Bayesian Quantile Trees for Sales Management

Mauro Bernardi and Paola Stolfi

Abstract Sales management is a fundamental issue in retail commerce, being one of the main ingredient to maximise stores revenue. The huge amount of data accessible nowadays makes this task challenging. This paper proposes a new nonparametric method for predicting sales quantiles at different confidence levels, conditional on several explanatory variables such as the type of store, the location of the store, the type of product and its the price, etc, thereby providing a complete picture of the relation between the response and the covariates. Moreover, predicting extreme sales quantiles provide valuable information for building automatic stock management systems and for the sales monitoring. As concerns the methodology, we propose to approximate the conditional quantile at level $\tau \in (0,1)$ of the response variable using bayesian additive nonparametric regression trees. Decision trees and their additive counterparts are promising alternatives to linear regression methods because of their superior ability to characterise nonlinear relationships and interactions among explanatory variables that is of fundamental relevance to get accurate predictions.

Key words: Regression tree, bayesian methods, quantile regression, prediction.

1 Introduction

In empirical studies, researchers are often interested in analysing the behaviour of a response variable given the information on a set of covariates. The typical answer is to specify a linear regression model where unknown parameters are estimated using the Ordinary Least Squares (OLS). The OLS method estimates unknown parameters by minimising the sum of squared errors leading to the approximation of the mean function of the conditional distribution of the response variable. Although the mean represents the average behaviour of the response variable, it provides little or no information about the behaviour of the conditional distribution on the tails. As far as the entire distribution is concerned, quantile regression methods [Koenker and Bassett, 1978] adequately characterise the behaviour of the response variable at different con-

Mauro Bernardi
Department of Statistical Sciences, University of Padova, Via Cesare Battisti, 241, 35121, Padova, e-mail: mauro.bernardi@unipd.it

Paola Stolfi
Institute for applied mathematics “Mauro Picone” (IAC) - CNR, Rome, Italy, e-mail: p.stolfi@iac.cnr.it

fidence levels. Moreover, the quantile analysis is particularly suitable when the conditional distribution is heterogeneous, non–Gaussian, skewed or fat–tailed, see, e.g., [Lum and Gelfand, 2012] and [Koenker, 2005].

Quantile models admitting a linear representation have been extensively applied in different areas, see, e.g., [Yu et al., 2003], such as, finance, as direct approach to estimate the Value–at–Risk, i.e., the loss–level a financial institution may suffer with a given confidence [Bassett Jr. and Chen, 2001], economics and social sciences [Hendricks and Koenker, 1992], medicine [Heagerty and Pepe, 1999], survival analysis [Koenker and Geling, 2001] and environmetrics [Pandey and Nguyen, 1999]. Futhermore, linear quantile models have been theoretically investigated from both a Bayesian, [Sriram et al., 2013] and a frequentist point of view, and the properties of the resulting estimates has been deeply studied. See [Koenker et al., 2017] and [Davino et al., 2014] for an extensive and up to date review latest theoretical results on quantile methods and their interesting applications. However, despite their relevance and widespread application in empirical studies, linear quantile models provide only a rough "first order" approximation of the relationship between the τ–level quantile of the response variable and the covariates. Indeed, as first recognised by [Koenker, 2005], quantiles are linear functions only within a Gaussian world, thereby stimulating many recent attempts to overcome this limitation. [Chen et al., 2009], [Faugeras, 2009], [De Backer et al., 2017] and [Kraus and Czado, 2017], for example, consider the copula–based approach to formalise nonlinear and parametric conditional quantile relationships. The copula approach, although quite flexible in fitting marginal data, forget to consider nonlinear interactions among the covariates. This paper try to overcome the traditional limitations of linear quantile methods by extending the quantile approach to the promising field of decision trees. Decision trees are regression techniques that are very popular within the machine learning community that try to mitigate the relevant problem of parsimoniously modelling interactions among covariates. Indeed, decision trees partition the space of relevant covariates into pieces, usually hyper–rectangles, where observations are homogeneous. For example, when the objective is to model the average response, observations with the same unconditional variance are clustered together, while for classification problems observations are partitioned according to the Gini index. Since their introduction several theoretical and applied papers have contributed to the diffusion of such idea in different contexts. Both the machine learners and the statistics communities contributed to the development of tools and methods. An up to date and comprehensive review of the methods developed in the machine learning literature can be found for example in [Loh, 2014a], [Strobl, 2014], [Ciampi, 2014], [Ahn, 2014], [Song and Zhang, 2014], [Rusch and Zeileis, 2014], [Loh, 2014b]. The main drawback of decision trees is related to the high variance of the resulting forecasts. The most promising alternative approach [Breiman, 2001], namely, the random forest, is an ensemble of bootstrap decision trees that reduces the variance and provides also a straightforward way to assess the relevant covariates. On the likelihood–based side, the Bayesian estimation of decision trees have been considered in [Chipman et al., 1998], [Denison et al., 1998], [Sha, 2002] and [Wu et al., 2007] and extended to additive trees by [Chipman et al., 2010]. The main novelty of this latter approach relies on exploiting the likelihood of parametric models where regressors splitting rules play the role of hard thresholding operators that partition the overall model into local models.

