Travaux Dirigés d'Econométrie Appliquée

Licence Analyse Economique

Professeur Georges Bresson

Université Paris II / Sorbonne Universités

TD 1 - Modèles à retards échelonnés
TD 2 - Modèles de régression non linéaires
TD 3 - Modèles binaires Logit et Probit
TD 4 - Modèles multinomiaux
TD 5 - Révision
TD 6 - Le modèle Tobit et les modèles de sélection
TD 7 - Les modèles de comptage
TD 8 - Données de panel: les modèles de base
TD 9 - Données de panel: variables instrumentales et GMM
TD 10 - Révision

TD 1 - Modèles à retards échelonnés

On considère des données macroéconomiques trimestrielles pour les USA sur la période 1950-2000 (fichier *"consumption_ USA.dta"*). On dispose, entre autres, des variables suivantes: *realcons* (real consumption expenditures, in billions US$), *realdpi* (real disposable personal income, in billions US$) et *pop* (population, in millions). On souhaite estimer la relation consommation par tête - revenu par tête. Soit $C_t = 1000(\text{real}\,cons/pop)_t$ et $R_t = 1000(\text{real}\,pi/pop)_t$ la consommation réelle par tête et le revenu réel par tête en US$.

1. Tracez sur un même graphique les courbes de $\log C_t$ et $\log R_t$ sur la période 1950q1-2000q4. Que constatez-vous?

2. On souhaite estimer un modèle à retards échelonnés $DL(p)$:

$$\log C_t = \alpha + \sum_{i=0}^{p} \beta_i \log R_{t-i} + \varepsilon_t$$

   A l'aide du critère AIC, déterminer l'ordre optimal p^* de p. On testera le modèle pour $p = 1, 2, 3, ..., 30$. Pourquoi un modèle DL(p) pour p fini est-il inapproprié? Quelle autre spécification suggérez-vous?

3. On souhaite estimer un modèle autorégressif à retards échelonnés $ARDL(p, q)$:

$$\log C_t = \alpha + \sum_{i=1}^{p} \gamma_i \log C_{t-i} + \sum_{j=0}^{q} \beta_j \log R_{t-j} + \varepsilon_t$$

   A l'aide du critère AIC, déterminer les ordres optimaux p^* de p et q^* de q. On testera le modèle pour $p = 1, 2, 3$ et $q = 0, 1, 2, 3$. Pourquoi ne peut-on pas uniquement se fier au critère AIC pour choisir la meilleure spécification?

4. On estime un $ARDL(1, 1)$. Faites le graphique des séries observée et estimée de $\log C_t$. Faites le graphique des résidus estimés. Que constatez-vous? Calculez le multiplicateur de long terme et les multiplicateurs dynamiques (cumulés) pour un horizon de 15 trimestres. Comment interprétez-vous ces multiplicateurs? Représentez graphiquement les multiplicateurs cumulés.

5. A partir de l'$ARDL(1,1)$, on estime un ECM:

$$\Delta \log C_t = \alpha_0 + \beta_0 \Delta \log R_t + \eta_1 \log C_{t-1} + \delta_1 \log R_{t-1}$$

Que concluez-vous ? On souhaite tenir compte des taux de croissance annuels et non plus trimestriels. Pour cela, on différencie saisonnièrement les séries: $\Delta_4 \log C_t$ $(= \log C_t - \log C_{t-4})$ et $\Delta_4 \log R_t$ $(= \log R_t - \log R_{t-4})$. On estime un nouvel ECM (ECM "*à la Hendry*")

$$\Delta_4 \log C_t = \alpha_0 + \beta_0 \Delta_4 \log R_t + \eta_1 \log C_{t-4} + \delta_1 \log R_{t-4}$$

Montrez que la propension marginale à consommer de long terme est unitaire. Combien faut-il de trimestres pour réaliser 100% de l'ajustement?

