Régressions de quantiles

Pauline Givord

CREST, INSEE

2011/2012

Introduction

- 95% des études empiriques portent sur l'estimation d'effets moyens
- Si le revenu est continu (salaire, niveau scolaire...), la moyenne ne contient qu'une petite partie de l'information.

Enrichir le diagnostic

Des informations plus complètes sur la distribution sont souvent utiles :

- analyse des inégalités
  ex : stabilité du revenu moyen aux US mais forte progression du dernier décile (Buchinsky, 1998,...)
- en terme d'évaluation des politiques publiques :
  mesure peut avoir un impact moyen nul
  mais être jugée "souhaitable" si affecte positivement suffisamment de personnes, ou suffisamment de certaines personnes.
  Exemples : échec scolaire, exclusion...

Régression de quantiles : étendre les techniques des régressions linéaires aux quantiles

Répondre aux problèmes soulevés par la nature de certaines variables

- Moindre sensibilité que la moyenne à la présence de valeurs extrêmes.
  Exemple : données $=(1, 3, 3, 5, 7, 9, 10, 10, 300)$
  médiane $= 7$, moyenne$=38$
- Peut arriver en cas d'erreur de codage, ou du fait de certaines distributions (log-logistique par exemple), pour lequel la probabilité d'avoir une valeur très élevée reste importante

Répondre aux problèmes soulevés par la nature de certaines variables

- Données censurées, modèle Tobit... Une propriété intéressante des quantiles est l'équivariance par transformation monotone : si h est une fonction croissante, $Q_\tau(h(Y)|X) = h(Q_\tau(Y|X))$ (ce qui n'est pas le cas pour la moyenne !).
  Modèle Tobit (on observe uniquement les valeurs sup > 0) : $h(x) = max(0, x)$
- Utilisation pour éviter de spécifier la forme paramétrique de certaines distributions (modèles de durée, de comptage)

Plan du cours

- introduction (succincte) aux régressions de quantiles
  Pour plus de détails : cf. Buchinsky (1998) et Koenker et Hallock (2002)
- application(s) au cadre de l'évaluation : quantile treatment effect

Définition

Rappel : Pour une v.a. Y de distribution F $(F(y) = P(Y < y))$,
τ^{ieme} quantile : $Q_\tau(Y) = inf\ \{y : F(y) \geq \tau\}$.
Remarque : ie. q tq $P(Y < q) = \tau$, ou $F(q) = \tau$, ou encore
$q = F^{-1}(\tau)$ si F "bonnes" propriétés.

Quantiles usuels :
médiane $(\tau = 0.5)$, premier et dernier décile $(\tau = 0.1$ et $\tau = 0.9)$,
premier et dernier quartile $(\tau = 0.25$ et $\tau = 0.75)$.

Minimisation

Pour la suite, utile de voir les quantiles comme la solution d'un programme de minimisation.

- cas le plus simple : Médiane
  solution de $argmin_b \sum_i |Y_i - b|$: Least Absolute Deviation (LAD)

Intuition

Dem : Conditions du premier ordre

$$\frac{\partial S(y,b)}{\partial b} = \sum_i (1(u_i > 0)(+1) + 1(u_i < 0)(-1))$$

avec $u_i = y_i - b$. Donc min pour b tel que autant de résidus négatifs que positifs : autant de y_i plus grands que b que de y_i plus petits : médiane par définition.
Remarque : ne dépend que du signe des résidus $(y_i - b)$, donc robuste à la présence de points extrêmes

check function

- cas général : pondérations différentes des observations :

$$argmin_b \sum_{i:Y_i \geq b} \tau|Y_i - b| + \sum_{i:Y_i < b} (1-\tau)|Y_i - b|$$

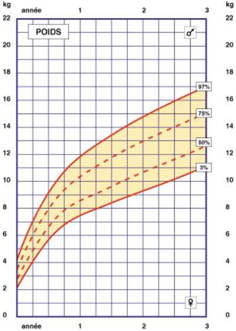
- ou encore : $argmin_b \sum_i \rho_\tau(Y_i - b)$ la fonction de pondération s'appelle "check function" :

$$\rho_\tau(u) = u(\tau - 1(u \neq 0))$$

Régression de quantile

On s'intéresse à la distribution conditionnelle selon des caractéristiques X.

