EIEF WORKING PAPER SERIES

EINAUDI INSTITUTE FOR ECONOMICS AND FINANCE

EIEF Working Paper 15/11

October 2015

Comparing Distribution and Quantile Regression

by

Samantha Leorato

(University of Rome "Tor Vergata")

Franco Peracchi

(Georgetown University, University of Rome "Tor Vergata" and EIEF)

Comparing Distribution and Quantile Regression*

Samantha Leorato
Department of Economics and Finance, University of Rome Tor Vergata

Franco Peracchi
Department of Economics, Georgetown University
Department of Economics and Finance, University of Rome Tor Vergata
Einaudi Institute for Economics and Finance

October 21, 2015

Abstract

We study the sampling properties of two alternative approaches to estimating the conditional distribution of a continuous outcome Y given a vector X of regressors. One approach – distribution regression – is based on direct estimation of the conditional distribution function; the other approach – quantile regression – 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 or, to guarantee monotonicity, their rearranged versions. We provide a systematic comparison of the asymptotic and finite sample performance of monotonic estimators obtained from the two approaches, considering both cases when the underlying linear-in-parameter models are correctly specified and several types of model misspecification of considerable practical relevance.

KEYWORDS: Distribution regression; quantile regression; linear location model; nonseparable models.

JEL CLASSIFICATIONS: C1, C21, C25.

* We thank the Editor and three anonymous referees for their comments. We also thank Jana Jurečková, Blaise Melly, Frank Vella, and especially Roger Koenker for useful discussions. Corresponding author: Samantha Leorato (samantha.leorato@uniroma2.it).

1 Introduction

In this paper we compare the sampling properties of two alternative approaches to estimating the conditional distribution of a continuous outcome Y given a vector X of regressors.

One approach – the distribution regression or DR approach – models parametrically the conditional distribution function (CDF) $F(y\,|\,x) = \Pr\{Y \leq y\,|X = x\}$ locally in $y \in \mathscr{Y}$, and obtains estimates $\hat{F}(\cdot\,|\,x)$ of the CDF by fitting a sequence of binary regression models at a finite number of cutoff values $y_1, \ldots, y_J$, each model corresponding to the conditional mean of the binary indicator $D_j = 1\{Y \leq y_j\}$. In fact, a convenient strategy is to model not $F(y\,|\,x)$ directly, but rather the conditional log-odds function (CLF) $t(y\,|\,x) = \ln F(y\,|\,x) - \ln\big(1 - F(y\,|\,x)\big)$. While this strategy leaves the range of t completely unrestricted, it guarantees that estimates of $F(y\,|\,x)$ obtained by inverting estimates of $t(y\,|\,x)$ are bounded between zero and one. An important special case is the linear-in-parameters specification $t(y\,|\,x) = P(x)^\top \boldsymbol{\theta}(y)$, where $P(x)$ is a vector of known transformations of the regressors, $\boldsymbol{\theta}(y)$ is a point in some finite-dimensional parameter space and all elements of $\boldsymbol{\theta}$ may vary with y. Although restrictive, this specification leads to models that are easy to estimate and to interpret. Further, any smooth CLF can be approximated arbitrarily well by one that is linear in parameters. This approach, first proposed by Foresi and Peracchi (1995), has recently been considered by Fortin, Lemieux and Firpo (2011), Rothe (2012), Chernozhukov, Fernández-Val and Melly (2013) and Hothorn, Kneib and Bühlmann (2014).

The other approach – the quantile regression or QR approach – models parametrically the conditional quantile function (CQF) $Q(p\,|\,x) = \inf\{y \in \mathscr{Y}\colon F(y\,|\,x) \geq p\}$ locally in $p \in (0,1)$, and obtains estimates of the CQF by fitting a sequence of asymmetric least absolute deviation regressions at a finite number of quantile levels $p_1, \ldots, p_J$. Here again, an important special case is the linear-in-parameter specification $Q(p\,|\,x) = P(x)^\top \boldsymbol{\gamma}(p)$, where $P(x)$ is a vector of known transformations of the regressors, $\boldsymbol{\gamma}(p)$ is a point in some finite-dimensional parameter space and all elements of $\boldsymbol{\gamma}$ may vary with p. This approach, first proposed by Koenker and Bassett (1978), has been generalized in many directions, including penalized likelihood methods, semi-parametric methods, methods for nonidentically distributed or dependent observations, extremal quantile regression, and weighted quantile regression (see Koenker 2005 for a review).

The CDF and the CQF are equivalent characterizations of the conditional distribution of Y given X, as they are generalized inverses of each other, that is, $Q(F(y\,|\,x)\,|\,x) \leq y$ and $F(Q(p\,|\,x)) \geq p$, which implies that $F(y\,|\,x) \geq p$ if and only if $Q(p\,|\,x) \geq y$. Thus, by analogy with this relationship, given a direct estimate $\hat{F}(\cdot\,|\,x)$ of the CDF, one may obtain an indirect estimate of the CQF by taking its generalized inverse $\tilde{Q}(p\,|\,x) = \inf\big\{y\colon \hat{F}(y\,|\,x) \geq p\big\}$. Similarly, given a direct estimate $\hat{Q}(\cdot\,|\,x)$ of the CQF, one may obtain an indirect estimate of the CDF by taking its generalized inverse $\tilde{F}(y\,|\,x) = \inf\{p\colon\, \hat{Q}(p\,|\,x) \geq y\}$.

In the unconditional case, where no parametric assumption is needed, the empirical distribution function and the empirical quantile function are generalized inverses of each other, so the DR and QR approaches are equivalent. When conditioning on a set of regressors, however, the DR and QR approaches are generally not equivalent because a convenient parametric model for the CQF, such as a linear-in-parameter specification, need not imply an equally convenient or easily interpretable model for the CDF, and viceversa.

Further, estimates based on linear-in-parameter specifications, while easy to obtain, need not be proper since they need not satisfy the key monotonicity properties of the CDF and the CQF, namely $F(y'\,|\,x) \geq F(y\,|\,x)$ whenever $y' > y$ and $Q(p'\,|\,x) \geq Q(p\,|\,x)$ whenever $p' > p$, for any x. Non-monotonicity of the estimates may just be a finite-sample problem or may reflect, more fundamentally, misspecification of the underlying parametric model for the CDF or the CQF. In any case, far from being a minor technical problem, the issue of how to guarantee monotonicity is of central importance, as indirect estimates can only be proper if they are obtained by inverting a proper estimate of the CDF or the CQF, and only in this case their asymptotic properties can be derived by means of the functional delta method.

Among the various ways of guaranteeing monotonicity proposed in the literature, in this paper we concentrate on the rearrangement procedure suggested by Chernozhukov, Fernández-Val and Galichon (2010), which is particularly attractive for its general nature and computational simplicity. We provide a systematic study of the asymptotic and finite sample performance of monotonic estimators obtained from the DR and QR approaches, considering both cases when the underlying linear-in-parameter models are correctly specified and several types of model misspecification of considerable practical relevance. We focus on bias, precision (as measured by the variance) and efficiency (as measured by the mean squared error) of the various estimators, both in finite samples and asymptotically. Of course, efficiency is only one of the many theoretical and practical criteria for comparing estimators, and other criteria may be taken into account, such as statistical robustness, computational ease, etc.

We assume throughout the paper that the available data $\{(X_i, Y_i)\}_{i=1}^n$ are a sample from the joint distribution of the random vector (X, Y) with support $\mathscr{X} \times \mathscr{Y}$, where $\mathscr{X} \subseteq \mathbb{R}^k$ and $\mathscr{Y} \subseteq \mathbb{R}$. This restrictive assumption helps simplify the presentation, but our results can easily be generalized to the case of heterogeneous or dependent observations. We also assume that the distribution of X has a finite nonsingular second moment matrix and that, for any $x \in \mathscr{X}$, the CDF of Y given $X = x$ is continuous and strictly increasing, which implies that the conditional density $f(y\,|\,x)$ of Y given $X = x$ exists and is finite and bounded away from zero. This in turn implies that the CDF and the CQF are inverses of each other, that is, $F(Q(p\,|\,x)\,|\,x) = p$ and $Q(F(y\,|\,x)\,|\,x) = y$. We further assume that our linear-in-parameter models for the CDF and the CQF always include an intercept, and to simplify notation we set $P(X) = (1, X^\top)^\top \equiv \boldsymbol{X}$ and $P(x) = (1, x^\top)^\top \equiv \boldsymbol{x}$. Finally, we denote by $l^\infty(\mathscr{S})$ the space of bounded and measurable real-valued functions defined on $\mathscr{S}$.

The remainder of the paper is organized as follows. Section 2 introduces the direct DR estimator of the CDF and the indirect estimator obtained by rearrangement. Section 3 introduces the direct QR estimator of the CQF and the indirect estimator obtained by rearrangement. Section 4 compares the asymptotic properties of estimators obtained under the DR and the QR approach, both when the linear-in-parameter models on which they are based are correctly specified and when they are not. Section 5 compares the finite sample properties of the various estimators considered via a set of Monte Carlo experiments. Finally, Section 6 summarizes and offers some conclusions.

2 DR estimators

Given a random sample $\{(X_i, Y_i)\}_{i=1}^n$ from the distribution of (X, Y) and a linear-in-parameter model $t(y\,|\,x) = \boldsymbol{x}^\top \boldsymbol{\theta}(y)$ for the CLF, an estimate $\hat{\boldsymbol{\theta}}_n(y)$ of $\boldsymbol{\theta}(y)$ may be obtained by maximizing over the parameter space the average pseudo log-likelihood

$$L_n(\boldsymbol{\theta}; y) = n^{-1} \sum_{i=1}^n \left[D_{yi} \boldsymbol{X}_i^\top \boldsymbol{\theta} - \ln\left(1 + \exp \boldsymbol{X}_i^\top \boldsymbol{\theta}\right)\right],$$

where $D_{yi} = \mathbb{1}\{Y_i \leq y\}$. Given $\hat{\boldsymbol{\theta}}_n(y)$, the direct DR estimate of the population CDF at the cutoff value y is $\hat{F}_n^\ddagger(y\,|\,x) = \Lambda\left(\boldsymbol{x}^\top \hat{\boldsymbol{\theta}}_n(y)\right)$, where $\Lambda(u) = e^u/(1 + e^u)$ is the standard logistic distribution function with density $\lambda(u) = \Lambda(u)(1 - \Lambda(u))$.

The population analog of $\hat{F}_n^\ddagger(y\,|\,x)$ is denoted by $F^\ddagger(y\,|\,x) = \Lambda\left(\boldsymbol{x}^\top \boldsymbol{\theta}(y)\right)$, where $\boldsymbol{\theta}(y)$ maximizes the expected pseudo log-likelihood $L(\boldsymbol{\theta}; y) = \mathbb{E}\left[D_y \boldsymbol{X}^\top \boldsymbol{\theta} - \ln\left(1 + \exp(\boldsymbol{X}^\top \boldsymbol{\theta})\right)\right]$ over the parameter space. If the assumed model for the CLF is correctly specified, then $F^\ddagger(y\,|\,x) = F(y\,|\,x)$ for almost all x and y values, so a consistent indirect estimator of the CQF may simply be obtained by taking the generalized inverse of $\hat{F}_n^\ddagger$. However, if the assumed model is misspecified, then $\hat{F}_n^\ddagger$ converges to a limit function $F^\ddagger$ that differs from F on a subset of $\mathscr{X} \times \mathscr{Y}$ with positive measure. This has two consequences. First, the direct estimator $\hat{F}_n^\ddagger$ is inconsistent for F, so the indirect estimator obtained by taking the generalized inverse of $\hat{F}_n^\ddagger$ is inconsistent for Q. Second, although bounded between zero and one, the limit function $F^\ddagger$ need not be a proper CDF because it need not be nondecreasing in y for all x. This implies that the generalized inverse of $F^{\ddagger}$ is not a continuous function, which prevents one from using the functional delta method to study the asymptotic properties of the CQF estimator obtained by inverting $\hat{F}_n^{\ddagger}$.

A simple way to guarantee monotonicity is the rearrangement procedure proposed by Chernozhukov, Fernández-Val and Galichon (2010).[1] Their procedure relies on the fact that, even when $\hat{F}_n^{\ddagger}$ is nonmonotonic, a proper estimate of the CDF is $\hat{F}_n^{+}(y \mid x) = \inf\{p : \hat{Q}_n^{+}(p \mid x) \geq p\}$, where

$$\hat{Q}_n^{+}(p \mid x) = \int_0^{\infty} \mathbb{1}\{\hat{F}_n^{\ddagger}(y \mid x) \leq p\} dy - \int_{-\infty}^{0} \mathbb{1}\{\hat{F}_n^{\ddagger}(y \mid x) > p\} dy \tag{1}$$

can be shown to be a proper estimate of the CQF (the proof is in Chernozhukov, Fernández-Val and Galichon 2007). Notice that while the DR approach does not directly estimate the CQF, rearrangement produces joint estimates of both the CDF and the CQF. If $\hat{F}_n^{\ddagger}$ is monotone, then $\hat{F}_n^{+}$ and $\hat{F}_n^{\ddagger}$ coincide. In general, $\hat{F}_n^{+}(y \mid x) = \hat{F}_n^{\ddagger}(y \mid x)$ at all points where $\hat{F}_n^{\ddagger}(y \mid x)$ is increasing in y and the equation $\hat{F}_n^{\ddagger}(y \mid x) = p$ has a unique solution. The same rearrangement procedures applied to the limit function $F^{\ddagger}$ gives both its rearranged version of F^{+} and its generalized inverse Q^{+}.

Rearrangement offers two main advantages. First, the rearranged estimator $\hat{F}_n^{+}$ is the continuous and Hadamard differentiable inverse of $\hat{Q}_n^{+}$, so its asymptotic properties can be derived via the functional delta method. Second, as shown by Chernozhukov, Fernández-Val and Galichon (2010) in their Monte Carlo experiments, $\hat{F}_n^{+}$ has a smaller bias than the original estimator $\hat{F}_n^{\ddagger}$.

Chernozhukov, Fernández-Val and Melly (2013) derive the asymptotic properties (as $n \to \infty$) of the stochastic processes $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n(y) - \boldsymbol{\theta}(y)\right)$, $\sqrt{n}\left(\hat{F}_n^{\ddagger}(y \mid x) - F^{\ddagger}(y \mid x)\right)$ and $\sqrt{n}\left(\hat{F}_n^{+}(y \mid x) - F^{+}(y \mid x)\right)$. Their results, summarized in Theorem 1 below for the case when the assumed model is linear in parameters, rely on the following two assumptions:

A.1: There exist $\underline{y} < \overline{y}$ in the interior of $\mathscr{Y}$ such that, for any $y \in [\underline{y}, \overline{y}]$, $\boldsymbol{\theta}(y)$ uniquely maximizes $L(\boldsymbol{\theta}; y)$ on a compact subset Θ of the parameter space.

