EIEF WORKING PAPER SERIES

EIEF

Einaudi Institute for Economics and Finance

EIEF Working Paper 29/13

December 2013

Distributional vs. Quantile Regression

by

Roger Koenker

(University of Illinois at Urbana-Champaign)

Samantha Leorato

(University of Rome "Tor Vergata")

Franco Peracchi

(University of Rome "Tor Vergata" and EIEF)

Distributional vs. Quantile Regression*

Roger Koenker
University of Illinois at Urbana-Champaign

Samantha Leorato
University of Rome Tor Vergata

Franco Peracchi
University of Rome Tor Vergata and EIEF

December 10, 2013

Abstract

Given a scalar random variable Y and a random vector X defined on the same probability space, the conditional distribution of Y given X can be represented by either the conditional distribution function or the conditional quantile function. To these equivalent representations correspond two alternative approaches to estimation. One approach, distributional regression (DR), is based on direct estimation of the conditional distribution function; the other approach, quantile regression (QR), is instead based on direct estimation of the conditional quantile function. Indirect estimates of the conditional quantile function and the conditional distribution function may then be obtained by inverting the direct estimates obtained from either approach. Despite the growing attention to the DR approach, and the vast literature on the QR approach, the link between the two approaches has not been explored in detail. The aim of this paper is to fill-in this gap by providing a better understanding of the relative performance of the two approaches, both asymptotically and in finite samples, under the linear location model and certain types of heteroskedastic location-scale models.

KEYWORDS. Quantile regression, distributional regression, functional Delta-method, asymptotic relative efficiency, linear location model, location-scale models.

JEL CLASSIFICATION. C1, C21, C25.

* We thank Blaise Melly for comments and Jana Jurečková for many useful discussions. Corresponding author: Franco Peracchi (franco.peracchi@uniroma2.it).

1 Introduction

In this paper we consider two alternative approaches to the problem of estimating the conditional distribution of a scalar random variable Y given a random vector X when the available data are a sample from the joint distribution of (Y, X). One approach estimates the conditional distribution function (CDF) $F(y \mid x) = \Pr\{Y \leq y \mid X = x\}$, $y \in \mathbb{R}$ (distributional regression or DR approach). The other approach estimates the conditional quantile function (CQF) $Q(p \mid x) = \inf\{y \in \mathbb{R}\colon\ F(y \mid x) \geq p\}$, $p \in (0,1)$ (quantile regression or QR approach).

The CDF and the CQF are equivalent characterizations of the conditional distribution of Y given X. In particular, if $F(y \mid x)$ is continuous and strictly increasing in y, then $Q(p \mid x)$ is also continuous and strictly increasing in p. Further, for all y, p and x, $Q(F(y \mid x) \mid x) = y$ and $F(Q(p \mid x)) = p$, that is, each function is the inverse of the other. Given a continuous and strictly increasing estimate $\hat{F}(\cdot \mid x)$ of the CDF, one can then obtain an indirect estimate $\hat{Q}(\cdot \mid x)$ of the CQF by solving the implicit equation $p = \hat{F}(Q(p \mid x))$ for all p and x or, equivalently, by setting $\hat{Q}(p \mid x) = \inf\left\{y\colon\ \hat{F}(y \mid x) \geq p\right\}$. In the unconditional case, an early example of this approach is Nadaraya (1964). Similarly, given a continuous and strictly increasing estimate $\tilde{Q}(\cdot \mid x)$ of the CQF, one can obtain an indirect estimate $\tilde{F}(\cdot \mid x)$ of the CDF by solving the implicit equation $y = \tilde{Q}(F(y \mid x))$ for all y and all x or, equivalently, by setting $\tilde{F}(y \mid x) = \inf\left\{p\colon\ \tilde{Q}(p \mid x) \geq y\right\}$.

Although, any estimation procedure that directly estimates the CDF can be included in the DR approach, in this paper and in most of the recent literature, the term DR refers to estimation of the CDF at a finite number of cutoffs $y_1, \ldots, y_J$ by estimating a model for the conditional mean of the binary indicators $D_j = 1\{Y \leq y_j\}$, $j = 1, \ldots, J$. This approach, first suggested by Foresi and Peracchi (1995), has recently been explored by Fortin, Lemieux and Firpo (2011), Rothe (2012), Rothe and Wied (2012), and Chernozhukov, Fernández-Val and Melly (2013).

Direct estimation of the CQF at a finite number of quantile levels $p_1, \ldots, p_J$ was first addressed by Koenker and Bassett (1978), who introduced the asymmetric LAD estimator. Since then, several generalizations and variants have been proposed,including nonparametric and semiparametric estimators, penalized likelihood methods, methods for non identically distributed or dependent observations, isotonic estimators, extremal quantile regression, and weighted quantile regression (see Koenker 2005 for a review).

Despite the growing attention to the DR approach, and the vast literature on the QR approach, the link between the two approaches has not been explored in detail, although some considerations on the choice between them appear in Peracchi (2002) and Chernozhukov, Fernández-Val and Melly (2013).

The aim of this paper is to fill-in this gap by providing a better understanding of the relative performance of the two approaches, both asymptotically and in finite samples, under the linear location model and certain types of heteroskedastic location-scale models.

Of course, the choice between the two approaches may be based on aspects other than their statistical properties. One such aspect is possible generalization to the problem of estimating the conditional distribution of a random vector Y. Here the CDF approach appears to be more natural. Another aspect is interpretability. For example, if Y is household income and X is a binary indicator representing educational attainments (higher or lower), then the DR approach allows one to directly estimate differences across educational groups in the probability that income will fall below a given threshold. On the other hand, the QR approach allows one to directly estimate income differences between people ranked the same in the two groups. Thus, if the aim of the analysis is to study differences in the poverty rate between the two groups, then the DR approach may be more natural. If the aim is instead to compare income differentials within the two groups, then the QR approach may be more natural.

One important property that characterizes the CDF $F(y \mid x)$ and the CQF $Q(p \mid x)$ is monotonicity, in y and p respectively, for all x. Thus, a relevant issue for both approaches is how to impose this property on the estimates. For the CDF, Foresi and Peracchi (1995) propose an isotonic estimation procedure. For the CQF, Chernozhukov, Fernández-Val and Galichon (2010) propose a method based on rearrangement of preliminary CQF estimates. A similar approach is followed by Dette and Volgushev (2008). >From an asymptotic perspective, the question of non-monotonicity becomes largely irrelevant, provided that both the DR and QR estimators are consistent for the CDF and CQF respectively. This provides a justification for focusing here on non-isotonic QR and DR estimators.

The remainder of the paper is organized as follows. Section 2 introduces our basic statistical model, the linear location model. Section 3 compares the QR and DR estimators of the parameters of the linear location model, focusing on their asymptotic relative efficiency. Section 4 studies the asymptotic properties of the QR and DR estimators of the CDF and the CQF, focusing on their asymptotic relative efficiency under the linear location model. Section 5 considers the more general class of location-scale models. Section 6 presents the results of a set of Monte Carlo experiments. Finally, Section 7 summarizes and offers some conclusions.

2 The linear location model

We begin by assuming that the available data $\{(X_i, Y_i), i = 1, \ldots, n\}$ are a sample from the linear location model

$$Y = \alpha + X^\top \beta + \omega U, \tag{1}$$

where β is a vector of k parameters, X is a random k-vector with finite variance, ω is a strictly positive scale parameter, and U is a random error distributed independently of X with a smooth and strictly increasing distribution function G. To keep notation simple, we assume without loss of generality that ω is known and equal to one.

The CDF for this model is

$$F(y \mid x) = G(y - \alpha - x^\top \beta) = G(\gamma_y + x^\top \delta),$$

where $\gamma_y = y - \alpha$ and $\delta = -\beta$, while the CQF is

$$Q(p \mid x) = \alpha + x^\top \beta + G^{-1}(p) = \xi_p + x^\top \beta,$$

where $\xi_p = \alpha + G^{-1}(p)$. Notice that $\xi_p = \alpha$ if $G^{-1}(p) = 0$. In particular, $\xi_{1/2} = \alpha$ if the distribution of U has zero median. Also notice that $\partial F(y \mid x)/\partial y = g(\gamma_y + x^\top \delta)$ and $\partial F(y \mid x)/\partial x = \delta\, g(\gamma_y + x^\top \delta)$, where $g(u) = G'(u)$ is the density function of the error U. Thus, the linear location model (1) implies the following relationship

$$-\frac{\partial F(y \mid x)/\partial x}{\partial F(y \mid x)/\partial y} = \beta = \frac{\partial Q(p \mid x)}{\partial x},$$

for all x, y and p. This relationship justifies the common practice of focusing attention on the slope coefficient β, treating the intercept α as a nuisance parameter.

3 Estimating the parameters of the linear location model

The problem of estimating the vector of parameters $\theta = (\alpha, \beta)$ in the linear location model (1) may be regarded as a preliminary step towards the final goal of estimating the conditional distribution of Y given X.

For a given cutoff y, the DR approach first obtains an estimate $(\hat{\gamma}_y, \hat{\delta}_y)$ of (γ_y, δ) in the binary regression model $\mathrm{E}(D_y \mid X = x) = G(\gamma_y + x^\top \delta)$, where $D_y = 1\{Y \leq y\}$ is a 0-1 indicator. It then estimates θ by $\hat{\theta}_y = (\hat{\alpha}_y, \hat{\beta}_y)$, where $\hat{\alpha}_y = y - \hat{\gamma}_y$ and $\hat{\beta}_y = -\hat{\delta}_y$. When G is known, asymptotically efficient estimation is typically based on the method of maximum likelihood (ML). When G is unknown, the semi-parametric method proposed by Klein and Spady (1993) may instead be used.

For a given quantile level p, the QR approach first obtains an estimate $(\tilde{\xi}_p, \tilde{\beta}_p)$ of (ξ_p, β) in the model $Q(p\,|\,x) = \xi_p + x^\top \beta$. It then estimates θ by $\tilde{\theta}_p = (\tilde{\alpha}_p, \tilde{\beta}_p)$, where $\tilde{\alpha}_p = \tilde{\xi}_p - G^{-1}(p)$. Notice that the QR approach allows one to estimate the slope vector β without any knowledge of G. This is an important advantage relative to the DR approach, where consistent estimation of β is only possible if G is known or is replaced by a preliminary consistent estimate. Also notice that, unlike the model parameters α and β which do not vary with y and p, the DR estimates vary with y and the QR estimates vary with p because they are solutions to optimization problems that depend on y and p respectively.

3.1 DR approach

For a given cutoff $y \in \mathbb{R}$, a ML estimator $(\hat{\gamma}_y, \hat{\delta}_y)$ maximizes the average log-likelihood

$$L(a, b; y) = n^{-1} \sum_{i=1}^{n} \left[D_{iy} \ln \frac{\pi_i(a, b)}{1 - \pi_i(a, b)} + \ln(1 - \pi_i(a, b)) \right],$$

where $D_{iy} = 1\{Y_i \leq y\}$ and $\pi_i(a, b) = G(a + X_i^\top b)$. If the linear location model (1) holds and $n \to \infty$ then, under mild regularity conditions, $\hat{\theta}_y = (y - \hat{\gamma}_y, -\hat{\delta}_y)$ is consistent for θ and the rescaled difference $\sqrt{n}(\hat{\theta}_y - \theta)$ is asymptotically Gaussian with asymptotic variance

$$V_D(y) = \left[\mathrm{E}\, w_i(y) \boldsymbol{X}_i \boldsymbol{X}_i^\top \right]^{-1}, \tag{2}$$

where $\boldsymbol{X}_i^\top = (1, X_i)$, $w_i(y) = w(G(\gamma_y + X_i^\top \delta))$ and

$$w(p) = \frac{g(G^{-1}(p))^2}{p(1-p)}, \qquad 0 < p < 1. \tag{3}$$

For example, if G is equal to the logistic distribution with zero mean and unit variance (the "logit link" case), then $w(p) = c^2 p(1-p)$, where $c = \pi/\sqrt{3}$.

3.2 QR approach

For a given quantile level $p \in (0, 1)$, a QR estimator $(\tilde{\xi}_p, \tilde{\beta}_p)$ solves the problem

$$\min_{a,b} n^{-1} \sum_{i=1}^{n} \rho_p \left(Y_i - a - X_i^\top b \right),$$

where $\rho_p(u) = u[p - 1\{u \leq 0\}]$ is the asymmetric absolute loss function. If the linear location model (1) holds and $n \to \infty$ then, under mild regularity conditions, $\tilde{\theta}_p = (\tilde{\xi}_p - G^{-1}(p), \tilde{\beta}_p)$ is consistent for θ and the rescaled difference $\sqrt{n}(\tilde{\theta}_p - \theta)$ is asymptotically Gaussian with asymptotic variance equal to

$$V_Q(p) = \left[\mathrm{E}\, w(p) \boldsymbol{X}_i \boldsymbol{X}_i^\top \right]^{-1} = \frac{1}{w(p)} P_X^{-1}, \tag{4}$$

where the function $w(\cdot)$ is defined in (3) and $P_X = \mathrm{E}\, \boldsymbol{X}_i \boldsymbol{X}_i^\top$. Notice that P^{-1} is the asymptotic variance of the OLS estimator of θ.

The asymptotic variances $V_D(y)$ and $V_Q(p)$ look quite similar, with the function w appearing in both. However, while $V_Q(p)$ depends on the constant weight $w(p)$, $V_D(y)$ depends on the stochastic weights $w_i(y)$. When $\beta = 0$, or equivalently the regression R^2 is equal to 0, the two asymptotic variances coincide as $w_i(y) = w(p)$ for all p and y such that $p = G(y - \alpha)$.

3.3 Relative efficiency

Consider now the performance of the QR estimator of θ relative to the DR estimator under the linear location model (1). Since both estimators are $\sqrt{n}$-consistent and asymptotically Gaussian, comparison may be based on their asymptotic relative efficiency (ARE).

Recall that if $\hat{\theta}_1$ and $\hat{\theta}_2$ are two $\sqrt{n}$-consistent and asymptotically Gaussian estimators of the same scalar parameter θ, with asymptotic variances V_1 and V_2 respectively, then the ARE of $\hat{\theta}_2$ to $\hat{\theta}_1$ is the ratio $ARE(\hat{\theta}_2, \hat{\theta}_1) = V_1/V_2$. If $\hat{\theta}_2$ is used to estimate θ with a sample of size n, then $n\,ARE(\hat{\theta}_2, \hat{\theta}_1)$ is approximately equal to the number of observations needed for $\hat{\theta}_1$ to be as precise as $\hat{\theta}_2$. The concept of ARE extends to the k-dimensional case by taking the ratio of the determinants of the asymptotic variance matrices raised to the power $1/k$ (Serfling 1980).

First notice that, from (4), the ARE of QR to OLS depends only on p and is equal to $w(p)$. On the other hand, the ARE of QR to DR depends not only on the error distribution in (1), but also on the distribution of the regressors. General results are not available, so we focus on a few specific examples. We confine attention to the case of a single regressor X and consider three distributions that are particularly relevant in empirical applications: the Bernoulli, the uniform and the Gaussian.

1. Bernoulli regressor. If $X \sim \mathscr{B}(\mu_X)$, the Bernoulli distribution with mean μ_X, we have simple closed-form expressions for the estimators of the parameter β and for the matrix V_D.

The DR estimator of β at a given threshold y is $\hat{\beta}_y = G^{-1}(n_{0y}/n_1) - G^{-1}(n_{1y}/n_0)$, where n_{xy}/n_x is the fraction of observations for which $Y_i \leq y$ and $X_i = x$, with $x = 0, 1$. On the other hand, the QR estimator of β at a given quantile level p is $\tilde{\beta}_p = Y^1_{(p)} - Y^0_{(p)}$, where $Y^x_{(p)}$ is the pth sample quantile in subsample $x = 0, 1$. Although the DR and QR estimators of β differ under the two approaches, it is easily shown that the direct and the indirect estimators of the CDF and the CQF coincide, namely $\hat{F}(y \,|\, X = x) = \tilde{F}(y \,|\, X = x) = n_{xy}/n_x$ and $\tilde{Q}(p \,|\, X = x) = \hat{Q}(p \,|\, X = x) = \inf\{Y_i : n_{x,Y_i}/n_x \geq p\}$.

The asymptotic variance of the DR estimator of θ under the linear location model (1) is

$$V_D(y) = \frac{1}{w(G(\gamma_y))\,\sigma_X^2}\begin{bmatrix} \mu_X & -\mu_X \\ -\mu_X & \mu_X + (1-\mu_X)\frac{w(G(\gamma_y))}{w(G(\gamma_y-\beta))} \end{bmatrix}, \tag{5}$$

where $\sigma_X^2 = \mu_X(1-\mu_X)$ is the variance of X. Hence, the ARE of QR to DR is

$$ARE(\tilde{\theta}_p, \hat{\theta}_y) = \left(\frac{|V_D(y)|}{|V_Q(p)|}\right)^{1/2} = \frac{w(p)}{\sqrt{w(G(\gamma_y))\,w(G(\gamma_y-\beta))}}.$$

In the special case when $y = y_p = \alpha + G^{-1}(p)$ we have

$$ARE(\tilde{\theta}_p, \hat{\theta}_{y_p}) = \sqrt{\frac{w(p)}{w(G(G^{-1}(p)-\beta))}}. \tag{6}$$

When the function $w(p)$ is symmetric around $p = .5$ and nondecreasing in the interval $(0, .5)$, (6) implies that $ARE(\tilde{\theta}_p, \hat{\theta}_{y_p}) > 1$ if and only if $\beta > 0$ for $p < .5$, or $\beta < 0$ for $p > .5$. As shown in Figure 1, both the logistic and the Gaussian distributions satisfy this condition. Other distributions share a similar behavior, for example the Student's t, the Laplace, the Cauchy, and the Generalized Tukey for particular values of its parameters. In all these cases, $ARE(\tilde{\theta}_{.5}, \hat{\theta}_{y_{.5}}) = \sqrt{w(0)/w(G(-\beta))} \geq 1$, so QR does better than DR irrespective of β unless $\beta = 0$, in which case the asymptotic variances of the two estimators coincide.

Notice that, in this example, the conditional distributions of Y_i when $X_i = 0$ and $X_i = 1$ differ only in location, not scale or shape. The general two-sample model, which may be represented as $Y_i = \alpha + \beta X_i + U_i$, where the distribution function of U_i is G_0 in the subsample with $X_i = 0$ and G_1 in the subsample with $X_i = 1$, does not satisfy (1) unless $G_0 = G_1$. The case when $G_1(u) = G_0(u/(1+\eta))$ is a special case of the linear location-scale model considered in Section 5. When G_0 and G_1 are fully unrestricted, the direct and the indirect estimators of the CDF and the CQF again coincide with the empirical distribution functions and the empirical quantile functions in the two subsamples, but identification and consistent estimation of the slope parameter β requires additional restrictions. In the DR approach one needs knowledge of both G_0 and G_1 to consistently estimate β by $\hat{\beta} = G_0^{-1}(n_{0y}/n_0) - G_1^{-1}(n_{1y}/n_1)$. In the QR approach, since $Q(p\,|\,x=0) = \alpha + G_0^{-1}(p)$, while $Q(p\,|\,x=1) = \alpha + \beta + G_1^{-1}(p)$, one can identify β only if $G_1^{-1}(p) - G_0^{-1}(p)$ is known for some p.

2. Uniform regressor. Computing the ARE of QR to OLS for non-binary regressors requires specifying the distribution of the random error U in (1). Here and in the next example, we focus on the logit link as the function $w(p)$ takes a particularly simple form. Even in this case, however, a closed-form expression for $V_D(y)$ is generally not available.

