

NHH



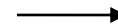
Master Thesis Presentation

CAN WE UNVEIL THE PRESENCE OF SECRECY HAVENS?

Agenda

1

Web Crawling & Text Mining
from Edgar Server



All Public Listed US
Companies 2018

2

Data Analysis, Categorization and Detecting Patterns

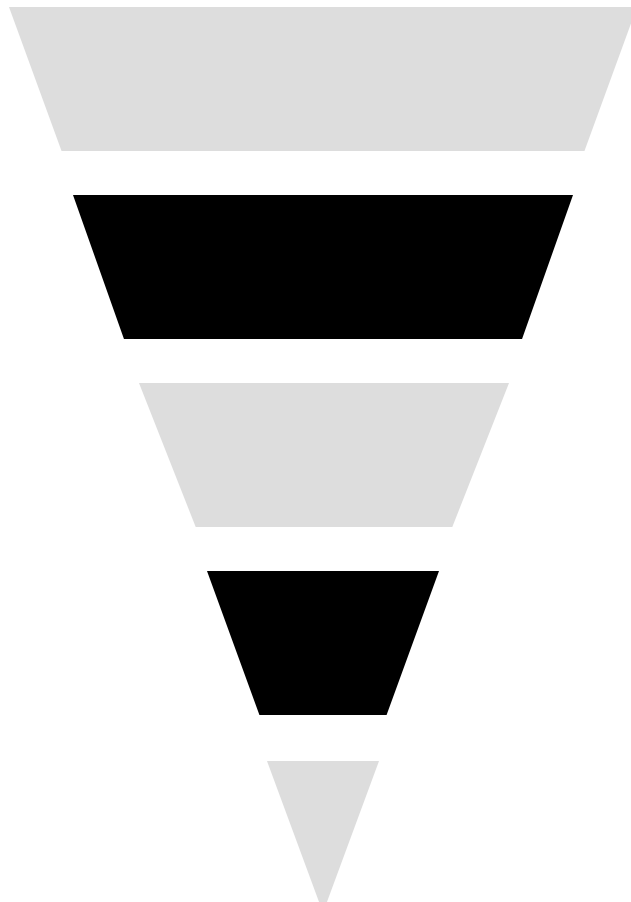
3

Predicting Secrecy Haven Intensity

Problem Statement

- Thesis research delves into interaction between **financial reporting, tax avoidance and secrecy exploitations** by public listed US companies
- Public listed companies are required to file and disclose all operations and subsidiary locations with U.S. Securities and Exchange Commission (SEC)
- Companies have to disclose all their affiliate operations but they are not fact checked and they can fail to disclose subsidiaries in secrecy havens
- We **extracted** a novel dataset of all the location data for declared US companies in 2018 from the Edgar Server by web crawling
- The Secrecy Haven location data across all the companies was collected, quantified and analyzed
- Can we predict the intensity of secrecy havens in case firms fail to disclose?
- We will address the issue by creating (1) Bins according to Secrecy Haven Intensity (2) Weighted bins with number of companies

Data Extracted



Total CIK = 7,093

55 %

Total CIK With Ex 21 = 3,880

44 %

CIK Extracted = 3,140

36 %

CIK joined with WRDS = 2,540

18 %

Observations after
removal of missing data = 1,258



Explaining Data Extraction

sec.gov/Archives/edgar/data/1439222/000143922218000004/agio-ex211x123117.htm

Apps Bookmarks iPad til skoletechste... WE - Webcruiter Inbox (2,366) - dani... What is Azure Powe...

