La régression quantile en pratique *

Xavier D'Haultfœuille† Pauline Givord‡

1er juillet 2014

Résumé

Les régressions quantiles sont des outils statistiques permettant de décrire l'impact de variables explicatives sur une variable d'intérêt. Elles permettent une description plus riche que les régressions linéaires classiques, puisqu'elles s'intéressent à l'ensemble de la distribution conditionnelle de la variable d'intérêt et non seulement à la moyenne de celle-ci. En outre, elles peuvent être plus adaptées pour certains types de données (variables censurées ou tronquées, présence de valeurs extrêmes, modèles non linéaires...). Ce document propose une introduction à ces outils, en insistant sur les détails de leur implémentation pratique par les logiciels statistiques standards (SAS, R, Stata). Il peut également être utilisé comme guide d'interprétation des études mobilisant ces méthodes, grâce aux deux applications concrètes exposées en détail. Enfin, il présente, pour un public plus averti, des extensions récentes, autour en particulier du problème d'endogénéité (variables instrumentales, données de panel...).

Classification JEL : C1, C2.

Mots clés : Régression quantile, hétérogénéité, robustesse, instruments, panels.

*Les auteurs remercient les très nombreuses personnes qui ont contribué par leurs commentaires sur les premières versions successives de ce document à l'améliorer significativement, et plus particulièrement Cédric Afsa, Didier Blanchet, Pascale Breuil, Elise Coudin, Jean-Michel Floch, Marine Guillerm, Jérôme Le, Simon Quantin et Olivier Sautory. Nous restons seuls responsables des erreurs et approximations qui pourraient demeurer dans ce document.

†CREST. xavier.dhaultfoeuille@ensae.fr

‡INSEE, Département des Méthodologies Statistiques, direction de la méthodologie et de la coordination internationale. pauline.givord@insee.fr

Introduction

La majorité des études économiques empiriques se concentrent sur la modélisation de la moyenne. Celle-ci apporte une information essentielle mais néanmoins limitée. Le revenu moyen n'informe pas, par exemple, sur la répartition plus ou moins inégale de ces revenus dans la population. Par exemple, on a observé sur les dernières années une très faible augmentation du revenu moyen, alors que le revenu du dernier décile et surtout du dernier percentile a fortement augmenté (voir les travaux de Landais, 2007, et pour une actualisation Solard, 2010). L'une des préconisations du rapport Stiglitz-Sen-Fitoussi appelle donc à sortir de la « dictature de la moyenne » en présentant plus souvent des analyses sur la répartition des revenus. L'intérêt d'analyser l'ensemble de la distribution de la variable d'intérêt et de ses déterminants ne se limite pas à la mesure des inégalités. Pour évaluer l'effet d'une politique publique, il est souvent pertinent d'aller au-delà des effets moyens de celle-ci. Il peut ainsi être souhaitable de mettre en œuvre une politique éducative qui permet de réduire la proportion d'élèves en grande difficulté, même si elle n'a qu'un effet négligeable sur le niveau moyen de l'ensemble des élèves.

Par ailleurs, la moyenne conditionnelle s'avère parfois difficile à modéliser. Cela peut être le cas en présence de valeurs extrêmes ou aberrantes (dues par exemple à des erreurs de mesures), auxquelles la moyenne est bien plus sensible que les quantiles. Lorsque la distribution de la variable d'intérêt est très étalée, ce qui est par exemple le cas des revenus, la moyenne pourra beaucoup varier en fonction de l'échantillon utilisé. L'estimation de la moyenne est également compromise en présence de données censurées, c'est-à-dire lorsqu'on n'observe la variable d'intérêt qu'au-delà ou en deçà d'un seuil fixe. Ainsi, pour des raisons de confidentialité, les données individuelles de revenus sont parfois diffusées en écrêtant ceux qui sont supérieurs à un certain niveau. Il n'est pas possible d'estimer la moyenne d'une variable censurée de la sorte, sauf à faire des hypothèses paramétriques sur la distribution de cette variable au-dessus du seuil. En revanche, en-dessous de ce niveau, les quantiles de la variable censurée coïncident avec ceux de la variable d'intérêt (cf. Buchinsky, 1994, pour une application à l'évolution des revenus aux Etats-Unis).

La régression quantile est un outil dont dispose l'économètre pour répondre à ces limites inhérentes à la moyenne. Elle permet d'avoir une description plus précise de la distribution d'une variable d'intérêt conditionnelle à ses déterminants qu'une simple régression linéaire, qui se focalise sur la moyenne conditionnelle. Si son principe est ancien [1], elle a connu récemment un regain d'intérêt. Un ensemble de procédures préprogrammées en font aujourd'hui un outil simple d'utilisation. L'objet de cet article est de préciser les principes de cette méthode et l'utilisation qui peut en être faite, en particulier pour l'analyse microé-

1. La méthode des moindres déviations absolues (Least Absolue Deviation, LAD), qui revient à s'intéresser à la médiane plutôt qu'à la moyenne comme dans les moindres carrés ordinaires, remonte à Gauss et Laplace.

conomique [2]. Le lecteur intéressé pourra trouver une présentation plus détaillée en anglais dans Koenker & Hallock (2001), Cade & Noon (2003) ou encore Koenker (2005).

Ce document est destiné à plusieurs usages. Les lecteurs souhaitant simplement disposer d'un guide d'interprétation pour les articles mettant en œuvre ce type de méthode tireront surtout profit, après la partie 1 qui présente les principaux intérêts de la régression quantile, de la dernière partie qui détaille deux applications. Un lecteur néophyte souhaitant mettre en œuvre des régressions quantiles sous sa forme classique complétera cette lecture par la partie 2 qui rappelle les principes d'estimation sous-jacents et les propriétés statistiques des estimateurs. Enfin, un lecteur plus averti trouvera dans la partie 3 plusieurs extensions, en particulier sur la prise en compte de l'endogénéité et les modèles non-linéaires. Les parties les plus avancées en termes économétriques et pouvant être sautées en première lecture sont suivies d'un astérisque.

1 L'intérêt des régressions quantiles

Pour poser les notations, nous nous intéressons à une variable aléatoire Y, de fonction de répartition F_Y ($F_Y(y) = P(Y \leq y)$). Rappelons que le quantile d'ordre τ est généralement défini par

$$q_\tau(Y) = \inf\{y : F_Y(y) \geq \tau\}.$$

Si F_Y est continue, on retrouve la propriété intuitive $P(Y < q_\tau(Y)) = \tau$ [3]. Les quantiles les plus couramment utilisés sont la médiane ($\tau = 0,5$), les premier et dernier déciles ($\tau = 0,1$ et $\tau = 0,9$), et les premier et dernier quartiles ($\tau = 0,25$ et $\tau = 0,75$). Dans le langage courant, il y a parfois confusion entre la valeur du quantile et les personnes pour lesquelles la valeur de Y se situe en-dessous de ce quantile. Par exemple, on parle du « premier décile » pour désigner les 10% de la population les moins riches. Cette désignation est incorrecte. Le premier décile désigne stricto sensu le seuil de revenus en-dessous duquel 10% exactement de la population se situe.

1.1 Modéliser la dépendance de la distribution de la variable d'intérêt aux explicatives

Les régressions quantiles tentent d'évaluer comment les quantiles conditionnels $q_\tau(Y|X)$, définis par $q_\tau(Y|X) = \inf\left\{y : F_{Y|X}(y) \geq \tau\right\}$, se modifient lorsque les déterminants $X =$

2. On trouvera dans Cornec (2010) une application pour prévoir la conjoncture économique (voir également la note de conjoncture de l'Insee de décembre 2011) à partir de données temporelles, qui n'est pas développé dans cet article.

3. Cette égalité n'est cependant pas vraie dans le cas général, voir l'annexe pour une discussion de ce point et de quelques propriétés utilisées par la suite.

$(1, X_2, ..., X_p)'$ de Y varient. Il n'y a pas de raison en effet de supposer que l'impact d'une de ces caractéristiques X_k soit le même aux différents quantiles de la distribution conditionnelle de Y. On peut en trouver une illustration dans les classiques courbes de croissance utilisées dans les carnets de santé. Elles montrent comment la distribution du poids, ou de la taille, varie en fonction de l'âge. Plus précisément, elles représentent certains percentiles (traditionnellement les $3^{\text{ème}}$, $25^{\text{ème}}$, $75^{\text{ème}}$ et $97^{\text{ème}}$) de ces distributions conditionnelles à l'âge. Elles permettent ainsi de vérifier que la croissance d'un enfant est « normale » en le situant dans la distribution correspondant à son âge. On observe sur de telles courbes que les différents percentiles des poids augmentent de manière à peu près linéaire avec l'âge sur les premiers mois. Mais les taux de croissance correspondants diffèrent suivant le percentile : les droites ne sont pas parallèles.

Ce type de modélisation graphique est possible et utile lorsque l'on s'intéresse à un seul déterminant, mais atteint vite ses limites pour étudier simultanément l'effet de plusieurs caractéristiques sur la variable d'intérêt. Les régressions quantiles permettent justement d'étudier ce cadre multivarié : plus précisément, elles tentent de déterminer comment les quantiles de la distribution conditionnelle $F_{Y|X}$ varient en fonction de X.

Dans la régression quantile standard, on suppose que ces quantiles de la distribution conditionnelle ont une forme linéaire :

$$q_\tau(Y|X) = X'\beta_\tau \tag{1}$$

où à chaque τ correspond donc un vecteur de coefficients $\beta_\tau = (\beta_{1\tau}, ..., \beta_{p\tau})'$ correspondant aux p variables explicatives (dont la constante)[4]. Pour la suite, il peut être utile de remarquer que cette expression peut s'écrire de manière équivalente :

$$Y = X'\beta_\tau + \epsilon_\tau, \text{ avec } q_\tau(\epsilon_\tau|X) = 0 \tag{2}$$

La condition 1 est à rapprocher de celle effectuée dans la régression linéaire standard, dans laquelle on modélise la moyenne conditionnelle de la variable d'intérêt Y comme une expression linéaire des variables explicatives X : $E(Y|X) = X'\beta$. Une différence importante est qu'ici, on autorise les coefficients à différer d'un quantile à l'autre. Ceci apporte une information supplémentaire qui ne ressort pas d'une simple régression linéaire. Pour bien comprendre les implications de ce dernier point, considérons quelques exemples.

4. Cette dépendance linéaire n'exclut pas une dépendance plus compliquée des quantiles par rapport à certaines variables explicatives. Par exemple, dans l'exemple des courbes de croissance, les quantiles conditionnels à l'âge ne sont pas linéaires au-delà de la première année. En revanche, cette dépendance serait probablement bien approchée en utilisant non pas l'âge en niveau mais en logarithme, ou encore en utilisant une forme polynomiale de cette variable.

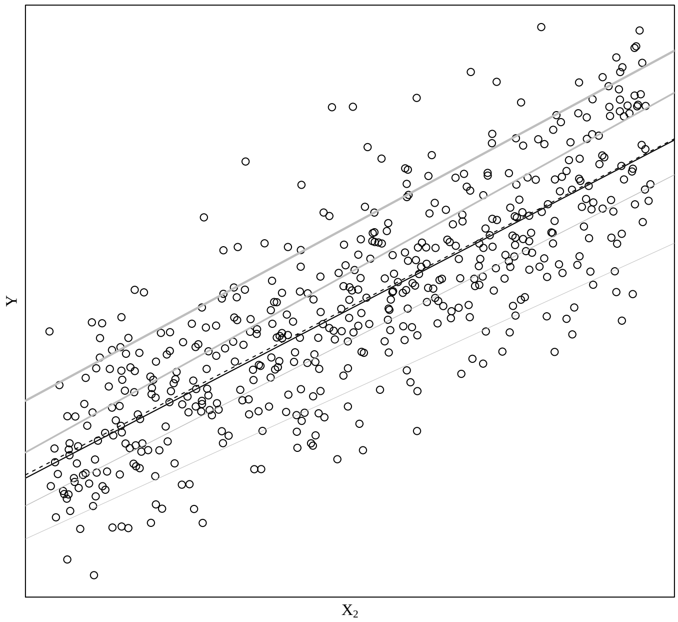
1.1.1 Le modèle de translation simple

Le premier exemple suppose que les variables explicatives n'ont d'impact que sur la moyenne de la variable d'intérêt (et pas sur sa variance par exemple). Il s'agit du modèle de translation linéaire :

$$Y = X'\gamma + \varepsilon \tag{3}$$

où ε est indépendant de X et de moyenne nulle. Sous cette hypothèse, les résidus sont en particulier homoscédastiques, i.e., $V(\varepsilon|X) = \sigma^2$. Dans ce modèle, les fonctions quantiles correspondant à différents τ sont parallèles puisque $q_\tau(Y|X) = X'\gamma + q_\tau(\varepsilon)$. On est donc bien dans le cadre de l'hypothèse 1, mais ici seul le coefficient correspondant à la constante varie en fonction de τ : $\beta_{1,\tau} = \gamma_1 + q_\tau(\varepsilon)$, tandis que $\beta_{k,\tau} = \gamma_k$ pour $k > 1$. On parle dans ce cas d'*homogénéité des pentes*. On en trouvera une illustration dans le graphique 1, dans le cas simple d'une régression univariée ($X = (1, X_2)$). Les droites correspondant aux régressions quantiles sont parallèles. Leur pente commune, $\beta_{2,\tau} = \gamma_2$, correspond à l'effet, constant ici, d'une augmentation de X_2 de x_2 à $x_2 + 1$, à ε fixé. 0

Cela signifie également que pour tout τ, les coefficients $(\beta_{k,\tau})_{k=2,\dots,p}$ sont les mêmes que ceux correspondant à la modélisation de la moyenne conditionnelle $E(Y|X) = X'\gamma$. Les résultats obtenus par régression quantile ou par une régression linéaire estiment donc les mêmes paramètres, par des méthodes différentes. Il peut parfois être intéressant d'utiliser les estimateurs des régressions quantiles plutôt que ceux des moindres carrés ordinaires. Ils sont plus robustes par exemple à la présence de valeurs aberrantes, et peuvent être plus précis pour certaines distributions de ε (cf. partie 2.3 ci-dessous). Enfin, on ne sait pas en général si Y suit réellement un modèle de translation. Les résultats quantiles permettent de tester cette restriction. Elle implique en effet que les estimateurs de régressions quantiles effectuées pour différents τ doivent être très proches.



Lecture : droites correspondant aux régressions quantiles pour les déciles d'ordre 1,3,7 et 9 (en gris), la médiane (en noir) et à une régression linéaire classique (en noir pointillé).

