Unbiasedness in Least Quantile Regression

Dirk Tasche[⋆]

Technische Universität München, Zentrum Mathematik, e-mail: `tasche@ma.tum.de`

First version: April 27, 2001
This revision: September 6, 2001

Abstract We develop an abstract notion of regression which allows for a non-parametric formulation of unbiasedness. We prove then that least quantile regression is unbiased in this sense even in the heteroscedastic case if the error distribution has a continuous, symmetric, and uni-modal density. An example shows that unbiasedness may break down even for smooth and symmetric but not uni-modal error distributions. We compare these results to those for least MAD and least squares regression.

Key words Least Quantile – Regression – Unbiasedness – Fisher consistency – Quantile Derivative – Lord's paradox

1 Introduction

Least median of squares (LMS) regression was introduced in [8] as a regression methodology which leads to a high breakdown point and hence is quite robust with respect to outliers in the data. Later on, in [9], this notion was generalized to that of *least quantile of squares* regression. In the present paper, we shall examine how robust the methodology is with respect to unbiasedness when the distribution of the errors varies.

By unbiasedness, we understand here that a disturbed input function will be recovered by means of the regression or is at least among the minimizers of the decision function (cf. Definition 2). Essentially, this is a weak form of Fisher consistency. Our main result (Theorem 1) states that least quantile regression is unbiased even in the heteroscedastic case if the error distribution has a continuous, symmetric, and uni-modal density. In the special case of LMS regression, Theorem 1 extends a result from [2] on the

⋆ Zentrum Mathematik (SCA), TU München, 80290 München, Germany

Fisher consistency of S-estimators to the heteroscedastic case. The unbiasedness property may be lost if the error distribution is asymmetric or not uni-modal (Example 4). This is in contrast to the situation for least squares regression (Proposition 1) or least MAD regression (Proposition 2) which are both much more robust concerning unbiasedness. The paper concludes with some annotations on the connection between unbiasedness of least mean absolute deviation (MAD) regression and the so-called Lord's paradox (cf. [4]).

The discussion on p. 126 in [9] reveals that taking the square in the original definition of least quantile regression has a certain importance when dealing with sample data. However, we drop the "squares" in this paper since with our approach results will not depend on taking the square or not.

We choose here an unusual formulation of regression problems which involves σ-algebras instead of finite-dimensional random vectors. We do this mainly because it represents a convenient and quite general way to include possible dependence of the heteroscedasticity factor and the explanatory variables into the model.

2 An abstract view on regression analysis

Fix a measurable space $(\Omega, \mathcal{F})$, a sub-σ-algebra $\mathcal{A} \subset \mathcal{F}$, and an $\mathcal{F}$-measurable real random variable Y. The σ-algebra $\mathcal{F}$ is the class of all possible events, $\mathcal{A}$ is the σ-algebra of the *explanatory* events, and Y is the *response* variable. The goal in the regression problem is to explain Y the best possible way by some $\mathcal{A}$-measurable real function $\hat{Z}$. In other words: we want to determine the part of Y which is influenced by events in $\mathcal{A}$.

A natural interpretation of "best possible" would be to fix a decision criterion and to search for $\hat{Z}$ under all $\mathcal{A}$-measurable real functions. In general, this problem might be too difficult. Thus, in the sequel we allow to restrict the search on strict subclasses of the class of all $\mathcal{A}$-measurable functions. As the title of the paper suggests, we shall focus on the "least quantile" decision criterion. Nevertheless, we give a general definition of the term "decision function" in order to make the results comparable.

