Économétrie

Chapitre 2

Modèle de Régression Linéaire Simple

Le modèle de régression simple

- La régression simple est le modèle le plus simple: une variable endogène est expliquée par une variable exogène
- Soit la fonction de production keynésienne:

  $C=a_0+a_1Y$
- C=consommation
- Y=revenu
- a1=propension marginale à consommer
- a0=consommation autonome ou incompressible

Le modèle de régression simple

- La variable consommation est appelée variable à expliquer ou variable endogène
- La variable revenu est appelée variable explicative ou exogène
- a_0 et a_1 sont les paramètres du modèle ou encore les coefficients de régression

Le modèle de régression simple

- Nous pouvons distinguer deux types des spécifications
- Les modèles en série temporelle, les variables représentent des phénomènes observés à intervalles de temps réguliers
- Par exemple la consommation et le revenu annuel de 1985 et 2005 pour un pays donné:

$$C_t = a_0 + a_1 Y_t \qquad t = 1985, \ldots, 2005$$

Le modèle de régression simple

- Les modèles en coupe instantanée, le variables représentent des phénomènes observé au même instant mais concernant divers individus, par exemple la consommation et le revenu observe sur un échantillon de 20 pays

$$C_i = a_0 + a_1 Y_i \qquad i=1,\ldots,20$$

- C_i=consommation pour le pays i en 2005
- Y_i=revenu pour le pays i en 2005

Le modèle de régression simple

- Le modèle qu'il vient d'être spécifié n'est qu'un caricature de la réalité.
- En effet, ne retenir que le revenu pour l'explication de la consommation est à l'évidence même insuffisant
- Il existe un multitude d'autre facteurs susceptible d'expliquer la consommation

Le modèle de régression simple

- C'est pourquoi nous ajoutons un terme (ε_t) qui synthétise l'ensemble de ces informations non explicitées dans le modèle
- $C_t = a_0 + a_1 Y_t + \varepsilon_t$ série temporelle
- $C_i = a_0 + a_1 Y_i + \varepsilon_i$ coupe instantanée
- Où ε représente l'erreur de spécification du modèle, c'est-à-dire l'ensemble des phénomènes explicatifs de la consommation non liés au revenu

