Inhaltsverzeichnis

[Introduction 2](#_Toc168855419)

[Literature review 2](#_Toc168855420)

[Methodological literature overview 2](#_Toc168855421)

[Role of uncertainty in marketing 5](#_Toc168855422)

[Methods to derive prediction intervals 5](#_Toc168855423)

[Conformal prediction 5](#_Toc168855424)

[Ensemble method 6](#_Toc168855425)

[Bayesian method 6](#_Toc168855426)

[Quantile regression 7](#_Toc168855427)

[Applied methods 7](#_Toc168855428)

[Implementation of PI methods 7](#_Toc168855429)

[Conformal prediction 7](#_Toc168855430)

[Ensemble (former Method 1) 9](#_Toc168855431)

[Bayesian method 10](#_Toc168855432)

[Quantile regression 10](#_Toc168855433)

[Conceptual comparison of the methods 11](#_Toc168855434)

[Measures to assess reliability and sharpness 11](#_Toc168855435)

[(True) coverage / PICP 11](#_Toc168855436)

[ACE 11](#_Toc168855437)

[Upper coverage 11](#_Toc168855438)

[Lower coverage 12](#_Toc168855439)

[MIS 12](#_Toc168855440)

[MSIW (Mean Scaled Interval Width) 12](#_Toc168855441)

[MSIWW (Mean Scaled Interval Width Weighted) 12](#_Toc168855442)

[SWR (Sharpness Width Ratio) 12](#_Toc168855443)

[Bias 13](#_Toc168855444)

[Results 13](#_Toc168855445)

[Managerial / research discussion 15](#_Toc168855446)

[Conclusion 15](#_Toc168855447)

[References 15](#_Toc168855448)

[Appendix 16](#_Toc168855449)

# Introduction

…

# Literature review

## Methodological literature overview

The underlying CLVTools package make use of two models to predict customer lifetime value, the pnbd model and the gamma-gamma model. As models in general are incomplete representations of the reality, they are never achieving a correct result and hence there is some amount of uncertainty in the forecasts. For users of models, it is hence vital to understand and quantify this uncertainty. To achieve this for the underlying model and its context, first must be determined where the uncertainty in the prediction of customer lifetime value comes from. To achieve this, it follows a non-exhaustive literature overview about the sources of uncertainty. It is split into two parts where the first covers uncertainty sources related to “outside of the model”, i.e. customer behavior and the second will focus on uncertainty coming from the modeling part.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 14PI | 20 | 24 | 16 | 9 | 12 | 23 | 27 | 1 | 28 | 29 | 30 | 31 | 32 |
| Customer behavior / model extern |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Campaigns of competitors |  | x |  |  |  |  |  |  |  |  |  |  |  |  |
| Marketing contacts |  |  |  | x |  |  |  |  |  |  |  |  |  |  |
| State of the economy |  |  |  | x |  |  |  |  |  |  |  |  |  |  |
| Customer retention or churn (not observable) |  | x | x |  |  |  |  |  |  |  |  | x[[1]](#footnote-1) | x1 | X1 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Model (related) errors (named as such) / model intern | x |  |  |  | x | x |  |  |  |  |  |  |  |  |
| Parameter estimation errors | x |  |  |  | x |  | x | x |  | x |  |  |  |  |
| Wrong form of the point  forecasting model | x |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Random variations in data  generating process | x |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Data uncertainty |  |  |  |  | x | x |  |  |  |  | x |  |  |  |
| Epistemic and aleatory uncertainty |  |  |  |  |  | x |  |  |  |  |  |  |  |  |

In the following, the most important points are discussed in more detail, starting with customer behavior. As the whole CLV calculation attempts to model customer behavior, it is self-explaining that uncertainty is based here. Influences found in the literature that increase the uncertainty of customer behavior and hence the CLV do so because they are not considered in the model. Examples are campaigns of competitors, marketing contacts in the past, presence and future and state of the economy in a sense that people change their consumption behavior between recession and boom times. What is also a dominant issue that creates uncertainty is the possibility of a customer leaving the company forever either to switch to a competitor or stop consuming. The problem here is that the probability of being “alive” is included in the model but still, most customers won’t notify the company when they churn, so it stays a mere probability. Considering a customer still alive who churned quickly at the very beginning may obviously lead to overestimation of the CLV and vice versa, hence uncertainty. The second part considers uncertainty that comes from inside the model. It is to note that the papers quoted here are not necessarily concerned with CLV estimation but treat forecasting models in general or in other contexts, often time series or wind forecasting. Nevertheless, since the model concerned here suffers from similar issues as other forecasting models, these aspects are relevant here as well. Especially often addressed is the fact that a model based on parameters needs its parameters to be estimated first which is connected to some amount of uncertainty. This issue will be addressed in this work as well in the following part. Also often mentioned is data uncertainty that might not be a problem here as the process how these data are obtained and treated before the model is employed is not too complex. Nevertheless, problems as mentioned in (29), i.e. inconsistencies in data integration from various sources (e.g. different branches), any data corruption in the process or the employment privacy policies. The latter can become an issue when mistakes are made when anonymizing the records of single customers.

