

Progress Report

Ying Jin

2023-05-12

Contents

Visualization of weights in Incident sensitivity	1
Model overfit	1
Data contamination	3
Simulation with contaminated data	3
TV-AUC	4
Spread of TV-AUC	4
Concordance	4
Simulation with noise signals	5
TV-AUC	5
Spread of TV-AUC	5
Concordance	6
Data Application	6
Summary table	6
Main text	7

Visualization of weights in Incident sensitivity

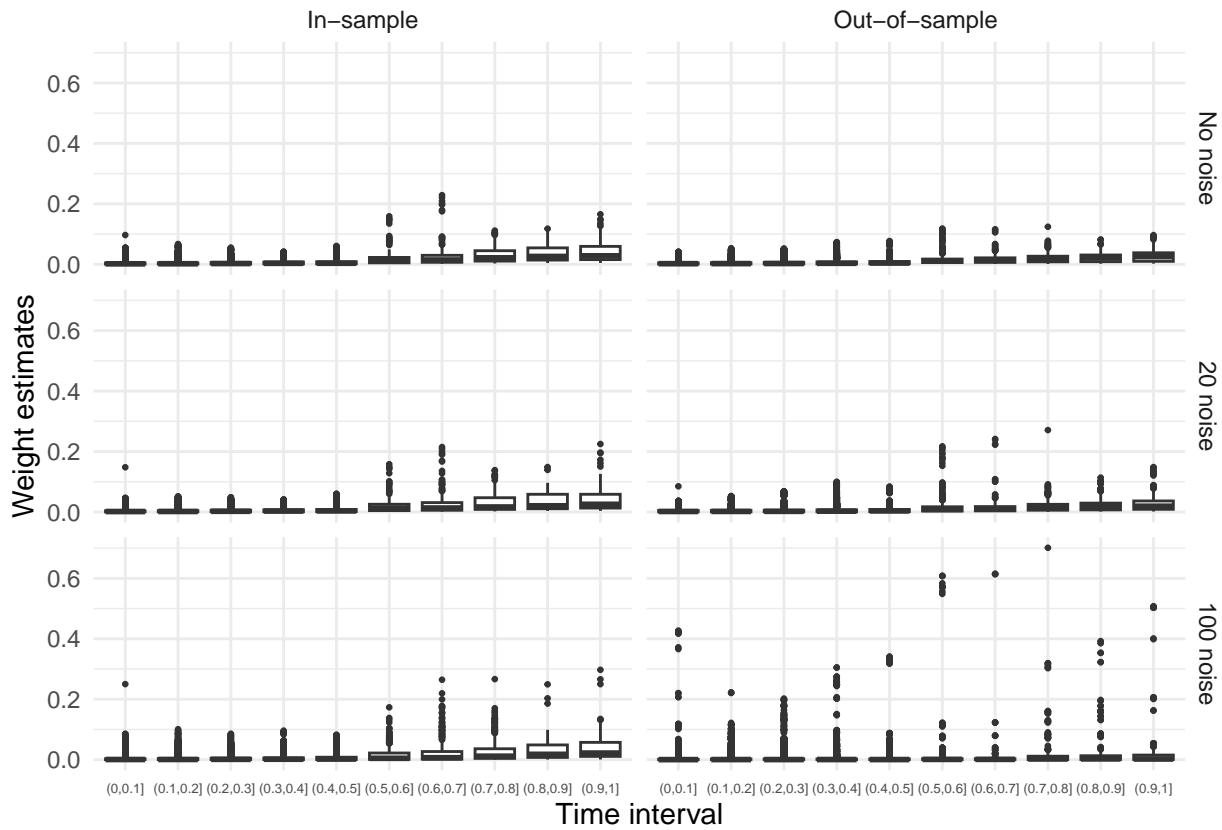
Model overfit

```
## # A tibble: 30 x 6
## # Groups:   sample, model [6]
##   tid    weight  time model    tbin      sample
##   <chr>  <dbl> <dbl> <fct>    <fct>    <chr>
## 1 X51     0.229 0.665 No noise (0.6,0.7] In-sample
## 2 X43     0.221 0.665 No noise (0.6,0.7] In-sample
## 3 X141    0.209 0.664 No noise (0.6,0.7] In-sample
## 4 X40     0.200 0.655 No noise (0.6,0.7] In-sample
```