A.2: For any $x \in \mathscr{X}$, the number of critical points $\{y : \partial_y F^{\ddagger}(y \mid x) = 0\}$ is finite.

Assumption A.1 and the assumption that X has finite nonsingular second moments guarantee that $L_n(\boldsymbol{\theta}; y)$ is twice differentiable in $\boldsymbol{\theta}$ and that its first and second partial derivatives have finite second moments. It also implies that the matrix $H(y) = \mathbb{E}\left[\lambda\left(\boldsymbol{X}^{\top}\boldsymbol{\theta}(y)\right)\boldsymbol{X}\boldsymbol{X}^{\top}\right]$ is finite and negative definitive for all $y \in [\underline{y}, \overline{y}]$, and that $\boldsymbol{\theta}(y)$ is continuously differentiable in y. In fact, applying the implicit function theorem to the system of equations $\partial L(\boldsymbol{\theta}; y)/\partial \boldsymbol{\theta} = 0$ shows that $\boldsymbol{\theta}(y)$ is continuous and differentiable

[1] Other procedures that guarantee monotonicity have been proposed by Foresi and Peracchi (1996), Hall, Wolff and Yao (1999), Hall and Müller (2003), and Dette and Volgushev (2008) among others.

in y, with derivative $\boldsymbol{\theta}'(y) = [H(y)]^{-1}\left(\mathbb{E} f(y\,|\,X)\boldsymbol{X}\right)$, so it is also continuously differentiable in y. As a consequence, $F^{\ddagger}(y\,|\,x)$ is continuously differentiable in both its arguments.

Assumption A.2 guarantees that, for all x, y and p, the equation $F^{\ddagger}(y\,|\,x) = p$, or equivalently the equation $\boldsymbol{x}^{\top}\boldsymbol{\theta}(y) = \ln[p/(1-p)]$, has a finite number $N(p\,|\,x)$ of roots which we denote by $y_j(p\,|\,x)$.

We also denote by $\mathscr{U}_x^* \subset (0,1)$ the set of regular values of the function $F^{\ddagger}(y\,|\,x)$, that is, the subset of the codomain of $F^{\ddagger}(\cdot\,|\,x)$ whose preimage does not contain critical points, and define $(0,1)\mathscr{X}^* = \{(p,x)\colon p \in \mathscr{U}_x^*, x \in \mathscr{X}\}$.

Theorem 1 *If A.1 holds, then the process $\hat{\boldsymbol{\theta}}_n(\cdot)$ is uniformly consistent for $\boldsymbol{\theta}(\cdot)$, that is, $\sup_{\underline{y}\le y\le \overline{y}} \|\hat{\boldsymbol{\theta}}_n(y) - \boldsymbol{\theta}(y)\| = o_p(1)$, and the process $H(\cdot)\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n(\cdot) - \boldsymbol{\theta}(\cdot)\right)$ converges weakly on $l^{\infty}([\underline{y},\overline{y}])$ to a zero-mean multivariate Gaussian process $B_D(\cdot)$ with covariance function*

$$\Sigma_D(y,y') = \mathbb{E}\left[\left(D_y - \Lambda(\boldsymbol{X}^{\top}\boldsymbol{\theta}(y))\right)\left(D_{y'} - \Lambda(\boldsymbol{X}^{\top}\boldsymbol{\theta}(y'))\right)\boldsymbol{X}\boldsymbol{X}^{\top}\right], \quad y \le y'.$$

In addition, for any compact subset $\mathscr{K} \subset [\underline{y},\overline{y}] \times \mathscr{X}$, the process $\sqrt{n}\left(\hat{F}_n^{\ddagger}(y\,|\,x) - F^{\ddagger}(y\,|\,x)\right)$, indexed by (y,x), converges weakly on $l^{\infty}(\mathscr{K})$ to a zero-mean Gaussian process W defined as

$$W(y\,|\,x) = \lambda\left(\boldsymbol{x}^{\top}\boldsymbol{\theta}(y)\right)x^{\top}H(y)^{-1}B_D(y). \tag{2}$$

If A.1–A.2 hold then, for any compact subset $\mathscr{K} \subset (0,1)\mathscr{X}^*$, the process $\sqrt{n}\left(\hat{Q}_n^{+}(p\,|\,x) - Q^{+}(p\,|\,x)\right)$, indexed by (p,x), converges weakly on $l^{\infty}(\mathscr{K})$ to a zero-mean Gaussian process C_W defined as

$$C_W(p\,|\,x) = -\sum_{j=1}^{N(p\,|\,x)} \frac{W(y_j(p\,|\,x)\,|\,x)}{\left|\partial_y F^{\ddagger}(y_j(p\,|\,x)\,|\,x)\right|}.$$

Finally, letting $\mathscr{K}^* = \left\{(y,x) \in [\underline{y},\overline{y}] \times \mathscr{X}\colon \left(F^{+}(y\,|\,x), x\right) \in \mathscr{K}\right\}$, the process $\sqrt{n}\left(\hat{F}_n^{+}(y\,|\,x) - F^{+}(y\,|\,x)\right)$, indexed by (y,x), converges weakly on $l^{\infty}(\mathscr{K}^*)$ to a zero-mean Gaussian process D_W defined as

$$D_W(y\,|\,x) = -\left(\sum_{j=1}^{N\left(F^{\ddagger}(y|x)\,|\,x\right)} \frac{1}{\left|\partial_y F^{\ddagger}\left(y_j(F^{+}(y\,|\,x)\,|\,x)\,\middle|\,x\right)\right|}\right)^{-1} C_W\left(F^{+}(y\,|\,x)\,\middle|\,x\right).$$

The function $C_W(p\,|\,x)$ in Theorem 1 is the Hadamard differential of Q^{+} at W tangentially to the space of continuous functions defined on $\mathscr{Y}\mathscr{X} = \left\{(y,x)\colon \left(F^{+}(y\,|\,x), x\right) \in (0,1)\mathscr{X}^*\right\}$ (see e.g. van der Waart 1998). If $F^{\ddagger}$ is strictly increasing in y then the equation $F^{\ddagger}(y\,|\,x) = p$ has a unique root and $F^{+}(y\,|\,x) = F^{\ddagger}(y\,|\,x)$ for all x, y and p, so $C_W(p\,|\,x) = -W\left(Q^{+}(p\,|\,x)\,\middle|\,x\right)/\partial_y F^{\ddagger}\left(Q^{+}(p\,|\,x)\,\middle|\,x\right)$ and $D_W(y\,|\,x) = W(y\,|\,x)$.

It follows from Theorem 1 that the asymptotic variance of $\hat{F}_n^{\ddagger}(p\,|\,x)$ is equal to $\mathbb{V}\left(\hat{F}_n^{\ddagger}(p\,|\,x)\right) = \lambda\left(\boldsymbol{x}^\top\boldsymbol{\theta}(y)\right)^2 \boldsymbol{x}^\top V_D(y)\boldsymbol{x}$, where $V_D(y) = H(y)^{-1}\Sigma_D(y,y)H(y)^{-1}$ denotes the asymptotic variance of $\hat{\boldsymbol{\theta}}_n(y)$. If the assumed linear-in-parameter model for the CLF is correctly specified, then $\Sigma_D(y,y) = H(y)$ and $V_D(y) = H(y)^{-1}$, so the asymptotic variance of $\hat{F}_n^{\ddagger}(y\,|\,x)$ simplifies to $\mathbb{V}\left(\hat{F}_n^{\ddagger}(y\,|\,x)\right) = \lambda\left(\boldsymbol{x}^\top\boldsymbol{\theta}(y)\right)^2 \boldsymbol{x}^\top H(y)$

3 QR estimators

An alternative to directly estimate the CDF is to first estimate the CQF and then obtain indirect estimates of the CDF by taking the generalized inverse or by rearrangement. To fix the ideas, consider the linear location model

$$Y = \alpha + X^\top\beta + U, \tag{3}$$

where U is a random error distributed independently of X with a strictly increasing distribution function G. Its CQF is $Q(p\,|\,x) = \boldsymbol{x}^\top\boldsymbol{\gamma}(p)$, where $\boldsymbol{\gamma}(p) = \left(\alpha + G^{-1}(p), \beta^\top\right)^\top$. This model has the restrictive feature that $Q(p\,|\,x) - Q(p'\,|\,x) = G^{-1}(p) - G^{-1}(p')$, that is, conditional quantiles corresponding to different values of p are at a constant distance from each other. A straightforward generalization retains linearity in the parameters but allows all elements of $\boldsymbol{\gamma}(p)$ to depend on p, leading to the linear-in-parameter specification $Q(p\,|\,x) = \boldsymbol{x}^\top\boldsymbol{\gamma}(p)$, where $\boldsymbol{\gamma}(p)$ is a point in some finite-dimensional parameter space and all elements of $\boldsymbol{\gamma}$ may vary with p.

Given a random sample $\{(X_i, Y_i)\}_{i=1}^n$ from the distribution of (X, Y) and a linear-in-parameter model for the CQF, an estimate $\hat{\boldsymbol{\gamma}}_n(p)$ of $\boldsymbol{\gamma}(p)$ may be obtained by minimizing over the parameter space the objective function

$$\ell_n(\boldsymbol{\gamma}; p) = n^{-1}\sum_{i=1}^n \rho_p\left(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\gamma}\right),$$

where $\rho_p(u) = u\,[p - \mathbb{1}\{u \le 0\}]$ is the asymmetric absolute loss function. Given $\hat{\boldsymbol{\gamma}}_n(p)$, the direct QR estimate of the population CQF at the quantile level p is $\hat{Q}_n^*(p\,|\,x) = \boldsymbol{x}^\top\hat{\boldsymbol{\gamma}}_n(p)$.

The population analog of $\hat{Q}_n^*(p\,|\,x)$ is denoted by $Q^*(p\,|\,x) = \boldsymbol{x}^\top\boldsymbol{\gamma}(p)$, where $\boldsymbol{\gamma}(p)$ is the minimizer over the parameter space of $\ell(\boldsymbol{\gamma}; p) = \mathbb{E}\,\rho_p\left(Y - \boldsymbol{X}^\top\boldsymbol{\gamma}\right)$, the population analog of $\ell_n(\boldsymbol{\gamma}; p)$. If the assumed model for the CQF is correctly specified, then $Q^*(y\,|\,x) = Q(p\,|\,x)$ for almost all x and p values, so a consistent indirect estimator of the CDF may simply be obtained by taking the generalized inverse of $\hat{Q}_n^*$. However, if the assumed model is misspecified, then $\hat{Q}_n^*$ converges to a limit function Q^* that differs from Q on a subset of $\mathscr{Y} \times \mathscr{X}$ with positive measure. This poses the same problems discussed in Section 2. As suggested by Chernozhukov, Fernández-Val and Galichon (2010), a possible solution is again rearrangement. When $\hat{Q}_n^*$ is nonmonotonic, a proper estimate of the CQF is $\hat{Q}_n^\circ(p\,|\,x) = \inf\{y\colon \hat{F}_n^\circ(y\,|\,x) \ge p\}$, where

$$\hat{F}_n^\circ(y \mid x) = \int_0^1 \mathbb{1}\{\hat{Q}_n^*(p \mid x) \leq y\} dp \tag{4}$$

is a proper estimate of the CDF. The same rearrangement procedures applied to the limit function Q^* gives both its rearranged version Q° and its generalized inverse F°. Notice that Q^* and Q° coincide if Q^* is monotone. Moreover, $Q^\circ(p \mid x) = Q^*(p \mid x)$ provided that $Q^*(p \mid x)$ is increasing at p and $Q^*(p \mid x) = p$ has a unique solution for $y = Q^\circ(p \mid x)$.

Chernozhukov, Fernández-Val and Melly (2013)) also provide the QR counterpart of Theorem 1. Their results, summarized in Theorem 2 below, rely on the following two assumptions:

B.1: There exist $\underline{p} < \overline{p}$ in the interior of $(0,1)$ such that, for any $p \in [\underline{p}, \overline{p}]$, $\boldsymbol{\gamma}(p)$ uniquely minimizes $\ell(\boldsymbol{\gamma}; p)$ on a compact subset Γ of the parameter space.

B.2: For any $x \in \mathscr{X}$, the number of critical points $\{p : \partial_p Q^*(p \mid x) = 0\}$ is finite.

Assumptions B.1–B. 2 play the same role as Assumptions A.1–A.2 in Section 2. In particular, Assumption B.1 implies that the matrix $J(p) = \mathbb{E}\left[f\left(\boldsymbol{X}^\top \boldsymbol{\gamma}(p) \mid X\right) \boldsymbol{X}\boldsymbol{X}^\top\right]$ is finite and positive definite for all p in the closed interval $[\underline{p}, \overline{p}]$, and that the function $\boldsymbol{\gamma}(p)$ is continuously differentiable on $[\underline{p}, \overline{p}]$ with derivative $\boldsymbol{\gamma}'(p)$, while Assumption B.2 guarantees that, for all x, y and p, the equation $Q^*(p \mid x) = y$, or equivalently the equation $\boldsymbol{x}^\top \boldsymbol{\gamma}(p) = y$, has a finite number $N(y \mid x)$ of roots which we denote by $p_j(y \mid x)$. We also denote by $\mathscr{Y}_x^*$ the subset of the codomain of $Q^*(\cdot \mid x)$ whose preimage does not contain critical points. Thus, $\partial_p Q^*(p \mid x) \neq 0$ for all p such that $Q^*(p \mid x) \in \mathscr{Y}_x^*$.