One exception is when X is distributed as $\mathscr{U}(a,b)$, the uniform distribution on the interval (a,b). The formula for $V_D(y)$ is not particularly illuminating but, when $X \sim \mathscr{U}(0,1)$ and $\alpha = 0$, it simplifies to

$$V_D(y) = \beta^2 \begin{bmatrix} C_{y,0} - C_{y,\beta} & C_{y,\beta} - \ln \frac{G_{y,0}}{G y,\beta} \\ C_{y,\beta} - \ln \frac{G_{y,0}}{G_{y,\beta}} & C_{y,\beta} + 2 \ln G_{y,\beta} + \frac{2}{c\beta} \left[\mathrm{Li}_2(-e^{-cy}) - \mathrm{Li}_2(-e^{-c(y-\beta)}) \right] \end{bmatrix}^{-1}, \qquad (7)$$

where $C_{y,\beta} = c\beta\,[1 - G(y-\beta)]$, $G_{y,\beta} = G(y-\beta)$ and $\mathrm{Li}_2(x) = -\int_0^x t^{-1} \log(1-t)\,dt$ is the dilogarithmic function.

Figure 2 shows the contour plots of the asymptotic efficiency gain of the QR estimator relative to DR estimator as a function of both p and y. Specifically, the figure shows the contour plots of $ARE(\tilde{\theta}_p, \hat{\theta}_y) - 1$ when $ARE(\tilde{\theta}_p, \hat{\theta}_y) - 1 \geq 0$, and of $1 - ARE(\hat{\theta}_y, \tilde{\theta}_p)$ when $ARE(\tilde{\theta}_p, \hat{\theta}_y) - 1 < 0$. For example, a value of 5 means a 500% gain in asymptotic efficiency of QR relative to DR, while a value of -5 means a 500% gain in asymptotic efficiency of DR relative to QR. The blue (orange-red) area corresponds to the pairs of p and y values for which the QR (DR) estimator is asymptotically more efficient than the DR (QR) estimator. The darker the color, the higher the asymptotic efficiency gain. The figure shows that the asymptotic efficiency of QR relative to DR tends to deteriorate when either p or $1-p$ are small. On the other hand, the further away is y from the median of Y, the larger is the asymptotic efficiency of QR relative to DR. As already pointed out, the behavior of the two estimators tends to converge as the regression R^2 tends to 0. This is highlighted in Panel (d) of Figure 2, where we plot $ARE(\tilde{\theta}_p, \hat{\theta}_y)$ for $R^2 = 0$. The curves that separate the blue and the orange-red areas, where QR and DR respectively are more efficient, satisfy $ARE(\tilde{\theta}_p, \hat{\theta}_y) = 1$ and correspond to the equations $y = \alpha + G^{-1}(p)$ and $y = \alpha + G^{-1}(1-p)$.

3. Gaussian regressor. When X is Gaussian, the asymptotic variance of DR has no closed-form expression, not even when U has a logistic distribution, but is easily obtained by numerical methods. The results in this case are qualitatively similar to those in Figure 2, so we do not show them here although they are available upon request.

4 Estimating the conditional distribution

In this section, we consider the statistical properties of the DR and QR approaches when the goal is estimating the conditional distribution of Y given X.

4.1 DR approach

For simplicity, we focus on the logit link and denote by $\Lambda(u)$ the distribution function of the unit logistic distribution, namely $\Lambda(u)=[1+\exp(-\pi u/\sqrt{3})]^{-1}$.

The direct estimator of the CDF at a given cutoff y is $\hat{F}(y\,|\,x)=\Lambda(\hat{\gamma}_y+x^\top\hat{\delta}_y)=\Lambda(\boldsymbol{x}^\top\hat{\nu}(y))$, where $\boldsymbol{x}=(1,\,x^\top)^\top$ and $\hat{\nu}(y)=(\hat{\gamma}_y,\hat{\delta}_y^\top)^\top$ maximizes the average log-likelihood

$$L(\nu;y)=n^{-1}\sum_{i=1}^{n}\left[D_{iy}\ln\frac{\Lambda(\boldsymbol{X}_i^\top\nu)}{1-\Lambda(\boldsymbol{X}_i^\top\nu)}+\ln(1-\Lambda(\boldsymbol{X}_i^\top\nu))\right]. \tag{8}$$

The next theorem, which establishes the asymptotic properties of the stochastic process $\hat{\nu}(\cdot)$ defined on $\mathbb{R}$, follows from Theorem 5.2 in Chernozhukov, Fernàndez-Val and Melly (2013). Its proofs is in Appendix A, along with all other proofs. >From now on, all convergence results hold for $n\to\infty$.

Theorem 1 *Suppose that: (i) $\{(Y_i,X_i),i=1,\dots,n\}$ is a sample from the joint distribution of (Y,X); (ii) $\mathrm{E}\,|X|^{2+\varepsilon}<\infty$ for some $\varepsilon>0$; (iii) for any $y\in\mathbb{R}$, $\hat{\nu}(y)$ maximizes the average logit log-likelihood $L(\nu;y)$ on a compact subset Θ of $\mathbb{R}^2$; (iv) for any $y\in\mathbb{R}$, the expected logit log-likelihood*

$$\ell(\nu;y)=\mathrm{E}\left[D_y\ln\frac{\Lambda(\boldsymbol{X}^\top\nu)}{1-\Lambda(\boldsymbol{X}^\top\nu)}+\ln(1-\Lambda(\boldsymbol{X}^\top\nu))\right]$$

attains a unique maximum on Θ at $\nu_D(y)$; (v) the matrix $V_D(y)$ is positive definite for all y in a closed interval $[\underline{y},\overline{y}]$ of $\mathbb{R}$. Then, $\hat{\nu}(\cdot)$ is uniformly consistent for $\nu_D(\cdot)$, that is, $\sup_{\underline{y}\le y\le\overline{y}}\|\hat{\nu}(y)-\nu_D(y)\|=o_p(1)$. Further, the process $V_D^{-1}(\cdot)\sqrt{n}[\hat{\nu}(\cdot)-\nu_D(\cdot)]$ converges weakly to a zero-mean multivariate Gaussian process $B_D(\cdot)$ defined on $[\underline{y},\overline{y}]$ with covariance function

$$\Sigma_D(y,y')=\mathrm{E}\left[\varphi(y)\,\varphi(y')\boldsymbol{X}\boldsymbol{X}^\top\right], \tag{9}$$

where $\varphi(y)=c(D_y-\Lambda(\boldsymbol{X}^\top\nu_D(y)))$ and $c=\pi/\sqrt{3}$.

If model (1) with logistic errors is correctly specified, then the maximizer of the expected log-likelihood $\ell(\nu;y)$ is equal to $\nu_D(y)=(y-\alpha,-\beta^\top)^\top$ and the covariance function of the process in Theorem 1 simplifies to

$$\Sigma_D(y,y')=c^2\,\mathrm{E}\left[\Lambda(\boldsymbol{X}^\top\nu_D(y))\left(1-\Lambda(\boldsymbol{X}^\top\nu_D(y))\right)\boldsymbol{X}\boldsymbol{X}^\top\right],\qquad y\le y',$$

so $\Sigma_D(y,y)=V_D(y)^{-1}$. In this case, the asymptotic variance of $\hat{F}(y\,|\,x)$ is

$$\begin{aligned}V(\hat{F}(y\,|\,x))&=\Lambda'(\boldsymbol{x}^\top\nu_D(y))^2\,\boldsymbol{x}^\top V_D(y)\,\Sigma_D(y,y)\,V_D(y)\boldsymbol{x}\\&=\Lambda'(\boldsymbol{x}^\top\nu_D(y))^2\,\boldsymbol{x}^\top V_D(y)\boldsymbol{x}\\&=c^{-2}\,w(F(y\,|\,x))^2\,\boldsymbol{x}^\top V_D(y)\boldsymbol{x}.\end{aligned} \tag{10}$$

Because $w(0) = w(1) = 0$, formula (10) implies that, for all finite x, the asymptotic variance of $\hat{F}(y \mid x)$ tends to zero as $|y| \to \infty$.

If model (1) with logistic errors is misspecified, then maximizing the expected log-likelihood $\ell(\nu; y)$ gives the best logistic approximation (in the Kullback-Leibler sense) $F^+(y \mid x) = \Lambda(\boldsymbol{x}^\top \nu_D(y))$ to the true CDF. Now both components of $\nu_D(y)$ depend, possibly nonlinearly, on y. Clearly, $F^+(y \mid x) = F(y \mid x)$ for all y if the logit link is correctly specified. Theorem 1 then implies that, for all x, $\sqrt{n}[\hat{F}(\cdot \mid x) - F^+(\cdot \mid x)]$ converges weakly to a zero-mean Gaussian process defined on $\mathbb{R}$ with covariance function

$$\operatorname{cov}(\hat{F}(y \mid x), \hat{F}(y' \mid x)) = \Lambda'\left(\boldsymbol{x}^\top \nu_D(y)\right) \Lambda'\left(\boldsymbol{x}^\top \nu_D(y')\right) \boldsymbol{x}^\top V_D(y)\, \Sigma_D(y, y')\, V_D(y')\boldsymbol{x}. \tag{11}$$

The indirect estimator of the CQF based on the DR approach, obtained by inverting the estimator $\hat{F}(y \mid x)$ of the CDF, is $\hat{Q}(p \mid x) = \min\{y \colon \hat{F}(y \mid x) \geq p\}$. We also denote by $Q^+(p \mid x)$ the inverse of the best logistic approximation $F^+(y \mid x)$ to $F(y \mid x)$, namely $Q^+(p \mid x) = \inf\{y \colon F^+(y \mid x) \geq p\}$. To derive the asymptotic properties of $\hat{Q}(p \mid x)$ we use the functional Delta-method.

Theorem 2 *Suppose that the assumptions of Theorem 1 hold, and let $\underline{p} = F^+(\underline{y} \mid x)$ and $\overline{p} = F^+(\overline{y} \mid x)$. If $\nu_D(\cdot)$ is continuously differentiable on $[\underline{y}, \overline{y}]$ then, for all x, $\hat{Q}(p \mid x)$ is uniformly consistent for $Q^+(p \mid x)$ and the process $\sqrt{n}[\hat{Q}(\cdot \mid x) - Q^+(\cdot \mid x)]$ converges to a zero-mean Gaussian process $M_D(\cdot \mid x)$ defined on $[\underline{p}, \overline{p}]$ with covariance function*

$$\operatorname{cov}(M_D(p \mid x), M_D(p' \mid x)) = [\boldsymbol{x}^\top \nu_D'(y_p)][\boldsymbol{x}^\top \nu_D'(y_{p'})]\, \boldsymbol{x}^\top V_D(y_p)\, \Sigma_D(y_p, y_{p'})\, V_D(y_{p'})\boldsymbol{x},$$

where $y_p = \alpha + x^\top \beta + G^{-1}(p)$, $\nu_D'(y) = \partial \nu_D(y)/\partial y$, and $\Sigma_D(\cdot, \cdot)$ is defined in (9).

Theorem 2 implies that, under the linear location model (1), the asymptotic variance of $\hat{Q}(p \mid x)$ is equal to $V(\hat{Q}(p \mid x)) = \boldsymbol{x}^\top V_D(y_p)\boldsymbol{x}$. Just like $V(\tilde{Q}(p \mid x))$, this asymptotic variance is U-shaped in p with a minimum at $p = .5$ and diverges as p tends to 0 or 1. However, unlike $V(\tilde{Q}(\mid x))$, it also depends on the model parameters through y_p.

4.2 QR approach

A linear QR estimator of the CQF is of the form $\tilde{Q}(p \mid x) = \boldsymbol{x}^\top \tilde{\nu}(p)$, where $\tilde{\nu}(p) = (\tilde{\xi}_p, \tilde{\beta}_p^\top)^\top$. Angrist, Chernozhukov and Fernández-Val (2006) establish the asymptotic properties of the process $\{\tilde{\nu}(p), p \in (0,1)\}$ when the CQF $Q(p \mid x)$ is arbitrary, not necessarily linear. Following their notation, let $\nu_Q(p) = (\xi_p, \beta_p^\top)^\top$ solve the problem

$$\min_{a,b} \operatorname{E} \rho_p(Y - a - X^\top b),$$

and let $Q^*(p\,|\,x) = \boldsymbol{x}^\top \nu_Q(p)$ denote the best linear approximation to $Q(p\,|\,x)$ (in the ρ_p sense). In the special case of the linear location model (1), we have that $\xi_p = \alpha + G^{-1}(p)$, $\beta_p = \beta$ for any p, and $Q^*(p\,|\,x) = Q(p\,|\,x)$. For convenience of the reader, we restate the main result of Angrist, Chernozhukov and Fernández-Val (2006).

Theorem 3 *Suppose that: (i) $\{(Y_i, X_i), i = 1, \ldots, n\}$ is a sample from the joint distribution of (Y, X); (ii) the conditional density $f(y\,|\,x)$ exists and is bounded and uniformly continuous in y, uniformly in x over the support of X; (iii) $\|X\|^{2+\epsilon} < \infty$ for some $\epsilon > 0$; (iv) the matrix $J(p) = \mathrm{E}[f(\boldsymbol{X}^\top \nu_Q(p)|X)\boldsymbol{X}\boldsymbol{X}^\top]$ is symmetric and positive definite for all p in a closed interval $[\underline{p}, \overline{p}]$ of $(0,1)$. Then, $\tilde{\nu}(\cdot)$ is uniformly consistent for $\nu_Q(\cdot)$, that is, $\sup_{\underline{p}\le p \le \overline{p}} \|\tilde{\nu}(p) - \nu_Q(p)\| = o_p(1)$. Further, the process $J(\cdot)\sqrt{n}[\tilde{\nu}(\cdot) - \nu_Q(\cdot)]$ converges weakly to a zero-mean multivariate Gaussian process $B_Q(\cdot)$ defined on $[\underline{p}, \overline{p}]$ with covariance function*

$$\Sigma_Q(p, p') = \mathrm{E}\left[\left(p - 1\{Y < Q^*(p\,|\,X)\}\right)\left(p' - 1\{Y < Q^*(p'\,|\,X)\}\right) \boldsymbol{X}\boldsymbol{X}^\top\right]. \tag{12}$$

Theorem 3 implies that, for all x,

$$\sqrt{n}[\tilde{Q}(\cdot\,|\,x) - Q^*(\cdot\,|\,x)] \Rightarrow J^{-1}(\cdot)\, B_Q(\cdot)^\top \boldsymbol{x}. \tag{13}$$

Under the linear location model (1), the covariance function (12) simplifies to $\Sigma_Q(p, p') = p(1 - p')P_X$, $p \le p'$, and the asymptotic variance of $\tilde{Q}(p\,|\,x)$ is just $V(\tilde{Q}(p\,|\,x)) = \boldsymbol{x}^\top V_Q(p)\boldsymbol{x}$. Since $V_Q(p)$ is inversely proportional to $w(p)$, the asymptotic variance of $\tilde{Q}(p\,|\,x)$, viewed as function of p for x given, is U-shaped with a minimum at $p = .5$ and diverges as p tends to 0 or 1.

The indirect estimator of the CDF based on the QR approach, obtained by inverting the estimator $\tilde{Q}(p\,|\,x)$ of the CQF, is $\tilde{F}(y\,|\,x) = \inf\{p\colon \boldsymbol{x}^\top \tilde{\nu}(p) \ge y\}$. We also denote by $F^*(y\,|\,x)$ the inverse of the best linear approximation $Q^*(p\,|\,x)$ to $Q(p\,|\,x)$, namely $F^*(y\,|\,x) = \inf\{p\colon \boldsymbol{x}^\top \nu_Q(p) \ge y\}$. If $Q(p\,|\,x) = Q^*(p\,|\,x)$ for all p, then $F^*(y\,|\,x) = F(y\,|\,x)$ for all y. Notice that our indirect QR estimator of $F(y\,|\,x)$ is constructed differently from Chernozhukov, Fernàndez-Val and Melly (2013) who rely instead on the integral transformation $F(y\,|\,x) = \int_0^1 1\{Q(u\,|\,x) \le y\}\,du$. To derive the asymptotic properties of $\tilde{F}(y\,|\,x)$ we again use a functional Delta-method argument.

Theorem 4 *Suppose that the assumptions of Theorem 3 hold, and let $\underline{y} = Q^*(\underline{p}\,|\,x)$ and $\overline{y} = Q^*(\overline{p}\,|\,x)$. If $\nu_Q(\cdot)$ is continuous on $(0,1)$ then, $\tilde{F}(\cdot\,|\,x)$ is uniformly consistent for $F^*(\cdot\,|\,x)$ for all x, and the process $\sqrt{n}[\tilde{F}(\cdot\,|\,x) - F^*(\cdot\,|\,x)]$ converges to the zero-mean Gaussian process*

$$M_Q(\cdot\,|\,x) = \frac{\boldsymbol{x}^\top J^*(\cdot\,|\,x)^{-1} B^*(\cdot\,|\,x)}{\boldsymbol{\mu}^\top J^*(\cdot\,|\,x)^{-1}\boldsymbol{x}}$$

defined on $[\underline{y},\overline{y}]$, *where* $B^*(\cdot\,|\,x) = B_Q(F^*(\cdot\,|\,x))$, $J^*(\cdot\,|\,x) = J(F^*(\cdot\,|\,x))$ *and* $\boldsymbol{\mu} = \mathrm{E}\,\boldsymbol{X} = (1,\,\mu_X^\top)^\top$. *The process* M_Q *has covariance function*

$$\operatorname{cov}(M_Q(y\,|\,x), M_Q(y'\,|\,x)) = \frac{\boldsymbol{x}^\top J^*(y\,|\,x)^{-1}\Sigma_Q(F^*(y\,|\,x), F^*(y'\,|\,x))J^*(y'\,|\,x)^{-1}\boldsymbol{x}}{[\boldsymbol{\mu}^\top J^*(y\,|\,\boldsymbol{x})^{-1}\boldsymbol{\mu}]^2}.$$

Theorem 4 implies that, under the linear location model (1),

$$\Sigma_Q(F^*(y\,|\,x), F^*(y'\,|\,x)) = G(y-\alpha-x^\top\beta)[1-G(y'-\alpha-x^\top\beta)]\,P_X, \qquad y \le y',$$

and

$$J^*(y\,|\,x) = \mathrm{E}\left[f\left(\boldsymbol{X}^\top \nu_Q(F(y\,|\,x))\,|\,X\right)\boldsymbol{X}\boldsymbol{X}^\top\right].$$

Since in this case

$$\boldsymbol{X}^\top \nu_Q(F(y\,|\,x)) = \alpha + X^\top\beta + G^{-1}(G(y-\alpha-x^\top\beta)) = y - \beta^\top(x-X)$$

and $f(y-(x-X)^\top\beta\,|\,X) = g(y-\alpha-x^\top\beta)$, it follows that $J^*(y\,|\,x) = g(y-\alpha-x^\top\beta)\,P_X$. Hence, under the linear location model (1), the asymptotic variance of $\tilde{F}(y\,|\,x)$ simplifies to

$$V(\tilde{F}(y\,|\,x)) = G(y-\alpha-x^\top\beta)[1-G(y-\alpha-x^\top\beta)]\,\frac{\boldsymbol{x}^\top P_X^{-1}\boldsymbol{x}}{(\boldsymbol{\mu}^\top P_X^{-1}\boldsymbol{x})^2}.$$

Next notice that, by the inversion formula for block matrices,

$$P_X^{-1} = \begin{pmatrix} 1 & \mu_X^\top \\ \mu_X & \mathrm{E}\,XX^\top \end{pmatrix}^{-1} = \begin{pmatrix} 1+\mu_X^\top\Sigma_{XX}^{-1}\mu_X & -\mu_X^\top\Sigma_{XX}^{-1} \\ -\Sigma_{XX}^{-1}\mu_X & \Sigma_{XX}^{-1} \end{pmatrix},$$

where Σ_{XX} denotes the variance of X. This implies that $\boldsymbol{\mu}^\top P_X^{-1} = (1,0)$, so $\boldsymbol{\mu}^\top P_X^{-1}\boldsymbol{x} = 1$ for all x. Therefore

$$V(\tilde{F}(y\,|\,x)) = G(y-\alpha-x^\top\beta)[1-G(y-\alpha-x^\top\beta)]\,\boldsymbol{x}^\top P_X^{-1}\boldsymbol{x}. \tag{14}$$

Formula (14) shows that the asymptotic variance of $\tilde{F}(y\,|\,x)$, viewed as a function of y for x given, has an inverted U-shape with a maximum at $y = Q(.5\,|\,x)$ and tends to 0 as $|y| \to \infty$. Also notice that, if $x = \mu_X$, then $\boldsymbol{\mu}^\top P_X^{-1}\boldsymbol{\mu} = 1$, so $V(\tilde{F}(y\,|\,\mu_X)) = G(y-\alpha-\mu_X^\top\beta)[1-G(y-\alpha-\mu_X^\top\beta)]$. Thus, when $x = \mu_X$, $\tilde{F}(\cdot\,|\,x)$ is asymptotically equivalent to the empirical distribution function based on a random sample from the distribution of the unobservable random variable $Y-\alpha-\mu_X^\top\beta$. However, since $\boldsymbol{x}^\top P_X^{-1}\boldsymbol{x} \ge \boldsymbol{\mu}^\top P_X^{-1}\boldsymbol{\mu}$ by Cauchy-Schwartz inequality, the asymptotic variance of $\tilde{F}(y\,|\,x)$ exceeds that of $\tilde{F}(y\,|\,\mu_X)$ for all other values of x.

