Abnormality in Non-Profit Financial Data

Rameen Mahdavi

Github: https://github.com/rmahdavi

Email:Ramahdavi@gmail.com

Determining the right Non-Profit is hard.

- Historically Limited Data Available
- Little Incentive to release or collect data
- Difficult to measure performance

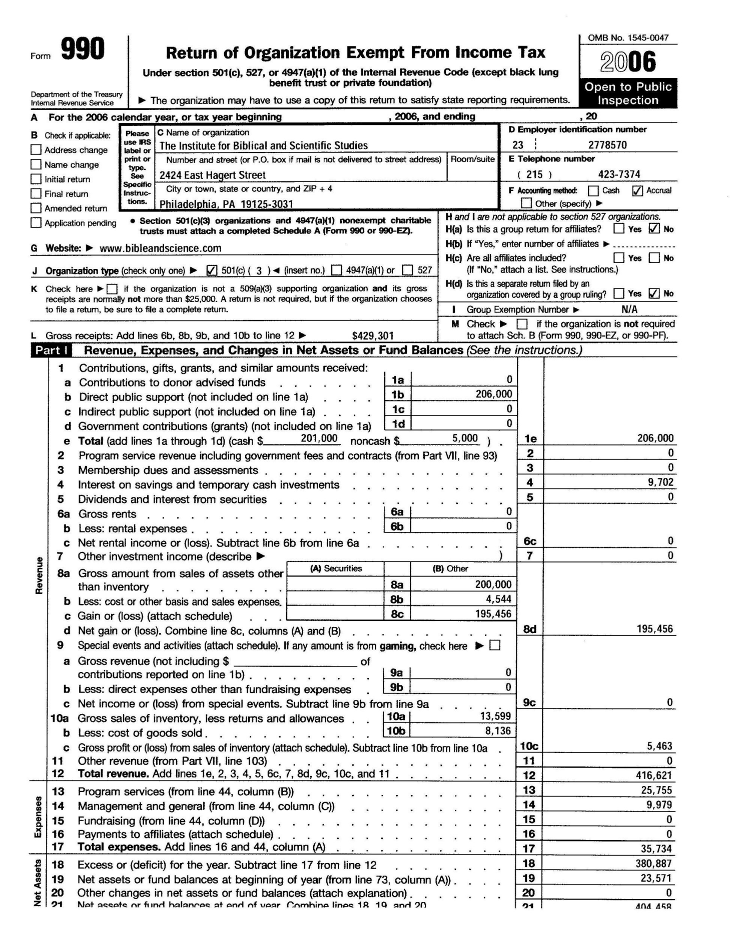
But determining the organization not to invest in might be easier.

Objective

Create an outlier model that will detect higher risk organizations that might need additional scrutiny and see if we can make any inferences.

Data

- Non-Profit Organization are required to file form 990 with the IRS to keep their tax exempt status
- Recently the IRS released the full form for all 220,000+ nonprofits that filed and posted parsable xmls on Amazon Web Services



Form **990**

Department of the Treasury
Internal Revenue Service

Return of Organization Exempt From Income Tax

Under section 501(c), 527, or 4947(a)(1) of the Internal Revenue Code (except black lung benefit trust or private foundation)

▶ The organization may have to use a copy of this return to satisfy state reporting requirements.

OMB No. 1545-0047

2006

Open to Public Inspection

A For the 2006 calendar year, or tax year beginning , 2006, and ending , 20

B Check if applicable:
□ Address change
□ Name change
□ Initial return
□ Final return
□ Amended return
□ Application pending

Please use IRS label or print or type. See Specific Instructions.

C Name of organization
The Institute for Biblical and Scientific Studies

Number and street (or P.O. box if mail is not delivered to street address) | Room/suite
2424 East Hagert Street

City or town, state or country, and ZIP + 4
Philadelphia, PA 19125-3031

D Employer identification number
23 | 2778570

E Telephone number
(215) 423-7374

F Accounting method: □ Cash ☑ Accrual
□ Other (specify) ▶

• **Section 501(c)(3) organizations and 4947(a)(1) nonexempt charitable trusts must attach a completed Schedule A (Form 990 or 990-EZ).**

H *and* **I** *are not applicable to section 527 organizations.*

H(a) Is this a group return for affiliates? □ Yes ☑ No

H(b) If "Yes," enter number of affiliates ▶

H(c) Are all affiliates included? □ Yes □ No
(If "No," attach a list. See instructions.)

