

Quantile regression: A short story on how and why

Elisabeth Waldmann¹

¹Department of Medical Informatics, Biometry and Epidemiology, Friedrich-Alexander-Universität, Erlangen-Nürnberg, Germany.

Abstract: Quantile regression quantifies the association of explanatory variables with a conditional quantile of a dependent variable without assuming any specific conditional distribution. It hence models the quantiles, instead of the mean as done in standard regression. In cases where either the requirements for mean regression, such as homoscedasticity, are violated or interest lies in the outer regions of the conditional distribution, quantile regression can explain dependencies more accurately than classical methods. However, many quantile regression papers are rather theoretical so the method has still not become a standard tool in applications. In this article, we explain quantile regression from an applied perspective. In particular, we illustrate the concept, advantages and disadvantages of quantile regression using two datasets as examples.

Key words: quantile regression, tutorial article, additive regression models, gradient boosting

Received May 2017; revised December 2017; accepted December 2017

1 Introduction

Mathematically speaking, the probable [...] and the improbable [...] are not different in kind, but only in frequency, whereby the more frequent appears a priori more probable. But the occasional occurrence of the improbable does not imply the intervention of a higher power, something in the nature of a miracle, as the layman is so ready to assume. The term probability includes improbability at the extreme limits of probability, and when the improbable does occur this is no cause for surprise, bewilderment or mystification.

(Frisch, 1959)

This is a quote from Max Frisch's most famous book *Homo Faber* in which the main character meets his daughter, whose existence he did not know of, and falls in love with her. This is only one of the many coincidences in this book, or—as he puts it—one of the improbable events. Another one is a plane crash, which leads him to the earlier quoted reasoning. This example is not very close to our everyday life experience. As

Address for correspondence: Elisabeth Waldmann, Department of Medical Informatics, Biometry and Epidemiology, Waldstraße 6, D-91054 Erlangen, Germany.
E-mail: elisabeth.waldmann@faulde

an introduction to the topic of this tutorial article, we hence use a dataset on the body mass index (BMI) of Dutch males aged 0 to 21 (from now on referred to as the ‘Dutch boys’ dataset). The data will be described in detail in Section 2. The average BMI of the 7 294 observations is 18.027. Thus, we can say ‘the expected BMI for a dutch male between 0 and 21 years is’ 18.027. However, reducing an experiment to its expectation is exactly what the main character of our novel criticizes the ‘layman’ for, who is surprised by the occurrence of events at the ‘the extreme limits of probability’. Furthermore, those extreme limits are highly relevant for the BMI data since obesity and (in case of the Dutch population maybe less relevant) underweight are more relevant than the simple question for the average weight. But what are those ‘limits of probability’ and how can we capture them? A histogram of the data, together with a fitted Gaussian distribution is displayed in the left panel of Figure 1. Distributions can be described quite accurately by their moments, that is, expectation, variance, skewness and so on, as well as by a list of quantiles.

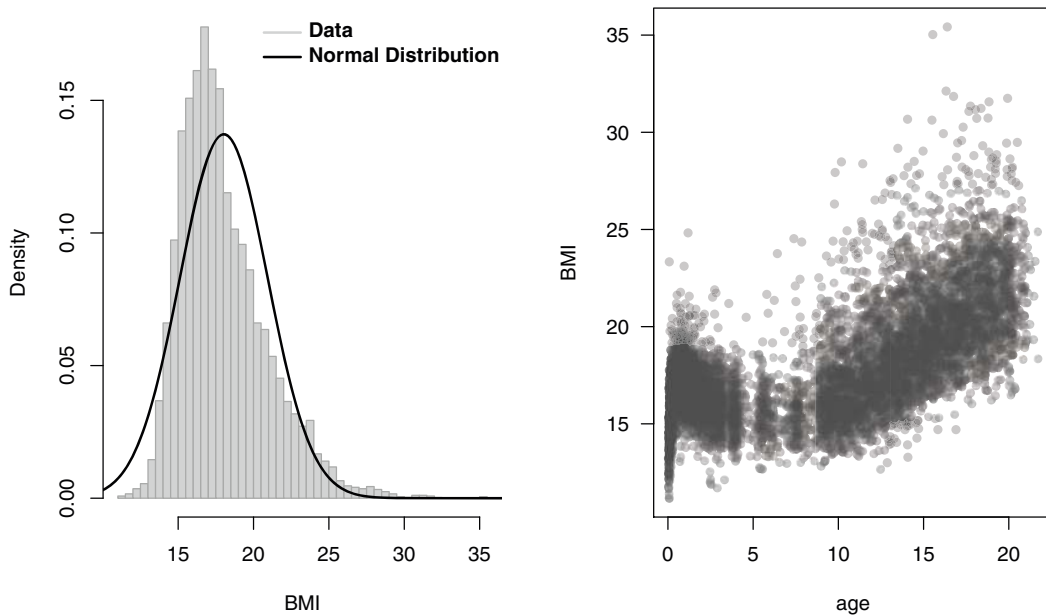


Figure 1 Histogram of the BMI from the Dutch boys dataset with the Gaussian distribution with mean and standard deviation taken from the dataset

In the case of the Gaussian distribution, just as for most distributions from a parametric family, those statistics are known and can be calculated for arbitrary

parameters. As we can see in the left panel of Figure 1, however, the shape of the fitted Gaussian distribution is quite different from that of the observed distribution of the data, as the histogram is quite asymmetric in comparison to the distribution curve. When one is interested in the extreme values of the dataset (e.g., the 95% quantile), we would hence refrain from using the quantiles calculated via the distribution and simply use the quantile of the data itself: while the 95% quantile of the fitted normal distribution is 22.809, the 95% quantile of the data is 23.612. This is a relevant difference; the approximation with the Gaussian distribution would substantially misinform us. Researchers, however, hardly ever only deal with the simple description of the univariate dataset, but are more interested in the influence of covariates on a dependent variable or the prediction based on independent variables. In our case, we could imagine wanting to predict the BMI depending on the age of an individual. The plot in the right panel of Figure 1 displays the scatterplot of age and BMI. The asymmetry already visible in the univariate representation also shows in the scatterplot. When doing classical regression on this relation however, we would assume exactly the same thing as when plainly approximating our histogram with a Gaussian distribution: that the BMI conditional on the age is normally distributed and this assumption hence underlies all statistics we derive from this (such as the conditional 95% quantile). The solution to this problem is to use quantile regression, that is, calculating the impact of covariates on quantiles directly, rather than assuming an underlying conditional distribution. This tutorial article explains what quantile regression is and how it can be calculated.

