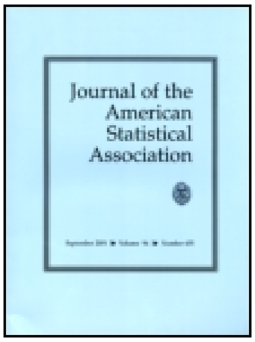

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Quantiles for Counts

José A. F Machado & J. M. C. Santos Silva

To cite this article: José A. F Machado & J. M. C. Santos Silva (2005) Quantiles for Counts, Journal of the American Statistical Association, 100:472, 1226-1237, DOI: 10.1198/016214505000000330

To link to this article: https://doi.org/10.1198/016214505000000330

Published online: 01 Jan 2012.

Quantiles for Counts

José A. F. MACHADO and J. M. C. SANTOS SILVA

This article studies the estimation of conditional quantiles of counts. Given the discreteness of the data, some smoothness must be artificially imposed on the problem. 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 estimators are illustrated by simulations and an application.

KEY WORDS: Jittering; Quantile regression; Smoothing.

1. INTRODUCTION

Since the publication of Jorgenson's (1961) seminal work on regression analysis of a Poisson process, the research on count data regression has gained great popularity, being now the subject of two monographs (Cameron and Trivedi 1998; Winkelmann 2003) 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 Nelder and Wedderburn (1972) and Gourieroux, Monfort, and Trognon (1984) (see also McCullagh and Nelder 1989 and the references therein) and concentrates on the semiparametric estimation of the conditional mean of the count variate based on the pseudolikelihood 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 that are often of interest in applied research. In fact, the pseudolikelihood approach permits only 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 Hausman, Hall, and Griliches (1984), several fully parametric probabilistic models have been developed to describe particular features of count datasets often found in applications. These models completely describe the conditional distribution of interest and thus 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.

The estimation of conditional quantile functions or quantile regressions (QRs) was originally advocated and studied by Koenker and Bassett (1978) and is becoming increasingly popular. [See, e.g., 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 exists a rapidly expanding set of results for other setups. The seminal departure was the work of Manski (1975, 1985) on median regression for binary and multinomial models, later extended by Horowitz (1992). Powell (1984, 1986) studied QR for censored data, and Lee (1992) analyzed median estimation of ordered discrete responses. More recently, QR was applied to the study of duration or transition data by Koenker and Geling (2001), Koenker and Bilias (2001), and Machado and Portugal (2002).

This article studies the possibility of estimating conditional quantiles of count data. The estimation of conditional quantiles requires assumptions that are comparable to those underlying the pseudolikelihood approach, while allowing the researcher to obtain most of the results that otherwise can be obtained only using more structured models. In particular, using QR, 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 estimating conditional quantiles for counts stems from the conjunction of a nondifferentiable sample objective function with a discrete dependent variable. Because there are points of positive mass, the nonsmoothness 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. This is the problem faced in the estimation of conditional quantiles for binary data, which ultimately yields the nonstandard rate of convergence of the maximum score estimator (see Manski 1975, 1985). Huber (1981, pp. 50–51) discussed the interplay between smoothness of the objective function and of the density of the data (see also Simpson, Carroll, and Ruppert 1987; Knight 1998).

To be able to apply quantile regression to counts, some degree of smoothness must be artificially imposed on the problem. Two different ways of doing this, which are briefly surveyed in the next section, have been available for the past 10 years. But these methods are not particularly attractive, and to our knowledge, they have never been used in practice and were not referred to in the monographs of Cameron and Trivedi (1998) and Winkelmann (2003).

The approach explored in this article is based on the artificial smoothing of the data using a specific form of jittering introduced by Stevens (1950) in a different context. The necessary smoothness is achieved by adding a uniformly distributed noise to the count variable. In this way we construct a continuous variable with conditional quantiles that have a one-to-one relationship with the conditional quantiles of the counts, and

José A. F. Machado is Professor, Faculdade de Economia, Universidade NOVA de Lisboa, 1099-03 Lisboa, Portugal (E-mail: *jafm@fe.unl.pt*). J. M. C. Santos Silva is Professor, Department of Mathematics, ISEG/Universidade Técnica de Lisboa and CEMAPRE, 1200-781 Lisboa, Portugal (E-mail: *jmcss@iseg.utl.pt*). The authors thank the editor, an associate editor, and two referees for their important and constructive comments and suggestions. They also thank L. Catela Nunes, J. F. Correia Guedes, Russell Davidson, Josef Falkinger, M.-J. Lee, Paulo Parente, and Rainer Winkelmann for helpful discussions and suggestions. The usual disclaimer applies. The data used in Section 5 were kindly made available by Winfried Pohlmeier and are from the public-use version of the German Socio-Economic Panel Study. The authors thank the Deutsches Institut für Wirtschaftsforschung for authorizing the use of these data. This study was supported in part by the Fundação para a Ciência e Tecnologia, program POCTI, partially funded by FEDER. Machado was also supported by a NOVA FORUM research grant.

Journal of the American Statistical Association
December 2005, Vol. 100, No. 472, Theory and Methods
DOI 10.1198/016214505000000330

use this artificially constructed continuous variable as a base for inference.

The remainder of the article is organized as follows. Section 2 overviews the two approaches currently available for the estimation of QRs with count data, and Section 3 presents the new solution to this problem. Sections 4 and 5 evaluate the proposed approach using simulation experiments and a well-known dataset. Finally, Section 6 presents some concluding remarks. Proofs of all theorems are included in the Appendix.

2. EXISTING ALTERNATIVES

Consider the random variables Y and $\mathbf{X}$, and let $Q_Y(\alpha|\mathbf{x})$ denote the 100αth quantile of the conditional distribution of Y given $\mathbf{X}=\mathbf{x}$; recall that the 100αth quantile of Y given $\mathbf{x}$ is defined by $Q_Y(\alpha|\mathbf{x}) = \min\{\eta|P(Y \le \eta|\mathbf{x}) \ge \alpha\}$. Koenker and Bassett (1978) gave sufficient conditions for asymptotically valid inference on the parameters of $Q_Y(\alpha|\mathbf{x})$. Among these conditions, the conditional probability density function $f(Y|\mathbf{x})$ is required to be continuous and positive at $Q_Y(\alpha|\mathbf{x})$. If Y results from a count, then its support is the set of the nonnegative integers, and those sufficient conditions are not satisfied.

A possible approach to estimating quantiles for counts is to view the data as the discretized result of a continuous underlying process. Then, as in the binary choice case, the functional form of the conditional median 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. Although we are not aware of such 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 that of Horowitz (1992), as described by Melenberg and van Soest (1994, 1996) (see also Horowitz 1998). This entails replacing all of the indicator functions with integrated kernels. But the final smooth objective function depends on the specification of smoothing parameters for which there are no optimal selection rules at present. Moreover, for typical kernels, this estimator does not achieve the usual $\sqrt{n}$ convergence rate, and its implementation is still computationally very expensive, which may explain why it is seldom used in practice. Finally, by its own nature, this approach lacks an interpretation in terms of an underlying count process, which may be an unappealing feature in some situations.

An alternative approach to modeling count data was pioneered by Efron (1992), who proposed the asymmetric maximum likelihood estimator (AML). This ingenious estimator requires assumptions that are comparable to those underlying the pseudolikelihood approach and can be interpreted as resulting from smoothing the objective function defining the QR estimator. The AML estimates conditional location functions for count data that are akin to the conditional expectiles proposed by Newey and Powell (1987) for the linear model. In a count data context, the main advantage of the expectiles is that they are easy to estimate, because they are not restricted to be integers, and asymptotically valid inference can be performed using standard methods. However, strictly speaking, this method does not permit the estimation of conditional quantiles for counts, and, as in the case of the linear model (see Koenker 1992, 1993), the estimated location measures are difficult to interpret.

Using the subscript i to denote realizations of random variables for the ith sample observation, the AML estimator is defined by

$$\hat{\boldsymbol{\beta}}_w^{\mathrm{AML}} = \arg\max_{\mathbf{b}} \sum_{i=1}^{n} \bigl(Y_i \mathbf{x}_i'\mathbf{b} - \exp(\mathbf{x}_i'\mathbf{b}) - \ln(Y_i!)\bigr) w^{I(Y_i > \exp(\mathbf{x}_i'\mathbf{b}))},$$

where $w > 0$ and $I(A)$ is the usual indicator function that equals 1 when A is true and 0 otherwise. For $w = 1$, this estimator is identical to the Poisson pseudolikelihood estimator. Efron (1992) defined that the 100αth AML regression percentile is obtained for w such that $\alpha = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \le \exp(\mathbf{x}_i'\hat{\boldsymbol{\beta}}_w^{\mathrm{AML}}))$. Thus, because $I(Y_i \le \exp(\mathbf{x}_i'\hat{\boldsymbol{\beta}}_w^{\mathrm{AML}}))$ is necessarily equal to 1 when $Y_i = 0$, the AML regression percentiles cannot be computed for values of α smaller that the proportion of 0's in the sample.

3. JITTERING

3.1 Quantile Functions

The main problem with the estimation of QR when Y results from counts is that because Y has a discrete distribution, $Q_Y(\alpha|\mathbf{x})$ 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 random variable independent of Y and $\mathbf{X}$ uniformly distributed in the interval [0, 1), leading to $Z = Y + U$. This approach uses a specific form of jittering proposed by Stevens (1950) (see also Pearson 1950) to introduce smoothness into the problem, leading to a conditional quantile function that is continuous in α. Indeed, it is possible to show that

$$Q_Z(\alpha|\mathbf{x}) = Q_Y(\alpha|\mathbf{x}) + \frac{\alpha - \sum_{y=0}^{Q_Y(\alpha|\mathbf{x})-1} \Pr(Y = y|\mathbf{x})}{\Pr(Y = Q_Y(\alpha|\mathbf{x})|\mathbf{x})}. \tag{1}$$

Therefore, continuity is achieved by interpolating each jump in the conditional quantile function of the counts using an integrated kernel, much in the same way as Horowitz (1992) smoothed the conditional median of binary data. The difference is that here the uniform distribution is used because of the important historical precedent and because in this case the uniform distribution allows important algebraic and computational simplifications. However, using the uniform noise to jitter the data is by no means a necessity. The smoothing noise may be generated by any continuous distribution with support on [0, 1) and a density bounded away from 0; for instance, any member of the beta family with density bounded away from 0 could be used. This issue is retaken in Section 3.5.

