economical

MULTIVARIATE SURVIVAL ANALYSIS

A Legal Representation Model

Craig A. Sloss, PhD, FCAS, ACIA
Senior Technical Specialist – Economical Insurance

SESSION OVERVIEW

Based on joint work between Craig Sloss and Sunny Xu

Agenda for the Session:

- Background on the business problem
- Construction of multivariate survival models
- Model validation using censored data


ACCIDENT BENEFITS
COVERAGE

ONTARIO ACCIDENT BENEFITS

Overview of the Coverage

- First party no-fault auto injury insurance
- Medical and income replacement benefits with standard application forms
- Injuries classified as minor, non-minor, or catastrophic
- Regulatory limits on medical payments for each class

economical

CENSORED DATA CHALLENGES

Will Claim 2 eventually become represented?

CLAIM ID	CLAIM STATUS	DAYS OPEN	LEGAL REPRESENTATION?	WHEN DID LEGAL REP JOIN?
1	Open	200	Yes	Day 7
2	Open	10	No	NA
3	Closed	450	Yes	Day 30
4	Closed	250	No	NA

POLLING QUESTION

You are assembling lists of eligible predictors for two models by considering ASOP 12 criteria.

One model will be used for pricing, and one will be used to support claims operations.

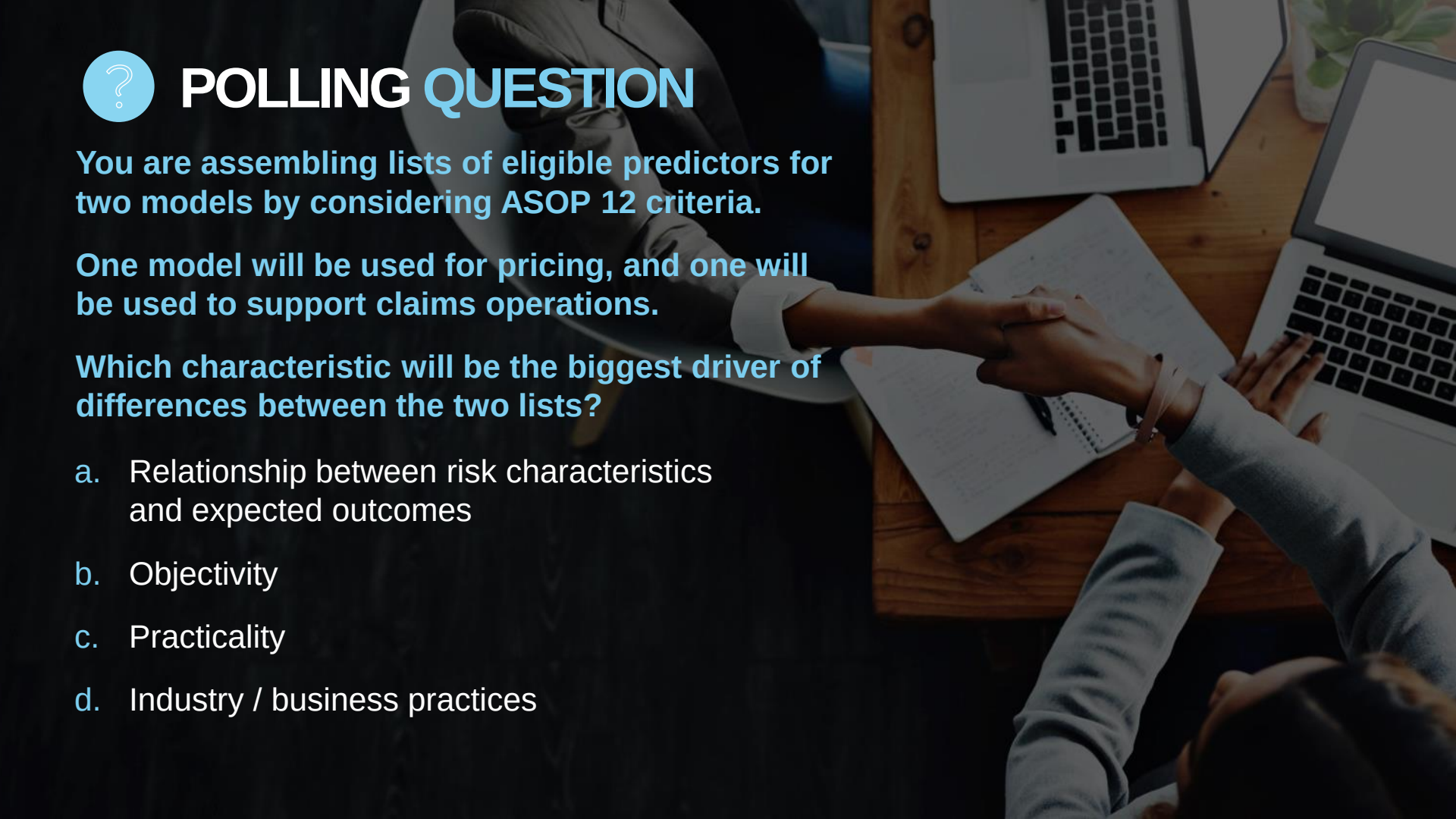
Which characteristic will be the biggest driver of differences between the two lists?