H(d) Is this a separate return filed by an organization covered by a group ruling? □ Yes ☑ No

I Group Exemption Number ▶ N/A

M Check ▶ □ if the organization is **not** required to attach Sch. B (Form 990, 990-EZ, or 990-PF).

G Website: ▶ www.bibleandscience.com

J Organization type (check only one) ▶ ☑ 501(c) (3) ◀ (insert no.) □ 4947(a)(1) or □ 527

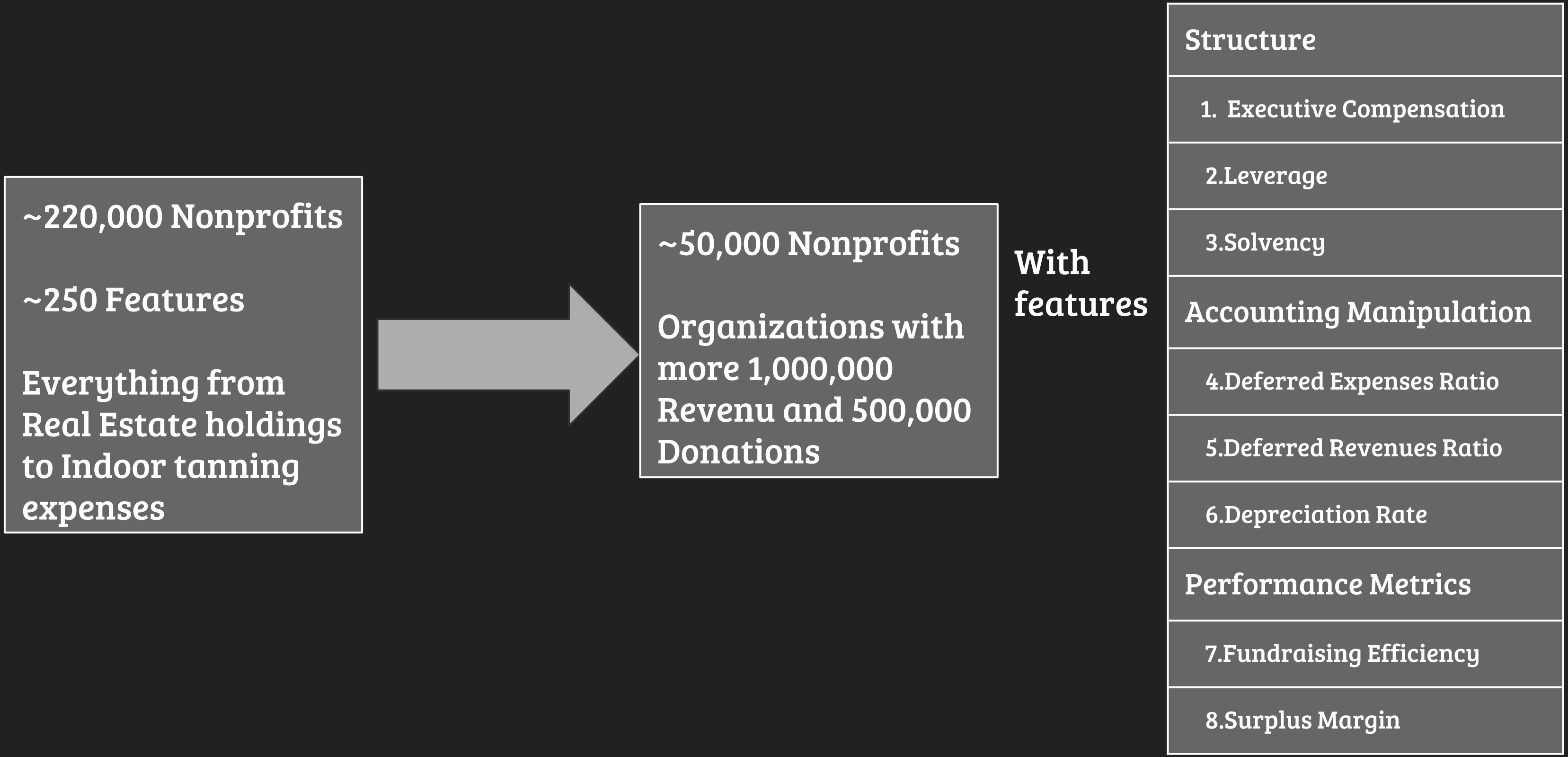
K Check here ▶ □ if the organization is not a 509(a)(3) supporting organization **and** its gross receipts are normally **not** more than $25,000. A return is not required, but if the organization chooses to file a return, be sure to file a complete return.