With these points raised, it is evident that mere point forecasts are an insufficient statement about the likely future values as they do not provide information about uncertainty (2PI,6PI,9PI,12PI). Hence, point estimates are often accompanied or even replaced by intervals, so-called prediction intervals in the context of forecasts or confidence intervals in the context of e.g. parameter estimation. (5PI) The advantages shall be briefly discussed in the following. An interval (forecast) is offering a range of possible values of future outcomes with a specific level of confidence. (11PI) This means that the true value of the prediction will fall into this declared interval with a specific percentage value, e.g. 95%. (4PI) point out 4 main points why PIs are of such importance: 1. They assess future uncertainty and 2. hence enable the user to plan “different strategies for the range of possible outcomes”. This should mean that one can prepare a strategy in case a high value inside the interval is realized and one for a low value but in the context of CLV, this could also mean to realize that a customer has a high variability in their CLV and therefore target them. There are 2 rationales behind this approach: First, (16L) state that there is often a right tail distribution for CLV, and it hence makes sense to target a customer with high variability to realize that potential. Second, (33L) state one should focus on those as it offers the opportunity to learn and reduce uncertainty. The 3. point (4PI) makes about the importance of PIs is that they enable to assess different methods of forecasting more thoroughly and 4. PIs allow to compare forecasts made under different assumptions more carefully. Another point, made by (32PI) “forecasts cannot be expected to be perfect, and intervals emphasize this” which underlines maybe the most important characteristic of PIs, namely pointing out to the user of forecasts that they are most probably wrong and hence treat them appropriately. Thinking one step beyond PIs, a more sophisticated option are density predictions, which are comparable with PIs but they assign probabilities to each area in the interval and provide so more information about uncertainty. (6PI)

As the importance of PIs is justified now, it shall be revisited how such intervals may be obtained. In different contexts, methods are applicable that are not in other contexts. To begin with the context classical regression, (33PI) summarizes 4 methods where 3 of them might not be applicable in the specific CLV framework, Bayes methods (where “training” observations are taken into account step by step and with each step the parameters and with each step the parameters and predictions recalculated), direct interval estimation methods (which are especially made for interval estimations, like quantile regression) and conformal prediction. An applicable method from this paper might be the ensemble method what would include calculating the CLV for each customer with several methods/models and taking the predicted values as observations from which to calculate mean and variance and from this setting, assuming a normal distribution, calculate a classic prediction interval. In the environment of time-series, prediction intervals for forecast are usually computed in a different way, using the observed forecasting errors observed in the calibration period. The procedure that is inter alia suggested in (2PI) includes to first observe the errors between true value and prediction of the employed model and the “bootstrap” (randomly choose) from these observations 1 error for the first prediction. From this first prediction, do the second prediction and so on, incorporating all the previous predictions. In this way, a whole future is being constructed. Repeating this process e.g. 1000 times will create 1000 possible futures from which one can get for each time step the 95% central values which can then be treated as prediction interval. Instead of randomly draw from the distribution of errors, one can also assume that the errors are normally distributed and draw from a normal distribution. Unfortunately, in the context of the pnbd model, no such error observations from a training period are available which makes it impossible to apply this approach. Nevertheless, the CLVTools still applies a bootstrap method in terms of randomly choosing which customers to include in a model parameter calculation. Deriving a significant number of models (and getting the CLV for each customer for each of the models) allows to retrieve a prediction interval. Approaches to derive prediction intervals for this CLV model will be introduced later in this work.

To eventually come to the purpose of this work, the derived PIs shall be evaluated. The following table gives an overview about possible measures found in the literature and measures that will be deployed in this work.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2L | 15 | 16 | 20 | 31 | 35 | 34 | 37 | 38 | 39 | This work | Reliability and Sharpness | Reliability | Sharpness | Downside risk | Upside risk | Context independent/Generalizability | Simplicity |
| MIS | (x) |  |  |  | (x) |  |  |  |  | (x) | x | x |  |  |  |  |  |  |
| (True) coverage / PICP / ECP | x | x | x | x |  | x |  |  | x |  | x |  | x |  |  |  | x | x |
| Upper coverage | x |  |  |  |  |  |  |  |  |  | x |  | x |  |  | x | x | x |
| Lower coverage |  |  |  |  |  |  |  |  |  |  | x |  | x |  | x |  | x | x |
| ACE |  |  | x | x | x |  |  |  | x |  | x |  | x |  |  |  | x | x |
| MSIW |  |  |  |  |  | (x) | (x) | (x) | (x) |  | x |  |  | x |  |  |  |  |
| MSIWW |  |  |  |  |  |  |  |  |  |  | x |  |  | x |  |  |  |  |
| SWR |  |  |  |  |  |  |  |  |  |  | x | x |  |  |  |  | x | x |
| Bias | (x) |  |  |  |  |  |  |  |  |  | x |  |  |  |  |  | x | x |

This table is not exhaustive in a sense that it does not contain all papers that dealt with evaluating their PIs but contains a collection of recent studies that employed inter alia methods that are applicable for the CLV context as well. The measures listed above are also employed in this work. Additional measures like MSIS (2L, 31, 39), NMPIL (34,35), PINAW (37) or Bias (2L) were not directly translatable to the CLV context[[2]](#footnote-2), but their idea is used in a slightly changed measure, hence they are noted in brackets. All the measures used will be explained in the section below. The bias has a different position as it does not evaluate PIs but the point forecasts.

