

B.12.2 Réutiliser un calcul : la fonction

Code B.33. Une fonction pour le R² ajusté

```
function scalar r2aj(series y, list X)
    ols y X --quiet
    return 1 - (1 - $rsq) * ($nobs - 1) / $df    # R2 ajusté
end function

open data7-11 --quiet
series lprice = log(price)
list base = const sqft age lajolla
list plus = base pool view
printf "R2 ajusté, modèle de base  : %.4f\n", r2aj(lprice, base)
printf "R2 ajusté, modèle étendu   : %.4f\n", r2aj(lprice, plus)
```

```
R2 ajusté, modèle de base  : 0,7687
R2 ajusté, modèle étendu   : 0,7945
```

La fonction `r2aj` reçoit une série et une liste de régresseurs, estime le modèle sans affichage et renvoie le R^2 ajusté. On la rappelle ensuite pour comparer deux spécifications : l'ajout de la piscine et de la vue relève le R^2 ajusté de 0,769 à 0,795. Une fonction écrite une fois peut être mise dans un fichier séparé et chargée par la commande `include`.

B.13 Exporter les résultats

Un atout de Gretl est de fabriquer directement les tableaux d'un article ou d'un mémoire. Après une estimation, `tabprint` écrit le tableau des coefficients en LaTeX et `eqnprint` l'équation estimée ; `outfile` redirige vers un fichier tout ce qui s'afficherait à l'écran ; `store` enregistre les données au format natif ou en CSV ; `gnuplot ... --output` exporte un graphique (code B.34).

Code B.34. Exporter un tableau, une équation, un journal

```
open data7-11 --quiet
series lprice = log(price)
ols lprice const sqft age lajolla pool view --robust --quiet
tabprint --output="modele_prix.tex"     # tableau LaTeX
eqnprint --output="equation_prix.tex"   # équation estimée
outfile "journal.txt"                   # les affichages vont au fichier
    summary price sqft
end outfile
store "maisons_san_diego.gdt"           # données au format natif
```

```
Modèle imprimé vers modele_prix.tex
Modèle imprimé vers equation_prix.tex
Écriture en cours au fichier '~/gretl/journal.txt'
Fichier d'output '~/gretl/journal.txt' fermé
maisons_san_diego.gdt écrit
```

Le fichier `modele_prix.tex` contient un environnement `tabular` prêt à être inséré dans un document, avec la virgule décimale puisque la sortie suit la langue de l'interface ; `equation_prix.tex` contient l'équation estimée avec les écarts-types entre parenthèses (le paquet `amsmath` est nécessaire). La commande `modeltab`, qui assemble plusieurs modèles côte à côte, n'existe que dans l'interface graphique (fenêtre des icônes de la session) et n'a pas été testée ici. Dans la fenêtre d'un modèle, le menu `LaTeX` offre les mêmes services.

Messages d'erreur courants

- « *Le symbole 'sqftt' n'est pas défini* » : faute de frappe dans un nom de variable ; vérifiez avec `varlist` (les noms respectent la casse).
- « *Omis pour cause de multicolinéarité parfaite : d2* » : une variable est combinaison linéaire des autres, par exemple l'ensemble des indicatrices d'une modalité avec la constante ; retirez l'indicatrice de référence ou la constante.
- « *Désolé, commande non disponible pour cet estimateur* » : on a lancé une commande de test qui ne s'applique pas à l'estimateur courant ; c'est le cas de `hausman`, qui n'existe pas (le test est affiché automatiquement après `panel --random-effects`).
- « *Erreur lors de la tentative d'ouverture du fichier* » : le fichier n'est pas dans le dossier de travail ; donnez le chemin complet ou changez de dossier avec `Fichier > Dossier de travail...`.
- « *Matrice non définie positive* » : l'optimisation d'un GARCH ou d'un modèle à vraisemblance a échoué, souvent parce que la série est mal échelonnée ou inadaptée ; essayez un autre ordre ou une autre échelle.
- R^2 *anormalement proche de 1* : la constante a été oubliée, le R^2 non centré n'est pas comparable.

B.14 Récapitulatif des commandes essentielles

Le tableau B.5 met en regard chaque méthode du manuel, le menu qui l'ouvre et la commande qui la lance. Les mêmes méthodes sont présentées avec R à l'Annexe D (voir aussi Kleiber & Zeileis, 2008).

