# 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 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 then estimate the components of this expression by suitable sample analogues. The last step generally involves (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 the distribution of bidders' valuations in structural auction models.

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

---

First Version: October 3, 2016. This Version: July 2, 2018. Christoph Rothe, Department of Economics, University of Mannheim, email: `rothe@vwl.uni-mannheim.de`. Dominik Wied, Institute for Econometrics and Statistics, University of Cologne, email: `dwied@uni-koeln.de`.

## 1. INTRODUCTION

Estimating the conditional distribution of a dependent variable $Y$ given covariates $X$ 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 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

[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 discussions of the relative merits of QR and DR models, and Rothe and Wied (2013) for a formal specification test.

function-valued parameter in QR and DR models. 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.

Our general approach to estimating the derivative of the function-valued parameter $\theta_o(u)$ is the same for both QR and DR models, as these models share a common structure. In both cases, the parameter $\theta_o(u)$ is identified for every value of $u$ in some index set as the solution to a moment condition or estimating equation. We exploit this structure to obtain a general and explicit expression for $\theta_o^u(u)$ from the Implicit Function Theorem, and estimate the components of this expression by suitable sample analogues.

This approach does not only apply to QR and DR models, but all setups that share the same general features. The implementation details of our estimation strategy depend on the exact properties of the respective moment condition though. For 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 the derivative $\theta_o^u(u)$ is asymptotically normal and has bias and variance whose order of magnitude is analogous to that of a one-dimensional nonparametric kernel estimator of the *level* of a regression function. 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) also obtain many 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 APPROACH

While our primary interest is in estimating the derivative of the function-valued parameters in QR and DR models, it is useful to motivate our approach in a more general setting that covers both cases, and also potentially other interesting ones.

### 2.1. Framework

Suppose that 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, 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 $\theta$ and $u$, but the underlying moment function $m(Z,\theta,u)$ can potentially be non-differentiable. The data consist of an i.i.d. sample $\{Z_i\}_{i=1}^n$ from the distribution of $Z$, and 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, as pointed put above, plays an important role in several applications. Such an estimator cannot be obtained by simply taking the an analytical derivative of the function $u \mapsto \widehat{\theta}(u)$, as this random function is generally not differentiable. Our proposed approach is based on direct sample-analogue estimation of $\theta_o^u(u)$. To this end, the next proposition gives primitive conditions for the existence of this derivative, and derives an explicit expression.

**Proposition 1.** *Suppose that the function $(\theta, u) \mapsto M(\theta, u)$ is 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 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.*

This result follows directly from the Implicit Function Theorem. 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 $u$. Strengthening the conditions of the proposition to $(\theta, u) \mapsto M(\theta, u)$ being $k$ times continuously differentiable for some integer $k$ yields that $u \mapsto \theta_o(u)$ is $k$ times continuously differentiable; and one can develop an explicit formula analogous to 2.4 for derivatives of

$\theta_o(u)$ up to order $k$ in this case. However, we do not pursue estimation and inference for such objects in this paper.

### 2.3. Estimation approach

Proposition 1 motivates constructing an estimator $\widehat{\theta^u}(u)$ of $\theta_o^u(u)$ 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

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

We argue in more detail in Section 2.4 below that this general approach to estimating $\theta_o^u(u)$ is attractive for both theoretical and practical reasons.

The details of how to obtain "suitable" sample analogues of $M^\theta(\theta, u)$ and $M^u(\theta, u)$ depend on the particular setup. If the function $m(Z, \theta, u)$ is differentiable with respect to $\theta$ and/or $u$, and the respective derivative is available analytically, such estimators take the form of a simple sample mean. Specifically, with differentiability with respect to $\theta$, we 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.

If $m(Z, \theta, u)$ is not differentiable with respect to $\theta$ or $u$, it is necessary to use alternative methods to estimate $M^\theta(\theta, u)$ or $M^u(\theta, u)$, respectively. This issue occurs in both the QR and the DR model: in each case, the respective moment function $m(Z, \theta, u)$ is only differentiable with respect to one of the two arguments $\theta$ 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.