a. Relationship between risk characteristics and expected outcomes

b. Objectivity

c. Practicality

d. Industry / business practices



MULTIVARIATE SURVIVAL MODELS

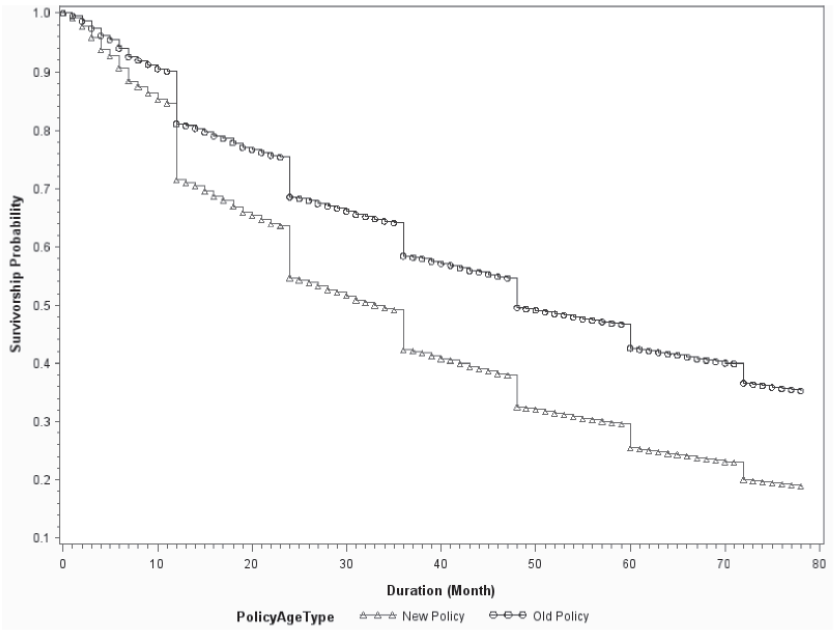
The Cox proportional hazards method

SOURCE MATERIAL

Policy Retention Analysis

- Builds on *Estimating Insurance Attrition Using Survival Analysis* by Luyang Fu and Hongyuan Wang
- Model the probability that a policy will be in force greater than X days.
- Right censoring: if a policy has not been cancelled, and has been in force for Y days, its cancellation time is greater than Y.

Figure 4. Survival curves for new vs. 5-year policies



economical

PROPORTIONAL HAZARDS

Approach due to Cox (1972)

- Survival function: $S(t) = P(T \geq t)$
- Hazard rate: $h(t) = -\frac{S\prime(t)}{S(t)}$
- $S(t) = \exp(-\int_0^t h(t)dt)$
- Cox: $h(t) = h_0(t)\exp(\boldsymbol{\beta x})$