4.3 Relative efficiency

We now compare the asymptotic efficiency of the direct and the indirect approaches to estimating $F(y \mid x)$ and $Q(p \mid x)$.

First consider the case when the linear location model (1) holds. For a given quantile level p, the asymptotic variances of the QR (direct) and DR (indirect) estimators of the CQF are, respectively, $V(\tilde{Q}(p \mid x)) = \boldsymbol{x}^\top V_Q(p)\boldsymbol{x}$ and $V(\hat{Q}(p \mid x)) = \boldsymbol{x}^\top V_D(y_p)\boldsymbol{x}$, where $y_p = \alpha + x^\top \beta + G^{-1}(p)$. The ARE of $\tilde{Q}(p \mid x)$ to $\hat{Q}(p \mid x)$ is therefore

$$ARE(\tilde{Q}(p \mid x), \hat{Q}(p \mid x)) = \frac{\boldsymbol{x}^\top V_D(y_p)\boldsymbol{x}}{\boldsymbol{x}^\top V_Q(p)\boldsymbol{x}}. \tag{15}$$

Notice that this ARE is greater than one if and only if $\boldsymbol{x}^\top [V_D(y_p \mid x) - V_Q(p \mid x)]\boldsymbol{x} > 0$, and that the matrix difference $V_D - V_Q$ is positive definite whenever all the eigenvalues of the matrix $V_Q V_D^{-1}$ are smaller than one. A necessary but not sufficient condition is that $\det(V_Q V_D^{-1}) = ARE(\hat{\theta}_{y_p}, \tilde{\theta}_p)^{1+k} < 1$.

As before, we consider the case of a single regressor X. Except for the binary regressor case, we also assume that U has a unit logistic distribution.

1. Binary regressor. In this case, $\boldsymbol{x}^\top V_Q(p)\boldsymbol{x}$ is equal to $\mu_X/[w(p)\sigma_X^2]$ when $x = 0$ and to $(1-\mu_X)/[w(p)\sigma_X^2]$ when $x = 1$. If $y_p = \alpha + \beta x + G^{-1}(p)$, then $\gamma_{y_p} = \beta x + G^{-1}(p)$, so

$$\boldsymbol{x}^\top V_D(y_p)\boldsymbol{x} = \begin{cases} \dfrac{\mu_X}{w\left(G(\gamma_{y_p})\right)\sigma_X^2} = \dfrac{\mu_X}{w(p)\,\sigma_X^2}, & \text{if } x = 0, \\ \dfrac{1-\mu_X}{w\left(G(\gamma_{y_p} - \beta)\right)\sigma_X^2} = \dfrac{1-\mu_X}{w(p)\,\sigma_X^2}, & \text{if } x = 1. \end{cases}$$

Thus, in line with the results in Section 3.3, $ARE(\tilde{Q}(p \mid x), \hat{Q}(p \mid x)) = 1$ for all p.

2. Uniform regressor. If $X \sim \mathscr{U}(0,1)$, then the matrix $V_D(y_p)$ takes the complicated expression (7). It is instead easy to compute $\boldsymbol{x}^\top V_Q(p)\boldsymbol{x}$, as

$$\boldsymbol{x}^\top V_Q(p)\boldsymbol{x} = \frac{\mathrm{E}\,X^2 - 2x\mu_X + x^2}{w(p)\,\sigma_X^2} = \frac{12(x^2 - x + 1/3)}{w(p)}.$$

Figure 3 presents the contour plots of asymptotic efficiency gain of $\tilde{Q}(p \mid x)$ relative to $\hat{Q}(p \mid x)$, for three values of the regression R^2 (high, medium and low) and different values of p and x. As in Figure 2, the blue area corresponds to pairs (p, x) for which $ARE(\tilde{Q}(p \mid x), \hat{Q}(p \mid x)) > 1$, while the orange-red area corresponds to pairs for which $ARE(\tilde{Q}(p \mid x), \hat{Q}(p \mid x)) \leq 1$. The three panels in the figure show that, as R^2 decreases, the region where $ARE(\tilde{Q}(p \mid x), \hat{Q}(p \mid x)) \leq 1$ becomes larger and larger, starting from the extreme values of p, until covering the whole area when $R^2 = 0$ ($\beta = 0$), as $ARE(\tilde{Q}(p\,|\,x), \hat{Q}(p\,|\,x)) = 1$ in this case.

The maximum percentage gain of $\tilde{Q}$ relative to $\hat{Q}$ among the three different models is around 560% (attained when $R^2 = 0.9$), while the maximum percentage gain of $\hat{Q}$ relative to $\tilde{Q}$ is around 360% (attained when $R^2 = 0.5$).

3. Gaussian regressor. Figure 4 presents the contour plots of $ARE(\tilde{Q}(p\,|\,x), \hat{Q}(p\,|\,x))$ when $X \sim \mathcal{N}(0,1)$. We again consider three different values of the regression R^2 and different values of p and x. As for the case when X is uniform, the region where $ARE(\tilde{Q}(p\,|\,x), \hat{Q}(p\,|\,x)) \leq 1$ (white in the figure) becomes larger and larger as the R^2 decreases. Thus, the cases of X uniform or Gaussian both support efficiency of QR relative to DR except for small value of R^2 or extreme quantile levels.

Using formula (10), and rewriting (14) as

$$V(\tilde{F}(y\,|\,x)) = G(y - \alpha - \beta x)[1 - G(y - \alpha - \beta x)]\,\boldsymbol{x}^\top P_X^{-1}\boldsymbol{x} = g(y - \alpha - \beta x)^2\,\boldsymbol{x}^\top V_Q(p_y)\boldsymbol{x},$$

where $p_y = G(y - \alpha - \beta x)$, we get the following expression for the ARE of the estimators of the CDF based on the two approaches

$$ARE(\tilde{F}(y\,|\,x), \hat{F}(y\,|\,x)) = \frac{\boldsymbol{x}^\top V_D(y)\boldsymbol{x}}{\boldsymbol{x}^\top V_Q(p_y)\boldsymbol{x}}.$$

Since this expression coincides with (15) when $y = y_p$, the ARE of $\tilde{F}(y\,|\,x)$ to $\hat{F}(y\,|\,x)$ and the ARE of $\tilde{Q}(p\,|\,x)$ to $\hat{Q}(p\,|\,x)$ are the same. Thus, the performance of the direct estimator of the CDF from the DR approach relative to its indirect estimator from the QR approach is asymptotically the same as the performance of the indirect estimator of the CQF from the DR approach relative to its direct estimator from the QR approach. This is a consequence of the following key result.

Theorem 5 *Let $\hat{F}(y\,|\,x)$ be a uniformly consistent estimator of a CDF $F(y\,|\,x)$ with a continuous density $f(y\,|\,x)$ and let $\tilde{Q}(p\,|\,x)$ be a uniformly consistent estimator of the CQF $Q(p\,|\,x) = F^{-1}(p\,|\,x)$. Suppose that, for all x, $\sqrt{n}[\hat{F}(\cdot\,|\,x) - F(\cdot\,|\,x)]$ converges weakly to a zero-mean Gaussian process defined on a closed interval $[\underline{y}, \overline{y}]$ of $\mathbb{R}$, and that $\sqrt{n}[\tilde{Q}(\cdot\,|\,x) - Q(\cdot\,|\,x)]$ converges weakly to a zero-mean Gaussian process defined on a closed interval $[\underline{p}, \overline{p}]$ of $(0,1)$. Let $\hat{Q}(p\,|\,x) = \hat{F}^{-1}(p\,|\,x)$ be the indirect estimator of $Q(p\,|\,x)$ obtained by inverting $\hat{F}(\cdot\,|\,x)$ and let $\tilde{F}(y\,|\,x) = \tilde{Q}^{-1}(y\,|\,x)$ be the indirect estimator of $F(y\,|\,x)$ obtained by inverting $\tilde{Q}(\cdot\,|\,x)$. Then, for all x, all $p \in [\underline{p}, \overline{p}]$ and all $y \in [\underline{y}, \overline{y}]$ such that $F(y\,|\,x) = p$,*

$$\frac{V(\hat{F}(y\,|\,x)))}{V(\hat{Q}(F(y\,|\,x)\,|\,x)} = \frac{V(\tilde{F}(y\,|\,x))}{V(\tilde{Q}(F(y\,|\,x)\,|\,x))} = f(y\,|\,x)^2. \tag{16}$$

Theorem 5 has three important implications. First, for all $p \in [\underline{p}, \overline{p}]$ and all $y \in [\underline{y}, \overline{y}]$ such that $F(y\,|\,x) = p$,

$$ARE(\tilde{F}(y\,|\,x), \hat{F}(y\,|\,x)) = ARE(\tilde{Q}(p\,|\,x), \hat{Q}(p\,|\,x)).$$

Thus, the relative performance of the DR and QR approaches in estimating the CDF is asymptotically the same as their relative performance in estimating the CQF. Consistently with this result, Azzalini (1981) found that the approximate MSE of the direct kernel estimator $\hat{F}$ of an unconditional distribution function (obtained by integrating a kernel density estimator) relative to the MSE of the empirical distribution function is about the same as the MSE of the indirect estimator of the quantile function, obtained by inverting $\hat{F}$, relative to the MSE of the sample quantile function.

Second, if the conditional density of Y has vanishing tails then, irrespective of the approach (either direct or indirect), the estimator of the CDF for values of y in the tails of the conditional distribution of Y has smaller asymptotic variance than the corresponding estimator of the CQF.

Third, for $p = F(y\,|\,x)$, let $\mathbb{V}_{dir}$ denote the asymptotic variance of the direct estimator $(\hat{F}(y\,|\,x), \tilde{Q}(p\,|\,x))$ of the vector $(F(y\,|\,x), Q(p\,|\,x))$, and let $\mathbb{V}_{ind}$ denote the asymptotic variance of the indirect estimator $(\tilde{F}(y\,|\,x), \hat{Q}(p\,|\,x))$. It follows from the asymptotic representations of $\sqrt{n}(\hat{Q} - Q)$ and $\sqrt{n}(\tilde{F} - F)$ in the proof of Theorem 5 that the asymptotic covariance between $\hat{F}(y\,|\,x)$ and $\tilde{Q}(p\,|\,x))$ is the same as the asymptotic covariance between $\tilde{F}(y\,|\,x)$ and $\hat{Q}(p\,|\,x))$ for all $p = F(y\,|\,x)$. Formula (16) then implies that

$$ARE(\text{indirect}, \text{direct}) = \frac{|\mathbb{V}_{dir}|^{1/2}}{|\mathbb{V}_{ind}|^{1/2}} = 1,$$

that is, the direct and the indirect estimators of the vector $(F(y\,|\,x), Q(p\,|\,x))$ are asymptotically equivalent.

Now consider the case when the linear location model (1) does not hold. In this case, neither $\hat{F}(y\,|\,x)$ nor $\tilde{Q}(p\,|\,x)$ are guaranteed to be consistent. If they are not consistent then, under the assumptions of Theorems 1 and 3, it follows that $V(\hat{F}(y\,|\,x)))/V(\hat{Q}(F^{+}(y\,|\,x)\,|\,x)) = f^{+}(y\,|\,x)^2$, while $V(\tilde{F}(y\,|\,x)))/V(\tilde{Q}(F^{*}(y\,|\,x)\,|\,x)) = f^{*}(y\,|\,x)^2$, where $f^{*}(y\,|\,x)$ and $f^{+}(y\,|\,x)$ are the conditional densities associated with $F^{*}(y\,|\,x)$ and $F^{+}(y\,|\,x)$ respectively. These two quantities need not be the same for all y. Thus,

$$\frac{V(\hat{F}(y\,|\,x)))}{V(\tilde{F}(y\,|\,x))} = \frac{V(\hat{Q}(F^{+}(y\,|\,x)\,|\,x))}{V(\tilde{Q}(F^{*}(y\,|\,x)\,|\,x))} \frac{f^{+}(y\,|\,x)^2}{f^{*}(y\,|\,x)^2}.$$

5 The linear location-scale model

The linear location model (1) is one of the workhorses of empirical analysis but has the very restrictive implication that its conditional quantiles have the same slope. Correspondingly, in the binary regression model underlying the DR approach, only the intercept, not the slope, changes with the cutoff y.

One natural extension of model (1) is a heteroskedastic location-scale model where both the location and the scale parameters depend on regressors. To keep thing simple, we focus on the following subclass of location-scale models

$$Y = \alpha + Z^\top \beta + (1 + Z^\top \eta)U, \tag{17}$$

where Z is a random vector whose elements are functions of X and possibly other regressors, η is a vector of parameters such that $1 + Z^\top \eta > 0$ with probability one, and U is a random error distributed independently of Z with smooth distribution function G and density function g. Model (17), henceforth referred to as the linear location-scale model, encompasses some of the most popular location-scale models. The case when $Z = X$ has been discussed extensively by Koenker and Xiao (2002). When $\eta = 0$, we obtain the linear location model (1).

5.1 DR approach

Under (17), the binary regression model underlying the DR approach is

$$\mathrm{E}(D_y \,|\, Z = z) = P\left((1 + z^\top \eta)U \leq y - \alpha - z^\top \beta\right) = G\left(\frac{\gamma_y + z^\top \delta}{1 + z^\top \eta}\right).$$

The unknown parameters $\gamma_y = y - \alpha$, $\delta = -\beta$ and η are typically estimated by maximizing an average log-likelihood $L(\nu; y)$ of the same form as (8) with $\pi_i(\theta)$ replaced by $\pi_i(\nu) = G(r(Z_i, \nu))$, where $\nu = (\gamma, \delta^\top, \eta^\top)^\top$ and $r(z; \nu) = (\gamma + z^\top \delta)/(1 + z^\top \eta)$. The following theorem establishes weak convergence of the ML estimator of ν under model (17).

Theorem 6 *Suppose that: (i) $\{(Y_i, Z_i), i = 1, \ldots, n\}$ is a sample from the joint distribution of (Y, Z); (ii) $\mathrm{E}\,\|Z\|^{4+\varepsilon} < \infty$ for some $\varepsilon > 0$ and there exists $0 < M < \infty$, such that $1 + Z^\top \eta > 1/M$; (iii) for all y, $\hat{\nu}(y)$ maximizes the average log-likelihood $L(\nu; y)$ on a compact parameter space; (iv) for all y in a closed interval $[\underline{y}, \overline{y}]$ of $\mathbb{R}$, the expected log-likelihood*

$$\ell(\nu; y) = \mathrm{E}\left[D_y \ln \frac{G(r(Z; \nu))}{1 - G(r(Z; \nu))} - \ln(1 - G(r(Z; \nu)))\right]$$

attains a unique maximum at $\nu_D(y)$; (v) the matrix

$$V_D(y) = \left[\mathrm{E}\, w(F(y \,|\, Z)) R(y \,|\, Z) R(y \,|\, Z)^\top\right]^{-1},$$

where $R(y\,|\,z) = \partial r(z;\nu)/\partial \nu|_{\nu=\nu_D(y)}$, is positive definite for all y. Then, $\hat{\nu}(\cdot)$ is uniformly consistent for $\nu_D(\cdot)$ and the process $\sqrt{n}V_D^{-1}(\cdot)[\hat{\nu}(\cdot) - \nu_D(\cdot)]$ converges weakly to a zero-mean multivariate Gaussian process $B_D(\cdot)$ defined on $[\underline{y}, \overline{y}]$ with covariance function

$$\Sigma_D(y,y') = c^2\,\mathrm{E}\left[F(y\,|\,Z)(1-F(y'\,|\,Z))R(y\,|\,Z)R(y'\,|\,Z)^\top\right], \qquad y \le y'.$$

Moreover, for all z, the process $\sqrt{n}[\hat{F}(\cdot\,|\,z) - F(\cdot\,|\,z)]$ converges weakly to a zero-mean Gaussian process defined on $[\underline{y}, \overline{y}]$ with covariance function

$$c^{-2}w(F(y\,|\,z))w(F(y'\,|\,z))R(y\,|\,z)^\top V_D(y)\,\Sigma_D(y,y')\,V_D(y')R(y'\,|\,z). \tag{18}$$

The assumptions of Theorem 6 are the same as those of Theorem 1 except for condition (ii), which guarantees that $\mathrm{E}\,\|\partial r(Z;\nu)/\partial \nu\|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$ and all values of ν in the parameter space.

The proof of Theorem 6 is in the Appendix and consists of two steps. First we establish uniform consistency of the regression process $y \mapsto \hat{\nu}(y) = (\hat{\gamma}_y, \hat{\delta}(y)^\top, \hat{\eta}(y)^\top)$. Then we prove asymptotic normality of $\sqrt{n}[\hat{\nu}(y) - \nu_D(y)]$ by representing it as an empirical process indexed by a Donsker class of functions.

Theorem 6 implies that the asymptotic variance of $\hat{F}(y\,|\,z)$ is equal to

$$c^{-2}w(F(y\,|\,z))^2R(y\,|\,z)^\top V_D(y)R(y\,|\,z).$$

By repeating the arguments of Theorem 2 after noting that $H'(y) = c^{-1}w(z;r(\nu_D(y)))\nu_D'(y)R(y\,|\,z)$, the asymptotic variance of the indirect estimator $\hat{Q}(p\,|\,x)$ is

$$\begin{aligned} V(\hat{Q}(p\,|\,z)) &= c^2[w(p)\nu_D'(y_p)R(y_p\,|\,z)]^{-2}c^{-2}w(p)^2R(y_p\,|\,z)^\top V_D(y_p)R(y_p\,|\,z) \\ &= [\nu_D'(y_p)R(y_p\,|\,z)]^{-2}R(y_p\,|\,z)^\top V_D(y_p)R(y_p\,|\,z) \\ &= (1+z^\top\eta)^2R(y_p\,|\,z)^\top V_D(y_p)R(y_p\,|\,z), \end{aligned} \tag{19}$$

where we used the fact that, under (17), $\nu_D'(y) = (1,0,0)$, $\nu_D'(y_p)R(y_p\,|\,z) = (1+z^\top\eta)^{-1}$ and, if $y_p = Q(p\,|\,z)$, then $w(G(z;r(\nu_D(y_p)))) = w(p)$.

