Conformal prediction, or Conformal Inference is a relatively young method to derive prediction intervals with attractive empirical guarantees and few assumptions about the data and the model form. (1CP) It was first introduced by 8CP in 2005 and gained a lot of attention in recent years. Conformal prediction has two main ways of implementation, Full conformal prediction (Transductive Conformal Prediction) and Split Conformal prediction (Inductive Conformal Prediction) where Full CP has been developed first and Split CP has emerged as an important special case (1CP). This comes due to the high computational costs associated with the full version but also sacrifices statistical efficiency (1CP). In this work, the focus will be exclusively put on the split version, because of the mentioned computational efficiency but also, which is much more central, because Full CP is not applicable for the pnbd-model. This reason for this shall be roughly discussed in following before going into detail with the applied split version.

Following 1CP and 9CP, full conformal prediction is implemented as follows: Assume there are 250 records of 1. predictors Xi = 1:250 and observed outcomes Yi = 1:250. From these 250 records, 1 record is taken out. Assuming to not know what the true Y for this record is, one can only state that is lives in **Y.** The approach is to take n values as possible outcomes and reunite each record with the 249 unchanged records, ending up with n sets of records. For each set, a new model is being fitted which is computationally costly. Predicting with each of these different models the value that was left out, and applying a score function to this outcome, one ends up with n score values. From here one would go on and create prediction intervals. But in the context of the pnbd-model, it is not possible to continue because this model does not consider the true outcomes, the Ys as it is exclusively focused on the purchase history of a customer. Therefore, fitting n models by supplying n different outcomes for Y would not result in different models and would not allow to form PIs.

Even though CP is applicable to both, regression and classification, the focus of this work will be put exclusively on regression. The procedure in general of split conformal prediction for regression works as follows (1CP, 9CP)

1. Split the data in training and calibration set and fit the prediction model on the calibration set
2. Define a heuristic notion of uncertainty, e.g. |y-f^(x)|
3. Define a score function (A score function can be chosen arbitrarily if it has the right orientation, i.e. lower values are better) 9CP
4. Compute the quantile as see 1CP, p.5 of the calibration scores
5. Use this quantile to form prediction intervals (in this work, add/subtract the quantile from the point prediction)

Regardless of the score function, these intervals have the validity property, defined in (23CP).

See 1CP, p.6

The only condition that must hold for this coverage guarantee is exchangeability in a sense that records, from training, validation and test (what is being predicted) are exchangeable which is weaker than i.i.d. data because exchangeability can be expressed with the following formula.

(Y1 Yn+1) d = (Y (1) Y (n+1)) for all permutations, directly taken from 9CP, p.3

The concrete implementation of CP for the pnbd case will be done in the next chapter.

As indicated earlier, only Split Conformal prediction will be implemented. This implementation follows in principle the steps from the general description but the are several modifications to be made.

1. Heteroskedasticity of the outcomes: It appears that customers have a very different repurchasing behavior and might buy again 0 or 50 times. When the model is off by 3, say for the first customer, it predicts 3 and for the second customer it predicts 53, the absolute delta would be equal, but the model would have done a bad job for the 1st customer and good job for the 2nd customer. Assuming that model is equally good at all levels, it is reasonable to employ adaptive confidence intervals. Otherwise, the method would suffer from an over coverage for and under coverage for large values. A solution to that issue is given in (1CP, 9CP), as they suggest scaling the residuals by their standard deviation, “studentization”. The approximate standard deviation for customers of one company can be approximated as a linear function of their CET level, retrieving sd(CET\_i). This process can be summarized as.
   1. Fit pnbd-model on the data
   2. Make prediction
   3. Get the absolute differences for each prediction to their true value
   4. Fit a linear model sd(CET) ~ CET

For every CET, there is now a reasonable scale.

1. As before, it was before the problem with Quantile regression, Conformal prediction need the true data as well to work and again, one could make the assumption that customers behavior for one firm is approximately constant over time. The whole process of model fitting and derivation of the quantile could therefore be conducted on old, known data and then the quantile forwarded to the current period.

Assuming that the standard deviations for scaling have been derived already, the whole process can be summarized:

For the old data set

1. Split the old data set into a train and a test set (split customer wise)
2. Train the pnbd-model on the training data set
3. Make point predictions on this train and test data set
4. Take the deviations from the training set and scale them (divide by their estimated standard deviation)
5. Take the desired x%-quantile of the scaled residuals
6. Rescale the quantile with the CETs of the test data set
7. Add and subtract these individually scaled residuals to the point predictions on the test set
8. Evaluate the average coverage on the test set

Theoretically, it is not necessary to split the old data set but it is reasonable to check if the quantile works at least with old data, without having it transferred to the new data.

For the new data set

1. Train the pnbd-model on the new data set
2. Make point predictions on this data set
3. Take the quantile that was previously derived
4. Rescale the quantile with the predicted standard deviation for each customer i (quantile \* sd(CET\_i))
5. Subtract and add the individually scaled quantile for each point prediction

Two additional remarks on the procedure:

1. If the old data set is not very large and there is a potential bias when splitting the old dataset, one may want to repeat this procedure several times and average over the retrieved quantiles. This is done in this work as well.
2. Typically, one would use a separate validation set in Conformal prediction to get realistic quantiles from new data which the model has not seen yet. Since the model is fitted on the new data anyway, the use of a calibration set becomes redundant.