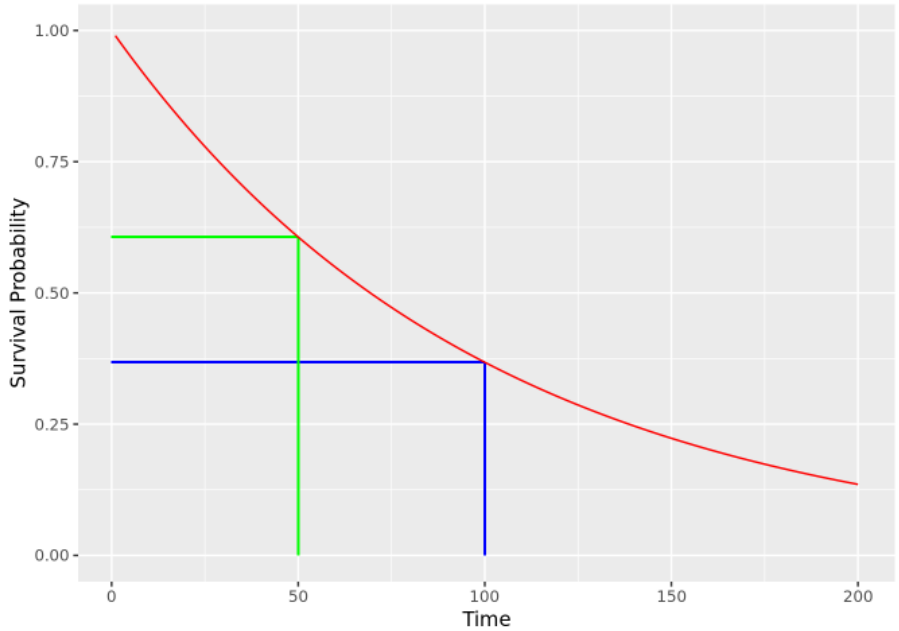
Think: Kaplan-Meier

Think: GLM

economical

ADVANTAGES OF COX MODELS

- Producing S(t) provides flexibility in how we define “prediction”
- More responsive to recent data
- Similar to familiar actuarial techniques



economical

IMPLEMENTATION OPTIONS

- R “survival” package
  (+ “survminer” for plots)
- Python “lifelines” package
- SAS “PHREG” procedure
- SPSS

MODEL CONSTRUCTION RECIPE

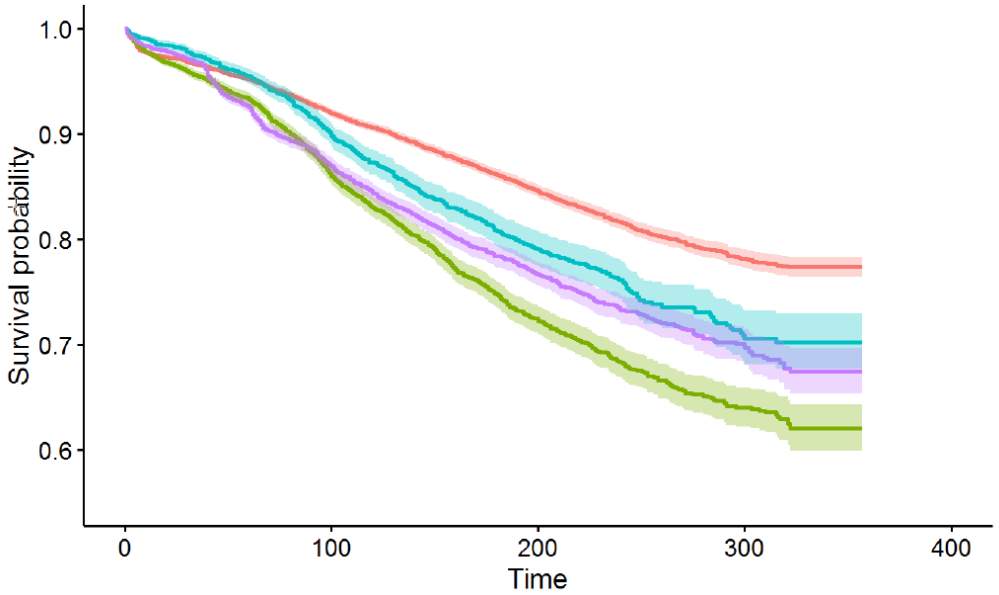
$$\boldsymbol{h(t) = h_0(t)\exp(\beta x)}$$

1. Select x and β using “the usual” linear modelling approaches
2. Test proportionality assumption
3. If not proportional: fit a *strata* (different $h_0(t)$ for each level of the variable)

PROPORTIONALITY CHECK

The quick check

- Produce a survival curve for each level of a variable
- Look for *qualitative* differences in the shape of the curve (e.g. crossing) which indicate non-proportionality