FIGURE 1 – Exemple de données distribuées selon un modèle de translation

1.1.2 Le modèle de translation-échelle

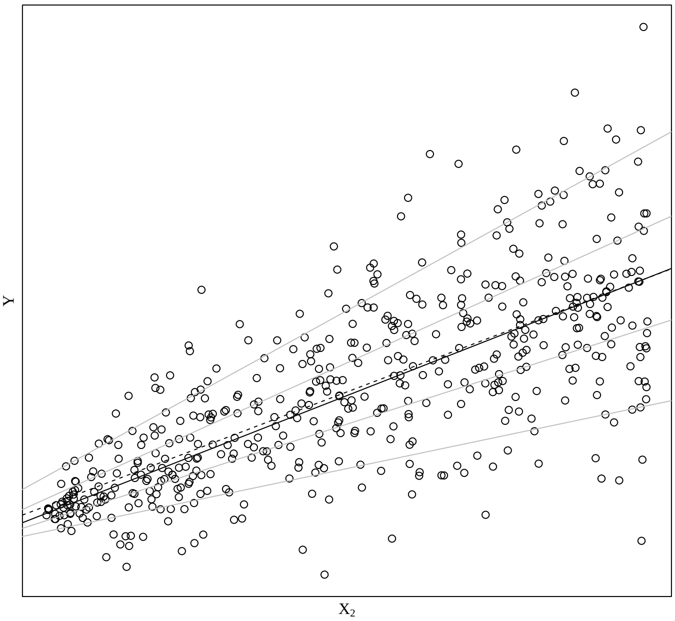
Le deuxième exemple, un peu plus général que le modèle de translation, suppose que ces déterminants ont non seulement un impact sur la moyenne mais aussi sur la variance de la variable d'intérêt. Ces modèles, appelés « translation-échelle », correspondent à une certaine forme d'hétéroscédasticité :

$$Y = X'\gamma + (X'\alpha)\varepsilon \tag{4}$$

avec encore une fois ε indépendant de X, de moyenne nulle et $X'\alpha > 0$. Dans un tel modèle, la dispersion de la variable dépendante est plus importante pour certaines valeurs de X. Un exemple classique est celui des salaires, qui sont plus dispersés pour les diplômés du supérieur que pour les personnes sans diplôme (cf. l'exemple de la partie 4.1). Le modèle de translation-échelle correspondant à l'équation (4) implique que $q_\tau(Y|X) = X'(\beta + q_\tau(\varepsilon)\alpha)$. Ainsi, l'hypothèse (1) est bien vérifiée, avec $\beta_\tau = \beta + q_\tau(\varepsilon)\alpha$. L'impact des variables explicatives ne sera pas le même pour les différents quantiles, et il n'y a plus homogénéité des pentes. Dans l'exemple des salaires, l'effet du diplôme est faible pour les premiers quantiles ($\beta_{k,\tau}$ petit pour τ proche de 0) et plus fort pour les derniers quantiles ($\beta_{k,\tau}$ grand pour τ proche de 1).

Le modèle de translation-échelle est illustré par le graphique 2, dans le cas univarié. Ici les pentes $\beta_{2,\tau}$ correspondant aux différentes régressions quantiles sont croissantes avec τ (soit

$\alpha_2 > 0$), ce qui traduit une dispersion d'autant plus grande que X_2 est élevé (comme dans l'exemple du diplôme). Cette information ne ressort pas d'une régression linéaire standard, qui se contente d'estimer le coefficient γ : on a en effet toujours $E(Y|X) = X'\gamma$.



Lecture : droites correspondant aux régressions quantiles pour les déciles d'ordre 1, 3, 7 et 9 (en gris), la médiane (en noir) et à une régression linéaire classique (en noir pointillé).

FIGURE 2 – Exemple de données distribuées selon un modèle de translation échelle

1.1.3 Un cadre plus général : le modèle à coefficients aléatoires*

Une étape supplémentaire dans la généralisation est franchie par le modèle à coefficients aléatoires, qui s'écrit :

$$Y = X'\beta_U, \quad \text{où } U \text{est indépendant de } X \text{ et de loi uniforme sur } [0,1]. \tag{5}$$

On suppose également que la fonction $u \mapsto x'\beta_u$ est strictement croissante pour tout x [5]. Dans ce modèle, U peut s'interpréter comme une composante individuelle inobservée qui positionne l'individu dans la distribution de Y. Par exemple si on veut modéliser le niveau de salaire en fonction de caractéristiques observables comme le niveau d'étude, U pourrait correspondre à un niveau de productivité « intrinsèque » du salarié, qui implique en particulier que le niveau d'études ou d'autres caractéristiques ont des effets variant d'une personne à l'autre. Ce modèle à coefficients aléatoires repose sur des hypothèses très

5. Ce modèle vérifie bien l'hypothèse (1) puisque par indépendance de U et X et croissance de $u \mapsto x'\beta_u$,

$$P(Y \leq x'\beta_\tau | X = x) = P(x'\beta_U \leq x'\beta_\tau) = P(U \leq \tau) = \tau.$$

flexibles sur la dépendance en U. Il généralise les deux exemples précédents. Le modèle de translation linéaire correspond à un cas où le coefficient correspondant à la variable explicative $k > 1$, $\beta_{k,U}$, est indépendant de U. Dans le modèle de translation-échelle, on a $\beta_U = \gamma + q_U(\varepsilon)\alpha$.

Ce modèle à coefficients aléatoires fournit une interprétation intéressante du coefficient β_τ dans (1). Il implique en effet que si l'on modifie marginalement la variable observable X à U constant, l'effet sur la variable d'intérêt Y est égal à β_U. Ainsi, β_τ correspond à l'effet marginal de X pour les individus au $\tau^{\text{ième}}$quantile de la distribution des caractéristiques inobservées U. Si l'on s'intéresse par exemple à la modélisation des salaires en fonction du niveau d'études, la composante de $\beta_{0.5}$ correspondant au nombre d'années d'études mesurera l'effet marginal d'une augmentation de celui-ci, pour les salariés dont la productivité intrinsèque U se situe à un niveau médian.

Cette interprétation n'est possible ici que parce qu'on a supposé l'existence d'un unique U, quelle que soit la valeur de X. Ceci implique la propriété dite d'invariance des rangs, qui signifie dans notre exemple que la productivité « intrinsèque » mentionnée plus haut est la même quel que soit le niveau d'études ou les autres variables observables utilisées. Il est important de comprendre que cette propriété restrictive, sur laquelle nous revenons plus loin, n'est pas requise dans les régressions quantiles. Toujours dans le cadre d'un modèle à coefficients aléatoires, on pourrait supposer par exemple que $Y = X'\beta_{U_X}$, avec U_x uniforme sur $[0, 1]$ pour tout x (la fonction $u \mapsto x'\beta_u$ étant toujours supposée strictement croissante). Dans ce cas, l'hypothèse (1) serait encore vérifiée, mais plus la propriété d'invariance des rangs.

Notons enfin que le modèle à coefficients aléatoires, bien que plus général que les modèles de translation ou de translation-échelle, repose encore sur une forme très spécifique de dépendance entre Y, X et l'effet inobservable. En toute généralité, cette dépendance peut s'écrire $Y = g(X, U)$ où g est une fonction quelconque. Avec des hypothèses similaires à celle du modèle (5), le quantile conditionnel de Y correspond donc à $q_\tau(Y|X) = g(X, \tau)$. Le modèle de régression quantile (1) peut donc se voir comme une approximation linéaire de cette fonction g. Cette approximation est une des raisons pour laquelle on peut obtenir des estimations incohérentes. En principe, la fonction $\tau \mapsto g(x, \tau)$ doit être strictement croissante pour tout x, puisque $\tau \mapsto q_\tau(Y|X = x)$ est strictement croissante. Mais la forme linéaire implique que si les pentes ne sont pas parallèles et si X a un large support, les différentes courbes $x \mapsto x'\beta_\tau$ indicées par τ se croiseront nécessairement pour certains quantiles $\tau \neq \tau'$ [6].

6. Notons que même si les courbes théoriques ne se croisent pas, ceci peut également se produire sur les courbes estimées $x \mapsto x'\widehat{\beta}_\tau$, en particulier lorsque l'échantillon sur lequel se base l'estimation est petit. Plusieurs solutions ont été proposées dans la littérature pour traiter de cette propriété indésirable (voir par exemple Chernozhukov et al., 2010).

1.1.4 Interprétation des régressions quantiles

La manière dont les distributions conditionnelles se modifient en fonction des variables explicatives renvoie à plusieurs questions. La première est simplement de décrire comment les quantiles conditionnels se modifient en fonction d'un changement d'un de ces déterminants, sans tenter de savoir si ce sont des personnes « comparables » aux différents quantiles conditionnels. Les régressions quantiles sont des outils développés pour répondre à cette question. Prenons l'exemple de l'évaluation d'un programme spécifique d'apprentissage de la lecture. On cherche en particulier à comparer les scores à des épreuves standardisées des élèves dans des classes ayant bénéficié de ce programme avec ceux d'autres classes n'en ayant pas bénéficié. Si les deux groupes d'élèves étaient similaires initialement, ce qui est le cas par exemple si les classes ayant bénéficié du programme étaient tirées au sort, les régressions quantiles permettent de conclure sur l'impact de ce programme sur le premier décile. Autrement dit, on peut affirmer que grâce à ce programme, le niveau minimum atteint par 90% des élèves a augmenté de x points.

A condition de faire l'hypothèse simple mais restrictive d'invariance des rangs, elles peuvent également permettre de répondre à une deuxième question, qui est de savoir comment Y varie suite à ce même changement d'un déterminant, pour les personnes se trouvant initialement à un certain niveau de la distribution conditionnelle de Y. Dans l'exemple du programme d'apprentissage de la lecture, il s'agit d'évaluer de combien augmente le score individuel des élèves dont le niveau en l'absence du programme expérimental se situait au niveau du premier décile. L'hypothèse d'invariance des rangs signifie ici qu'à l'intérieur d'une classe, les élèves sont toujours « classés » de la même manière entre eux, qu'ils bénéficient ou non du programme d'apprentissage. D'une manière générale, cette hypothèse n'est pas testable, et il convient d'évaluer au cas par cas dans quelle mesure elle semble plausible pour pouvoir interpréter les résultats obtenus de cette manière.

Enfin, toujours sous l'hypothèse d'invariance de rang, on peut également répondre à une troisième question, qui est la détermination de la distribution des effets pour chaque élève du programme d'apprentissage de la lecture sur le score individuel. Cette distribution permet de savoir, par exemple, que pour 10% des élèves, le niveau du score a augmenté d'au moins x points.

Pour bien comprendre que les réponse à ces trois questions peuvent être différentes, il peut être utile de considérer une situation très simple. Supposons que la population d'élèves est composée de cinq types en proportion identique, qu'on identifiera par les lettres de A à E. Par ailleurs, $X = (1, X_2)$, où $X_2 = 1$ si l'élève a bénéficié du programme spécifique de lecture, 0 sinon. X_2 est supposé tiré au sort, si bien qu'on observe les différents types dans les mêmes proportions lorsque $X_2 = 0$ et $X_2 = 1$. Suivant la valeur de X_2, les élèves peuvent avoir des valeurs différentes de Y dont les valeurs sont détaillées dans le tableau 1. Les effets du programme d'apprentissage correspondent pour un élève à un

Table 1 – Caractéristiques de la population fictive

Type	Valeur de Y si $X_2 = 0$ (1)	Valeur de Y si $X_2 = 1$ (2)	ΔY (2)-(1)
A	1	4	*3*
B	2	*6*	4
C	*4*	5	1
D	5	11	6
E	9	10	1
Mediane	*4*	*6*	*3*
Moyenne	4.2	7.2	3

Note : Les cinq "types" sont supposés en proportion égale dans la population. Lecture : conditionnellement au fait d'avoir $X_2 = 0$ (respectivement $X_2 = 1$), la valeur médiane de Y est 4 (resp. 6). Le résultat d'une régression quantile de l'effet de X_2 sur la médiane serait donc $6 - 4 = 2$. Ce résultat est donc différent de la médiane de l'effet de X_2 sur Y (ΔY), qui vaut 3.

passage de $X_2 = 0$ à $X_2 = 1$. Une régression quantile d'ordre 0.5 (*regression médiane*) de Y sur X_2 mesure l'écart entre la médiane de la distribution de Y conditionnelle à $X_2 = 0$ et la médiane de la distribution conditionnelle à $X_2 = 1$. Il vaut donc ici 2 (6- 4). Cette valeur est différente de l'effet pour les élèves médians quand $X_2 = 0$ (i.e., tels que $Y = q_{0.5}(Y|X_2 = 0)$) de passer de $X_2 = 0$ à $X_2 = 1$. Ces élèves sont ceux de types C, et pour eux $\Delta Y^C = 1$.

Cette différence vient du fait qu'ici, les élèves ne sont pas ordonnés (en termes de Y) de la même manière lorsque $X_2 = 0$ et $X_2 = 1$, autrement dit que l'hypothèse d'invariance des rangs n'est pas vérifiée. Dans notre exemple artificiel, certains élèves sont plus à l'aise avec la méthode spécifique d'apprentissage de la lecture, d'autres moins. Ainsi, les élèves de type C sont « devant » ceux de type B lorsque $X_2 = 0$ mais derrière eux lorsque $X_2 = 1$. L'hypothèse d'invariance des rangs n'est donc pas vérifiée, et on ne peut pas interpréter le coefficient de la régression quantile d'ordre 0.5 comme l'effet d'un passage de $X_2 = 0$ à $X_2 = 1$ pour les élèves médians quand $X_2 = 0$. Enfin, l'effet individuel du changement de méthode, ΔY, a une médiane de 3. On constate ici que la médiane de la distribution de ΔY ne correspond pas à la différence des médianes des distributions de la variable d'intérêt, ou en termes mathématiques, $q_{0.5}(\Delta Y) \neq q_{0.5}(Y|X_2 = 1) - q_{0.5}(Y|X_2 = 0)$.

Ces remarques permettent de bien cadrer les usages qui peuvent être faits, ou non, des résultats d'une régression quantile. Il est utile de faire le lien avec ceux obtenus par une régression linéaire classique. L'objet principal de celle-ci est de modéliser la manière dont la moyenne conditionnelle varie en fonction de déterminants : sur notre exemple, cela correspondrait à 3 ($= 7,2 - 4,2$). Du fait de la linéarité de la moyenne, la différence des moyennes conditionnelles correspond également à l'effet moyen de l'augmentation de X_2, c'est-à-dire à la moyenne de ΔY . En revanche, cette différence ne correspond pas à ce que gagnerait un élève, dont la valeur de Y est proche de la moyenne conditionnelle lorsque $X_2 = 0$ (dans notre exemple, il s'agit encore de C), s'il passait à $X_2 = 1$. Contrairement aux régressions quantiles, la condition d'invariance des rangs ne suffirait pas ici pour obtenir une telle interprétation.

1.2 Des estimateurs plus adaptés à certaines situations

1.2.1 Robustesse aux valeurs aberrantes ou à la dispersion des résidus

Au-delà de leur apport en termes d'interprétation, la régression quantile devrait être préférée à la régression linéaire dans certains cas. La première raison est qu'elle est robuste aux valeurs aberrantes de la variable d'intérêt ou à la présence d'erreurs très dispersées. Intuitivement, cette propriété est due au fait que les quantiles sont moins sensibles que la moyenne à la présence de valeurs très grandes [7]. Supposons que la variable Y^* vérifie le modèle de translation simple (3), mais que dans de très rares cas, les données observées ne correspondent pas à la variable d'intérêt Y^* mais à une variable erronée, éventuellement corrélée avec les variables explicatives X. Formellement, on observe $Y = AX'\delta + (1-A)Y^*$, où A est une variable inobservée valant 1 lorsque Y est aberrant, 0 sinon, avec $P(A = 1|X, \varepsilon) = p$ petit. Une régression linéaire de la variable observée (avec erreur) Y sur X donnera une estimation biaisée de notre paramètre d'intérêt γ, puisqu'elle sera égale, pour un échantillon de taille tendant vers l'infini, à $\gamma_{MCO} = \gamma + p(\delta - \gamma)$ [8]. Si δ est très différent de γ, le terme de biais peut être important même lorsque la probabilité d'observer des valeurs erronées p est faible. En revanche, on montre dans l'annexe A.1.2 que si les valeurs aberrantes $X'\delta$ sont très grandes, l'estimateur de l'effet de X_k ($k > 1$) obtenu par une régression quantile vaut bien γ_k. En d'autres termes, la présence de valeurs aberrantes n'affecte pas les résultats de la régression quantile, sauf les coefficients de la constante. Dans le cas plus général, si on suppose que Y^* vérifie (5), cette propriété n'est plus exactement vérifiée, mais elle reste une bonne approximation dès lors que la proportion p de valeur aberrante est faible. On obtient en effet par la régression quantile le coefficient $\beta_{\frac{\tau}{1-p}}$ au lieu de β_τ (voir encore l'annexe A.1.2).