3.2 Quantile Regression With Jittered Data

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. The main idea is that, except for the set of $\mathbf{x}$ for which $Q_Z(\alpha|\mathbf{x})$ is an integer, conditions for standard inference will hold. However, this set will have measure 0 if one assumes that there exists at least one continuously distributed covariate and that the conditional quantiles of Z, $Q_Z(\alpha|\mathbf{x})$, are measurable functions of that covariate.

The asymptotic distribution of the 100αth QR estimator is derived under the following assumptions:

(A1) Y is a discrete random variable with support in $\mathbb{N}_0$, the set of the nonnegative integers, and $\mathbf{X}$ is a random vector in $\mathbb{R}^k$; the conditional probability function of Y given $\mathbf{X}$ at $Q_Y(\alpha|\mathbf{x})$, $f_{Y|\mathbf{x}}(Q_Y(\alpha|\mathbf{x}))$, is uniformly bounded away from 0 for almost every realization of $\mathbf{X}$.

(A2) The regressors $\mathbf{X}$ are such that:

(a) $E(\mathbf{X}\mathbf{X}')$ is finite and nonsingular and

(b) $\mathbf{X}' = (X_1, \dots, X_k)$ can be partitioned as $(\mathbf{X}^{(d)\prime}\ \mathbf{X}^{(c)\prime})$ with $X_1^{(d)} = 1$ and $\mathbf{X}^{(c)} \in \mathbb{R}^{k_c}$, $1 \le k_c \le k-1$, satisfying $P(\mathbf{X}^{(c)} \in C) = 0$ for any countable subset C of $\mathbb{R}^{k_c}$.

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

$$Q_{T(Z;\alpha)}(\alpha|\mathbf{x}) = \mathbf{x}'\boldsymbol{\gamma}(\alpha) \quad \text{for } \alpha \in (0,1), \tag{2}$$

and $\boldsymbol{\gamma}(\alpha) \in \Gamma$, a compact subset of $\mathbb{R}^k$. Furthermore, if $\boldsymbol{\gamma}^{(c)}(\alpha)$ denotes the components of $\boldsymbol{\gamma}(\alpha)$ corresponding to the continuous covariates $\mathbf{X}^{(c)}$, then $\boldsymbol{\gamma}^{(c)}(\alpha) \neq \mathbf{0}$.

Most of these assumptions are standard in the QR 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 $\mathbf{x}$ is $\mathbb{N}_0$. Defining $T^{-1}(\cdot)$ as the inverse of the transformation $T(Z;\alpha)$, assumptions (A2b) and (A3) ensure that $P(T^{-1}(\mathbf{x}'\boldsymbol{\gamma}(\alpha)) \in \mathbb{N}_0) = 0$ and, consequently, for almost every realization of $\mathbf{X}$, the conditional density of the regressand at the quantile of interest will be continuous. Assumption (A3) restricts the conditional quantile functions to be single-index models of the form $Q_Z(\alpha|\mathbf{x}) = T^{-1}(\mathbf{x}'\boldsymbol{\gamma}(\alpha), \alpha)$. But the model is quite flexible, because the transformation may vary from quantile to quantile (see, e.g., Machado and Mata 2000 for an application). Alternatively, the whole procedure may be regarded as fitting the "best" linear quantile predictors of $T(Z;\alpha)$ in the sense of Angrist, Chernozhukov, and Fernandez-Val (2004). The specification $T^{-1}(\mathbf{x}'\boldsymbol{\gamma}(\alpha), \alpha) = \alpha + \exp(\mathbf{x}'\boldsymbol{\gamma}(\alpha))$ is proposed in Section 3.6. The results given in Sections 4 and 5 reveal that this specification works as a reasonable approximation in many situations.

We now present the main result of the article.

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

$$\min_{\mathbf{c} \in \mathbb{R}^k} \sum_{i=1}^n \rho_\alpha\big(T(z_i;\alpha) - \mathbf{x}_i'\mathbf{c}\big),$$

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

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

with

$$\mathbf{A} = \alpha(1-\alpha)E(\mathbf{X}\mathbf{X}') \qquad \text{and} \qquad \mathbf{D} = E\big(f_T\big(\mathbf{X}'\boldsymbol{\gamma}(\alpha)|\mathbf{X}\big)\mathbf{X}\mathbf{X}'\big),$$

where $f_T(\cdot|\cdot)$ denotes the conditional density of $T(Z;\alpha)$ given $\mathbf{x}$.

3.3 Identification of Partial Effects on the Conditional Quantiles of the Counts

Theorem 1 shows that, using conventional methods, it is possible to perform inferences about $Q_Z(\alpha|\mathbf{x})$, the conditional quantile functions of the smoothed data. The object of ultimate interest is, however, the quantile function of the count data Y. The next theorem, which is proved in Appendix A.2 for a general noise U with support on the interval $[0, 1)$, relates the quantiles of the two random variables.

Theorem 2. $Q_Y(\alpha|\mathbf{x}) = \lceil Q_Z(\alpha|\mathbf{x}) - 1\rceil$, where $\lceil a\rceil$ denotes the ceiling function that returns the smallest integer greater than, or equal to, a.

Note that when a is an integer, $\lceil a - 1\rceil$ is not the integer part of a. Consequently, $Q_Y(\alpha|\mathbf{x})$ cannot simply be defined as the integer part of $Q_Z(\alpha|\mathbf{x})$.

Let $x_{(j)}$ denote the jth element of $\mathbf{x}$ and let $\gamma_j(\alpha)$ denote the corresponding element of $\boldsymbol{\gamma}(\alpha)$. From Theorem 2, it is possible to conclude that $\gamma_j(\alpha) = 0$ implies that $Q_Y(\alpha|\mathbf{x})$ does not depend on $x_{(j)}$, but the converse is not true. Equation (1) shows that if $\gamma_j(\alpha) \neq 0$, then it must be the case that the probability distribution at or below $Q_Y(\alpha|\mathbf{x})$ depends on $x_{(j)}$, but it is possible that for a given α (α_0, say), $\gamma_j(\alpha_0) \neq 0$, and yet changes in $x_{(j)}$ fail to impact *that very same* quantile of the count. For example, in a dataset with $100 \times \theta$ percent of zero inflation, all of the quantiles of Y up to $\alpha = \theta$ will be identically 0, even if the corresponding quantiles of Z depend on $x_{(j)}$. Consequently, it is easier to detect dependence of the distribution of Y on $\mathbf{X}$ by looking at $Q_Z(\alpha|\mathbf{x})$ than by looking at $Q_Y(\alpha|\mathbf{x})$. This may be called a "magnifying glass effect" of $Q_Z(\alpha|\mathbf{x})$.

This raises an interesting question: Is it possible to define a set of conditions under which $\gamma_j(\alpha_0) \neq 0$ is equivalent to $Q_Y(\alpha_0|\mathbf{x})$ depending on $x_{(j)}$? The answer is positive. Under (A2) and (A3), if it is possible to choose values $\mathbf{x}_*' = (x_*^{(d)\prime}\ \mathbf{x}_*^{(c)\prime})$ of the discrete and continuous covariates such that $\mathbf{x}_*^{(c)}$ is an interior point of the support of $\mathbf{X}^{(c)}$ and $Q_Z(\alpha_0|\mathbf{x}_*) \in \mathbb{N}$, then $\gamma_j(\alpha_0) \neq 0$ implies that there must exist subpopulations, namely those in the neighborhood of $\mathbf{x}_*$, for which the $100\alpha_0$th quantile of the count depends on $x_{(j)}$. The requirements on the support of $\mathbf{X}$ needed to identify the effects of covariates on the quantiles of the counts, albeit less stringent, parallel those of Manski's (1975, 1985) maximum score estimator, which also requires that one regressor be continuous and that a "large support" assumption holds.

Theorems 1 and 2 show that it is possible to consistently estimate the quantiles of the count variable. Indeed, under the conditions of Theorem 1, $\lceil Q_Z(\alpha|\mathbf{x}) - 1\rceil$, as a function of $\mathbf{x}$, is continuous almost everywhere, and thus.

Corollary. Under (A1), (A2), and (A3), for any given $\alpha \in (0, 1)$, $Q_Y(\alpha|\mathbf{x})$ is consistently estimated by $\lceil T^{-1}(\mathbf{x}'\hat{\boldsymbol{\gamma}}(\alpha), \alpha) - 1\rceil$.

It is also possible to make inferences about the effect on $Q_Y(\alpha_0|\mathbf{x})$ of a particular variation of a regressor, evaluated at a given value of the covariates. Let

$$\Delta_j Q_Y\big(\alpha_0\big|\xi, x^0_{(j)}, x^1_{(j)}\big) \equiv Q_Y\big(\alpha_0\big|\xi, x_{(j)} = x^1_{(j)}\big) - Q_Y\big(\alpha_0\big|\xi, x_{(j)} = x^0_{(j)}\big)$$

define the partial effect of $\Delta x_j = x^1_{(j)} - x^0_{(j)}$ on the $100\alpha_0$th quantile of the count when the remaining covariates are fixed at ξ. A consistent point estimate of $\Delta_j Q_Y(\alpha_0|\xi, x^0_{(j)}, x^1_{(j)})$ can be obtained by direct application of the foregoing corollary.

