Master's Thesis

Testing fraud-detection tests

Validation of the Z-score, the M-score and the Sloan ratio, and development and validation of a New Model

Albert Vall Garcia

Master's Degree in Accounting and Financial Management

UPF Barcelona School of Management

Academic Year 2020 – 2021

Professor: Oriol Amat i Salas

June 17th, 2021

Abstract

Truthfulness in accounting is fundamental to ensure effective decision-making and avoid negative spillovers on society. This thesis validates, with respect to calibration and discrimination, the Z-score, the M-score and the S-ratio as accounting fraud detection models, using the Hosmer – Lemeshow goodness-of-fit test and a receiver-operation-curve respectively, and finds that these are unable to properly discriminate between companies manipulating their accounts and companies not doing so. I have developed and validated a New Model that discriminates better than existing ones using changes in Accounts Payable, Current Assets and Leverage. These tackle the different red-flags associated with accounting fraud: working capital changes, deteriorating margins and sudden changes in leverage.

Keywords

Accounting Fraud · Earnings Management · Earnings Manipulation

Accounts Manipulation · Financial Reporting · Fraud · Validation

Assignment carried out as a part of the program Master's Degree in Accounting and Financial Management given by UPF Barcelona School of Management, a center affiliated to Universitat Pompeu Fabra

Acknowledgements

Firstly, I would like to thank Oriol Amat for having agreed to tutor me and for having facilitated the database and literature needed to prepare this Master's Thesis.

Secondly, I would like to thank Judith Garcia for her insights in statistics, teaching me basing STATA programming and helping me approach the statistical analysis in the best possible way.

Thirdly, I would like to thank Raffaele Manini for his comments, guiding and helping me polish the drafting of this thesis.

Fourthly, I would like to thank Mariona Gastó for her help with the English revision.

1. Introduction

This thesis focuses on one of the most pressing issues in business: the reliability of the financial statements as effective decision-making documents. While accounting remains highly subjective in minor aspects, especially in IFRS jurisdictions, being able to detect major accounts manipulations can prevent stakeholders from incurring large losses.

To identify accounting fraud, some authors, such as Beneish or Amat, have focused on providing scores that help external parties detect, with relatively few amounts of easily obtainable data, if the accounts are at risk of having been manipulated. However, no comprehensive revisions of these models have been carried out. In fact, most papers only focus on the *ex-post* usage of these models to test *ex-ante* manipulation detection. Few papers point at how well these models identify manipulators in the general population.

Therefore, this thesis aims to bring attention to the need to validate fraud-detection formulas to ensure their effectiveness. Additionally, it provides a New Model to detect accounting fraud that discriminates better than existing ones. Furthermore, it raises the important question of recognizing the different costs associated with false positives and false negatives.

In first place, this thesis tests three major models: the Z-score, the M-score and the Sloan ratio, and finds that, even if these include all the relevant variables to detect fraud (the Hosmer – Lemeshow test is non-significant), these are not effective binary classifiers when it comes to determining if a company is manipulating its accounts or not [the area under the ROC curve (AUC) is close to 0.5].

In second place, this thesis generates a New Model including changes in Accounts Payable and Current Assets and the Leverage Index (LVGI) as main variables for predicting earnings management. These variables gather information about the main determinants of accounting fraud: working capital variations, highly levered capital structures and weakening economic results. The New Model performs better than existing ones on the validation cohort of the database.

The main limitations arise with respect to (i) the small sample size of the database, (ii) its external validity (since companies have not been randomly selected) and (iii) construction errors (given that information indirectly obtained is prone to mistakes).

2. Objectives

The objective of this thesis is to validate, with respect to calibration and discrimination, the Sloan ratio, the M-score and the Z-score as formulas to detect fraud in publicly listed companies in the Spanish stock market.

As a secondary objective, this thesis aims to propose and validate a New Model to detect accounting fraud.

3. Theoretical background

"Earnings management occurs when managers use judgment in financial reporting […] to either mislead some stakeholders about the underlying economic performance of the company, or to influence contractual outcomes that depend on reported accounting numbers" (Wahlen & Healy, 1999). However, this definition falls short in offering an applicable framework, forcing us to narrow down accounts manipulation "as an instance where management violates Generally Accepted Accounting Principles in order to beneficially represent the firm's financial performance" (Beneish, 1999).

As Dye (1988) and Jensen & Meckling (1976) suggest, manipulative behavior is fostered by an external demand to price the company and by an internal demand to maximize the manager's payoff conditioned by the principal-agent relationship. Therefore, the obligation to publish information and the strong unaligned incentives between stockholders and managers (e.g. pay linked to short term financial indicators) push them to "make it appear that the firm is doing better than its true performance" (Akerlof & Shiller, 2009).

Accounts manipulation, that is earnings management, is extremely wide in scope and generally entails (i) favoring from multiple legal treatments available for a single transaction, (ii) relying on highly subjective accounting criteria, (iii) optimizing the timing of revenues and expenses, (iv) drafting transactions to mislead stakeholders and (v) "using investment and production decisions to manage earnings" (Vladu, Amat, & Cuzdriorean, 2017). The problem arises when the accounts are altered within the confines of legality given that they are useless for decision-making purposes and misled stakeholders and related parties but no legal actions against managers can be taken (Beasley, 1996).

Authors such as Ueno, Amat, Johnson and Huang *inter alia* consider accounts manipulation *per se* as unethical and even intolerable (Loomis and Grant). Others, such as Arya or Parfet, however, advocate that these can be, to a certain extent, positive for the firm, especially when it comes to profit smoothing (Arya, Glover, & Delgado, 2003) or the impairment of failed investments to avoid market overreaction.

The prevalence of fraud among companies is extremely hard to measure given its pervasive nature and its many nuances. However, authors focus on the legal processes against manipulators initiated by the competent regulatory authorities to determine to what extent is accounting fraud widespread. In the case of Spanish listed companies, authors like Amat suggest that over 25% of Spanish listed companies manage their earnings, with up to 20% of reported earnings having a creative origin (Amat, Gowthorpe, & Perramon, 2003).

Additionally, given that the most important monitors for fraud are the media, industry regulators and employees, and that only 6% and 14% of fraud is detected by the SEC and auditors respectively (Dyck, Morse, & Zingales, 2010), regulators need to find ways to protect whistleblowers (U.S. Securities and Exchange Comission, 2021) to promote truthfulness of accounting.

