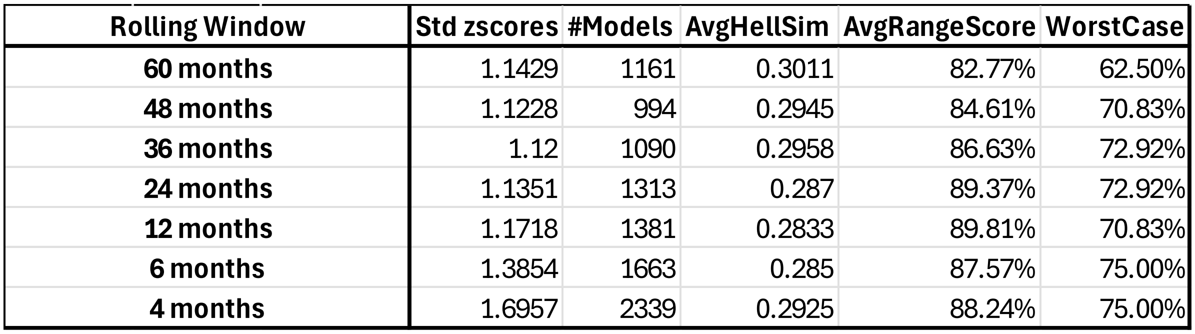
Rolling Window and Lag Value Selection

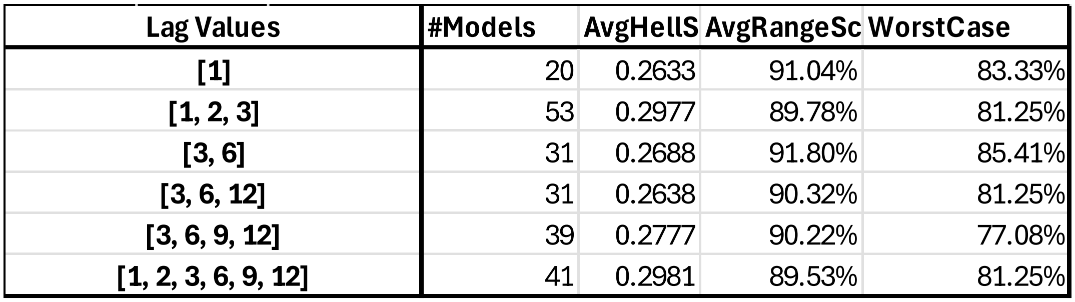
Once the optimal hyperparameters have been identified, we will proceed to determine other crucial parameters, such as the length of the rolling window to use for the estimation of z-scores. This value is important because it determines how far back we look to calculate the z-score of the returns. The metrics obtained for different rolling window lengths are presented below.

It can be observed that as the rolling window decreases, the interpolated range score increases, reaching its optimal value at 12 months. Additionally, the Hellinger similarity is optimal in this case. The values obtained using this window are not as extreme; in contrast, rolling windows of 6 and 4 months can produce values as high as 8 or 9, which could end up contaminating the data in some leaves. Therefore, a 12-month rolling window is chosen.



Rolling Window	Std zscores	#Models	AvgHellSim	AvgRangeScore	WorstCase
60 months	1.1429	1161	0.3011	82.77%	62.50%
48 months	1.1228	994	0.2945	84.61%	70.83%
36 months	1.12	1090	0.2958	86.63%	72.92%
24 months	1.1351	1313	0.287	89.37%	72.92%
12 months	1.1718	1381	0.2833	89.81%	70.83%
6 months	1.3854	1663	0.285	87.57%	75.00%
4 months	1.6957	2339	0.2925	88.24%	75.00%

It is also important to define the lag values that will be used to construct the models. A large number of lag values can result in numerous combinations, which may lead to very long execution times. Therefore, it is ideal to select a moderate number of lag values that yield very good results. The results obtained for different sets of lag values are presented below.