For $\Delta_j Q_Y(\alpha_0|\xi, x^0_{(j)}, x^1_{(j)})$ to be different from 0, the absolute value of $\gamma_j(\alpha_0)$ must be large enough so that Δx_j is sufficient to change the integer part of $Q_Z(\alpha_0|\xi, x^0_{(j)})$. That is, $\Delta_j Q_Y(\alpha_0|\xi, x^0_{(j)}, x^1_{(j)})$ is different from 0 if $\Delta_j Q_Z(\alpha_0|\xi, x^0_{(j)}, x^1_{(j)})$ is not contained in the interval

$$\big[-\mathrm{frac}\big(Q_Z\big(\alpha_0\big|\xi, x^0_{(j)}\big)\big) I\big(Q_Y\big(\alpha_0\big|\xi, x^0_{(j)}\big) > 0\big); \quad 1 - \mathrm{frac}\big(Q_Z\big(\alpha_0\big|\xi, x^0_{(j)}\big)\big)\big], \quad (3)$$

where $\mathrm{frac}(Q_Z(\alpha|\mathbf{x}))$ denotes the fractionary part of $Q_Z(\alpha|\mathbf{x})$.

The statistical significance of $\Delta_j Q_Y(\alpha_0|\xi, x^0_{(j)}, x^1_{(j)})$ can be evaluated coupling the foregoing expressions with standard procedures. For instance, asymptotically valid confidence intervals for $\Delta_j Q_Z(\alpha_0|\xi, x^0_{(j)}, x^1_{(j)})$ can be obtained using Theorem 1 and the delta method. Conditional on the estimated value of $Q_Z(\alpha_0|\xi, x^0_{(j)})$, a $100(1-p)\%$ interval for $\Delta_j Q_Z(\alpha_0|\xi, x^0_{(j)}, x^1_{(j)})$ implies, using the corollary, a set of values for $Q_Y(\alpha_0|\xi, x^1_{(j)})$ and thus for $\Delta_j Q_Y(\alpha_0|\xi, x^0_{(j)}, x^1_{(j)})$ that has a confidence level of $100(1-p)\%$. This approach is illustrated in detail in Section 5.

3.4 Estimation of the Covariance Matrix of $\hat{\boldsymbol{\gamma}}(\alpha)$

For the proposed estimator of $\boldsymbol{\gamma}(\alpha)$ to be useful in practice, it is necessary to obtain a consistent estimator for the covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\gamma}}(\alpha) - \boldsymbol{\gamma}(\alpha))$. Given that $\hat{\mathbf{A}} = \alpha(1-\alpha)\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i'$ is a consistent estimator of $\mathbf{A}$, the main difficulty is to find a consistent estimator for $\mathbf{D}$.

The approach followed here for the estimation of $\mathbf{D}$ is akin to the one proposed by Powell (1984), but it takes advantage of some specific characteristics of this problem. In particular, we explore the fact that, given the way in which 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 that returns the largest integer smaller than or equal to a. Therefore, the conditional expectation of the indicator

$$I\big\{\big\lfloor Q_Z(\alpha|\mathbf{x}_i)\big\rfloor \le Z_i < \big\lfloor Q_Z(\alpha|\mathbf{x}_i) + 1\big\rfloor\big\} \quad (4)$$

equals the conditional density of Z at $Q_Z(\alpha|\mathbf{x}_i)$. Of course, what is needed for the covariance estimator is not this, but rather an estimate of the density of $T(Z;\alpha)$ at $\mathbf{x}_i'\boldsymbol{\gamma}(\alpha)$, which can be obtained from the density of Z at $Q_Z(\alpha|\mathbf{x}_i)$, multiplying it by the Jacobian of the transformation. The implementation of this approach is complicated by the fact that the floor function is discontinuous; so this function is replaced in (4) by a continuous approximation. To obtain a consistent estimator for $\mathbf{D}$, three further assumptions are needed:

(A4) $T^{-1}(\cdot)$, 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:

(a) $E[|T^{-1}(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha))|\|\mathbf{x}_i\|^2]$ and
(b) $E[\sup_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}(\alpha)\|\le\delta} |T^{(1)}(\mathbf{x}_i'\boldsymbol{\gamma})|^l \|\mathbf{x}_i\|^2]$ for $l = 1, 2$ and some $\delta > 0$.

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

$$\frac{\sup_{1\le i\le n} |T^{-1}(\mathbf{x}_i'\hat{\boldsymbol{\gamma}}(\alpha)) - T^{-1}(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha))|}{c_n} = o_p(1),$$

as $n \to \infty$.

Theorem 3. Let

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

with

$$\hat{\omega}_i \equiv T^{(1)}(\mathbf{x}_i'\hat{\boldsymbol{\gamma}}(\alpha)) I\big\{F_n\big(\hat{Q}_{Z_i}(\alpha|\mathbf{x})\big) \le Z_i < F_n\big(\hat{Q}_{Z_i}(\alpha|\mathbf{x}) + 1\big)\big\},$$

where $\hat{Q}_{Z_i}(\alpha|\mathbf{x}) \equiv T^{-1}(\mathbf{x}_i'\hat{\boldsymbol{\gamma}}(\alpha))$ is the estimated 100αth conditional quantile of Z and, for a sequence c_n,

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

Then, under (A1)–(A6)

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

Assumption (A6) warrants a few comments. As is shown in the proof of Theorem 3 in the Appendix, $\sup_{1\le i\le n} |\hat{Q}_{Z_i}(\alpha|\mathbf{x}) - Q_{Z_i}(\alpha|\mathbf{x})| = o_p(1)$, 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 (A3). Suppose that the regressors are taken to have bounded support, that is, $P[\|\mathbf{x}_i\| < K] = 1$, for some K. This is the case considered in the simulation study presented in the next section. In this context, Theorem 1 implies that $\sup_{1\le i\le n} |\mathbf{x}_i'(\hat{\boldsymbol{\gamma}}(\alpha) - \boldsymbol{\gamma}(\alpha))| = O_p(1/\sqrt{n})$; on the other hand, the continuity of $T^{(1)}(\cdot)$ and the compactness of Γ imply that $T^{(1)}(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha))$ is bounded uniformly in i. Therefore,

$$\sup_{1\le i\le n} \big|T^{-1}(\mathbf{x}_i'\hat{\boldsymbol{\gamma}}(\alpha)) - T^{-1}(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha))\big| = O_p(1/\sqrt{n}),$$

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

The condition of bounded support may appear rather extreme. However, it is quite natural in linear QR models with heterogeneously distributed "errors" because otherwise, it would be impossible for the conditional quantiles to be linear for all values of $\mathbf{x}$ without crossing.

Alternatively, one may prefer the common assumption that the regressors satisfy

$$\sup_{1\le i\le n} \|\mathbf{x}_i\| = O_p\left(\frac{n^{1/4}}{\ln(n)}\right)$$

(see, e.g., Koenker and Machado 1999). Now specific rates for c_n must be derived on a case-by-case basis. For instance, for inverse transformations in the power family, $T^{-1}(\mathbf{x}) = \mathbf{x}^{\tau}$ with $\tau \le 2$, a simple mean-value expansion yields

$$\sup_{1\le i\le n} \left|T^{-1}(\mathbf{x}_i'\hat{\boldsymbol{\gamma}}(\alpha)) - T^{-1}(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha))\right| = O_p\left(\frac{n^{\tau/4}}{\ln(n)^{\tau}}\right) O_p\left(\frac{1}{\sqrt{n}}\right) = O_p\left(\frac{1}{n^{(2-\tau)/4}\ln(n)^{\tau}}\right),$$

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

3.5 Averaging Out the Uniform Noise

The estimate of the QR coefficients for the jittered data depends not only on the sample information $\{y_i, \mathbf{x}_i\}_{i=1}^n$, but also on the specific random sample $\{u_i\}_{i=1}^n$ drawn from a uniform in [0, 1) population. Because the latter represents "noise" introduced by mere technical reasons, it is natural to look for estimates that are less dependent on the specific realization of the random sample of U.

We consider an alternative, that we call "average-jittering," which averages the QR estimates for m "jittered" samples $\{y_i + u_i^{(l)}, \mathbf{x}_i\}_{i=1}^n$, $l = 1, \dots, m$, constructed from m independent random samples of size n from a uniform distribution. Formally, the average-jittering estimator ($\hat{\boldsymbol{\gamma}}_m^{\mathrm{A}}(\alpha)$) is

$$\hat{\boldsymbol{\gamma}}_m^{\mathrm{A}}(\alpha) = \frac{1}{m}\sum_{l=1}^{m} \hat{\boldsymbol{\gamma}}^{(l)}(\alpha),$$

where $\hat{\boldsymbol{\gamma}}^{(l)}(\alpha)$ is the QR estimator based on $\{y_i + u_i^{(l)}, \mathbf{x}_i\}$.

The asymptotic properties of this estimator follow from the results presented earlier.