Notwithstanding, it is likely that accounting fraud, especially if the company does not undergo financial stress, goes undetected and can represent, as Beneish (1999) indicates, a significant problem for investors. To identify *ex-ante* companies that might be carrying out accounts' manipulation authors have focused on qualitative and, to a lesser extent, quantitative information. While qualitative indicators usually rely on the personal traits and environmental factors prompting managers to manipulate the accounts [starting with Cressey (1953) and (1986)], quantitative indicators look for traces of manipulation in the financial statements of the company. Some of the most relevant quantitative indicators are the M-score (Beneish, 1999), the Sloan ratio (Sloan, 1996) and the Z-score (Vladu, Amat, & Cuzdriorean, 2017) inspired, to a certain extent, by Altman (1968).

When it comes to the validation of the quantitative indicators, most authors have focused on the *ex-ante* detection of companies that have *ex-post* been identified as manipulators. This falls short on ensuring the quality of existing models given that scandals are usually characterized for being extreme cases with many qualitative and quantitative red flags.

Similarly, authors have focused on studying the discriminative power of certain financial indicators to separate manipulators from non-manipulators. This provides information about how a large value in a certain index may indicate that the company is manipulating its accounts but does not clearly state how likely the company is to be a manipulator nor accounts for the multivariate nature of fraud-detection models.

Based on the abovementioned, an exhaustive validation of existing fraud-detection models is needed. The validation should ensure that models include all relevant variables to detect fraud and effectively differentiate between manipulators and non-manipulators. Additionally, if none of the existing models discriminate appropriately a new model will be in order.

4. Methodology

4.1. Study design and company data

This Master's Thesis is based on a sample of 63 companies listed in the Spanish stock market, of which 35 have manipulated their accounts and 28 have not, for which I have information about their financial statements between years 2005 and 2012. The sample has been acquired from Amat (2017). Table 3 shows the main characteristics of the sample.

4.2. Variables

Companies are classified as manipulators or non-manipulators based on the criteria of the stock exchange authorities (Comisión Nacional del Mercado de Valores). For each manipulating company there is information exclusively for the period during which accounts were manipulated. For non-manipulating companies, the same applies in the opposite direction.

For every year there is information about the company's: property plant and equipment (PPE), inventories (INV), accounts receivable (AR), accounts payable (AP), current assets (CA), total assets (TA), current liabilities (CL), total liabilities (TL), equity (E), sales (S), cost of goods sold (COGS), discretionary expenses, operating income, extraordinary income, provisions, depreciation and amortization (D&A), capitalization, earnings before extraordinary income, net income (NI), and cash from operations (OCF). Each company is given an ID.

4.3. Data cleaning and generation

The data cleaning phase consists in understanding the different variables and preparing the dataset for the statistical analysis. Said understanding entails ensuring no duplicate observations, managing outliers and handling missing data. In other words, this first phase makes sure the statistical analysis is performed on the highest quality data possible.

Firstly, all the variables have been described using univariate statistical analysis, especially focusing on the distribution and extreme values.

Secondly, impossible values, that is, variables inconsistent with theoretical principles or disproportionate in light of the company to which they belonged, have been transformed to missing. Additionally, I checked that the following theoretical concepts held in the database:

$$total\ assets = total\ liabilities + equity \qquad total\ assets \geq current\ assets + PPE$$

$$total\ liabilities \geq current\ liabilites \qquad current\ assets \geq inventories + accounts\ receivable$$

Thirdly, the variables needed for the Sloan ratio (Sloan, 1996), the M-score (Beneish, 1999) and the Z-score (Vladu, Amat, & Cuzdriorean, 2017) were generated, as defined by the authors, specifically: the receivables index (also defined as the days sales in receivables index) (RI & DSRI), the inventories index (II), the depreciation index (proposed by Amat) (DI), the leverage index (LVGI), the gross margin index (GMI), the asset quality index (AQI), the depreciation index (proposed by Beneish) (DEPI), the sales growth index (SGI), the sales general and administrative expenses index (SGAI) and the total accruals to total assets (TATA).

For the sales general and administrative expenses index the cost of goods sold to sales was used, given that I did not have information on administrative expenses.

For the total accruals to total assets, due to not having the data defined by Beneish, I have tried (i) change in working capital (current assets and current liabilities) to total assets, (ii) change in operating working capital (inventories, accounts payable and accounts receivable) to total assets and (iii) setting TATA equal to 1. Option (iii) has been kept given that it yields the best results in the calibration and discrimination phase. However, (ii) is conceptually closer to Beneish's proposal.

$$RI = DSRI = \frac{AR\ t/S\ t}{AR\ t-1/S\ t-1}$$

$$AQI = \frac{1-[(CA\ t + PPE\ t)/TA\ t]}{1-[(CA\ t-1 + PPE\ t-1)/TA\ t-1]}$$

$$II = \frac{INV\ t/COGS\ t}{INV\ t-1/COGS\ t-1}$$

$$SGI = \frac{S\ t}{S\ t-1}$$

$$DI = \frac{D\&A\ t/PPE\ t}{D\&A\ t-1/PPE\ t-1}$$

$$DEPI = \frac{D\&A\ t-1/(D\&A\ t-1 + PPE\ t-1)}{D\&A\ t/(D\&A\ t + PPE\ t)}$$

$$LI = \frac{TL\ t/TA\ t}{TL\ t-1/TA\ t-1} \approx LVGI$$

$$GMI = \frac{(S\ t-1 - COGS\ t-1)/S\ t-1}{(S\ t - COGS\ t)/S\ t}$$

$$SGAI \approx \frac{COGS\ t/S\ t}{COGS\ t-1/S\ t-1}$$

$$TATA\ (i) \approx \frac{(CA\ t - CL\ t) - (CA\ t-1 - CL\ t-1)}{TA\ t}$$

$$TATA\ (ii) \approx \frac{(AR\ t + INV\ t - AP\ t) - (AR\ t-1 + INV\ t-1 - AP\ t-1)}{TA\ t}$$

To avoid extreme data points from inducing non-existent relations, outliers above p-95 in distributions highly skewed to the right have been eliminated [in variable TATA I have cut below p-1 and above p-99 in options (i) and (ii)]. Winsorization has been considered as literature suggests (Beneish, 1999), but I believe trimming is more appropriate given that extreme values are due to errors in the construction of the database or extreme events (such as the financial crisis) unrelated to accounts manipulation.