EX-21.1 2 agio-ex211x123117.htm EXHIBIT 21.1

Exhibit 21.1

Entity

Agios Securities Corporation
Agios International Sarl
Agios Limited

SUBSIDIARIES

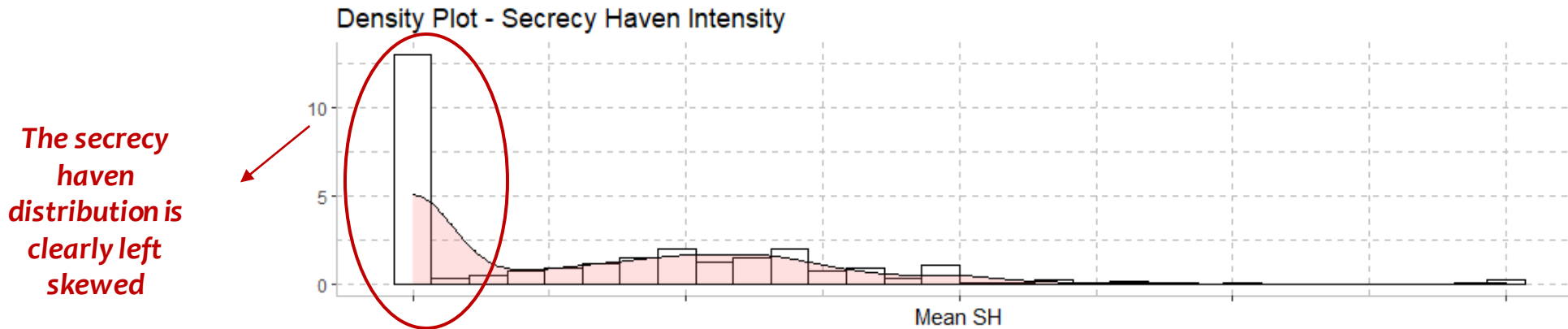
State or other Jurisdiction of Incorporation or Organization
Massachusetts
Switzerland
<u>Bermuda</u>

Secrecy Haven Jurisdiction

Subsidiary	Jurisdiction	Secrecy Haven
Agios Security Corporation	Massachusetts	0
Agios International Sarl	Switzerland	1
Agios Limited	Bermuda	1
		$2/3 = 67\%$

Mean Secrecy Haven

Data View & Methodology



Dependent Variable:

Haven Aggressiveness

- Haven Aggressiveness is a factor variable made from Mean Secrecy haven Score dividing Mean Secrecy Haven Score into 5 bins of 0.2
- 0.4** on Haven Aggressiveness means 40% of the jurisdictions are secrecy havens

Independent Variables:

- (1) Assets (2) Liabilities (3) Pre Tax Domestic Income (4) Pre Tax Foreign Income
(5) Property Plant Equipment (6) Tax Foreign (7) Tax Total

Haven Aggressiveness	1 %	—	1.0
	1 %	—	0.8
	8 %	—	0.6
	30 %	—	0.4
	60 %	—	0.2
		—	0.0

Top 10 Jurisdictions – Overall / Category

Rank	Location	# CIK
1	Netherlands	792
2	China	743
3	Singapore	671
4	Hong Kong	641
5	Japan	518
6	Switzerland	477

- **Netherlands, China** and **Singapore** came out to be the top secrecy haven places
- **Cayman Islands** and **Bermuda** makes into top 10 list

7	Cayman Islands	357
8	Bermuda	314
9	Jersey	314
10	Thailand	248

- Netherlands and Hong Kong are also havens, but they foster a lot of real activity, so they have reason for plausible deniability of being havens unlike Cayman Islands and Bermuda

#	Construction	FIR	Manufacturing	Mining	PA	Retail Trade	Services	TCEGS	WT
1	Jersey	Jersey	China	Netherlands	Switzerland	Hong Kong	Netherlands	Netherlands	Netherlands
2	Netherlands	Cayman Islands	Netherlands	Cayman Islands	Japan	China	Singapore	Jersey	Singapore
3	Chile	Hong Kong	Singapore	Singapore	Netherlands	Puerto Rico	China	Cayman Islands	China
4	Panama	Bermuda	Hong Kong	Bermuda	Singapore	Netherlands	Hong Kong	Bermuda	Switzerland
5	Indonesia	Singapore	Japan	British Virgin Islands	China	Bermuda	Japan	Hong Kong	Hong Kong
6	Singapore	China	Switzerland	Bahamas	Bermuda	Jersey	Switzerland	Singapore	Japan
7	Bermuda	Netherlands	Thailand	Venezuela	Cayman Islands	Japan	Jersey	Japan	Indonesia
8	Cayman Islands	Japan	Turkey	Indonesia	Hong Kong	Singapore	Israel	China	Puerto Rico
9	Japan	Switzerland	Cayman Islands	Switzerland	Jersey	Cayman Islands	Cayman Islands	Switzerland	Turkey
10	Puerto Rico	Puerto Rico	Taiwan	Ghana	Aruba	Switzerland	Philippines	Chile	Jersey