Theorem 4. The data $\{y_i, \mathbf{x}_i, u_i\}_{i=1}^n$ are a random sample of $(Y, \mathbf{X}, U)$ satisfying (A1)–(A5). Then

$$\sqrt{n}\big(\hat{\boldsymbol{\gamma}}_m^{\mathrm{A}}(\alpha) - \boldsymbol{\gamma}(\alpha)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{V}^{\mathrm{A}}),$$

where

$$\mathbf{V}^{\mathrm{A}} = \frac{1}{m}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1} + \left(1 - \frac{1}{m}\right)\mathbf{D}^{-1}\mathbf{B}\mathbf{D}^{-1}, \tag{5}$$

with $\mathbf{A}$ and $\mathbf{D}$ as in Theorem 1,

$$\mathbf{B} = E\{[\alpha(1-\alpha) - \eta(\mathbf{X})]\mathbf{X}\mathbf{X}'\}$$

and

$$\eta(\mathbf{X}) = f_{Y|\mathbf{x}}(Q_Y(\alpha|\mathbf{x}))\{[Q_Z(\alpha|\mathbf{x}) - Q_Y(\alpha|\mathbf{x})] \times [Q_Y(\alpha|\mathbf{x}) + 1 - Q_Z(\alpha|\mathbf{x})]\}.$$

The average-jittering estimator is necessarily more efficient than the estimator obtained with a single sample. This may also be confirmed directly by noting that $\mathbf{A} - \mathbf{B} = E[\eta(\mathbf{X})\mathbf{X}\mathbf{X}']$ is a positive-definite matrix because $\eta(\mathbf{X})$ is positive with probability 1. The asymptotic covariance of the average-jittering estimator is a weighted average of the matrices $\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}$ (the covariance of the QR estimator for a single sample) and $\mathbf{D}^{-1}\mathbf{B}\mathbf{D}^{-1}$. As the number of replications of the jittering procedure (m) increases, more and more weight is placed on $\mathbf{D}^{-1}\mathbf{B}\mathbf{D}^{-1}$, the "smaller" matrix, and consequently the average-jittering estimator gets more precise. Even with a moderate number of repetitions (say $m = 10$), the contribution of $\mathbf{A}$ is relatively minor (10%).

Given these efficiency improvements, it seems natural to allow m to pass to ∞. The next theorem deals with this case, and reveals $\mathbf{D}^{-1}\mathbf{B}\mathbf{D}^{-1}$ to be the limiting (in m and n) covariance matrix of the average-jittering estimator.

Theorem 5. The dataset $\{y_i, \mathbf{x}_i, u_i\}_{i=1}^n$ is a random sample of $(Y, \mathbf{X}, U)$ satisfying (A1)–(A5). Then, when n and m go to infinity at any rate,

$$\sqrt{n}\big(\hat{\boldsymbol{\gamma}}_m^{\mathrm{A}}(\alpha) - \boldsymbol{\gamma}(\alpha)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{D}^{-1}\mathbf{B}\mathbf{D}^{-1}).$$

Other alternatives exist to integrate out the noise. For instance, it is possible to define an estimator by $\min_{\mathbf{b}} \sum_{i=1}^n \int_0^1 \rho_\alpha(y_i + u - T^{-1}(\mathbf{x}_i'\mathbf{b}))\,du$. Explicit evaluation of the integral yields the program $\min_{\mathbf{b}} \sum_{i=1}^n \rho_\alpha^\star(y_i, \mathbf{x}_i; \mathbf{b})$ with

$$\begin{aligned}\rho_\alpha^\star(y_i, \mathbf{x}_i; \mathbf{b}) &= [\alpha - I\{y_i \le T^{-1}(\mathbf{x}_i'\mathbf{b}) - 1\}][y_i + \tfrac{1}{2} - T^{-1}(\mathbf{x}_i'\mathbf{b})] \\ &\quad + \tfrac{1}{2}[y_i - T^{-1}(\mathbf{x}_i'\mathbf{b})]^2 I\{T^{-1}(\mathbf{x}_i'\mathbf{b}) - 1 < y_i \le T^{-1}(\mathbf{x}_i'\mathbf{b})\}.\end{aligned}$$

This approach is not presented here in detail, because the estimators so produced do not dominate—either in finite samples or asymptotically—the average-jittering estimator, and are more cumbersome to implement (details available on request).

3.6 Implementation Issues

To implement the results in Theorems 1–5, it is necessary to specify the form of $Q_Z(\alpha|\mathbf{x})$ and the associated transformation $T(Z;\alpha)$. We now describe a possible approach to this problem. Noting that $Q_Z(\alpha|\mathbf{x})$ is bounded from below by α, and keeping in line with what is traditionally assumed in count data models, a parametric representation of $Q_Z(\alpha|\mathbf{x})$ can be specified as

$$Q_Z(\alpha|\mathbf{x}) = \alpha + \exp(\mathbf{x}'\boldsymbol{\gamma}(\alpha)). \tag{6}$$

Note 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 that linear regression is used to approximate unknown mean regression functions (see White 1980; Angrist et al. 2004). The simulations that follow suggest that this specification is reasonable for a wide class of models. In any case, the adequacy of (6) is an empirical issue that should be checked in each particular application. Of course, in presence of misspecification of the regression quantiles, appropriate estimators of the covariance matrix must be used (see Chamberlain 1994; Kim and White 2003; Angrist et al. 2004). Further details on this are given in Section 4.

Using (6), a simple algorithm to estimate $\boldsymbol{\gamma}(\alpha)$ can be designed taking advantage of well-known properties of the quantile functions. In fact, $\boldsymbol{\gamma}(\alpha)$ can be estimated by running a linear QR of

$$T(Z;\alpha) = \begin{cases} \log(Z-\alpha) & \text{for } Z > \alpha \\ \log(\varsigma) & \text{for } Z \le \alpha \end{cases} \tag{7}$$

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

Having estimated $\boldsymbol{\gamma}(\alpha)$, the average-jittering estimator can be obtained as the sample average of the estimates from each of the m jittered samples. Its covariance matrix can be estimated using $\mathbf{V}^{\mathrm{A}}$ as defined in (5), replacing $\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}$ and $\mathbf{D}^{-1}\mathbf{B}\mathbf{D}^{-1}$ by the sample averages of their estimates in the m samples.

A final implementation issue worth considering concerns the choice of the distribution for U. As mentioned earlier, the data smoothing can be done with a noise that is not uniform, as long as the random variables used have support on $[0, 1)$ and density $h(\cdot)$ bounded away from 0. Noting that the variance of the estimator depends on $h(\Lambda_i)$, where $\Lambda_i = \hat{Q}_{Z_i}(\alpha|\mathbf{x}) - \lfloor \hat{Q}_{Z_i}(\alpha|\mathbf{x}) \rfloor$, it would be natural to choose $h(\cdot)$ so that efficiency is maximized. Because for the uniform noise $h(\Lambda_i) = 1$ for all observations, efficiency certainly could be improved by smoothing the data with a noise such that $h(\Lambda_i) \geq 1$, $i = 1, \dots, n$. However, this is generally impossible to achieve with a common noise for all data points. Indeed, different observations generate values for Λ_i that correspond to different quantiles of U and thus to different heights of the density $h(\cdot)$; these in turn cannot be all above 1. Without previous knowledge of the differences Λ_i, it is not possible to ensure that using a different distribution for the noise will lead to efficiency gains relative to the estimator based on the uniform jittering. Therefore, it is natural to chose $h(\cdot)$ in a way that ensures that $h(\Lambda_i)$ never falls below 1.

In any case, from a theoretical viewpoint there is little loss of generality in assuming that U is drawn from a uniform distribution, because most of the foregoing results are valid if the smoothing noise is generated from any continuous distribution with support on $[0, 1)$ and $h(\cdot)$ bounded away from 0. The main exception to this is Theorem 3, in which the weights $\hat{\omega}_i$ should be multiplied by $h(\Lambda_i)$. Finally, it should be noted that the choice of the distribution used to generate the smoothing noise also has implications for the functional form of the model. With the uniform noise, continuity is achieved by linear interpolation of each jump in the conditional quantile function of the counts [cf. eq. (1)], whereas other distributions would imply nonlinear interpolation. Naturally, the efficiency of the QR estimators could also be improved using appropriate weights (see Newey and Powell 1990; Koenker and Zhao 1994; Zhao 2001), but that idea is not pursued here.

4. SIMULATION RESULTS

This section reports the results of a pilot simulation study on the finite-sample behavior of hypotheses tests based on $\hat{\boldsymbol{\gamma}}(\alpha)$ and $\hat{\boldsymbol{\gamma}}_m^{\mathrm{A}}(\alpha)$ and on the proposed estimators of their covariance matrices. As a byproduct, the relative efficiency of the estimators is also investigated.

In these experiments, the counts $Y_1, Y_2, \dots, Y_n$ were generated according to four different processes. In case 1, the counts are Poisson random variables with conditional mean $\mu_i = \exp(\lambda_0 + \lambda_1 x_{1i} + \lambda_2 x_{2i})$, where the x_{1i}'s were obtained as random draws from a beta-distributed random variable with parameters $5/3$ and $5/3$, centered and scaled to have support in the interval $(-5/3, 5/3)$, and x_{2i} is a dummy variable that equals 1 with probability .2 and 0 otherwise. In case 2, the conditional distribution of the Y's is negative binomial with mean μ_i as in case 1 and variance $\mu_i + .5\mu_i^2$. Cases 3 and 4 are zero-inflated versions of the first two, with a proportion of zero inflation of .2.

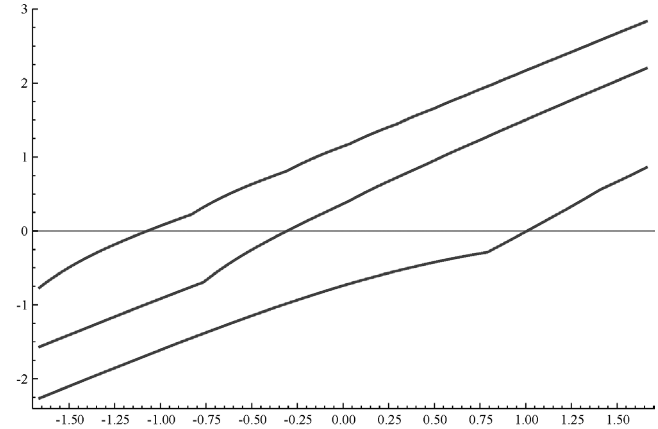

Figure 1. The Three Quartiles of $T(Z;\alpha)$ Against $\mathbf{x}_1$ for Case 4.