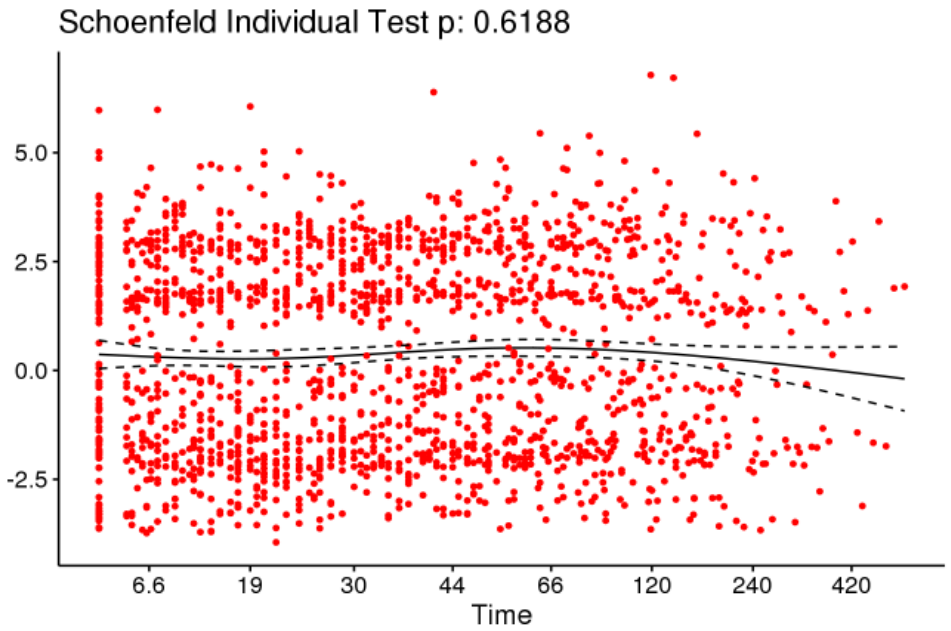


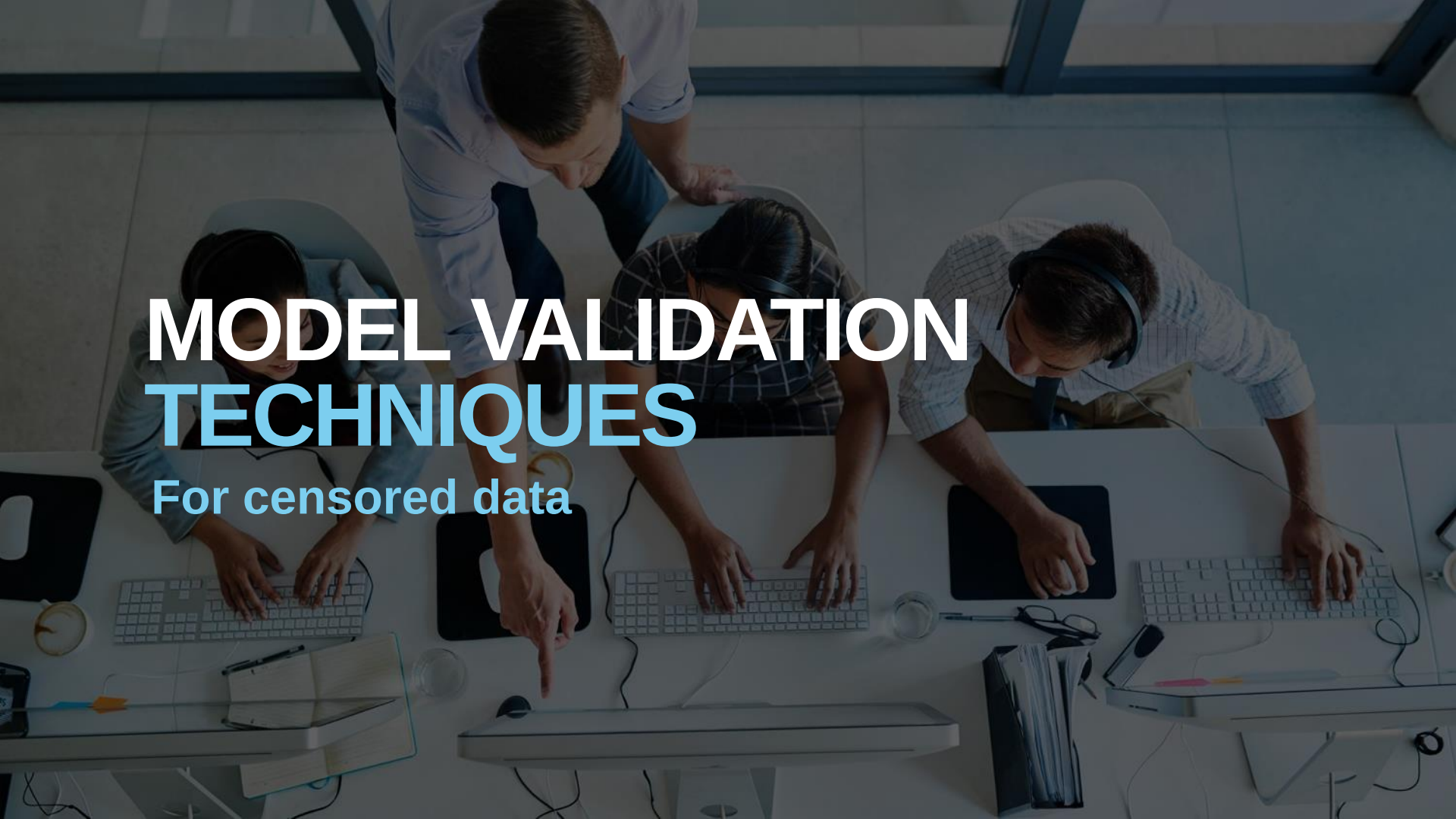
economical

PROPORTIONALITY CHECK

The rigorous check

- Schoenfeld Residuals Test
- Plot residuals vs. time
- Patterns in the residuals indicate non-proportionality





MODEL VALIDATION TECHNIQUES

For censored data