The remainder of the article is structured in the following way: the second section deals with univariate quantile regression model illustrated by the aforementioned dataset on the BMI of Dutch boys. The dataset will be analysed with both standard regression method as well as with quantile regression and the differences will be discussed. The third section will present a second example dataset, which is then used to show an additive quantile regression model, containing different types of covariates. In the fourth section different estimation methods and related models will be introduced. The last section consists of a short summary and a guideline of when to use quantile regression.

The supplementary material to this article includes the commented code for both examples. They are estimated with component-wise gradient boosting (Hofner et al., 2017); for more detailed explanation for this method please see the tutorial on boosting (Mayr and Hofner, 2018).

2 Basic quantile regression

2.1 Data example I: The BMI in the Netherlands

As a first illustrative example, we present the Fourth Dutch Growth Study which was already mentioned in the introduction. The dataset originates from a cross-sectional study that measures growth and development of the Dutch population between 0 and 21 years. We use a subset of the data available in the R package `gamlss`

(Stasinopoulos et al., 2017), which includes 7 294 observations of male Dutch children. In the following, we will refer to the dataset as ‘Dutch boys’ dataset.

The variable of interest in this example will be the BMI of the individuals with a special focus on those who are overweight. Recall Figure 1, where the fitted Gaussian distribution yields $P(\text{BMI} > 30) = 1F(\text{BMI} < 30) = 1.91 \cdot 10^{-5}$, while the empirical probability (i.e., the proportion of participants with $\text{BMI} > 30$) is 0.0021, which is two orders of magnitude larger. This is something that already shows in the comparison of the histogram to the theoretical distribution: the data deviates substantially from a Gaussian distribution. But what if we ask questions such as: Which BMI values should we consider to be comparatively high? Where on the scale are the 5% of children with the highest BMI? Which value does y have to have such that

$$P(\text{BMI} < y) = 95\%$$

holds? The correct answer to those questions is the 95% quantile of the distribution, which can be defined in the following way: ‘the τ quantile is the value y , for which 100 τ % of the values are smaller or equal than this X ’ or (in a formula):

$$Q_\tau(y) = F_y^{-1}(\tau). \quad (2.1)$$

The 95% quantile of the fitted normal distribution is 22.809, the 95% quantile of the data is 23.612. Hence, 5% have a higher BMI, and the rest has a lower BMI. Again theoretical and empirical result differs a lot, the distribution can thus not be considered to be Gaussian.

When only analysing the BMI, we ignore the fact that its conditional distribution varies with age, as can be seen in Figure 1 on the right side. The questions asked should not be restricted to the simple quantile (2.1) but be a quantile conditioned on age:

$$Q_\tau(y|x) = F_{y|x}^{-1}(\tau|x) = y_\tau(x), \quad (2.2)$$

where x is age. The most simple explanation for this formula is that for each τ it determines a function $y_\tau(x)$ bisecting the data such that, for any age x , 100 τ % of the data lie below the value of the function at that age. The function defined in this way splits the data into a τ and a $1 - \tau$ proportion, which is based on the covariate age. This model was first established by Koenker and Bassett (1978) by introducing a linear quantile regression and explaining a first algorithm that is able to calculate the parameters desired to choose a β such that

$$Q_\tau(y|X) = X\beta_\tau,$$

with X being the covariate matrix just as in usual mean regression. In order to cut off 100 τ % proportions sensibly in the case of the plot in the right panel of Figure 1, the quantile regressions cannot simply consist of linear functions, since the relation between age and BMI does not seem to be a linear one. In the following, we will use

penalized splines (P-splines) (Eilers and Marx, 1996) in order to model the different regression curves. P-splines are a special way of modelling non-linear effects. For a detailed explanation see Fahrmeir et al. (2013).

2.2 Modelling the BMI of Dutch boys

In the following, both mean regression and quantile regression models will be applied to the presented dataset.

2.2.1 Mean regression

The conventional non-linear regression model estimating the association of age with the expected value of BMI was calculated with R-package *mgcv* (Wood, 2014). As can be seen in Figure 2, the resulting curve lies in the centre of the data, but is unaffected by the asymmetry of the data, which is especially pronounced at higher ages.