All experiments were performed with $\lambda_0 = \lambda_1 = 1$ and $\lambda_2 = 0$. Thus μ_i does not depend on $\mathbf{x}_2$, which thus does not affect any aspect of the distributions of the counts. This implies that the effect of $\mathbf{x}_2$ on $Q_Y(\alpha|\mathbf{x})$ is zero for every α. Using TSP 4.5 (Hall and Cummins 1999), 5,000 simulations for each case were performed for $\alpha \in \{.25, .50, .75\}$ and $n \in \{500, 2{,}000\}$. A new set of covariates was drawn for each of the 5,000 simulations.

In all cases considered, $Q_Z(\alpha|\mathbf{x}) \in \mathbb{N}$ for an interior point of the support of $\mathbf{x}_1$. As discussed before, in these conditions it is possible to construct asymptotically valid tests for the hypothesis that $Q_Y(\alpha|\mathbf{x})$ does not depend on a given covariate by testing that the element of $\boldsymbol{\gamma}(\alpha)$ corresponding to that regressor is 0. To evaluate the proposed tests for the hypothesis that $Q_Y(\alpha|\mathbf{x})$ does not depend on $\mathbf{x}_2$, the counts were jittered as described in Section 3, and the quantiles of the artificial data were estimated using the two procedures described in the previous section. Specifying $Q_Z(\alpha|\mathbf{x})$ as in (6), Z was then transformed to obtain $T(Z;\alpha)$ according to (7), with $\varsigma = 10^{-5}$. The average-jittering estimator was implemented with $m = 50$.

Before proceeding, it is interesting to evaluate the specification of the models. A graphical analysis of the three quartiles of $T(Z;\alpha)$ as functions of $\mathbf{x}_1$ reveals that these are clearly nonlinear, especially for the zero-inflated cases. Figure 1 graphs the three quartiles of $T(Z;\alpha)$ as functions of $\mathbf{x}_1$ for case 4, the case where nonlinearities are more clearly present. (Similar graphs for the other cases considered in these experiments were given in Machado and Santos Silva 2002.) However, overall, the linearity assumption seems a reasonable approximation, at least over this range of $\mathbf{x}_1$. Therefore, the parameters of interest were estimated by running the usual QR of $T(Z;\alpha)$ on $\mathbf{x}_1$, $\mathbf{x}_2$, and a constant, and the significance of the parameter associated with $\mathbf{x}_2$ was tested using an asymptotic t-ratio.

Given the misspecification of the functional form of the quantiles being estimated, the misspecification-robust versions of the covariance matrices were used in constructing the t-ratios (see Chamberlain 1994; Kim and White 2003; Angrist et al. 2004). Specifically, the estimators used here are based on the asymptotically valid covariance matrices presented in Section 3, which in this particular example are constructed using the following estimators of **A**, **B**, and **D**:

$$\hat{\mathbf{A}} = \frac{1}{n}\sum_{i=1}^{n}\big[\alpha - I\big(T(Z_i;\alpha) \le \mathbf{x}_i'\hat{\boldsymbol{\gamma}}(\alpha)\big)\big]^2 \mathbf{x}_i\mathbf{x}_i',$$

$$\hat{\mathbf{B}} = \frac{1}{n}\sum_{i=1}^{n}\big[\alpha^2 + (1-2\alpha)I\big\{Y_i \le \hat{z}_{\alpha_i} - 1\big\} + \big[\big(\hat{z}_{\alpha_i} - Y_i\big)I\big\{\hat{z}_{\alpha_i} - 1 < Y_i \le \hat{z}_{\alpha_i}\big\}\big]\big(\hat{z}_{\alpha_i} - Y_i - 2\alpha\big)\big]\mathbf{x}_i\mathbf{x}_i',$$

and

$$\hat{\mathbf{D}} = \frac{1}{n}\sum_{i=1}^{n}\exp(\mathbf{x}_i'\hat{\boldsymbol{\gamma}}(\alpha))I\big(F_n\big(\hat{z}_{\alpha_i}\big) \le Z_i < F_n\big(\hat{z}_{\alpha_i}+1\big)\big)\mathbf{x}_i\mathbf{x}_i',$$

where the function $F_n(\cdot)$ is as defined in Theorem 3 and $\hat{z}_{\alpha_i} = T^{-1}(\mathbf{x}_i'\hat{\boldsymbol{\gamma}}(\alpha)) = \alpha + \exp(\mathbf{x}_i'\hat{\boldsymbol{\gamma}}(\alpha))$. This misspecification-robust estimator of **A** was given by Chamberlain (1994) and Kim and White (2003). The matrix $\hat{\mathbf{D}}$ is the one proposed in Theorem 3. Finally, the form of the misspecification-robust estimator of **B** can be easily obtained from the proof of Theorem 4 in the Appendix. For the average-jittering estimator, in each replication of the simulations, the covariance matrix estimator was constructed averaging $\hat{\mathbf{D}}^{-1}\hat{\mathbf{A}}\hat{\mathbf{D}}^{-1}$ and $\hat{\mathbf{D}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{D}}^{-1}$ across the m jittered samples.

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 asymptotic t-test is not very sensitive to the choice of this parameter. The results presented here were obtained setting $c_n = .5\ln(\ln(n))/\sqrt{n}$.

The rejection frequencies of the null at the 1%, 5%, and 10% nominal sizes for the tests based on $\hat{\boldsymbol{\gamma}}(\alpha)$ and $\hat{\boldsymbol{\gamma}}_{50}^{\mathrm{A}}(\alpha)$ were computed for all of the cases considered. Because the results obtained with $\hat{\boldsymbol{\gamma}}(\alpha)$ are generally inferior to those obtained with $\hat{\boldsymbol{\gamma}}_{50}^{\mathrm{A}}(\alpha)$ (especially for $n = 500$), only the results obtained with the latter are presented in Table 1. (The full set of results is available in Machado and Santos Silva 2002.)

Overall, it is encouraging to find that the results obtained using this estimator are reasonably accurate. Of course, the rejection frequencies are further away from the nominal significance level for the tail of the distribution and for the smaller sample size, but even in these cases the results are reasonable. 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 favorably with, for example, those reported by Horowitz (1992) for models estimated with the smoothed maximum score estimator.

Finally, it is interesting to consider the efficiency of $\hat{\boldsymbol{\gamma}}(\alpha)$ relative to $\hat{\boldsymbol{\gamma}}_{50}^{\mathrm{A}}(\alpha)$. In these experiments the estimated relative efficiency for the two upper quartiles is generally between .85 and .90. However, for the first quartile, the efficiency gain provided by $\hat{\boldsymbol{\gamma}}_{50}^{\mathrm{A}}(\alpha)$ is more substantial. For $\alpha = .25$, the estimated efficiency of $\hat{\boldsymbol{\gamma}}(\alpha)$ relative to $\hat{\boldsymbol{\gamma}}_{50}^{\mathrm{A}}(\alpha)$ is always below .80, sometimes dropping below .60. Therefore, at least in this setup, the small additional computational cost of the average-jittering estimator can be compensated by reasonable efficiency gains, especially for the lower quantiles. (See Machado and Santos Silva 2002 for more results on the efficiency of the proposed estimators.)

5. AN EMPIRICAL ILLUSTRATION

This section illustrates the application of QR for counts using the dataset on demand for health care previously studied by Pohlmeier and Ulrich (1995) and by Santos Silva and Windmeijer (2001). These data consist of 5,096 observations for employed individuals and are taken from the 1985 wave of the German Socioeconomic Panel (SOEP). In this study the demand for health care is measured by the number of visits to a specialist (except gynecology or pediatrics) in the last quarter. The covariates used in the analysis are described in Table A.1 in the Appendix and correspond to those originally used by Pohlmeier and Ulrich (1995), who gave more detailed information.

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 whether or not to seek medical care, and in a second stage, the individual and the health care provider decide on the total number of visits needed to complete the treatment. 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 was clearly illustrated in Winkelmann's (2004) recent study of the effects of German health care reform.

Generally, heavily parameterized models are used to study how the effects of the covariates vary across different regions of the conditional distribution of interest. Here QR is used for the same purpose. Given that the purpose of this section is purely to illustrate the application of the proposed methodology, we restrict the analysis to the three conditional quartiles.

Preliminary experiments revealed that in this example, some coefficients are quite sensitive to the particular sample of uniform random variables used to jitter the data, especially for $\alpha = .25$. Specifically, for $\alpha = .25$, the coefficient of the variable PHYSICIAN DENSITY has a variance across the m jittered samples of about 1.2. Therefore, only the results for the

Table 1. Rejection Frequencies for the Four Different Cases

	$n = 500$			$n = 2{,}000$		
α	Rejection at 1%	Rejection at 5%	Rejection at 10%	Rejection at 1%	Rejection at 5%	Rejection at 10%
Case 1: Poisson						
.75	1.54	5.36	10.68	1.36	5.30	10.38
.50	1.14	5.24	10.36	.86	5.14	10.64
.25	1.60	5.56	10.70	1.18	5.18	9.90
Case 2: Negative binomial						
.75	2.14	6.56	11.22	1.24	5.26	10.04
.50	1.60	6.34	11.56	1.40	5.22	10.06
.25	1.50	5.66	10.18	1.30	5.38	10.08
Case 3: Zero inflated Poisson						
.75	1.40	6.02	10.80	1.36	5.74	10.02
.50	1.36	5.30	10.06	.94	4.72	10.48
.25	.98	4.24	7.78	.84	4.38	9.20
Case 4: Zero inflated negative binomial						
.75	1.78	5.90	10.34	1.34	5.68	10.74
.50	1.90	5.56	9.86	1.38	5.46	11.04
.25	1.42	5.04	10.02	1.02	4.88	10.22

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
$\widehat{Q}_Y(.25\|\mathbf{x})$	5,027	67	2	0	0	0	0	0	0	0	0	0	0
$\widehat{Q}_Y(.50\|\mathbf{x})$	4,145	773	122	38	15	0	3	0	0	0	0	0	0
$\widehat{Q}_Y(.75\|\mathbf{x})$	1,714	2,245	432	220	126	99	49	56	29	20	13	15	78

average-jittering estimator are reported. Moreover, this estimator was implemented using 10,000 jittered samples to ensure that the results reported here can be closely replicated by other researchers if a suitably large value of m is used.