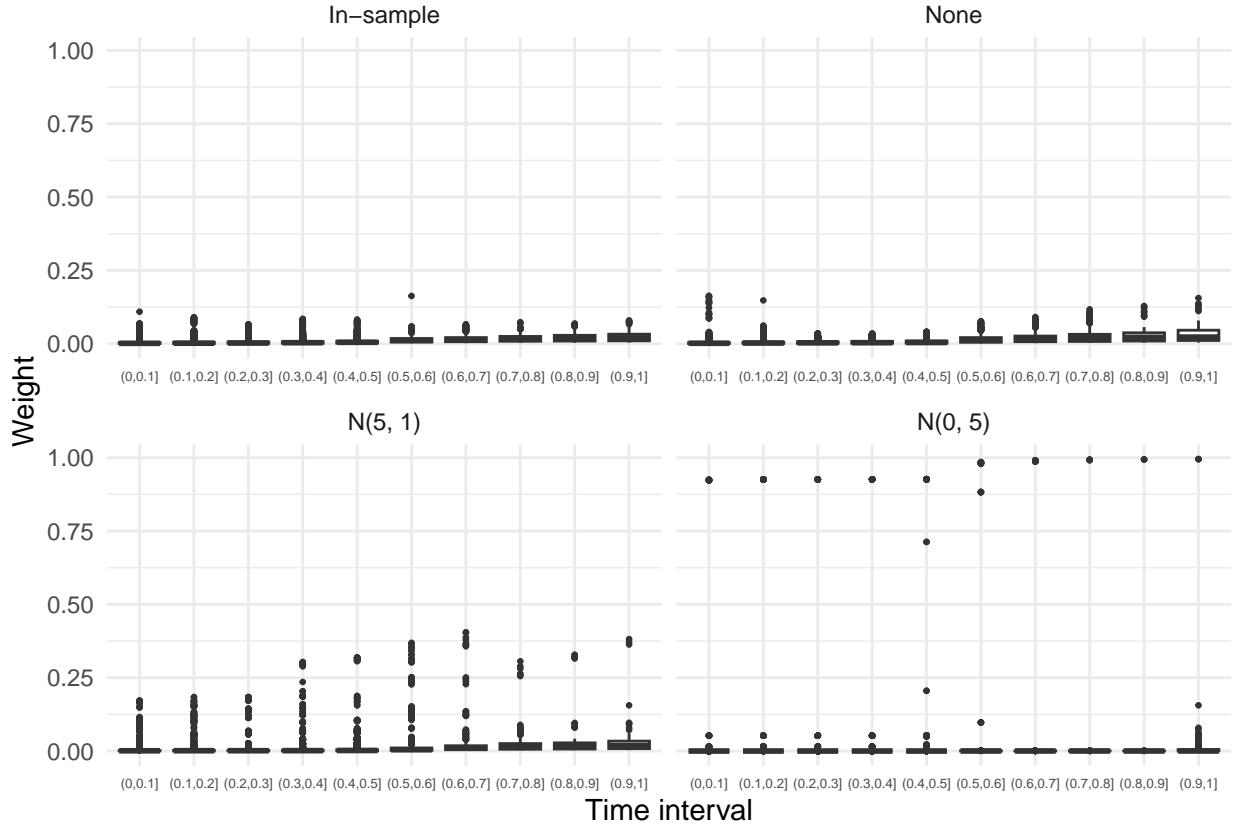
```

## 5 X19    0.197 0.646 No noise (0.6,0.7] In-sample
## 6 X117    0.225 0.977 20 noise (0.9,1]   In-sample
## 7 X51     0.215 0.665 20 noise (0.6,0.7] In-sample
## 8 X43     0.209 0.665 20 noise (0.6,0.7] In-sample
## 9 X141    0.201 0.664 20 noise (0.6,0.7] In-sample
## 10 X136   0.197 0.963 20 noise (0.9,1]   In-sample
## # i 20 more rows

