



ELSEVIER

Computational Statistics & Data Analysis 38 (2002) 433–447

**COMPUTATIONAL  
STATISTICS  
& DATA ANALYSIS**

www.elsevier.com/locate/csda

# On estimating conditional quantiles and distribution functions

Franco Peracchi

*Department of Statistics, Tor Vergata University, Rome, Italy*

---

## Abstract

This paper looks at ways of estimating the conditional distribution of a random variable  $Y$  given a vector  $X$  of covariates. We focus on cases where the investigator would like to avoid strong parametric assumptions but the set of covariates is large enough, relative to the available data, to make it problematic to apply standard nonparametric methods because of the well known “curse of dimensionality” problem. In these cases, estimating the conditional quantile function has become increasingly popular. An alternative is to estimate instead the conditional distribution function. Although the choice between these two equivalent representations depends to a large extent on which is more easily interpretable given the purposes of the analysis, considerations of statistical convenience should also be taken into account. We argue that such considerations tend to favor estimating the conditional distribution function directly. © 2002 Elsevier Science B.V. All rights reserved.

*Keywords:* Conditional distribution functions; Conditional log-odds; Logistic regression; Conditional quantile function; Quantile regression

---

## 1. Introduction

This paper looks at ways of estimating the conditional distribution of a random variable  $Y$  given a vector  $X$  of covariates. We focus on cases where the investigator would like to avoid strong parametric assumptions but the set of covariates is large enough, relative to the available data, to make it problematic to apply standard nonparametric methods because of the well known “curse of dimensionality” problem.

In these cases, estimating the conditional quantile function has become increasingly popular (see Buchinsky, 1997 for a survey). An alternative is to estimate instead the

---

*E-mail address:* franco.peracchi@uniroma2.it (F. Peracchi).

conditional distribution function. Although the choice between these two equivalent representations of the conditional distribution of  $Y$  given  $X$  depends to a large extent on which is more easily interpretable given the purposes of the analysis, considerations of statistical convenience should also be taken into account. We argue that such considerations tend to favor estimating the conditional distribution function directly.

The paper is organized as follows. Section 2 reviews the definitions and the relationships between conditional distribution functions and conditional quantile functions, and discusses a number of problems associated with quantile regression. Section 3 describes our approach to estimate conditional distribution functions. Section 4 provides an illustration by analyzing the dependence of the distribution of earnings on education, labor market experience and other observable personal characteristics using micro-economic data from the European Community Household Panel. Finally, Section 5 offers some conclusions.

## 2. Quantile regression

The distribution function (d.f.) and the quantile function (q.f.) are equivalent ways of characterizing the probability distribution  $P$  of a real-valued random variable  $Z$ . To simplify the presentation, let  $Z$  have an absolutely continuous distribution with strictly positive density. In this case, the distribution function is a bounded, continuous and strictly increasing function defined on  $\mathfrak{R}$  by  $F(z) = P\{Z \leq z\}$ , whereas the quantile function is a continuous and strictly increasing function defined on  $(0, 1)$  by  $Q(u) = \{z \in \mathfrak{R}: F(z) = u\}$ . Clearly,  $Q(F(z)) = z$  and  $F(Q(u)) = u$ , that is,  $Q$  and  $F$  are the inverse of each other. One can show that  $Q(u)$  may also be characterized as the unique solution to the problem

$$\min_{z \in \mathfrak{R}} E \ell_u(Z - z),$$

where  $\ell_u$  denotes the asymmetric absolute loss function

$$\ell_u(v) = [u - 1\{v < 0\}]v$$

and  $1\{A\}$  denotes the indicator function of the event  $A$ . If  $u = \frac{1}{2}$ , then  $\ell_u(v) = |v|/2$  is proportional to the familiar symmetric absolute loss function and  $Q(u)$  is the median of  $Z$ .

The generalization to conditional distributions is straightforward. Consider a random vector  $(X, Y)$ , where  $X$  takes values in  $\mathfrak{R}^k$  and  $Y$  is a real-valued continuous random variable with strictly positive density. The conditional probability distribution  $P_x$  of  $Y$  given  $X = x$  may be characterized either through the conditional distribution function (c.d.f.), defined on  $\mathfrak{R} \times \mathfrak{R}^k$  by  $F(y|x) = P_x\{Y \leq y\}$ , or through the conditional quantile function (c.q.f.), defined on  $(0, 1) \times \mathfrak{R}^k$  by  $Q(u|x) = \{y \in \mathfrak{R}: F(y|x) = u\}$ . It is easily verified that  $Q(u|x) = F^{-1}(u|x)$  and  $F(y|x) = Q^{-1}(y|x)$ .

Notice that, for any fixed  $x$ ,  $F(y|x)$  and  $Q(u|x)$  must satisfy all properties of a d.f. and a q.f. respectively. In particular, must have: (i)  $0 < F(y|x) < 1$  for all  $y$ , (ii)  $F(y'|x) > F(y|x)$  whenever  $y' > y$ , and (iii)  $Q(u'|x) > Q(u|x)$  whenever

$u' > u$ . The latter two properties imply that neither the functions  $F(y' | \cdot)$  and  $F(y | \cdot)$  can cross, nor the functions  $Q(u' | \cdot)$  and  $Q(u | \cdot)$ .

