Correcting Reporting Delays in Cyber Events at Industry Level

Seema Sangari (Presenter)
PhD Candidate
Kennesaw State University
SSangar1@students.Kennesaw.edu
linkedin.com/in/seemasangari

Dr. Eric Dallal
AIR Worldwide
Verisk Cyber Solutions

Scott Stransky
AIR Worldwide
Verisk Cyber Solutions

KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING
School of Data Science and Analytics

AIR

Outline

- Business Problem
- Solution in Development
- Key Take-aways

Business Problem

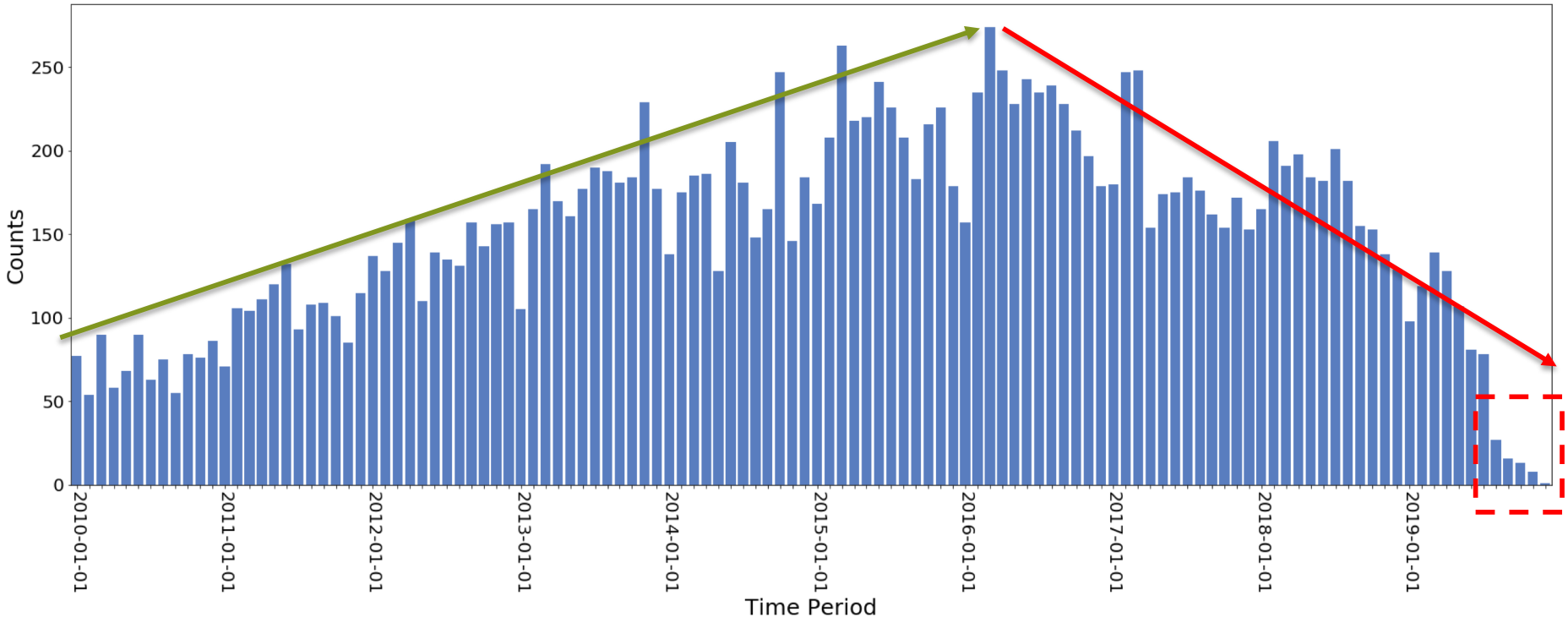
- Reporting Delays: Takes up to 10 years for the event to be entered into the database
- Cyber Models with such incomplete information would be questionable
- Business Problem: To correct for the false diminishing trend in counts due to reporting delays

Business Problem: Data

AIR Worldwide Proprietary data

When?		Where?	Who?
Incident Date	Reporting Date	Country	Industry

Business Problem : Monthly Cyber Event Counts (reported) US Finance and Insurance Industry

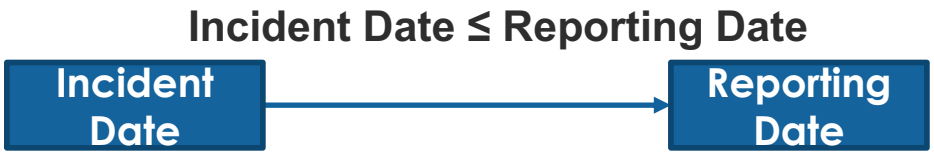


KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING
School of Data Science and Analytics