In this paper, we consider the Bayesian approach, namely, we extend the Bayesian Additive Regression Tree (BART) of [Chipman et al., 2010] to model the quantile of the response variable. The Bayesian Additive Quantile Regression Trees (BAQRT) exploits the data augmentation approach that relies on the Asymmetric Laplace working likelihood, see [Bernardi et al., 2015], to provide a Metropolis–within–Gibbs sampling method that efficiently explores the regressors space. The data augmentation approach allows to effectively marginalise out the leaf parameters of the trees when changing the tree structures. Section 2 formalises the likelihood function and the prior structure of the BAQRT method. Quantile random forest (QRF) methods have been previously introduced in the machine learning literature by [Meinshausen, 2006]. [Meinshausen, 2006] exploits the original version of random forest for modelling the conditional mean to infer the structure of the tree, while assigning the empirical quantile of the observations falling into each terminal leaf instead of the mean value. Therefore, the QRF algorithm is not highly flexible to adapt the structure of the generated trees according to the modelled quantiles.

Given their appealing perspective of discriminating observations below and above a given quantile threshold, the quantile approach is a promising method for solving many statistical problems usually encountered in business and industry. In Section 3 we apply BAQRT to strategic sales management in retail stores. Sales management is one of the main issues in retail commerce. Indeed, it is one of the main ingredient to maximise the income of stores. The huge amount of data accessible nowadays makes this task challenging. Indeed, there are many potentially predictors that could be useful to predict and monitor sales but there is not any model to reach this task. This is the usual situation in which machine learning algorithms represent a powerful instruments to extract insight from such heterogeneous data. We consider the dataset provided by BigMart, an international brand with both free home delivery services and outlet store of food and grocery, to show how quantile regression trees could be useful in selecting the most relevant variables and analyse their impact both for predicting and monitoring tasks.

2 Quantile regression tree

The linear quantile regression framework for independent and identically distributed data models the conditional τ–level quantile of the response variable Y, with $\tau \in (0,1)$, as a linear function of the vector of dimension $(q \times 1)$ of exogenous covariates $\mathbf{X}$, i.e., $\mathcal{Q}_\tau(Y \mid \mathbf{X} = \mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$, thereby avoiding any explicit assumptions about the conditional distribution of $Y \mid \mathbf{X} = \mathbf{x}$. From a frequentist perspective, within a likelihood–based framework, this is equivalent to assume an additive stochastic error term ε for the conditional regression function $\mu(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$ to be independent and identically distributed with zero τ–th quantile, i.e, $\mathcal{Q}_\tau(\varepsilon \mid \mathbf{x}) = 0$, and constant variance. Following [Yu and Moyeed, 2001] and [Bernardi et al., 2015], the previous condition is implicitly satisfied by assuming that the conditional distribution of the response variable Y follows an Asymmetric Laplace (AL) distribution located at the true regression function $\mu(\mathbf{x})$, with constant scale $\sigma > 0$ and shape parameter τ, i.e., $\varepsilon \sim \mathsf{AL}(\tau, \mu(\mathbf{x}), \sigma)$, with probability density function