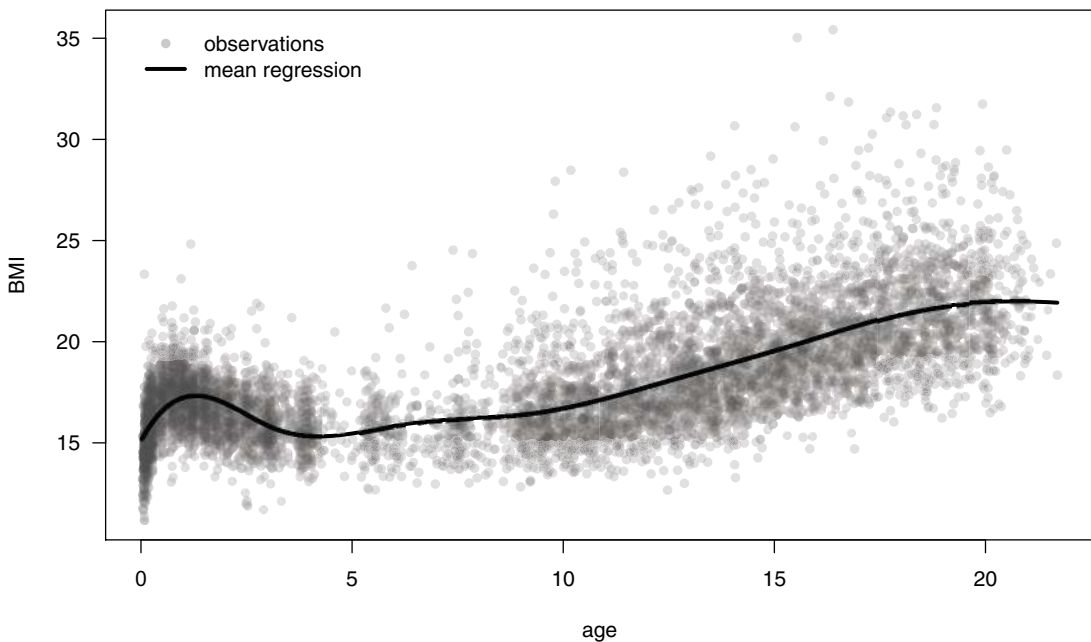


Figure 2 Scatterplot of age and BMI of the dataset Dutch boys with the regression line from a non-linear Gaussian regression model

The assumptions for conventional mean regression models which are violated quite obviously in this dataset are homoscedasticity and the symmetry of the Gaussian distribution. In addition to the mean curve, the quantiles which are

implied by the assumption of a Gaussian distribution are displayed for $\tau = (0.01, 0.1, 0.2, 0.3, 0.4, 0.6, 0.7, 0.8, 0.9, 0.99)$. These are calculated based on the variance obtained, when assuming a Gaussian distribution for the errors. These quantile curves do not fit the data very well for low and high BMI, which makes it even clearer that the assumption of parallel quantiles implied by the constant variance assumed in conventional regression is not met at all in the data, yet it is assumed in the requirements of the mean regression model. The BMI of the boys varies more the older they get. This increase in variation is not symmetrical. While the data is only slightly less dense for low BMIs, it is quite sparse in higher BMIs. It is unfavourable that exactly the part of the distribution which is of medical interest (i.e., potentially overweight children) is captured so badly.

2.2.2 Quantile regression

When modelling the quantiles independently of distributional assumptions yet conditional on the data (i.e., $Q(Y|X)$), we get the model as displayed in the right panel of Figure 3. The quantile-specific regression curves show differences to the ones extracted from the aforementioned Gaussian model in both aspects. The asymmetry of the distribution is captured by the approach. If the distribution was symmetrical, the distance between the median regression and the curve for $\tau = 0.6$ should be the same as the difference between the median regression and the curve for $\tau = 0.4$. The distance between $\tau = 0.4$ and $\tau = 0.3$ should be the same as the difference between $\tau = 0.6$ and $\tau = 0.7$, etc. This assumption does not hold in any of the quantile curves. The difference which is most obvious can be seen in the distance between the first and second (i.e., the $\tau = 0.01$ and the $\tau = 0.1$) quantile and the distance between the second to last and the last (i.e., the $\tau = 0.9$ and the $\tau = 0.99$) quantile. This reflects the positive skewness of the data which was already visible in Figure 1 in the left panel. The second central difference to the model shown earlier is the capturing of the heteroscedasticity. Again, the effect is most pronounced in the extremal quantiles: the increasing difference between the $\tau = 0.01$ quantile and the $\tau = 0.99$ quantile shows drastically how the variance in the BMI increases with increasing age. The difference in behaviour of the curves ultimately shows the importance of the approach: deviations from a conditional normal distribution could not have been captured without it. A second visualization of the fact that the conditional distribution is skewed is Figure 4, in which we can see that the mean is slightly above the median.

A third feature of quantile regression is the robustness against outliers. As can be seen in Figure 4, the median regression simply divides the data into two 50% parts. This figure also shows that the values scattered further away from the dense central cloud (e.g., those close to 35) only have an influence on the 99% quantile regression curve and not the more central curves. This, however, is not a problem since those are the data points that should be captured and described when looking at the 1% of the population with the highest BMI. The fact that the mean regression is above the 50% quantile regression curve is due to the long tail of the conditional distribution of the BMI.

3 Additive quantile regression

3.1 Data example II: Stunting in India

For the second data example, a dataset called `india` from the R package `gamboostLSS` (Hofner et al., 2017) was used (please note that the original dataset is publicly available on <http://www.measuredhs.com> after registration). It includes a number of outcomes from an Indian study on malnutrition in children up to the age of 3 which was conducted in the years 1998 and 1999. Since the process of growth is inhibited by malnutrition, stunting itself can be used as indicator for nutritional status of children. The outcome of interest in this case is hence stunted growth represented by a z-score measuring growth ranging from -6 to 6 , where negative values mean that a child is too short for its age, which is an indicator of prolonged malnutrition. Children with values below -2 are considered stunted (height-for-age). Since India has a severe problem with malnutrition among children, the interest in this dataset is very different from the the Dutch dataset. While the focus in the previous section was on analysing the higher quantiles due to the interest in risk of overweight, in this case,