TD 2 - Modèles de régression non linéaires

On considère le nombre d'articles écrits sur Wikipédia (version anglaise) entre Janvier 2001 et Août 2010, soit 116 observations mensuelles (fichier *"Wikipedia.dta"*). On dispose de 2 séries $t = 1, ..., T(= 116)$ (série *obs*) et Y_t (série *count*), le cumul des articles publiés chaque mois depuis Janvier 2001.

1. Représentez graphiquement l'évolution du nombre d'articles publiés (Y_t) (échelle linéaire, puis échelle semi-log). Représentez graphiquement l'évolution du nombre de nouveaux articles publiés chaque mois ($\Delta Y_t = Y_t - Y_{t-1}$) (échelle linéaire). Qu'en déduisez-vous?

2. On souhaite spécifier et estimer différents modèles non linéaires de croissance (avec la commande Stata *nl*). Les spécifications retenues sont:
   * un modèle quadratique

$$\begin{aligned} Y_t &= b_0 + b_1.\left(\frac{t}{100}\right) + b_2.\left(\frac{t}{100}\right)^2 + \varepsilon_t \ , \ t = 1, ..., T \qquad (1) \\ &= b_0 + b_1.\tau + b_2.\tau^2 + \varepsilon_t \ , \ \tau = 0.1, ..., T/100 \end{aligned}$$

   * un modèle exponentiel à deux paramètres

$$Y_t = b_1.\,(b_2)^\tau + \varepsilon_t \qquad (2)$$

   avec comme valeurs initiales $b_1 = 10^6$, $b_2 = 1$ pour accélérer la convergence et 100 itérations maximum.
   *un modèle exponentiel à trois paramètres

$$Y_t = b_0 + b_1.\,(b_2)^\tau + \varepsilon_t \qquad (3)$$

   avec comme valeurs initiales $b_0 = 10^6$, $b_1 = 10^3$, $b_2 = 1$.
   *un modèle exponentiel négatif à trois paramètres

$$Y_t = b_0 + b_1.\,[1 - (b_2)^\tau] + \varepsilon_t \qquad (4)$$

   avec comme valeurs initiales $b_0 = 10^6$, $b_1 = 2$, $b_2 = 1$.
   *un modèle logistique à trois paramètres

$$Y_t = \frac{b_1}{1 + \exp\left(-b_2.\,(\tau - b_3)\right)} + \varepsilon_t \qquad (5)$$

avec comme valeurs initiales $b_1 = 10^6$, $b_2 = 2$, $b_3 = 1$.
*un modèle de Gompertz à trois paramètres

$$Y_t = b_1 \exp\left[-b_2.\left(-b_3.\tau\right)\right] + \varepsilon_t \tag{6}$$

avec comme valeurs initiales $b_1 = 10^6$, $b_2 = 10$, $b_3 = 10$.
Commentez les résultats.

3. Tracez sur un même graphique les séries observée Y_t et calculées $\widehat{Y}_t$ pour toutes les spécifications. Tracez sur un même graphique les séries observée Y_t et calculées $\widehat{Y}_t$ pour les spécifications logistique et de Gompertz. Que concluez-vous?

4. A l'aide de la commande Stata *kdensity*, tracez sur un même graphique les densités non paramétriques des séries observée Y_t et calculées $\widehat{Y}_t$ pour toutes les spécifications. Puis tracez sur un même graphique les densités non paramétriques des résidus estimés pour toutes les spécifications. Que concluez-vous?

TD 3 - Modèles binaires Logit et Probit

On dispose d'un échantillon de 753 observations relatives à des femmes mariées aux USA dont 428 ont choisi de travailler (fichier "*mroz.dta*"). On souhaite estimer un modèle de participation au marché du travail (modèle d'offre de travail):

$$inlf = f\left(educ,\ exper,\ expersq,\ age,\ kidslt6,\ kidsge6,\ huswage,\ constant\right)$$

où $inlf$ est la variable dichotomique (1/0) de participation (ou non) au marché du travail. $educ$ est le niveau d'éducation de la femme mariée (nombre d'années d'études), age, son âge, $exper$, son expérience professionnelle potentielle ($exper = age - educ - 6$), $expersq$, le carré de l'expérience, $kidslt6$, le nombre d'enfants de moins de 6 ans, $kidsge6$, le nombre d'enfants de 6 ans et plus, $huswage$, le salaire du mari en dollars 1975.

