

The econometrics of inequality and poverty
*Lecture 10: Explaining poverty and inequality using
econometric models*

Michel Lubrano

February 2010

Contents

1	Introduction	2
2	Decomposing poverty and inequality	2
2.1	FGT indices	2
2.2	Oaxaca decomposition	3
2.3	Explaining the income-to-needs ratio	5
3	Dynamic models of income poverty	9
4	A first introduction to quantile regression	10
5	Quantile regressions in full	12
5.1	Introduction	12
5.2	Applications	13
6	Statistical inference	16
6.1	Inférence Bayésienne	17
6.2	Non-parametric inference	17
6.2.1	L'estimation nonparamétrique des quantiles	18
6.2.2	Régression quantile non-paramétrique	18
6.2.3	Les apports de Yu et Jones (1998)	19

1 Introduction

Nous nous sommes intéressés jusqu'à présent à la description de la distribution des revenus. On a vu comment comparer deux distributions, soit dans deux pays différents, soit entre deux points du temps pour un même pays. Mais tout cela est resté très descriptif et l'on n'a pas essayé de rendre compte de la formation de la distribution des revenus ou de l'explication du phénomène de pauvreté. Et en cela nous avons été fidèles à cette dichotomie qui existe dans la littérature entre la mesure des inégalités et de la pauvreté et la théorie de la formation des revenus. Le revenu des ménages se décompose en plusieurs parties à savoir les salaires (la plus grosse part), les revenus locatifs ou financiers, et les impôts et les transferts. Les économistes du travail se sont penchés dans les années 1980 sur l'explication de la dispersion des salaires individuels, mais n'ont jamais relié cette recherche à la mesure des inégalités de revenus des ménages. Le but de cette section n'est pas de combler cette lacune, on regardera pour cela par exemple Atkinson (2003). Mais on va par contre s'attacher à présenter certains des outils économétriques qui permettent d'expliquer la formation et l'évolution de la distribution des revenus et des phénomènes de pauvreté.

2 Decomposing poverty and inequality

The idea is to split the inequality or the poverty measured by an index into different and mutually exclusive groups. Which group in the population is more subject to poverty? This principle can be extended to the decomposition of inequality, most of the time wage inequality, between two groups. For instance is wage differential between male and females or black and white due to intrinsic differences or to a mere discrimination?

2.1 FGT indices

The index of Foster, Greer, and Thorbecke (1984) is decomposable because of its linear structure. Let us consider the decomposition of a population between rural and urban. If X represents all income of the population, the partition of X is defined as $X = X^U + X^R$. Let us call p the proportion of X^U in X . Then the total index can be decomposed into

$$\begin{aligned} P_\alpha &= p \frac{1}{n} \sum_{i=1}^{n_U} \left(\frac{z - x_i^U}{z} \right)^\alpha \mathbf{1}(x_i \leq z) + (1-p) \frac{1}{n} \sum_{i=1}^{n_R} \left(\frac{z - x_i^R}{z} \right)^\alpha \mathbf{1}(x_i \leq z) \\ &= p P_\alpha^U + (1-p) P_\alpha^R. \end{aligned} \quad (1)$$

where P_α^U is the index computed for the urban population and P_α^R the index computed for the rural population.

We illustrate this decomposability using the FES data for 1996. We have defined a poverty line as 50% of the mean income for the total sample. We can divide this sample into mutually exclusive groups, depending on the status of the head of the household. Poverty is concentrated

Table 1: Decomposing poverty in the 1996 UK

	Retired	Working	Unemployed	Others	Total
n	1806	2355	949	933	6043
%	0.299	0.390	0.157	0.154	1.000
P_0	4.23	0.46	16.13	2.50	4.36
$P_0 n_i / n$	1.27	0.18	2.53	0.39	4.36

among the unemployed followed by the retired group.

Other indices, in particular some inequality indices are decomposable. In this class we find the Atkinson index and the family of Generalised Entropy indices. *This would be a good exercise to show that these indices are decomposable. Give the expression of the two indices. Detail and explain the decomposability of the Gini using the results of Jacques Silber.*

2.2 Oaxaca decomposition

In the previous section, we have decomposed a poverty rate according to mutually exclusive groups of the population. But, we provided no explanation on the reason of this decomposition, what made a person belong to one of these groups. Oaxaca (1973) was the first to try to give an explanation on the sources, the causes of inequality, using a regression model.

Oaxaca (1973) took interest in wage inequality between males and females. Suppose that we have divided our sample in two groups, one group of males, one group of females. For each group we estimate a wage equation which relates the log of the wage to a number of characteristics, among which we find experience and years of schooling. Other variables can include regional location and city size for instance.

$$\log(W_i) = X_i \beta_i + u_i, \quad i = m, f.$$

Once these two equations are estimated, we have a $\hat{\beta}_m$ for males and a $\hat{\beta}_f$ for females. We are going to try to explain wages differences between males and females as follows. We can say that a part of this difference can be explained by different characteristics. For instance if males have more experience. These objective differences are measured by $X_h - X_f$. But another part of the wage differences can be explained simply by the different yield of these characteristics: for an identical experience, a female is paid less than a male. These differences in yields are at the root of the discrimination existing between males and females on the labour market. In a regression model, the mean of the endogenous variable is given by

$$\log(\bar{W}_i) = \bar{X}_i \hat{\beta}_i,$$

because of the zero mean assumption on the residuals. Using this property, Oaxaca proposed the following decomposition:

$$\log(\bar{W}_m) - \log(\bar{W}_f) = (\bar{X}_m - \bar{X}_f)\hat{\beta}_m + \bar{X}_f(\hat{\beta}_m - \hat{\beta}_f).$$

In this decomposition, the difference in percentage between the average male and female wages is explained first by the difference in average characteristics. As a second term comes the difference in yield of female average characteristics expressed by $\hat{\beta}_m - \hat{\beta}_f$.

This decomposition is very popular in the literature. The original paper is cited more than 3171 times (using GoogleScholar). It gave birth to many subsequent developments. For instance, Juhn, Murphy, and Pierce (1993) generalised the previous result to the framework of quantile regression. Radchenko and Yun (2003) provide a Bayesian that make easier significance tests.