We first consider the issue of estimating $M_j^u(\theta, u)$, the first derivative with respect to $u$ of the $j$th component of the vector-valued function $M(\theta, u)$. To motivate our approach, note that it follows from the assumed differentiability of the *population* moment condition that

$$M_j(\theta, v) \approx M_j(\theta, u) + M_j^u(\theta, u)(v-u) \text{ for } v \text{ close to } u.$$

Thus, if differentiability of $m(Z, \theta, u)$ with respect to $u$ fails, a natural estimate of $M_j^u(\theta, u)$ is given by a variant of local linear smoothing:

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

Here $K$ is a density function with mean zero and compact support, say $[-1, 1]$, that is bounded, symmetric, and vanishes at the boundary of its support; $h$ is a bandwidth chosen by the analyst; and $K_h(s)$ is a shorthand notation for $K(s/h)/h$. Computing the solution of (2.5) is simple and does not require the use of numerical optimization methods. Indeed, simple algebra shows that

$$\widehat{M_j^u}(\theta, u) = \frac{1}{h\kappa_{2,h}(u)} \left( \int_{(u_*-u)/h}^{(u^*-u)/h} \widehat{M}_j(\theta, u+vh)vK(v)dv - \widehat{M}_j(\theta, u)\kappa_{1,h}(u) \right),$$

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

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

Finally, we note that in many applications we can either take $\mathcal{U} = \mathbb{R}$, or focus on values of $u$ that are well in the interior of some bounded index set $\mathcal{U}$. In such cases, the estimator further simplifies to

$$\widehat{M_j^u}(\theta, u) = \frac{1}{h\kappa_2} \int_{-1}^{1} \widehat{M}_j(\theta, u+vh)vK(v)dv,$$

with $\kappa_s = \int_{-1}^{1} v^s K(v)dv$.

The construction of an estimator $\widehat{M^\theta}(\theta,u)$ in cases where differentiability of $m(Z,\theta,u)$ with respect to $\theta$ fails proceeds similarly. That is, we estimate the $(j,k)$-entry of the $(p\times p)$ matrix $M^\theta(\theta,u)$ by

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

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 $j$th component is equal to $t$, and whose remaining components are equal to the corresponding component of $\theta$. The bandwidth $h$ in (2.6) can, and generally should, be different from the one in (2.5). As above, we can write $\widehat{M^\theta_{jk}}(\theta,u)$ more explicitly as

$$\widehat{M^\theta_{jk}}(\theta,u) = \frac{1}{h\kappa_{2,h}(\theta_j)}\left(\int_{(\theta_{j*}-\theta_j)/h}^{(\theta^{j*}-\theta_j)/h} \widehat{M}_k(\theta_{-j}(\theta_j+th),u)tK(t)dt - \widehat{M}(\theta,u)\kappa_{1,h}(\theta_j)\right),$$

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. In applications, we can often take $\Theta = \mathbb{R}^d$, or focus on values of $\theta$ that are well in the interior of $\Theta$. In such cases, the estimator simplifies to

$$\widehat{M^\theta_{jk}}(\theta,u) = \frac{1}{h\kappa_2}\int_{-1}^{1} \widehat{M}_k(\theta_{-j}(\theta_j+th),u)tK(t)dt.$$

We complete this subsection with a number of remarks.

**Remark 1** (Properties under general conditions)**.** It is difficult to provide a full analysis of the asymptotic properties of our smoothing-based estimators of $M^\theta(\theta,u)$ and $M^u(\theta,u)$ under easily interpretable low-level conditions in the context of an abstract moment condition model. One can show however, that these estimators have bias of order $O(h^2)$ at interior points, bias of order $O(h)$ at the boundary, and variance of order $O((nh)^{-1})$, all under rather general conditions.[2] Moreover, one can derive an explicit formula for the leading bias term. Since our main interest in this paper is in specific applications, we present these results in Appendix B. More detailed results, such as asymptotic normality and explicit expressions for the asymptotic variance, are derived for the special cases of QR and DR models in Section 3 and 4 below.

**Remark 2** (Finite-Difference Methods I)**.** An alternative to using local linear smoothing to estimate derivatives would be to use finite-difference methods, which are routinely used in many software packages for numerical differentiation. In its simplest (two-sided) form, such an estimator of, say, $M_j^u(\theta, u)$ is given by

$$\frac{\widehat{M}(\theta, u + \epsilon) - \widehat{M}(\theta, u - \epsilon)}{2\epsilon}$$

for some $\epsilon > 0$ that is chosen by the analyst. Hong, Mahajan, and Nekipelov (2015) provide consistency results for such estimators and their generalizations, but no distributional results of the kind needed for our analysis below.

