Introduction aux modèles économétriques

Laurent Ferrara

Janvier 2019

Plan de la présentation

1. **Concepts**

2. **Modèles économétriques**

3. **Exemples de modèles**

1. Concepts

Objectif

Prendre des décisions à partir de l'observation d'un ensemble de données

Méthode

Construction d'un modèle économétrique pour chaque type d'étude

Outils

1) <u>Théorie de l 'information et concepts d'optimalité</u>

→ construction d'une population exhaustive : pas tjs facile ….

→ Solution : échantillonnage

2) <u>Modèles économétriques:</u>

Modèles probabilistes : loi de distribution de l’échantillon

Modèles paramétriques : modèles de régression linéaire et non-linéaire, modèles de séries chronologiques, ….

→ Chaque type de modèle fait appel à des paramètres (de la loi et / ou du modèle), *a priori* inconnus qu’il faudra estimer.

On peut identifier de 2 types de paramètres : paramètres de loi de distribution ou paramètres de structure.

3) <u>Estimation et tests</u> :

Inférence statistique basée sur l’échantillon observé

→ Contrôle de la qualité de l ’information et de la décision prise associée au test

Exemples :

- Etude d'une population par sondage (élections, enquêtes d'opinion auprès des ménages et des industriels…)

- Explication de phénomènes macro-économiques et micro-économiques

- Prévision (variables macro, taux de change, actifs financiers, …)

- Prise de décision de politique économique par le gouvernement, la banque centrale, les institutions internationales

2. Modèles économétriques

Définition formelle

On appelle <u>*modèle au sens statistique*</u> *la donnée d'un triplet*

$$(\Omega, F, IP)$$

où :

Ω est l'ensemble (les données)

F est une tribu sur Ω

IP est une famille de lois de proba. sur (Ω, F) tq : $IP \rightarrow (IP_\theta)_{\theta \in \Theta}$

$$\theta \in \Theta \subset \Re^k$$

qui dépend d'un paramètre vectoriel

Hypothèse clé de travail :

Les individus interrogés sont identifiés à des variables aléatoires $X_1, \ldots, X_n$, *à valeurs dans* Ω ($\Omega = R$ *ou* R^d), *indépendantes et de même loi de distribution* P_θ *(i.i.d.)*

Remarques :

R1: En général, $\Omega \subset R^d$, on travaille alors avec la tribu des boréliens B_{R^d}

R2 : Quand $\Theta \subset (R^d, B_{R^d})$ on parle de modèle paramétrique probabiliste. On suppose connaître la loi P_θ mais θ inconnu.

$\rightarrow$ On va donc se servir de l'*échantillon* qu 'on aura construit à partir des individus pour identifier ce paramètre θ.

Définition:

On appelle <u>n-échantillon</u>, le vecteur aléatoire :

$X = (X_1, \ldots, X_n)$ de loi P_θ^n, suite finie de v.a. indépendantes et identiquement distribuées (iid) de loi P_θ.

Définition:

On appelle <u>observation</u> une réalisation du vecteur aléatoire X, notée : $x = (x_1, \ldots, x_n)$.

Remarques :

R1: Le modèle statistique associé à X est : $(\Omega^n, F^n, IP_\theta^n)$

R2: La famille de lois P_θ est supposée posséder une densité continue $f_\theta(x)$ ou discrète $p(x,\theta)$ (abus de language)

R3 : $\Delta_\theta = \{x; \quad f_\theta(x) > 0\}$

est le <u>support</u> de la loi P_θ .

R4 : Sous l'hypothèse d'indépendance : $P_\theta^n = P_\theta \times ... \times P_\theta$

Définition:

On appelle statistique toute fonction mesurable f tq:

$f: (X_1, \ldots, X_n) \rightarrow f(X_1, \ldots, X_n) \in R^k$,

k étant la dimension de la statistique.