$$\mathsf{AL}(Y \mid \mathbf{X}, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\frac{1}{\sigma}\rho_\tau(Y - \mu(\mathbf{x}))\right\} \boldsymbol{I}_{(-\infty,\infty)}(Y), \tag{1}$$

where $\mu(\mathbf{x})$ is the regression function and $\rho_\tau(u) = u\left(\tau - \boldsymbol{I}_{(-\infty,0)}(u)\right)$ denotes the quantile check function at level τ. The quantile regression model postulated in equation (1) assumes the AL distribution as a misspecified working likelihood that correctly identify the conditional quantile function.

Similarly to the Bayesian Additive Regression Tree approach of [Chipman et al., 2010] for modelling the conditional mean of the response variable, the quantile regression tree approach extends the linear quantile model defined in equation (1) by assuming a sum–of–trees ensemble for the regression function $\mu(\mathbf{x})$. Specifically, the Bayesian Additive Quantile Regression Tree (BAQRT) model can be expressed as

$$Y = \mu(\mathbf{x}) + \varepsilon \tag{2}$$

$$\approx \mathcal{T}_1^{\mathcal{M}}(\mathbf{x}) + \mathcal{T}_2^{\mathcal{M}}(\mathbf{x}) + \cdots + \mathcal{T}_m^{\mathcal{M}}(\mathbf{x}) + \varepsilon, \tag{3}$$

where $\varepsilon \sim \mathsf{AL}(\tau, 0, \sigma)$. The assumption about the error term in equation (3) implies that $\mu(\mathbf{x}) = \mathcal{Q}_\tau(Y \mid \mathbf{X} = \mathbf{x})$. Furthermore, in equation (3) we assume that the quantile of the response variable is an additive function of

$m \geq q$ regression trees, each composed by a tree structure, denoted by $\mathscr{T}$, and the parameters of the terminal nodes (also called leaves), denoted by $\mathscr{M}$. Therefore, the j–th tree for $j = 1, 2, \dots, m$, denoted by $\mathscr{T}_j^{\mathscr{M}}$, represents a specific combination of tree structure $\mathscr{T}_j$ and tree parameters $\mathscr{M}$, i.e., the regression parameters associated to its terminal nodes. The tree structure $\mathscr{T}_j$ contains information on how any observation y_i, in a set of n independent and identically distributed observations $\mathbf{y} = (y_1, y_2, \dots, y_n)$, recurses down the tree specifying a splitting rule for each non–terminal (internal) node. The splitting rule has the form $x_k \leq c$ and consists of the splitting variable x_k and the splitting value $c \in \mathbb{R}$. The observation y_i is assigned to the left child if the splitting rule is satisfied and to the right child, otherwise, until a terminal node is reached and the value of the leaf of that terminal node is assigned as its predicted value. Therefore, the quantile prediction corresponding to y_i assigned by the sum of regression tree specified in equation (3) is the sum of the m leaf values. Hereafter, we denote by $\mathscr{M}_j = \left\{\mu_{j,1}, \mu_{j,2}, \dots, \mu_{j,b_j}\right\}$ the set of parameters associated to the b_j terminal nodes of the j–th tree, where $\mu_{j,l}$, for $l = 1, 2, \dots, b_l$ denotes the conditional quantile predicted by the model.

The additive quantile regression tree specified in equation (3) provides a natural framework for likelihood–based inference on the set of quantile regression parameters, i.e., the location parameters associated to the terminal nodes of each tree belonging to the ensemble. However, additional prior information should be imposed in order to infer the structure of the each tree. The next Section discusses the likelihood and the prior structure for both the model parameters and the trees.