**Remark 3** (Finite-Difference Methods II)**.** Our derivative estimators can be interpreted as a weighted average of a continuum of simple "one-sided" finite-difference estimators. To see this, define

$$\widehat{\beta}_\epsilon = \frac{\widehat{M}(\theta, u + \epsilon) - \widehat{M}(\theta, u)}{\epsilon}$$

for any $\epsilon \in \mathbb{R} \setminus \{0\}$, set $\widehat{\beta}_\epsilon$ to an arbitrary constant for $\epsilon = 0$, note that the problem (2.5) can then be stated equivalently as

$$\min_{\beta \in \mathbb{R}} \int_{u_* - u}^{u^* - u} (\widehat{\beta}_\epsilon - \beta)^2 \epsilon^2 K_h(\epsilon) d\epsilon,$$

[2]Note that the bias properties are analogous to those of the local linear estimator of the *derivative* of a conditional expectation function in a univariate nonparametric regression problem (Fan and Gijbels, 1996). If boundary bias was a primary concern, one could obtain a bias of order $O(h^2)$ by switching to a local quadratic estimator. However, as shown below, neither the QR nor the DR model naturally give rise to boundary issues, and thus we do not formally investigate such estimators in this paper.

whose solution is given by

$$\widehat{M_j^u}(\theta, u) = \int_{u_*-u}^{u^*-u} \widehat{\beta}_\epsilon \cdot \frac{\epsilon^2 K_h(\epsilon)}{\int_{u_*-u}^{u^*-u} \epsilon^2 K_h(\epsilon) d\epsilon} d\epsilon.$$

**Remark 4** (Intercepts)**.** There is no need to include an additional "intercept" parameter into the least squares problem (2.5). Simple algebra shows that if we were to estimate $M_j^u(\theta, u)$, for example, by the second component of

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

the resulting estimator would be identical to ours at interior points $u$, and have the same general bias properties.

**Remark 5** (Local Constant Estimation)**.** Another alternative to our local linear derivative estimator would be one based on a local constant procedure. For example, a local constant smoother of $\widehat{M}_j(\theta, u)$ with respect to $u$ would be of the form

$$\int_{u_*}^{u^*} \widehat{M}_j(\theta, v) K_h(v-u) dv.$$

This function is differentiable with respect to $u$ if the kernel function is smooth, and the derivative is

$$h^{-2} \int_{u_*}^{u^*} \widehat{M}_j(\theta, v) K'((v-u)/h) dv = h^{-1} \int_{(u_*-u)/h}^{(u^*-u)/h} \widehat{M}_j(\theta, u+th) K'(t) dt,$$

with $K'$ the derivative of $K$. Some algebra shows that this estimator has less favorable bias properties than the one we propose above for commonly used kernel functions, and hence we focus on local linear estimation in this paper.

**Remark 6** (Existing Estimators of $M^\theta(\theta, u)$)**.** Estimators of $M^\theta(\theta, u)$ have been proposed in the context of many specific modles, since such estimates are needed to construct a plug-in estimator of the asymptotic variance of $\widehat{\theta}(u)$; see equation (2.3). For example, for the special case of quantile regression an estimator of $M^\theta(\theta, u)$ was proposed by Powell (1986). We are not aware of a paper that has proposed our specific procedure. Moreover, papers that consider estimating $M^\theta(\theta, u)$ to obtain an estimate of the asymptotic variance of $\widehat{\theta}(u)$ typically only provide consistency result, but do not develop distribution theory of the kind we need for our analysis in Sections 3 and 4.

**Remark 7** (Bandwidth Choice)**.** We leave a formal study of the question how to best choose the bandwidth $h$ in applications to future research. In simulations, the following heuristic bootstrap-based algorithm tends to work well if the sample is at least of moderate size. Starting with some initial bandwidth value, the approach is to use the nonparametric bootstrap to estimate the bias and variance of $\widehat{\theta^u}(u)$, and use these quantities to estimate the MSE optimal bandwidth. This can be repeated using the updated bandwidth value until some convergence criterion is satisfied.