Exemples :

- $(X_1, \ldots, X_n)$ statistique de dimension n
- $(X_{(1)}, \ldots, X_{(n)})$ statistique de dimension n
- X_1 statistique de dimension 1
- $1/n \sum_i X_i$ statistique de dimension 1

3. Exemples de modèles

Modèle Binomial

Ω = {données} = {réponses à une question binaire (oui/non)}

On interroge n individus, X_i est la réponse de l 'individu i :

$X_i = 1$ si oui et $X_i = 0$ si non

$\rightarrow$ X_i est une v.a. qui suit une loi de Bernouilli de paramètre θ inconnu tq :

θ = probabilité que l'individu réponde oui

Le modèle statistique associé au vecteur aléatoire X est :

$$(\{0,1\}^n, F^n, B(\theta, n))$$

Exemples ? θ ?

Modèle Log-Normal

Sur n individus, on mesure une variable R_i .

On suppose qu 'il s 'agit d 'une variable continue positive

$$R_i \sim LogN\,(m,\sigma^2)$$

$$ie \qquad Log(R_i) \sim N\,(m,\sigma^2)$$

On a 2 paramètres d 'intérêt m et σ^2

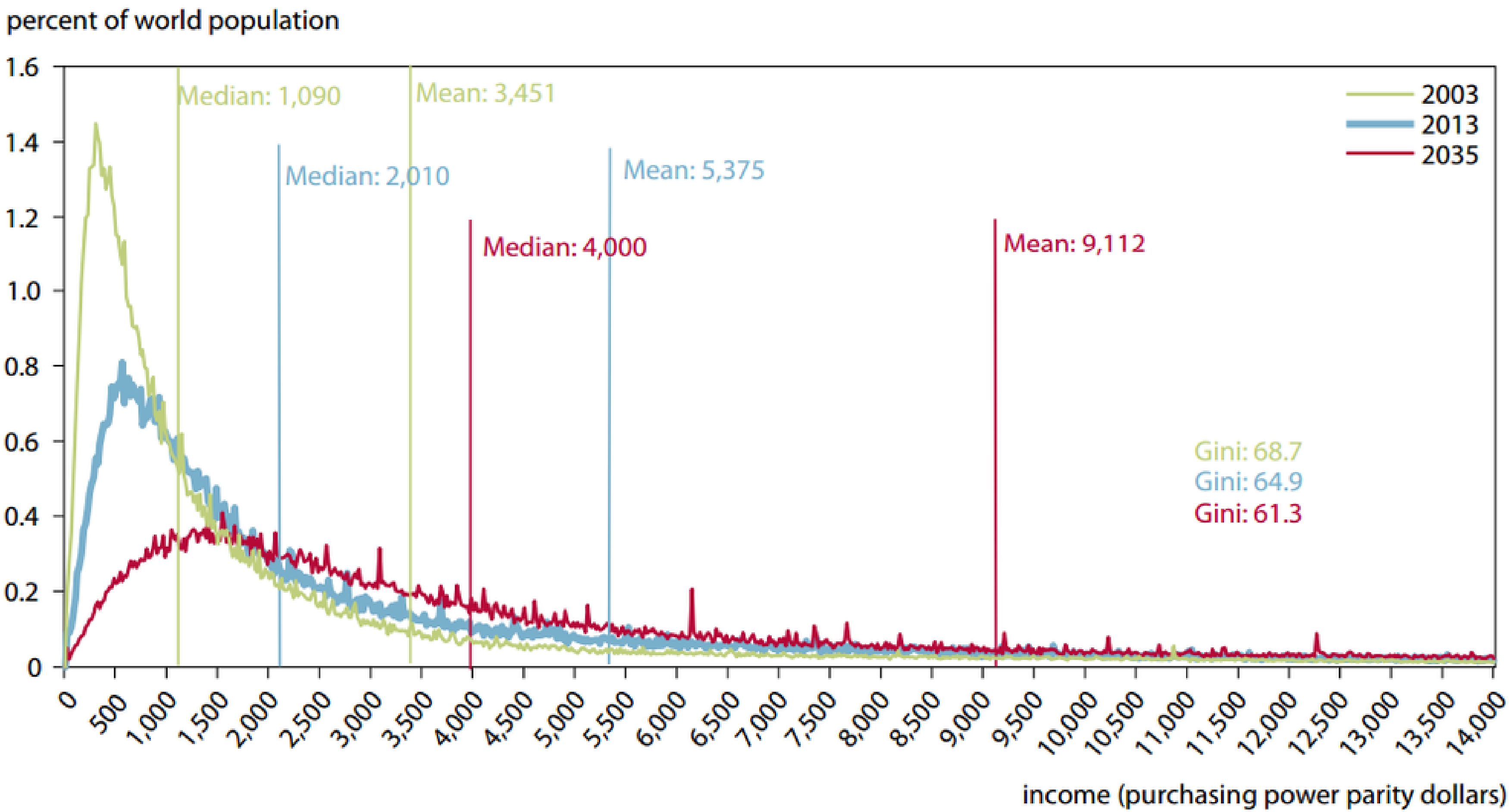
Le modèle statistique associé au vecteur aléatoire X est :