The quantile regression problem is how to use the information contained in a sample in order to estimate the c.q.f.  $Q(u|x)$ . In what follows, we focus on the case when  $X$  is a continuous random  $k$ -vector and the sample  $(X_1, Y_1), \dots, (X_n, Y_n)$  consists of  $n$  independent copies of  $(X, Y)$ . A number of approaches may be followed. One possibility is to notice that  $Q(u|x)$ , viewed as a function of  $x$  for a given  $u$ , may also be characterized as the unique solution to the problem

$$\min_{g \in \mathcal{E}} E \ell_u(Y_i - g(X_i)), \quad 0 < u < 1, \tag{1}$$

where  $\mathcal{E}$  is the class of real-valued functions defined on  $\mathfrak{R}^k$ . This suggests estimating  $Q(u|x)$  (for a given  $u$ ) by choosing a function of  $x$ , out of a suitable family  $\mathcal{G} \subseteq \mathcal{E}$ , so as to solve the sample analogue of (1)

$$\min_{g \in \mathcal{G}} n^{-1} \sum_{i=1}^n \ell_u(Y_i - g(X_i)), \quad 0 < u < 1.$$

To recover  $Q(u|x)$ , now viewed as a function of both  $u$  and  $x$ , it is customary to select  $J$  distinct probability values  $0 < u_1 < \dots < u_J < 1$ , and then estimate  $J$  distinct functions  $Q_1, \dots, Q_J$ , each defined on  $\mathfrak{R}^k$ , with  $Q_j(x) = Q(u_j|x)$ ,  $j = 1, \dots, J$ . There may be as many such values as one wishes. By suitably choosing their number and position, one may hope to get a reasonably accurate description of the c.q.f. Given an estimate  $\hat{Q}(u|x)$  of the c.q.f., the c.d.f. may then be estimated by inversion

$$\hat{F}(y|x) = \sup\{u \in (0, 1) : \hat{Q}(u|x) \leq y\}.$$

This is a proper c.d.f. if and only if  $\hat{Q}$  is a proper c.q.f.

The various approaches differ in the choice of the family of functions  $\mathcal{G}$ . In the original parametric approach of Koenker and Bassett (1978),  $\mathcal{G}$  is the class of linear functions of  $x$ , that is,  $Q(u|x) = x^\top \beta(u)$ . In this case, problem (1) reduces to estimating the  $k$ -dimensional parameter  $\beta(u)$  by any solution  $\hat{\beta}(u)$  to the problem

$$\min_{b \in \mathfrak{R}^k} n^{-1} \sum_{i=1}^n \ell_u(Y_i - X_i^\top b), \quad 0 < u < 1. \tag{2}$$

The resulting estimate of  $Q(u|x)$  is  $\hat{Q}(u|x) = x^\top \hat{\beta}(u)$ . The solution to this problem has the nice feature that, just like the univariate case, a fraction  $u$  of the data points lies on or below the estimated quantile regression hyperplane  $\hat{Q}(u|x)$ .

Problem (2) has a nice representation as a linear program for which simple and fast computational algorithms are available. This enables one to efficiently compute the whole set of conditional quantiles for  $u \in (0, 1)$  (see for example, Portnoy and Koenker, 1997). The linear programming formulation also makes it particularly easy to impose linear restrictions on the model parameters.

By suitably redefining the elements of the covariate vector  $X_i$ , one may easily generalize problem (2) to cases in which  $\mathcal{G}$  is a class of functions that depend linearly on a finite dimensional parameter vector, such as the class of polynomial functions of  $X$  of a given degree.

Linear quantile regression estimators can be shown to be  $\sqrt{n}$ -consistent and asymptotically normal under general conditions (see e.g. Ruppert and Carroll, 1980). This result may be used to construct asymptotic tests of hypotheses about the effects of covariates on the spread, the symmetry, the tail behavior and other characteristics of the conditional distribution of  $Y$  (see e.g. Koenker and Bassett, 1982).

Although increasingly used in empirical work to describe the conditional distribution of a variable of interest, linear quantile regression models have several drawbacks. Two crucial issues are: (i) the validity of the linearity assumption, and (ii) the form of the asymptotic variance of linear quantile regression estimators when this assumption does not hold and linear models are simply viewed as convenient approximations.

To illustrate the first issue, let  $X$  be a scalar random variable and let the c.d.f. of  $Y$  given  $X = x$  be of the form  $F(y|x) = F(y - \mu(x))$ , that is, the c.d.f. of  $Y$  depends on  $x$  only through the location parameter  $\mu(x)$ . By definition,  $Q(u|x)$  must satisfy

$$u = F(Q(u|x) - \mu(x)).$$

Inverting this relationship gives

$$Q(u|x) = Q(u) + \mu(x),$$

where  $Q(u) = F^{-1}(u)$  is the  $u$ th quantile of  $F$ . In this case,  $Q(u'|x) - Q(u|x) = Q(u') - Q(u)$  for any  $x$ , that is, the distance between any pair of conditional quantiles is independent of  $x$ . In particular, if  $\mu(x) = \alpha + x\beta$ , then  $Q(u|x) = [\alpha + Q(u)] + x\beta$ , that is, the set of conditional quantiles of  $Y$  is a family of parallel lines with slope equal to  $\beta$ .

Now consider the case when the conditional distribution of  $Y$  depends on  $x$  through both a linear location parameter  $\mu(x) = \alpha + x\beta$  and a scale parameter  $\sigma(x) > 0$ . In this case

$$Q(u|x) = \alpha + \sigma(x)Q(u) + x\beta, \quad (3)$$

which implies that the conditional quantiles of  $Y$  are no longer linear in  $x$ . In addition, the distance between any pair of them varies with  $x$ . Partial exceptions to this general rule are the case when  $F$  is symmetric about zero, implying that the conditional median (but not the other quantiles) is linear in  $x$ , and the case when  $\sigma(x)$  is linear in  $x$ , implying that regression quantiles are linear in  $x$ , although no longer with a common slope.

Result (3) is troubling, for it means that linear quantile regression models may be a poor approximation when data are conditionally heteroskedastic and the square root of the conditional variance function is far from being linear in  $x$ . More seriously, in the general case when  $\mu(x)$  is an arbitrary function, the conditional quantiles of  $Y$  are of the form  $Q(u|x) = \mu(x) + \sigma(x)Q(u)$ . In the absence of prior information, it is now impossible to determine whether nonlinearity of  $Q(u|x)$  is due to nonlinearity of  $\mu(x)$ , heteroskedasticity, or both. From the practical point of view, the presence of heteroskedasticity implies that estimated linear models for conditional quantiles may cross each other, thus violating one of the fundamental properties of quantiles and complicating the interpretation of the results of a statistical analysis.

