

Using Gradient Boosting to Find Optimal RDD Bandwidth: DUI Punishment and Recidivism

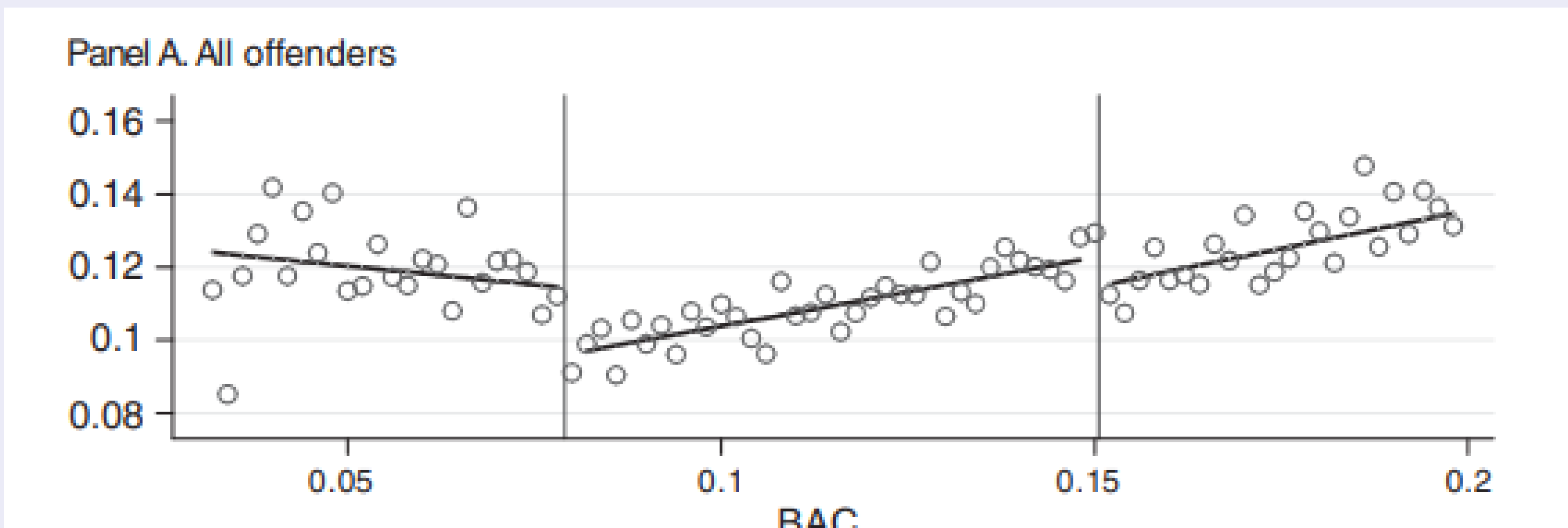
David Jones, Paul Dahm, Jonas Ater, Espen Hovland
Texas A&M University
April 29, 2024

Abstract

Using gradient boosting and cross validation to determine optimal bandwidth, we enhanced the study, "Punishment and Deterrence: Evidence From Drunk Driving" by Benjamin Hansen. The study uses the Washington state legal thresholds for DUI's (0.08 BAC) and aggravated DUI's (0.15 BAC) as cutoffs for a Regression Discontinuity Design analysis in order to examine recidivism (the tendency of a criminal to re-offend). Hansen used an arbitrary RDD bandwidth of 0.05 and 0.025 around the BAC cutoffs.

Machine Learning Method

- We used gradient boosting to model the relationship between BAC levels and recidivism within varying bandwidths around both DUI and Aggravated DUI thresholds
- We then used cross validation (through a grid search) to optimize the gradient boosting model, this systematically tested different combinations of the G-Boost parameters across multiple folds to find the combination that minimizes error.



Findings

Table: Bandwidth Comparison for effect of DUI/ADUI on Recidivism.