```



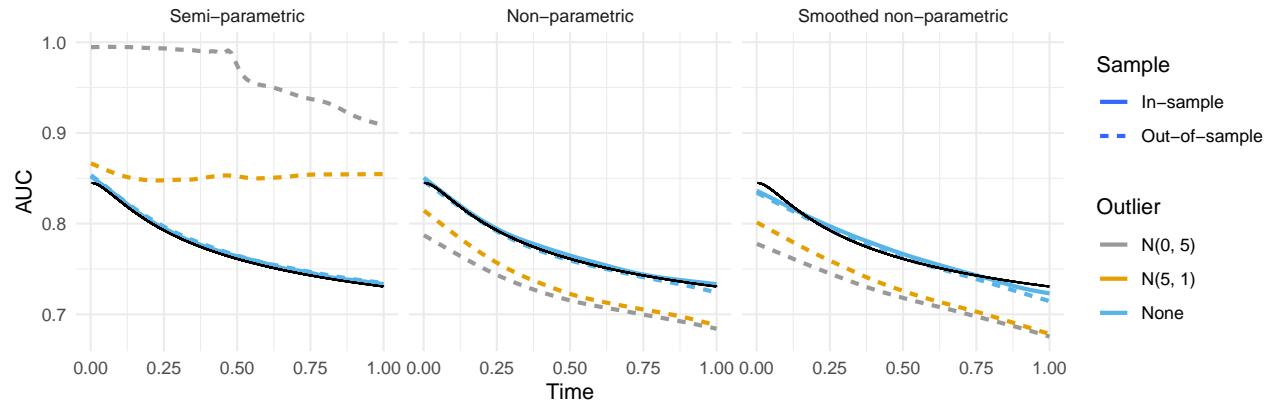
Data contamination



Simulation with contaminated data

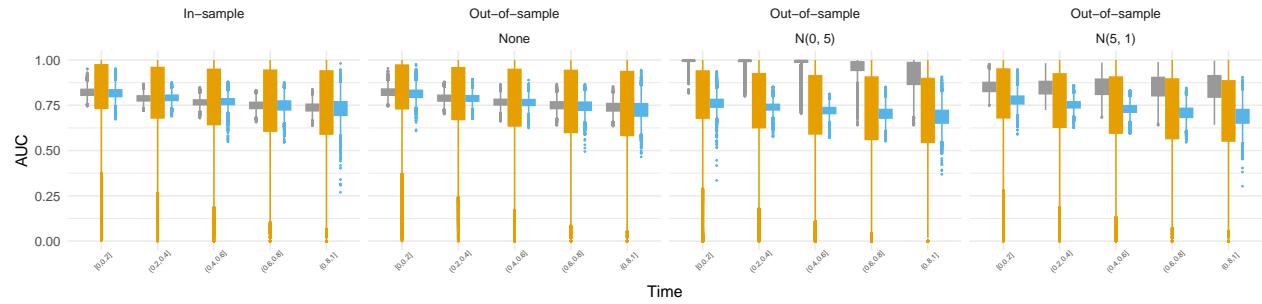
- Introduce outliers to test datasets (not noise in the model)
- Modified from last time: instead of arbitrarily change covariate value, generate from different random distributions instead
- Within each simulation, two scenarios:
 1. 10% contaminated subjects, $N(5, 1)$
 2. 10% contaminated subjects, $N(0, 5)$
- Some unexpected coding issue: couldn't find old.alpha parameter in scam function (bfgs_gcv.ubre). Needed to reduce k (k=10).