Turning now to the second issue, if conditional quantiles are nonlinear in  $x$ , then the asymptotic variance matrix of  $\hat{\beta}$  becomes rather complicated (see e.g. Powell, 1984). Inference in this case is typically carried out by bootstrap methods but this trends to be computationally burdensome.

To overcome the problems arising with the linearity assumption, several nonparametric estimators of conditional quantiles have recently been proposed which leave  $\mathcal{G}$  essentially unrestricted, except for smoothness. Among them are various estimators based on kernel or nearest neighbor methods (Antoch and Janssen, 1989; Samanta, 1989; Truong, 1989; Bhattacharya and Gangopadhyay, 1990; Chaudhuri, 1991), regression splines with a fixed number of knots (Hendricks and Koenker, 1992), smoothing splines (Koenker et al., 1994) and penalized likelihood.

There are some problems with all these approaches. First, it is not clear how to impose the “no-crossing” condition on estimated regression quantiles. Second, because these estimators are all nonlinear, it is hard to represent them in ways that facilitate comparisons. For example, it is not clear how to generalize the concepts of equivalent kernel and equivalent degrees of freedom that prove so useful for linear smoothers. Third, it is not clear how to generalize these estimators to cases in which the number of covariates is more than one or two.

### 3. Estimation of the conditional distribution function

Much attention has been devoted in the literature to estimation of a set of c.q.f. and to estimates of the c.d.f. obtained by inverting the c.q.f. However, when the interest is not merely in a few quantiles but in the whole conditional distribution, why not estimating the c.d.f.  $F(y|x)$  directly?

Given a random sample  $(X_1, Y_1), \dots, (X_n, Y_n)$  from the joint distribution of  $(X, Y)$ , Stone (1977) was among the first to suggest nearest neighbor-type estimates of  $F(y|x)$  of the form

$$\hat{F}(y|x) = n^{-1} \sum_{i=1}^n W_i(x) 1\{Y_i \leq y\}, \quad -\infty < y < \infty,$$

where  $W_i(x) = W_i(x; X_1, \dots, X_n)$  weights more heavily  $Y$ -values for which  $X_i$  is closer to  $x$ . Stute (1986) establishes the asymptotic properties of two estimators of this type. The first one is based on kernel weights of the form

$$W_i(x) = \frac{1}{a_n} K \left( \frac{\hat{H}(x) - \hat{H}(X_i)}{a_n} \right),$$

where  $K$  is a smooth kernel with bounded support,  $a_n$  is a bandwidth converging to zero as  $n \rightarrow \infty$  at an appropriate rate, and  $\hat{H}(x)$  is the empirical d.f. of  $X_1, \dots, X_n$ . The resulting estimator of  $F(y|x)$  is nonnegative and nondecreasing but, because the kernel weights do not add to one, it is not a proper d.f. The second estimator overcomes the problem by choosing normalized weights  $W_i^*(x) = W_i(x) / \sum_{j=1}^n W_j(x)$ .

Unfortunately, estimators of this type tend to do rather poorly when data are sparse, which is typically the case with more than two or three covariates. The rest of this

section describes an approach that appears to perform well even when the dimension of  $X$  is large relative to the available data.

### 3.1. A semi-nonparametric approach

Following Foresi and Peracchi (1995), select  $J$  distinct values  $-\infty < y_1 < \dots < y_J < \infty$  in the range of  $Y$ , and consider estimating the  $J$  functions  $F_1(x), \dots, F_J(x)$ , with  $F_j(x) = F(y_j | x)$ ,  $j = 1, \dots, J$ . There may be as many such values as one wishes. By suitably choosing their number and position, one may hope to get a reasonably accurate description of  $F(y | x)$ .

If the conditional distribution of  $Y$  is continuous with support on the whole real line then, at any  $x$  in the support of  $X$ , the sequence of functions  $\{F_j(x)\}$  must satisfy the following two conditions:

$$0 < F_j(x) < 1, \quad j = 1, \dots, J, \quad (4)$$

$$0 < F_1(x) < \dots < F_J(x) < 1. \quad (5)$$

One way of automatically imposing condition (4) is to model not  $F_j(x)$  directly, but rather the log-odds  $\eta_j(x) = \ln [F_j(x)/(1 - F_j(x))]$ . Given an estimate  $\hat{\eta}_j(x)$  of  $\eta_j(x)$ , one may then estimate  $F_j(x)$  by

$$\hat{F}_j(x) = \frac{\exp \hat{\eta}_j(x)}{1 + \exp \hat{\eta}_j(x)}.$$

Let  $\mathcal{H}$  be the class of functions of  $x$  that are possible candidates for the log-odds. Because the random variable  $1\{Y \leq y_j\}$  has a Bernoulli distribution with parameter  $F_j(x)$ , one may define the best Kullback–Leibler approximation  $\eta_j^*(x)$  to  $\eta_j(x)$  in the class  $\mathcal{H}$  as the minimizer of  $\mathcal{K}(\eta, \eta_j) = \ell(\eta_j) - \ell(\eta)$ , with

$$\begin{aligned} \ell(\eta) &= E[1\{Y \leq y_j\}\eta(X) - \ln(1 + \exp \eta(X))] \\ &= E[F_j(X)\eta(X) - \ln(1 + \exp \eta(X))], \end{aligned}$$

where the first expectation is taken with respect to the joint distribution of  $(X, Y)$  and the second with respect to the marginal distribution of  $X$ . Equivalently, the function  $\eta_j^*$  maximizes  $\ell(\eta)$  over  $\mathcal{H}$ . If  $\eta_j \in \mathcal{H}$ , then  $\eta_j^* = \eta_j$ . For example, if  $X$  is a scalar random variable and  $\mathcal{H}$  is the class of functions that are linear in  $x$ , then the best Kullback–Leibler approximation to  $\eta_j(x)$  is of the form  $\eta_j^*(x) = \gamma_j + x\delta_j$ , where  $(\gamma_j, \delta_j)$  are such that the approximation error

$$F_j(X) - \frac{\exp \eta_j^*(X)}{1 + \exp \eta_j^*(X)}$$

has mean zero and is uncorrelated with  $X$ .

