Travaux Dirigés d'Econométrie

Licence Analyse Economique

Professeur Georges Bresson

Université Paris II / Sorbonne Universités

TD 1 et 2 - Le modèle de régression simple

1. Soit le modèle de régression simple:

$$Y_i = \alpha + \beta X_i + u_i \ , \ i = 1, ..., n$$

Vérifiez les propriétés numériques suivantes des estimateurs des Moindres Carrés Ordinaires (MCO)

$$\begin{aligned} \sum_{i=1}^{n} e_i &= 0 \ , \ \sum_{i=1}^{n} e_i X_i = 0 \ , \ \sum_{i=1}^{n} e_i \widehat{Y}_i = 0, \\ \sum_{i=1}^{n} \widehat{Y}_i &= \sum_{i=1}^{n} Y_i \text{ avec } e_i = \widehat{u}_i = Y_i - \widehat{Y}_i \end{aligned}$$

2. En utilisant l'équation de la question 1),

   (a) Montrez que $\widehat{\alpha}_{MCO} = \alpha + \left(\beta - \widehat{\beta}_{MCO}\right) \overline{X} + \overline{u}$. En déduire que: $E\left(\widehat{\alpha}_{MCO}\right) = \alpha$.

   (b) Sachant que: $\left(\widehat{\beta}_{MCO} - \beta\right) = \frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2}$ où $x_i = X_i - \overline{X}$, utilisez le résultat précédent pour montrer que:

$$Var\left(\widehat{\alpha}_{MCO}\right) = \sigma^2 \left[\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n} x_i^2}\right] = \sigma^2 \frac{\sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n} x_i^2}$$

   (c) Montrez que $\widehat{\alpha}_{MCO}$ est un estimateur convergent de α.

   (d) Montrez que $Cov\left(\widehat{\alpha}_{MCO}, \widehat{\beta}_{MCO}\right) = -\overline{X}.Var\left(\widehat{\beta}_{MCO}\right) = -\sigma^2 \overline{X} / \sum_{i=1}^{n} x_i^2$. Que concluez-vous si $\overline{X} > 0$?

3. Sachant que le coefficient de détermination est donné par:

$$R^2 = \frac{\sum_{i=1}^{n} \widehat{y}_i^2}{\sum_{i=1}^{n} y_i^2} \text{ où } \widehat{y}_i = \widehat{\beta}_{MCO} x_i \text{ et } \widehat{\beta}_{MCO} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

(a) Montrez que $R^2 = r_{xy}^2$ où r_{xy}^2 est le carré du coefficient de corrélation entre x_i et y_i

$$r_{xy}^2 = \frac{\left(\sum_{i=1}^{n} x_i y_i\right)^2}{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right)}$$

(b) Comme $y_i = \widehat{y}_i + e_i$, montrez que:

$$\sum_{i=1}^{n} \widehat{y}_i y_i = \sum_{i=1}^{n} \widehat{y}_i^2$$

et que

$$r_{y\widehat{y}}^2 = \frac{\left(\sum_{i=1}^{n} \widehat{y}_i y_i\right)^2}{\left(\sum_{i=1}^{n} \widehat{y}_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right)} = R^2$$

4. On souhaite prévoir Y_0 à partir de l'observation X_0. La nouvelle observation Y_0 est générée à partir de la relation: $Y_0 = \alpha + \beta X_0 + u_0$.

(a) Montrez que: $E(Y_0) = \alpha + \beta X_0$ et $\widehat{Y}_0$ est une estimation sans biais de $E(Y_0)$.

(b) Montrez que:

$$Var\left(\widehat{Y}_0\right) = Var\left(\widehat{\alpha}_{MCO}\right) + X_0^2 Var\left(\widehat{\beta}_{MCO}\right) + 2X_0 Cov\left(\widehat{\alpha}_{MCO}, \widehat{\beta}_{MCO}\right)$$

En déduire que:

$$Var\left(\widehat{Y}_0\right) = \sigma^2\left[\frac{1}{n} + \frac{\left(X_0 - \overline{X}\right)^2}{\sum_{i=1}^{n} x_i^2}\right]$$

(c) On considère un prédicteur linéaire de $E(Y_0)$, noté $\widetilde{Y}_0 = \sum_{i=1}^{n} a_i Y_i$.

Montrez que $\sum_{i=1}^{n} a_i = 1$ et $\sum_{i=1}^{n} a_i X_i = X_0$ pour que $\widetilde{Y}_0$ soit un prédicteur linéaire sans biais de $E(Y_0)$.

5. Dans le fichier TD1.XLS, on considère des données énergétiques relatives à 20 pays européens pour l'année 1980. On dispose pour chaque pays de la consommation d'énergie EN (exprimée en milliers de tonnes équivalent charbon) et du produit intérieur brut réel GDP (en millions de dollars 1975). On souhaite estimer des relations entre la consommation énérgétique et la production de chaque pays.

   (a) Estimer par les MCO la relation:

   $$\log(EN) = \alpha + \beta \log(GDP) + u$$

   (b) Interprétez les coefficients estimés. Quelle est la valeur du R^2?

   (c) Tracez le graphique des résidus. Que montre-t-il?

   (d) Testez l'hypothèse nulle: $H_0 : \beta = 1$.

   (e) On suppose que la valeur de la consommation énergétique de l'Allemagne de l'Ouest (352.677) est aberrante et qu'elle aurait due être multipliée par 1000, soit 352677. Recommencez l'estimation et les questions a) à d). Qu'en déduisez-vous?