Theorem 2 *If B.1 holds, then the process* $\hat{\boldsymbol{\gamma}}_n(\cdot)$ *is uniformly consistent for* $\boldsymbol{\gamma}(\cdot)$*, that is,* $\sup_{\underline{p} \leq p \leq \overline{p}} \|\hat{\boldsymbol{\gamma}}_n(p) - \boldsymbol{\gamma}(p)\| = o_p(1)$*, and the process* $J(\cdot)\sqrt{n}\left(\hat{\boldsymbol{\gamma}}_n(\cdot) - \boldsymbol{\gamma}(\cdot)\right)$ *converges weakly on* $l^\infty([\underline{p}, \overline{p}])$ *to a zero-mean multivariate Gaussian process* $B_Q(\cdot)$ *with covariance function*

$$\Sigma_Q(p, p') = \mathbb{E}\left[\left(p - \mathbb{1}\{Y < \boldsymbol{X}^\top \boldsymbol{\gamma}(p)\}\right)\left(p' - \mathbb{1}\{Y < \boldsymbol{X}^\top \boldsymbol{\gamma}(p')\}\right) \boldsymbol{X}\boldsymbol{X}^\top\right], \quad p \leq p'.$$

In addition, for any compact subset $\mathscr{H} \subset [\underline{p}, \overline{p}] \times \mathscr{X}$*, the process* $\sqrt{n}\left(\hat{Q}_n^*(p \mid x) - Q^*(p \mid x)\right)$*, indexed by* (p, x)*, converges weakly on* $l^\infty(\mathscr{H})$ *to the zero-mean Gaussian process* Z *defined as*

$$Z(p \mid x) = \boldsymbol{x}^\top J(p)^{-1} B_Q(p).$$

If B.1–B.2 hold then, for any compact subset $\mathscr{K} \subset (0,1)\mathscr{X}^*$*, the process* $\sqrt{n}\left(\hat{Q}_n^+(p \mid x) - Q^+(p \mid x)\right)$*, indexed by* (p, x)*, converges weakly on* $l^\infty(\mathscr{K})$ *to a zero-mean Gaussian process* C_W *defined as*

$$C_Z(y \mid x) = -\sum_{j=1}^{N(y \mid x)} \frac{Z\left(p_j(y \mid x) \mid x\right)}{\left|\partial_p Q^*\left(p_j(y \mid x) \mid x\right)\right|}.$$

Finally, letting $\mathscr{K}^* = \left\{(y,x) \in [\underline{y}, \overline{y}] \times \mathscr{X} : \left(F^+(y \mid x), x\right) \in \mathscr{K}\right\}$*, the process* $\sqrt{n}\left(\hat{F}_n^+(y \mid x) - F^+(y \mid x)\right)$*, indexed by* (y,x)*, converges weakly on* $l^\infty(\mathscr{K}^*)$ *to a zero-mean Gaussian process* D_W *defined as*

$$D_Z(p \mid x) = \left(\sum_{j=1}^{N(y \mid x)} \frac{1}{\left|\partial_p Q^*\left(p_j(y \mid x) \mid x\right)\right|} \right)^{-1} C_Z(y \mid x) \Bigg|_{y = Q^\circ(p \mid x)} .$$

The function $C_Z(p \mid x)$ in Theorem 2 is the Hadamard differential of F° at Z tangentially to the space of continuous functions defined on $(0,1)\mathscr{X}$. If Q^* is strictly increasing in p then the equation $Q^*(p \mid x) = y$ has a unique root and $Q^\circ(p \mid x) = Q^*(p \mid x)$ for all x, y and p, so $C_Z(y \mid x) = -Z\left(p(y \mid x)\right)/\partial_p Q^*\left(p(y \mid x) \mid x\right)$ and $D_Z(p \mid x) = Z(p \mid x)$.

It follows from Theorem 2 that the asymptotic variance of $\hat{Q}_n^*(p \mid x)$ is $\mathbb{V}\left(\hat{Q}_n^*(p \mid x)\right) = \boldsymbol{x}^\top V_Q(p) \boldsymbol{x}$, where $V_Q(p) = J(p)^{-1} \Sigma_Q(p,p) J(p)^{-1}$ denotes the asymptotic variance of $\boldsymbol{\gamma}(p)$. If the assumed linear-in-parameter model is correctly specified, then $J(p) = \mathbb{E}\left[f\left(\boldsymbol{X}^\top \boldsymbol{\gamma}(p) | X\right) \boldsymbol{X}\boldsymbol{X}^\top\right]$. In particular, under the linear location model (3), $\Sigma_Q(p,p) = p(1-p) P_X$ and $J(p) = g_p P_X$, with $g_p = g(G^{-1}(p))$ and $g = G'$, so the asymptotic variance of $\boldsymbol{\gamma}(p)$ simplifies to $V_Q(p) = [p(1-p)/g_p^2] P_X^{-1}$.

4 Asymptotic relationships

In this section we compare the asymptotic properties of estimators obtained under the two approaches, both when the assumed linear-in-parameter models for the CLF and the CQF are correctly specified and when they are not. Figure 1 summarizes the relationships between the various estimators considered and their population counterparts.

4.1 Correct specification

If the assumed linear-in-parameter models for the CDF and the CQF are both correctly specified, which is essentially equivalent to assuming that the data satisfy the linear location model (3) with logistic errors, then Theorems 1 and 2 imply that, for any x and all $y \in [\underline{y}, \overline{y}]$ such that $\underline{p} \le F(y \mid x) \le \overline{p}$, the asymptotic variances of all estimators considered are linked by the following relationships

$$\mathbb{V}\left(\hat{F}_n^\ddagger(y \mid x)\right) = \mathbb{V}\left(\hat{F}_n^+(y \mid x)\right) = f(y \mid x)^2\, \mathbb{V}\left(\hat{Q}_n^+(F(y \mid x) \mid x)\right)$$

and

$$\mathbb{V}\left(\hat{F}_n^\circ(y \mid x)\right) = f(y \mid x)^2\, \mathbb{V}\left(\hat{Q}_n^*(F(y \mid x) \mid x)\right) = f(y \mid x)^2\, \mathbb{V}\left(\hat{Q}_n^\circ(F(y \mid x) \mid x)\right).$$

This set of results implies that, for any x and all $p \in [\underline{p}, \overline{p}]$ and $y \in [\underline{y}, \overline{y}]$ such that $F(y \mid x) = p$ and $Q(p \mid x) = y$, we have

$$ARE\left(\hat{F}_n^{\circ}(y \mid x), \hat{F}_n^{+}(y \mid x)\right) = ARE\left(\hat{Q}_n^{\circ}(p \mid x), \hat{Q}_n^{+}(p \mid x)\right).$$

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 mean squared error (MSE) of the direct kernel estimator $\hat{F}$ of a 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.

4.2 Misspecification

If the assumed linear-in-parameter model for the CLF is misspecified, then the DR approach leads to inconsistent estimates, as $F^{\ddagger}$ no longer coincides with the true CDF. The same is true for the QR approach if the assumed linear-in-parameter model for the conditional CQF is misspecified. In these cases, asymptotic comparison of the various estimators may be based on their MSE, which is asymptotically dominated by bias.

In this section we focus on a fairly general type of misspecification, namely the case when the assumed models for the CLF and the CQF are linear in parameters but the sample observations are generated from the nonseparable data generating process (DGP)

$$Y = \alpha + X^{\top}\beta + \psi_{\delta}(X, U), \tag{5}$$

where U is distributed independently of X as standard logistic, δ is a scalar parameter, and $\psi_{\delta}(X, U)$ is a term that captures the particular way in which the logistic linear location model may be misspecified. We assume that the function $\psi_{\delta}(x, u)$ varies smoothly with δ for all x and u, and that $\psi_{\delta}(x, u) = u$ only when $\delta = 0$. Thus, when $\delta = 0$, model (5) reduces to the logistic linear location model, in which case our linear-in-parameter models for the CLF and the CQF are both correct.[2]

The DGP (5) is quite general and encompasses several important types of departure from the logistic linear location model. Notice that the degree to which the logistic linear location model is misspecified

[2] Notice that we are assuming that the DGP is the same for all observations. An alternative is to allow a fraction δ of the observations to deviate from the assumed model. Asymptotic results for this case are presented in Leorato and Peracchi (2015).

only depends on the value of the scalar parameter δ, which we will term the "degree of misspecification". When $\delta = 0$ there is no misspecification, but the precise meaning of δ depends on the type of misspecification considered.

Under the assumption that the function $\psi_\delta(x,u)$ is strictly increasing in u for all x and δ, with inverse function $\varphi_\delta(x,u)$, the CDF implied by (5) is

$$F_\delta(y \mid x) = \Lambda\left(\varphi_\delta(x, y - \alpha - x^\top \beta)\right),$$

while its CQF is

$$Q_\delta(p \mid x) = \alpha + x^\top \beta + \psi_\delta\left(x, \Lambda^{-1}(p)\right),$$

where $\Lambda^{-1}(p) = \ln[p/(1-p)]$. Putting $\delta = 0$ gives the CDF and the CQF of the logistic linear location model, namely $F_0(y \mid x) = \Lambda\left(y - \alpha - x^\top \beta\right)$ and $Q_0(p \mid x) = \alpha + x^\top \beta + \Lambda^{-1}(p)$.

The remainder of this section discusses in more detail four types of departure from the logistic linear location model that are of considerable practical relevance and represent the main focus of the Monte Carlo study described in Section 5.

(i) Omitted variables: The GDP takes the following form

$$Y = \alpha + X^\top \beta + \delta \phi(X) + U,$$

so $\psi_\delta(x,u) = \delta\phi(x) + u$ and $\varphi_\delta(x,u) = u - \delta\phi(x)$. By suitable defining X and the function $\phi(x)$, this formulation includes both the case of omitted variables and the case of nonlinearity of the conditional mean of Y.

(ii) Heteroskedasticity: The GDP takes the following form

$$Y = \alpha + X^\top \beta + \sigma_\delta(X)\, U,$$

where $\sigma_\delta(x) = 1 + \delta\phi(x)$ is a positive scale function, so $\psi_\delta(x,u) = \sigma_\delta(x)\, u$ and $\varphi_\delta(x,u) = u/\sigma_\delta(x)$. By suitable defining X and the function $\phi(x)$, this formulation also includes the case when the conditional mean and the conditional variance of Y depend on different sets of regressors.

(iii) Nonlogistic models: The GDP takes the following form

$$Y = \alpha + X^\top \beta + G_\delta^{-1}\left(\Lambda(U)\right),$$

where G_δ is a strictly increasing distribution function such that $G_0 = \Lambda$, so $\psi_\delta(x,u) = G_\delta^{-1}\left(\Lambda(u)\right)$ and $\varphi_\delta(x,u) = \Lambda^{-1}\left(G_\delta(u)\right)$, which depends only on u, not on x.

(iv) Transformation models: The GDP takes the following form

$$\phi_\delta(Y) = \alpha + X^\top \beta + U,$$

where $\phi_0(y) = y$ and $\phi_\delta(y)$ is strictly monotone in y for every δ with inverse function ϕ_δ^{-1}. In this case $\psi_\delta(x,u) = \phi_\delta^{-1}(\alpha + x^\top \beta + u) - \alpha - x^\top \beta$ and $\varphi_\delta(x,u) = \phi_\delta(\alpha + x^\top \beta + u) - \alpha - x^\top \beta$. A leading example is when Y is a nonnegative random variable and $\phi_\delta(Y) = Y^{(1-\delta)} + 1$, where $Y^{(1-\delta)}$ is the Box-Cox transform of Y, that is, $Y^{(1-\delta)} = (Y^{1-\delta} - 1)/(1-\delta)$ if $\delta \neq 1$ and $Y^{(1-\delta)} = \ln Y$ if $\delta = 1$.[3]

If we increase the sample size keeping fixed the degree of misspecification δ, then we eventually reach a situation where the bias completely dominates the MSE. The usual approach in order to strike a balance between asymptotic precision and bias is to allow the standard error of estimation and the degree of misspecification to vanish asymptotically at the same rate. Following this approach, Leorato and Peracchi (2015) derive the bias of estimators based on linear models for the CLF or the CQF when the logistic linear location model is locally misspecified, that is, the DGP is of the form (5) but $\delta = c/n^{-1/2}$ for some constant c. Their results require two conditions: the function $\psi_\delta(x,u)$ must be strictly increasing and differentiable in u for all x and δ, and the function

$$\Psi(x,u) = \lim_{\delta \to 0} \frac{\psi_\delta(x,u) - u}{\delta}$$

must exist and be square integrable in x, uniformly in u. Since all types of misspecification considered in this paper satisfy these two conditions, we can use their results to compute the local asymptotic bias of all our estimators, namely their asymptotic bias under local misspecification.

When $\delta = c/\sqrt{n}$, it follows from Corollary 1 in Leorato and Peracchi (2015) that, for any $x \in \mathscr{X}$ and all $y \in [\underline{y}, \overline{y}]$, the direct DR estimator $\hat{F}_n^\ddagger$ of the CDF has the same local asymptotic bias as the rearranged DR estimator $\hat{F}_n^+$, namely

$$\mathbb{B}\left(\hat{F}_n^+(y \mid x)\right) = c\, \lambda_y(x) \Big[\Psi_y(x) - \boldsymbol{x}^\top \left(\mathbb{E}\, \lambda_y(X) \boldsymbol{X}\boldsymbol{X}^\top \right)^{-1} \left(\mathbb{E}\, \lambda_y(X) \Psi_y(X) \boldsymbol{X} \right) \Big], \tag{6}$$

where $\lambda_y(x) = \lambda(y - \alpha - x^\top \beta)$, $\Psi_y(x) = \Psi(x, y - \alpha - x^\top \beta)$, $\lambda_y(X) = \lambda(y - \alpha - X^\top \beta)$ and $\Psi_y(X) = \Psi(X, y - \alpha - X^\top \beta)$. Similarly, for any $x \in \mathscr{X}$ and all $p \in [\underline{p}, \overline{p}]$, the local asymptotic bias of the rearranged DR estimator $\hat{Q}_n^+$ of the CQF is

$$\mathbb{B}\left(\hat{Q}_n^+(p \mid x)\right) = -c \Big[\Psi_{y(p|x)}(x) - \boldsymbol{x}^\top \left(\mathbb{E}\, \lambda_{y(p|x)}(X) \boldsymbol{X}\boldsymbol{X}^\top \right)^{-1} \left(\mathbb{E}\, \lambda_{y(p|x)}(X) \Psi_{y(p|x)}(X) \boldsymbol{X} \right) \Big], \tag{7}$$

[3] This particular specification of the Box-Cox transformation model guarantees that (3) holds when $\delta = 0$.

where $y(p\,|\,x) = \alpha + \Lambda^{-1}(p) + x^\top\beta$, $\Psi_{y(p|x)}(x) = \Psi(x, \Lambda^{-1}(p))$, $\lambda_{y(p|x)}(X) = \lambda(\Lambda^{-1}(p) + (x - X)^\top\beta)$, and $\Psi_{y(p|x)}(X) = \Psi(X, \Lambda^{-1}(p) + (x - X)^\top\beta)$. Further, for any $x \in \mathscr{X}$ and all $p \in [\underline{p}, \overline{p}]$, the direct QR estimator $\hat{Q}_n^*$ of the CQF has the same local asymptotic bias as the rearranged QR estimator $\hat{Q}_n^\circ$, namely

$$\mathbb{B}\left(\hat{Q}_n^\circ(p\,|\,x)\right) = -c\left[\Psi_{y(p|x)}(x) - \boldsymbol{x}^\top(\mathbb{E}\,\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\left(\Psi_{y(p|X)}(X)\boldsymbol{X}\right)\right], \tag{8}$$

where $\Psi_{y(p|X)}(X) = \Psi(X, \Lambda^{-1}(p))$. Finally, for any $x \in \mathscr{X}$ and all $y \in [\underline{y}, \overline{y}]$, the local asymptotic bias of the rearranged QR estimator $\hat{F}_n^\circ$ of the CDF is

$$\mathbb{B}\left(\hat{F}_n^\circ(y\,|\,x)\right) = c\,\lambda_y(x)\left[\Psi_y(x) - \boldsymbol{x}^\top(\mathbb{E}\,\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\left(\mathbb{E}\,\Psi(X, y - \alpha - x^\top\beta)\boldsymbol{X}\right)\right]. \tag{9}$$

Given these results and Theorems 1–2, we compute the local asymptotic MSE of each estimator as the sum of its asymptotic variance and its local asymptotic squared bias.