5.2 QR approach

The CQF for model (17) is $Q(p\,|\,z) = \boldsymbol{z}^\top\nu_Q(p)$, where $\boldsymbol{z} = (1, z^\top)^\top$, $\nu_Q(p) = (\xi_p, \beta_p^\top)^\top$, $\xi_p = \alpha + G^{-1}(p)$, and $\beta_p = \beta + \eta G^{-1}(p)$. Notice that this function is still linear in the model parameters, but now both its intercept and its slope vary with the quantile level p.

The asymptotic properties of the QR estimator $\tilde{\nu}(p)$, first derived by Koenker and Xiao (2002), follow immediately from Theorem 3 by replacing X with Z and $\boldsymbol{X}$ with $\boldsymbol{Z} = (1, Z^\top)^\top$, and by letting

$J(p) = g(G^{-1}(p))K(\eta)$, where $K(\eta) = \mathrm{E}[(1+Z^\top\eta)^{-1}\boldsymbol{ZZ}^\top]$. Thus, $\tilde{\nu}(\cdot)$ is uniformly consistent for $\nu_Q(\cdot)$ and the process $J(\cdot)\sqrt{n}[\tilde{\nu}(\cdot) - \nu_Q(\cdot)]$ converges weakly to a zero-mean multivariate Gaussian process with covariance function

$$\begin{aligned}\Sigma_Q(p,p') &= \mathrm{E}\left[\big(p - 1\{Y < Q(p\,|\,\boldsymbol{Z})\}\big)\big(p' - 1\{Y < Q(p'\,|\,\boldsymbol{Z})\big)\boldsymbol{ZZ}^\top\right] \\ &= p(1-p')P_Z, \qquad p \le p',\end{aligned} \tag{20}$$

where $P_Z = \mathrm{E}\,\boldsymbol{ZZ}^\top$. The asymptotic variance of the direct estimator $\tilde{Q}(p\,|\,z) = \boldsymbol{z}^\top\tilde{\nu}(p)$ of $Q(p\,|\,z)$ is therefore $V(\tilde{Q}(p\,|\,z)) = w(p)^{-1}\boldsymbol{z}^\top K(\eta)^{-1}P_Z K(\eta)^{-1}\boldsymbol{z}$.

Since $\partial\boldsymbol{z}^\top\nu_Q(p)/\partial p = (1+z^\top\eta)/[g(G^{-1}(p))]$ under model (17), Theorem 4 implies that the indirect estimator $\tilde{F}(y\,|\,z)$ is $\sqrt{n}$-consistent for $F(y\,|\,x)$ and asymptotically normal with asymptotic variance

$$V(\tilde{F}(y\,|\,z)) = \frac{F(y\,|\,z)(1-F(y\,|\,z))}{(1+z^\top\eta)^2}\, g\left(\frac{y-\alpha-z^\top\beta}{1+z^\top\eta}\right)^2 \boldsymbol{z}^\top J^*(y\,|\,z)^{-1} P_Z\, J^*(y\,|\,z)^{-1}\boldsymbol{z}.$$

In particular,

$$J^*(y\,|\,z) = \mathrm{E}\left[f\left(\nu_Q(F(y\,|\,z))\boldsymbol{Z}\,|\,Z\right)\boldsymbol{ZZ}^\top\right] = g\left(\frac{y-\alpha-z^\top\beta}{1+z^\top\eta}\right)K(\eta),$$

and therefore

$$\begin{aligned}V(\tilde{F}(y\,|\,z)) &= \frac{F(y\,|\,z)(1-F(y\,|\,z))}{(1+z^\top\eta)^2}\,\boldsymbol{z}^\top K(\eta)^{-1}P_Z K(\eta)^{-1}\boldsymbol{z} \\ &= \frac{w(F(y\,|\,z))}{c^2(1+z^\top\eta)^2}\,\boldsymbol{z}^\top K(\eta)^{-1}P_Z K(\eta)^{-1}\boldsymbol{z},\end{aligned}$$

where again $c^2 = \pi^2/3$.

5.3 Relative efficiency

It follows from the results in Section 5.1 and 5.2 that, under the linear location-scale model (17), the ARE of $\tilde{Q}$ to $\hat{Q}$ and the ARE of $\tilde{F}$ to $\hat{F}$ are again the same, namely

$$\begin{aligned}ARE(\tilde{Q}(p\,|\,z), \hat{Q}(p\,|\,z)) &= ARE(\tilde{F}(y_p\,|\,z), \hat{F}(y_p\,|\,z)) \\ &= w(p)\,\frac{(1+z^\top\eta)^2 R(y_p\,|\,z)^\top V_D(y_p) R(y_p\,|\,z)}{\boldsymbol{z}^\top K(\eta)^{-1}P_Z K(\eta)^{-1}\boldsymbol{z}}.\end{aligned}$$

As an illustration, consider a stratified population where the conditional variance of Y given a scalar regressor X is different between groups but constant within group. To fix the ideas, consider the case of only two groups and write the model as

$$Y = \alpha + \beta X + (1+\eta W)U, \tag{21}$$

where W is a binary indicator equal to zero if a population unit belongs to the first group and to one if it belongs to the second group. Theorem 6 applies directly to model (21) and, in this case,

$$R(y\,|\,x,w) = \frac{1}{1+\eta w}\begin{pmatrix} 1 \\ -x \\ -r(z;\,\nu(y))\,w \end{pmatrix}.$$

Figure 5 presents the behavior of the ARE when $\alpha = 0$, $\beta = 3.364$, and X and W are independent distributed, respectively as $\mathscr{U}(0,1)$ and $\mathscr{B}(.5)$, for different values of p, x, w and η. The panels in the figure show the contour plots of $ARE(\tilde{Q}(p\,|\,x,w),\hat{Q}(p\,|\,x,w))-1$ when $ARE(\tilde{Q}(p\,|\,x,w),\hat{Q}(p\,|\,x,w)) > 1$, and of $1-ARE(\hat{Q}(p\,|\,x,w),\tilde{Q}(p\,|\,x,w))$ when $ARE(\tilde{Q}(p\,|\,x,w),\hat{Q}(p\,|\,x,w)) \le 1$. As for Figures 3 and 4, the blue (orange-red) area corresponds to the pairs of p and x values for which $\tilde{Q}$ ($\hat{Q}$) is asymptotically more efficient than $\hat{Q}$ ($\tilde{Q}$). The darker the color, the higher the asymptotic efficiency gain. For all the values of w and η, the contour plots show a bimodal pattern of the asymptotic efficiency gain for certain values of x and p.

6 Monte Carlo results

In this section we present the results of a set of Monte Carlo experiments aimed at exploring the finite sample performance of estimators obtained under the DR and the QR approaches. We consider the linear location model (1) and the grouped-heteroskedasticity model (21), both with $\alpha = 0$, a single regressor X and U unit logistic. To keep things simple, we only present the results for the case when $X \sim \mathscr{U}(0,1)$. The results for the case when $X \sim \mathscr{N}(0,1)$ are qualitatively very similar and are available upon request. For each Monte Carlo experiment, we generate $1,000$ samples of increasing size ($n =$ 100, 500, and 2,500).

The direct estimate of the CQF is first computed at a grid of equally spaced points $\mathscr{P} = \{p_1,\ldots,p_J\}$, where J increases with the sample size. For the linear location model, we set $J = 39$ when $n = 100$, $J = 78$ when $n = 500$, and $J = 156$ when $n = 2,500$. Thus, we allow J to increase approximately with the square root of the increase in n. The estimate of the CQF at any arbitrary point $p \in (p_{j-1},p_j)$, $j = 2,\ldots,J$, is then obtained by linear interpolation.

Similarly, the direct estimate of the CDF is first computed at a grid of points $\mathscr{Y} = \{y_1,\ldots,y_J\}$, where $y_j = Q(p_j\,|\,x)$ depends on both p_j and x. The estimate of the CDF at any arbitrary point $y \in (y_{j-1},y_j)$, $j = 2,\ldots,J$, is then obtained by linear interpolation. Finally, the indirect estimates of the CDF and the CQF are defined as $\tilde{F}(y\,|\,x) = \min\{p \in (p_1,p_J)\colon \tilde{Q}(p\,|\,x) \ge y\}$ and $\hat{Q}(p\,|\,x) = \min\{y \in (y_1,y_J)\colon \hat{F}(y\,|\,x) \ge p\}$.

6.1 Linear location model

We consider three different choices for the slope parameter β, namely $\beta = 10.392$, 3.464 and 1.155. Since β and the population regression R^2 are linked through the relationship $\beta^2\sigma_X^2 = R^2/(1-R^2)$, these three values correspond respectively to a high (.90), medium (.50) and low (.10) value of the R^2.

The panels in Figure 6 show the Monte Carlo root mean squared error (RMSE) of, respectively, the direct estimator $\tilde{Q}(p \mid x)$ and the indirect estimator $\hat{Q}(p \mid x)$ of the CQF for different sample sizes and different values of x (columns) and β (rows). For simplicity, we only present the results for R^2 equal to .90 and .10, and for x equal to the upper and lower quartiles of X. Not surprisingly, the RMSE falls with n for both estimators and, viewed as a function of p, has a U-shaped profile with a minimum around $p = .5$, in line with the asymptotic results.

The panels in Figure 7 show the Monte Carlo RMSE of, respectively, the indirect estimator $\tilde{F}(y \mid x)$ and the direct estimator $\hat{F}(y \mid x)$ of the CDF for different sample sizes and different values of x (columns) and β (rows). The RMSE again falls with n for both estimators and, viewed as a function of y, has an approximately bell-shaped profile with a maximum around the median of Y, also in line with the asymptotic results.

Table 1 presents the ratio of the Monte Carlo MSE of the direct estimator $\tilde{Q}(p \mid x)$ to that of the indirect estimator $\hat{Q}(p \mid x)$ for different values of p and different sample sizes, while Table 2 presents the ratio of the Monte Carlo MSE of the indirect estimator $\tilde{F}(y \mid x)$ to that of the direct estimator $\hat{F}(y \mid x)$ for different values of y and different sample sizes. Each table is divided in nine blocks, corresponding to various combinations of x and β. Both tables show that, for all sample sizes, the relative efficiency of QR to DR decreases with the value of the R^2, in line with the asymptotic results in Section 4. Further, except possibly for the case when $n = 100$, the ratio of Monte Carlo MSE's and the ARE have roughly the same profiles as functions of p.

6.2 Grouped heteroskedasticity

In this case we generate the data from model (21) with $W = 1\{X > .4\}$, so grouping depends on X. For comparability, we choose the same values of β as in Section 6.1. We consider three different choices for the parameter η, namely .633, .354 and .08. These values are chosen to attain specific values of the variance of $(1 + \eta W)U$ which is just equal to the ratio of the variances of the unobservables in the heteroskedastic and the homoskedastic model. Since W and U are independent and U has unit variance, the variance of $(1 + \eta W)U$ is equal to 2 when $\eta = .633$, to 1.50 when $\eta = .354$, and to 1.10 when $\eta = .08$. To prevent the squared bias from becoming the dominant part of the MSE, we set the number of grid points in $\mathscr{P}$ and $\mathscr{Y}$ to be twice the number of points used for the linear location model. Although $\mathscr{P}$ and $\mathscr{Y}$ have the same number of grid points, the contribution of the squared bias to the MSE is about twice larger for DR estimators compared to QR estimators, a phenomenon that also occurs in the pure location model. This suggests that, in order to keep the bias down, the DR approach requires a finer grid than the QR approach. For each point in $\mathscr{Y}$, we use the method of ML to obtain a DR estimator of α, β and $\log \eta$, and then invert the estimate of $\log \eta$ to guarantee that the resulting estimate of η is strictly positive.

The panels in Figures 8 and 9 show the Monte Carlo RMSE of $\tilde{Q}(p \mid x, w)$, $\hat{Q}(p \mid x, w)$, $\tilde{F}(y \mid x, w)$ and $\hat{F}(y \mid x, w)$ for different sample sizes and different values of x (columns) and η (rows). For simplicity, we only present results for the intermediate case of $\beta = 3.464$. Results for the other values of β are available upon request. The profiles of the RMSE of both the direct and the indirect estimators of the CQF and the CDF are similar to those for the location model but, all else being equal, their values are generally larger. This is especially true for the DR estimators, due to the extra variability caused by estimation of the additional parameter η.

Table 3 presents the ratio of the Monte Carlo MSE of the direct estimator $\tilde{Q}(p \mid x)$ to that of the indirect estimator $\hat{Q}(p \mid x)$ for different values of p and different sample sizes, while Table 4 presents the ratio of the Monte Carlo MSE of the indirect estimator $\tilde{F}(y \mid x)$ to that of the direct estimator $\hat{F}(y \mid x)$ for different values of y and sample sizes. Each table is divided in nine blocks, corresponding to various combinations of x and η. For simplicity, we only present results for the intermediate case of $\beta = 3.464$. Both tables show that, compared to the linear location case with the same sample size and the same values of x and β, the relative efficiency of the QR approach is greatly reduced, especially when x is equal to the median of X.

Finally Figure 10 shows Monte Carlo 90% pointwise confidence bands for $F(p \mid x)$, based on $\hat{F}(y \mid x)$ and $\tilde{F}(y \mid x)$ respectively, for increasing sample sizes (from lighter to darker color) and different values of x. Each band is constructed by taking the smallest interval that contains 90% of the estimates of $F(y \mid x)$ resulting from a particular Monte Carlo experiment. The solid line inside each band is the true value of the CDF. Figure 11 presents the corresponding Monte Carlo 90% pointwise confidence bands for $Q(p \mid x)$ based on $\hat{Q}(p \mid x)$ and $\tilde{Q}(y \mid x)$. The four panels of each figure correspond to various combinations of β and η. The figures nicely show how the Monte Carlo confidence bands cover the population feature of interest and how they shrink as the sample size increases from $n = 100$ to $n = 500$.

7 Conclusions

This paper compares the statistical properties of the DR and the QR approaches to estimating the conditional distribution of Y given X. While the DR approach provides a direct estimate of the CDF, the QR approach derives an indirect estimate by inverting its estimate of the CQF. Similarly, while the QR approach provides a direct estimate of the CQF, the DR approach derives an indirect estimate by inverting its estimate of the CDF.

If interest focuses on the parameters of an underlying linear location model for Y, then we show that no approach dominates the other. The DR estimator is asymptotically more or less efficient than the QR estimator depending on the cutoff y, the quantile level p, and the R^2 of the underlying linear location model. In general, given y, its asymptotic relative efficiency tends to be higher for quantile levels near 0 or 1. Given p, it tends to be higher for cutoff values near the center of the distribution of Y. It is generally greater than 1 when p is near the boundaries of the unit interval, y is near the median of Y, and the R^2 is small.

If interest focuses instead on estimating the conditional distribution of Y then the two approaches tend to behave very similarly when the regression R^2 is small, as expected from the results for the unconditional case. We show that, under the linear location model, as the R^2 increases, the QR approach is generally more efficient than the DR approach except for quantile levels near 0 or 1. When the R^2 is large, the QR approach actually outperforms the DR approach except for extreme quantile levels. Under the kind of heteroskedasticity considered in this paper, however, the relative efficiency of the QR approach is greatly reduced and may actually be reversed, especially when the CDF and the CQF are estimated at values of x near the center of the distribution of X.

More generally we show that asymptotically, if the models for the CDF and the CQF are correctly specified, then the performance of the direct estimator of the CDF relative to its indirect estimator is the same as the performance of the indirect estimator of the CQF relative to its direct estimator. In other words, the relative efficiency of the DR and QR approaches in estimating the CDF is asymptotically the same as their relative efficiency in estimating the CQF. Further, the direct and the indirect estimators of the vector $(F(y \mid x), Q(p \mid x))$ are asymptotically equivalent whenever $p = F(y \mid x)$.

References

Angrist J., Chernozhukov V., and Fernández-Val I. (2006). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica*, 74: 539–563.

Azzalini A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, 68: 326–328.

Chernozhukov V., Fernández-Val I., and Galichon A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78: 1093–1125.

Chernozhukov V., Fernández-Val I., and Melly B. (2013). Inference on counterfactual distributions. *Econometrica*, 81: 2205–2268.

Dette H., and Volgushev S. (2008). Non-crossing nonparametric estimates of quantile curves. *Journal of the Royal Statistical Society, Series B*, 70: 609–627.

Foresi S., and Peracchi F. (1995). The conditional distribution of excess returns: An empirical analysis. *Journal of the American Statistical Association*, 90: 451–466.

Fortin N., Lemieux T., and Firpo S. (2011). Decomposition methods in economics. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Vol. 4 A, 1–102.

Klein R.W., and Spady R.H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica*, 61: 387–421.

Koenker R., and Bassett G. (1978). Regression quantiles. *Econometrica*, 46: 33–50.

Koenker R. (2005). *Quantile Regression*. Cambridge University Press: New York.

Koenker R., and Xiao Z. (2002). Inference on the quantile regression process, *Econometrica*, 70: 1583–1612.

Nadaraya E.A. (1964). Some new estimates for distribution functions. *Theory of Probability and its Applications*, 15: 497–500.

Peracchi F. (2002). On estimating conditional quantiles and distribution functions. *Computational Statistics & Data Analysis*, 38: 433–447.

Ramberg J.S., Tadikamalla P.R., Dudewicz E.J., and Mykytka E.F. (1979). A probability distribution and its uses in fitting data. *Technometrics*, 21: 201–214.

Rothe, C. (2012). Partial distributional policy effects. *Econometrica*, 80: 2269–2301.

Rothe C., and Wied D. (2012). Misspecification testing in a class of conditional distributional models. IZA Working Paper 6364.

Serfling R. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley: New York.

van der Vaart A.W., and Wellner J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer: New York.

	$\beta = 10.392$				$\beta = 3.464$				$\beta = 1.155$			
	n =100	n =500	n =2500	$n = \infty$	n =100	n =500	n =2500	$n = \infty$	n =100	n =500	n =2500	$n = \infty$
$x = .25$												
$p = .1$	2.433	2.755	2.460	2.628	1.071	1.112	1.050	1.035	.839	.928	.887	.931
$p = .2$	2.574	2.908	2.477	2.772	1.031	1.116	.990	1.072	.924	.936	.918	.961
$p = .3$	2.662	2.789	2.747	2.737	1.216	1.151	1.025	1.125	.961	1.000	.935	.988
$p = .4$	2.525	2.536	2.880	2.689	1.154	1.091	1.197	1.189	1.009	1.010	1.005	1.013
$p = .5$	2.240	2.385	3.012	2.680	1.083	1.160	1.327	1.253	1.013	.973	1.109	1.033
$p = .6$	2.233	2.417	3.044	2.718	1.067	1.223	1.408	1.307	1.059	1.032	1.104	1.047
$p = .7$	2.570	2.777	2.959	2.770	1.250	1.385	1.384	1.331	.994	1.093	1.017	1.054
$p = .8$	2.911	2.727	2.742	2.745	1.391	1.310	1.219	1.296	1.068	1.089	1.038	1.051
$p = .9$	2.422	2.406	2.274	2.403	1.170	1.210	1.049	1.134	.962	1.020	.951	1.034
$x = .50$												
$p = .1$	4.245	4.334	4.064	4.273	1.834	1.717	1.803	1.857	1.126	1.161	1.158	1.140
$p = .2$	4.614	4.480	4.840	4.844	1.693	1.596	1.683	1.841	1.073	1.022	1.072	1.118
$p = .3$	4.374	4.622	4.663	4.883	1.677	1.554	1.692	1.784	1.043	1.049	1.045	1.103
$p = .4$	4.282	4.161	4.607	4.814	1.609	1.496	1.681	1.737	1.063	.995	1.070	1.093
$p = .5$	4.296	4.310	4.616	4.783	1.615	1.594	1.776	1.720	1.090	1.089	1.111	1.090
$p = .6$	4.411	4.174	4.794	4.831	1.667	1.640	1.871	1.738	1.029	1.067	1.172	1.094
$p = .7$	4.453	4.312	4.662	4.911	1.751	1.704	1.831	1.785	1.071	1.026	1.141	1.104
$p = .8$	4.467	4.777	4.378	4.870	1.785	1.833	1.726	1.841	1.070	1.132	1.063	1.121
$p = .9$	4.204	3.972	3.557	4.270	1.682	1.791	1.720	1.854	1.012	1.099	1.066	1.143
$x = .75$												
$p = .1$	2.415	2.506	2.463	2.446	1.074	1.116	1.087	1.140	.994	1.042	1.019	1.038
$p = .2$	2.685	2.786	2.862	2.814	1.311	1.317	1.337	1.302	1.009	1.026	1.084	1.054
$p = .3$	2.570	2.786	2.781	2.841	1.261	1.294	1.339	1.337	1.029	1.076	1.044	1.056
$p = .4$	2.518	2.543	2.760	2.784	1.252	1.246	1.349	1.312	1.098	1.020	1.007	1.048
$p = .5$	2.432	2.770	2.833	2.738	1.214	1.280	1.294	1.258	1.042	1.050	1.020	1.033
$p = .6$	2.633	2.692	3.086	2.740	1.194	1.125	1.261	1.194	.993	.935	.986	1.012
$p = .7$	2.669	2.559	3.064	2.785	1.117	1.029	1.176	1.130	.939	.933	1.003	.987
$p = .8$	2.728	2.875	2.912	2.822	1.074	1.133	1.040	1.077	.941	.998	.987	.959
$p = .9$	2.435	2.566	2.508	2.682	.961	1.007	.984	1.042	.825	.914	.915	.929

Table 1: Ratio of the Monte Carlo MSE of $\hat{Q}(p\,|\,x)$ to that of $\tilde{Q}(p\,|\,x)$ in the linear location model with $\alpha = 0$, $X \sim \mathscr{U}(0,1)$ and U unit logistic for different sample sizes and different values of p, x and β.