Definition 1 *Let $(\Omega, \mathcal{F})$ be a measurable space and $\mathcal{A} \subset \mathcal{F}$ a sub-σ-algebra of $\mathcal{F}$. Consider two classes $M_{\mathcal{F}}$ and $M_{\mathcal{A}}$ of $\mathcal{F}$- and $\mathcal{A}$-measurable real functions.*

(i) Any function $d : M_{\mathcal{F}} \times M_{\mathcal{A}} \to [0, \infty)$ is a decision function *for the regressions of the elements of $M_{\mathcal{F}}$ on $M_{\mathcal{A}}$.*
(ii) Fix $Y \in M_{\mathcal{F}}$. Then $\hat{Z} \in M_{\mathcal{A}}$ is a solution *for the regression of Y on $M_{\mathcal{A}}$ according to the decision function d if*

$$d(Y, \hat{Z}) \quad = \quad \min\{d(Y, Z) : \ Z \in M_{\mathcal{A}}\}. \qquad \square$$

Evidently, reasonable choices for d should somehow measure the difference between the response variable Y and the elements of $M_{\mathcal{A}}$. We illustrate this by three examples. For any real random variable ξ on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ and any confidence level $\alpha \in (0,1)$, define the *α-quantile* of ξ by

$$q_\alpha(\xi) \quad = \quad \inf\{z \in \mathbb{R} : \; \mathrm{P}[\xi \leq z] \geq \alpha\}\,. \tag{1}$$

Example 1 (Least Quantile Regression, [9])
Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and $\mathcal{A}$ be a sub-σ-algebra of $\mathcal{F}$. Let $M_{\mathcal{F}} = \mathcal{L}_0(\mathcal{F})$ be the class of all $\mathcal{F}$-measurable real random variables and $M_{\mathcal{A}}$ be a class of $\mathcal{A}$-measurable real random variables. Fix some $\alpha \in (0,1)$. Then

$$d_\alpha(Y,Z) \quad = \quad q_\alpha(|Y-Z|), \qquad Y \in \mathcal{L}_0(\mathcal{F}),\, Z \in M_{\mathcal{A}}, \tag{2}$$

is the *least quantile* decision function. □

Example 2 (Least MAD Regression)
Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and $\mathcal{A}$ be a sub-σ-algebra of $\mathcal{F}$. Let $M_{\mathcal{F}} = \mathcal{L}_1(\mathcal{F})$ be the class of all integrable real random variables on $(\Omega, \mathcal{F}, \mathrm{P})$ and $M_{\mathcal{A}}$ be a class of integrable real random variables. Then

$$d_{\mathrm{MAD}}(Y,Z) \quad = \quad \mathrm{E}[|Y-Z|], \qquad Y \in \mathcal{L}_1(\mathcal{F}),\, Z \in M_{\mathcal{A}}, \tag{3}$$

is the *least mean absolute deviation (MAD)* decision function. □

Example 3 (Least Squares Regression)
Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and $\mathcal{A}$ be a sub-σ-algebra of $\mathcal{F}$. Let $M_{\mathcal{F}} = \mathcal{L}_2(\mathcal{F})$ be the class of all square integrable real random variables on $(\Omega, \mathcal{F}, \mathrm{P})$ and $M_{\mathcal{A}}$ be a class of square integrable real random variables. Then

$$d_{\mathrm{SQ}}(Y,Z) \quad = \quad \mathrm{E}[(Y-Z)^2], \qquad Y \in \mathcal{L}_2(\mathcal{F}),\, Z \in M_{\mathcal{A}}, \tag{4}$$

is the *least squares* decision function. □

Common choices for $M_{\mathcal{A}}$ in Examples 1, 2, and 3 are

$$\begin{aligned} &M_{\mathcal{A}} = \mathcal{L}_i(\mathcal{A}), \quad i = 0,1,2, \qquad \text{and} \\ &M_{\mathcal{A}} = \mathrm{span}(1, X_1, \ldots, X_k) \end{aligned} \tag{5}$$

for some random variables $X_1, \ldots, X_k \in \mathcal{L}_i(\mathcal{A}),\ i = 0,1,2$. The second case is called *linear regression*.

Note that in Example 3 when $M_{\mathcal{A}} = \mathcal{L}_2(\mathcal{A})$ and $M_{\mathcal{A}} = \mathrm{span}(1, \ldots, X_k)$ with $X_1, \ldots, X_k \in \mathcal{L}_2(\mathcal{A})$, the solutions to the regression problems are well-known. Since

$$d_{\mathrm{SQ}}(Y, \mathrm{E}[Y|\mathcal{A}]) \quad = \quad \min\{d_{\mathrm{SQ}}(Y,Z) : \; Z \in \mathcal{L}_2(\mathcal{A})\}\,, \tag{6}$$

for any $Y \in \mathcal{L}_2(\mathcal{F})$, the solution is the conditional expectation in the first case. For the second case, an explicit solution formula is known (see, e.g., [12], Proposition A.1). For the case $M_\mathcal{A} = \mathcal{L}_1(\mathcal{A})$ from Example 2 the solution for the regression problem is known to be the conditional median of Y given $\mathcal{A}$.