TV-AUC

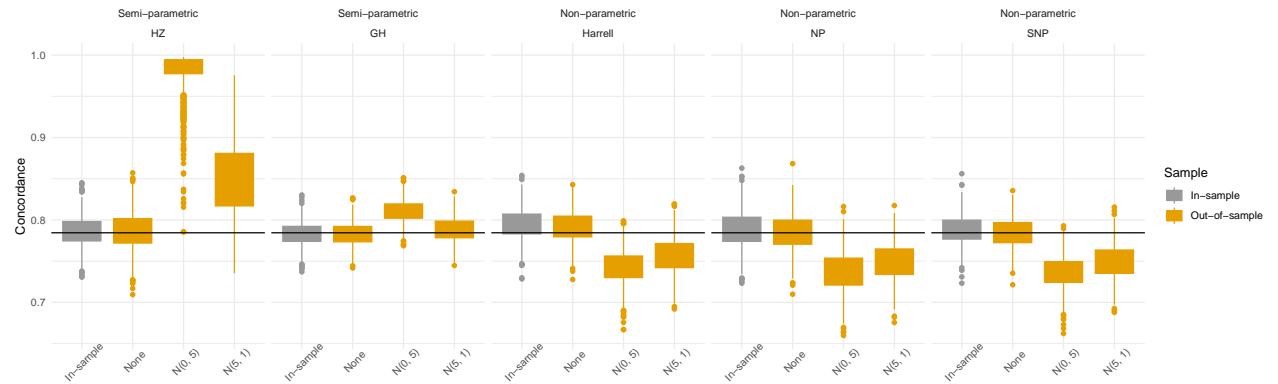


Spread of TV-AUC

Do we have alternative ways to present the instability of TV-AUC estimates? Perhaps remove the figures but only use specific examples:

