Estimating Derivatives of Function-Valued Parameters in a Class of Moment Condition Models

Christoph Rothe and Dominik Wied

Abstract

We develop a general approach to estimating the derivative of a function-valued parameter $\theta_o(u)$ that is identified for every value of u as the solution to a moment condition. This setup in particular covers many interesting models for conditional distributions, such as quantile regression or distribution regression. Exploiting that $\theta_o(u)$ solves a moment condition, we obtain an explicit expression for its derivative from the Implicit Function Theorem, and estimate the components of this expression by suitable sample analogues, which requires the use of (local linear) smoothing. Our estimator can then be used for a variety of purposes, including the estimation of conditional density functions, quantile partial effects, and structural auction models in economics.

Keywords: Quantile Regression, Distribution Regression, Local Linear Smoothing, Conditional Density Estimation, Quantile Partial Effects

First Version: October 3, 2016. This Version: January 5, 2017. Christoph Rothe, Columbia University, Department of Economics, email: cr2690columbia.edu. Dominik Wied, Institute for Econometrics and Statistics, University of Cologne, email: dwied@uni-koeln.de. Research is supported by Deutsche Forschungsgemeinschaft (SFB 823, project A1).

1. INTRODUCTION

Estimating the conditional distribution of a dependent variable Y given covariates X in $\mathbb{R}^p$ is an important problem in many areas of applied economics. For example, studies of changes in income inequality often involve estimation of the conditional distribution of workers' wages given their observable characteristics (e.g. Machado and Mata, 2005; Autor, Katz, and Kearney, 2008). Such applications require models that on the one hand are flexible enough to capture the potentially highly heterogeneous impact of covariates on the dependent variable at different points in the distribution, but on the other hand can also be estimated using computationally and theoretically attractive methods. A model that is particularly popular in such contexts is the linear quantile regression (QR) model (Koenker and Bassett, 1978), which specifies the conditional quantile function $Q_{Y|X}(u,x)$ of Y given X as

$$Q_{Y|X}(u,x) = x'\theta_o(u).$$

Another model that has received much attention recently is the linear distribution regression (DR) model (Foresi and Peracchi, 1995), which specifies the conditional c.d.f. $F_{Y|X}(u,x)$ of Y given X as

$$F_{Y|X}(u,x) = \Lambda(x'\theta_o(u)),$$

where $\Lambda(\cdot)$ is a known link function that, for convenience, is often taken to be the Logit function.[1] A common feature of these two models, and other models for conditional distributions, is that the respective specification depends on a function-valued parameter $\theta_o(u)$ for which at every appropriate value of u there exist an asymptotically normal estimator that

[1] Which of the two models, if any, is suitable for a particular empirical applications depends on the specific context. See Chernozhukov, Fernández-Val, and Melly (2013) and Leorato and Peracchi (2015) for comparisons of the relative merits of QR and DR models, and Rothe and Wied (2013) for a formal specification test.

converges at the usual parametric rate.

In this paper, we consider the problem of estimating the derivative $\theta_o^u(u) = \partial_u \theta_o(u)$ of the function-valued parameter in a class of models that covers both QR and DR. This derivative plays an important role in estimating many interesting functionals of a conditional distribution. One application is the estimation of conditional density functions. In the QR model, for instance, the conditional density $f_{Y|X}(y,x)$ of Y given X is

$$f_{Y|X}(y,x) = \frac{1}{x'\theta_o^u(F_{Y|X}(y,x))},$$

with $F_{Y|X}(y,x) = \int_0^1 \mathbb{I}\{x'\theta_o(u) \leq y\}du$ the conditional c.d.f. of Y given X implied by the QR model. Similarly, in the DR model, the conditional density of Y given X is

$$f_{Y|X}(u,x) = \lambda(x'\theta_o(u))x'\theta_o^u(u),$$

with $\lambda(u) = \partial_u \Lambda(u)$ the derivative of the link function. Other applications we consider in this paper include estimating the distribution of bidders' private valuations of auctioned objects based on a QR specification of the distribution of observed bids, and estimating Quantile Partial Effects (QPEs) in a DR model.

Motivated by these applications, we develop a general approach to estimating the derivative of a function-valued parameter $\theta_o(u)$ in an abstract class of models in which this parameter is identified for every value of u in some index set as the solution to a moment condition or estimating equation. As mentioned above, this setup covers both QR and DR models. We exploit that $\theta_o(u)$ solves a moment condition to obtain an explicit expression for $\theta_o^u(u)$ from the Implicit Function Theorem, and estimate the components of this expression by suitable sample analogues. The details of the last step depend on the exact properties of the moment condition, and in both QR and DR models some form of smoothing is needed. We use local linear smoothing in this case, which leads to a computationally simple estimator with attractive theoretical properties. For both QR and DR, we show that our estimator of

$\theta_o^u(u)$ is asymptotically normal and has bias and variance whose order of magnitude is analogous to that of a one-dimensional nonparametric kernel regression. These properties then carry over to the above-mentioned applications like density estimation via the Continuous Mapping Theorem.

Our paper is connected to a well-established literature on quantile regression, surveyed for example in Koenker (2005). It also contributes to an emerging literature on distribution regression, which was originally proposed by Foresi and Peracchi (1995) and further studied by Chernozhukov, Fernández-Val, and Melly (2013). See also Rothe (2012, 2015) for examples of applications of distribution regression in economics, Rothe and Wied (2013) for specification testing, and Leorato and Peracchi (2015) for a comparison with quantile regression. Chernozhukov, Fernández-Val, and Melly (2013) obtain general results regarding the properties of estimators of function-valued regular parameters. Our paper seems to be the first to address the estimation of derivatives of such parameters in general settings, but at least the specific problem of estimating the derivative $Q^u_{Y|X}(u,x) = \partial_u Q_{Y|X}(u,x)$ of a conditional quantile function with respect to the quantile level has been studied before. In particular, Parzen (1979), Xiang (1995), and Guerre and Sabbah (2012) propose methods based on smoothing an estimate of the function $u \mapsto Q_{Y|X}(u,x)$, whereas Gimenes and Guerre (2013) propose an estimator based on an augmented quantile regression problem with a locally smoothed criterion function. Both approaches differ conceptually from the one we propose in this paper; we explain this in more detail below.

The remainder of this paper is structured as follows. In Section 2, we describe a general approach to estimating the derivative of the function-valued parameters in a class of models that give rise to a moment condition or estimating equation of a particular form. In Sections 3 and 4, we apply this approach to QR and DR models, respectively, and study some applications. Section 5 reports the results of a simulation study, and Section 6 concludes. All proofs are contained in the appendix. Throughout the paper, we use repeated superscripts to denote the partial derivatives of functions up to various orders. That is, with $g(y, x)$ a generic function, we write $g^y(y, x) = \partial_y g(y, x)$, $g^{yy}(y, x) = \partial_y^2 g(y, x)$, $g^{yyy}(y, x) = \partial_y^3 g(y, x)$, etc., for the first, second, third, etc., partial derivative with respect to y.

2. GENERAL SETUP

While we are mostly interested in estimating the derivative of the function-valued parameters in QR and DR models, we find it useful to motivate our approach in a more general setting that also covers other interesting cases. This section describes the approach, obtains some results on bias properties, and discusses its merits relative to alternative methods.

2.1. Framework

We consider a model in which there is a function-valued parameter $u \mapsto \theta_o(u)$, with $u \in \mathcal{U} = [u_*, u^*] \subset \mathbb{R}$ and $\theta_o(u) \in \Theta = \times_{j=1}^{p} [\theta_{j*}, \theta^{j*}] \subset \mathbb{R}^p$, that is identified for every $u \in \mathcal{U}$ through a moment condition. That is, we assume that there exists a function $M(\theta, u) = \mathbb{E}(m(Z, \theta, u))$, with m a known function taking values in $\mathbb{R}^p$ and Z an observable random vector, such that

$$M(\theta, u) = 0 \text{ if and only if } \theta = \theta_o(u) \tag{2.1}$$

for every $u \in \mathcal{U}$. The moment condition $M(\theta, u)$ is assumed to be smooth with respect to both θ and u, but the underlying function $m(Z, \theta, u)$ can potentially be be non-differentiable (this notion is made more formal below). We also assume that the data consist of an i.i.d. sample $\{Z_i\}_{i=1}^n$ from the distribution of Z, and that there is an estimator $\widehat{\theta}(u)$ of $\theta_o(u)$ satisfying

$$\left\| \widehat{M}(\widehat{\theta}(u), u) \right\|^2 = \inf_{\theta \in \Theta} \left\| \widehat{M}(\theta, u) \right\|^2 + o_P(n^{-1/2}), \tag{2.2}$$

uniformly over $u \in \mathcal{U}$, where $\widehat{M}(\theta, u) = n^{-1} \sum_{i=1}^n m(Z_i, \theta, u)$ is the sample version of the moment condition. Under regularity conditions (e.g. Chernozhukov, Fernández-Val, and

Melly, 2013), the random function $u \mapsto \sqrt{n}(\widehat{\theta}(u) - \theta_o(u))$ then converges to a mean zero Gaussian process with almost surely continuous paths, and for fixed $u \in \mathcal{U}$ it holds that

$$\sqrt{n}(\widehat{\theta}(u) - \theta_o(u)) \xrightarrow{d} \mathcal{N}\left(0, M^\theta(\theta_o(u), u)^{-1} S(\theta_o(u), u) M^\theta(\theta_o(u), u)^{-1}\right), \tag{2.3}$$

where $S(\theta_o(u), u) = \mathbb{E}(m(Z, \theta_o(u), u) m(Z, \theta_o(u), u)^\top)$. Many flexible models for conditional distributions, including QR and DR, fit into this framework.

