EIEF WORKING PAPER SERIES

EINAUDI INSTITUTE FOR ECONOMICS AND FINANCE

EIEF Working Paper 06/15
July 2015

Shape Regressions

by
Samantha Leorato
(University of Rome "Tor Vergata")
Franco Peracchi
(Georgetown University,
University of Rome "Tor Vergata" and EIEF)

Shape Regressions*

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

July 10, 2015

Abstract

Learning about the shape of a probability distribution, not just about its location or dispersion, is often an important goal of empirical analysis. Given a continuous random variable Y and a random vector X defined on the same probability space, the conditional distribution function (CDF) and the conditional quantile function (CQF) offer two equivalent ways of describing the shape of the conditional distribution of Y given X. To these equivalent representations correspond two alternative approaches to shape regression. One approach – distribution regression – is based on direct estimation of the conditional distribution function (CDF); the other approach – quantile regression – is instead based on direct estimation of the conditional quantile function (CQF). Since the CDF and the CQF are generalized inverses of each other, indirect estimates of the CQF and the CDF may be obtained by taking the generalized inverse of the direct estimates obtained from either approach, possibly after rearranging to guarantee monotonicity of estimated CDFs and CQFs. The equivalence between the two approaches holds for standard nonparametric estimators in the unconditional case. In the conditional case, when modeling assumptions are introduced to avoid curse-of-dimensionality problems, this equivalence is generally lost as a convenient parametric model for the CDF need not imply a convenient parametric model for the CQF, and vice versa. Despite the vast literature on the quantile regression approach, and the recent attention to the distribution regression approach, no systematic comparison of the two has been carried out yet. Our paper fills-in this gap by comparing the asymptotic properties of estimators obtained from the two approaches, both when the assumed parametric models on which they are based are correctly specified and when they are not.

KEYWORDS: Distribution regression; quantile regression; functional delta-method; non-separable models; influence function.

JEL CLASSIFICATIONS: C1, C21, C25.

*We thank John Rust, Frank Vella, and especially Roger Koenker for useful discussions. The second author acknowledges the hospitality of the Department of Economics at Georgetown University while writing this paper. Corresponding author: Samantha Leorato (samantha.leorato@uniroma2.it).

1 Introduction

Learning about the shape of a probability distribution, not just about its location or dispersion, is often an important goal of empirical analysis. In this paper we consider the relationships between alternative approaches to the problem of shape regression, namely how to estimate the conditional distribution of a continuous random variable Y given a random vector X, when the data are a sample from the joint distribution of (X, Y). If X is discrete and takes only a small number of values, the standard nonparametric approach is to compute either the empirical distribution function or the empirical quantile function of Y for any given value of X. If instead X takes many possible values or is continuous, several nonparametric alternatives are available but all face the same curse-of-dimensionality problem.

A simple way out is local parametric modeling. One approach – distribution regression (DR) – models parametrically the conditional distribution function (CDF) $F(y \mid x) = \Pr\{Y \leq y \mid X = x\}$ at a finite number of cutoff values $-\infty < y_1 < \cdots < y_J < \infty$, and obtains direct estimates of the CDF by fitting a sequence of binary regression models, each corresponding to the conditional mean of the binary indicator $D_j = 1\{Y \leq y_j\}$. This approach, first proposed by Foresi and Peracchi (1995), has recently been considered by Fortin, Lemieux and Firpo (2011), Rothe (2012), Rothe and Wied (2012), and Chernozhukov, Fernández-Val and Melly (2013). Another approach – quantile regression (QR) – models parametrically the conditional quantile function (CQF) $Q(p \mid x) = \inf\{y \in \mathbb{R} \colon F(y \mid x) \geq p\}$ at a finite number of quantile levels $0 < p_1 < \cdots < p_J < 1$, and obtains direct estimates of the CQF by fitting a sequence of quantile regression models. This approach, first proposed by Koenker and Bassett (1978), has been generalized in a variety of directions (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, $Q(F(y \mid x) \mid x) \leq y$ and $F(Q(p \mid x)) \geq p$, which implies that $F(y \mid x) = \inf\{p \colon Q(p \mid x) \geq y\}$ and $Q(p \mid x) = \inf\{y \colon F(y \mid x) \geq p\}$. This nice relationship also holds for standard nonparametric estimates in the unconditional case, as the empirical distribution function and the empirical quantile function are generalized inverses of each other. It is generally lost, however, when modeling assumptions are introduced, since a convenient parametric model for the CDF need not imply an equally convenient parametric model for the CQF, and viceversa. Thus, the statistical properties of a direct estimator of the CDF may be quite different from those of the derived indirect estimators of the CQF. Similarly, the statistical properties of a direct estimator of the CQF may be quite different from those of the derived indirect estimators of the CDF.

Despite the vast literature on the QR approach, and the growing attention to the DR approach, the relationship between the two approaches has not been studied in detail, although some considerations on the choice between them appear in Peracchi (2002), Chernozhukov, Fernández-Val and Melly (2013) and Koenker, Leorato and Peracchi (2013). Our paper fills-in this gap by comparing the performance of estimators obtained from the two approaches, both when the assumed parametric models on which they are based are correctly specified and when they are not.

Of course, when choosing between the two approaches, other aspects may be taken into account. One is the possibility of generalizing to cases when Y is discrete, subject to censoring, or multivariate. In these cases, the DR approach looks more natural. Another is interpretability. Suppose for example that Y is income and X is a vector of socio-economic variables. If one wants to study differences in poverty rates between population subgroups, then the DR approach may be more natural. If instead one wants to compare income differentials within population groups, then the QR approach may be more natural.

We assume throughout the paper that the available data $\{(X_i, Y_i), i = 1, \ldots, n\}$ are a sample from the distribution of the random vector (X, Y) with support $\mathscr{X} \times \mathscr{Y}$, where $\mathscr{X} \subseteq \mathbb{R}^k$ and $\mathscr{Y} \subseteq \mathbb{R}$. Although this assumption is restrictive, our results can easily be generalized to the case of heterogeneous or dependent observations. We also assume that X has finite nonsingular second moment matrix and that the CDF of Y is continuous and strictly increasing in y for any $x \in \mathscr{X}$. This implies that the conditional density of Y exists and is finite and bounded away from zero, and that the CDF and the CQF are inverses of each other. 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 our direct estimators of the CDF and the CQF. Section 3 introduces indirect estimators based on inversion or rearrangement of the direct estimators. Section 4 compares the asymptotic properties (as $n \to \infty$) of the different estimators, both when the model on which they are based is correctly specified and when it is not. Finally, Section 5 summarizes and concludes.

2 Direct estimators

The direct DR approach relies on the fact that, for any fixed cutoff value y, the CDF $F(y \,|\, x)$ is equal to the conditional mean of the binary indicator $D_y = \mathbb{1}\{Y \leq y\}$. Since $F(\cdot \,|\, x)$ takes values in the unit interval, a convenient strategy is to model not $F(y \,|\, x)$ directly, but rather the conditional log-odds function (CLF)

$$\ln \frac{F(y \,|\, x)}{1 - F(y \,|\, x)} = t(y \,|\, x),$$

where $t(\cdot \,|\, x)$ is a function with values on the whole real line. This leaves the range of $t(\cdot \,|\, x)$ completely unrestricted but guarantees that the estimates of the CDF obtained by inverting an estimate of the CLF are bounded between zero and one. An example is the linear location model

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

where α and β are unknown parameters and U is a random error distributed independently of X with a smooth and strictly increasing distribution function G. The CDF for this model is $F(y \mid x) = G(y - \alpha - x^\top \beta)$ and its CLF is $t(y \mid x) = \ln G(y - \alpha - x^\top) - \ln\big[1 - G(y - \alpha - x^\top \beta)\big]$. In particular, if U has the standard logistic distribution function $\Lambda(u) = [1 + \exp(-u)]^{-1}$, then $t(y \mid x) = y - \alpha - x^\top \beta$, so the CLF is linear in the parameters. Notice however that the logistic linear location model has the restrictive feature that $t(y \mid x) - t(y' \mid x) = y - y'$, that is, log-odds corresponding to different cutoff values are at a constant distance from each other. More generally, one may consider a smooth parametric model for the CLF, such as the linear-in-parameter model $t(y \mid x) = \boldsymbol{x}^\top \boldsymbol{\theta}(y)$, where $\boldsymbol{x}$ is a finite-dimensional vector consisting of known transformations of the elements of x and all elements of the parameter vector $\boldsymbol{\theta}(y)$ may depend on the cutoff value y. Although restrictive, linear-in-parameter models are useful because easy to interpret and to estimate. Further, any smooth CLF can be approximated arbitrarily well by a linear-in-parameter model.

An alternative is to directly model the CQF. To fix ideas, consider again the linear location model (1). Its CQF is $Q(p \mid x) = \alpha + x^\top \beta + G^{-1}(p)$, which is also linear in the parameters but has the restrictive feature that conditional quantiles corresponding to two different quantile levels p and p' are at a constant distance $|G^{-1}(p) - G^{-1}(p')|$ from each other. Again, a straightforward generalization is a smooth parametric model for the CQF, such as the linear-in-parameter model $Q(p \mid x) = s(x; \boldsymbol{\theta}(y))$, where all elements of the parameter vector $\boldsymbol{\gamma}(p)$ now may depend on the quantile level p.

2.1 The direct DR estimator

Given a smooth parametric model $h(x; \boldsymbol{\theta}(y))$ for the CLF, the direct DR approach first estimates $\boldsymbol{\theta}(y)$ by maximizing over a finite-dimensional parameter space the average pseudo log-likelihood

$$L_n(\boldsymbol{\theta}; y) = n^{-1} \sum_{i=1}^{n} \big[D_{yi} h(X_i; \boldsymbol{\theta}) - \ln\big(1 + \exp h(X_i; \boldsymbol{\theta})\big)\big],$$

where $D_{yi} = \mathbb{1}\{Y_i \leq y\}$. Given an estimate $\hat{\boldsymbol{\theta}}_n(y)$ of $\boldsymbol{\theta}(y)$, the direct DR estimate of the population CDF at the cutoff value y is $\hat{F}_n^{\ddagger}(y \mid x) = \Lambda\Big(h(x; \hat{\boldsymbol{\theta}}_n(y))\Big)$. Its population analog is denoted by $F^{\ddagger}(y \mid x) = \Lambda\big(h(x; \boldsymbol{\theta}(y))\big)$, where $\boldsymbol{\theta}(y)$ maximizes the expected pseudo log-likelihood $L(\boldsymbol{\theta}; y) = \mathbb{E}\big[D_y\, h(X; \boldsymbol{\theta}) - \ln\big(1 + \exp h(X; \boldsymbol{\theta})\big)\big]$ over the parameter space.

We derive the asymptotic properties of the processes $\sqrt{n}\Big(\hat{\boldsymbol{\theta}}_n(y) - \boldsymbol{\theta}(y)\Big)$ and $\sqrt{n}\big(\hat{F}_n^{\ddagger}(y \mid x) - F^{\ddagger}(y \mid x)\big)$, indexed by y and (y, x) respectively, under the following assumptions:

A.1: The function $h(x;\boldsymbol{\theta})$ is continuously differentiable in x and square integrable for all $\boldsymbol{\theta}$; it is also injective and twice differentiable in $\boldsymbol{\theta}$ for all x with continuous first and second derivatives $\partial_\theta h$ and $\partial_\theta^2 h$ uniformly bounded by square integrable functions.

A.2: 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.

Assumption A.1 guarantees that $L_n(\boldsymbol{\theta}; y)$ is twice differentiable in $\boldsymbol{\theta}$ and that its first and second partial derivatives have finite second moments, while Assumptions A.1–A.2 together imply that the function $\boldsymbol{\theta}(y)$ is continuously differentiable in y, the function $F^{\ddagger}(y \,|\, x)$ is continuously differentiable in both y and x, and the matrix

$$H(y) = \mathbb{E}\left[F^{\ddagger}(y \,|\, X)\left(1 - F^{\ddagger}(y \,|\, X)\right) \partial_\theta h(X)\, \partial_\theta h(X)^\top - \left(D_y - F^{\ddagger}(y \,|\, X)\right) \partial_\theta^2 h(X)\right]$$

is finite and positive definite for all $y \in [\underline{y}, \overline{y}]$, where $\partial_\theta h(X)$ and $\partial_\theta^2 h(X)$ are abbreviations for $\partial_\theta h(X; \boldsymbol{\theta}(y))$ and $\partial_\theta^2 h(X; \boldsymbol{\theta}(y))$.