## Role of uncertainty in marketing

…

## Methods to derive prediction intervals

* Methods shall be introduced in general, explicit implementation with CLV-specific adaptations later
* Give overview over the method, not (only) about the concrete procedure
* Focus on Conformal prediction and give alternatives

### Conformal prediction

**Introduction**

* …
* Full conformal prediction and ICP (Both introduced first by (8CP, e.g. p.110 for ICP)
* Full conformal prediction less used in practice because of computation costs (…) -> Focus on ICP
* “explicit, non-asymptotic guarantees even without distributional assumptions or model assumptions” (1CP)
* “generate prediction sets for any model” (1CP)

**Split conformal prediction**

Regression

* “define the nonconformity of a new example (x, y) as the absolute value of the difference between y and the predicted value of the calculated from x and the old examples” (8CP)
* Address computational “inefficiency” with ICP (8CP)
* General procedure: (1CP) these steps can be found in the concrete implementation

1. Define heuristic notion of uncertainty (f^(y | x))
2. Define score function
3. Compute q
4. Use quantiles to make a prediction set

* Explain a bit more in words (about the procedure), use 1CP, e.g. p.8, 9CP, e.g p.4
* Real implementation for the CLV context will be done below

Classification

* …
* “main difference between applying conformal prediction on classification problems versus regression problems is the choice of score” (1CP)
* “define the nonconformity of a new example as the absolute value of its difference from the average of the old examples” (8CP)

**Full conformal prediction (for regression)** (good explanation in 1CP, p.27, 9CP, p.9)

* “sacrifices statistical efficiency because it requires splitting the data into training and calibration datasets” (1CP)
* “avoids data splitting at the cost of many more model fits” (1CP)
* … from introduction for full conformal prediction in (1CP)
* Procedure: (1CP, 9CP)

1. The dataset has 250 values, take 249 and 1 value separately
2. For the one separate value, take x values that could be the outcome of the prediction
3. For each of the outcome values, fit a model, taking the 249 values and the first of the x possible predictions
4. Go through all x
5. Calculate for each run the residuals of the 249 training values and the 1 “test” value separately
6. When all x are finished for the 1 value: Take the 90% residual quantile of the 249 values, that might be 1.298
7. Find the smallest value of the x residuals that is below 1.298
8. That should be the desired quantile

**Time series**

* CLV calculation based on time series of purchase behaviours
* Model does not predict the behaviour in the next periods but an aggregated summary
* Most time series uncertainty quantification tries to assess the next periods
* Still necessary to look at how conformal prediction is used in time series PI estimation
* Overview about the most recent and influential papers (not exhaustive)

*See* ***literature table for CP applications in time series in excel****, have to change it, does not fit here*

### Ensemble method

* Hard to tell who “invented” it
* Idea: Train many models, derive outcomes from them and calculate statistics over the outcome distribution
* Either directly or assume normal distribution and calculate the interval under normal distribution
* Special form suggested in ([[3]](#footnote-3)):
  + Fit 1 model
  + Get x models by replicating x parameter combinations from the covariance matrix
  + Predict x values for each outcome
  + Take statistics (α and 1-α intervals) from this distribution

### Bayesian method

* Foundation in the 18th century by Thomas Bayes
* Hard to tell who was the first one to use it in constructing PIs
* To my knowledge: 55PI introduced the idea of using the Bayesian approach’s strengths in forming “tolerance regions”, see especially 55PI, p.9
* Also, good explanation in 51PI, p.4, 33PI, p.4
* Idea: Have a probability distribution over the parameter(s) and calculate based on these parameters a distribution of the outcome and take the desired statistic
* Process, following 51PI and 55PI
  + Have information about the parameter distribution before the parameters are observed (or take an uninformative one): *Prior distribution*
  + Have a likelihood function that states how likely it is to observe the data that are revealed step by step under the current parameter distribution
  + “Update” the prior beliefs about the parameter distribution accordingly to derive the posterior distribution, using Bayes’ rule
  + Take the quantile of the parameters to get a confidence interval for parameters (55PI)
  + Go on, predict the outcomes (posterior predictive distribution) based on the posterior parameter distribution
  + From here, statistics like quantiles can be retrieved
  + Parameter estimation (getting posterior distribution) and outcome predictions (posterior predictive distribution) can be retrieved approximately by applying the mcmc method

### Quantile regression

* “invented” by 54PI, p.7 (according to 47PI, p.1)
* Direct interval estimation (33PI, p.12)
* Find quantiles through an optimization problem
* Comparable to minimizing sum of squared residuals (47PI, p.3, good explanation)
* Don’t estimate mean but median by balancing the number of residuals >0 and <0
* Same applies to quantiles: Weigh residuals with respective alpha and (1-alpha) and include it in a loss function
* Run an optimization problem to find the parameters that minimize this loss function
* Should yield quantiles

# Applied methods

* Concrete implementation of methods to derive PIs
* Concrete implementation of measures to assess PI performance