2.2. Parameter of interest

We are interested in estimating the derivative $\theta_o^u(u) = \partial_u \theta_o(u)$ of the function-valued parameter $\theta_o(u)$, which plays an important role in several applications. Such an estimator cannot be obtained by simply taking the (analytical or numerical) derivative of the function $u \mapsto \widehat{\theta}(u)$, as this random function is generally not differentiable. The next proposition, which follows directly from the Implicit Function Theorem, gives primitive conditions for the derivative $\theta_o^u(u)$ to exist in the first place, and derives an explicit expression.

Proposition 1. *Suppose that the function $(\theta, u) \mapsto M(\theta, u)$ is k times continuously differentiable over $\Theta \times \mathcal{U}$, and that the matrix $M^\theta(\theta_o(u), u)$ is invertible for all $u \in \mathcal{U}$. Then the function $u \mapsto \theta_o(u)$ is k times continuously differentiable over $\mathcal{U}$, and*

$$\theta_o^u(u) = -M^\theta(\theta_o(u), u)^{-1} M^u(\theta_o(u), u). \tag{2.4}$$

is its first derivative.

Heuristically, the formula for $\theta_o^u(u)$ can be obtained by taking the total derivative of $u \mapsto M(\theta_o(u), u)$, and noting that because of the identification condition (2.1) this derivative is equal to zero for every value of u. Analogous reasoning also leads to a formula for second- and higher-order derivatives of $\theta_o(u)$, but we do not pursue estimation and inference for such objects any further in this paper.

2.3. Estimation approach

Proposition 1 provides a motivation for constructing an estimator $\widehat{\theta^u}(u)$ of $\theta^u_o(u)$ as a sample analogue of the representation (2.4). That is, with $\widehat{M^\theta}(\theta, u)$ and $\widehat{M^u}(\theta, u)$ suitable sample analogues of $M^\theta(\theta, u)$ and $M^u(\theta, u)$, respectively, we put

$$\widehat{\theta^u}(u) = -\widehat{M^\theta}(\widehat{\theta}(u), u)^{-1}\widehat{M^u}(\widehat{\theta}(u), u).$$

This approach is attractive relative to taking the (numerical) derivative of a smoothed version of $\widehat{\theta}(u)$, for example, as it requires computing the estimate $\widehat{\theta}(u)$ for a single value of u only (we discuss the relative merits of the approach more extensively in Section 2.5 below). Moreover, obtaining sample analogues of $M^\theta(\theta, u)$ and $M^u(\theta, u)$ is typically not very difficult, although the details depend on the particular application. If the function $m(Z, \theta, u)$ is differentiable with respect to θ, we can simply put

$$\widehat{M^\theta}(\theta, u) = \frac{1}{n}\sum_{i=1}^{n} m^\theta(Z_i, \theta, u);$$

and if $m(Z, \theta, u)$ is differentiable with respect to u we define

$$\widehat{M^u}(\theta, u) = \frac{1}{n}\sum_{i=1}^{n} m^u(Z_i, \theta, u).$$

Being sample means of simple transformations of i.i.d. data, these two quantities are both easy to compute and straightforward to analyze.

In the QR and DR models that we consider below, however, the function $m(Z, \theta, u)$ is only differentiable with respect to one of the two arguments θ and u, and non-differentiable with respect to the other. Such non-smoothness occurs even though the moment condition $M(\theta, u)$ is smooth with respect to both arguments in QR and DR models. This property motivates considering a different estimation approach for the case where $m(Z, \theta, u)$ is not smooth. Specifically, if differentiability of $m(Z, \theta, u)$ with respect to u fails, we propose to estimate the jth component of $M^u(\theta,u)$ by the local linear approximation

$$\widehat{M_j^u}(\theta,u) = \underset{\beta\in\mathbb{R}}{\operatorname{argmin}} \int_{u_*}^{u^*} \left(\widehat{M}_j(\theta,v) - \widehat{M}_j(\theta,u) - \beta(v-u)\right)^2 K_h(v-u)dv, \tag{2.5}$$

where K is a bounded and symmetric density function with mean zero and compact support, say $[-1,1]$, h is a "small" bandwidth chosen by the analyst, $K_h(s)$ is a shorthand notation for $K(s/h)/h$, and $\widehat{M}_j(\theta,u)$ denotes the jth component of the vector $\widehat{M}(\theta,u)$. The least squares problem in (2.5) has a unique solution as long as the function $v \mapsto \widehat{M}_j(\theta,v)$ is non-constant over $(u-h,u+h)$. Note that computing this solution does not require the use of numerical optimization methods. Indeed, simple algebra shows that

$$\widehat{M_j^u}(\theta,u) = \frac{1}{nh\kappa_{2,h}(u)} \sum_{i=1}^n \left(\int_{(u_*-u)/h}^{(u^*-u)/h} m_j(Z_i,\theta,u+vh)vK(v)dv - m_j(Z_i,\theta,u)\kappa_{1,h}(u) \right),$$

where for any integer s and $u \in \mathcal{U}$ the constant $\kappa_{s,h}(u)$ is defined as

$$\kappa_{s,h}(u) = \int_{(u_*-u)/h}^{(u^*-u)/h} v^s K(v)dv.$$

Similarly, if differentiability of $m(Z,\theta,u)$ with respect to θ fails, we propose to estimate $M^\theta(\theta,u)$ by the $(p\times p)$ matrix $\widehat{M^\theta}(\theta,u)$ whose (j,k)-entry is given by the local linear approximation

$$\widehat{M_{jk}^\theta}(\theta,u) = \underset{\beta\in\mathbb{R}}{\operatorname{argmin}} \int_{\theta_{j*}}^{\theta^{j*}} \left(\widehat{M}_k(\theta_{-j}(t),u) - \widehat{M}_k(\theta,u) - \beta(t-\theta_j)\right)^2 K_h(t-\theta_j)dt. \tag{2.6}$$

Here $\theta_{-j}(t) = (\theta_1,\ldots,\theta_{j-1},t,\theta_{j+1},\ldots,\theta_p)'$ is a shorthand notation for the p-dimensional vector whose jth component is equal to t, and whose remaining components are equal to the corresponding component of θ. As before, the term $\widehat{M_{jk}^\theta}(\theta,u)$ can be expressed more explicitly as

$$\begin{aligned}&\widehat{M_{jk}^\theta}(\theta,u)\\ &\quad= \frac{1}{nh\kappa_{2,h}(\theta_j)} \sum_{i=1}^n \left(\int_{(\theta_{j*}-\theta_j)/h}^{(\theta^{j*}-\theta_j)/h} m_k(Z_i,\theta_{-j}(\theta_j+th),u)tK(t)dt - m_k(Z_i,\theta,u)\kappa_{1,h}(\theta_j) \right),\end{aligned}$$

where for all integers s, t and $\theta \in \Theta$ the constant $\kappa_s(\theta_t)$ is defined as

$$\kappa_{s,h}(\theta_t) = \int_{(\theta_{t*}-\theta_t)/h}^{(\theta^{t*}-\theta_t)/h} v^s K(v)dv.$$

Note that we distinguish the kernel functionals $\kappa_{s,h}(u)$ and $\kappa_{s,h}(\theta_k)$ through the name of their argument only, which is a slight abuse of notation. Also note that for many models estimation of $M^\theta(\theta, u)$ might already be implemented in many software packages, since such estimates are needed to construct a plug-in estimator of the asymptotic variance of $\widehat{\theta}(u)$; see (2.3).

2.4. Some results on bias under general conditions

At the current level of generality, it is difficult to conduct a full asymptotic analysis of the estimator $\widehat{\theta^u}(u)$ in the "non-smooth" case, where either $\widehat{M^u}(\theta, u)$ or $\widehat{M^\theta}(\theta, u)$ are constructed as described in (2.5) or (2.6), respectively. The following lemma gives a useful intermediate bias result under the assumption that the moment condition $M(\theta, u)$ satisfies suitable differentiability conditions.

Lemma 1. *Suppose that the function $(\theta, u) \mapsto M(\theta, u)$ is three times continuously differentiable over $\Theta \times \mathcal{U}$, and that the derivatives are uniformly bounded. Then*

$$\mathbb{E}(\widehat{M^u}(\theta, u)) - M^u(\theta, u) = \frac{h}{2}\frac{\kappa_{3,h}(u)}{\kappa_{2,h}(u)} M^{uu}(\theta, u) + \frac{h^2}{6}\frac{\kappa_{4,h}(u)}{\kappa_{2,h}(u)} M^{uuu}(\theta, u) + o(h^2)$$

if $\widehat{M^u}(\theta, u)$ is constructed as described in (2.5)*; and*

$$\mathbb{E}(\widehat{M^\theta_{jk}}(\theta, u)) - M^\theta_{jk}(\theta, u) = \frac{h}{2}\frac{\kappa_{3,h}(\theta_j)}{\kappa_{2,h}(\theta_j)} M^{\theta\theta}_{jk}(\theta, u) + \frac{h^2}{6}\frac{\kappa_{4,h}(\theta_j)}{\kappa_{2,h}(\theta_j)} M^{\theta\theta\theta}_{jk}(\theta, u) + o(h^2)$$

if the estimator $\widehat{M^\theta_{jk}}(\theta, u)$ is constructed as described in (2.6).