For comparison, results for the conditional expectiles based on the AML estimator proposed by Efron (1992) are also presented. Because the AML regression percentiles cannot be computed for values of α smaller than the proportion of 0's in the sample, and because this sample has 3,456 observations for which the individuals report no visits, the AML approach allows only 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 quantile regression is the one leading to $\frac{1}{n}\sum_{i=1}^{n} I(Y_i \leq \exp(\mathbf{x}_i'\hat{\boldsymbol{\beta}}_w^{\mathrm{AML}})) = .75$, which is obtained with $w = .875$.

Table 2 presents the absolute frequencies of the estimated quantiles for the number of visits to specialists. Given that almost 70% of the individuals in the sample have no visits, it is not surprising to find that the estimate of the first conditional quartile of the distribution of the number of visits is 0 for 5,027 individuals, which is almost 99% of the sample. Even the higher quartiles are flat at 0 for a large proportion of the sample. In fact, the results in Table 2 show that the probability of having no visits to a specialist during the observation period is at least .75 for about one-third (1,714 of 5,096) of the individuals in the sample.

Table 3 presents estimates of the partial effects of the covariates on $Q_Z(\alpha|\bar{\mathbf{x}})$, where $\bar{\mathbf{x}}$ is a vector containing the mean value of the continuous covariates and 0's for the dummy variables. For the continuous variables, the effects of the regressors on $Q_Z(\alpha|\bar{\mathbf{x}})$ are defined as the appropriate derivative, and for the dummies the effects are $\exp(\bar{\mathbf{x}}'\boldsymbol{\gamma}(\alpha) + \gamma_j(\alpha)) - \exp(\bar{\mathbf{x}}'\boldsymbol{\gamma}(\alpha))$. Table 3 also includes the estimated effects obtained with the AML estimator for $w = .875$. The effects of the covariates in the AML regression are computed in a way analogous to that described earlier.

In interpreting these results, it is important to keep in mind that $Q_Z(\alpha|\bar{\mathbf{x}})$ is a function of α. Therefore, in general, a variable with the same effect for all of the quantiles will have a proportional effect that varies with α. Using the results in Table 3, the proportional effects can be easily computed with the following set of additional results: $\widehat{Q}_Z(.25|\bar{\mathbf{x}}) = .311$ [standard error (SE) $= .007$], $\widehat{Q}_Z(.50|\bar{\mathbf{x}}) = .616$ (SE $= .012$), $\widehat{Q}_Z(.75|\bar{\mathbf{x}}) = .954$ (SE $= .022$), and $\exp(\bar{\mathbf{x}}'\hat{\boldsymbol{\beta}}_{.875}^{\mathrm{AML}}) = .554$ (SE $= .053$).

Starting with the analysis of the QR results, it is interesting to note that for some covariates, the point estimates of the partial effects do not vary much across the estimated quartiles. But this happens mostly with variables associated with statistically insignificant parameters, whose estimated effects are often close to 0 (e.g., PRIVATE INSURANCE, UNEMPLOYMENT, and HOSPITALIZED). The effects of the regressors related to job characteristics (i.e., HEAVY LABOR, STRESS, VARIETY ON JOB, SELF–DETERMINED, and CONTROL) are also close to 0 and not individually significant in any of the estimated quantiles. This evidence suggests that these covariates have little or no impact on the shape of the conditional distribution, a result in line with the findings of Pohlmeier and Ulrich (1995) and Santos Silva and Windmeijer (2001).

Table 3. Estimation Results

	Quantile regression results						Efron's AML	
	1st quartile		2nd quartile		3rd quartile		$w = .875$	
Regressors	Effect	SE	Effect	SE	Effect	SE	Effect	SE
FEMALE	.094	.012	.223	.028	.700	.084	.458	.072
SINGLE	−.017	.005	−.034	.011	−.061	.021	−.160	.045
AGE	−.005	.002	−.009	.005	−.016	.009	−.017	.019
INCOME	.013	.010	.026	.035	.080	.040	.053	.047
EDUCATION	.002	.001	.003	.002	.008	.004	.001	.007
UNEMPLOYMENT	−.002	.010	−.002	.003	−.007	.008	−.018	.016
PRIVATE INSURANCE	.016	.011	.039	.022	.041	.034	−.009	.071
CHRONIC COMPLAINTS	.104	.016	.259	.043	.744	.122	.766	.124
HOSPITALIZED	.019	.012	.038	.027	.082	.048	.168	.097
SICK LEAVE	.037	.011	.095	.026	.307	.078	.590	.129
DISABILITY	.021	.013	.067	.033	.149	.059	.135	.085
HEAVY LABOR	−.003	.006	−.001	.012	−.002	.023	−.062	.050
STRESS	.004	.005	.010	.011	.042	.026	.053	.051
VARIETY ON JOB	.009	.005	.014	.011	.027	.020	.106	.050
SELF–DETERMINED	.005	.005	.011	.011	.027	.021	−.012	.045
CONTROL	.000	.006	.000	.012	.017	.026	.069	.059
POP-0/5	−.031	.006	−.063	.010	−.124	.022	−.258	.052
POP-5/20	−.025	.005	−.048	.010	−.078	.019	−.146	.048
POP-20/100	−.015	.005	−.031	.010	−.052	.021	−.127	.044
PHYSICIAN DENSITY	.091	.044	.134	.082	.203	.164	.529	.344

Despite these exceptions, for most regressors, the estimated effects vary substantially with α. Leading examples of this are the variables FEMALE, CHRONIC COMPLAINTS, SICK LEAVE, and DISABILITY, whose effects sharply increase from the first quartile to the third quartile. For example, using the estimated effect of FEMALE and the estimates of $\widehat{Q}_Z(\alpha|\bar{\mathbf{x}})$ presented earlier, it is possible to conclude that, keeping all of the other variables constant, the estimated quantiles of Y for $\alpha = .25$ and $\alpha = .50$ are 0 for both men and women, but the third quartile is estimated to be 0 for men and 1 for women. That is, evaluated at this point, the ceteris paribus effect of the variable FEMALE appears to be more pronounced in the upper tail of the distribution.

For the subpopulation under analysis, the effect of gender on the third quartile of the number of visits is estimated very precisely as 1. Indeed, conditional on $\widehat{Q}_Z(.75|\bar{\mathbf{x}}) = .954$, a 99% confidence interval for the 75th quantile of Z for an otherwise equal female ranges from 1.42 to 1.89. Within this range, the corresponding count is always estimated as 1, and because $\widehat{Q}_Y(.75|\bar{\mathbf{x}}) = 0$, we may estimate that $\{1\}$ constitutes a 99% confidence set for the effect of gender in this subpopulation. In contrast, evaluated at this point, the effect of gender is clearly insignificant for $\alpha = .25$ and $\alpha = .5$, because the corresponding point estimates .094 (SE = .012) and .223 (SE = .028) lie more than five standard errors away from the relevant boundaries of the intervals given by (3), [0; .689] and [−.616; .384].

In contradistinction to what happens with FEMALE, the effect of PHYSICIAN DENSITY is estimated more precisely in the lower tail of the distribution, being statistically significant only for the first quartile. By itself, this effect is not strong enough to impact the 25th quantile of the number of visits of the reference subpopulation (characterized by attributes $\bar{\mathbf{x}}$). Indeed, the ceteris paribus partial effect on the 25th quantile of Z resulting from of changing PHYSICIAN DENSITY from its sample mean to the sample maximum is estimated to be .210 (SE = .197). However, because the corresponding coefficient is significantly different from 0, in principle it is possible to find subpopulations for which variations of the physicians density would change the 25th quantile of the count from 0 to 1. Thus it can be argued that this variable is especially important in determining 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).

Using $w = .875$, the AML estimator leads to 75% of negative residuals, and thus $\hat{\boldsymbol{\beta}}^{\mathrm{AML}}_{.875}$ is the vector of parameters for the 75th AML regression percentile, as defined by Efron (1992). For some covariates, like CHRONIC COMPLAINTS, this AML estimator leads to results that are reasonably close to those obtained for the third quartile. Moreover, as in the three quartiles, the covariates related to job characteristics are not individually significant. The only exception is VARIETY ON JOB, which is statistically significant in the AML regression but not in any of the quantile regressions. Despite these similarities, there are some notable differences between the two sets of results. A clear example of this is provided by the variables FEMALE and SINGLE, which are statistically significant in both the AML regression and the third quartile but have very different estimated effects in the two regressions. Moreover, the variable DISABILITY is highly significant in the third quartile but statistically insignificant in this AML regression. These important differences suggest that using 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 how the shape of the distribution is affected by the covariates. Fully parametric models achieve this by modeling the effect of the covariates on a few key aspects of the distribution. However, 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 QR method used here provides a more graphical description of the effect of the regressors on the shape of the conditional distribution of interest and is less sensitive to distributional assumptions.