The difference in means can also be expressed as

$$R = (\bar{X}_m - \bar{X}_f)\beta_* + [\bar{X}_m(\hat{\beta}_m - \beta_*) + \bar{X}_f(\beta_* - \beta_f)],$$

the first part being the explained part, while the term in squared brackets being the unexplained part. We recover the previous decomposition for $\beta_* = \hat{\beta}_m$ while the Blinder decomposition is found for $\beta_* = \hat{\beta}_f$. Other decomposition found in the literature choose β_* as the average between the two regression coefficients. Jann (2008) proposes to compute standard errors for this decomposition. The question is to know if the regressors are stochastic or not. If the regressors are fixed, then

$$\text{Var}(\bar{X}\hat{\beta}) = \bar{X}'\text{Var}(\hat{\beta})\bar{X}$$

If the regressors are stochastic, but however uncorrelated, Jann shows that this variance becomes

$$\text{Var}(\bar{X}\hat{\beta}) = \bar{X}'\text{Var}(\hat{\beta})\bar{X} + \hat{\beta}'\text{Var}(\bar{X})\hat{\beta} + \text{tr}(\text{Var}(\bar{X})\text{Var}(\hat{\beta}))$$

From these expression, he derives the variance of the Oaxaca decomposition. This is simple, but tedious algebra. A command exists in STATA, but apparently not in R. Jann illustrates his method for decomposing the gender wage gap on the Swiss labour market using the Swiss Labour Force Survey 2000 (SLFS; Swiss Federal Statistical Office). The sample includes Employees aged 20-62, working fulltime, having only one job. The dependent variable is the Log of hourly wages. The explanatory variables are the number of years of schooling, the number of years of experience, its square divided by 100, two dummy variables concerning Tenure and the gender of the supervisor. There are 3383 males and 1544 females. Results are as follows From the estimates reported in Table 2, we can compute the original Oaxaca decomposition The bootstrap and the stochastic regressor assumption give very comparable standard deviations. Assuming fixed regressors under evaluate the standard deviations. Wage differentials is more explained by discrimination than by differences in characteristics. These differences are significant. There are both differences in characteristics and discrimination.

Further developments: Bourguignon, Ferreira, and Leite (2008) *for next year*.

Table 2: Wage equations for Switzerland 2000

Log wages	Men		Women	
	Coef.	Mean	Coef.	Mean
Education	0.0754 (0.0023)	12.0239 (0.0414)	0.0762 (0.0044)	11.6156 (0.0548)
Experience	0.0221 (0.0017)	19.1641 (0.2063)	0.0247 (0.0031)	14.0429 (0.2616)
Exp ²	-0.0319 (0.0036)	5.1125 (0.0932)	-0.0435 (0.0079)	3.0283 (0.1017)
Tenure	0.0028 (0.0007)	10.3077 (0.1656)	0.0063 (0.0014)	7.6729 (0.2013)
Supervisor	0.1502 (0.0113)	0.5341 (0.0086)	0.0709 (0.0193)	0.3737 (0.0123)
Constant	2.4489 (0.0332)		2.3079 (0.0564)	
R^2	0.3470		0.2519	

Table 3: Oaxaca decomposition for Switzerland 2000

	Value	Bootstrap	Stochastic	Fixed
Differential	0.2422	0.0122	0.0126	0.0107
Explained	0.1091	0.0076	0.0075	0.0031
Unexplained	0.1331	0.0113	0.0112	0.0111

2.3 Explaining the income-to-needs ratio

Let us consider a poverty line z and the income y_i of an household. The ratio y/z is known to be the **income-to-needs ratio** in the literature. It can be used to explain the probability that this household has of getting in a state of poverty. $\log(y_i/z)$ is negative if the household is poor, positive otherwise. We can then estimate a regression

$$\log(y_i/z) = x_i'\beta + u_i$$

where x_i is a set of characteristics of the household. If we suppose that u_i is normal, we can compute the probability that an household is poor by mean of

$$P_0 = \Pr(x_i \hat{\beta} < 0) = \Phi(-x_i \beta / \sigma)$$

where σ^2 is the variance of the residuals and $\Phi(\cdot)$ the normal cumulative distribution. When n tends to infinity, the estimated variance tends to zero so that this probability approaches the head count measure.

We can now extend the approach of Oaxaca to explain the difference that there exist of being poor between two groups: white and black households in the US or between Serbs and Albanian households in Kosovo. Yun (2004) propose a generalisation of Oaxaca decomposition for non-linear models and in particular for probit models. Let us call A and B the two groups we consider. The decomposition proposed by Yun (2004) is as follows:

$$P_A^0 - P_B^0 = [\overline{\Phi(-X_A \beta_A / \sigma_A)} - \overline{\Phi(-X_B \beta_A / \sigma_A)}] + [\overline{\Phi(-X_B \beta_A / \sigma_A)} - \overline{\Phi(-X_B \beta_B / \sigma_B)}],$$

ce qui correspond à la différence entre les caractéristiques et la différence entre les coefficients. L'application de la procédure de Yun (2004) permet de pondérer cette décomposition en fonction du poids de chacune des k caractéristiques individuelles

$$P_A^0 - P_B^0 = \sum_{i=1}^k W_{\Delta X}^i [\overline{\Phi(-X_A \beta_A / \sigma_A)} - \overline{\Phi(-X_B \beta_A / \sigma_A)}] + \sum_{i=1}^k W_{\Delta \beta}^i [\overline{\Phi(-X_B \beta_A / \sigma_A)} - \overline{\Phi(-X_B \beta_B / \sigma_B)}]$$

où les poids $W_{\Delta X}^i$ et $W_{\Delta \beta}^i$ sont donnés dans Bhaumik, Gang, and Yun (2006a) sur un argument de linéarisation de Yun (2004).

Bhaumik, Gang, and Yun (2006b) use the 2001 Living Standards Measurement Survey (LSMS) data for Kosovo to decompose the difference in the average likelihood of poverty incidence between Serb and Albanian households. The survey, which was carried out between September and December of 2000, collected data from 2,880 households. After accounting for missing values, the survey provides information on 2101 Kosovo Albanian households and 416 Kosovo Serbian households. The ratio $R = y_i/z$ is computed using the world bank poverty line for Kosovo. The differences in the average probability of being poor between groups A and B , $(\bar{P}_A - \bar{P}_B)$, can be algebraically decomposed into two components which represent the characteristics and coefficients effects. The predicted poverty rate for Serbs is 55.98% while it is only of 45.41% for Albanian. There is a gap of 10.56%. How can we explain this gap? Bhaumik, Gang, and Yun (2006b) provide in their Table 2 (reproduced here) an estimation for the two equations. In their Table 3 (reproduced here), they analyse the differences in poverty between the two communities.