2.1 Likelihood and prior

As discussed in [Yu and Moyeed, 2001], due to the complexity of the quantile likelihood function in equation (1), the resulting posterior density for the regression parameters does not admit a closed form representation for the full conditional distributions, and needs to be sampled by using MCMC–based algorithms. Following [Kozumi and Kobayashi, 2011] and [Bernardi et al., 2015], we instead adopt the well–known representation (see, e.g., [Kotz et al., 2001] and [Park and Casella, 2008]) of $\varepsilon \sim \mathsf{L}(\tau, 0, \sigma)$ as a location–scale mixture of Gaussian distributions:

$$\varepsilon = \zeta\omega + \varsigma\sqrt{\sigma\omega}\epsilon, \tag{4}$$

where $\omega \sim \mathsf{Exp}\left(\sigma^{-1}\right)$ and $\epsilon \sim \mathsf{N}(0,1)$ are independent random variables and $\mathsf{Exp}(\cdot)$ denotes the Exponential distribution. Moreover, the parameters ζ and ς^2 are fixed equal to

$$\zeta = \frac{1-2\tau}{\tau(1-\tau)}, \qquad \varsigma^2 = \frac{2}{\tau(1-\tau)}, \tag{5}$$

in order to ensure that the τ–th quantile of ε is equal to zero. The previous representation in equation (4) allows us to use a Gibbs sampler algorithm for sampling the trees parameters $\boldsymbol{\mu}_j$ of tree $j = 1, 2, \dots, m$, detailed in the next subsection. Exploiting the augmented data structure defined in equation (4), the additive quantile regression tree in (3) admits, conditionally on the latent factor ω, the following Gaussian representation:

$$Y \mid \omega = \mu(\mathbf{x}) + \zeta\omega + \varsigma\sqrt{\sigma\omega}\epsilon, \tag{6}$$

$$\approx \mathscr{T}_1^{\mathscr{M}}(\mathbf{x}) + \mathscr{T}_2^{\mathscr{M}}(\mathbf{x}) + \cdots + \mathscr{T}_m^{\mathscr{M}}(\mathbf{x}) + \zeta\omega + \varsigma\sqrt{\sigma\omega}\epsilon \tag{7}$$

$$\omega \sim \mathsf{Exp}\left(\sigma^{-1}\right). \tag{8}$$

The hierarchical model representation in equations (6)–(8) has the advantage of being conditionally Gaussian, leading to a conjugate Bayesian analysis for the parameters associated to the terminal nodes and the scale parameter as well.

The Bayesian inferential procedure requires the specification of the prior distribution for the unknown vector of model parameters $(\boldsymbol{\mu}, \sigma)$ and the structure of the tree. In principle, as discussed in the seminal paper of [Yu and Moyeed, 2001], non informative priors can be specified for the vector of regression parameters, i.e., $\pi(\boldsymbol{\mu}) \propto 1$. Alternatively, as in [Bernardi et al., 2015], the usual Normal–Inverse Gamma prior can be specified for regression and scale parameters, respectively, i.e.,

$$\mu_i \sim \mathsf{N}_1\left(\mu_0, \sigma_\mu^2\right) \tag{9}$$

$$\sigma \sim \mathsf{IG}\left(\frac{\eta_\sigma}{2}, \frac{\eta_\sigma \lambda_\sigma}{2}\right) \tag{10}$$

$$\mathbb{P}(\mathscr{T}) \propto \alpha (1+d)^{-\beta}, \tag{11}$$

where $\alpha \in (0,1)$ and $\beta \in [0, \infty)$ and d is the depth of the tree defined as the distance from the root. Here, N_1 denotes the univariate Normal density while IG is the Inverse Gamma distribution and $\left(\mu_0, \sigma_\mu^2, \eta_\sigma, \lambda_\sigma\right)$ are fixed hyperparameters, with $\eta_\sigma > 0$ and $\lambda_\sigma > 0$.