	$\beta = 10.392$				$\beta = 3.464$				$\beta = 1.155$			
	n =100	n =500	n =2500	$n = \infty$	n =100	n =500	n =2500	$n = \infty$	n =100	n =500	n =2500	$n = \infty$
$x = .25$												
$y = Q(.1 \mid x)$	2.572	2.942	2.622	2.628	1.309	1.191	1.134	1.035	1.077	.980	.976	.931
$y = Q(.2 \mid x)$	2.977	2.918	2.569	2.772	1.425	1.151	1.007	1.072	1.015	1.014	.962	.961
$y = Q(.3 \mid x)$	3.082	2.841	2.773	2.737	1.268	1.123	1.034	1.125	1.034	.999	.950	.988
$y = Q(.4 \mid x)$	3.196	2.864	2.899	2.689	1.244	1.146	1.180	1.189	1.069	.985	1.003	1.013
$y = Q(.5 \mid x)$	3.273	2.642	3.044	2.680	1.255	1.243	1.418	1.253	1.076	1.040	1.116	1.033
$y = Q(.6 \mid x)$	3.413	2.722	3.100	2.718	1.322	1.342	1.377	1.307	1.057	1.107	1.106	1.047
$y = Q(.7 \mid x)$	3.299	2.813	2.998	2.770	1.308	1.384	1.416	1.331	1.101	1.105	1.056	1.054
$y = Q(.8 \mid x)$	3.247	2.917	2.695	2.745	1.367	1.330	1.246	1.296	1.074	1.057	1.049	1.051
$y = Q(.9 \mid x)$	2.815	2.539	2.529	2.403	1.309	1.246	1.119	1.134	1.073	1.078	1.036	1.034
$x = .50$												
$y = Q(.1 \mid x)$	4.161	4.150	4.182	4.273	1.907	1.738	1.782	1.857	1.172	1.155	1.115	1.140
$y = Q(.2 \mid x)$	5.458	4.913	4.909	4.844	2.012	1.650	1.677	1.841	1.229	1.069	1.114	1.118
$y = Q(.3 \mid x)$	5.737	4.636	4.683	4.883	1.836	1.544	1.730	1.784	1.130	1.036	1.088	1.103
$y = Q(.4 \mid x)$	6.236	4.624	4.609	4.814	1.859	1.625	1.647	1.737	1.121	1.055	1.081	1.093
$y = Q(.5 \mid x)$	6.047	4.412	4.813	4.783	1.788	1.641	1.884	1.720	1.101	1.119	1.177	1.090
$y = Q(.6 \mid x)$	6.238	4.637	4.793	4.831	1.830	1.689	1.875	1.738	1.168	1.094	1.149	1.094
$y = Q(.7 \mid x)$	5.499	4.740	4.995	4.911	1.875	1.663	1.895	1.785	1.104	1.023	1.169	1.104
$y = Q(.8 \mid x)$	5.311	4.914	4.402	4.870	1.910	1.841	1.714	1.841	1.230	1.140	1.069	1.121
$y = Q(.9 \mid x)$	3.977	4.046	3.855	4.270	1.797	1.687	1.812	1.854	1.225	1.148	1.147	1.143
$x = .75$												
$y = Q(.1 \mid x)$	2.386	2.818	2.393	2.446	1.261	1.187	1.048	1.140	1.074	1.083	.980	1.038
$y = Q(.2 \mid x)$	2.923	2.741	3.030	2.814	1.275	1.325	1.415	1.302	1.050	1.042	1.164	1.054
$y = Q(.3 \mid x)$	3.390	2.905	2.813	2.841	1.322	1.310	1.362	1.337	1.072	1.054	1.039	1.056
$y = Q(.4 \mid x)$	3.545	3.033	2.771	2.784	1.368	1.370	1.365	1.312	1.110	1.124	1.043	1.048
$y = Q(.5 \mid x)$	3.903	3.110	3.036	2.738	1.391	1.342	1.356	1.258	1.136	1.084	1.067	1.033
$y = Q(.6 \mid x)$	3.658	2.849	3.167	2.740	1.398	1.232	1.280	1.194	1.101	.999	1.029	1.012
$y = Q(.7 \mid x)$	3.170	2.748	3.214	2.785	1.249	1.075	1.175	1.130	.993	.966	1.041	.987
$y = Q(.8 \mid x)$	3.254	2.994	3.025	2.822	1.280	1.158	1.063	1.077	1.042	1.033	.995	.959
$y = Q(.9 \mid x)$	2.660	2.499	2.490	2.682	1.309	1.064	1.031	1.042	1.077	.981	.953	.929

Table 2: Ratio of the Monte Carlo MSE of $\hat{F}(y \mid x)$ to that of $\tilde{F}(y \mid x)$ in the linear location model with $\alpha = 0$, $X \sim \mathscr{U}(0,1)$ and U unit logistic for different sample sizes and different values of p, x and β.

$\beta = 3.464$	$\eta = .633$				$\eta = .354$				$\eta = .08$			
	n =100	n =500	n =2500	$n = \infty$	n =100	n =500	n =2500	$n = \infty$	n =100	n =500	n =2500	$n = \infty$
$x = .25$												
$p = .1$	1.054	1.228	1.281	1.362	1.022	1.091	1.223	1.397	1.036	.982	1.143	1.433
$p = .2$	1.055	1.179	1.339	1.355	.982	1.134	1.368	1.387	.977	1.119	1.058	1.418
$p = .3$	1.090	1.141	1.237	1.311	1.057	1.152	1.205	1.341	.971	.944	1.099	1.372
$p = .4$	1.042	1.144	1.204	1.241	.968	1.048	1.146	1.266	1.029	1.009	1.033	1.293
$p = .5$	1.040	1.060	1.178	1.148	1.013	1.011	1.001	1.159	.948	.961	1.107	1.170
$p = .6$	.980	.993	1.045	1.036	1.070	.978	.972	1.025	.883	.984	1.010	1.007
$p = .7$	.893	.997	.982	.933	.917	.870	.970	.912	.920	.933	.938	.881
$p = .8$	.836	1.014	.989	.921	.955	.831	.949	.925	.893	.927	.833	.928
$p = .9$	.988	1.061	1.047	1.010	1.022	.944	1.044	1.033	.969	.946	1.002	1.056
$x = .50$												
$p = .1$	.579	.593	.643	.646	.523	.650	.585	.632	.545	.622	.609	.646
$p = .2$	.437	.454	.441	.456	.410	.505	.406	.459	.608	.564	.506	.511
$p = .3$	.378	.333	.365	.355	.321	.483	.410	.444	.570	.540	.557	.621
$p = .4$	.473	.515	.549	.590	.440	.698	.736	.756	.685	.625	.717	.923
$p = .5$	1.025	.979	1.056	1.053	.860	.975	1.145	1.090	.905	.859	.921	1.126
$p = .6$	.996	1.144	1.285	1.247	.918	1.079	1.244	1.212	.853	.869	.932	1.193
$p = .7$	1.018	1.235	1.345	1.255	.925	1.132	1.177	1.210	.835	.948	.980	1.178
$p = .8$	.980	1.260	1.281	1.204	.907	1.047	1.152	1.150	.860	.912	.989	1.106
$p = .9$	.972	1.157	1.154	1.121	.840	.996	1.108	1.048	.872	.869	.927	.975
$x = .75$												
$p = .1$	1.078	1.088	1.116	1.142	1.207	1.222	1.122	1.225	1.145	1.249	1.133	1.378
$p = .2$	1.030	1.155	1.172	1.219	1.291	1.324	1.249	1.271	1.480	1.408	1.385	1.363
$p = .3$	1.130	1.187	1.135	1.154	1.273	1.267	1.119	1.213	1.365	1.339	1.279	1.308
$p = .4$	1.141	1.149	1.040	1.126	1.052	1.246	1.168	1.178	1.337	1.268	1.269	1.263
$p = .5$	1.074	1.087	1.096	1.108	1.042	1.208	1.081	1.151	1.241	1.265	1.188	1.227
$p = .6$	1.031	1.157	1.061	1.092	1.068	1.210	1.180	1.131	1.141	1.227	1.135	1.204
$p = .7$	1.043	1.110	1.044	1.079	1.049	1.167	1.090	1.118	1.125	1.095	1.100	1.194
$p = .8$	1.036	1.012	1.133	1.072	1.088	1.119	1.127	1.115	1.183	1.081	1.106	1.199
$p = .9$	1.008	1.088	1.150	1.072	1.064	1.106	1.120	1.121	1.164	1.077	1.127	1.218

Table 3: Ratio of the Monte Carlo MSE of $\hat{Q}(p\,|\,x)$ to that of $\tilde{Q}(p\,|\,x)$ in the location-scale model with $\alpha = 0$, $\beta = 3.464$, $X \sim \mathscr{U}(0,1)$, $W = 1\{X > .4\}$ and U unit logistic for different sample sizes and different values of η and x.

β =**3.464**	$\eta = .633$				$\eta = .354$				$\eta = .08$			
	n =100	n =500	n =2500	$n = \infty$	n =100	n =500	n =2500	$n = \infty$	n =100	n =500	n =2500	$n = \infty$
$x = .25$												
$y = Q(.1 \mid x)$	1.114	1.292	1.375	1.362	1.200	1.192	1.298	1.397	1.154	1.049	1.137	1.433
$y = Q(.2 \mid x)$	1.196	1.231	1.290	1.355	1.254	1.134	1.372	1.387	1.243	1.049	1.066	1.418
$y = Q(.3 \mid x)$	1.179	1.181	1.251	1.311	1.270	1.194	1.278	1.341	1.202	1.043	1.105	1.372
$y = Q(.4 \mid x)$	1.164	1.112	1.268	1.241	1.222	1.145	1.154	1.266	1.165	1.022	1.016	1.293
$y = Q(.5 \mid x)$	1.144	1.128	1.179	1.148	1.138	1.034	1.044	1.159	1.186	.979	1.087	1.170
$y = Q(.6 \mid x)$	1.178	1.011	1.029	1.036	1.153	.988	.972	1.025	1.243	1.046	1.057	1.007
$y = Q(.7 \mid x)$	1.185	1.027	.982	.933	1.275	.908	.965	.912	1.350	.984	.904	.881
$y = Q(.8 \mid x)$	1.059	1.015	.985	.921	1.232	.917	.935	.925	1.316	1.047	.874	.928
$y = Q(.9 \mid x)$	.979	1.059	1.036	1.010	1.150	.961	.997	1.033	.917	.966	1.050	1.056
$x = .50$												
$y = Q(.1 \mid x)$	.610	.634	.660	.646	.519	.657	.599	.632	.672	.668	.612	.646
$y = Q(.2 \mid x)$	.463	.459	.455	.456	.388	.531	.402	.459	.524	.565	.530	.511
$y = Q(.3 \mid x)$	.434	.360	.357	.355	.362	.470	.411	.444	.501	.504	.539	.621
$y = Q(.4 \mid x)$	.452	.526	.583	.590	.395	.677	.745	.756	.591	.578	.684	.923
$y = Q(.5 \mid x)$	.682	.906	1.054	1.053	.589	.955	1.135	1.090	.720	.715	.860	1.126
$y = Q(.6 \mid x)$	1.010	1.217	1.282	1.247	.852	1.073	1.215	1.212	.888	.923	.964	1.193
$y = Q(.7 \mid x)$	1.142	1.305	1.331	1.255	1.092	1.095	1.168	1.210	.996	.925	.999	1.178
$y = Q(.8 \mid x)$	1.043	1.262	1.332	1.204	1.111	1.148	1.180	1.150	.945	.965	.974	1.106
$y = Q(.9 \mid x)$	1.086	1.145	1.158	1.121	.989	1.036	1.127	1.048	.890	.960	.968	.975
$x = .75$												
$y = Q(.1 \mid x)$	1.043	1.111	1.147	1.142	1.145	1.183	1.174	1.225	1.322	1.363	1.159	1.378
$y = Q(.2 \mid x)$	1.209	1.159	1.175	1.219	1.190	1.358	1.207	1.271	1.274	1.343	1.351	1.363
$y = Q(.3 \mid x)$	1.141	1.180	1.104	1.154	1.191	1.343	1.121	1.213	1.301	1.355	1.241	1.308
$y = Q(.4 \mid x)$	1.086	1.193	1.032	1.126	1.192	1.271	1.180	1.178	1.252	1.275	1.294	1.263
$y = Q(.5 \mid x)$	1.130	1.133	1.093	1.108	1.077	1.264	1.150	1.151	1.261	1.328	1.198	1.227
$y = Q(.6 \mid x)$	1.169	1.125	1.137	1.092	1.055	1.230	1.168	1.131	1.301	1.194	1.142	1.204
$y = Q(.7 \mid x)$	1.071	1.075	1.068	1.079	1.210	1.224	1.159	1.118	1.370	1.097	1.106	1.194
$y = Q(.8 \mid x)$	1.137	1.063	1.117	1.072	1.299	1.211	1.114	1.115	1.281	1.157	1.101	1.199
$y = Q(.9 \mid x)$	1.234	1.116	1.168	1.072	1.285	1.183	1.181	1.121	1.430	1.175	1.123	1.218

Table 4: Ratio of the Monte Carlo MSE of $\hat{F}(y \mid x)$ to that of $\tilde{F}(y \mid x)$ in the location-scale model with $\alpha = 0$, $\beta = 3.464$, $X \sim \mathscr{U}(0,1)$, $W = 1\{X > .4\}$ and U unit logistic for different sample sizes and different values of η and x.

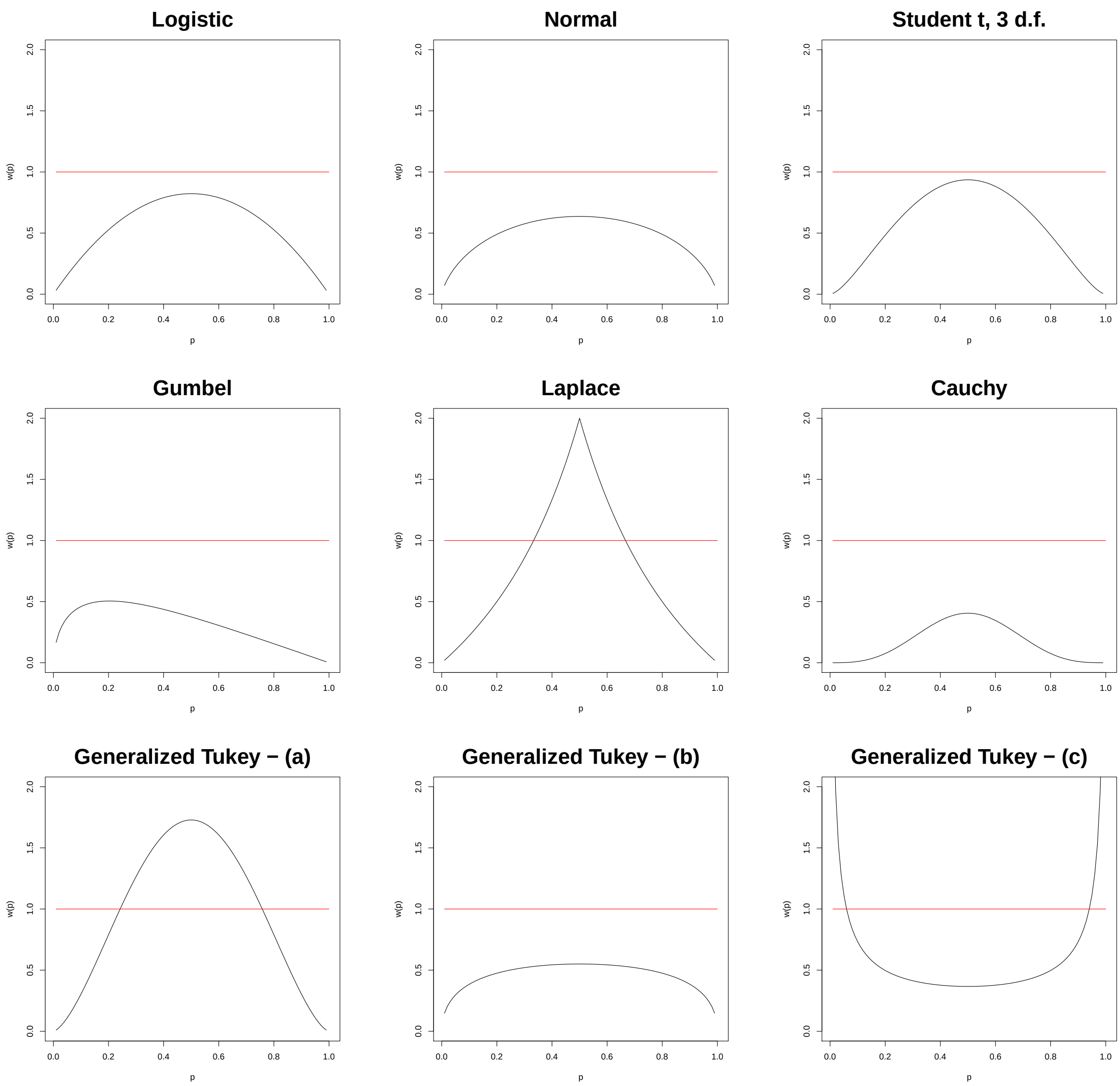


Figure 1: $ARE(\tilde{\theta}_p, \hat{\theta}_{OLS})$ for different error distributions in the linear location model. The parameters for the generalized Tukey distribution are (a): $\lambda_3 = -.25$; (b): $\lambda_3 = .25$; (c): $\lambda_3 = .75$.