Theorem 1 *If Assumptions A.1–A.2 hold, then:*

(i) The process $\hat{\boldsymbol{\theta}}_n(\cdot)$ is uniformly consistent for $\boldsymbol{\theta}(\cdot)$, that is, $\sup_{\underline{y} \le y \le \overline{y}} \|\hat{\boldsymbol{\theta}}_n(y) - \boldsymbol{\theta}(y)\| = o_p(1)$.

(ii) 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

$$\Sigma_D(y, y') = \mathbb{E}\left[\left(D_y - F^{\ddagger}(y \,|\, X)\right)\left(D_{y'} - F^{\ddagger}(y' \,|\, X)\right) \partial_\theta h(X)\, \partial_\theta h(X)^\top\right], \quad y \le y'.$$

(iii) 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

$$W(y \,|\, x) = F^{\ddagger}(y \,|\, x)\left(1 - F^{\ddagger}(y \,|\, x)\right) \partial_\theta h(x)^\top H(y)^{-1} B_D(y).$$

Theorem 1 is more general than Theorem 5.2 in Chernozhukov, Fernández-Val and Melly (2013) because does not require the assumed model for the CLF to be correctly specified. It implies that the asymptotic variance of $\hat{F}_n^{\ddagger}(p \,|\, x)$ is equal to $\mathbb{V}(\hat{F}_n^{\ddagger}(p \,|\, x)) = \left[F^{\ddagger}(y \,|\, x)(1 - F^{\ddagger}(y \,|\, x)\right]^2 \partial_\theta h(x)^\top\, \mathbb{V}(\hat{\boldsymbol{\theta}}_n(y))\, \partial_\theta h$, where $\mathbb{V}(\hat{\boldsymbol{\theta}}_n(y)) = H(y)^{-1}\, \Sigma_D(y, y) H(y)^{-1}$ denotes the asymptotic variance of $\hat{\boldsymbol{\theta}}_n(y)$. If the assumed model for the CLF is correctly specified, then $F^{\ddagger}(y \,|\, x) = F(y \,|\, x)$ so $\Sigma_D(y, y) = \mathbb{E}\left[\sigma_y^2(X)\, \partial_\theta h(X)\, \partial_\theta h(X)^\top\right] =$

$H(y)$, where $\sigma_y^2(x) = F(y\,|\,x)\big(1-F(y\,|\,x)\big)$. In this case, the asymptotic variance of $\hat{F}_n^{\ddagger}(y\,|\,x)$ simplifies to $\mathbb{V}\left(\hat{F}_n^{\ddagger}(y\,|\,x)\right) = \left[\sigma_y^2(x)\right]^2 \partial_\theta h(x)^\top H(y)^{-1}\,\partial_\theta h(x)$. In particular, if the linear location model (1) holds, then $\mathbb{V}\left(\hat{F}_n^{\ddagger}(y\,|\,x)\right) = \left[g(\boldsymbol{x}^\top\boldsymbol{\theta}(y))\right]^2 \boldsymbol{x}^\top H(y)^{-1}\boldsymbol{x}$, with $\boldsymbol{x} = (1, x^\top)^\top$ and

$$H(y) = \mathbb{E}\left[\frac{g(\boldsymbol{X}^\top\boldsymbol{\theta}(y))^2}{\sigma_y^2(X)}\boldsymbol{X}\boldsymbol{X}^\top\right],$$

where $g = G'$ denotes the density of U and $\boldsymbol{X} = (1, X^\top)^\top$. Finally, if h is correctly specified as linear in parameter, then $\sigma_y^2(x) = \lambda(\boldsymbol{x}^\top\boldsymbol{\theta}(y))$, where $\lambda = \Lambda'$ denotes the logistic density and $\boldsymbol{x}$ and $\boldsymbol{X}$ now denote vectors of known transformations of the elements of x and X, so $H(y) = \mathbb{E}\left[\lambda(\boldsymbol{X}^\top\boldsymbol{\theta}(y))\boldsymbol{X}\boldsymbol{X}^\top\right]$.

2.2 The direct QR estimator

Given a smooth parametric model $s(x;\boldsymbol{\gamma}(p))$ for the CQF, an estimate $\hat{\boldsymbol{\gamma}}_n(p)$ of $\boldsymbol{\gamma}(p)$ may be obtained by minimizing over a finite-dimensional parameter space the objective function

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

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 $\hat{Q}_n^*(p\,|\,x) = s(x;\hat{\boldsymbol{\gamma}}_n(p))$. Its population analog is denoted by $Q^*(p\,|\,x) = s(x;\boldsymbol{\gamma}(p))$, where $\boldsymbol{\gamma}(p)$ minimizes the population analog $\ell(\boldsymbol{\gamma};p) = \mathbb{E}\,\rho_p\left(Y - s(X;\boldsymbol{\gamma})\right)$ of $\ell_n(\boldsymbol{\gamma};p)$ over the parameter space.

Our next result represents the QR counterpart of Theorem 1. It extends Theorem 3 in Angrist, Chernozhukov and Fernández-Val (2006) to the case of a general misspecified parametric model for the CQF and relies on the following two assumptions:

- B.1: The function $s(x;\boldsymbol{\theta})$ is continuously differentiable in x and square integrable for all $\boldsymbol{\gamma}$; it is also injective and twice differentiable in $\boldsymbol{\gamma}$ for all x with continuous first and second derivatives $\partial_\gamma s$ and $\partial_\gamma^2 s$ uniformly bounded by square integrable functions.

- B.2: 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.

Assumptions B.1–B.2 represent the QR counterparts of A.1–A.2. Together, they imply that the function $\boldsymbol{\gamma}(p)$ is continuously differentiable in p, the function $Q^*(p\,|\,x)$ is continuously differentiable in both p and x, and the matrix

$$J(p) = \mathbb{E}\left[f\left(s(X;\boldsymbol{\gamma}(p))\,|\,X\right)\partial_\gamma s(X)\,\partial_\gamma s(X)^\top - \left(p - \mathbb{1}\{Y \le s(X;\boldsymbol{\gamma}(p))\}\right)\partial_\gamma^2 s(X)\right]$$

is finite and positive definite for all p in the closed interval $[\underline{p}, \overline{p}]$,[1] where $f(y \mid x)$ denotes the conditional density of Y and $\partial_\gamma s(X)$ and $\partial_\gamma^2 s(X)$ are abbreviations for $\partial_\gamma s(X; \boldsymbol{\gamma}(p))$ and $\partial_\gamma^2 s(X; \boldsymbol{\gamma}(p))$.

Theorem 2 *If Assumptions B.1–B.2 hold, then:*

(i) The process $\hat{\boldsymbol{\gamma}}_n(\cdot)$ is uniformly consistent for $\boldsymbol{\gamma}(\cdot)$, that is, $\sup_{\underline{p} \le p \le \overline{p}} \|\hat{\boldsymbol{\gamma}}_n(p) - \boldsymbol{\gamma}(p)\| = o_p(1)$.

(ii) 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

$$\Sigma_Q(p, p') = \mathbb{E}\left[\left(p - \mathbb{1}\{Y < Q^*(p \mid X)\}\right)\left(p' - \mathbb{1}\{Y < Q^*(p' \mid X)\}\right) \partial_\gamma s(X)\, \partial_\gamma s(X)^\top\right], \quad p \le p'.$$

(iii) 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 a zero-mean Gaussian process Z defined as

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

The theorem implies that the asymptotic variance of $\hat{Q}_n^*(p \mid x)$ is $\mathbb{V}\left(\hat{Q}_n^*(p \mid x)\right) = \partial_\gamma s(x)^\top\, \mathbb{V}(\hat{\boldsymbol{\gamma}}_n(p))\, \partial_\gamma s(x)$, where $\mathbb{V}(\hat{\boldsymbol{\gamma}}_n(p)) = J(p)^{-1}\, \Sigma_Q(p, p) J(p)^{-1}$ denotes the asymptotic variance of $\hat{\boldsymbol{\gamma}}_n(p)$. If the assumed model for the CQF is linear in parameters, as in Theorem 3 of Angrist, Chernozhukov and Fernández-Val (2006), then $\partial_\gamma s(x) = \boldsymbol{x}$ and the covariance function of $B_Q(\cdot)$ simplifies to

$$\Sigma_Q(p, p') = \mathbb{E}\left[\left(p - \mathbb{1}\{Y < Q^*(p \mid X)\}\right)\left(p' - \mathbb{1}\{Y < Q^*(p' \mid X)\}\right) \boldsymbol{X}\boldsymbol{X}^\top\right], \quad p \le p',$$

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

3 Indirect estimators

When the assumed CLF is correctly specified, a consistent indirect estimator of the population CQF Q may simply be obtained by taking the generalized inverse of $\hat{F}_n^\ddagger$. When it is misspecified, $\hat{F}_n^\ddagger$ converges to a limit function $F^\ddagger$ that differs from the population CDF F on a subset of $\mathscr{Y} \times \mathscr{X}$ with positive measure. This has two consequences. First, the direct estimator $\hat{F}_n^\ddagger$ is inconsistent for F, so the indirect

[1] Assumption B.2 is also sufficient for conditions C2 and C3 in Oberhofer and Haupt (2015), who study the asymptotic distribution of nonlinear QR estimator in the case of weakly dependent data.

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 taking the generalized inverse of $\hat{F}_n^{\ddagger}$. The same problem arises with the direct QR estimator, since there is no guarantee that Q^* is nondecreasing in p for all x.

One way to guarantee monotonicity is to adopt the rearrangement procedure suggested by Chernozhukov, Fernández-Val and Galichon (2010). This relies on the fact that, given an estimate $\hat{F}_n^{\ddagger}$, not necessarily monotonic, a proper estimate of the CQF is

$$\hat{Q}_n^{+}(p \,|\, x) = \int_0^{\infty} \mathbb{1}\{\hat{F}_n^{\ddagger}(y \,|\, x) \leq p\}\, dy - \int_{-\infty}^{0} \mathbb{1}\{\hat{F}_n^{\ddagger}(y \,|\, x) > p\}\, dy. \tag{2}$$

Similarly, given an estimate $\hat{Q}_n^{*}$, not necessarily monotonic, a proper estimate of the CDF is

$$\hat{F}_n^{\circ}(y \,|\, x) = \int_0^{1} \mathbb{1}\{\hat{Q}_n^{*}(p \,|\, x) \leq y\}\, dp. \tag{3}$$

Thus, taking the generalized inverses of $\hat{Q}_n^{+}$ and $\hat{F}_n^{\circ}$ gives proper estimates $\hat{F}_n^{+}(y \,|\, x) = \inf\{p : \hat{Q}_n^{+}(p \,|\, x) \geq p\}$ and $\hat{Q}_n^{\circ}(p \,|\, x) = \inf\{y : \hat{F}_n^{\circ}(y \,|\, x) \geq p\}$ of the CDF and the CQF respectively. Notice that $\hat{F}_n^{+}$ coincides with $\hat{F}_n^{\ddagger}$ whenever the latter is strictly increasing in y, while $\hat{Q}_n^{\circ}$ coincides with $\hat{Q}_n^{+}$ whenever the latter is strictly increasing in p. Also notice that, unlike the direct approach, rearrangement always produces joint estimates of both the CDF and the CQF.[2]

Rearrangement offers two main advantages. First, the rearranged estimators $\hat{F}_n^{+}$ and $\hat{Q}_n^{\circ}$ are the continuous and Hadamard differentiable inverses of $\hat{Q}_n^{+}$ and $\hat{Q}_n^{*}$ respectively, so their asymptotic properties can be derived via the functional delta method. Second, as shown in Proposition 4 of Chernozhukov, Fernández-Val and Galichon (2010), the rearranged estimators have a smaller bias than the original direct estimators.

3.1 Rearranged DR estimators

To derive the asymptotic properties of $\hat{F}_n^{+}$ and $\hat{Q}_n^{+}$, we make the following additional assumption:

A.3: For all (x, y, p), the equation $F^{\ddagger}(y \,|\, x) = p$ has a finite number $N(p \,|\, x)$ of roots, denoted by $y_j(p \,|\, x)$, $j = 1, \ldots, N(p \,|\, x)$.

[2] Rearrangement is not the only way to guarantee monotonicity, and other alternatives are discussed in Foresi and Peracchi (1996), Hall, Wolff and Yao (1999), Hall and Müller (2003), and Dette and Volgushev (2008) among others.