1. On estime le modèle à l'aide du Logit et du Probit. Donnez les résultats sous la forme d'un tableau unique avec les coefficients, les écarts-type et les t-stat. Interprétez les résultats.

2. Calculez les valeurs prédites par le Logit ainsi que le coefficient de corrélation entre les valeurs observées et prédites.

3. Pour le modèle Logit, calculez l'effet estimé sur la probabilité de participer au marché du travail si le niveau d'éducation passe de 12 ans (bac) à 17 ans (Master), puis si le nombre d'enfants en bas âge passe de 0 à 3, les autres variables étant prises à leurs valeurs moyennes.

4. Calculez la probabilité de participer au marché du travail ainsi que les effets marginaux pour une femme mariée sans enfant ayant 36 ans, un niveau d'éducation de 12 années d'études, une expérience professionnelle de 18 ans, un mari dont le salaire horaire est de 10 dollars de l'heure. Calculez la probabilité de participer au marché du travail et les effets marginaux pour une femme ayant les mêmes caractéristiques mais avec 3 enfants en bas âge. Que concluez-vous?

5. On introduit deux autres variables d'environnement: le taux de chômage du lieu de résidence ($unem$) et le taux marginal de la taxe fédérale pour le second salaire dans le ménage (mtr). Ré-estimez le modèle

Logit avec ces deux variables supplémentaires. Calculez les effets marginaux pour une femme mariée sans enfant ayant 36 ans, un niveau d'éducation de 12 années d'études, une expérience professionnelle de 18 ans, un mari dont le salaire horaire est de 10 dollars de l'heure, un taux de chômage de 5 points et un taux marginal de taxe fédérale de 0.5 point. Interprétez les résultats.

TD 4 - Modèles multinomiaux

On s'intéresse au choix de modes de transport entre Sydney et Melbourne, selon 4 modes de transport: avion, train, bus et voiture (fichier *"travel.dta"*). On a interrogé 210 groupes de voyageurs (de 1 à 6 voyageurs par groupe) et on dispose des variables suivantes: *mode* (choix du mode de transport (1/0) parmi les 4 modes de transport), *ttme* (temps d'attente (en min.) au terminal pour les avions, trains ou bus), *invc* (coût du transport (en dollars australiens)), *invt* (temps de transport (en min.)), *gc* (coût généralisé du transport (en dollars)), *hinc* (revenu du ménage (en milliers de dollars)), *psize* (nombre de voyageurs dans le groupe), *grp* (groupe) et *travel* (*air*, *train*, *bus*, *car* (selon le rang de '*mode*')). Le coût généralisé du transport est la somme du coût monétaire du transport et du produit entre le temps de transport et la valeur du temps de l'individu. Cette valeur du temps corrrespond à la désutilité marginale du temps. Donc $gc = invc + (invt *$valeur du temps épargnée).

1. Donnez les statistiques (moyenne et écart-type) du coût généralisé selon les 4 modes de transport. Donnez les moyennes des variables *gc*, *psize*, *invt*, *ttme* et *hinc* selon les 4 modes de transport.

2. Soit le modèle Logit multinomial suivant:

$$p\left(travel_j\right) = f\left(gc, ttme, psize, hinc\right) \ , \ j = 1, 2, 3, 4$$

   (a) On estime ce modèle avec *bus* comme catégorie de référence. Interprétez.

   (b) Faites un test joint de type Wald de nullité du coût généralisé. Interprétez.

   (c) Déterminez les risques relatifs (*odds ratios*) de ce modèle. Interprétez.

   (d) Calculez les probabilités prédites du choix de chaque mode de transport et comparez avec les fréquences observées.

   (e) Calculez les effets marginaux et les élasticités de chaque régresseur sur la probabilité de choix des 2 modes de transport *avion* et *train*. Interprétez les résultats.

