cemmap
centre for microdata methods and practice

QUANTILES FOR COUNTS

José A F Machado
J M C Santos Silva

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP22/02

Quantiles for Counts*

José A. F. Machado
Faculdade de Economia, Universidade Nova de Lisboa
J. M. C. Santos Silva
ISEG, Universidade Técnica de Lisboa

October 3, 2002

Abstract

This paper studies the estimation of conditional quantiles of counts. Given the discreteness of the data, some smoothness has to be artificially imposed on the problem. The methods currently available to estimate quantiles of count data either assume that the counts result from the discretization of a continuous process, or are based on a smoothed objective function. However, these methods have several drawbacks. We show that it is possible to smooth the data in a way that allows inference to be performed using standard quantile regression techniques. The performance and implementation of the estimator are illustrated by simulations and an application.

JEL classification code: C13, C25.

Key words: Asymmetric maximum likelihood, Jittering, Maximum score estimator, Quantile regression, Smoothing.

---

*We are indebted to J.F. Correia Guedes, Paulo Parente and participants in seminars at ISEG and Universidade Católica for many helpful comments. The usual disclaimer applies. The data used in section 5 was kindly made available by Winfried Pohlmeier and are from the public use version of the German Socio-Economic Panel Study. We thank the Deutsches Institut für Wirtschaftsforschung for authorizing its use. The authors gratefully acknowledge the partial financial support from Fundação para a Ciência e a Tecnologia, program POCTI, partially funded by FEDER

Address for correspondence: João Santos Silva, ISEG, R. do Quelhas 6, 1200 Lisboa, Portugal. Fax: 351 213922781. E-mail: jmcss@iseg.utl.pt.

1. INTRODUCTION

Since the publication of the two seminal papers on the econometrics of count data (Hausman, Hall and Griliches, 1984, and Gourieroux, Monfort and Trognon, 1984), this research area has gained great popularity, being now the subject of two monographs (Cameron and Trivedi, 1998, and Winkelmann, 2000) and of countless theoretical and applied journal articles. This vast literature is roughly divided into two main strands.

Part of the literature follows the lead of Gourieroux, Monfort and Trognon (1984) and concentrates on the semiparametric estimation of the conditional mean of the count variate based on the pseudo-likelihood framework. Despite its elegance and attractiveness, this approach is limited by its own nature because it does not provide information on many aspects of the distribution of the counts which are often of interest in applied research. In fact, the pseudo-likelihood approach only permits the estimation of the conditional expectation of the variate of interest, which gives very little information about a conditional distribution in which features other than location can depend on the regressors in a complex way. Given the limitations of this approach, it is not surprising to find that, following the early work of Mullahy (1986), several fully parametric probabilistic models have been developed to describe particular features of count data sets often found in applications. These models completely describe the conditional distribution of interest and therefore allow the researcher to study the impact of the covariates on every aspect of the conditional distribution, at the cost of strong parametric assumptions and consequent lack of robustness.

An alternative approach to modelling count data was pioneered by Efron (1992) who proposed the asymmetric maximum likelihood estimator. This approach requires hypotheses that are comparable to those underlaying the pseudo-likelihood approach, while allowing the researcher to study the impact of the regressors on different regions of conditional distribution of interest. However, the estimated location measures are difficult to interpret.

The estimation of conditional quantile functions or quantile regressions (QR) was originally advocated and studied by Koenker and Bassett (1978) and is becoming increasingly popular (see, for instance, the special issue of *Empirical Economics* on this subject edited by Fitzenberger, Koenker and Machado, 2001). Although typical applications of QR assume random sampling from absolutely continuous populations, there exist a rapidly expanding set of results for other set-ups. The seminal departure was, of course, the work of Manski (1975 and 1985) on median regression for binary and multinomial models, later extended by Horowitz (1992). Powell (1984 and 1986) studied QR for censored data and Lee (1992) analysed median estimation of ordered discrete responses. More recently, QR techniques were applied to the study of duration or transition data by Koenker and Geling (1999), Koenker and Bilias (2001) and Machado and Portugal (2002).

This paper studies the possibility of estimating conditional quantiles of count data. Like the asymmetric maximum likelihood estimator, the estimation of conditional quantiles requires hypotheses that are comparable to those underlaying the pseudo-likelihood approach, while allowing the researcher to obtain most of the results that so far were only possible to obtain using more structured models. In particular, using quantile regression it is possible to study the impact of the regressors on each quantile of the distribution and it is also possible to produce some probabilistic statements about the counts. The main problem with the estimation of conditional quantiles with count data stems from the conjunction of a non-differentiable sample objective function with a discrete dependent variable. Since there are points of positive mass, the non-smoothness of the objective function is not necessarily averaged away and, consequently, the usual strategies based on Taylor expansions cannot be used to obtain the asymptotic distribution of the conditional quantiles (see, Huber, 1981, pp. 50-51 for a discussion about the interplay between smoothness of the objective function and of the density of the data). The problem is analogous to that faced in the estimation of conditional quantiles for binary data using the maximum score estimator, which ultimately yields the non-standard rate of convergence of the estimator (see Manski, 1975, and 1985). The objective function of the maximum score estimator is non-differentiable both because it depends on the absolute value of a residual and because the conditional median of Y given x, $\text{med}\,(Y|x)$, is an indicator function. However, as shown by Horowitz (1992), both sources of non-differentiability can be eliminated by smoothing the indictor function, leading to an estimator with a much more tractable asymptotic distribution and to convergence rates of practical utility.

In order to be able to apply quantile regression to counts, some degree of smoothness has to be artificially imposed on the problem. A possible approach is to view the count data as the discretized result of a continuous underlying model. Then, as in the binary choice case, the functional form of $\text{med}(Y|x)$ results from the specification of a latent model and the results of Lee (1992) on median regression for ordinal data can be extended to estimate the parameters of interest. To achieve rates of convergence of practical interest, the objective function defining the quantiles will have to be smoothed as in Horowitz (1992) and Melenberg and van Soest (1994 and 1996). However, this approach has several drawbacks. By its own nature, this approach lacks an interpretation in terms of an underlying count process and this may be an unappealing feature in some situations. Moreover, this approach requires the introduction of a new parameter for each count actually observed. On the other hand, the smoothed version of this estimator proposed by Melenberg and van Soest (1994 and 1996) depends on the specification of smoothing parameters for which, at present, there are no optimal selection rules and, for typical kernels, does not achieve the usual $\sqrt{n}$ convergence rate. Finally, the implementation of this estimator is still computationally very expensive, which may explain why it is seldom used in practice.

An alternative approach explored in this paper is based on the artificial smoothing of the data using a specific form of jittering introduced by Stevens (1950) in a different context. The idea is to construct a continuous variable whose conditional quantiles have a one-to-one relationship with the conditional quantiles of the counts, and use this artificially constructed continuous variable as a base for inference.

The remainder of the paper is organized as follows. Section 2 overviews the issues involved in the estimation of quantiles for discrete data and provides a flavour of the approach based on the artificial smoothing of the data. Section 3 presents the proposed quantile regression estimator as well as two alternative approaches currently available.

Sections 4 and 5 evaluate the proposed approach using simulation experiments and a well known data set. Finally, section 6 presents some concluding remarks.

2. A BIRD'S-EYE VIEW OF THE PROBLEM

Although this paper focuses on the estimation of quantile regression for count data, it is useful to introduce the main features of this problem, and the solution proposed here, in a much simpler context. To this end, this section considers the estimation of the quantiles of a discrete random variable in the simplest possible case. Let Y denote a Bernoulli random variable with $\Pr(Y=0)=q$, and $0<q<1$. The α-quantile of Y can be written as[1]

$$Q_Y(\alpha) = 1(\alpha > q),$$

where $1(A)$ is the usual indicator function for the event A. Given a sample $\{Y_i\}_{i=1}^n$, the maximum likelihood estimator of q is given by $\hat{q} = 1 - \frac{1}{n}\sum_{i=1}^n Y_i$, leading to the following estimator for $Q_Y(\alpha)$

$$\hat{Q}_Y(\alpha) = 1(\alpha > \hat{q}),$$

which is the solution of $\min_{\eta\in[0,1]} \sum_{i=1}^n \rho_\alpha(Y_i - \eta)$, with $\rho_\alpha(v) = v(\alpha - I(v<0))$ being the usual check function.

Notice that, since $\hat{Q}_Y(\alpha)$ is not a continuous function of $\hat{q}$ at $\alpha = \hat{q}$, Slutsky's theorem cannot be invoked to claim consistency for $\hat{Q}_Y(\alpha)$. However, the asymptotic behaviour of $\hat{Q}_Y(\alpha)$ is easy to analyse. In fact, $\hat{Q}_Y(\alpha)$ is also a Bernoulli random variable with $\Pr\left(\hat{Q}_Y(\alpha)=0\right) = \Pr(\alpha \leq \hat{q})$. Therefore, by the De Moivre-Laplace theorem, for large n the probability $\Pr\left(\hat{Q}_Y(\alpha)=0\right)$ can be approximated by $1-\Phi\left(\sqrt{n}\frac{\alpha-q}{\sqrt{q(1-q)}}\right)$.

This result has several interesting implications.

1. For $\alpha > q$, $\Pr\left(\hat{Q}_Y(\alpha)=0\right)$ tends to 0 as n passes to ∞. Therefore $P\lim\left(\hat{Q}_Y(\alpha)\right) = Q_Y(\alpha)$. Similarly, for $\alpha < q$, $\Pr\left(\hat{Q}_Y(\alpha)=0\right)$ tends to 1, and again $P\lim\left(\hat{Q}_Y(\alpha)\right) = Q_Y(\alpha)$. Therefore, for $\alpha \neq q$, the estimator $\hat{Q}_Y(\alpha)$ is consistent for $Q_Y(\alpha)$, a result that can also be obtained using Slutsky's theorem.

[1] Recall that the α-quantile of Y is defined by $Q_Y(\alpha) = \min\{\eta | P(Y \leq \eta) \geq \alpha\}$.

2. If $\alpha = q$, $\Pr\left(\hat{Q}_Y(\alpha) = 0\right) = 1 - \Phi(0) = 0.5$. Consequently, as the sample size goes to infinity, $\hat{Q}_Y(\alpha)$ converges to a Bernoulli distribution with parameter 0.5, and therefore the estimator is not consistent.