Finally, the fraud-detection formulas have been generated, as defined by the authors:

$$Sloan\ ratio = \frac{(NI - OCF)}{TA}$$

$$M - score = (-4.84) + 0.92 * DSRI + 0.528 * GMI + 0.404 * AQI + 0.892 * SGI + 0.115 * DEPI - 0.172 * SGAI + 4.679 * TATA - 0.327 * LVGI$$

$$Z - score = (-4.5) + 0.03 * RI + 0.15 * II - 0.17 * DI + 4.23 * LI$$

I assumed that missing data was completely random and used a complete case approach, that is, I only used cases for which there were no missing variables to carry out the statistical analysis.

4.4. Statistical analysis

To maximize statistical power (1-β) of the analysis, every accounting period was used as a unit of analysis. The database compares the financial information of each company at opening (t-1) and at closing (t) of the accounting period for all the years about which information is available, yielding 7 pairs per company.

To validate fraud detection formulas, tests were run to determine whether the formulas include all the relevant variables to explain accounting fraud (calibration) and how well these differentiate manipulators from non-manipulators (discrimination). Differentiating manipulators (positive) from non-manipulators (negative) entails correctly identifying manipulators (avoid false negatives) as well as non-manipulators (avoid false positives).

The existing models were calibrated using the Hosmer – Lemeshow test, which estimates the goodness-of-fit of logistic regression models using fraud as the outcome and each fraud-detection formula as the exposure continuous variable. If the test is significant, it means that the model could be improved to explain systematic variability that is not gathered by the existing variables and coefficients.

Each fraud-detection formula was discriminated with a receiver-operation-curve (ROC) to test the ability of the model to act as binary classifier. The larger the surface under the curve, the better the prediction of the model given that there is a higher True Positive Rate (sensitivity) and a lower False Positive Rate (1 – specificity).

Because existing models did not exhibit good calibration nor discrimination properties, a new fraud-detection formula was generated (New Model). To predict a dichotomic outcome variable (manipulator = 1 and non-manipulator = 0) with continuous exposure variables (the different variables and indexes) a logistic regression model was used. To this aim, the dataset was randomly split into two groups, one for generation and the other for validation.

A univariable logistic regression model was run in the generation cohort with fraud as the outcome and each of the indexes potentially related to manipulation, as described by existing literature (Vladu, Amat, & Cuzdriorean, 2017) & (Beneish, 1999), as exposure continuous variables. The change in all the variables provided for in the database (section 4.2) was also tested. Trimming was applied to the latter at p-95 in distributions highly skewed to the right and at p-1 – p-99 in variables normally distributed.

Then, with the variables associated (p-value < 0.10) with the outcome from the previous models a multivariable logistic regression model was built. The exposure variables included were tested for correlation, using a spearman test, to avoid incurring multicollinearity (exposure variables that are linearly related). Variables with a correlation coefficient >0.7 were eliminated.

With the remaining variables a multivariable logistic regression model was built with the least number of variables that still explained the data (Hosmer & Lemeshow, 1980) and had significant or marginally significant coefficients (p-value<0.15).

To interpret de model correctly, given that a logistic regression was used, the following transformation was applied so the outcome would be a variable between 0 (less risk of being a manipulator) and 1 (higher risk of being a manipulator).

$$expon = exp\left(constant + \sum_{i}^{n} coef\ i * variable\ i\right)$$

$$risk\ of\ being\ a\ manipulator = {}^{expon}/_{(1 + expon)}$$

Finally, the New Model was calibrated and discriminated with the validation subset using the approach previously defined.

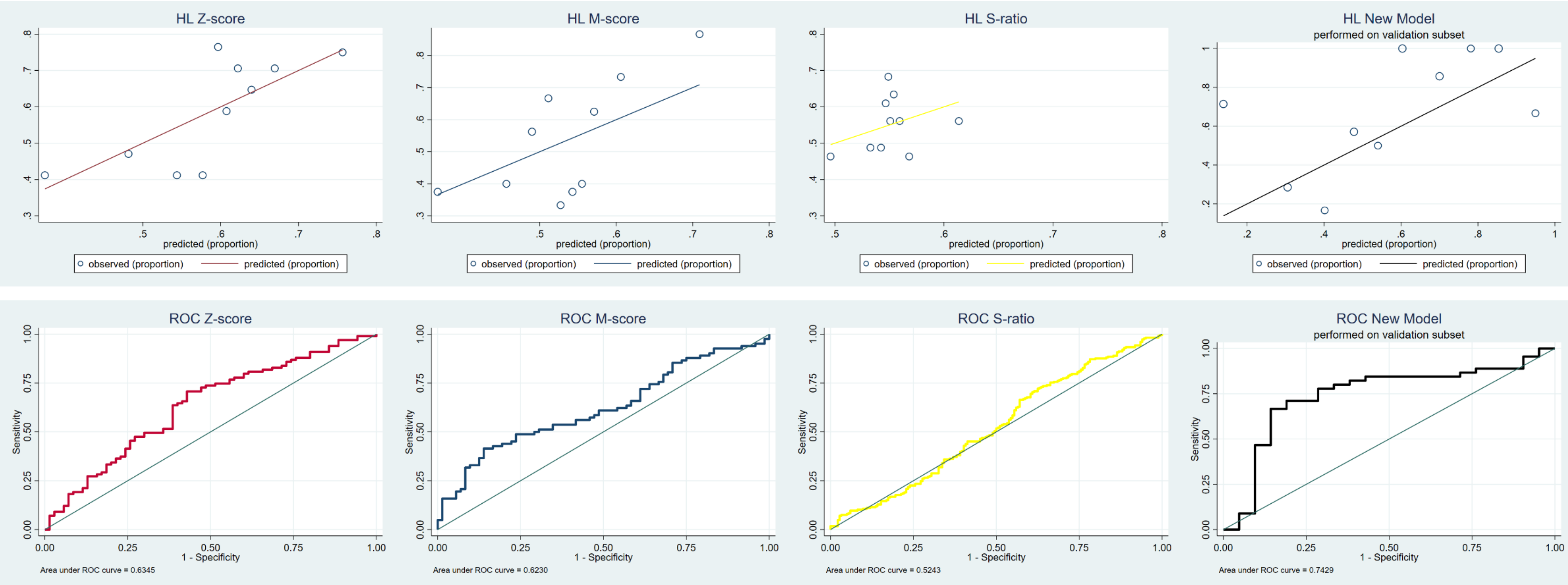

Table 1: Calibration and Discrimination of the different models.

5. Results

5.1. Validation of the Z-score, the M-score and the Sloan ratio

With regards to the Z-score, the Hosmer – Lemeshow calibration performed on 169 observations suggests that no extra systemic variability can be explained by adding additional variables (Prob>chi2=0.8283).