3. On souhaite estimer un modèle Logit emboîté dégénéré. A partir de la variable $travel$, on définit un arbre de décision à deux membres (*limbs*): fly et $ground$ ($j = 1, 2$). Le membre fly n'a qu'une branche (*branch*): air ($k = 1$) tandis que le membre $ground$ en a trois: $train$, bus et car ($k = 1, 2, 3$). On suppose que la catégorie de référence au niveau du membre (resp. de la branche) est $ground$ (resp. car). Les 2 régresseurs spécifiques aux alternatives (*alternative-specific variables*) sont $X_{jk} = [gc, ttme]$. Les 2 régresseurs invariants aux alternatives (*cases-specific variables*) sont $Z_j = [hinc, psize]$. On introduit également des constantes pour chaque alternative des branches. Donc $p\,(travel_{jk}) = p_j \times p_{k|j}$, la probabilité jointe de l'alternative (jk) (*i.e.*, d'être sur la branche k du membre j) est le produit de la probabilité de choisir le membre j par la probabilité de choisir la branche k sachant que l'on est sur le membre j. C'est donc un modèle dégénére (ou contraint) dans lequel le paramètre de dissimilarité du choix fly vaut 1.

   (a) Donnez la structure de l'arbre et la répartition des groupes selon les alternatives à chaque niveau (fly, $ground$) et ($air, train, bus, car$).

   (b) Estimez le modèle Logit emboîté dégénéré. Quelle est la valeur du paramètre de dissimilarité du membre $ground$. L'hypothèse d'indépendance des choix non pertinents (non retenus) (IIA) est-elle vérifiée? Interprétez les résultats.

   (c) Estimez, pour chaque groupe de voyageurs, les probabilités p_j de chaque membre et les probabilités conditionnnelles $p_{k|j}$ de chaque branche et comparez avec les fréquences observées. Affichez ces probabilités pour les groupes 1 à 3, puis 16 à 18, puis 41 à 43 et 176 à 178.

TD 6 - Le modèle Tobit et les modèles de sélection

On souhaite répliquer les résultats de la célèbre étude de Ray Fair (1978), A Theory of Extramarital Affairs, *Journal of Political Economy*, 86, 1, 45-61. Dans cet article, l'auteur développe un modèle d'allocation du temps des femmes mariées, temps consacré au conjoint et temps consacré à l'amant. Les données (fichier "*Fair.dta*") concernent les réponses à un questionnaire de 6366 femmes interrogées en 1974. La variable expliquée $extraconjug\_time$ est une mesure du temps consacré aux relations extraconjugales. C'est le ratio $partners\_relat/yrs\_married$. La variable $partners\_relat$ est le produit du nombre de partenaires par le nombre de relations et la variable $yrs\_married$ est le nombre d'années de mariage. Ces deux variables ($extraconjug\_time$ et $partners\_relat$) sont des variables censurées comprises entre $[0; 57.6]$ et $[0; 144]$ respectivement.

Les variables explicatives du modèle sont:

- $occupation$, variable codée de 1 à 6 (1: student, 2: unskilled, 3: white collar (secretarial), 4: teacher, social worker, artist, ..., 5: managerial, business, 6: professional with advanced degree),

- $education$ variable codée de 9 à 20 (9: grade school, 12: high school, 14: some college, 16: college graduate, 17: some graduate school, 20: advanced degree),

- $husb\_occupation$, l'activité professionnelle du mari a la même codification que la variable $occupation$,

- $marital\_happiness$, variable codée de 1 à 5 (1: very poor, 2: poor, 3: fair, 4: good, 5: very good),

- age,

- $yrs\_married$, nombre d'années de mariage,

- $numb\_children$ nombre d'enfants (0, 1, 2, 3, 4 et 5.5 pour 5 enfants et plus),

- $religiosity$, variable codée de 1 à 4 (1: not, 2: mildly, 3: fairly, 4: strongly).