Table 2
Determinants of Ratio of Per Capita Expenditure to Poverty Line (ML estimation)

	Albanians		Serbs	
	Estimate	S.E.	Estimate	S.E.
Constant	- 0.33***	(0.09)	- 1.10***	(0.21)
<i>Demographic characteristics of households</i>				
Proportion aged 15 or below	- 0.58***	(0.06)	- 0.17	(0.12)
Proportion aged above 65	- 0.10	(0.11)	- 0.06	(0.13)
Proportion of adults who are male	0.04	(0.09)	0.23	(0.16)
Households with male head	- 0.06	(0.05)	0.06	(0.09)
<i>Education</i>				
Proportion of adults with primary education	0.18**	(0.08)	0.31	(0.19)
Proportion of adults with secondary education	0.58***	(0.08)	0.92***	(0.20)
Proportion of adults with vocational training	0.52***	(0.10)	0.91***	(0.23)
Proportion of adults with tertiary education	0.75***	(0.10)	1.46***	(0.21)
<i>Labor market characteristics</i>				
Proportion of working adults	0.45***	(0.06)	0.22**	(0.11)
Proportion of households with members working in family farms & businesses	- 0.00	(0.07)	- 0.04	(0.11)
<i>Wealth/Assets</i>				
Acreage of land household owns (000)	0.17	(0.15)	0.01	(0.01)
Value of animals household owns (000 DM)	0.03	(0.02)	0.04	(0.03)
<i>Transfers</i>				
Households at least one of whose members has a disability card	0.02	(0.04)	- 0.10	(0.07)
Household at least one of whose members receive private transfers	0.09***	(0.02)	0.33***	(0.11)
<i>Geographic Characteristics</i>				
Urban households	0.05	(0.03)	0.06	(0.06)
Standard deviation of error term (σ)	0.46***	(0.01)	0.46***	(0.03)
Log-likelihood (L)	-150785.98		-19300.24	
Constrained Log-likelihood ($L0$)	-180607.61		-25300.89	
Number of households	2101		416	

Note: *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. Weights are used in estimation. Standard errors which are robust to mis-specification are reported. Constrained log-likelihood is calculated only when constant and standard deviation of error term are estimated. Likelihood ratio test, $2*(L - L0)$, rejects the null hypothesis that coefficients except for the constant are zero for both Serbs and Albanians.

Table 3
Decomposing Difference in Poverty Rates of 10.56% between Serbs and Albanians using
Estimates of Per Capita Expenditure Regression Equations

	Characteristics Effect		Coefficients Effect	
	Estimate	Share	Estimate	Share
<i>Aggregate Effect</i>	- 0.035	- 33.55	0.141***	133.55
<i>Aggregate Effect Without Constants</i>	- 0.035	- 33.55	- 0.429**	- 405.66
Constant			0.570***	539.21
<i>Demographic characteristics of households</i>	- 0.016	- 15.04	- 0.244**	- 231.24
Proportion aged 15 or below	- 0.021*	- 19.52	- 0.095***	- 90.19
Proportion aged above 65	0.003	2.90	- 0.001	- 0.97
Proportion of adults who are male	- 0.001	- 1.33	- 0.068	- 64.11
Proportion with male head	0.003	2.92	- 0.080	- 75.96
<i>Education</i>	- 0.113***	- 106.66	- 0.191	- 180.66
Proportion of adults with primary education	0.034*	32.64	- 0.044	- 41.81
Proportion of adults with secondary education	- 0.165***	- 155.94	- 0.076*	- 72.17
Proportion of adults with vocational training	0.006***	5.33	- 0.024	- 22.42
Proportion of adults with tertiary education	0.012***	11.32	- 0.047***	- 44.25
<i>Labor market characteristics</i>	- 0.008*	- 7.64	0.074**	70.41
Proportion of working adults	- 0.010*	- 9.67	0.067*	63.30
Proportion of households with members working in family farms & businesses	0.002	2.02	0.008	7.11
<i>Wealth/Assets</i>	0.003	2.46	0.003	2.88
Acreage of land household owns (000)	- 0.000	- 0.36	0.008	7.39
Value of animals household owns (000 DM)	0.003	2.82	- 0.005	- 4.51
<i>Transfers</i>	0.106**	100.00	- 0.068*	- 64.28
Proportion of households at least one of whose members has a disability card	- 0.000	- 0.03	0.009	8.65
Proportion of household at least one of whose members receive private transfers	0.106**	100.03	- 0.077**	- 72.92
<i>Geographic Characteristics</i>				
Urban households	- 0.007	- 6.67	- 0.003	- 2.77

Note: Share is the ratio of the contribution of each factor to the “predicted” overall difference in poverty rate (10.56%) between Serbs (55.98%) and Albanians (45.41%), in percentage terms. The observed overall difference in poverty rate are 11.87% between Serbs (57.38%) and Albanians (45.52%). The predicted poverty rate is computed using estimates from the per capita expenditure regression. The details of the computation using the per capita expenditure regression is explained in the main text. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

The overall characteristics effect is -0.035. This means that of the 10.56 percentage point gap in poverty rate, -3.54 percentage points are due to the characteristics effect, or $-3.54/10.56 = -33.55\%$ of the gap in poverty incidence is due to characteristics differences.

The overall coefficients effect (or discrimination effect) is 0.141. Of the 10.56 percentage

point gap, 14.11 percentage points or $14.11/10.56 = 133.55\%$ of the gap in poverty incidence.

In other words, Serbs would be worse off if the differences between their characteristics and those of the Albanian households disappear, and Serbs would be better off if there is no difference in the poverty mitigating effectiveness of those characteristics between the Serbian and Albanian households. When we look at detailed decomposition, it becomes clear that the main reason why Serbs have higher poverty incidence is due to coefficients effect of constant term. Even though Serbs have better characteristics which can lower poverty incidence, and enjoy stronger poverty mitigating effect of these characteristics relative to Albanians, there is huge baseline gap in poverty incidence between the two ethnic groups, captured by the coefficients effect of the constant term.

3 Dynamic models of income poverty

In his presidential address to the European Society for Population Economics, Jenkins (2000) underlines that the income distribution in the UK has experienced great changes during the eighties, but that since 1991, this distribution seems to have remained relatively stable. If the poverty line is defined as half the mean income, the percentage of poor remains relatively stable, while if it is defined as half the mean of 1991 in real term, this percentage decreases steadily. The Gini coefficient remains extremely stable around 0.31-0.32. These figures characterise a cross-section stability in income.