3. The variance of the distribution of $\hat{Q}_Y(\alpha)$ is

$$V\left[\hat{Q}_Y(\alpha)\right] = \Pr(\alpha \leq \hat{q}) \Pr(\alpha > \hat{q}),$$

   which, invoking again De Moivre-Laplace's theorem, for $\alpha \neq q$, vanishes as n passes to infinity. It is interesting to notice that the rate of convergence is not the usual since $nV\left[\hat{Q}_Y(\alpha)\right]$ also converges to zero. Using well known results on the rate of decay of the tails of the normal distribution (see, e.g., Pollard, 1984, appendix B), it is possible to establish that $V\left[\hat{Q}_Y(\alpha)\right]$ converges to zero like $n^{-1/2}\exp(-n)$.

4. Because the estimated quantiles are random variables with the same support as Y, their asymptotic distribution cannot be normal when the data is discrete.

To sum up, in this very simple case none of the standard results on the asymptotic properties of the estimators holds.

Consider now that a new random variable Z is constructed by adding to Y a random variable U uniformly distributed in the interval $[0, 1)$. That is, $Z = Y + U$. Obviously, Z is a continuous random variable with positive density on the interval $[0, 2)$ and $Q_Z(\alpha)$ is defined by

$$Q_Z(\alpha) = \begin{cases} \frac{\alpha}{q} & q > \alpha \\ 1 + \frac{\alpha - q}{(1-q)} & q < \alpha \end{cases}.$$

The interesting feature of the random variable Z is that there is a one-to-one relationship between $Q_Y(\alpha)$ and $Q_Z(\alpha)$. Proposition 1 in section 3 shows that $Q_Y(\alpha) = \lceil Q_Z(\alpha) - 1 \rceil$, where $\lceil a \rceil$ denotes the so-called *ceiling* function which returns the smallest integer greater than, or equal to, a.

Therefore, information about the quantiles of Y, the variate of interest, can be obtained from $Q_Z(\alpha)$. It is now interesting to study the properties of the following estimator of

$Q_Z(\alpha)$

$$\hat{Q}_Z(\alpha) = \begin{cases} \frac{\alpha}{\hat{q}} & \hat{q} \geq \alpha \\ 1 + \frac{\alpha - \hat{q}}{(1-\hat{q})} & \hat{q} < \alpha \end{cases},$$

which, like $\hat{Q}_Y(\alpha)$, only depends on $\hat{q}$ and α.

The first thing that can be noticed is that $Q_Z(\alpha)$ is a continuous function of q and therefore, by Slutsky's theorem, $\hat{Q}_Z(\alpha)$ is consistent for $Q_Z(\alpha)$. Moreover, for $\alpha \neq q$, $\hat{Q}_Z(\alpha)$ is a continuous and differentiable function of $\hat{q}$ and in this case the delta method can be used to obtain the usual normal approximation to the distribution of $\hat{Q}_Z(\alpha)$.

However, for $\alpha = q$, the derivative of $Q_Z(\alpha)$ with respect to q only exists when $\alpha = q = 0.5$. This means that, except in this case, there will be a "discontinuity in the variance" of the asymptotic distribution of $\hat{Q}_Z(\alpha)$ (cf. Huber, 1981, pp. 50-51 and also Simpson, Carroll and Ruppert, 1987 and Knight, 1998). Therefore, the standard results will hold for the estimation of any quantile when $q = 0.5$. Otherwise, the usual asymptotic theory is valid only for quantiles with $\alpha \neq q$.

As noticed, the reason for this breakdown of the asymptotic theory when $\alpha = q$ and $q \neq 0.5$ is that in this case the derivative of $Q_Z(\alpha)$ with respect to q is not continuous at $\alpha = q$. Equivalently, this results from the fact that the probability density function of Z is not continuous at $Q_Z(\alpha)$ when $\alpha = q$, unless $q = 0.5$. What is interesting to notice is that, whatever the values of q and α, this discontinuity occurs for $Q_Z(\alpha) = 1$. Using the results of Knight(1998) it is possible to show that

$$\sqrt{n}\left(\hat{Q}_Z(\alpha) - 1\right) \xrightarrow{\mathcal{D}} \frac{1}{1-q} V I\{V \geq 0\} + \frac{1}{q} V I\{V < 0\}$$

where V is standard normal. When $q = 1/2$ the limiting distribution is $\mathcal{N}(0, 1/4)$; otherwise is a member of the two-piece normal family of distributions (John, 1982), with parameters $(0, (1-q)^{-1}, q^{-1})$.

Consider now the regression problem in which q is a continuous function of a set of regressors, say x, and the researcher is interested in the estimation of the conditional quantiles of Y, denoted $Q_Y(\alpha|x)$. As before, $Q_Y(\alpha|x)$ is either 0 or 1 and its estimator is marred by problems which are similar to those found in the simpler case discussed above.

However, smoothing Y by adding to it the uniform random variable, partially solves these problems. In particular, except for the set in which x is such that $Q_Z(\alpha|x) = 1$, standard inference holds. However, as it will be argued below, this set will have measure zero if one assumes that there exists at least on continuously distributed covariate and the conditional quantiles of Z, $Q_Z(\alpha|x)$ are a continuous function (measurable would suffice) of that covariate. This approach is explored in the following section.

3. SMOOTH QUANTILES FOR COUNTS

Some of the problems resulting from the discreteness of the data were highlighted in the previous section. In the remainder of this section, we explore different ways around this problem. Two of the methods considered have been available for the last 10 years. However, these methods are not particularly attractive and, to our knowledge, they have never been used in practice to model count data, and are not ever referred to in the monographs of Cameron and Trivedi (1998) and Winkelmann (2000). We start by surveying the existing alternatives and then we will present the new method whose essence was sketched in the previous section.

3.1 Smoothing the Process

One possible approach consists in modelling the discrete data as the result of a continuous process crossing some thresholds. Despite the already mentioned drawbacks of treating count data as the result of a continuous process, for completeness, we now present a sketch of this approach.

The dependent variable Y takes values in the set $\{0, 1, \ldots, M\}$ according to the model

$$Y - \sum_{j=1}^{M} 1(Y^\star > \xi_j) = 0$$

with $\xi_1 < \xi_2 < \ldots < \xi_M$ and where $Y^\star$ is a latent variable satisfying, for a given set of covariates x,

$$Y^\star = x'\beta(\alpha) + \varepsilon_\alpha \ , \ Q_{\varepsilon_\alpha}(\alpha|x) \equiv 0.$$

Since the function $\sum_{j=1}^{M} 1(Y^{\star} > \xi_j)$ is monotone, the equivariance property of the quantiles implies that,

$$Q_Y(\alpha|x) = \sum_{j=1}^{M} 1(Q_{Y^\star}(\alpha|x) > \xi_j) = \sum_{j=1}^{M} 1(x'\beta(\alpha) > \xi_j).$$

Lee (1992) uses this result to extend Manski's maximum score estimator to ordered discrete data models that is, in our setting, to the estimation of the conditional median. The extension to general quantiles would be direct,

$$\min_b \sum_{i=1}^{n} \rho_\alpha \left(Y_i - \sum_{j=1}^{M} 1(x_i' b > \xi_j) \right)$$

where, as before, $\rho_\alpha(v)$ is the check function (for simplicity of exposition we are conditioning on the ξ_j's).

Although we are not aware of such a proof, it is natural to anticipate that the estimator will share Manski's major drawback, the $n^{-1/3}$ rate of convergence. This remark suggests a smoothing alternative similar to Horowitz's (1992) and Melenberg and van Soest (1994 and 1996) (see also Horowitz, 1998). That would entail replacing all the indicator functions with integrated kernels. Notice, however, that the final smooth objective function would result quite messy because things do not simplify as nicely as in the binary case.[2]

3.2 Smoothing the objective function

As noted in the introduction, the discreteness of the count data makes the sample objective function defining the quantile regression estimator non-differentiable, leading to all sorts of problems in obtaining asymptotically valid inference. Efron's (1992) asymmetric maximum likelihood (AML) estimator can be interpreted as resulting from smoothing the objective function defining the quantile regression estimator. The AML estimates conditional location functions for count data which are akin to the conditional expectiles proposed by Newey and Powell (1987) for the linear model. Therefore, strictly speaking,

[2] A simpler parametric alternative would be the rank regression approach of Fortin and Lemieux (2000).

this method does not permit the estimation of conditional quantiles for counts, but it allows the estimation of different location functions for which asymptotically valid inference can be performed using standard methods.

This approach requires hypotheses that are comparable to those underlaying the pseudo-likelihood approach, while allowing the researcher to study the impact of the regressors on different location functions. Moreover, this method is flexible enough to accommodate the problems of under or overdispersion that so often afflict empirical models.

The asymmetric maximum likelihood estimator is defined by

$$\hat{\beta}_w^{\text{AML}} = \arg\max_b \sum_{i=1}^{n} \left(Y_i x_i' b - \exp\left(x_i' b\right) - \ln\left(Y_i!\right)\right) w^{I\left(Y_i > \exp\left(x_i' b\right)\right)},$$

and reduces to the usual Poisson pseudo-likelihood estimator for $w = 1$.

Efron (1992) defines that the 100α-th AML regression percentile is obtained for w such that $\alpha = \frac{1}{n}\sum_{i=1}^{n} I\left(Y_i \leq \exp\left(x_i'\hat{\beta}_w^{\text{AML}}\right)\right)$. This definition reveals an important limitation of this approach. Since the indicator function $I\left(Y_i \leq \exp\left(x_i'\hat{\beta}_w^{\text{AML}}\right)\right)$ is necessarily equal to one when $Y_i = 0$, the AML regression percentiles cannot be computed for values of α smaller that the proportion of zeros in the sample. Furthermore, as in the case of the linear model (see Koenker 1992 and 1993), the estimated location measures are difficult to interpret and, like the pseudo-likelihood approach, this modelling strategy does not allow the researcher to make probability statements about the counts. In a count data context, the main advantage of the expectiles is that they are generally easier to estimate because are not restricted to be integers.

3.3 Smoothing the Data

Consider the random variables Y and X, and let $Q_Y(\alpha|x)$ denote the α-th quantile of the conditional distribution of Y given $X = x$. Koenker and Bassett (1978) give sufficient conditions for asymptotically valid inference on the parameters of $Q_Y(\alpha|x)$. Among these conditions, the conditional probability density function $f(Y|x)$ is required to be continuous and positive at $Q_Y(\alpha|x)$. If Y results from a count, its support is the set of the non-negative integers and those sufficient conditions are not satisfied.