e.g. : courbe de Quetelet



Quantile regression

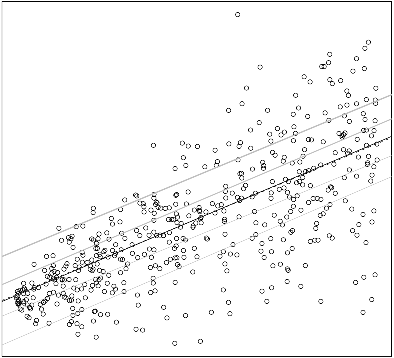
Motivation : les variables observables X affectent l'ensemble de la distribution
L'impact sur les quantiles les plus élevés peut être très différent que pour le milieu de la distribution

- on suppose une dépendance linéaire de chaque quantile dans ces observables.
- on remplace b dans le programme précédent par une fonction linéaire dans les covariables
- on estime :

$$\beta_\tau = \arg min_\beta \sum \rho_\tau(Y_i - X_i\beta)$$

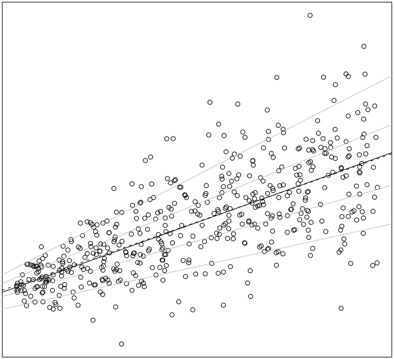
Exemple : modéle de translation

On suppose que le modèle sous-jacent est simplement :
$Y = X'\beta + U$



Exemples : modéle de translation/échelle

Si on ajoute un peu d'homoscédasticité : $Y = X'\beta + (X'\gamma)U$



estimation

- fonction objectif non différentiable, donc procédure du gradient classique non utilisable
- peut s'écrire comme solution d'un modèle de programmation linéaire (Koenker et Bassett, 1978).
- implémentation sous stata (qreg, sqreg) ou R (rq), sas (quantreg).

Résultats

- on peut faire une régression de quantile pour chaque τ :

$$Q_{Y|X}(\tau) = \beta_\tau X$$

- l'impact peut être résumé par la fonction :

$$\tau \to \beta_\tau$$

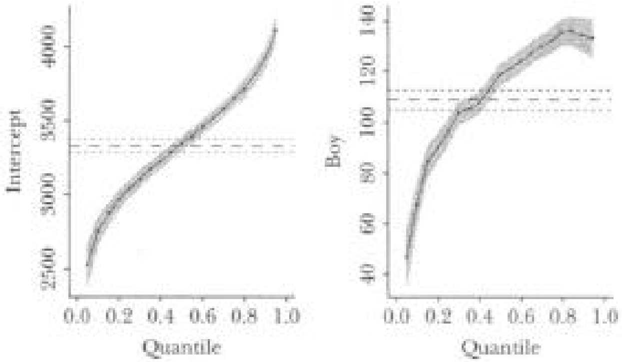
- i.e. il y a un vecteur de coefficient différent pour chaque valeur de τ
- présentation des résultats : par exemple graphe des coefficients comme une fonction des quantiles

Exemple : Koenker and Hallock

- Poids de naissance
- fonction de caractéristiques de la mère
- situation familiale, éducation, age...
- pratiques (fumeuse, suivi maternel)

