Corporate Goverance and Company Performance

Project Goals + Status

☑ An initial goal is to acquire the dataset, whether by contacting MM or using Bloomberg terminals available in UCD Quinn and Smurfit, and reproduce some of MM's results.

Done

☑ A central goal is to apply (regularized) regression, and classification-on-regression (i.e. thresholding on the real value predicted by regression) rather than straight classification, as the main analysis, and follow this with a careful discussion of the pros and cons, and the findings (e.g. correlation analysis, both linear and rank correlation).

Done

☑ A side goal is to consider alternative features and measures of performance, beyond Z and Q.

Tried Beneish M Score as a dependent variable. Various ESG variables from Bloomberg as independent variables

☐ A stretch goal is to investigate modern work in causality (Pearl, 2009, King et al., 2016) and attempt to apply it here. Any contribution to the task of proving causality, as opposed to correlation, would be a large contribution.

Implementing akelleh, and causalTree. Work needed on specific implementation details.

M&M - Classification

Taking the data, algorithms and hypothesis put forward by MM and replicate results. Data used as is or with manipulations made by M&M. Results as per Table 1 and 2.

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.)

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?
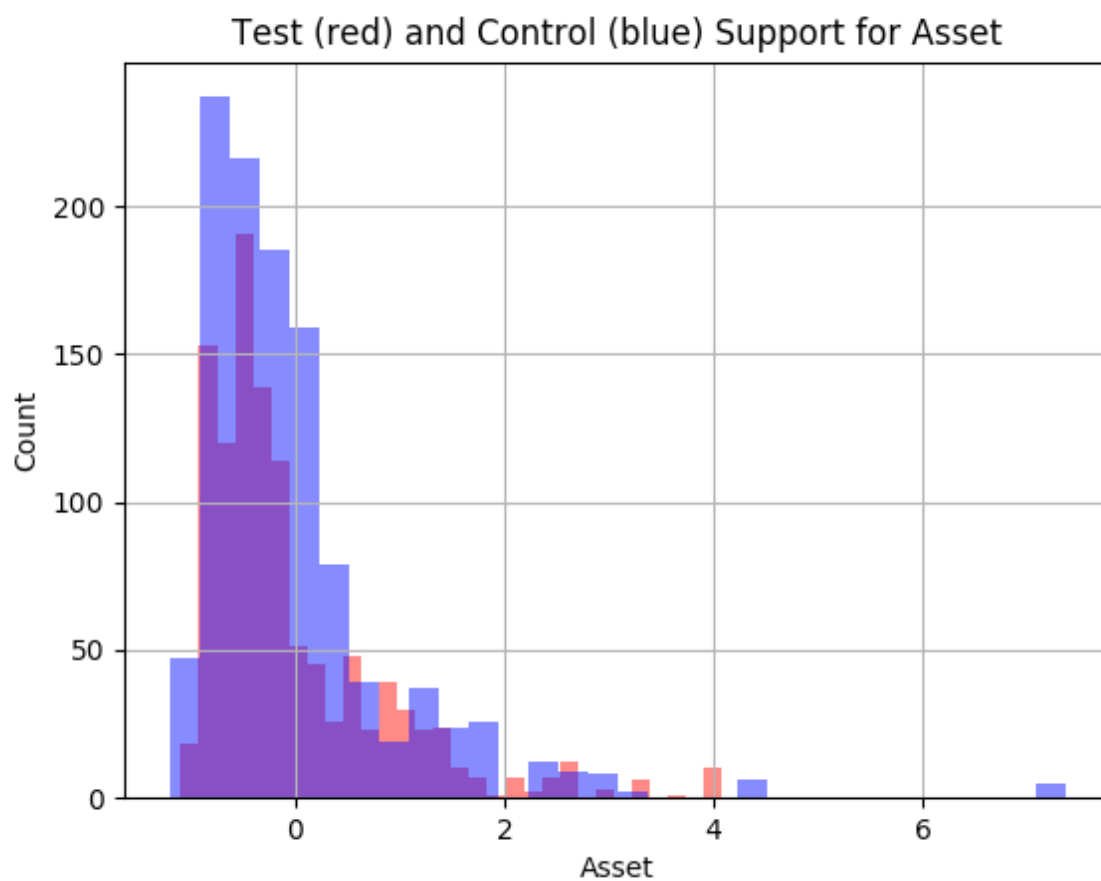
Figure 1: PSM testing - Asset