However, since 1991, the UK started the British Household Panel Survey. This means that the same household are interviewed between 1991 and 1996 each year. It then become possible to study income dynamics. Jenkins provide an estimation for a transition matrix between income groups at a distance of one year. These groups are defined by reference to a fraction of the mean, fraction taken between 0.5 and 1.5 In line, we have groups for wave t , and in column groups for

Table 4: Transition probabilities in percentage

Income group	Period t					
	< 0.5	0.5-0.75	0.75-1.0	1.0-1.25	1.25-1.5	> 1.5
Period $t - 1$						
< 0.5	54	30	9	4	2	2
0.5-0.75	15	56	21	5	1	2
0.75-1.0	5	19	48	20	5	3
1.0-1.25	3	6	20	44	20	7
1.25-1.5	2	3	8	25	35	27
> 1.5	1	2	4	6	12	75

wave $t - 1$. If we except the very rich who have a probability of 0.75 to remain rich, the other groups have in general a probability less than 0.50 to stay in their original group and a probability of going to the neighbouring group of 0.20 on average. Consequently, there was a large income mobility in dynamics. The percentage of poor remained the same, but the persons in a state of poverty were not the same along the 6 years of the panel.

Household do not stay all the time in poverty. They have poverty spells, they enter into poverty and get out of it. Stevens (1999) got interest in explaining the duration of these poverty spells for the USA. In her paper, she propose several models. We keep only one which explains again the logarithm of the income-to-needs ratio as a function exogenous variables but also of dynamic errors. The model is then used to make judgement about the persistance of poverty spells in the USA in order to evaluate the economic situation of an household. Il est calculé en prenant le revenu du ménage sans tenir compte des diverses allocations et en le divisant par le seuil de pauvreté officiel correspondant à la composition de ce ménage. Le modèle de base est le suivant:

$$\begin{aligned} \log\left(\frac{y_{it}}{z}\right) &= x_{it}\beta + \delta_i + v_{it} \\ \delta_i &\sim N(0, \sigma_\delta^2) \\ v_{it} &= \gamma v_{it-1} + \eta_{it} \end{aligned}$$

The log of the income to needs ratio is explained by individual variables that are time independent as sex and education level, and by individual variables that are time varying. There is a random individual effect δ_i for unobserved heterogeneity. Le paramètre γ modélise une permanence commune à tous les individus. On peut dire que les individus reçoivent des shocks v_{it} qui perdurent. Sous une hypothèse de normalité pour δ_i et η_{it} , Stevens (1999) simule ce modèle pour 20 années et calcule le temps moyen passé sous la ligne de pauvreté. Estimé sur les données du PSID, ce modèle indique que ce temps moyen est en général très légèrement supérieur si le chef de ménage est noir ou si c'est une femme.

4 A first introduction to quantile regression

Un modèle de régression donne le lien linéaire ou non-linéaire qui existe entre une variable endogène et une ou plusieurs variables explicatives. Dans le cas le plus simple, la ligne de régression représente une espérance conditionnelle linéaire. Une régression non-paramétrique va expliquer une espérance conditionnelle de façon non-linéaire. Mais aucune de ces régressions ne se préoccupe d'expliquer tous les quantiles de la distribution de la variable endogène. La régression quantile, introduite en économétrie par Koenker and Basset (1978) fournit l'instrument recherché pour expliquer l'évolution complète de la variable endogène. Le principe est simple, mais sa réalisation numérique est plus complexe, en particulier pour calculer des écart-types.

Considérons un modèle linéaire de régression $y = x'\beta + e$ à propos duquel on ne fait pour le moment aucune hypothèse sur la distribution conditionnelle de y . On va noter $q_\alpha(x)$ le quantile

de niveau α qui se définit par

$$q_\alpha(x) = F^{-1}(\alpha)$$

On veut expliquer ce quantile par une combinaison linéaire des x

$$q_\alpha(x) = x'\beta$$

Il faut tout d'abord noter que si F est la distribution normale, nous n'apprenons pas grand chose par ce modèle, car premièrement pour cette densité la moyenne et la médiane sont confondues et qu'ensuite les quantiles conditionnels sont des droites. Il faut s'écarter de ce cadre traditionnel pour que le modèle prenne tout son sens. Une première possibilité consiste à avoir des erreurs hétéroscédastiques, c'est à dire que si les erreurs sont normales de variance σ^2 , ce paramètre va être fonction des variables exogènes. Une hypothèse plus extrême consiste à ne faire aucune hypothèse paramétrique sur la distribution F et à travailler dans un cadre semi-paramétrique. Pour cela on va définir une fonction d'erreurs

$$\rho_\alpha(u) = \begin{cases} u\alpha & \text{si } u > 0 \\ u(\alpha - 1) & \text{sinon.} \end{cases}$$

On va chercher ensuite à minimiser en β la quantité $\sum \rho_\alpha(y_i - x'_i\beta)$, ce qui revient à résoudre un problème de programmation quadratique. En utilisant un cadre Bayésien, Yu and Moyeed (2001) montrent que l'estimation du quantile de y peut être rendue équivalente à l'estimation du paramètre de localisation d'une distribution asymétrique de Laplace. Ceci permet d'écrire une fonction de vraisemblance avec

$$L(\beta|y, x) \propto \alpha^n (1 - \alpha)^n \exp\left\{-\sum_i \rho_\alpha(y_i - x'_i\beta)\right\}$$

pour évaluer ensuite la densité a posteriori des paramètres. Il devient alors facile d'estimer des écart-types, ce qui n'est par contre pas du tout évident dans le cadre initial proposé par Koenker and Bassett (1978).

La régression quantile a été très souvent employée pour rendre compte de l'évolution des inégalités salariales à partir d'une équation Mincerienne de salaire, ce qui permet de mesurer les différences de rendements de l'éducation et de l'expérience. Voir par exemple Machado and Mata (2005) pour une étude récente. La régression quantile a été utilisée pour rendre compte des phénomènes de pauvreté dans certains pays du tiers monde en expliquant l'écart de revenu entre les villes et les campagnes. Ainsi Nguyen, Albrecht, Vroman, and Westbrook (2007) montrent que pour le Vietnam l'écart ville-campagne dans la distribution des revenus est dû à des différences d'éducation pour les quantiles les plus faibles, mais que pour le reste de la distribution, cet écart est principalement dû à des différences dans le rendement des attributs. On peut déduire de ces résultats qu'une politique de réduction de la pauvreté dans les campagnes doit se centrer sur l'éducation.