$$(R^n, B_{R^n}, (LogN)^n)$$

<u>Exemples</u> ?

Distribution des revenus mondiaux à différentes dates

Figure 5 Frequency plot of global income distribution, 2003, 2013, and 2035



Modèle linéaire

Sur n individus, on mesure les variables X_i et Y_i ,

On suppose qu'il existe une relation linéaire entre elles, ie:

pour chaque i :

$$Y_i = a X_i + b$$

Les paramètres (a,b) sont inconnus

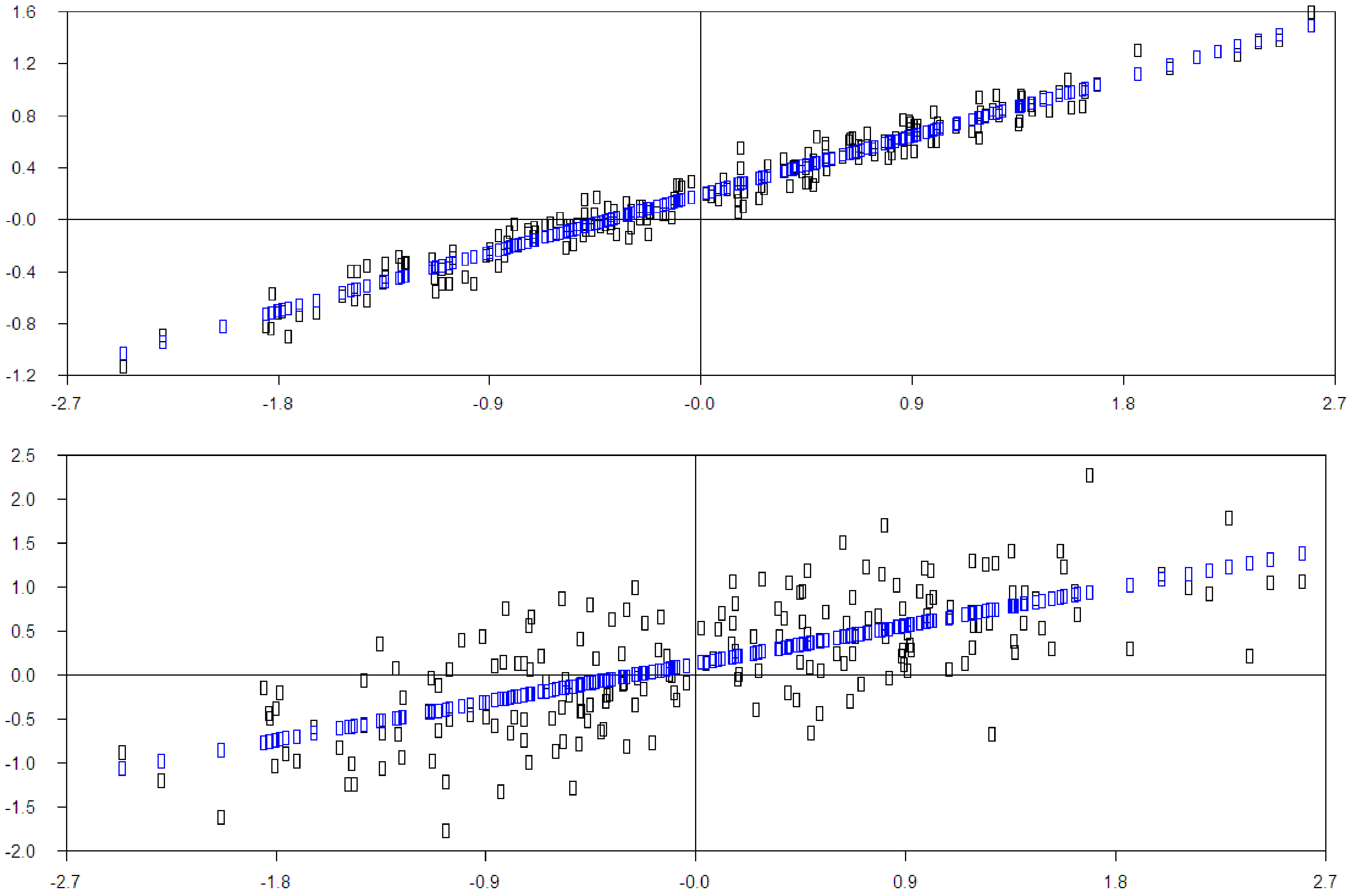
La relation n'est pas forcément déterministe