3.3.1 Jittering

The main problem with the estimation of QR when Y results from counts is that, because Y has a discrete distribution, $Q_Y(\alpha|x)$ cannot be a continuous function of the parameters of interest. This limitation can be overcome by constructing a continuous random variable whose quantiles have a one-to-one relation with the quantiles of Y. A variable satisfying this requirement can be constructed by adding to Y, the count variate of interest, U, a uniform random variable in the interval $[0,1)$, leading to $Z = Y + U$. This approach uses a specific form of jittering proposed by Stevens (1950) (see also Anscombe, 1948, and Pearson, 1950) to introduce smoothness into the problem, leading to a conditional quantile function that is continuous in α. Continuity is achieved by interpolating each jump in the conditional quantile function of the counts using an integrated kernel, much in the same way Horowitz (1992) smooths the conditional median of binary data. The difference is that here the uniform distribution is used, not only due to the important historical precedent, but specially because in the case of count data the kernel has to have bounded support. Also, in this case, the uniform distribution allows important algebraic and computational simplifications.

The next proposition, whose proof is given in the appendix, relates the quantiles of the two random variables.

Proposition 1 $Q_Y(\alpha|x) = \lceil Q_Z(\alpha|x) - 1 \rceil$, *where* $\lceil a \rceil$ *denotes the ceiling function which returns the smallest integer greater than, or equal to,* a.

As we show below, its is passible to perform inferences about the conditional quantile functions $Q_Z(\alpha|x)$ and, in particular, about the "regression coefficients" $\gamma_j(\alpha;x) \equiv \partial Q_Z(\alpha|x)/\partial x_j$. These quantities are is interesting in themselves as the distribution of Z is closely related to the distribution of interest. However, $\gamma(\alpha;x)$ also provides information about the way $Q_Y(\alpha|x)$ depends on the regressors. Proposition 1 above makes clear that if a given $\gamma_j(\alpha;x)$ is zero that implies that $Q_Y(\alpha|x)$ does not depend on the corresponding covariate. Therefore, it is possible to test the significance of the regressors in $Q_Y(\alpha|x)$ by testing their significance in $Q_Z(\alpha|x)$.

But it is possible to say more about the interpretation of $\gamma_j(\alpha;x)$. Of course, it does not make sense to talk about derivatives of $Q_Y(\alpha|x)$ with respect to a particular x_j. But its certainly possible to define "inverse sensitivities" as the minimum variation in x_j needed to ensure that $\Delta Q_Y(\alpha) = Q_Y(\alpha|x_j + \Delta x_j) - Q_Y(\alpha|x_j) = 1$, taking all other covariates as fixed. First notice that if a given quantile of Z vary by an integer amount so does the corresponding quantile of Y; in symbols, $\Delta Q_Z(\alpha) = i \Rightarrow \Delta Q_Y(\alpha) = i$, i integer. Thus, under the smoothness conditions of Taylor's theorem, the variation in the j-th regressor sufficient to ensure that the α-th conditional quantile (at x) of the count process increases by one unit can be approximated by $1/\gamma_j(\alpha;x)$. Therefore, the interpretation of $\gamma_j(\alpha;x)$ in terms of $Q_Y(\alpha|x)$ is not very different from the interpretation of the parameters in any standard non-linear model.

3.3.2 Asymptotics

Although the distribution function of Z is continuous, it is not smooth over its entire support. In fact, it does not have continuous derivatives for integer values of Z. The problems for the theory developed by Koenker and Bassett (1978) would occur for inferences about quantiles that turn out to be integers. However, under mild assumptions, valid asymptotic inference is still possible, as we now proceed to show.

The asymptotic distribution of the QR estimator will be derived under the following assumptions:

(A1) Y is a discrete random variable, with support in $\mathbb{N}_0$, the set of the non-negative integers; in its (minimum) support the conditional probability function of Y given X is uniformly bounded away from zero for almost every realization of X;

(A2) X is a random vector in $\mathbb{R}^k$ (with $X_1 = 1$) satisfying

- (a) $P(X'\eta = 0) = 0$ for any $\eta \in \mathbb{R}^k$,
- (b) $E(XX')$ is finite and non-singular;

(A3) Make $Z = Y + U$, where U is a Uniform in $[0, 1)$ random variable, independent of X and Y. For some known monotone transformation $T(\cdot; \alpha)$, possibly depending on α, the following restriction on the quantile process of Z given X holds

$$Q_{T(Z;\alpha)}(\alpha|x) = x'\gamma(\alpha), \text{ for } \alpha \in (0, 1).$$

and $\gamma(\alpha) \in \Gamma$, a compact subset of $\mathbb{R}^k$.

Most of these assumptions are standard in the quantile regression literature (see, e.g., Pollard, 1991). The only "usual" assumption that is missing is the continuity of the conditional density of the regressand at the quantile of interest. By construction, the set of discontinuity points of the density of Z given x is $\mathbb{N}_0$. Assumption (A2a) ensures that $P(X'\gamma(\alpha) \in \mathbb{N}_0) = 0$ and, consequently, for almost every realization of X the conditional density of the regressand at the quantile of interest will be continuous. It is clear that the same result could be obtained under conditions weaker than (A2a). For instance, it would suffice to suppose that the quantile being estimated depends on a continuously distributed covariate. That is, it is enough to assume that in (A3) the component of $\gamma(\alpha)$ corresponding to the continuous covariate is not zero for the α of interest.

We now present the main results of the paper, whose proofs can be found in the appendix.

Proposition 2 *The data $\{(y_i, x_i, u_i)\}_{i=1}^n$ is a random sample of (Y, X, U), satisfying (A1), (A2) and (A3). If $\hat{\gamma}(\alpha)$ is the estimator of $\gamma(\alpha)$ defined by*

$$\min_{c \in \mathbb{R}^k} \sum_{i=1}^{n} \rho_\alpha \left(T(Z_i; \alpha) - x_i' c \right)$$

where $\rho_\alpha(v) = v(\alpha - I(v < 0))$, then

$$\sqrt{n}\left(\hat{\gamma}(\alpha) - \gamma(\alpha)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, D^{-1} A D^{-1})$$

with

$$A = \alpha(1 - \alpha)E(XX')$$

and

$$D = E\left(f_{T|x}(X'\gamma(\alpha)|X)XX'\right)$$

with $f_{T|x}(\cdot|\cdot)$ denoting the conditional density of $T(Z; \alpha)$ given x.

For this estimator to be useful in practice, it is necessary to obtain a consistent estimator for the covariance matrix of $\sqrt{n}\left(\hat{\gamma}\left(\alpha\right)-\gamma\left(\alpha\right)\right)$. Given that $\hat{A}=\alpha\left(1-\alpha\right)\frac{1}{n}\sum_{i=1}^{n}x_{i}x_{i}'$ is a consistent estimator of A, the main difficulty in constructing an asymptotically valid covariance matrix is to find a consistent estimator for D.

The approach followed here for the estimation of D is akin to the one proposed by Powell (1984), but takes advantage of some specific characteristics of this problem. In particular, we explore the fact that, given the way Z is constructed, its density at $Z=z$ equals the probability that Z is greater or equal to $\lfloor z\rfloor$ and smaller than $\lfloor z+1\rfloor$, where $\lfloor a\rfloor$ denotes the floor function which returns the largest integer smaller than, or equal to, a. Therefore, the conditional expectation of

$$I\{\lfloor Q_{Z_i}(\alpha|x)\rfloor\leq Z_i<\lfloor Q_{Z_i}(\alpha|x)+1\rfloor\}\tag{1}$$

equals the conditional density of Z_i at $Q_{Z_i}(\alpha|x)$. Of course, what is needed for the covariance estimator is not this, but an estimate of the density of $T(Z;\alpha)$ at $x'\gamma\left(\alpha\right)$, which can be obtained from the density of Z at $Q_{Z_i}(\alpha|x)$ multiplying it by the Jacobian of the transformation. The implementation of this approach is complicated by the fact that the floor function is discontinuous. In order to be able to prove the consistency of the proposed estimator, this function is replaced in (1) by a continuous function that approaches the floor function as the sample size grows. To construct a consistent estimator for D, three further assumptions are needed.

(A4) The inverse of transformation $T(Z;\alpha)$, $T^{-1}(\cdot)$, is twice continuously differentiable, with derivatives denoted by $T^{(j)}\equiv\partial^{j}T^{-1}(v;\alpha)/\partial v^{j}$, $j=1,2$.

(A5) The following expectations exist

1. $E\left[\left|T^{-1}(x_i'\gamma\left(\alpha\right))\right|\left\|x_i\right\|^2\right]$
2. $E\left[\sup_{\|\gamma-\gamma(\alpha)\|\leq\delta}\left|T^{(j)}(x_i'\gamma)\right|\left\|x_i\right\|^2\right]$, for $j=1,2$ and some $\delta>0$.

(A6) $\{c_n\}$ is a sequence of real numbers in $(0,1/2)$, such that $c_n=o(1)$ and

$$\frac{\sup_{1\leq i\leq n}\left|T^{-1}\left(x_i'\hat{\gamma}\left(\alpha\right)\right)-T^{-1}\left(x_i'\gamma\left(\alpha\right)\right)\right|}{c_n}=o_p(1)$$

as $n \to \infty$.