Now, let $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ be the vector of observations on the response variable Y and let $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$ be the associated matrix of covariates of dimension $(n \times q)$, then the joint prior distribution of the tree structure and tree parameters augmented by the latent factor ω, can be factorised as follows:

$$\begin{aligned}
\mathbb{P}\left(\mathscr{T}_1^{\mathscr{M}}, \mathscr{T}_1^{\mathscr{M}}, \ldots, \mathscr{T}_1^{\mathscr{M}}, \boldsymbol{\omega}, \sigma\right) &= \left[\prod_{j=1}^{m} \mathbb{P}\left(\mathscr{T}_j^{\mathscr{M}} \mid \boldsymbol{\omega}\right)\right] \mathbb{P}(\sigma) \\
&= \left[\prod_{j=1}^{m} \mathbb{P}(\mathscr{M}_j \mid \mathscr{T}_j, \boldsymbol{\omega}) \mathbb{P}(\mathscr{T}_j \mid \boldsymbol{\omega})\right] \mathbb{P}(\boldsymbol{\omega}) \mathbb{P}(\sigma) \\
&= \left[\prod_{j=1}^{m} \prod_{l=1}^{b_j} \mathbb{P}\left(\mu_{j,\ell} \mid \mathscr{T}_j, \boldsymbol{\omega}\right) \mathbb{P}(\mathscr{T}_j \mid \boldsymbol{\omega})\right] \mathbb{P}(\boldsymbol{\omega}) \mathbb{P}(\sigma) \\
&= \left[\prod_{j=1}^{m} \prod_{l=1}^{b_j} \phi\left(\mu_{j,\ell} \mid \mathscr{T}_j, \boldsymbol{\omega}\right) \mathbb{P}(\mathscr{T}_j \mid \boldsymbol{\omega})\right] \mathbb{P}(\boldsymbol{\omega}) \mathbb{P}(\sigma),
\end{aligned} \tag{12}$$

where $\mu_{j,\ell}$ denotes the parameter (conditional quantile) associated to the ℓ–th terminal nodes of tree $j = 1, 2, \ldots, m$, for $\ell = 1, 2, \ldots, b$, $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_n, \sigma)$ is the vector of auxiliary variables $\omega_i \sim \mathsf{Exp}\left(\sigma^{-1}\right)$, $\sigma \in \mathbb{R}$ is the scale parameter which is common to all the trees, and $\phi(\cdot)$ denotes the gaussian probability density function.

3 Application

Retail stores invest much effort in high level strategy to maximise their income. The type of store, its location, the furnitures and the product proposal are some of the main ingredients driving strategic decisions. Sales product prediction is therefore one of the most challenging problem in retail commerce, fundamental for

Variable Name	Description
Item Identifier	Unique product ID
Item Weight	Weight of product
Item Fat Content	Whether the product is low fat or not
Item Visibility	The % of total display area of all products in a store allocated to the particular product
Item Type	The category to which the product belongs
Item MRP	Maximum Retail Price of the product (list price)
Outlet Identifier	Unique store ID
Outlet Estabilishment Year	The year in which store was estabilished
Outlet Size	The size of the store in terms of ground area covered
Outlet Location Type	The type of city in which the store is located
Outlet Type	Whether the outlet is just a grocery store or some sort of supermarket
Item Outlet Sales	Sales of the product in the particular store. This is the outcome variable.

Table 1: Variables and their description for the BigMart dataset.

instance for commercialising new products, opening new stores or monitor the income performance of the stores.

In this work we apply BAQRT to analyse data coming from BigMart. It is an international brand with both free home delivery services and outlet store of food and grocery. Data scientists at BigMart created a dataset containing sales for 1559 products in 10 different stores located in several cities. They also reported many features related both to products and stores, in table 1 we provide a list of all the variables together with a description. In particular, there are eleven predictors and a scalar response function that is the "Outlet Sales". Eight of the eleven predictors are categorical while the remaining are continuous. The main interest consists in the identification of the variables that mostly influence the sales and if the relevance of these features changes by considering the different quantiles of the response variable. Indeed, the tail behaviour of the sales is also useful in stock management.