AIR

Solution in Development: Delay and Age

- Delay:



- Age :

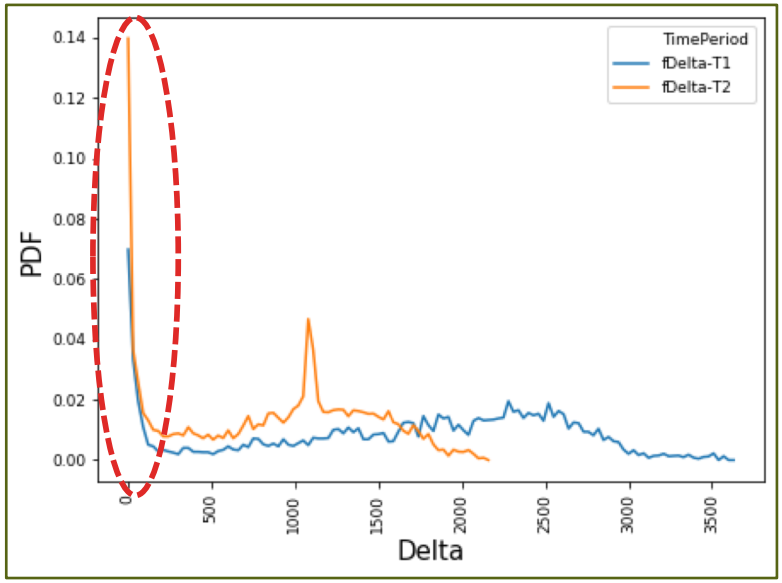
Solution in Development: Delay Distribution

Delay Distribution generated based on Delay Ratio

$$\frac{\#Events\ with\ delay}{Estimated\ True\ \#Events\ where\ age \geq delay}$$

Solution Development: Problem with distribution generated from Delay ratio

- Non-stationarity observed
- Delay Distribution does not estimate beyond maximum Delay



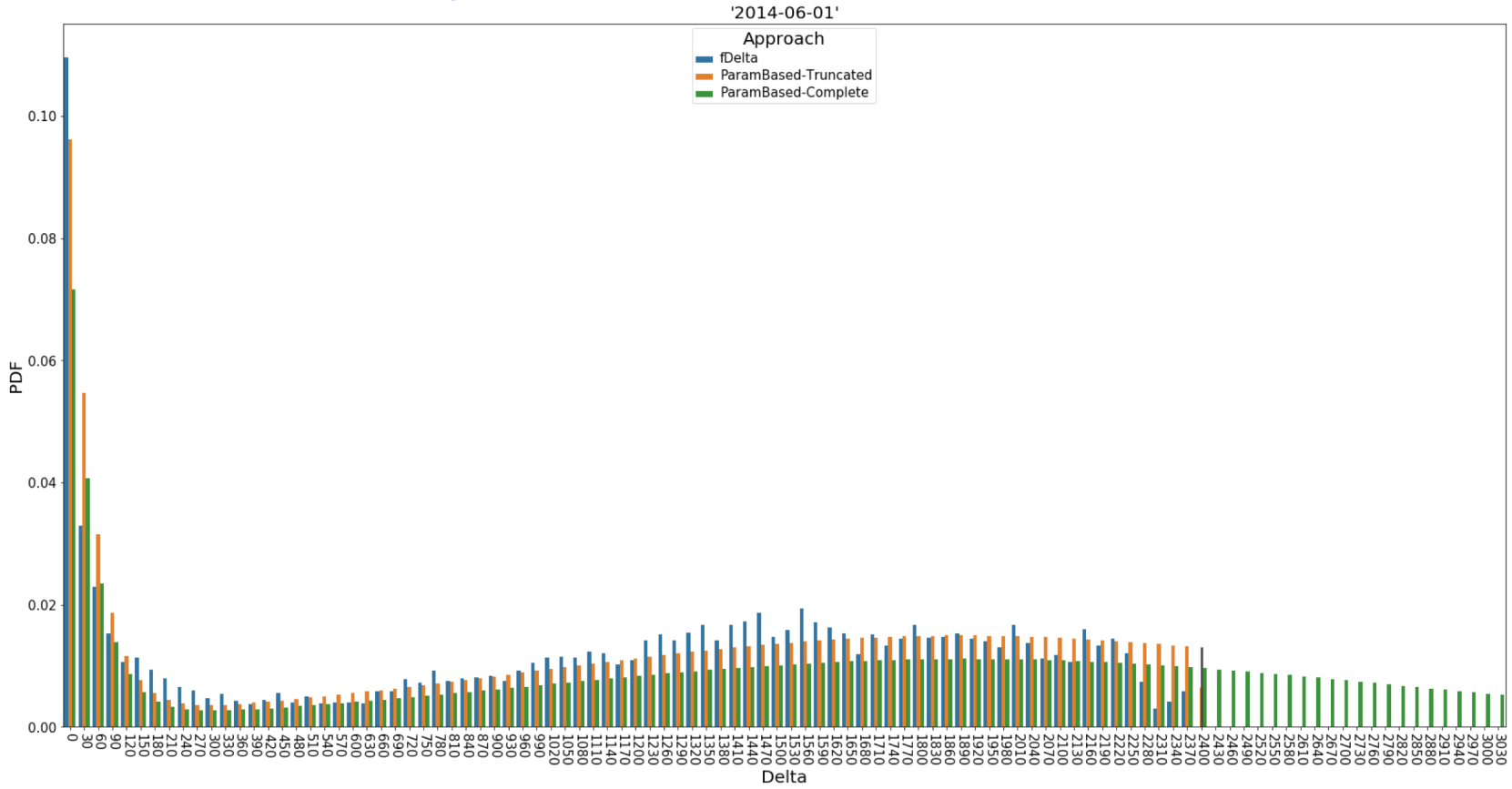
Modeled distribution is a mixture of an Exponential and a Normal distribution