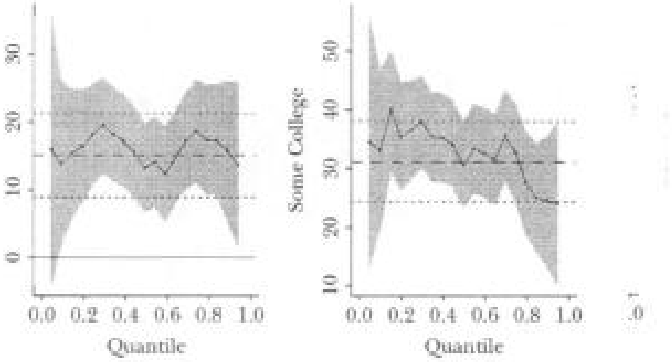
Exemple : Koenker and Hallock

Moyenne et sexe du bébé



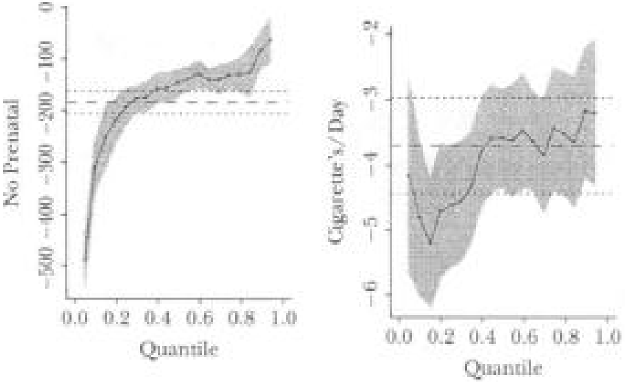
Exemple : Koenker and Hallock

Education de la mère



Exemple : Koenker and Hallock

Fumeuse, suivi prénatal



Exemple 2 : Moreira et Pita Barros

"Double coverage and demand for health care : evidence from quantile regression", JHE 2010

- évaluation de l'impact d'une mutuelle complémentaire sur la consommation médicale
- système portuguais :
  - assurance maladie (presque) universelle, mais implique franchise et délai pour accès aux soins
  - systèmes complémentaires pour certains salariés (public et certaines entreprises privées)
- quel impact sur le nombre de consultations médicales ?

Exemple 2 : Moreira et Pita Barros

- modélisation de la variable d'intérêt (comptage) par régression de quantiles
- deux avantages :
  - alternative moins paramétrique à une spécification de cette variable
  - plutôt qu'un effet moyen (sur une variable qui peut être très "étalée), permet de répondre à des questions type : est-ce que la double couverture santé (=mutuelle) a des effets différenciés pour les plus grandes consommations importantes de soin vs les consommations plutôt basses (dans la distribution)

Exemple 2 : Moreira et Pita Barros

Table 2: Empirical distribution of the dependent variable

	TOTAL	NHS	Public sub.	Private sub.
y		relative frequency		
0	50.31	50.88	48.82	41.91
1	26.94	26.53	28.54	29.83
2	10.78	10.61	11.37	12.61
3	6.77	6.82	6.15	8.72
4	1.99	2.02	1.69	2.84
5	1.12	1.06	1.25	2.10
6	0.98	0.95	1.17	0.95
7	0.20	0.21	0.14	0.21
8	0.22	0.23	0.18	0.42
9	0.08	0.07	0.13	0.11
10	0.19	0.17	0.30	0.11
11-15	0.25	0.28	0.15	0.11
16-20	0.04	0.04	0.06	0.11
21-25	0.06	0.06	0.06	0.00
26-30	0.06	0.07	0.02	0.00

Exemple 2 : Moreira et Pita Barros

- données administratives
- variables : état de santé (capté par maladie), variables socio-demo standard (age, sexe...), saisonnière et régionale
- présence d'une double couverture

Résultats : estimations

	$\widehat{\beta(0.25)}$	$\widehat{\beta(0.50)}$	$\widehat{\beta(0.60)}$	$\widehat{\beta(0.70)}$	$\widehat{\beta(0.80)}$	$\widehat{\beta(0.90)}$
Health insurance status variables						
pubsub	$0.078^{\dagger}$	0.088	0.095	0.096	0.073	$0.055^{\dagger}$
privsub	0.200	0.229	0.247	0.232	0.185	0.148