For the classes $M_\mathcal{A} = \mathrm{span}(1, X_1, \ldots, X_k)$ from Example 2 and $M_\mathcal{A} = \mathcal{L}_0(\mathcal{A})$, $M_\mathcal{A} = \mathrm{span}(1, X_1, \ldots, X_k)$ from Example 1 no fully general explicit solution formulae for the regression problems are known. It is just this aspect which makes the question of unbiasedness interesting in least quantile regression.

We will conclude this section by a general, non-parametric definition of unbiasedness for regression problems. Still before, it might be useful to comment on the relation between the abstract notion of regression from Definition 1 and the more common sample based, fixed or random design regression problems.

Remark 1
The usual setting for a (uni-variate) regression problem is a sample of data points $(x_{11}, \ldots, x_{1k}, y_1), \ldots, (x_{n1}, \ldots, x_{nk}, y_n)$. Here, the y_i are always considered realizations of some random variables $Y_1, \ldots, Y_n$ whereas the $(x_{i1}, \ldots, x_{ik})$ may be realizations of random vectors $(X_{i1}, \ldots, X_{ik})$ (*random design*) or deterministic (*fixed design*).
Random design. In Definition 1, each element Z of $M_\mathcal{A}$ may be a function $Z = f_Z(X_1, \ldots, X_k)$ of some random vector $(X_1, \ldots, X_k)$. In particular, this is the case in linear regression when f is an affine function $\mathbb{R}^k \to \mathbb{R}$. The $(X_{i1}, \ldots, X_{ik})$ are then copies of $(X_1, \ldots, X_k)$. The Y_i are regarded to be copies of some $Y \in M_\mathcal{F}$. In practice, then $d(Y, Z) = d(Y, f_Z(X_1, \ldots, X_k))$ will be replaced by an estimation based on the sample. How to do this, is obvious in Examples 1, 2, and 3.
Fixed design. Suppose that $Y_1, \ldots, Y_n$ are copies of a random variable Y on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$. We interpret then the sample $(x_{11}, \ldots, x_{1k}, y_1)$, $\ldots, (x_{n1}, \ldots, x_{nk}, y_n)$ as a purely random sample from a random vector $(X_1, \ldots, X_k, Y)$ on an enlarged probability space $(\Omega', \mathcal{F}', \mathrm{P}')$ with $\Omega' = \{1, \ldots, n\} \times \Omega$, $\mathcal{F}' = \mathcal{P}(\{1, \ldots, n\}) \otimes \mathcal{F}$ and P' such that the distribution of $(X_1, \ldots, X_k)$ is the empirical distribution corresponding to $(x_{11}, \ldots, x_{1k})$, $\ldots, (x_{n1}, \ldots, x_{nk})$ and the distribution of Y remains the same as under P, i.e.

$$\begin{aligned} \mathrm{P}'[(X_1, \ldots, X_k) \in A] = n^{-1} \sum_{i=1}^{n} \delta_{(x_{i1}, \ldots, x_{ik})}(A) \qquad \text{and} \\ \mathrm{P}'[Y \in B] = \mathrm{P}[Y \in B] \end{aligned} \tag{7}$$

for any Borel-measurable sets $A \subset \mathbb{R}^k$ and $B \subset \mathbb{R}$. Thus, we are again in a random design situation. Obviously, the specification of the marginal distributions in (7) does not determine uniquely the probability P'. But if Y is – as usual – some function of $(X_1, \ldots, X_k)$ plus an independent error with given distribution, P' becomes unique. □