ie: il existe une v.a. (e) telle que : $e_i \sim N(0,\sigma^2)$ et

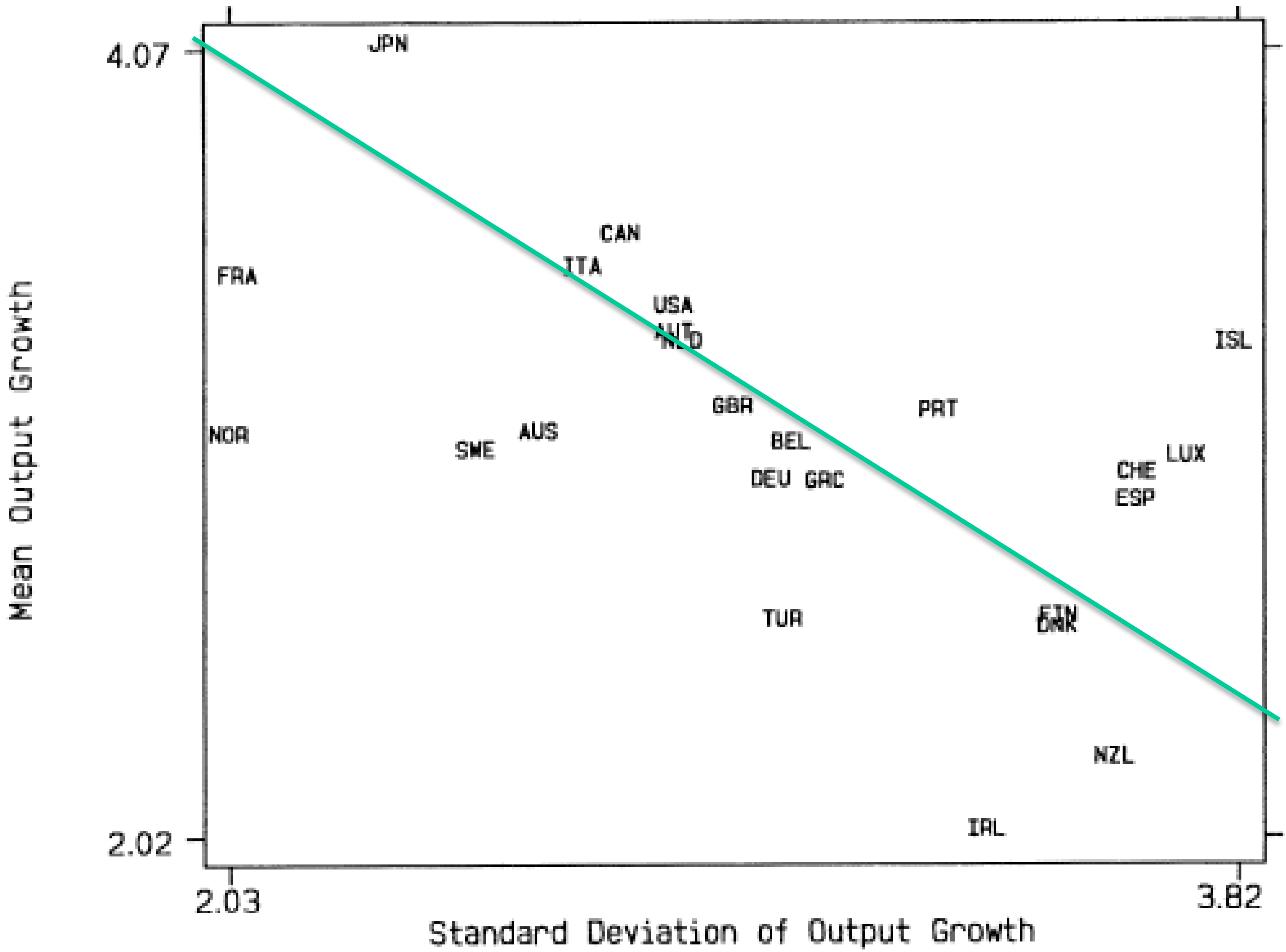
$$Y_i = a X_i + b + e_i$$

Les paramètres (a, b, σ^2) sont inconnus

Effet de la variance résiduelle sur un nuage de points (voir simulations sous RATS)