   (f) Est-il légitime d'inverser la relation fonctionnelle $\log(GDP) = \gamma + \delta \log(EN) + \varepsilon$?

      i. Economiquement, cela change-t-il l'interprétation de l'équation? Expliquez.

      ii. Estimez cette nouvelle relation et comparez le R^2 avec le précédent. Vérifiez si $\widehat{\delta} = 1/\widehat{\beta}$. Pourquoi sont-ils différents?

      iii. Statistiquement, en inversant la relation, quelles sont les hypothèses classiques que l'on ne respecte plus?

      iv. Montrez que $\widehat{\delta}\widehat{\beta} = R^2$.

   (g) A la place du modèle log-linéaire $\log(EN) = \alpha + \beta \log(GDP) + u$, on teste maintenant un modèle linéaire: $EN = \alpha_0 + \beta_0 GDP + u$. Calculez les élasticités pour Malte et pour l'Allemagne de l'Ouest. Sont-elles différentes l'une de l'autre et de celle du modèle log-linéaire?

TD 3 et 4 - Le modèle de régression multiple

1. Soit la régression multiple:

$$Y_i = a + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \ , \ i = 1, ..., n$$

et les régressions auxiliaires

$$\begin{aligned} X_{2i} &= \widehat{\alpha} + \widehat{b} X_{3i} + \widehat{\nu}_{2i} \\ X_{3i} &= \widehat{c} + \widehat{d} X_{2i} + \widehat{\nu}_{3i} \end{aligned}$$

On sait que l'estimateur des MCO $\widehat{\beta}_2$ de β_2 peut être interprété comme celui d'une régression simple de Y_i sur les résidus MCO $\widehat{\nu}_{2i}$. De même pour l'estimateur des MCO $\widehat{\beta}_3$ de β_3 qui peut être interprété comme celui d'une régression simple de Y_i sur les résidus MCO $\widehat{\nu}_{3i}$. Mais $\widehat{\beta}_2$ n'est pas nécessairement équivalent à $\widehat{\delta}_2$, l'estimateur des MCO de δ_2 obtenu à partir de la régression:

$$Y_i = \gamma + \delta_2 \widehat{\nu}_{2i} + \delta_3 \widehat{\nu}_{3i} + w_i \ , \ i = 1, ..., n$$

Prouvez cette assertion en déterminant une relation entre les β et les δ.

2. Soit la régression multiple à K_1 régresseurs:

$$Y_i = \alpha + \sum_{k=2}^{K_1 - 1} \beta_k X_{ki} + u_i \ , \ i = 1, ..., n$$

et soit SCR_1 la somme des carrés des résidus de l'estimation MCO. Ajoutons K_1 régresseurs, de telle sorte que le nombre total de régresseurs soit: $K = K_1 + K_2$ et soit SCR_2 la somme des carrés des résidus de l'estimation MCO de cette nouvelle régression.

   (a) Montrez que $SCR_2 \leq SCR_1$ et que les R^2 correspondants satisfont l'inégalité: $R_2^2 \geq R_1^2$.

   (b) Montrez que

$$\left(1 - \overline{R}_2^2\right) = \left(1 - R_2^2\right) \frac{n-1}{n-K}$$

3. Soit la régression:

$$Y_t = \alpha + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \ , \ t = 1, ..., T$$

et soit $\widehat{u}_t$ les résidus MCO estimés et $\widehat{Y}_t$ les valeurs estimées de Y_t.

   (a) Déterminez le R^2 de la régression de $\widehat{u}_t$ sur une constante, X_{2t} et X_{3t}.

   (b) Si on régresse Y_t sur une constante et sur $\widehat{Y}_t$, quels sont les estimateurs de l'ordonnée à l'origine et de la pente? Quelle est la relation entre le R^2 de cette régression et celui de la régression précédente?

   (c) Si on régresse Y_t sur une constante et sur $\widehat{u}_t$, quels sont les estimateurs de l'ordonnée à l'origine et de la pente? Quelle est la relation entre le R^2 de cette régression et celui de la régression initiale?

4. Soit la régression multiple à K_1 (> 6) régresseurs:

$$Y_i = \alpha + \sum_{k=2}^{K_1 - 1} \beta_k X_{ki} + u_i \ , \ i = 1, ..., n$$

Déterminez la statistique F de Fisher pour les tests d'hypothèses nulles:

$$\begin{aligned} H_0 &: \beta_2 = \beta_4 = \beta_6 \\ H_0 &: \beta_2 = -\beta_3 \text{ et } \beta_5 - \beta_6 = 1 \end{aligned}$$

5. On dispose d'une base de données relatives aux rémunérations de 595 individus en 1982 aux USA. Cette base de données est issue du "*Panel Study of Income Dynamics*". Les variables sont:
(1) EXP = années d'expérience d'emploi à plein-temps
(2) EXP2 = carré des années d'expérience.
(3) WKS = nombre de semaines travaillées.
(4) OCC = (OCC=1, si l'individu a un emploi d'ouvrier (*blue-collar*)).
(5) IND = (IND=1, si l'individu travaille dans une industrie manufacturière).
(6) SOUTH = (SOUTH=1, si l'individu réside dans le Sud des USA).
(7) SMSA = (SMSA=1, si l'individu réside dans une zone urbaine (*standard metropolitan statistical area*)).

(8) MS = (MS=1, si l'individu est marié).
(9) FEM = (FEM=1, si l'individu est une femme).
(10) UNION = (UNION=1, si le salaire de l'individu est défini par un contrat syndical).
(11) ED = années d'études.
(12) BLK = (BLK=1, si l'individu est noir).
(13) LWAGE = Logarithme du salaire.
(14) M = (M=1, si l'individu est un homme).
(15) F_EDC = années d'études (uniquement pour les femmes).