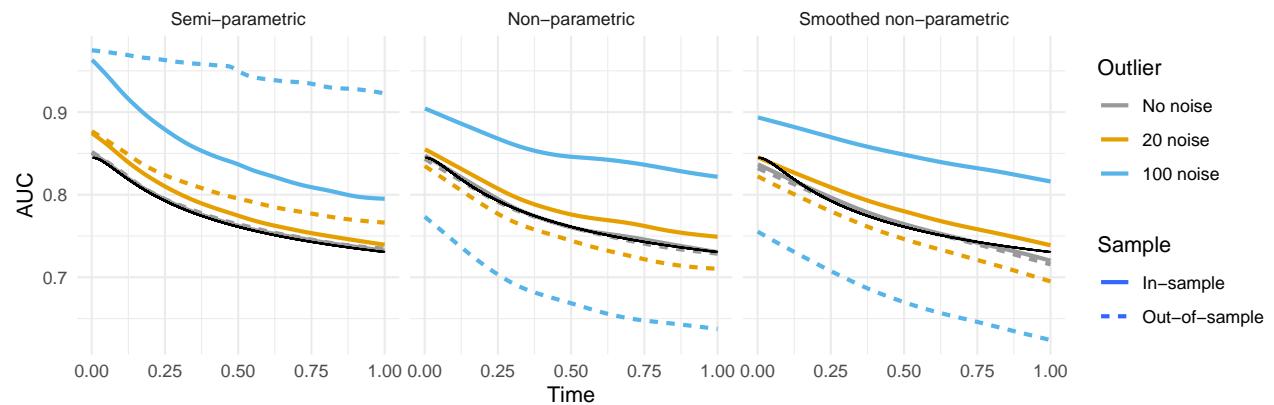


Concordance

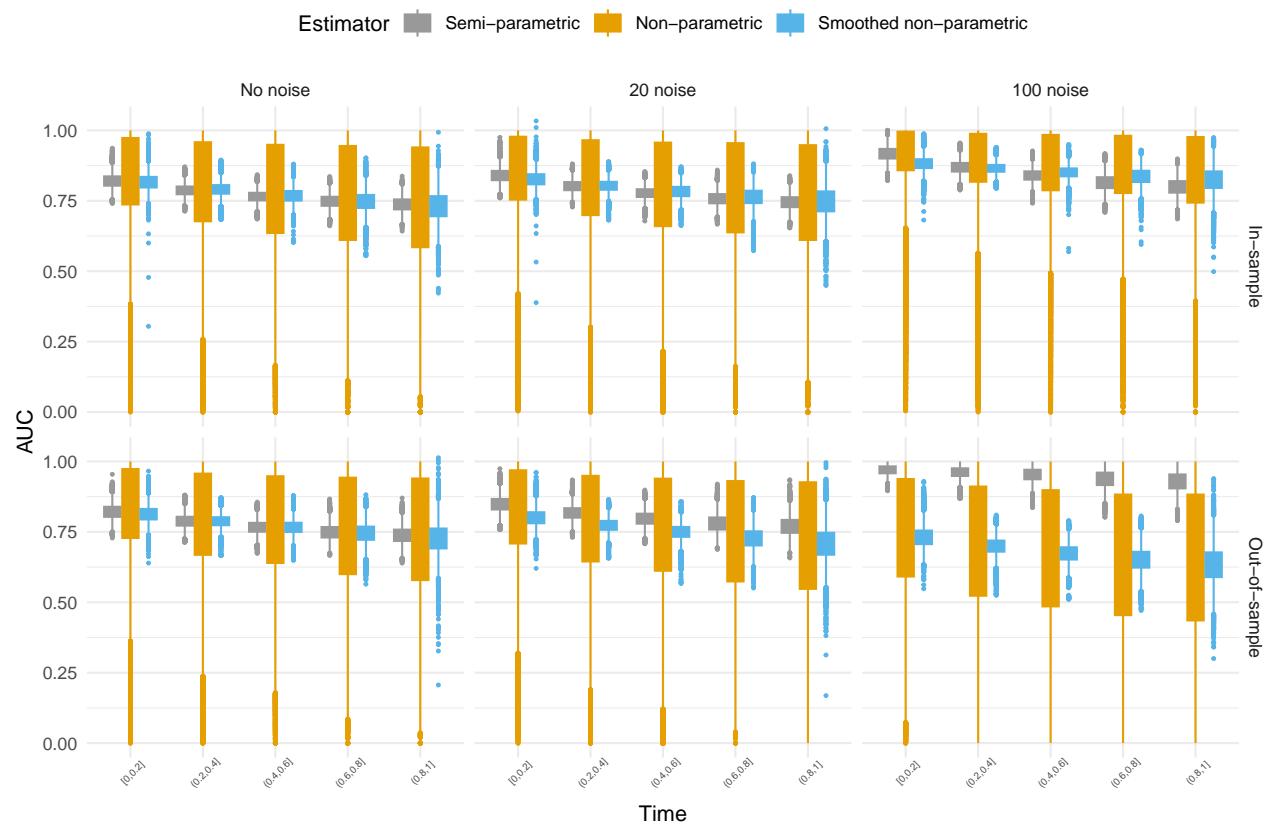


Simulation with noise signals

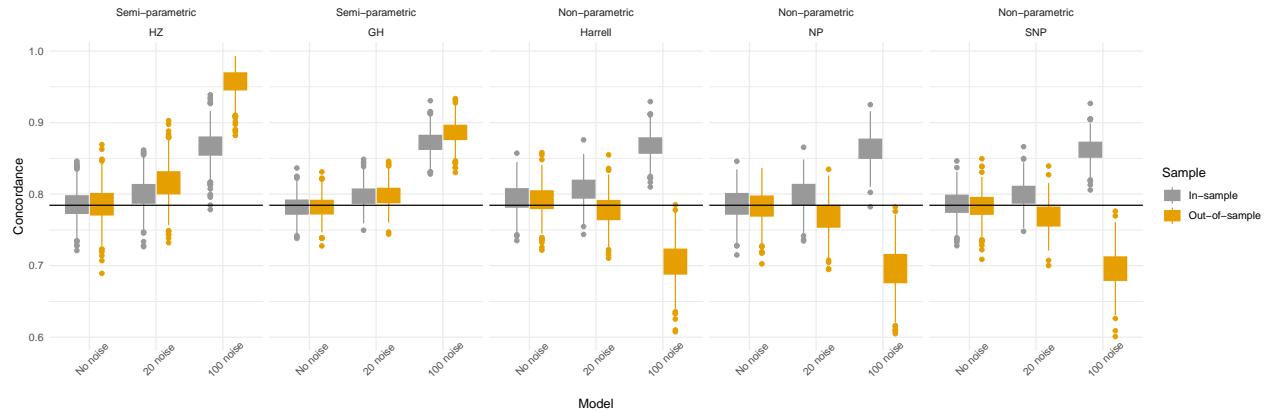
TV-AUC



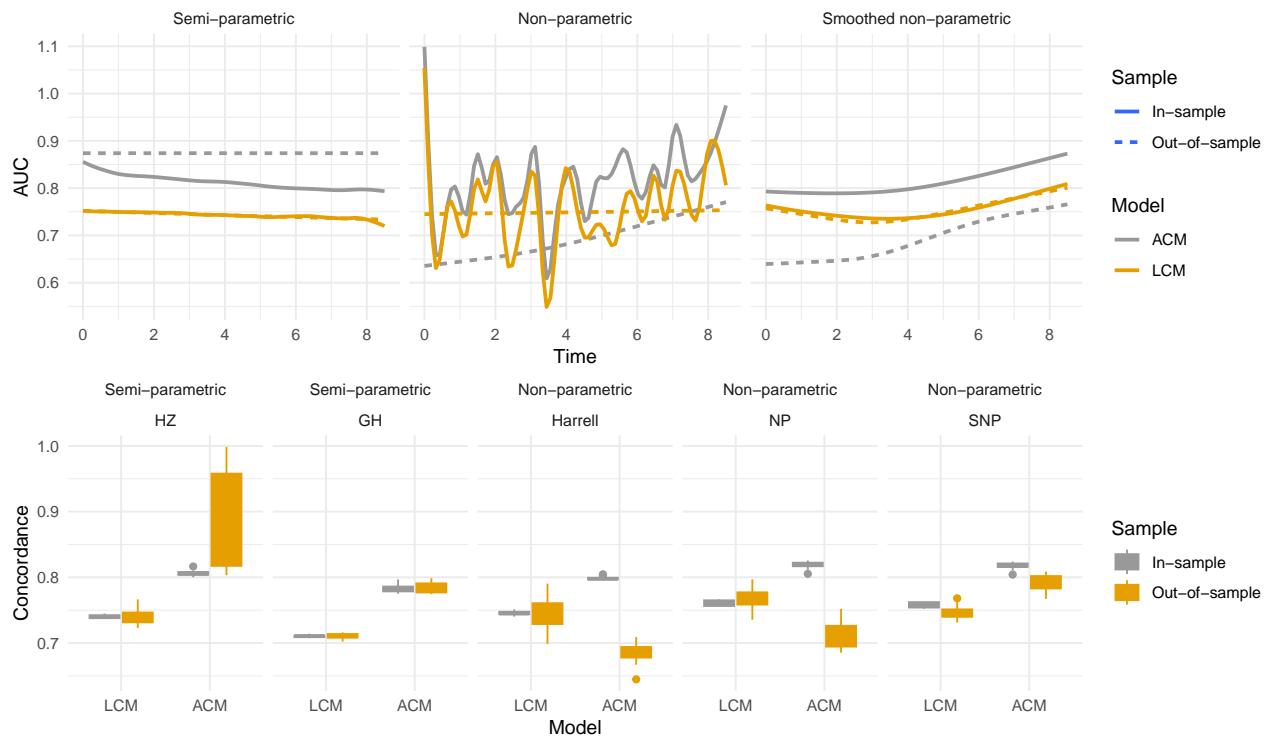
Spread of TV-AUC



Concordance



Data Application



Summary table

```
## Warning in !is.null(rmarkdown::metadata$output) && rmarkdown::metadata$output
## %in% : 'length(x) = 2 > 1' in coercion to 'logical(1)'

##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
```

Estimator	Bias	Variability	Out-of-sample behavior
Time-dependent AUC			
Semi-parametric	Unbiased	Low	Over-optimistic
Non-parametric	Unbiased	High	Appropriate
Smoothed non-parametric	Slightly biased	Low	Appropriate
Concordance (semi-parametric)			
Heagerty-Zheng	Unbiased	Low	Over-optimistic
Gonen-Heller	Unbiased	Low	Over-optimistic
Concordance (non-parametric)			
Non-parametric	Unbiased	High	Appropriate
Smoothed non-parametric	Unbiased	Low	Appropriate
Harrell	Biased upwards	Low	Appropriate

```
##  
##     group_rows
```

Main text

- Figure 1: outlier example
- Figure 2: Simulation case overfit: a.time-varying AUC; b.concordance
- Figure 3: Simulation case contamination: a. time-varying AUC; b.concordance
- Figure 4: Sensitivity weights: a. overfit; b.contamination
- Figure 5: Data application results: a. tv-auc; b. concordance