1. Calculez les statistiques descriptives de toutes les variables de la base de données. Pour la variable $extraconjug\_time$, donnez en plus les coefficients de kurtosis et d'asymétrie ainsi que les centiles (de 10 en 10). Tracez son histogramme. Interprétez.

2. Donnez les statistiques descriptives de l'ensemble des variables pour le groupe des épouses ayant des relations extraconjugales. A l'aide de

tests de différences de moyennes pour toutes les variables (autres que $extraconjug\_time$ et $partners\_relat$) entre les deux groupes (sans relation et ayant des relations extraconjugales), pouvez-vous brosser un "portrait type" (âge, éducation, années de mariage, ...) de la femme mariée ayant des relations extraconjugales?

3. Estimez le modèle Tobit

$$\begin{aligned} y\,(= extraconjug\_time) &= f(X) \\ &= f(occupation, education, husb\_occupation, \\ &\quad marital\_happiness, age, yrs\_married, \\ &\quad numb\_children, religiosity) \end{aligned}$$

   (a) Commentez et calculez les effets marginaux sur la variable tronquée $E\left[y \mid X, y > 0\right]$.

4. Sur la variable $y = partners\_relat$, on introduit la transformation logarithmique suivante:

$$\ln\_y = \begin{cases} -1.10^{-5} & \text{si} \quad y = 0 \\ \ln(y) & \text{si} \quad y > 0 \end{cases}$$

   (a) La distribution de $\ln\_y$ est-elle gaussienne pour $y > 0$?

   (b) On spécifie un modèle "*Two-Part*". La première équation (équation de participation) du modèle "*Two-Part*" est une equation Probit $Pr[y > 0]$. La seconde équation (équation de résultat) du modèle "*Two-Part*" est une équation de régression linéaire qui modélise $E[ln(y) \mid y > 0]$. Les régresseurs sont les mêmes X qu'à la question 3. Estimez le modèle et calculez la log-vraisemblance du modèle. Tester les hypothèses d'homoscédasticité et de normalité des résidus du modèle.

   (c) On suppose que les erreurs homoscédastiques sont corrélées et suivent une distribution normale jointe (modèle de sélection "*à la Heckman*"). Estimez le modèle à l'aide du maximum de vraisemblance. Interprétez et comparez les résultats avec ceux du modèle "*Two-Part*".

TD 7 - Les modèles de comptage

On considère la base de données relative au nombre d'articles publiés par 915 étudiants en thèse de biochimie (fichier "*articles_PhD.dta*"). Les variables de la base sont:
- art, nombre d'articles publiés dans les 3 dernières années de thèse,
- fem, le genre (femme = 1, homme=0)
- mar, le statut marital (marié(e)=1, 0 sinon)
- $kid5$, codé 1 si présence d'enfants de moins de 6 ans, 0 sinon
- phd, prestige du programme de PhD
- $ment$, nombre d'articles publiés par le mentor (directeur de thèse) dans les 3 dernières années.

1. Calculez les statistiques descriptives de toutes les variables et commentez-les. Donnez le tableau des fréquences et des fréquences cumulées du nombre d'articles (art) publiés par les étudiants en thèse.

2. Tracez l'histogramme de la distribution du nombre d'articles publiés par les étudiants en thèse ainsi que la densité non paramétrique estimée. Commentez.

3. On souhaite estimer un modèle de Poisson du nombre d'articles publiés par les étudiants en thèse. La spécification retenue est:

$$art = f(X) = f(fem, mar, kid5, phd, ment)$$

   (a) Estimez le modèle par le maximum de vraisemblance puis par le pseudo-maximum de vraisemblance. Interprétez les résultats.

   (b) Tester l'hypothèse de sur-dispersion. Que concluez-vous?

   (c) Calculez les effets marginaux au point moyen de l'échantillon (on pourra utiliser la commande Stata *margeff*).

4. On souhaite estimer un modèle binomial négatif sur la même spécification.

   (a) Estimez le modèle et interprétez les résultats. Comparez-les avec ceux du modèle de Poisson.