Par ailleurs, même lorsque les données sont toutes correctement observées, la distribution sous-jacente peut être telle qu'une régression linéaire n'est pas adaptée. C'est le cas en particulier lorsque ε, dans le modèle de translation simple, peut prendre des valeurs très grandes avec une probabilité non négligeable (on parle de distribution à queue épaisse).

7. Par exemple, les deux échantillons (1,5,6,8,11) et (1,5,6,8,999999) ont la même médiane mais des moyennes très différentes.

8. On néglige ici les erreurs d'estimation liées au fait que l'échantillon est de taille finie. Autrement dit, nos résultats doivent se comprendre comme étant les valeurs limites des estimateurs lorsque la taille de l'échantillon tend vers l'infini.

Ainsi, lorsque l'on mesure les patrimoines ou les très hauts revenus, les résidus peuvent être très dispersés au sens où l'on peut observer des valeurs très élevées, qui ne sont pas des valeurs aberrantes. C'est la raison pour laquelle on les modélise en général par des lois de Cauchy ou certaines lois de Pareto, qui n'ont pas d'espérance. Dans ce cas, l'estimateur des moindres carrés ordinaires n'est pas convergent : même pour des échantillons énormes, il pourra prendre des valeurs très différentes du vrai paramètre β. A l'inverse, l'estimateur obtenu par régression quantile sera convergent car intuitivement, cet estimateur ne dépend pas des queues de distribution de ε mais seulement de la distribution de ε autour du quantile considéré (cf. partie 2.3).

1.2.2 Etude des inégalités

Par ailleurs, dans la mesure où elles permettent d'étudier l'ensemble de la distribution d'intérêt, les régressions quantiles constituent un outil privilégié pour l'étude des inégalités. Les estimations fournies par les régressions quantiles permettent par exemple de déterminer si les salaires sont plus ou moins dispersés parmi les détenteurs d'un niveau de diplôme par rapport à un autre niveau, en comparant les coefficients obtenus par des régressions quantiles pour les différents quantiles (voir par exemple l'illustration dans l'analyse des résultats de la régression quantile des salaires) : une mesure plus synthétique des inégalités intra groupes sera fournis par l'étude des différences interdéciles $(q_{0.9}(Y|X) - q_{0.1}(Y|X))$ ou interquartiles $(q_{0.75}(Y|X) - q_{0.25}(Y|X))$ que l'on peut reconstituer simplement grâce à l'approximation linéaire du quantile conditionnel. En évolution, on peut s'intéresser à l'évolution des determinants des inégalités de salaires. L'un des articles pionniers sur cette question est Buchinsky (1994) qui s'intéresse à l'évolution des inégalités aux Etats-Unis sur longue période. Ce type d'analyse permet en particulier d'étudier comment les rendements de l'éducation et de l'expérience, les principaux facteurs explicatifs du revenu, ont évolués sur la période (par exemple du fait du progrès technique). Plus récemment, Charnoz et al. (2011*b*) reproduisent ce type d'analyse pour la France. L'estimation à plusieurs dates permet par exemple d'étudier comment les inégalités de salaires ont évolué au cours du temps à l'intérieur d'un groupe de diplôme ou entre les groupes, en modélisant les différents quantiles prédits à partir des estimations obtenues, et à caractéristiques observables fixées. L'évolution temporelle des coefficients obtenus (correspondant aux rendements de l'éducation et de l'expérience dans une équation de Mincer classique) est en elle-même intéressante pour fournir des pistes d'interprétation des conséquences en termes d'inégalités des évolutions institutionnelles (salaire minimum en particulier) et économiques (progrès technique biaisé par exemple).

En revanche, les régressions quantiles ne sont pas adaptées directement pour répondre à une question connexe, qui met l'accent moins sur les facteurs économiques expliquant l'évolution des inégalités (comment les rendements de l'éducation évoluent avec le temps) que sur les effets de composition (les diplômés sont de plus en plus nombreux, or les salaires des diplômés sont plus dispersés). Ce type d'approche a été popularisé sous le nom de décomposition d'Oaxaca-Blinder. En pratique, il s'agit de quantifier la part des différences observées dans les salaires de deux groupes (ici, deux dates différentes, mais ces méthodes sont souvent utilisées pour détecter d'éventuelles discriminations salariales, entre les hommes et les femmes par exemple) qui ne s'explique pas simplement par des différences de composition. Il ne s'agit pas ici de décrire ces méthodes, mais pour mémoire cette décomposition repose sur le fait que pour deux groupes (noté A et B), en supposant que les rendements des facteurs observables puissent être différents, on a :

$$E(Y_A) - E(Y_B) = (E(X_A) - E(X_B))'\beta_A + E(X_B)'(\beta_B - \beta_A)$$

où $E(Z_i)$ correspond à la moyenne de la variable Z dans le groupe i. La première partie correspond à la part « expliquée » des inégalités (différences de structure entre les deux groupes), et le reste la part inexpliquée (voir par exemple **?** pour une utilisation sur données françaises). En pratique, il s'agit donc d'estimer l'équation sur un des groupes, et d'utiliser les coefficients obtenus et les moyennes des caractéristiques observables dans les deux groupes pour obtenir la part expliquée.

Il n'est pas possible d'appliquer cette méthode à partir de régressions quantiles. La décomposition repose en effet sur la propriété de linéarité de l'espérance, que ne partage pas la fonction quantile : $q_\tau(Y) \neq E(q_\tau(Y|X))$. Dit autrement, le quantile de la variable d'intérêt sur l'ensemble de la population ($q_\tau(Y)$) ne correspond pas à la moyenne sur la population des quantiles conditionnels ($E(q_\tau(Y|X))$). Sur les années récentes, de nombreuses solutions permettant d'étendre les méthodes de décomposition aux différents quantiles ont été proposées. On en trouvera une synthèse dans Fortin et al. (2011).

1.2.3 Estimation de modèles non-linéaires

Enfin, une propriété importante des quantiles est qu'ils sont invariants par une transformation monotone : si g est une fonction croissante continue à gauche, on a $q_\tau(g(Y)) = g(q_\tau(Y))$ (voir l'annexe pour une preuve). Cette propriété n'est bien sûr pas vérifiée par l'espérance. Ceci rend les régressions quantiles plus naturelles et simples à utiliser dans des modèles non-linéaires [9] comme les modèles à censure fixe (cf. la partie 3.4 pour davantage de détails), ou les modèles de durées (voir par exemple Bilias & Koenker, Fitzenberger & Wilke, 2001, 2005). On peut trouver aussi des applications aux modèles de comptage (Machado & Silva, 2005).

9. Cette propriété signifie aussi qu'on pourra facilement déduire l'effet marginal d'un déterminant du salaire (par exemple) sur un de ces quantiles conditionnels à partir d'une régression quantile modélisant le logarithme du salaire. On aura en effet $q_\tau(W|X) = \exp(X\beta_\tau)$ où β_τ correspond au coefficient estimé par la régression quantile de $\log(W)$ sur X.

2 Principes statistiques et mise en œuvre pratique

2.1 Définition de l'estimateur et propriétés statistiques*

Pour bien comprendre le principe des régressions quantiles, il est utile de détailler comment on peut estimer les quantiles d'une variable d'intérêt Y à partir d'un échantillon $(Y_i)_{i=1...n}$ de variables supposées i.i.d. La manière la plus intuitive de calculer l'estimateur standard $\widehat{q}_\tau(Y)$ consiste à ordonner ces n variables, le quantile d'ordre τ étant fourni par la $[n\tau^{ieme}]$ observation où $[n\tau]$ est le plus petit entier supérieur ou égal à $n\tau$ [10]. Mais il est plus utile, pour le passage aux régressions quantiles, de remarquer qu'on a également [11] (voir l'annexe pour une preuve) :

$$\widehat{q}_\tau(Y) = \arg\min_b \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(Y_i - b). \tag{6}$$

où $\rho_\tau(.)$ est une « fonction test » définie par $\rho_\tau(u) = (\tau - \mathbb{1}\{u < 0\})u$. Par exemple, pour $\tau = 1/2$, c'est-à-dire si l'on s'intéresse à la médiane, la fonction test correspond simplement à la (demi-) valeur absolue.

L'intérêt de cette définition est de s'étendre simplement au cadre conditionnel qui nous intéresse, où l'on modélise le quantile conditionnel de la variable d'intérêt Y comme une fonction des variables explicatives X. Il suffit en effet de remplacer $\widehat{q}_\tau(Y)$ et b dans (6) par respectivement $q_\tau(Y|X)$ et une fonction $b(X)$. Dans le cas des régressions quantiles classiques, on peut se limiter aux fonctions linéaires puisqu'on suppose que $q_\tau(Y|X) = X'\beta_\tau$. On a alors :

$$\beta_\tau = \arg\min_\beta E[\rho_\tau(Y - X'\beta)]. \tag{7}$$

On peut noter l'analogie avec le modèle de régression linéaire classique, qui modélise l'espérance conditionnelle de Y par une forme linéaire en X : $E(Y|X) = X'\beta_0$. L'espérance d'une variable aléatoire pouvant être obtenue par $E(Y) = \arg\min_a E[(Y-a)^2]$, le coefficient β_0 est défini par $\beta_0 = \arg\min_\beta E[(Y - X'\beta)^2]$. La fonction de perte quadratique qui est utilisée dans une régression linéaire par les moindres carrés ordinaires est donc remplacée, dans la régression quantile, par la fonction test $\rho_\tau(.)$. Celle-ci augmentant de manière linéaire et non quadratique avec le résidu, les très grands écarts sont beaucoup moins pénalisés, ce qui explique la robustesse de la régression quantile aux valeurs extrêmes ou aberrantes discutées plus haut.

L'estimateur de la régression quantile est alors obtenu en remplaçant l'espérance dans (7)

10. Cet estimateur n'est pas le seul utilisé. Le logiciel R propose ainsi rien moins que 9 estimateurs différents, basés sur les définitions données dans Hyndman & Fan (1996).

11. En toute rigueur, il n'y a pas toujours unicité au programme de minimisation $\min_a E[\rho_\tau(Y-a)]$, cf. l'annexe pour une discussion. On néglige ici ces complications.

par la moyenne sur l'échantillon :

$$\widehat{\beta}_\tau = \arg\min_\beta \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(Y_i - X_i'\beta). \tag{8}$$

Dans le cas où $\tau = 1/2$, l'estimateur revient à minimiser la somme des valeurs absolues des erreurs $Y_i - X_i'\beta$. On parle de l'estimateur des moindres déviations absolues ("Least Absolute Deviation Estimator" en anglais), qui est de fait utilisée depuis longtemps comme une alternative robuste aux moindres carrés ordinaires.

L'estimation se fait en utilisant l'*ensemble* de l'échantillon. Cela ne peut donc consister à diviser l'échantillon en fonction des quantiles de la variable d'intérêt, et à effectuer des régressions linéaires sur les sous-échantillons ainsi obtenus. Cette pratique n'est de fait pas cohérente : elle revient à d'un côté « figer » les valeurs des quantiles de la variable d'intérêt, et de l'autre à faire varier cette même variable d'intérêt en fonction de variables explicatives. L'intuition qui la sous-tend repose sans doute sur la confusion qu'il existe entre le niveau des quantiles (les bornes de l'intervalle) et les personnes dont la valeur de la variable d'intérêt se situent dans ces intervalles.

Cette estimation peut se faire pour tout quantile d'ordre τ, où $\tau \in [0, 1]$. Il existe donc en principe une infinité de régressions quantiles possibles. En pratique, le nombre de quantiles qu'on estime dépendra de la taille de l'échantillon. Il est bien entendu illusoire de tenter d'approcher très finement une distribution avec un nombre fini d'observations : le nombre de quantiles empiriques distincts est restreint (de l'ordre de $n\ln(n)$, où n désigne la taille de l'échantillon, voir Portnoy, 1992). Le choix de modéliser l'ensemble des percentiles ou simplement les quartiles et la médiane dépendra non seulement du degré de précision souhaitée pour décrire la distribution mais aussi des données disponibles.

Notons enfin qu'un indicateur de la qualité de l'ajustement d'une régression quantile a été proposé par Koenker & Machado (1999). Il est défini par

$$R^1(\tau) = 1 - \frac{\min_{b\in\mathbb{R}^p} \rho_\tau(Y_i - X_i'b)}{\min_{b_0\in\mathbb{R}} \rho_\tau(Y_i - b_0)}.$$

Comme le R^2, ce critère est compris entre 0, lorsque l'estimateur des coefficients relatifs à $(X_2, ..., X_p)$ vaut 0, et 1, lorsque Y est une fonction linéaire déterministe de X. Il augmente également lorsqu'on ajoute des variables explicatives au modèle.

2.2 Algorithmes utilisés*

Il n'existe pas de solution explicite à (8), si bien qu'il faut résoudre ce programme numériquement. Un problème est que la fonction objectif n'est pas différentiable, puisque la fonction ρ_τ n'est pas dérivable en 0. Les algorithmes standards tels que celui de Newton

Raphson ne peuvent donc pas être utilisés directement. Cependant, ((8)) peut se réécrire comme un programme linéaire :

$$\min_{(\beta,u,v)\in\mathbb{R}^p\times\mathbb{R}_+^{2n}} \tau\mathbf{1}'u+(1-\tau)\mathbf{1}'v \quad s.t.\ \mathbf{X}\beta+u-v-\mathbf{Y}=0,$$

où $\mathbf{X}=(X_1,...,X_n)'$, $\mathbf{Y}=(Y_1,...,Y_n)'$ et $\mathbf{1}$ est un vecteur de 1 de taille n. Les variables supplémentaires u et v désignent respectivement les parties positives et négatives des résidus. Cette reformulation, a priori peu naturelle car elle augmente la dimension du vecteur à optimiser, est très utile car de nombreux algorithmes permettent de résoudre de tels programmes linéaires. La méthode du simplexe a été jusqu'à récemment la méthode la plus classique pour résoudre ce type de problèmes linéaires. Cependant, elle devient coûteuse en temps de calcul lorsque le nombre d'observations augmente et elle n'est donc indiquée que pour de petits échantillons. Pour des échantillons plus conséquents, les méthodes de points intérieurs sont plus performantes pour résoudre ces problèmes (Portnoy & Koenker, 1997). On trouvera dans **?** une description pratique de l'implémentation de ces méthodes dans les logiciels statistiques standards.

2.3 Propriétés asymptotiques de l'estimateur et estimation de la précision*

Les propriétés asymptotiques de $\widehat{\beta}_\tau$ sont délicates à établir car, contrairement à l'estimateur des moindres carrés, il n'existe pas de forme explicite pour $\widehat{\beta}_\tau$. Pour plus de détails, on se référera par exemple à l'ouvrage de Koenker (2005). Nous nous contentons ici du résultat principal sur la loi asymptotique de $\widehat{\beta}_\tau$.