L Gross receipts: Add lines 6b, 8b, 9b, and 10b to line 12 ▶ $429,301

Part I Revenue, Expenses, and Changes in Net Assets or Fund Balances *(See the instructions.)*

Revenue	1	Contributions, gifts, grants, and similar amounts received:					
	a	Contributions to donor advised funds		1a	0		
	b	Direct public support (not included on line 1a)		1b	206,000		
	c	Indirect public support (not included on line 1a)		1c	0		
	d	Government contributions (grants) (not included on line 1a)		1d	0		
	e	**Total** (add lines 1a through 1d) (cash $ 201,000 noncash $ 5,000)				1e	206,000
	2	Program service revenue including government fees and contracts (from Part VII, line 93)				2	0
	3	Membership dues and assessments				3	0
	4	Interest on savings and temporary cash investments				4	9,702
	5	Dividends and interest from securities				5	0
	6a	Gross rents		6a	0		
	b	Less: rental expenses		6b	0		
	c	Net rental income or (loss). Subtract line 6b from line 6a				6c	0
	7	Other investment income (describe ▶)				7	0
			(A) Securities		**(B)** Other		
	8a	Gross amount from sales of assets other than inventory		8a	200,000		
	b	Less: cost or other basis and sales expenses.		8b	4,544		
	c	Gain or (loss) (attach schedule)		8c	195,456		
	d	Net gain or (loss). Combine line 8c, columns (A) and (B)				8d	195,456
	9	Special events and activities (attach schedule). If any amount is from **gaming,** check here ▶ □					
	a	Gross revenue (not including $ ____ of contributions reported on line 1b)		9a	0		
	b	Less: direct expenses other than fundraising expenses		9b	0		
	c	Net income or (loss) from special events. Subtract line 9b from line 9a				9c	0
	10a	Gross sales of inventory, less returns and allowances		10a	13,599		
	b	Less: cost of goods sold		10b	8,136		
	c	Gross profit or (loss) from sales of inventory (attach schedule). Subtract line 10b from line 10a				10c	5,463
	11	Other revenue (from Part VII, line 103)				11	0
	12	**Total revenue.** Add lines 1e, 2, 3, 4, 5, 6c, 7, 8d, 9c, 10c, and 11				12	416,621
Expenses	13	Program services (from line 44, column (B))				13	25,755
	14	Management and general (from line 44, column (C))				14	9,979
	15	Fundraising (from line 44, column (D))				15	0
	16	Payments to affiliates (attach schedule)				16	0
	17	**Total expenses.** Add lines 16 and 44, column (A)				17	35,734
Net Assets	18	Excess or (deficit) for the year. Subtract line 17 from line 12				18	380,887
	19	Net assets or fund balances at beginning of year (from line 73, column (A))				19	23,571
	20	Other changes in net assets or fund balances (attach explanation)				20	0
	21	Net assets or fund balances at end of year. Combine lines 18, 19, and 20				21	404,458