Definition 2 *Let $(\Omega, \mathcal{F})$ be a measurable space and $\mathcal{A}$ be a sub-σ-algebra of $\mathcal{F}$. Let $M_{\mathcal{A}}$ be a class of $\mathcal{A}$-measurable real functions on Ω. Let an $\mathcal{F}$-measurable function η* (error) *and an $\mathcal{A}$-measurable function H* (heteroscedasticity factor) *be given. Assume that $M_{\mathcal{F}}$ is a class of $\mathcal{F}$-measurable real functions on Ω that contains the (η, H)-perturbation of class $M_{\mathcal{A}}$, i.e.*

$$M_{\mathcal{F}} \quad \supset \quad \{Z + H\,\eta : \ Z \in M_{\mathcal{A}}\}\,. \tag{8}$$

We say that the decision function $d : M_{\mathcal{F}} \times M_{\mathcal{A}}$ is unbiased with respect to (η, H) *if every $Z \in M_{\mathcal{A}}$ is a solution for the regression of $Z + \eta\,H$ on $M_{\mathcal{A}}$ in the sense of Definition 1 (ii).*

Of course, Definition 2 is adapted to the most common regression models with the regressand regarded as function of the regressors plus some additive, perhaps heteroscedastic error. Definition 2 then could be figured out as the property that the regression procedure is verifiable in the sense of the question: "Is it possible to recover a known input variable Y?" By revisiting Examples 1, 2, and 3, we will see below that the crucial point in Definition 2 is the question for which error distributions the required solution property holds.

3 Least Quantile Regression

We shall examine on which kind of error distributions Definition 2 applies. Concerning the positive part of our results, we need the following lemma on differentiability of quantiles. It is a modification of Theorem 3.3 in [11], adapted for the purpose of the paper at hand. Recall the definition of the α-quantile q_α in (1).

Lemma 1 *Let $Z, H,$ and η be real random variables on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Assume that (Z, H) and η are independent, that H is positive P-a.s., and that H^{-1} and Z/H are integrable. Assume further that η has a positive and continuous density f. Fix $\alpha \in (0, 1)$. Then the function*

$$Q_\alpha : \ t \mapsto q_\alpha(|t\,Z + H\,\eta|), \ \mathbb{R} \to \mathbb{R} \tag{9}$$

is continuously differentiable on $\mathbb{R}$ with

$$Q'_\alpha(t) \quad = \quad \frac{\mathrm{E}\left[\frac{Z}{H}\left(f\left(\frac{Q_\alpha(t) - t\,Z}{H}\right) - f\left(\frac{-Q_\alpha(t) - t\,Z}{H}\right)\right)\right]}{\mathrm{E}\left[\frac{1}{H}\left(f\left(\frac{Q_\alpha(t) - t\,Z}{H}\right) + f\left(\frac{-Q_\alpha(t) - t\,Z}{H}\right)\right)\right]}. \tag{10}$$

Proof Define

$$F(z, h, t, x) \quad = \quad \mathrm{P}[|t\,z + h\,\eta| \le x] \tag{11}$$

and note that by independence of (Z, H) and η

$$\mathrm{P}[|t\,Z + H\,\eta| \le x] \quad = \quad \mathrm{E}[F(Z, H, t, x)]\,. \tag{12}$$

From the representations (for $x, h > 0,\ t, z \in \mathbb{R}$)

$$F(z,h,t,x) = h^{-1} \int_0^x f\left(\frac{u - t\,z}{h}\right) + f\left(\frac{-u - t\,z}{h}\right) du \qquad (13)$$
$$= \int_{\frac{-x-t\,z}{h}}^{\frac{x-t\,z}{h}} f(u)\,du,$$

we see that F has partial derivatives in $x > 0$ and $t \in \mathbb{R}$ which are jointly continuous in (x,t):

$$\begin{aligned} \frac{\partial F}{\partial x}(z,h,t,x) &= h^{-1}\left(f\left(\tfrac{x-t\,z}{h}\right) + f\left(\tfrac{-x-t\,z}{h}\right)\right), \\ \frac{\partial F}{\partial t}(z,h,t,x) &= -\frac{z}{h}\left(f\left(\tfrac{x-t\,z}{h}\right) - f\left(\tfrac{-x-t\,z}{h}\right)\right). \end{aligned} \qquad (14)$$