### 2.4. Discussion of alternative approaches

Our approach to estimating $\theta_o^u(u)$ based on the representation (2.4) is of course not the only possible one. A simple alternative would be to compute a finite-difference numerical derivative of the estimated parameter function $u \mapsto \widehat{\theta}(u)$. This would yield an estimator of the form $(\widehat{\theta}(u+\epsilon) - \widehat{\theta}(u-\epsilon))/2\epsilon$ for some $\epsilon > 0$. Since $\widehat{\theta}(u)$ is non-smooth in $u$ this estimator could be sensitive to the choice of $\epsilon$. Moreover, since $\widehat{\theta}(u)$ is the solution of an optimization problem, the theoretical properties of such a derivative estimator are not covered by the results in Hong, Mahajan, and Nekipelov (2015) on finite-difference numerical derivatives of estimated functions.[3]

Another possible approach would be to compute the derivative of a smoothed version of the function $u \mapsto \widehat{\theta}(u)$. If local linear smoothing is used, for instance, the $j$th component of

[3] As pointed out in Remark 2 above, Hong, Mahajan, and Nekipelov (2015) also focus on consistency results, and do discuss asymptotic normality of finite-difference numerical derivatives, for instance.

the corresponding estimator is given by

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

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 $\theta_o^u(u)$ for one particular value of $u$. This is important because computation of $\widehat{\theta}(u)$ can be expensive in many settings. In contrast, our procedure is computationally much less expensive, as we only require an estimate of $\theta_o(u)$ to estimate $\theta_o^u(u)$.

A further alternative 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 non-standard optimization problem, whereas our estimator can be computed using routines that are implemented in most 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

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

In the QR model, the derivatives of $M(\theta,u)$ with respect to $\theta$ 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 $\theta$ 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

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

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 $\theta$. 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.[4]

The final estimator of the derivative $\theta^u_o(u)$ 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* $Q_{Y|X}(u,x) = x'\theta_o(u)$

[4]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.

*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* $nh/\log(n) \to \infty$ *as* $n \to \infty$*.*

Assumption 1 collects conditions that are mostly standard in the literature on QR models. Part (a) assumes that the QR model is correctly specified. This is strictly speaking not necessary, but facilitates the interpretation of our results.[5] Part (b) is a regularity condition on the conditional density function $f_{Y|X}(y,x)$. Part (c) implies that the conditional density $f_{Y|X}(y,x)$ is bounded away from zero over an appropriate range of $(y,x)$ values. Part (d) is a technical condition on the moments of the covariates. Finally, part (e) imposes restrictions on the rate at which the bandwidth tends to zero in large samples. This are such that certain uniform convergence arguments can be used in our proofs.

Under Assumption 1, both $\widehat{\theta}(u)$ and $\widehat{M^u}$ are $\sqrt{n}$-consistent, whereas each element of the matrix $\widehat{M^\theta}(\theta_o(u))$ converges to its population counterpart at a slower nonparametric rate. Heuristically, 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

[5]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 $M(\theta,u) = 0$.

covariance between the $(j,k)$ and the $(l,m)$ element is

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

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 $p\times p$ 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*

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

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 $u\in\mathcal{U}$. Note that since the bias function depends in the inverse of $f_{Y|X}$, the bias can potentially be large if $u$ corresponds to a quantile level for which the conditional density is low. Choosing $h\sim n^{-1/5}$ minimizes the order of the asymptotic mean squared error, and choosing $h$ such that $nh^5\to0$ 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

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

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. Specifically, a natural conditional density estimator in this context is

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

Note that since the quantile regression model is only assumed to be correctly specified for quantile levels $u \in [u_*, u^*] \subset (0,1)$, density estimation is naturally restricted to values of $(y, x)$ such that $Q_{Y|X}(u_*, x) < y < Q_{Y|X}(1-u^*, x)$. To obtain some intuition for the theoretical properties of our density estimator, first note that $\widehat{F}_{Y|X}(y, x)$ is $\sqrt{n}$-consistent, as it is a linear transformation of the $\sqrt{n}$-consistent process $\widehat{\theta}(u)$. One can then prove that the asymptotic behavior of $\widehat{f}_{Y|X}(y, x)$ is driven by that of the slower-converging derivative estimator $\widehat{\theta}^u_o(u)$ at the point $u = F_{Y|X}(y, x)$. The following corollary formally states the result.

**Corollary 1.** *Suppose that Assumption 1 holds, and that $(y, x) \in \mathbb{R}^{1+d}$ is such that $Q_{Y|X}(u_*, x) < y < Q_{Y|X}(1-u^*, x)$. Then*

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

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 $x'\theta^u_o(u)$ is close to zero. This is because the delta method used in its proof is well-known to yield inexact approximations for an inverse function that is evaluated close to zero with high probability.