Cayman Islands



Panama



Bermuda



Bahamas

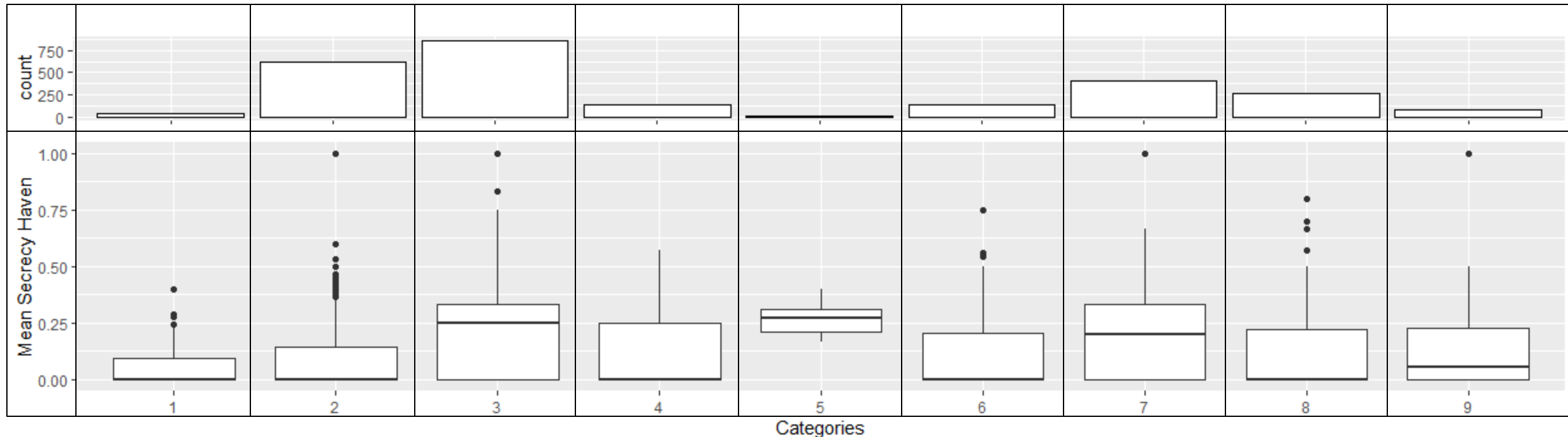
*Categories were created using SIC code list (ref. Slide 'References'); Data for Jersey was not totally accurate, thus ignoring Jersey

FIR: Finance, Insurance, And Real Estate; **PPE:** Property, Plant, Equipment; **PA:** Public Administration;

WT: Wholesale Trade; **TCEGS:** Transportation, Communications, Electric, Gas, And Sanitary Services

Diagnostic Analysis of Secrecy Haven Score

- The box plot gives a view of Mean Secrecy Haven versus Industrial Category and the plot above it gives count of companies in the category
- **Wholesale Trade, Services, Manufacturing & FIR** are industries where companies even operate 100% in secrecy haven locations
- Companies working in PA seem on average to be most SH intensive but number of firms in the category is low. Companies in **Manufacturing** and **Services** seem to have the biggest Secrecy Haven Score inter quartile range



*1 : Construction; 2 : Finance, Insurance, And Real Estate; 3 : Manufacturing; 4 : Mining; 5 : Public Administration; 6 : Retail Trade; 7 : Services;
8 : Transportation, Communications, Electric, Gas, And Sanitary Services; 9 : Wholesale Trade;

Prediction & Model Comparison

- Train – Test split = 70% - 30%
- Stratified sampling as the categories had unequal observations
- Predictions on test sample were realized from 5 different models including Random Forest, Gradient Boosting Machine, K Nearest Neighbor, Support Vector Machine, Multinomial Classification
- 10K cross validation was used in all the models
- Confusion tables from the models were concluded to determine model accuracy and comparison. **Gradient Boosting Machine** has the highest prediction accuracy of all the models


	CIK
Train Sample	1,780
Test Sample	760
	2,540

→ **Total Sample**

Random Forest	Actual (Accuracy: 68.7%)				
Prediction	1	2	3	4	5
1	387	92	31	5	6
2	70	135	27	3	2
3	1	1	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0

GBM	Actual (Accuracy: 71.2%)				
Prediction	1	2	3	4	5
1	409	93	25	5	6
2	47	132	32	3	1
3	0	3	0	0	0
4	1	0	1	0	1
5	1	0	0	0	0

