Corporate Goverance and Company Performance

M&M - Classification

Created a new dataset, as a copy of the orignal but subsetting to include only variables listed in Appendix A of the orginal paper. Run classification results again and compare to non-subsetted dataset.

M&M - Regression

Taking same data, but not thresholding on dependent variables leaving it as a regression problem. Imputing (multiple, m=5) the data using the mice package. Lasso wont work on data with missing values. Running with just complete cases is an option, but could introduce bias? Also, zero rows in eebp are complete, very small number in sxxp. Tried for complete cases in spx (56 rows).

All results in table 3.

Included in those results is a regression on the Beneish MScore, which is a measure of how likely it is that the reported earnings of the company have been manipulated. Results are poor in terms of R^2. There is room for more work here (other dependent variables, and include in causal analysis as described below.)

```
mysql> mysql> select * from altman_z_results;
+---------------------+----------+---------+----------+-------------------------------+------------------+--------------------+--------------------+--------------------+--------------------+
| DateStamp           | Algorithm | DataSet | Target   | Correctly Classified Instances | Coverage Of cases | Precision Class 0  | Precision Class 1  | Precision Class 2  | ROC area           |
+---------------------+----------+---------+----------+-------------------------------+------------------+--------------------+--------------------+--------------------+--------------------+
| 2018-07-09 18:08:23 | Adaboost | spx     | Altman.Z | 0.827338129496403             | NULL | 0.714285714285714  | 0.631578947368421  | 0.888888888888889  | 0.85568667987481   |
| 2018-07-09 18:08:23 | Adaboost | sxxp    | Altman.Z | 0.777070063694268             | NULL | 0.866666666666666  | 0.657894736842105  | 0.807692307692308  | 0.818698360570282  |
| 2018-07-09 18:08:23 | Adaboost | eebp    | Altman.Z | 0.692307692307692             | NULL | 0.333333333333333  | 0.515151515151515  | 0.857142857142857  | 0.776              |
| 2018-07-09 18:08:24 | J48      | spx     | Altman.Z | 0.676258992805755             | NULL | 0.551724137931035  | 0.260869565217391  | 0.827586206896552  | 0.731647940074906  |
| 2018-07-09 18:08:24 | J48      | sxxp    | Altman.Z | 0.726114649681529             | NULL | 0.541666666666667  | 0.558823529411765  | 0.828282828282828  | 0.835045513088276  |
| 2018-07-09 18:08:24 | J48      | eebp    | Altman.Z | 0.6794871794871718            | NULL |              NULL  | 0.666666666666667  | 0.681818181818182  | 0.69047619047619   |
+---------------------+----------+---------+----------+-------------------------------+------------------+--------------------+--------------------+--------------------+--------------------+
6 rows in set (0.01 sec)
```

Figure 1: PSM testing - Asset

Causality

akelleh

Methodology - Causal Estimation

M&M make 8 statements about the effects of various corporate governance features on either Tobins Q or the Altman Z Score. This is the basis for my work on causal estimation, following this guide - https://github.com/akelleh/causality/tree/master/causality/estimation. The goal is to pick a treatment and outcome, and measure the magnitude of the effect of the former on the latter. Uses propensity matching.

I've built a MySQL table per statement to test. The manipulations I carry out to prepare each are:

- Impute the data to remove missing values. Taking just complete cases is infeasable especially for sxxp and eebp data-sets, since there are so few cases without missing data.

- Scale all columns apart from the treatment and target. Speeds up algorithm run-time. Leaving treatment and target as-is so that the resulting estimated causal effect is in the same units.

Called using something like;

```
ATE_results = matcher.estimate_ATE(
  data,
  treatment,
  target,
  {'P.B': 'c', 'Asset':'c', 'Tax':'c', 'P.E':'c'}, #to control for
  bootstrap=True
)
```

The main issue is how to pick out variables to control for. When I control for all in the dataset, I get an error about perfect separation in the matching stage.

Results for each of M&M's statements are in table 4.