6. CONCLUDING REMARKS

Direct estimation of QRs for count data is not feasible, because of the combination of the discreteness of the data and the nondifferentiability of the sample objective function defining the estimator. We propose applying QR 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 QR estimator has a 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 dataset on demand for health care, it was shown that the proposed methodology is very easy to implement and leads to interesting results.

Naturally, QR cannot replace the more structured and well-proven models for count data analysis. However, it can be a valuable additional tool that aids in the construction of more complex models, providing insights on how the regressors affect not only the location of the conditional distribution, but also its entire shape.

APPENDIX: PROOFS

A.1 Proof of Theorem 1

The results follow as described by Pollard (1991), once we show that one can do without the continuity of $f_{T|\mathbf{x}}(\cdot|\cdot)$ at $\mathbf{x}'\boldsymbol{\gamma}(\alpha)$. The strategy of Pollard's proof is to develop a quadratic approximation to the sample objective function defining $\hat{\boldsymbol{\gamma}}(\alpha)$ whose minimizing value has a limiting normal distribution and is asymptotically equivalent to $\hat{\boldsymbol{\gamma}}(\alpha)$. In that process, the "continuity assumption" is only required to ensure that the "limiting" objective function $M(t;\mathbf{x}) \equiv E(\rho_\alpha(\epsilon_\alpha - t)|\mathbf{x})$, with $\epsilon_\alpha \equiv T(Z;\alpha) - \mathbf{X}'\boldsymbol{\gamma}(\alpha)$ [note that $Q_{\epsilon_\alpha}(\alpha|\mathbf{x}) = 0$], has a second-order Taylor expansion around $t = 0$.

Assumptions (A2b) and (A3) imply that $f_{\epsilon|\mathbf{x}}(0|\mathbf{x})$ is continuous for almost every x (a.e. $\mathbf{x}$) and, consequently, $M''(t;\mathbf{x})$ is finite at $t = 0$ and equals

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

Therefore, the warranted second-order Taylor expansion,

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

exists a.e. $\mathbf{x}$ [cf. Pollard 1991, eq. (1)].

A.2 Proof of Theorem 2

Let $Z = Y + U$, where U, is any random variable with support on $[0, 1)$ and $Y = i$, $i = 0, 1, \dots,$ with probability conditional on $\mathbf{X} = \mathbf{x}$, $p_i(\mathbf{x})$. Clearly, $Y - 1 \le Z - 1 < Y$ and, because the quantile function is nondecreasing, $Q_Y(\alpha|\mathbf{x}) - 1 \le Q_Z(\alpha|\mathbf{x}) - 1 \le Q_Y(\alpha|\mathbf{x})$. The results now follow because $Q_Y(\alpha|\mathbf{x})$ is an integer.

A.3 Proof of Theorem 3

Let

$$\tilde{\mathbf{D}}_n \equiv \frac{1}{n}\sum_{i=1}^{n} \omega_i \mathbf{x}_i \mathbf{x}_i',$$

with

$$\omega_i \equiv T^{(1)}(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha)) I\big\{F_n(z_{\alpha_i}) \le Z_i < F_n(z_{\alpha_i}+1)\big\},$$

where z_{α_i} is used as shorthand for $T^{-1}(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha)) = Q_Z(\alpha|\mathbf{x}_i)$. Notice that as n passes to ∞, $F_n(x) \to \lfloor x \rfloor$. Thus, by dominated convergence, as $n \to \infty$,

$$E[\omega_i \mathbf{x}_i \mathbf{x}_i'] \to E\big[T^{(1)}(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha)) I\big\{\lfloor z_{\alpha_i}\rfloor \le Z_i < \lfloor z_{\alpha_i}+1\rfloor\big\}\mathbf{x}_i\mathbf{x}_i'\big],$$

which a simple change of variable shows equals $E[f_T(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha))\mathbf{x}_i\mathbf{x}_i']$. The law of large numbers then yields

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

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

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

with

$$B_{1ni} \equiv \big|T^{(1)}(\mathbf{x}_i'\hat{\boldsymbol{\gamma}}(\alpha)) - T^{(1)}(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha))\big|,$$

and

$$B_{2ni} \equiv \big|T^{(1)}(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha))\big|\Big(\Big|I\big\{F_n(\hat{z}_{\alpha_i}) \le Z_i < F_n(\hat{z}_{\alpha_i}+1)\big\} - I\big\{F_n(z_{\alpha_i}) \le Z_i < F_n(z_{\alpha_i}+1)\big\}\Big|\Big),$$

with $\hat{z}_{\alpha_i} = T^{-1}(\mathbf{x}_i'\hat{\boldsymbol{\gamma}}(\alpha))$. Therefore, $\|\hat{\mathbf{D}}_n - \tilde{\mathbf{D}}_n\|$ can be bounded above by the sum of two terms, the first being

$$\frac{1}{n}\sum \|\mathbf{x}_i\|^2 B_{1ni}.$$

Consider B_{1ni} to be a function of $\hat{\boldsymbol{\gamma}}(\alpha)$, $B_{1ni} = B_{1ni}(\hat{\boldsymbol{\gamma}}(\alpha))$, and note that, by (A4), it is continuous at $\boldsymbol{\gamma}(\alpha)$. Also, $\sup_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}(\alpha)\|\le\delta} \|\mathbf{x}_i\|^2 \times B_{1ni}(\boldsymbol{\gamma}) \le 2\|\mathbf{x}_i\|^2 \sup_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}(\alpha)\|\le\delta} |T^{(1)}(\mathbf{x}_i'\boldsymbol{\gamma})|$, whose expectation is finite by (A5). Thus a direct application of lemma 4.3 of Newey and McFadden (1994) yields

$$\frac{1}{n}\sum \|\mathbf{x}_i\|^2 B_{1ni} \xrightarrow{P} E\big\{\|\mathbf{x}_i\|^2 B_{1n1}(\boldsymbol{\gamma}(\alpha))\big\} = 0.$$

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

$$\begin{aligned}&\frac{1}{n}\sum \|\mathbf{x}_i\|^2 \big|T^{(1)}(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha))\big| I\big\{\big|Z_i - F_n(z_{\alpha_i})\big| \le \big|F_n(\hat{z}_{\alpha_i}) - F_n(z_{\alpha_i})\big|\big\} \\ &\quad + \frac{1}{n}\sum \|\mathbf{x}_i\|^2 \big|T^{(1)}(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha))\big| \\ &\qquad \times I\big\{\big|Z_i - F_n(z_{\alpha_i}+1)\big| \le \big|F_n(\hat{z}_{\alpha_i}+1) - F_n(z_{\alpha_i}+1)\big|\big\}.\end{aligned}$$

We show that the first of these terms is $o_p(1)$; the second term is analogous. To simplify notation, let $\Delta_i \equiv F_n(\hat{z}_{\alpha_i}) - F_n(z_{\alpha_i})$. For any $\eta > 0$ and any ϵ,

$$\begin{aligned}&P\bigg[\frac{1}{n}\sum \|\mathbf{x}_i\|^2 \big|T^{(1)}(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha))\big| I\big\{\big|Z_i - F_n(z_{\alpha_i})\big| \le |\Delta_i|\big\} > \eta\bigg] \\ &\quad \le P\bigg[\frac{1}{n}\sum \|\mathbf{x}_i\|^2 \big|T^{(1)}(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha))\big| I\big\{\big|Z_i - F_n(z_{\alpha_i})\big| \le \epsilon\big\} > \eta\bigg] \\ &\qquad + P\Big[\sup_{1\le i\le n} |\Delta_i| > \epsilon\Big].\end{aligned}$$

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

$$E\big[I\big\{\big|Z_i - F_n(z_{\alpha_i})\big| \le \epsilon\big\}\big] \le 2\epsilon,$$

show that the first term goes to 0 with ϵ.

To complete the proof, we show that $\Delta_i = o_p(1)$ uniformly in i. First, note the fact that by (A5), $E[\sup_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}(\alpha)\|\le\delta} |T^{(1)}(\mathbf{x}_i'\boldsymbol{\gamma})|^2 \times \|\mathbf{x}_i\|^2] < \infty$, for some $\delta > 0$, implies that

$$\sup_{1\le i\le n}\ \sup_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}(\alpha)\|\le\delta} \big|T^{(1)}(\mathbf{x}_i'\boldsymbol{\gamma})\big|\|\mathbf{x}_i\| = o_p(\sqrt{n}).$$

Therefore, by a mean-value expansion, $\sup_{1\le i\le n}|\hat{z}_{\alpha_i} - z_{\alpha_i}| = o_p(1)$. Assumption (A6) states that this convergence must be faster than c_n; that is, for n sufficiently large,

$$P\Big[\sup_{1\le i\le n}\big|\hat{z}_{\alpha_i} - z_{\alpha_i}\big| < \epsilon c_n\Big] \ge 1 - \delta$$

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,

$$P\big[\big|F_n(\hat{z}_{\alpha_i}) - F_n(z_{\alpha_i})\big| < \epsilon, \forall\, i \le n\big] \ge 1 - \delta,$$

as n passes to ∞.

A.4 Proof of Theorem 4

For fixed m, the asymptotic normality follows from Theorem 1. Because $\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}$ is the asymptotic covariance matrix of each $\sqrt{n}(\hat{\boldsymbol{\gamma}}^{(l)}(\alpha) - \boldsymbol{\gamma}(\alpha))$, $l = 1, \dots, m$, it remains to evaluate the $m(m-1)$ covariances of the estimators for different jittered samples.