(a) Estimez le modèle:

$$LWAGE_i = \alpha_F FEM_i + \alpha_M M_i + u_i \ , \ i = 1, ..., n$$

  i. Montrez que $\alpha_F = \overline{LWAGE_{F,i}}$ est le gain moyen des femmes et que $\alpha_M = \overline{LWAGE_{M,i}}$ est le gain moyen des hommes.
  ii. Soit la régression: $LWAGE_i = \alpha + \beta FEM_i + u_i$, montrez que $\widehat{\alpha} = \widehat{\alpha}_M$ et que $\widehat{\beta} = \widehat{\alpha}_F - \widehat{\alpha}_M$.
  iii. Substituez $M_i = 1 - FEM_i$ et montrez que $\alpha = \alpha_M$ et que $\beta = \alpha_F - \alpha_M$.

(b) Pour le modèle:

$$LWAGE_i = \alpha + \beta_F FEM_i + \beta_B BLK_i + u_i \ , \ i = 1, ..., n$$

  i. Montrez que $E[LWAGE_i / BlackFemale] = \alpha + \beta_F + \beta_B$ et que $E[LWAGE_i / BlackMale] = \alpha + \beta_B$.
  ii. En conclure que $\beta_F = E[LWAGE_i / BlackFemale] - E[LWAGE_i / BlackMale]$.

(c) Soit le modèle:

$$LWAGE_i = \alpha + \alpha_F FEM_i + \beta . ED_i + \gamma \left(FEM \times ED \right)_i \ + u_i, \ i = 1, ..., n$$

  i. Ecrivez le test F de Fisher et estimez les régressions non contraintes et contraintes pour les hypothèses suivantes:
    A. Les coefficients de pente et d'ordonnées à l'origine pour les hommes et les femmes sont les mêmes ($H_0 : \alpha_F = \gamma = 0$).
    B. Il y a égalité des ordonnées à l'origine sachant que les coefficients de pente des hommes et des femmes sont les mêmes ($H_0 : \alpha_F = 0 \mid \gamma = 0$)

C. Il y a égalité des ordonnées à l'origine sachant que les coefficients de pente des hommes et des femmes sont différents ($H_0 : \alpha_F = 0 \mid \gamma \neq 0$)

(d) Soit le modèle général:

$$\begin{aligned} LWAGE_i &= \alpha + \alpha_e EXP_i + \alpha_{e2} EXP_i^2 + \alpha_w WKS_i + \alpha_o OCC_i \\ &\quad + \alpha_i IND_i + \alpha_s SOUTH_i \alpha_{sm} SMSA_i + \alpha_{ms} MS_i \\ &\quad + \alpha_F FEM_i + \alpha_U UNION_i + \alpha_{ed} ED_i + \alpha_B BLK_i + u_i \end{aligned}$$

i. Estimez ce modèle par les MCO.
ii. Testez l'hypothèse nulle: $H_0 : LWAGE_i = \alpha + u_i$
iii. Testez l'hypothèse nulle: $H_0 : \alpha_F = 0$ et $\alpha_B = 0$
iv. Testez l'hypothèse nulle: $H_0 : \alpha_{ms} = 0$ et $\alpha_U = 0$.

TD 5 - Hétéroscédasticité

1. Soit le modèle de régression simple:

$$Y_i = \alpha + \beta X_i + u_i \ , \ i = 1, ..., n \tag{1}$$

On suppose la présence d'hétéroscédasticité $E\left(u_i^2\right) = \sigma_i^2$. Montrez que $E\left(s^2\right)$ est une fonction de σ_i^2.

2. Pour le modèle de régression linéaire simple avec une hétéroscédasticité de la forme: $E\left(u_i^2\right) = \sigma_i^2 = bx_i^2$ où $b > 0$, montrez que $E\left(s^2 / \sum\limits_{i=1}^{n} x_i^2\right)$ sous-estime la variance de $\widehat{\beta}_{MCO}$ qui est: $\left(\left(\sum\limits_{i=1}^{n} \sigma_i^2 X_i^2\right) / \left(\sum\limits_{i=1}^{n} x_i^2\right)^2\right)$

3. Soit le modèle:

$$Y_i \left(\sigma/\sigma_i\right) = \alpha \left(\sigma/\sigma_i\right) + \beta \left(\sigma/\sigma_i\right) X_i + u_i^* \ , \ i = 1, ..., n \text{ avec } u_i^* = u_i \left(\sigma/\sigma_i\right) \tag{2}$$

   (a) Résoudre les équations normales et montrez que:

$$\begin{aligned} \widetilde{\alpha} &= \overline{Y}^* - \widetilde{\beta}\overline{X}^* \\ \widetilde{\beta} &= \frac{\sum\limits_{i=1}^{n} w_i^* \left(X_i - \overline{X}^*\right)\left(Y_i - \overline{Y}^*\right)}{\sum\limits_{i=1}^{n} w_i^* \left(X_i - \overline{X}^*\right)^2} \end{aligned}$$

   avec

$$w_i^* = \frac{1}{\sigma_i^2} \ , \ \overline{X}^* = \frac{\sum\limits_{i=1}^{n} w_i^* X_i}{\sum\limits_{i=1}^{n} w_i^*} \text{ et } \overline{Y}^* = \frac{\sum\limits_{i=1}^{n} w_i^* Y_i}{\sum\limits_{i=1}^{n} w_i^*}$$

   (b) Montrez que:

$$Var\left(\widetilde{\beta}\right) = \frac{1}{\sum\limits_{i=1}^{n} w_i^* \left(X_i - \overline{X}^*\right)^2}$$

4. On considère le modèle linéaire simple avec une hétéroscédasticité de la forme: $\sigma_i^2 = \sigma^2 X_i^\delta$ où $X_i = 1, 2, ..., 10$.