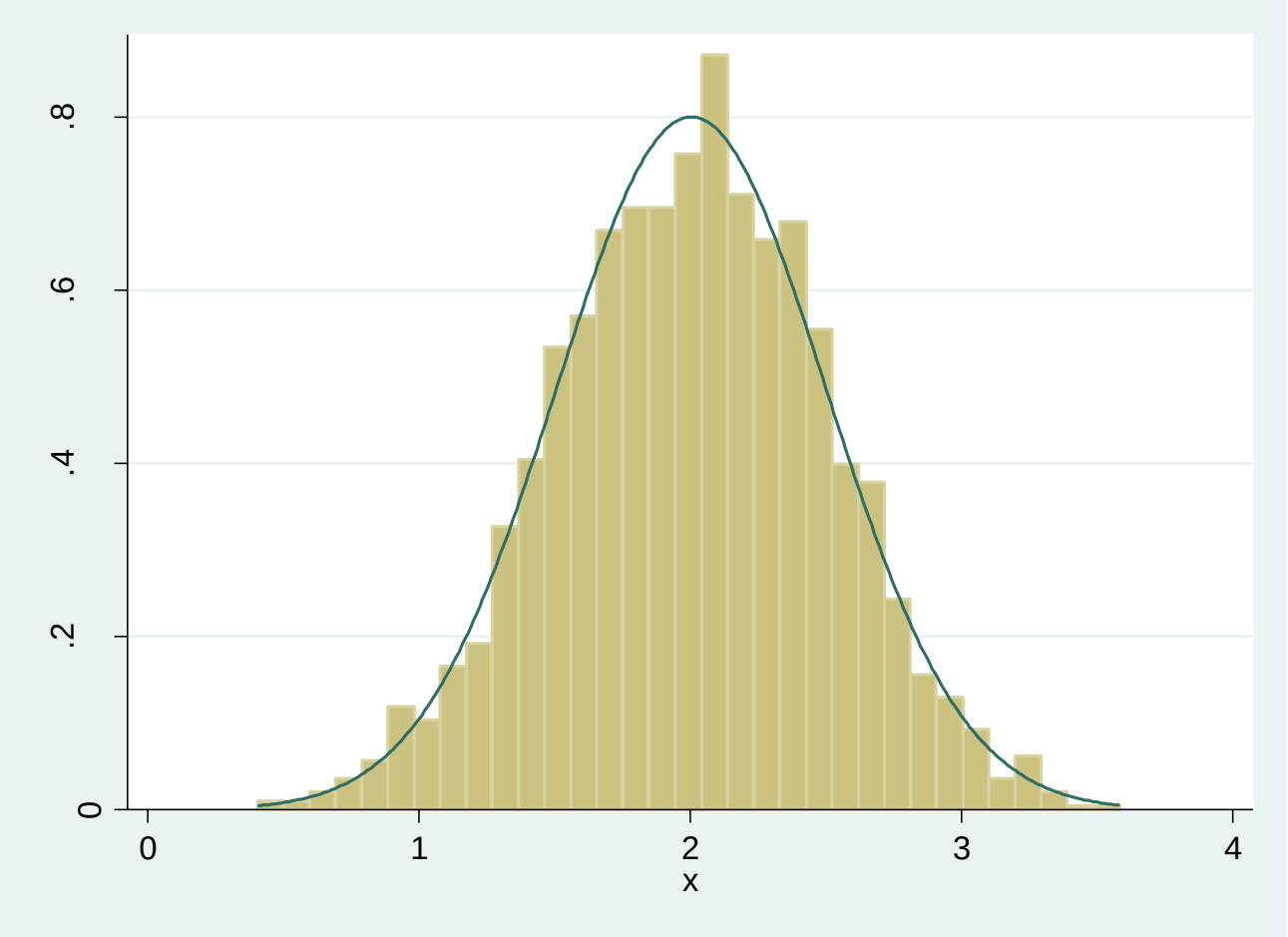
Le modèle de régression simple

- En pratique, le terme ε mesure le différence entre les valeurs réellement observées de C_t et les valeurs qui auraient été observées si la relation spécifiée avait été rigoureusement exacte
- Le terme ε regroupe donc trois types d 'erreurs:
  - Erreur de spécification, c.à.d. le fait que la variable explicative n'est pas suffisante
  - Erreur de mesure: le données ne représentent pas exactement le phénomène
  - Erreur de fluctuation d'échantillonnage

Le modèle de régression simple

- Dans la réalité nous ne connaissons pas les valeurs vrais des coefficients
- On peut seulement observer le valeurs de C et de R
- Les estimateurs de coefficients sont notés respectivement: $\widehat{a_0}$ $\widehat{a_1}$
- ces sont des variables aléatoires, qui suivent les mêmes lois de probabilité, celle de ε, puisque ils sont fonction de ε