By evoking now an appropriate result on differentiation under the integral (e.g. Theorem A.(9.1) of [3]), we obtain from (12) and (14) that also $\mathrm{P}[|t\,Z + H\,\eta| \le x]$ has partial derivatives in $x > 0$ and $t \in \mathbb{R}$ which are jointly continuous in (x,t):

$$\begin{aligned} \frac{\partial}{\partial x}\mathrm{P}[|t\,Z + H\,\eta| \le x] &= \mathrm{E}\left[H^{-1}\left(f\left(\tfrac{x-t\,Z}{H}\right) + f\left(\tfrac{-x-t\,Z}{H}\right)\right)\right], \\ \frac{\partial}{\partial t}\mathrm{P}[|t\,Z + H\,\eta| \le x] &= -\mathrm{E}\left[\tfrac{Z}{H}\left(f\left(\tfrac{x-t\,Z}{H}\right) - f\left(\tfrac{-x-t\,Z}{H}\right)\right)\right]. \end{aligned} \qquad (15)$$

Since by (12) and (13) the distribution of $|t\,Z + H\,\eta|$ is continuous for every $t \in \mathbb{R}$, we have

$$\alpha \quad = \quad \mathrm{P}[|t\,Z + H\,\eta| \le Q_\alpha(t)], \qquad t \in \mathbb{R}. \qquad (16)$$

Observe that the partial derivative in (15) is positive because f is positive. Hence, by the implicit function theorem, (10) follows from (15) and (16). □

With Lemma 1, we are in a position to derive the main result of the paper:

Theorem 1 *Let Z, H, and η be real random variables on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Assume that (Z,H) and η are independent, that H is positive* P*-a.s., and that H^{-1} and Z/H are integrable. Assume further that η has a continuous and symmetric density f which is non-increasing on $(0,\infty)$. Then for any $\alpha \in (0,1)$ we have*

$$q_\alpha(|Z + H\,\eta|) \quad \ge \quad q_\alpha(H\,|\eta|). \qquad (17)$$

Remark 2

(i) Theorem 1 implies that the least quantile decision function d_α from Example 1 is unbiased with respect to (η, H), in the sense of Definition 2, if the conditions of Theorem 1 hold for η and H, $\mathcal{A}$ and η are independent, and for each $Z \in M_{\mathcal{A}}$ the ratio Z/H is integrable.

Note that in case of constant H this result follows also from Theorem 4.6 in [2] on the Fisher consistency of S-estimates (introduced first in [10]). However, there is no obvious way to generalize the proof from [2] for the case of non-constant H.

(ii) If the density f of η in Theorem 1 is strictly decreasing on $(0, \infty)$, then $q_\alpha(|Z + H\,\eta|) = q_\alpha(H\,|\eta|)$ only if $Z = 0$ P-a.s. $\square$

Proof of Theorem 1 We distinguish two cases:

1. The density f is positive on $\mathbb{R}$.
2. There is a number $r > 0$ such that f is positive on $(-r, r)$ and 0 on $\mathbb{R}\backslash(-r, r)$.

The proof for case 1 is short. From the mean value theorem and Lemma 1, and by symmetry of f, we obtain the existence of a number $t \in (0, 1)$ with

$$q_\alpha(|Z+H\,\eta|)-q_\alpha(H\,|\eta|) \;=\; \frac{\mathrm{E}\left[\frac{Z}{H}\left(f\left(\frac{Q_\alpha(t)-t\,Z}{H}\right)-f\left(\frac{Q_\alpha(t)+t\,Z}{H}\right)\right)\right]}{\mathrm{E}\left[\frac{1}{H}\left(f\left(\frac{Q_\alpha(t)-t\,Z}{H}\right)+f\left(\frac{Q_\alpha(t)+t\,Z}{H}\right)\right)\right]}. \tag{18}$$

Observe that by symmetry and monotonicity of f we have for all $a \geq 0$ and $b \in \mathbb{R}$

$$b\Big(f(a-b) - f(a+b)\Big) \quad \geq \quad 0\,. \tag{19}$$

This implies the assertion in case 1. The assertion for case 2 is a conclusion from that of case 1. To see this, introduce a standard normal random variable ξ that is independent of (Z, H, η) and define

$$\eta_h \quad = \quad \eta + h\,\xi, \qquad h > 0. \tag{20}$$

Obviously, $Z + H\,\eta_h \to Z + H\,\eta$ in distribution as $h \to 0$. The first step in order to prove (17) in case 2 is to show that case 1 may be applied to η_h.

