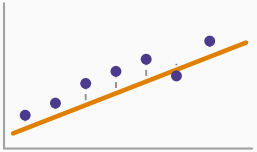
Introduction à l'économétrie

Chapitre 3 : Le modèle économétrique



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Plan du chapitre

Ajouter l’aléa

Le modèle économétrique

$$C_i = \beta_0 + \beta_1 R_i + \varepsilon_i$$

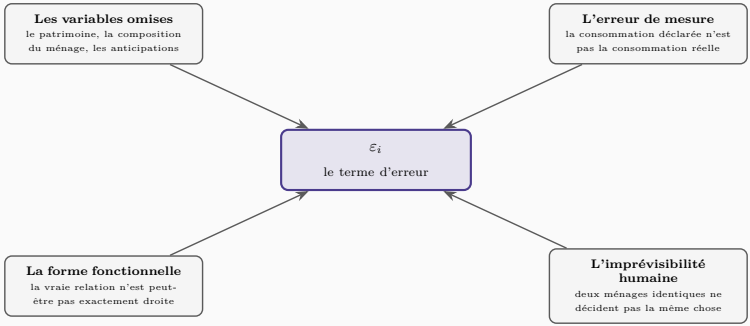
Le terme ε_i est l'**erreur** — ou perturbation, ou aléa. Il transforme une identité fausse en modèle statistique vrai.

Le modèle n'affirme plus la même chose

Sans ε : « ce ménage consomme exactement cela ». Avec ε : « il consomme cela **en moyenne**, à un écart près ».

La première proposition est réfutée par la première observation venue. La seconde est testable.

Ce que contient le terme d'erreur



Le terme d'erreur n'est pas une faute : c'est l'aveu, écrit dans le modèle, de tout ce qu'il ne contient pas.

Les quatre sources n'ont pas le même statut. Les trois dernières sont *inévitables* et sans gravité : elles ajoutent du bruit, l'estimation reste juste en moyenne.

La première est d'une tout autre nature. Si une variable omise est corrélée à celles du modèle, l'erreur l'est aussi — et l'estimation devient **biaisée**. C'est le sens de l'hypothèse $E(\varepsilon_i \mid X_i) = 0$, et c'est la seule que l'économétrie ne sait pas réparer par une formule. Le chapitre 17 lui est consacré.

Les quatre sources n'ont pas le même statut

Trois sont bénignes

L'**erreur de mesure**, la **forme fonctionnelle** approchée et l'**imprévisibilité humaine** ajoutent du bruit.

Elles rendent l'estimation moins précise. Elles ne la rendent pas fausse.

La quatrième ne l'est pas

Les **variables omises** sont d'une tout autre nature.

Si une variable absente du modèle est liée à celles qui y figurent, elle contamine l'erreur — et l'estimation devient **biaisée**. Tout le cours tourne autour de cette distinction.

L’hypothèse décisive

L'exogénéité

$$E(\varepsilon_i \mid X_i) = 0$$

Connaissant la valeur de la variable explicative, l'erreur vaut zéro en moyenne. Autrement dit : **l'erreur ne contient rien de systématiquement lié à X.**

Ce qu'elle interdit

Elle n'interdit pas que l'erreur soit grande — le modèle peut expliquer peu.

Elle interdit qu'elle soit **orientée** selon X. Si les ménages à fort revenu étaient aussi ceux qui consomment davantage pour d'autres raisons — une famille plus nombreuse, par exemple — la pente estimée capterait les deux effets et attribuerait au revenu ce qui ne lui revient pas.

Pourquoi elle porte tout le reste

La conséquence

Sous cette hypothèse, et sous elle seule, les estimateurs des moindres carrés sont **sans biais** :

$$E(\hat{\beta}_1) = \beta_1$$

Le contraste avec les autres hypothèses

Les hypothèses du chapitre 12 concernent la **précision** : elles décident si les écarts-types sont justes. Quand elles tombent, une correction technique répare tout.

L'exogénéité concerne la **justesse elle-même**. Quand elle tombe, aucune formule ne répare : **il faut changer de modèle ou de méthode.** C'est la ligne de partage de l'économétrie entière.

Erreur et résidu

Deux mots que l'on confond souvent

$$\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i) \qquad e_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

L'**erreur** se mesure à la vraie droite : elle est **inconnue**. Le **résidu** se mesure à la droite estimée : il est **calculé**.

C'est la distinction du cours d'inférence, appliquée point par point

m et $\overline{x}$, σ et s, β_1 et $\hat{\beta}_1$: à chaque fois, un paramètre inconnu et son estimation.

Ici, la même chose pour chaque observation. **Et c'est le résidu, seule quantité observable, qui servira à diagnostiquer l'erreur** — tout le chapitre 13 repose là-dessus.

Ce que le modèle affirme, exactement

La lecture rigoureuse

$$E(C \mid R) = \beta_0 + \beta_1 R$$

La droite ne prétend pas passer par les points : elle décrit **la moyenne de C pour chaque valeur de R.**

Une régression est une moyenne conditionnelle

C’est exactement la notion du chapitre 12 des probabilités — la distribution conditionnelle et sa moyenne.

La nouveauté n’est pas le concept : c’est qu’on suppose ces moyennes **alignées**, ce qui permet de les résumer par deux nombres au lieu d’une infinité.

Une conséquence pratique

Ce qu'on peut et ne peut pas prédire

Le modèle prédit correctement la consommation **moyenne** des ménages à revenu donné.

Il ne prédit pas la consommation d'un ménage particulier — l'écart individuel lui échappe par construction. **C'est pourquoi le chapitre 7 donnera deux intervalles de prévision, et non un seul.**

La modestie du modèle est sa force

Un modèle qui prétendrait tout expliquer serait invérifiable. Celui-ci annonce d'emblée ce qu'il ignore, et le range dans ε.

C'est cette honnêteté d'écriture qui rend possible de mesurer sa fiabilité.

Ce qu’il faut retenir

Le résumé du chapitre

Six points

1. $C_i = \beta_0 + \beta_1 R_i + \varepsilon_i$: le terme d'erreur transforme une identité fausse en modèle testable.
2. Il contient variables omises, erreurs de mesure, approximation de forme, et imprévisibilité.
3. Seules les **variables omises** sont dangereuses ; les trois autres sources ne font que du bruit.
4. $E(\varepsilon_i \mid X_i) = 0$ est l'hypothèse décisive : sans elle, l'estimation est biaisée.
5. **Erreur** inconnue, **résidu** calculé — comme m et $\overline{x}$.
6. La droite décrit une **espérance conditionnelle**, non les points eux-mêmes.

La suite

Le modèle est posé. Il reste à en estimer les deux paramètres à partir des données.

La méthode est celle du cours de statistique descriptive — les moindres carrés — mais on va enfin dire **pourquoi** ce critère-là plutôt qu'un autre.