   (a) Calculez $Var\left(\widehat{\beta}_{MCO}\right)$ pour $\delta = 0.5, 1, 1.5$ et 2.

   (b) Calculez $Var\left(\widetilde{\beta}_{BLUE}\right)$ pour $\delta = 0.5, 1, 1.5$ et 2.

   (c) Déterminez l'efficacité relative de l'estimateur des MCO sous l'hypothèse d'hétéroscédasticité:
   $$\frac{Var\left(\widetilde{\beta}_{BLUE}\right)}{Var\left(\widehat{\beta}_{MCO}\right)}$$
   pour $\delta = 0.5, 1, 1.5$ et 2. Que se passe-t-il quand δ augmente?

5. Dans le fichier Cigarettes.txt, on considère des données de consommation de cigarettes dans 46 états des USA pour l'année 1992. On souhaite estimer la fonction de demande:
   $$LNC_i = \alpha + \beta LNY_i + \gamma LNP_i + u_i \ , \ i = 1, ..., 46$$
   où LNC_i (resp. LNY_i et LNP_i) est le logarithme de la consommation de cigarettes (en paquets) par tête dans l'état (i) pour des individus agés de plus de 16 ans (resp.le logaritme du revenu disponible par tête et le logarithme du prix relatif des cigarettes).

   (a) Estimez par les MCO cette relation.

   (b) Tracez le graphique des résidus par rapport à LNY_i.

   (c) Effectuez le test de Glejser en régressant la valeur absolue des résidus MCO $|e_i|$ sur $(LNY_i)^\delta$ avec $\delta = 1, -1, -0.5$ et 0.5. Que concluez-vous?

   (d) Effectuez le test de Goldfeld et Quandt en classant les observations en fonction de LNY_i et en omettant les 12 valeurs centrales de l'échantillon ainsi classé. Faites les deux régressions basées sur les 17 premières et dernières observations. Que concluez-vous?

   (e) Déterminez le coefficient de corrélation des rangs de Spearman basé sur le rang de LNY_i et sur le rang de $|e_i|$. Que concluez-vous?

(f) Faites le test d'hétéroscédasticité multiplicative proposé par Harvey et basé sur la régression de $\log(e_i^2)$ sur $\log(LNY_i)$. Que concluez-vous?

(g) Faites le test de Breusch-Pagan basé sur la régression de $(e_i/\widehat{\sigma}_i)$ sur LNY_i. Que concluez-vous?

(h) Faites le test de White. Que concluez-vous?

(i) Faites le test de normalité de Jarque-Bera. Que concluez-vous?

(j) Estimez l'équation de demande de cigarettes et utilisez la correction de White de la variance des estimateurs.

TD 6 - Autocorrélation

1. Soit un processus autorégressif d'ordre un (AR(1)) sur les perturbations:
$$u_t = \rho u_{t-1} + \varepsilon_t \ , \ t = 1, ..., T \text{ et } \varepsilon_t \sim iid\left(0, \sigma_\varepsilon^2\right) \tag{3}$$
Montrez que $E\left(u_t u_s\right) = \rho^s \sigma_u^2$ où $\sigma_u^2 = Var\left(u_t\right).$

2. Pour le modèle de régression linéaire simple sans constante:
$$y_t = \beta x_t + u_t \text{ où } u_t = \rho u_{t-1} + \varepsilon_t \ , \ t = 1, ..., T \text{ et } \varepsilon_t \sim iid\left(0, \sigma_\varepsilon^2\right)$$
   (a) Montrez que
$$Var\left(\widehat{\beta}_{MCO}\right) = \frac{\sigma_u^2}{\sum\limits_{t=1}^{T} x_t^2} \left[1 + 2\rho \frac{\sum\limits_{t=1}^{T-1} x_t x_{t+1}}{\sum\limits_{t=1}^{T} x_t^2} + 2\rho^2 \frac{\sum\limits_{t=1}^{T-2} x_t x_{t+2}}{\sum\limits_{t=1}^{T} x_t^2} + ... + 2\rho^{T-1} \frac{x_1 x_T}{\sum\limits_{t=1}^{T} x_t^2}\right]$$
   et que l'estimateur de Prais-Winsten $\widehat{\beta}_{PW}$ a pour variance:
$$Var\left(\widehat{\beta}_{PW}\right) = \frac{\sigma_u^2}{\sum\limits_{t=1}^{T} x_t^2} \left[\frac{1 - \rho^2}{1 + \rho^2 - 2\rho \left(\frac{\sum\limits_{t=1}^{T-1} x_t x_{t+1}}{\sum\limits_{t=1}^{T} x_t^2}\right)}\right]$$

3. Supposons un processus AR(1) sur x_t: $x_t = \lambda x_{t-1} + v_t$, $t = 1, ..., T$ et $T \to \infty$.

   (a) Montrez que l'efficacité relative asymptotique de l'estimateur des MCO est:
$$asy.eff\left(\widehat{\beta}_{MCO}\right) = \lim_{T \to \infty} \frac{Var\left(\widehat{\beta}_{BW}\right)}{Var\left(\widehat{\beta}_{MCO}\right)} = \frac{\left(1 - \rho^2\right)\left(1 - \rho\lambda\right)}{\left(1 + \rho^2 - 2\rho\lambda\right)\left(1 + \rho\lambda\right)}$$

   (b) Tabulez cette efficacité relative pour $\lambda = [-0.9; +0.9]$ et $\rho = [0; +0.9]$ avec un pas de 0.1 Que concluez-vous? La perte d'efficacité des MCO est-elle importante?