Proposition 3 *Let,*

$$\hat{D}_n \equiv \frac{1}{n} \sum_{i=1}^{n} \hat{\omega}_i x_i x_i'$$

with

$$\hat{\omega}_i \equiv T^{(1)} \left(x_i' \hat{\gamma} (\alpha) \right) I \left\{ F_n \left(\hat{Q}_{Z_i} (\alpha|x) \right) \leq Z_i < F_n \left(\hat{Q}_{Z_i} (\alpha|x) + 1 \right) \right\}$$

where $\hat{Q}_{Z_i}(\alpha|x) \equiv T^{-1}(x_i' \hat{\gamma}(\alpha))$ *is the estimated* α*-th conditional quantile of* Z *and, for a sequence* c_n*,*

$$F_n(x) = \begin{cases} \lfloor x \rfloor - 1/2 + (x - \lfloor x \rfloor) / (2c_n) & x - \lfloor x \rfloor < c_n \text{ and } x \geq 1 \\ \lfloor x \rfloor & c_n \leq x - \lfloor x \rfloor < 1 - c_n \text{ or } x < 1 \\ \lfloor x \rfloor + 1/2 + (x - \lfloor x \rfloor - 1) / (2c_n) & x - \lfloor x \rfloor \geq 1 - c_n \end{cases} .$$

Then, under (A1) to (A6)

$$\hat{D}_n \xrightarrow{P} D.$$

Assumption (A6) warrants a few comments. As it is shown in the proof of Proposition 3, $\sup_{1 \leq i \leq n} \left| \hat{Q}_{Z_i} (\alpha|x) - Q_{Z_i} (\alpha|x) \right| = o_p(1)$ and, thus, (A6) restricts c_n to converge to 0 not too fast (slower than the numerator). The precise nature of this restriction depends on the regressors and on the transformation $T(\cdot)$ defined in Proposition 2. Suppose first that the regressors are taken to have bounded support, i.e., $P[\|x_i\| < K] = 1$, for some K. This is the case considered in the simulation study presented in the next section. In this context Proposition 2 implies that $\sup_{1 \leq i \leq n} |x_i' (\hat{\gamma}(\alpha) - \gamma(\alpha))| = O_p(\sqrt{n})$; on the other hand, the continuity of $T^{(1)}(\cdot)$ implies that, in a closed neighbourhood of $\gamma(\alpha)$, $T^{(1)}(x_i' \gamma(\alpha))$ is bounded, uniformly in i. Therefore,

$$\sup_{1 \leq i \leq n} \left| T^{-1} \left(x_i' \hat{\gamma} (\alpha) \right) - T^{-1} \left(x_i' \gamma (\alpha) \right) \right| = O_p(1/\sqrt{n}),$$

and c_n must converge to 0 slower than $1/\sqrt{n}$ that is, $c_n \sqrt{n} \to \infty$.

The condition of bounded support is rather extreme. However, it is quite natural in linear quantile regression models with heterogeneously distributed "errors" since, otherwise, it would be impossible for the conditional quantiles to be linear for all values of x without crossing. Alternatively, one may prefer the common assumption that the regressors satisfy

$$\sup_{1\leq i\leq n}\|x_i\| = O_p\left(\frac{n^{1/4}}{\ln(n)}\right)$$

(see, e.g., Koenker and Machado, 1999). If one further assumes that $T^{(1)}(\cdot)$ is Lipschitz continuous,

$$\sup_{1\leq i\leq n}\left|T^{-1}\left(x_i'\hat{\gamma}(\alpha)\right) - T^{-1}\left(x_i'\gamma(\alpha)\right)\right| = O_p\left(\frac{n^{1/4}}{\ln(n)}\right)O_p\left(\frac{1}{\sqrt{n}}\right) = O_p\left(\frac{1}{n^{1/4}\ln(n)}\right).$$

Here, c_n must be such that $c_n n^{1/4}\ln(n) \to \infty$. The Lipschitz-continuity assumption may be too stringent. For instance, it is not met when $T^{(1)}(\cdot) = \exp(\cdot)$. Specific rates for c_n must be derived on a case by case basis. For inverse transformations in the power family, $T^{(1)}(x) = x^\tau$ with $\tau \leq 2$, a simple mean-value expansion yields

$$\sup_{1\leq i\leq n}\left|T^{-1}\left(x_i'\hat{\gamma}(\alpha)\right) - T^{-1}\left(x_i'\gamma(\alpha)\right)\right| = O_p\left(\frac{n^{\tau/4}}{\ln(n)^\tau}\right)O_p\left(\frac{1}{\sqrt{n}}\right) = O_p\left(\frac{1}{n^{(2-\tau)/4}\ln(n)^\tau}\right)$$

and, consequently, in this case c_n will be restricted to go to 0 slower than $1/n^{(2-\tau)/4}\ln(n)^\tau$.

3.3.3 Implementation issues

In order to implement the results in propositions 2 and 3, it is necessary to specify the form of $Q_Z(\alpha|x)$ and the associated transformation $T(Z;\alpha)$. A possible approach to this problem is now described. Noting that $Q_Z(\alpha|x)$ is bounded from below by α, and keeping in line with what is traditionally assumed in count data models (including the work of Efron, 1992, on conditional expectiles for count data), a parametric representation of $Q_Z(\alpha|x)$ can be specified as

$$Q_Z(\alpha|x) = \alpha + \exp\left(x'\gamma(\alpha)\right). \tag{2}$$

Notice that the count data models commonly used in practice do not lead to conditional quantiles of Z of this form. However, this specification permits great computational simplifications and provides an approximation to the unknown conditional quantile functions, much in the same way linear regression is used to approximate unknown mean regression functions (see White, 1980). Of course, in presence of misspecification of the regression quantiles, appropriate estimators of the covariance matrix have to be used (see Chamberlain, 1994, and Kim and White, 2002). Further details on this are given in section 4.

Using (2), a simple algorithm to estimate $\gamma(\alpha)$ can be designed taking advantage of well known properties of the quantile functions. In fact, $\gamma(\alpha)$ can be estimated by running a quantile regression of

$$T(Z;\alpha) = \begin{cases} \log(Z-\alpha) & \text{for} \quad Z > \alpha \\ \log(\varsigma) & \text{for} \quad Z \leq \alpha \end{cases} \tag{3}$$

on x, with ς being a suitably small positive number. This is so because quantiles are invariant to monotonic transformations, and because quantiles are also invariant to censoring from below up to the quantile of interest.

4. SIMULATION RESULTS

The previous section suggests a simple procedure to estimate conditional quantiles of a count variable Y from the quantiles of an artificially constructed auxiliary variable Z. Moreover, it was argued that it is possible to construct an asymptotically valid test for the hypothesis that $Q_Y(\alpha|x)$ does not depend on a given covariate by testing that the element of $\gamma(\alpha)$ corresponding to that regressor is zero. This section reports the results of a pilot simulation study on the finite sample behaviour of such test.

In these experiments, the counts $Y_1, Y_2, \ldots, Y_n$ were generated according to four different models. In the first model, the counts are Poisson random variables with conditional mean $\lambda_i = \exp(\theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i})$, where $x_{1i} = \Phi^{-1}\left(0.05 + 0.9\frac{i-1}{n-1}\right)$ and x_{2i} is a dummy variable that equals 1 when i is a multiple of 5, being 0 otherwise. In the second model the conditional distribution of the Y's is negative binomial with mean λ_i as in model 1, and overdispersion parameter $\sigma = 0.5$. The third and fourth models are zero inflated versions of the first two models, with a proportion of zero inflation of 0.2.

All experiments were performed with $\theta_0 = \theta_1 = 1$ and $\theta_2 = 0$. Thus, λ_i does not depend on x_2, which therefore does not affect any aspect of the distributions of the counts. This implies that the effect of x_2 on $Q_Y(\alpha|x)$ is zero for every α. Using TSP 4.5 (Hall and Cummins, 1999), 5000 simulations for each model were performed for $\alpha \in \{0.25, 0.50, 0.75\}$ and $n \in \{500, 1000, 2000\}$.

In order to evaluate the proposed test for the hypothesis that $Q_Y(\alpha|x)$ does not depend on x_2, the counts were jittered as described in section 3 and the quantiles of the artificial data were estimated using the procedure described in the previous section. Specifying $Q_Z(\alpha|x)$ as in (2), Z was then transformed to obtain $T(Z;\alpha)$ according to (3), with $\varsigma = 10^{-5}$.

Figures 1 to 4 graph the first there quartiles of $T(Z;\alpha)$ against x_1 for each of the models considered. These quantiles are clearly non-linear, especially for the zero inflated cases. However, overall, the assumption of linearity seems a reasonable approximation, at least over this range of x_1. Therefore, the parameters of interest were estimated by running the usual QR of $T(Z;\alpha)$ on x_1, x_2 and a constant, and the significance of the parameter associated with x_2 was tested using a t-ratio. Given the misspecification of the functional form of the quantiles being estimated, the misspecification robust version of the covariance matrix was used in the construction of the t-ratio (see Chamberlain, 1994 and Kim and White, 2002). Specifically, the estimator used is based on the asymptotically valid covariance matrix presented in section 3, which in this particular example has the form

$$\hat{D}^{-1}\hat{A}\hat{D}^{-1},$$

with

$$\hat{D} = \frac{1}{n}\sum_{i=1}^{n} \exp\left(x_i'\hat{\gamma}(\alpha)\right) I\left(F_n\left(\alpha + \exp\left(x_i'\hat{\gamma}(\alpha)\right)\right) \leq Z_i < F_n\left(\alpha + \exp\left(x_i'\hat{\gamma}(\alpha)\right) + 1\right)\right) x_i x_i',$$