VALIDATION CHALLENGES

How would you validate the model on holdout data?

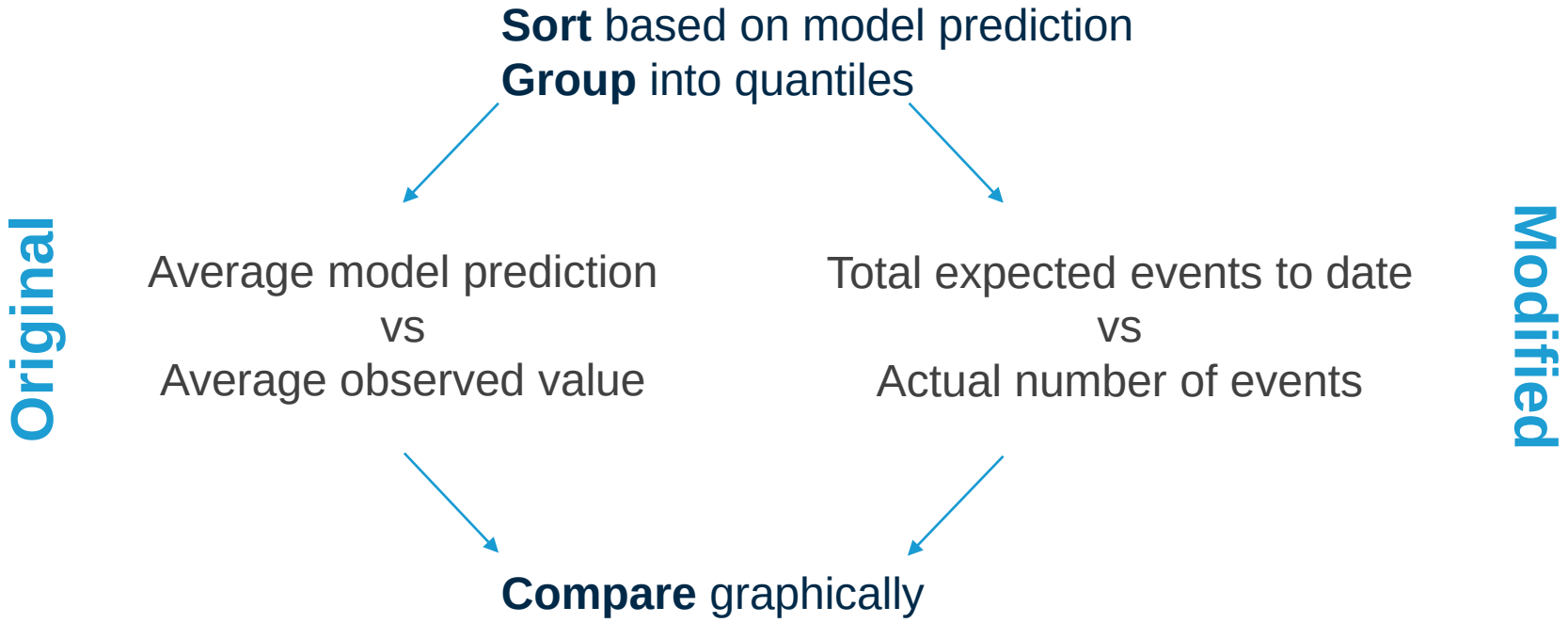
CLAIM ID	CLAIM STATUS	DAYS OPEN	LEGAL REP?	MODEL PREDICITION	FLAGGED BY MODEL?
1	Open	200	Yes	0.75	Yes
2	Open	10	No	0.65	Yes
3	Closed	450	Yes	0.3	No
4	Closed	250	No	0.2	No