Notice that the term in square brackets in (8) is the error in approximating $\Psi_{y(p|x)}(x)$ using the linear least-squares projection of $\Psi_{y(p|X)}(X)$ on $\boldsymbol{X}$, while the corresponding term in (7) is the error in approximating $\Psi_{y(p|x)}(x)$ using the weighted linear least-squares projection of $\Psi_{y(p|x)}(X)$ on $\boldsymbol{X}$ with weights equal to $\lambda_{y(p\,|\,x)}(X)$. The term in square brackets in (9) is instead the error in approximating $\Psi_y(x)$ using the linear least-squares projection of $\Psi(X, y - \alpha - x^\top\beta)$ on $\boldsymbol{X}$, while the corresponding term in (6) is the error in approximating $\Psi_y(x)$ using the weighted linear least-squares projection of $\Psi_y(X)$ on $\boldsymbol{X}$ with weights equal to $\lambda_y(X)$.

As shown in the Appendix, the local asymptotic bias of the various estimators depends on the properties of the function $\Psi(x, u)$. In the case of omitted variables, where Ψ depends only on x, all estimators are generally inconsistent, but the asymptotic bias of $\hat{Q}_n^\circ$ does not depend on p. In the case of heteroskedasticity, again all estimators are generally inconsistent, but the asymptotic bias of $\hat{Q}_n^\circ$ is proportional to $\Lambda^{-1}(p)$, so it vanishes when $p = 1/2$. In the case of nonlogistic models, Ψ depends only on u, say $\Psi(x, u) = \zeta(u)$ for some function ζ, so the asymptotic biases of $\hat{F}_n^\circ$ and $\hat{Q}_n^\circ$ are both proportional to $\boldsymbol{x}^\top P_X^{-1}(\mathbb{E}\,\boldsymbol{X}) - 1$. Since $\boldsymbol{x}^\top P_X^{-1}(\mathbb{E}\,\boldsymbol{X}) = 1$ for any $\boldsymbol{x}$, the QR estimators have no asymptotic bias. The situation is just the opposite for transformation models. In this case $\Psi(x, u) = \chi(\alpha + x^\top\beta + u)$ for some function χ, so the bias of $\hat{F}_n^+$ is proportional to $\boldsymbol{x}^\top H(y)^{-1}\,\mathbb{E}\left[\lambda_y(X)\boldsymbol{X}\right] - 1$, where $H(y) = \mathbb{E}\left[\lambda_y(X)\boldsymbol{X}\boldsymbol{X}^\top\right]$. Since $\boldsymbol{x}^\top H(y)^{-1}\,\mathbb{E}[\lambda_y(X)\boldsymbol{X}] = 1$ for any $\boldsymbol{x}$ and y, $\hat{F}_n^+$ has no asymptotic bias. A similar argument holds for $\hat{Q}_n^+$. Thus, the DR estimators have no asymptotic bias.

By combining two or more sources of misspecification, the four types of departures from the logistic linear location model considered in this section can be used to construct more complex scenarios.[4]

[4] For example, in the case of a Box-Cox transformation model, a logistic error distribution is generally misspecified because it is unbounded from below: the combination in this case of the two forms of misspecification (iii) and (iv) produces a function of the form $\Phi(x, u) = \zeta(u)\,\chi(\alpha + x^\top\beta + u)$, and both approaches lead to inconsistent estimates.

5 Finite-sample properties

In this section we present the results of a set of Monte Carlo experiments aimed at comparing the finite-sample properties of DR and QR estimators based on simple linear-in-parameter models for the CLF and the CQF.

5.1 Monte Carlo design

We compute DR and QR estimators from samples of size n generated from a DGP of the form (5), where X is a single regressor uniformly distributed on the interval $(0,1)$, U has a standard logistic distribution, the function $\psi_\delta(x,u)$ varies smoothly with δ for all x and u, is strictly increasing in u for all x and δ, and is such that $\psi_\delta(x,u) = u$ for all x and u only when $\delta = 0$. In our benchmark case we set $\delta = 0$, which implies that the DGP is a linear location model with logistic errors, so a linear specification is correct for both the CLF and the CQF.[5]

We also consider the four types of departure from this benchmark discussed in Section 4.2. The first is the omitted variables case, where

$$F_\delta(y \mid x) = \Lambda\big(y - \alpha - \beta x - \delta\phi(x)\big),$$
$$Q_\delta(p \mid x) = \alpha + \Lambda^{-1}(p) + \beta x + \delta\phi(x).$$

We set $\phi(x) = x^2$, so the true CLF and CQF are both quadratic in x. In this case, estimators based on linear specifications are always inconsistent.

The second is the case of heteroskedasticity, where

$$F_\delta(y \mid x) = \Lambda\left(\frac{y - \alpha - \beta x}{\sigma_\delta(x)}\right),$$
$$Q_\delta(p \mid x) = \alpha + \beta x + \sigma_\delta(x)\,\Lambda^{-1}(p).$$

We set $\sigma_\delta(x) = 1 + \delta x^2$, so estimators based on a linear specification of the CLF are inconsistent. Estimators based on a linear specification of the CQF are also inconsistent, except when $p = .50$.[6]

The third case is when the distribution function of the error in the linear location model (3) is a convex combination $G_\delta = (1-\delta)\Lambda + \delta G$ of Λ and another strictly increasing distribution function G, with mixing probability $0 \le \delta < 1$. In this case

$$F_\delta(y \mid x) = G_\delta\big(y - \alpha - \beta x\big),$$
$$Q_\delta(p \mid x) = \alpha + \beta x + G_\delta^{-1}(p),$$

[5] In this case, the parameter β is linked to the population regression R^2 through the relationship $\beta^2\sigma_X^2/\sigma_U^2 = R^2/(1-R^2)$.
[6] This is because $\Lambda^{-1}(.50) = 0$ by symmetry of the standard logistic distribution, so $Q_\delta(.50 \mid x) = \alpha + \beta x$ for any δ.

so estimators based on a linear specification of the CQF are consistent, while those based on a linear specification of the CLF are not. We consider various values of δ and different choices of G, such as the Student t with 3 degrees of freedom, which is symmetric about zero but has no moments beyond the second, and the standard Gumbel (Type 1 extreme value), which is not symmetric about zero but has moments of all order.

The fourth is the case when the logistic linear location model holds after applying a Box-Cox transform, that is, the DGP is $Y^{(1-\delta)} + 1 = \alpha + \beta X + U$, with U distributed as standard logistic. In this case,

$$F_\delta(y \,|\, x) = \Lambda\big(y^{(1-\delta)} + 1 - \alpha - \beta x\big)$$

and

$$Q_\delta(p \,|\, x) = \begin{cases} \big[(1-\delta)\big(\alpha + \beta x + \Lambda^{-1}(p)\big) + \delta\big]^{1/(1-\delta)}, & \text{if } \delta \neq 1, \\ \exp\big(\alpha + \beta x + \Lambda^{-1}(p) - 1\big), & \text{if } \delta = 1. \end{cases}$$

Unlike the previous case, now estimators based on a linear specification of the CLF are consistent, while those based on a linear specification of the CQF are not.

For each Monte Carlo experiment, we generate 1000 samples of size n, with $n = 900$, 1600 and 3600 (so $\sqrt{n} = 30$, 40 and 60), and estimate the CDF at a grid $\{y_j \in \mathbb{R}, j = 1, \ldots, J\}$ of cutoff values and the CQF at a grid $\{p_j \in (0,1), j = 1, \ldots, J\}$ of quantile levels. For the p-grid we take $J = 199$ equally spaced quantile levels, while for the y-grid we take the empirical marginal quantiles of Y at level p_j, $j = 1, \ldots, J$.[7] As for the values of x, we consider a grid of 999 equally spaced values ranging from .001 to .999. Estimates of $F(y \,|\, x)$ and $Q(p \,|\, x)$ at points not in the p-, x- or y-grids, needed to compute the generalized inverse, are obtained by linear interpolation.

5.2 Monte Carlo results

This section summarizes the results of our set of Monte Carlo experiments, first for the benchmark linear location model (3) with logistic errors, and then for the four types of deviation from this benchmark discussed in Section 4.2, namely heteroskedasticity, omitted variables, nonlogistic models, and transformation models. We present both graphical displays and tabular evidence.

Figures 2–6 present, for each type of DGP, a few summaries of the Monte Carlo distribution of the monotonic estimators $\hat{F}_n^+$ and $\hat{F}_n^\circ$ as functions of the cutoff level y, and of the monotonic estimators

[7] In principle, J should increase with the sample size, but for our Monte Carlo experiment we found that a fine enough fixed grid was a good choice. Our choice of a grid of $J = 199$ points is the result of some experimentation. Reducing the grid size increases the bias, increasing it slows down the computations. To check the validity of our choice of J, we performed 1000 simulations from the logistic linear location model with finer grid sizes for the largest sample ($J = 397$ and $J = 793$) and found no difference in the bias and variance of all estimators. Results are available upon request.

$\hat{Q}_n^+$ and $\hat{Q}_n^\circ$ as functions of the quantile level p. The summaries considered are the squared bias, the variance and the MSE, all averaged over the distribution of X. Each panel in a figure compares the results obtained from samples of increasing size ($n = 900$, 1600 and 3600) drawn from the same DGP.

Tables 1–5 focus on our two monotonic estimators of the CQF, namely the DR estimator $\hat{Q}_n^+$ and the QR estimator $\hat{Q}_n^\circ$, and present their average squared bias and average variance at various quantile levels ($p = .10$, .25, .50, .75 and .90) for different sample sizes.[8] The last two columns of each table present asymptotic calculations based on the results in Sections 2–4. To facilitate comparisons between the Monte Carlo results and the asymptotic calculations, we rescale all estimates multiplying by $\sqrt{n}$. Thus, care is needed when making comparisons across columns of a table corresponding to different sample sizes.

The results for the logistic linear location model are presented in Table 1, separately for $\beta = 2\pi$ and $\beta = 6\pi$, corresponding respectively to a medium (.50) and a high (.90) value of the regression R^2. To save space, the results for the other DGPs discussed in Section 4.2 are presented in Tables 2–5 for $\beta = 2\pi$ only. In line with our asymptotic framework, we allow the degree of misspecification to change with the sample size by setting $\delta = c/\sqrt{n}$ for different values of c.[9]

Before discussing specific results for each DGP, we briefly summarize some findings that are common across DGPs. First, the profiles of the average variance are different for estimators of the CDF and the CQF: for the former they have an inverted U-shape with evidence of asymmetry and bimodality, for the latter they instead have a nice symmetric U-shape with a minimum near $p = .50$. Second, $\hat{F}_n^\circ$ and $\hat{Q}_n^+$ have smoother average variance and MSE profiles than $\hat{F}_n^+$ and $\hat{Q}_n^\circ$, especially when the sample size is relatively small ($n = 900$), which reflects the fact that the former are obtained by integration, the latter by inversion. Third, the DR estimators $\hat{F}_n^+$ and $\hat{Q}_n^+$ are always less precise (i.e., have higher average variance) than the QR estimators $\hat{F}_n^\circ$ and $\hat{Q}_n^\circ$. Thus, the DR estimators are more efficient (i.e., have lower average MSE) than the QR estimators only in a few cases when they are substantially less biased (i.e., have lower average squared bias) than the QR estimators. Fourth, the asymptotic approximations to bias and variance are quite accurate, i.e., close to the Monte Carlo biases and variances, even for relatively small samples, except perhaps when y is near the tails of Y or p is close to 0 or 1. In particular, the average squared bias of all estimators and the ratio of their average squared bias to their average variance are roughly proportional to c^2 for any value of n. Finally, changes in the value of β affect

[8] The corresponding tables for our two monotonic estimators of the CDF, $\hat{F}_n^+$ and $\hat{F}_n^\circ$, are available upon request. Tables are also available for the squared bias and the variance of $\hat{Q}_n^+$ and $\hat{Q}_n^\circ$ at specific value of x. Qualitatively, the results are very similar to those for the average squared bias and the average variance.

[9] We choose the values of c in such a way that δ is constant along the main anti-diagonal of the 3×3 table corresponding to the nine different combinations of c and n that we consider.

heavily the variance of the DR estimators, but only have a very small effect on the variance of the QR estimators. The effects of changes in the value of β on the bias of the different estimators, and of changes in the value of δ and c on their bias and variance, vary instead with the form of misspecification.

5.2.1 Logistic linear location model

Figure 2 refers to the benchmark case with $\beta = 2\pi$. Since a linear specification is correct for both the CLF and the CQF, it is not surprising that all our estimators shows little evidence of bias. As a result, their average variance and MSE have essentially the same profiles.

Table 1 shows the average squared bias and the average variance of the DR estimator $\sqrt{n}\hat{Q}_n^+$ and the QR estimator $\sqrt{n}\hat{Q}_n^\circ$ at various quantile levels, separately for $\beta = 2\pi$ and $\beta = 6\pi$. The squared bias of $\hat{Q}_n^+$ is almost always slightly larger than the bias of $\hat{Q}_n^\circ$, while its variance is always larger than the variance of $\hat{Q}_n^\circ$. Notice that the squared bias and the variance of $\hat{Q}_n^+$ both increase with β, especially in smaller samples ($n = 900$). On the contrary, those of $\hat{Q}_n^\circ$ do not change with β. This is a consequence of the shift equivariance property of linear QR estimators (see e.g. Koenker 2005, p. 39), as the QR estimates of the intercept and the slope of model (3) when $(\alpha, \beta) = (0, 6\pi)$ are linked to those when $(\alpha, \beta) = (0, 2\pi)$ by the relationships $\hat{\alpha}(0, 6\pi) = \hat{\alpha}(0, 2\pi)$ and $\hat{\beta}(0, 6\pi) = \hat{\beta}(0, 2\pi) + 4\pi$. Since $\hat{Q}_n^\circ$ is always more precise and is almost always less biased than $\hat{Q}_n^+$, it emerges clearly as the estimator of choice. Further, consistently with earlier findings in Koenker, Leorato and Peracchi (2013), the relative advantage of the QR estimator in terms of efficiency increases with β.