Featurize and Clean



~220,000 Nonprofits

~250 Features

Everything from Real Estate holdings to Indoor tanning expenses

~50,000 Nonprofits

Organizations with more 1,000,000 Revenu and 500,000 Donations

With features

Structure

1. Executive Compensation
2. Leverage
3. Solvency

Accounting Manipulation

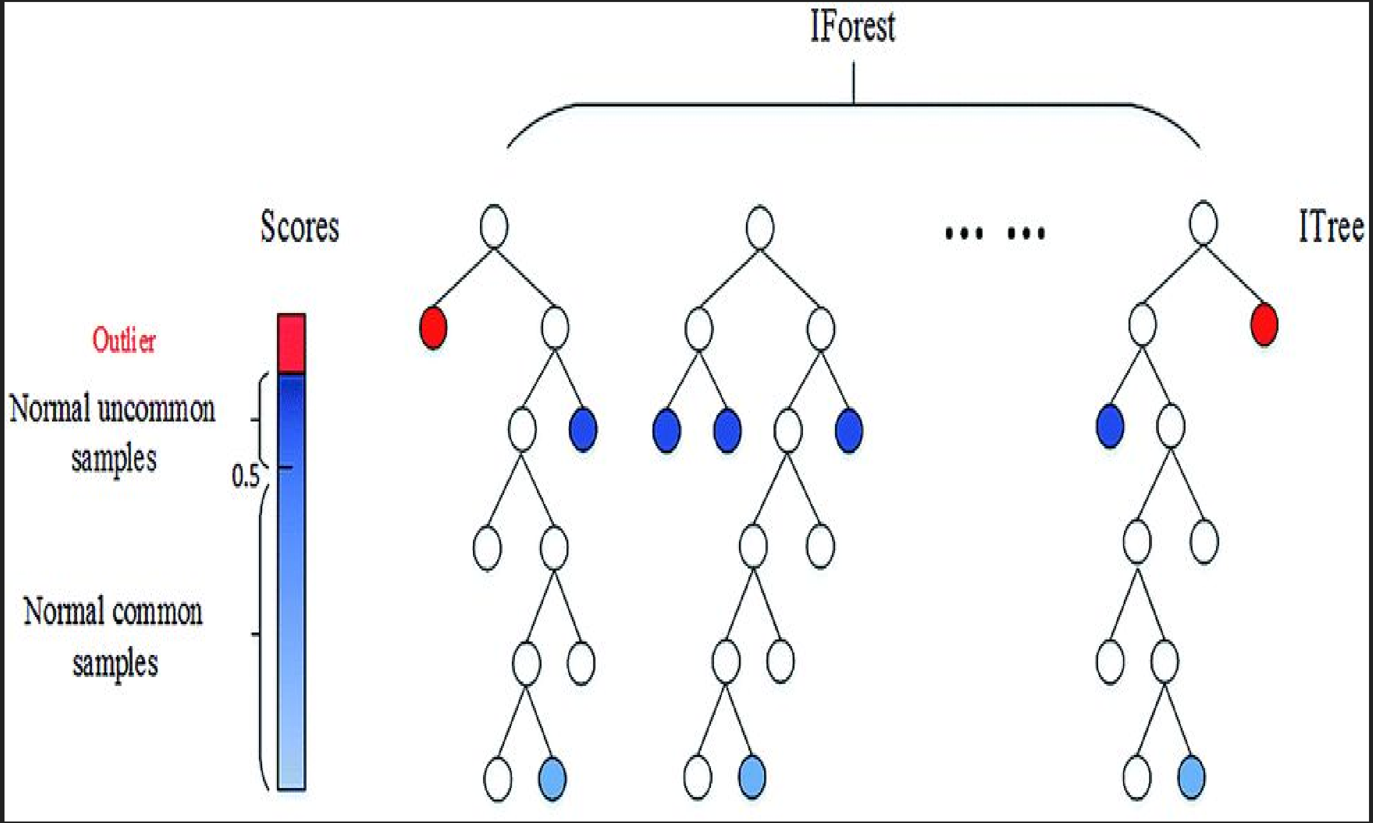
4. Deferred Expenses Ratio
5. Deferred Revenues Ratio
6. Depreciation Rate