The lemma shows that $\widehat{M^u}(\theta, u)$ has a bias of order $O(h)$ for values of u close to the boundary of the index set $\mathcal{U}$, and, since $\kappa_{3,h}(u) = 0$ for $u \in [u_* + h, u^* - h]$, a bias of order $O(h^2)$ for values of u sufficiently far in the interior of $\mathcal{U}$. These bias properties are analogous to those of the Nadaraya-Watson estimator in a standard nonparametric regression problem. A similar statement applies to the elements of the estimator $\widehat{M^\theta}(\theta,u)$ in principle. As explained below, however, we can take $\Theta=\mathbb{R}^p$ in both the QR and the DR model, and thus the bias is of the order $O(h^2)$ there for all values of θ. Obtaining further results, such as ones about the variance of $\widehat{M^u}(\theta,u)$ and $\widehat{M^\theta_{jk}}(\theta,u)$, does not seem possible in non-smooth cases without being more specific about the exact nature of the non-smoothness. We therefore only derive such results in the context of our to main applications, QR and DR, below.

2.5. Discussion of alternative estimation approaches

Our approach to estimating $\theta^u_o(u)$ based on the representation (2.4) is by no means the only possible one. An obvious alternative would be to compute the derivative of a smoothed version of the function $u\mapsto\widehat{\theta}(u)$. Recall that directly taking a numerical derivative is generally not feasible, as the function $u\mapsto\widehat{\theta}(u)$ is non-smooth in many important applications. If local linear smoothing is used, the jth component of the corresponding estimator $\widetilde{\theta^u}(u)$ is given by

$$\widetilde{\theta^u_j}(u)=\underset{\beta\in\mathbb{R}}{\operatorname{argmin}}\int\left(\widehat{\theta}_j(v)-\widehat{\theta}_j(u)-\beta(v-u)\right)^2K_h(v-u)dv,\quad j=1,\ldots,p.$$

This approach is similar to the ones used by Parzen (1979), Xiang (1995), and Guerre and Sabbah (2012) for estimating derivatives of quantile functions with respect to the quantile level. Proceeding like this has the disadvantage that it requires computing $\widehat{\theta}(u)$ for many values of u over a sufficiently fine mesh in order to approximate the integral with sufficient numerical accuracy, even if one is only interested in $\theta^u_o(u)$ for one particular value of u. In contrast, our procedure is computationally much less expensive, as we only require an estimate of $\theta_o(u)$ to estimate $\theta^u_o(u)$.

Another alternative approach is due to Gimenes and Guerre (2013), who proposed an Augmented Quantile Regression estimator for the derivative of the function-valued parameter in a QR model. Adapted to our general setting, their approach amounts to estimating the pair $(\theta_o(u), \theta_o^u(u))$ jointly by solving a linearly augmented and smoothed version of the moment condition:

$$\left(\widetilde{\theta}(u), \widetilde{\theta^u}(u)\right) = \operatorname*{argmin}_{\theta \in \mathbb{R}^p, \beta \in \mathbb{R}^p} \left\| \int \widehat{M}(\theta + \beta(v-u), v) K_h(v-u) dv \right\|^2$$

The downside of proceeding like this is that it requires solving a higher-dimensional and slightly non-standard optimization problem, whereas our estimator can be computed using routines that are implemented in standard software packages. Moreover, augmented regression as described in the last equation has the disadvantage that it gives rise to an unnecessary bias term when estimating the function $\theta_o(u)$ itself.

3. QUANTILE REGRESSION

In this section, we study our approach in the context of a QR model, and consider applications to conditional density and density-quantile estimation, and to recovering bidders' valuations from auction data.

3.1. Setup and estimators

In a linear QR model (Koenker and Bassett, 1978; Koenker, 2005), the conditional quantile function $Q_{Y|X}(u,x)$ of a dependent variable $Y \in \mathcal{Y}$ given a vector of covariates $X \in \mathcal{X} \subset \mathbb{R}^p$ is specified for a range of quantile levels $u \in \mathcal{U} = [u_*, u^*] \subset (0,1)$ as $Q_{Y|X}(u,x) = x'\theta_o(u)$, and the parameter vector $\theta_o(u)$ is estimated by

$$\widehat{\theta}(u) = \operatorname*{argmin}_{\theta \in \mathbb{R}^p} \sum_{i=1}^{n} (u - \mathbb{I}\{Y_i \leq X_i'\theta\})(Y_i - X_i'\theta).$$

Under regularity conditions stated formally below, this model fits into our general setup with

$$Z = (Y, X')', \quad \Theta = \mathbb{R}^p,$$
$$m(Z, \theta, u) = (\mathbb{I}\{Y \leq X'\theta\} - u)X$$
$$M(\theta, u) = \mathbb{E}((F_{Y|X}(X'\theta, X) - u)X).$$

In the QR model, the derivatives of $M(\theta, u)$ with respect to θ and u are therefore given by

$$M^{\theta}(\theta, u) = \mathbb{E}(f_{Y|X}(X_i'\theta, X_i)X_iX_i') \quad \text{and} \quad M^{u}(\theta, u) = -\mathbb{E}(X_i),$$

respectively. Since $M^{\theta}(\theta, u)$ does not depend on u, and $M^{u}(\theta, u)$ does not depend on either θ or u, we denote these objects by $M^{\theta}(\theta)$ and M^{u}, respectively, for the remainder of this section to simplify the notation. We then estimate M^u by

$$\widehat{M^u} = -\frac{1}{n}\sum_{i=1}^{n} X_i,$$

and construct an estimator $\widehat{M^{\theta}_{jk}}(\theta)$ of the (j,k) element of $M^{\theta}(\theta)$ as in described in (2.6). The last step yields the expression

$$\widehat{M^{\theta}_{jk}}(\theta) = \frac{1}{nh\kappa_2}\sum_{i=1}^{n} X_{k,i}\left(\int_{-\infty}^{\infty} \mathbb{I}\{Y_i \leq X_i'\theta + X_{j,i}th\}tK(t)dt\right),$$

with $\kappa_s = \int_{-1}^{1} v^s K(v)dv$. Note that since $\Theta = \mathbb{R}^p$, the area of integration in the last equation does not require a boundary adjustment irrespective of the value of θ. With some algebra, we can write this estimator a bit more efficiently as

$$\widehat{M^{\theta}_{jk}}(\theta) = \frac{1}{n\kappa_2}\sum_{i=1}^{n} X_{k,i}\,\mathrm{sign}(X_{j,i})\bar{K}_h\left(\frac{Y_i - X_i'\theta}{|X_{j,i}|}\right),$$

where $\bar{K}(s) = \int_s^1 tK(t)dt$ is a new "pseudo-kernel" function (it is a symmetric, positive function, but generally does not integrate to one), $\bar{K}_h(s) = \bar{K}(s/h)/h$, and $\mathrm{sign}(x) = \mathbb{I}\{x > 0\} - \mathbb{I}\{x < 0\}$ is the sign function. Note that our notation here is to be understood such that

$$\mathrm{sign}(X_{j,i})\bar{K}_h\left((Y_i - X_i'\theta)/|X_{j,i}|\right) = 0 \text{ if } X_{j,i} = 0.$$

The new expression for $\widehat{M^\theta_{jk}}(\theta)$ is convenient for the derivation of asymptotic properties, and highlights the similarities with objects commonly studied in the context of kernel-based nonparametric regression.[2]

The final estimator of the derivative $\theta_o^u(u)$ of the function-valued parameter $\theta_o(u)$ in the QR model is then given by

$$\widehat{\theta^u}(u) = -\widehat{M^\theta}(\widehat{\theta}(u))^{-1}\widehat{M^u}.$$

To derive the asymptotic properties of $\widehat{\theta^u}(u)$, we make the following assumption.

Assumption 1. *(a) The conditional quantile function takes the form* $Q_{Y|X}(u,x) = x'\theta_o(u)$ *for all* $u \in \mathcal{U}$ *and all* $x \in \mathcal{X}$*; (b) the conditional density function* $f_{Y|X}(y,x)$ *exists, is uniformly continuous over the support of* (Y,X)*, uniformly bounded, is twice continuously differentiable with respect to its first argument, and its derivatives are uniformly bounded over the support of* (Y,X)*; (c) The minimal eigenvalue of* $M^\theta(\theta_o(u))$ *is bounded away from zero uniformly over* $u \in \mathcal{U}$*; (d)* $\mathbb{E}(\|X\|^{4+\delta}) < \infty$ *for some* $\delta > 0$*; (e) the bandwidth* h *satisfies* $h \to 0$ *and* $nh \to \infty$ *as* $n \to \infty$*.*

Assumption 1 collects conditions that are mostly standard in the literature on QR models. Assuming that the QR model is correctly specified is strictly speaking not necessary, but facilitates the interpretation of our results.[3] Note that part (c) implies that the conditional density $f_{Y|X}(y,x)$ is bounded away from zero over an approriate range of (y,x) values. Under Assumption 1, both $\widehat{\theta}(u)$ and $\widehat{M^u}$ are $\sqrt{n}$-consistent, whereas each element of the matrix

[2]Note that while the matrix $M^\theta(\theta)$ is symmetric under the QR model, the estimator $\widehat{M^\theta}(\theta)$ is generally not. To improve finite-sample properties, one could therefore consider the "symmetrized" estimator $(\widehat{M^\theta}(\theta) + \widehat{M^\theta}(\theta)')/2$ instead.

[3]Under misspecification, our estimator $\widehat{\theta^u}(u)$ is an estimate of the derivative of the "pseudo-true" parameter $\theta_o(u)$ that solves the moment condition $M(\theta, u) = 0$.