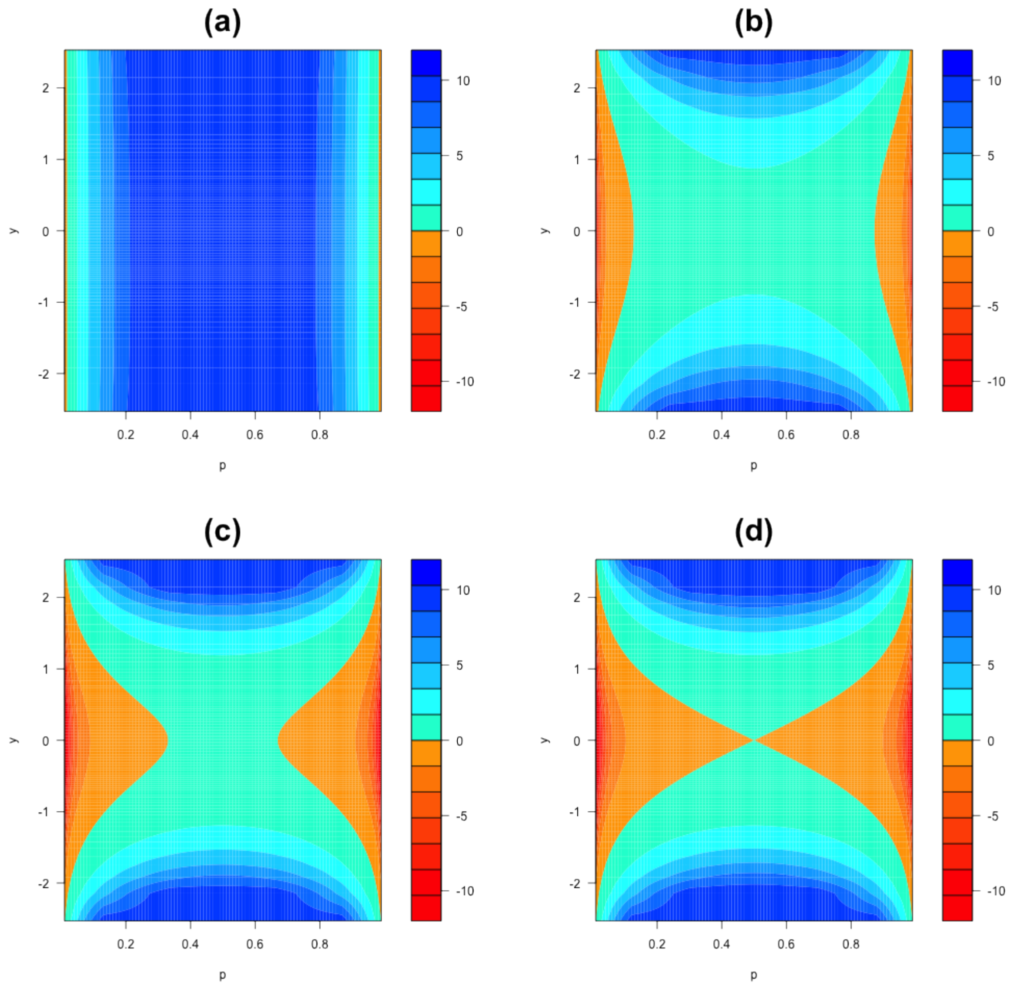


Figure 2: Contour plots of $ARE(\tilde{\theta}_p, \hat{\theta}_y) - 1$ (blue area) and $1 - ARE(\hat{\theta}_y, \tilde{\theta}_p)$ (orange-red area) in the linear location model with $\alpha = 0$ and $X \sim \mathscr{U}(0, 1)$ for different values of the regression R^2. Panel (a): $R^2 = .90$; Panel (b): $R^2 = .50$; Panel (c): $R^2 = .10$. Panel (d): $R^2 = 0$.

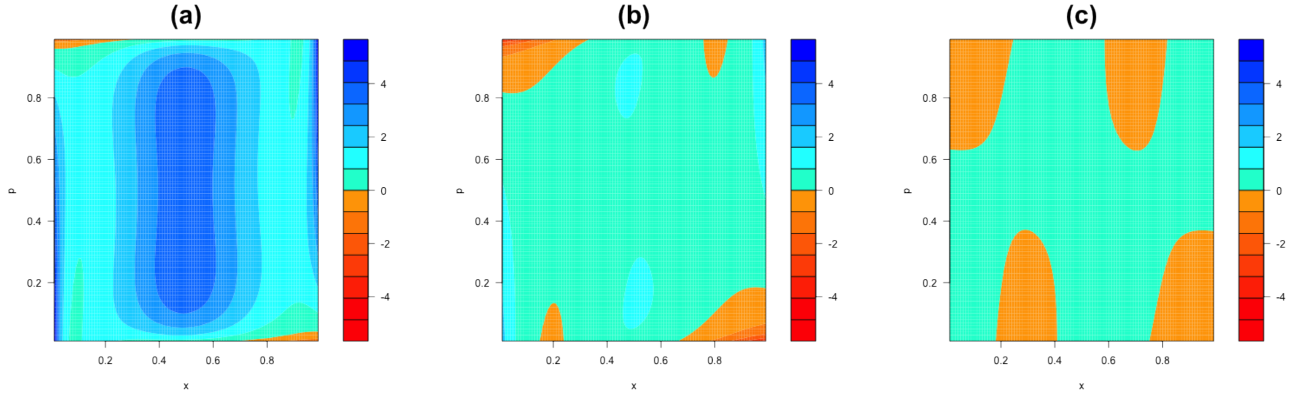


Figure 3: Contour plots of $ARE(\tilde{Q}(p\,|\,x), \hat{Q}(p\,|\,x)) - 1$ (blue area) and $1 - ARE(\hat{Q}(p\,|\,x), \tilde{Q}(p\,|\,x))$ (orange-red area) in the linear location model with $\alpha = 0$ and $X \sim \mathscr{U}(0,1)$ for different values of p and x and three values of the regression R^2. Panel (a): $R^2 = .90$; Panel (b): $R^2 = .50$; Panel (c): $R^2 = .10$.

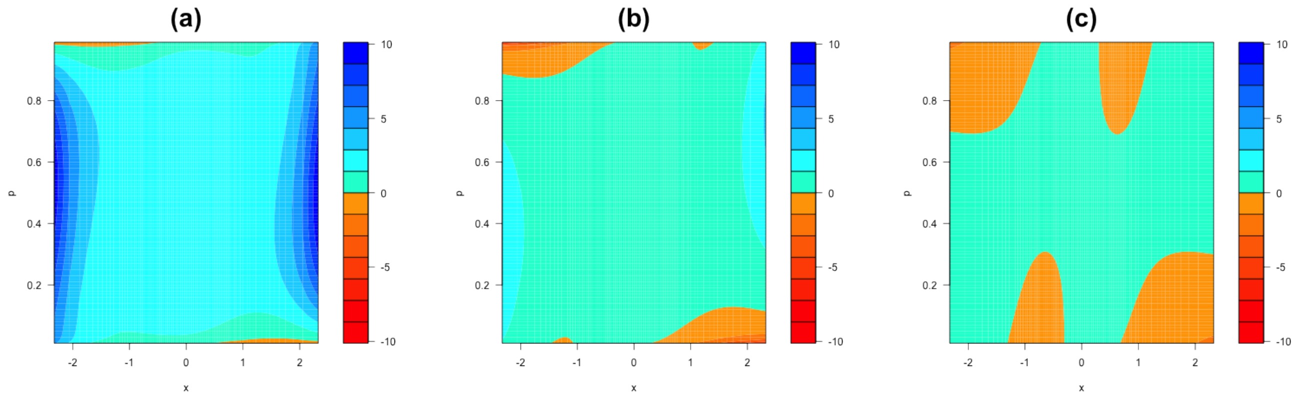


Figure 4: Contour plots of $ARE(\tilde{Q}(p\,|\,x), \hat{Q}(p\,|\,x)) - 1$ (blue area) and $1 - ARE(\hat{Q}(p\,|\,x), \tilde{Q}(p\,|\,x))$ (orange-red area) in the linear location model with $\alpha = 0$ and $X \sim \mathscr{N}(0,1)$ for different values of p and x and three values of the regression R^2. Panel (a): $R^2 = .90$; Panel (b): $R^2 = .50$; Panel (c): $R^2 = .10$.

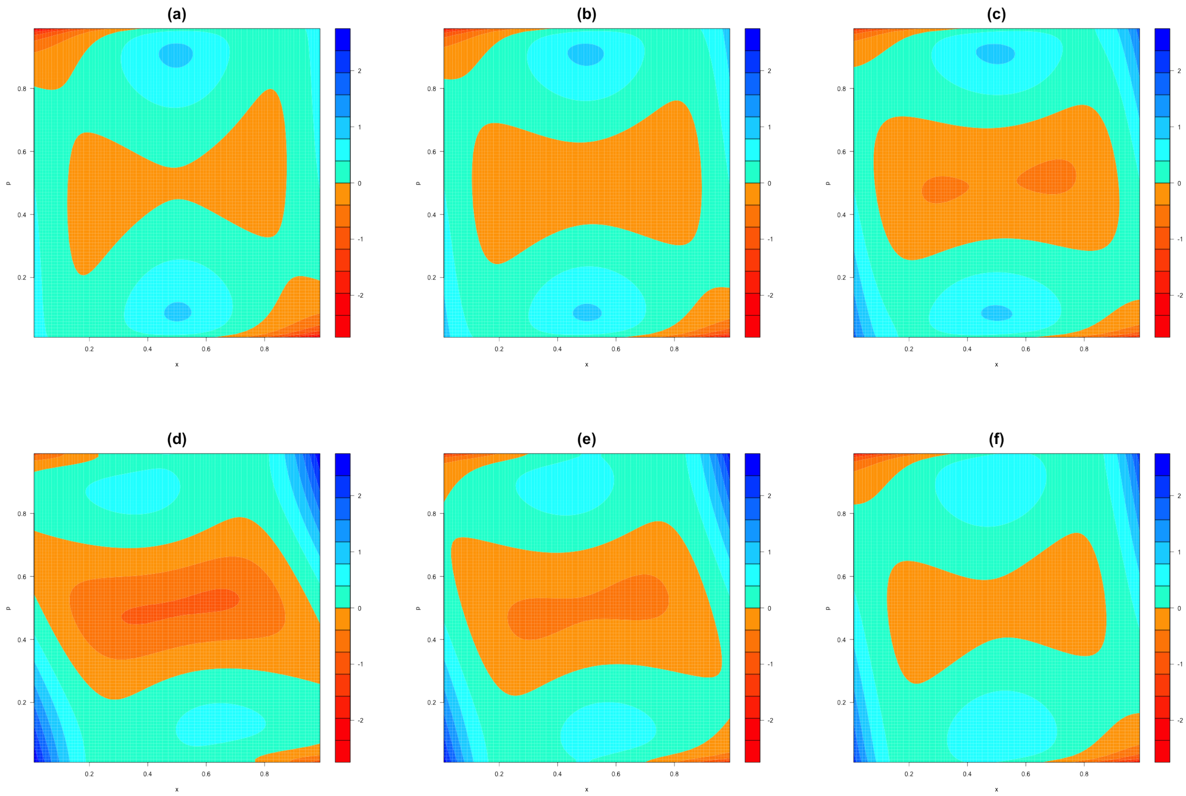


Figure 5: Contour plots of $ARE(\tilde{Q}(p \mid x, w), \hat{Q}(p \mid x, w)) - 1$ (blue area) and $1 - ARE(\hat{Q}(p \mid x, w), \tilde{Q}(p \mid x, w))$ (orange-red area) in the linear location-scale model with $\alpha = 0$, $\beta = 3.364$, $X \sim \mathscr{U}(0, 1)$ and $W \sim \mathscr{B}(.5)$ for different values of p, x, and w and three values of η. Panel (a): $\eta = 0.63$, $w = 0$; Panel (b): $\eta = 0.35$, $w = 0$; Panel (c): $\eta = 0.08$, $w = 0$; Panel (d): $\eta = 0.63$, $w = 1$; Panel (e): $\eta = 0.35$, $w = 1$; Panel (f): $\eta = 0.08$, $w = 1$

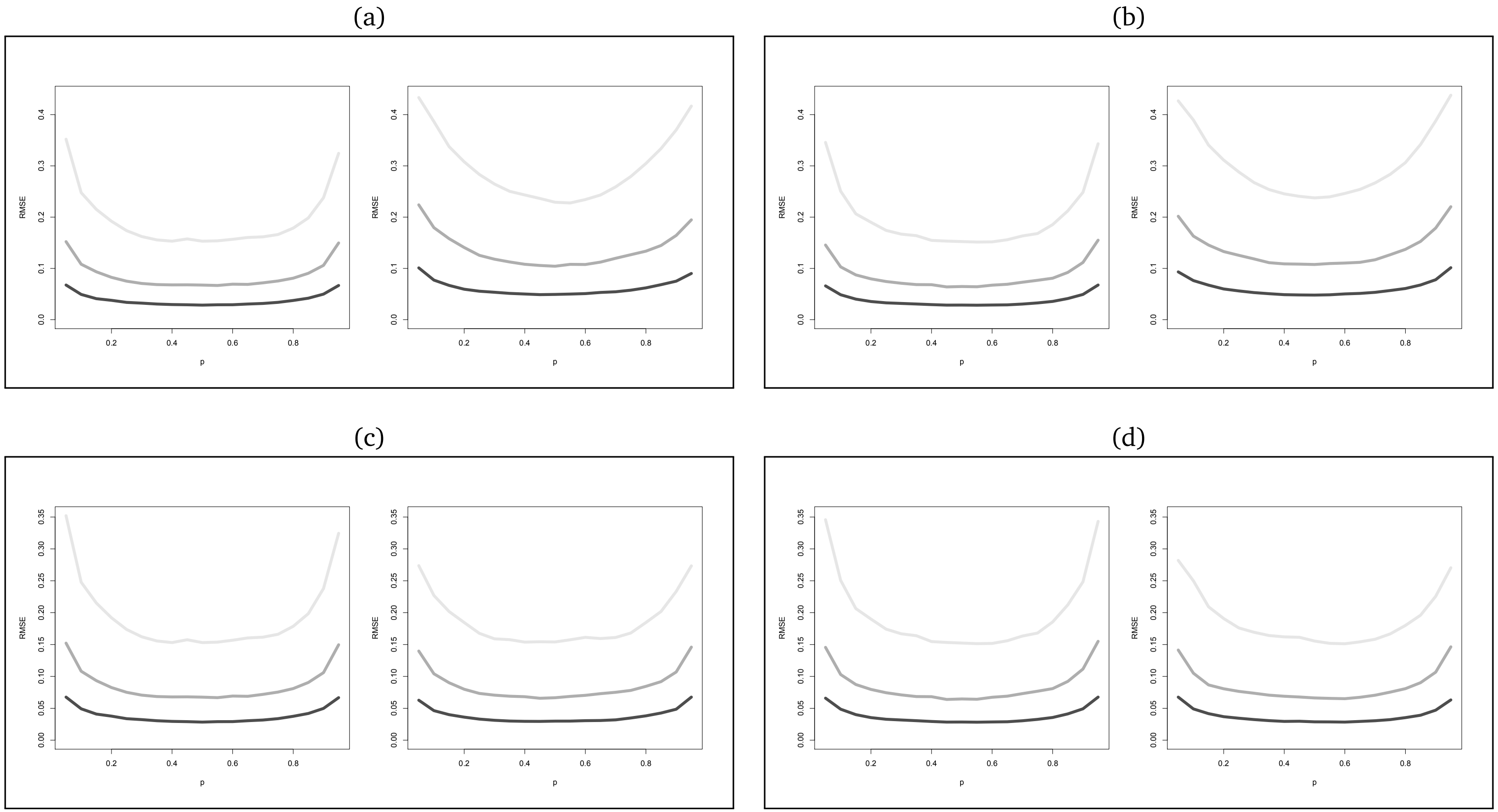


Figure 6: Monte Carlo RMSE of $\tilde{Q}(p \mid x)$ (left sub-panels) and $\hat{Q}(p \mid x)$ (right sub-panels) in the linear location model with $\alpha = 0$ and $X \sim \mathscr{U}(0,1)$ for different sample sizes (from light to dark gray $n = 100, 500, 2500$) and different values of β and x. Panel (a): $\beta = 10.392$, $x = .25$; Panel (b): $\beta = 10.392$, $x = .75$; Panel (c): $\beta = 1.155$, $x = .25$ Panel (d): $\beta = 1.155$, $x = .75$

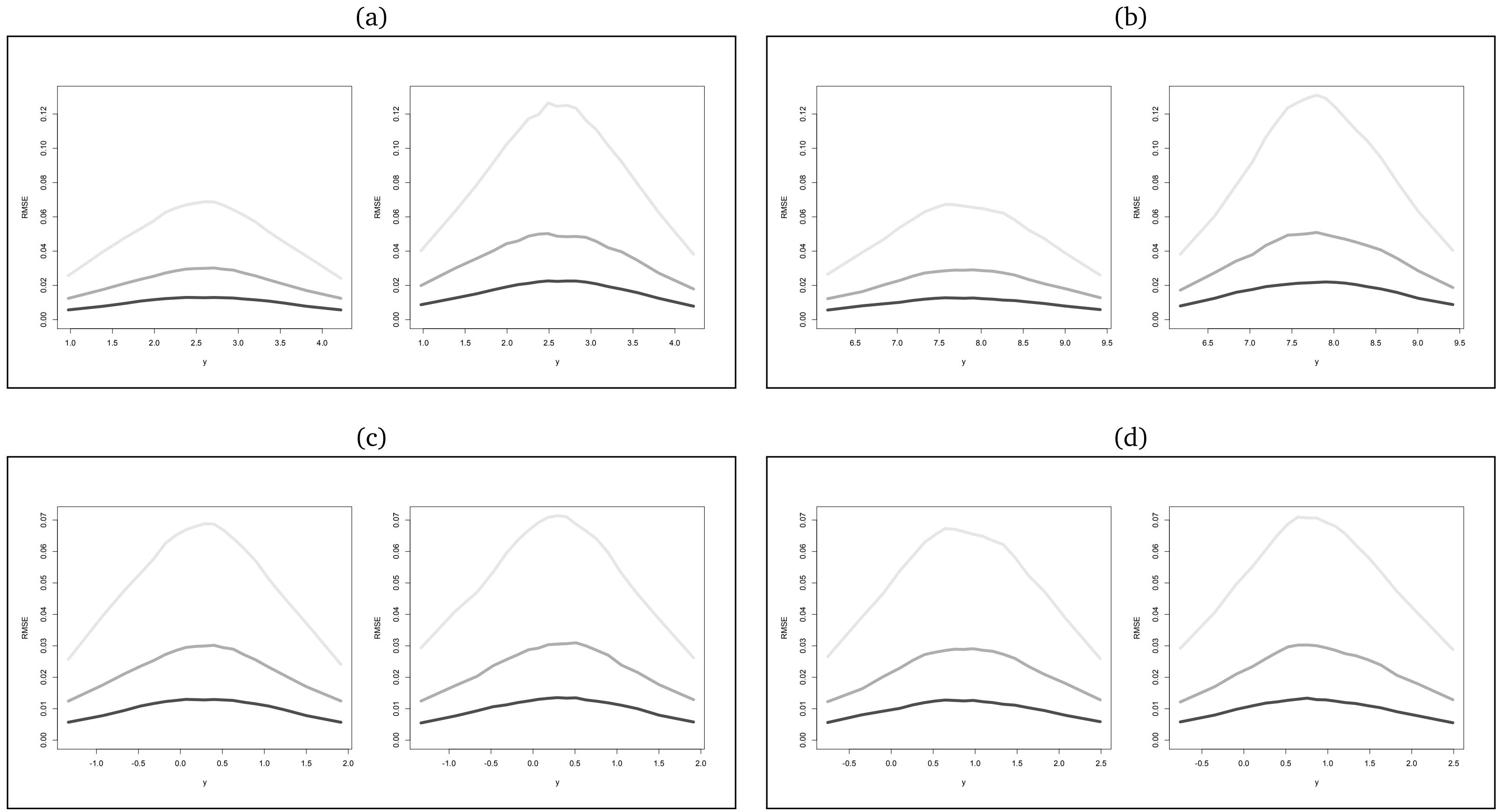


Figure 7: Monte Carlo RMSE of $\tilde{F}(y\,|\,x)$ (left sub-panels) and $\hat{F}(y\,|\,x)$ (right sub-panels) in the linear location model with $\alpha = 0$ and $X \sim \mathscr{U}(0,1)$ for different sample sizes (from light to dark gray n =100, 500, 2500) and different values of β and x. Panel (a): $\beta = 10.392$, $x = .25$; Panel (b): $\beta = 10.392$, $x = .75$; Panel (c): $\beta = 1.155$, $x = .25$ Panel (d): $\beta = 1.155$, $x = .75$