Let $\boldsymbol{\omega}_{il} \equiv \boldsymbol{\omega}(y_i, u_i^{(l)}, \mathbf{x}_i) \equiv [\alpha - I\{y_i + u_i^{(l)} \le T^{-1}(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha))\}]\mathbf{x}_i$, so that

$$\sqrt{n}\big(\hat{\boldsymbol{\gamma}}^{(l)}(\alpha) - \boldsymbol{\gamma}(\alpha)\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbf{D}^{-1}\boldsymbol{\omega}_{il} + o_P(1).$$

Because $E\boldsymbol{\omega}_{il}\boldsymbol{\omega}_{jk} = \mathbf{0}$ for $i \ne j$,

$$E\Bigg(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{D}^{-1}\boldsymbol{\omega}_{il}\Bigg)\Bigg(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{D}^{-1}\boldsymbol{\omega}_{ik}\Bigg)' = \mathbf{D}^{-1}E(\boldsymbol{\omega}_{il}\boldsymbol{\omega}_{ik}')\mathbf{D}^{-1}.$$

It remains to evaluate $E(\boldsymbol{\omega}_{il}\boldsymbol{\omega}_{ik})$ for $l \ne k$. To simplify the notation, put $z_{\alpha_i} \equiv T^{-1}(\mathbf{x}_i'\boldsymbol{\gamma}(\alpha)) = Q_Z(\alpha|\mathbf{x}_i)$ and $y_{\alpha_i} \equiv Q_{Y|\mathbf{x}_i}(\alpha) = \lceil Q_Z(\alpha|\mathbf{x}_i) - 1\rceil$. Conditional on $\mathbf{X} = \mathbf{x}$, the relevant factor in the expectation is

$$\begin{aligned}&E\big[\big(\alpha - I\big\{y + u^{(l)} \le z_\alpha\big\}\big)\big(\alpha - I\big\{y + u^{(k)} \le z_\alpha\big\}\big)|\mathbf{x}\big] \\ &\quad = \alpha^2 - \alpha E\big[I\big\{y + u^{(l)} \le z_\alpha\big\}|\mathbf{x}\big] - \alpha E\big[I\big\{y + u^{(k)} \le z_\alpha\big\}|\mathbf{x}\big] \\ &\qquad + E\big[I\big\{y + u^{(l)} \le z_\alpha\big\} I\big\{y + u^{(k)} \le z_\alpha\big\}|\mathbf{x}\big] \\ &\quad = E\big[I\big\{y + u^{(l)} \le z_\alpha\big\} I\big\{y + u^{(k)} \le z_\alpha\big\}|\mathbf{x}\big] - \alpha^2 \\ &\quad = E_{Y|\mathbf{x}}\big[E\big[I\big\{u^{(l)} \le z_\alpha - y\big\} I\big\{u^{(k)} \le z_\alpha - y\big\}|\mathbf{x}, y\big]\big] - \alpha^2 \\ &\quad = E_{Y|\mathbf{x}}\big[F_U(z_\alpha - y)^2\big] - \alpha^2,\end{aligned}$$

where $F_U(u)$ represents the uniform in $[0, 1)$ cdf. Consequently,

$$F_U(z_\alpha - y)^2 = I\{y \le z_\alpha - 1\} + (z_\alpha - y)^2 I\{z_\alpha - 1 < y \le z_\alpha\}$$

and

$$E_{Y|\mathbf{x}}[F_U(z_\alpha - y)^2] = P(Y \le z_\alpha - 1|\mathbf{x}) + (z_\alpha - y_\alpha)^2 P(Y = y_\alpha|\mathbf{x})$$
$$= \alpha - P(Y = y_\alpha|\mathbf{x})(z_\alpha - y_\alpha)(1 - z_\alpha + y_\alpha),$$

where the last equality follows from expressing $P(Y \le z_\alpha - 1|\mathbf{x})$ as $\alpha - P(Y = y_\alpha|\mathbf{x})(z_\alpha - y_\alpha)$ [see eq. (1)]. Subtracting α^2, we get the desired result.

A.5 Proof of Theorem 5

Let $\mathbf{e}(y, \mathbf{x}) = E(\boldsymbol{\omega}(y, u, \mathbf{x})|y, \mathbf{x})$, with $\boldsymbol{\omega}(\cdot)$ as in the proof of Theorem 4. As we have seen,

$$\mathbf{e}(y, \mathbf{x}) = \mathbf{x}E\big[(\alpha - I\{u \le z_\alpha - y\})|\mathbf{x}, y\big] = \mathbf{x}\big(\alpha - F_U(z_\alpha - y)\big),$$

where $F_U(u)$ represents the uniform in [0, 1) cdf.

The average-jittering estimator has the representation

$$\sqrt{n}\big(\hat{\boldsymbol{\gamma}}_m^{\mathrm{A}}(\alpha) - \boldsymbol{\gamma}(\alpha)\big)$$
$$= \mathbf{D}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{e}(y_i, \mathbf{x}_i)$$
$$+ \mathbf{D}^{-1}\frac{1}{\sqrt{n}}\frac{1}{m}\sum_{i=1}^{n}\sum_{l=1}^{m}\big(\boldsymbol{\omega}\big(y_i, u_i^{(l)}, \mathbf{x}_i\big) - \mathbf{e}(y_i, \mathbf{x}_i)\big) + o_P(1).$$

Because $\boldsymbol{\omega}(y, u, \mathbf{x})$ is obviously square-integrable and, by construction, $E[\boldsymbol{\omega}(y, u, \mathbf{x}) - \mathbf{e}(y, \mathbf{x})|y, \mathbf{x}] = \mathbf{0}$, lemma 2 of Gourieroux and Monfort (1991) implies that the second term on the right side is $O_P(1/\sqrt{m})$. The result then follows by the central limit theorem, noting that (see the proof of Thm. 4), $E(F_U(z_\alpha - y)|\mathbf{x}) = \alpha$, and thus $E\mathbf{e}(y, \mathbf{x}) = \mathbf{0}$, and that $E\mathbf{e}(y, \mathbf{x})\mathbf{e}'(y, \mathbf{x}) = E\{\mathbf{x}\mathbf{x}'E[F_U(z_\alpha - y)^2|\mathbf{x}]\} = \mathbf{B}$.

A.6 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 1 year
PRIVATE INSURANCE	1 if had private medical insurance in the previous year
EDUCATION	Number of years in education after age 16
HEAVY LABOR	1 if has a job in which physically heavy labor 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 5,000 inhabitants
POP-5/20	1 if place of residence has between 5,000 and 20,000 inhabitants
POP-20/100	1 if place of residence has between 20,000 and 100,000 inhabitants
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 7 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%

[Received January 2004. Revised February 2005.]

REFERENCES

Angrist, J. D., Chernozhukov, V., and Fernandez-Val, I. (2004), "Quantile Regression Under Misspecification, With an Application to the U.S. Wage Structure," Working Paper 10428, National Bureau of Economic Research.

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

Chamberlain, G. (1994), "Quantile Regression, Censoring and the Structure of Wages," in *Advances in Econometrics*, ed. C. A. Sims, Cambridge, U.K.: Cambridge University Press, pp. 171–209.

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.

Gourieroux, C., and Monfort, A. (1991), "Simulation-Based Inference in Models With Heterogeneity," *Annales d'Économie et de Statistique*, 20, 69–107.

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*, Palo-Alto, CA: TSP International.

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.

——— (1998), *Semiparametric Methods in Econometrics*, New York: Springer-Verlag.

Huber, P. J. (1981), *Robust Statistics*, New York: Wiley.

Jorgenson, D. W. (1961), "Multiple Regression Analysis of a Poisson Process," *Journal of the American Statistical Association*, 56, 235–245.

Kim, T.-H., and White, H. (2003), "Estimation, Inference and Specification Testing for Possibly Misspecified Quantile Regression," in *Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later*, eds. T. Fomby and R. C. Hill, New York: Elsevier, pp. 107–132.

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.

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

Koenker, R., and Bassett, G. S., Jr. (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.

Koenker, R., and Zhao, Q. (1994), "L-Estimation for Linear Heteroscedastic Models," *Journal of Nonparametric Statistics*, 3, 223–235.

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

Machado, J. A. F., and Mata, J. (2000), "Box–Cox Quantile Regression and the Distribution of Firm Sizes," *Journal of Applied Econometrics*, 15, 289–310.

Machado, J. A. F., and Portugal, P. (2002), "Exploring Transition Data Through Quantile Regression Methods: An Application to U.S. Unemployment Duration," in *Statistical Data Analysis Based on the L_1-Norm and Related Methods*, ed. Y. Dodge, Basel: Birkhäuser-Verlag, pp. 77–94.

Machado, J. A. F., and Santos Silva, J. M. C. (2002), "Quantiles for Counts," Centre for Microdata Methods and Practice CWP22/02, available at *http://cemmap.ifs.org.uk/docs/cwp2202.pdf*.

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

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

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), London: Chapman & Hall.

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

——— (1996), “Measuring the Costs of Children: Parametric and Semiparametric Estimators,” *Statistica Neerlandica*, 50, 171–192.

Nelder, J. A., and Wedderburn, R. W. (1972), “Generalized Linear Models,” *Journal of the Royal Statistical Society*, Ser. A, 135, 370–384.

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

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

——— (1990), “Efficient Estimation of Linear and Type I Censored Regression Models Under Conditional Quantile Restrictions,” *Econometric Theory*, 6, 295–317.

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. (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.

——— (1986), “Censored Regression Quantiles,” *Journal of Econometrics*, 32, 143–155.

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

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.

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. (2003), *Econometric Analysis of Count Data* (4th ed.), Berlin: Springer-Verlag.

——— (2004), “Health Care Reform and the Number of Doctor Visits: An Econometric Analysis,” *Journal of Applied Econometrics*, 19, 455–472.

Zhao, Q. (2001), “Asymptotically Efficient Median Regression in the Presence of Heteroskedasticity of Unknown Form,” *Econometric Theory*, 17, 765–784.