We denote by F^+ and Q^+ the rearranged versions of $F^\ddagger$. For any $x \in \mathscr{X}$, we also denote by $\mathscr{U}_x^* \subset (0,1)$ the subset of the codomain of $F^\ddagger(\cdot \mid x)$ whose preimage does not contain critical points, namely points where $\partial_y F^\ddagger(y \mid x) = 0$. Further, we define $(0,1)\mathscr{X}^* = \{(p,x) \colon p \in \mathscr{U}_x^*, x \in \mathscr{X}\}$.

Under Assumptions A.1–A.3, it follows by simply adapting the argument in Proposition 5 of Chernozhukov, Fernández-Val and Galichon (2010) that, 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

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

The function $C_W(p \mid x)$ is the Hadamard differential of Q^+ at W, tangentially to the space of continuous functions defined on $\mathscr{Y}\mathscr{X} = \left\{(y,x) \colon \left(F^+(y \mid x), x\right) \in (0,1)\mathscr{X}^*\right\}$. In addition, letting $\mathscr{K}^* = \left\{(y,x) \in [\underline{y}, \overline{y}] \times \mathscr{X} \colon \left(F^+(y \mid x), x\right) \in \mathscr{K}\right\}$, 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

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

If $F^\ddagger$ is strictly increasing in y, then the equation $F^\ddagger(y \mid x) = p$ has a unique root and $F^+(y \mid x) = F^\ddagger(y \mid x)$ for all (x,y,p), so $C_W(p \mid x) = -W\left(y(p \mid x) \,\middle|\, x\right) / \partial_y F^\ddagger\left(y(p \mid x) \,\middle|\, x\right)$ and $D_W(y \mid x) = W(y \mid x)$.

3.2 Rearranged QR estimators

As for the rearranged QR estimators $\hat{F}_n^\circ$ and $\hat{Q}_n^\circ$, we make the following additional assumption:

B.3: For all (x,y,p), the equation $Q^*(p \mid x) = y$ has a finite number $N(y \mid x)$ of roots, denoted by $p_j(y \mid x)$, $j = 1, \ldots, N(y \mid x)$.

We denote by Q° and F° the rearranged versions of Q^*. For any $x \in \mathscr{X}$, we also denote by $\mathscr{Y}_x^* \subset \mathscr{Y}$ the subset of the codomain of $Q^*(\cdot \mid x)$ whose preimage does not contain critical points, namely points where $\partial_p Q^*(p \mid x) = 0$. Further, we define $\mathscr{Y}\mathscr{X}^* = \{(y,x) \colon y \in \mathscr{Y}_x^*, x \in \mathscr{X}\}$.

Under Assumptions B.1–B.3,[3] it follows from Proposition 5 in Chernozhukov, Fernández-Val and Galichon (2010) that, for any compact subset $\mathscr{H} \subset \mathscr{Y}\mathscr{X}^*$, the process $\sqrt{n}\left(\hat{F}_n^\circ(y \mid x) - F^\circ(y \mid x)\right)$, in-

[3] Assumptions B.1–B.3 imply Assumptions 1(a) and 1(b) in Chernozhukov, Fernández-Val and Galichon (2010). In particular, Assumption B.3 is equivalent to their Assumption 1(b), while Assumptions B.1–B.2 imply that the function $Q^*(p \mid x)$ is continuously differentiable in both its arguments, so their Assumption 1(a) is satisfied.

dexed by (y,x), converges weakly on $l^\infty(\mathscr{H})$ to a zero-mean Gaussian process C_Z defined as

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

The function $C_Z(y\,|\,x)$ is the Hadamard differential of F° at Z tangentially to the space of continuous functions defined on $(0,1)\mathscr{X} = \{(p,x)\colon (Q^\circ(p\,|\,x),x) \in \mathscr{Y}\mathscr{X}^*\}$. In addition, letting $\mathscr{H}^* = \{(p,x)\colon (Q^\circ(p\,|\,x),x) \in \mathscr{K}\}$, the process $\sqrt{n}\left(\hat{Q}_n^\circ(p\,|\,x) - Q^\circ(p\,|\,x)\right)$, indexed by (p,x), converges weakly on $l^\infty(\mathscr{H}^*)$ to a zero-mean Gaussian process D_Z defined as

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

If Q^* is strictly increasing in p, then $N(y\,|\,x) = 1$ and $Q^\circ(p\,|\,x) = Q^*(p\,|\,x)$ for all (x,y,p), so $C_Z(y\,|\,x) = -Z\left(p(y\,|\,x)\right)/\partial_p Q^*\left(p(y\,|\,x)\mid x\right)$ and $D_Z(p\,|\,x) = Z(p\,|\,x)$.

4 Asymptotic comparisons

In this section we compare the asymptotic properties of direct and indirect estimators obtained under the DR and the QR approach when the assumed parametric models on which they are based are correctly specified and when they are not.

4.1 Asymptotic relative efficiency

From Theorem 1 and the results in Section 3.1, $\hat{F}_n^\ddagger$ and $\hat{F}_n^+$ are asymptotically equivalent if the assumed model for the CLF is correctly specified. Similarly, from Theorem 2 and the results in Section 3.2, $\hat{Q}_n^*$ and $\hat{Q}_n^\circ$ are asymptotically equivalent if the assumed model for the CQF is correctly specified. If both models are correctly specified then, for any x and all $y \in [\underline{y}, \overline{y}]$ such that $\underline{p} \le F(y\,|\,x) \le \overline{p}$, the asymptotic variances of all estimators considered are linked by the following relationships

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

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

It follows from these relationships that, for all x, $p \in [\underline{p}, \overline{p}]$ and $y \in [\underline{y}, \overline{y}]$ such that $F(y\,|\,x) = p$ and $Q(p\,|\,x) = y$, the asymptotic relative efficiency of $\hat{F}_n^\circ(y\,|\,x)$ to $\hat{F}_n^+(y\,|\,x)$ is

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

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 taking the generalized inverse of $\hat{F}$, relative to the MSE of the empirical quantile function.

Since rearrangement produces estimates of the pair $\big(F(y\,|\,x), Q(p\,|\,x)\big)$, we can also compare the asymptotic performances of the direct estimator $\hat{T}_n = \big(\hat{F}_n^{\ddagger}(y\,|\,x), \hat{Q}_n^{*}(p\,|\,x)\big)$ and the indirect estimator $\tilde{T}_n = \big(\hat{F}_n^{\circ}(y\,|\,x), \hat{Q}_n^{+}(p\,|\,x)\big)$. The next theorem shows that the direct estimator is asymptotically as efficient as the indirect estimator.

Theorem 3 *Let $\hat{F}_n^{\ddagger}(y\,|\,x)$ be a uniformly consistent estimator of a CDF $F(y\,|\,x)$ with a continuous density $f(y\,|\,x)$, and let $\hat{Q}_n^{*}(p\,|\,x)$ be a uniformly consistent estimator of the associated CQF $Q(p\,|\,x)$. Suppose that, for any $x \in \mathscr{X}$, $\sqrt{n}\big(\hat{F}_n^{\ddagger}(\cdot\,|\,x) - F(\cdot\,|\,x)\big)$ converges weakly to a zero-mean Gaussian process defined on a closed interval $[\underline{y}, \overline{y}]$ of $\mathscr{Y}$ and $\sqrt{n}\big(\hat{Q}_n^{*}(\cdot\,|\,x) - Q(\cdot\,|\,x)\big)$ converges weakly to a zero-mean Gaussian process defined on a closed interval $[\underline{p}, \overline{p}]$ of $(0,1)$. Let $\tilde{T}_n = \big(\hat{F}_n^{\circ}(y\,|\,x), \hat{Q}_n^{+}(p\,|\,x)\big)$ be the estimator of $(F(y\,|\,x), Q(p\,|\,x))$ obtained from $\hat{T}_n = \big(\hat{F}_n^{\ddagger}(y\,|\,x), \hat{Q}_n^{*}(p\,|\,x)\big)$ by rearrangement. Then $ARE(\tilde{T}_n, \hat{T}_n) = 1$ for any $x \in \mathscr{X}$ and all $p \in [\underline{p}, \overline{p}]$ and $y \in [\underline{y}, \overline{y}]$ such that $F(y\,|\,x) = p$.*

Although Theorem 3 only holds for values of (x, y, p) that satisfy the relationship $F(y\,|\,x) = p$, it does not depend on the particular choice of direct estimators, provided they are consistent and asymptotically Gaussian.

4.2 Asymptotic bias

If the assumed model for the CLF is misspecified, the DR approach gives inconsistent estimates of the CDF. This is also true for the QR approach if the assumed model for the conditional CQF is misspecified. In such cases, asymptotic comparisons between estimators may be based on their MSE, which is asymptotically dominated by bias.

In this section we consider the case when the assumed models for the CLF and the CQF are linear in parameters but the data are generated from the non-separable model

$$Y = \alpha + X^{\top}\beta + \psi_{\varepsilon}(X, U), \tag{4}$$

where U is a random error distributed independently of X as standard logistic, ε is a scalar and $\psi_\varepsilon(X,U)$ is a term that captures potential misspecification and is equal to U only when $\varepsilon = 0$. Thus, when $\varepsilon = 0$, model (4) reduces to a logistic linear location model, in which case our linear-in-parameter specifications of the CLF and the CQF are both correct.[4] The reason we focus on (4) is that it encompasses several important types of departure from the logistic linear location model (see Section 4.3 below).

We assume that the function $\psi_\varepsilon(x,\cdot)$ is strictly increasing for all $x \in \mathscr{X}$ and $\varepsilon \geq 0$, with inverse $\varphi_\varepsilon(x,\cdot)$. Under this assumption, the CDF and the CQF implied by model (4) are, respectively, $F_\varepsilon(y \,|\, x) = \Lambda\big(\varphi_\varepsilon(x, y-\alpha-x^\top\beta)\big)$ and $Q_\varepsilon(p \,|\, x) = \alpha + x^\top\beta + \psi_\varepsilon\big(x, \Lambda^{-1}(p)\big)$, where $\Lambda^{-1}(p) = \ln p - \ln(1-p)$. Evaluating $F_\varepsilon(y \,|\, x)$ and $Q_\varepsilon(p \,|\, x)$ at $\varepsilon = 0$ returns the CDF $F_0(y \,|\, x) = \Lambda(y-\alpha-x^\top\beta)$ and the CQF $Q_0(p \,|\, x) = \alpha + x^\top\beta + \Lambda^{-1}(p)$ of a logistic linear location model.

It is convenient to associate to the DR estimator $\hat{\boldsymbol{\theta}}_n(y)$ and the QR estimator $\hat{\boldsymbol{\gamma}}_n(p)$, based on linear-in-parameter specifications of the CLF and the CQF, the statistical functionals $\boldsymbol{\theta}_y(\mathfrak{F})$ and $\boldsymbol{\gamma}_p(\mathfrak{F})$ defined on a convex space of distribution functions for (X,Y) which contains all empirical distribution functions $\hat{\mathfrak{F}}_n(x,y) = n^{-1}\sum_{i=1}^n \mathbb{1}\{X_i \leq x, Y_i \leq y\}$. Evaluating these functionals at $\hat{\mathfrak{F}}_n$ returns $\hat{\boldsymbol{\theta}}_n(y)$ and $\hat{\boldsymbol{\gamma}}_n(p)$, while evaluating them at the distribution function $\mathfrak{F}_\varepsilon(x,y) = F_\varepsilon(y \,|\, x)\,F(x)$ of (X,Y) under (4), where $F(x)$ is the marginal distribution function of X, returns the limits in probability of $\hat{\boldsymbol{\theta}}_n(y)$ and $\hat{\boldsymbol{\gamma}}_n(p)$ under (4). We also associate to the direct estimators $\hat{F}_n^\ddagger(y \,|\, x)$ and $\hat{Q}_n^*(p \,|\, x)$ the functionals $T_y^\ddagger(\mathfrak{F} \,|\, x) = \Lambda\big(\boldsymbol{x}^\top\boldsymbol{\theta}_y(\mathfrak{F})\big)$ and $T_p^*(\mathfrak{F} \,|\, x) = \boldsymbol{x}^\top\boldsymbol{\gamma}_p(\mathfrak{F})$, where $\boldsymbol{x} = (1, x^\top)^\top$, which we write more compactly as $T^\ddagger$ and T^*. Evaluating $T^\ddagger$ and T^* at $\hat{\mathfrak{F}}_n$ returns $\hat{F}_n^\ddagger$ and $\hat{Q}_n^*$, evaluating them at $\mathfrak{F}_\varepsilon$ returns $F_\varepsilon^\ddagger$ and Q_ε^*, the limits in probability of $\hat{F}_n^\ddagger$ and $\hat{Q}_n^*$ under (4), while evaluating them at $\mathfrak{F}_0(x,y) = \Lambda(y-\alpha-x^\top\beta)\,F(x)$ returns F_0 and Q_0, the limits in probability of $\hat{F}_n^\ddagger$ and $\hat{Q}_n^*$ under the logistic linear location model.