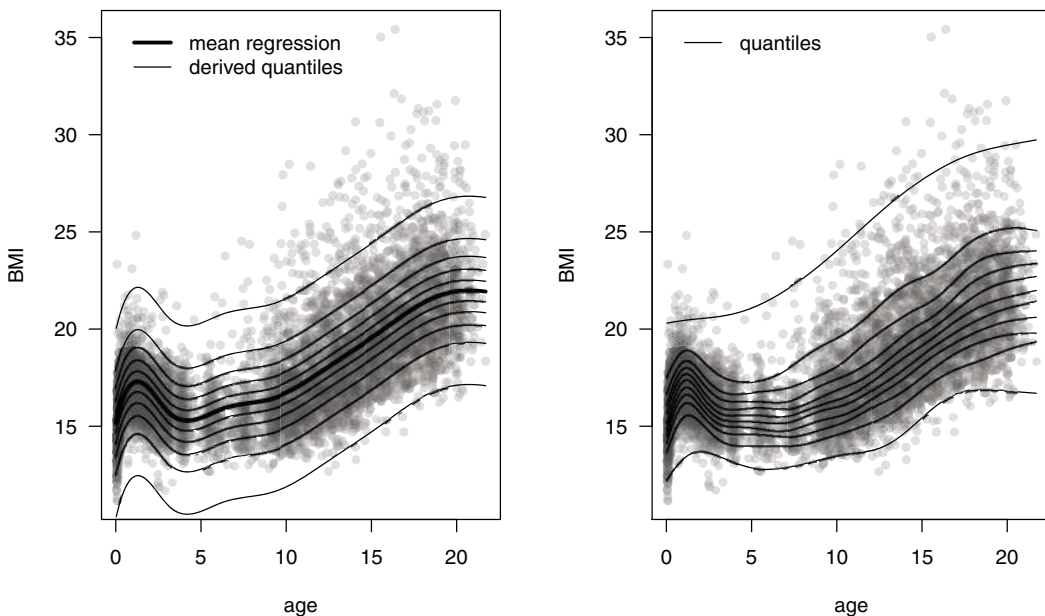


Figure 3 Scatterplot of age and BMI of the dataset Dutch boys. Left hand side: the bold line is the regression line from a non-linear Gaussian regression model, the other lines depict the $\tau = (0.01, 0.1, 0.2, 0.3, 0.4, 0.6, 0.7, 0.8, 0.9, 0.99)$ quantiles of the mean regression model. Right hand side: the lines depict the $\tau = (0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.99)$ quantile regression models

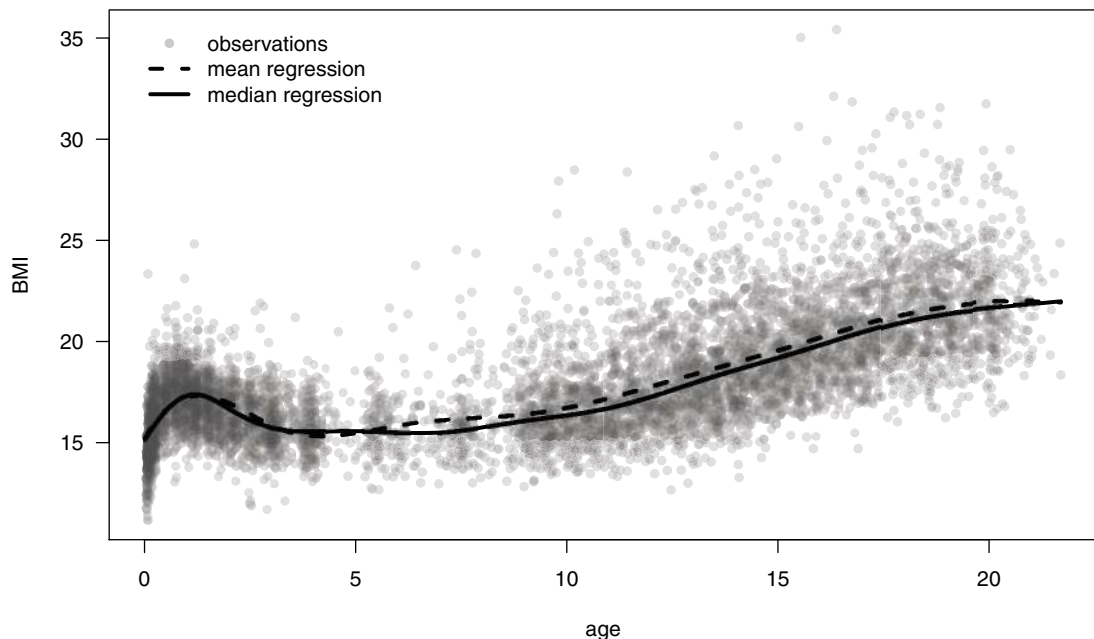


Figure 4 Scatterplot of age and BMI of the dataset Dutch boys. The solid line displays the median regression and the dashed line displays the mean regression

we are mainly interested in analysing the lower part of the distribution. The dataset contains the ‘z-score’ of 4 000 children, and the covariate ‘age’ in months, the ‘BMI of the mother’ and the ‘district’ the child was living in at the time of the study. The model described in the following section will measure the impact of the covariates on the z-score. For a further analysis of this dataset with a more content-related interpretation, see for example Fenske et al. (2013). The same data example is used in the tutorial article on boosting (Mayr and Hofner, 2018).