$\widehat{M^\theta}(\theta_o(u))$ converges to its population counterpart at a slower nonparametric rate. This means that

$$\widehat{\theta^u}(u) - \theta_o^u(u) \cong M^\theta(\theta_o(u))^{-1}\left(\widehat{M^\theta}(\theta_o(u)) - M^\theta(\theta_o(u))\right) M^\theta(\theta_o(u))^{-1}M^u,$$

and that the stochastic properties of $\widehat{M^\theta}(\theta_o(u))$ drive the asymptotic behavior of $\widehat{\theta^u}(u)$. To state this result formally, we introduce some notation. For every $\theta \in \Theta$, let $\mathbf{A}(\theta)$ be a random $p \times p$ matrix whose elements are jointly normal, have mean zero, and are such that the covariance between the (j, k) and the (l, m) element is

$$\begin{aligned}&\operatorname{Cov}(\mathbf{A}_{jk}(\theta), \mathbf{A}_{lm}(\theta))\\ &\quad = \kappa_2^{-2}\mathbb{E}\left(X_{k,i}X_{m,i}f_{Y|X}(X_i'\theta, X_i)\operatorname{sign}(X_{j,i}X_{l,i})\int \bar{K}(s)\bar{K}(s|X_{j,i}|/|X_{l,i}|)ds\right).\end{aligned}$$

The distribution of the random matrix $\mathbf{A}(\theta)$ then implicitly defines a positive-definite matrix $V_o(u)$ that is such that

$$M^\theta(\theta_o(u))^{-1}\mathbf{A}(\theta_o(u))M^\theta(\theta_o(u))^{-1}M^u \sim \mathcal{N}(0, V_o(u)).$$

A more explicit expression of the matrix $V_o(u)$ could be derived, but this would require notation that is cumbersome and not very insightful. We also define the bias function

$$B_o(u) = M^\theta(\theta_o(u))^{-1}A(\theta_o(u))M^\theta(\theta_o(u))^{-1}M^u,$$

with $A(\theta)$ the fixed $p \times p$ matrix whose (j, k) element is equal to

$$A_{jk}(\theta) = \frac{1}{6}\frac{\kappa_4}{\kappa_2}\mathbb{E}(f_{Y|X}^{yy}(X_i'\theta, X_i)X_{k,i}X_{j,i}^3)$$

With this notation, we obtain the following result.

Theorem 1. *Suppose that Assumption 1 holds. Then*

$$\sqrt{nh}(\widehat{\theta^u}(u) - \theta_o^u(u) - h^2 B_o(u)) \xrightarrow{d} \mathcal{N}(0, V_o(u)).$$

The theorem shows that $\widehat{\theta^u}(u)$ has bias of order $O(h^2)$ and variance of order $O((nh)^{-1})$ for every value of $u \in \mathcal{U}$. These properties are analogous to those of the local linear estimator in a univariate nonparametric regression problem. Choosing $h \sim n^{-1/5}$ minimizes the order of the asymptotic mean squared error, and choosing h such that $nh^5 \to 0$ as $n \to \infty$ ensures that the bias of $\widehat{\theta^u}(u)$ is asymptotically negligible. In the latter case, we can also conduct inference using a consistent estimator of the asymptotic variance $V_o(u)$. Such an estimator is difficult to express explicitly, but can be obtained as follows. First, note that a simple consistent estimator of the covariance between the (j,k) and the (l,m) element of $\mathbf{A}(\theta_o(u))$ is

$$\frac{1}{n\kappa_2^2}\sum_{i=1}^{n}\left(X_{k,i}X_{m,i}\widehat{d}_{Y|X}(u,X_i)\,\mathrm{sign}(X_{j,i}X_{l,i})\int \bar{K}(s)\bar{K}(s|X_{j,i}|/|X_{l,i}|)ds\right)$$

with $\widehat{d}_{Y|X}(u,x) = 1/x'\widehat{\theta^u}(u)$ the estimator of the density-quantile function $d_{Y|X}(u,x) \equiv f_{Y|X}(Q_{Y|X}(u,x),x)$ studied in the subsection after the next one. We can then simulate draws $\widehat{\mathbf{A}}_s$, $s = 1,\ldots,S$, from the distribution of a Gaussian random matrix with mean zero and the just-estimated covariance structure. Finally, we obtain an estimate $\widehat{V}(u)$ of $V_o(u)$ as

$$\widehat{V}(u) = \frac{1}{S}\sum_{s=1}^{s}\widehat{T}_s\widehat{T}_s' \quad \text{with} \quad \widehat{T}_s = \widehat{M^\theta}(\widehat{\theta}(u))^{-1}\widehat{\mathbf{A}}_s\widehat{M^\theta}(\widehat{\theta}(u))^{-1}\widehat{M^u}.$$

This estimator is consistent as $S \to \infty$, and can thus be expected to perform reasonably well if the number of simulation draws S is sufficiently large.

3.2. Application to density estimation

We can use the structure implied by a linear QR model to estimate the conditional density function $f_{Y|X}(y,x)$ of Y given X. This is an important application because certain distributional features, such as the location of modes, are easier to detect on a density graph than on the graph of a quantile function. In a QR model, we have that

$$f_{Y|X}(y,x) = \frac{1}{x'\theta_o^u(F_{Y|X}(y,x))}, \quad \text{with} \quad F_{Y|X}(y,x) = \int_0^1 \mathbb{I}\{x'\theta_o(u) \le y\} du$$

the conditional c.d.f. of Y given X implied by the QR model. By exploiting this structure, we can circumvent the "curse of dimensionality" that makes fully nonparametric estimation of conditional densities infeasible in settings with many covariates. In particular, we propose the density estimator

$$\widehat{f}_{Y|X}(y,x) = \frac{1}{x'\widehat{\theta^u}(\widehat{F}_{Y|X}(y,x))} \quad \text{with} \quad \widehat{F}_{Y|X}(y,x) = \epsilon + \int_\epsilon^{1-\epsilon} \mathbb{I}\{x'\widehat{\theta}(u) \le y\} du,$$

for some small constant $\epsilon > 0$. Trimming by ϵ was proposed for example by Chernozhukov, Fernández-Val, and Melly (2013) to prevent unstable estimates of very low or very high quantiles to exert an unduly effect on the estimate of the conditional c.d.f. For empirical applications, we follow the recommendation of Chernozhukov, Fernández-Val, and Melly (2013) and put $\epsilon = .01$.

Since $\widehat{\theta}(u)$ is $\sqrt{n}$-consistent so is $\widehat{F}_{Y|X}(y,x)$, and thus the asymptotic behavior of $\widehat{f}_{Y|X}(y,x)$ is driven by that of the slower-converging derivative estimator $\widehat{\theta}_o^u(u)$ at the point $u = F_{Y|X}(y,x)$.

Corollary 1. *Suppose that Assumption 1 holds. Then*

$$\sqrt{nh}\left(\widehat{f}_{Y|X}(y,x) - f_{Y|X}(y,x) + h^2 \frac{x'B_o(F_{Y|X}(y,x))}{f_{Y|X}(y,x)^2}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{x'V_o(F_{Y|X}(y,x))x}{f_{Y|X}(y,x)^4}\right).$$

Note that the limiting distribution in the previous corollary can be a poor approximation to the actual finite-sample distribution of the density estimate $\widehat{f}_{Y|X}$ in areas where the conditional quantile function is rather flat, and thus $x'\theta_o^u(u)$ is close to zero. This can be seen for example by noting that both the bias and the asymptotic variance explode in this case.

3.3. Application to density-quantile estimation

An application that is closely related to density estimation is that of estimating the density-quantile function $d_{Y|X}(u,x) = f_{Y|X}(Q_{Y|X}(u,x),x)$ of Y given X. Parzen (1979) highlights the role of this function for exploratory data analysis, but it also plays are role for estimating the asymptotic variance of the quantile regression estimator $\widehat{\theta}(u)$, which is given by

$$u(1-u)\mathbb{E}(d_{Y|X}(u,X_i)X_iX_i')^{-1}\mathbb{E}(X_iX_i')\mathbb{E}(d_{Y|X}(u,X_i)X_iX_i')^{-1};$$

see Koenker (2005). In the QR model, the density-quantile function and its natural estimator are easily seen to be

$$d_{Y|X}(u,x) = \frac{1}{x'\theta_o^u(u)} \quad \text{and} \quad \widehat{d}_{Y|X}(u,x) = \frac{1}{x'\widehat{\theta^u}(u)},$$

respectively; and the theoretical properties of the estimator are straightforward to establish.

Corollary 2. *Suppose that Assumption 1 holds. Then*

$$\sqrt{nh}\left(\widehat{d}_{Y|X}(u,x) - d_{Y|X}(u,x) + h^2\frac{x'B_o(u)}{d_{Y|X}(u,x)^2}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{x'V_o(u)x}{d_{Y|X}(u,x)^4}\right).$$

Substituting $\widehat{d}_{Y|X}$ for $d_{Y|X}$ in the asymptotic variance formula above, and replacing expectations with appropriate sample averages, then leads to a new estimator for the asymptotic variance of the quantile regression estimator. This estimator could then be used for example in place of the popular estimator proposed by Powell (1986).

3.4. Application to estimating bidders' valuations in auctions

Another interesting way to exploit the structure of a QR model occurs in the analysis of auction data in economics. In a first-price sealed-bid auction with independent private values (e.g. Guerre, Perrigne, and Vuong, 2000), an object with observable characteristics $X \in \mathbb{R}^p$ is auctioned among $b > 2$ bidders. Each bidder submits a bid Y_j, $j = 1,\ldots,b$, without knowing the bids of the others, and the object is sold to the highest bidder at the price

$\max_{j=1,\ldots,b} Y_j$. Each bidder also has a private (unobserved) valuation V_j, $j = 1,\ldots,b$ for the object, and these valuations are modeled as independent draws from an unknown c.d.f. $F_{V|X}(\cdot, X)$. Guerre, Perrigne, and Vuong (2009) show that if bidders are risk-neutral the quantiles of the distribution of valuations can be written in terms of the quantiles of the observed bids as

$$Q_{V|X}(u,x) = Q_{Y|X}(u,x) + \frac{uQ^u_{Y|X}(u|x)}{b-1}.$$

See Haile, Hong, and Shum (2003), Marmer and Shneyerov (2012) and Gimenes and Guerre (2013) for related results. Using a linear QR specification for the conditional quantile function of observed bids given the object's characteristics, we find that

$$Q_{V|X}(u,x) = x'\theta(u) + \frac{ux'\theta^u(u)}{b-1}.$$

A natural estimator of $Q_{V|X}(u,x)$ is thus given by

$$\widehat{Q}_{V|X}(u|x) = x'\widehat{\theta}(u) + \frac{ux'\widehat{\theta^u}(u)}{b-1}.$$

Since $\widehat{\theta}(u)$ converges faster than $\widehat{\theta}^u_o(u)$, the asymptotic properties of $\widehat{Q}_{V|X}(u,x)$ are again driven by that of the derivative estimator. This is shown formally by the next result.