### 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

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

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 Distribution Regression (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 as $F_{Y|X}(u,x) = \Lambda(x'\theta_o(u))$, where $\Lambda(\cdot)$ is a known link function. We assume that this specification holds for all $u \in \mathcal{Y}$, so that we can take $\mathcal{U} = \mathbb{R}$ here. For notational simplicity, we also 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

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

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

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

In the DR model, the derivatives of $M(\theta, u)$ with respect to $\theta$ 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 $\theta$, 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):

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

This estimator can be written a bit more efficiently as

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

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 $nh/\log(n) \to \infty$ as $n \to \infty$.*

Assumption 2 collects conditions that are mostly standard in the literature on DR models, and largely analogous to Assumption 1 in the previous section on QR models. 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. Heuristically, 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 introduce the positive-definite variance matrix

$$V_o(u,c) = \frac{\int_{-1}^{1} \bar{K}(s)^2 ds}{\kappa_2^2} \cdot M^\theta(\theta_o(u))^{-1} \mathbb{E}(f_{Y|X}(u, X_i) X_i X_i') M^\theta(\theta_o(u))^{-1},$$

where $\kappa_s = \int_{-1}^{1} t^s K(t) dt$; and we introduce the bias function

$$B_o(u) = \frac{1}{6} \frac{\kappa_4}{\kappa_2} \mathbb{E}(f_{Y|X}^{uu}(u, X_i) X_i).$$

We then obtain the following result.

**Theorem 2.** *Suppose that Assumption 2 holds. Then*

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

The theorem shows that $\widehat{\theta^u}(u)$ has bias of order $O(h^2)$ and variance of order $O((nh)^{-1})$ for all values of $u$.[6] 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. The algorithm for bandwidth choice described in the section on quantile regression can also be used here with the obvious modifications. Finally, a simple consistent estimator of the asymptotic variance $V_o(u)$ is given by

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

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

[6] If $\mathcal{U}$ is chosen as a compact set, one can show that the estimator $\widehat{\theta^u}(u)$ has bias of order $O(h)$ for values of $u$ on the boundary, but maintains the order $O((nh)^{-1})$ for the variance; see our Appendix B. Choosing $\mathcal{U}$ in such a way is often difficult to justify in practice, and hence we do not provide a detailed treatment of boundary issues here.

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 $\tau$. 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. NUMERICAL EVIDENCE

In this section, we present some numerical results regarding the performance of our proposed estimation procedures. Specifically, we present the results of two simulation studies, and an empirical illustration.

### 5.1. Simulation performance: comparison with other procedures

To illustrate the finite sample properties of our proposed procedure, and to show how these compare to other methods one could use, we report the results of a small-scale Monte Carlo study. For brevity, we focus on the QR model.[7] We generate data as $Y = X + (1 + X)U$, where $X$ follows a $\chi^2$ distribution with 1 degree of freedom, $U$ follows a standard Logistic

[7] 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.

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

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

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. These results can be compared to those for two alternative approaches to estimating $\theta_o^u(u)$ 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 its advantages go beyond computational simplicity. All estimators 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 the functions that are being smoothed are conceptually different across the three procedures we consider here.