Interprétation

- on mesure comment les quantiles de la distribution conditionnelle changent avec les covariables
- paramètre d'intérêt : $\frac{\partial EQ_{Y|X}(\tau)}{\partial X_j}$
- i.e. changement marginal dans le τ^{th} quantile conditionnelle après un changement marginal dans X_j.
- si x_j est linéaire, c'est juste β_j

quantile conditionnel versus inconditionnel

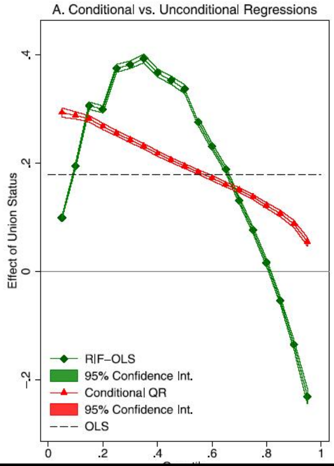
- Les estimations comparent les quantiles des distributions conditionnelles entre eux
- dans le cadre linéaire classique (MCO), suffisant pour obtenir des estimateurs convergents de l'impact de X sur la moyenne *inconditionnelle* du revenu Y :
  $E[Y|X] = X\beta$ implique que $E[Y] = E[X]\beta$
- mais on n'a pas cette propriété pour les quantiles conditionnels :
  $$q_Y(\tau) \neq E[q_{Y|X}(\tau)]$$
- Il sera plus complexe d'obtenir de telles distributions contrefactuelles. On peut préférer mobiliser des méthodes spécifiques.

quantile conditionnel versus inconditionnel : Exemple

Firpo, Fortin and Lemieux (2009) "Unconditional quantile regressions ", Econometrica

- question empirique : "Quel est l'impact d'augmenter la proportion de salariés syndiqués sur l'ensemble de la distribution de revenus, toute choses égales par ailleurs"
- les estimateurs classiques de régression de quantile ne permettent pas de répondre à cette question simple
- FFL proposent une estimation de l'effet du traitement sur les quantiles inconditionnels (sous l'hypothèse d'exogénéité sur T)

cond. versus incond. quantile impact du statut syndiqué ou pas sur les revenus



impact du statut sur les quantile de revenus cond. versus incond.

- syndicat réduit la dispersion intra-groupe (“groupe” = salariés avec les mêmes caractéristiques observables X)
- mais augmente la moyenne conditionnelle des salaires des syndiqués (i.e. inégalités entre les groupes)
- donc tendance à augmenter les salaires pour les quantiles inconditionnels faibles (les deux effets vont dans la même direction)
- mais décroître les salaires pour les hauts quantiles inconditionnels (effets en sens opposés)

définition

Cadre de l'évaluation : on s'intéresse à l'effet du traitement binaire T sur un revenu Y.
Soit Y_0 et Y_1 les revenus avec et sans traitement, F_{Y_0} et F_{Y_1} les distributions correspondantes.
On s'intéresse au τ^{ieme} quantile treatment effect (QTE) :

$$\delta_\tau = F_{Y_1}^{-1}(\tau) - F_{Y_0}^{-1}(\tau)$$

"Distance" horizontale entre les deux distributions (Lehmann, 1974 et Doksum, 1974).
De même, on peut définir sa restriction aux traités (QTET) :

$$\delta_{\tau|T=1} = F_{Y_1}^{-1}(\tau|T=1) - F_{Y_0}^{-1}(\tau|T=1)$$

Interprétation

- On estime la différence des quantiles et non le quantile de la différence (effet du traitement) $Y_1 - Y_0$
  $\rightarrow$ Pas d'interprétation individuelle.
- pour passer à l'interprétation individuelle, il faut faire une hypothèse d'invariance des rangs : une personne dont le revenu sans traitement se trouve dans un certain quantile se trouve dans le même quantile de la distribution des revenus avec traitement.
  Si i tq $Y_{0i} < Q_{Y_0}(\tau)$, alors $Y_{1i} < Q_{Y_1}(\tau)$