To illustrate our approach, consider the direct QR estimator $\hat{Q}_n^*(p \,|\, x) = \boldsymbol{x}^\top\hat{\boldsymbol{\gamma}}_n(p)$. Since $\hat{\boldsymbol{\gamma}}_n(p)$ is a regular M-estimator, its associated functional $\boldsymbol{\gamma}_p$ is Gateaux differentiable and its Gateaux differential at $\mathfrak{F}$ in the direction of some other distribution function $\mathfrak{G}$ is a functional $\dot{\boldsymbol{\gamma}}_p(\mathfrak{F};\mathfrak{G}-\mathfrak{F})$, linear in $\mathfrak{G}-\mathfrak{F}$, with the integral representation $\dot{\boldsymbol{\gamma}}_p(\mathfrak{F};\mathfrak{G}-\mathfrak{F}) = \int_{\mathscr{Y}}\int_{\mathscr{X}} \vartheta(x',y';\mathfrak{F})\, d(\mathfrak{G}-\mathfrak{F})(x',y')$, where $\vartheta(x',y';\mathfrak{F})$ is called the influence function associated with $\boldsymbol{\gamma}_p$ (see e.g. Hampel 1974). Since $\hat{Q}_n^*$ is a linear transformation of $\hat{\boldsymbol{\gamma}}_n(p)$, its associated functional T^* is also Gateaux differentiable and its Gateaux differential at $\mathfrak{F}$ in the direction $\mathfrak{G}$ is $\dot{T}^*(\mathfrak{F};\mathfrak{G}-\mathfrak{F}) = \boldsymbol{x}^\top\dot{\boldsymbol{\gamma}}_p(\mathfrak{F};\mathfrak{G}-\mathfrak{F})$. Now suppose that there exists a function $\mathfrak{H}(x,y) = H(y \,|\, x)\,dF(x)$ such that the distribution function of (X,Y) under (4) can be represented, at least for small ε, as $\mathfrak{F}_\varepsilon = \mathfrak{F}_0 + \varepsilon\mathfrak{H}$. Then we can approximate $Q_\varepsilon^* - Q_0 = T^*(\mathfrak{F}_\varepsilon) - T^*(\mathfrak{F}_0)$ by $\varepsilon\,\dot{T}^*(\mathfrak{F}_0;\mathfrak{H})$ with an

[4] An alternative way of modeling misspecification is to allow a fraction ε of the observations to depart from the logistic linear location model. This case is considered in Appendix A.

approximation error of smaller order than ε, so

$$Q^*_\varepsilon - Q_0 = \varepsilon\, \boldsymbol{x}^\top \dot{\boldsymbol{\gamma}}_p(\mathfrak{F}_0;\mathfrak{H}) + o(\varepsilon) = \varepsilon\, \boldsymbol{x}^\top \int_{\mathscr{Y}}\int_{\mathscr{X}} \vartheta(x',y';\mathfrak{F}_0)\, dH(y'\,|\,x')\, dF(x') + o(\varepsilon).$$

A similar approximation holds for the difference $F^\ddagger_\varepsilon - F_0 = T^\ddagger(\mathfrak{F}_\varepsilon) - T^\ddagger(\mathfrak{F}_0)$. We can then approximate the differences $Q^+_\varepsilon - Q_0$ and $F^\circ_\varepsilon - F_0$ by applying the chain rule to the integral transforms (2) and (3) of $T^\ddagger$ and T^* respectively. Finally, we can approximate the differences $F^+_\varepsilon - F_0$ and $Q^\circ_\varepsilon - Q_0$ by repeatedly applying the chain rule to the generalized inverses of the integral transforms of $T^\ddagger$ and T^*.

Our next result collects all these approximations.

Theorem 4 *Suppose that model (4) holds with $\psi_\varepsilon(x,\cdot)$ strictly increasing and differentiable for any $x \in \mathscr{X}$ and all ε. Also suppose that $\Psi(x,u) = \lim_{\varepsilon\downarrow 0}[\psi_\varepsilon(x,u) - u]/\varepsilon$ exists and is square integrable in x, uniformly in $u \in (0,1)$. If Assumptions A.2–A.3 and B.2–B.3 hold then, for any $x \in \mathscr{X}$ and all $y \in [\underline{y},\overline{y}]$ and $p \in [\underline{p},\overline{p}]$,*

$$F^\ddagger_\varepsilon(y\,|\,x) - F_0(y\,|\,x) = F^+_\varepsilon(y\,|\,x) - F_0(y\,|\,x) = -\varepsilon\, \lambda_y(x)\, \boldsymbol{x}^\top \left[\mathbb{E}\, \lambda_y(X) \boldsymbol{X}\boldsymbol{X}^\top\right]^{-1} \left[\mathbb{E}\, \lambda_y(X)\, \Psi_y(X) \boldsymbol{X}\right] + o(\varepsilon),$$
$$Q^+_\varepsilon(p\,|\,x) - Q_0(p\,|\,x) = \varepsilon\, \boldsymbol{x}^\top \left[\mathbb{E}\, \lambda_{Q_0(p|x)}(X) \boldsymbol{X}\boldsymbol{X}^\top\right]^{-1} \left[\mathbb{E}\, \lambda_{Q_0(p|x)}(X)\, \Psi_{Q_0(p|x)}(X) \boldsymbol{X}\right] + o(\varepsilon),$$
$$Q^*_\varepsilon(p\,|\,x) - Q_0(p\,|\,x) = Q^\circ_\varepsilon(y\,|\,x) - Q_0(y\,|\,x) = \varepsilon\, \boldsymbol{x}^\top P_X^{-1} \left[\mathbb{E}\, \Psi(X, \Lambda^{-1}(p)) \boldsymbol{X}\right] + o(\varepsilon),$$
$$F^\circ_\varepsilon(y\,|\,x) - F_0(y\,|\,x) = -\varepsilon\, \lambda_y(x)\, \boldsymbol{x}^\top P_X^{-1} \left[\mathbb{E}\, \Psi(X, y - \alpha - x^\top\beta) \boldsymbol{X}\right] + o(\varepsilon),$$

where $\lambda_y(x) = \lambda(y - \alpha - x^\top\beta)$, $\Psi_y(x) = \Psi(x, y - \alpha - x^\top\beta)$, $Q_0(p\,|\,x) = \alpha + \Lambda^{-1}(p) + x^\top\beta$, and all expectations are with respect to the marginal distribution of X.

Notice that Assumptions A.1 and B.1 are not needed because automatically satisfied by our linear-in-parameter specifications. Also notice that the assumption about $\psi_\varepsilon(x,u)$ implies that, for ε small enough, we can rewrite the data generating process (4) as $Y = \alpha + X^\top\beta + \varepsilon\,\Psi(X,U) + U$. Thus, $\Psi(X,U)$ is the "control function" that should be added to either the logistic linear location model or the linear regression quantile model in order to consistently estimate α and β (see e.g. Blundell, Newey and Vella 2015).

4.3 Bias and MSE under local misspecification

One way of striking a balance between asymptotic precision and bias is to consider a sequence of data generating processes of the form (4) with $\varepsilon = n^{-1/2}$, so the logistic linear location model is misspecified but its degree of misspecification and the sampling uncertainty both vanish asymptotically at exactly the same rate. From Theorem 4 we can derive the local asymptotic bias of estimators based on linear-in-parameter specification of the CLF or the CQF, namely their asymptotic bias under this sequence of data generating processes. We can then compute their local asymptotic MSE as the sum of their asymptotic variance and their squared local asymptotic bias.

Thus, let $T_n = T(\hat{\mathfrak{F}}_n)$ be any of $\hat{F}_n^{\ddagger}$, $\hat{F}_n^{+}$, $\hat{Q}_n^{+}$, $\hat{Q}_n^{*}$, $\hat{Q}_n^{\circ}$ or $\hat{F}_n^{\circ}$, and let τ be the functional that returns either the CDF or the CQF of Y. Notice that $\tau(\mathfrak{F}_\varepsilon) \neq T(\mathfrak{F}_\varepsilon)$, except when $\varepsilon = 0$. The asymptotic local bias of T_n is

$$\mathbb{B}(T_n) = \lim_{\varepsilon \to 0} \frac{T(\mathfrak{F}_\varepsilon) - \tau(\mathfrak{F}_\varepsilon)}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{T(\mathfrak{F}_\varepsilon) - \tau(\mathfrak{F}_0)}{\varepsilon} - \lim_{\varepsilon \to 0} \frac{\tau(\mathfrak{F}_\varepsilon) - \tau(\mathfrak{F}_0)}{\varepsilon},$$

where $\varepsilon = n^{-1/2}$. Theorem 4 provides expressions for $\lim_{\varepsilon \to 0} \big(T(\mathfrak{F}_\varepsilon) - \tau(\mathfrak{F}_0)\big)/\varepsilon$. The assumptions of the theorem also imply that $\lim_{\varepsilon \to 0} \big(\tau(\mathfrak{F}_\varepsilon) - \tau(\mathfrak{F}_0)\big)/\varepsilon = -\lambda_y(x)\,\Psi_y(x)$ if τ is the CDF functional, while $\lim_{\varepsilon \to 0} \big(\tau(\mathfrak{F}_\varepsilon) - \tau(\mathfrak{F}_0)\big)/\varepsilon = \Psi_{Q_0(p|x)}(x)$ if τ is the CQF functional. This gives the following corollary.

Corollary 1 *If the conditions of Theorem 4 hold with $\varepsilon = n^{-1/2}$ then, for any $x \in \mathscr{X}$ and all $y \in [\underline{y}, \overline{y}]$ and $p \in [\underline{p}, \overline{p}]$,*

$$\mathbb{B}\left(\hat{F}_n^{\ddagger}(y\,|\,x)\right) = \mathbb{B}\left(\hat{F}_n^{+}(y\,|\,x)\right) = \lambda_y(x)\Big[\Psi_y(x) - \boldsymbol{x}^\top \left(\mathbb{E}\,\lambda_y(X)\boldsymbol{X}\boldsymbol{X}^\top\right)^{-1}\left(\mathbb{E}\,\lambda_y(X)\,\Psi_y(X)\boldsymbol{X}\right)\Big], \tag{5}$$

$$\mathbb{B}\left(\hat{Q}_n^{+}(p\,|\,x)\right) = -\Big[\Psi_{Q_0(p|x)}(x) - \boldsymbol{x}^\top \left(\mathbb{E}\,\lambda_{Q_0(p|x)}(X)\boldsymbol{X}\boldsymbol{X}^\top\right)^{-1}\left(\mathbb{E}\,\lambda_{Q_0(p|x)}(X)\,\Psi_{Q_0(p|x)}(X)\boldsymbol{X}\right)\Big], \tag{6}$$

$$\mathbb{B}\left(\hat{Q}_n^{*}(p\,|\,x)\right) = \mathbb{B}\left(\hat{Q}_n^{\circ}(p\,|\,x)\right) = -\left[\Psi_{Q_0(p|x)}(x) - \boldsymbol{x}^\top(\mathbb{E}\,\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\left(\mathbb{E}\,\Psi_{Q_0(p|X)}(X)\boldsymbol{X}\right)\right], \tag{7}$$

$$\mathbb{B}\left(\hat{F}_n^{\circ}(y\,|\,x)\right) = \lambda_y(x)\left[\Psi_y(x) - \boldsymbol{x}^\top(\mathbb{E}\,\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\left(\mathbb{E}\,\Psi(X, y - \alpha - x^\top\beta)\boldsymbol{X}\right)\right]. \tag{8}$$

Notice that the term in square brackets in (7) is the error in approximating $\Psi_{Q_0(p|x)}(x) = \Psi(x, \Lambda^{-1}(p))$ using the linear least-squares projection of $\Psi_{Q_0(p|X)}(X) = \Psi(X, \Lambda^{-1}(p))$ on $\boldsymbol{X}$, while the corresponding term in (6) is the error in approximating $\Psi_{Q_0(p|x)}(x)$ using the weighted linear least-squares projection of $\Psi_{Q_0(p|x)}(X) = \Psi(X, \Lambda^{-1}(p) + (x - X)^\top\beta)$ on $\boldsymbol{X}$ with weights equal to $\lambda_{Q_0(p\,|\,x)}(X)$. The term in square brackets in (8) 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 (5) 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 $\lambda_y(X)$.