(c) Si on ignore cette autocorrélation, on utilisera $\left(\sigma_u^2 / \sum_{t=1}^{T} x_t^2\right)$ pour $Var\left(\widehat{\beta}_{MCO}\right)$. La différence entre cette formule fausse et la formule dans la question (a) nous donne le bias de l'estimation de la variance. Montrez quand $T \rightarrow \infty$ que ce biais est donné par: $2\rho\lambda / (1 + \rho\lambda)$.

(d) Tabulez ce biais asymptotique pour $\lambda = [-0.9; +0.9]$ et $\rho = [0; +0.9]$ avec un pas de 0.1 Que concluez-vous?

(e) Montrez que:

$$E\left(s^2\right) = \sigma_u^2 \left\{ T - \left(1 + 2\rho \frac{\sum_{t=1}^{T-1} x_t x_{t+1}}{\sum_{t=1}^{T} x_t^2} + 2\rho^2 \frac{\sum_{t=1}^{T-2} x_t x_{t+2}}{\sum_{t=1}^{T} x_t^2} + ... + 2\rho^{T-1} \frac{x_1 x_T}{\sum_{t=1}^{T} x_t^2} \right) \right\} / (T-1)$$

Vérifiez que si $\rho = 0$, alors $E\left(s^2\right) = \sigma_u^2$. Si x_t suit un AR(1): $x_t = \lambda x_{t-1} + v_t$ et pour T grand, montrez que:

$$E\left(s^2\right) = \sigma_u^2 \left(T - \frac{1 + \rho\lambda}{1 - \rho\lambda} \right) / (T-1)$$

(f) Calculez cette expression pour $T = 101$ et $\lambda = \rho = 0.9$. Que concluez-vous?

4. Dans le fichier Consumption.txt, on considère des données de consommation et de revenu aux USA entre 1950 et 1971. On souhaite estimer la fonction de demande:

$$C_t = \alpha + \beta Y_t + u_t \ , \ t = 1950, ..., 1971$$

où C_t (resp. Y_t) est la consommation réelle par tête en dollars 1987 (resp. le revenu disponible réel par tête en dollars 1987).

(a) Estimez par les MCO cette relation et calculez la statistique de Durbin-Watson. Testez l'hypothèse de non-autocorrélation des perturbations.

(b) Appliquez la méthode de Cochrane-Orcutt, puis la procédure à deux étapes de Prais-Winsten. Itérez la procédure de Prais-Winsten.

(c) Appliquez la méthode de Durbin.

(d) Testez l'autocorrélation des résidus avec la méthode de Breusch et Godfrey.

5. Dans le fichier Orange.txt, on dispose de données trimestrielles (1961:1 - 1983:4) pour le comté d'Orange aux USA et on s'intéresse à la relation emploi-production:

$$EMP_t = \alpha + \beta RGNP_t + u_t$$

où EMP_t (resp. $RGNP_t$) est l'emploi (resp. le PNB réel). On estime ce modèle par les MCO.

   (a) Qu'est-ce que l'inspection des résidus et la statistique de Durbin-Watson suggèrent?

   (b) Supposons que $u_t = \rho u_{t-1} + \varepsilon_t$, $t = 1, ..., T$ et $\varepsilon_t \sim iid\,(0, \sigma_\varepsilon^2)$, utilisez la procédure de Cochrane-Orcutt pour estimer ρ et β. Comparez ces estimateurs et leurs écarts-type avec ceux des MCO.

   (c) Même question mais avec la procédure de Prais-Winsten.

   (d) Appliquez le test de Breusch et Godfrey pour une autocorrélation du 1er ordre et du second ordre. Que concluez-vous?

TD 7 - Le modèle linéaire général

1. Soit le modèle linéaire géneral:

$$y = X\beta + u$$

où y est de taille $(n, 1)$, X est de taille (n, k) et β est de taille $(k, 1)$. On post-multiplie (multiplication à droite) les variables explicatives par une transformation non singulière C telle que: $X^* = XC$ et soit $y = X^*\beta^* + u$.

   (a) Montrez que: $P_{X^*} = P_X$ et $\overline{P}_{X^*} = \overline{P}_X$ où $P_X = X\left(X'X\right)^{-1} X'$ et $\overline{P}_X = I_n - P_X$.

   (b) En conclure que la régression de y sur X a les mêmes valeurs estimées et les mêmes résidus que la régression de y sur X^*.

   (c) Supposons que chaque X soit multiplié par une constante. Les valeurs estimées et les résidus changent-ils si l'on estime cette régression par rapport à la précédente?

   (d) Supposons que X soit constitué de 2 régresseurs X_1 et X_2. Si on régresse y sur $(X_1 - X_2)$ et y sur $(X_1 + X_2)$, cela conduira-t-il aux mêmes résultats que ceux de la régression originale de y sur X_1 et X_2?

2. Théorème de Frisch-Waugh-Lovell (FWL). Soit le modèle: $y = X\beta + u = X_1\beta_1 + X_2\beta_2 + u$. En utilisant la formule de l'inverse partitionnée[1],

   (a) montrez que:

$$\widehat{\beta}_{2,MCO} = \left(X_2'\overline{P}_{X_1}X_2\right)^{-1} X_2'\overline{P}_{X_1}y$$

[1] L'inverse partitionnée de la matrice:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

est:

$$A^{-1} = \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}B_{22}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}B_{22} \\ -B_{22}A_{21}A_{11}^{-1} & B_{22} \end{pmatrix}$$

où $B_{22} = \left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)$.