The illustrations have to be developed.

5 Quantile regressions in full

5.1 Introduction

Considérons un échantillon d'une variable aléatoire Y et sa densité $f(y)$. On va définir la moyenne comme

$$\hat{\mu} = \int yf(y)dy$$

Si $F(\cdot)$ est la distribution de Y , alors la médiane sera

$$q_{0.50}(y) = F_y^{-1}(0.50)$$

On peut définir de la même manière les autres quantiles.

Considérons maintenant un échantillon bivarié de deux variables aléatoires Y et X distribués conjointement selon $f(y, x)$. Si $f(y|x)$ est la distribution conditionnelle de y si x , alors l'espérance conditionnelle $E(y|x)$ se définit comme

$$E(y|x) = \int yf(y|x)dy$$

qui va prendre autant de valeurs différentes que x . Il s'agit donc d'une fonction. Si $F(\cdot)$ est la distribution Normale de moyenne (μ_y, μ_x) et de variance $\sigma_y^2, \sigma_{yx}, \sigma_x^2$, alors la fonction de régression se note simplement par propriétés de Normale bivariée

$$E(y|x) = \mu_y + \frac{\sigma_{yx}}{\sigma_x^2}(x - \mu_x).$$

On exprime donc l'espérance conditionnelle comme une fonction linéaire de x . Si l'on s'intéresse maintenant aux quantiles conditionnels dans cette même normale

$$q_p(x) = \mu_y + \frac{\sigma_{yx}}{\sigma_x^2}(x - \mu_x) + \Phi^{-1}(p)\sqrt{\sigma_y^2 - \frac{\sigma_{x,y}^2}{\sigma_x^2}}.$$

Le cas de la Normale est très particulier car $q_{0.50}(x) = E(y|x)$ et les autres fonctions quantiles sont des droites parallèles étant donné que $\Phi^{-1}(0.50) = 0$. Le quantile conditionnel est la moyenne conditionnelle corrigée par la valeur du quantile de la normale standardisée multiplié par la racine carrée de la variance conditionnelle.

L'intérêt de la régression quantile introduite par Koenker and Basset (1978) c'est que dès que l'on sort du cadre normal, les fonctions quantiles ne sont plus des fonctions linéaires de X . On prendra comme exemple le modèle hétéroskédastique

$$\begin{aligned} Y_t &= 2 + X_t + \exp(-X_t)\epsilon_t \\ X &\sim N(0, 1) \\ \epsilon_t &\sim N(0, 1) \end{aligned}$$

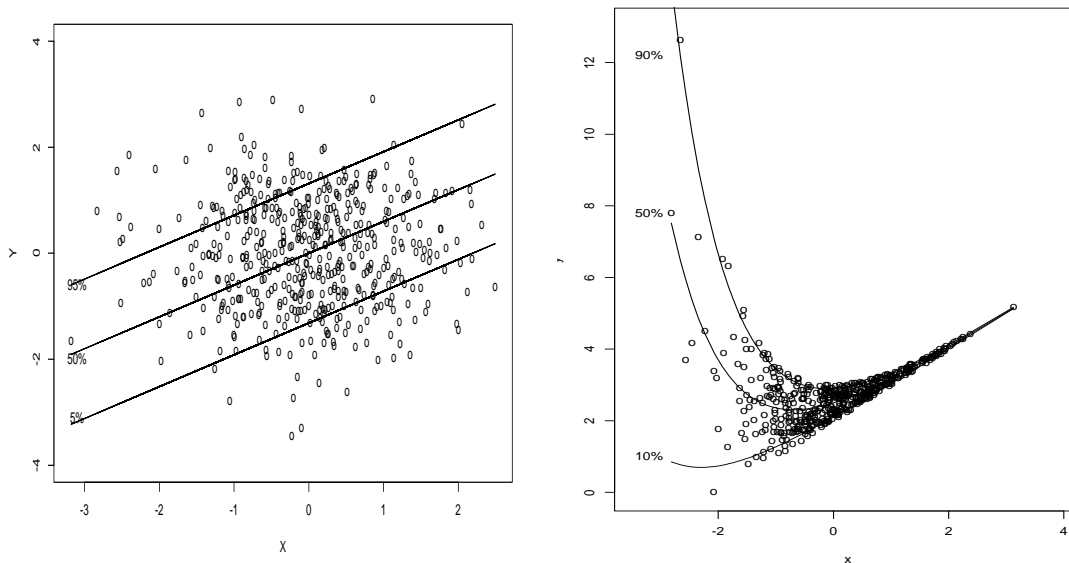


Figure 1: Quantile in a standard and in a heteroskedastic regression

que l'on a utilisé pour simuler un échantillon. Alors on peut comparer les deux types de régression dans le cas normal et dans le cas hétéroskédastique sur un échantillon simulé. On a des résultats analytiques dans des cas particuliers comme

$$Y = m(t) + m(t)\epsilon_t \quad \epsilon_t \sim N(0, 1).$$

Alors

$$q_p(t) = m(t) + m(t)\Phi^{-1}(p).$$

On peut remarquer que si le modèle n'était pas hétéroskédastique tous les quantiles seraient parallèles.

5.2 Applications

L'analyse de la dispersion des salaires en économie du travail et l'analyse de la distribution des revenus. La raison c'est que l'influence d'une variable, ne serait-ce que le temps, peut être très différente sur les groupes à faible ou fort salaire/revenus. Implémenter une politique fiscale ou sociale quand on veut cibler certains groupes.

Bailar (1991) a étudié l'évolution du salaire de 459 professeurs de statistique en prenant comme variable explicative le nombre d'années depuis laquelle ils avaient la tenure. On constate que les plus riches (quantile 0.75) sont devenu plus riches au cours du temps alors que les autres ont eu un revenu plus stationnaire.