## Implementation of PI methods

### Conformal prediction

* Full conformal prediction is computationally intense and does not work because the model does not take true values into account while fitting the parameters, see e.g. 9CP, p.10 red marked
* -> Only split conformal prediction

Split conformal prediction

* Approach follows the one from the literature review but with adaptations
* 2 major problems in advance:
* **1. True CET/PTS is unknown** for the respective customers (CP relies on having the true values at some point)
  + Potential solution in reality: Get the quantile and standard deviations from an old cohort/panel, fit the model for the new cohort/panel and use the old quantile and standard deviations for the PIs of the new data
  + Strong assumption: New customers behave similar to the old ones
  + Not yet implemented and I don’t know if it will work
* **2. Heteroskedasticity:** “Standard deviations” (how strong true and predicted values are differing) are necessary because in CLV context, you cannot use absolute residuals:
  + customer with true total spending of $1, PTS of 21$ -> delta = $20 is bad
  + customer with true total spending of $1000, PTS of 1021$ -> delta = $20 is good
  + high values would suffer under-coverage, low values over-coverage
  + Solution (1CP, 9CP): Take sd for each customer (or any useful measure of uncertainty) and scale residuals with it
  + Procedure for getting sds
    - For (x in 1:80) do (x chosen arbitrarily, 80 is ok)
      * Split the data[[4]](#footnote-4) randomly (train and test)
      * Train the pnbd model incl. spending
      * Predict for test customers the CET and PTS
      * Take the differences to the true values

end

* + - average over the difference for each customer
    - average over the CET and PTS predictions for each customer
    - it is useful to make sds dependent on the predicted values and thus make it applicable to any customer
      * -> sd\_CET(customer\_i) = f(CET\_i)
    - f is not generally defined and can be chosen more or less arbitrarily (1CP, p.9)
    - linear coherence of CET\_i seems reasonable
      * -> fit linear model: avg(sd\_i) ~ avg(pred\_i)
* Another issue: Bias: When doing split conformal prediction then the choice of customers for training and calibration changes the result a lot, thus do it many times and average over the results, again 80 seems reasonable

**Complete Procedure:**

1. Get sd(CET\_i)

For (x in 1:80) do

* Random data split (training and validation)
* Train the model with training
* Predict CET and PTS for validation
* Collect the differences between true value and CET and the predictions

End

Calculate the mean of the collected sds, PTS and CET predictions for each customer

Fit linear model: Regress mean(sd\_i) on mean(CET\_i)

1. Get quantile q (and PIs)

* q = ceiling(((n + 1) \* (1 - alpha)))/n (formula often seen, e.g. 1CP, p.5)

For (x in 1:80) do

* Random data split (training, validation, test)
* Transfer model parameters from the trained model to the calibration and test models (to imitate the normal machine learning procedure[[5]](#footnote-5))
* Predict CET (PTS) on the calibrate set
* Collect the absolute residuals (1.)
* Scale them (residual/sd\_i) by the respective sd\_i (2.)
* Take the q-quantile of the scaled residuals (3.)
* Make point predictions on the test data set
* Get test data CET intervals by: CET point prediction\_i +- (quantile\_x \* sd\_i) (4.)
* Collect quantile\_x

End

1. Average for each customer their retrieved PI values (from when they were in the test data)
2. Managerial version:
   1. Average over all collected quantile\_x
   2. Fit a pnbd model on the new data and retrieve PIs:

PI\_i = CET point prediction\_i +- (quantile \* sd\_i)

#### Validity

The measure validity assesses how exact the prediction intervals perform. They must cover on average the true value in (1-α)\*100% of the cases to be called valid. The current approach delivers in this regard (and with the simulated test data) satisfying results when repeated often enough (what is typically not the case when only as many runs are conducted as are necessary to sample each customer at least once as described in the approach above).

#### Exchangeability

### Ensemble (former Method 1)

**Procedure** (based on [[6]](#footnote-6), online book, section 16.5: https://www.openforecast.org/adam/adamRefitted.html)

1. Fit pnbd and gg model and receive parameter estimates and respective covariance matrices
2. From the estimates and covariance matrices, simulate n draws of the parameters
3. Run the prediction with each draws
4. Receive n values for CLV prediction
5. Take the central 90% interval of the predictions

### Bayesian method

1. Estimate the pnbd model with the Bayesian approach:
   * Use the mcmc method to get the parameter posterior distribution
   * Use the mcmc method to get the draws of the posterior predictive distribution (outcomes, only CET because gamma model for spending not implemented)
2. Until here everything done by BTYDplus
3. Take the desired quantiles of the posterior predictive CET distribution for each customer (Predicted Total Spending PTS is not implemented)

### Quantile regression

* Actually: Use an optimization problem to find the parameters that minimize loss function and yield the number of predictions above/below the desired quantile
* Issue: Complex problem, solving algorithms fail
* Alternative approach:
  + Do not: Implement the loss function optimization during model estimation
  + Do: Use grid of possible parameter combinations and calculate the necessary measures (distances to the desired quantile) for each parameter combination