Lag Values	#Models	AvgHellS	AvgRangeSc	WorstCase
[1]	20	0.2633	91.04%	83.33%
[1, 2, 3]	53	0.2977	89.78%	81.25%
[3, 6]	31	0.2688	91.80%	85.41%
[3, 6, 12]	31	0.2638	90.32%	81.25%
[3, 6, 9, 12]	39	0.2777	90.22%	77.08%
[1, 2, 3, 6, 9, 12]	41	0.2981	89.53%	81.25%

Due to their good results and the size of the set, the lag values set of 3 months, 6 months, and 12 months will be selected. The lag values set of 3 months and 6 months also presents good results.

Once these parameters have been determined, we will proceed to select the models by analyzing the properties of the associated distributions.

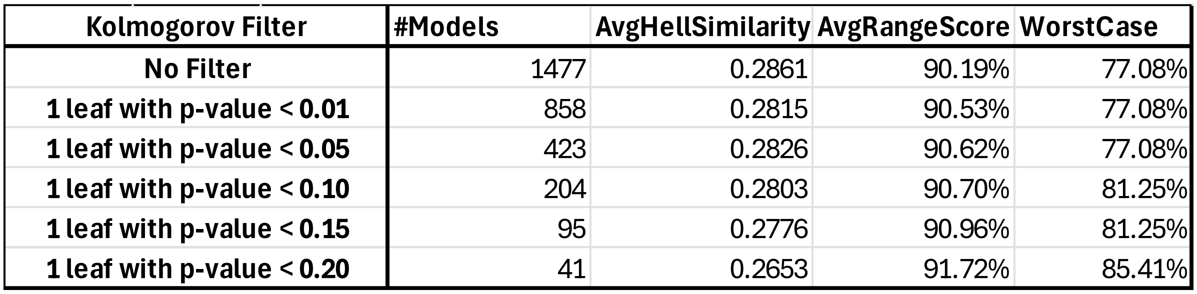
Tree Selection

Several trees will be constructed using the selected hyperparameters and various random states. Based on the samples located in each of the leaves, probability distributions will be constructed using Gaussian KDE. The Silverman parameter will be used to calculate the bandwidth of the KDE. This parameter allows the bandwidth to be determined based on the number of samples present, which is ideal for our data since it is common to find leaves with few data points included in them.

The objective is to identify decision trees whose leaves exhibit representative probability distributions. However, due to the scarcity of data, some of these distributions may be biased or unrepresentative. Therefore, it is crucial to discard decision trees that contain such biased distributions. By working with more representative distributions, a more accurate estimation of extreme values can be achieved, which in turn allows for better risk assessment.

To identify biased distributions, the Kolmogorov-Smirnov goodness-of-fit test will be used on the validation set. This statistical test determines whether a constructed theoretical distribution adequately fits a set of observed data. If the p-values obtained from the test are low, it can be concluded with some certainty that the observed data do not fit the assumed theoretical distribution. Consequently, such distributions will be discarded, and only trees containing well-fitting distributions will be selected. The performance of these selected trees will then be evaluated on the test set.

Below is a table summarizing the results obtained. In general, it is observed that as the p-value increases, the average range score also increases. Additionally, the Hellinger similarity measure is better in these cases. Lastly, it is worth noting that when the p-value is low, range scores below 81% can be obtained.



Kolmogorov Filter	#Models	AvgHellSimilarity	AvgRangeScore	WorstCase
No Filter	1477	0.2861	90.19%	77.08%
1 leaf with p-value < 0.01	858	0.2815	90.53%	77.08%
1 leaf with p-value < 0.05	423	0.2826	90.62%	77.08%
1 leaf with p-value < 0.10	204	0.2803	90.70%	81.25%
1 leaf with p-value < 0.15	95	0.2776	90.96%	81.25%
1 leaf with p-value < 0.20	41	0.2653	91.72%	85.41%

Following this approach, models will be selected using a p-value of 0.20. With this threshold, a sufficiently large number of models (41) is obtained, and the range scores for all these models are above 85%.