$$CDF\ at\ \delta, F_\theta = \alpha\left(F_{Exp}(\delta, 0, Scale)\right) + (1-\alpha)\left(F_N(\delta, \mu, \sigma)\right)$$

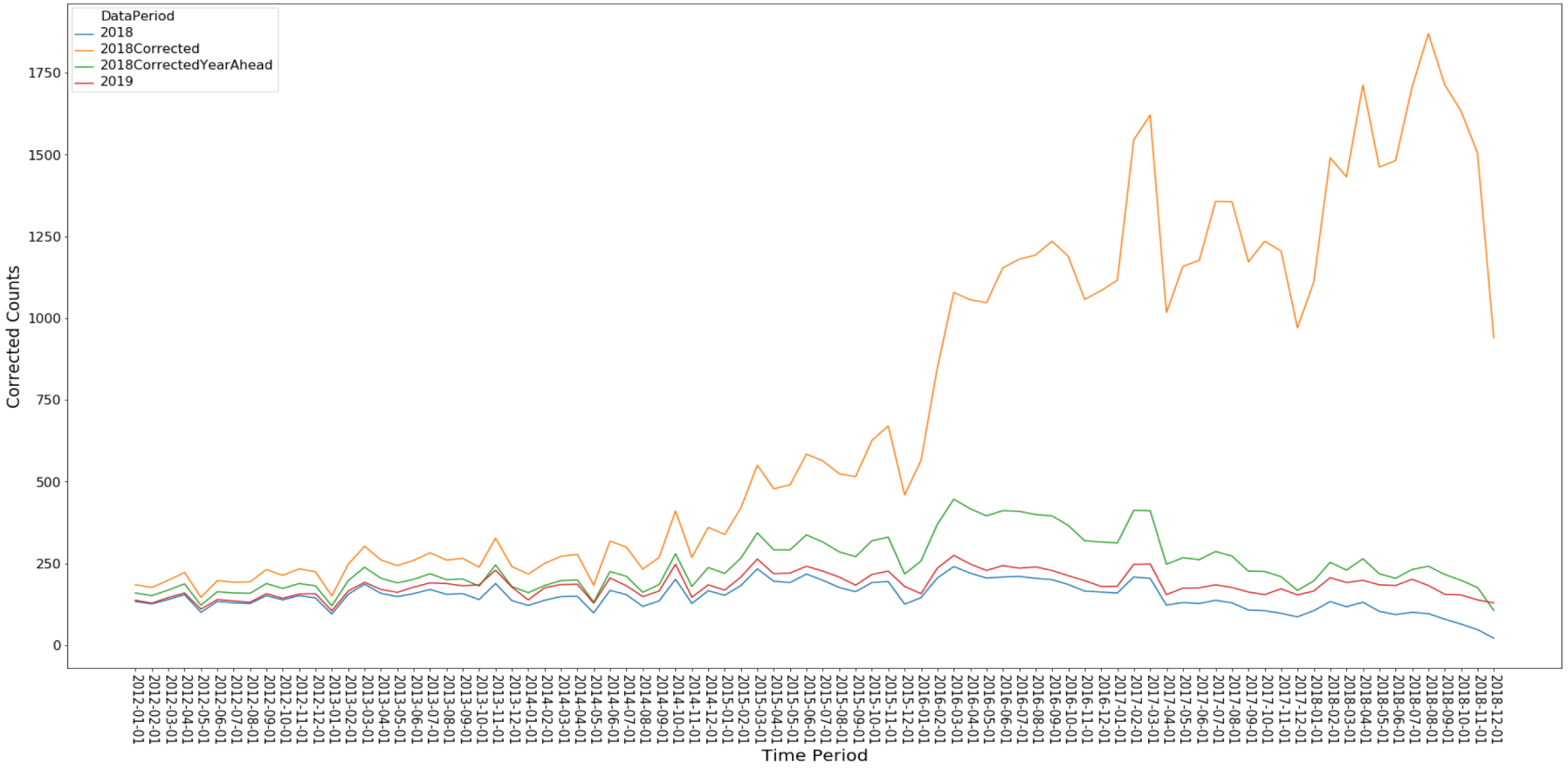
KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING
School of Data Science and Analytics