Illustration

- Supposons que la population ne comprend que cinq individus (A,B,C,D,E)
- avec une distribution des revenus potentiels sans traitement :
  $(Y_0^A = 1, Y_0^B = 2, Y_0^C = 4, Y_0^D = 5, Y_0^E = 9)$
  et pour les revenus potentiels avec traitement :
  $(Y_1^A = 4, Y_1^B = 6, Y_1^C = 5, Y_1^D = 11, Y_1^E = 10)$
- la médiane du revenu potentiel sans traitement est 4
  la médiane du revenu potentiel sans traitement est 6

Illustration

- l'effet du traitement pour la médiane est 2
- ce qui est différent de la médiane des effets individuels du traitement
  $(\Delta Y^A = 3, \Delta Y^B = 3, \Delta Y^C = 1, \Delta Y^D = 6, \Delta Y^E = 1)$ soit 3
  ce qui est aussi différent du gain de l'individu dont le revenu potentiel sans traitement correspond à la médiane de cette distribution

- sinon, Heckman, Smith et Clements (1997) proposent des bornes, selon plusieurs hypothèses sur la manière dont les individus sont réordonnés
- Pas vraiment un problème en terme d'évaluation de politique publique : intéressant d'avoir des informations sur les différences entre les distributions avec et sans traitement (et pas forcément sur distributions des effets individuels) : suffisant par exemple pour avoir diagnostic sur la modification de la médiane, du bas de la distribution, du Gini...

Identification

- on peut utiliser les régressions de quantiles précédentes (en utilisant le traitement T comme une variable explicative particulière)
- mais on est évidemment confronté aux mêmes problèmes d'endogénéité.
- à part dans le cas de données expérimentales, effets de sélection $\rightarrow$ Extension des méthodes d'identification du cadre classique.
- littérature très récente...

identification : CIA

Conditionnement par des observables : Firpo (2007)
Hypothèses

1. indépendance conditionnelle (CIA) :

$$Y_0, Y_1 \perp T|X$$

   où X sont des caractéristiques observables

2. support commun : $p(X) = p(T = 1|X) \in]0, 1[$

équivalent hyp. Matching

Pb pour estimer l'effet du traitement au niveau du quantile :

- on sait estimer un effet du traitement *conditionnel* au niveau du τ^{ieme} quantile par régression de quantiles
- mais ici pas de loi des espérances itérées : moyenne des quantiles n'est pas le quantile de la moyenne.

$$q_Y(\tau) \neq E[q_Y(\tau|X)]$$

- ici, Firpo propose méthode(s) pour l'estimer directement de manière semi-paramétrique

Identification

Firpo montre que sous ces hypothèses le quantile $Q_{Y_1}(\tau)$ est identifié par (Y, T, X) :

$$\tau = E\left[\frac{T1(Y \leq Q_{Y_1}(\tau))}{p(X)}\right]$$

Par définition, $\tau = P(Y_1 \leq Q_{Y_1}(\tau))$: application loi espérance itérées + hypothèses indépendances...
Résultat important : dans la dernière expression, la seule fonction *conditionnelle* à estimer est le score $p(X) = P(T = 1|X)$ (et plus distribution de y_1).

quantiles conditionnels

De même, les quantiles conditionnels $Q_{Y_1}(\tau|T=1)$ et $Q_{Y_1}(\tau|T=1)$ sont identifiés par :

$$\tau = E\left[\frac{T1(Y \leq Q_{Y_1}(\tau|T=1))}{p}\right]$$

et

$$\tau = E\left[\frac{p(X)}{1-p(X)}\frac{(1-T)1(Y \leq Q_{Y_0}(\tau|T=1))}{p}\right]$$

avec $p = P(T=1)$

estimation

QTET : $\hat{\Delta}(\tau|T=1) = \hat{Q}_{Y_1}(\tau|T=1) - \hat{Q}_{Y_0}(\tau|T=1)$
Procédure en deux étapes :