Performance Metrics

7. Fundraising Efficiency
8. Surplus Margin

Isolation Forest Model

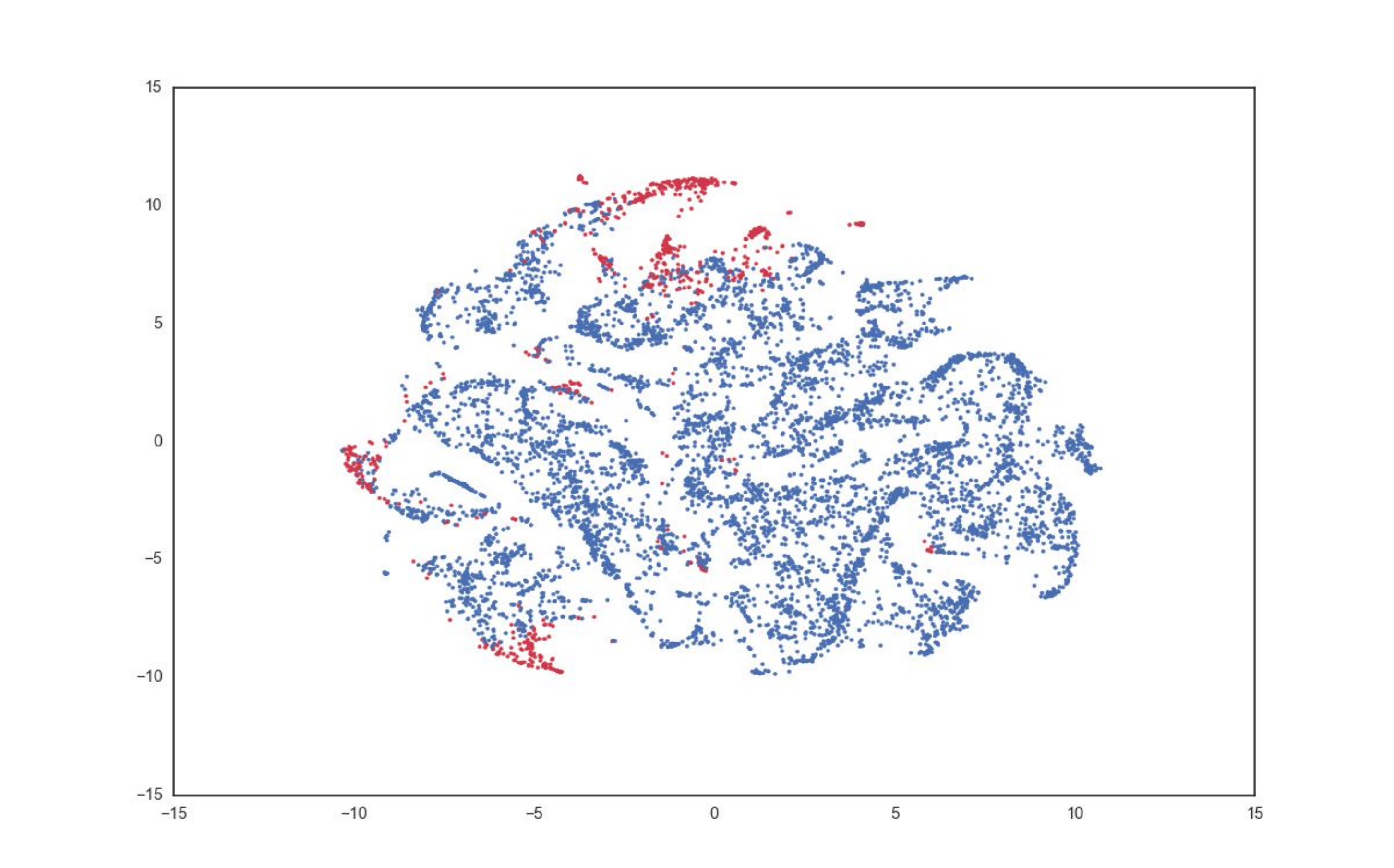
- Good way to identify outliers in Multidimensional data sets



How it works

- Randomly select a feature and Randomly select a split between the min and max of the feature
- Split until observation is isolated
- Closer the terminal node is root node the more abnormal the observation is
- Perform on a number of trees and average the distances to get an abnormality score

TSNE Display showing the different clusters of outlier organizations



What are we detecting?

Structure

1. Executive Compensation

2.Leverage

3.Solvency

Accounting Manipulation

4.Deferred Expenses Ratio

5.Deferred Revenues Ratio

6.Depreciation Rate

Performance Metrics

7.Fundraising Efficiency

8.Surplus Margin

Seems we are identifying non-profits that share similar characteristics and can be classified into distinct clusters