Notably, significant differences exist between the expected and realized observations for some HL groups (groups 3 to 5). This is consistent with the theoretical approach of establishing a grey area where manipulators and non-manipulators cannot be differentiated. For the highest and lowest groups, the model discriminates well (there are no significant differences between observed and expected values).

In the case without trimming the model worsens given that the 1st group expects 50% of manipulators and the 10th 74% whereas when trimming it is 37 and 76% respectively.

When it comes to the discrimination capacity of the model, the AUC is of 0.6345, suggesting that the model discriminates poorly between categories. However, this result should be interpreted in the light of the abovementioned gray area, meaning that the model identifies "extreme" cases correctly but fails to do so in "mid-table" ones.

With regards to the M-score, the HL test on 154 observations is not significant (Prob>chi2=0.3943).

As with the Z-score, significant differences arise in groups 4 to 7. This is coherent with the fact that there are ranges in which the model is unable to discriminate correctly. The test performed on option (ii) yielded a Prob>chi2=0.2869 and had a similar number of expected manipulators per group (40% in the 1st and 70% in the 10th).

In terms of discrimination, the model performs poorly with an area of 0.62 under the ROC curve. In relation to the definition of variable TATA in section 4.3, options (i) (0.51) and (ii) (0.58) offer smaller AUC.

With regards to the Sloan ratio, the HL test on 410 observations is not significant (Prob>chi2=0.6266). Notwithstanding, the model performs poorly given that the 1st group expects a 46% presence of manipulators and the 10th a 55%.

This fact is supported by an area of 0.52 under the ROC curve, implying that the prediction capacity of the model is close to random (AUC = 0.5).

5.2. Generation and validation of the proposed model

Of all the variables and indices only AP, TL, E, LVGI, CA and GMI, having trimmed outliers, are significantly related to being or not manipulator in the generation cohort. When performing the spearman correlation test, LVGI is highly correlated ($|corr| > 0.7$) with TL and E. This is consistent with the definition of LVGI. TL and E have been eliminated also given the limitation with regards to the consistency of TL and CL.

Performing a logistic regression with the remaining variables we find a marginally significant model (Prob > chi2 = 0.12) with a Pseudo R2 of 0.13 and non-significant coefficients. This model is thus disregarded.

The most parsimonious model, that is the model with the least number of variables that still explains the data, is the one including AP, LVGI and CA. It is significant, explains 16% of the data and has significant (AP and LVGI) and marginally significant (CA) coefficients.

$$expon = exp[-4.64 + (-0.015) * AP + 4.673 * LVGI + (-0.018) * CA]$$

$$risk = \frac{expon}{(1 + expon)}$$

The HL test on the New Model (55 observations) shows that systematic variability could be explained by adding other variables. However, adding additional variables would likely result in non-significant coefficients in the logistic regression given that the dataset is small. It can be presumed that the significance of the HL test is due to 1 or 2 companies that manipulate their accounts in a way that is undetectable by this model.

The area under the ROC curve is 0.74, implying that the model discriminates between better than existing models.

6. Discussion

The results obtained indicate that existing models include all variables relevant to determine whether a company is a manipulator or not but fail to correctly classify them. Bearing in mind the limitations, existing models perform poorly given that AUC < 0.7 (Lemeshow & Hosmer, 2013).

This statement can, however, be qualified taking Beneish's approach to sensibility and specificity. It is better, from the stakeholder perspective, that, considering that a significant number of substitutes are available on the market, fraud-detection formulas prioritize minimizing the false negative rate, given that when a party interacts with a company that manipulates its accounts, it has more at stake than when it forgoes dealing with a company that was incorrectly classified as a manipulator.

In contrast with the previous, the New Model offers acceptable discrimination properties AUC ∈ (0.7 – 0.8).

However, the HL test is significant, meaning extra systemic variability is not gathered by the model. Several facts might influence said outcome:

(i) The small sample size of the generation subset may yield non-significant coefficients for variables that are significant at a population level. In this situation, the variables are not included in the New Model but are recognized in the validation subset as relevant systemic variability.

(ii) Similarly, sample errors and sample size of the validation cohort may condition the HL test as is mentioned in section 5.2.

(iii) The period at hand includes the financial crisis, during which non-manipulating companies saw their profits drastically reduced without there being the incentive to "make it appear the firm is doing better".

I expect that with a better, larger database a formula including more variables could be found.

With regards to the variables significantly associated with accounting fraud at an individual level, having eliminated highly correlated variables, change in Current Assets and Accounts Payable, LVGI and GMI were obtained. These variables can be classified in three groups: working capital accounts (CA and AP), capital structure (LVGI) and operative result (GMI). Each of these categories have different underlying motivations related to accounting fraud (Tirole, 2006).

Working capital accounts are the most susceptible to creative accounting given the role subjectivity plays it their determination. Timing of operations, changes in accounting policies or sales to doubtful customers can be used to boost the accounting result and are thus indicators of fraud (Vladu, Amat, & Cuzdriorean, 2017).

Tackling change in Accounts Payable, we find that it is inversely associated with manipulation (Odds Ratio = 0.9860 and $P>|z|= 0.000$ for the whole database and Odds Ratio = 0.9919 and $P>|z|= 0.081$ for the generation subset). In existing literature, no authors have considered AP as a potential indicator of accounting fraud. The intuition behind said association can be that a decrease in AP is a sign that the company is not performing well (it is reducing its operating activities). When a company is not performing as expected, managers can be tempted to manipulate the accounts to show a better performance and keep their jobs. This would support why the GMI is not significant in the multivariate model given that both explain a similar effect.

Tackling change in Current Assets, it includes Inventories and Accounts Receivable, both widely known to be associated with accounting fraud. From fictious sales in AR and non-existent inventories to sales to poor clients and insufficient impairment, many techniques allow managers to offer a biased financial result. Against what could be expected, the change in CA is inversely associated with manipulation. Notwithstanding, this could be challenged given that the coefficient in the New Model is only marginally significant (p-value = 0.12). Note however that the inverse association is significant at the level of the whole database (Odds Ratio = 0.9945 and P>|z|= 0.031).

The change in CA can replace the difference between EBITDA and OCF in a model created with a small sample size because both work on a similar direction: changes in working capital can be used to distort the accounting result. We could expect to include INV and AR separately in a model generated with a larger sample size.