POLLING QUESTION

Which of the following holdout testing methods will need to change to reflect censored data?

a. False positive / false negative rates
b. Gini coefficient
c. Quantile plots

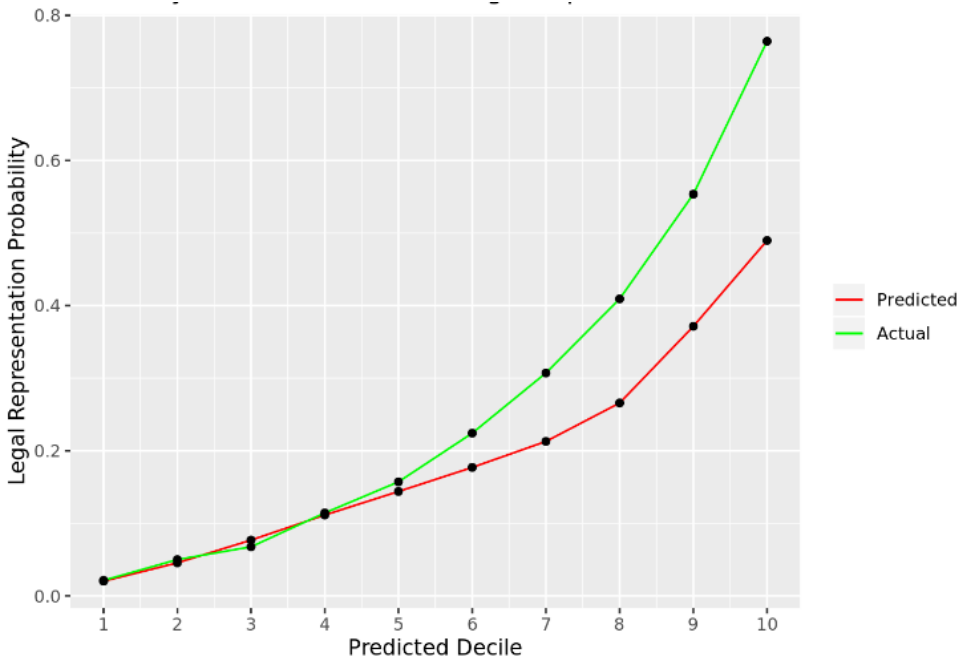
MODIFIED QUANTILE PLOT



MODIFIED QUANTILE PLOT

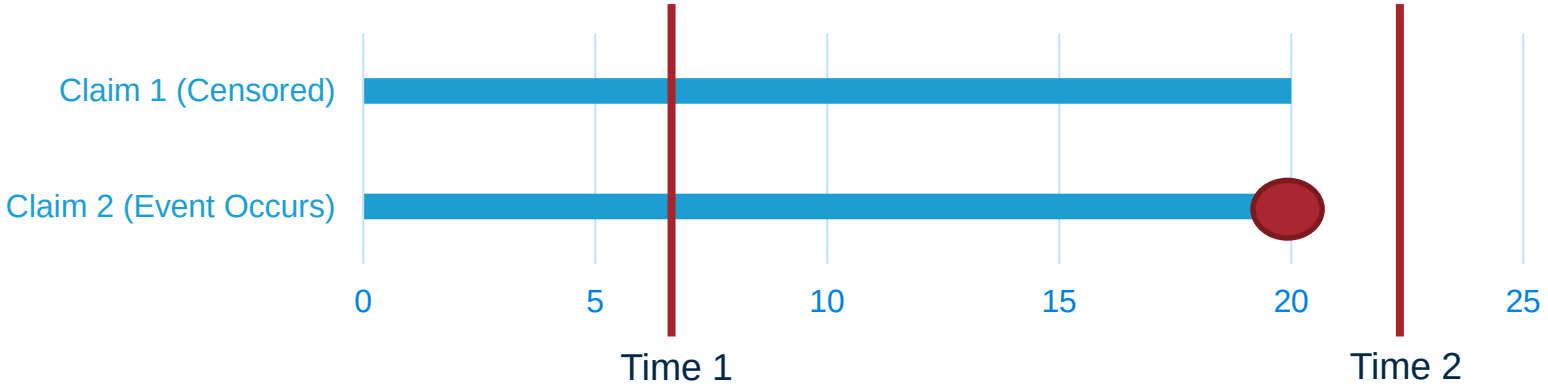
Calculated on holdout data

- Good segmentation of high vs. low risk of legal representation
