Cut parts

# Appendix B: Cut Parts

In this paper, I have demonstrated the value in bringing theoretical knowledge to bear on predictive models of enrollment. Community colleges construct such models under unique constraints. Moreover, the financial consequences of inaccurate models is more acutely felt than at other types of institutions of higher education. While the extant literature has sought improvements to predictive accuracy by using 'brute-force' methods - e.g. increasing the number of features in models, implementing more model types, exhaustively searching across model hyper-parameters - I argue that these efforts essentially re-invent the wheel insofar as they 'rediscover' relationships that are already well established. Our theoretical knowledge of the predictors of enrollment and persistence is robust. That knowledge should be reflected in our empirical models.

Though the differences between stacked and linear lagged models are statistically small, their substantive importance should not be dismissed. Such differences can easily mean missing predictions by several hundred students at a community college with a student body in the tens of thousands. Moreover, it should be noted that, while the linear model with a lagged term is more parsimonious than the stacked models, parsimony for its own sake is dubious in circumstances where predictive accuracy is paramount

Numerous dissertations (see: x, y, and z) have reproduced and expanded this work in varying local contexts, adding important contextual knowledge but few explicitly addressing modeling choices.