(b) En partitionnant le modèle $y = X_1\beta_1 + X_2\beta_2 + u$, les équations normales sont:

$$\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} \begin{pmatrix} \widehat{\beta}_{1,MCO} \\ \widehat{\beta}_{2,MCO} \end{pmatrix} = \begin{pmatrix} X_1'y \\ X_2'y \end{pmatrix}$$

Ecrivez ce système comme un système de deux équations à deux inconnues ($\widehat{\beta}_{1,MCO}$ et $\widehat{\beta}_{2,MCO}$). En éliminant $\widehat{\beta}_{1,MCO}$, résolvez ce système et montrez que le résultat est:

$$\widehat{\beta}_{2,MCO} = \left(X_2'\overline{P}_{X_1}X_2\right)^{-1} X_2'\overline{P}_{X_1}y.$$

(c) En utilisant le théorème de FWL, montrez que si $X_1 = e_n$, un vecteur unitaire de taille $(n, 1)$ indiquant la présence d'une constante dans le modèle et si X_2 est un ensemble de variables économiques, alors:

i. $\widehat{\beta}_{2,MCO}$ peut être obtenu en régressant $(y_i - \overline{y})$ sur l'ensemble des variables X_2 exprimées en écarts à leurs moyennes $\left(X_2 - \overline{X}_2\right)$ où $\overline{X}_2 = e_n'X_2/n$;

ii. l'estimateur de la constante $\widehat{\beta}_{1,MCO}$ peut être obtenu par: $\overline{y} - \overline{X}_2\widehat{\beta}_{2,MCO}$.

3. Soit la Log-vraisemblance du modèle linéaire général:

$$LogL\left(\beta, \sigma^2\right) = -\left(n/2\right)\log\left(2\pi\sigma^2\right) - \left(y - X\beta\right)'\left(y - X\beta\right)/2\sigma^2$$

(a) Dérivez les conditions du premier ordre de la maximisation et montrez que:

$$\widehat{\beta}_{MV} = \widehat{\beta}_{MCO} \text{ et } \widehat{\sigma}^2_{MV} = SCR/n$$

(b) Calculez les dérivées secondes et vérifiez que la matrice d'information se réduit à:

$$I\left(\beta, \sigma^2\right) = \begin{pmatrix} X'X/\sigma^2 & 0 \\ 0 & n/2\sigma^4 \end{pmatrix}$$

4. Sachant que $u \sim N\left(0, \sigma^2 I_n\right)$ et $\left(n-k\right)s^2/\sigma^2 \sim \chi^2_{n-k}$, montrez que:

(a) s^2 est un estimateur sans biais de σ^2.

(b) $Var\left(s^2\right) = 2\sigma^4 / \left(n-k\right)$ sachant que $E\left(\chi_r^2\right) = r$ et $Var\left(\chi_r^2\right) = 2r$.

5. On considère l'horizon de prévision $T_{0.}$ Soit la spécification:

$$y^* = X^*\delta + u^* \quad \leftrightarrow \quad \begin{pmatrix} y \\ y_0 \end{pmatrix} = \begin{pmatrix} X & 0 \\ X_0 & I_{T_0} \end{pmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \begin{pmatrix} u \\ u_0 \end{pmatrix}$$

(a) Montrez que:

$$\begin{aligned} \widehat{\delta}'_{MCO} &= \left(\widehat{\beta}'_{MCO}, \widehat{\gamma}'_{MCO}\right) \\ \text{où } \widehat{\beta}_{MCO} &= \left(X'X\right)^{-1} X'y \\ \text{et } \widehat{\gamma}_{MCO} &= y_0 - X_0\widehat{\beta}_{MCO} \end{aligned}$$

(b) Montrez que:

$$e^*_{MCO} = \left(e'_{MCO}, 0'\right)' \text{ et } s^{*2} = s^2$$

(c) Montrez que:

$$s^{*2}\left(X^{*\prime}X^*\right)^{-1} = s^2 \begin{pmatrix} \left(X'X\right)^{-1} & 0 \\ 0 & \left[I_{T_0} + X_0\left(X'X\right)^{-1}X_0'\right] \end{pmatrix}$$

6. Montrez que l'estimateur contraint:

$$\widehat{\beta}_r = \widehat{\beta}_{MCO} + \left(X'X\right)^{-1} R' \left[R\left(X'X\right)^{-1} R'\right]^{-1} \left(r - R\widehat{\beta}_{MCO}\right)$$

est sans biais sous l'hypothèse nulle: $H_0 : R\beta = r$.

(a) Montrez que:

$$Var\left(\widehat{\beta}_r\right) = Var\left(A\left(X'X\right)^{-1} X'u\right)$$

où

$$A = I_k - \left(X'X\right)^{-1} R' \left[R\left(X'X\right)^{-1} R'\right]^{-1} R$$

(b) Prouvez que: $A^2 = A$ mais que $A' \neq A$. En conclure que:

$$\begin{aligned} Var\left(\widehat{\beta}_r\right) &= \sigma^2 A\left(X'X\right)^{-1} A' \\ &= \sigma^2 \left\{\left(X'X\right)^{-1} - \left(X'X\right)^{-1} R' \left[R\left(X'X\right)^{-1} R'\right]^{-1} R\left(X'X\right)^{-1}\right\} \end{aligned}$$

(c) Montrez que $Var\left(\widehat{\beta}_{MCO}\right) - Var\left(\widehat{\beta}_r\right)$ est une matrice semi-définie positive.

7. Soit le modèle linéaire général:

$$y = X\beta + u = X_1\beta_1 + X_2\beta_2 + u \ , \ u \sim N\left(0, \sigma^2 I_T\right)$$

   (a) Ecrivez la log-vraisemblance du modèle et déterminez l'estimateur du maximum de vraisemblance de β et de σ^2.