3.2 Modelling the Z-score of Indian children

The model presented in the previous section (i.e., a model with one continuous explanatory variable) is a comparatively simple explanatory model. Quantile regression methods were extended, however, in many different ways towards more complex model classes, just like mean regression, such as for models with measurement errors by Wei and Carroll (2009), time series by Kley et al. (2016), or additive, partly non-linear models by Lee et al. (2010). For an overview over the progress that has been made, see the paper of Koenker (2017); for the range of types of effects that can be included in structured additive quantile regression, see Kneib (2013). In the following, we present an additive model for the dataset on stunting in

India described earlier. The model includes three covariates, being modelled in two different ways. While the influence of the age of the child—just as for the Dutch boys dataset—and the BMI of the mother will be modelled by P-splines (Eilers and Marx, 1996), the geographical differences will be modelled by a Markov random field. This method allows all districts to have their own effect on the dependent variable (or in our case the quantile of interest of the dependent variable), but it does not allow neighbouring districts to vary too much from each other. This leads to a similar type of penalization as in P-splines and thus also to a smooth surface (Fahrmeir et al., 2004). The model formula hence looks like this:

$$Q_{\tau}(y_{\text{stunting}}|\mathbf{X}) = \beta_0 + f_{\tau\text{nl}}(\mathbf{x}_{\text{age}}) + f_{\tau\text{nl}}(\mathbf{x}_{\text{BMI}}) + f_{\tau\text{spat}}(\mathbf{x}_{\text{district}}),$$

where β_0 is an intercept and the f stands for non-linear (nl) and spatial (spat) functions. Due to the predictor being a sum of different functions, this type of model is called ‘additive model’. There are many different ways to estimate the effects of those kinds of models, three of which will be listed in the next section. For this example, the component-wise gradient ‘boosting’ approach (Fenske et al., 2011) with the following call was used:

```
qr_boost <- gamboost(stunting ~
  age + BMI +
  bmr(f(district, "markov", bnd=india.bnd, center=F),
  family = QuantReg(tau = 0.3),
  data = india)
```

The call `gamboost` combined with the specification `family = QuantReg(tau = 0.3)` produces an additive quantile regression model for the $\tau = 0.3$ quantile. The algorithm is programmed such that it recognizes age and BMI to be continuous variables and thus automatically assigns them smooth non-linear effects. The inner function `bmr(f(district, "markov", bnd=india.bnd, center=F))` uses the regions in the variable `districts` which are linked to an additional data object `india.bnd`. This object is a so-called ‘boundary file’ which includes information on the districts and their boundaries. The “markov” tells the function to use the aforementioned Markov properties (i.e., values $f_{\tau\text{spat}}(\mathbf{x}_{\text{district}})$ from neighbouring districts to be dependent in certain ways). The boosting algorithm can, besides estimating the effects, also carry out variable selection. This feature will not be described here in detail, but we note that it can be useful when dealing with a large number of possibly influential variables and being interested in variable selection in general. The models obtained by this algorithm have the advantage of a very good interpretability.

A set of models with different quantile levels $\tau = (0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.99)$ was calculated. The estimated effects for these models are illustrated by Figures 5–7. In Figure 5, the effect of age on the different quantiles

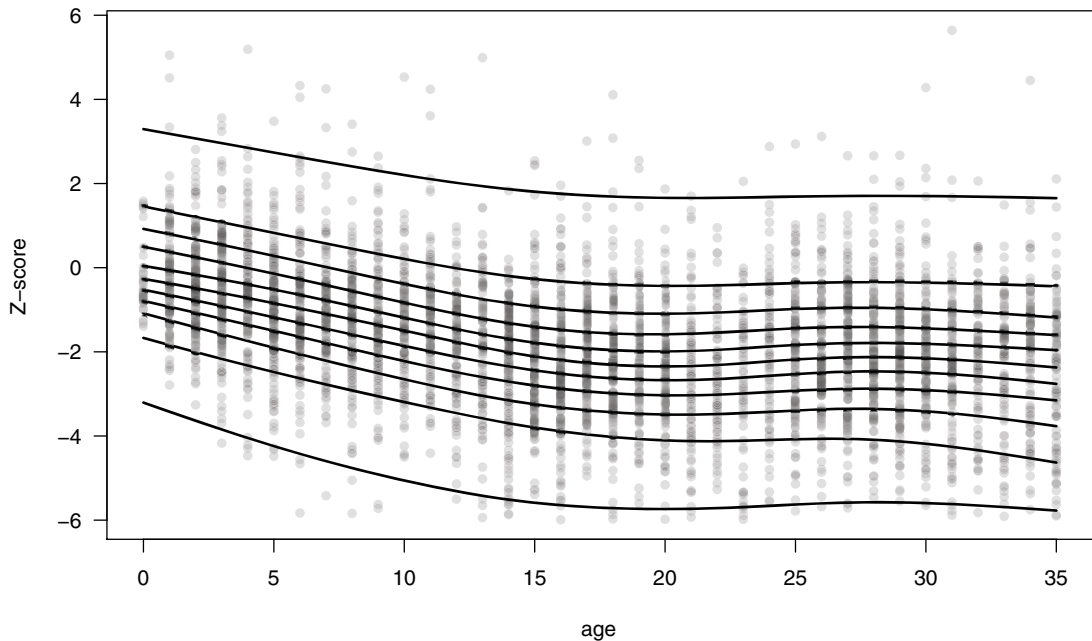


Figure 5 Scatterplot of age and z-score of the dataset on stunting in India. The lines depict the $\tau = (0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.99)$ quantile regression models adjusted for the rest of the data

is displayed. In order to be able to plot the non-linear effects into one graph, each quantile is shifted to the level of the predicted values for this quantile.

To this end, we use a centred version of the effect and then add the mean of the conditional quantile (i.e., the average over the prediction for all individuals) of interest. The different shape of the curves for different quantiles can be seen, just like in Section 2. Especially, the steeper decline for the lowest quantile ($\tau = 0.01$) in the early months attracts the attention. This indicates that malnutrition worsens with increasing age in the lower part of the distribution, that is, for children who are already in a more severe state. Figure 6 shows how in certain data situations, problems can occur. In this case, the effect of the BMI of the mother on the z-score is displayed. The estimation of the curves is increasing nearly linearly for lower BMI values, in most quantiles. In high BMI values, the central quantile, curves are flattening out. In the highest quantile, the impact of the BMI of the mother is rising until a BMI of 25 and then falling again. Estimations for higher levels of BMI of the mother are less accurate, since there are fewer observations.

