

# MULTICOLINEARITE

Source : D.N.Gujarati, "Basic Econometrics", Third Ed., McGraw Hill, 1995

## ***Le concept***

Le terme multicollinéarité vise les phénomènes d'interdépendance (de corrélation) entre variables explicatives. On distinguera la multicollinéarité parfaite de la multicollinéarité partielle. Dans le premier cas, une variable explicative est une combinaison linéaire parfaite des autres variables explicatives. Dans le second cas, une variable explicative est fortement corrélée à une ou plusieurs variables explicatives (ou à l'une de leur combinaison).

## ***L'estimation en présence de multicollinéarité***

**Multicollinéarité parfaite** : l'estimation des paramètres est impossible, la matrice des données étant singulière. Les coefficients du modèle sont indéterminés et leur variance est infinie.

**Multicollinéarité partielle** : les coefficients du modèle de régression peuvent être déterminés (ils gardent leur propriété BLUE) mais l'écart-type de leur estimation est important. Les coefficients ne peuvent donc pas être estimés avec beaucoup de précision.

## ***Conséquences pratiques***

La détection de la présence de multicollinéarité n'est pas toujours évidente. On peut bien entendu calculer les coefficients de corrélations entre les variables explicatives prises deux à deux mais cela ne teste que l'existence du phénomène à un niveau d'ordre 1. Le même problème peut se poser au niveau deux (une variable explicative est fortement corrélée avec la combinaison de deux autres variables), au niveau trois, ... Une solution alternative est de réaliser une analyse en composantes principales et d'étudier l'évolution de variance expliquée en fonction du nombre de facteurs retenus.

En cas de multicollinéarité parfaite, la sanction est simple : l'algorithme des moindres carrés "plante". En cas de multicollinéarité classique, on constate en générale :

- que la variance de l'estimation des paramètres tend à être très forte;
- que par conséquent, l'intervalle de confiance autour des paramètres s'élargit considérablement;
- que les tests en t tendent à devenir peu significatifs;
- que malgré cela, le coefficient de détermination peut être très élevé;
- que l'estimation des paramètres est très sensible à la constitution de l'échantillon.

Face à ce problème, il n'y a pas d'autre choix que de revoir la construction du modèle (une autre approche possible est de passer par l'analyse en composantes principales pour orthogonaliser les variables mais on perd alors l'interprétabilité directe des coefficients associés aux variables explicatives).