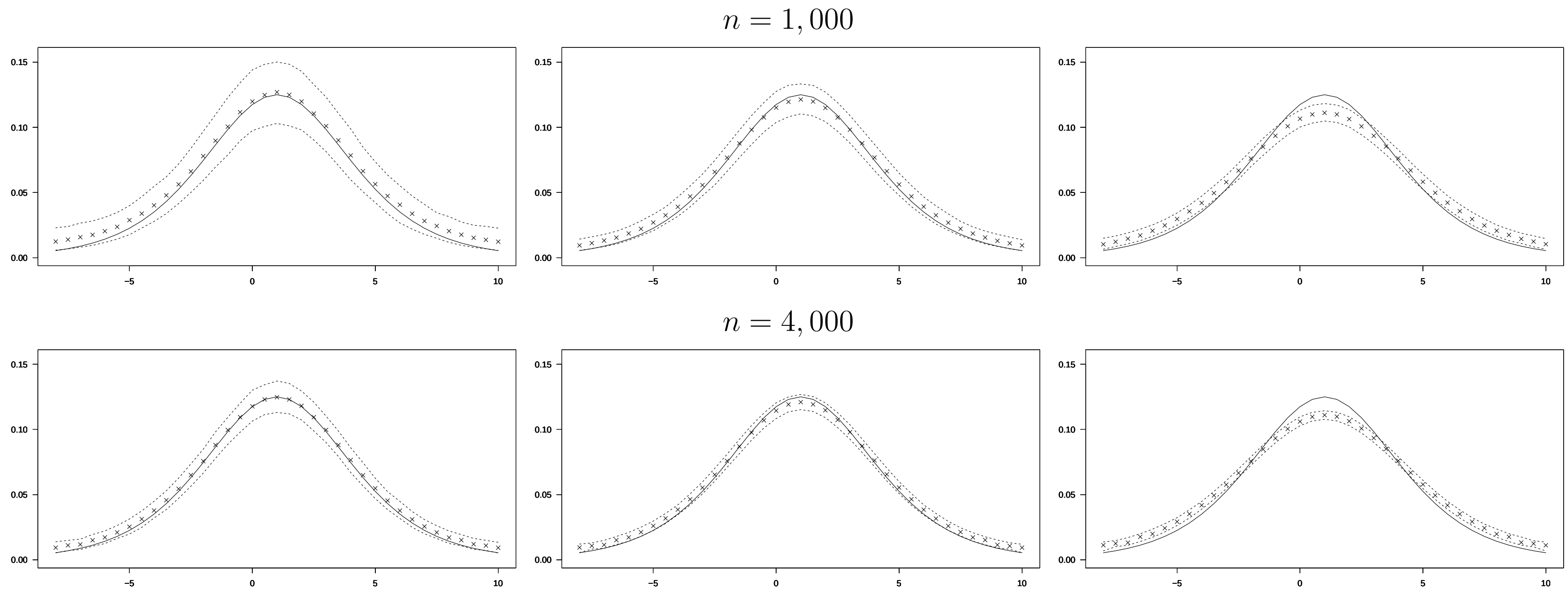


Figure 5.1: Simulation results: Figures show true conditional density function $f_{Y|X}(y, x)$ for $x = 1$ and various values of $y$ (solid line) along with the pointwise median (crosses) and pointwise 95% and 5% quantiles (dashed line) of the corresponding QR based conditional density estimator for $n = 1,000$ (top panel) and $n = 4,000$ (bottom panel); and for bandwidth values $h = .5$ (left column), $h = 1.5$ (middle column), and $h = 3$ (right column).

### 5.2. Simulation performance: conditional density estimation

We conduct a second simulation study to illustrate the performance of our method in the context of conditional density estimation. We generate data in exactly the same way as described in the previous subsection, and use a linear QR specification and the same implementation details as before to estimate the conditional density function $f_{Y|X}(y,x)$ for $x = 1$ and various values of $y$. The panels of Figure 5.1 then show the true density function (which is a location-scale transformation of a standard logistic density), together with the pointwise 5%, 50% and 95% quantiles of our density estimates across 10,000 simulation runs for various sample sizes and bandwidth values.

Our estimator generally captures the overall shape of the density very well. The graphs also clearly show the bias-variance trade-off involved in choosing the bandwidth: larger bandwidths reduce sampling variability, but tend to drive up the bias. Our pictures also confirm that bias can be a bigger issue for our density estimator in tails of the estimated distribution than it is in the center (recall that this is the case because the leading bias term depends in the inverse of $f_{Y|X}$). This suggests that one should use different bandwidths depending on which values of $(y,x)$ on is interested in.

### 5.3. Empirical illustration

As a final illustration, we present an application of our method to real data. Specifically, we use our approach 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.[8] 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 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.

## 6. 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 applications for our method include conditional density estimation, estimation of Quantile Partial Effects, and estimation of auction models. 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 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

[8] 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.