5.2.2 Omitted variables

Figure 3 refers to the case when $\phi(x) = x^2$ and $\delta = 10$. Since our estimators omit a quadratic term, they are all biased. The average squared bias of $\hat{F}_n^+$ and $\hat{F}_n^\circ$ has an inverted U-shape with a peak at the center of the distribution of Y, while that of $\hat{Q}_n^+$ and $\hat{Q}_n^\circ$ changes little with p. The DR estimators $\hat{F}_n^+$ and $\hat{Q}_n^+$ have less bias than the QR estimators $\hat{F}_n^\circ$ and $\hat{Q}_n^\circ$, but they are also less precise, so their MSE is actually larger than that of the QR estimators when $n = 900$ and is only slightly smaller for larger sample sizes.

Table 2 shows the values of the average squared bias and variance of $\sqrt{n}\hat{Q}_n^+$ and $\sqrt{n}\hat{Q}_n^\circ$ when the degree of misspecification is $\delta = c/\sqrt{n}$, for different values of c and n. Notice that, in line with the asymptotic calculations, the average squared bias of $\hat{Q}_n^\circ$ changes very little with p. Increasing c increases by about c^2 the average squared bias of both estimators, and therefore also the ratio between their average squared bias and their average variance. For example, when $p = .50$ and $n = 1600$, doubling

c from 30 to 60 increases the average squared bias from 3.51 to 14.05 for $\hat{Q}_n^\circ$ and from .73 to 2.56 for $\hat{Q}_n^+$. It also increases the ratio between the average squared bias and the average variance from .48 to 1.91 for $\hat{Q}_n^\circ$ and from .07 to .23 for $\hat{Q}_n^+$.

From the table we can also compute the average squared bias and variance of $\hat{Q}_n^+$ and $\hat{Q}_n^\circ$ at any given quantile level p when δ is kept constant but n grows. The results are again in line with the asymptotic calculations. For example, when $\beta = 2\pi$ and $\delta = c/\sqrt{n} = 1$, reading along the anti-diagonal shows that the average squared bias of $\hat{Q}_n^\circ$ at $p = .50$ changes only marginally from $3.494/900 = .00388$ when $n = 900$ to $13.866/3600 = .00385$ when $n = 3600$. On the contrary, the average variance of $\hat{Q}_n^\circ$ at $p = .50$ falls from $8.179/900 = .0091$ when $n = 900$ to $7.433/3600 = .0021$ when $n = 3600$. The latter value represents 22.7 percent of the average variance when $n = 900$, very close to the value of 25 percent based on the asymptotic calculations.

The table shows that the relative advantage of $\hat{Q}_n^+$ in terms of lower bias falls with n but increases with c. In line with the asymptotic calculations, it also falls as p moves towards 0 or 1. On the other hand, the relative disadvantage of $\hat{Q}_n^+$ in terms of lower precision changes little with n or c, but falls as p moves towards 0 or 1.

5.2.3 Heteroskedasticity

Figure 4 refers to the case when $\sigma_\delta(x) = 1 + \delta x^2$, with $\delta = 1/6$, that is, the scale function is quadratic in x. Except for $\hat{Q}_n^\circ$ when $p = .50$, all estimators are biased, as a linear specification of the CLF or the CQF is generally incorrect under this form of heteroskedasticity.[10] However, the DR estimators $\hat{F}_n^+$ and $\hat{Q}_n^+$ are both more biased and less precise than the QR estimators $\hat{F}_n^\circ$ and $\hat{Q}_n^\circ$, so they are clearly dominated in terms of MSE.

Table 3 shows the values of the average squared bias and variance of $\sqrt{n}\hat{Q}_n^+$ and $\sqrt{n}\hat{Q}_n^\circ$ when the degree of misspecification is $\delta = c/\sqrt{n}$, for different values of c and n. As in the omitted variables case, increasing c increases by about c^2 the average squared bias of both estimators, and therefore also the ratio between their average squared bias and their average variance. For example, when $p = .10$ and $n = 900$, doubling c from 5 to 10 increases the average squared bias from .48 to 1.91 for $\hat{Q}_n^\circ$ and from 2.66 to 11.21 for $\hat{Q}_n^+$. It also increases the ratio between the average squared bias and the average variance from .02 to .08 for $\hat{Q}_n^\circ$ and from .10 to .38 for $\hat{Q}_n^+$.

The relative disadvantage of $\hat{Q}_n^+$ in terms of bias changes little with n but, in line with the asymptotic calculations, changes a lot with p and c, and is especially large when $p = 1/2$ (the quantile level at which

[10] Results for other forms of heteroskedasticity where one of the two approaches (either DR or QR) remains consistent are available upon request.

the QR estimator $\hat{Q}_n^\circ$ is consistent by the symmetry of the logistic distribution). On the other hand, the relative disadvantage of $\hat{Q}_n^+$ in terms of precision changes little with n, p and c.

5.2.4 Nonlogistic models

Figure 5 refers to the case when the DGP is of the form (3) but the error is a mixture $(1-\delta)*U+\delta*T$, where $\delta = 1/4$, U is distributed as standard logistic, and T is distributed as Student t with 3 degrees of freedom. The DR estimators $\hat{F}_n^+$ and $\hat{Q}_n^+$ are biased, as a linear specification is correct for the CQF but not for the CLF, and their average squared bias has a bimodal profile. This bimodality is in line with the asymptotic calculations as in this case $\Psi(x,u) = \zeta(u)$, where $\zeta(u) = (\Lambda(u) - G(u))/\lambda(u)$ and $G(u)$ denotes the distribution function of a Student t, and the square of $\zeta(u)$ is bimodal. Notice that there is also some evidence of bias for $\hat{F}_n^\circ$. Overall, the DR estimators are both more biased and less precise than the QR estimators, so they are clearly dominated in terms of MSE.

Table 4 shows the values of the average squared bias and variance of $\sqrt{n}\hat{Q}_n^+$ and $\sqrt{n}\hat{Q}_n^\circ$ when the mixing probability is $\delta = c/\sqrt{n}$, for different values of c and n. Since the bias of $\hat{Q}_n^\circ$ falls with n, the relative disadvantage of $\hat{Q}_n^+$ in terms of bias increases with n. It also varies a lot with p and c, in accordance with the asymptotic calculations.

We obtain qualitatively similar results when the error in model (3) is distributed asymmetrically as a mixture of the standard logistic and the standard Gumbel, the main difference being that now the bias and the variance of the DR estimators are no longer symmetric.[11]

5.2.5 Transformation models

Figure 6 refers to the case when the DGP is a Box-Cox transformation model of the form $Y^{(1-\delta)} + 1 = \alpha + \beta X + U$, with $\delta = 1/4$ and $\beta = 2\pi$. We also set $\alpha = 4\pi$ to guarantee that $Z = \alpha + \beta X + U$ is almost always nonnegative so, for any $\delta \neq 1$, the inverse transformation $Y = [\delta + (1-\delta)Z]^{1/(1-\delta)}$ returns almost always a real number. Now the QR estimators $\hat{F}_n^\circ$ and $\hat{Q}_n^\circ$ are biased, as a linear specification is correct for the CLF but not for the CQF, and the average squared bias of $\hat{F}_n^\circ$ has an inverted U-shape with a peak at the center of the distribution of Y, while that of $\hat{Q}_n^\circ$ is relatively small and changes little with p. Although the DR estimators $\hat{F}_n^+$ and $\hat{Q}_n^+$ have almost no bias, they are always much less precise than the QR estimators, so their MSE is actually larger than the MSE of the QR estimators.

Table 5 shows the values of the average squared bias and variance of $\sqrt{n}\hat{Q}_n^+$ and $\sqrt{n}\hat{Q}_n^\circ$ when the degree of misspecification is $\delta = c/\sqrt{n}$, for different values of c and n. The results support the graphical

[11] Results for this case are available upon request.

evidence from Figure 6. In particular, even in the worst case for $\hat{Q}_n^\circ$ (namely $c = 2$, $p = .10$ and $n = 900$), the ratio of its MSE to the MSE of $\hat{Q}_n^+$ is equal to .851.

6 Conclusions

In this paper we presented asymptotic results and Monte Carlo evidence on the sampling properties of monotonic CDF and CQF estimators obtained from the DR approach under the assumption that the CLF is linear in parameters and from the QR approach under the assumption that the CQF is linear in parameters. We considered both cases when the underlying linear-in-parameter models are correctly specified and several types of model misspecification of considerable practical relevance.

Our results may be summarized as follows. First, the profiles of the average variance are different for estimators of the CDF and the CQF: for the former they have an inverted U-shape, with evidence of asymmetry and bimodality, for the latter they instead have a nice symmetric U-shape even for relatively small samples ($n = 900$).

Second, estimators obtained by rearrangement ($\hat{F}_n^\circ$ and $\hat{Q}_n^+$) have smoother average variance and MSE profiles than estimators obtained by inversion ($\hat{F}_n^+$ and $\hat{Q}_n^\circ$), especially in smaller samples.

Third, the main advantage of the DR approach relative to the QR approach is that it produces estimators that are less biased (i.e., have lower average squared bias) in some settings. These include the cases when the assumed model ignores a quadratic term in the conditional mean or the need of monotonically transforming the outcome of interest. On the other hand, QR estimators are less biased when the assumed model ignores the presence of heteroskedastic or nonlogistic errors.

Fourth, DR estimators are always less precise (i.e., have higher average variance) than QR estimators. Thus, the only case when they are more efficient (i.e., have lower average MSE) than QR estimators is when they have substantially less bias. In our Monte Carlo experiments this only occurs when the assumed models omit a quadratic term in the conditional mean.

Fifth, the asymptotic approximations to bias and variance are quite accurate, even for relatively small samples, except perhaps when y is near the tails of Y or p is close to 0 or 1.

We hope that our results provide guidance to practitioners about the choice between the DR and the QR approach. Of course, when it comes to choosing between the two approaches, other aspects may also matter besides the sampling properties in finite or in large samples. One important aspect is the possibility of generalizing to the case when Y is discrete, or subject to censoring, or multivariate. Another is the presence of censoring or mass points in the distribution of the outcome of interest. In both these cases, which we leave for future research, the DR approach may look more natural.

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. (2007). Quantile and probability curves without crossing. CeMMAP Working Paper CWP 10/07.

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.

Hall P., and Müller P.-G. (2003). Order-preserving nonparametric regression, with application to conditional distribution and quantile function estimation. *Journal of the American Statistical Association*, 98: 598–608.

Hall P., Wolff R.C.L., and Yao Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, 94: 154–163.

Hothorn T., Kneib T., and Bühlmann P. (2014). Conditional transformation models. *Journal of the Royal Statistical Society–Series B*, 76: 3–37.

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

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

Koenker R., Leorato S., and Peracchi F. (2013). Distributional vs. quantile regression. EIEF Working Paper 13/29.

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

Leorato S., and Peracchi F. (2015). Shape regressions. EIEF Working Paper 15/07.

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

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

van der Vaart A.W. (1998). *Asymptotic Statistics*. Cambridge University Press: New York.

	$n=900$		$n=1600$		$n=3600$		$n=\infty$	
				$\beta=2\pi$				
	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$
$p=.10$	.004	.057	.012	.038	.028	.027	.000	.000
$p=.25$	.000	.014	.007	.013	.008	.009	.000	.000
$p=.50$	.001	.008	.001	.008	.004	.012	.000	.000
$p=.75$	.003	.015	.006	.035	.004	.019	.000	.000
$p=.90$	.013	.068	.095	.071	.021	.067	.000	.000
	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$
$p=.10$	22.159	25.407	21.712	24.891	20.117	24.261	20.149	24.616
$p=.25$	10.578	13.792	9.968	13.358	9.572	12.961	9.672	13.028
$p=.50$	8.101	10.618	7.173	10.050	7.480	10.122	7.254	10.091
$p=.75$	10.233	13.528	9.883	12.994	9.998	13.007	9.672	13.028
$p=.90$	22.114	25.630	20.372	24.580	20.012	24.326	20.149	24.616
				$\beta=6\pi$				
	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$
$p=.10$	.004	.311	.012	.152	.023	.061	.000	.000
$p=.25$	.001	.078	.008	.033	.009	.017	.000	.000
$p=.50$	.001	.029	.001	.013	.004	.018	.000	.000
$p=.75$	.003	.117	.005	.078	.005	.055	.000	.000
$p=.90$	.012	.390	.088	.207	.020	.131	.000	.000
	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$
$p=.10$	22.035	49.833	21.197	51.167	19.977	49.572	20.149	50.617
$p=.25$	10.472	27.631	9.781	27.716	9.609	26.973	9.672	27.489
$p=.50$	8.107	20.626	7.174	20.294	7.480	20.061	7.254	20.457
$p=.75$	10.126	27.358	9.791	27.202	9.999	27.080	9.672	27.489
$p=.90$	22.061	50.146	20.304	49.999	19.892	49.488	20.149	50.617

Table 1: Average squared bias $\mathbb{B}^2$ and variance $\mathbb{V}$ of the rescaled CQF estimators $\sqrt{n}\hat{Q}_n^\circ$ and $\sqrt{n}\hat{Q}_n^+$ for different sample sizes at 5 different quantile levels. The DGP is the logistic linear location model $Y=\beta X+U$, with $X\sim U(0,1)$ and $U\sim$ standard logistic.

	$n = 900$		$n = 1600$		$n = 3600$		$n = \infty$	
				$c = 60$				
	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$
$p = .10$	13.539	4.116	13.613	4.650	13.518	5.755	14.047	7.218
$p = .25$	13.600	2.221	13.718	2.570	13.647	2.825	14.047	4.074
$p = .50$	13.936	2.331	14.053	2.560	13.866	2.823	14.047	4.053
$p = .75$	14.195	1.796	14.277	2.219	14.030	2.682	14.047	4.074
$p = .90$	14.403	3.339	14.606	4.354	13.912	4.955	14.047	7.218
	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$
$p = .10$	22.836	28.894	21.416	27.523	20.714	26.154	20.149	24.616
$p = .25$	10.832	16.038	9.899	14.910	9.725	14.054	9.672	13.028
$p = .50$	8.269	12.102	7.372	11.212	7.433	10.880	7.254	10.091
$p = .75$	9.838	15.352	9.436	14.442	9.862	13.813	9.672	13.028
$p = .90$	21.374	28.631	18.899	27.137	19.318	25.939	20.149	24.616
				$c = 40$				
	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$
$p = .10$	6.068	2.336	6.059	2.459	5.985	2.888	6.243	3.208
$p = .25$	6.082	1.237	6.101	1.377	6.059	1.402	6.243	1.810
$p = .50$	6.184	1.217	6.230	1.250	6.142	1.315	6.243	1.801
$p = .75$	6.232	.974	6.263	1.147	6.198	1.305	6.243	1.810
$p = .90$	6.249	1.968	6.474	2.467	6.105	2.573	6.243	3.208
	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$
$p = .10$	22.141	27.642	21.339	26.521	20.411	25.481	20.149	24.616
$p = .25$	10.642	15.256	9.661	14.386	9.615	13.677	9.672	13.028
$p = .50$	8.206	11.621	7.330	10.817	7.487	10.632	7.254	10.091
$p = .75$	9.710	14.684	9.404	13.943	9.962	13.527	9.672	13.028
$p = .90$	21.489	27.546	19.130	26.166	19.471	25.345	20.149	24.616
				$c = 30$				
	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$
$p = .10$	3.428	1.502	3.392	1.521	3.357	1.738	3.512	1.804
$p = .25$	3.427	.789	3.420	.856	3.395	.835	3.512	1.018
$p = .50$	3.494	.741	3.507	.726	3.440	.744	3.512	1.013
$p = .75$	3.487	.595	3.520	.693	3.467	.758	3.512	1.018
$p = .90$	3.469	1.298	3.672	1.595	3.404	1.586	3.512	1.804
	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$
$p = .10$	22.065	27.048	21.411	26.008	20.379	25.151	20.149	24.616
$p = .25$	10.646	14.887	9.599	14.133	9.663	13.491	9.672	13.028
$p = .50$	8.179	11.380	7.266	10.633	7.498	10.508	7.254	10.091
$p = .75$	9.743	14.394	9.373	13.682	9.925	13.384	9.672	13.028
$p = .90$	21.630	27.043	19.194	25.728	19.692	25.016	20.149	24.616

Table 2: Average squared bias $\mathbb{B}^2$ and variance $\mathbb{V}$ of the rescaled CQF estimators $\sqrt{n}\hat{Q}_n^\circ$ and $\sqrt{n}\hat{Q}_n^+$ for different sample sizes at 5 different quantile levels. The DGP is the quadratic model $Y = 2\pi X + \delta X^2 + U$, with $X \sim U(0,1)$, $U \sim$ standard logistic, and $\delta = c/\sqrt{n}$.

