Quantile regression is the last method that shall be used in the course of this work. It was first introduced by 54PI in 1978 and is a form of direct interval estimation what means a method that does not model a distribution of outcomes but is designed to directly output an interval. (33PI)

Following (33PI, 47PI, 54PI) Quantile regression does so by optimizing the parameters with respect to a loss function that is employed while model fitting. The idea is that one combination of parameters yields x overestimations and y underestimations of outcomes where it is the objective to balance outcomes above and below when this symmetric distribution is the objective. 47PI state that this is possible with any other quantile than the 50% quantile as well as the same principle applies: The loss function penalizes deviation of the desired above-below ratio, which 1:1 in the 50% (median case). Targeting e.g. the symmetric 90% intervals would require finding the 5%- and 95% quantile, so now, two parameter combinations must be found. Applying the same principle as above, the loss function would penalize according to the desired quantiles differently, yielding the parameter combinations that come closest to the objective.

This procedure requires interrupting and changing the model fitting procedure. As this would be out of scope of this work, a modified version will be implemented for the CLV context in the Applied Methods Chapter that keeps the core idea but simplifies the procedure. A step-by-step guide for the implementation will be provided in this later chapter.

As indicated before, the implementation of Quantile regression will be a modified of the “original” approach but keeps the idea direct interval estimation. To avoid introducing an optimization in the model fitting process, the optimization is 1. Conducted after the model fitting and 2. Broken down into to a grid search. More precisely,

1. Build a grid of parameter combinations
2. Fit a model with each parameter combination
3. Predict the CET for each customer, using each model
4. Introduce a distance measure what serves as a loss function for
   1. The parameters for the Upper bound parameters
   2. The parameters for the Upper bound parameters

When the central (1-a)-quantile is requested, a/2 of the observations should a be above and a/2 below. The first part measures exactly this coverage and then a/2 is deducted.

1. Collect the differences
2. Select the parameter combinations that yield the lowest absolute differences, one combination for the upper, one combination for the lower bound.

It is apparent that the true values are required in this method in order to figure out the performance of each combination. In reality, one does either not have these values (because they lay in future, and they shall be predicted) or they are known because they lay in the past. Then, there would be no point in predicting them. To overcome this issue, one could assume that customer behavior for one firm will not change a lot. In this case, there is no reason to assume that the optimal parameter combinations that yield the desired quantile would change. So, the first part could be run on old data of the company to figure out the optimal combination and then use them on the new, interesting data to construct prediction intervals.

A few notes on the first part of the procedure:

* For setting up the grid, it is helpful to have some prior knowledge where parameters should be approximately located. This was done manually, running several attempts for a rough orientation and then giving several alternatives around. It turns out that, regardless of the dataset (and in the next chapter, regardless of the learning and holdout period lengths), nearly the same parameters are selected. This is a very convenient situation for the application in practice as one does not have to search for them anymore and can run the approach with a smaller grid.
* Many customers don’t buy again after their initial purchase. Regardless, which parameter combination is selected, the model will never predict exactly 0 but something very close to 0. It makes sense to introduce a small tolerance and set those to 0 to give method a chance to perform well. This seems arbitrary but in practice, one could argue that a managerial decision for a customer will barely differ if CET = 0 or CET = 0.1 (what is the used tolerance). Also, one could argue that it is “unfair” against the other methods. Quantile regression is the only method that suffers from this issue and adding this tolerance to increases at the same time the QR-interval’' widths, so it comes at a cost. Doing this trade-off for other methods would only decrease their performance.