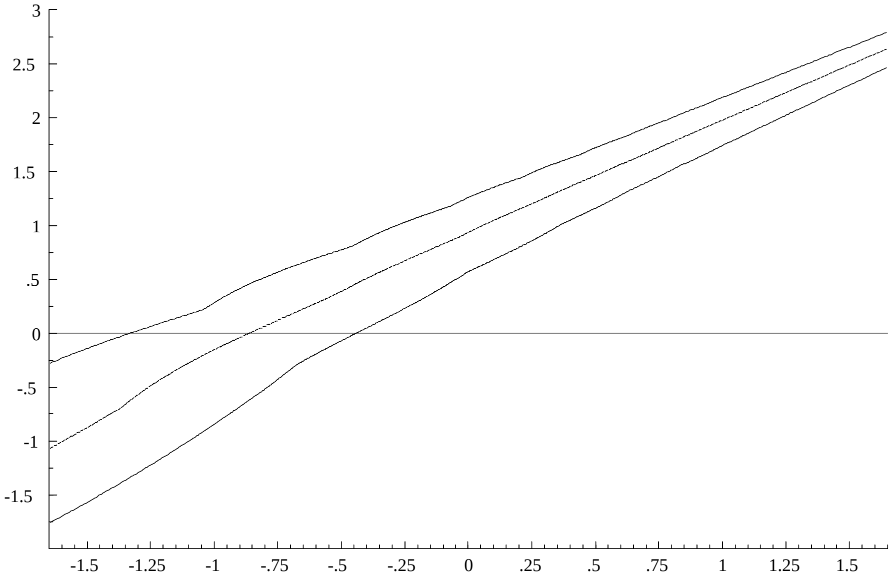

Fig. 1: The first three quartiles of $T(Z;\alpha)$ against x_1 for model 1

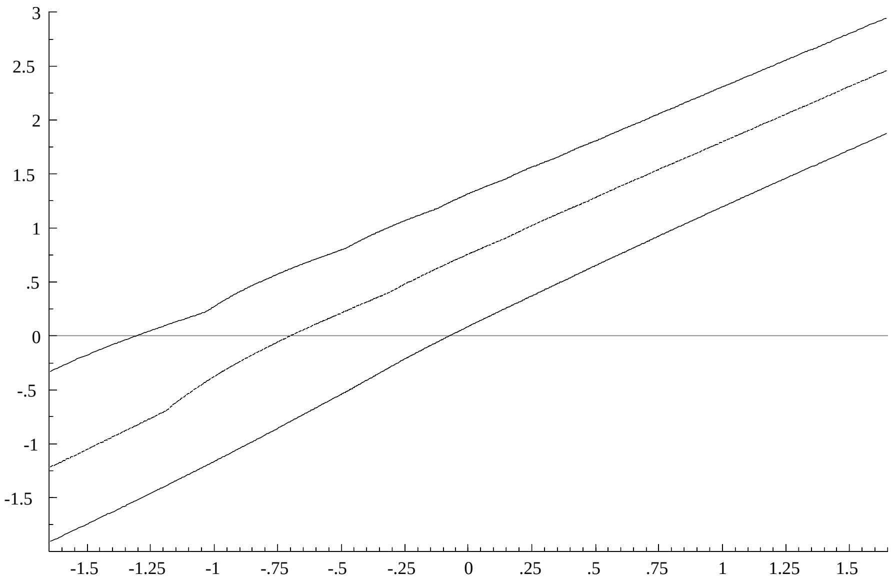

Fig. 2: The first three quartiles of $T(Z;\alpha)$ against x_1 for model 2

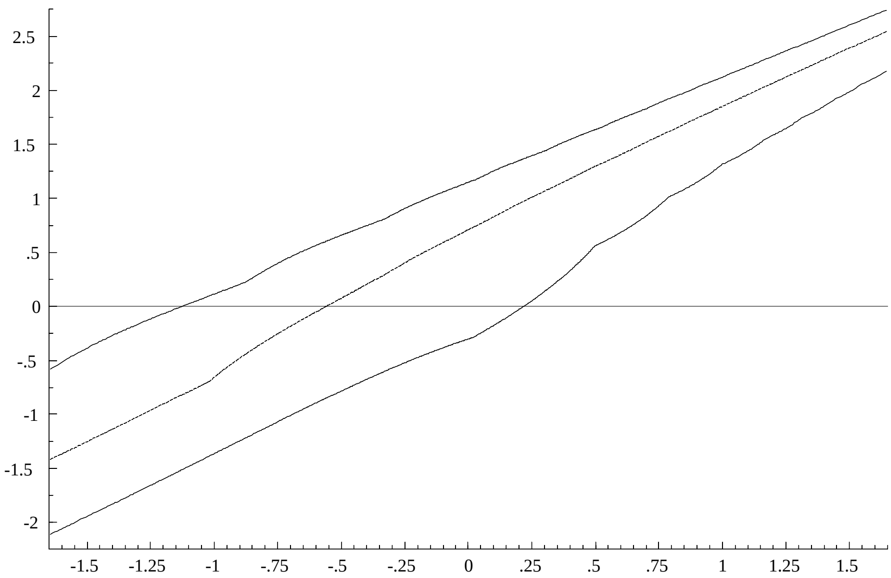

Fig. 3: The first three quartiles of $T(Z;\alpha)$ against x_1 for model 3

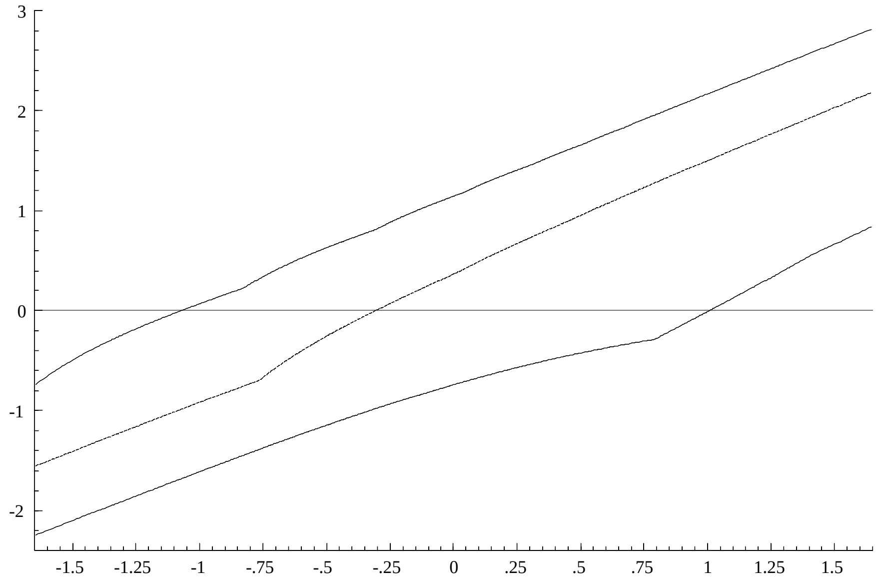

Fig. 4: The first three quartiles of $T(Z;\alpha)$ against x_1 for model 4

and

$$\hat{A} = \frac{1}{n}\sum_{i=1}^{n}\left[\alpha - I(T(Z_i;\alpha) \leq x_i'\hat{\gamma}(\alpha))\right]^2 x_i x_i',$$

where the function $F_n(\cdot)$ is as defined in Proposition 3.

Several preliminary experiments were performed to choose the value of c_n and to evaluate the sensitivity of the results to the choice of this smoothing parameter. The results of these experiments indicate that the performance of the t test is not very sensitive to the choice of this parameter. The results presented here were obtained setting $c_n = 0.5\ln(\ln(n))/\sqrt{n}$.

Table 1 presents the rejection frequencies of the null at the 1, 5 and 10 percent nominal sizes for all the cases considered. To facilitate their interpretation, each of the reported results was tested for compatibility with the nominal size using the usual asymptotic test for proportions. Overall, it is encouraging to find that the results are reasonably accurate. In fact, considering that the density of the data is discontinuous, that the distribution of the errors is far from being independent of the regressors, and that the models are often clearly misspecified, the results obtained are surprisingly accurate and compare very favourably, for example, with those reported by Horowitz (1992) for models estimated with the smoothed maximum score estimator. Of course, the rejection frequencies are closer to the nominal significance level for the larger sample sizes, but even for $n = 500$ the results are relatively reasonable.

5. AN EMPIRICAL ILLUSTRATION

This section illustrates the application of quantile regression for counts using the data set on demand for health care previously studied by Pohlmeier and Ulrich (1995) and by Santos Silva and Windmeijer (2001). These data consist of 5096 observations for employed individuals, and they are taken from the 1985 wave of the German Socioeconomic Panel (SOEP). In this data set the demand for health care services is measured by the number of visits to a general practitioner in the last quarter, and by the number of visits to a specialist (except gynaecology or paediatrics) in the same period. Here, we focus exclusively on the case of visits to specialists. The variables used in the analysis are described in Table A.1

Table 1: Simulation results for the four different models

	$n = 500$			$n = 1000$			$n = 2000$		
α	Reject. at 1%	Reject. at 5%	Reject. at 10%	Reject. at 1%	Reject. at 5%	Reject. at 10%	Reject. at 1%	Reject. at 5%	Reject. at 10%
Model 1: Poisson									
0.75	1.78*	6.76*	12.30*	1.82*	6.28*	11.64*	1.18	5.04	10.00
0.50	1.76*	6.54*	11.58*	1.56*	5.94*	10.58	1.20	5.02	9.90
0.25	1.90*	6.56*	11.64*	1.68*	5.60	10.62	1.18	5.54	10.28
Model 2: Negative binomial									
0.75	3.18*	8.32*	13.50*	2.12*	6.48*	11.46*	1.42*	6.26*	11.08*
0.50	2.14*	6.42*	11.24*	1.66*	6.26*	11.32*	1.02	5.56	10.86*
0.25	1.80*	6.54*	11.28*	1.26*	5.26	10.14	1.00	5.00	9.58
Model 3: Zero inflated Poisson									
0.75	1.94*	6.66*	11.50*	1.38*	6.32*	11.42*	1.02	4.88	10.14
0.50	2.02*	6.12*	10.88*	1.30*	5.34	10.58	1.54*	5.72*	10.48
0.25	1.08	4.68	8.40*	0.90	4.64	9.50	0.90	4.68	9.66
Model 4: Zero inflated negative binomial									
0.75	3.06*	7.94*	13.00*	2.00*	6.36*	11.26*	1.12	5.14	10.10
0.50	1.70*	6.72*	11.46*	1.28*	5.76*	10.48	1.12	5.00	10.02
0.25	1.14	4.10*	8.08*	0.84	4.14*	9.00*	0.78	4.00*	8.98*

* The result is not compatible with the nominal size at the 5% level.

in the appendix and correspond to those originally used by Pohlmeier and Ulrich (1995). For more detailed information on the sample and on the variables used, see Pohlmeier and Ulrich (1995).

Both Pohlmeier and Ulrich (1995) and Santos Silva and Windmeijer (2001) suggested probabilistic models to describe the demand for health care that take into consideration the fact that this demand is derived by two different decision processes: in a first stage the individual decides to seek medical care and, in a second stage, the individual and the health care provider decide on the total number of visits that are needed to complete the treatment. Pohlmeier and Ulrich (1995) argued that the parameters of the two decision processes can be separately estimated using a hurdle model. Santos Silva and Windmeijer (2001) studied the conditions under which it is possible to identify the two sets of parameters using only data on the total number of visits, and concluded that this can only be achieved under reasonably strong hypotheses. In any case, the fact that the two tails of the distribution are generated by two different processes means that it is often important to assess the effect of the covariates on different regions of the distribution. This point is clearly illustrated in Winkelmann's (2001) recent study of the effects of the German Health Care reform.