Figures 1 and 2 illustrate these results for the case when $\alpha = 0$, $\beta = 2\pi$ and X is a single regressor distributed uniformly on the unit interval, so $\boldsymbol{X} = (1, X)^\top$. Figure 1 refers to the direct DR estimator $\hat{F}_n^{\ddagger}$ and the rearranged QR estimator $\hat{F}_n^{\circ}$, while Figure 2 refers to the direct QR estimator $\hat{Q}_n^{*}$ and the rearranged DR estimator $\hat{Q}_n^{+}$, all based on simple linear specifications of the CLF and the CQF. The rows of each figure plot the local asymptotic squared bias, the asymptotic variance and the local asymptotic

MSE of the various estimators, integrated over the distribution of X, under four types of departure from a logistic linear location model. Our asymptotic calculations are in line with the Monte Carlo evidence in Koenker, Leorato and Peracchi (2013) and Leorato and Peracchi (2015).

The first row is for a heteroskedastic model where $\psi_\varepsilon(x,u) = \big(1 + \varepsilon\phi(x)\big)u$ and $\Psi(x,u) = \phi(x)u$, with $\phi(x) = 10x^2$. Under this particular form of heteroskedasticity both approaches lead to asymptotically biased estimators, but the QR approach dominates the DR approach in terms of MSE because of its smaller bias and variance. Since $\mathbb{B}\left(\hat{Q}_n^*(y\,|\,x)\right) = \Lambda^{-1}(p)\big[\boldsymbol{x}^\top P_X^{-1}\left(\mathbb{E}\,\phi(x)\boldsymbol{X}\right) - \phi(x)\big]$, the bias of $\hat{Q}_n^*$ actually vanishes when $p = 1/2$, reflecting the linearity-in-parameters of the conditional median of Y.

The second row is for an omitted variable model where $\psi_\varepsilon(x,u) = \varepsilon\phi(x) + u$ and $\Psi(x,u) = \phi(x)$, with $\phi(x) = 50x^2$. Both approaches again lead to asymptotically biased estimators, but now the smaller bias of DR estimators offsets their larger variance, which results in a smaller MSE than QR estimators. Notice that $\mathbb{B}\left(\hat{Q}_n^*(y\,|\,x)\right) = \boldsymbol{x}^\top P_X^{-1}\left(\mathbb{E}\,\phi(x)\boldsymbol{X}\right) - \phi(x)$ does not depend on p, so the integrated squared bias of $\hat{Q}_n^*$ is the same for all p.

The third row is for the case when the distribution of the error in the linear location model (1) is not logistic. Specifically, we consider a mixture of standard logistic and Student t with 3 degrees of freedom, with weights equal to $1-\varepsilon$ and ε respectively. In this case $\Psi(x,u)$ depends only on u, say $\Psi(x,u) = \zeta(u)$, so $\mathbb{B}\left(\hat{Q}_n^*(p\,|\,x)\right) = \zeta\left(\Lambda^{-1}(p)\right)\left(\boldsymbol{x}^\top P_X^{-1}\mu_X - 1\right)$ and $\mathbb{B}\left(\hat{F}_n^\circ(y\,|\,x)\right) = \lambda_y(x)\,\zeta(y-\alpha-x^\top\beta)\left(1-\boldsymbol{x}^\top P_X^{-1}\mu_X\right)$, where $\mu_X = \mathbb{E}\,\boldsymbol{X}$. The asymptotic bias of the DR estimators $\hat{F}_n^\ddagger$ and $\hat{Q}_n^+$ is surprisingly small. On the other hand, $\boldsymbol{x}^\top P_X^{-1}\mu_X - 1 = 0$ for any $\boldsymbol{x}$, so the QR estimators $\hat{F}_n^\circ$ and $\hat{Q}_n^*$ have no asymptotic bias. Since they also have smaller asymptotic variance than $\hat{F}_n^\ddagger$ and $\hat{Q}_n^+$, the QR approach clearly dominates.

The fourth row is for a monotonic transformation model of the form $Y^{(1-\varepsilon)} + 1 = \alpha + \beta X + U$, where $Y^{(1-\varepsilon)}$ is the Box-Cox transform of Y, that is, $Y^{(1-\varepsilon)} = (Y^{1-\varepsilon} - 1)/(1-\varepsilon)$ for $\varepsilon \neq 1$. In this case, $\varphi_\varepsilon(x,u) - u = [1 + (1-\varepsilon)(\alpha + x^\top\beta + u - 1)]^{1/(1-\varepsilon)} - \alpha - x^\top\beta - u$, so $\Psi(x,u) = \phi(\alpha + x^\top\beta + u)$ with $\phi(z) = z\ln(z) + 1 - z$. This is a case where the DR estimators are not affected by asymptotic bias.

Notice that, locally near the linear location model (1), QR estimators have smaller integrated asymptotic variance than DR estimators, as already pointed out by Koenker, Leorato and Peracchi (2013). So, the main advantage of the DR approach is the lower integrated asymptotic bias, at least in some cases.

5 Conclusions

If the assumed parametric models for the CLF and the CQF are correctly specified, then the relative efficiency of the DR and the QR approach in estimating the population CDF F is asymptotically the same as their relative efficiency in estimating the population CQF Q. Further, the direct estimator of the pair (F, Q) is asymptotically as efficient as the indirect estimator based on rearrangement.

If the assumed models for the CLF and the CQF are incorrectly specified as linear in parameters, then the relative performance of the various estimators may be judged by their asymptotic local MSE, which depends on both asymptotic variance and asymptotic local bias. In general, estimators obtained from the QR approach have smaller integrated asymptotic variance than those obtained from the DR approach. Their local asymptotic bias under the relatively general non-separable model (4) depends instead on how well the regressors in X help predict the value of the "control function" $\Psi(x, u)$ in the asymptotically equivalent representation of the data generating process as $Y = \alpha + X^\top \beta + \varepsilon \Psi(X, U) + U$, where U is standard logistic error. In the case of a linear location model with non-logistic errors QR estimators have no asymptotic bias, while in the case of a monotonic transformation model DR estimators have no asymptotic bias. Under heteroskedasticity or omitted variables, the relative bias of the various estimators depends instead on the precise nature of model misspecification.

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.

Blundell R., Newey W.K., and Vella F. (2015). Control functions. Mimeo.

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.

Horowitz J.L., Manski C.F. (1995). Identification and robustness with contaminated and corrupted data. *Econometrica*, 63: 281–302.

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.

Leorato S., and Peracchi F. (2015). Comparing distribution and quantile regression. Mimeo.

Oberhofer W. and Haupt H. (2015). Asymptotic theory for nonlinear quantile regression under weak dependence. *Econometric Theory*, available on CJO2015. doi:10.1017/S0266466615000031.

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.

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

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

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

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

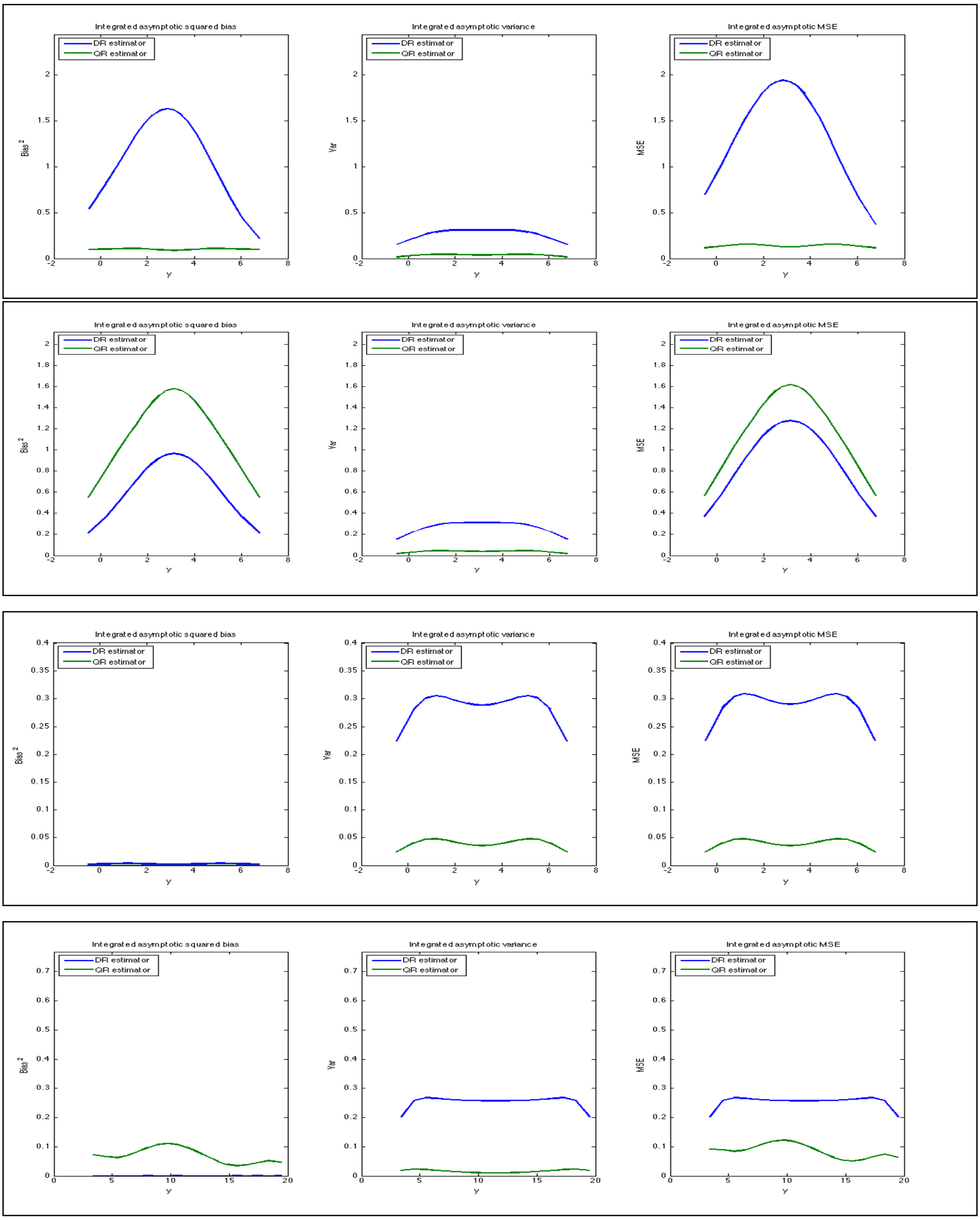

Figure 1: Integrated local asymptotic squared bias, integrated asymptotic variance and integrated local asymptotic MSE of the direct DR estimator $\hat{F}_n^{\ddagger}$ and the rearranged QR estimator $\hat{F}_n^{\circ}$ under model (4) with a single covariate $X \sim \mathscr{U}(0,1)$, $\alpha = 0$, $\beta = 2\pi$ and standard logistic error. First row: heteroskedasticity, $\Psi(x,u) = (1+10x^2)u$. Second row: omitted variables, $\Psi(x,u) = 50x^2$. Third row: non-logistic error, $\Psi(x,u) \approx (G(u) - \Lambda(u))/\Lambda'(u)$ and $G \sim t_3$. Fourth row: Box-Cox transform, $\alpha = 2$, $\beta = 6\pi$ and $\Psi(x,u) = \ln(\alpha + \beta x + u)(\alpha + \beta x + u)(1 - \alpha - \beta x - u)$.

