# Third-Party Litigation Finance Derivative Pricing

Johanan Anton Pranesh & Gregory Radu Colonescu
JohananAnton@outlook.com & Gregory.Colonescu@gmail.com

#### Introduction

- What is TPLF?
- What are we trying to do?

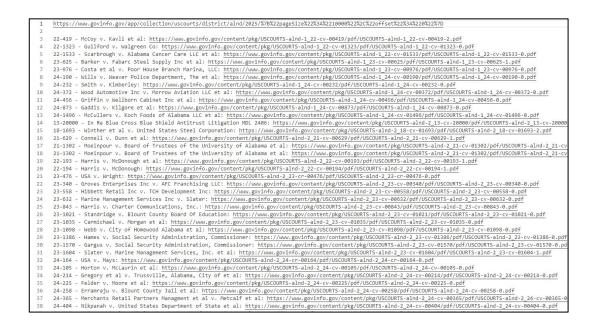
- Funding litigation for a share of the earnings
- We are trying to more accurately price these lawsuits and TPLF contracts

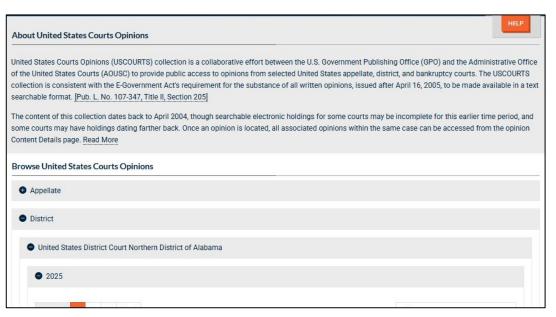
### Methodology

- Web Scraping
- PDF Parsing (LLMs)
- Plug into models:
  - TFIDF + LegalBERT
  - Neural Network
  - Discriminant Analysis
  - PDE based on Black-Scholes-Merton

### Web Scraping for Case Documents

- Identified U.S. district court cases and settlements
- Built code to scrape class action PDFs
- Extracted case outcomes, case duration, judge names
- Curated dataset of class actions





## Signal Generation for Feature Engineering

- Analyzed case documents using NLP models
- Generated predictive signals based on case structure
- Examples: defendant industry, type of allegation, judge political party
- Integrated signals into dataset for modeling

Case 1:18-cv-05291-ARR-RML Document 1 Filed 09/20/18 Page 1 of 12 PageID #: 1

#### UNITED STATES DISTRICT COURT EASTERN DISTRICT OF NEW YORK

DANIEL ABRAHAMOV, on behalf of himself and all others similarly situated,

Distriction

CLASS ACTION COMPLAINT AND DEMAND FOR JURY TRIAL

-against

PORTFOLIO RECOVERY ASSOCIATES

Defendant

Plaintiff DANIEL ABRAHAMOV (hereinafter, "Plaintiff"), a New York resident, brings this class action complaint by and through his attorneys, Cohen & Mizrahi LLP, against Defendant PORTFOLIO RECOVERY ASSOCIATES, LLC ("Defendant"), individually and on behalf of a class of all others similarly situated, pursuant to Rule 23 of the Federal Rules of Civil Procedure, based upon information and belief of Plaintiff's counsel, except for allegations specifically pertaining to Plaintiff, which are based upon Plaintiff's personal knowledge.

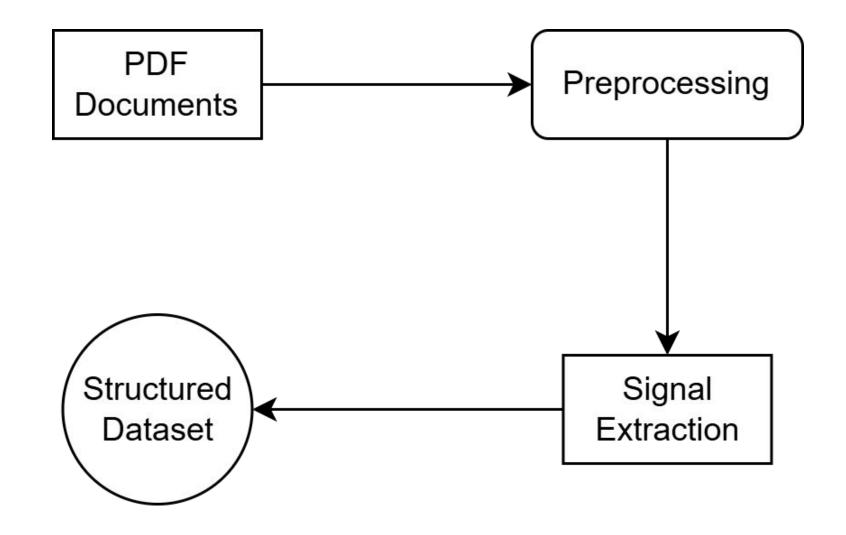
#### INTRODUCTION/PRELIMINARY STATEMENT

1. Congress enacted § 1692 et seq. of Title 15 of the United States Code, commonly referred to as the Fair Debt Collections Practices Act ("FDCPA") in 1977 in response to the "abundant evidence of the use of abusive, deceptive, and unfair debt collection practices by many debt collectors." 15 U.S.C. § 1692(a). At that time, Congress was concerned that "abusive debt collection practices contribute to the number of personal bankruptcies, to marital instability, to the loss of jobs, and to invasions of individual privacy." Id. Congress concluded that "existing laws...

