# Conclusion

Jusqu'à présent, on a considéré que lors de l'estimation du modèle linéaire, les hypothèses suivantes étaient respectées :

*Hypothèses probabilistes (stochastiques)*

- les X sont observés sans erreur (non aléatoires)
- en moyenne, le modèle est bien spécifié : $E(u_i) = 0$
- *homoscédasticité* : la variance de l'erreur est constante :

$$\forall i = 1, \ldots, T ; V(u_i) = \sigma^2$$

- *Absence d'autocorrélation des erreurs* : $E(u_i, u_j) = 0$
- L'erreur est indépendante des variables explicatives *(exogénéité)* :

$$Cov(u_i, X_i) = E(u_i X_i) = 0$$

Sous ces hypothèses, nous avons montré que l'estimateur des moindres carrés ordinaires est le meilleur estimateur possible parmi les estimateurs linéaires sans biais.

Or, dans la pratique, plusieurs questions se posent :

- comment être sûr que ces hypothèses soient vraiment respectées ? (tests de ces hypothèses)
- si elles ne sont pas respectées, que faire ? Quelles sont les conséquences statistiques et existe-t-il des procédures d'estimation plus fiables que les MCO ?

# Quelques exemples

## *1. Erreur sur la spécification du modèle*

Ce problème recouvre en fait plusieurs sources possibles de difficultés dont les conséquences ne sont pas identiques. On trouvera ainsi :

- l'omission de variables explicatives pertinentes,
- la prise en compte de variables explicatives non pertinentes.

Il est encore possible de ranger dans cette catégorie les aspects concernant les erreurs de mesure sur les variables.

# Quelques exemples

## *1. Erreur sur la spécification du modèle*

- Prise en compte de variables non pertinentes

Supposons que le modèle pertinent fasse appel à *k* variables explicatives :

$$y_i = \beta_1 + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i$$

Supposons que le modèle soumis à estimation retienne *n* variables explicatives supplémentaires :

$$y_i = \beta_1 + \beta_2 x_{i2} + \ldots. + \beta_k x_{ik} + \beta_{k+1} x_{ik+1} + \ldots. + \beta_{k+n} x_{ik+n} + u_i$$

Quelles sont alors les conséquences de la prise en compte de ces variables non pertinentes sur les propriétés des estimateurs par MCO ?

# Quelques exemples

## *1. Erreur sur la spécification du modèle*

- Prise en compte de variables non pertinentes

On peut montrer que les estimateurs MCO sont non-biaisés mais ne sont plus à variance minimale : l'inclusion de variables non pertinentes provoque une augmentation de la variance des estimateurs des coefficients des variables pertinentes.

En pratique, l'inclusion de variables non pertinentes dans la liste des variables explicatives affectent donc la qualité des tests d'hypothèses. Cela revient à travailler avec un seuil de risque réel inférieur au seuil explicitement pris. On sera donc amené à accepter l'hypothèse nulle plus souvent qu'il ne le faudrait pour respecter ce seuil explicite.

# Quelques exemples

## *1. Erreur sur la spécification du modèle*

- Omission de variables pertinentes

Ici, un certain nombre de variables pertinentes sont absentes de la liste des variables explicatives dans l'équation estimée. Le vrai modèle est :

$$y_i = \beta_1 + \beta_2 x_{i2} + .... + \beta_k x_{ik} + \beta_{k+1} x_{ik+1} + .... + \beta_{k+n} x_{ik+n} + u_i$$

Mais seulement *k* variables pertinentes sont présentes dans l'estimation réalisée :

$$y_i = \beta_1 + \beta_2 x_{i2} + ... + \beta_k x_{ik} + u_i$$

Quelles sont alors les conséquences de cette omission sur les propriétés des estimateurs des k coefficients ?

# Quelques exemples

## *1. Erreur sur la spécification du modèle*

- Omission de variables pertinentes

On peut montrer qu'en général, ces estimateurs sont biaisés.

=> il apparaît que l'oubli de variables pertinentes est plus dommageable que la prise en compte de variables non pertinentes.

- Erreurs de mesure sur les variables

On peut montrer qu'en général, des erreurs de mesure sur les variables explicatives entraînent un biais des estimateurs.

La solution la plus employée consiste à recourir à la technique des variables instrumentales

# Quelques exemples

## 2. *Autocorrélation des erreurs*

Il y a autoccorléation des erreurs lorsque $E(u'u) \neq \sigma^2 I$

L'autocorrélation des erreurs peut être observée pour plusieurs raisons :

- absence d'une variable explicative importante
- mauvaise spécification du modèle.

L'autocorrélation des erreurs se rencontre essentiellement dans les modèles en séries temporelles où l'influence d'une erreur (due à une mauvaise spécification) d'une période sur l'autre est plausible.

# Quelques exemples

## 2. *Autocorrélation des erreurs*

En présence d'autocorrélation, les estimateurs MCO sont sans biais mais ne sont plus à variance minimale.

La solution consiste à recourir à la technique des moindres carrés généralisés.

$\hat{\beta}_{MCG}$ est sans biais et à variance minimale : il est meilleur que l'estimateur MCO dans le cas d'autocorrélation des erreurs

# Quelques exemples

## 3. *Hétéroscédasticité*

Il y a hétéroscédasticité lorsque les variances des erreurs ne sont plus constantes

$$\Omega_u = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_T^2 \end{bmatrix} \neq \sigma^2 I$$

Les variances des erreurs ne sont plus constantes sur la première diagonale.

Ce problème se rencontre plus fréquemment pour les modèles spécifiés en coupe instantanée ou bien sur données groupées où chaque observation est une moyenne pour un groupe et les groupes de tailles différentes.

# Quelques exemples

## *3. Hétéroscédasticité*

Les causes de l'hétéroscédasticité sont multiples :

- observations représentant des moyennes

- répétition d'une même valeur de la variable à expliquer pour des valeurs différentes d'une variable explicative (ex: regroupement en tranches)

- lorsque les erreurs sont liées aux valeurs prises par une variable explicative.

Ex : les familles à faibles revenus dépensent relativement peu en loisirs. Les variations de ces dépenses entre ces familles sont donc faibles. Mais pour les familles avec des revenus importants, le montant moyen dépensé en loisirs sera plus élevé, mais il y aura une plus grande variabilité entre de telles familles. D'où l'hétéroscédasticité.

# Quelques exemples

## 3. *Hétéroscédasticité*

Les conséquences de l'hétéroscédasticité sont identiques à celles de l'autocorrélation des erreurs : les estimateurs MCO sont sans biais mais ne sont plus à variance minimale.

Avec la méthode des MCO, on donne des poids égaux à toutes les observations, quand en fait, les observations avec la variance la plus grande contiennent moins d'information que les observations avec les plus faibles variances.

L'estimateur BLUE du modèle hétéroscédastique est celui des moindres carrés généralisés (MCG) :

$$\boxed{\begin{aligned} \hat{\beta}_{MCG} &= (X'\Omega_u^{-1}X)^{-1}(X'\Omega_u^{-1}Y) \\ \Omega_{\hat{\beta}_{MCG}} &= (X'\Omega_u^{-1}X)^{-1} \end{aligned}}$$