Théorème 2.1. *Supposons que $\varepsilon_\tau = Y - X'\beta_\tau$ admette, conditionnellement à X, une densité en 0 $f_{\varepsilon_\tau|X}(0|X)$ et que $J_\tau = E\left[f_{\varepsilon_\tau|X}(0|X)XX'\right]$ soit inversible. Alors*

$$\sqrt{n}\left(\widehat{\beta}_\tau-\beta_\tau\right)\xrightarrow{d}\mathcal{N}\left(0,\tau(1-\tau)J_\tau^{-1}E\left[XX'\right]J_\tau^{-1}\right) \tag{9}$$

Dans le cas du modèle de translation simple (3), la variance asymptotique prend une forme particulièrement simple. On a en effet, $\varepsilon_\tau=\varepsilon-q_\tau(\varepsilon)$ et la variance asymptotique V_{as} s'écrit plus simplement

$$V_{\mathrm{as}}=\frac{\tau(1-\tau)}{f_\varepsilon(q_\tau(\varepsilon))^2}E\left[XX'\right]^{-1}.$$

Cette variance est très proche de celle des MCO avec résidus homoscédastiques, si ce n'est que $\sigma^2=V(\varepsilon)$ est remplacé par $\tau(1-\tau)/f_\varepsilon(q_\tau(\varepsilon))^2$. Le terme de densité est logique : seuls les résidus autour de $q_\tau(\varepsilon)$ vont apporter de l'information sur la valeur du quantile conditionnel de Y. Ce résultat explique que même dans le cas de translation simple, il peut être parfois préférable d'utiliser une régression quantile pour certaines distributions des termes inobservés ε. L'estimation par régression quantile sera plus précise qu'une estimation par MCO lorsque $\tau(1-\tau)/f_\varepsilon(q_\tau(\varepsilon))^2 < \sigma^2$.

En dehors du modèle restrictif de translation simple, la variance asymptotique est plus complexe à estimer que dans le cadre d'un modèle de régression linéaire simple. Plusieurs méthodes d'inférence ont été proposées pour construire des tests ou des intervalles de confiance sur β_τ, et il n'existe pas à l'heure actuelle de consensus sur la méthode à utiliser. On trouvera dans Kocherginsky et al. (2005) une présentation générale de ces méthodes et une discussion pratique des cas où certaines sont plus ou moins indiquées. Le choix dépend des hypothèses plus ou moins restrictives qu'on accepte de faire sur le modèle sous-jacent (modèle de translation...), de la taille de l'échantillon ou du nombre de variables du modèle.

Certaines méthode s'appuient sur une estimation directe de la variance asymptotique en partant de la formule (9). La difficulté principale de cette approche est la présence de la densité conditionnelle $f_{\varepsilon_\tau|X}(0|X)$, qui est délicate à estimer (cf. l'annexe A.2.1 pour plus de détails). Dans le cadre restrictif d'un modèle de translation-échelle, une méthode basée sur les tests de rang est parfois utilisée (cf. Koenker, 2005). Il est surtout courant de s'appuyer sur des méthodes de bootstrap. Elles consistent à générer des échantillons « factices » par des tirages avec remise à partir de l'échantillon initial, et à effectuer une régression quantile sur ces échantillons (cf. l'annexe A.2.2). L'inconvénient de ces méthodes est qu'elles sont souvent coûteuses en temps de calcul. Ce dernier augmente à la fois avec la taille de l'échantillon et le nombre de variables explicatives. Une solution récente (« Markov Chain Marginal Bootstrap » , ou MCMB) a été proposée par He & Hu (2002) pour résoudre en partie ce problème quand le nombre de variables explicatives est important.

Enfin, l'un des intérêts de la régression quantile étant de ne pas supposer *a priori* que les variables explicatives ont un effet homogène sur l'ensemble de la distribution de la variable d'intérêt, il est tout à fait possible de tester cette hypothèse à partir des estimations obtenues. Par exemple, l'homogénéité de l'effet de l'une des variables X_k correspond à l'égalité des coefficients $\beta_{k,\tau_1}, ..., \beta_{k,\tau_m}$ (où $(\tau_1, ...\tau_m)$ peuvent être par exemple l'ensemble des déciles), ce qui peut se tester simplement. Un tel test s'appuie sur la distribution jointe asymptotique de $(\widehat{\beta}_{\tau_1}, ..., \widehat{\beta}_{\tau_m})$, donnée par le résultat suivant :

$$\sqrt{n}\left(\widehat{\beta}_{\tau_k} - \beta_{\tau_k}\right)_{k=1}^{m} \xrightarrow{d} \mathcal{N}(0, V), \tag{10}$$

où V est une matrice par bloc dont le bloc $V_{k,l}$ vérifie

$$V_{k,l} = [\min(\tau_k, \tau_l) - \tau_k \tau_l]\, J_{\tau_k}^{-1} E[XX']\, J_{\tau_l}^{-1}.$$

Ce résultat est donc une généralisation du théorème 2.1 à plusieurs quantiles.

3 Extensions*

3.1 Les régressions quantiles instrumentales

Comme en régression linéaire, il arrive fréquemment que certaines composantes des variables X soient a priori endogènes. Par exemple, dans une étude sur l'impact d'un dispositif de formation sur le salaire, le fait de participer à ce dispositif peut être lié à des caractéristiques inobservées qui influent également le salaire. Dans ce cas, l'estimateur $\widehat{\beta}_\tau$ défini par (8) ne mesure pas l'effet causal du dispositif de formation.

En revanche, on peut disposer d'instruments affectant ces variables mais pas directement les composantes inobservées de la variable d'intérêt (représentées par le résidu ε_τ). Plus précisément, si l'on se place dans le cadre de la régression quantile classique, définie par (2),

$$Y = X'\beta_\tau + \varepsilon_\tau,$$

mais on suppose que certaines des variables explicatives, notées $X_1 \in \mathbb{R}^q$, sont endogènes, c'est-à-dire telles que $q_\tau(\varepsilon_\tau|X_1) \neq 0$. Les autres variables explicatives, notées X_2, sont supposées exogènes, i.e. vérifier $q_\tau(\varepsilon_\tau|X_2) = 0$. On suppose disposer par ailleurs d'instruments, notés $Z_1 \in \mathbb{R}^r$ (avec $r \geq q$). Z_1 doit être corrélé aux variables explicatives endogènes mais pas aux résidus, si bien que :

$$q_\tau(\varepsilon_\tau|Z) = 0 \tag{11}$$

avec $Z = (Z_1, X_2)$. Cette hypothèse est l'équivalent à l'hypothèse $E(\varepsilon|Z) = 0$ en régression linéaire instrumentale. Attention néanmoins, il serait tout à fait incorrect de transposer la méthode classique des doubles moindres carrés, qui consiste à régresser la variable d'intérêt non sur la variable endogène, mais sur sa « projection » sur les variables instrumentales obtenues par une première étape. Cette méthode repose de manière essentielle sur la propriété de linéarité de l'espérance, que ne vérifie pas les quantiles. En revanche, on peut proposer une méthode s'appuyant sur la condition d'exclusion (11). En effet, cette condition implique que :

$$q_\tau(Y - X_1'\beta_{1\tau}|Z) = q_\tau(X_2'\beta_{2\tau} + \varepsilon_\tau|Z_1, X_2) = X_2'\beta_{2\tau} + Z_1'\, 0_r, \tag{12}$$

où 0_r est un vecteur de 0 de taille r. L'équation (12) est à la base d'une méthode proposée récemment par Chernozhukov & Hansen (2008). Elle signifie que dans une régression quantile de $Y - X_1'\beta_{1\tau}$ sur X_2 et Z_1, les coefficients de Z_1 sont égaux à 0. Par ailleurs, si l'on se trompe en considérant un coefficient $\beta \neq \beta_{1\tau}$, la variable modifiée $Y - X_1'\beta$ sera toujours liée à X_1. Donc, si X_1 et Z_1 sont dépendants, ce qui correspond à la condition de rang dans les régressions linéaires, la régression quantile de $Y - X_1'\beta$ sur Z conduira a priori à un coefficient non nul sur Z_1 [12]. L'idée de Chernozhukov et Hansen est alors

12. La condition de dépendance exacte nécessaire entre X_1 et Z est plus difficile à expliciter que dans les modèles linéaires, voir Chernozhukov & Hansen (2008) pour plus de détails à ce sujet.

d' « inverser » la régression quantile, en estimant $\beta_{1\tau}$ par le paramètre $\widehat{\beta}_{1\tau}$ permettant d'obtenir, dans la régression quantile de $Y - X_1'\widehat{\beta}_{1\tau}$ sur Z, un coefficient égal à 0 pour Z_1. En pratique, les auteurs proposent l'algorithme suivant :

1. Définir une grille sur $\beta_{1\tau}$, $\{b_1, ..., b_J\}$.
2. Pour $j = 1$ à J :
   - Calculer les estimateurs de régression quantile de $Y - X_1'b_j$ sur (Z_1, X_2). Soit $(\widehat{\gamma}(b_j), \widehat{\beta}_{2\tau}(b_j))$ les estimateurs correspondants.
   - Calculer la statistique de Wald correspondant au test de $\gamma(b_j) = 0$:

$$W_n(b_j) = n\widehat{\gamma}(b_j)'\widehat{V}_{\text{as}}^{-1}(\widehat{\gamma}(b_j))\widehat{\gamma}(b_j).$$

3. Définir l'estimateur de $\beta_\tau = (\beta_{1\tau}, \beta_{2\tau})$ par

$$\widehat{\beta}_{1\tau} = \arg\min_{j=1...J} W_n(b_j),\ \widehat{\beta}_{2\tau} = \widehat{\beta}_{2\tau}(\widehat{\beta}_{1\tau}).$$

La commande `Stata ivqreg`, introduite récemment (voir **?**, **?**, utilise cet algorithme pour estimer β_τ dans un tel modèle. Même en l'absence de procédure préprogrammée, cet algorithme a l'intérêt de ne s'appuyer que sur des régressions quantiles classiques. Il peut donc être mis en œuvre simplement avec des logiciels standards. En pratique, la grille doit être suffisamment fine pour ne pas altérer les propriétés asymptotiques de l'estimateur (cf. Chernozhukov & Hansen, 2008 pour plus de détails). Pour que le temps de calcul reste raisonnable, le nombre de variables endogènes doit donc être petit ($q = 1$ ou 2).

Notons enfin que d'autres solutions existent pour estimer des régressions quantiles instrumentales. Abadie et al. (2002), en particulier, proposent de recourir à une approche par régression quantile pondérée pour les cas où la variable endogène X_1 et l'instrument Z_1 sont binaires. Quelle que soit la méthode retenue, la difficulté principale est évidemment de trouver un instrument valide, c'est-à-dire vérifiant (11). Nous proposons dans la partie 4.2 un exemple d'instrument utilisant une expérimentation sociale.

3.2 Les régressions quantiles avec des données de panel

L'utilisation de données de panel, c'est-à-dire de données répétées pour les mêmes unités, peut être également une manière de traiter la présence d'hétérogénéité individuelle inobservée. Ces données sont en effet plus souvent disponibles que des variables instrumentales valides. Sous l'hypothèse, certes restrictive, que l'hétérogénéité individuelle est fixe dans le temps et indépendante des termes résiduels, une simple différenciation permet dans le cadre des régressions linéaires de se débarrasser de ces effets individuels fixes dans le temps. Cependant, ces méthodes ne s'appliquent plus directement dans le cas des quantiles. Ces derniers n'ont pas de propriété de linéarité comme la moyenne. Les quantiles des variables différenciées ne correspondent pas directement aux quantiles d'intérêt.

Sur la période très récente, de nombreux estimateurs permettant de tenir compte de la présence d'effets individuels dans des régressions quantiles ont été proposés. L'estimateur de Canay (2011) a l'avantage de s'appuyer sur des techniques standards utilisant des procédures couramment disponibles dans les logiciels statistiques. Pour en comprendre le principe, il est utile d'utiliser les notations du modèle à coefficients aléatoires présenté en introduction, dans lequel on prend en compte également un effet individuel α_i fixe dans le temps :

$$Y_{it} = X_{it}\beta_{U_{it}} + \alpha_i, \tag{13}$$

où α_i et U_{it} sont inobservables et U_{it} est indépendant de (α_i, X_{it}) et suit une distribution uniforme sur $[0, 1]$. X_{it} représentent les variables explicatives tandis que α_i représente des caractéristiques fixes dans le temps. Notons qu'aucune hypothèse n'est imposée ici quant à la dépendance entre X_{it} et α_i. Ce modèle permet donc de résoudre en partie le problème d'endogénéité puisqu'il autorise des corrélations entre X_{it} et les facteurs inobservables individuels, pourvu que ces derniers soient stables dans le temps.

En introduisant $e_{it} = X_{it}(\beta_{U_{it}} - \beta_\tau)$, on peut réécrire le modèle comme :

$$Y_{it} = X'_{it}\beta_\tau + \alpha_i + e_{it}, \text{ avec } \; q_\tau(e_{it}|X_i) = 0. \tag{14}$$

Le terme inobservé U_{it} est indépendant des caractéristiques individuelles observées et inobservées (X_{it}, α_i). On suppose également que les effets individuels α_i ont un effet de translation simple sur la distribution. Sous des hypothèses techniques détaillées par Canay (2011), on peut montrer que les termes β_τ sont identifiés et proposer un estimateur convergent lorsque T tend vers l'infini. Plus précisément, Canay (2011) propose un estimateur basé sur les deux étapes suivantes :

1. Estimation par un estimateur within classique de la régression linéaire

   $$Y_{it} = X'_{it}\beta_\mu + \alpha_i + u_{it}, \text{ avec } E(u_{it}|X_{it}, \alpha_i) = 0 \text{ et } \beta_\mu = E[\beta_{U_{it}}].$$

   A partir de l'estimation de β_μ, on peut obtenir des estimations des effets individuels : $\hat{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T}(Y_{it} - X_{it}\hat{\beta}_\mu)$.

2. Régression quantile classique de la variable transformée $\hat{Y}_{it} = Y_{it} - \hat{\alpha}_i$ sur les régresseurs X_{it}.

Canay montre également que ces estimateurs sont asymptotiquement normaux. Cependant, la matrice de variance-covariance ne correspond pas à celle produite par défaut dans la deuxième étape. Canay (2011) fournit un estimateur de la matrice de variance-covariance, mais propose également d'utiliser une procédure de bootstrap compte tenu de sa complexité. Dans cette procédure, il s'agit de répliquer, pour chacun des échantillons bootstrap, les deux étapes précédentes de l'estimation. Il n'existe pas encore dans les logiciels statistiques standard de procédure ou de package utilisant cette méthode. On trouvera cependant sur le site d'Ivan Canay, un programme `R` permettant d'utiliser cette procédure en deux étapes [13].

Canay n'établit la convergence de cet estimateur que lorsque le nombre de périodes T tend vers l'infini. Comme il le montre sur des données simulées, l'estimateur n'est pas toujours performant sur des panels contenant un petit nombre de périodes. D'autres méthodes ont également été proposées, toujours sous l'hypothèse (13) que les effets individuels ont un simple effet de translation. Koenker (2004) propose d'estimer simultanément les effets de X_{it} correspondant à q quantiles différents, $(\beta_{\tau_k})_{k=1\ldots q}$, et les effets individuels $(\alpha_i)_{i=1\ldots n}$ (voir aussi Lamarche, 2010) :

$$\begin{aligned}((\widehat{\beta}_{\tau_k})_{k=1\ldots q}, (\widehat{\alpha}_i)_{i=1\ldots n}) \quad &= \quad \arg \min_{\substack{(\alpha_i)_{i=1\ldots n} \\ (\beta_{\tau_k})_{k=1\ldots q}}} \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{k=1}^{q} \rho_{\tau_k}(Y_{it} - X_{it}'\beta_{\tau_k} - \alpha_i) \\ &\qquad + \lambda \sum_{i=1}^{n} |\alpha_i|.\end{aligned}$$

L'ajout du terme de pénalisation $\lambda \sum_{i=1}^{n} |\alpha_i|$ permet d'éviter une trop grande dispersion des nombreux termes liés aux différents effets individuels. Ce coefficient de régularisation λ doit être fixé pour l'estimation : lorsqu'il est nul, aucune contrainte n'est donnée pour l'estimation des effets individuels et le risque existe d'obtenir des effets très dispersés. Lorsqu'il est très grand, la solution au programme de minimisation fournira des effets individuels estimés très faibles. Le package R `rqpd` permet de faire cette estimation. Cette procédure est cependant très coûteuse en temps de calcul lorsque le nombre d'individus est élevé, et soulève la question du choix du coefficient de régularisation λ. Comme celui de Canay, cet estimateur n'est convergent qu'asymptotiquement avec T. Notons que l'hypothèse commune à ces deux estimateurs est que les effets individuels n'ont qu'un effet de translation sur la distribution d'intérêt, ce qui constitue une restriction importante. Des articles récents proposent l'utilisation d'estimateurs non-paramétriques, mais ces méthodes sont difficiles à mettre en œuvre et peu adaptées lorsque le nombre de variables explicatives est important. L'utilisation des données de panel pour des régressions quantiles est aujourd'hui encore un champ de recherche actif, et il est probable que d'autres estimateurs seront proposés dans les prochaines années.