The variables considered in this example are a small subset of all the possible variables that could be analysed by retail stores. The importance of the variables are however based on some hypothesis and there is not any model that can be used both for sales prediction and monitoring tasks. This motivates the application of our method to investigate the type of relations occurring between the predictors and the response variable.

In figure 1 we report the predictor importance for the confidence levels $\tau = (0.1, 0.5, 0.9)$. The first two quantiles show similar variables importance ranking, in particular, the most relevant variables are "Item MRP", "Item Type", "Item Visibility" and "Item Weight". The first variable represents the price, its relevance supports the fact that promotional offers makes the customers much more inclined to buy products. The second one is the "Item Type" that represents the category to which the product belongs to. This finding supports the idea that one of the main indicator of product's sale is its utility, that is daily use products have much more sales rate than others. The third one, the "Item Visibility", confirms the fact that the position of the products in the stores is fundamental for their sales. Finally, the last "Item Weight" supports the hypothesis that lighter (and often smaller) products are easier to carry and so people are more inclined to buy them even if they are not needed.

The higher quantile instead show a different situations for the predictors' importance. Indeed, the "Item Visibility" becomes the most relevant variable, while all the others have a quite similar importance.

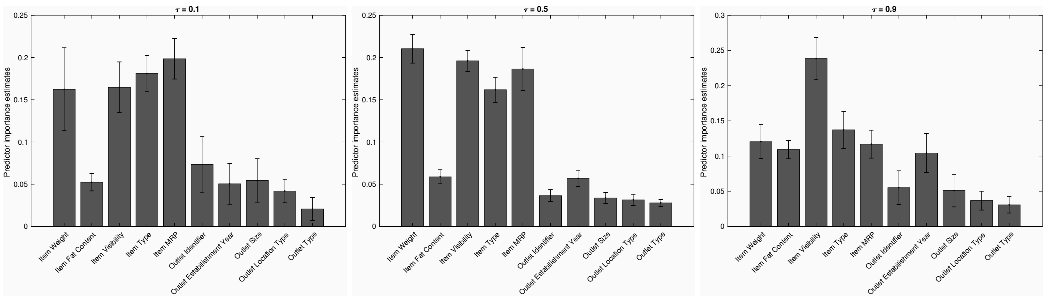

Fig. 1: Predictors' importance for different quantiles. The bottom figure in the left column refers to the mean behaviour.

References

[Ahn, 2014] Ahn, H. (2014). Discussion: "Fifty years of classification and regression trees" [mr3280974]. *Int. Stat. Rev.*, 82(3):357–359.

[Bassett Jr. and Chen, 2001] Bassett Jr., G. W. and Chen, H.-L. (2001). Portfolio style: Return-based attribution using quantile regression. *Empirical Economics*, 26(1):293–305.

[Bernardi et al., 2015] Bernardi, M., Gayraud, G., and Petrella, L. (2015). Bayesian tail risk interdependence using quantile regression. *Bayesian Anal.*, 10(3):553–603.

[Breiman, 2001] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

[Chen et al., 2009] Chen, X., Koenker, R., and Xiao, Z. (2009). Copula-based nonlinear quantile autoregression. *Econometrics Journal*, 12:S50–S67.

[Chipman et al., 1998] Chipman, H. A., George, E. I., and McCulloch, R. E. (1998). Bayesian cart model search. *Journal of the American Statistical Association*, 93(443):935–948.

[Chipman et al., 2010] Chipman, H. A., George, E. I., and McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *Ann. Appl. Stat.*, 4(1):266–298.

[Ciampi, 2014] Ciampi, A. (2014). Discussion: "Fifty years of classification and regression trees" [mr3280974]. *Int. Stat. Rev.*, 82(3):352–357.

[Davino et al., 2014] Davino, C., Furno, M., and Vistocco, D. (2014). *Quantile regression*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester. Theory and applications.

[De Backer et al., 2017] De Backer, M., El Ghouch, A., and Van Keilegom, I. (2017). Semiparametric copula quantile regression for complete or censored data. *Electron. J. Statist.*, 11(1):1660–1698.