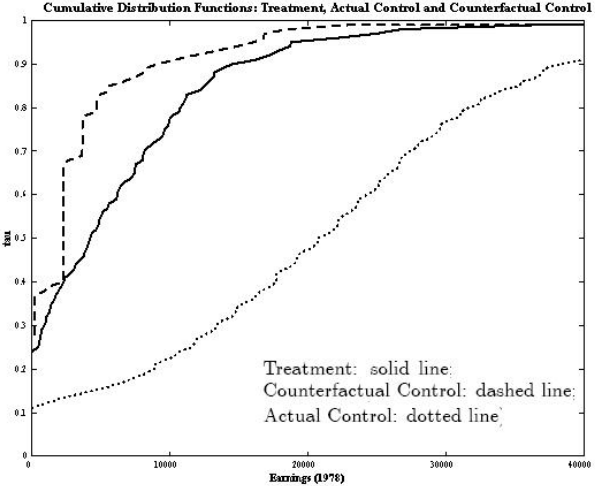
1. estimation (non paramétrique) de $p(X)$
2. des expressions ci-dessus, on a des estimateurs cv de $\hat{Q}_{Y_t}(\tau)$ $(t = 0, 1)$ par : $argmin_b \sum \hat{\omega}_{t,i|T=1} \rho_\tau(Y_i - b)$ avec $\hat{\omega}_{1,i|T=1} = \frac{T_i}{\sum_l T_l}$(i.e. analogue de T/p dans l'expression ci-dessus) et $\hat{\omega}_{0,i|T=1} = \frac{\hat{p}(X_i)}{1-\hat{p}(X_i)} \frac{(1-T_i)}{\sum_l T_l}$.

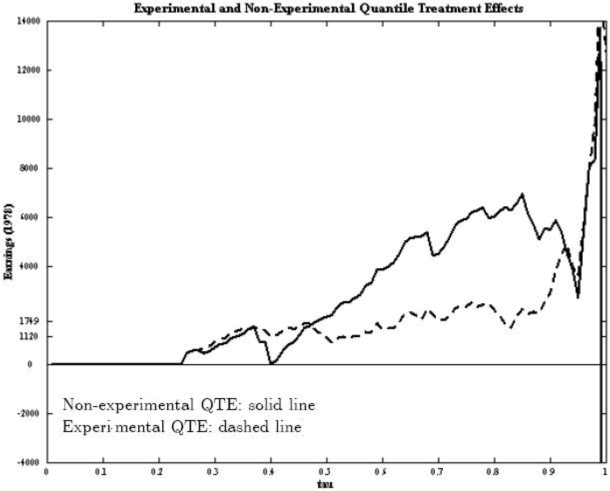
Par rapport à la régression de quantiles classique, les poids sont modifiés pour tenir compte des effets de sélection.

Application

- Application aux données de Lalonde (1986) : comparaison de données expérimentales (aide retour à l'emploi NSW) et données administratives.
- Dehejia et Whaba (1999) : application de procédures de matching pour estimer l'effet moyen.
- estimation du score qui "équilibre" les covariables entre le groupe traité et non traité
- effet moyen correspond exactement à $\sum_i (\hat{\omega}_{1,i|T=1} - \hat{\omega}_{0,i|T=1}) Y_i$.

Résultats


Cumulative Distribution Functions: Treatment, Actual Control and Counterfactual Control
tau
Earnings (1978)
0
10000
20000
30000
40000
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Treatment: solid line;
Counterfactual Control: dashed line;
Actual Control: dotted line


Experimental and Non-Experimental Quantile Treatment Effects
Earnings (1978)
14000
12000
10000
8000
6000
4000
1749
1120
0
-2000
-4000
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
tau
Non-experimental QTE: solid line
Experimental QTE: dashed line

Remarque

- effet moyen du traitement avec données expérimentales : 1794$
- effet moyen du traitement avec données non expérimentales, Dehejia et Wabba : 1129$
- effet médian du traitement avec données non expérimentales, Firpo : 1927$

Variable instrumentale

Abadie, Angrist et Imbens (2002) : extension cadre AIR aux régressions de quantile