The capital structure is the most relevant indicator of fraud given the large incentive-linked implications it has (Tirole, 2006). Growing leverage is not only significantly associated (Odds Ratio = 5.6177 in the generation subset) with the deterioration of the firm's financial situation but also indicates unaligned interests between shareholders and debtholders as well as between managers and shareholders. In the first situation, a clear principal–agent conflict, shareholders are prone to taking riskier projects that increase their expected payoff given that they are the residual claimers of a smaller share of the firm's value and they have fewer capital at risk. This has implications in the second situation, yet another principal–agent conflict, where incentives for managers to provide financial results above what could be expected from a low-levered firm motivate managers to manipulate the accounts to meet their targets and boost their personal payoff (Akerlof & Shiller, 2009).

The operative result signals when the company is not performing well. Given that manipulations can yield a "desired" net income, indicators such as the GMI give information about how the company is improving its operations. A reduction of the GMI is thus inversely associated with manipulation on the grounds that worse prospects "force" mangers to manipulate the accounts (Odds Ratio = 0.4457 and P>|z|= 0.065 in the generation subset). Note however this variable is not included in the most parsimonious model because it yielded a non-significant coefficient and was, from a theoretical perspective, partially explained by other variables.

7. Conclusions

This thesis provides reasonable evidence that existing fraud-detection models fail to discriminate between companies manipulating their accounts and companies not doing so in the context of Spanish companies listed in the stock exchange in the 2005 – 2012 period. The proposed New Model discriminates better than existing ones in the validation cohort and should thus be employed by authorities in said context to approach the likelihood of accounting fraud.

The most relevant indicator of accounting fraud remains the leverage of the firm for the incentives it entails for managers. Indicators and variables related to working capital and the economic results of the firm remain relevant but do not have the same stark impact on the outcome of the models.

Models can be improved to consider the expected true prevalence of manipulators in society and account for the largest economic cost of a false negative than of a false positive. This would help market and tax authorities optimize their inspection efforts and protect investors from investing in companies manipulating their accounts.

Further research on this topic is needed to ensure existing formulas perform fairly good while requiring the least amount of information, that should be publicly available, as possible.

8. Limitations

The results of the thesis are constrained on the following grounds:

8.1. Sample size

Sample size has the largest effect on the generation and validation phase of the New Model. Regressors are more likely to be non-significant and calibration and discrimination can significantly vary depending on the randomly selected data for the subsets. Errors and outliers are more likely to influence the different analysis, inducing non-existent relations.

8.2. External validity

The prevalence of manipulators in the sample is larger than in the general population and the database has not been constructed randomly picking companies from the market. The results obtained might thus not be applicable to the general population given that neither the sample represents the population nor the prevalence of manipulators on the sample reflects the true prevalence of such on the overall population. In fact, if there are no changes in any of the accounts, the model predicts that there is a 50% chance that the company manipulates its accounts (which is consistent with a ≈50% prevalence of manipulators in the generation cohort but not with the numbers provided for in the literature).

8.3. Missing values

There are missing variables in many of the companies and years.

8.4. Construction errors

The database has a poor quality:

(i) The database has been constructed mechanically so it is prone to human mistakes.

(ii) There are 42 (out of 300) observations in which current liabilities are larger than total liabilities. These values have not been transformed to missing given that, based on prudence, all the data form those companies for those years should be eliminated and the impact on the sample size would be considerable.

(iii) As seen in the data cleaning phase, many variables have abnormal values (e.g. AR grow more than 10x fold or AR are around 10m.u. while CA for the same company and year are above 5.000m.u.). Given that this is frequent I assumed it was correct and have not set these variables as missing.

9. Annex

Model		D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	Prob > chi2
Z-score	obs	41	47	41	41	77	59	71	65	71	75	0.8283
	exp	37	48	54	58	60	61	62	64	67	76	
M-score	obs	36	40	56	67	33	38	40	63	73	87	0.3943
	exp	37	46	49	51	53	54	56	57	61	71	
S-ratio	obs	46	49	49	61	68	56	63	56	46	56	0.6266
	exp	50	53	54	55	55	55	55	56	57	61	
New Model	obs	71	29	17	57	50	100	85	100	100	67	0.0000
	exp	14	30	40	48	54	60	70	78	85	95	

Table 2: Hosmer-Lemeshow test on the different models. The New Model is tested on the validation subset.

Sector	Manipulators	Non-manipulators	Total
Basic Materials, Industry and Construction	7	10	17
Consumer Goods	7	5	12
Consumer Services	7	5	12
Financial Services and Real Estate	8	0	8
Petrol and Power	3	5	8
Technology and Telecommunications	3	3	6

Table 3: Sample description.

constant	AP	LVGI	CA	Prob > chi2	Pseudo R2
-4.640318 (0.066)	-0.0152856 (0.050)	4.673196 (0.057)	-0.018379 (0.120)	0.0078	0.1604

Table 4: Logistic regression for the New Model.

10. Bibliography

Akerlof, G., & Shiller, R. (2009). Animal spirits. *Pinceton University Press*.

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 589-609.

Amat, O., & Singh, N. (2020). *Detecting accounting fraud using quantitative techniques.* Department of Economics and Business, Universitat Pompeu Fabra: Economics Working Papers.

Amat, O., Gowthorpe, C., & Perramon, J. (2003). Earnings management in Spain: an assessment of the effect on reported earnings of larger listed companies 1999-2001. *Economic Working Paper Series*.

Arya, A., Glover, J., & Delgado, M. (2003). Are unmanaged earnings always better for shreholders? *The Accounting Horizon*, 111-116.

Beasley, M. (1996). An empirical analysis of the relation between the board of director composition and financial statement fraud. *The Accounting Review*, 443-465.

Beneish, M. (1999). The detection of earnings manipulation. *Financial Analysts Journal*, 24-36.

Cressey, D. (1953). *Other people's money: a study in the social psychology of embezzlement.* Glencoe: The Free Press.

Cressey, D. (1986). Why managers commit fraud. *Australian & New Zealand Journal of Criminology*, 195-209.

Dyck, A., Morse, A., & Zingales, L. (2010). Who blows the whistle on corporate fraud. *Journal of Finance*, 2213-2253.

Dye, R. (1988). Earnings management in an overlapping generations model. *Journal of Accounting Research*, 195-235.

Healy, P. (1985). The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics*, 85-107.

Hosmer, & Lemeshow. (1980). A Goodness-of-Fit Tests for the Multiple Logistic Regression Model. *Communications in Statistics*, 1043-1069.

Jensen, M., & Meckling, W. (1976). Theory of the firm: managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics*, 305-360.

Lemeshow, & Hosmer. (2013). *Applied logistic regression.*

Sloan, R. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings. *The Accounting Review*, 289-315.