- Underestimates the absolute probability of legal representation
- Appropriate for use cases involving flagging the top risks



economical

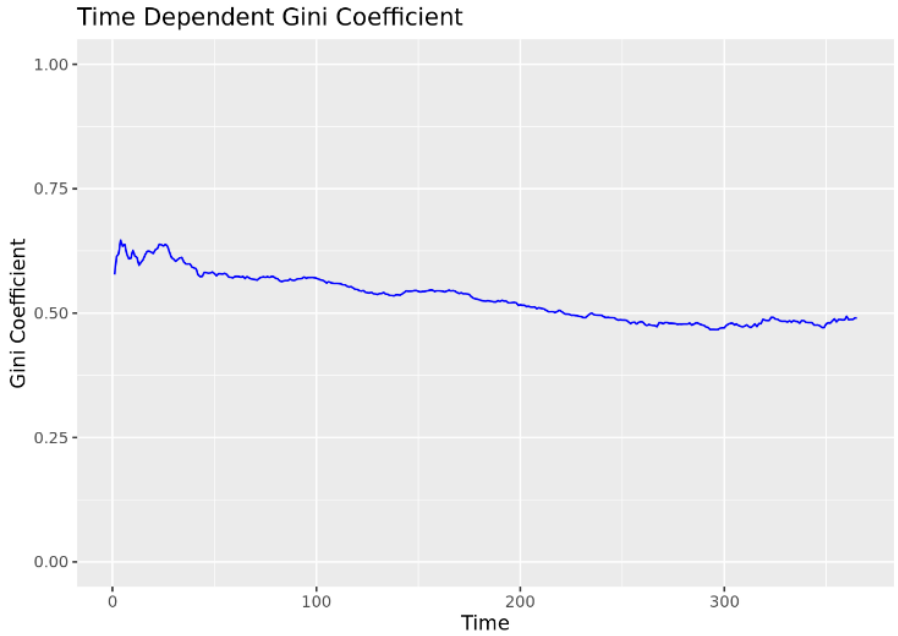
TIME-DEPENDENT
SENSITIVITY AND SPECIFICITY



	Time 1	Time 2
Cumulative sensitivity	Claim 2 Negative	Claim 2 Positive
Dynamic specificity	Claim 1 Negative	Claim 1 Excluded

