

Yule-Simon: Application Note

Loss Prediction

In this application note, we use the Yule-Simon process to predict the expected loss of Microsoft (MSFT) conditioned on there being a tail event with probability of 0.05 and 0.01. We use a sliding window approach to first sample the Yule-Simon posterior on a given set of log returns, and then predict the distribution for the next unobserved log return. The expected loss (also known as the cVaR) is the conditional expectation of the predictive distribution given the observation is less than 0.05 or 0.01 (both run separately). This process is continually repeated by sliding the window one sample ahead after each prediction, updating the state of the Gibbs sampler, and predicting the next cVaR for ~32 years of MSFT daily closing price log returns (3/13/1986 to 01/02/2020).

Note that this takes several hours to run, so the results have been saved to results.mat so the analysis plots can be recreated immediately (see SCRIPT.m to change this and rerun the entire simulation).

We use the following configuration

- Yule-Simon model: default parameters
- Sliding window size: 500 samples
- Predictive samples: 50000/step.

In addition, we also compute the “Plain-Vanilla” expected loss by taking the sample mean of the tails in sliding window below the 1st and 5th percentiles.

After the simulation is complete, we begin by counting the number of actual log-returns less than the predicted percentiles:

```
Percent of Data < 1% Threshold: 1.612%  
Percent of Data < 5% Threshold: 5.575%
```

Here we see the values are a little high, perhaps the Yule-Simon model parameters can be tuned to improve this.

Next, we plot the predicted losses for each log-return that exceeded the predicted 0.05 and 0.01 thresholds (respectively) using the Yule-Simon model and the Plain Vanilla model. From the plots, we can see the Yule-Simon model does a good job following the trend of the losses, while the Plain vanilla model seems to have a lag. There are several large losses which neither model catch, however it's unclear whether these losses are predictable.