   (b) Ecrivez le score $S(\beta) = \partial LogL(\beta, \sigma^2)/\partial\beta$ et montrez que la matrice d'information est bloc-diagonale.

   (c) Dérivez les tests du rapport de vraisemblance (LR), de Wald (W) et du multiplicateur de Lagrange (ML) pour tester:

$$H_0 : \beta_1 = \beta_1^0 \text{ versus } H_1 : \beta_1 \neq \beta_1^0$$

   (d) Comme $X = [X_1, X_2]$, montrez que:

$$\begin{aligned} W &= \left(\beta_1^0 - \widehat{\beta}_1\right)' \left[X_1' \overline{P}_{X_2} X_1\right] \left(\beta_1^0 - \widehat{\beta}_1\right) / \widehat{\sigma}^2 \\ LR &= T.\log\left(\widetilde{u}'\widetilde{u}/\widehat{u}'\widehat{u}\right) \\ ML &= \widetilde{u}' X_1 \left[X_1' \overline{P}_{X_2} X_1\right]^{-1} X_1' \widetilde{u}/\widetilde{\sigma}^2 \end{aligned}$$

   où $\widehat{u} = y - X\widehat{\beta}$, $\widetilde{u} = y - X\widetilde{\beta}$, $\widehat{\sigma}^2 = \widehat{u}'\widehat{u}/T$ et $\widetilde{\sigma}^2 = \widetilde{u}'\widetilde{u}/T$. $\widehat{\beta}$ (resp. $\widetilde{\beta}$) est l'estimateur du maximum de vraisemblance non contraint (resp. contraint).

8. Dans le fichier Cigarettes.txt (cf. TD 5), on considère des données de consommation de cigarettes dans 46 états des USA pour l'année 1992. On souhaite estimer la fonction de demande:

$$\log C = \alpha + \beta \log P + \gamma \log Y + u$$

   où $\log C$ (resp.$\log Y$ et $\log P$) est le vecteur $(46, 1)$ du logarithme de la consommation de cigarettes (en paquets) par tête dans l'état (i, $i = 1..46$) pour des individus agés de plus de 16 ans (resp.le logaritme du revenu disponible par tête et le logarithme du prix relatif des cigarettes).

(a) Calculez les tests du rapport de vraisemblance (LR), de Wald (W) et du multiplicateur de Lagrange (ML) pour tester:

$$H_0 : \beta = -1$$

(b) Calculez la statistique de Wald pour:

$$H_0^A : \beta = -1 \ , \ H_0^B : \beta^5 = -1 \text{ et } H_0^C : \beta^{-5} = -1$$

TD 8 - Les moindres carrés généralisés

1. Soit le modèle linéaire géneral:

$$y = X\beta + u$$

où y est de taille $(n, 1)$, X est de taille (n, K) et β est de taille $(K, 1)$. On suppose que $u \sim (0, \sigma^2\Omega)$ où $\Omega \neq I_n$.

   (a) Montrez que:

$$Var\left(\widehat{\beta}_{MCO}\right) = \sigma^2 \left(X'X\right)^{-1} X'\Omega X \left(X'X\right)^{-1}$$

   (b) Montrez que:

$$Var\left(\widehat{\beta}_{MCO}\right) - Var\left(\widehat{\beta}_{MCG}\right) = \sigma^2 A\Omega A'$$

   où

$$A = \left[\left(X'X\right)^{-1} X' - \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}\right]$$

   et vérifiez que cette différence est semi-définie positive.

   (c) Montrez que:

$$E\left(s^2\right) = \sigma^2 \mathrm{tr}\left(\Omega \overline{P}_X\right) / (n - K) \text{ où } \overline{P}_X = I_n - P_X \text{ et } P_X = X\left(X'X\right)^{-1} X'$$

2. Soit le modèle général: $y = X\beta + u$ où le vecteur des résidus u suit un processus autorégressif d'ordre 1 ($AR(1)$):

$$u_t = \rho u_{t-1} + \varepsilon_t \text{ où } |\rho| < 1 \text{ et } \varepsilon_t \sim \left(0, \sigma_\varepsilon^2\right) \text{ pour } t = 1, ..., T$$

la matrice de variances-covariances des perturbations s'écrit:

$$\Omega = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{pmatrix}$$

et son inverse est:

$$\Omega^{-1} = \frac{1}{(1-\rho^2)} \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

(a) Vérifier que: $\Omega\Omega^{-1} = I_T$

(b) Montrez que: $P^{-1}\left(P^{-1}\right)' = (1-\rho^2)\,\Omega^{-1}$ pour:

$$P^{-1} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

(c) En conclure que $Var\left(P^{-1}u\right) = \sigma^2 I_T$.

3. On sait que sous l'hypothèse nulle: $H_0 : R\beta = r$, l'estimateur des MCO contraint est:

$$\widehat{\beta}_r = \widehat{\beta}_{MCO} + \left(X'X\right)^{-1} R' \left[R\left(X'X\right)^{-1} R'\right]^{-1} \left(r - R\widehat{\beta}_{MCO}\right)$$

si $u \sim \left(0, \sigma^2 I_n\right)$. Donnez l'expression de l'estimateur contraint quand $u \sim \left(0, \sigma^2\Omega\right)$.

4. Pour l'observation future $(T+s)$, on a:

$$y_{T+s} = x'_{T+s}\beta + u_{T+s}$$

et on considère un prédicteur linéaire de la forme:

$$\widehat{y}_{T+s} = c'y$$

où $u \sim (0, \Sigma)$ et $\Sigma = \sigma^2\Omega$.