AIR

Solution in Development: Function Performance



Solution in Development: Validation Plot

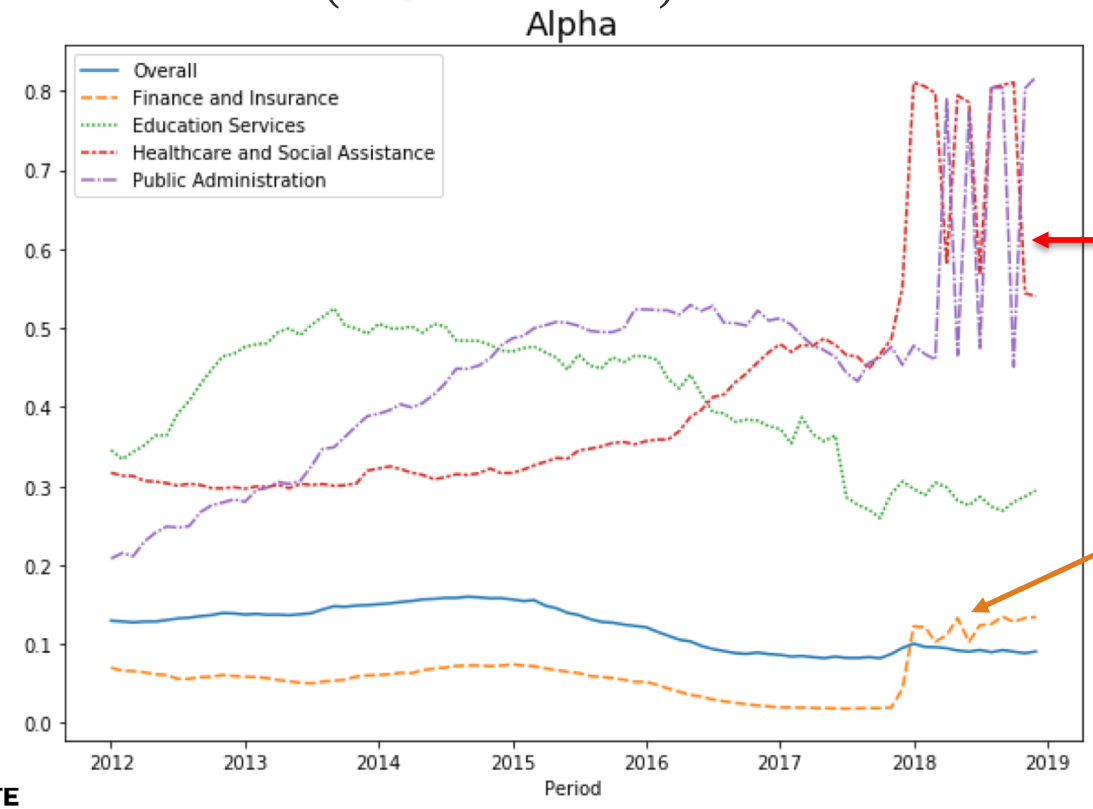


KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING
School of Data Science and Analytics

AIR

Key Take-away: Alpha - Comparison at Industry level

$$CDF\ at\ \delta, F_\theta = \boldsymbol{\alpha}\left(F_{Exp}(\delta, 0, Scale)\right) + (1 - \boldsymbol{\alpha})\left(F_N(\delta, \mu, \sigma)\right)$$



