

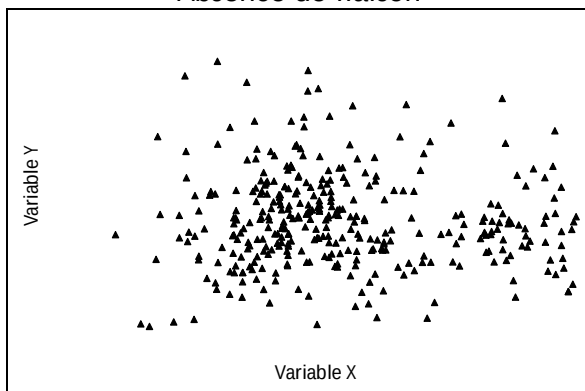
Une brève introduction à l'économétrie¹

L'économétrie est un ensemble de techniques qui servent à identifier et à quantifier des liaisons entre variables économiques. De ce point de vue, elle ne fait que perfectionner et objectiver les méthodes élémentaires de la statistique descriptive.

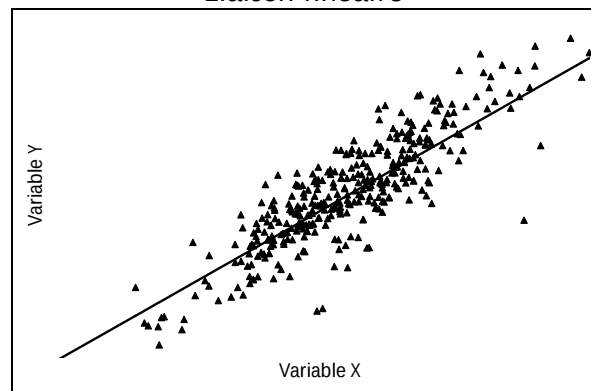
Imaginons que l'on dispose de plusieurs observations d'un couple de variables X et Y, représentant par exemple la consommation et le revenu. Chaque observation peut correspondre à un individu, et on parlera alors de données individuelles. Mais ces observations peuvent aussi porter sur l'ensemble des ménages et correspondre à des années successives : on parlera alors de séries temporelles. Quand on combine ces deux dimensions, individuelles et chronologiques, on parle alors de panel.

Si l'on porte sur un graphique les points observés, deux situations sont possibles. Soit on trouve une « patate » qui ne fait apparaître aucune liaison entre Y et X (graphique 1) ; soit les points adoptent au contraire une disposition régulière plus ou moins vague qui fait apparaître une liaison entre les variables. La forme la plus simple d'une telle liaison, à laquelle on peut souvent se ramener en transformant les variables, est une liaison linéaire : les points sont presque alignés, et l'on est tenté alors de tracer « à la main » une droite résumant cette relation (graphique 2).

Graphique 1
Absence de liaison



Graphique 2
Liaison linéaire



L'économétrie est au fond une technique pour limiter l'arbitraire de cet exercice. Puisqu'il s'agit de trouver la meilleure droite, l'économétrie objective les moyens d'y parvenir. Dans sa version la plus simple, la technique économétrique choisit de faire passer la droite par le point moyen (de coordonnées X moyen et Y moyen) et ensuite de minimiser la somme des carrés des écarts de chaque point à la droite. Autrement dit, on fait pivoter une droite autour du point moyen et on calcule une sorte de « distance de la droite au nuage de points » que l'on cherche à rendre la plus petite possible. C'est pourquoi cette droite, que l'on a tracée sur le graphique 2, s'appelle « droite des moindres carrés ». En pratique, au lieu d'utiliser le papier millimétré, on charge les données sur un logiciel adéquat, qui calcule les paramètres de la droite, c'est-à-dire la pente et l'ordonnée à l'origine.

Tout cela revient à poser un « modèle » linéaire $Y = aX + b$ où Y est appelée variable dépendante (ou « expliquée »), et X variable indépendante (ou « explicative »). La technique économétrique est ce qui permet d'« estimer » les valeurs des paramètres a et b, pour obtenir par exemple une liaison de la forme $Y = 0,8 X + 10$. On peut ensuite se libérer de la représentation graphique et passer dans un espace à plusieurs dimensions, ce

¹ Extrait de Michel Husson, [Les ajustements de l'emploi](#), Editions Page Deux, 1999

qui permet de multiplier le nombre des variables indépendantes ou « explicatives ». Avec deux variables explicatives, le nuage de points occupe un espace à trois dimensions ; mais au-delà, la représentation graphique devient évidemment impossible.

Le danger du calcul des coefficients de la droite des moindres carrés est cet exercice se fait « en aveugle », puisqu'il peut être effectué, même dans le cas de la « patate », où il n'a manifestement plus grand sens. On sent bien qu'il existe dans le choix de la meilleure droite une imprécision d'autant plus grande que le nuage de points est dispersé, mal aligné, etc. Il ne suffit donc pas d'estimer des coefficients mais de tester la validité même de cette opération. Là encore, la technique statistique propose une batterie de tests permettant d'évaluer objectivement la « précision » et la « significativité » du modèle.

Petit guide de lecture des résultats économétriques

La présentation d'une équation économétrique doit être accompagnée d'un certain nombre d'informations qui permettent d'en apprécier la pertinence. L'équation économétrique donnée ci-dessous à titre d'illustration permet d'estimer un modèle qui « explique » V, le nombre d'heures de travail, par deux variables « explicatives », la production Q et le temps T.

V = 0,481 Q - 2,998 T + 8,367		
(4,8)	(20,1)	(16,0)
1971-1993	R ² =0,983	SE=0,022
DW=0,36		
V = Nombre d'heures de travail		
Q = Production		
T = Temps		

Sur la première ligne figure le modèle identifié, avec la valeur estimée des coefficients. Au-dessous de chacun des coefficients figure entre parenthèse une grandeur que l'on appelle le « t de Student » et qui permet d'apprécier la significativité de chacune des variables. L'interprétation approximative en est simple : si t est trop bas en valeur absolue (inférieur à 2 en pratique), alors la variable correspondante est faiblement significative, parce que la valeur estimée se situe au milieu d'une plage qui incorpore zéro. Puisque son coefficient pourrait aussi bien être nul, la variable doit donc être retirée du modèle.

Le R² est le coefficient de corrélation qui mesure la significativité globale du modèle explicatif. Ce coefficient rapporte la variance expliquée à la variance totale : plus il est voisin de l'unité, meilleure est l'explication. Ce coefficient est cependant trompeur car il dépend en partie de la spécification du modèle : si la variable expliquée suit une tendance temporelle, la variance associée à ce *trend* est facilement « expliquée » par le temps de telle sorte que le R² est plus élevé que celui qu'on obtiendrait sur un modèle semblable mais spécifié par exemple en taux de croissance.

Deux autres valeurs (et bien d'autres correspondant à des tests plus sophistiqués) peuvent être fournies. SE (pour *Standard Error*) désigne l'écart résiduel moyen entre valeur observée et valeur simulée et permet donc de mesurer la précision du modèle. Quant au DW (Durbin et Watson) il permet de tester la bonne répartition des résidus qui sont censés être distribués au hasard donc indépendants les uns des autres. Quant le DW s'écarte trop de 2, on considère que cette hypothèse n'est pas corroborée, et il convient alors de modifier le modèle. Dans l'exemple donné plus haut, on cherchera par exemple à introduire la valeur de V à la période précédente dans la liste des variables explicatives.