**Procedure:**

1. Build a grid of parameters (use prior knowledge of where you expect the parameters to be)
2. For each parameter combination
   1. Predict the CET and PTS with this parameter combination
   2. Calculate for CET (PTS) 2 “distance measures”, see 54PI, p.7 (alternatively, one can take differences to the desired quantile directly and achieve similar performance)
      1. CET Upper:
      2. CET Lower[[7]](#footnote-7):
3. Select for each desired quantile (CET/PTS Lower/Upper) the model (parameter combination) that reaches the distance closest to 0
4. Predict with the 4 selected models for all customers their intervals

* Managerial implementation: “Recycle” old models (their parameters) and apply them to the new data to get PIs (not implemented yet, I don’t know if it works)

### Conceptual comparison of the methods

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Focused uncertainty** | **Assumptions** | **True values needed** | **Computational  effort** | **Approach complexity** | **Frequentist approach** |
| **Bootstrap** | | Data uncertainty/Bias | None | No | High | Low | Yes |
| **Ensemble** | | Model and parameter uncertainty | Normal distribution of re-sampled parameters | No | High | Medium | Yes |
| **Bayesian** | | Model and parameter uncertainty | Non assumptions but set priors for parameters | No | Medium | High | No? |
| **Quantile regression** | | (Distributional uncertainty of the response variable), model misspecification | None | Yes | Depends on prior knowledge on parameters | Medium | Yes |
| **Conformal prediction** | **Academic** | Model uncertainty | Exchangeability | Yes | High | Medium | (Yes) |
| **Managerial** | Model uncertainty | Exchangeability | (Yes) | Low | Low | (Yes) |

## Measures to assess reliability and sharpness

Besides the measure in table … above, some additional ones are created or taken from the literature and adopted to this very specific CLV context to ensure a reasonable analysis.

### (True) coverage / PICP

This measure holds the percentage of cases when the true value lays indeed in the PI:

With n being the number of customers, yi the true observation for a customer’s CLV and PIi the respective prediction interval.

### ACE

This measure indicates how much on average the intended nominal prediction coverage, PINC, e.g. 90% and the true coverage, PICP differ

### Upper coverage

This measure indicates the percentage of times the upper prediction limit was not exceeded by the true value

With ULi being the upper limit of the prediction interval.

### Lower coverage

This measure indicates the percentage of times the lower prediction limit was not exceeded by the true value

With ULi being the upper limit of the prediction interval.

### MIS

The Mean Interval Score is based on the interval score, proposed by (35) and averaged over all customers instead of timesteps as it originates from time series forecasting. Due to the scale difference between small and large values for CLV, the differences in the initial formula were scaled by the estimated value (scaling by the true value would result in division by 0 if the customer would not buy again). Also, it is not to be confounded with MSIS which would in addition be scaled by mean absolute seasonal difference. (31) This measure assesses reliability and sharpness at the same time because it penalized interval width and true values outside of the CI at the same time.

### MSIW (Mean Scaled Interval Width)

This measure assesses the interval width with respect to the level of the true value. This takes account for the special CLV situation where an interval having a width of $10 is of different value for a CLV of $20 or $200 what is a realistic case for different customers. The actual MSIW is the mean over all customers.

### MSIWW (Mean Scaled Interval Width Weighted)

Lower CLVs tend to have a higher (worse) MSIW for all methods (see appendix). At the same time, firms may want to put more emphasize on higher CLV customers. This measure will therefore weigh customers with higher CLV more, i.e. the weight for a customer will be its CLV relative to the sum of all CLVs.

### SWR (Sharpness Width Ratio)

This measure evaluates reliability per width achieved by the intervals. A higher value is better than a lower.

### Bias

The bias is not calculated for the intervals but for the point predictions. It measures if the predictions are systematically wrong in one or the other direction, compared to the actual value. Therefore, it will not be different for different methods to derive prediction intervals.

With yEsti being the estimate. The formula was taken from (2L) and was slightly adapted.

# Results

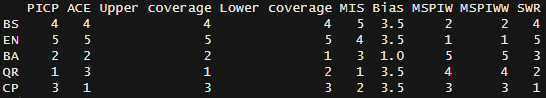
## Methods’ Performance - Benchmarking Bootstrap

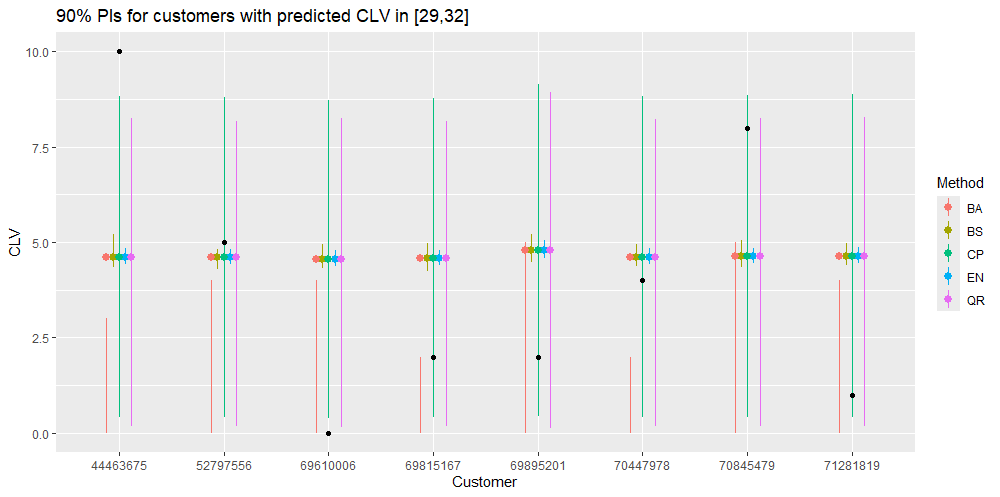
Overview about performance with real situation data (Training only with data which have been available by this point in time). Select training and testing period so that: … All the data from this paragraph come from this 1 combination