Tirole, J. (2006). *Theory of Corporate Finance.* Princeton University Press.

U.S. Securities and Exchange Comission. (2021, February 9). *Office of the Whistleblower*. Retrieved from U.S. Securities and Exchange Comission: https://www.sec.gov/whistleblower/retaliation

Vladu, A., Amat, O., & Cuzdriorean, D. (2017). Truthfulness in accounting: how to discriminate accounting manipulators from non-manipulators. *Journal of Business Ethics*, 633-648.

Wahlen, M., & Healy, P. (1999). A review of the earnings management literature and its implications for standard setting. *Accounting Horizons*, 365-383.

11. Code used for the statistical analysis

```
*********************
* select variables   *
*********************

desc
summ

* I remove varibles that will not be used
drop lossgainonsaleofnoncuure
drop ar

*********************
* clean variables    *
*********************

* convert to missing impossible values
* check minumum and maximum
* make sure no missing in relevant variables

sort id t

tab id
tab man
tab man, mis
tab t

summ art1, d
*negative value don't make sense
list id if art1<0
list t art1 if id==39
replace art1=. if id==39 & t==1112
summ art1, d
* 0 is not a possible value
replace art1=. if art1==0
summ art1, d
hist art1, bin(200)
* some few very high values but likely
hist art1 if art1<50000, bin(200)
hist art1 if art1<10000, bin(200)
hist art1 if art1<1000, bin(200)
* under 1000 looks strange
list id if art1<100
list id t art1 if art1<200
list id t art1 if id==11
* I checked back to the excel and there is no mistake in
data conversion
* because strange values are frequent, I assume they are
correct
* NOTE: this distribution is highly skewed to the right
summ art, d
list id t if art<0
replace art=. if id==39 & t==1011
replace art=. if art==0
summ art, d

summ st1 st, d
replace st1=. if st1==0
replace st=. if st==0
summ st1 st, d
hist st1 , bin(200)
* NOTE: this distribution is highly skewed to the right
replace st1=. if st1==6
replace st=. if st==6

summ invt1 invt, d
replace invt1=. if invt1==0
replace invt=. if invt==0
summ invt1 invt, d
list id invt1 invt if (invt1<100) | (invt<100)
* in most cases small values are followed by . so OK
list id t invt1 invt if id==31
hist invt1
* NOTE: this distribution is skewed to the right
replace invt1=. if invt1==42503
replace invt=. if invt==42503

summ cogst1 cogst, d
replace cogst1=cogst1*(-1) if cogst1<0
replace cogst=cogst*(-1) if cogst<0
summ cogst1 cogst, d
replace cogst1=. if cogst1==0
replace cogst=. if cogst==0
summ cogst1 cogst, d
hist cogst1 , bin(200)
hist cogst , bin(200)
* NOTE: this distribution is skewed to the right
replace cogst1=. if cogst1==378439
replace cogst=. if cogst==378439

summ dat1 dat, d
replace dat1=dat1*(-1) if dat1<0
replace dat=dat*(-1) if dat<0
replace dat1=. if dat1==0
replace dat=. if dat==0
summ dat1 dat, d
hist dat1 , bin(200)
hist dat , bin(200)
* NOTE: this distribution is skewed to the right

summ ppet1 ppet, d
replace ppet1=. if ppet1==0
replace ppet=. if ppet==0
summ ppet1 ppet, d
hist ppet1 , bin(100)
hist ppet , bin(100)
* NOTE: this distribution is skewed to the right

summ det1 det , d
replace det1=det1*(-1) if det1<0
replace det=det*(-1) if det<0
replace det1=. if det1==0
replace det=. if det==0
summ det1 det , d
hist det1 , bin(100) norm
hist det , bin(100)
* NOTE: this distribution is skewed to the right

summ tlt1 tlt , d
replace tlt1=. if tlt1==0
replace tlt=. if tlt==0
summ tlt1 tlt , d
hist tlt1 , bin(100)
hist tlt , bin(100)
* NOTE: this distribution is skewed to the right

summ tat1 tat  , d
replace tat1=. if tat1==0
replace tat=. if tat==0
summ tat1 tat  , d
hist tat1 , bin(100)
hist tat , bin(100)
* NOTE: this distribution is skewed to the right
```

```
gen testtat1=tat1-tlt1-et1
summ testtat1, d
hist testtat1
gen testtat=tat-tlt-et
summ testtat, d
hist testtat
* this should be 0, therefore everything OK

summ cat1 cat  , d
replace cat1=. if cat1==0
replace cat=. if cat==0
summ cat1 cat  , d
hist cat1 , bin(100)
hist cat , bin(100)
* NOTE: this distribution is skewed to the right

gen testcat1=tat1-cat1
summ testcat1
gen testcat=tat-cat
summ testcat
* it should be positive, therefore OK

summ ebitdat1 ebitdat , d
replace ebitdat1=. if ebitdat1==0
replace ebitdat=. if ebitdat==0
summ ebitdat1 ebitdat , d
hist ebitdat1 , bin(100) norm
hist ebitdat , bin(100) norm
* NOTE: this distribution is normal but HIGH SD, slightly skewed to the right

summ nit1 nit  , d
replace nit1=. if nit1==0
replace nit=. if nit==0
summ nit1 nit  , d
hist nit1 , bin(100) norm
hist nit , bin(100) norm
* NOTE: this distribution is normal but HIGH SD, slightly skewed to the right

summ pt1 pt, d
replace pt1=. if pt1==0
replace pt=. if pt==0