The problem can be illustrated by a simple example: there are 7 women that have a BMI higher than 35. Thus, between 35 and 40, all 11 functions are estimated based on those 7 data points, which is very little information. We hence try to construct functions, which split 7 data points into 12 groups, without even informing one

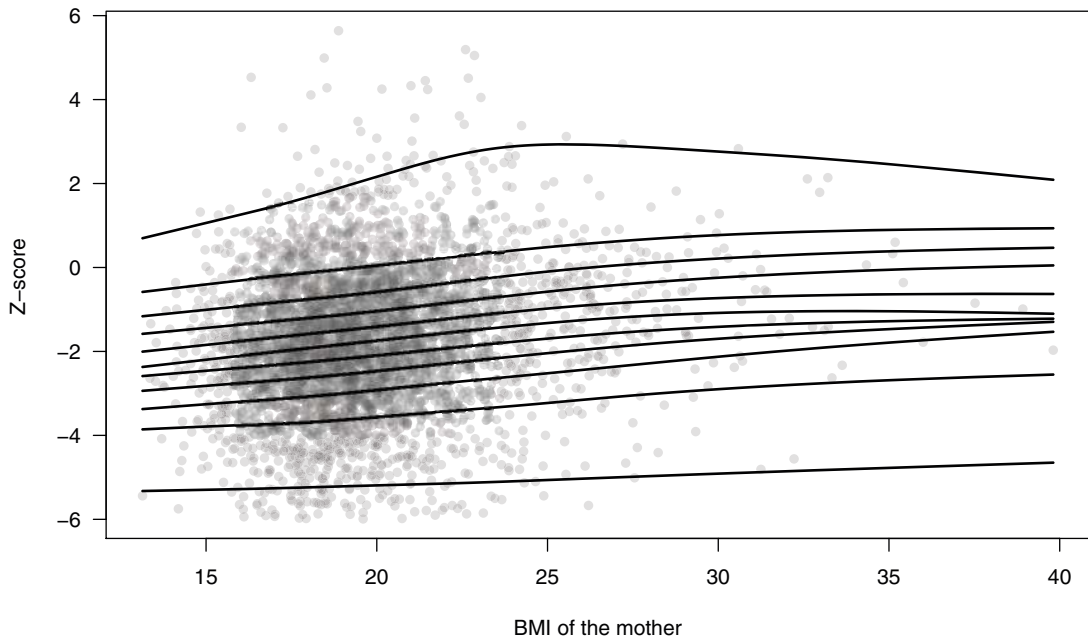


Figure 6 Scatterplot of the BMI of the mother and z-score of the dataset on stunting in India. The lines depict the $\tau = (0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.99)$ quantile regression models adjusted for the rest of the data

function about the position of the rest. When estimating the models independently for the different quantiles (as we did here), this can even lead to the quantiles to cross. For references to methods avoiding crossing see Section 4. Figure 7 displays the spatial effects for the quantiles $\tau = 0.2$, $\tau = 0.5$ and $\tau = 0.8$. In all three plots, the estimations are smooth in the sense that neighbouring districts do not differ very much in terms of effects. Between quantiles, however, a few differences are present. As an exemplary region Bihar was chosen. This region has a small frame around it in the upper row of the figure and is displayed in more detail in the lower row of the figure. While a negative trend can be spotted in the 0.2 and 0.5 quantiles, this effect is not visible at all in the 0.8 quantile.

4 Implementations and related models

In this article a gradient boosting approach was used to produce the results for the illustrative example. This could however have been done with many different approaches, which will be explained in the following.

Quantile regression has many advantages, but a major disadvantage is that parameters are harder to estimate than in Gaussian or generalized regression.