By the analogy principle,  $\eta_j^*$  may be estimated by maximizing the sample log-likelihood

$$L(\eta) = n^{-1} \sum_{i=1}^n [1\{Y_i \leq y_j\}\eta(X_i) - \ln(1 + \exp \eta(X_i))]$$

over  $\mathcal{H}$ . This entails fitting  $J$  separate logistic regressions, one for each 0–1 random variable  $1\{Y_i \leq y_j\}$ ,  $j = 1, \dots, J$ . Boundedness of  $1\{Y_i \leq y_j\}$  ensures good robustness properties with respect to outliers in the  $Y$ -space.

Alternative specifications of  $\mathcal{H}$  correspond to alternative estimation methods. As for quantile regression, one may distinguish between parametric methods, nonparametric methods with  $\mathcal{H}$  unrestricted, and nonparametric methods with  $\mathcal{H}$  restricted. The latter class includes projection pursuit, neural networks, additive modelling and semi-additive modelling.

The simplest case is when  $\eta_j(x) = x^\top \theta_j$ , that is, the log odds-ratio is indeed linear in  $x$ . For  $j = 1, \dots, J$ , let  $\hat{\theta}_{jn}$  be the logit estimator of  $\theta_j$  corresponding to a sample of size  $n$  and let  $\hat{\eta}_{jn}(x) = x^\top \hat{\theta}_{jn}$  be the implied estimator of  $\eta_j(x)$ . Under mild regularity conditions,  $\sqrt{n}(\hat{\theta}_{jn} - \theta_j) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}_j^{-1})$  as  $n \rightarrow \infty$ , where

$$\mathcal{I}_j = E[V_{jj}(X)XX^\top]$$

and  $V_{jj}(x) = F_j(x)[1 - F_j(x)]$ . Now let

$$F(x) = \begin{pmatrix} F_1(x) \\ \vdots \\ F_J(x) \end{pmatrix}, \quad \eta(x) = \begin{pmatrix} \eta_1(x) \\ \vdots \\ \eta_J(x) \end{pmatrix} = (I_J \otimes x^\top)\theta, \quad \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_J \end{pmatrix},$$

where  $I_J$  is the unit matrix of order  $J$ . Let  $\hat{F}_n(x)$ ,  $\hat{\eta}_n(x)$  and  $\hat{\theta}_n$  be the estimators of  $F(x)$ ,  $\eta(x)$  and  $\theta$  respectively. Then  $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}^{-1})$  for  $n \rightarrow \infty$ , where

$$\mathcal{I} = E[V(X) \otimes XX^\top]$$

and  $V(x)$  is the  $kJ \times kJ$  matrix with generic element

$$V_{jk}(x) = \min(F_j(x), F_k(x)) - F_j(x)F_k(x).$$

Hence  $\sqrt{n}[\hat{\eta}_n(x) - \eta(x)] \xrightarrow{d} \mathcal{N}(0, A(x))$  as  $n \rightarrow \infty$ , where

$$A(x) = (I_J \otimes x^\top)\mathcal{I}^{-1}(I_J \otimes x),$$

and therefore  $\sqrt{n}[\hat{F}_n(x) - F(x)] \xrightarrow{d} \mathcal{N}(0, V(x)A(x)V(x)^\top)$  as  $n \rightarrow \infty$ . If the log-odds ratio is not linear in  $x$ , then the asymptotic distribution of  $\hat{\theta}_n$  is still Gaussian but its asymptotic variance matrix has the slightly more complicated “sandwich” form.

### 3.2. Imposing monotonicity

While the approach outlined in the previous section automatically imposes condition (4), it does not guarantee that the monotonicity condition (5) holds. Since  $\eta_j(x)$  is strictly increasing in  $F_j(x)$ , monotonicity is equivalent to the condition that  $-\infty < \eta_1(x) < \dots < \eta_J(x) < \infty$  for all  $x$  in the support of  $X$ . Writing  $\eta_j(x) = \gamma_j + \mu_j(x)$ , it is apparent that monotonicity holds if  $\gamma_j > \gamma_{j-1}$  and  $\mu_j(x) \geq \mu_{j-1}(x)$ . One case where these two conditions are satisfied is the ordered logit model, where  $\gamma_j = y_j$  and  $\mu_j(x) = \mu(x)$  for all  $j$ . This model is restrictive, for it implies that changes in the covariate vector  $X$  affect the conditional distribution of  $Y$  only through a location shift.

One alternative is to model  $F_1(x)$  and the conditional probabilities (or discrete hazards)

$$\lambda_j(x) = P\{Y \leq y_j \mid Y > y_{j-1}, X = x\} = \frac{S_{j-1} - S_j(x)}{S_{j-1}(x)}, \quad j = 2, \dots, J,$$

where  $S_j(x) = 1 - F_j(x)$  is the survivor function evaluated at  $y_j$ . Using the recursion

$$S_j(x) = [1 - \lambda_j(x)]S_{j-1}, \quad j = 2, \dots, J,$$

it is easily seen that

$$F_j(x) = 1 - S_1(x) \prod_{m=2}^j [1 - \lambda_m(x)].$$

If we now model the log-odds  $\eta_1(x) = \ln[1 - S_1(x)]/S_1(x)$  and  $\psi_j(x) = \ln[\lambda_j(x)/(1 - \lambda_j(x))]$ , then both monotonicity and the constraint (4) are automatically satisfied.