	$n=900$		$n=1600$		$n=3600$		$n=\infty$	
				$c=10$				
	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$
$p=.10$	1.914	11.209	1.977	10.707	2.003	10.475	1.884	10.168
$p=.25$	.478	5.883	.518	5.907	.514	5.940	.471	5.587
$p=.50$	.001	4.291	.001	4.849	.004	4.882	.000	4.666
$p=.75$	.456	3.808	.457	3.997	.454	3.990	.471	3.979
$p=.90$	1.818	5.998	1.951	6.078	1.800	6.226	1.884	6.818
	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$
$p=.10$	26.928	29.269	26.170	28.014	22.207	25.745	20.149	24.616
$p=.25$	13.032	16.110	11.677	15.054	10.579	13.958	9.672	13.028
$p=.50$	10.075	12.761	8.592	11.481	8.397	11.040	7.254	10.091
$p=.75$	12.535	16.696	11.836	15.303	11.335	14.525	9.672	13.028
$p=.90$	26.998	32.370	24.008	29.249	22.329	27.464	20.149	24.616
				$c=6.67$				
	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$
$p=.10$	.855	4.826	.908	4.610	.929	4.528	.837	4.519
$p=.25$	.215	2.588	.244	2.592	.241	2.655	.209	2.483
$p=.50$	.001	1.962	.001	2.263	.004	2.270	.000	2.074
$p=.75$	.201	1.763	.201	1.835	.202	1.822	.209	1.768
$p=.90$	.791	2.778	.921	2.731	.788	2.797	.837	3.030
	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$
$p=.10$	25.352	27.897	23.920	26.882	21.412	25.217	20.149	24.616
$p=.25$	12.169	15.296	11.104	14.502	10.245	13.621	9.672	13.028
$p=.50$	9.396	12.014	8.110	10.966	8.093	10.733	7.254	10.091
$p=.75$	11.705	15.589	11.128	14.519	10.883	13.997	9.672	13.028
$p=.90$	25.519	30.025	22.185	27.663	21.511	26.386	20.149	24.616
				$c=5$				
	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$
$p=.10$	.484	2.664	.527	2.543	.552	2.495	.471	2.542
$p=.25$	.124	1.445	.147	1.445	.146	1.510	.118	1.397
$p=.50$	.001	1.126	.001	1.325	.004	1.332	.000	1.166
$p=.75$	.113	1.026	.112	1.063	.113	1.047	.118	.995
$p=.90$	.438	1.603	.542	1.534	.441	1.589	.471	1.705
	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$
$p=.10$	24.445	27.216	22.651	26.329	21.040	24.985	20.149	24.616
$p=.25$	11.672	14.890	10.459	14.221	10.068	13.450	9.672	13.028
$p=.50$	9.068	11.653	7.868	10.729	7.936	10.573	7.254	10.091
$p=.75$	11.220	15.053	10.483	14.132	10.640	13.745	9.672	13.028
$p=.90$	24.499	28.884	20.898	26.851	21.095	25.843	20.149	24.616

Table 3: Average squared bias $\mathbb{B}^2$ and variance $\mathbb{V}$ of the rescaled CQF estimators $\sqrt{n}\hat{Q}_n^\circ$ and $\sqrt{n}\hat{Q}_n^+$ for different sample sizes at 5 different quantile levels. The DGP is the heteroskedastic model $Y = 2\pi X + (1+\delta X^2)U$, with $X \sim U(0,1)$, $U \sim$ standard logistic, and $\delta = c/\sqrt{n}$.

	$n = 900$		$n = 1600$		$n = 3600$		$n = \infty$	
				$c = 15$				
	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$
$p = .10$	.005	1.546	.146	1.618	.012	1.495	.000	1.405
$p = .25$	.001	2.910	.010	3.677	.009	3.585	.000	4.458
$p = .50$	.004	.243	.008	.380	.025	.357	.000	.515
$p = .75$	.023	3.133	.018	3.930	.033	3.841	.000	4.458
$p = .90$	.012	1.542	.081	1.747	.031	1.733	.000	1.405
	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$
$p = .10$	19.819	23.149	19.586	23.927	20.072	23.604	20.149	24.616
$p = .25$	7.738	10.976	7.763	11.354	8.620	11.498	9.672	13.028
$p = .50$	4.641	8.016	5.337	8.321	5.449	8.649	7.254	10.091
$p = .75$	7.773	10.695	7.573	11.014	8.312	11.653	9.672	13.028
$p = .90$	19.549	22.489	19.088	22.798	20.479	24.295	20.149	24.616
				$c = 10$				
	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$
$p = .10$	.018	.426	.061	.538	.005	.624	.000	.624
$p = .25$	.002	1.367	.004	1.715	.019	1.642	.000	1.981
$p = .50$	.006	.115	.008	.180	.035	.161	.000	.229
$p = .75$	.026	1.608	.011	1.901	.024	1.868	.000	1.981
$p = .90$	.005	.517	.036	.669	.013	.829	.000	.624
	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$
$p = .10$	19.762	23.372	19.911	23.562	20.229	23.789	20.149	24.616
$p = .25$	8.553	11.605	8.298	11.600	8.832	11.810	9.672	13.028
$p = .50$	5.221	8.461	5.677	8.691	5.921	8.986	7.254	10.091
$p = .75$	7.950	11.054	7.939	11.434	8.622	11.979	9.672	13.028
$p = .90$	19.086	22.912	19.034	23.101	20.739	24.123	20.149	24.616
				$c = 7.5$				
	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$
$p = .10$	.027	.157	.067	.281	.008	.359	.000	.351
$p = .25$	.006	.713	.006	1.025	.022	.917	.000	1.114
$p = .50$	.005	.066	.012	.116	.033	.091	.000	.129
$p = .75$	.030	.923	.009	1.146	.021	1.059	.000	1.114
$p = .90$	.001	.232	.029	.325	.016	.494	.000	.351
	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$
$p = .10$	19.673	23.585	20.341	23.618	19.938	23.975	20.149	24.616
$p = .25$	8.655	11.844	8.827	11.983	8.871	11.996	9.672	13.028
$p = .50$	5.492	8.795	6.040	9.122	6.135	9.114	7.254	10.091
$p = .75$	8.101	11.507	8.542	11.946	8.794	12.065	9.672	13.028
$p = .90$	19.213	23.401	19.981	23.451	20.319	24.084	20.149	24.616

Table 4: Average squared bias $\mathbb{B}^2$ and variance $\mathbb{V}$ of the rescaled CQF estimators $\sqrt{n}\hat{Q}_n^\circ$ and $\sqrt{n}\hat{Q}_n^+$ for different sample sizes at 5 different quantile levels. The DGP is the nonlogistic model $Y = 2\pi X + V$, with $X \sim U(0,1)$ and $V \sim (1-\delta) * U + \delta * T$, where $U \sim$ standard logistic, $T \sim$ Student t_3, and $\delta = c/\sqrt{n}$.

	$n=900$		$n=1600$		$n=3600$		$n=\infty$	
				$c=2$				
	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$
$p=.10$	0.055	0.081	0.051	0.049	0.066	0.032	0.034	0.000
$p=.25$	0.045	0.020	0.042	0.017	0.039	0.011	0.029	0.000
$p=.50$	0.044	0.011	0.035	0.011	0.031	0.014	0.025	0.000
$p=.75$	0.034	0.022	0.030	0.047	0.027	0.022	0.022	-0.000
$p=.90$	0.037	0.101	0.114	0.095	0.035	0.082	0.019	0.000
	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$
$p=.10$	31.259	36.396	27.221	32.497	23.961	28.907	20.149	24.616
$p=.25$	15.128	20.008	12.676	17.608	11.462	15.542	9.672	13.028
$p=.50$	11.847	15.579	9.573	13.355	9.032	12.204	7.254	10.091
$p=.75$	14.781	20.048	12.818	17.392	12.118	15.766	9.672	13.028
$p=.90$	32.661	38.377	26.006	33.147	24.122	29.631	20.149	24.616
				$c=1.33$				
	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$
$p=.10$	0.021	0.072	0.027	0.045	0.045	0.030	0.015	0.000
$p=.25$	0.017	0.017	0.020	0.015	0.021	0.010	0.013	0.000
$p=.50$	0.017	0.010	0.014	0.010	0.014	0.014	0.011	0.000
$p=.75$	0.013	0.019	0.012	0.042	0.012	0.021	0.010	0.000
$p=.90$	0.020	0.088	0.090	0.086	0.023	0.076	0.008	-0.000
	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$
$p=.10$	27.634	32.171	24.845	29.674	22.556	27.244	20.149	24.616
$p=.25$	13.308	17.608	11.524	16.025	10.783	14.616	9.672	13.028
$p=.50$	10.416	13.656	8.681	12.121	8.470	11.456	7.254	10.091
$p=.75$	12.929	17.513	11.603	15.746	11.348	14.772	9.672	13.028
$p=.90$	28.446	33.404	23.530	29.933	22.572	27.717	20.149	24.616
				$c=1$				
	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$	$\mathbb{B}^2(\hat{Q}_n^\circ)$	$\mathbb{B}^2(\hat{Q}_n^+)$
$p=.10$	0.011	0.068	0.019	0.043	0.038	0.029	0.008	0.000
$p=.25$	0.010	0.016	0.015	0.015	0.015	0.010	0.007	0.000
$p=.50$	0.010	0.009	0.007	0.010	0.009	0.013	0.006	0.000
$p=.75$	0.007	0.018	0.007	0.040	0.007	0.020	0.005	-0.000
$p=.90$	0.015	0.082	0.082	0.082	0.019	0.074	0.005	0.000
	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$	$\mathbb{V}(\hat{Q}_n^\circ)$	$\mathbb{V}(\hat{Q}_n^+)$
$p=.10$	26.021	30.287	23.744	28.378	21.873	26.457	20.149	24.616
$p=.25$	12.499	16.542	10.992	15.301	10.466	14.179	9.672	13.028
$p=.50$	9.773	12.805	8.266	11.557	8.204	11.103	7.254	10.091
$p=.75$	12.110	16.394	11.048	14.995	10.987	14.304	9.672	13.028
$p=.90$	26.596	31.215	22.394	28.470	21.842	26.818	20.149	24.616

Table 5: Average squared bias $\mathbb{B}^2$ and variance $\mathbb{V}$ of the rescaled CQF estimators $\sqrt{n}\hat{Q}_n^\circ$ and $\sqrt{n}\hat{Q}_n^+$ for different sample sizes at 5 different quantile levels. The DGP is the Box-Cox transformation model $Y^{(1-\delta)} = 4\pi + 2\pi X + U$, with $X \sim \mathscr{U}(0,1)$, $U \sim$ standard logistic, and $\delta = c/\sqrt{n}$.

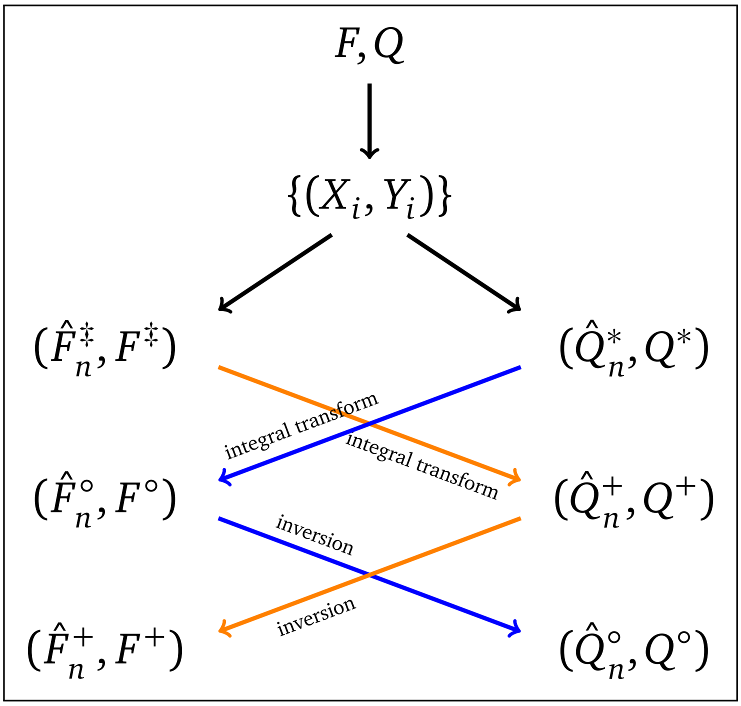

Figure 1: Scheme of the DR and QR approaches.