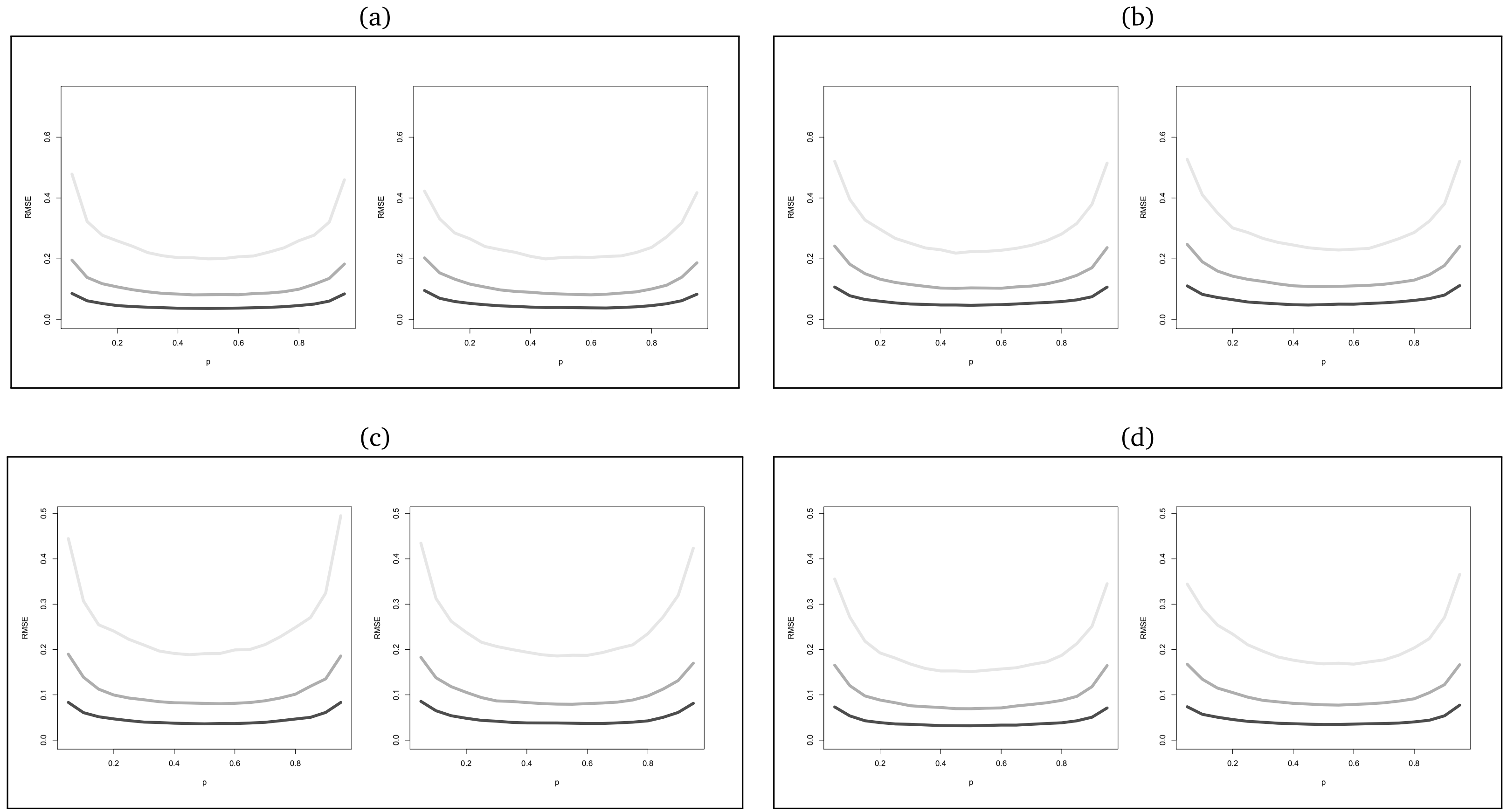


Figure 8: Monte Carlo RMSE of $\tilde{Q}(p \mid x, w)$ (left sub-panels) and $\hat{Q}(p \mid x, w)$ (right sub-panels) in the linear location-scale model with $\alpha = 0$, $\beta = 3.464$, $X \sim \mathscr{U}(0, 1)$ and $W = 1\{X > .4\}$ for different sample sizes (from light to dark gray $n = 100$, 500, 2500) and different values of η and x. Panel (a): $\eta = 0.633$, $x = .25$; Panel (b): $\eta = 0.633$, $x = .75$; Panel (c): $\eta = 0.08$, $x = .25$ Panel (d): $\eta = 0.08$, $x = .75$

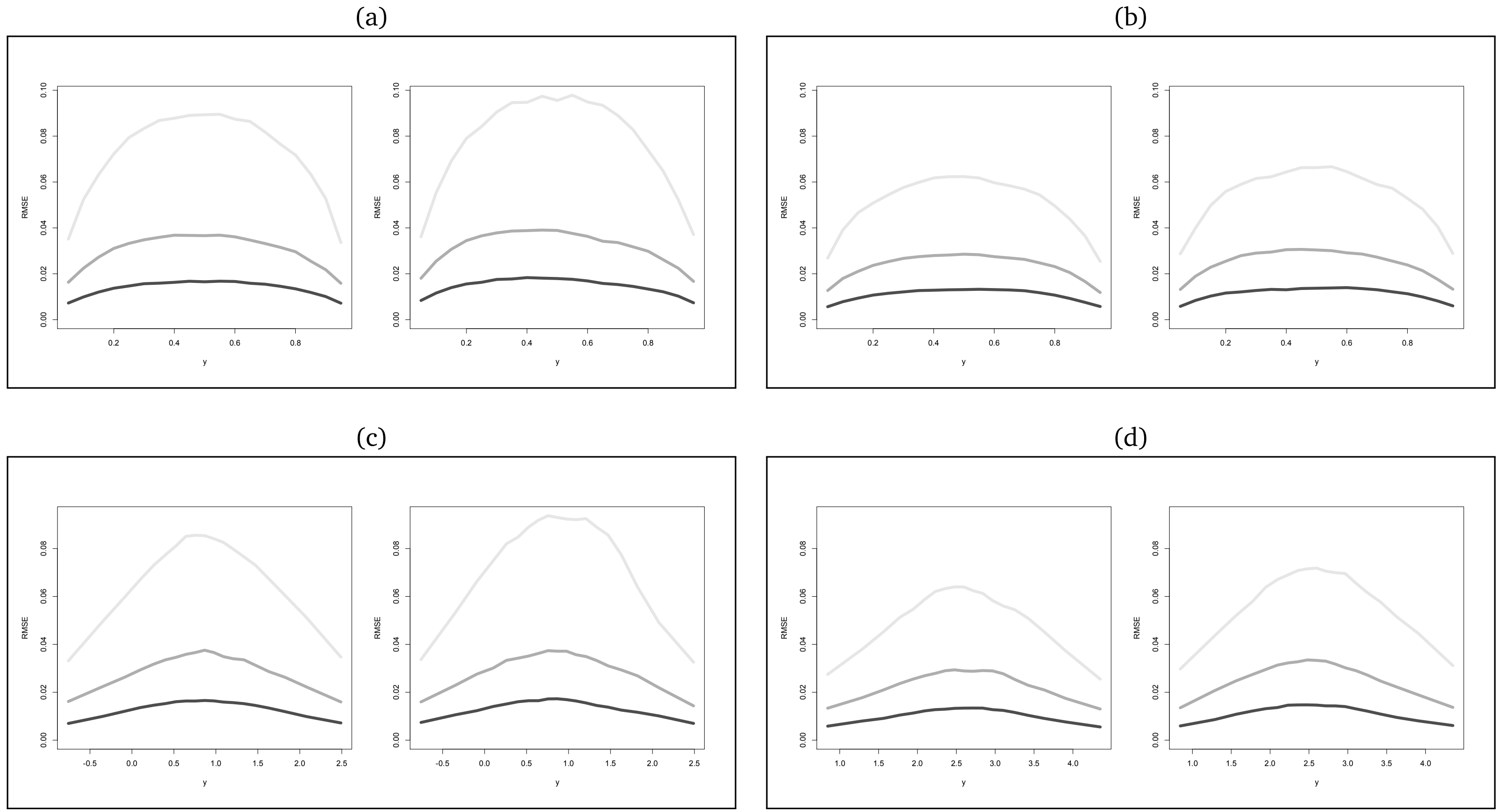


Figure 9: Monte Carlo RMSE of $\tilde{F}(y \mid x, w)$ (left sub-panels) and $\hat{F}(y \mid x, w)$ (right sub-panels) in the linear location-scale model with $\alpha = 0$, $\beta = 3.464$, $X \sim \mathscr{U}(0, 1)$ and $W = 1\{X > .4\}$ for different sample sizes (from light to dark gray $n = 100, 500, 2500$) and different values of η and x. Panel (a): $\eta = 0.633$, $x = .25$; Panel (b): $\eta = 0.633$, $x = .75$; Panel (c): $\eta = 0.08$, $x = .25$ Panel (d): $\eta = 0.08$, $x = .75$

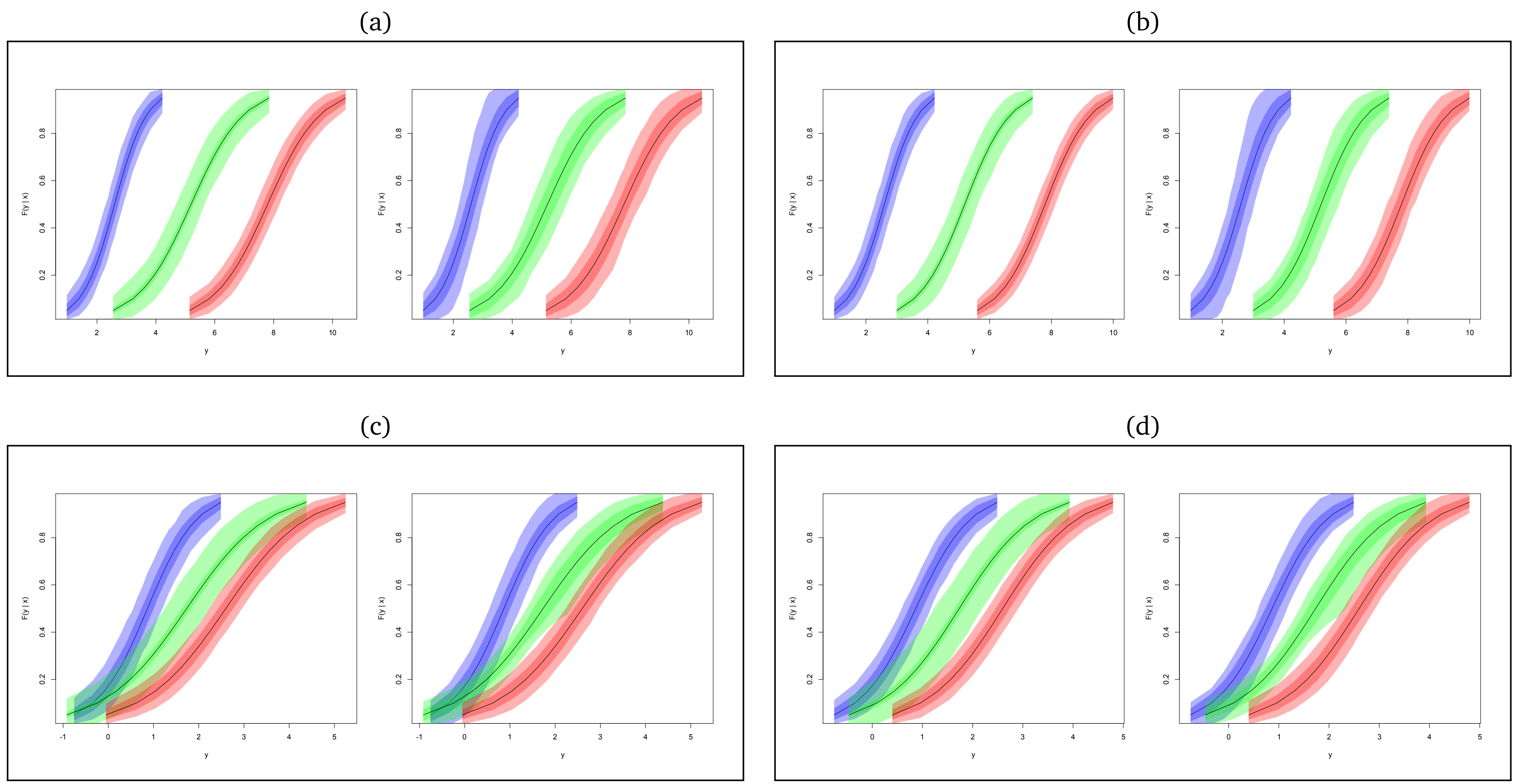
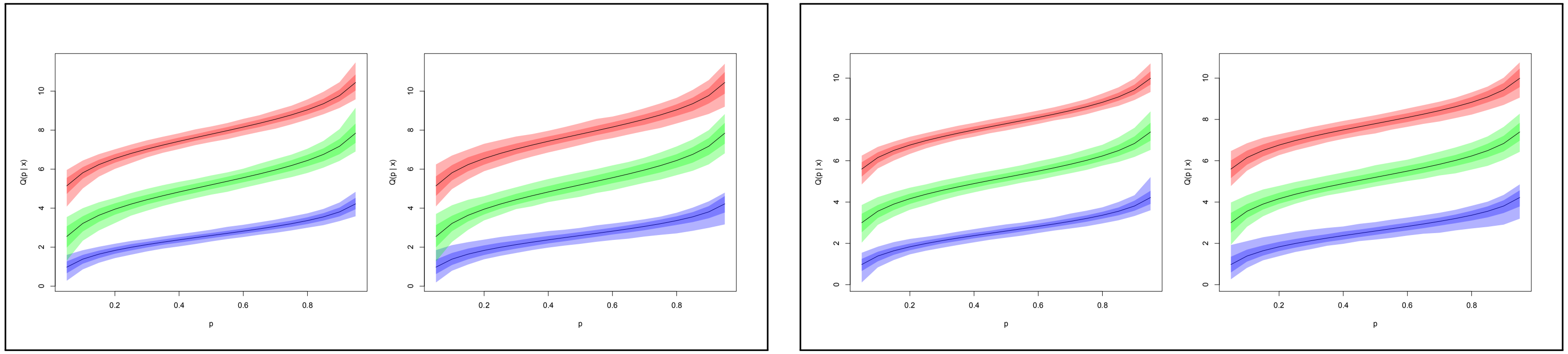


Figure 10: Monte Carlo pointwise confidence bands for $F(y\,|\,x)$ based on $\hat{F}(y\,|\,x,w)$ (left-subpanel) and $\tilde{F}(y\,|\,x,w)$ (right sub-panel) in the linear location-scale model with $\alpha = 0$ and $X \sim \mathscr{U}(0,1)$ for different sample sizes (n = 100, 500) and x equal to the first (blue), second (green) and third (red) quartile of X. Panel (a): $\beta = 10.382$, $\eta = 0.633$; Panel (b): $\beta = 10.382$, $\eta = 0.354$; Panel (c): $\beta = 3.464$, $\eta = 0.633$; Panel (d):$\beta = 3.464$, $\eta = 0.354$

(a) (b)



(c) (d)

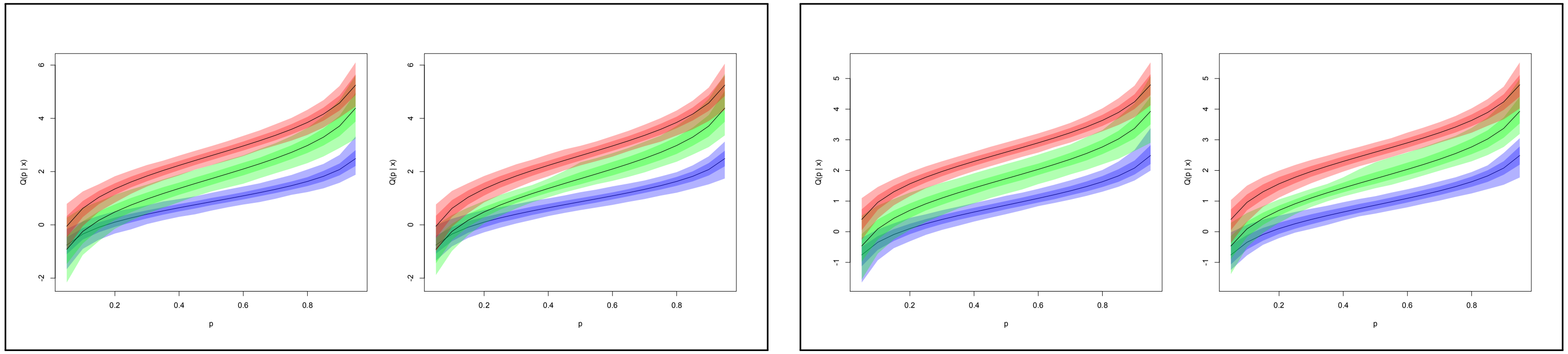


Figure 11: Monte Carlo pointwise confidence bands for $Q(p\,|\,x)$ based on $\hat{Q}(p\,|\,x,w)$ (left sub-panel) and $\tilde{Q}(p\,|\,x,w)$ (right sub-panel) in the linear location-scale model with $\alpha = 0$ and $X \sim \mathscr{U}(0,1)$ for different sample sizes ($n = 100$, 500) and x equal to the first (blue), second (green) and third (red) quartile of X. Panel (a): $\beta = 10.382$, $\eta = 0.633$; Panel (b): $\beta = 10.382$, $\eta = 0.354$; Panel (c): $\beta = 3.464$, $\eta = 0.633$; Panel (d):$\beta = 3.464$, $\eta = 0.354$

A Proofs

Proof of Theorem 2

The proof uses the functional Delta method. Recall that, given two normed space D and E, a map $\tau: D \mapsto E$ is Hadamard differentiable at $H \in C \subset D$ if there exists a continuous linear map $\tau'_H: D \mapsto E$ such that $r_n^{-1}[\tau(H + r_n Z_n) - \tau(H)] \to \tau'_H(Z)$ for all $r_n \to 0$ and all $Z_n \to Z$ such that $H + r_n Z_n \in C$. Let H_n be a sequence of stochastic processes in D, and let the function $H \in C$ be such that $r_n(H_n - H) \Rightarrow Z$, where Z is stochastic process in C. If τ is Hadamard differentiable, then $r_n[\tau(H_n) - \tau(H)] \Rightarrow \tau'_H(Z)$.