Turning to estimation,  $\psi_j(x)$ ,  $j = 2, \dots, J$ , may be estimated by maximizing the sample log-likelihood

$$L(\psi) = n^{-1} \sum_{i=1}^n 1\{Y_i > y_{j-1}\} [1\{Y_i \leq y_j\} \psi(X_i) - \ln(1 + \exp \psi(X_i))]$$

over some class of functions  $\Psi$ . This entails fitting  $J - 1$  separate logistic regressions, one for each 0–1 random variable  $1\{Y_i \leq y_j\}$ ,  $j = 2, \dots, J$ , using only the data points for which  $Y_i > y_{j-1}$ .

### 3.3. Generalizations

While it is not even clear how to define quantiles when  $Y$  is vector-valued, the approach in Section 3.1 generalizes very easily. Consider, for simplicity, the bivariate case  $Y = (Y_1, Y_2)$  and let  $F(y \mid x)$  now denote the bivariate d.f. of  $Y$  conditional on  $X = x$ . Given  $J_1$  distinct points in the range of  $Y_1$  and  $J_2$  distinct points in the range of  $Y_2$ , we propose to estimate  $J_1 \times J_2$  functions  $F_{jk}(x)$ , where

$$F_{jk}(x) = F((y_j, y_k) \mid x) = P\{Y_1 \leq y_j, Y_2 \leq y_k \mid X = x\}, \\ j = 1, \dots, J_1, \quad k = 1, \dots, J_2.$$

Given a model for the log-odds  $\eta_{jk}(x) = \ln[F_{jk}(x)/(1 - F_{jk}(x))]$ , this approach entails fitting  $J_1 \times J_2$  separate logistic regressions, one for each binary random variable  $1\{Y_1 \leq y_j, Y_2 \leq y_k\}$ . Once an estimate  $\hat{\eta}_{jk}(x)$  has been obtained,  $F_{jk}(x)$  may be estimated by

$$\hat{F}_{jk}(x) = \frac{\exp \hat{\eta}_{jk}(x)}{1 + \exp \hat{\eta}_{jk}(x)}.$$



Our approach may also provide an alternative to standard autoregressive models for stationary time series, which are essentially models of the conditional mean of the process. Consider, for example, the discrete-time univariate AR(1) process  $Y_t = \rho Y_{t-1} + \sigma U_t$ , where  $|\rho| < 1$ ,  $\sigma > 0$  and  $\{U_t\}$  is an innovation process with zero mean and marginal d.f. equal to  $G$ . The c.d.f. of  $Y_t$  given  $Y_{t-1} = x$  is

$$F(y|x) = P\{Y_t \leq y | Y_{t-1} = x\} = G\left(\frac{y - \rho x}{\sigma}\right). \quad (6)$$

By the stationarity assumption, this is also the conditional distribution of  $Y_{t+h}$  given  $Y_{t+h-1} = x$  for any  $|h| = 0, 1, 2, \dots$ . The assumptions implicit in (6) are strong. As an alternative, one may retain the assumption that  $F(y|x)$  is time-invariant and apply the results of the previous section letting  $X_t = Y_{t-1}$ .

Extensions to more general finite-order Markov processes are straightforward. For example, in the univariate stationary  $m$ th autoregression, what is time-invariant is

$$F(y|x) = P\{Y_t \leq y | X_t = x\},$$

where  $X_t = (Y_{t-1}, \dots, Y_{t-m})$  and  $x = (x_1, \dots, x_m)$ , while in the bivariate stationary  $m$ th autoregression, what is time-invariant is

$$F((y_j, y_k)|x) = P\{Y_{1t} \leq y_j, Y_{2t} \leq y_k | X_t = x\},$$

where  $X_t = (Y_{1,t-1}, \dots, Y_{1,t-m}, Y_{2,t-1}, \dots, Y_{2,t-m})$  and  $x = (x_{11}, \dots, x_{1m}, x_{21}, \dots, x_{2m})$ .

#### 4. Empirical example

As an illustration, we now present the results obtained by using the two approaches modelling the c.q.f. and modelling the c.d.f. — to analyze the dependence of the distribution of earnings on education, labor market experience (defined as the difference between a worker's age and the age when he started working) and other observable personal characteristics in a cross section of individuals.

##### 4.1. The data

Our data are from the public-use files of the European Community Household Panel (ECHP). This data set is a standardized multi-purpose annual longitudinal survey carried out at the level of the European Union (EU). It is centrally designed and coordinated by the Statistical Office of the European Communities (Eurostat), and covers demographics, labor force behavior, income, health, education and training, housing, migration, etc. The first wave of the ECHP, conducted in 1994, included all current members of the EU except Austria, Finland and Sweden. Austria was added in the second wave, conducted in 1995, Finland in the third wave, conducted in 1996, and Sweden in the fourth wave, conducted in 1997. We refer to Peracchi (2001) for a detailed description of the survey.

In this paper, we pool the data from the second and the third wave. To reduce the amount of heterogeneity, we focus on male workers who are full-time private-sector nonagricultural employees, full-time being defined as working at least 30 h per week.