The methods used to study how the effects of the covariates vary across different regions of the distribution are generally based on complex and heavily parametrized models. Here, quantile regression is used to study how the impact of the covariates varies across different regions of the conditional distribution of interest. Of course, imposing less structure on the data, the results obtained with quantile regression are necessarily more impressionistic, but also less sensitive to incidental assumptions, than those obtained by models that attempt to fully describe the probabilistic structure of the data.

Given that the purpose of this section is purely to illustrate the application of the proposed methodology, we restrict the analysis to the first three conditional quartiles. Of course, in a more detailed analysis of the demand for health care using quantile regression it would be interesting to estimate other quantiles, especially on the upper tail of the distribution where the effect of the covariates changes more rapidly. For comparison, results for the conditional expectiles based on the asymmetric maximum likelihood (AML) estimator proposed by Efron (1992) are also presented. Given that the AML regression percentiles cannot be computed for values of α smaller that the proportion of zeros in the sample, and noting that in the case under study the sample has 3456 observations for which the individuals report zero visits to a specialist, the AML approach only allows the estimation of regression percentiles with $\alpha > 67.8$. Therefore, in this example, the only AML regression percentile that is comparable with the quartiles estimated by the quantile regression approach is the one leading to $\frac{1}{n}\sum_{i=1}^{n} I\left(Y_i \leq \exp\left(x_i'\hat{\beta}_w^{\text{AML}}\right)\right) = 0.75$, which is obtained with $w = 0.875$.

Table 2 presents the absolute frequencies of the estimated quantiles for the number of visits to specialists. Given that almost 70 percent of the individuals in the sample have zero visits, it is not surprising to find that the estimate of the first conditional quartile of the distribution of the number of visits is zero for 4955 individuals, which is more than 97 percent of the sample. Even the higher quartiles are flat at zero for a large proportion of the sample. In fact, the results in Table 2 show that the probability of having zero visits to a specialist during the observation period is at least 0.75 for about on third (1693/5096) of the individuals in the sample.

We now turn to a brief analysis of the effects of the covariates. Table 3 gives the parameter estimates for the first three quartiles, together with standard errors estimated as described in the simulation study of the previous section, as well as the estimates obtained with the AML estimator with $w = 0.875$. Given the non-linearity of the specification of $Q_Z(\alpha|x)$ defined by (2), the values of the parameter estimates give an incomplete picture of the effect of the covariates on the shape of the distribution. To facilitate the comparison of the effects of the regressors across the different models, Table 4 presents estimates of the semi-elasticities of $Q_Z(\alpha|x)$ with respect to the covariates, evaluated at the mean value of the continuous covariates and setting the dummy variables to zero. Defining $\bar{x}$ as the point where they are evaluated, the semi-elasticity of $Q_Z(\alpha|x)$ with respect to $\bar{x}_j$, the j-th element of $\bar{x}$, is computed as the appropriate derivative of $\ln(Q_Z(\alpha|x))$ for the continuous variables, or as $\left(\exp\left(\gamma_j(\alpha)\right) - 1\right)\left(1 - \alpha/Q_Z(\alpha|\bar{x})\right)$ for the dummies. For the AML regression the semi-elasticities are computed in a similar way.

Table 2: Frequencies of estimated quantiles for the number of visits to specialists

	0	1	2	3	4	5	6	7	8	9	10	11	≥12
$\widehat{Q_Y}(0.25\|x)$	4955	127	13	1	0	0	0	0	0	0	0	0	0
$\widehat{Q_Y}(0.50\|x)$	4171	721	134	39	21	6	1	1	1	1	0	0	0
$\widehat{Q_Y}(0.75\|x)$	1693	2250	453	221	131	88	64	42	30	22	13	13	76

Table 3: Estimation results for visits to specialists

	Quantile regression results						Efron's AML	
Regressors	$\hat{\gamma}(0.25)$	s.e.	$\hat{\gamma}(0.50)$	s.e.	$\hat{\gamma}(0.75)$	s.e.	$\hat{\beta}^{\text{AML}}_{\mathbf{0.875}}$	s.e.
Constant	−2.372	1.092	−2.157	0.868	−1.450	0.740	−0.378	0.505
Female	0.913	0.172	1.094	0.122	1.511	0.097	0.602	0.075
Single	−0.194	0.242	−0.308	0.196	−0.353	0.159	−0.342	0.105
Age	−0.908	0.533	−0.339	0.406	−0.213	0.350	−0.215	0.239
Age^2	0.936	0.649	0.306	0.492	0.136	0.438	0.232	0.288
Income	0.291	0.069	0.192	0.484	0.380	0.056	0.095	0.084
Chronic complaints	0.979	0.168	1.127	0.123	1.515	0.130	0.868	0.087
Private insurance	0.480	0.268	0.401	0.196	0.194	0.157	−0.016	0.130
Education	0.046	0.023	0.032	0.019	0.033	0.017	0.003	0.013
Heavy labour	−0.144	0.252	−0.002	0.160	0.006	0.126	−0.120	0.099
Stress	0.043	0.179	0.001	0.131	0.154	0.115	0.091	0.085
Variety on job	0.198	0.179	0.214	0.123	0.187	0.101	0.174	0.080
Self determined	0.156	0.185	0.195	0.124	0.151	0.104	−0.021	0.082
Control	0.332	0.199	0.030	0.155	0.184	0.136	0.118	0.097
Pop-0/5	−1.400	0.678	−0.737	0.220	−0.850	0.197	−0.627	0.137
Pop-5/20	−0.578	0.249	−0.689	0.173	−0.521	0.128	−0.305	0.108
Pop-20/100	−0.139	0.199	−0.136	0.139	−0.280	0.132	−0.260	0.088
Physicians density	4.654	0.999	2.172	0.858	1.066	0.780	0.954	0.640
Unemployment	0.015	0.055	0.015	0.035	−0.059	0.044	−0.033	0.028
Hospitalized	−0.004	0.249	0.369	0.210	0.277	0.175	0.265	0.130
Sick leave	0.895	0.189	0.757	0.155	0.921	0.136	0.725	0.105
Disability	0.212	0.276	0.523	0.200	0.579	0.169	0.218	0.122

Table 4: Semi-elasticities for visits to specialists

	1st Quartile		2nd Quartile		3rd Quartile		AML ($w = 0.875$)	
Regressors	S.-elast.	s.e.	S.-elast.	s.e.	S.-elast.	s.e.	S.-elast.	s.e.
FEMALE	0.227	0.057	0.325	0.052	0.734	0.090	0.826	0.101
SINGLE	−0.027	0.031	−0.043	0.024	−0.062	0.024	−0.289	0.089
AGE	0.998	0.659	0.343	0.564	0.181	0.650	1.637	2.063
INCOME	0.044	0.015	0.031	0.080	0.079	0.013	0.095	0.084
CHRONIC COMPLAINTS	0.253	0.075	0.341	0.070	0.737	0.127	1.381	0.134
PRIVATE INSURANCE	0.094	0.065	0.081	0.047	0.045	0.039	−0.016	0.129
EDUCATION	0.007	0.004	0.005	0.003	0.007	0.004	0.003	0.013
HEAVY LABOUR	−0.020	0.034	−0.000	0.026	0.001	0.026	−0.113	0.093
STRESS	0.007	0.028	0.000	0.021	0.035	0.029	0.095	0.089
VARIETY ON JOB	0.033	0.032	0.039	0.024	0.043	0.024	0.191	0.087
SELF DETERMINED	0.026	0.032	0.035	0.024	0.034	0.025	−0.021	0.081
CONTROL	0.060	0.040	0.005	0.026	0.042	0.034	0.125	0.103
POP-0/5	−0.115	0.036	−0.085	0.022	−0.119	0.023	−0.466	0.100
POP-5/20	−0.007	0.027	−0.081	0.020	−0.084	0.020	−0.263	0.092
POP-20/100	−0.020	0.028	−0.021	0.021	−0.051	0.023	−0.229	0.078
PHYSICIANS DENSITY	0.708	0.178	0.355	0.136	0.222	0.160	0.954	0.640
UNEMPLOYMENT	0.002	0.008	0.002	0.006	−0.012	0.009	−0.033	0.028
HOSPITALIZED	−0.001	0.038	0.073	0.050	0.066	0.049	0.303	0.148
SICK LEAVE	0.220	0.071	0.185	0.056	0.314	0.078	1.065	0.151
DISABILITY	0.036	0.052	0.112	0.057	0.163	0.061	0.244	0.136

Starting with the quantile regression results, it is noticeable that, although some covariates have very similar effects on the three estimated quartiles (e.g. EDUCATION and SICK LEAVE) there are also regressors whose effects vary substantially with α. A leading example of this is the variable PHYSICIANS DENSITY, whose effect is halved from the first to the second quartile, decreasing again substantially from the median to the third quartile. Considering that $\widehat{Q_Y}(0.25|x)$ is zero for most of the individuals in the sample, it can be concluded that this variable is especially important to determine whether or not the individual visits a specialist, but is much less important in explaining the length of the treatment, conditional on having at least one visit. This result is in accordance with the findings of Pohlmeier and Ulrich (1995) and Santos Silva and Windmeijer (2001).

In contradistinction, the effects of FEMALE and CHRONIC COMPLAINTS increase with α, and are statistically significant for the three quartiles. This means that the distribution of visits to specialists for individuals for which these variables equal one is displaced upwards and more spread-out. At first sight these results are at odds with the findings of Santos Silva and Windmeijer (2001) who concluded that these regressors had a positive effect on the decision to contact the specialist, but a negative coefficient (statistically insignificant for CHRONIC COMPLAINTS) in the model for the number of referrals. However, given that for most individuals in the sample the distribution of the number of visits has most of its probability mass concentrated at 0 and 1, the fact that the coefficients on these variables increase with α may just indicate a shift in the probability mass from 0 to 1, and this is not incompatible with the results previously reported.

Another conclusion of the results in Tables 3 and 4, which is in line with the results of Pohlmeier and Ulrich (1995) and Santos Silva and Windmeijer (2001), is that the regressors related to job characteristics (HEAVY LABOUR, STRESS, VARIETY ON JOB, SELF DETERMINED and CONTROL) are not individually significant in any of the quantiles estimated.