KNN	Actual (Accuracy: 61.5%)				
Prediction	1	2	3	4	5
1	436	197	56	8	8
2	22	31	2	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0

 Only values highlighted in confusion tables represents correct predictions

SVM	Actual (Accuracy: 63.3%)				
Prediction	1	2	3	4	5
1	456	203	54	8	8
2	1	24	3	0	0
3	1	1	1	0	0
4	0	0	0	0	0
5	0	0	0	0	0

Multinomial	Actual (Accuracy: 62.4%)				
Prediction	1	2	3	4	5
1	453	204	56	8	8
2	4	21	2	0	0
3	1	1	0	0	0
4	0	0	0	0	0
5	0	2	0	0	0

Actual : Category 5
Predicted: Category 1

Calculating Accuracy:
 $(453+21+0+0+0)$
 $(453+204+56+8+8+4+21+2+1+1+2)$
 $= 62.4\%$

Model Accuracy & Variable Importance

- **'Foreign Tax'** is the most important variable for predicting secrecy haven intensity
- For rest of the predictor variables, various models assign differing importance whilst predicting

	Models				
	Random Forest	GBM	KNN	SVM	Multinomial
Prediction Accuracy	68.7%	71.2%	61.5%	63.3%	62.4%
	Variable by Importance				
1	Foreign Tax		Foreign Tax	Foreign Tax	Foreign Tax
2	PI For		PPE	PPE	PPE
3	PI Dom		Assets	Assets	PI Dom
4	Total Tax		Liabilities	Total Tax	Total Tax
5	Liabilities		PI Dom	Liabilities	PI For
6	Assets		Total Tax	PI Dom	Liabilities
7	PPE		PI For	PI For	Assets

GBM Prediction – Equal Bins

- A common pattern that comes out is all the models are **less accurate** while predicting companies with **higher intensiveness** in secrecy havens. A probable explanations can be **lesser data points** for the high intensive companies and therefore the model are weak learners in that area
- Instead, we divided the companies into more equal bins by number of companies
- Gradient Boosting Machine proved to be the best performer for this dataset and we predicted values with GBM
- The prediction accuracy came out to be 49.7%

Confusion Table
with Equal Bins

GBM	Actual (Accuracy: 49.7%)				
Prediction	1	2	3	4	5
1	276	36	26	27	38
2	31	22	11	10	10
3	16	18	30	23	15
4	5	20	20	28	20
5	13	9	18	17	22

Bin		Secrecy Haven Score Range
5	13.9 %	0.36 -1.0
4	13.8 %	0.29 – 0.36
3	13.8 %	0.20 – 0.29
2	13.8 %	0.02 – 0.20
1	44.8%	0.0 – 0.0

Testing Predictions

- I. From the predictions concluded we randomly picked some of the companies in Natural Resources and looked at how the model fared in predicting Secrecy Haven Intensity
- II. 6 out of 10 predictions turned out to be true.

Name	Secrecy Heaven Score	Category Distribution	Predicted Category	Prediction Status
Baker Hughes	0.143	2	4	
Cheniere Energy	0.429	5	1	
Adam Resources and Energy	0	1	1	
EP Energy Corp	0	1	1	
Superior Energy Services Inc.	0.125	2	3	
Chaparral Energy Inc.	0	1	1	
National Oil Well Varco	0.323	4	4	
Parsley Energy	0	1	1	
PBF Energy	0	1	1	
Berkshire Hathway Energy Co	0	1	4	

Summary

- I. Secrecy Heaven Score data was left skewed and the variations across the firm was not as well accounted for with Haven Intensity formation of bins
- II. We made equally weighed bins with respect to the number of companies
- III. Prediction Accuracy rate dropped and therefore we see room for improvement in predicting Secrecy Haven operations and intensity

Glossary

SH – Secrecy Haven

FIR – Finance, Insurance, And Real Estate

PPE – Property, Plant, Equipment

TCEGS – Transportation, Communications, Electric,
Gas, And Sanitary Services

Manu – Manufacturing

WT – Whole Sale Trade

Min – Mining

Cons – Construction

RT – Retail Trade

Serv – Services

SH – Secrecy Haven

SVM – Support Vector Machine

KNN – K Nearest Neighbor

GBM – Gradient Boosting Machine

WRDS – Wharton Research Database

References

¹<https://www.naics.com/sic-codes-industry-drilldown/>