Table 1  
Sample statistics

	Sample size	Quantiles of log earnings			Median experience	Fraction of			
		25th	50th	75th		High ed.	Low ed.	Married	1996
Austria	2442	6.826	7.049	7.316	19	0.045	0.138	0.599	0.487
Belgium	1545	6.928	7.099	7.341	18	0.345	0.283	0.742	0.475
Denmark	1853	6.934	7.116	7.318	21	0.316	0.215	0.571	0.480
Finland	1044	6.775	6.931	7.149	20	0.298	0.211	0.645	1.00
France	2941	6.781	7.019	7.345	20	0.188	0.337	0.657	0.495
Germany	3171	6.932	7.180	7.459	22	0.272	0.209	0.709	0.488
Greece	1835	6.270	6.517	6.773	15	0.187	0.461	0.655	0.470
Ireland	1860	6.829	7.088	7.372	15	0.189	0.390	0.544	0.460
Italy	3794	6.751	6.910	7.073	17	0.056	0.538	0.650	0.487
Luxembourg	715	7.235	7.476	7.804	20	0.171	0.575	0.731	0.483
Netherlands	2931	6.954	7.144	7.350	19	0.158	0.212	0.691	0.508
Portugal	3298	6.042	6.264	6.522	18	0.028	0.879	0.623	0.503
Spain	4008	6.662	6.892	7.203	20	0.183	0.622	0.690	0.482
Uk	2271	6.891	7.199	7.484	21	0.264	0.367	0.715	0.413
Total	33708	6.710	6.988	7.284	19	0.174	0.413	0.658	0.498

We drop observations with missing values on earnings or any of the covariates considered. After our sample selection criteria, the available country subsamples range from a minimum of 715 observations in Luxembourg to a maximum of 4408 observations in Spain, for a total of 33,708 observations. Basic sample statistics are presented in Table 1.

#### 4.2. Model specification

The response variable is the natural logarithm of monthly wage and salary earnings in the current job, net of income and payroll taxes. The original data are in national currencies and have been converted to a common scale by using purchasing power parities. Very similar results are obtained by converting national currencies to ECUs.

Fig. 1 shows the estimated density of the response variable for three experience groups: low (0–15 y), medium (15–30 y) and high (more than 30 y). The densities have been estimated separately by country using the kernel method with a Gaussian kernel and bandwidth equal to  $1.059n\hat{\sigma}$ , where  $n$  is the sample size and  $\hat{\sigma}$  is the estimated standard error of log earnings for each country/experience combination. This bandwidth is an estimate of the optimal one for a Gaussian kernel with Gaussian data. Unlike the density of earnings, which is highly skewed, the one of log earnings resembles a bell-shaped curve, although it is still somewhat skewed. As we move from low to high levels of labor market experience, the density of log earnings shifts to the right and tends to become more dispersed.

Fitting of models for the c.d.f. and the c.q.f. is done separately by country. The probability values for quantile regression range from 0.15 to 0.85, with a constant

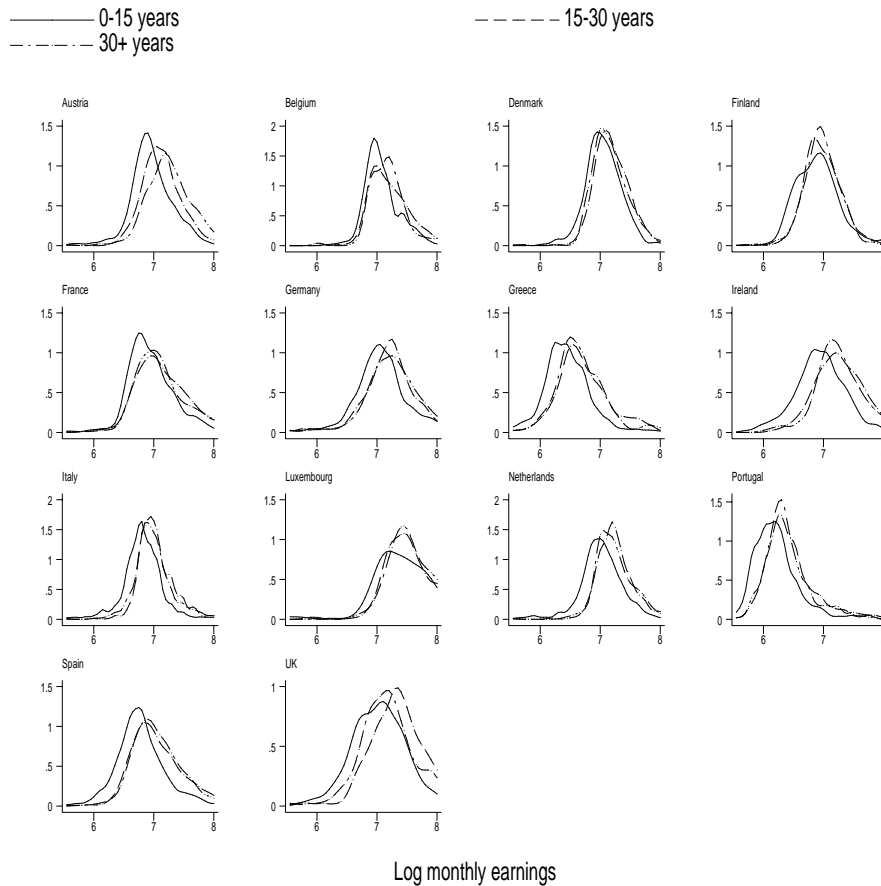


Fig. 1. Estimated densities of log monthly earnings by labor market experience.

increment of 0.05, whereas the thresholds for logistic regression correspond to the unconditional percentiles of log earnings for the pooled data, from the 15th to the 95th with a constant increment of five percentage points. This gives a total of 15 evaluation points for both functions.

The fitted models are all parametric and include as covariates a quadratic function of labor market experience, fully interacted with a set of indicators for the highest level of education completed (one indicator for those with less than the second stage of secondary level education and one for those with a college degree). They also include an indicator for not being married and one for the second (1995) wave. For all models, the intercept corresponds to a baseline worker who is married, has 20 y of labor market experience, and only completed the second stage of secondary level education.

Standard errors for quantile regression estimates were obtained by 30 replications of the non-parametric bootstrap, while standard errors for logistic regression estimates were obtained using the infinitesimal jackknife or “sandwich” estimator.