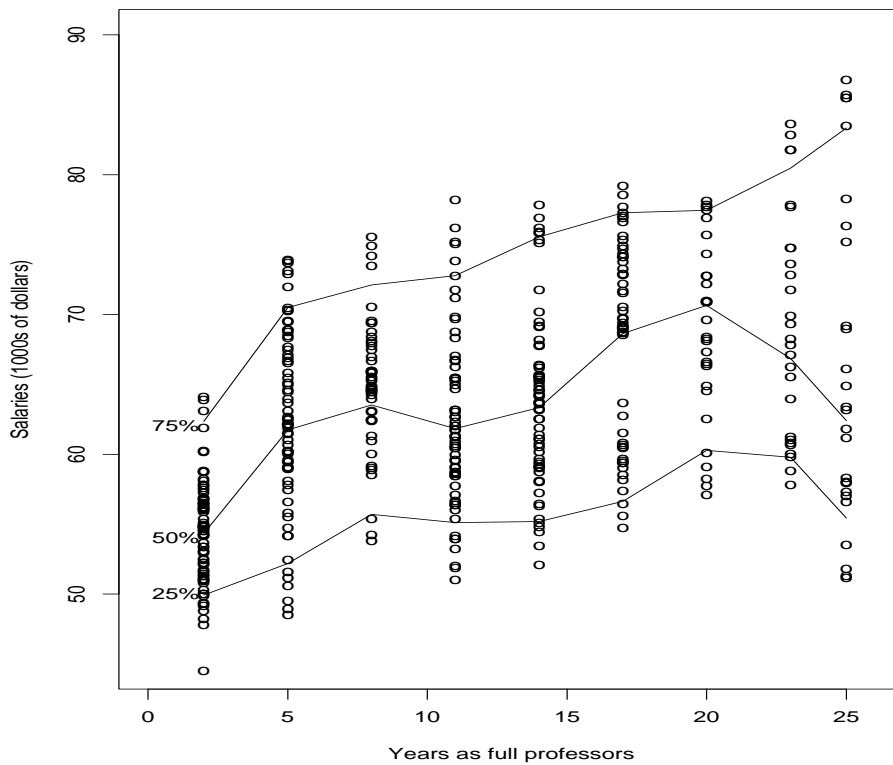


Figure 2: Wage distribution as a function of experience since tenure

Trede (1998) make use of a non-parametric quantile regression for explaining the income distribution of year t as a function of the income distribution of year $t - 1$. The two case study are Germany and the USA.

Trede has considered three sample period for each country. The samples concern household income in both countries. The reference year is 1984. It serves to normalise the other years, using the median. If the distribution does not change, all the quantiles will be identical to the 45° line. The distribution of the first period completely determines the distribution in the second period. If on the contrary, the distribution of the second period is independent of the distribution in the first period, the quantile will be horizontal. This is a sign of income mobility.

The first conclusion is that income mobility is more important in the long term than in the short term. This is a natural finding. What is more unexpected is that income mobility is greater in Germany than what it is in the USA.

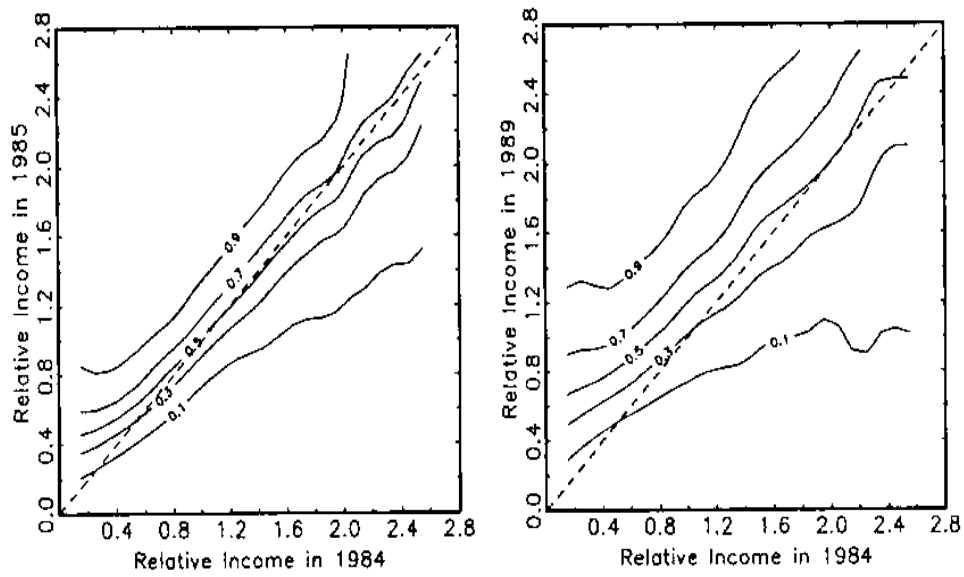


Fig. 1. Quantile regression of (relative) income in 1985 and 1989 on (relative) income in 1984, Germany.

Figure 3: Income mobility in Germany: 1984-1985 and 1984-1989

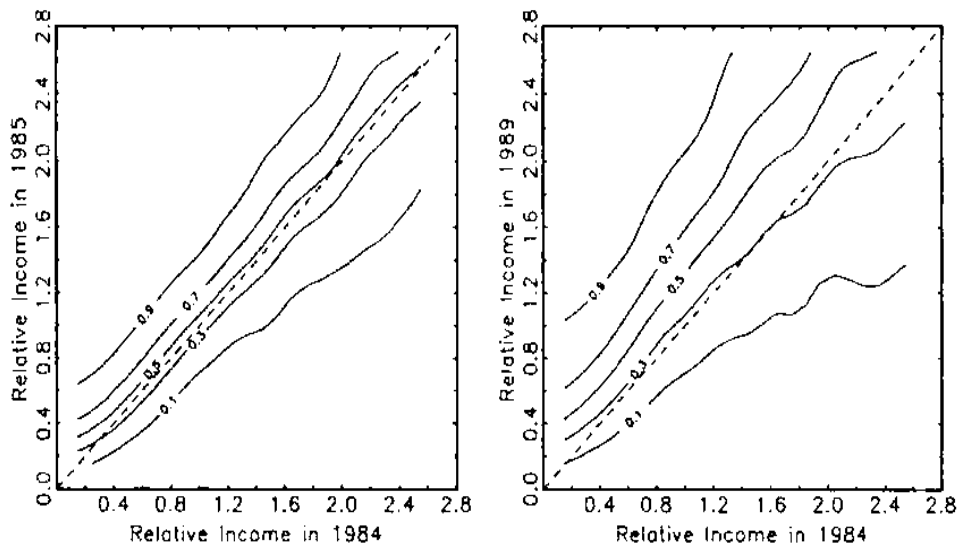


Fig. 2. Quantile regression of (relative) income in 1985 and 1989 on (relative) income in 1984, United States.