[we]re inadequate to protect consumers," and that "the effective collection of debts" does not require "misrepresentation or other abusive debt collection practices." 15 U.S.C. §§ 1692(b) & (c).

General	,	Class Action		
Signal	Where you can find it	Signal	Where you can fin	
Breyers		Breyers		
Date Filed	Top of Page	Mention of Mission Statement	Introduction	
Law Firm	Top of Page	Mention of Monopoly	Introduction	
Law Firm State	Top of Page	Mention of Unaccountability	Introduction	
Court District	Top of Page	Actions Resulting in Competitory Exclus Introduction		
Court State	Top of Page	Federal Trade Commission Involvement		
Number of Plaintiffs	Top of Page	Federal Law Violations		
Defendent Name (Company)	Top of Page	Death of Plaintiffs		
Defendent designing of products or services	Intro	Violation of EULA/SLA		
Primary Business or Secondary Business	Intro	Additional Notes		

#### From Scraping to Structure

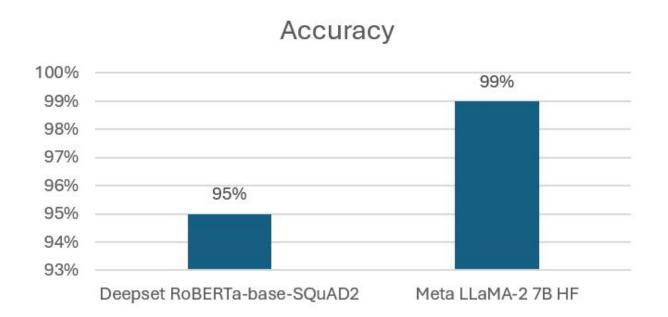


### **Modular Signal Extraction**

Feature Type	Technique		
Prices & %	Regex		
Legal Themes	WordNet keywords		
Product vs Service	Zero-shot model check		
Entities	spaCy NER		
Structure	Heuristics (TOC, Delaware)		

#### **Verifying Signals with LLMs**

- BERT used for extraction
- RoBERTa & LLaMA-2 used for validation
- Asked LLMs: "Is this value correct?"
- 95–99% accuracy on clean documents



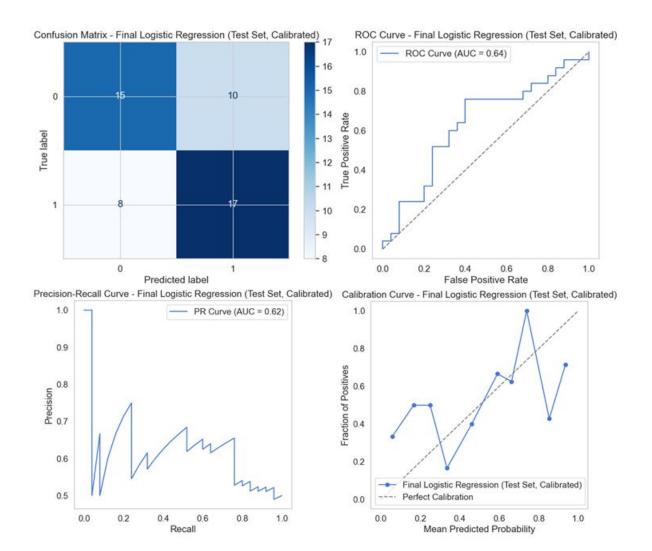
#### **Neural Network Model**

- Coding-wise, a relatively simple ReLU, or Rectified Linear Unit, neural network using the tensorflow library
- Errors as low as 12% in Diff 0%1 column (0% is incorrectly calculated)

	AmountWon	AmtWonPred	Diff0	Diff0%0	Diff0%1
Unnamed: 0					
foglesong-v-marquis-metal-works-llc.pdf	1.900000e+03	-21921770.00	21923670	0	115
ranieri-v-premier-fire-alarms-and-integration-systems-inc.pdf	1.740000e+04	34581264.00	-34563864	0	-19
schneider-v-natera-inc-et-al.pdf	3.980000e+04	50581504.00	-50541704	0	-12
doyle-v-florida-health-solution-inc.pdf	1.000000e+04	19870360.00	-19860360	0	-19
cummins-et-al-v-the-united-states-of-america.pdf	1.350000e+09	-34673616.00	1384673616	0	0
husar-v-dolgen-midwest-llc.pdf	1.000000e+02	30322638.00	-30322538	0	-3032
koller-v-harley-davidson-motor-company-group-llc-et-al.pdf	0.000000e+00	-58415244.00	58415244	0	584152440000
sharon-v-cac-financial-corp.pdf	2.018790e+05	6824042.00	-6622163	0	0
boone-et-al-v-amazon-com-services-llc.pdf	5.500000e+08	-4794697.50	554794697	-1	0
williams-et-al-v-centene-corporation-et-al.pdf	4.000000e+02	43753944.00	-43753544	0	-1093
dettmering-et-al-v-vbit-technologies-corp-et-al.pdf	1.500000e+03	-74215488.00	74216988	0	494
houch-v-maricopa-county.pdf	3.000000e+02	3617314.75	-3617014	0	-120
edge-v-roundpoint-mortgage-servicing-corporation.pdf	1.280000e+07	-60838948.00	73638948	0	0