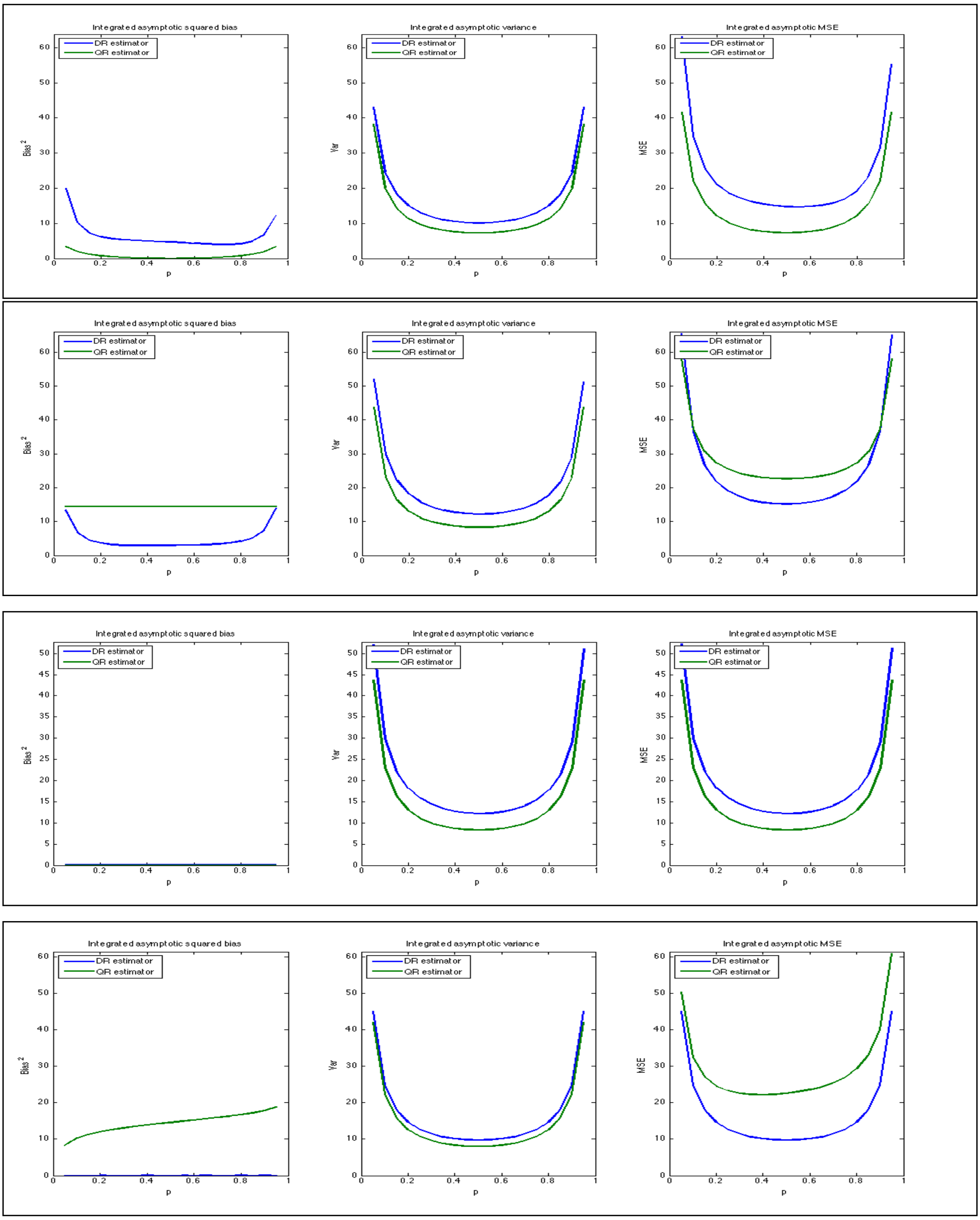

Figure 2: Integrated local asymptotic squared bias, integrated asymptotic variance and integrated local asymptotic MSE of the direct QR estimator $\hat{Q}_n^*$ and the rearranged DR estimator $\hat{Q}_n^+$ under model (4) with a single covariate $X \sim \mathscr{U}(0,1)$, $\alpha = 0$, $\beta = 2\pi$ and standard logistic error. First row: heteroskedasticity, $\Psi(x,u) = (1+10x^2)u$. Second row: omitted variables, $\Psi(x,u) = 50x^2$. Third row: non-logistic error, $\Psi(x,u) \approx (G(u) - \Lambda(u))/\Lambda'(u)$ with $G \sim t_3$. Fourth row: row: Box-Cox transform, $\alpha = 2$, $\beta = 6\pi$ and $\Psi(x,u) = \ln(\alpha + \beta x + u)(\alpha + \beta x + u)(1 - \alpha - \beta x - u)$.

A Model contamination

In this section we study the asymptotic bias of QR and DR estimators under the following contamination model

$$Y = \alpha + X^\top \beta + U + D\,\Psi(X,U), \tag{9}$$

where U is a random error distributed independently of X as standard logistic and D is a Bernoulli random variable distributed independently of X and U with probability of success $0 \le \varepsilon < 1$. This model differs from that in Section 4.2 because now the linear location model (1) is misspecified only for a fraction ε of the observations drawn from a contaminating distribution (see e.g. Horowitz and Manski 1995).

Using the same notation of Section 4.2, we denote the joint distribution function of (X,Y) under model (9) by $\mathfrak{F}_\varepsilon(x,y) = F(x)\,F_\varepsilon(y\,|\,x)$, where

$$F_\varepsilon(y\,|\,x) = (1-\varepsilon)\,\Lambda(y-\alpha-x^\top\beta) + \varepsilon\,\Pr\{U + \Psi(x,U) \le y-\alpha-x^\top\beta\}$$

denotes the CDF of Y. We also denote by $F_\varepsilon^\ddagger$, F_ε^+, Q_ε^+, Q_ε^*, Q_ε° and F_ε° the limits in probability of $\hat{F}_n^\ddagger$, $\hat{F}_n^+$, $\hat{Q}_n^+$, $\hat{Q}_n^*$, $\hat{Q}_n^\circ$ and $\hat{F}_n^\circ$ under model (9), and by F_0 and Q_0 the CDF and the CQF of the logistic linear location model corresponding to $\varepsilon = 0$. Then we have the following counterpart of Theorem 4.

Theorem 5 *Suppose that model (9) holds with D distributed as Bernoulli with parameter ε and that the function $\psi(x,u)\colon u \mapsto u + \Psi(x,u)$ is strictly increasing in u for all $x \in \mathscr{X}$ with inverse $\varphi(x,\cdot)$. If Assumptions A.1–A.3 and B.1–B.3 hold then, for any $x \in \mathscr{X}$ and all $p \in [\underline{p},\overline{p}]$ and $y \in [\underline{y},\overline{y}]$,*

$$\begin{aligned}
F_\varepsilon^\ddagger(y\,|\,x) - F_0(y\,|\,x) &= F_\varepsilon^+(y\,|\,x) - F_0(y\,|\,x) = \\
&= -\varepsilon\,\lambda_y(x)\,\boldsymbol{x}^\top \left[\mathbb{E}\,\lambda_y(X)\boldsymbol{X}\boldsymbol{X}^\top\right]^{-1} \left(\mathbb{E}\,\Delta(X, y-\alpha-X^\top\beta)\boldsymbol{X}\right) + o(\varepsilon), \\
Q_\varepsilon^+(p\,|\,x) - Q_0(p\,|\,x) &= \varepsilon\,\boldsymbol{x}^\top \left[\mathbb{E}\,\lambda_{Q_0(p\,|\,x)}(X)\boldsymbol{X}\boldsymbol{X}^\top\right]^{-1} \left(\mathbb{E}\,\Delta(X, Q_0(p\,|\,x)-X^\top\beta)\boldsymbol{X}\right) + o(\varepsilon), \\
Q_\varepsilon^*(p\,|\,x) - Q_0(p\,|\,x) &= Q^\circ(y\,|\,x) - Q_0(y\,|\,x) = \\
&= \varepsilon\,[p(1-p)]^{-1}\boldsymbol{x}^\top P_X \left(\mathbb{E}\,\Delta(X,\Lambda^{-1}(p))\boldsymbol{X}\right) + o(\varepsilon), \\
F_\varepsilon^\circ(y\,|\,x) - F_0(y\,|\,x) &= -\varepsilon\,\boldsymbol{x}^\top P_X \left(\mathbb{E}\,\Delta(X, y-\alpha-x^\top\beta)\boldsymbol{X}\right) + o(\varepsilon),
\end{aligned}$$

where $Q_0(p\,|\,x) = \alpha + x^\top\beta + \Lambda^{-1}(p)$, $\Delta(x,u) = \Lambda(u) - \Lambda(\varphi(x,u))$.

Using Theorem 5 we immediately obtain the counterpart of Corollary 1, namely expressions for the asymptotic bias of estimators based on linear-in-parameter specifications of the CLF and the CQF under a sequence of data generating processes of the form (9) with $\Pr\{D=1\} = \varepsilon = n^{-1/2}$.

B Proofs

Theorem 1

Part (i) establishes uniform consistency of the process $y \mapsto \hat{\boldsymbol{\theta}}_n(y)$; part (ii) establishes asymptotic normality of $\sqrt{n}[\hat{\boldsymbol{\theta}}_n(y) - \boldsymbol{\theta}(y)]$ by a Bahadur representation; finally, part (iii) establishes convergence of the process $\sqrt{n}(\hat{F}_n^{\ddagger} - F^{\ddagger})$ by a Taylor expansion argument.

(i) Uniform consistency. Assumption A.1 implies that h is Lipschitz with a square integrable Lipschitz bound $K(x)$. Since $\mathbb{E}\,|h(X;\boldsymbol{\theta})|^2 < \infty$, then also $\mathbb{E}\left|D_y h(X;\boldsymbol{\theta}) - \ln\left(1+e^{h(X;\boldsymbol{\theta})}\right)\right| < \infty$ because $\ln(1+x) \leq \max(\ln 2, \ln(2x))$. Next we show that $L_n(\boldsymbol{\theta};y) = L(\boldsymbol{\theta};y) + o_p(1)$ uniformly in $(\boldsymbol{\theta}, y) \in \Theta \times \mathscr{Y}$. This is because $L_n(\boldsymbol{\theta};y) = L(\boldsymbol{\theta};y) + o_p(1)$ for all fixed $(\boldsymbol{\theta}, y)$, by Khintchine law of large numbers, and the process $(\boldsymbol{\theta}, y) \mapsto L_n(\boldsymbol{\theta};y)$ is stochastically equicontinuous. In fact, for every $(\boldsymbol{\theta}, y)$ and $(\boldsymbol{\theta}', y')$ we have

$$
\begin{aligned}
\left|L_n(\boldsymbol{\theta};y) - L_n(\boldsymbol{\theta}';y')\right| &= \left|\frac{1}{n}\sum_i \left[(D_{yi} - D_{y',i})h(X_i;\boldsymbol{\theta}') - D_{yi}\left(h(X_i;\boldsymbol{\theta}') - h(X_i;\boldsymbol{\theta})\right) + \ln\frac{1+e^{h(X_i;\boldsymbol{\theta})}}{1+e^{h(X_i;\boldsymbol{\theta}')}}\right]\right| \\
&= \left|\mathbb{E}\left((D_y - D_{y'})h(X;\boldsymbol{\theta}')\right) - \mathbb{E}\left(D_y\left(h(X;\boldsymbol{\theta}') - h(X;\boldsymbol{\theta})\right)\right) + \mathbb{E}\ln\frac{1+e^{h(X;\boldsymbol{\theta})}}{1+e^{h(X;\boldsymbol{\theta}')}} + o_p(1)\right| \\
&\leq \left(\left|\mathbb{E}\left((F(y\,|\,X) - F(y'\,|\,X))h(X;\boldsymbol{\theta}')\right)\right| + 2\left|\mathbb{E}\left((\boldsymbol{\theta}-\boldsymbol{\theta}')^\top K(X) + O_p(\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|^2)\right)\right|\right)(1+o_p(1)),
\end{aligned}
$$

where the last inequality follows from the fact that h is twice differentiable and

$$
\ln\frac{1+e^{h(X;\boldsymbol{\theta})}}{1+e^{h(X;\boldsymbol{\theta}')}} \leq (\boldsymbol{\theta}-\boldsymbol{\theta}')^\top \partial_\theta h(X;\boldsymbol{\theta}) + O_p\left(\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|^2\right).
$$

In turn, this implies stochastic equicontinuity of the process $(\boldsymbol{\theta}, y) \mapsto L_n(\boldsymbol{\theta};y)$ because

$$
\begin{aligned}
&\left|L_n(\boldsymbol{\theta};y) - L_n(\boldsymbol{\theta}';y')\right| \leq \\
&\leq \left(\sup_x |F(y\,|\,x) - F(y'\,|\,x)|\,\mathbb{E}\,|K(X)| + \|\boldsymbol{\theta}-\boldsymbol{\theta}'\|_\infty\,\mathbb{E}\,|K(X)| + O\left(\mathbb{E}\,|(\boldsymbol{\theta}'-\boldsymbol{\theta})^\top K(X)|^2\right)\right)(1+o_p(1)).
\end{aligned}
$$

Convexity of L in h and Assumption A.1 then imply that, uniformly in y, $L(\boldsymbol{\theta}(y);y) > \sup_{\theta\notin\mathscr{B}} L(\boldsymbol{\theta};y)$, where $\mathscr{B}$ is an open subset containing $\boldsymbol{\theta}(y)$. It follows from Corollary 3.2.3 of van der Vaart and Wellner (1996) that $\|\boldsymbol{\theta}(y) - \hat{\boldsymbol{\theta}}_n(y)\| \to 0$ in (outer) probability, uniformly in y.