Figure 4: Income mobility in the USA: 1984-1985 and 1984-1989

6 Statistical inference

Dans le papier original de Koenker and Bassett (1978), l'expression des quantiles d'une distribution est tirée d'un exercice de Fergusson (1967) qui demande de préciser les paramètres d'une fonction de perte en valeur absolue dont la perte espérée associée est minimum pour le quantile. C'est un exercice classique dans la littérature Bayésienne.

Il est bon de rappeler qu'un estimateur Bayésien est un estimateur qui minimise la perte espérée a posteriori. Si la fonction de perte est quadratique, l'estimateur Bayésien sera l'espérance a posteriori. Si la fonction de perte est de la forme

$$l(x, \theta) = \begin{cases} c_1(x - \theta) & \text{si } x \geq \theta \\ c_2(\theta - x) & \text{si } x < \theta \end{cases}$$

l'estimateur Bayésien sera le fractile $c_2/(c_1 + c_2)$ de la distribution a posteriori de θ . Voir Bauwens, Lubrano, and Richard (1999).

Dans l'article de Koenker et Bassett on a la définition suivante pour calculer la valeur b du quantile p :

$$Q_p(y) = \underset{[b]}{\text{ArgMin}} \left[\sum_{y_t \geq b} p|y_t - b| + \sum_{y_t < b} |y_t - b| \right]$$

On cherche donc bien un estimateur b qui va minimiser une fonction de perte espérée. Pour la régression quantile, on généralise cette expression à:

$$R_p(y) = \underset{[b]}{\text{ArgMin}} \left[\sum_{y_t \geq x'_t b} p|y_t - x'_t b| + \sum_{y_t < x'_t b} |y_t - x'_t b| \right]$$

Le seul problème, et il est de taille, c'est que maintenant on doit faire de l'optimisation en dimension k , la taille de b alors que la fonction de perte n'est pas différentiable. Le premier résultat de Koenker et Bassett c'est que cet estimateur est unique quand X est de rang plein. Le second, c'est que l'estimateur trouvé a une distribution asymptotique Normale qui est tirée de celle des estimateurs naturels des quantiles.

La littérature a ensuite parlé de check function. Ainsi dans l'article qui nous occupe, de la fonction de perte $\rho_p(z) = p|z|$, on passe à l'expression équivalente

$$\begin{aligned} \rho_p(z) &= pz\mathbf{1}(z > 0) - (1 - p)z\mathbf{1}(z \leq 0) \\ &= z \times (p - \mathbf{1}(z < 0)) \end{aligned}$$

L'estimation du fractile p se fera en minimisant la perte espérée

$$E_{y|x}[\rho_p(y - x'_t\beta)]$$

ou

$$\min_{[\beta]} \sum \rho_p(y_t - x'_t\beta)$$

Comme la fonction $\rho_p(z)$ n'est pas différentiable en zero, on est obligé de passer par un programme linéaire de la forme

$$\begin{aligned} z &= \text{ArgMin } c'z \\ Az &= y \\ z &\geq 0 \end{aligned}$$

qui n'est pas très efficace quand on a un grand nombre d'observations. Le problème vient de la contrainte de positivité. On va alors remplacer le programme original par

$$\text{Min}_{[x]} c'x - \mu \sum_m \ln x_m$$

Ces procédures sont implémentées dans la librairie QreG de Thierry Roncalli sous Gauss. Il semble que les écart-types ne soient pas disponibles.

6.1 Inférence Bayésienne

Une régression linéaire s'écrit

$$y = \theta(x) + \epsilon \quad \text{E}(\epsilon) = 0$$

Estimer $q_p(x)$ (le p-quantile de la distribution conditionnelle $f(y|x)$) revient à estimer la régression

$$y = \theta(x) + \epsilon \quad q_p(\epsilon) = 0$$

On se rend compte alors que pour conduire l'inférence dans ce type de modèle, il suffit de considérer une distribution asymétrique adéquate pour ϵ . Yu and Moyeed (2001) adoptent une distribution de Laplace asymétrique qui conduit à la fonction de vraisemblance

$$f_p(y|\beta) = p^n (1-p)^n \exp\{-\sum \rho_p(y_i - x'_i \beta)\}$$

L'inférence Bayésienne conduit à intégrer cette fonction sous une a priori possiblement uniforme. Ceci se fait très bien par Monte Carlo et peut se généraliser à des fonctions non-linéaires pour les quantiles.

6.2 Non-parametric inference

Les solutions classiques que l'on a décrites reposent sur la programmation linéaire. Les quantiles sont des fonctions linéaires de x . Une généralisation que l'on souhaite immédiatement apporter, c'est que ceux-ci soient des fonctions non-linéaires de x . On peut aussi vouloir adopter une approche non-paramétrique.

6.2.1 L'estimation nonparamétrique des quantiles

Considérons un échantillon X de distribution F . On définit l' α quantile $Q(\alpha)$ par l'inverse à gauche de F

$$Q(\alpha) = \inf\{x : F(x) \geq \alpha\}$$

L'estimateur traditionnel que l'on note SQ_α est défini à partir des statistiques d'ordre

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

qui n'est rien d'autre que l'échantillon ordonné. Le quantile empirique est obtenu au moyen de

$$SQ_\alpha = X_{[n\alpha]+1}$$

Cet estimateur n'est pas très efficace à cause de la variance d'échantillonnage qui fait varier les statistiques d'ordre. On va donc chercher à lisser par un kernel pour réduire cette variation. L'estimateur générique par noyau est

$$KQ_\alpha = \sum_{i=1}^n \left[\int_{i-1/n}^{i/n} K_h(t - \alpha) dt \right] X_i$$

On peut simplifier cet estimateur de différentes manières dont la plus intuitive est

$$KQ_{\alpha,1} = \sum_{i=1}^n \left[\frac{1}{n \times h} K_h(i/n - \alpha) \right] X_i$$

Il est bien entendu que cette formule suppose que l'échantillon est ordonné, comme pour l'estimateur naturel. On rappelle qu'un noyau est une densité qui s'intègre à 1. On peut prendre un noyau uniforme, triangulaire ou normal. Par exemple

$$K_h(i/n - \alpha) = \phi\left(\frac{i/n - \alpha}{h}\right)$$

avec $h = c * \sigma/n^{1/5}$ et ϕ la normale standardisée.