Methodology - Causal DAG

Efforts at constructing a causal DAG as per https://github.com/akelleh/causality, with arrows inferring causal influence. Under the hood, the algorithm is Pearls IC.

Requires the selection of variables to feed into the algorithm. I haven't figured out a smarter way of doing this, other than picking important variables from the regression analysis plus one of [Tobins Q / Altman Z]. Results looks like

[('Tax', 'P.E', 'arrows': ['Tax', 'P.E', 'P.E'], 'marked': False), ('Tax', 'Asset', 'arrows': ['Tax', 'Asset'], 'marked': False), ('P.E', 'P.B', 'arrows': ['P.E', 'P.E'], 'marked': False), ('P.E', 'P.EBITDA', 'arrows': ['P.E', 'P.E'], 'marked': False), ('Asset', 'Tobins.Q', 'arrows': ['Asset', 'Tobins.Q'], 'marked': False), ('P.EBITDA', 'Tobins.Q', 'arrows': ['Tobins.Q', 'Tobins.Q'], 'marked': False), ('Tobins.Q', 'P.B', 'arrows': ['Tobins.Q'], 'marked': False)]

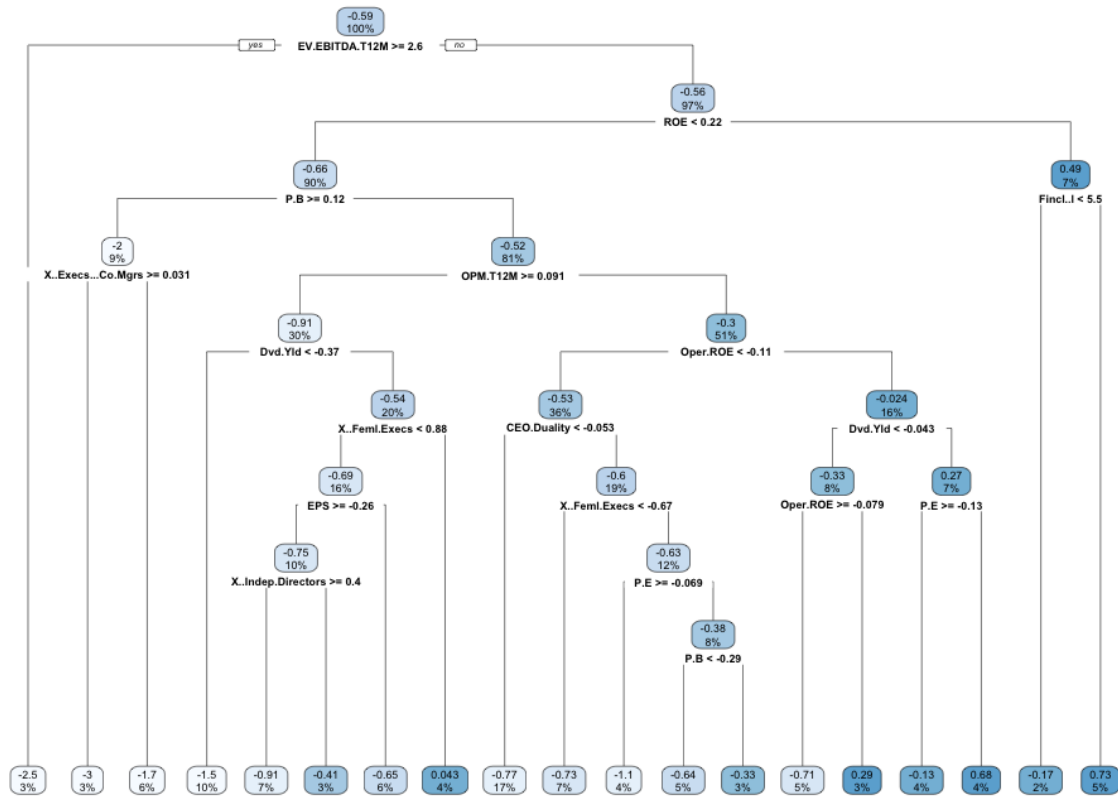I have yet to find a edge that meets Pearls criteria for causal inference.