- estimation des QTE avec un instrument
- application à une expérimentation

notation

- Affectation au traitement (formation) : $Z = 0, 1$
- traitement (formation) $T = 0, 1$, dépend instrument T_z (soit T_0 et T_1)
- revenu Y, dépend traitement Y_t (soit Y_0 et Y_1)
- caractéristiques observables X

hypothèses

1. indépendance : (Y_1, Y_0, T_1, T_0) indep de Z cond. X
2. assignation non triviale : $P(Z = 1|X) \in]0, 1[$
3. première étape $E[T_1|X] \neq E[T_0|X]$
4. monotonicité : $P(T_1 \geq T_0|X) = 1$ (pas de defiers)

Par analogie avec AIR, on appelle compliers les personnes dont le comportement est affecté par le traitement : $T_1 > T_0$
Conséquence : indépendance des revenus au traitement pour les compliers :

$$(Y_1, Y_0) \perp T|X, T_1 > T_0$$

estimation du QTEc

Paramètre d'intérêt :

$$Q_\tau(Y|X, T, T_1 > T_0) = \delta_\tau T + X\beta_\tau$$

i.e. estimation du QTE *pour les compliers*
estimation de

$$(\hat{\delta}_\tau, \hat{\beta}_\tau) = argminE(\rho_\tau(Y - \delta T - X\beta)|T_1 > T_0)$$

Pb : population des compliers inconnue

- pour revenir à quantité connue, AAD montre qu'on peut utiliser la fonction de poids :

$$\kappa(T, Z, X) = 1 - \frac{T(1-Z)}{(1-\pi_0(X))} - \frac{(1-T)Z}{\pi_0(X)}$$

avec $\pi_0(X) = P(Z = 1|X)$.

- κ vaut 1 si $T = Z$ (correspond aux compliers)
- montrent que pour toute fonction $h(Y, T, X)$:

$$E[h(Y, T, X)|T_1 > T_0] = \frac{1}{P(T_1 > T_0)} E[\kappa h(Y, T, X)]$$

- on peut alors estimer de

$$(\hat{\delta}_\tau, \hat{\beta}_\tau) = argminE[\kappa\rho_\tau(Y - \delta T - X\beta)]$$

- en pratique, utilisent
$\kappa_\nu = E[\nu|Y,T,X] = P(T_1 > T_0|Y,T,X)$

Application : JTPA

- données expérimentales
- Job Training Partnership Act (JTPA) : programme de formation pour les personnes rencontrant des difficultés d'insertion
- assignation aléatoire (20000 personnes), mais seulement 60% des personnes affectées en ont profité.
- à l'inverse, peu de contamination : seulement 2% des personnes du groupe de contrôle ont suivi le programme

Application

- variable d'intérêt : revenu après 30 mois
- affectation au traitement fournit un instrument valide
- contrôlent aussi des caractéristiques individuelles : noir ou hispanique, éducation, age, emploi récemment, marié ou pas... : utilité ?

Résultats

		Quantile				
	2SLS	0.15	0.25	0.50	0.75	0.85
A. Men						
Training	1,593	121	702	1,544	3,131	3,378
	(895)	(475)	(670)	(1,073)	(1,376)	(1,811)
% Impact of Training	8.55	5.19	12.0	9.64	10.7	9.02
B. Women						
Training	1,780	324	680	1,742	1,984	1,900
	(532)	(175)	(282)	(645)	(945)	(997)
% Impact of Training	14.6	35.5	23.1	18.4	10.1	7.39

Remarque :

- Froelich et Melly (2008) proposent un estimateur de quantile non conditionnel
- intéressant quand on a un instrument valide uniquement si on conditionne par des observables (ce qui n'est pas le cas ici).
- voir aussi Chernozhukov et Hansen (2001)

Papiers très récents qui étendent techniques classiques au cas des quantiles :

- Lamarche (2007) : introduction d'effets fixes pour traiter des biais d'endogénéité, et application à l'évaluation de l'expérience des vouchers (Milwaukee).
- Froelich et Melly (2008) étendent aux régressions sur discontinuités
- Athey et Imbens (2003) : application aux différences de différences.
- application aux modèles de durée : voir Koenker et Bilias (2002) : réévaluation du "Bonus Experiment"