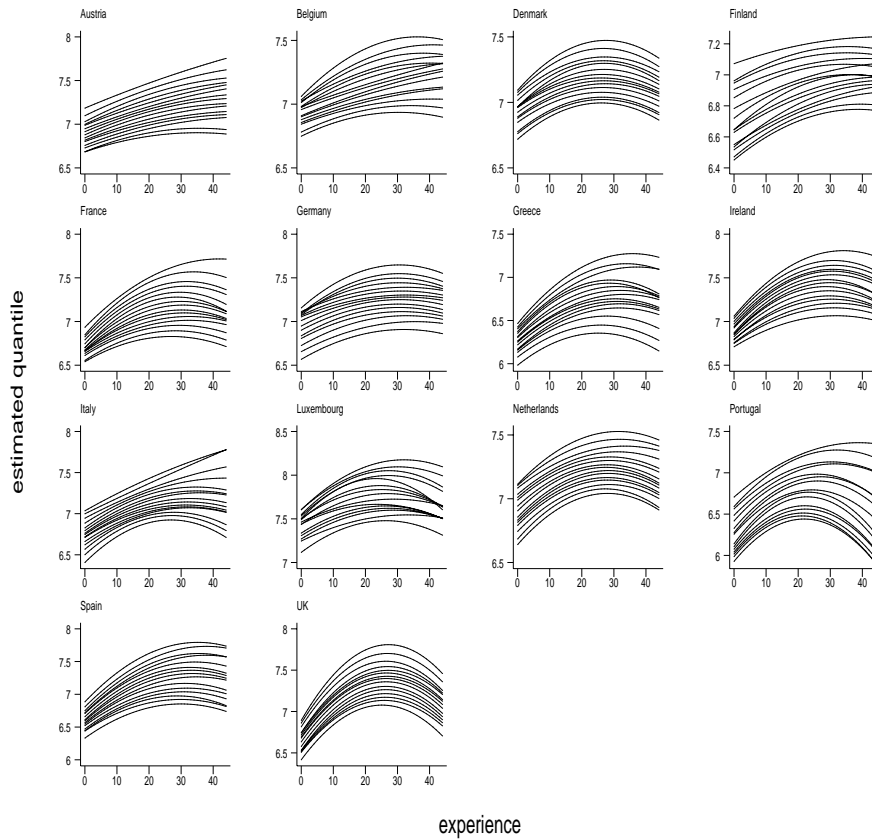


Fig. 2. Estimated conditional quantiles of log monthly earnings.

### 4.3. Empirical results

For brevity, we only present a few graphical displays. Estimated coefficients and standard errors are not reported but are available for the interested reader. In general, the two approaches produce similar results. In particular, other things being equal, earnings increase with education, are higher for married workers, and are substantially lower in Greece and Portugal relative to the other countries.

Fig. 2 shows, for each of the 14 countries considered, the estimated profile of the conditional quantiles of earnings of a baseline worker as a function of labor market experience. The relationship between estimated quantiles and labor market experience is concave with a peak between 20 and 30 y of experience. There is a moderate amount of heteroskedasticity in the data, signalled by some statistically significant differences in estimated coefficients across quantiles. This creates problems with quantiles crossing each other in a few countries. The problem is most serious for Finland and Luxembourg.

Figs. 3 and 4 show the estimated profile of the conditional distribution function of earnings of a baseline worker as a function of labor market experience. Each

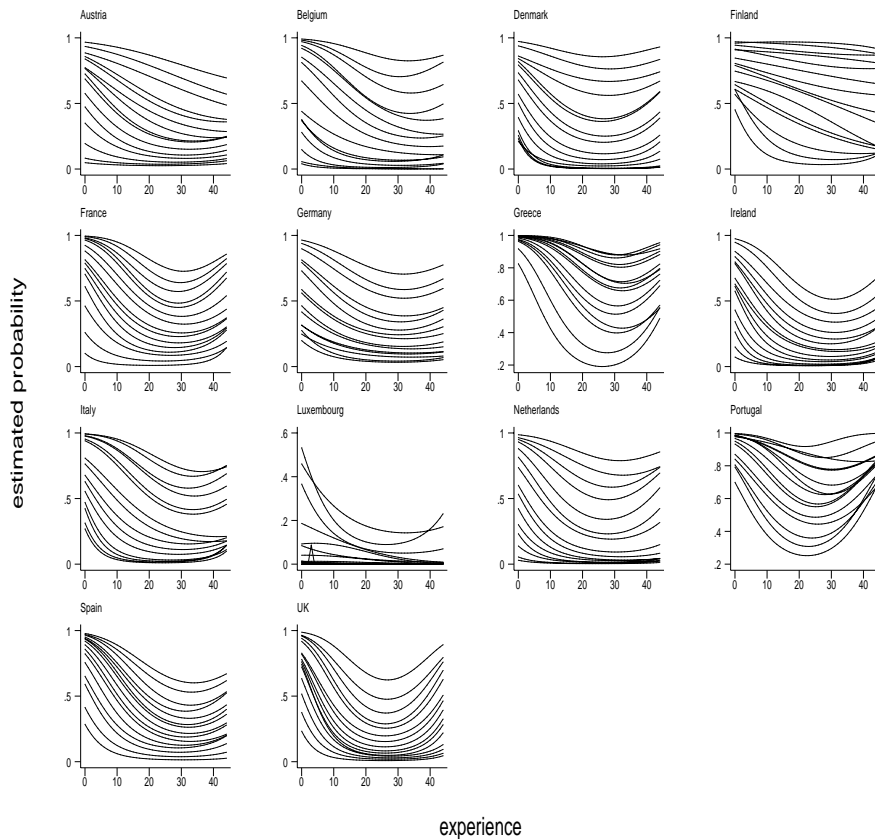


Fig. 3. Estimated conditional distribution functions of log monthly earnings.

of the lines in a quadrant traces the estimated conditional probability of earnings falling below a fixed threshold. Going from bottom to top corresponds to higher and higher thresholds. A negative slope in one of the lines implies a shift to the right in the conditional distribution, a positive slope a shift to the left. Fig. 3 presents the result of the approach outlined in Section 3.1, while, Fig. 4 presents the results of the approach outlined in Section 3.2 in order to impose monotonicity. Qualitatively, the results agree with the ones from Fig. 2. While Fig. 3 shows some evidence of nonmonotonicity, the estimates in Fig. 4 corresponds to a proper c.d.f.