Using $w = 0.875$, the AML estimator leads to 75 percent of negative residuals and therefore $\hat{\beta}_{0.875}^{\text{AML}}$ is the vector of parameters for the 75-th AML regression percentile, as defined by Efron (1992). For some covariates, this AML estimator leads to results that are reasonably close to those obtained for the third quartile. This is the case of the variables Female, Income and Education. Moreover, as in the three quartiles, most of the covariates related to job characteristics (Heavy labour, Stress, Self determined and Control) are not individually significant, and the dummies related to the number of inhabitants in the place of residence are all individually significant. However, there are some notable differences between the two sets of results. A clear example of this is provided by the semi-elasticities corresponding to Chronic complaints and Sick leave, which are statistically significant both in the AML regression and in the third quartile, but are much larger in the AML regression. Moreover, the variable Disability is highly significant in the third quartile but statistically insignificant in this AML regression. On the other hand, Hospitalized is statistically significant in the AML regression, but not in any of the quantile regressions. Therefore, using the expectiles estimated by the AML approach to approximate the regression quantiles can be somewhat unsatisfactory.

To sum up, this example makes clear that in count data models it is interesting to study not only how the location of the conditional distribution changes with the regressors, but also to analyse how the shape of the distribution is affected by the covariates. Fully parametric models, like those used by Pohlmeier and Ulrich (1995) and Santos Silva and Windmeijer (2001) to study these data, achieve this by modelling the effect of the covariates on a few key aspects of the distribution. However, by looking at the estimation results from that sort of models, it is not always obvious how the shape of the conditional distribution is affected by the covariates. The quantile regression method used here provides a more graphical description of the effect of the regressors on the shape of the conditional distribution of interest. Of course, quantile regressions are not a substitute for carefully constructed probabilistic models. However, they can be a valuable tool which helps in the construction and understanding of more complex models.

6. CONCLUDING REMARKS

Direct estimation of quantile regressions for count data is not feasible due to the combination of the discreteness of the data with the non-differentiability of the sample objective function defining the estimator. We propose the application of quantile regression to jittered count data as a way to make inference about relevant aspects of the conditional quantiles of the counts. This is possible because the quantiles of the randomly perturbed data have a one-to-one relation with the quantiles of the original data. Despite the discontinuity of the conditional density of the jittered counts, the regression quantile estimator has the usual normal distribution and standard inference methods are asymptotically valid. The simulation results presented in Section 4 are promising in that they show that the approximation to the asymptotic distribution is reasonable for moderate sample sizes. Furthermore, using a well known data set on demand for health care, it was shown that the proposed methodology is very easy to implement and leads to interesting results.

Naturally, quantile regression cannot replace the more structured and well proven models for count data analysis. However, it can be a valuable additional tool that can help to understand how the regressors affect, not only the location of the conditional distribution, but also the shape of the entire distribution.

The results presented here can be extended in a number of ways. For example, the functional form of the conditional regression quantiles can be generalized by considering members of the inverse Box-Cox (1964) family. Another possible development of the technique presented here is the use of simulation methods to integrate-out the added uniform noise to increase the efficiency of the estimator. These developments are currently being investigated.

APPENDIX

A.1. Proof of Proposition 1

Let $p_i(x)$, $i = 0, 1, \cdots$ denote the conditional probability function of Y given $X = x$, that is

$$Y = i \text{ with probability } p_i(x),\ i = 0, 1, \cdots$$

and $p_i(x) > 0$ and $\sum_{i=0}^{\infty} p_i(x) = 1$.

Standard convolution arguments imply that the conditional distribution of $Z = Y + U$ (U being uniform in $[0, 1)$ and independent of X)is

$$\begin{aligned} F_{Z|X}(z|x) \equiv P(Z \leq z|x) &= \sum_{i=0}^{\infty} p_i(x) P(U \leq z - i) \\ &= \sum_{i=0}^{\infty} p_i(x) \min(z - i, 1)^+ \end{aligned}$$

where $\min(a)^+ = 0$ whenever $a < 0$.

Let $i_\alpha \equiv Q_Y(\alpha|x)$, i.e., the value in the support satisfying,

$$\sum_{i=0}^{i_\alpha} p_i(x) \geq \alpha \text{ and } \sum_{i=0}^{i_\alpha - 1} p_i(x) < \alpha.$$

Denote also by z_α the α-th conditional quantile of Z given x. Then, omitting the dependence on x to simplify notation,

$$\alpha = F_{Z|X}(z_\alpha) = \sum_{i=0}^{\lceil z_\alpha - 1 \rceil - 1} p_i + p_{\lceil z_\alpha - 1 \rceil}[z_\alpha - \lceil z_\alpha - 1 \rceil].$$

Therefore,

$$\sum_{i=0}^{\lceil z_\alpha - 1 \rceil - 1} p_i(x) < \alpha$$

and

$$\sum_{i=0}^{\lceil z_\alpha - 1 \rceil} p_i(x) \geq \alpha$$

with equality when z_α is integer. The result follows from the definition of i_α. Q.E.D.

A.2. Proof of Proposition 2

The results follow as in Pollard (1991), once we show that one can do without the continuity of $f_{T|x}(\cdot|\cdot)$ at $x'\gamma(\alpha)$. The strategy of Pollard's proof is to develop a quadratic approximation to the sample objective function defining $\hat{\gamma}(\alpha)$ whose minimizing value has a limiting normal distribution and is asymptotically equivalent to $\hat{\gamma}(\alpha)$. It that process, the "continuity assumption" is only required to ensure that the "limiting" objective function

$$M(t;x) \equiv E(\rho_\alpha(\epsilon_\alpha - t)|x)$$

with $\epsilon_\alpha \equiv T(Z;\alpha) - X'\gamma(\alpha)$ (note that, $Q_{\epsilon_\alpha}(\alpha|x) = 0$), possesses a second order Taylor expansion around $t = 0$.

Assumption (A2a) implies that $f_{\epsilon|x}(0|x)$ is continuous for almost every x (a.e.-x) and, consequently, $M''(t;x)$ is finite at $t = 0$, and equals

$$\frac{\partial \int I(e < t) f_{\epsilon|x}(e|x)de}{\partial t} = f_{\epsilon|x}(0|x), \text{ a.e.-}x.$$

Therefore, the warranted second order Taylor expansion

$$M(t;x) = M(0;x) + \frac{1}{2} f_{\epsilon|x}(0|x)t^2 + o(t^2),$$

exists a.e.-x (c.f. Pollard, 1991, equation (1)). Q.E.D.

A.3. Proof of Proposition 3

Let,

$$\tilde{D}_n \equiv (1/n) \sum_{i=1}^{n} \omega_i x_i x_i'$$

with

$$\omega_i \equiv T^{(1)}(x_i'\gamma(\alpha)) I\{F_n(Q_{Z_i}(\alpha)) \leq Z_i < F_n(Q_{Z_i}(\alpha) + 1)\},$$

where $Q_{Z_i}(\alpha)$ is used as shorthand for $Q_{Z_i}(\alpha|x)$.

Notice that, as n passes to ∞, $F_n(x) \to \lfloor x \rfloor$. Thus, by dominated convergence, as $n \to \infty$,

$$E[\omega_i x_i x_i'] \to E\left[T^{(1)}(x_i'\gamma(\alpha)) I\{\lfloor Q_{Z_i}(\alpha) \rfloor \leq Z_i < \lfloor Q_{Z_i}(\alpha) + 1 \rfloor\} x_i x_i'\right]$$

which, a simple change of variable shows to equal

$$E\left[f_{T|x}(x_i'\gamma(\alpha))x_ix_i'\right].$$

The law of large numbers then yields

$$\tilde{D}_n \xrightarrow{P} D.$$

It remains to prove that $\hat{D}_n - \tilde{D}_n \xrightarrow{P} 0$. First notice that

$$|\hat{\omega}_i - \omega_i| \leq B_{1ni} + B_{2ni}$$

with

$$B_{1ni} \equiv \left|T^{(1)}(x_i'\hat{\gamma}(\alpha)) - T^{(1)}(x_i'\gamma(\alpha))\right|$$

and,

$$\begin{aligned} B_{2ni} \equiv\ & |T^{(1)}(x_i'\gamma(\alpha))|\Big(\Big|I\left\{F_n\left(\hat{Q}_{Z_i}(\alpha)\right) \leq Z_i < F_n\left(\hat{Q}_{Z_i}(\alpha)+1\right)\right\} \\ & -I\left\{F_n\left(Q_{Z_i}(\alpha)\right) \leq Z_i < F_n\left(Q_{Z_i}(\alpha)+1\right)\right\}\Big|\Big). \end{aligned}$$

Therefore $||\hat{D}_n - \tilde{D}_n||$ can be bounded above by the sum of two terms the first being,

$$(1/n)\sum \|x_i\|^2 B_{1ni} = o_p(1).$$

This order of magnitude follows from a mean value expansion of B_{1ni} and the fact that for any $\hat{\gamma}(\alpha) \xrightarrow{P} \gamma(\alpha)$

$$(1/n)\sum \|x_i\|^2 T^{(2)}(x_i'\hat{\gamma}(\alpha)) \xrightarrow{P} E\left\{\|x_i\|^2 T^{(2)}(x_i'\gamma(\alpha))\right\} = O_p(1)$$

(Newey and McFadden, 1994, Lemma 4.3) and that

$$\sup_{1\leq i\leq n} |x_i'(\hat{\gamma}(\alpha) - \gamma(\alpha))| = o_p(1).$$

The last result follows as $E\left[\|x_i\|^2\right] < \infty$ implies that $\sup_{1\leq i\leq n}\|x_i\| = o_p(\sqrt{n})$.