3.3 Les « Quantile Treatment Effects »

En pratique, on est souvent intéressé par l'effet non pas de l'ensemble des variables explicatives, mais plus spécifiquement de l'une d'entre elles. On peut par exemple s'intéresser

13. L'adresse actuelle du site est `http://faculty.wcas.northwestern.edu/~iac879/research.htm`. Le programme, nommé QRPanel.R, contient la fonction `pqr.estimator(dataframe,eqformula,tau)` qui estime β_τ pour $\tau =$ `tau`. `eqformula` permet de spécifier Y et les X, et `dataframe` correspond au fichier de données. Une estimation de la précision est ensuite fournie par la fonction `pqr.se`

à l'effet d'avoir suivi une formation professionnelle sur les revenus, d'une politique éducative sur la réussite scolaire... L'objectif est alors d'évaluer l'effet causal du fait d'avoir bénéficié du programme évalué, qu'on peut noter par une indicatrice binaire T. Pour disposer d'un cadre pour traiter de ces questions, on peut utiliser le formalisme des méthodes issues de la littérature sur l'évaluation empirique de politique publique. Dans ce cadre, on suppose que chaque personne a deux revenus « potentiels » , Y_0 (celui qu'il peut espérer en l'absence du programme) et Y_1 (celui qu'il peut espérer avec le programme). A ces revenus potentiels sont associées les deux distributions F_{Y_0} et F_{Y_1}. On peut alors définir le $\tau^{\text{ième}}$« quantile treatment effect » (QTE) comme la « distance » horizontale entre les deux distributions (Doksum (1974)) :

$$\delta_\tau = q_\tau(Y_1) - q_\tau(Y_0).$$

De même, on peut définir sa restriction aux personnes qui ont effectivement bénéficié du programme (*quantile treatment effect on the treated*, QTET) :

$$\delta_{\tau|T=1} = q_\tau(Y_1|T = 1) - q_\tau(Y_0|T = 1).$$

On peut ici s'attarder sur l'interprétation de ces paramètres. L'effet du traitement sur le quantile pour les traités (traduction très littérale du quantile treatment effect on the treated) correspond à la différence entre le τ^{eme} quantile de la distribution de la variable d'intérêt parmi les personnes qui ont bénéficié du programme T, et le quantile équivalent de la distribution de cette variable parmi ces mêmes personnes, si elles n'avaient pas bénéficié de ce programme. L'effet du traitement sur le quantile (quantile treatment effect) est plus général, puisqu'il correspond à la différence entre les quantiles des distributions qu'on s'attend à observer dans la population respectivement si le programme est généralisé à tous ou au contraire en son absence. Comme on l'a déjà souligné, ces paramètres ne correspondent pas en général à l'effet de ce programme pour les personnes qui se trouvent au niveau du $\tau^{\text{ième}}$quantile de la distribution de Y en l'absence de ce programme. Ce paramètre ne permet donc en principe pas de dire si ce sont les personnes initialement les plus (dé)favorisées qui bénéficieraient du programme qu'on évalue. Pour passer à cette interprétation, il est ici encore nécessaire de faire une hypothèse d'invariance des rangs (les personnes seraient « ordonnées » de la même manière dans la distribution de la variable d'intérêt Y en l'absence ou en présence du programme). Enfin, du fait de la non linéarité des quantiles, ces paramètres ne correspondent pas à la distribution de l'effet du programme ($q_\tau(Y_1) - q_\tau(Y_0) \neq q_\tau(Y_1 - Y_0)$). On trouvera dans Clements et al. (1997) une discussion des conditions sous lesquelles il est possible de borner cette distribution des effets.

Au-delà de ces questions d'interprétation spécifiques aux régressions quantiles, la difficulté classique pour estimer ces effets du programme sur les quantiles tient à ce que pour une personne donnée, on n'observe en fait que $Y = TY_1 + (1 - T)Y_0$, c'est-à-dire le revenu potentiel avec traitement (Y_1) si elle a bénéficié du programme et le revenu potentiel sans traitement sinon. On pourrait être tenté d'estimer simplement δ_τ (ou $\delta_{\tau|T=1}$) par la différence de quantiles conditionnels des revenu observés, $q_\tau(Y|T=1) - q_\tau(Y|T=0)$. En général cependant, cette différence ne correspond pas au paramètre d'intérêt. En effet, dès qu'il existe une (auto)sélection dans l'entrée dans le programme (par exemple, lorsque les personnes qui ont choisi d'en bénéficier sont celles pour lesquelles il sera le plus efficace), la distribution des revenus observés parmi les bénéficiaires $F_{Y|T=1}$ n'est pas représentative de la distribution du revenu potentiel avec le programme de l'ensemble de la population F_{Y_1}.

Plusieurs méthodes ont été proposées pour identifier les effets moyens d'un programme en présence d'effets de sélection (on en trouvera une description par exemple dans Givord, 2010). Des extensions de ces méthodes à l'analyse des quantiles ont été proposées récemment. Nous ne présentons ici qu'une possibilité, qui est facilement implémentable. Elle repose sur l'hypothèse d'indépendance conditionnelle (Conditional Independance Assumption, ou CIA) suivante :

$$(Y_0, Y_1) \perp\!\!\!\perp T|X. \tag{15}$$

Cette hypothèse correspond à l'exogénéité conditionnelle de T (on parle aussi de sélection sur observables, c'est-à-dire que toute la sélection dans le programme peut être expliquée par les variables observées X). Elle est à la base des méthodes d'appariement (*matching*) ou simplement des régressions linéaires lorsque l'on s'intéresse à la moyenne. On pourrait donc envisager d'estimer l'impact du programme T par une régression quantile en « contrôlant » de l'effet des observables X. Cependant, une telle régression quantile ne permet pas d'estimer directement le paramètre d'intérêt. Lorsque l'on inclut des variables de contrôles supplémentaires X, la régression quantile estime en effet le paramètre $\widetilde{\delta}_\tau = q_\tau(Y_1|X=x) - q_\tau(Y_0|X=x)$. Du fait de la non-linéarité des quantiles, ce paramètre ne correspond pas à δ_τ ni même à $\delta_{\tau|T=1}$ en général. Les quantiles de la distribution des revenus potentiels Y_0 et Y_1 ne sont pas les mêmes que ceux des distributions de ces revenus potentiels conditionnelles aux observables.

Firpo (2007) propose une méthode pour résoudre ces deux problèmes. Celle-ci s'apparente aux méthodes d'appariement utilisées pour estimer l'effet moyen du traitement $E(Y_1 - Y_0)$ sous l'hypothèse 15 d'indépendance conditionnelle. On fait tout d'abord une hypothèse de support commun, nécessaire également dans les méthodes d'appariement :

$$p(X) = P(T=1|X) \in]0,1[\tag{16}$$

Cette hypothèse signifie que pour chaque bénéficiaire du programme, il existe une personne qui n'en a pas bénéficié et présente les mêmes caractéristiques observables (et inversement). Firpo montre que sous les hypothèses ci-dessus il est possible d'identifier les deux quantiles $q_\tau(Y_1)$ et $q_\tau(Y_0)$, à partir des seules données observées(Y, T, X). Il utilise pour cela les relations[14] :

$$\tau = E\left[\frac{T\mathbb{1}\{Y \leq q_\tau(Y_1)\}}{p(X)}\right] = E\left[\frac{(1-T)\mathbb{1}\{Y \leq q_\tau(Y_0)\}}{1-p(X)}\right] \tag{17}$$

Comme on observe (T, Y, X) et que l'on peut identifier le score de propension $p(X)$, ces relations permettent d'estimer $q_\tau(Y_1)$ et $q_\tau(Y_0)$. En pratique, Firpo montre que l'on peut estimer $\delta_\tau = q_\tau(Y_1) - q_\tau(Y_0)$ par une procédure en deux étapes :

1. estimer le score $p(X)$. Notons $\widehat{p}(X)$ un tel estimateur ;
2. estimer $q_\tau(Y_0)$ puis $q_\tau(Y_1)$ en utilisant une régression quantile sur la seule constante : $\widehat{q}_\tau(Y_t) = \arg\min_b \sum \widehat{\omega}_{t,i}\rho_\tau(Y_i - b)$ $(t = 0, 1)$, avec des pondérations $\widehat{\omega}_{1,i} = \frac{T_i}{\widehat{p}(X_i)}$ et $\widehat{\omega}_{0,i} = \frac{1-T_i}{1-\widehat{p}(X_i)}$.

Il s'agit donc d'une régression quantile simple, mais pondérée afin de tenir compte des effets de sélection. Intuitivement, on pondère ainsi parmi les personnes non traitées celles qui ont néanmoins une probabilité plus grande de l'être. Firpo propose également un jeu de pondérations pour estimer l'effet du traitement sur les seuls traités $\delta_{\tau|T=1}$: dans ce cas on utilise comme pondération respectivement $\hat{\omega}_{0,i|T=1} = \frac{\frac{\hat{p}(X_i)}{1-\hat{p}(X_i)}(1-T_i)}{\sum_1^l T_l}$ pour estimer $q_\tau(Y_0|T=1)$ et $\hat{\omega}_{1,i|T=1} = \frac{T_i}{\sum_1^l T_l}$ pour estimer $q_\tau(Y_1|T=1)$.

Cette méthode peut donc être implémentée simplement en utilisant des régressions quantiles standards, en pondérant les observations par le poids estimé correspondant au score de propension. Elle n'est en revanche valide que lorsque l'hypothèse d'indépendance conditionnelle aux observables est vérifiée, ce qui doit être discuté avec attention en fonction des modes de sélection dans le traitement. Lorsque cette hypothèse ne paraît pas plausible, il faut se tourner vers d'autres méthodes d'identification : selon le cas et les données disponibles, l'utilisation de variables instrumentales, de données de panel peuvent être des options. Frandsen et al. (2012) proposent une extension aux régressions quantiles des méthodes de régression sur discontinuités (cas où l'affectation au traitement dépend d'une variable annexe de manière discontinue).

Au-delà du cadre strict de l'évaluation de politique publique, la méthode de Firpo (2007) est utile lorsqu'on souhaite isoler l'effet d'une variable explicative binaire T sur la distribution d'une variable d'intérêt, en ayant contrôlé des autres caractéristiques observables X. Si l'on s'intéresse plutot à une variable continue, on peut citer une méthode développée par ces auteurs (Firpo et al., 2009), qui permet d'étudier l'effet d'une augmentation infinitésimale d'une seule variable explicative continue (disons X_2) sur la distribution de Y,

14. Ce résultat se montre comme suit (ici pour Y_1) :

$$\begin{aligned} E\left[\frac{T\mathbb{1}\{Y \leq q_\tau(Y_1)\}}{p(X)}\right] &= E\left[\frac{\mathbb{1}\{Y_1 \leq q_\tau(Y_1)\}}{p(X)}E(T|Y_1, X)\right] = E\left[\frac{\mathbb{1}\{Y_1 \leq q_\tau(Y_1)\}}{p(X)}E(T|X)\right] \\ &= E\left[\mathbb{1}\{Y_1 \leq q_\tau(Y_1)\}\right] = \tau. \end{aligned}$$

les autres covariables $(X_3, ..., X_p)$ restant inchangées. On parle dans ce cas de régression quantile inconditionnelle dans la mesure où l'objet d'intérêt est $q_\tau(Y)$ (ou plus précisément sa variation suite à une variation infinitésimale de X_2), plutôt que $q_\tau(Y|X)$.

3.4 Les régressions quantiles dans les modèles non linéaires

Nous considérons ici des extensions de la régression linéaire quantile aux modèles non-linéaires de la forme

$$Y = g(X'\beta_0 + \varepsilon), \tag{18}$$

où g est une fonction non-linéaire connue. Deux exemples importants sont le modèle binaire, pour lequel $g(x) = \mathbb{1}\{x > 0\}$, et le modèle à censure fixe, pour lequel $g(x) = \max(s, x)$ (ou $g(x) = \min(s, x)$) avec s une constante connue. Ce dernier modèle est souvent utilisé pour modéliser la consommation d'un bien, qui prend la valeur nulle quand il n'est pas consommé (pour plus de détails, cf. par exemple Wooldridge, 2002, chap. 16). Ceci peut être rationalisé par l'existence d'une valuation implicite du bien c^* par les consommateurs éventuels, qui ne consomment ce bien que lorsque cette valuation est strictement positive. On observe donc la consommation $c = max(0, c^*)$. Dans ces modèles, il est difficile d'utiliser des restrictions de la forme $E(\varepsilon|X) = 0$ car en général, $E(Y|X) \neq g(X'\beta_0)$. L'approche standard consiste alors à imposer des hypothèses paramétriques sur la distribution des résidus. Par exemple, il est fréquent de supposer l'indépendance entre X et ε et la normalité de ces derniers (on parle alors de modèle probit lorsque $g(x) = \mathbb{1}\{x > 0\}$ et de modèle tobit lorsque $g(x) = \max(0, x)$). Ces hypothèses sont cependant restrictives et souvent difficiles à justifier.

Une approche alternative à ces hypothèses paramétriques est de recourir à des restrictions sur les quantiles. En effet, on peut facilement étendre les restrictions sur les quantiles des termes de perturbations ϵ à une transformation non linéaire, grâce à la propriété d'invariance déjà évoquée en partie 1.2 :

$$g(q_\tau(U)) = q_\tau(g(U)),$$

valable pour toute variable aléatoire U et toute fonction g croissante et continue à gauche [15]. Ainsi, si l'on impose dans le modèle non linéaire (18) la restriction $q_\tau(\varepsilon|X) = 0$ et que g est croissante continue à gauche, on obtient

$$q_\tau(Y|X) = g(q_\tau(X'\beta_0 + \varepsilon|X)) = g(X'\beta_0).$$

Par le même argument que celui développé dans la partie 3, il s'ensuit que

$$\beta_0 \in \arg\min_\beta E\left[\rho_\tau\left(Y - g(X'\beta)\right)\right].$$

15. Rappelons qu'une fonction g est continue à gauche si pour tout x, $\lim_{u\uparrow x} g(u) = g(x)$. Les fonctions $g(x) = \mathbb{1}\{x > 0\}$ et $g(x) = \max(0, x)$ sont donc continues à gauche.

Comme précédemment, on estime alors β_0 par

$$\widehat{\beta} = \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau \left(Y_i - g(X_i'\beta) \right). \tag{19}$$

L'estimateur défini par (19) est très proche de celui de la régression quantile linéaire, la différence étant simplement l'ajout dans le programme de la fonction g.