Corollary 3. *Suppose that Assumption 1 holds. Then*

$$\sqrt{nh}\left(\widehat{Q}_{V|X}(u,x) - Q_{V|X}(u,x) - h^2\frac{ux'B_o(u)}{b-1}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{u^2x'V_o(u)x}{(b-1)^2}\right).$$

4. DISTRIBUTION REGRESSION

In this section, we study our approach in the context of a DR model, and consider applications to estimating conditional densities and Quantile Partial Effects (QPEs).

4.1. Setup and estimators

In a DR model (Foresi and Peracchi, 1995), the conditional c.d.f. $F_{Y|X}(u,x)$ of $Y \in \mathcal{Y} \subset \mathbb{R}$ given $X \in \mathcal{X} \subset \mathbb{R}^p$ is specified for a range of threshold values $u \in \mathcal{U} = [u_*, u^*] \subset \mathbb{R}$ as $F_{Y|X}(u,x) = \Lambda(x'\theta_o(u))$, where $\Lambda(\cdot)$ is a known link function. For notational simplicity, we postulate for this paper that the Logit link $\Lambda(u) = 1/(1+\exp(-u))$ is used, but alternative ones such as Probit are of course possible as well. For every $u \in \mathcal{U}$, the parameter vector $\theta_o(u)$ is estimated by

$$\widehat{\theta}(u) = \operatorname*{argmin}_{\theta \in \mathbb{R}^p} \sum_{i=1}^{n} \left(\mathbb{I}\{Y_i \leq y\} \log(\Lambda(X_i'\theta)) + \mathbb{I}\{Y_i > y\} \log(1 - \Lambda(X_i'\theta))\right),$$

which amounts to fitting a Logistic regression for each $u \in \mathcal{U}$ with $\mathbb{I}\{Y_i \leq u\}$ as the dependent variable. Under regularity conditions stated below, this model fits into our general setup with

$$\begin{aligned} Z &= (Y, X')', \quad \Theta = \mathbb{R}^p, \\ m(Z,\theta,u) &= (\mathbb{I}\{Y \leq u\} - \Lambda(X'\theta))X \\ M(\theta,u) &= \mathbb{E}\left((F_{Y|X}(u,X_i) - \Lambda(X_i'\theta))X_i\right). \end{aligned}$$

In the DR model, the derivatives of $M(\theta,u)$ with respect to θ and u are therefore given by

$$M^\theta(\theta,u) = -\mathbb{E}(\lambda(X_i'\theta)X_iX_i') \quad \text{and} \quad M^u(\theta,u) = \mathbb{E}(f_{Y|X}(u,X_i)X_i),$$

respectively, where $\lambda(u) = \partial_u\Lambda(u)$ is the derivative of the Logit link function. Since $M^\theta(\theta,u)$ does not depend on u, and $M^u(\theta,u)$ does not depend on θ, we denote these objects by $M^\theta(\theta)$ and $M^u(u)$, respectively, for the remainder of this section to simplify the notation. We then estimate $M^\theta(\theta)$ by

$$\widehat{M^\theta}(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \lambda(X_i'\theta)X_iX_i',$$

and construct an estimator of $M^u(u)$ as described in (2.5):

$$\widehat{M^u}(u) = \frac{1}{nh\kappa_{2,h}(u)} \sum_{i=1}^{n} X_i \left(\int_{u_*}^{u^*} \mathbb{I}\{Y_i \leq u + th\} tK(t)dt - \mathbb{I}\{Y_i \leq u\}\kappa_{1,h}(u) \right).$$

This estimator can be written a bit more efficiently as

$$\widehat{M^u}(u) = \frac{1}{nh\kappa_{2,h}(u)} \sum_{i=1}^{n} X_i \left(\bar{K}\left(\frac{Y_i - u}{h}\right) - \mathbb{I}\{Y_i \leq u\}\kappa_{1,h}(u) \right),$$

where $\bar{K}(s) = \int_s^1 tK(t)dt$ as in the previous section; and for values of u such that $u_* + h < u < u^* - h$ we obtain the even simpler representation

$$\widehat{M^u}(u) = \frac{1}{n\kappa_2} \sum_{i=1}^{n} X_i \bar{K}_h(Y_i - u),$$

In any case, we estimate $\theta^u(u)$ by

$$\widehat{\theta^u}(u) = -\widehat{M^\theta}(\widehat{\theta}(u))^{-1}\widehat{M^u},$$

and study its asymptotic properties under the following assumption.

Assumption 2. *(a) The conditional c.d.f. takes the form $F_{Y|X}(u, x) = \Lambda(x'\theta_o(u))$ for all $u \in \mathcal{U}$ and all $x \in \mathcal{X}$; (b) the conditional density function $f_{Y|X}(y, x)$ exists, is uniformly continuous over the support of (Y, X), uniformly bounded, is twice continuously differentiable with respect to its first argument, and its derivatives are uniformly bounded over the support of (Y, X); (c) The minimal eigenvalue of $M^\theta(\theta_o(u))$ is bounded away from zero uniformly over $u \in \mathcal{U}$; (d) $\mathbb{E}(\|X\|^{2+\delta}) < \infty$ for some $\delta > 0$; (e) the bandwidth h satisfies $h \to 0$ and $nh \to \infty$ as $n \to \infty$.*

Assumption 2 collects conditions that are mostly standard in the literature on DR models. As in the case of QR models in the previous section, correct specification of the DR model is only assumed to facilitate the interpretation of the results. The asymptotic properties of $\widehat{\theta^u}(u)$ then follow from arguments that are analogous to but simpler than the ones used in the context of the QR model in the previous section. In particular, Assumption 2 guarantees that both $\widehat{\theta}(u)$ and $\widehat{M^\theta}(\theta)$ are $\sqrt{n}$-consistent, whereas each element of the vector $\widehat{M^u}(u)$ converges to its population counterpart at a slower nonparametric rate. This means that

$$\widehat{\theta^u}(u) - \theta_o^u(u) \cong M^\theta(\theta_o(u))^{-1}\left(\widehat{M^u}(u) - M^u(u)\right),$$

and that the stochastic properties of $\widehat{M^u}(u)$ drive the asymptotic behavior of $\widehat{\theta^u}(u)$. To formally state the result, we have to introduce notation that allows us to distinguish the behavior of $\widehat{\theta^u}(u)$ for u in the interior and close to the boundary of $\mathcal{U}$. We define the positive-definite variance matrix

$$V_o(u,c) = \Gamma(c) \cdot M^\theta(\theta_o(u))^{-1}\mathbb{E}(f_{Y|X}(u,X_i)X_iX_i')M^\theta(\theta_o(u))^{-1},$$

where $\Gamma(c) = (\int_{-c}^1 \bar{K}(s)^2 ds + c(\kappa_1(c)^2 - 2\kappa_1(c))/\kappa_2(c)^2$ and $\kappa_s(c) = \int_{-c}^1 t^s K(t)dt$ are constants that depend on the kernel function only. Note that for $c \geq 1$ the kernel constant $\Gamma(c) = \int_{-1}^1 \bar{K}(s)^2 ds/\kappa_2^2$ does not depend on c, and we thus write $V_o(u) = V_o(u,1)$. We also define the bias functions

$$B_{o,int}(u) = \frac{1}{6}\frac{\kappa_4}{\kappa_2}\mathbb{E}(f_{Y|X}^{uu}(u,X_i)X_i) \quad \text{and} \quad B_{o,bnd}(c,u) = \frac{1}{2}\frac{\kappa_3(c)}{\kappa_2(c)}\mathbb{E}(f_{Y|X}^{u}(u,X_i)X_i).$$

We then obtain the following result.

Theorem 2. *Suppose that Assumption 2 holds. Then it holds for fixed $u \in int(\mathcal{U})$ that*

$$\sqrt{nh}(\widehat{\theta^u}(u) - \theta_o^u(u) - h^2 B_{o,int}(u)) \overset{d}{\to} \mathcal{N}(0, V_o(u));$$

for $u = u_* + ch$ with $c \in (0,1)$ it holds that

$$\sqrt{nh}(\widehat{\theta^u}(u) - \theta_o^u(u) - hB_{o,bnd}(u_*,c)) \overset{d}{\to} \mathcal{N}(0, V_o(u_*,c));$$

and for $u = u^* - ch$ with $c \in (0,1)$ it holds that

$$\sqrt{nh}(\widehat{\theta^u}(u) - \theta_o^u(u) - hB_{o,bnd}(u^*,c)) \overset{d}{\to} \mathcal{N}(0, V_o(u^*,c)).$$

The theorem shows that $\widehat{\theta^u}(u)$ has bias of order $O(h^2)$ for values of u in the interior of $\mathcal{U}$, and bias of order $O(h)$ for values of u on the boundary. The variance is of the order $O((nh)^{-1})$ for every value of $u \in \mathcal{U}$. These properties are analogous to those of the Nadaraya-Watson estimator in a univariate nonparametric regression problem. Choosing $h \sim n^{-1/5}$ minimizes the order of the asymptotic mean squared error u in the interior, and choosing h such that $nh^5 \to 0$ as $n \to \infty$ ensures that the bias of $\widehat{\theta^u}(u)$ is asymptotically negligible (and analogously for values of u on the boundary). In the latter case, we can also conduct inference using a consistent estimator of the asymptotic variance $V_o(u)$, such as

$$\widehat{V}(u) = \frac{\int_{-1}^{1} \bar{K}(s)^2 ds}{\kappa_2^2} \cdot \widehat{M^\theta}(\widehat{\theta}(u))^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \widehat{f}_{Y|X}(u, X_i) X_i X_i' \right) \widehat{M^\theta}(\widehat{\theta}(u))^{-1},$$

with $\widehat{f}_{Y|X}(u,x) = \lambda(x'\widehat{\theta}(u))x'\widehat{\theta^u}(u)$ the density estimator studied in the next subsection.