Lemma 2 *The density g_h of η_h is continuous, symmetric, and non-increasing on $(0, \infty)$.*

Proof of Lemma 2 Denote by $\phi(t) = (2\,\pi)^{-1}\,e^{-1/2\,t^2}$ the standard normal density. The density g_h of η_h can be written as

$$g_h(x) \quad = \quad \frac{1}{h}\int_{-\infty}^{\infty}\phi\Big(\frac{t}{h}\Big)\,f(x-t)\,d\,t\,. \tag{21}$$

By assumption on f, (21) immediately yields continuity and $g_h > 0$. Symmetry is obvious from (21), too. Concerning monotonicity, fix non-negative numbers $x_1 < x_2$. Then

$$\begin{aligned} g_h(x_2) - g_h(x_1) &= \frac{1}{h}\int_{-\infty}^{\infty} \phi\Big(\frac{t}{h}\Big)\,(f(x_2 - t) - f(x_1 - t))\,dt \\ &= \frac{1}{h}\int_0^{\infty}\left(\phi\left(\frac{\frac{x_1+x_2}{2}+t}{h}\right) - \phi\left(\frac{\frac{x_1+x_2}{2}-t}{h}\right)\right) \\ &\qquad \times \left(f\left(\frac{x_2-x_1}{2}-t\right) - f\left(\frac{x_2-x_1}{2}+t\right)\right)dt\,. \end{aligned}$$

Inequality (19) implies for all positive t

$$\begin{aligned} \phi\left(\frac{\frac{x_1+x_2}{2}+t}{h}\right) - \phi\left(\frac{\frac{x_1+x_2}{2}-t}{h}\right) &\leq 0 \quad \text{and} \\ f\left(\frac{x_2-x_1}{2}-t\right) - f\left(\frac{x_2-x_1}{2}+t\right) &\geq 0\,. \end{aligned}$$

Hence we have obtained $g_h(x_2) - g_h(x_1) \leq 0$. $\square$

Continuing the proof for case 2 of Theorem 1, we have by Lemma 2

$$q_\alpha(|Z + H\,\eta_h|) \quad \geq \quad q_\alpha(H\,|\eta_h|), \qquad h > 0. \tag{22}$$

We will use the following well-known fact on the convergence of quantiles (see, e.g. [7], Proposition 0.1):

Lemma 3 *Let $(\mu_n)_{n\geq 0}$ be a sequence of probability measures on $\mathbb{R}$ such that $\mu_n \to \mu_0$ weakly for $n \to \infty$. Denote by $q_\alpha(\mu_n) = \inf\{x \in \mathbb{R} : \mu_n((-\infty, x]) \geq \alpha\}$ for $n \geq 0$ and $\alpha \in (0,1)$ the α-quantile of μ_n. Then for any $\alpha \in (0,1)$ with*

$$|\{x \in \mathbb{R} :\ \mu_0((-\infty, x]) = \alpha\}| \quad \leq \quad 1 \tag{23}$$

the limit $\lim_{n\to\infty} q_\alpha(\mu_n)$ exists and is equal to $q_\alpha(\mu_0)$.

Observe that (23) holds for arbitrary $\alpha \in (0,1)$ when μ_0 is the distribution of $H\,|\eta|$ and for all α in a dense subset of $(0,1)$ when μ_0 is the distribution of $|Z + H\,\eta|$. Hence, by left-continuity of $\alpha \mapsto q_\alpha(|Z + H\,\eta|)$ and Lemma 3 we obtain (17) from (22). This completes the proof of Theorem 1. $\square$

We conclude this section with an example which shows that symmetry and uni-modality of the error density are crucial for Theorem 1.