6.2.2 Régression quantile non-paramétrique

On va commencer par exposer une solution simple, celle d'Abberger (1997). Soient y_1 deux variables aléatoires. On considère la densité jointe $f(y_1, y_2)$ et la densité conditionnelle de y_2 si y_1 $f(y_2|y_1) = f(y_1, y_2)/f(y_1)$. On peut alors définir la cumulative de y_2 conditionnelle à y_1

$$F(y_2|y_1) = \int_{-\infty}^{y_2} \frac{f(y_1, t)}{f(y_1)} dt$$

Le α quantile de cette distribution qui dépend de y_1 peut se calculer comme solution en y_2 de

$$F(q_\alpha(y_2)|y_1) = \alpha$$

Sous certaines conditions de régularité, on peut inverser F et calculer alors directement

$$q_\alpha(y_1) = F^{-1}(\alpha|y_1)$$

Mais ici il faut tout d'abord estimer cette densité conditionnelle de façon non-paramétrique. L'estimateur le plus simple est

$$\hat{F}(y_2|y_1) = \frac{\sum K_h(y_1 - y_{1,i}) \mathbf{1}(y_{2,i} \leq y_2)}{\sum K_h(y_1 - y_{1,i})}$$

Cette solution conduit à un estimateur qui n'est pas très lisse. Il peut conduire à trouver des quantiles conditionnels qui se coupent. Considérons un estimateur plus élaboré, où la fonction indicatrice est remplacé par un noyau

$$G(z) = \int_{-\infty}^z K(t) dt$$

ce qui conduit à

$$\hat{F}(y_2|y_1) = \frac{\sum K_h(y_1 - y_{1,i}) G_h(y_2 - y_{2,i})}{\sum K_h(y_1 - y_{1,i})}$$

Il est sans doute plus facile d'inverser numériquement l'estimateur lissé que l'autre.

6.2.3 Les apports de Yu et Jones (1998)

Tout d'abord, il semble que le double noyau de la première méthode de ce papier soit identique à celle exposée plus haut, à ceci près que l'on autorise deux fenêtres de lissage.

Une régression non-paramétrique classique de y sur x est obtenue au moyen de

$$m(x) = \frac{\sum_i K_h(x_i - x) y_i}{\sum_i K_h(x_i - x)}$$

Si l'on veut utiliser la check function dans cette écriture, on sera conduit à utiliser l'algorithme

$$m_p(x) = \underset{[a]}{\text{Min}} \sum_i K_h(x_i - x) \rho_p(y_i - a)$$

Yu and Jones (1998) donnent un algorithme itératif qui passe par une série de moindres carrés pondérés.

6.3 Quantile regression using R

There is a package for computing non-parametric quantile regression in R. The general package needs `library(np)`. Then the routine is `npqreg`. There is an example using an Italian income panel which should be investigated seriously. The package `quantreg` contains all the necessary tools for parametric quantile regression. The basic command is `rq`.

References

- ATKINSON, A. (2003): "Income Inequality in OECD Countries: Data and Explanations," Working Paper 49: 479–513, CESifo Economic Studies, CESifo.
- BAILAR, B. A. . (1991): "Salary survey of U.S. colleges and universities offering degrees in statistics," *Amstat News*, 182, 3–10.
- BAUWENS, L., M. LUBRANO, AND J.-F. RICHARD (1999): *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, Oxford.
- BHAUMIK, S. K., I. N. GANG, AND M.-S. YUN (2006a): "A Note on Decomposing Differences in Poverty Incidence Using Regression Estimates: Algorithm and Example.," Working Paper 2006-33, Department of Economics, Rutgers University, Rutgers University.
- (2006b): "A Note on Decomposing Differences in Poverty Incidence Using Regression Estimates: Algorithm and Example," Discussion Paper No. 2262, IZA, Bonn, Germany, Available at SSRN: <http://ssrn.com/abstract=928808>.
- BOURGUIGNON, F., F. H. G. FERREIRA, AND P. G. LEITE (2008): "Beyond Oaxaca-Blinder: Accounting for differences in household income distributions," *Journal of Economic Inequality*, 6(2), 117–148.
- FERGUSON, T. S. (1967): *Mathematical Statistics: a decision theoretic approach*. Academic Press, New York.
- FOSTER, J., J. GREER, AND E. THORBECKE (1984): "A class of decomposable poverty measures," *Econometrica*, 52, 761–765.
- JANN, B. (2008): "The Blinder-Oaxaca decomposition for linear regression models," *The Stata Journal*, 8(4), 453–479, Standard errors for the Blinder-Oaxaca decomposition. Handout for the 3rd German Stata Users Group Meeting, Berlin, April 8 2005.
- JENKINS, S. P. (2000): "Modelling Household Income Dynamics," *Journal of Population Economics*, 13, 529–567.
- JUHN, C., K. M. MURPHY, AND B. PIERCE (1993): "Wage Inequality and the Rise in Returns to Skill," *The Journal of Political Economy*, 101(3), 410–442.
- KOENKER, R., AND G. BASSET (1978): "Regression quantiles," *Econometrica*, 46(1), 33–50.
- MACHADO, J. A. F., AND J. MATA (2005): "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of Applied Econometrics*, 20(4), 445 – 465.
- NGUYEN, B. T., J. W. ALBRECHT, S. B. VROMAN, AND M. D. WESTBROOK (2007): "A quantile regression decomposition of urban–rural inequality in Vietnam," *Journal of Development Economics*, 83(2), 466–490.

- OAXACA, R. (1973): "Male-female wage differentials in urban labor markets," *International Economic Review*, 14, 693–709.
- RADCHENKO, S. I., AND M.-S. YUN (2003): "A Bayesian approach to decomposing wage differentials," *Economics Letters*, 78(3), 431–436.
- STEVENS, A. H. (1999): "Climbing out of poverty, falling back in," *The Journal of Human Resources*, 34(3), 557–588.
- TREDE, M. (1998): "Making mobility visible: a graphical device," *Economics Letters*, 59, 77–82.
- YU, K., AND M. JONES (1998): "Local linear regression quantile estimation," *Journal of the American Statistical Association*, 93, 228–238.
- YU, K., AND R. A. MOYEED (2001): "Bayesian quantile regression," *Statistics and Probability Letters*, 54, 437–447.
- YUN, M.-S. (2004): "Decomposing differences in the first moment," *Economics Letters*, 82(2), 275–280.