Le modèle de régression simple



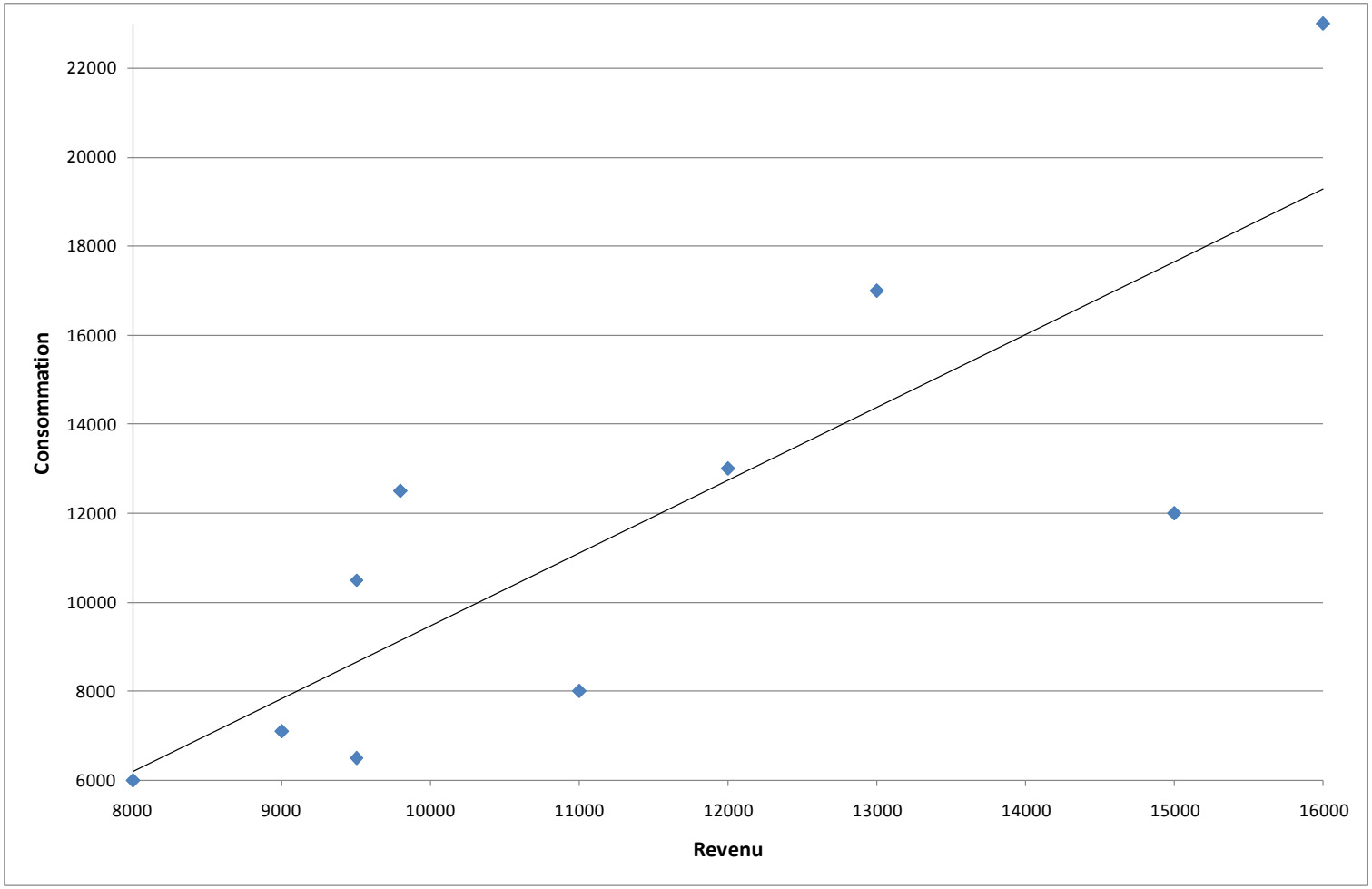
Le modèle de régression simple

- Estimation des parametres
- $y_t = a_0 + a_1 x_t + \varepsilon_t$ pour $t=1,\ldots,n$
- Avec:
  - y_t = variable à expliquer au temps t
  - x_t = variable explicative au temps t
  - a_0 a_1 = paramètres du modèle
  - ε_t = erreur de spécification
  - n = nombre d'observations

Le modèle de régression simple

- Hypothèses:
- H1: Le modèle est linéaire en x
- H2: Le valeurs de x sont observés sans erreur
- H3: $E(\varepsilon)=0$, l'espérance mathématique de l'erreur est nulle
- H4: $E(\varepsilon^2)=\sigma^2$, la variance de l'erreur est constante (homoscédasticité)
- H5: $E(\varepsilon_t \varepsilon_{t+1})$, les erreurs sont non corrélées (ou indépendants)
- H6: $Cov(x_t\ \varepsilon_t)$, l'erreur est indépendante de la variable explicative

Le modèle de régression simple



Le modèle de régression simple

- Les estimateurs des coefficients a_0 et a_1 est obtenu en minimisant la distance au carré entre chaque observation et la droit
- D'où le nom d'estimateur des moindres carrés ordinaires (MCO)
- La résolution analytique est la suivante:

$$Min \sum_{t=1}^{T} \varepsilon^2 = Min \sum_{t=1}^{T} \left(y_t - \hat{a}_0 - \hat{a}_1 x_t \right)^2 = Min\ S$$