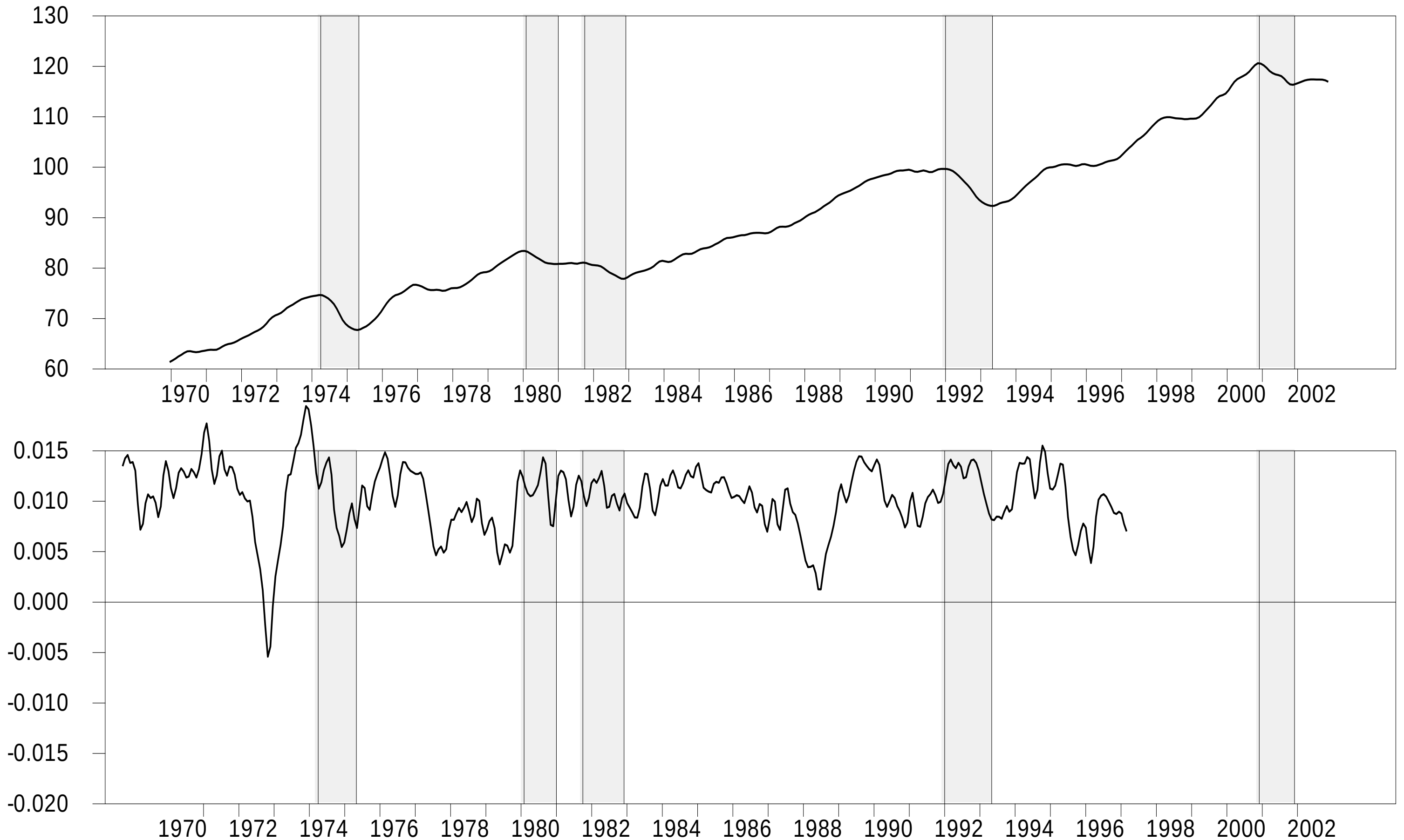


Exemple: Ramey and Ramey (AER)



Modèles de séries chronologiques

Indice de la production industrielle en zone euro



Modèles de séries chronologiques

Les n individus *i* deviennent n dates *t*,

On mesure une variable R_t pour t = 1, …, T

Pb: Qu’en est-il de l’hypothèse i.i.d. ?

1/ Indépendance : Hypothèse pas raisonnable

2/ Identiquement distribué : Hypothèse nécessaire

Modèles de séries chronologiques

Les n individus *i* deviennent n dates *t*,

On suppose qu'il existe une relation linéaire entre elles, ie:

pour chaque t :

$$Y_t = a\, Y_{t-1} + b$$

Les paramètres (a,b) sont inconnus

La relation n'est pas forcément déterministe, ie: il existe la va e telle que : $e_t \sim N(0,\sigma^2)$ et

$$Y_t = a\, Y_{t-1} + b + e_t$$

Les paramètres (a,b,σ^2) sont inconnus

Modèles de séries chronologiques déterministes

On suppose qu'il existe une relation non forcément linéaire entre elles, ie:

pour chaque i :