summ pt1 pt, d
hist pt1 , bin(100) norm
hist pt , bin(100) norm
* NOTE: this distribution is skewed to the right

summ  ocft1 ocft, d
replace ocft1=. if ocft1==0
replace ocft=. if ocft==0
summ  ocft1 ocft, d
hist ocft1 , bin(100) norm
hist ocft , bin(100) norm
* NOTE: this distribution looks a bit normal around zero

summ  capt1 capt, d
replace capt1=. if capt1==0
replace capt=. if capt==0
summ  capt1 capt, d
hist capt1 , bin(100) norm
hist capt , bin(100) norm
* NOTE: this distribution is skewed to the right

summ apt1 apt, d
replace apt1=. if apt1==0
replace apt=. if apt==0
summ  apt1 apt, d
list id if apt==0.001
list id t apt1 apt if id==53
replace apt=. if apt==0.001
hist apt1 , bin(100) norm
hist apt , bin(100) norm
* NOTE: this distribution is skewed to the right

summ clt1 clt, d
replace clt1=. if clt1==0
replace clt=. if clt==0
summ clt1 clt, d
hist clt1 , bin(100) norm
hist clt , bin(100) norm
* NOTE: this distribution is skewed to the right

gen cltcheck=tlt-clt
summ cltcheck if cltcheck<0, d
list id t tlt clt cltcheck if cltcheck<0
* this is WRONG

gen tacheck=tat-tlt-et
summ tacheck if tacheck<0, d
*this is ok

gen tatcheck=tat-cat-ppet
summ tatcheck if tatcheck<0, d
*this is ok

gen cacheck=cat-invt-art
summ cacheck if cacheck<0, d
*this is ok

************************
* generate variables   *
************************

* (1a) generate the variables required to estimate z

gen ri=(art/st)/(art1/st1)
summ ri, d
egen ri_p95 = pctile(ri), p(95)

gen ii=(invt/cogst)/(invt1/cogst1)
summ ii, d
hist ii
egen ii_p95 = pctile(ii), p(95)


gen di=(dat/ppet)/(dat1/ppet1)
summ di, d
list id if di>50 & di!=.
egen di_p95 = pctile(di), p(95)


gen li=(tlt/tat)/(tlt1/tat1)
summ li, d
list id if li>50 & li!=.
egen li_p95 = pctile(li), p(95)


* (1b) generate z continuous and categorical

gen z=(-4.5) + 0.03*ri + 0.15*ii - 0.17*di + 4.23*li
summ z, d
gen z_p95=(-4.5) + 0.03*ri + 0.15*ii - 0.17*di + 4.23*li
if li<li_p95 & ri<ri_p95 & ii<ii_p95 &   di<di_p95
summ z_p95, d
```

```
* (2) generate the variables required to estimate M

gen dsri=(art/st)/(art1/st1)
summ dsri, d
egen dsri_p95 = pctile(dsri), p(95)

gen lvgi=(tlt/tat)/(tlt1/tat1)
summ lvgi, d
egen lvgi_p95 = pctile(lvgi), p(95)

gen gmi=((st1-cogst1)/st1)/((st-cogst)/st)
summ gmi, d
egen gmi_p95 = pctile(gmi), p(95)
egen gmi_p1 = pctile(gmi), p(1)
summ gmi if gmi>gmi_p1 & gmi<gmi_p95, d

gen aqi=(1-(cat+ppet)/tat)/(1-(cat1+ppet1)/tat1)
summ aqi, d
egen aqi_p95 = pctile(aqi), p(95)

gen sgi=st/st1
summ sgi, d
egen sgi_p95 = pctile(sgi), p(95)

gen depi=(dat1/(dat1+ppet1))/(dat/(dat+ppet))
summ depi, d
egen depi_p95 = pctile(depi), p(95)

gen sgai=(cogst/st)/(cogst1/st1)
summ sgai, d
egen sgai_p95 = pctile(sgai), p(95)

*Option (i)
gen tata=((cat-clt)-(cat1-clt1))/tat
summ tata, d
egen tata_p99 = pctile(tata), p(99)
egen tata_p1 = pctile(tata), p(1)
summ tata if tata>tata_p1 & tata<tata_p99, d

*Option (ii)
gen tata1=(art-art1+invt-invt1-apt+apt1-dat)/tat
summ tata1, d


* (2) generate M

gen m2=(-4.84)+0.92*dsri+0.528*gmi+0.404*aqi+0.892*sgi+0.115*depi-0.172*sgai+4.679*tata1-0.327*lvgi
summ m2, d

gen m6=(-4.84)+0.92*dsri+0.528*gmi+0.404*aqi+0.892*sgi+0.115*depi-0.172*sgai+4.679*1-0.327*lvgi
summ m5, d

gen m4=(-4.84)+0.92*dsri+0.528*gmi+0.404*aqi+0.892*sgi+0.115*depi-0.172*sgai+4.679*tata1-0.327*lvgi if dsri<dsri_p95 & gmi<gmi_p95 & gmi>gmi_p1 & aqi<aqi_p95 & sgi<sgi_p95 & depi<depi_p95 & sgai<sgai_p95 & lvgi<lvgi_p95
summ m4, d

gen m5=(-4.84)+0.92*dsri+0.528*gmi+0.404*aqi+0.892*sgi+0.115*depi-0.172*sgai+4.679*1-0.327*lvgi if dsri<dsri_p95 & gmi<gmi_p95 & gmi>gmi_p1 & aqi<aqi_p95 & sgi<sgi_p95 & depi<depi_p95 & sgai<sgai_p95 & tata>tata_p1 & tata<tata_p99 & lvgi<lvgi_p95
summ m5, d


* (3) generate S

gen s=(nit-ebitdat)/tat
summ s, d

* (4) generate indexes for all variables on the model on the basis of ((year t - tear (t-1))/yeat (t-1)*100)

gen ppe=((ppet-ppet1)/ppet1)*100
summ ppe,d
hist ppe
lowess ppe ppet1
egen ppe_p95 = pctile(ppe), p(95)


gen inv=((invt-invt1)/invt1)*100
summ inv, d
hist inv
egen inv_p95 = pctile(inv), p(95)

gen ar=100*(art-art1)/art1
summ ar, d
hist ar
egen ar_p95 = pctile(ar), p(95)

gen ap=100*(apt-apt1)/apt1
summ ap, d
hist ap
egen ap_p95 = pctile(ap), p(95)


gen ca=100*(cat-cat1)/cat1
summ ca, d
egen ca_p95 = pctile(ca), p(95)


gen ta=100*(tat-tat1)/tat1
summ ta, d
egen ta_p95 = pctile(ta), p(95)


gen cl=100*(clt-clt1)/clt1
summ cl, d
egen cl_p95 = pctile(cl), p(95)


gen tl=100*(tlt-tlt1)/tlt1
summ tl, d
egen tl_p95 = pctile(tl), p(95)


gen e=100*(et-et1)/et1
summ e, d
egen e_p95 = pctile(e), p(95)


gen sg=100*(st-st1)/st1
summ sg, d
egen sg_p95 = pctile(sg), p(95)


gen cogs=100*(cogst-cogst1)/cogst1
summ cogs, d
```

```
egen cogs_p95 = pctile(cogs), p(95)