Example 4
Symmetry. Let η be a random variable as in Remark 2 (ii). Then we have

$$q_\alpha(\,|\,\eta - 1\,|\,) \quad > \quad q_\alpha(\,|\,\eta\,|\,)\,.$$

By inspection of $\eta' = \eta - 1$ this shows that (17) may fail if the distribution of the error is not symmetric.
Uni-modality. For $\epsilon \in [0, 1/4)$, define the density f_ϵ by

$$f_\epsilon(x) \quad = \quad \begin{cases} \epsilon + (1/3 - 4/3\,\epsilon)\,|x|\,, & |x| \leq 1 \\ 1/3\,(1-\epsilon)\,e^{-(|x|-1)}\,, & |x| > 1\,. \end{cases}$$

For each fixed $\epsilon \in [0, 1/4)$ the function f_ϵ is a bimodal continuous symmetric density which is positive for $\epsilon > 0$.

Denote by η a random variable with density f_ϵ and let $\alpha = \mathrm{P}[|\eta| \leq 1] = 1/3 + 2/3\,\epsilon$. We then compute

$$\mathrm{P}[|\eta - 1| \leq 1] \quad = \quad 1/6 + 1/3\,(1 - e^{-1}) + 1/3\,\epsilon\, e^{-1}\,,$$

which shows in contrast to (17) that here

$$q_\alpha(|\eta|) \quad = \quad 1 \quad > \quad q_\alpha(|\eta - 1|) \tag{24}$$

for some small positive ϵ since $1 - e^{-1} > 1/2$. Observe that by continuity (24) remains true for all confidence levels in some interval containing α. □

4 Least MAD and Least Squares Regression

By Theorem 1 least quantile regression is unbiased if the error distribution has a continuous, symmetric and uni-modal density. Example 4 shows that symmetry and uni-modality are crucial for unbiasedness. In this section we show that in contrast unbiasedness holds for least MAD and least squares regression under much weaker conditions. For least squares regression even independence of the error may be replaced by an appropriate kind of orthogonality. For reasons of comparability, here we consider only the case of independence.

Proposition 1 *Let Z, H, and η be square integrable real random variables on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Assume that (Z, H) and η are independent and that $\mathrm{E}[\eta] = 0$. Then*

$$\mathrm{E}[(Z + H\,\eta)^2] \quad \geq \quad \mathrm{E}[H^2]\,\mathrm{E}[\eta^2] \tag{25}$$

with equality if and only if $Z = 0$ P-a.s.

Proof Immediate from $\mathrm{E}[(Z + H\,\eta)^2] = \mathrm{E}[Z^2] + 2\,\mathrm{E}[Z\,H\,\eta] + \mathrm{E}[H^2\,\eta^2]$. □

Proposition 1 just states that the decision function d_{SQ} from Example 3 is unbiased as long as the error distribution is square integrable with mean 0. Hence, we observe a certain trade-off between robustness with respect to very heavy tails of the errors in case of least quantile regression and robustness with respect to "ugly" shapes of the error distributions in case of least squares regression. This impression is strengthened by a result in case of least MAD regression. Proposition 2 generalizes Lemma 3, ch. 2, of [1].

Proposition 2 *Let Z, H, and η be integrable real random variables on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Assume that (Z, H) and η are independent, that $H > 0$ P-a.s. and that $\mathrm{P}[\eta < 0] \leq 1/2 \leq \mathrm{P}[\eta \leq 0]$. Then*

$$\mathrm{E}[|Z + H\,\eta|] \quad \geq \quad \mathrm{E}[H]\,\mathrm{E}[|\eta|] \tag{26}$$

with equality if and only if $q_{1/2}(\eta) \leq Z \leq \sup\{s \in \mathbb{R} : \mathrm{P}[\eta < s] \leq 1/2\}$ P-a.s.