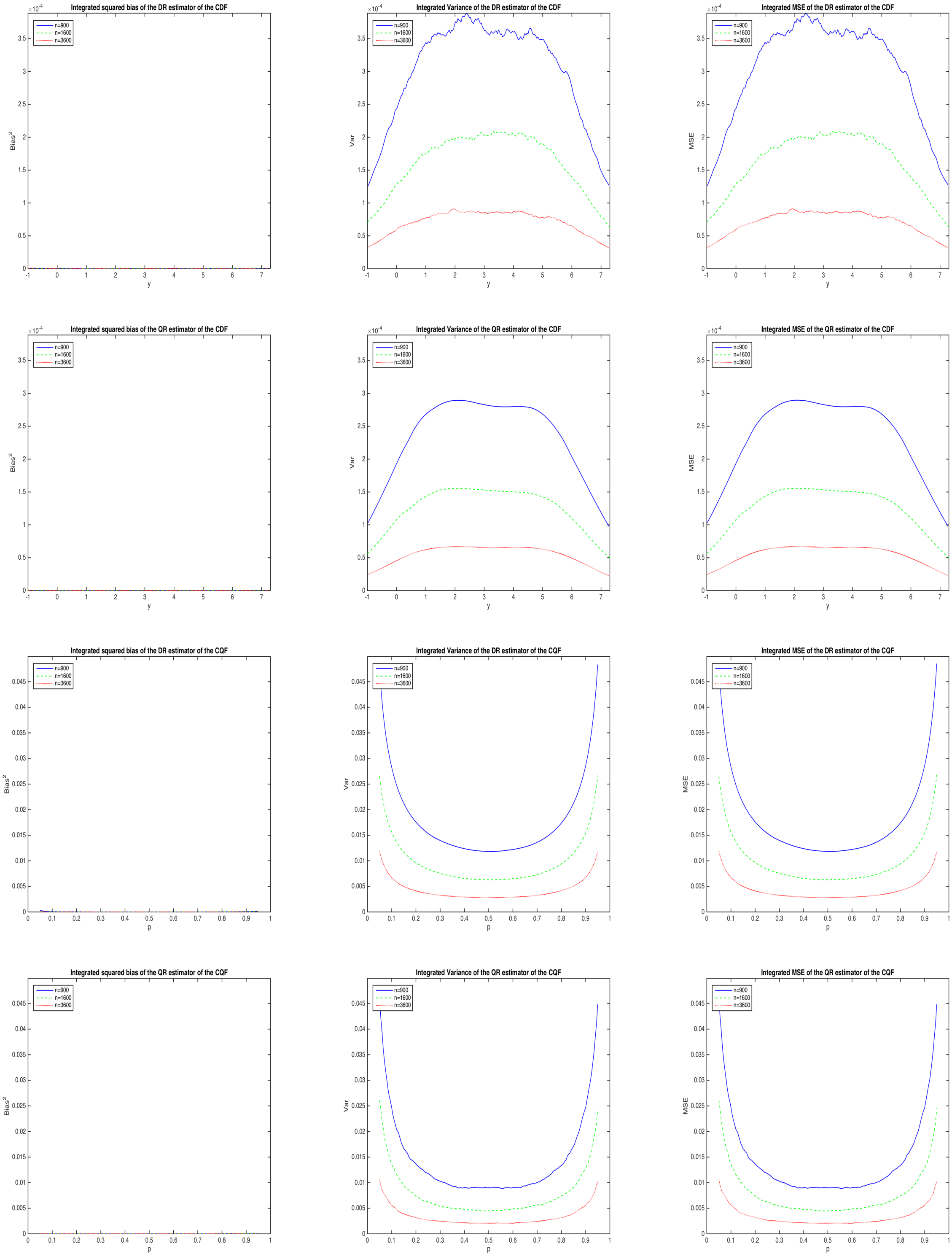

Figure 2: Average squared bias (first column), variance (second column) and MSE (third column) of $\hat{F}_n^{+}(y\,|\,x)$, $\hat{F}_n^{\circ}(y\,|\,x)$, $\hat{Q}_n^{+}(p\,|\,x)$ and $\hat{Q}_n^{\circ}(p\,|\,x)$ (first to fourth row). The DGP is the logistic linear location model $Y = 2\pi X + U$, with $X \sim \mathscr{U}(0,1)$ and $U \sim$ standard logistic.

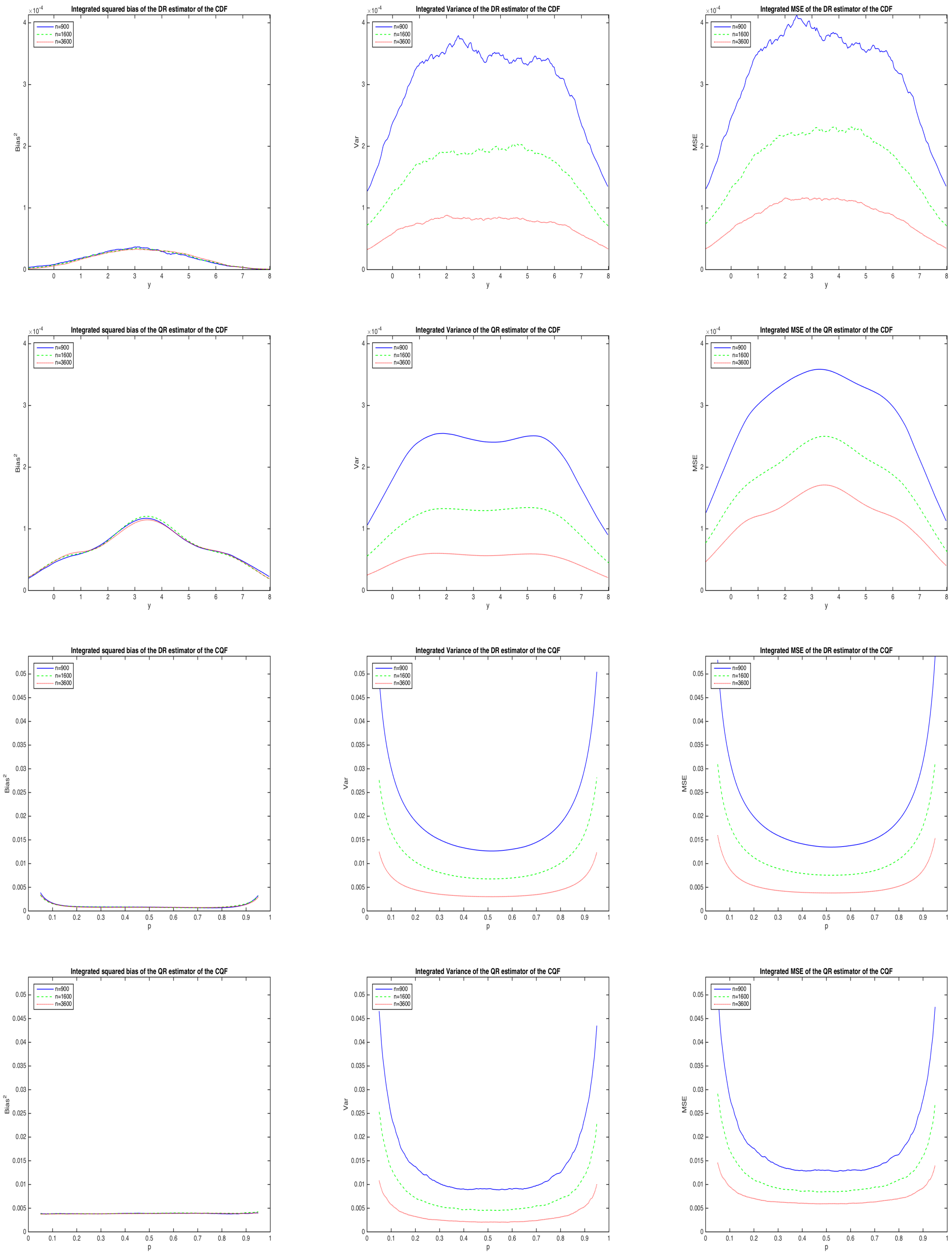

Figure 3: Average squared bias (first column), variance (second column) and MSE (third column) of $\hat{F}_n^{+}(y\,|\,x)$, $\hat{F}_n^{\circ}(y\,|\,x)$, $\hat{Q}_n^{+}(p\,|\,x)$ and $\hat{Q}_n^{\circ}(p\,|\,x)$ (first to fourth row). The DGP is the quadratic model $Y = 2\pi X + X^2 + U$, with $X \sim \mathscr{U}(0,1)$ and $U \sim$ standard logistic.

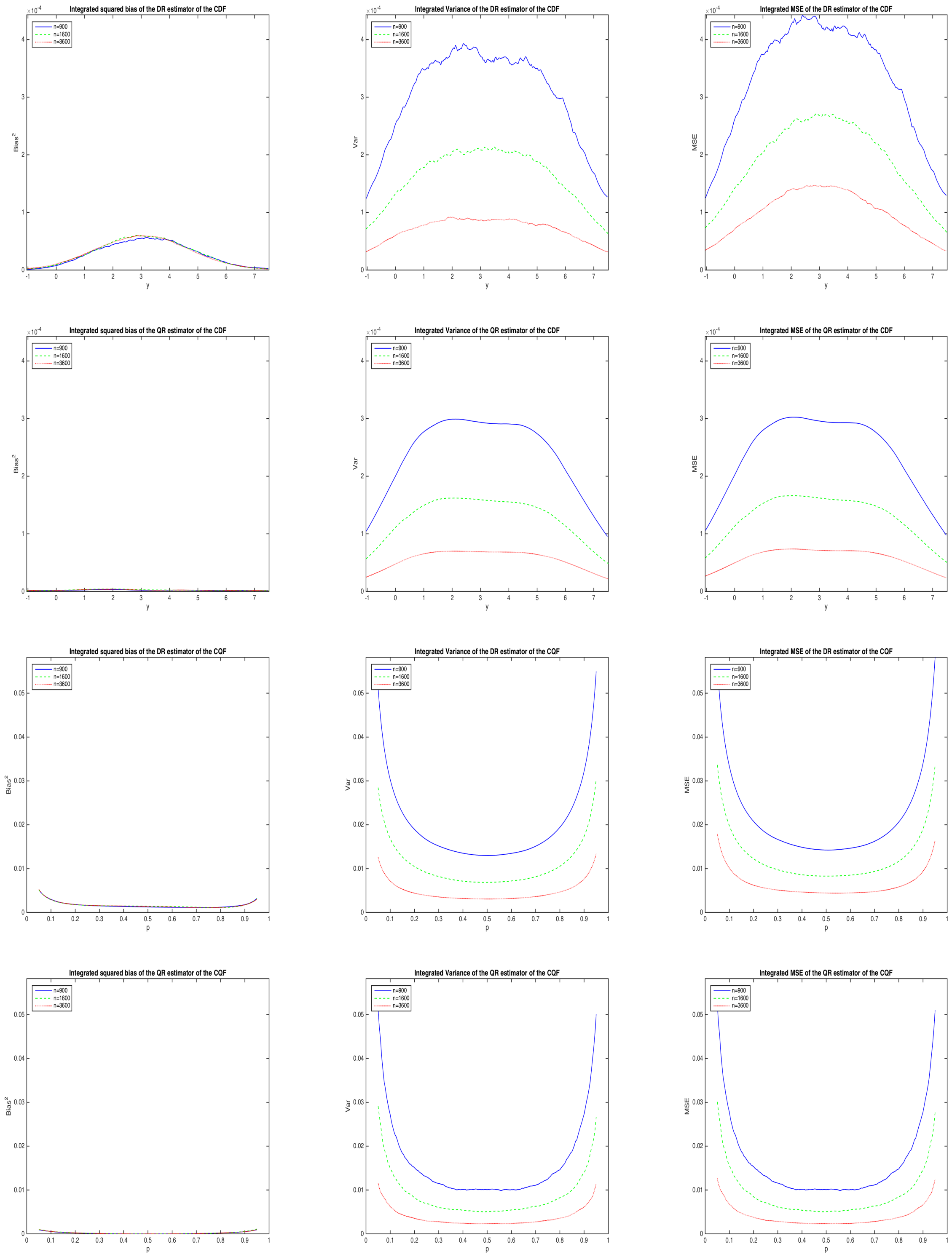

Figure 4: Average squared bias (first column), variance (second column) and MSE (third column) of $\hat{F}_n^{+}(y\,|\,x)$, $\hat{F}_n^{\circ}(y\,|\,x)$, $\hat{Q}_n^{+}(p\,|\,x)$ and $\hat{Q}_n^{\circ}(p\,|\,x)$ (first to fourth row). The DGP is the heteroskedastic model $Y = 2\pi X + (1 + X^2/6)U$, with $X \sim \mathscr{U}(0,1)$ and $U \sim$ standard logistic.

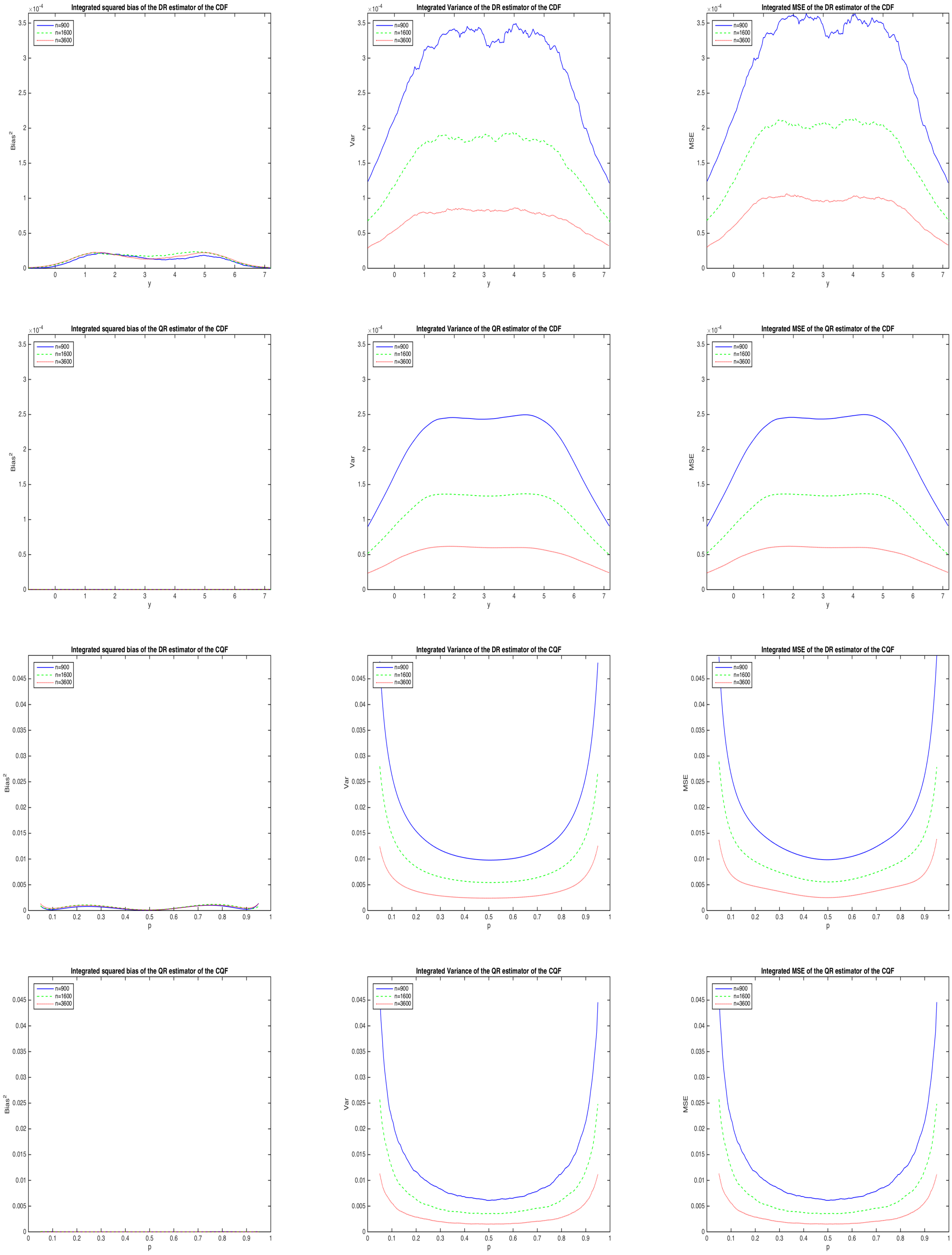

Figure 5: Average squared bias (first column), variance (second column) and MSE (third column) of $\hat{F}_n^{+}(y\,|\,x)$, $\hat{F}_n^{\circ}(y\,|\,x)$, $\hat{Q}_n^{+}(p\,|\,x)$ and $\hat{Q}_n^{\circ}(p\,|\,x)$ (first to fourth row). The DGP is the nonlogistic model $Y = 2\pi X + V$, with $X \sim \mathscr{U}(0,1)$ and $V \sim (3/4) * U + (1/4) * T$, where $U \sim$ standard logistic and $T \sim$ Student t_3.