Gets detected at shorter delays

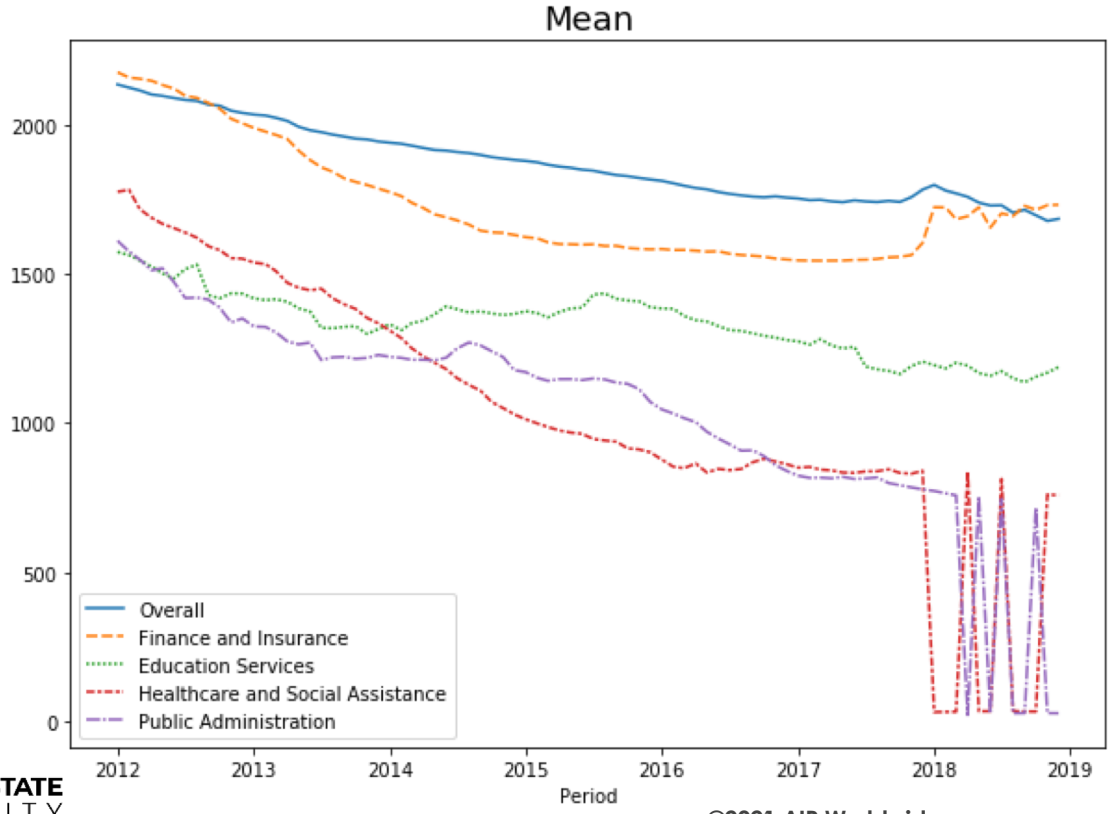
Not being detected at shorter delays

KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING
School of Data Science and Analytics

AIR

Key Take-away: Mean - Comparison at Industry level

$$CDF\ at\ \delta, F_\theta = \alpha\left(F_{Exp}(\delta, 0, Scale)\right) + (1-\alpha)\left(F_N(\delta, \boldsymbol{\mu}, \sigma)\right)$$



Key Take-away

In order to obtain credible cyber model

- Despite most recent cyber database, data requires debiasing
  - Reporting delays need to be addressed
- Each industry has its own modeled delay distribution
  - Different industries had large differences
- Better decision made with better data

Questions

Solution in Development: Optimization Function

Optimization Function

1. Compares empirical Delay Distribution with Modeled Delay Distribution up to maximum delay
2. Compares Modeled Delay Distribution beyond maximum delay for two consecutive months
3. Penalizes negative delays

$$CDF\ at\ \delta, F_\theta = \alpha\left(F_{Exp}(\delta, 0, Scale)\right) + (1-\alpha)\left(F_N(\delta, \mu, \sigma)\right)$$

$$\theta = (\alpha, Scale, \mu, \sigma)$$

$$\boldsymbol{\theta_{Opt}} = \underset{\theta=(a,Scale,\mu,\sigma)}{\operatorname{argmin}} \underbrace{\frac{1}{n}\|\log_{10} F_\theta - \log_{10} F_\Delta\|^2}_{Truncated\ until\ \delta_{max}} + \underbrace{(F'_{\theta'} - F'_\theta)^2}_{Beyond\ \delta_{max}} + \underbrace{F_N^2(0, \mu, \sigma)}_{Below\ \delta=0}$$

$$F_\Delta = Monthly\ Delay\ Distribution\ rolling\ over\ 2-year\ window$$

$$F_\theta = F_\theta\ defined\ over\ (0, \infty)$$

Thanks