(ii) Asymptotic normality. Assumption A.1 implies that h is continuously differentiable in $\boldsymbol{\theta}$ with square integrable Lipschitz bound $K(x) \geq \partial_\theta h$. Then, the class $\mathscr{H} = \{h(x;\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta\}$ satisfies the bracketing inequality $N_{[]}(2\varepsilon\|K\|, \mathscr{H}, \|\cdot\|) \leq N(\varepsilon, \Theta, d)$ for any norm $\|\cdot\|$, where $N_{[]}(\varepsilon, \mathscr{F}, \|\|)$ is the minimum number of ε-brackets and $N(\varepsilon, \mathscr{F}, \|\|)$ is the number of ε-balls with which $\mathscr{F}$ can be covered (van der Vaart and Wellner 1996, Theorem 2.7.11). Moreover, because the parameter space Θ is a bounded

subset of a finite-dimensional Euclidean space, the right-hand-side of the inequality is bounded above by $C(1/\varepsilon)^2$ if d is the Euclidean distance. It follows from Theorem 2.5.2 in van der Vaart and Wellner (1996) that $\mathscr{H}$ is a Donsker class. Further, by the permanence of the Donsker property, both $\Lambda(\mathscr{H}) = \left\{e^h/(1+e^h),\ h \in \mathscr{H}\right\}$ and $\mathscr{D} = \{\mathbb{1}\{Y \le y\},\ y \in \mathscr{Y}\}$ are bounded Donsker classes, and so is $\{\mathscr{D} - \Lambda(\mathscr{H})\}$. The class $\mathscr{G} = \{(\mathscr{D} - \Lambda(\mathscr{H}))\,\mathscr{H}'\} = \{g(\boldsymbol{\theta}, y),\ (\boldsymbol{\theta}, y) \in \Theta \times \mathscr{Y}\}$ is also Donsker, where $\mathscr{H}'$ is the class of partial derivatives $\mathscr{H}' = \{\partial_\theta h(x;\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\} \subseteq \prod_j \mathscr{H}'_j$ and every $\mathscr{H}'_j$ is Donsker.

Given the vector valued function $g(\boldsymbol{\theta}, y) = \left[D_y - \left(1+\exp(-h(x;\boldsymbol{\theta}))\right)^{-1}\right]\partial_\theta h(x;\boldsymbol{\theta})$, it follows that the mapping $(\boldsymbol{\theta}, y) \mapsto n^{-1/2}\sum_i [g(\boldsymbol{\theta}, y) - \mathbb{E}\,g(\boldsymbol{\theta}, y)]$, is stochastically equicontinuous over $\Theta \times \mathscr{Y}$ with respect to the pseudometric $d\left((\boldsymbol{\theta}, y), (\boldsymbol{\theta}', y')\right)^2 = \max_j \mathbb{E}\left[(g_j(\boldsymbol{\theta}, y) - g_j(\boldsymbol{\theta}', y'))\right]^2$, where g_j is the jth component of g. Further, since $\sup_y \|\hat{\boldsymbol{\theta}}_n(y) - \boldsymbol{\theta}(y)\| = o_p(1)$ and each g_j is Lipschitz in $\boldsymbol{\theta}$, we have that $\sup_y d\left((\boldsymbol{\theta}(y), y), (\hat{\boldsymbol{\theta}}_n(y), y)\right)^2 = o_p(1)$ in outer measure. We conclude that

$$\mathbb{U}_n\left[g(\hat{\boldsymbol{\theta}}_n(y), y)\right] = \mathbb{U}_n\left[g(\boldsymbol{\theta}(y), y)\right] + o_p(1)$$

in $l^\infty(\mathbb{R})$, where $\mathbb{U}_n$ denotes the empirical process, i.e. $\mathbb{U}_n[f] = \sqrt{n}\left[n^{-1}\sum_i f(X_i) - \mathbb{E}\,f(X)\right]$.

A Taylor expansion of $\mathbb{E}[g(\boldsymbol{\theta}, y)]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n(y)}$ around $\boldsymbol{\theta}(y)$ gives

$$\begin{aligned}\mathbb{E}\left[g(\boldsymbol{\theta}, y)\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n(y)} &= \mathbb{E}\left[g(\boldsymbol{\theta}(y), y)\right] + \frac{\partial}{\partial \boldsymbol{\theta}}\left\{\mathbb{E}\left[g(\boldsymbol{\theta}, y)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}(y)}\right\}\left(\hat{\boldsymbol{\theta}}_n(y) - \boldsymbol{\theta}(y)\right) + o_p\left(\|\hat{\boldsymbol{\theta}}_n(y) - \boldsymbol{\theta}(y)\|\right)\\ &= \left[-H(y) + o_p(1)\right]\left(\hat{\boldsymbol{\theta}}_n(y) - \boldsymbol{\theta}(y)\right).\end{aligned} \tag{10}$$

The first order condition for $\hat{\boldsymbol{\theta}}_n(y)$ implies that

$$\sqrt{n}\,\mathbb{E}_n\,g(\hat{\boldsymbol{\theta}}_n(y), y) = n^{-1/2}\sum_{i=1}^n\left(D_{i,y} - \frac{1}{1+\exp(-h(X_i;\boldsymbol{\theta}))}\right)\partial_\theta h(X_i;\boldsymbol{\theta}) \le o_p(1).$$

Multiplying (10) by $\sqrt{n}$ and subtracting $\sqrt{n}\,\mathbb{E}_n\,g(\hat{\boldsymbol{\theta}}_n(y), y))$ gives

$$o_p(1) = \left[H(y) + o_p(1)\right]\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n(y) - \boldsymbol{\theta}(y)\right) + \mathbb{U}_n\left[g(\boldsymbol{\theta}(y), y)\right] + o_p(1).$$

Positive definiteness of $H(y)$ then implies $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n(y) - \boldsymbol{\theta}(y)\right) = -\mathbb{U}_n\left[H(y)^{-1}g(\boldsymbol{\theta}(y), y)\right] + o_p(1)$.

(iii) Weak convergence of $\hat{F}^{\ddagger}_n$. Follows from (ii) by the application of the continuous mapping theorem and the continuity of $F^{\ddagger}(y\,|\,x)$ in both arguments. In particular, by (ii), for every $x \in \mathscr{X}$,

$$\begin{aligned}\sqrt{n}\left(h(x;\hat{\boldsymbol{\theta}}_n(y)) - h(x;\boldsymbol{\theta}(y))\right) &= \sqrt{n}\,\partial_\theta h(x;\boldsymbol{\theta}(y))^\top\left(\hat{\boldsymbol{\theta}}_n(y) - \boldsymbol{\theta}(y)\right) + o_p\left(\sqrt{n}\|\hat{\boldsymbol{\theta}}_n(y) - \boldsymbol{\theta}(y)\|^2\right)\\ &\quad - \mathbb{U}_n\left[\partial_\theta h(x;\boldsymbol{\theta}(y))^\top H(y)^{-1}\,g(\boldsymbol{\theta}(y), y)\right] + o_p(1).\end{aligned}$$

A Taylor expansion of $\hat{F}^{\ddagger}_n(y\,|\,x) = \left(1+\exp(-h(x;\hat{\boldsymbol{\theta}}_n(y)))\right)^{-1}$ around $h(x;\boldsymbol{\theta}(y))$ gives

$$\hat{F}^{\ddagger}_n(y\,|\,x) - F^{\ddagger}(y\,|\,x) = F^{\ddagger}(y\,|\,x)(1 - F^{\ddagger}(y\,|\,x))\left(h(x;\hat{\boldsymbol{\theta}}_n(y)) - h(x;\boldsymbol{\theta}(y))\right) + o_p(n^{-1/2}),$$

so $\sqrt{n}\left(\hat{F}^{\ddagger}_n(y\,|\,x) - F^{\ddagger}(y\,|\,x)\right) = F^{\ddagger}(y\,|\,x)(1 - F^{\ddagger}(y\,|\,x))\,\mathbb{U}_n\left[\partial_\theta h(x;\boldsymbol{\theta}(y))^\top H(y)^{-1}\,g(\boldsymbol{\theta}(y), y)\right] + o_p(1)$.

Theorem 2

The proof follows the same steps of the proof of Theorem 1, as they are both M-estimators.

Theorem 3

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

The functional delta method applied to the functionals $\hat{Q}_n^{+} = \tau(\hat{F}_n^{\ddagger}) = \hat{F}_n^{\ddagger -1}$ and $\hat{F}_n^{\circ} = \tau(\hat{Q}_n^{*}) = \hat{Q}_n^{*-1}$ gives

$$\sqrt{n}\left(\tau(\hat{F}_n^{\ddagger}) - \tau(F)\right) = \sqrt{n}(\hat{Q}_n^{+} - Q) \Rightarrow -\frac{Z_F(F^{-1})}{F'(F^{-1})},$$

with $\Rightarrow$ denoting weak convergence, and

$$\sqrt{n}\left(\tau(\hat{Q}_n^{*}) - \tau(Q)\right) = \sqrt{n}(\hat{F}_n^{\circ} - F) \Rightarrow -\frac{Z_Q(Q^{-1})}{Q'(Q^{-1})} = -Z_Q(F)\, f(Q(Q^{-1})),$$

which implies that $V(\hat{Q}_n^{+}) = \sigma_F^2(F^{-1})/F'(F^{-1})^2$ and $V(\hat{F}_n^{\circ}(y)) = f^2(y)\,\sigma_Q^2(F(y))$. Setting $p = F(y)$, we then have the following relationship between the determinants of the asymptotic variances of $\hat{T}$ and $\tilde{T}$

$$\begin{aligned}
|V(\hat{T})(y)| &= V\left(\hat{F}_n^{\ddagger}(y)\right) V\left(\hat{Q}_n^{*}(F(y))\right) - C\left(\hat{F}_n^{\ddagger}(y), \hat{Q}_n^{*}(F(y))\right)^2 \\
&= \sigma_F^2(y)\,\sigma_Q^2(F(y)) - C\left(Z_F(y), Z_Q(F(y))\right) \\
&= \frac{\sigma_F^2(y)}{f^2(y)} V\left(\sigma_Q^2(F(y))\right) f^2(y) - C\left(\frac{Z_F(y)}{f(y)}, f(y)\, Z_Q(F(y))\right) = |V(\tilde{T})(y)|.
\end{aligned}$$

Theorem 4

Recall that, given a regular M-estimator T, its Gateaux differential at $\mathfrak{F}$ in some direction $\mathfrak{G} = \mathfrak{H} + \mathfrak{F}$ is $\dot{T}(\mathfrak{F}; \mathfrak{H}) = \int_{\mathcal{X}} \int_{\mathcal{Y}} \vartheta(x, y, \mathfrak{F})\, d\mathfrak{H}$, where $\vartheta(x, y, \mathfrak{F}_0)$ is the influence function of the estimator (see e.g. Serfling 1980, Chapter 7). Thus, the Gateaux differential for the DR functional $\boldsymbol{\theta}_y$ is

$$\dot{\boldsymbol{\theta}}_y(\mathfrak{F}; \mathfrak{H}) = \left[\mathbb{E}\, \Lambda'(y - \alpha - X^\top \beta) \boldsymbol{X}\boldsymbol{X}^\top\right]^{-1} \int_{\mathcal{X}} \int_{\mathcal{Y}} \left[\mathbb{1}\{y' \leq y\} - \Lambda(y - \alpha - x^\top \beta)\}\right] \boldsymbol{x}\, d\mathfrak{H}(x, y'), \quad (11)$$

while for the QR functional $\boldsymbol{\gamma}_p$ is

$$\dot{\boldsymbol{\gamma}}_p(\mathfrak{F}; \mathfrak{H}) = \frac{1}{p(1-p)} P_X^{-1} \int_{\mathcal{X}} \int_{\mathcal{Y}} \left[p - \mathbb{1}\{y \leq \alpha + x^\top \beta + \Lambda^{-1}(p)\}\right] \boldsymbol{x}\, d\mathfrak{H}(x, y). \quad (12)$$

Both differentials are continuous, in (x, y) and (x, p) respectively.