Pour le modèle de censure fixe pour lequel $g(x) = \max(s, x)$, $\widehat{\beta}$ est $\sqrt{n}$−convergent, et peut être estimé par une application itérative de régressions quantiles linéaires (cf. par exemple Buchinsky, 1994). Pour la médiane ($\tau = 1/2$), l'estimateur proposé par Powell (1984) (« censored LAD estimator », i.e. l'estimateur des moindres valeurs absolues censuré) est implémenté sous Stata via la commande `clad`. Hong & Chernozhukov (2002) proposent par ailleurs un estimateur simple en trois étapes qui peut être utilisé pour l'ensemble des quantiles (voir Fack & Landais, 2009 pour une utilisation de leur méthode sur données françaises).

4 Exemples d'application

4.1 Comment lire les résultats d'une régression quantile ?

A titre d'illustration, on estime une équation de salaire classique à partir de l'enquête Emploi 2008. Cet exercice n'a d'autre prétention que d'illustrer les résultats issus d'une régression quantile sur un cas pratique (et ne traite en particulier pas des problèmes de sélection dans l'emploi, voir par exemple Buchinsky, 1998 pour une discussion de cette question). Pour une étude plus complète de la question des rendements salariaux de l'expérience et de l'éducation, et de leur évolution en France, on se reportera par exemple à Charnoz et al. (2011*a*).

La variable d'intérêt est ici le salaire, ou plus précisément son logarithme. Les variables explicatives sont les caractéristiques observables du salarié, à savoir le nombre d'années d'études, le sexe, sa nationalité, le nombre d'années d'expérience potentielle ainsi que le carré de celui-ci. Les estimations ont été faites pour chaque décile de la distribution conditionnelle du logarithme du salaire. On suppose donc, pour chaque décile :

$$\text{décile}_j(\ln(\text{salaire}|X)) = X'\beta_j$$

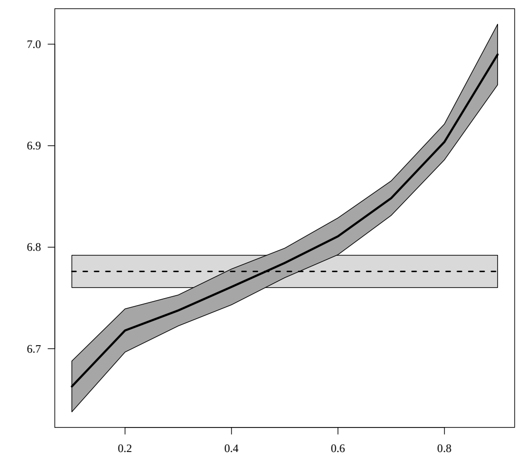
Les régressions quantiles permettent de déterminer comment varie chaque décile en fonction des déterminants auxquels on s'intéresse. Par exemple, le paramètre β_{kj} dans la régression $\text{décile}_j(\ln(\text{salaire})|X) = X'\beta_j$ vérifie

$$\beta_{kj} = \text{décile}_j(\ln(\text{salaire})|X_{-k}, X_k = x_k) - \text{décile}_j(\ln(\text{salaire})|X_{-k}, X_k = x_k + 1).$$

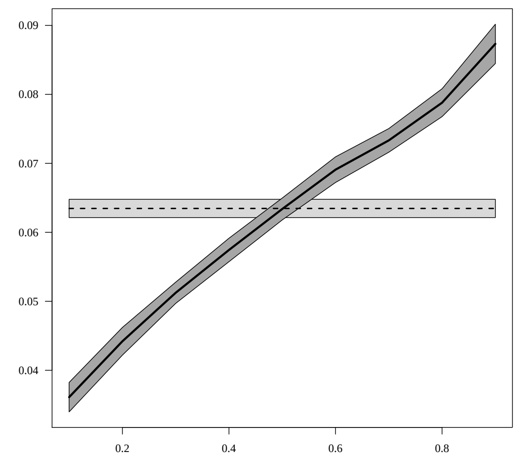
Il s'agit du changement du $j^{\text{ième}}$ décile de la distribution conditionnelle de salaire suite à une augmentation d'une unité de X_k, par exemple une augmentation d'une année d'études, toutes choses égales par ailleurs (i.e., les autres variables X_{-k} restent constantes). Dans le cas d'une variable explicative binaire, comme le fait d'être un homme pour un salarié, β_{kj} mesure simplement l'écart entre le $j^{\text{ième}}$ décile de la distribution des salaires des hommes (conditionnelle à l'ensemble des autres variables explicatives X_{-k}) et le $j^{\text{ième}}$ décile de la distribution des salaires des femmes (également conditionnelle à X_{-k}).

En termes de présentation, on notera qu'on a un jeu de coefficients estimés pour chaque quantile auquel on s'intéresse. Les résultats sont donc plus lourds à présenter. Dans la littérature, on les trouve présentés sous forme d'un tableau regroupant l'ensemble des coefficients, ou de manière peut être plus parlante sous forme de graphiques. C'est la solution que nous avons retenue ici (cf. figure 3).

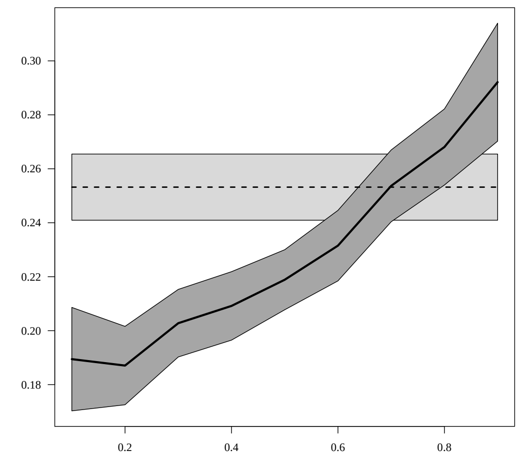
Nous avons choisi de représenter les estimations des coefficients pour les différents déciles, avec l'intervalle de confiance à 95% (zone grisée), ainsi, à titre de comparaison, que la valeur du coefficient des moindres carrés ordinaires (en pointillé).



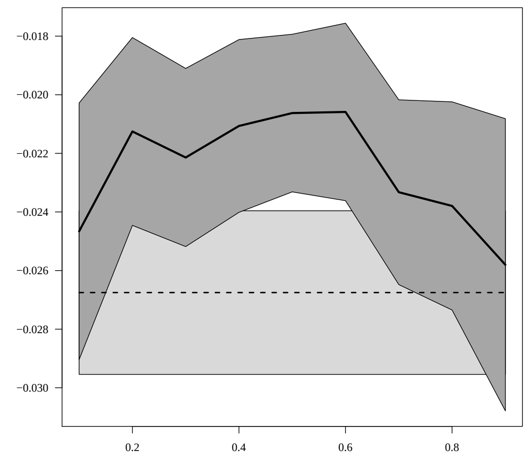
Constante



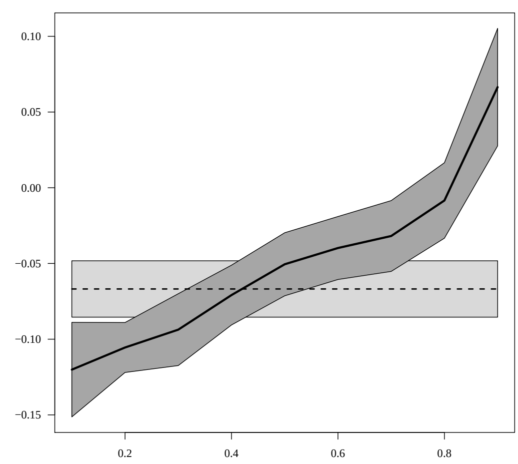
Nombre d'années d'étude



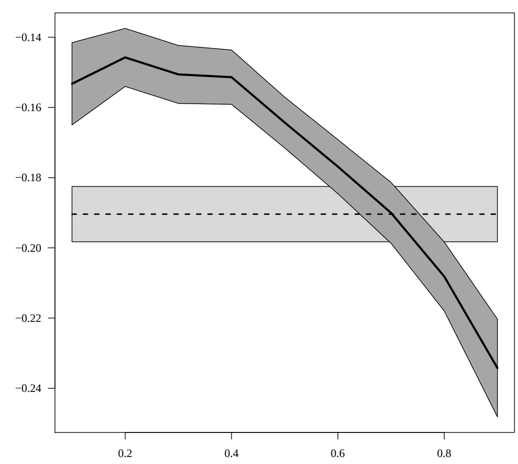
Expérience potentielle



Expérience potentielle au carré



Etranger



Femme

FIGURE 3 – Estimation des coefficients par régressions quantiles (en pointillé : estimation par les moindres carrés ordinaires).

Le coefficient correspondant à la constante peut être considéré comme le décile des salariés ayant les modalités de référence (ici, le fait d'être un homme salarié avec la nationalité française, sans expérience potentielle et dont le niveau d'étude est minimal). Il est sans surprise croissant avec le décile (premier graphique en haut à gauche). On passe ainsi de 6,5 pour le premier décile à 7,0 pour le neuvième décile. Le coefficient estimé par les moindres carrés ordinaires, qui correspond pour sa part donc au salaire moyen des salariés aux modalités de référence, sans expérience potentielle et de niveau d'étude minimal, est plus proche des premiers déciles (autour de 6,7), ce qui exprime bien que la distribution du logarithme des salaires est asymétrique.

Le coefficient correspondant au nombre d'années d'étude est toujours positif, reflétant le fait que le niveau d'étude décale globalement la distribution des salaires vers le haut. Son effet est très nettement croissant avec le décile. Ainsi, l'augmentation du premier décile de la distribution des salaires conditionnelle suite à une augmentation d'une année d'études est, une fois que l'on contrôle des autres variables observées, de 4%, contre 8% pour le dernier décile. Une autre manière de présenter ce résultat est d'observer que la dispersion des salaires augmente avec le nombre d'années d'études, ou encore que les distributions de salaires des plus diplômés sont plus inégales que celles des moins diplômés. C'est également le cas pour l'expérience potentielle [16].

Les salaires des femmes sont systématiquement inférieurs à ceux des hommes, mais ces différences sont d'autant plus fortes que l'on s'élève dans la distribution : conditionnellement aux autres caractéristiques observables, le neuvième décile de la distribution des salaires des femmes est ainsi inférieur de 24% au neuvième décile de la distribution des salaires des hommes, tandis que cette différence n'est « que » de 15% pour le premier décile. Ces différences peuvent s'expliquer par exemple par la présence de « plafonds de verre ». A l'inverse, le fait de ne pas disposer de la nationalité française a un impact négatif pour le bas de la distribution, mais les deux distributions conditionnelles se rapprochent ensuite. Le coefficient augmente avec le décile, il n'est plus significatif au niveau des septième et huitième déciles et même positif au niveau du neuvième décile.

Ces résultats sont à interpréter avec la même prudence que ceux que donnerait la modélisation du salaire conditionnel moyen. La régression quantile est un outil qui permet d'estimer les effets de variables explicatives sur l'ensemble de la distribution d'une variable d'intérêt, mais elle ne règle aucun des éventuels problèmes d'endogénéité de certaines de ces variables. Par exemple, le fait d'avoir fait des études longues peut être lié à des compétences spécifiques, des réseaux familiaux qui ont également un effet positif sur le salaire. Ces éléments ne sont pas observés dans l'enquête. Dans ce cas, le coefficient des années

16. Du fait du terme quadratique, on peut interpréter l'augmentation marginale de l'expérience potentielle (exppot) sur le décile j de la distribution de salaire comme $\beta_{1j} + 2\beta_{2j}$ exppot, où β_{1j} est le coefficient relatif à exppot et β_{2j} le coefficient du carré d'exppot.

d'études reflète également l'effet positif de ces caractéristiques inobservées, et pas uniquement l'effet causal d'une augmentation d'une année d'étude. De même, l'interprétation du coefficient correspondant au fait d'être étranger est délicate. Les coefficients négatifs puis positifs ne peuvent s'interpréter comme la présence d'une discrimination négative puis positive envers les étrangers, sauf à considérer que la distribution des caractéristiques ayant un effet sur le salaire de l'ensemble des salariés étrangers travaillant en France est identique à cette même distribution pour les salariés français. Cependant, les étrangers décidant de travailler en France ont sans doute des profils professionnels particuliers, qui peuvent expliquer ces coefficients. Cette difficulté explique le recours au *testing* pour mettre en évidence la discrimination. Au final, exactement les mêmes précautions d'interprétation que dans le cas d'une régression linéaire s'imposent. En cas d'endogénéité de certaines variables explicatives, il est nécessaire, pour obtenir une interprétation causale des coefficients, de mobiliser par exemple la méthode des variables instrumentales décrite dans la partie 3 (mais il peut être difficile d'obtenir de tels instruments).

D'autre part, les estimations obtenues correspondent à l'effet des variables explicatives sur les distributions de la variable d'intérêt conditionnelles à ces variables. Elles renseignent sur les écarts entre les quantiles d'ordre τ de la distribution de salaire des salariés conditionnelle à ces variables. Comme expliqué en partie 1.2.2, elle ne permet pas d'évaluer directement comment se modifierait le quantile d'ordre τ de la distribution de salaires de *l'ensemble de la population* si la distribution de ces variables explicatives, par exemple la proportion de diplômés du supérieur, était différente. Ceci vient de la propriété de non-linéarité des quantiles déjà évoquée : les quantiles de la population entière ne s'obtiennent pas simplement en intégrant les quantiles conditionnels qui sont modélisés dans les régressions quantiles.

Enfin, comme expliqué dans la partie 1.1.4, les résultats quantiles, tout comme ceux obtenus par une régression linéaire, n'ont pas directement d'interprétation individuelle. Le principe des régressions quantiles est *stricto sensu* de comparer des distributions conditionnelles entre elles. Elles permettent par exemple de dire que le premier décile des salaires des salariés étrangers est inférieur de 12% à celui des salariés ayant la nationalité française, toutes choses égales par ailleurs. En revanche, elles ne permettent pas a priori de dire que le salaire d'un salarié étranger qui se trouve au niveau du premier décile de la distribution de salaire de cette population augmenterait d'autant s'il acquérait la nationalité française. Pour pouvoir ainsi interpréter les résultats obtenus, il faut supposer que ce salarié occupe le même rang dans les deux distributions (correspondant respectivement aux salaires des salariés de nationalité française et aux salaires des salariés ne disposant pas de la nationalité française). Comme expliqué plus haut, cette hypothèse d'invariance des rangs, qui a le mérite de fournir une interprétation simple des coefficients des régressions quantiles, est restrictive.

Cette mise en garde dans l'interprétation des résultats est de même nature que celle qu'on peut avoir pour les résultats de la modélisation de la moyenne du logarithme des salaires par une régression linéaire. Les estimations ainsi obtenues montrent que la moyenne du logarithme des salaires des étrangers est inférieur de 7% à celui des salariés français. Cela ne signifie pas qu'un salarié étranger dont le salaire se trouve au niveau au salaire moyen verrait son salaire augmenter d'autant s'il acquérait la nationalité française.

4.2 Un exemple de régression quantile instrumentale

Ce deuxième exemple, qui reprend l'application d'Abadie et al. (2002), permet d'illustrer l'utilisation des méthodes de régressions quantiles instrumentales développées dans la partie 3.1. Il utilise les données issues de l'expérimentation de l'efficacité d'un programme de formation de chômeur, le « Job Training Partnership Act (JTPA) » , mis en place à partir de 1983 aux Etats-Unis. Il s'agit d'un ensemble de programmes de formation et d'assistance destinés aux jeunes défavorisés. L'évaluation de l'efficacité de ce genre de programme est souvent rendue difficile par les effets d'auto-sélection : en général, ce sont les personnes qui peuvent en retirer le plus grand bénéfice qui choisissent de rentrer dans le dispositif.

Une expérimentation a cependant été mise en place entre 1987 et 1989 dans 16 structures locales auprès d'un échantillon initial de 20 000 jeunes environ. Les programmes de formation correspondant au JTPA n'ont été proposés qu'à deux tiers de ces jeunes, tirés aléatoirement. Des données ont ensuite été collectées sur la séquence de revenus de l'ensemble des jeunes de l'échantillon initial.