summ cogs if cogs<cogs_p95, d


gen de=100*(det-det1)/det1
summ de, d
egen de_p95 = pctile(de), p(95)


gen p=100*(pt-pt1)/pt1
summ p, d
egen p_p95 = pctile(p), p(95)


gen da=100*(dat-dat1)/dat1
summ da, d
egen da_p95 = pctile(da), p(95)


gen ni=100*(nit-nit1)/nit1
summ ni, d
egen ni_p95 = pctile(ni), p(95)
egen ni_p5 = pctile(ni), p(5)


gen ocf=100*(ebitdat-ebitdat1)/ebitdat1
summ ocf, d
hist ocf, bin (100)
egen ocf_p95 = pctile(ocf), p(95)
egen ocf_p5 = pctile(ocf), p(5)
summ ocf if ocf<ocf_p95 & ocf>ocf_p5, d


**************
* Validation *
**************

* calibration

logistic man z_p95
hl
predict p_z_p95
hl man p_z_p95 , plot xla(0.5 (.1) 0.8) yla(0.3 (.1) 0.8)

logistic man m5
hl
predict p_m5
hl man p_m5 , plot xla(0.5 (.1) 0.8) yla(0.3 (.1) 0.8)
*m5 is trimmed already (tata=1)

logistic man s
hl
predict p_s
hl man p_s , plot xla(0.5 (.1) 0.8) yla(0.3 (.1) 0.8)

*discrimination

logistic man z_p95
lroc

logistic man m5
lroc

logistic man s
lroc

roccomp man z_p95 s m5, graph

******************************************

gen random=runiform(0,1)
summ random, d

gen r=0 if random<0.5
replace r=1 if random>=0.5
tab r

* if r==0 generation
* if r==1 validation

******************************************


tab r
drop random

logistic man ppe if ppe<ppe_p95 & r==0 /*not related*/
logistic man inv if inv<inv_p95 & r==0 /*not related/
logistic man ar if ar<ar_p95 & r==0 /*not related*/
logistic man ap if ap<ap_p95 & r==0
logistic man ca if ca<ca_p95 & r==0 /*margin*/
logistic man ta if ta<ta_p95 & r==0 /*not related*/
logistic man cl if cl<cl_p95 & r==0 /*not related*/
logistic man tl  if tl<tl_p95 & r==0
logistic man e if e<e_p95 & r==0
logistic man sg if sg<sg_p95 & r==0 /*not related*/
logistic man cogs if cogs<cogs_p95 & r==0 /*not related*/
logistic man de if de<de_p95  & r==0 /*not related*/
logistic man p if p<p_p95 & r==0 /*not related*/
logistic man da if da<da_p95 & r==0 /*not related*/
logistic man ni if ni<ni_p95 & ni>ni_p5 & r==0 /*not related*/
logistic man ocf if ocf<ocf_p95 & ocf>ocf_p5 & r==0 /*not related*/

logistic man dsri if dsri<dsri_p95 & r==0 /*not related*/
logistic man lvgi if lvgi<lvgi_p95 & r==0
logistic man di if di<di_p95 & r==0 /*not related*/
logistic man ii if ii<ii_p95 & r==0 /*not related*/
logistic man gmi if gmi<gmi_p95 & gmi>gmi_p1 & r==0 /*margin related/
logistic man aqi if aqi<aqi_p95 & r==0 /*not related*/
logistic man sgi if sgi<sgi_p95 & r==0 /*not related*/
logistic man depi if depi<depi_p95 & r==0 /*not related*/
logistic man sgai if sgai<sgai_p95 & r==0 /*not related*/
logistic man tata if tata<tata_p99 & tata>tata_p1 & r==0 /*not related*/
logistic man s if r==0 /*not related*/

* list of variables/indices related to man (if excluding outliers): ap tl e lvgi ca gmi

spearman ap tl e lvgi ca gmi if r==0 & ap<ap_p95 & tl<tl_p95 & e<e_p95 & lvgi<lvgi_p95 & ca<ca_p95 & gmi<gmi_p95 & gmi>gmi_p1

* give the high correlation between lcgi and tl/e we likely need to choose


****************************
* development of new model *
****************************
```

```
logistic man ap tl e lvgi ca gmi if r==0 & ap<ap_p95 &
tl<tl_p95 & e<e_p95 & lvgi<lvgi_p95 & ca<ca_p95 &
gmi<gmi_p95 & gmi>gmi_p1

logistic man ap lvgi ca gmi if r==0 & ap<ap_p95  &
lvgi<lvgi_p95 & ca<ca_p95 & gmi<gmi_p95 &
gmi>gmi_p1

logistic man ap lvgi ca if r==0 & ap<ap_p95  &
lvgi<lvgi_p95 & ca<ca_p95

logistic man  lvgi ca gmi if r==0  & lvgi<lvgi_p95 &
ca<ca_p95 & gmi<gmi_p95 & gmi>gmi_p1 /*no*/

logistic man ap lvgi gmi if r==0 & ap<ap_p95  &
lvgi<lvgi_p95 & gmi<gmi_p95 & gmi>gmi_p1 /*no*/

* final model
logistic man ap lvgi ca if r==0 & ap<ap_p95  &
lvgi<lvgi_p95 & ca<ca_p95

logistic man ap lvgi ca if r==0 & ap<ap_p95  &
lvgi<lvgi_p95 & ca<ca_p95, coef

gen expon= exp(-4.640318 + (( -.0152856)*ap) +
((4.673196)*lvgi) + (( -.018379)*ca)) if ap<ap_p95  &
lvgi<lvgi_p95 & ca<ca_p95

gen risk=expon/(1+expon)
summ risk
```

```
***************************
* Validation of new model *
***************************

* validity in development cohort
hl man risk if r==0 & ap<ap_p95  & lvgi<lvgi_p95 &
ca<ca_p95 , plot

logistic man risk if r==0 & ap<ap_p95  & lvgi<lvgi_p95
& ca<ca_p95
lroc

* validity in validation cohort
hl man risk if r==1 & ap<ap_p95  & lvgi<lvgi_p95 &
ca<ca_p95 , plot

logistic man risk if r==1 & ap<ap_p95  & lvgi<lvgi_p95
& ca<ca_p95
lroc

hl man risk if r==1 & ap<ap_p95  & lvgi<lvgi_p95 &
ca<ca_p95 , plot xla(0.5 (.1) 0.8) yla(0.3 (.1) 0.8)

* validity all cohorts
hl man risk if ap<ap_p95  & lvgi<lvgi_p95 & ca<ca_p95
, plot

logistic man risk if ap<ap_p95  & lvgi<lvgi_p95 &
ca<ca_p95
lroc

hl man risk if ap<ap_p95  & lvgi<lvgi_p95 & ca<ca_p95
, plot xla(0.5 (.1) 0.8) yla(0.3 (.1) 0.8
```