#### TFIDF + LegalBERT

- Information from the Initial Filings alone can predict the outcome with 64% accuracy
  - With external features about the plaintiffs, the defendants, and more, it could increase further
- Consistent across multiple machine learning models

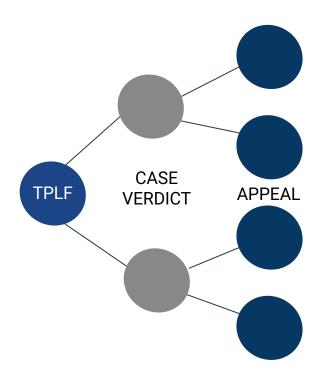


#### **Discriminant Analysis**

- Discriminant analysis is a way of creating result categorization
- Linear DA Accuracy for Amount Won: 96%
  - Equation has 379 variables
- Quadratic DA Accuracy: 77%
- Both are on skewed data

#### Why Black-Scholes?

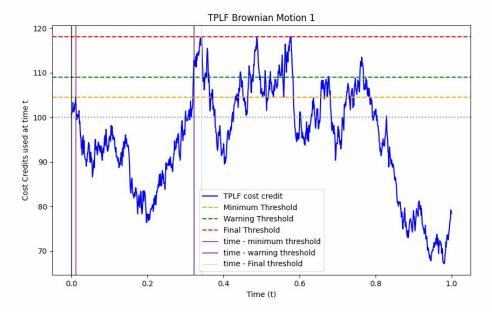
- Structured way to evaluate uncertain future outcomes
- Consistency and standardization
- Mathematical models are transparent and interpretable
- Don't require massive datasets to make good estimates

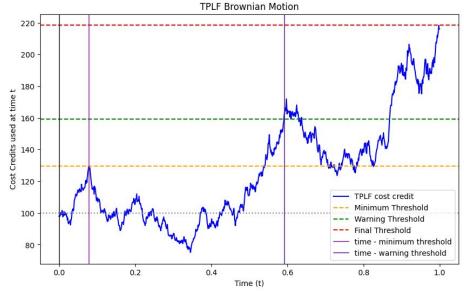


# TPLF PDE (Partial Differential Equation)

The PDE is modelled as on the underlying cost per unit time utilized for the case. The cost per unit time is an effective way to model an asset that is not traded actively and is discontinuous in nature.

The PDE was developed with the intention to be sensitive to excessive risks thus it moves with the expectation of potential risk outcome. Uncorrelated risks are anticipated using the cumulative risk measures for better understanding.





#### **TPLF PC Formula**

C - Capital committed for litigation

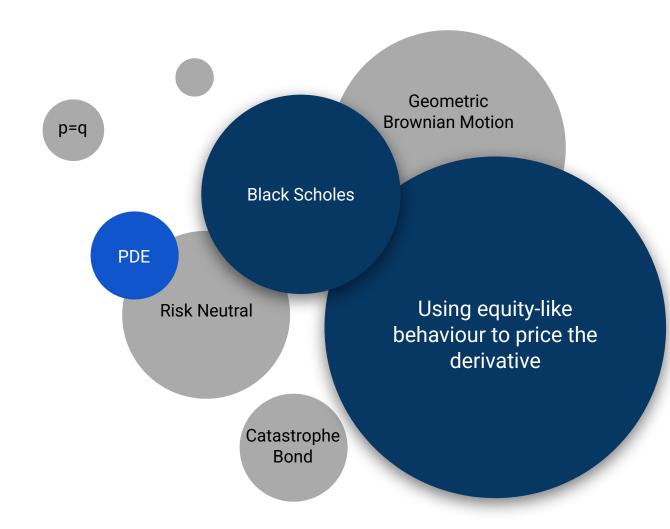
C<sub>t</sub>- Cost per unit time

α - Firm risk measure

β - negotiated rate of return

The PC formula uses the discounted C<sub>t</sub> and the probabilities for expected payout and expectation that the current cost doesn't exceed total cost.

The alpha and beta are case driven and helps control firm risk. In fact, alpha is correlated with beta.



#### **Key Takeaways**

- Using Black Scholes framework for pricing TPLF
   Derivatives, we are fully capture all the aspects of credit risk as well as litigation risk
- The PC formula has shown effectiveness using synthetic data and it is shown to be 20% more effective in pricing
- The framework allows for the asset to be traded on the longer duration as well as on the shorter end



#### Thank you

To answer any further questions, please contact:

Johanan Anton Pranesh johanananton@outlook.com

Gregory Radu Colonescu Gregory.Colonescu@gmail.com