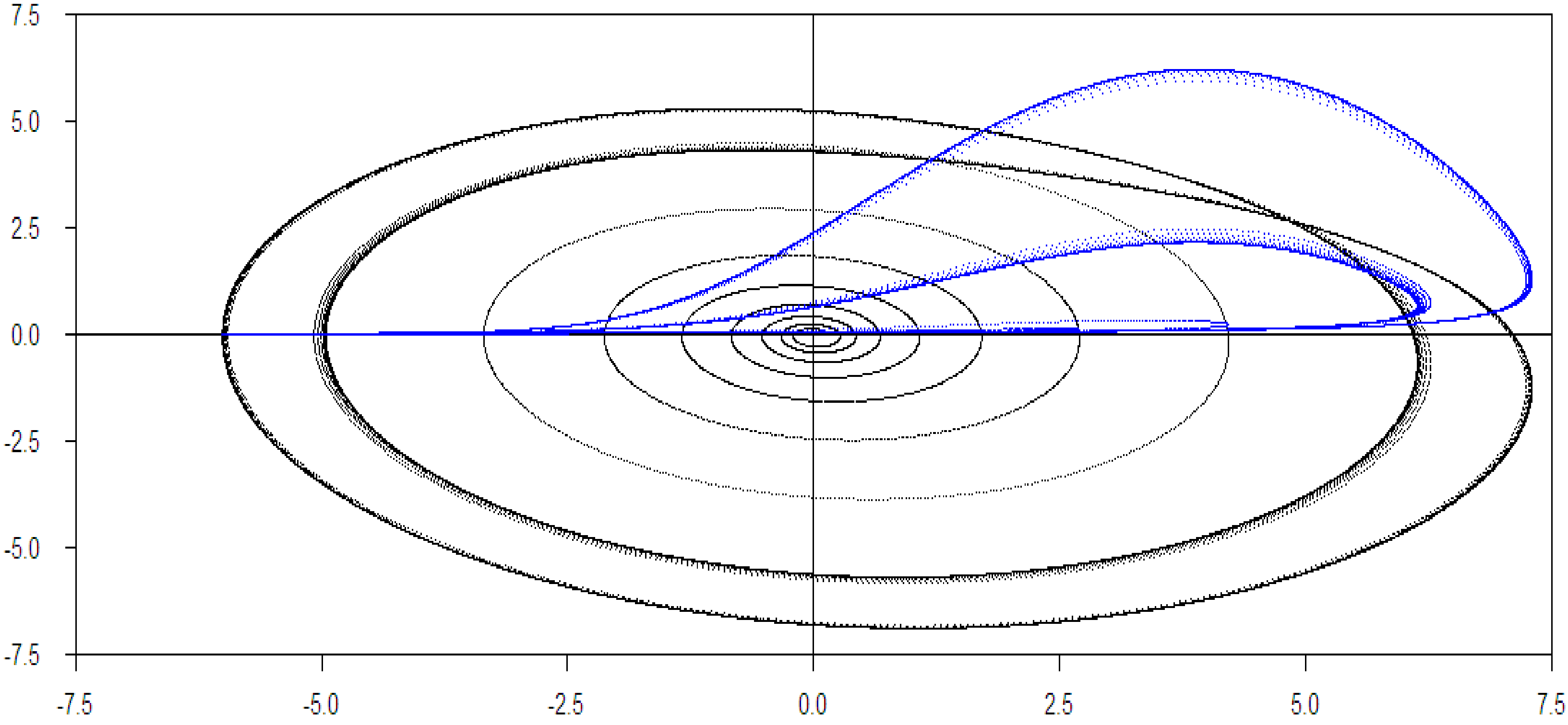
$$Y_t = f(\theta, Y_{t-1})$$

Les paramètres (θ) sont inconnus

Exemples: Fonctions Chaotiques: logistique, Rossler, Henon

```
com nn = 10000
com aa = 3.75
com xx(1) = 0.8
do i=2,nn
com xx(i) = aa*xx(i-1)*(1-xx(i-1))
enddo i
```

Squelette de séries chronologiques déterministes (Rossler)



Remarques

- Quel est le but du jeu de toute tentative de modélisation d'une variable Y ?

$\rightarrow$ **Minimiser la variance résiduelle**

Y = partie déterministe + partie aléatoire

$Y = f(X) + \varepsilon$

Par indépendance, $\mathbf{V(Y) = V(f(X)) + V(\varepsilon)}$

Conclusion

$\rightarrow$ Un modèle économétrique permet de faire un choix économique à partir d'un ensemble d'information

Algorithme de modélisation statistique:

- Définir l'ensemble d'information
- Spécifier le modèle statistique (= identifier P_θ^n)
- Estimation du modèle
- Validation / Contrôle
- Prise de décision / Prévision

Schéma de mise en œuvre d'une modélisation

Analyse des données — *Distributions, lien linéaire,…*

↓

Modèle de régression linéaire — *Choix des variables*

↓

Estimation des paramètres du modèle — *MCO, MLE*

↓

Validation du modèle — *Tests d'hypothèses, Analyse de la variance*

↓

Utilisation du modèle en prévision — *Points et IC*

Conclusion

Outils:

- Eléments de la théorie des probabilités
- Construction d'échantillon
- Choix de la classe de modèle paramétrique
- Méthodes de spécification du modèle
- Estimation des paramètres
- Tests d'hypothèses