(b) Calculez les probabilités prédites pour la publication de 0, 1, 2, ..., 9, 10 articles pour le modèle de Poisson et pour le modèle négatif binomial (on pourra utiliser la commande Stata *countfit*). Tracez les graphiques des résidus ($art - \widehat{art}$) et comparez les résultats issus des deux modèles. Que concluez-vous?

(c) Calculez les probabilités prédites pour la publication de 0, 1, 2, ..., 9, 10 articles pour le modèle négatif binomial pour une femme mariée ayant des enfants de moins de 6 ans puis pour un homme célibataire et sans enfant. Comparez les résultats. Que concluez-vous?

TD 8 - Données de panel: les modèles de base

On souhaite estimer une fonction de gains "*à la Mincer*". On dispose de données relatives à 595 individus suivis sur la période 1976-1982, données issues du fichier américan PSID (*Panel Study of Income Dynamics*) (fichier "*psid_wages.dta*"). La fonction de gains (*earnings equation*) est définie par la relation log-linéaire:

$$\ln w_{it} = f(X_{it}) + u_{it} = f\left(\exp_{it}, \exp 2_{it}, wks_{it}, ed_i\right) + u_{it},\ i = 1, .., N(= 595), t = 1, ..., T(= 7)$$

où w_{it} est le salaire mensuel (en dollars) de l'individu i à la date t, ed est le niveau d'éducation (nombre d'années d'études) de cet individu, exp est son expérience professionnelle potentielle ($exp = age - ed - 6$), $exp2$ est le carré de exp et wks est le nombre de semaines travaillées dans l'année. La base de données contient bien d'autres variables.

1. Statistiques descriptives.

   (a) Donnez les statistiques générales de ces variables (moyenne, écart-type, ...), les variabilités totale, *within* et *between* ainsi que la matrice de corrélation de ces 5 variables. Commentez les résultats.

   (b) Calculez les corrélations entre $\ln w_{it}$, $\ln w_{it-1}$, $\exp_{it}$ et $\exp_{it-1}$. Commentez les résultats.

   (c) Tracez le nuage de points ($\ln w_{it}$, $\exp_{it}$) et sa courbe d'ajustement. Tracez le nuage de points des transformations *within* de ces 2 variables ainsi que sa courbe d'ajustement (on pourra utiliser la commade Stata *xtdata*). Commentez.

2. On suppose la présence d'effets spécifiques $u_{it} = \alpha_i + \varepsilon_{it}$ dans la spécification précédente. Estimez le modèle à l'aide des estimateurs MCO, *Within*, *Between* et *FGLS* (effets aléatoires). Affichez les résultats dans un tableau unique avec les coefficients, écarts-type, t-stat, R^2, variances des effets individuels et des perturbations, ... Commentez les résultats.

3. Pour comparer les estimateurs *Within* et *FGLS*, calculez la statistique d'Hausman. Que concluez-vous?

4. Dans l'estimation *Within*, on calcule les valeurs estimées ($\widehat{\ln w_{it}} = X_{it}\widehat{\beta} + \widehat{\alpha}_i$) ainsi que les erreurs $\widehat{\varepsilon}_{it} = \ln w_{it} - \widehat{\ln w_{it}}$. Comparez les valeurs observées et estimées (moyennes, écarts-types, ...) et calculer le coefficient de corrélation entre $\ln w_{it}$ et $\widehat{\ln w_{it}}$. Commentez les résultats.

5. A l'aide des MCO, on estime le modèle en différences premières: $\Delta \ln w_{it} = f(\Delta \exp_{it}, \Delta \exp 2_{it}, \Delta wks_{it}) + \Delta\varepsilon_{it}$ où $\Delta \ln w_{it} = \ln w_{it} - \ln w_{it-1}$. Que concluez-vous?