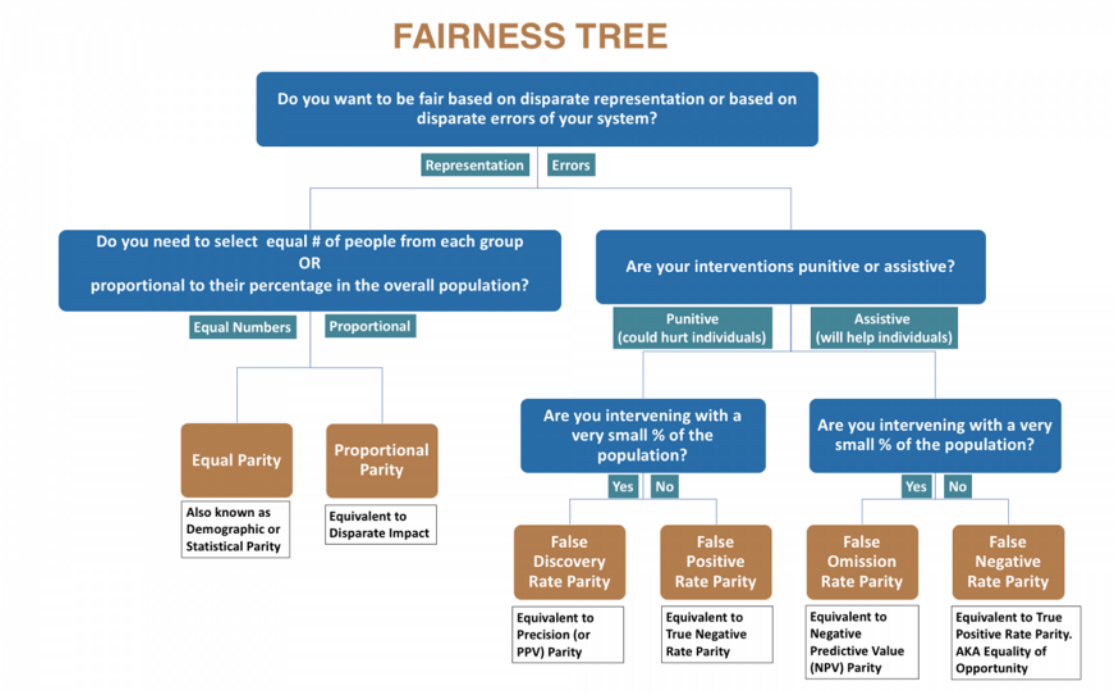
TIME-DEPENDENT GINI / AUROC

- Fix a time *t*
- Apply the CS / DS Rules
- Calculate Gini / AUROC
- Repeat for each *t*



economical

BIAS AND FAIRNESS AUDIT



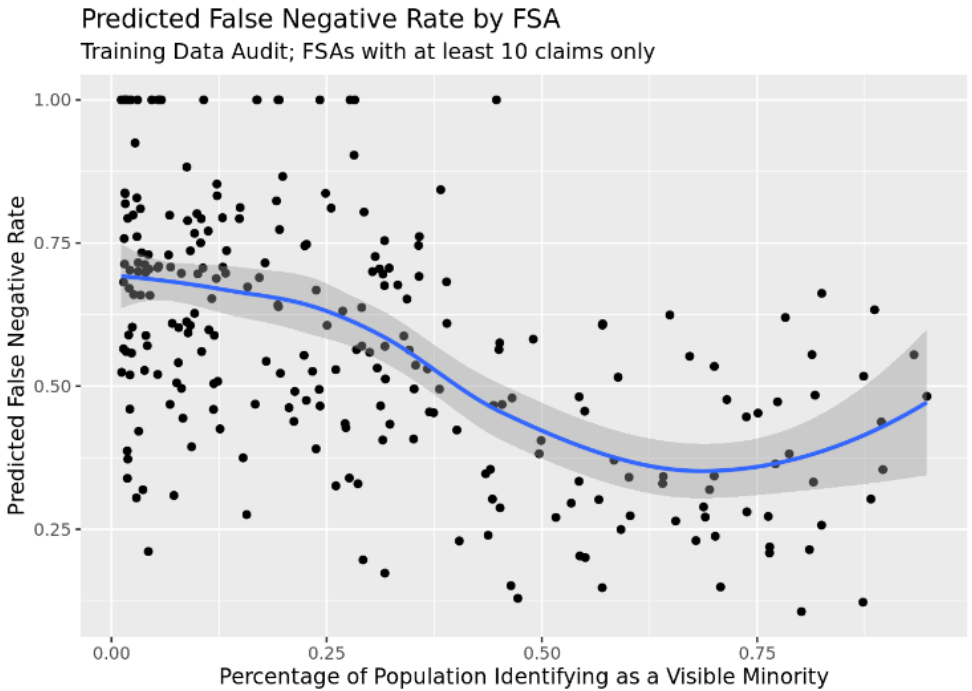
Source: https://dsapp.uchicago.edu/projects/aequitas/

economical

BIAS AND FAIRNESS AUDIT

Our modifications

- Census averages
- Scatterplot should not show an increasing trend
- Used *predicted* false negative rate



economical

Thank you