My analysis also includes some plotting to show how good the matching process was, in terms of overlap in 1-D between the test and control groups. They need to overlap on the x axis. I think this is a good way to show whether propensity score matching is valid using those variables, but I can't just try every combination to see how the matching performs (I don't think?). Diagrams such as <over the page>. I need to find a better way to pick these variables, presumably by picking out ones that are marked as important in the literature? Could also use those marked as important in the classification / regression phase?

Table 1: MM Classification Results - Tobins Q as target

Algorithm	DataSet	Target	Correctly.Classified.Instances	Coverage.Of.cases	Precision.Class.0	Precision.Class.1	ROC.area
Adaboost	spx	Tobins.Q.class	0.7891566	NA	0.7710843	0.8072289	0.7891566
Adaboost	sxxp	Tobins.Q.class	0.8442211	NA	0.8500000	0.8383838	0.8441919
Adaboost	eebp	Tobins.Q.class	0.6363636	NA	0.7200000	0.5510204	0.6355102
J48	spx	Tobins.Q.class	0.7891566	NA	0.7831325	0.7951807	0.7891566
J48	sxxp	Tobins.Q.class	0.7889447	NA	0.8673469	0.7128713	0.7901091
J48	eebp	Tobins.Q.class	0.6767677	NA	0.8148148	0.5111111	0.6629630

Table 2: MM Classification Results - Altman Z as target

Algorithm	DataSet	Target	Correctly.Classified.Instances	Coverage.Of.cases	Precision.Class.0	Precision.Class.1	Precision.Class.2	ROC.area
Adaboost	spx	Altman.Z	0.7913669	NA	0.7647059	0.5454545	0.8500000	0.8217260
Adaboost	sxxp	Altman.Z	0.7133758	NA	0.8181818	0.4772727	0.8021978	0.8451599
Adaboost	eebp	Altman.Z	0.7051282	NA	0.5000000	0.5769231	0.8043478	0.7709630
J48	spx	Altman.Z	0.7338129	NA	0.7333333	0.1363636	0.8627451	0.7031046
J48	sxxp	Altman.Z	0.6624204	NA	0.6666667	0.4000000	0.7400000	0.7923221
J48	eebp	Altman.Z	0.5897436	NA	0.2000000	0.4545455	0.6862745	0.7004762

Table 3: Reg. Regression using Lasso Results

dataset	target	r2
spx - complete cases	tobins.q	0.9858568
spx - complete cases	asz	0.9822047
spx	tobins.q	0.7370037
spx	asz	0.5131296
spx	MScore-EightVar	0.1644398
spx	MScore-FiveVar	0.1983800
sxxp	tobins.q	0.9378874
sxxp	asz	0.4952553
eebp	tobins.q	0.7275647
eebp	asz	0.5240328

Table 4: akelleh Estimation Results

dataset	treatment	target	results	mm
spx	Indep.Lead.Dir.Fincl..l	AZS	(-0.48126852399043463, -0.43435592738302015, -0.38972494514119399)	...but also the pres
spx	Feml.CEO.or.Equiv	Tobins.Q	(-0.10126279335439194, 0.012848126155105874, 0.12824127721356948)	This is my own
sxxp	Indep.Lead.Dir.Feml.CEO.or.Equiv	Tobins.Q	(-0.074067123529340029, -0.028348354835763347, 0.012416082088477042)	the presence of an
sxxp	X..Women.on.Bd	Tobins.Q	(-0.18509495129967421, -0.13146459101319122, -0.085012004750609257)	A large percentage
eebp	BOD.Age.Rng	Tobins.Q	(-0.17513949067428736, -0.06049739831209239, 0.044424480674329032)	we found that a sn
eebp	Indep.Chrprsn.Feml.CEO.or.Equiv	AZS.class.Binary	(0.37714899918755695, 0.4801957873096407, 0.58600681649399455)	...to be on the safe
eebp	Fincl.l.treatment	Tobins.Q	(-0.23141294665514928, -0.12380378577028565, -0.016707652349402049)	...and that a finance
spx	X..Women.on.Bd	Tobins.Q	(-0.083227561878907452, -0.037628411530611736, 0.0040058740830202881)	For the American