4.2. Application to density estimation

Similarly to the way we used the QR model above, we can use the structure implied by a DR model to estimate the conditional density function $f_{Y|X}(u,x)$ of Y given X. The density and its natural estimator are given by

$$f_{Y|X}(u,x) = \lambda(x'\theta_o(u))x'\theta_o^u(u) \quad \text{and} \quad \widehat{f}_{Y|X}(u,x) = \lambda(x'\widehat{\theta}(u))x'\widehat{\theta^u}(u),$$

respectively. Since $\widehat{\theta}(u)$ is $\sqrt{n}$-consistent and $\widehat{\theta}_o^u(u)$ converges as a slower rate, the asymptotic properties of $\widehat{f}_{Y|X}(y,x)$ are driven by those of our derivative estimator.

Corollary 4. *Suppose that Assumption 2 holds. Then*

$$\sqrt{nh}(\widehat{f}_{Y|X}(u,x) - f_{Y|X}(u,x) - h^2\lambda(x'\theta_o(u))x'B_o(u)) \xrightarrow{d} \mathcal{N}(0, \lambda(x'\theta_o(u))^2 x'V_o(u)x).$$

Through similar arguments, one could also obtain an estimator of the density-quantile function (see the section on QR models above). Since this function is less useful in a DR context, we omit the details in the interest of brevity.

4.3. Application to estimating quantile partial effects

The vector of Quantile Partial Effects (QPEs) of the conditional distribution of Y given X is formally defined as $\pi(\tau, x) \equiv \partial_x Q_{Y|X}(\tau, x)$ for any quantile level $\tau \in (0,1)$. QPEs are widely used and easily interpretable summary measures in many areas of applied statistics. In the QR model, the function-valued parameter coincides with the QPE. This means that the parametrization is easily interpretable, but also imposes the restriction that the function $x \mapsto \pi(\tau, x)$ is constant for every τ. Fully nonparametric estimation of QPEs has been considered by Chaudhuri (1991), Lee and Lee (2008), or Guerre and Sabbah (2012); but such methods become practically infeasible with many covariates due to the "curse of dimensionality".

Here we study the use of the DR model as an alternative way to estimate QPEs. This is particularly attractive in economic application involving wage data, for which Rothe and Wied (2013) argue DR often provides a better fit than QR models. An application of the Implicit Function Theorem yields that under a DR specification

$$\pi(\tau, x) = -\frac{\theta_o(Q_{Y|X}(\tau, x))}{x'\theta_o^u(Q_{Y|X}(\tau, x))}, \quad \text{with} \quad Q_{Y|X}(\tau, x) = \inf\{u : \Lambda(x'\theta_o(u)) \geq \tau\}$$

the conditional quantile function of Y given X implied by the DR model. This representation of the QPE suggest the estimator

$$\widehat{\pi}(\tau, x) = -\frac{\widehat{\theta}(\widehat{Q}_{Y|X}(\tau, x))}{x'\widehat{\theta^u}(\widehat{Q}_{Y|X}(\tau, x))}, \quad \text{with} \quad \widehat{Q}_{Y|X}(\tau, x) = \inf\{u : \Lambda(x'\widehat{\theta}(u)) \geq \tau\}.$$

Since $\widehat{\theta}(u)$ is $\sqrt{n}$-consistent, so is $\widehat{Q}_{Y|X}(\tau, x)$; and since $\widehat{\theta}_o^u(u)$ converges as a slower rate, the asymptotic properties of $\widehat{\pi}(\tau, x)$ are again driven by those of our derivative estimator.

Corollary 5. *Suppose that Assumption 2 holds. Then*

$$\sqrt{nh}\left(\widehat{\pi}(\tau, x) - \pi(\tau, x) - h^2 \frac{\theta_o(Q_{Y|X}(\tau, x))x'B_o(Q_{Y|X}(\tau, x))}{(x'\theta_o^u(Q_{Y|X}(\tau, x)))^2}\right)$$
$$\xrightarrow{d} \mathcal{N}\left(0, \theta_o(Q_{Y|X}(\tau, x))\theta_o(Q_{Y|X}(\tau, x))' \cdot \frac{x'V_o(Q_{Y|X}(\tau, x))x}{(x'\theta_o^u(Q_{Y|X}(\tau, x)))^4}\right).$$

5. SIMULATIONS

To illustrate the finite sample properties of our proposed procedures, we report the results of a small-scale Monte Carlo study. For brevity, we focus on the QR model.[4] We generate data as $Y = X + (1+X)U$, where X follows a χ^2 distribution with 1 degree of freedom, U follows a standard Logistic distribution, and X and U are stochastically independent. This means that the conditional quantile function of Y given X is $Q_{Y|X}(u,x) = \Lambda^{-1}(u) + (\Lambda^{-1}(u)+1)x$, which in turn means that the linear QR model is correctly specified, with

$$\theta_o(u) = \left(\Lambda^{-1}(u), \Lambda^{-1}(u)+1\right)' \quad \text{and} \quad \theta_o^u(u) = \left(\frac{1}{\Lambda(\lambda(u))}, \frac{1}{\Lambda(\lambda(u))}\right)'.$$

We consider estimating $\theta_o^u(u) = (\theta_{o,0}^u(u), \theta_{o,1}^u(u))'$ using the procedure proposed in this paper for the quantile level $u = .5$, sample sizes $n \in \{1000, 4000\}$, and various bandwidth values. We also use a triangular kernel, and set the number of replications to 10,000. Results on the estimator's finite sample bias, variance and mean squared error are given in Table 1. To have a point of reference for these findings, we also consider estimating $\theta_o^u(u)$ using the two alternative approaches discussed in Section 2.5: smoothing the estimated quantile regression process and the using Augmented Quantile Regression estimator of Gimenes and Guerre (2013). The corresponding results are reported in Tables 2 and 3, respectively.

[TABLES 1–3 ABOUT HERE]

Overall, our approach compares favorably to the two competing procedures. While the minimal MSE for estimating the "intercept" parameter $\theta_{o,0}^u(u)$ is similar across the three estimator, our procedure has a substantially smaller MSE when estimating the "slope" parameter $\theta_{o,1}^u(u)$. Indeed, MSE is reduced by about one third to one quarter, depending on the sample size. This shows the potential usefulness of our proposed procedure for applications, and shows that it advantages go beyond computational simplicity. All estimators

[4] We conducted a Monte Carlo study analogous reported to the one in this section for the DR model, and obtained results that are qualitatively very similar.

are sensitive with respect to the choice of the bandwidth parameter, and the range of values that produces reasonable results is very different for our procedure than it is for the two competitors. This is because for our procedure smoothing is with respect to θ, whereas for the two competing procedures smoothing is with respect to u.

6. EMPIRICAL ILLUSTRATION

In this section, we apply our methods to estimate the conditional density of US workers' wages given various explanatory variables. The data are taken from 1988 wave of the Current Population Survey (CPS), an extensive survey of US households. The same data set was previously used in DiNardo, Fortin, and Lemieux (1996), to which we refer for details of its construction. It contains information on 74,661 males that were employed in the relevant period, including the hourly wage, years of education and years of potential labor market experience. We fit linear QR and DR models for the conditional distribution of the *natural logarithm* of wages given education and experience, and then estimate the corresponding conditional density function as described above.[5] In Figure 6.1, we plot the result for a worker with 12 years of education and 16 years of experience, the respective median values of the two variables. For comparison, we also plot the standard Rosenblatt-Parzen kernel estimator of the density of log-wages, computed from the only 948 observations in our data with exactly 12 years of education and 16 years of experience. Both the QR and DR based estimates avoid this type of "localization", and make use of the entire sample.

[5]Results in Rothe and Wied (2013) suggest that the DR specification is more suitable than the QR specification for this type of data. This is because linear QR specifications have difficulties capturing the effect of the minimum wage and the substantial amount of heaping in wage data. We report results for both specifications here nonetheless for the purpose of illustration.