In our case, for any given x, let $H_n(y) = \hat{F}(y \,|\, x) = \Lambda(\boldsymbol{x}^\top \hat{\nu}(y))$ and $H(y) = \Lambda(\boldsymbol{x}^\top \nu_D(y))$, and let τ be the inverse map $\tau(H)(p) = H^{-1}(p) = \inf\{y : \Lambda(\boldsymbol{x}^\top \nu_D(y)) \geq p\}$, which satisfies $\tau(H(y)) = y$. The map τ is Hadamard differentiable because $\nu_D(y)$ is a differentiable function of y and $\Lambda(\boldsymbol{x}^\top \nu)$ is a differentiable function of ν. The derivative of τ at H tangentially to $C[0,1]$ is

$$M_D = \tau'_H(Z_x) = -\frac{Z_x(H^{-1})}{H'(H^{-1})},$$

where $Z_x(y)$ is the zero mean Gaussian process with covariance function given by (11) and

$$\begin{aligned}
H'(y) = \frac{\partial \Lambda(\boldsymbol{x}^\top \nu_D(y))}{\partial y} &= \Lambda'(\boldsymbol{x}^\top \nu_D(y))\boldsymbol{x}^\top \nu'_D(y) \\
&= c\,\Lambda(\boldsymbol{x}^\top \nu_D(y))[1 - \Lambda(\boldsymbol{x}^\top \nu_D(y))]\,\boldsymbol{x}^\top \nu'_D(y) \\
&= c^{-1}\, w(H(y))\,\boldsymbol{x}^\top \nu'_D(y).
\end{aligned}$$

Therefore,

$$\tau'_H(Z_x) = \frac{c\, Z_x(H^{-1})}{w(H(H^{-1}))\, \nu'_D(H^{-1})\boldsymbol{x}}. \tag{22}$$

It then follows that the covariance function of the process M_D is

$$\operatorname{cov}(M_D(p), M_D(p')) = \frac{c^2\, w(p)w(p')\,\boldsymbol{x}^\top V_D(H^{-1}(p))\rho_Q(p,p')V_D(H^{-1}(p'))\boldsymbol{x}}{c^2\, w(p)w(p')\,\boldsymbol{x}^\top \nu'_D(H^{-1}(p))\;\boldsymbol{x}^\top \nu'_D(H^{-1}(p'))}, \tag{23}$$

where $\rho_Q(p,p') = \Sigma_D(y_p, y_{p'})$ and $H^{-1}(p) = y_p$.

In particular, under the linear location model (1), $\nu_D(y)$ depends on y only through $\gamma_y = y - \alpha$, so $\boldsymbol{x}^\top \nu'_D(y) = 1$ and $\rho_Q(p,p')$ simplifies to

$$\begin{aligned}
\rho_Q(p,p') &= \mathrm{E}\left[c^2\left(1\{Y \leq H^{-1}(p)\} - \Lambda(\gamma_{y_p} + \delta^\top X)\right)\left(1\{Y \leq H^{-1}(p')\} - \Lambda(\gamma_{y_{p'}} + \delta^\top X)\right)\boldsymbol{X}\boldsymbol{X}^\top\right] \\
&= \mathrm{E}\left[c^2\left(\Lambda(\gamma_{y_p} + \delta^\top X) - \Lambda(\gamma_{y_p} + \delta^\top X)\,\Lambda(\gamma_{y_{p'}} + \delta^\top X)\right)\boldsymbol{X}\boldsymbol{X}^\top\right],
\end{aligned}$$

with $p \leq p'$. Therefore, in this case, $\rho(p,p) = \mathrm{E}[w(H^{-1}(p))\boldsymbol{X}\boldsymbol{X}^\top] = V_D(y_p)^{-1}$ and the asymptotic variance of $\hat{Q}(p \,|\, x)$ simplifies to $V(\hat{Q}(p \,|\, x)) = \boldsymbol{x}^\top V_D(y_p)\boldsymbol{x}$.

Proof of Theorem 4

The proof follows the same arguments as the proof of Theorem 2.

For any given x, let $H_n(p) = \boldsymbol{x}^\top \hat{\nu}(p)$, $H(p) = \boldsymbol{x}^\top \nu_Q(p)$ and $Z(\cdot) = \boldsymbol{x}^\top J^{-1}(\cdot)B(\cdot)$. Let τ be the inverse map $\tau(H)(y) = H^{-1}(y) = \inf\{p : \boldsymbol{x}^\top \nu_Q(p) \geq y\}$, which satisfies $\tau(\boldsymbol{x}^\top \nu_Q(p)) = p$ for all p. Because $\nu_Q(p)$ is differentiable in p and H is linear in x, τ is Hadamard differentiable and its derivative at H, tangentially to $C = C[a, b]$, is

$$M_Q = \tau'_H(Z) = -\frac{Z(H^{-1})}{H'(H^{-1})}$$

(see Lemma 3.9.23 in van der Vaart and Wellner 1996). Moreover,

$$H'(p) = \frac{\partial \boldsymbol{x}^\top \nu_Q(p)}{\partial p} = \boldsymbol{x}^\top J(p)^{-1}\mu_X.$$

Therefore

$$\tau'_H(Z) = -\frac{Z(H^{-1})}{\boldsymbol{x}^\top J(H^{-1})^{-1}\mu_X} = -\frac{\boldsymbol{x}^\top J(H^{-1})^{-1}B(H^{-1})}{\boldsymbol{x}^\top J(H^{-1})^{-1}\mu_X}. \tag{24}$$

Proof of Theorem 5

For notational convenience, we drop the x argument in the CDF and the CQF, their estimates and the related limiting Gaussian processes, and simply write $F(y) = F(y \mid x)$, $Q(p) = Q(p \mid x)$, and so on. We denote by Z_F the limiting Gaussian process of $\sqrt{n}(\hat{F} - F)$ and by σ_F^2 its variance function. Similarly, we denote by Z_Q the limiting Gaussian process of $\sqrt{n}(\tilde{Q} - Q)$ and by σ_Q^2 its variance function.

The functional $\tau(F) = F^{-1}$ is Hadamard differentiable since F is continuously differentiable. Then, following the same arguments as in the proofs of Theorems 4 and 2,

$$\sqrt{n}[\tau(\hat{F}_n) - \tau(F)] = \sqrt{n}(\hat{Q}_n - Q) \Rightarrow -\frac{Z_F(F^{-1})}{F'(F^{-1})}.$$

Consequently, $V(\hat{Q}_n) = \sigma_F^2(F^{-1})/F'(F^{-1})^2$ and

$$\frac{V(\hat{F}_n(y))}{V(\hat{Q}_n(p))} = \frac{\sigma_F^2(y)}{\sigma_F^2(Q(p))/f(Q(p))^2} \tag{25}$$

Similarly, the fact that

$$\sqrt{n}[\tau(\tilde{Q}_n) - \tau(Q)] = \sqrt{n}(\tilde{F}_n - F) \Rightarrow -\frac{Z_Q(Q^{-1})}{Q'(Q^{-1})} = -Z_Q(F)f(Q(Q^{-1}))$$

implies that $V(\tilde{F}_n(y)) = f^2\sigma_Q^2(F(y))$ and

$$\frac{V(\tilde{F}_n(y))}{V(\tilde{Q}_n(p))} = \frac{f(y)^2\,\sigma_Q^2(F(y))}{\sigma_Q^2(p)} \tag{26}$$

Finally, setting $p = F(y)$ in (25) and (26) gives the result.

Proof of Theorem 6

Uniform consistency.

Because of assumption (ii) and of of $\ln(1+x) \le \max(\ln 2, \ln(2x))$,

$$\mathrm{E}\left| cD_y r(Z;\nu) - \ln(1+e^{cr(Z;\nu)})] \right| < \infty,$$

where $c = \pi/\sqrt{3}$. Next we show that $L(\nu; y) = \ell(\nu; y) + o_p(1)$ uniformly in $(\nu, y) \in \Theta \times \mathbb{R}$. This is because $L(\nu; y) = \ell(\nu; y) + o_p(1)$ for all fixed (ν, y), by Khintchine law of large numbers, and the process

$$(\nu, y) \mapsto L(\nu; y) = n^{-1}\sum_{i=1}^{n}\left[cD_{iy}r(Z_i;\nu) - \ln(1+e^{cr(Z_i;\nu)})]\right]$$

is stochastically equicontinuous.

In fact, for every (ν, y) and (ν', y'),

$$\begin{aligned}
&\left|L(\nu;y) - L(\nu';y')\right| \\
&\le\ c\left|\frac{1}{n}\sum_{i=1}^{n}(D_{iy} - D_{y',i})r(Z_i;\nu)\right| + c\left|\frac{1}{n}\sum_{i=1}^{n}D_{y',i}\left(r(Z_i;\nu) - r(Z_i;\nu')\right)\right| + \left|\frac{1}{n}\sum_{i=1}^{n}\ln\left(\frac{1+e^{cr(Z_i;\nu)}}{1+e^{cr(Z_i;\nu')}}\right)\right| \\
&\le\ c\sup_{\nu\in\Theta}\mathrm{E}\left|r(Z;\nu)\right|\mathrm{E}_Z\left|P(Y \le y \mid Z) - P(Y \le y' \mid Z)\right|(1+O_p(n^{-1/2})) + \\
&\qquad +c\ \mathrm{E}\left|r(Z;\nu) - r(Z;\nu')\right|(1+o_p(n^{-1/2})) + \frac{1}{n}\sum_{i=1}^{n}\left|\ln\left(\frac{1+e^{cr(Z_i;\nu)}}{1+e^{cr(Z_i;\nu')}}\right)\right| \\
&\le\ c\left[\sup_{\nu\in\Theta}\mathrm{E}\left|r(Z;\nu)\right|\mathrm{E}_Z\left|P(Y \le y \mid Z) - P(Y \le y' \mid Z)\right| + \mathrm{E}\left|r(Z;\nu) - r(Z;\nu')\right|\right](1+O_p(n^{-1/2})) \\
&\qquad +c\,\frac{1}{n}\sum_{i=1}^{n}\Lambda(r(Z_i;\nu'))\left|\frac{\partial r(Z_i;\nu')}{\partial \nu}^{\top}(\nu - \nu') + O(\|\nu - \nu'\|^2)\right|.
\end{aligned}$$

where the last inequality follows from the Taylor expansions:

$$\begin{aligned}
\log\left(1+e^{cr(Z_i;\nu)}\right) &= \log\left(1+e^{cr(Z_i;\nu')}\right) + \frac{\partial}{\partial \nu}\log\left(1+e^{cr(Z_i;\nu)}\right)^{\top}(\nu' - \nu) + O_p\left(\|\nu' - \nu\|^2\right) \\
&= \log\left(1+e^{cr(Z_i;\nu')}\right) + \frac{\Lambda'(r(Z_i;\nu'))}{1-\Lambda(r(Z_i;\nu'))}\frac{\partial r(Z;\nu')}{\partial \nu}^{\top}(\nu' - \nu) + O_p\left(\|\nu' - \nu\|^2\right)
\end{aligned}.$$

>From this chain of inequalities we get

$$\begin{aligned}
&\left|L(\nu;y) - L(\nu';y')\right| \\
&\le\ c\sup_{\nu\in\Theta}\mathrm{E}\left|r(Z;\nu)\right|\sup_{z}\left|F(y \mid z) - F(y' \mid z)\right|(1+O_p(n^{-1/2})) + \\
&\qquad +c\ \mathrm{E}\left|\frac{\partial r(Z;\nu')}{\partial \nu}^{\top}(\nu - \nu') + O(\|\nu - \nu'\|^2)\right|(1+o_p(1)).
\end{aligned}$$

Given stochastic equicontinuity, the fact that $\ell(\nu, y)$ reaches its maximum at $\nu_D(y)$ while $L(\hat{\nu}(y), y) \geq L(\nu_D(y), y) - o_p(1)$ implies that $o_P(1) \geq \ell(\nu_D(y), y) - \ell(\hat{\nu}(y), y) > 0$. If that was not true, in fact we could find, for an arbitrary large n, a constant $C > 0$ such that $\ell(\nu_D(y), y) - \ell(\hat{\nu}(y), y) > C$ that would imply

$$C + o_P(1) \leq \ell(\nu_D(y), y) - \ell(\hat{\nu}(y), y) = L(\nu_D(y), y) - L(\hat{\nu}(y), y) + o_P(1)$$

that contradicts $L(\hat{\nu}(y), y) \geq L(\nu_D(y), y) - o_p(1)$. Uniform consistency of $\hat{\nu}(\cdot)$ then follows from Corollary 3.2.3 in van der Vaart and Wellner (1996).

Asymptotic normality. First of all, under the assumptions of Theorem 6 the classe of functions $\mathscr{D} = \{D_y = 1\{Y \leq y\},\ y \in \mathbb{R}\}$ is a Donsker class.

Let the parameter space Θ be decomposed in $\Theta = \Theta_1 \times \Theta_2$, with $(\gamma, \delta) \in \Theta_1$ and $\eta \in \Theta_2$. The class $\mathscr{F}_1 = \left\{\gamma + z^\top \delta,\ (\gamma, \delta) \in \Theta_1\right\}$ is a VC-class of functions because the subgraphs are given by the set $\{z :\ \gamma + z^\top \delta \leq t\}$ for $t \in \mathbb{R}$ and $(\gamma, \delta) \in \Theta_1$ is a class of half-lines with VC-index equal to $k+2$, where k is the dimension of the vector of regressors z (see van der Vaart and Wellner 1996, Chapter 2.6, Problems and Complements No. 14). By Theorem 2.6.7 in van der Vaart and Wellner (1996), if F is a square integrable envelope function for $\mathscr{F}_1$, the covering numbers of $\mathscr{F}_1$ are bounded above by

$$N(\varepsilon\|F\|_2, \mathscr{F}_1, L_2(P)) \leq K\varepsilon^{-4},$$

where K only depends on the dimension of z. This implies that the entropy condition

$$\int_0^\delta \sqrt{\ln N(\varepsilon\|F\|_2, \mathscr{F}_1, L_2(P))}\, d\varepsilon < \infty$$

is satisfied, so $\mathscr{F}_1$ is a Donsker class.

Similar considerations hold for the class $\mathscr{F}_2 = \left\{1 + z^\top \eta,\ \eta \in \Theta_2\right\}$. Moreover, assumption (ii) implies also that $1/\mathscr{F}_2$ is Donsker (Example 2.10.9 in Van der Vaart and Wellner, 1996) and allows to apply Corollary 2.10.13 of Van der Vaart and Wellner (1996) and to conclude that the class $\mathscr{R} = \left\{r(Z; \nu),\ \nu = (\gamma, \beta^\top, \eta^\top)^\top \in \Theta\right\} = \mathscr{F}_1/\mathscr{F}_2$ is also Donsker.

Since $1/\mathscr{F}_2$ is Donsker, the class $\{z^\top/\mathscr{F}_2\}$ is also Donsker if $\mathrm{E}\,|Z|^4 < \infty$. The class $\mathscr{F}_2^2$ is Donsker because $\mathscr{F}_2$ satisfies the uniform entropy bound and $\mathrm{E}\,|Z|^4 < \infty$. Therefore, by (ii), also $z \cdot \mathscr{F}_1/\mathscr{F}_2^2$ is a Donsker class because the envelope $M^2 z \cdot F$ is square integrable. This implies the class

$$\partial(\mathscr{R}) = \left\{r'(\nu; Z) = \frac{\partial r(Z; \nu)}{\partial \nu},\ \nu \in \Theta\right\} = \left\{(1 + z^\top \eta)^{-1}(1, -z^\top, -z^\top r(z; \nu)),\ \nu \in \Theta\right\}$$

to be Donsker. Therefore Donsker classes are also $\Lambda(\mathscr{R})\partial\mathscr{R} = \{\Lambda(r(Z; \nu))\ \nu \in \Theta\}$ because $\Lambda(\mathscr{R})$ is Lipshitz continuous, and $\{\mathscr{D} - \Lambda(\mathscr{R})\}$ by the permanence of the Donsker property.

Finally, the product of the Donsker class $\partial\mathscr{R}$ with the bounded class $\{\mathscr{D}-\Lambda(\mathscr{R})\}$ gives the Donsker class $\Psi=\{\psi(\nu,y)=\big(1\{Y\leq y\}-\Lambda(r(Z;\nu))\big)\,r'(Z;\nu),\,\nu\in\Theta,y\in\mathbb{R}\}$.

The mapping $(\nu,y)\mapsto n^{-1/2}\sum_i(\psi_i(\nu,y)-\mathrm{E}\,\psi(\nu,y))$ is stochastically equicontinuous over $\mathbb{R}^{2k+2}$ with respect to the pseudometric $d\big((\nu,y),(\nu',y')\big)^2=\mathrm{E}\,\|(\psi(\nu,y)-\psi(\nu',y'))\|^2$. Further, $\sup_y\|\hat{\nu}(y)-\nu_D(y)\|=o_p(1)$ because of convergence with respect to the pseudometric

$$\sup_y d\big((\nu_D(y),y),(\hat{\nu}(y),y)\big)^2=o_p(1).$$

We conclude that

$$\mathbb{U}_n\big(\psi(\hat{\nu}(y),y))\big)=\mathbb{U}_n\big(\psi(\nu_D(y),y))\big)+o_p(1)\ \text{ in }\ l^\infty(\mathbb{R}),$$

where where $\mathbb{U}_n$ indicates the empirical process $\mathbb{U}_n(f)=\sqrt{n}\big(n^{-1}\sum_i f(X_i)-\mathrm{E}\,f(X)\big)$ and $l^\infty(\mathbb{R})$ is the set of uniformly bounded real functions on $\mathbb{R}$.

By a Taylor expansion of $\mathrm{E}[\psi(y,\nu)]_{\nu=\hat{\nu}(y)}$ around $\nu_D(y)$,

$$\begin{aligned}
&\mathrm{E}[\psi(y,\nu)]_{\nu=\hat{\nu}(y)}\\
&\quad=\mathrm{E}[\psi(\nu_D(y),y)]+\frac{\partial}{\partial\nu}\big\{\mathrm{E}[\psi(y,\nu)]_{\nu=\nu_D(y)}\big\}(\hat{\nu}(y)-\nu_D(y))+o_p(\|\hat{\nu}(y)-\nu_D(y)\|)\\
&\quad=\big[-V_D(y)^{-1}+o_p(1)\big](\hat{\nu}(y)-\nu_D(y)).
\end{aligned}$$

The first order condition for $\hat{\nu}(y)$ and the fact that $\mathrm{E}\,|Z|^{4+\varepsilon}<\infty$ imply that $n^{-1/2}\sum_i\Psi_i(\hat{\nu}(y),y)=o_p(1)$, so

$$\begin{aligned}
o_p(1)&=\big[V_D(y)^{-1}+o_p(1)\big]\sqrt{n}(\hat{\nu}(y)-\nu_D(y))^\top+\mathbb{U}_n(\psi(\hat{\nu}(y),y))\\
&=\big[V_D(y)^{-1}+o_p(1)\big]\sqrt{n}(\hat{\nu}(y)-\nu_D(y))^\top+\mathbb{U}_n(\psi(\nu_D(y),y))+o_p(1).
\end{aligned}$$

and the empirical process $\mathbb{U}_n(\psi(\nu_D(y),y))$ converges to a Gaussian process with mean zero and covariance function $\Sigma_D(y,y')$.

Asymptotic normality of the CDF estimator. A Taylor expansion of $\hat{F}(y\,|\,x)=\Lambda(r(z;\hat{\nu}(y)))$ around $\nu_D(y)$ gives

$$\hat{F}(y\,|\,x)-\Lambda(r(z;\nu_D(y)))=\Lambda'(r(z;\nu_D(y)))r'(z;\nu_D(y))^\top(\hat{\nu}(y)-\nu_D(y))+o_p(1),$$

so

$$\sqrt{n}\big(\hat{F}(y\,|\,x)-\Lambda(r(z;\nu_D(y)))\big)=\Lambda'(r(z;\nu_D(y)))R(y\,|\,z)^\top V_D(y)\,\mathbb{U}_n\big(-\psi(\nu_D(y),y)\big)+o_p(1).$$

that finally implies the result.