Figure 2: CausalTree - spx / AZS / indepdirfincl

causalTree

Efforts at implementing https://github.com/susanathey/causalTree.

Getting trees, which spx / AZS with the presense of an indep lead director and a financial leverage < 2.5)

I'm having the same issue with this, in that I don't know what to include in the tree building process. I also still need to go over the validation of these trees. For example,

Node number 878: 85 observations
causal effect=-0.6436914, error=13396.07
Node number 879: 60 observations
causal effect=-0.3289201, error=9465.907

Those errors seem very high, but I don't understand what unit's they're in.

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	0.8855422	NA	0.8674699	0.9036145	0.8855422
Adaboost	sxxp	Tobins.Q.class	0.8844221	NA	0.8300000	0.9393939	0.8846970
Adaboost	eebp	Tobins.Q.class	0.7878788	NA	0.7800000	0.7959184	0.7879592
J48	spx	Tobins.Q.class	0.8192771	NA	0.8607595	0.7816092	0.8211843
J48	sxxp	Tobins.Q.class	0.8793970	NA	0.8367347	0.9207921	0.8787634
J48	eebp	Tobins.Q.class	0.6767677	NA	0.8085106	0.5576923	0.6831015

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	0.8273381	NA	0.7142857	0.6315789	0.8888889	0.8556867
Adaboost	sxxp	Altman.Z	0.7770701	NA	0.8666667	0.6578947	0.8076923	0.8186984
Adaboost	eebp	Altman.Z	0.6923077	NA	0.3333333	0.5151515	0.8571429	0.7760000
J48	spx	Altman.Z	0.6834532	NA	0.3750000	0.3333333	0.7961165	0.6622580
J48	sxxp	Altman.Z	0.6878981	NA	0.5500000	0.4594595	0.8000000	0.7829983
J48	eebp	Altman.Z	0.6282051	NA	0.0000000	0.3333333	0.7413793	0.6992225

∞

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
spx	Indep.Lead.Dir.Fincl..l	AZS	(-0.48126852399043463, -0.43435592738302015, -0.38972494514119399)
spx	Feml.CEO.or.Equiv	Tobins.Q	(-0.10126279335439194, 0.012848126155105874, 0.12824127721356948)
spx	X..Women.on.Bd	Tobins.Q	(0.11442235071138444, 0.1601459138176336, 0.20621396523499541)
sxxp	Indep.Lead.Dir.Feml.CEO.or.Equiv	Tobins.Q	(-0.074067123529340029, -0.028348354835763347, 0.012416082088477042)
sxxp	X..Women.on.Bd	Tobins.Q	(-0.18509495129967421, -0.13146459101319122, -0.085012004750609257)
eebp	BOD.Age.Rng	Tobins.Q	(-0.17513949067428736, -0.06049739831209239, 0.044424480674329032)
eebp	Fincl.l.treatment	Tobins.Q	(-0.13761563029896531, -0.027686073658574183, 0.074156162497339778)
eebp	Indep.Chrprsn.Feml.CEO.or.Equiv	AZS.class.Binary	(0.3880723356677665, 0.48173832421819107, 0.58578958829450489)