

La régression de quantiles en pratique

Pauline GIVORD¹, Xavier D'HAULTFOEUILLE²

La très grande majorité des études empiriques économiques portent sur les effets moyens des variables auxquelles on s'intéresse. Cet état de fait s'explique par la diffusion d'un outil méthodologique, la régression linéaire, simple à utiliser et bien maîtrisé. La moyenne apporte une information essentielle, mais cependant limitée: le revenu moyen n'informe pas, par exemple, sur leur répartition plus ou moins inégale. En terme d'évaluation des politiques publiques, on peut considérer comme souhaitable une mesure qui n'aurait d'impact positif que pour les plus défavorisés, même si son effet moyen est négligeable (par exemple, une mesure permettant d'améliorer le niveau scolaire des élèves les plus en difficultés, sans élever le niveau moyen). Cette limitation est de plus en plus contestée, le public ne se reconnaissant pas toujours dans cette construction virtuelle que serait l'individu moyen. L'une des préconisations du rapport Stiglitz-Sen-Fitoussi appelle ainsi à sortir de la « dictature de la moyenne » en présentant plus souvent des analyses sur la répartition des revenus.

Les régressions de quantile est un outil dont dispose l'économètre pour répondre à ce type de demande. Elles permettent en effet d'étudier l'impact de différents facteurs sur l'ensemble de la distribution d'une variable d'intérêt. Si leur principe est ancien, elles ont connu récemment un regain d'intérêt (voir par exemple Buchinski, 1998 et Koenker et Hallock, 2001). Un ensemble de procédures préprogrammées en font aujourd'hui un outil simple d'utilisation.

Cet article se propose de présenter un guide d'utilisation pratique de ces méthodes. Il commence par décrire le principe de l'estimation par régression quantiles, ainsi que les propriétés statistiques des estimateurs obtenus et les procédures disponibles dans les logiciels statistiques standard. Le principe des régressions de quantile est de s'intéresser à l'impact des covariables sur les quantiles conditionnels de la variable d'intérêt. Un exemple d'application sur les déterminants des salaires en France permet de mieux comprendre l'interprétation des résultats qu'on peut obtenir par ces régressions, parfois plus délicate que dans le cas standard. On propose également plusieurs extensions. En particulier, on relie ces résultats avec la littérature d'économétrie de l'évaluation, qui a proposé l'étude d'un paramètre particulier, le « quantile treatment effect ». Ceci conduit naturellement à s'interroger sur le traitement de l'endogénéité dans les régressions de quantiles. On met en particulier l'accent sur deux méthodes: l'estimation sous hypothèse d'indépendance conditionnelle aux observables (Firpo, 2007), et l'estimation par variables instrumentales (Chernozukhov et Hansen, 2008). Ces résultats permettent de montrer, à travers l'exemple d'un programme de formation aux personnes sans emploi, que les effets moyens de ce type de programme masquent une grande hétérogénéité dans les bénéfices que les personnes peuvent en retirer.

¹ pauline.givord@insee.fr, INSEE, 15, boulevard Gabriel Péri 92 240 MALAKOFF

² xavier.dhautfoeuille@ensae.fr, INSEE, 15, boulevard Gabriel Péri 92 240 MALAKOFF