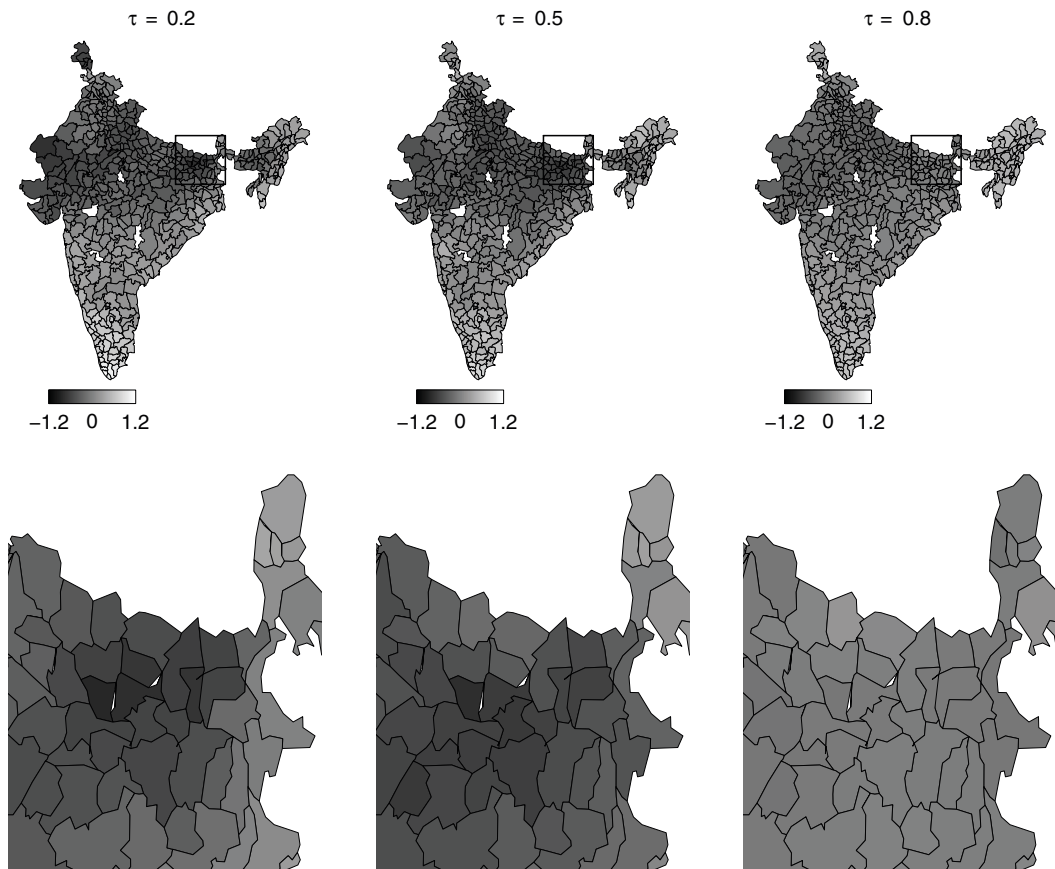


Figure 7 Effect of the district on the $\tau = 0.2$, the $\tau = 0.5$ and the $\tau = 0.8$ quantiles. The framed part contains the district of Bihar, which is further displayed in more detail

Inference on them can get complicated because the estimators for coefficients are not available in closed form. There are many ways to estimate quantile regression parameters; this will be sketched out in the following three ways. The first and original way is a linear optimization algorithm and was brought forward by Koenker and Bassett (1978) together with the original proposal. This approach is implemented in the R-package `quantreg` (Koenker, 2016), which includes linear as well as non-linear effects. Confidence intervals for the latter are based on piece-wise linear approximations and when using too many effects, the algorithm shows instabilities (Kneib, 2013). Model diagnostics rely on the construction of a statistic related to the R^2 in the mean regression case. The statistic is called R^1 and uses the minimizers of the check function for the models which have to be compared (Koenker and Machado, 1999). A second option is using boosting algorithms (Fenske et al., 2011) to

estimate sophisticated quantile regression models, which can handle a whole range of different effects, such as linear, non-linear and spatial effects; however, it lacks of a straight forward implementation of p-values and confidence intervals of the estimated regression parameters. This algorithm is implemented in the R package `mboost` (Hothorn et al., 2016). To find the optimal quantile regression model, however, appropriate tuning parameters like the number of iterations the algorithm has to perform have to be chosen. In the web-appendix, the example shown in this article is presented with comments on how to deal with those requirements. Boosting algorithms do not naturally produce p-values, but they can be generated post hoc. The gradient boosting algorithm is used in this article because the functions in the R package (Hothorn et al., 2016) are straightforward to apply. A third option is to use Bayesian inference (see e.g., Yu and Moyeed, 2001). An implementation of a structured additive model can be built in the stand-alone MCMC tool BayesX (Belitz et al., 2013). Due to the Bayesian structure, the algorithm delivers distributions of the estimated parameters. Uncertainty can hence be captured approximately, but variances of parameters are usually underestimated (Waldmann et al., 2013). Model diagnostics in Bayesian quantile regression do in general rely on the auxiliary distribution and not a real distributional assumption. Comparison between models is hence possible yet the comparison with models generated by other approaches is not possible. The Bayesian model involves a well-calibrated specification of the prior distributions and hyper parameters for optimal modelling. In most cases, they should be chosen as little informative as possible. For more information on how to chose appropriate priors and hyperprior parameters see, for example, Berger (2006). One major drawback of quantile regression is that most estimation techniques can lead to quantile crossing. This problem is mainly caused by the fact that those techniques estimate the models for all quantiles independently. There are, however, methods which prevent this problem, like for example the so-called quantile sheets Schnabel and Eilers (2013).

Another form of regression that exceeds the form of simply modelling the mean but does not have the issue of crossing quantiles is the generalized additive model for location, shape and scale (GAMLSS Rigby and Stasinopoulos, 2005). For a tutorial paper on this topic see Stasinopoulos et al. (2018). In this type of regression, all parameters of the assumed distribution can be modelled by covariates. The model is estimated as a whole, hence problems similar to the quantile crossing issue cannot occur in this case. This method is in some cases more stable than quantile regression (Kneib, 2013) and delivers similar results, as long as the assumed distribution is close enough to the real data distribution. This can be checked either graphically or by cross-validation. The Bayesian approach called distributional regression works analogously to this approach (Umlauf and Kneib, 2018). A related idea which has been explored less is the so-called Bayesian density regression (Dunson et al., 2007). In this case, the distribution of the dependent variable does not have to be specified beforehand but drawn in a Dirichlet process mixture scheme. A further model form closely related to this kind of model are the so-called conditional transformation models (Hothorn, 2018). Those models have the advantage of flexibly estimating the conditional distribution of the variable of interest.

There is one further model type which is very similar to quantile regression. Expectile regression is a generalization of the mean regression in the same way that quantile regression is a generalization of the median regression. The resulting regression functions minimize asymmetrically weighted squared residuals, where the weights are the same as the weights in quantile regression. The interpretation of those models is slightly harder than for quantile regression yet it is very useful in the analysis of financial risks, for example, the expected shortfall (Taylor, 2008).

5 Summary and Recommendations

Our goal in this article is to explain quantile regression in a simple way, in order to provide some guidance in the decision for which research questions it should be used. When only being interested in the mean and its characteristics and predictions, standard tools can already provide enough information. Quantile regression on the other hand is suitable for the following situations:

When being interested in events at the ‘limits of probability’: One example for this type of question was described in this article using the data from studies on stunting and obesity. Similar models were used in many publications such as in Fenske et al. (2011). In general, many medical questions might not search for answers on the centre of a distribution. Reliable results for extreme quantiles can however only be produced if large datasets are available.