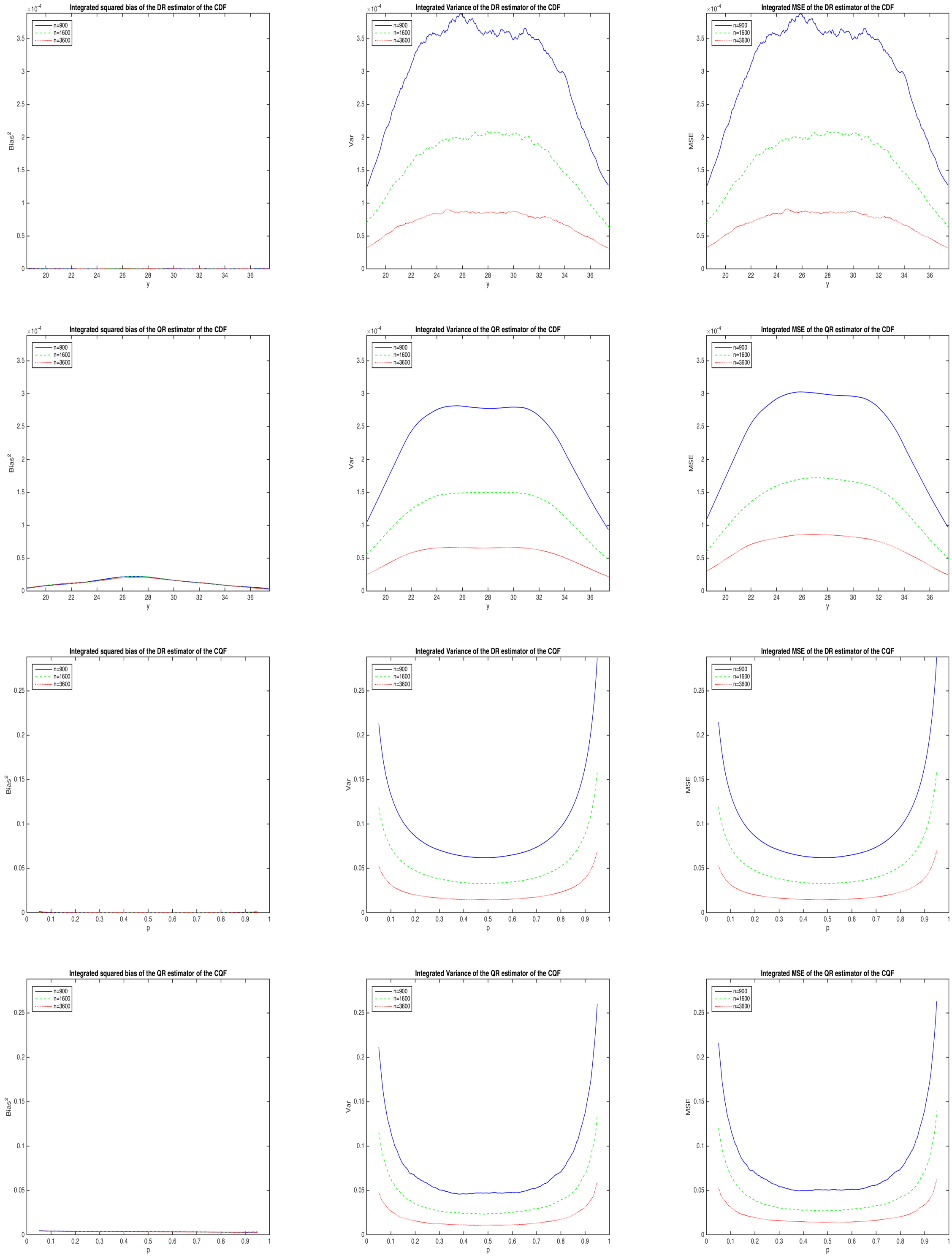

Figure 6: Average squared bias (first column), variance (second column) and MSE (third column) of $\hat{F}_n^+(y\,|\,x)$, $\hat{F}_n^\circ(y\,|\,x)$, $\hat{Q}_n^+(p\,|\,x)$ and $\hat{Q}_n^\circ(p\,|\,x)$ (first to fourth row). The DGP is the Box-Cox transformation model $Y^{(3/4)}+1=4\pi+2\pi X+U$, with $X\sim\mathscr{U}(0,1)$ and $U\sim$ standard logistic.

Appendix

This appendix provides expressions for the local asymptotic bias of estimators based on linear-in-parameter models for the CLF and the CQF when the DGP is of the form (5) for various specifications of the function $\psi_\delta(x,u)$ and $\delta = c/\sqrt{n}$ for some constant c. We denote by $\Lambda(u)$ and $\lambda(u)$ the distribution function and the density of the standard logistic distribution, and let $\boldsymbol{x} = (1, x^\top)^\top$, $\boldsymbol{X} = (1, X^\top)^\top$, $P_X = \mathbb{E}\, \boldsymbol{X}\boldsymbol{X}^\top$, $\lambda_y(X) = \lambda(y - \alpha - X^\top\beta)$, $H(y) = \mathbb{E}\left[\lambda_y(X)\boldsymbol{X}\boldsymbol{X}^\top\right]$, and $y(p\,|\,x) = \alpha + \Lambda^{-1}(p) + x^\top\beta$.

Omitted variables

Since $\psi_\delta(x,u) = \delta\phi(x) + u$ and $\Psi(x,u) = \phi(x)$, we have

$$
\begin{aligned}
\mathbb{B}\left(\hat{F}_n^\ddagger(y\,|\,x)\right) &= \mathbb{B}\left(\hat{F}_n^+(y\,|\,x)\right) = c\,\lambda_y(X)\left[\phi(x) - \boldsymbol{x}^\top H(y)^{-1}\,\mathbb{E}\left(\lambda_y(X)\,\phi(X)\boldsymbol{X}\right)\right],\\
\mathbb{B}\left(\hat{F}_n^\circ(y\,|\,x)\right) &= c\,\lambda_y(X)\left[\phi(x) - \boldsymbol{x}^\top P_X^{-1}\,\mathbb{E}\left(\phi(X)\boldsymbol{X}\right)\right],\\
\mathbb{B}\left(\hat{Q}_n^+(p\,|\,x)\right) &= -c\left[\phi(x) - c\,x^\top H\left(y(p\,|\,x)\right)^{-1}\mathbb{E}\left(\lambda_{y(p\,|\,x)}(X)\,\phi(X)\boldsymbol{X}\right)\right],\\
\mathbb{B}\left(\hat{Q}_n^*(p\,|\,x)\right) &= \mathbb{B}\left(\hat{Q}_n^\circ(p\,|\,x)\right) = -c\left[\phi(x) - x^\top P_X^{-1}\,\mathbb{E}\left(\phi(X)\boldsymbol{X}\right)\right].
\end{aligned}
$$

Heteroskedasticity

Since $\psi_\delta(x,u) = \left(1 + \delta\phi(x)\right)u$ and $\Psi(x,u) = \phi(x)u$, we have

$$
\begin{aligned}
\mathbb{B}\left(\hat{F}_n^\ddagger(y\,|\,x)\right) &= \mathbb{B}\left(\hat{F}_n^+(y\,|\,x)\right)\\
&= c\,\lambda_y(X)\left[\phi(x)(y - \alpha - x^\top\beta) - \boldsymbol{x}^\top H(y)^{-1}\,\mathbb{E}\left(\phi(X)\,\lambda_y(X)\,(y - \alpha - X^\top\beta)\boldsymbol{X}\right)\right],\\
\mathbb{B}\left(\hat{F}_n^\circ(y\,|\,x)\right) &= c\,\lambda_y(X)\left[\phi(x)(y - \alpha - x^\top\beta) - \boldsymbol{x}^\top P_X^{-1}\,\mathbb{E}\left(\phi(X)\,(y - \alpha - x^\top\beta)\boldsymbol{X}\right)\right],\\
\mathbb{B}\left(\hat{Q}_n^+(p\,|\,x)\right) &= -c\,\phi(x)\Lambda^{-1}(p) + c\,\boldsymbol{x}^\top H\left(y(p\,|\,x)\right)^{-1}\mathbb{E}\left[\phi(X)\,\lambda_{y(p\,|\,x)}(X)\left(y(p\,|\,x) - \alpha - X^\top\beta\right)\boldsymbol{X}\right],\\
\mathbb{B}\left(\hat{Q}_n^*(p\,|\,x)\right) &= \mathbb{B}\left(\hat{Q}_n^\circ(p\,|\,x)\right) = -c\,\Lambda^{-1}(p)\left[\phi(x) - \boldsymbol{x}^\top P_X^{-1}\,\mathbb{E}\left(\phi(X)\boldsymbol{X}\right)\right].
\end{aligned}
$$

Nonlogistic models

Since $\psi_\delta(x,u) = G_\delta^{-1}(\Lambda(u))$ and $\Phi(x,u) = \zeta(u)$ for some function ζ, we have

$$
\begin{aligned}
\mathbb{B}\left(\hat{F}_n^\ddagger(y\,|\,x)\right) &= \mathbb{B}\left(\hat{F}_n^+(y\,|\,x)\right)\\
&= c\,\lambda_y(X)\,\zeta(y - \alpha - x^\top\beta) - c\,\lambda_y(X)\,\boldsymbol{x}^\top H(y)^{-1}\,\mathbb{E}\left[\lambda_y(X)\,\zeta(y - \alpha - X^\top\beta)\boldsymbol{X}\right],\\
\mathbb{B}\left(\hat{F}_n^\circ(y\,|\,x)\right) &= c\,\lambda_y(X)\,\zeta(y - \alpha - x^\top\beta)\left(1 - \boldsymbol{x}^\top P_X^{-1}\mu_X\right) = 0,\\
\mathbb{B}\left(\hat{Q}_n^+(p\,|\,x)\right) &= -c\,\zeta\left(\Lambda^{-1}(p)\right) + c\,\boldsymbol{x}^\top H\left(y(p\,|\,x)\right)^{-1}\mathbb{E}\left[\lambda_{y(p\,|\,x)}(X)\,\zeta\left(y(p\,|\,x) - \alpha - X^\top\beta\right)\boldsymbol{X}\right],\\
\mathbb{B}\left(\hat{Q}_n^*(p\,|\,x)\right) &= \mathbb{B}\left(\hat{Q}_n^\circ(p\,|\,x)\right) = -c\,\zeta\left(\Lambda^{-1}(p)\right)\left(1 - \boldsymbol{x}^\top P_X^{-1}\mu_X\right) = 0,
\end{aligned}
$$

where $\mu_X = \mathbb{E}\,\boldsymbol{X}$ and we used the fact that $\boldsymbol{x}^\top P_X^{-1}\mu_X = 1$ for any $\boldsymbol{x}$.

Transformation models

Consider the monotone transformation model $\phi_\delta(Y) = \alpha + X^\top\beta + U$, where U is logistic, $\phi_\delta(y)$ is monotone for every δ and $\lim_{\delta\to 0}\phi_\delta(y) = y$. Then ϕ_δ is invertible for all δ and the model can equivalently be written as $Y = \alpha + X^\top\beta + \psi_\delta(X,U)$, where $\psi_\delta(x,u) = \phi_\delta^{-1}(\alpha + x^\top\beta + u) - \alpha - x^\top\beta$. Since

$$\Psi(x,u) = \lim_{\delta\to 0}\frac{\phi_\delta^{-1}(\alpha + x^\top\beta + u) - (\alpha + x^\top\beta + u)}{\delta} = \chi(\alpha + x^\top\beta + u),$$

for some function χ, we have

$$\begin{aligned}
\mathbb{B}\left(\hat{F}_n^{\ddagger}(y\,|\,x)\right) &= \mathbb{B}\left(\hat{F}_n^{+}(y\,|\,x)\right) = c\,\lambda_y(X)\,\chi(y)\left[1 - \boldsymbol{x}^\top H(y)^{-1}\,\mathbb{E}\left(\lambda_y(X)\boldsymbol{X}\right)\right] = 0,\\
\mathbb{B}\left(\hat{F}_n^{\circ}(y\,|\,x)\right) &= c\,\lambda_y(X)\left[\chi(y) - \boldsymbol{x}^\top P_X^{-1}\,\mathbb{E}\left(\chi(y - (x - X)^\top\beta)\boldsymbol{X}\right)\right],\\
\mathbb{B}\left(\hat{Q}_n^{+}(p\,|\,x)\right) &= -c\,\chi\left(y(p\,|\,x)\right)\left[1 - \boldsymbol{x}^\top H\left(y(p\,|\,x)\right)^{-1}\,\mathbb{E}\left(\lambda_{y(p\,|\,x)}(X)\boldsymbol{X}\right)\right] = 0,\\
\mathbb{B}\left(\hat{Q}_n^{*}(p\,|\,x)\right) &= \mathbb{B}\left(\hat{Q}_n^{\circ}(p\,|\,x)\right) = -c\left[\chi\left(y(p\,|\,x)\right) - \boldsymbol{x}^\top P_X^{-1}\,\mathbb{E}\left(\chi(y(p\,|\,x))\right)\right],
\end{aligned}$$

where we used the fact that $\boldsymbol{x}^\top H(y)^{-1}\,\mathbb{E}\left(\lambda_y(X)\boldsymbol{X}\right) = 1$ for any $\boldsymbol{x}$ and y.