Le modèle de régression simple

- En opérant par dérivation par rapport à a0 et a1 afin de trouver le minimum de cette fonction, on obtient les résultats suivants:

$$\frac{\partial S}{\partial \hat{a}_0} = 0$$

$$\frac{\partial S}{\partial \hat{a}_1} = 0$$

$$\hat{a}_1 = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{T}(x_t - \bar{x})^2} = \frac{\sum_{t=1}^{T} x_t y_t - T\,\bar{x}\,\bar{y}}{\sum_{t=1}^{T} x_t^2 - T\,\bar{x}^2}$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}$$

Le modèle de régression simple

- La spécification du modèle n'est pas neutre:
  - y=f(x) n'est pas équivalente à x=f(y)
  - Le coefficient a_1 représente la pente de la droite ou encore la propension marginale. On verra que lorsque les variables sont transformés en logs le coefficient représentera l'élasticité.
- Il y a des cases spéciaux où le terme constante est nul: pour exemple le cas d'un fonction de production où le facteur fixe n'intervienne pas.

Le modèle de régression simple

- Le modèle de régression linéaire simple peut s'écrire sous deux formes selon qu'il s'agit du modèle théorique spécifié par l'économiste ou du modèle estimé à partir d'un échantillon

$$y_t = a_0 + a_1 x_t + \varepsilon_t$$

$$y_t = \widehat{a_0} + \widehat{a_1} x_t + e_t = \widehat{y_t} + e_t$$

Le modèle de régression simple

- Le résidu observé et est donc la différence entre les valeurs observées de la variable à expliquer et les valeurs ajustées à l'aide des estimations de coefficients du modèle:

$$\hat{y}_t = \widehat{a_0} + \widehat{a_1} x_t$$

Le modèle de régression simple

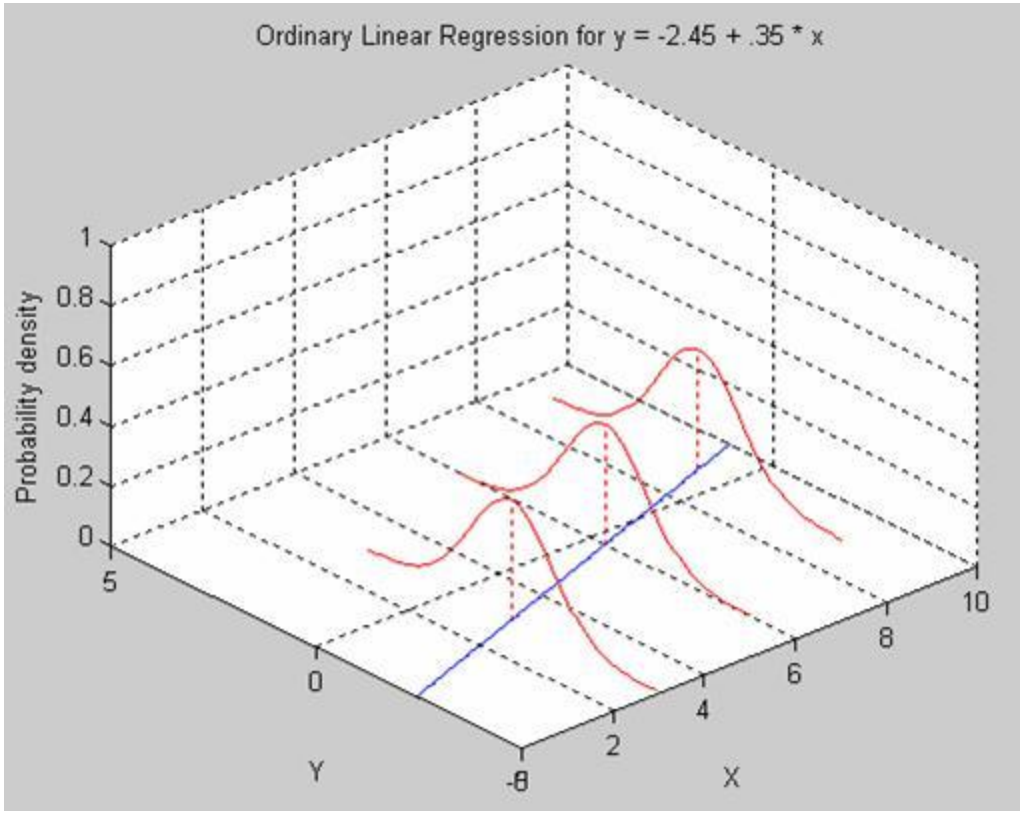
- Les estimateurs obtenus en utilisant le méthode de moindres carrés ordinaires ont deux proprieté importantes
- Ils sont sans biais: $E(\widehat{a_0}) = a_0 \qquad E(\widehat{a_1}) = a_1$
- Il sont convergents: $\lim_{n\to\infty} V(\widehat{a_1}) = 0$
- Ces types d'estimateurs sont dit 'BLUE' : best linear unbiased estimators.