TD 9 - Données de panel: variables instrumentales et GMM

On souhaite estimer une fonction de consommation de cigarettes relative à 46 états américains sur la période 1963-1992 (ficher "*cigarette.dta*"). La spécification retenue est dynamique:

$$\begin{aligned}
\ln C_{it} &= f(X_{it}) + u_{it},\ i = 1, .., N(=46), t = 1, ..., T(=30) \\
&= \beta_0 + \beta_1 \ln C_{it-1} + \beta_2 \ln P_{it} + \beta_3 \ln Pn_{it} + \beta_4 \ln Y_{it} + u_{it} \quad (7) \\
\text{avec } u_{it} &= \mu_i + \lambda_t + \varepsilon_{it}
\end{aligned}$$

où C_{it} est la consommation réelle par tête de cigarettes pour des individus âgés de 14 ans et plus. Cette consommation est mesurée en nombre de paquets de cigarettes par individu, P_{it} est le prix moyen au détail d'un paquet de cigarettes en termes réels, Pn_{it} est le prix minimum d'un paquet de cigarettes en termes réels dans un des états voisins. C'est une variable *proxy* de la contrebande occasionnelle. Y_{it} est le revenu disponible brut réel par tête. Dans la base de données, les variables $\ln C_{it}$, $\ln P_{it}$, $\ln Pn_{it}$ et $\ln Y_{it}$ sont nommées *logc*, *logp*, *logpn* et *logy*. Le terme de perturbations u_{it} est spécifié comme un modèle général à erreurs composés (*two-way error component*). μ_i désigne l'effet spécifique de chaque état américain et λ_t désigne un effet spécifique annuel. On suppose que λ_t est un paramètre fixe que l'on peut estimer à l'aide d'une variable *dummy* temporelle pour chaque année.

1. Construisez les 30 dummies ($dum1$, $dum2$, ..., $dum30$) associées à λ_1, λ_2, ...,λ_{30}.

2. On considère que la variable retardée $\ln C_{it-1}$ est endogène.

   (a) Estimez le modèle à l'aide des doubles moindres carrés ($2SLS$) sans tenir compte des effets spécifiques individuels μ_i. On utilisera comme instruments de $\ln C_{it-1}$ les variables "exogènes" retardées $\ln P_{it-1}$, $\ln Pn_{it-1}$ et $\ln Y_{it-1}$.

   (b) Estimez le modèle à l'aide de l'estimateur Within (FE) sans tenir compte de l'endogénéité de $\ln C_{it-1}$.

   (c) Estimez le modèle à l'aide des doubles moindres carrés Within ($FE - 2SLS$) qui tient compte des effets spécifiques individuels μ_i. Commentez et comparez les trois résultats d'estimation.

3. On écrit le modèle en différences premières (FD):

$$\begin{aligned} \Delta \ln C_{it} &= \ln C_{it} - \ln C_{it-1} \qquad (8) \\ &= \beta_1 \Delta \ln C_{it-1} + \beta_2 \Delta \ln P_{it} + \beta_3 \Delta \ln Pn_{it} + \beta_4 \Delta \ln Y_{it} + \Delta \varepsilon_{it} \end{aligned}$$

   Estimez le modèle à l'aide des doubles moindres carrés ($FD - 2SLS$). On utilisera comme instruments de $\Delta \ln C_{it-1}$ les variables "exogènes" retardées $\Delta \ln P_{it-1}$, $\Delta \ln Pn_{it-1}$ et $\Delta \ln Y_{it-1}$. Pourquoi cette estimation n'est-elle pas satisfaisante?

4. Estimez l'équation (8) à l'aide de l'estimateur GMM à deux étapes d'Arellano et Bond avec matrice de variance-covariance robuste (on pourra utiliser la commande Stata *xtdpd*). Calculez les autocorrélations des résidus du 1er et 2ème ordre (test d'Arellano et Bond). Calculez la statistique de Sargan de test de l'hypothèse nulle de validité des restrictions de sur-identification. Commentez les résultats.

5. On restreint l'ensemble des dummies temporelles à la série ($dum3$, $dum8$, $dum10$, ..., $dum29$). Ré-estimez le modèle avec l'estimateur des GMM à deux étapes. Re-calculez les autocorrélations des résidus du 1er et 2ème ordre. Re-calculez la statistique de Sargan. Que concluez-vous?