TABLE B.5. Correspondance entre méthodes, menus et commandes de GRETL

Méthode	Menu	Commande	Chap.
MCO	`Modèle > Moindres Carrés Ordinaires, MCO`	`ols y const x1 x2`	5
MCO robustes	case « Écarts-types robustes »	`ols y const x --robust`	5
Test de restrictions	`Tests > Omettre les variables, Restrictions linéaires`	`omit, restrict`	5
Diagnostics	`Tests > Hétéroscédasticité, Test du RESET de Ramsey`	`modtest --white, reset`	5
Variables instrumentales	`Modèle > Variables instrumentales > Doubles Moindres Carrés, 2MC`	`tsls y X ; Z`	6
Panel	`Modèle > Panel > Effets fixes ou aléatoires`	`panel y X --fixed-effects, --random-effects`	6
Probit, logit	`Modèle > Variable dépendante censurée > Probit, Logit`	`probit y X, logit y X`	8
Tobit, Heckit	`Modèle > Variable dépendante censurée > Tobit, Heckit`	`tobit y X, heckit y X ; s Z`	8
Comptage	`Modèle > Variable dépendante censurée > Données de comptage`	`poisson y X, negbin y X`	8
Régression quantile	`Modèle > Estimation robuste > Régression de quantile`	`quantreg tau y X`	8
Racine unitaire	`Variable > Tests de racine unitaire`	`adf k y, kpss k y`	7
Corrélogramme	`Variable > Corrélogramme`	`corrgm y k`	7
ARIMA	`Modèle > Série temporelle univariée > ARIMA`	`arima p d q ; y`	7
GARCH	`Modèle > Série temporelle univariée > GARCH`	`garch p q ; y`	7
VAR	`Modèle > Séries temporelles multivariées > Vecteur Autorégressif, VAR`	`var p liste`	7
Cointégration	`Modèle > Séries temporelles multivariées > Test de cointégration de Johansen`	`johansen p liste`	7
VECM	`Modèle > Séries temporelles multivariées > Modèle à correction d'erreurs, VECM`	`vecm p r liste`	7
Réponses impulsionnelles	fenêtre du VAR	`irf(cible, choc), fevd(cible, choc)`	7
Différences-en-différences	`Modèle > Moindres Carrés Ordinaires` sur la variation	`ols dy const traitement`	15

Dans toutes les commandes, `y` est la variable dépendante, `X` la liste des régresseurs, `Z` la liste des instruments, `s` la variable de sélection, `p` et `q` les ordres AR et MA, `k` un nombre de retards.

Points clés à retenir

- Chaque action de menu correspond à une commande *hansl*, visible dans le journal des commandes : on explore par les menus, on documente par un script.
- Un script type ouvre les données, déclare leur structure (`setobs`), transforme (`series`), estime, teste et exporte (`tabprint`, `gnuplot --output`).
- Les résultats d'une estimation sont accessibles par `$coeff`, `$stderr`, `$rsq`, `$uhat` ; le R^2 ajusté se calcule à la main, car `$adjrsq` n'existe pas.
- Le test de Hausman n'est pas une commande : il s'affiche après `panel --random-effects`. Les tests des variables instrumentales (Hausman, Sargan, instruments faibles) sont affichés par `tsls`.
- Pour les modèles de choix, Gretl donne les pentes à la moyenne ; l'effet marginal moyen se calcule en trois lignes. Il n'y a pas de commandes `mfx`, `ologit` ou `lrtest`.
- Dans un VAR, `irf` et `fevd` sont des fonctions à arguments numériques ; l'ordre des variables fixe la décomposition de Cholesky.
- Tous les scripts de cette annexe sont reproductibles avec la version 2023c et les jeux de données livrés avec le logiciel.

B.15 Exercices

Exercice B.1 : Lire une sortie de régression

Les sorties du code B.11 et du code B.12 portent sur les 14 maisons de `data4-1`. (a) Interprétez la pente de `sqft` dans la régression simple, en dollars par pied carré. (b) Calculez un intervalle de confiance à 95 % pour cette pente (la valeur critique de la loi de Student à 12 degrés de liberté est 2,1788). (c) Prévoyez le prix d'une maison de 2 000 pieds carrés. (d) Pourquoi le R^2 ajusté diminue-t-il quand on passe au modèle multiple alors que le R^2 augmente ?

Résultats attendus. (a) Un pied carré de plus augmente le prix prévu de 138,75 dollars. (b) $0{,}13875 \pm 2{,}1788 \times 0{,}0187329$, soit [0,0979 ; 0,1796]. (c) $52{,}3509 + 0{,}13875 \times 2000 = 329{,}85$ milliers de dollars. (d) $\bar{R}^2$ passe de 0,8056 à 0,7868, car la baisse de la somme des carrés des résidus ne compense pas la perte de deux degrés de liberté ; `bedrms` et `baths` ne sont pas significatives.

Exercice B.2 : Déclarer la structure d'un fichier de panel

Un fichier CSV contient 26 provinces observées 14 années, trié par province puis par année, sans colonne d'identifiants. (a) Quelle commande `setobs` le déclare comme panel ? (b) Que changer si le fichier est trié par année puis par province ? (c) Quelle commande utiliser si le fichier contient les colonnes `province_id` et `annee` ? (d) Quelle étendue Gretl affiche-t-il, et pourquoi le nombre d'observations vaut-il 364 ?