Pour évaluer l'efficacité du JTPA sur les revenus futurs, on ne peut cependant pas comparer directement les jeunes participant au programme de formation et ceux ne participant pas. En effet, malgré l'affectation aléatoire initialement créée par le dispositif expérimental, les personnes qui, par le tirage au sort, étaient affectées au programme pouvaient choisir de ne pas en profiter. Ainsi, seules 60% des personnes tirées au sort ont profité des programmes de formation. Inversement, l'entrée dans le programme n'était pas complètement fermée aux jeunes qui n'avaient pas été sélectionnés par le tirage au sort puisque 2% des jeunes non tirés au sort les ont néanmoins suivis. Ainsi, malgré le tirage au sort, il y a une part d'auto-sélection dans la participation effective au programme. On dispose ici cependant d'un instrument, l'affectation aléatoire au dispositif. Issue d'un tirage au sort, elle n'est évidemment pas corrélée aux déterminants inobservés du revenu. En revanche, elle explique fortement le fait d'avoir bénéficié ou non du programme.

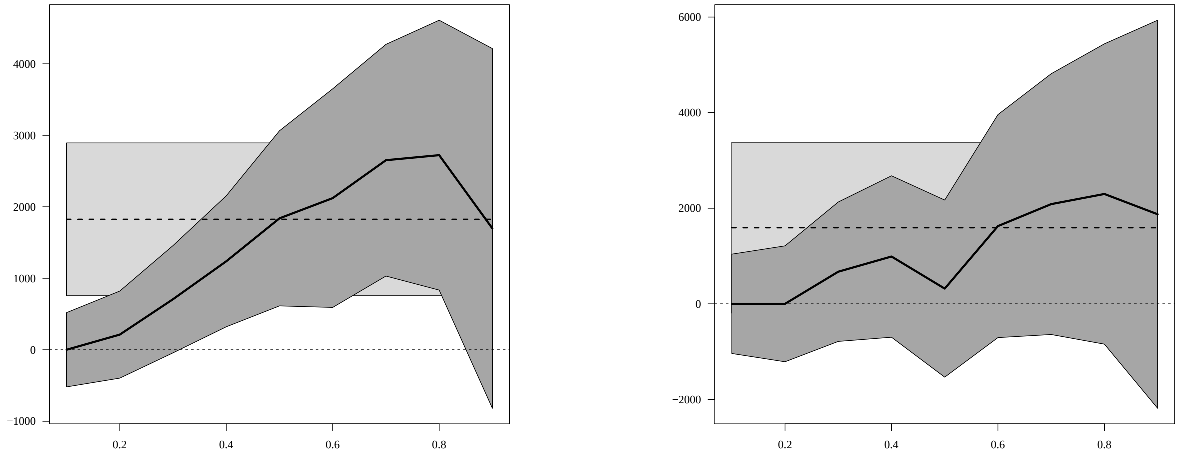
Nous appliquons donc la méthode de Chernozhukov & Hansen (2008) décrite dans la partie 3.1 pour évaluer l'impact de ce programme sur l'ensemble des jeunes [17].

Les estimations sont conduites séparément pour les hommes et les femmes. On a ici estimé les trois quartiles, le premier et le dernier décile. Les résultats sont présentés sur le graphique 4. Les régressions instrumentales fournissent des résultats près de deux fois plus faibles que les simples régressions linéaires (droites pointillées en gras), ce qui traduit bien l'auto-sélection dans le dispositif. Les régressions quantiles indiquent également que la moyenne masque de grandes disparités dans l'effet du programme. Pour les femmes, l'effet moyen du traitement, c'est-à-dire la différence entre la moyenne espérée des revenus en l'absence du programme et celle des revenus avec le programme est de 1 825 dollars. Mais il n'augmente en fait que de 390 dollars le premier quartile de la distribution de revenus, alors que l'augmentation atteint 2 800 dollars pour le dernier quartile (voir le graphique 4). Les différences sont également très marquées pour les hommes, mais les estimations sont bien plus imprécises et ne permettent jamais d'exclure leur nullité aux seuils ordinaires de significativité. Les intervalles de confiance sont en effet larges. De manière similaire aux régressions linéaires classiques, la précision des paramètres estimés à partir d'une méthode instrumentale est bien plus faible qu'avec une régression quantile standard, même si comme ici les instruments prédisent correctement les variables explicatives endogènes.

En termes d'interprétation, on rappelle que ces valeurs correspondent à la différence des quantiles des distributions de revenus qu'on s'attend à observer respectivement en l'absence ou en présence du programme de formation. Cela ne signifie pas a priori que les femmes dont le revenu serait au niveau du premier quartile en l'absence de la formation vont bénéficier d'une augmentation de 390 dollars grâce à celle-ci, sauf à faire l'hypothèse que la formation ne modifie pas l'ordre relatif des revenus des personnes. Cela pourrait ne pas être le cas si le programme bénéficie beaucoup plus à des personnes en bas de l'échelle des revenus qu'à ceux au dessus et que les revenus potentiels se croisent. Par ailleurs, si on peut interpréter le résultat obtenu par les doubles moindres carrés comme l'effet moyen de la formation (et pas seulement comme la différence des moyennes des revenus avec et sans la formation), il n'est pas en général possible d'interpréter des résultats obtenus par la régression quantile (ici instrumentée) comme ceux de la distribution des effets. Autrement dit, il n'est pas possible de dire qu'un quart des jeunes femmes va bénéficier d'au moins 390 dollars grâce à la formation, car la différence des premiers quartiles des distributions de revenus avec et sans formation ne correspond pas a priori avec le premier quartile des gains individuels liés à la formation. Les régressions quantiles permettent de décrire comment se déforme la distribution des niveaux de revenus (quelle répartition des richesses on observe grâce au programme), non celle des évolutions (comment se répartissent les

17. Les données sont disponibles à l'adresse `http://econ-www.mit.edu/faculty/angrist/data1/data/abangim02`.

gains). Cela n'en réduit pas l'intérêt. Si le gain moyen est une information essentielle pour l'évaluation, par exemple en regard des coûts engagés pour ce programme, la répartition finale des niveaux de revenus est également importante, dans la mesure où elle permet, par exemple, de déterminer si au moins x% des personnes sont au-dessus d'un seuil minimal de revenu. Cette information est souvent plus pertinente que l'identification précise des gagnants et des perdants.



Femmes Hommes

Lecture : les paramètres estimés sont la moyenne (en pointillé, estimation par les doubles moindres carrés), la médiane, le premier et le dernier décile, le premier et le dernier quartile. Les zones grisées correspondent à l'intervalle de confiance à 95%, estimé par une méthode de bootstrap. L'effet moyen de la formation sur les revenus des jeunes femmes est de 1825 dollars. L'impact de cette formation sur le premier décile de la distribution des revenus des jeunes femmes est proche de zéro.

FIGURE 4 – Estimation de l'impact du programme de formation, régression quantile instrumentée.

5 Pour conclure

En conclusion, la régression quantile est un outil, facile d'utilisation, qui permet d'enrichir la description quantitative des phénomènes économiques et sociaux. Cet article en présente les principes et précise l'utilisation qui peut en être fait. On insiste sur le fait que les problèmes, classiques, d'endogénéité éventuelle des variables explicatives doivent être traités ce qui peut nécessiter de mobiliser des outils d'identification spécifiques détaillés dans la partie 3 (à condition de disposer des instruments ou des données nécessaires). Par ailleurs, les régressions quantiles modélisent les quantiles conditionnels de la distribution d'intérêt. L'objet d'étude est donc avant tout cette distribution. Tout comme l'objet d'une régression linéaire n'est pas de renseigner sur l'effet de covariables sur un hypothétique "individu moyen", le premier objectif des régressions quantiles n'est pas de décrire l'effet de ces covariables sur des individus "assignés" à une place dans la distribution. Par exemple, on pourra déduire de l'évaluation des effets d'un programme éducatif par une régression quantile au niveau du décile qu'il permet d'augmenter de x points le niveau atteint par au moins 90% des élèves. En toute rigueur, on ne peut dire sans hypothèse supplémentaire qu'il permet d'augmenter d'autant le niveau des élèves les plus faibles en l'absence de ce programme. Cette interprétation suppose que les enfants les plus en difficulté avec à une méthode d'apprentissage le sont également face à une méthode différente. Cette hypothèse d'invariance des rangs, vraisemblable ici, n'est pas nécessaire pour mener une analyse de régression quantiles.

Expliciter ces précautions indispensables pour l'interprétation ne doit pas conduire le lecteur à conclure que la régression quantile est un outil sophistiqué mais peu utile pour l'analyse. Le bilan d'un programme ne se fait pas en rapport à son effet sur tel ou tel individu, mais sur sa capacité à améliorer ou non une situation générale. Ainsi, estimer qu'un programme augmente significativement les plus bas déciles mais n'a pas d'effet sur les déciles supérieurs signifie qu'il a permis de réduire les inégalités de résultats, sans diminution des ambitions scolaires (les résultats des têtes de classe sont aussi bons, même si - en principe- rien ne garantit que les têtes de classes sont les mêmes avec et sans programme). Quantiles et fonction de répartition étant naturellement liés, on peut aussi, au prix de quelques manipulations, tirer des conclusions sur le fait que le programme permet de ramener de tant à tant la proportion d'élèves en dessous d'un seuil (dans notre exemple, un niveau minimum d'acquisition des connaissances). De fait, de nombreux articles récents ont montré que l'analyse « au-delà de la moyenne » permet d'enrichir les évaluations quantitatives de politiques publiques. Ainsi Bitler et al. (2006) montre qu'une même réforme peut avoir des effets négligeables en moyenne mais modifier significativement la distribution des revenus. Casalone & Sonedda (2013) évaluent les effets d'une réforme fiscale sur la distribution des revenus, Jackson & Page (2013) l'impact d'un programme de réduction de la taille des classes (projet STAR) sur la réussite scolaire des enfants (voir aussi Bitler et al., 2014). Les régressions quantiles sont également des outils adaptés pour étudier l'évolution des inégalités de revenus, à la suite des travaux de Buchinsky sur le sujet. Enfin, même lorsque l'objet d'intérêt n'est pas l'ensemble de la distribution, les propriétés statistiques des régressions quantiles en font une alternative aux méthodes économétriques classiques dans certains cas (variable tronquée ou censurée, valeurs aberrantes...).

A Annexe : propriétés des quantiles et détails sur l'inférence

A.1 Quelques propriétés des quantiles

A.1.1 Définitions

Le quantile d'ordre $\tau \in (0,1)$ d'une variable aléatoire réelle U est défini par

$$q_\tau(U) = \inf\{x | F_U(x) \geq \tau\},$$

F_U étant la fonction de répartition de U. Dans le cas où F_U est continue et strictement croissante, on a simplement $q_\tau(U) = F_U^{-1}(\tau)$. Le graphique 5 illustre la définition des quantiles dans le cas général.

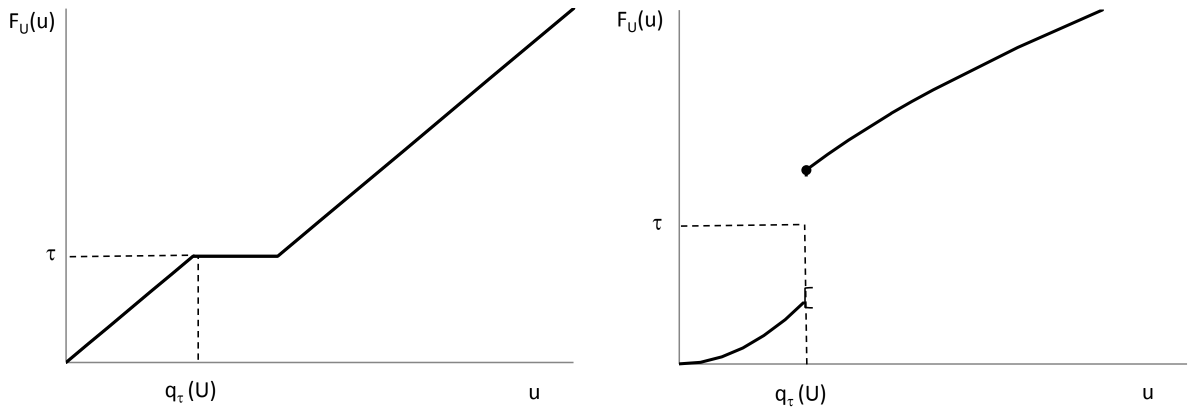


Figure 5 – Quantile d'une variable dans le cas général.

Pour deux variables aléatoires U et V, le quantile conditionnel $q_\tau(U|V)$ est défini de manière similaire par :

$$q_\tau(U|V) = \inf\{x | F_{U|V}(x) \geq \tau\},$$

où $F_{U|V}$ est la fonction de répartition de U conditionnelle à V.

A.1.2 Robustesse aux valeurs aberrantes

Nous démontrons tout d'abord les propriétés de robustesse des régressions quantiles aux valeurs aberrantes énoncées en partie 1.2.1. On suppose tout d'abord

$$Y = AX'\delta + (1-A)(X'\gamma + \varepsilon),$$

avec ε, A et X mutuellement indépendants. A est une variable inobservée valant 1 lorsque Y est aberrant, 0 sinon et $P(A=1) = p$. Le paramètre β_τ de la régression quantile vérifie, pour tout x, $\tau = P(Y \leq x'\beta_\tau | X = x)$. Par conséquent,

$$\begin{aligned}
\tau &= P(Y \leq X'\beta_\tau | X = x) \\
&= P(A = 1|X = x)P(Y < X'\beta_\tau | X = x, A = 1) + P(A = 0|X = x)P(Y \leq X'\beta_\tau | X = x, A = 0) \\
&= pP(x'\delta < x'\beta_\tau | X = x) + (1-p)P(x'\gamma + \epsilon < x'\beta_\tau | X = x, A = 0) \\
&= (1-p)P(\epsilon < x'(\beta_\tau - \gamma)),
\end{aligned}$$

où la dernière égalité vient du fait que $x'\delta$ est supposé très grand, et donc supérieur à $x'\beta_\tau$, et ϵ est indépendant de (A, X). Ainsi, pour tout x,

$$x'(\beta_\tau - \gamma) = q_{\frac{\tau}{1-p}}(\epsilon)$$

On en déduit que $\beta_{k,\tau} = \gamma_k$ pour tout $k > 1$ et $\beta_{1,\tau} = \gamma_1 + q_{\frac{\tau}{1-p}}(\epsilon)$. Si l'on excepte la constante, les coefficients de la régression quantile que l'on obtient sont donc égaux à ceux que l'on obtiendrait en l'absence de valeur aberrante.

Considérons maintenant le modèle plus général suivant :

$$Y = AX'\delta + (1-A)(X'\beta_U),$$

où U, A et X sont mutuellement indépendants et U suit une loi uniforme sur $[0,1]$. Notons $\widetilde{\beta}_\tau$ le coefficient de la régression quantile de Y sur X. En suivant le même raisonnement que précédemment, on obtient l'équation suivante en $\widetilde{\beta}_\tau$, valable pour tout x :

$$\tau = (1-p)P(x'\beta_U < x'\widetilde{\beta}_\tau). \tag{20}$$

Le paramètre $\beta_{\tau/(1-p)}$ est solution de cette équation. Dès que le modèle est identifié, il existe une seule solution vérifiant (20) pour tout x. Par conséquent, $\widetilde{\beta}_\tau = \beta_{\tau/(1-p)}$.

A.1.3 Invariance à une transformation monotone

Les quantiles satisfont l'importante propriété d'invariance suivante.