The joint distribution $\mathfrak{F}_\varepsilon$ of (X, Y) under model (4) cannot be written as a convex combination of two fixed distributions $\mathfrak{F}_0$ and $\mathfrak{G}$, because $\mathfrak{F}_\varepsilon = \mathfrak{F}_0 + \varepsilon\mathfrak{H}_\varepsilon$, with $\mathfrak{H}_\varepsilon(x, y) = F(x)H_\varepsilon(y \mid x)$ and

$$H_\varepsilon(y \mid x) = \frac{\Lambda\left(\psi_\varepsilon^{-1}(x, y-\alpha-x^\top\beta)\right) - \Lambda(y-\alpha-x^\top\beta)}{\varepsilon}.$$

However, the existence of $\Psi(x, u) = \lim_{\varepsilon\downarrow 0}[\psi_\varepsilon(x, u) - u]/\varepsilon$ implies that, for ε small enough, $\mathfrak{F}_\varepsilon = \mathfrak{F}_0 + \varepsilon\mathfrak{H} + o(\varepsilon)$, with $\mathfrak{H}(x, y) = F(x)H(y \mid x)$ and

$$H(y \mid x) = \lim_{\varepsilon\to 0} H_\varepsilon(y \mid x) = -\Lambda'(y-\alpha-x^\top\beta)\,\Psi(x, y-\alpha-x^\top\beta).$$

The approximation error $\mathfrak{F}_\varepsilon - \mathfrak{F}_0 - \varepsilon\mathfrak{H} = \varepsilon(\mathfrak{H}_\varepsilon - \mathfrak{H})$ is continuos and differentiable, as ψ_ε is differentiable for all ε and $\|H_\varepsilon - H\| = o(1)$. Thus, using linearity of the Gateaux differential in its second argument, we can write

$$\begin{aligned}
\dot{\boldsymbol{\gamma}}_p(\mathfrak{F}_0; \mathfrak{H}_\varepsilon) &= \dot{\boldsymbol{\gamma}}_p(\mathfrak{F}_0; \mathfrak{H}) + \dot{\boldsymbol{\gamma}}_p(\mathfrak{F}_0; \mathfrak{H}_\varepsilon - \mathfrak{H}) \\
&= \int_{\mathscr{X}}\int_{\mathscr{Y}} \vartheta(x, y, \mathfrak{F}_0)\, d\mathfrak{H} + \dot{\boldsymbol{\gamma}}_p(\mathfrak{F}_0; \mathfrak{H}_\varepsilon - \mathfrak{H}) \\
&= \frac{1}{p(1-p)} P_X^{-1} \int_{\mathscr{X}}\int_{\mathscr{Y}} \left[p - \mathbb{1}\{y \le \alpha + x^\top\beta + \Lambda^{-1}(p)\}\right] \boldsymbol{x}\, dH(y \mid x)\, dF(x) + o(1).
\end{aligned}$$

Therefore, under model (4),

$$\begin{aligned}
Q^*(p \mid x) - Q_0(p \mid x) &= T^*(\mathfrak{F}_\varepsilon) - T^*(\mathfrak{F}_0) = x^\top \dot{\boldsymbol{\gamma}}_p(\mathfrak{F}_0; \varepsilon\mathfrak{H}_\varepsilon) + o(\varepsilon) \\
&= \varepsilon\,\boldsymbol{x}^\top \dot{\boldsymbol{\gamma}}_p(\mathfrak{F}_0; \mathfrak{H}) + o(\varepsilon) = \varepsilon\,\boldsymbol{x}^\top P_X^{-1}\left[\mathbb{E}\,\Psi\left(X, \Lambda^{-1}(p)\right)\boldsymbol{X}\right] + o(\varepsilon).
\end{aligned}$$

Repeating the argument for the DR functional gives

$$\dot{\boldsymbol{\theta}}_y(\mathfrak{F}_0; \mathfrak{H}_\varepsilon) = -\left[\mathbb{E}\,\Lambda'(y-\alpha-X^\top\beta)\boldsymbol{X}\boldsymbol{X}^\top\right]^{-1}\left[\mathbb{E}\,\Lambda'(y-\alpha-X^\top\beta)\Psi(X, y-\alpha-X^\top\beta)\boldsymbol{X}\right] + o(1).$$

Thus, by applying the chain rule,

$$\begin{aligned}
F^\ddagger(y \mid x) - F_0(y \mid x) &= T^\ddagger(\mathfrak{F}_\varepsilon) - T^\ddagger(\mathfrak{F}_0) \\
&= \varepsilon \dot{T}^\ddagger(\mathfrak{F}_0; \mathfrak{H}) + o(\varepsilon) \\
&= -\varepsilon\,\lambda_y(x)\boldsymbol{x}^\top\left[\mathbb{E}\,\lambda_y(X)\boldsymbol{X}\boldsymbol{X}^\top\right]^{-1}\left[\mathbb{E}\,\lambda_y(X)\Psi(X, y-\alpha-X^\top\beta)\boldsymbol{X}\right] + o(\varepsilon),
\end{aligned}$$

where $\lambda_y(x) = \Lambda'(y-\alpha-x^\top\beta)$. The asymptotic biases of F° and Q^+ are instead obtained by applying the chain rule to the transformations

$$\begin{aligned}
F^\circ(y \mid x) &= \tau_1(T^*(\mathfrak{F})) = \int_0^1 \mathbb{1}\{T^*(\mathfrak{F})(u \mid x) \le y\}\, du, \\
Q^+(p \mid x) &= \tau_2(T^\ddagger(\mathfrak{F})) = \int_0^\infty \mathbb{1}\{T^\ddagger(\mathfrak{F})(y \mid x) \le p\}\, dy - \int_{-\infty}^0 \mathbb{1}\{T^\ddagger(\mathfrak{F})(y \mid x) > p\}\, dy.
\end{aligned}$$

For the former, we note that $\Lambda(\boldsymbol{x}^\top \boldsymbol{\theta}_y(\mathfrak{F}_0)) = F_0(y \mid x) = \Lambda(y-\alpha-x^\top\beta)$, so the root of $p = F_0(y \mid x)$ is unique, that is, $N_j(p \mid x) = 1$. Then, using the expression of the Hadamard differential of τ_2 (which can be derived symmetrically to that of the mapping τ_1), we get

$$\frac{\partial \tau_2(T^\ddagger(\mathfrak{F}_\varepsilon))}{\partial \varepsilon} = \dot{\tau}_2\Big(T^\ddagger(\mathfrak{F}_0); \dot{T}^\ddagger(\mathfrak{F}_0;\mathfrak{H}_\varepsilon)\Big) = \sum_{j=1}^{N_j(p \mid x)} C_{\dot{T}^\ddagger(\mathfrak{F}_0;\mathfrak{H}_\varepsilon)}(p \mid x)$$
$$= -\frac{\dot{T}^\ddagger(\mathfrak{F}_0;\mathfrak{H}_\varepsilon)}{\partial \Lambda(\boldsymbol{x}^\top\boldsymbol{\theta}_y(\mathfrak{F}_0))/\partial y} = -\frac{\dot{T}^\ddagger(\mathfrak{F}_0;\mathfrak{H})}{p(1-p)} + o(1).$$

By analogous reasoning we get

$$\frac{\partial \tau_1(T^*(\mathfrak{F}_\varepsilon))}{\partial \varepsilon}(y \mid x) = -\frac{\dot{T}^*(\mathfrak{F}_0;\mathfrak{H}_\varepsilon)}{\partial F_0^{-1}(y \mid x)/\partial p} = -\Lambda'(y-\alpha-x^\top\beta)\, \dot{T}^*(\mathfrak{F}_0;\mathfrak{H}) + o(1).$$

The result then follows from the fact that $\tau_1(\mathfrak{F}_\varepsilon) - \tau_1(\mathfrak{F}_0) = \varepsilon\big[\partial\tau_1(T^\ddagger(\mathfrak{F}_0))/\partial\varepsilon\big]_{\varepsilon=0} + o_p(\varepsilon)$ and $\tau_2(\mathfrak{F}_\varepsilon) - \tau_2(\mathfrak{F}_0) = \varepsilon\big[\partial\tau_2(T^*(\mathfrak{F}_0))/\partial\varepsilon\big]_{\varepsilon=0} + o_p(\varepsilon)$, with $y-\alpha-x^\top\beta = \Lambda^{-1}(p)$ when $\varepsilon = 0$.

Theorem 5

In this case, the distribution function of (X,Y) under model (9) is exactly of the form $\mathfrak{F}_\varepsilon = \mathfrak{F}_0 + \varepsilon(\mathfrak{G} - \mathfrak{F}_0)$ for some contaminating distribution function $\mathfrak{G}$, so we can directly use the Gateaux differentials (11) and (12) to derive the asymptotic bias of all our estimators after applying the chain rule in a way similar to Theorem 4. Thus,

$$\dot{T}^*(\mathfrak{F}_0;\mathfrak{H}) = \frac{1}{p(1-p)}\boldsymbol{x}_0^\top P_X^{-1}\int_{\mathscr{X}}\int_{\mathscr{Y}}\Big[p_0 - \mathbb{1}\{y - \boldsymbol{x}^\top\boldsymbol{\gamma}_{p_0}(F_0)\}\Big]\boldsymbol{x}\, d\mathfrak{H}$$

and

$$\dot{T}^\ddagger(\mathfrak{F}_0;\mathfrak{H}) = \left.\frac{\partial\Lambda\big(\boldsymbol{x}^\top\boldsymbol{\theta}_{y_0}(\mathfrak{F}_\varepsilon)\big)}{\partial\varepsilon}\right|_{\varepsilon=0} = \Lambda'\big(\boldsymbol{x}_0^\top\boldsymbol{\theta}_{y_0}(\mathfrak{F}_0)\big)\,\boldsymbol{x}_0^\top\dot{\boldsymbol{\theta}}_{y_0}(\mathfrak{F}_0;\mathfrak{H}).$$

where $\mathfrak{H} = \mathfrak{G} - \mathfrak{F}_0$. The results about $Q^* - Q_0$ and $F^\ddagger - F_0$ then follow by setting $\mathfrak{G}(x,y) = F(x)\,\Pr\{U + \Psi(x,U) \le y-\alpha-x^\top\beta\}$. In particular, the fact that $Q^*(p \mid x) - Q_0(p \mid x) = \boldsymbol{x}^\top\big(\boldsymbol{\gamma}(\mathfrak{F}_\varepsilon) - \boldsymbol{\gamma}(\mathfrak{F}_0)\big) = \varepsilon\,\boldsymbol{x}^\top\dot{\boldsymbol{\gamma}}(\mathfrak{F}_0,\mathfrak{H}) + o(\varepsilon)$ implies the first result, namely

$$Q^*(p \mid x_0) - Q_0(p \mid x_0) = \frac{\varepsilon}{p(1-p)}\,\boldsymbol{x}_0^\top P_X^{-1}\int_{\mathscr{X}}\int_{\mathscr{Y}}\big[p - \mathbb{1}\{y \le \alpha + x^\top\beta + \Lambda^{-1}(p)\}\big]\boldsymbol{x}\, d\mathfrak{H} + o(\varepsilon)$$
$$= \frac{\varepsilon}{p(1-p)}\big[p - \Lambda\big(\varphi(X,\Lambda^{-1}(p))\big)\big]\,\boldsymbol{x}_0^\top P_X\,\mu_X + o(\varepsilon),$$

where $\psi(x,u) = u + \Psi(x,u)$ and $\mu_X = \mathbb{E}\boldsymbol{X}$, while the fact that

$$\boldsymbol{x}^\top\dot{\boldsymbol{\theta}}(\mathfrak{F}_0;\mathfrak{H}) = \boldsymbol{x}^\top\big[\mathbb{E}\,\Lambda'(y-\alpha-X^\top\beta)\boldsymbol{X}\boldsymbol{X}^\top\big]^{-1}\,\mathbb{E}\boldsymbol{X}\big[\Lambda\big(\varphi(X,y-\alpha-X^\top\beta)\big) - \Lambda(y-\alpha-X^\top\beta)\big]$$

implies the second result. The asymptotic biases of F° and Q^+ are instead obtained by applying the chain rule to the transformations $\tau_1(T^*)$ and $\tau_2(T^\ddagger)$ defined in the proof of Theorem 4.