### General performance



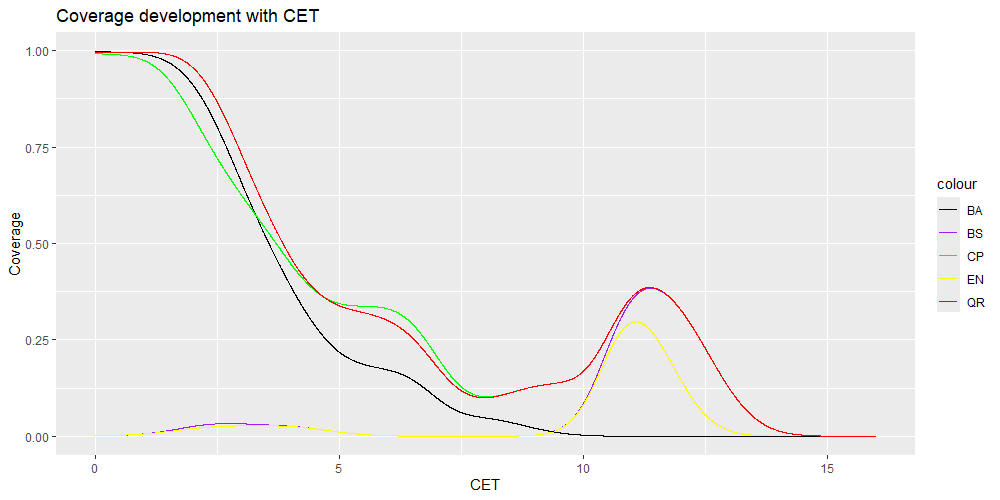
* BS an EN have very low coverage, BA, QR and CP have around 90%
* MIS: BS and EN have very high score, BA medium and QR and CP are comparably good
* BA and EN have very narrow intervals, BA, QR and CP have very wide intervals
* compared to the true value, all methods improve: Means the lower the true value, the higher the uncertainty, more certainty for higher values what is desirable
* SWR: CP delivers the most coverage per width

Balancing between low coverage and short range (BS, EN) or higher coverage for higher width (BA, QR, CP) difficult, see the ranking table

Contrast visible in the following plots

### Insight in the coverage

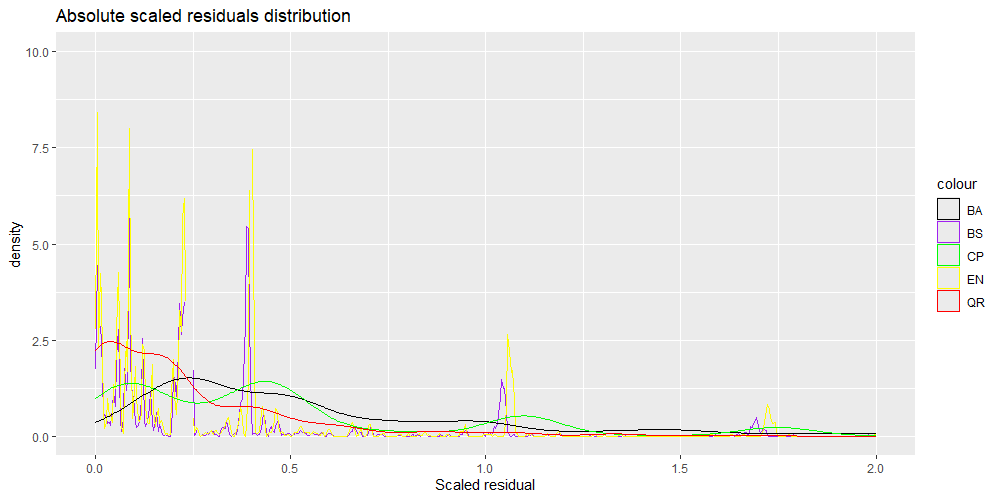
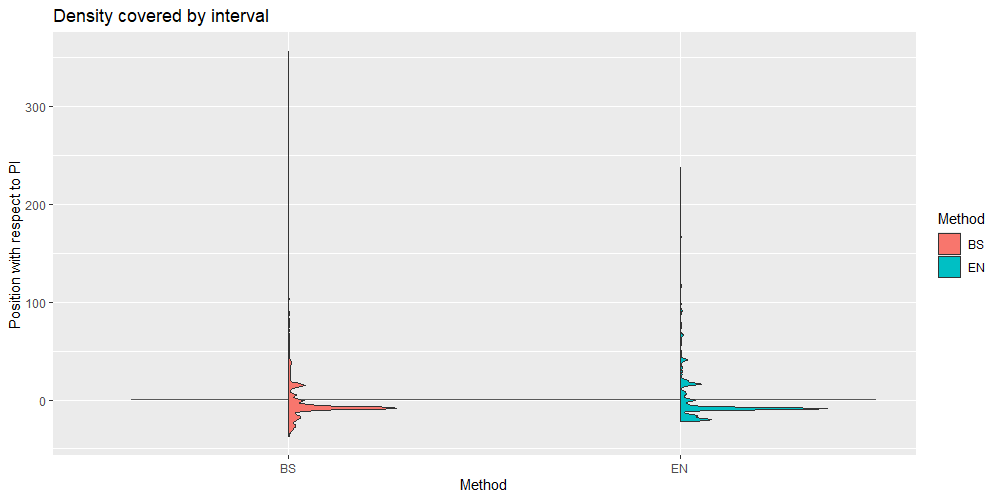
Important to note that coverage probability is dependent on the level of CET. BA, CP and QR perform well on lower levels and decrease rapidly for medium levels. CP and QR have another peak 70-80% of the maximum level. BS and EN have basically no coverage at low levels and have a peak like CP and QR at 70-80% of the maximum level.