Independent Variable	Estimate	St. Error	Confidence Interval
DUI bandwidth (0.05)	-0.0582	0.015	(-0.088,-0.028)
DUI bandwidth (0.025)	-0.0683	0.034	(-0.135,-0.001)
DUI Optimal bandwidth (0.02722)	-0.0421	0.031	(-0.103,0.019)
ADUI bandwidth (0.05)	-0.0042	0.002	(-0.007,-0.001)
ADUI bandwidth (0.025)	-0.0055	0.002	(-0.009,-0.002)
ADUI optimal bandwidth (0.027222)	-0.0057	0.002	(-0.009,-0.002)

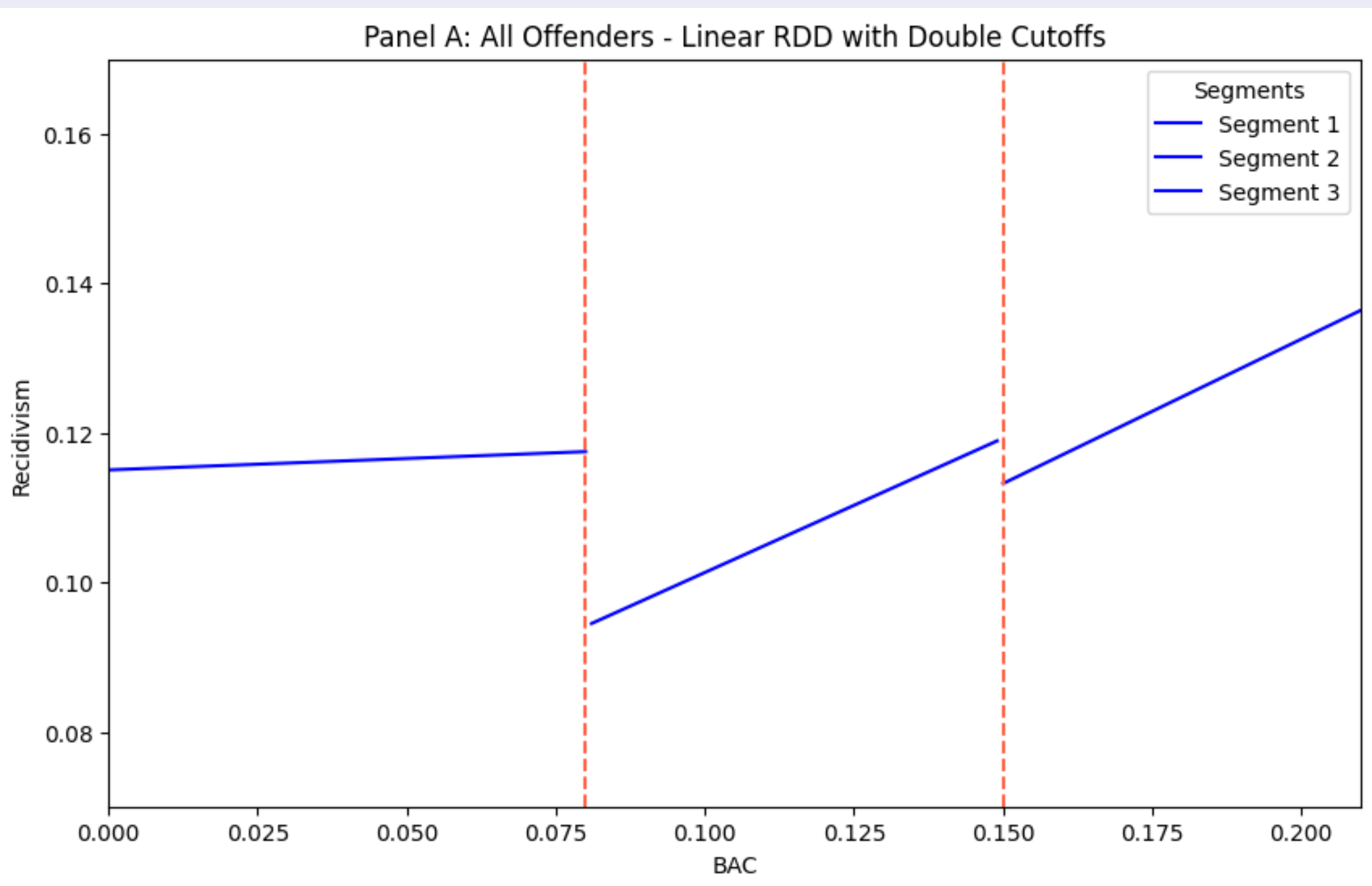


Figure: Original Cutoff

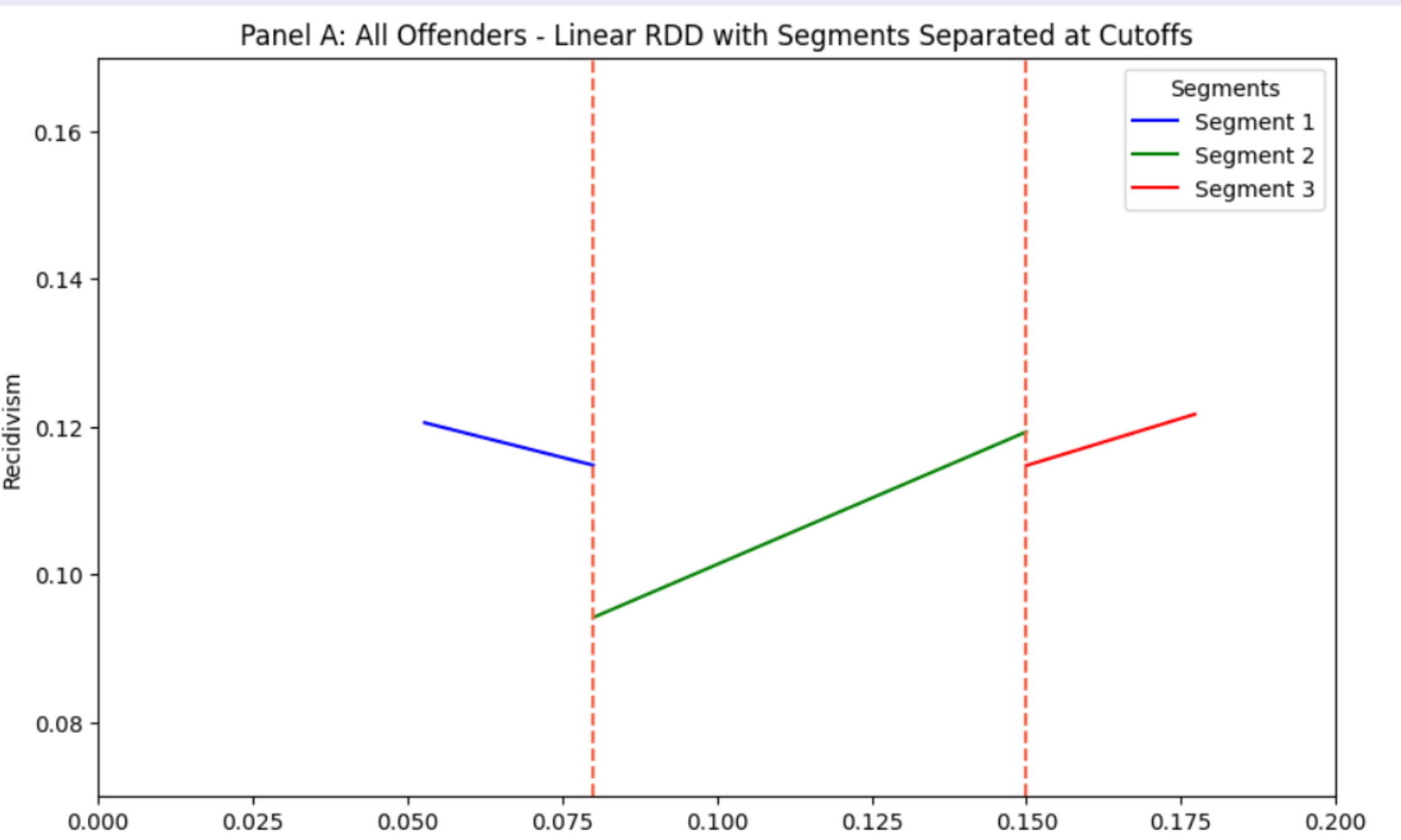


Figure: Optimal Cutoff

Literature Review

- Large-scale evaluation of k-fold cross-validation ensembles for uncertainty estimation (Dutschmann et al. 2023)
- Used cross validation on molecule datasets to help determine uncertainty with drug design.
- SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer (Prusty et al. 2022)
- Used Gradient boosting with Cross Validation to create a model that accurately predicts cervical cancer

Results

- Our results show the effect of the 0.08 threshold is stronger than the effect of the 0.15 threshold.
- After using the optimal bandwidth, only the estimate for the effect of Aggravated DUI was statistically significant.

Group References



(a) Colab



(b) Github