Le modèle de régression simple

- L'hypothèse de normalité des erreurs n'est pas nécessaire pour obtenir des estimateurs convergentes
- Il est en revanche importante pour construire des test statistiques concernent la validité du modèle estimé

$$\varepsilon_t \rightarrow N(0, \sigma_\varepsilon^2)$$

Le modèle de régression simple



Le modèle de régression simple

- On peut calculer les estimateurs de la variance de l'erreur et des estimateurs:

$$\hat{\sigma}_{\varepsilon}^2 = \frac{1}{n-2}\sum_t e^2$$

$$\hat{\sigma}_{a_1}^2 = \frac{\hat{\sigma}_{\varepsilon}^2}{\sum_t (x_t - \bar{x})^2}$$

$$\hat{\sigma}_{a_0}^2 = \hat{\sigma}_{\varepsilon}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_t (x_t - \bar{x})^2} \right)$$

Le modèle de régression simple

- En conséquence de l'hypothèse de normalité des erreurs, on peut observer que:

$$\frac{\widehat{a_1} - a_1}{\sigma_{a_1}} \rightarrow N(0,1) \qquad \frac{\widehat{a_0} - a_0}{\sigma_{a_0}} \rightarrow N(0,1)$$

- En utilisant ces formules, on peut mettre en place des test statistiques pour:
- Comparer un coefficient de régression par rapport à une valeur fixée
- Comparer deux coefficients provenant de deux échantillons différents
- Déterminer un intervalle de confiance pour un coefficient

Le modèle de régression simple

- L'analyse de la variance est importante pour évaluer dans quelle mesure le modèle estimé est capable de expliquer la réalité.
- La formule pour l'analyse de la variance est la suivante:

$$\sum_t (y_t - \bar{y})^2 = \sum_t (\hat{y}_t - \bar{\hat{y}})^2 + \sum_t e_t^2$$

- La variabilité totale est égale à la variabilité expliquée plus la variabilité des residus

Le modèle de régression simple

- Cette équation va nous permettre de juger de la qualité de l'ajustement d'un modèle
- Plus la variance expliquée est proche de la variance totale, meilleur est l'ajustement de la nuage de points par la droite de moindres carrés

$$R^2 = \frac{\sum_t (\hat{y}_t - \bar{y})^2}{\sum_t (y_t - \bar{y})^2} = 1 - \frac{\sum_t e_t^2}{\sum_t (y_t - \bar{y})^2}$$

- R^2 = Coefficient de détermination; R = corrélation multiple

Le modèle de régression simple

Source de la variation	Somme Des carrés	Degré de liberté	Carré moyens
x	SCE=$\sum_t \left(\hat{y}_t - \bar{\hat{y}}\right)^2$	1	SCE/1
Résidu	SCR=$\sum_t e_t^2$	n-2	SCR/(n-2)
Total	SCT=$\sum_t (y_t - \bar{y})^2$	n-1	

$$F^* = \frac{SCE/1}{SCR/(n-2)} = \frac{R^2}{(1-R^2)/(n-2)}$$