Proof Recall the following well-known characterization of quantiles (cf. Problem 25.9 of [5]):
Define $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$. If ξ is an integrable real random variable and $\alpha \in (0, 1)$ is fixed, then

$$\min_{s\in\mathbb{R}} \Big(\alpha\,\mathrm{E}[(\xi - s)^+] + (1-\alpha)\,\mathrm{E}[(\xi - s)^-]\Big) = \alpha\,\mathrm{E}[(\xi - s^*)^+] + (1-\alpha)\,\mathrm{E}[(\xi - s^*)^-] \tag{27}$$

for any

$$s^* \;\in\; \Big[q_\alpha(\xi),\, \sup\{s \in \mathbb{R} : \mathrm{P}[\xi < s] \leq \alpha\}\Big]. \tag{28}$$

Concerning (26), we obtain from (27) and by $\mathrm{P}[\eta < 0] \leq 1/2 \leq \mathrm{P}[\eta \leq 0]$ that

$$\begin{aligned}\mathrm{E}[|Z + H\,\eta|] &= \mathrm{E}\Big[H\,\mathrm{E}\Big[|\tfrac{z}{h} + \eta|\Big]\Big|_{(z,h)=(Z,H)}\Big] \\ &\geq \mathrm{E}[H]\,\mathrm{E}[|\eta|].\end{aligned}$$

The rest of the assertion follows from (28). □

Proposition 2 states that the decision function d_{MAD} from Example 2 is unbiased as soon as the error distribution is integrable with median 0. In contrast to Proposition 1 we have not obtained strict unbiasedness, because for error distributions with non-unique medians there might be more than one $\hat{Z}$ yielding the minimum in Definition 1 (ii).

Proposition 2 is also a new formulation of Lord's paradox ([6], [4], [13]). This paradox stems from the fact that (in the words of [4]) "adding an additional, independent source of randomness need not increase variability, as measured by MAD". The interesting point with this paradox is the exact sufficient and necessary conditions for equality in (26) (with $H = 1$). Proposition 2 implies both the theorem in [4] and Remark 1 of [13]. Note that Example 4 shows that adding a constant may even decrease variability when it is measured by a quantile.

Acknowledgements. The author thanks W. Hallerbach from Erasmus University Rotterdam for drawing his attention to Lord's paradox. Two anonymous referees gave helpful comments on an earlier version of this paper. While preparing this manuscript, the author was financially supported by RiskLab Switzerland.

References

1. P. Bloomfield and W. L. Steiger. *Least absolute deviations: Theory, applications, and algorithms*, volume 6 of *Progress in Probability and Statistics*. Birkhäuser Boston Inc., Boston, MA, 1983.
2. P. L. Davies. Aspects of robust linear regression. *Ann. Statist.*, 21(4):1843–1899, 1993.
3. R. Durrett. *Probability: theory and examples*. Duxbury Press, Belmont, CA, second edition, 1996.
4. D. K. Hildebrand. Lord's MAD paradox and Jensen's inequality. *Amer. Statist.*, 38(4):296–297, 1984.
5. K. Hinderer. *Grundbegriffe der Wahrscheinlichkeitstheorie*. Springer-Verlag, Berlin, 1972.
6. F. M. Lord. MAD query. *Amer. Statist.*, 37:343, 1983.
7. S. I. Resnick. *Extreme values, regular variation, and point processes*. Springer-Verlag, New York, 1987.
8. P. J. Rousseeuw. Least median of squares regression. *J. Amer. Statist. Assoc.*, 79(388):871–880, 1984.
9. P. J. Rousseeuw and A. M. Leroy. *Robust regression and outlier detection*. John Wiley & Sons Inc., New York, 1987.
10. P. J. Rousseeuw and V. Yohai. Robust regression by means of S-estimators. In *Robust and nonlinear time series analysis (Heidelberg, 1983)*, volume 26 of *Lecture Notes in Statistics*, pages 256–272. Springer, New York, 1984.
11. D. Tasche. Conditional expectation as quantile derivative. Technical report, Technische Universität München, 2000. http://www.ma.tum.de/stat/.
12. D. Tasche and L. Tibiletti. Approximations for the value-at-risk approach to risk-return analysis. Technical report, Technische Universität München, 2001. http://www.ma.tum.de/stat/.
13. N. N. Vakhania. Lord's paradox on mean absolute deviation. In *Probability theory and mathematical statistics (Kyoto, 1986)*, volume 1299 of *Lecture Notes in Math.*, pages 524–531. Springer, Berlin, 1988.