## 5. Conclusions

The approach proposed in this paper offers several advantages over modelling regression quantiles. First, imposing monotonicity is relatively easy, at least in principle, while it is not at all clear how to proceed in the case of regression quantiles. Second, if the log-odds are assumed to belong to a parametric family, evaluating the asymptotic variance of the resulting estimators is straightforward and does not require estimation of the density of the observations. Third, incorporating flexible

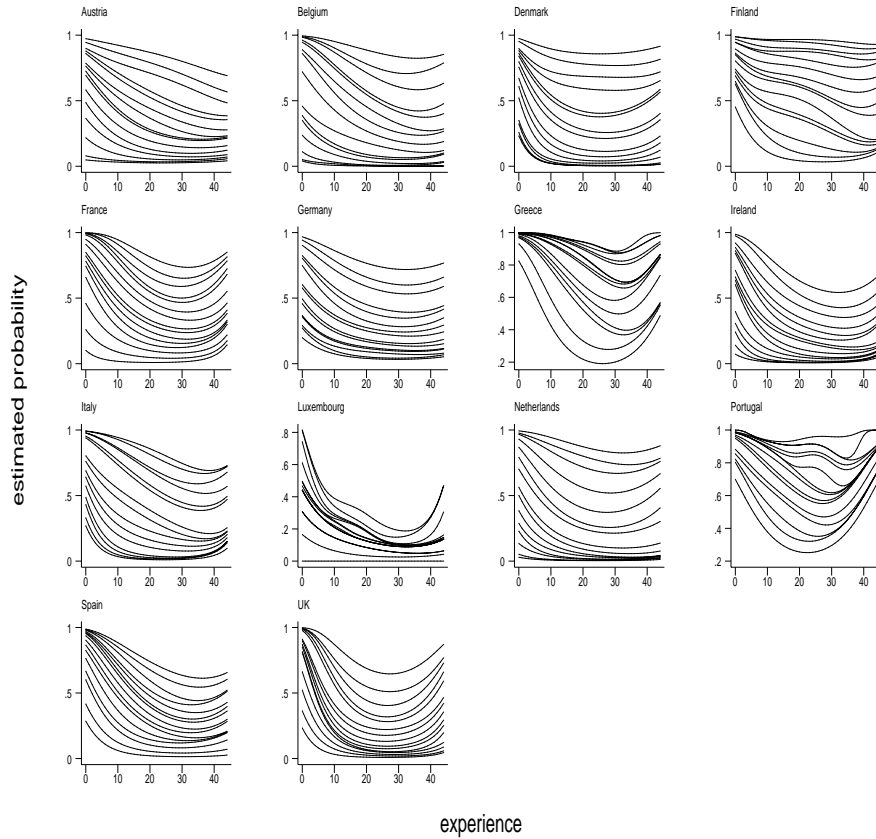


Fig. 4. Estimated conditional distribution functions of log monthly earnings. Discrete hazard modelling.

estimation methods is not a problem because the log-likelihood to maximize belongs to the exponential family of distributions. Fourth, computation may be carried out very quickly by gradient methods, such as the method of scoring. Finally, generalizations to the case in which the response is vector-valued are also straightforward.

### Acknowledgements

I would like to thank Manny Parzen, an anonymous referee and seminar participants at Universidad Carlos III de Madrid, Universitat Pompeu Fabra and Texas A&M University for helpful comments on an earlier version of this paper. Financial support from Consiglio Nazionale delle Ricerche is gratefully acknowledged.

### References

- Antoch, J., Janssen, P., 1989. Nonparametric regression M-quantiles. *Statist. Probab. Lett.* 8, 355–362.  
 Bhattacharya, P.K., Gangopadhyay, A.K., 1990. Kernel and nearest neighbor estimation of a conditional quantile. *Ann. Statist.* 18, 1400–1412.

- Buchinsky, M., 1997. Recent advances in quantile regression models. A practical guideline for empirical research. *J. Human Resources* 33, 89–126.
- Chaudhuri, P., 1991. Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* 19, 760–777.
- Foresi, S., Peracchi, F., 1995. The conditional distribution of excess returns: an empirical analysis. *J. Amer. Statist. Assoc.* 90, 451–466.
- Hendricks, W., Koenker, R., 1992. Hierarchical spline models for conditional quantiles and the demand for electricity. *J. Amer. Statist. Assoc.* 87, 58–68.
- Koenker, R., Bassett, G., 1978. Regression quantiles. *Econometrica* 46, 33–50.
- Koenker, R., Bassett, G., 1982. Robust tests for heteroskedasticity based on regression quantiles. *Econometrica* 50, 43–61.
- Koenker, R., Ng, P., Portnoy, S., 1994. Quantile smoothing splines. *Biometrika* 81, 673–680.
- Peracchi, F., 2001. The European Community Household Panel: a review. *Empirical Economics* forthcoming.
- Portnoy, S., Koenker, R., 1997. The Gaussian hare and the Laplacian tortoise: computability for squared-error versus absolute-error estimators (with discussion). *Statist. Sci.* 12, 279–300.
- Powell, J., 1984. Least absolute deviations estimation for the censored regression model. *J. Econometrics* 25, 303–325.
- Ruppert, D., Carroll, R.J., 1980. Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* 75, 828–838.
- Samanta, M., 1989. Nonparametric estimation of conditional quantiles. *Statist. Probab. Lett.* 7, 407–412.
- Stone, C.J., 1977. Consistent nonparametric regression (with discussion). *Ann. Statist.* 5, 595–645.
- Stute, W., 1986. Conditional empirical processes. *Ann. Statist.* 14, 638–647.
- Truong, Y.K., 1989. Asymptotic properties of kernel estimators based on local medians. *Ann. Statist.* 17, 606–617.