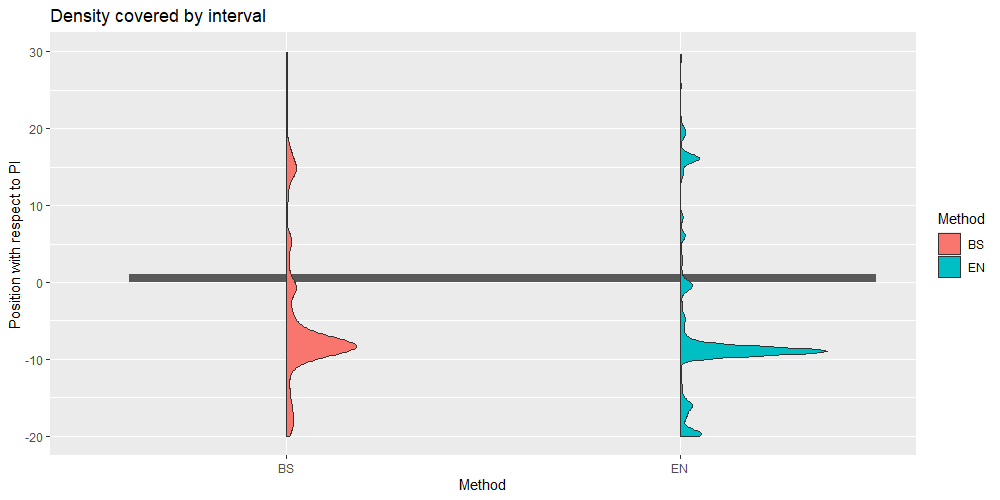


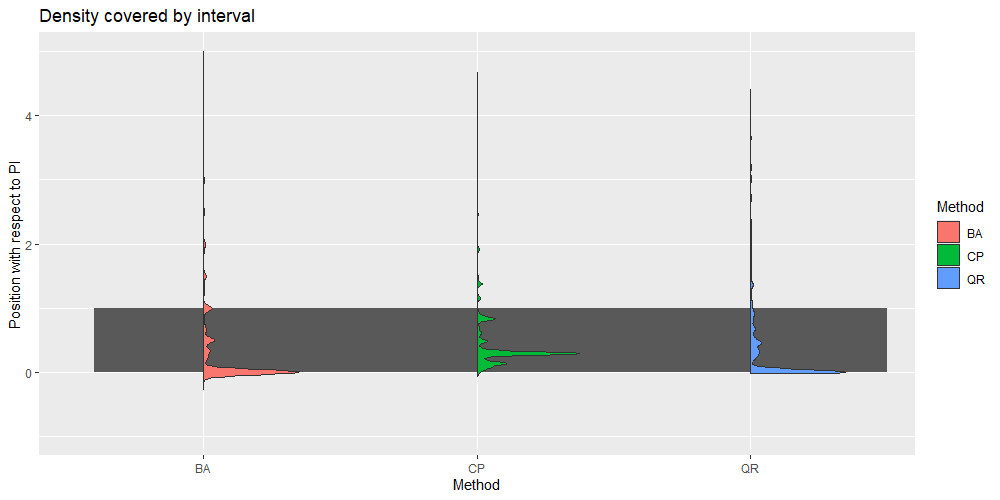
### Distribution of residuals and predictions

The following figure shows how residuals (distance of the uncovered points to their closest interval boundary) are distributed.

* very unsteady for BS and EN, with many peaks and outliers (y and x), some even needed to be cut to keep dimensions properly
* BA, CP and QR have a rather steady distribution

Also, interesting to see where the density is located in relation to the prediction interval. Different graphs for BS and EN and BA, CP and QR because of different scales due to coverage rate and interval width.



How do datasets differ?