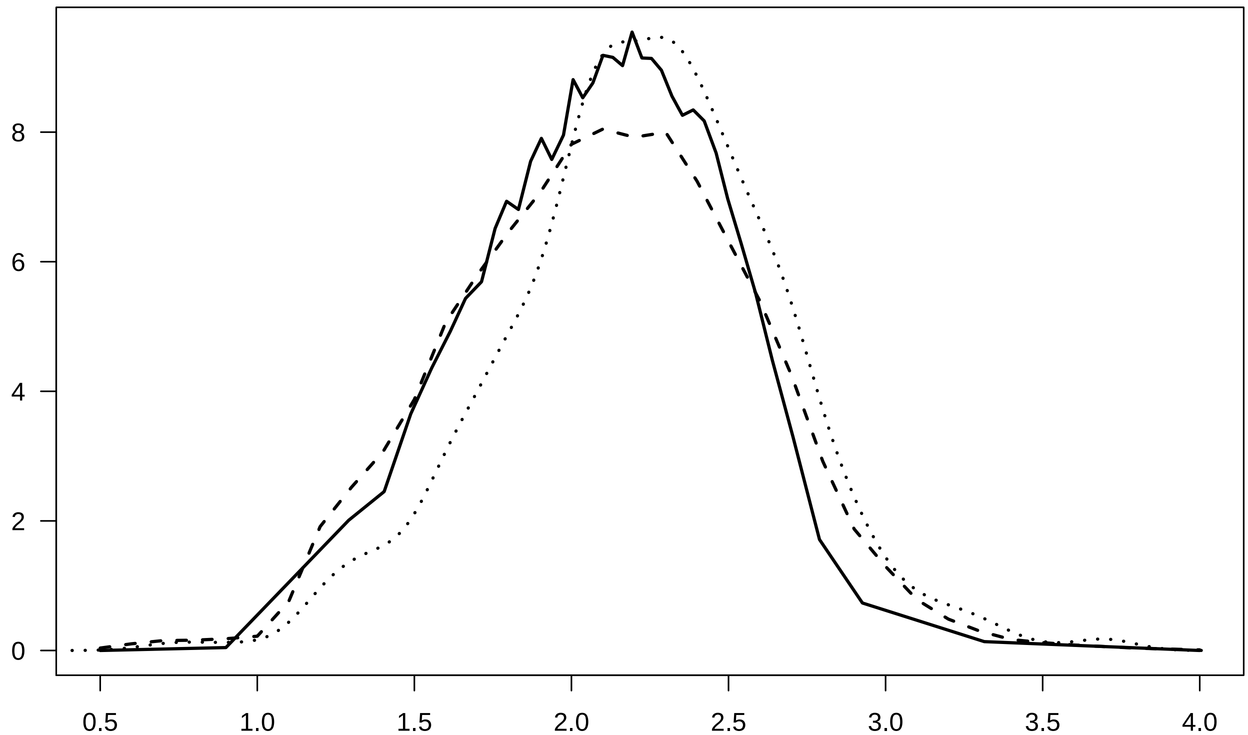


Figure 5.2: Estimated density of the natural logarithm of hourly wages given 12 years of education and 16 years of experience using a QR specification with $h = .03$ (solid line), DR specification with $h = .2$ (dashed line) and fully nonparametric specification with $h = .09$ (dotted line).

the matrix $\widehat{M^\theta}(\theta)$. Simple algebra shows that its $(j,k)$ element is

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

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

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

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 $u \in \mathcal{U}$; and it follows from standard results on uniform convergence of kernel-weighted averages (e.g. Masry, 1996) that $\widehat{M^\theta}_{jk}(\theta) = M^\theta_{jk}(\theta) + O(h^2 + (nh/\log(n))^{-1/2})$ uniformly over $\theta \in \Theta$. A result of Newey (1991, Theorem 1) then yields that $\widehat{M^\theta}_{jk}(\theta)$ satisfies a stochastic equicontinuity condition, which in turn implies that

$$\widehat{M^\theta}_{jk}(\widehat{\theta}(u)) = \widehat{M^\theta}_{jk}(\theta_o(u)) + O_P(n^{-1/2}).$$

We also 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.2. 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.

## B. SOME RESULTS FOR GENERAL MOMENT CONDITION MODELS

The theoretical results presented in the main body of the paper regarding the properties of $\widehat{\theta^u}(u)$ are for the special case of a quantile regression or a distribution regression specification. It is difficult to derive fully analogous results using only the general moment condition setup described in Section 2 without imposing either abstract regularity conditions that would be difficult to check, or postulating assumptions that essentially impose the special structure of QR and DR models. That does not mean, however, that nothing can be said about our estimation procedure outside the context of some special cases. In this appendix, we provide some results regarding the bias and the variance of $\widehat{M^u}(\theta, u)$ and $\widehat{M^\theta}(\theta, u)$ for the "non-smooth" case, where these estimators are constructed as described in (2.5) or (2.6), respectively. These results are fully in line with the findings in the main body of the paper.

