2L

* MSIS (Mean Squared Interval Score)
* Bias (“average signed error scaled by the in-sample mean”, L1 norm and L2 norm)
* Coverage
* Upper coverage
* Spread

3L

* Predicted vs. actual frequency of repeat transactions
* Compare predicted and actual (cumulative) repeat transactions

3PI

1. RMSFE
   * Calculate RMSFE on the residuals
   * Calculate a constant band size of p-value(95% prediction interval) \* RMSFE

6

* Also provide an interval and ask how likely it is for an observation to fall in this interval (bias reduction)

11

* “Rolling horizon methodology for evaluating the performances of prediction intervals over a range of samples”
* “Automatic model selection which removes the subjectivity from the model selection procedures”
* “the bias-corrected bootstrap prediction intervals perform most desirably, providing tight intervals with accurate coverage values“

14

* “We evaluate the finite sample performance of the empirical prediction intervals using Monte Carlo experiments”
* “We use two benchmarks for our study: theoretical prediction intervals and purely non-parametric prediction intervals.”

15

* Comprehensive performance analysis of PIs
* Compare Gaussian approximations and coherent PIs (count time series PIs)
* Metrics:
  + True coverage: Probability that an observation falls into the PI
  + Set of all coverage values and compute:
    - “Shortfall rate”: Proportion of intervals not meeting the requirement of e.g. 90%
    - “Average shortfall”, “average exceedance”
    - Sample standard deviation among all CIs

16

* PI nominal confidence (PINC)
* PI coverage probability (PICP)
* Average coverage rate
* Interval score

17

* Reliability
  + PINC, PIPC
  + Average coverage (ACE)
* Sharpness
* Reliability should be prioritized

20

* Interval score
* PINC
* ACE

31

* Absolute coverage difference ACD