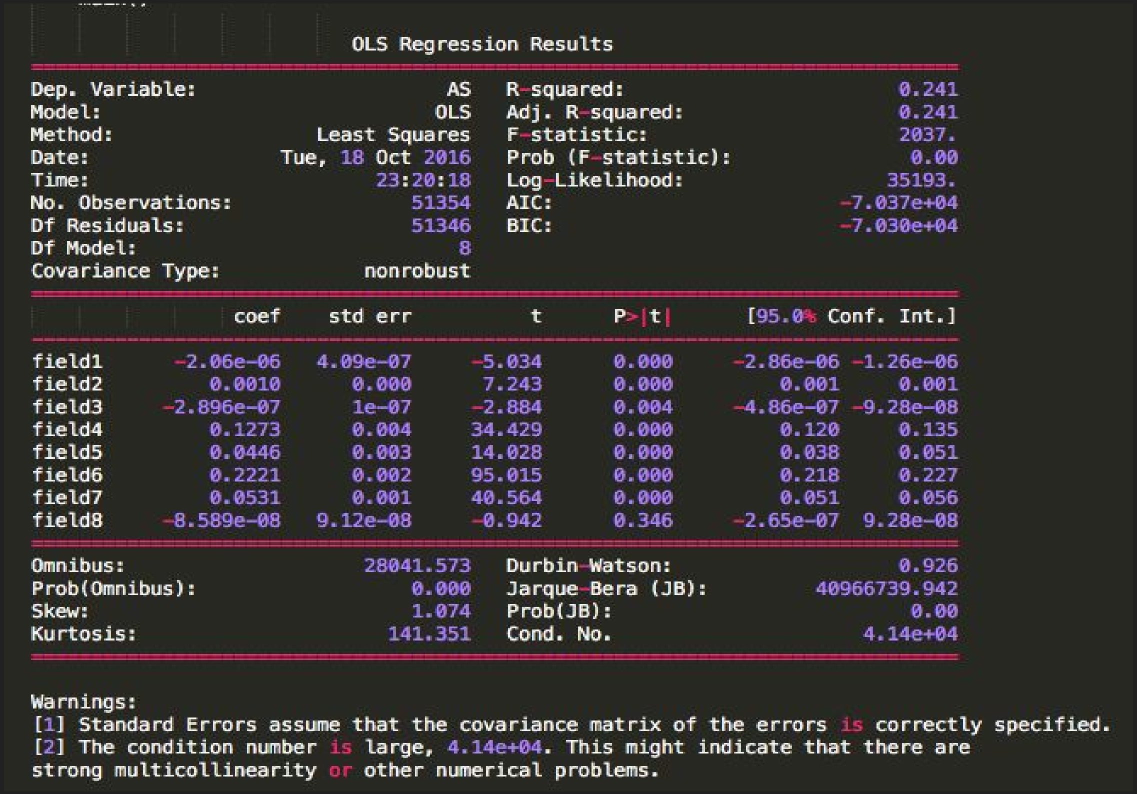
- Low (Assets, Liabilities), High (Revenue and Expenses) cluster - identified by structure metrics
- The Bad Cluster - identified by high judgment metrics

Thank You

Github: https://github.com/rmahdavi
Email:Ramahdavi@gmail.com

Appendix:

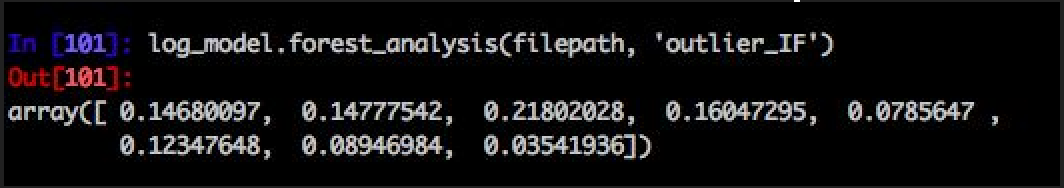
OLS Regression on Abnormality Score



```
                            OLS Regression Results
==============================================================================
Dep. Variable:                     AS   R-squared:                       0.241
Model:                            OLS   Adj. R-squared:                  0.241
Method:                 Least Squares   F-statistic:                     2037.
Date:                Tue, 18 Oct 2016   Prob (F-statistic):               0.00
Time:                        23:20:18   Log-Likelihood:                 35193.
No. Observations:               51354   AIC:                        -7.037e+04
Df Residuals:                   51346   BIC:                        -7.030e+04
Df Model:                           8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
field1      -2.06e-06   4.09e-07     -5.034      0.000     -2.86e-06 -1.26e-06
field2         0.0010      0.000      7.243      0.000         0.001     0.001
field3     -2.896e-07      1e-07     -2.884      0.004     -4.86e-07 -9.28e-08
field4         0.1273      0.004     34.429      0.000         0.120     0.135
field5         0.0446      0.003     14.028      0.000         0.038     0.051
field6         0.2221      0.002     95.015      0.000         0.218     0.227
field7         0.0531      0.001     40.564      0.000         0.051     0.056
field8     -8.589e-08   9.12e-08     -0.942      0.346     -2.65e-07  9.28e-08
==============================================================================
Omnibus:                    28041.573   Durbin-Watson:                   0.926
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         40966739.942
Skew:                           1.074   Prob(JB):                         0.00
Kurtosis:                     141.351   Cond. No.                     4.14e+04
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4.14e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Appendix:
Random Forest Feature Importance



```
In [101]: log_model.forest_analysis(filepath, 'outlier_IF')
Out[101]:
array([ 0.14680097,  0.14777542,  0.21802028,  0.16047295,  0.0785647 ,
        0.12347648,  0.08946984,  0.03541936])
```