Résultats attendus. (a) `setobs 14 1:1 --stacked-time-series` (14 périodes par bloc). (b) `setobs 26 1:1 --stacked-cross-section` (26 unités par bloc). (c) `setobs province_id`

`annee --panel-vars`. (d) L'étendue est `1:01 - 26:14`, soit $26 \times 14 = 364$ observations (vérifié avec `nulldata 364`).

Exercice B.3 : MCO empilés, effets fixes et effets aléatoires

Sur `grunfeld`, estimez `ols invest const value kstock` puis les effets fixes et aléatoires du code B.18. (a) Comparez les coefficients de `value` et de `kstock`. (b) Que montre la statistique de Durbin-Watson de la régression empilée ? (c) Quel estimateur le test de Hausman conduit-il à retenir ?

Résultats attendus. (a) MCO empilés : 0,1156 (`value`) et 0,2307 (`kstock`), constante −42,71, $R^2 = 0{,}812$; effets fixes : 0,1101 et 0,3101 ; effets aléatoires : 0,1098 et 0,3081. L'effet du capital est sous-estimé par les MCO empilés. (b) Durbin-Watson de 0,21 : forte autocorrélation, signe de l'oubli d'effets individuels. (c) $\chi^2(2) = 2{,}19$ ($p = 0{,}34$) : les effets aléatoires sont retenus.

Exercice B.4 : Racine unitaire du taux de chômage

Dans `greene5_1`, testez la stationnarité du taux de chômage `unemp` avec `adf 4 unemp --c --test-down=BIC` et `kpss -1 unemp`, puis de sa différence première `diff(unemp)`. Que concluez-vous ?

Résultats attendus. Pour `unemp`, l'ADF avec un retard retenu par le BIC donne $\tau = -3{,}61$ ($p = 0{,}006$) : rejet de la racine unitaire ; le KPSS donne 0,968, supérieur à la valeur critique de 0,739 à 1 % : rejet de la stationnarité. Les deux tests se contredisent, ce qui arrive pour des séries très persistantes mais peut-être stationnaires. Pour la différence première, ADF $\tau = -7{,}87$ ($p < 0{,}001$) et KPSS 0,057 (non significatif) : la différence est stationnaire. L'étude doit donc trancher à partir d'autres éléments (longueur de l'échantillon, théorie économique).

Exercice B.5 : Dépense alimentaire de huit ménages de Kinshasa

Huit ménages de Kinshasa ont un revenu mensuel (en dollars, valeurs d'énoncé purement illustratives) de 90, 110, 150, 180, 220, 260, 310 et 400 et une dépense alimentaire mensuelle de 70, 82, 105, 118, 140, 152, 175 et 205. (a) Écrivez un script qui crée ces deux séries avec `nulldata 8` et estime la dépense alimentaire en fonction du revenu. (b) Interprétez la pente. (c) Calculez l'élasticité-revenu de la dépense alimentaire aux moyennes. (d) Que dit la loi d'Engel de ce résultat ?

Résultats attendus. (a) Voir le code B.35 ci-dessous. (b) Pente 0,4364 ($t = 23{,}3$) : chaque dollar de revenu supplémentaire augmente la dépense alimentaire de 44 cents ; constante 37,06 ; $R^2 = 0{,}9891$. (c) Les moyennes sont 215 et 130,875, d'où une élasticité de $0{,}4364 \times 215/130{,}875 = 0{,}717$. (d) L'élasticité, inférieure à 1, est celle d'un bien de première nécessité : la part de l'alimentation dans le budget diminue quand le revenu augmente, conformément à la loi d'Engel.

Code B.35. Solution de l'exercice : estimation sur huit ménages

```
nulldata 8                      # jeu de données vide : 8 observations
series revenu = {90, 110, 150, 180, 220, 260, 310, 400}
series alim   = {70, 82, 105, 118, 140, 152, 175, 205}
ols alim const revenu --quiet
printf "pente = %.4f ; R2 = %.4f\n", $coeff(revenu), $rsq
```

```
pente = 0,4364 ; R2 = 0,9891
```

Références

Bollerslev, T., & Ghysels, E. (1996). Periodic autoregressive conditional heteroscedasticity. *Journal of Business & Economic Statistics*, *14*(2), 139–151.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, *47*(5), 1287–1294.

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge University Press.

Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*, *84*(4), 772–793.

Cottrell, A., & Lucchetti, R. (2023). *Gretl user's guide* (version 2023c). https://gretl.sourceforge.net

Greene, W. H. (2003). *Econometric analysis* (5e éd.). Prentice Hall.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press.

Kleiber, C., & Zeileis, A. (2008). *Applied econometrics with R*. Springer.

Kleiber, C., & Zeileis, A. (2010). The Grunfeld data at 50. *German Economic Review*, *11*(4), 404–417.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50.

Koenker, R., & Bassett, G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, *50*(1), 43–61.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, *54*(1–3), 159–178.

Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, *55*(4), 765–799.

Ramanathan, R. (2002). *Introductory econometrics with applications* (5e éd.). Harcourt.