Let us now turn to the second term of the upper bound of $\left\|\hat{D}_n - \tilde{D}_n\right\|$. This term, $(1/n)\sum \|x_i\|^2 B_{2ni}$, is bounded above by,

$$(1/n)\sum \|x_i\|^2 \left|T^{(1)}\left(x_i'\gamma(\alpha)\right)\right| I\left\{|Z_i - F_n(Q_{Z_i}(\alpha))| \leq \left|F_n\left(\hat{Q}_{Z_i}(\alpha)\right) - F_n(Q_{Z_i}(\alpha))\right|\right\}$$
$$+ (1/n)\sum \|x_i\|^2 \left|T^{(1)}\left(x_i'\gamma(\alpha)\right)\right| I\Big\{|Z_i - F_n(Q_{Z_i}(\alpha)+1)| \leq \Big|F_n\left(\hat{Q}_{Z_i}(\alpha)+1\right)$$
$$- F_n(Q_{Z_i}(\alpha)+1)\Big|\Big\}$$

We will show that the first of these terms is $o_p(1)$. The second is analogous. To simplify notation, put $\Delta_i \equiv F_n\left(\hat{Q}_{Z_i}(\alpha)\right) - F_n(Q_{Z_i}(\alpha))$. For any $\eta > 0$ and any ϵ,

$$P[(1/n)\sum \|x_i\|^2 \left|T^{(1)}\left(x_i'\gamma(\alpha)\right)\right| I\{|Z_i - F_n(Q_{Z_i}(\alpha))| \leq |\Delta_i|\} > \eta]$$
$$\leq P[(1/n)\sum \|x_i\|^2 \left|T^{(1)}\left(x_i'\gamma(\alpha)\right)\right| I\{|Z_i - F_n(Q_{Z_i}(\alpha))| \leq \epsilon\} > \eta]$$
$$+P[\sup_{1\leq i\leq n} |\Delta_i| > \epsilon]$$

Markov's inequality, and the fact that, conditional on x,

$$E[I\{|Z_i - F_n(Q_{Z_i}(\alpha))| \leq \epsilon\}] \leq 2\epsilon,$$

show that the first term goes to 0 with ϵ.

To complete the proof we will show that $\Delta_i = o_p(1)$ uniformly in i. First notice that $E\left[\sup_{\|\gamma-\gamma(\alpha)\|\leq\delta} \left|T^{(1)}\left(x_i'\gamma\right)\right| \|x_i\|^2\right] < \infty$, for some $\delta > 0$ (A5) implies that (e.g., Pollard, 1991), $\sup_{\|\gamma-\gamma(\alpha)\|\leq\delta} \left|T^{(1)}\left(x_i'\gamma\right)\right| \|x_i\| = o_p\left(\sqrt{n}\right)$. Therefore, by a mean value expansion, $\sup_{1\leq i\leq n} |\hat{Q}_{Z_i}(\alpha) - Q_{Z_i}(\alpha)| = o_p(1)$. Assumption (A6) states that this convergence must be faster than c_n, that is, for n sufficiently large,

$$P\left[\sup_{1\leq i\leq n} \left|\hat{Q}_{Z_i}(\alpha) - Q_{Z_i}(\alpha)\right| < \epsilon c_n\right] \geq 1-\delta$$

for any $\delta > 0$ and $\epsilon > 0$. But, from the definition of $F_n(\cdot)$, for any n and x_0, $|x - x_0| < \epsilon c_n \Rightarrow |F_n(x) - F_n(x_0)| < \epsilon$. Consequently, for arbitrary ϵ and δ

$$P\left[\left|F_n\left(\hat{Q}_{Z_i}(\alpha)\right) - F_n(Q_{Z_i}(\alpha))\right| < \epsilon,\ \forall i \leq n\right] \geq 1-\delta$$

as n passes to ∞. Q.E.D.

A.4. Description of Variables

Table A.1: Description of Variables

Female	1 if female
Single	1 if single
Age	age in decades
Income	net monthly household income
Chronic complaints	1 if has chronic complaints for at least one year
Private insurance	1 if had private medical insurance in the previous year
Education	number of years in education after the age of sixteen
Heavy labour	1 if has a job in which physically heavy labour is required
Stress	1 if has a job with high level of stress
Variety on job	1 if job offers a lot of variety
Self determined	1 if has a job where the individual can plan and carry out job tasks
Control	1 if has a job where work performance is strictly controlled
Pop-0/5	1 if place of residence has less than 5000 inhabitants
Pop-5/20	1 if place of residence has between 5000 and 20.000 inhabitants
Pop-20/100	1 if place of residence has between 20.000 and 100.000 residents
Physicians density	number of physicians per 100.000 inhabitants in the place of residence
Unemployment	number of months of unemployment in the previous year
Hospitalized	1 if was more than seven days hospitalized in the previous year
Sick leave	1 if missed more than 14 work days due to illness in the previous year
Disability	1 if the degree of disability is greater than 20%

REFERENCES

Anscombe, F.J. (1948). "Validity of Comparative Experiments" (with discussion), *Journal of the Royal Statistical Society A*, 109, 181-211.

Box, G.E.P. and Cox, D.R. (1964). "An Analysis of Transformations". *Journal of the Royal Statistical Society B*, 26, 211-243.

Cameron, A. C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data*, Cambridge University Press, Cambridge.

Chamberlain, G. (1994). "Quantile Regression, Censoring and the Structure of Wages", *in* Sims, C.A. (ed.) *Advances in Econometrics*, Cambridge University Press, Cambridge.

Efron, B. (1992). "Poisson Overdispersion Estimates Based on the Method of Asymmetric Maximum Likelihood", *Journal of the American Statistical Association*, 87, 98-107.

Fitzenberger, B.; Koenker, R. and Machado, J.A.F. (eds.) (2001). *Special Issue on Economic Applications of Quantile Regression*, *Empirical Economics*, 26, 1-324.

Fortin, N. and Lemieux (2000). "Rank Regressions, Wage Distributions, and the Gender Gap", *The Journal of Human Resources*, 33, 610-643.

Gourieroux, C.; Monfort, A. and Trognon, A. (1984). "Pseudo Maximum Likelihood Methods: Applications to Poisson Models", *Econometrica*, 52, 701-720.

Hall, B. H. and Cummins, C. (1999). *Time Series Processor Version 4.5 User's Guide*, TSP International, Palo-Alto (CA).

Hausman, J.; Hall, B.H. and Griliches, Z. (1984). "Econometric Models for Count Data With an Application to the Patents-R&D Relationship", *Econometrica*, 52, 909-938.

Horowitz, J.L. (1992). "A Smooth Maximum Score Estimator for the Binary Response Model", *Econometrica*, 60, 505-531.

Horowitz, J.L. (1998). *Semiparametric Methods in Econometrics* . Springer-Verlag, New York (NY).

Huber, P.J. (1981). *Robust Statistics*, John Wiley & Sons, New York (NY).

John, S. (1982). "The three-parameter two piece normal family of distributions and its fitting", *Communications in Statistics-Theory Methods*, 11, 879-885.

Kim, T.-H. and White, H. (2002). "Estimation, Inference and Specification Testing and Possibly Misspecified Quantile Regression", *Advances in Econometrics* (forthcoming).

Knight, K. (1998). "Asymptotics for the L_1 regression estimators under general conditions",*The Annals of Statistics*, 26, 755-770.

Koenker, R. (1992). "When Are Expectiles Percentiles?", Problem 92.3.2, *Econometric Theory*, 8, 423.

Koenker, R. (1993). "When Are Expectiles Percentiles?", Solution 92.3.2, *Econometric Theory*, 9, 526.

Koenker, R. and Bassett Jr., G.S. (1978). "Regression Quantiles", *Econometrica*, 46, 33-50.

Koenker, R. and Bilias, Y. (2001). "Quantile Regression for Duration Data: A Reappraisal of the Pennsylvania Reemployment Bonus Experiments", *Empirical Economics*, 26, 199-220.

Koenker, R. and Geling, O. (2001). "Reappraising Medfly Longevity: A Quantile Regression Survival Analysis", *Journal of the American Statistical Association*, 96, 458-468.

Koenker, R. and Machado, J. (1999). "Goodness of Fit and Related Inference Processes for Quantile Regression", *Journal of the American Statistical Association*, 94, 1296-1310.

Lee, M.-j. (1992). "Median Regression for Ordered Discrete Response", *Journal of Econometrics*, 51, 59-77.

Machado, J.A.F. and Portugal, P. (2002). *Quantile Regression Analysis of Transition Data*. Banco de Portugal, mimeo.

Manski, C.F. (1975). "Maximum Score Estimation of the Stochastic Utility Model of Choice", *Journal of Econometrics*, 3, 205-228.

Manski, C.F. (1985). "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator", *Journal of Econometrics*, 27, 313-333.

Melenberg, B. and van Soest, A. (1994). *Parametric and Semiparametric Estimation of Ordered Response Models with an Application to Satisfaction with Life*, Tilburg University, mimeo.

Melenberg, B. and van Soest, A. (1996). "Measuring the Costs of Children: Parametric and Semiparametric Estimators", *Statistica Neerlandica*, 50, 171-192.

Mullahy, J. (1986). "Specification and Testing in Some Modified Count Data Models", *Journal of Econometrics*, 33, 341-365.

Newey, W. and McFadden, D. (1994). "Large Sample Estimation and Hypothesis Testing". Chapter 36 *in* Engle, R. and McFadden, D. (eds.), *Handbook of Econometrics*, Vol. IV, North-Holland, New York.

Newey, W.K. and Powell, J.L. (1987). "Asymmetric Least Squares Estimation and Testing", *Econometrica*, 55, 819-847.

Pearson, E.S. (1950). "On Questions Raised by the Combination of Tests Based on Discontinuous Distributions", *Biometrika*, 37, 383-398.

Pohlmeier, W. and Ulrich, V. (1995). "An Econometric Model of the Two-Part Decision Process in the Demand for Health", *Journal of Human Resources*, 30, 339-361.

Pollard, D. (1984). *Convergence of Stochastic Process*, Springer-Verlag, New York (NY).

Pollard, D. (1991). "Asymptotics for Least Absolute Deviation Regression Estimators", *Econometric Theory*, 7, 186-199.

Powell, J.L. (1984). "Least Absolute Deviation Estimation for the Censored Regression Model", *Journal of Econometrics*, 25, 303-325.

Powell, J.L. (1986). "Censored Regression Quantiles", *Journal of Econometrics*, 32, 143-155.

Simpson, D.G.; Carroll R.J. and Ruppert, D. (1987). "M-Estimation for Discrete Data: Asymptotic Distribution Theory and Implications", *The Annals of Statistics*, 15, 657-669.

Santos Silva, J.M.C. and Windmeijer, F. (2001). "Two-Part Multiple Spell Models for Health Care Demand", *Journal of Econometrics*, 104, 67-89.

Stevens, W.L. (1950). "Fiducial Limits of the Parameter of a Discontinuous Distribution", *Biometrika*, 37, 117-129.

White, H. (1980). "Using Least Squares to Approximate Unknown Regression Functions". *International Economic Review*, 21, 149-170.

Winkelmann, R. (2000). *Econometric Analysis of Count Data*, 3rd ed., Springer-Verlag, Berlin.

Winkelmann, R. (2001). *Health Care Reform and the Number of Doctor Visits - An Econometric Analysis*, IZA Discussion Paper No. 317.