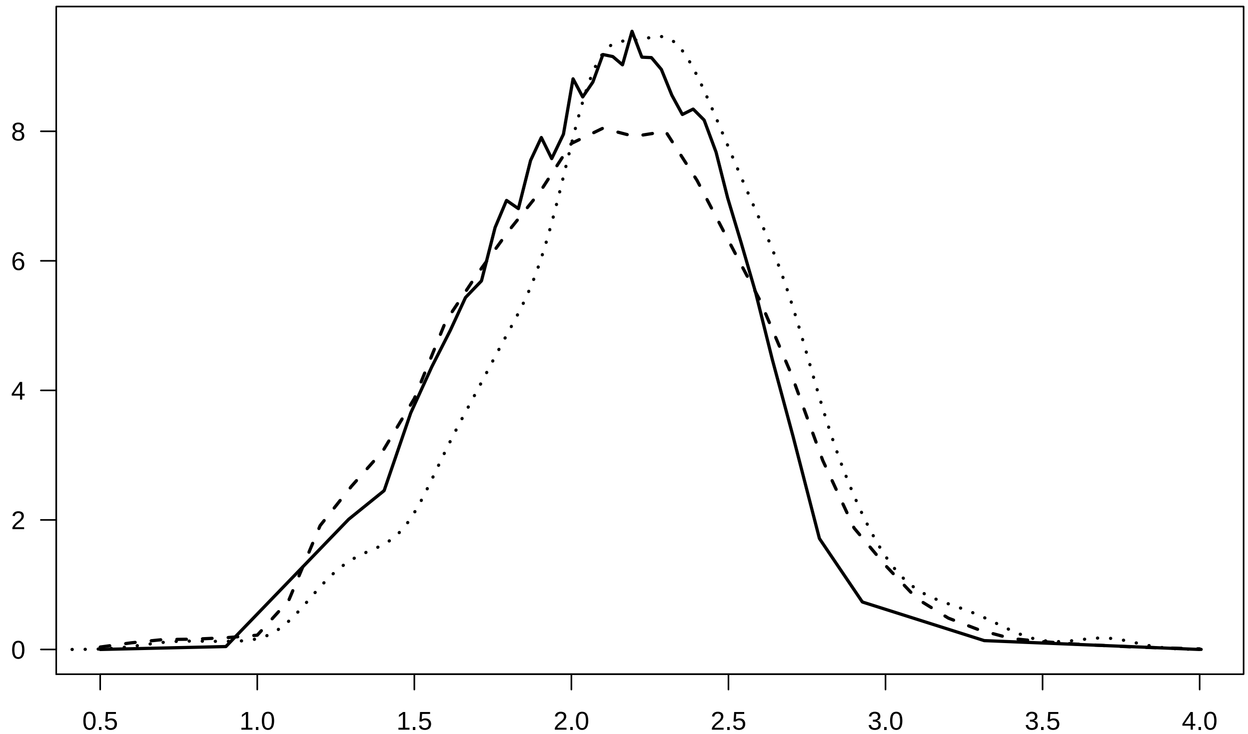

Figure 6.1: Estimated density of the natural logarithm of hourly wages given 12 years of education and 16 years of experience using a QR specification with $h = .03$ (solid line), DR specification with $h = .2$ (dashed line) and fully nonparametric specification with $h = .09$ (dotted line).

7. CONCLUSIONS

In this paper, we propose a new method for estimating the derivative of "regular" function-valued parameters in a class of moment condition models, and provide a detailed analysis of its theoretical properties for the special cases of Quantile Regression and Distribution Regression models. Possible statistical applications for our method include conditional density estimation, estimation of Quantile Partial Effects, and estimation of auction models in economics. Our simulation results suggests that the method compares favorably to alternative approaches that have been proposed in the literature.

A. PROOFS

A.1. Proof of Lemma 1

We only prove the first statement of the Lemma, as the second one follows from the same type of reasoning. Using the explicit expression of $\widehat{M_j^u}(\theta, u)$ given in the main text, and standard Taylor expansion arguments commonly used in the kernel smoothing literature, we find that

$$\begin{aligned}
\mathbb{E}(\widehat{M_j^u}(\theta, u)) &= \frac{1}{h\kappa_{2,h}(u)} \left(\int_{(u_*-u)/h}^{(u^*-u)/h} M_j(\theta, u + vh) v K(v) dv - M_j(\theta, u)\kappa_{1,h}(u) \right) \\
&= \frac{1}{h\kappa_{2,h}(u)} \Big(M_j^u(\theta, u) h \kappa_{2,h}(u) + \frac{1}{2} M_j^{uu}(\theta, u) h^2 \kappa_{3,h}(u) \\
&\quad + \frac{1}{6} M_j^{uuu}(\theta, u) h^3 \kappa_{4,h}(u) + o(h^3) \Big) \\
&= M_j^u(\theta, u) + \frac{h}{2} M_j^{uu}(\theta, u) \frac{\kappa_{3,h}(u)}{\kappa_{2,h}(u)} + \frac{h^2}{6} M_j^{uuu}(\theta, u) \frac{\kappa_{4,h}(u)}{\kappa_{2,h}(u)} + o(h^2),
\end{aligned}$$

as claimed.

A.2. Proof of Theorem 1

To simplify the exposition, we prove the Theorem for the special case that all components of X only take strictly positive values, with probability 1 (the general result follows from the same arguments with an additional case distinction). We begin by studying the properties of the matrix $\widehat{M^\theta}(\theta)$. Simple algebra shows that its (j, k) element is

$$\begin{aligned}
\widehat{M_{jk}^\theta}(\theta) &= \frac{1}{nh\kappa_2} \sum_{i=1}^n X_{k,i} \left(\int \mathbb{I}\{Y_i \leq X_i'\theta_{-j}(\theta_j + th)\} t K(t) dt \right) \\
&= \frac{1}{nh\kappa_2} \sum_{i=1}^n X_{k,i} \left(\int \mathbb{I}\{Y_i \leq X_i'\theta + X_{j,i} th\} t K(t) dt \right) \\
&= \frac{1}{nh\kappa_2} \sum_{i=1}^n X_{k,i} \left(\int_{t \geq (Y_i - X_i'\theta)/(X_{j,i}h)} t K(t) dt \right) \\
&= \frac{1}{nh\kappa_2} \sum_{i=1}^n X_{k,i} \bar{K} \left(\frac{Y_i - X_i'\theta}{X_{j,i}h} \right)
\end{aligned}$$

where $\bar{K}(s) = \int_s^1 tK(t)dt$. Standard kernel calculations involving a change of variables and a Taylor expansion of the conditional p.d.f. $f_{Y|X}$ then yield that

$$\mathbb{E}\left(\widehat{M^\theta_{jk}}(\theta)\right) = M^\theta_{jk}(\theta) + \frac{h^2}{6}\frac{\kappa_4}{\kappa_2}\mathbb{E}(f^{yy}_{Y|X}(X_i'\theta, X_i)X_{k,i}X^3_{j,i}) + o(h^2);$$

so that $\widehat{M^\theta_{jk}}(\theta)$ has bias of order $O(h^2)$. Similarly, we have that

$$\begin{aligned}
\mathbb{E}\left(X^2_{k,i}\bar{K}\left(\frac{Y_i - X_i'\theta}{X_{j,i}h}\right)^2\right) &= h\mathbb{E}\left(X^2_{k,i}\int \bar{K}(s)^2 f_{Y|X}(X_{j,i}sh + X_i'\theta, X_i)ds\right)\\
&= h\int \bar{K}(s)^2 ds \mathbb{E}\left(X^2_{k,i}f_{Y|X}(X_i'\theta, X_i)\right) + o(h) \quad \text{and}\\
\mathbb{E}\left(X_{k,i}\bar{K}\left(\frac{Y_i - X_i'\theta}{X_{j,i}h}\right)\right) &= h\mathbb{E}\left(X_{k,i}\int \bar{K}(s) f_{Y|X}(X_{j,i}sh + X_i'\theta, X_i)ds\right)\\
&= h\int \bar{K}(s) ds \mathbb{E}(X_{k,i}f_{Y|X}(X_i'\theta, X_i)) + o(h);
\end{aligned}$$

which means that $\widehat{M^\theta_{jk}}(\theta)$ has variance of order $O((nh)^{-1})$:

$$\mathbb{V}\left(\widehat{M^\theta}_{jk}(\theta)\right) = \frac{1}{nh}\frac{\int \bar{K}(s)^2 ds}{\kappa_2^2}\mathbb{E}(X^2_{k,i}f_{Y|X}(X_i'\theta, X_i)) + o((nh)^{-1}).$$

Note that the leading term in this variance does not depend on j. Next, we calculate the covariance between $\widehat{M^\theta_{jk}}(\theta)$ and $\widehat{M^\theta_{lm}}(\theta)$. Since, by the smoothness properties of the conditional density function $f_{Y|X}$, we have that

$$\begin{aligned}
&\mathbb{E}\left(X_{k,i}X_{l,i}\bar{K}\left(\frac{Y_i - X_i'\theta}{X_{j,i}h}\right)\bar{K}\left(\frac{Y_i - X_i'\theta}{X_{m,i}h}\right)\right)\\
&\quad = h\mathbb{E}\left(X_{k,i}X_{l,i}\int \bar{K}(s)\bar{K}(sX_{j,i}/X_{m,i}) f_{Y|X}(X_{j,i}sh + X_i'\theta, X_i)\right) ds\\
&\quad = h\mathbb{E}\left(X_{k,i}X_{l,i}\int \bar{K}(s)\bar{K}(sX_{j,i}/X_{m,i}) ds f_{Y|X}(X_i'\theta, X_i)\right) + o(h),
\end{aligned}$$

we find that

$$\begin{aligned}
&\mathrm{Cov}\left(\widehat{M^\theta}_{jk}(\theta), \widehat{M^\theta}_{lm}(\theta)\right)\\
&\quad = \frac{1}{nh\kappa_2^2}\mathbb{E}\left(X_{k,i}X_{l,i}\int \bar{K}(s)\bar{K}(sX_{j,i}/X_{m,i}) ds f_{Y|X}(X_i'\theta, X_i)\right) + o((nh)^{-1})
\end{aligned}$$