(a) Montrez que pour que $\widehat{y}_{T+s}$ soit un estimateur sans biais de $\widehat{y}_{T+s}$, il faut que: $c'X = x'_{T+s}$.

(b) Montrez que $Var(\widehat{y}_{T+s}) = c'\Sigma c + \sigma^2_{T+s} - 2c'\omega$ où $\sigma^2_{T+s} = Var(u_{T+s})$ et $\omega = E(u_{T+s}u)$.

(c) Minimisez $Var(\widehat{y}_{T+s})$ sous la contrainte $\left(c'X = x'_{T+s}\right)$ et montrez que:

$$\widehat{c} = \Sigma^{-1}\left[I_T - X\left(X'\Sigma^{-1}X\right)^{-1}X'\Sigma^{-1}\right]\omega + \Sigma^{-1}X\left(X'\Sigma^{-1}X\right)^{-1}x_{T+s}$$

(d) En conclure que:

$$\widehat{y}_{T+s} = x'_{T+s}\widehat{\beta}_{MCG} + \omega'\Sigma^{-1}e_{MCG} \text{ où } e_{MCG} = \left(y - X\widehat{\beta}_{MCG}\right)$$

TD 9 - Variables instrumentales, GMM et équations simultanées

1. Dans la base de données bovins.dta, on dispose de données macroéconomiques françaises pour le secteur agricole sur la période 1961-1998. Les données (aux prix de 1990, base 100 en 1990) sont:
q, la production bovine
ctot, la consommation totale
p, le prix à la production du boeuf
pb, le prix à la consommation du boeuf
w, le prix des aliments du bétail
b, le solde du commerce extérieur
s, le niveau des stocks publics en fin d'année
pm, le prix à la consommation du mouton
ipc, l'indice des prix à la consommation finale
r, le revenu disponible brut par tête
pop, la population française au 1er janvier
C, la consommation de viande bovine par tete
pr, le rapport du prix du boeuf à la production /prix a la consommation
d74, une dummy concernant le boycott de la viande de veau en 1974.
On souhaite estimer le modèle d'équilibre du secteur agricole bovin suivant

$$\begin{aligned} q_t &= \alpha_0 + \alpha_1 q_{t-1} + \alpha_2 p_{t-1} + \alpha_3 p_{t-2} + \alpha_4 p_{t-3} + \alpha_5 w_{t-1} + \alpha_6 d74_t + u_{1t} \quad \text{éq. d'offre} \\ C_t &= \beta_0 + \beta_1 C_{t-1} + \beta_2 pb_t + \beta_3 pm_t + \beta_3 r_t + u_{2t} \quad \text{éq. de demande} \\ pb_t &= \gamma_0 + \gamma_1 pr_t + u_{3t} \quad \text{éq. de marge} \\ b_t &= \delta_0 + \delta_1 s_{t-1} + u_{4t} \quad \text{eq. du commerce exterieur} \end{aligned}$$

   (a) Estimez chaque équation par les MCO et interprétez.
   (b) Estimez la fonction de demande par les 2SLS en prenant comme instruments les variables prédéterminées et les variables exogènes du modèle: p_{t-1}, p_{t-2}, p_{t-3}, w_{t-1}, $d74_t$, pm_t, r_t, s_t, s_{t-1}. Interprétez.
   (c) Avec les mêmes instruments, estimez le système complet par les 2SLS, puis par les 3SLS. Interprétez.
   (d) Avec les mêmes instruments, estimez l'équation de demande par les GMM avec correction de l'hétéroscédasticité et de l'autocorrélation des résidus par la méthode de Newey-West. Interprétez.

(e) Avec les mêmes instruments, estimez le système complet par les GMM avec correction de l'hétéroscédasticité et de l'autocorrélation des résidus par la méthode de Newey-West. Interprétez.

2. Dans la base de données textile.dta, on dispose de données (en écarts à leurs moyennes) sue la demande de textile pour 22 années. On souhaite estimer le modèle suivant:

$$
\begin{aligned}
y_{1t} &= a_{11}y_{2t} + a_{12}y_{3t} + a_{13}y_{3t-1} + a_{14}t + u_{1t} \quad \text{éq. de demande de textile} \\
y_{2t} &= a_{21}y_{1t} + a_{22}y_{4t} + a_{23}t + u_{2t} \quad \text{éq. de prix à la consom. textile} \\
y_{3t} &= a_{31}y_{3t-1} + a_{32}x_{1t} + u_{3t} \quad \text{éq. de revenu des ménages} \\
y_{4t} &= a_{41}y_{5t} + a_{42}y_{5t-1} + a_{43}t + u_{4t} \quad \text{éq. d'offre de textile} \\
y_{5t} &= a_{51}y_{2t} + a_{52}t + u_{5t} \quad \text{éq. de prix à la product. textile}
\end{aligned}
$$

où y_{1t} est la demande de produit textile, y_{2t} est l'indice relatif des prix à la consommation des produits textiles, y_{3t} est le revenu disponible par habitant, y_{4t} est la production de produits textiles, y_{5t} est l'indice relatif des prix à la production des produits textiles, x_{1t} est l'investissement par habitant et t la tendance.

(a) Etudier les conditions d'identification de ce modèle.

(b) Estimer le système complet par les 2SLS, puis par les 3SLS en utilisant les instruments suivants: $y_{3t-1}, y_{5t-1}, x_{1t-1}, t$. Interprétez.

(c) Avec les mêmes instruments, estimer le système complet par les GMM en supposant l'homoscédasticité et la non-autocorrélation des perturbations du modèle. Interprétez.