Finding the best method (for which purpose?)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Method | Adequate coverage | Coverage across outcomes (CET, PTS) | Coverage across levels | Performance across data sets | Upper and lower coverage | Usefulness |
| Bootstrap | No |  | No |  |  |  |
| Ensemble | No |  | No |  |  |  |
| Bayesian | Yes |  | Not really |  |  |  |
| Conformal prediction | Yes |  | Acceptable |  |  |  |
| Quantile regression | Yes |  | Acceptable |  |  |  |

## Performance over varying training and prediction periods

* Comparison between methods shifted from different measures to only PICP and from one period-combination to several combinations
* Address managerial question: Performance when only limited data available especially relevant for CP and QR
* Are the methods equally good across combinations? How do they vary?
* Which combinations are good, which aren’t? (reorder table and put in appendix)

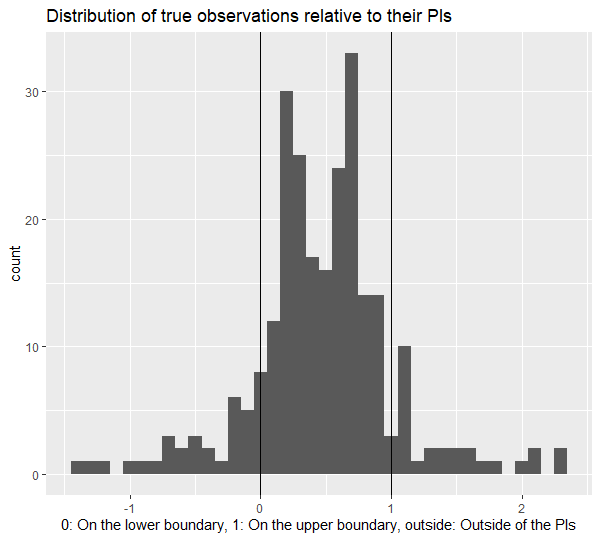
# Managerial / research discussion

…

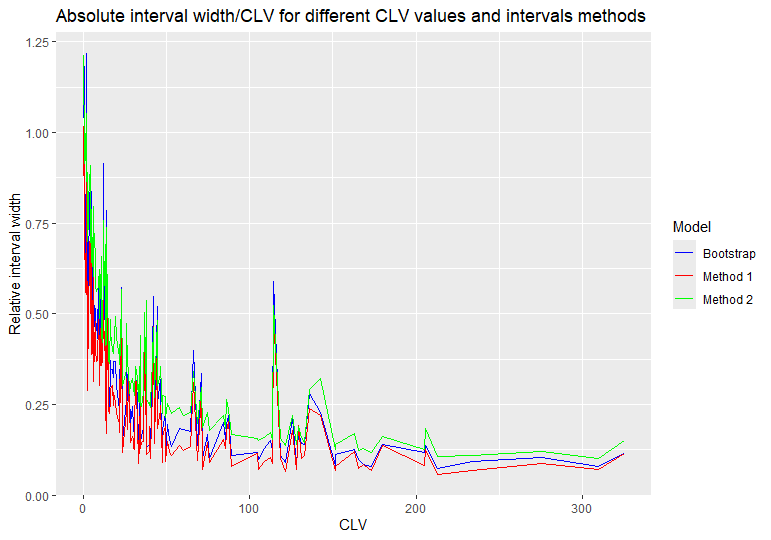
# Conclusion

…

# References

…

# Appendix



1. Even though those papers do not state explicitly that churn contributes to uncertainty in the CLV estimation, they do make the connection from churn to CLV estimation (20, p. 2) [↑](#footnote-ref-1)
2. Since the literature on benchmarking PIs in the CLV context is very limited, it was necessary to conduct research in other fields where it is more common to calculate and assess PIs. Typical examples are time series forecast competitions (31) or the prediction of wind (16, 20, 37) or electricity-related forecasts (35, 38) [↑](#footnote-ref-2)
3. Svetunkov, I. (2023). Forecasting and Analytics with the Augmented Dynamic Adaptive Model (ADAM) (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781003452652>

   (online book, section 16.5: https://www.openforecast.org/adam/adamRefitted.html) [↑](#footnote-ref-3)
4. A customer and its transactions can only be in either of the sets [↑](#footnote-ref-4)
5. Train the model on training data, use the model to make predictions on the test or calibration data. In CLVTools, I can to my knowledge not just predict a dataset with a model that has not been trained on these data, so I need to train on each data set the model first and then change it accordingly) [↑](#footnote-ref-5)
6. - Svetunkov, I., Pritularga, K., 2023. Incorporating Parameters Uncertainty in ETS. Department of Management Science Working Paper Series. 1–19.)

   - Svetunkov, I. (2023). Forecasting and Analytics with the Augmented Dynamic Adaptive Model (ADAM) (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781003452652> [↑](#footnote-ref-6)
7. Often the true value is 0 for both, CET and PTS, but predicting 0 is not achievable by any parameter combination but something like 0.0000214. To not let these cases destroy the approach, a small tolerance *CET\_tol* is added to the true value. It makes sense from a managerial point of view because one would barely treat a customer differently if the prediction is 0.0000214 or exactly 0. [↑](#footnote-ref-7)