It also follows from Lyapunov's central limit theorem and the restrictions on the bandwidth that the joint distribution of the

$$\sqrt{nh}\left(\widehat{M^\theta}_{jk}(\theta) - M^\theta_{jk}(\theta)\right), \quad (j,k) \in \{1,\dots,p\}^2$$

is asymptotically (as $n \to \infty$) multivariate normal, with the covariance structure given in the main part of the paper. From Chernozhukov, Fernández-Val, and Melly (2013), it follows that $\widehat{\theta}(u) = \theta_o(u) + O_P(n^{-1/2})$ uniformly over $u \in \mathcal{U}$, which means that

$$\widehat{M^\theta}_{jk}(\widehat{\theta}(u)) = \widehat{M^\theta}_{jk}(\theta_o(u)) + o((nh)^{-1});$$

and we clearly have that $\widehat{M^u} = M^u + O_P(n^{-1/2})$. From a first-order Taylor expansion of the inverse of a matrix, we then get that

$$\begin{aligned}
\widehat{\theta^u}(u) &= -\widehat{M^\theta}(\widehat{\theta}(u))^{-1}\widehat{M^u} \\
&= -\widehat{M^\theta}(\theta_o(u))^{-1}M^u + o_p(n^{-1/2}) \\
&= -M^\theta(\theta_o(u))^{-1}\left(\widehat{M^\theta}(\theta_o(u)) - M^\theta(\theta_o(u))\right)M^\theta(\theta_o(u))^{-1}M^u \\
&\quad - M^\theta(\theta_o(u))^{-1}M^u + o_P(n^{-1/2}).
\end{aligned}$$

Noting that $\theta^u(u) = -M^\theta(\theta_o(u))^{-1}M^u$, we then obtain the statement of the Theorem. □

A.3. Proof of Theorem 2

The proof of this theorem follows from arguments that are similar to those used to proof Lemma 1 and Theorem 2, but substantially simpler. We thus omit the details for brevity.

REFERENCES

Autor, D. H., L. F. Katz, and M. S. Kearney (2008): "Trends in US wage inequality: Revising the revisionists," *Review of Economics and Statistics*, 90(2), 300–323.

Chaudhuri, P. (1991): "Global nonparametric estimation of conditional quantile functions and their derivatives," *Journal of Multivariate Analysis*, 39(2), 246–269.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013): "Inference on counterfactual distributions," *Econometrica*, 81(6), 2205–2268.

DiNardo, J., N. Fortin, and T. Lemieux (1996): "Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach," *Econometrica*, 64(5), 1001–1044.

Foresi, S., and F. Peracchi (1995): "The Conditional Distribution of Excess Returns: An Empirical Analysis," *Journal of the American Statistical Association*, 90, 451–466.

Gimenes, N., and E. Guerre (2013): "Augmented quantile regression methods for first price auction," *Working paper*.

Guerre, E., I. Perrigne, and Q. Vuong (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68(3), 525–574.

——— (2009): "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions," *Econometrica*, 77(4), 1193–1227.

Guerre, E., and C. Sabbah (2012): "Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function," *Econometric Theory*, 28(01), 87–129.

Haile, P. A., H. Hong, and M. Shum (2003): "Nonparametric tests for common values at first-price sealed-bid auctions," Discussion paper, National Bureau of Economic Research.

Koenker, R. (2005): *Quantile regression*. Cambridge University Press.

Koenker, R., and G. Bassett (1978): "Regression quantiles," *Econometrica*, 46(1), 33–50.

Lee, Y. K., and E. R. Lee (2008): "Kernel methods for estimating derivatives of conditional quantiles," *Journal of the Korean Statistical Society*, 37(4), 365–373.

Leorato, S., and F. Peracchi (2015): "Comparing Distribution and Quantile Regression," Discussion paper, Einaudi Institute for Economics and Finance (EIEF).

Machado, J., and J. Mata (2005): "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of Applied Econometrics*, 20(4), 445–465.

Marmer, V., and A. Shneyerov (2012): "Quantile-based nonparametric inference for first-price auctions," *Journal of Econometrics*, 167(2), 345–357.

Parzen, E. (1979): "Nonparametric statistical data modeling," *Journal of the American Statistical Association*, 74(365), 105–121.

Powell, J. (1986): "Censored regression quantiles," *Journal of Econometrics*, 32(1), 143–155.

Rothe, C. (2012): "Partial distributional policy effects," *Econometrica*, 80(5), 2269–2301.

——— (2015): "Decomposing the Composition Effect: The Role of Covariates in Determining Between-Group Differences in Economic Outcomes," *Journal of Business & Economic Statistics*, 33(3), 323–337.

Rothe, C., and D. Wied (2013): "Misspecification Testing in a Class of Conditional Distributional Models," *Journal of the American Statistical Association*, 108, 314–324.

Xiang, X. (1995): "Estimation of conditional quantile density function," *Journal of Nonparametric Statistics*, 4(3), 309–316.

Table 1: Simulation results using new estimator for the QR model

			Bias		Variance		MSE	
u	n	h	$\widehat{\theta}^u_{o,0}(u)$	$\widehat{\theta}^u_{o,1}(u)$	$\widehat{\theta}^u_{o,0}(u)$	$\widehat{\theta}^u_{o,1}(u)$	$\widehat{\theta}^u_{o,0}(u)$	$\widehat{\theta}^u_{o,1}(u)$
.5	1000	0.5	-0.180	-0.001	0.490	1.038	0.522	1.038
		0.7	-0.110	0.022	0.328	0.706	0.340	0.707
		0.9	-0.053	0.044	0.267	0.523	0.270	0.525
		1.1	-0.010	0.077	0.208	0.415	0.208	0.421
		1.3	0.044	0.093	0.174	0.328	0.176	0.336
		1.5	0.089	0.130	0.147	0.267	0.155	0.283
		1.7	0.131	0.165	0.128	0.224	0.145	0.251
		2.0	0.208	0.221	0.114	0.182	0.157	0.231
		3.0	0.490	0.485	0.080	0.104	0.320	0.339
.5	4000	0.5	-0.028	0.010	0.121	0.245	0.122	0.245
		0.7	0.003	0.022	0.084	0.170	0.084	0.171
		0.9	0.026	0.048	0.063	0.127	0.064	0.130
		1.1	0.057	0.068	0.052	0.097	0.055	0.102
		1.3	0.097	0.094	0.042	0.080	0.052	0.089
		1.5	0.133	0.129	0.037	0.067	0.055	0.083
		1.7	0.171	0.167	0.032	0.056	0.061	0.084
		2.0	0.246	0.227	0.027	0.045	0.087	0.097
		3.0	0.522	0.486	0.019	0.026	0.292	0.262

Table 2: Simulation results using smoothed quantile regression coefficients

			Bias		Variance		MSE	
u	n	h	$\widehat{\theta}^u_{o,0}(u)$	$\widehat{\theta}^u_{o,1}(u)$	$\widehat{\theta}^u_{o,0}(u)$	$\widehat{\theta}^u_{o,1}(u)$	$\widehat{\theta}^u_{o,0}(u)$	$\widehat{\theta}^u_{o,1}(u)$
.5	1000	0.05	0.011	0.010	0.794	2.297	0.794	2.298
		0.10	0.030	0.021	0.368	1.053	0.369	1.053
		0.15	0.057	0.047	0.236	0.678	0.239	0.680
		0.20	0.099	0.084	0.172	0.497	0.182	0.504
		0.25	0.155	0.137	0.135	0.390	0.159	0.409
		0.30	0.228	0.209	0.111	0.319	0.163	0.363
		0.35	0.325	0.305	0.095	0.272	0.200	0.365
		0.40	0.453	0.433	0.083	0.239	0.289	0.426
		0.50	0.922	0.892	0.072	0.204	0.921	1.000
.5	4000	0.05	0.006	0.002	0.196	0.559	0.196	0.559
		0.10	0.022	0.018	0.090	0.256	0.091	0.256
		0.15	0.050	0.044	0.058	0.163	0.060	0.165
		0.20	0.091	0.085	0.042	0.118	0.050	0.125
		0.25	0.147	0.141	0.033	0.092	0.054	0.112
		0.30	0.220	0.214	0.027	0.076	0.075	0.122
		0.35	0.316	0.311	0.023	0.065	0.123	0.162
		0.40	0.443	0.440	0.020	0.057	0.217	0.251
		0.50	0.909	0.903	0.017	0.049	0.843	0.865

Table 3: Simulation results using augmented quantile regression

			Bias		Variance		MSE	
u	n	h	$\widehat{\theta}^u_{o,0}(u)$	$\widehat{\theta}^u_{o,1}(u)$	$\widehat{\theta}^u_{o,0}(u)$	$\widehat{\theta}^u_{o,1}(u)$	$\widehat{\theta}^u_{o,0}(u)$	$\widehat{\theta}^u_{o,1}(u)$
.5	1000	0.05	0.035	-0.321	0.776	2.182	0.778	2.285
		0.10	0.043	-0.146	0.372	1.047	0.373	1.069
		0.15	0.067	-0.064	0.241	0.685	0.245	0.689
		0.20	0.104	0.003	0.178	0.509	0.188	0.509
		0.25	0.155	0.071	0.142	0.406	0.166	0.411
		0.30	0.220	0.148	0.119	0.339	0.167	0.361
		0.35	0.302	0.238	0.103	0.295	0.194	0.351
		0.40	0.401	0.345	0.092	0.263	0.253	0.382
		0.50	0.672	0.624	0.081	0.231	0.533	0.620
.5	4000	0.05	0.013	-0.083	0.190	0.535	0.190	0.542
		0.10	0.025	-0.024	0.090	0.254	0.091	0.255
		0.15	0.052	0.017	0.059	0.164	0.061	0.164
		0.20	0.091	0.064	0.043	0.121	0.051	0.125
		0.25	0.143	0.120	0.034	0.096	0.055	0.111
		0.30	0.209	0.189	0.029	0.080	0.072	0.116
		0.35	0.290	0.274	0.025	0.070	0.109	0.145
		0.40	0.390	0.377	0.022	0.063	0.175	0.205
		0.50	0.661	0.650	0.020	0.056	0.456	0.479