[Denison et al., 1998] Denison, D. G. T., Mallick, B. K., and Smith, A. F. M. (1998). A Bayesian CART algorithm. *Biometrika*, 85(2):363–377.

[Faugeras, 2009] Faugeras, O. P. (2009). A quantile-copula approach to conditional density estimation. *Journal of Multivariate Analysis*, 100(9):2083 – 2099.

[Heagerty and Pepe, 1999] Heagerty, P. J. and Pepe, M. S. (1999). Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in us children. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48(4):533–551.

[Hendricks and Koenker, 1992] Hendricks, W. and Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association*, 87(417):58–68.

[Koenker, 2005] Koenker, B. (2005). *Quantile Regression*. Cambridge University Press, Cambridge.

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

[Koenker et al., 2017] Koenker, R., Chernozhukov, V., He, X., and Peng, L. (2017). *Handbook of Quantile Regression*. Chapman & Hall / CRC Handbooks of Modern Statistical Methods. CRC Press, Taylor & Francis Group.

[Koenker and Geling, 2001] Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: a quantile regression survival analysis. *J. Amer. Statist. Assoc.*, 96(454):458–468.

[Kotz et al., 2001] Kotz, S., Kozubowski, T., and Podgórski, K. (2001). *The Laplace distribution and generalizations. A revisit with Applications to Communications, Economics, Engineering, and Finance*. Birkhäuser.

[Kozumi and Kobayashi, 2011] Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81:1565–1578.

[Kraus and Czado, 2017] Kraus, D. and Czado, C. (2017). D-vine copula based quantile regression. *Computational Statistics & Data Analysis*, 110:1 – 18.

[Loh, 2014a] Loh, W.-Y. (2014a). Fifty years of classification and regression trees. *Int. Stat. Rev.*, 82(3):329–348.

[Loh, 2014b] Loh, W.-Y. (2014b). Rejoinder: "Fifty years of classification and regression trees" [mr3280975; mr3280976; mr3280977; mr3280978; mr3280979; mr3280974]. *Int. Stat. Rev.*, 82(3):367–370.

[Lum and Gelfand, 2012] Lum, K. and Gelfand, A. (2012). Spatial quantile multiple regression using the asymmetric laplace process. *Bayesian Analysis*, 7:235–258.

[Meinshausen, 2006] Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7(Jun):983–999.

[Pandey and Nguyen, 1999] Pandey, G. and Nguyen, V.-T.-V. (1999). A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology*, 225(1):92 – 101.

[Park and Casella, 2008] Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103:681–686.

[Rusch and Zeileis, 2014] Rusch, T. and Zeileis, A. (2014). Discussion: "Fifty years of classification and regression trees" [mr3280974]. *Int. Stat. Rev.*, 82(3):361–367.

[Sha, 2002] Sha, N. (2002). *Bolstering CART and Bayesian variable selection methods for classification*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–Texas A&M University.

[Song and Zhang, 2014] Song, C. and Zhang, H. (2014). Discussion: "Fifty years of classification and regression trees" [mr3280974]. *Int. Stat. Rev.*, 82(3):359–361.

[Sriram et al., 2013] Sriram, K., Ramamoorthi, R., and Ghosh, P. (2013). Posterior consistency of bayesian quantile regression based on the misspecified asymmetric laplace density. *Bayesian Analysis*, 8:479–504.

[Strobl, 2014] Strobl, C. (2014). Discussion: "Fifty years of classification and regression trees" [mr3280974]. *Int. Stat. Rev.*, 82(3):349–352.

[Wu et al., 2007] Wu, Y., Tjelmeland, H. k., and West, M. (2007). Bayesian CART: prior specification and posterior simulation. *J. Comput. Graph. Statist.*, 16(1):44–66.

[Yu et al., 2003] Yu, K., Lu, Z., and Stander, J. (2003). Quantile regression: applications and current research areas. *The Statistician*, 52(3):331–350.

[Yu and Moyeed, 2001] Yu, K. and Moyeed, R. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54:437–447.