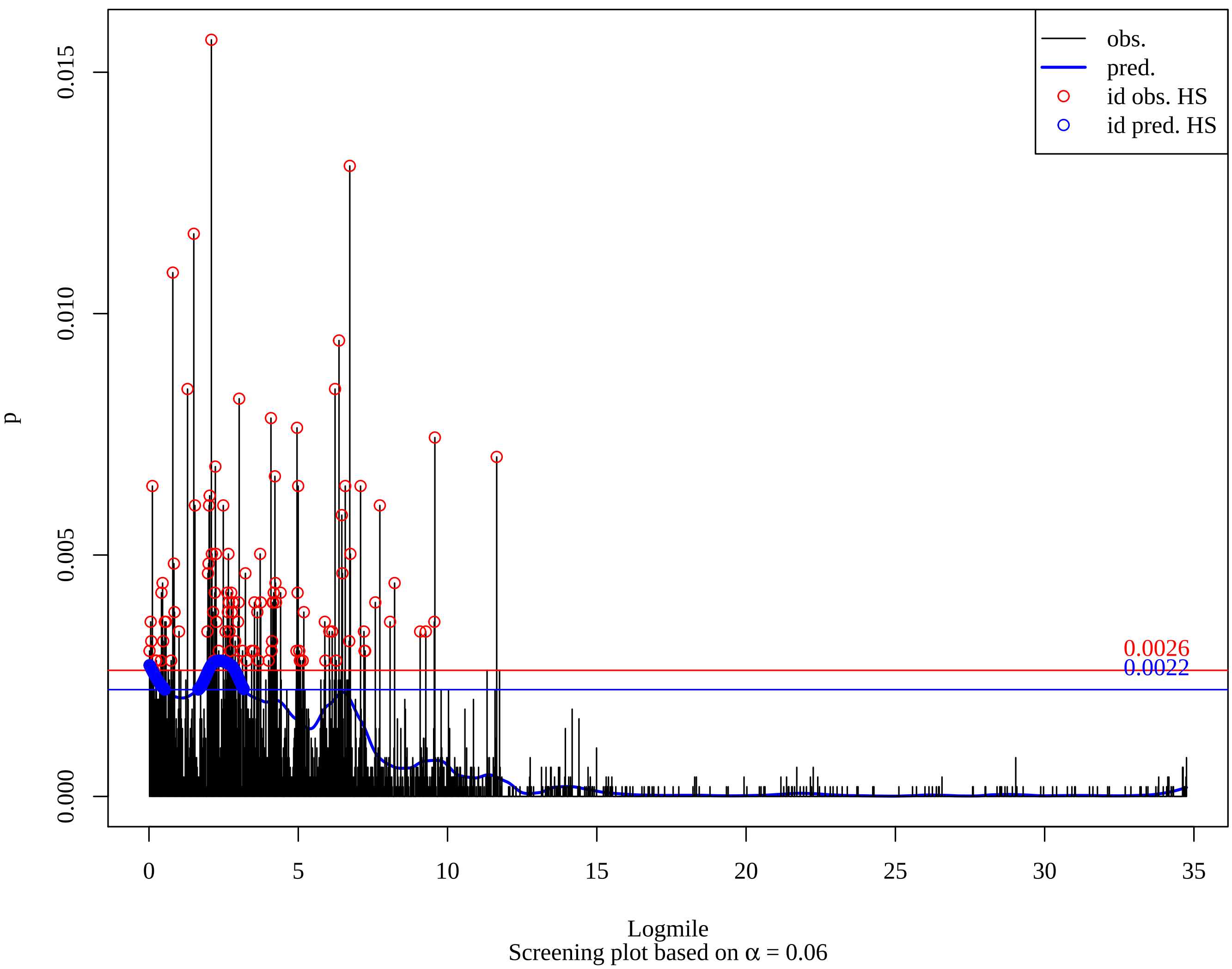
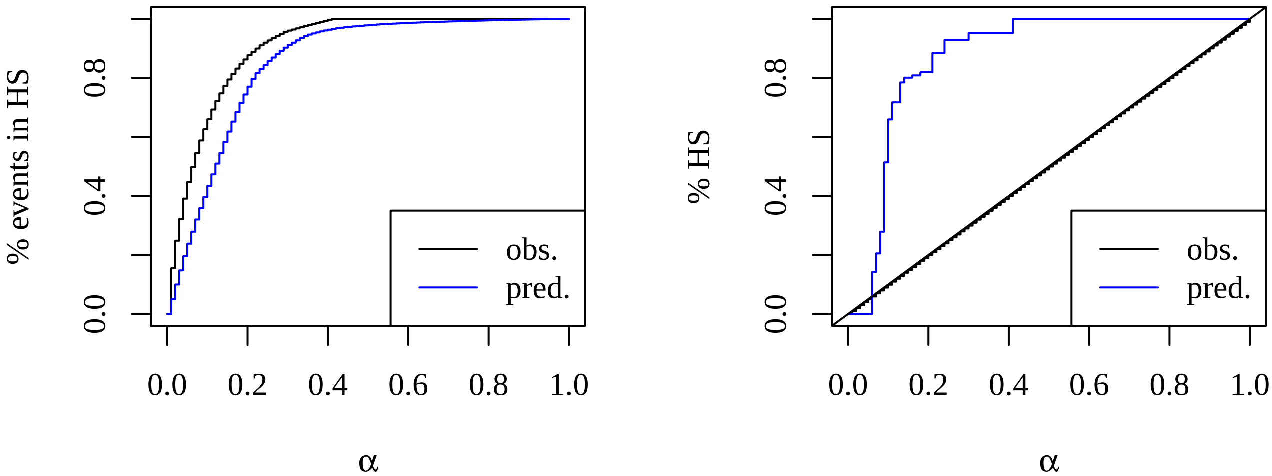
Level set setting

$\gamma_X(\alpha) = \sup_n\{X : \frac{\|C_X(n)\|}{N} \leq \alpha\}$, where n is the number of Logmile segments needed to satisfy the condition. $C_X(n) = \{X : f_X(x) \geq f_{X_{(N-n+1)}}\}$, where $f_{X_{(N-n+1)}}$ is the $(N-n+1)^{th}$ ordered crash rate (from minimal to maximal crash rate) and $f_{X_{(N-n+1)}} = \phi$, where ϕ is the threshold that satisfies $\frac{\|C_X(n)\|}{N} \approx \alpha$.

Suppose $\alpha = 0.06$, we can find out the threshold for observations is $\phi_{obs} = 0.0026$, whereas the threshold for predictions is $\phi_{pred} = 0.0022$ and the proportions of blues and reds are approximately α respectively (as plot showed below).





A sequence of α values is used to observe the pattern in different levels, $\alpha \in [0, 1]$ with width of interval 0.01. Y-axis is the cumulative capture crash rate in HS and X-axis is level of α. Suppose $\alpha = 0.1$, roughly 65% and 40% of crashes in observations and predictions respectively happened on the top 10% of Logmile segments (figure showed above left). Then, we observe the pattern under the same threshold for observations (figure showed above right). Since the proportion of observations above threshold is built based on α, it is similar to a straight line with slope of 1.