If the conditional distribution does not follow a known distribution: For approaches such as GAMLSS or even glms, it is important to know the conditional distribution. Datasets hardly ever follow those distributions perfectly. This is however only important if the deviation changes the answers and if the focus is on the entire conditional distribution. In case of doubts, it makes sense to simply run both types of analyses (GAMLSS and quantile regression) and then compare the results. In many cases, graphical analyses of the data will shed light on the question for the appropriate distribution.

If there are a lot of outliers in the conditional distribution: Quantile regression is able to cope with outliers better due to its robustness. Outliers only have an influence on the quantile curves close to them.

In the case of heteroscedasticity: If the variance depends on the covariates, quantile regression can capture this effect. Just like earlier, most datasets will slightly violate the assumption of homoscedasticity and the model should be chosen based on the effect this violation has. In this case, just like for violation of the distributional assumptions, graphical analyses should be done to find out whether there is heteroscedasticity in the dataset.

Acknowledgements

We want to thank the support of the Interdisciplinary Centre for Clinical Research (IZKF) of the Friedrich-Alexander-University Erlangen-Nürnberg (Project J61). Special thanks goes to Graeme Hickey for proofreading the article.

References

- Belitz C, Brezger A, Kneib T, Lang S and Umlauf N (2013) *BayesX: Software for Bayesian inference in structured additive regression models*. URL <http://www.BayesX.org/>. Version 2.1
- Berger J (2006) The case for objective Bayesian analysis (with discussion). *Bayesian Analysis*, **1**, 385–402.
- Dunson DB, Pillai N and Park J-H (2007) Bayesian density regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69**, 163–83. doi: 10.1111/j.1467-9868.2007.00582.x. URL <http://dx.doi.org/10.1111/j.1467-9868.2007.00582.x>
- Eilers PHC and Marx BD (1996) Flexible smoothing using B-splines and penalized likelihood. *Statistical Science*, **11**, 89–121.
- Fahrmeir L, Kneib T and Lang S (2004) Penalized structured additive regression for space-time data: A Bayesian perspective. *Statistica Sinica*, **14**, 731–61.
- Fahrmeir L, Kneib T, Lang S and Marx B (2013) *Regression: Models, Methods and Applications*. New York, NY: Springer-Verlag.
- Fenske N, Burns J, Hothorn T and Rehfues EA (2013) Understanding child stunting in India: A comprehensive analysis of socio-economic, nutritional and environmental determinants using additive quantile regression. *PLOS One* **8**.
- Fenske N, Kneib T and Hothorn T (2011) Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. *Journal of the American Statistical Association*, **106**, 494–510.
- Frisch M (1959) *Homo Faber*. London: Abelard-Schuman.
- Hofner B, Mayr A, Fenske N and Schmid M (2017) *gamboostLSS: Boosting methods for GAMLSS models*. R package version 2.0-0. URL <https://CRAN.R-project.org/package=gamboostLSS>
- Hothorn T (2018) Top-down transformation choice. *Statistical Modelling*, **18**, 274–98.
- Hothorn T, Buehlmann P, Kneib T, Schmid M and Hofner B (2016) *mboost: Model-based boosting, R package* version 2.7-0. URL <https://CRAN.R-project.org/package=mboost>
- Kley T, Volgushev S, Dette H and Hallin M (2016) Quantile spectral processes: Asymptotic analysis and inference. *Bernoulli*, **22**, 1770–1807. doi: 10.3150/15-BEJ711
- Kneib T (2013) Beyond mean regression. *Statistical Modelling*, **13**, 275–303.
- Koenker R (2016) *quantreg: Quantile regression*. R package version 5.24. URL <https://CRAN.R-project.org/package=quantreg>
- (2017) Quantile regression: 40 years on. *Annual Review of Economics*, **9**, 155–76.
- Koenker R and Bassett G (1978) Regression quantiles. *Econometrica*, **46**, 33–50.
- Koenker R and Machado JAF (1999) Goodness of t and related inference processes for quantile regression. *Journal of the American Statistical Association*, **94**, 1296–1310.
- Lee YK, Mammen E and Park BU (2010) Backfitting and smooth backfitting for additive quantile models. *The Annals of Statistics*, **38**, 2857–83.
- Mayr A and Hofner B (2018) Boosting for statistical modelling: A non-technical introduction. *Statistical Modelling*, **18**, 365–84.
- Rigby RA and Stasinopoulos DM (2005) Generalized additive models for location, scale and shape. *Journal of the Royal*

- Statistical Society: Series C (Applied Statistics)*, **54**, 507–54.
- Schnabel SK and Eilers PHC (2013) Simultaneous estimation of quantile curves using quantile sheets. *AStA Advances in Statistical Analysis*, **97**, 77–87.
- Stasinopoulos M, Rigby B, Voudouris V, Akantziliotou C, Enea M and Kiose D (2017) *gamlss: Generalised Additive Models for Location Scale and Shape*. R package version 5.0-6. URL <https://CRAN.R-project.org/package=gamlss>
- Stasinopoulos M, Rigby RA and de Bastiani F (2018) A distributional regression approach using *gamlss*. *Statistical Modelling*, **18**, 248–73.
- Taylor JW (2008) Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, **6**, 231–52.
- Umlauf N and Kneib T (2018) A primer on bayesian distributional regression. *Statistical Modelling*, **18**, 219–47.
- Waldmann E, Kneib T, Yue YR, Lang S and Flexeder C (2013) Bayesian semiparametric additive quantile regression. *Statistical Modelling*, **13**, 223–52.
- Wei Y and Carroll RJ (2009) Quantile regression with measurement error. *Journal of the American Statistical Association*, **104**, 1129–43.
- Wood S (2014) *mgcv: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation*. R package version 1.8-12. URL <https://CRAN.R-project.org/package=mgcv>
- Yu K and Moyeed RA (2001) Bayesian quantile regression. *Statistics & Probability Letters*, **54**, 437–47.