Proposition A.1. *Soit g une fonction croissante et continue à gauche. Alors :*

$$g(q_\tau(U)) = q_\tau(g(U)).$$

Preuve : Grâce à la monotonicité de g on a $P(U \leq q_\tau(U)) = P(g(U) \leq g(q_\tau(U)))$ et par définition de $q_\tau(U)$: $\tau \leq P(U \leq q_\tau(U))$. Or par définition on a aussi $q_\tau(g(U)) = \inf\{x \in \mathbb{R} | F_{g(U)}(x) \geq \tau\}$, donc $g(q_\tau(U)) \geq q_\tau(g(U))$. Réciproquement, en définissant $g^-(v) = \sup\{x | g(x) \leq v\}$, on a :

$$P(g(U) \leq q_\tau(g(U))) \leq P(U \leq g^-(q_\tau(g(U)))).$$

Par définition de $q_\tau(g(U))$ et $q_\tau(U) = \inf\{x \in \mathbb{R} | F_U(x) \geq \tau\}$ on en déduit que : $g^-(q_\tau(g(U))) \geq q_\tau(U)$. De la continuité de g à gauche, on a aussi que $g(g^-(q_\tau(g(U)))) \leq q_\tau(g(U))$. Donc $q_\tau(g(U)) \geq g(q_\tau(U))$, ce qui conclut la preuve. □

Ce résultat implique notamment que $q_\tau(aU+b) = aq_\tau(U)+b$, ou, de même, $q_\tau(a(X)U+b(X)|X) = a(X)U+b(X)$. Mais il implique également que $q_\tau(\max(s,U)) = \max(s, q_\tau(U))$, ou que $q_\tau(\mathbb{1}\{U>0\} = \mathbb{1}\{q_\tau(U)>0\}$. En revanche, et contrairement à l'espérance, la fonction quantile n'est pas linéaire : on a en général $q_\tau(U_1+U_2) \neq q_\tau(U_1)+q_\tau(U_2)$.

A.1.4 Quantiles et minimisation de perte

La propriété suivante est cruciale pour l'estimation.

Proposition A.2. *Soit* $\rho_\tau(u) = (\tau - \mathbb{1}\{u<0\})\, u$. *On a :*

$$q_\tau(U) \in \arg\min_a E[\rho_\tau(U-a)].$$

Preuve : Soit $m(a) = E[\rho_\tau(U-a)]$. Prenons $a \leq q_\tau(U)$ et montrons que $m(a) \geq m(q_\tau(U))$. On a, après quelques calculs,

$$m(a) - m(q_\tau(U)) = (q_\tau(U)-a)\tau + aF_U(a) - q_\tau(U)F_U(q_\tau(U)^-) + E\left[U\mathbb{1}\{U \in [a, q_\tau(U)[\}\right],$$

où $F_U(q_\tau(U)^-) = \lim_{x\uparrow q_\tau(U)} F_U(x)$. Par ailleurs,

$$E\left[U\mathbb{1}\{U \in [a, q_\tau(U)[\}\right] \geq aE\left[\mathbb{1}\{U \in [a, q_\tau(U)[\}\right] = a\left(F_U(q_\tau(U)^-) - F_U(a)\right).$$

Donc

$$m(a) - m(q_\tau(U)) \geq (q_\tau(U)-a)(\tau - F_U(q_\tau(U)^-)).$$

Par définition des quantiles, $F_U(q_\tau(U)^-) \leq \tau$. Donc $m(a) \geq m(q_\tau(U))$. On montre de même que $m(a) \leq m(q_\tau(U))$ pour tout $a \geq q_\tau(U)$. □

Notons que le minimum de $a \mapsto E[\rho_\tau(U-a)]$ n'est pas unique en général. Ceci provient du fait que l'équation $F_U(a) = \tau$ peut avoir plusieurs solutions. Le minimum est cependant unique si la fonction de répartition de U est strictement croissante.

A.2 Détails sur les méthodes d'inférence

A.2.1 Estimation directe

Cette approche consiste à estimer directement la variance asymptotique en partant de la formule 9. Dans le cas général, la difficulté principale est d'estimer $J_\tau = E(f_{\varepsilon_\tau|X}(0|X)XX')$. Pour ce faire, Powell (1991) propose de s'appuyer sur l'idée suivante :

$$J_\tau = \lim_{h\to 0} E\left[\frac{\mathbb{1}\{|\varepsilon_\tau| \leq h\}}{2h} XX'\right].$$

On estime alors J_τ par

$$\widehat{J}_\tau = \frac{1}{2nh_n}\sum_{i=1}^n \mathbb{1}\{|\widehat{\varepsilon}_{i\tau}| \leq h_n\}X_iX_i'. \qquad (21)$$

où $h_n \to 0$ et $\sqrt{n}h_n \to \infty$.

Cette formule est plus simple dans le cas du modèle de translation, puisque seule l'estimation de $1/f_\varepsilon(q_\tau(\varepsilon)$ est problématique. Soit $\widehat{\varepsilon}_{i\tau} = Y_i - X_i'\widehat{\beta}_\tau$, on peut alors estimer $1/f_\varepsilon(q_\tau(\varepsilon)$ [18] par $(\widehat{\varepsilon}_{(\lceil n(\tau+h_n)\rceil)\tau} - \widehat{\varepsilon}_{(\lceil n(\tau-h_n)\rceil)\tau})/2h_n$. L'estimateur de la variance asymptotique vaut alors :

$$\widehat{V}_{\text{as}} = \tau(1-\tau)\left(\frac{\widehat{\varepsilon}_{(\lceil n(\tau+h_n)\rceil)\tau} - \widehat{\varepsilon}_{(\lceil n(\tau-h_n)\rceil)\tau}}{2h_n}\right)^2 \left[\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right]^{-1}. \tag{22}$$

Cet estimateur est parfois proposé par défaut dans des logiciels standard. Il faut cependant garder à l'esprit qu'il n'est convergent que dans le très restrictif modèle de translation.

Une fois obtenu un estimateur convergent de V_{as}, l'inférence sur β_τ est aisée. Un intervalle de confiance de niveau $1-\alpha$ sur β_τ s'écrit ainsi :

$$IC_\alpha = \left[\widehat{\beta}_\tau - z_{1-\alpha/2}\sqrt{\widehat{V}_{\text{as}}}, \widehat{\beta}_\tau + z_{1-\alpha/2}\sqrt{\widehat{V}_{\text{as}}}\right],$$

où $z_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ d'une loi $\mathcal{N}(0,1)$. De même, la statistique de Wald T du test $\beta_\tau = 0$ s'écrit $T = n\widehat{\beta}_\tau' \widehat{V}_{\text{as}}^{-1}\widehat{\beta}_\tau$, avec T qui tend vers un χ^2_p sous l'hypothèse nulle, où p est le nombre de variables explicatives.

A.2.2 Bootstrap

Une autre possibilité pour faire de l'inférence est de recourir au bootstrap. Rappelons que le principe de bootstrap est de générer des échantillons "factices" par des tirages avec remises à partir de l'échantillon initial. Dans le cas du bootstrap standard, on applique l'algorithme suivant.

De $b = 1$ à B :
- Tirer avec remise un échantillon de taille n à partir de l'échantillon initial $(Y_i, X_i)_{i=1...n}$. Soit $(k^*_{b1}, ..., k^*_{bn})$ les indices correspondants aux observations tirées ;
- Calculer $\widehat{\beta}^*_{\tau b} = \arg\min_\beta \sum_{j=1}^n \rho_\tau(Y_{k^*_{bj}} - X'_{k^*_{bj}}\beta)$.

On peut alors estimer la variance asymptotique par

$$V^*_{\text{as}} = \frac{1}{B}\sum_{b=1}^{B}(\widehat{\beta}^*_{\tau b} - \widehat{\beta})^2.$$

Des intervalles de confiance ou tests peuvent être alors construits comme précédemment, en utilisant l'approximation normale. Pour construire des intervalles de confiance, on peut

18. On a en effet $\frac{1}{f_\varepsilon(q_\tau(\varepsilon))} = \frac{1}{f_\varepsilon(F_\varepsilon^{-1}(\tau))} = \frac{\partial F_\varepsilon^{-1}}{\partial \tau}(\tau) = \lim_{h\to 0}\frac{F_\varepsilon^{-1}(\tau+h) - F_\varepsilon^{-1}(\tau-h)}{2h}$

également s'appuyer sur le *percentile bootstrap*. Soit q_u^* le quantile empirique d'ordre u de $(\widehat{\beta}^*_{\tau 1}, ..., \widehat{\beta}^*_{\tau B})$, on construit simplement l'intervalle de confiance par

$$IC_{1-\alpha} = [q^*_{\alpha/2}, q^*_{1-\alpha/2}].$$

Par rapport à l'estimateur (22), les méthodes de bootstrap ont l'avantage de ne pas supposer que le vrai modèle est un modèle de translation. Elles évitent également de devoir choisir le paramètre de lissage h_n, sachant que les résultats peuvent être sensibles à ce choix.

Références

Abadie, A., Angrist, J. & Imbens, G. (2002), 'Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings', *Econometrica* **70**(1), 91–117.

Bilias, Y. & Koenker, R. (2001), 'Quantile regression for duration data : A reappraisal of the pennsylvania reemployment bonus experiments', *Empirical Economics* **26**(1), 199–220.

Bitler, M. P., Gelbach, J. B. & Hoynes, H. W. (2006), 'What Mean Impacts Miss : Distributional Effects of Welfare Reform Experiments', *American Economic Review* **96**(4), 988–1012.

Bitler, M. P., Gelbach, J. B. & Hoynes, H. W. (2014), Can Variation in Subgroups' Average Treatment Effects Explain Treatment Effect Heterogeneity ? Evidence from a Social Experiment, NBER Working Papers 20142, National Bureau of Economic Research, Inc.

Buchinsky, M. (1994), 'Changes in the U.S. Wage Structure 1963-1987 : Application of Quantile Regression', *Econometrica* **62**(2), 405–58.

Buchinsky, M. (1998), 'The dynamics of changes in the female wage distribution in the USA : a quantile regression approach', *Journal of Applied Econometrics* **13**(1), 1–30.

Cade, B. S. & Noon, B. R. (2003), 'A Gentle Introduction to Quantile Regression for Ecologists', *Frontiers in Ecology and The Environment* **1**, 412–420.

Canay, I. A. (2011), 'A simple approach to quantile regression for panel data', *The Econometrics Journal* **14**(3), 368–386.

Casalone, G. & Sonedda, D. (2013), 'Evaluating The Distributional Effects Of Fiscal Policies Using Quantile Regressions', *Review of Income and Wealth* **59**(2), 305–325.

Charnoz, P., Coudin, . & Gaini, M. (2011*a*), Wage inequalities in france 1976-2004 : a quantile regression analysis, Technical report.

Charnoz, P., Coudin, E. & Gaini, M. (2011*b*), Wage inequalities in France 1976-2004 : a quantile regression analysis, Documents de Travail de la DESE - Working Papers of the DESE g2011-06, Institut National de la Statistique et des Etudes Economiques, DESE.

Chernozhukov, V., Fernandez-Val, I. & Galichon, A. (2010), 'Quantile and Probability Curves Without Crossing', *Econometrica* **78**(3), 1093–1125.

Chernozhukov, V. & Hansen, C. (2008), 'Instrumental variable quantile regression : A robust inference approach', *Journal of Econometrics* **142**, 379–398.

Clements, N., Heckman, J. & Smith, J. (1997), 'Making the most out of programme evaluations and social experiments : Accounting for heterogeneity in programme impacts', *Review of Economic Studies* **64**(4), 487–535.

Cornec, M. (2010), Constructing a conditional gdp fan chart with an application to french business survey data. mimeo Insee.

Doksum, K. (1974), 'Empirical probability plots and statistical inference for nonlinear models in the two-sample case', *The Annals of Statistics* **2**(2), pp. 267–277.

Fack, G. & Landais, C. (2009), 'Les incitations fiscales aux dons sont-elles efficaces ?', *Économie et Statistique* **427**(1), 101–121.

Firpo, S. (2007), 'Efficient semiparametric estimation of quantile treatment effects', *Econometrica* **75**(1), 259–276.

Firpo, S., Fortin, N. M. & Lemieux, T. (2009), 'Unconditional quantile regressions', *Econometrica* **77**(3), 953–973.

Fitzenberger, B. & Wilke, R. A. (2005), Using quantile regression for duration analysis, ZEW Discussion Papers 05-65, ZEW / Center for European Economic Research.

Fortin, N., Lemieux, T. & Firpo, S. (2011), *Decomposition Methods in Economics*, Vol. 4 of *Handbook of Labor Economics*, Elsevier, chapter 1, pp. 1–102.

Frandsen, B. R., Frolich, M. & Melly, B. (2012), 'Quantile treatment effects in the regression discontinuity design', *Journal of Econometrics* **168**(2), 382–395.

Givord, P. (2010), Méthodes économétriques pour l'évaluation des politiques publiques, Documents de Travail de la DESE - Working Papers of the DESE g2010-08, Institut National de la Statistique et des Etudes Economiques, DESE.

He, X. & Hu, F. (2002), 'Markov chain marginal bootstrap', *Journal of the American Statistical Association* **97**(459), pp. 783–795.

Hong, H. & Chernozhukov, V. (2002), 'Three-step censored quantile regression and extramarital affairs', *Journal of the American Statistical Association* **97**, 872–882.

Hyndman, R. J. & Fan, Y. (1996), 'Sample quantiles in statistical packages', *The American Statistician* **50**(4), pp. 361–365.

Jackson, E. & Page, M. E. (2013), 'Estimating the distributional effects of education reforms : A look at Project STAR', *Economics of Education Review* **32**(C), 92–103.

Kocherginsky, M., He, X. & Mu, Y. (2005), 'Practical confidence intervals for regression quantiles', *Journal of Computational and Graphical Statistics* **14**(1), 41–55.

Koenker, R. (2004), 'Quantile regression for longitudinal data', *Journal of Multivariate Analysis* **91**(1), 74–89.

Koenker, R. (2005), *Quantile Regression*, Econometric Society Monograph Series, Cambridge University Press.

Koenker, R. & Hallock, K. F. (2001), 'Quantile regression', *Journal of Economic Perspectives* **15**(4), 143–156.

Koenker, R. W. & Machado, J. A. (1999), 'Goodness of fit and related inferences processes for quantile regression', *Journal of the American Statistical Association* **94**(448), pp. 1296–1310.

Lamarche, C. (2010), 'Robust penalized quantile regression estimation for panel data', *Journal of Econometrics* **157**(2), 396–408.

Landais, C. (2007), Les hauts revenus en France (1998-2006) : Une explosion des inégalités ? mimeo Paris School Economics.

Machado, J. A. & Silva, J. M. C. S. (2005), 'Quantiles for counts', *Journal of the American Statistical Association* **100**, 1226–1237.

Portnoy, S. (1992), 'Asymptotic behavior of the number of regression quantile breakpoints', *SIAM Journal on Scientific and Statistical Computing* **12**(4), 867–883.

Portnoy, S. & Koenker, R. (1997), 'The gaussian hare and the laplacian tortoise : Computability of squared- error versus absolute-error estimators', *Statistical Science* **12**(4), pp. 279–296.

Powell, J. (1984), 'Least absolute deviations estimation for the censored regression model', *Journal of Econometrics* **25**, 303–325.

Powell, J. L. (1991), *Estimation of monotonic regression models under quantile restrictions*, Cambridge : Cambridge University Press.

Solard, J. (2010), 'Les très hauts revenus : des différences de plus en plus marquées entre 2004 et 2007', *Insee Références* **Editions 2010**, 45–64.

Wooldridge, J. W. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press.