**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 $\mathcal{U}$. A similar statement applies to the estimator $\widehat{M^\theta}(\theta, u)$. As explained in the main body of the paper, however, in both the QR and the DR model we can generally chose $\Theta$ and $\mathcal{U}$ as unbounded sets, so that boundary bias is not an issue.[9]

To state a result about the variance, we introduce a definition from Hong, Mahajan, and Nekipelov (2015). A generic function $g(Z_i, v)$ is said to satisfy Lipschitz continuity in mean square with respect to $v$ if for all $\epsilon \in \mathbb{R}$ with $|\epsilon|$ sufficiently small there exists a constant $C$ such that $\mathbb{E}((g(Z_i, v+\epsilon) - g(Z_i, v))^2) \leq C|\epsilon|$. We then have the following result.

**Lemma 2.** *Suppose that the functions $u \mapsto m_j(Z_i, \theta, u)$ and $t \mapsto m_j(Z_i, \theta + te_k, u)$, where $e_k$ denotes the kth unit vector, are Lipschitz continuous in mean square. Then*

$$\mathbb{V}\left(\widehat{M^u_j}(\theta, u)\right) = O((nh)^{-1}) \quad \text{and} \quad \mathbb{V}\left(\widehat{M^\theta_{jk}}(\theta, u)\right) = O((nh)^{-1}).$$

The lemma shows that both variances are of the order $O((nh)^{-1})$. No distinction for boundary cases is necessary here. Lipschitz continuity in mean square can easily be shown to be satisfied in both the DR and the QR model for the respective moment function functions. Of course, Lemma 1 and 2 taken together imply that if $h \to 0$ and $nh \to \infty$ as $n \to \infty$ both $\widehat{M^u}(\theta, u)$ and $\widehat{M^\theta}(\theta, u)$ are (pointwise) consistent, with a rate of convergence of $O_P(h^2 + (nh)^{-1/2})$ in the interior, and $O_P(h + (nh)^{-1/2})$ at the boundary.

### B.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 $\widehat{M^u_j}(\theta, u)$ given in the main text, and standard

[9] If one is concerned about boundary bias, one could use a local quadratic instead of a local linear approximation to estimate either $\widehat{M^u}(\theta, u)$ or $\widehat{M^\theta}(\theta, u)$. Results in Fan and Gijbels (1996) suggest that in this case the bias would be of order $O(h^2)$ both in the interior and at the boundary.

Taylor expansion arguments commonly used in the kernel smoothing literature, we find that

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

as claimed.

## B.2. Proof of Lemma 2

We again 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, it follows an application of Cauchy-Schwarz, Lipschitz continuity in mean square, and the same arguments as those in the proof of Lemma 1, that for $h$ sufficiently small

$$\begin{aligned}\mathbb{V}\left(\widehat{M_j^u}(\theta,u)\right) &= \frac{1}{nh^2\kappa_{2,h}(u)^2}\left[\mathbb{E}\left(\left(\int_{\frac{u_*-u}{h}}^{\frac{u^*-u}{h}}\left(m_j(Z_i,\theta,u+vh)-m_j(Z_i,\theta,u)\right)vK(v)dv\right)^2\right)\right.\\ &\quad\left. -\left(\mathbb{E}\left(\int_{\frac{u_*-u}{h}}^{\frac{u^*-u}{h}}\left(m_j(Z_i,\theta,u+vh)-m_j(Z_i,\theta,u)\right)vK(v)dv\right)\right)^2\right]\\ &\le \frac{1}{nh^2\kappa_{2,h}(u)^2}\mathbb{E}\left(\int_{(u_*-u)/h}^{(u^*-u)/h}\left(m_j(Z_i,\theta,u+vh)-m_j(Z_i,\theta,u)\right)^2v^2K(v)^2dv\right)\\ &\quad +O(n^{-1})\\ &\le \frac{C}{nh\kappa_{2,h}(u)^2}\int_{\frac{u_*-u}{h}}^{\frac{u^*-u}{h}}|v|^3K(v)^2dv+O(n^{-1})\\ &= O((nh)^{-1}),\end{aligned}$$

as claimed.

## 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.

Fan, J., and I. Gijbels (1996): *Local polynomial modelling and its applications*. Chapman & Hall/CRC.

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.

Hong, H., A. Mahajan, and D. Nekipelov (2015): "Extremum estimation and numerical derivatives," *Journal of Econometrics*, 188(1), 250–263.

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.

Masry, E. (1996): "Multivariate local polynomial regression for time series: uniform strong consistency and rates," *Journal of Time Series Analysis*, 17(6), 571–599.

Newey, W. K. (1991): "Uniform convergence in probability and stochastic equicontinuity," *Econometrica*, 59(4), 1161–1167.

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 |