Earnings Manipulation and Excess Returns

A study on the Swedish stock exchange

Simon Tarvainen & Robert Rask

Bachelor Thesis in Finance

Stockholm School of Economics May 14, 2018

Tutor: Riccardo Sabbatucci

Abstract: This study aims to evaluate the efficacy of one of the most well-known quantitative methods for discovering earnings manipulation, the M-score model (Beneish 1999), as a tool for investors on the Swedish stock exchange by testing the model's predictive power on future stock returns. The data used in the study covers 126 firms during the years of 2005-2017. The M-score's predictive power is tested by regressing the firm's assigned M-score against the stock's excess returns during the period, also testing for potential lagged effects. In accordance with previous research on the American stock exchange, the hypotheses of this study are that M-score higher has explanatory power on stock returns, and that a higher M-score should yield lower stock returns. Results conclude however that the M-score does not have significant explanatory power on future stock returns. Companies with a high M-score were not found to have been performing worse than companies with a low M-score on the stock exchange during the sample period, contrary to the hypothesis of this study. A likely cause of these results stems from the ongoing evolution of more complicated financial manipulation schemes, and knowledge of the model becoming more common amongst manipulating firms. With the hypotheses being rejected, M-score is not deemed to be a reliable tool to investors on the Swedish stock exchange and will likely be even less so in the future.

***Keywords*:** M-score, earnings manipulation, financial statement manipulation, stock returns, Fama-French Three-Factor Model

Acknowledgments: We would like to thank Riccardo Sabbatucci for his valuable input and assistance.

Correspondence: to: Simon Tarvainen 23727@student.hhs.se; Robert Rask 23729@student.hhs.se

Table of Contents

1. Introduction

Financial market participants must always be wary of fraud and manipulation due to the high costs associated with it. This is illustrated by several famous corporate implosions such as Worldcom, Enron and the Swedish company Fermenta. Their egregious earnings manipulation ended up costing investors billions of dollars in lost savings. These examples highlight the interest of investors and markets participants to avoid firms which materially misstate their earnings with or without malicious intent. The shares of firms engaged in such activities are more likely to experience a sudden substantial loss of value compared to their peers not engaged in manipulations.

By separating companies which are deemed likely to manipulate their earnings from those that are unlikely to do so, investors are able to greatly enhance their chances of protecting their capital and by extension dramatically improve their prospective returns. However, discovering these potential manipulators has historically been considered very difficult due to the accounting expertise and the extensive experience required to reach any qualified judgement. Given the loss of capital associated with mistaking a manipulator for a non-manipulator the costs of misclassification can be highly punitive to inexperienced practitioners. The exercise of investigating financial statements with the purpose to uncover signs of fraud or manipulation has been termed "forensic accounting". This term commonly refers to the process of analysing financial statements and regulatory filings to uncover irregularities or inconsistencies which indicate fraudulent or intentionally misleading accounting. The meaning of "forensic" describes the aim of uncovering any dishonest or illegal activities through such an investigation.

Historically this endeavour has required a great amount of human labour and individual judgement making it a costly, time consuming activity. The case could be made that forensic accounting is more of an artform rather than an exact science. This circumstance coupled with the tectonic shift in the asset management industry away from active management of capital into passive or automated vehicles and strategies during the past years raises a profound concern. A quantitative or passive strategy is ill equipped to incorporate such an artisanal concept as forensic accounting. However, there are models which attempt to condense the essence of forensic accounting into readily usable quantitative formulas.

The M-score model originally created by Messod D. Beneish in 1999 is perhaps the most common of these models. It seeks to identify earnings manipulation by examining accounting information through which reflections of the most common forms of manipulation can be uncovered. Combining the primary factors of forensic accounting with metrics aiming to indicate a firm's propensity to engage in manipulation, the model has shown favourable results in previous research. This study seeks to investigate the efficacy of the M-score model in modern market conditions and draw conclusions on whether this quantitative model qualifies as a reliable tool for predicting stock returns in the future.

2. Theory and Previous Literature

There are several studies in areas related to the central premise of this paper, e.g. the connection between earnings manipulation and bankruptcy, and how accounting measures can be used to discover earnings manipulation and models for detecting manipulation. This study in particular will be similar to that of Beneish et al. (2012), which studies how manipulation of financial statements affect financial returns of American stocks.

This paper's contribution will be the examination of whether similar results, as achieved in previous studies where the M-score model was applied, can be achieved on a smaller and less diversified equity market, specifically the Swedish stock exchange. The following section presents the theoretical context of the study, primarily the Fama-French Three Factor Model and the M-score Model.

2.1 Economic Theories

2.1.1 The Efficient Market Hypothesis

The theory of market efficiency first introduced by Fama (1970) is an economic theory which stipulates that the prices of assets will incorporate information to the degree that they reflect the security´s actual intrinsic value. It further stipulates that investors are rational, can understand information correctly and act upon said information. There is no arbitrage opportunities available in the market according to this theory, except for limited time periods when investor irrationality might give rise to them. However, those instances of price discrepancy are quickly corrected by rational investors bidding for the securities until they converge to intrinsic value once again.

There are three levels of efficiency, the first of which is weak-form efficiency. It implies that current prices of assets will incorporate all historical information essentially rendering the data of past prices performance of a security useless to investors since it is already incorporated in the current price. The second level is semi-strong-form efficiency which is an extension of the weak-form efficiency. It states that asset prices also reflect all publicly available information such as the information contained in regulatory filings. Lastly there is the strong-form efficiency which suggests asset prices incorporate all information i.e. even information known only to company insiders.

2.1.2 Fama-French Three-Factor Model

Presented by Fama and French (1993), the Fama-French Three-Factor Model is a country specific asset pricing model that calculates the expected return of equities based on three factors. The Fama-French Three-Factor Model is based on earlier research by Fama and French (1992) which concludes that a firm's size, earnings/price-ratio, leverage, and book-to-market equity has explanatory power for the stock's return. In the Fama-French Three-Factor Model, these four factors are embedded in the two variables: SMB, which captures the risk of firms with small market capitalizations relative to larger firms and HML, which captures the risk of growth companies that have low equity book values compared to their market capitalizations. A market variable (MKT) was also added which captures the market's excess return, attributable to the risk of the market which includes all listed companies. In the study conducted by Fama and French (1993), the model proved to have significant explanatory power of excess stocks returns with an R^2 above 0.9 in 21 of its 25 portfolios during the time period of 1963-1991. The complete Fama-French Three-Factor Model is presented below:

$$R_{t_i} = \alpha + \beta_{MKT_i} * R_{MKT_i} + \beta_{SMB} * R_{SMB_i} + \beta_{HML} * R_{HML_i} \qquad (1)$$

The SMB factor is calculated as the average return of three portfolios containing small sized companies sorted by book value of equity relative to market value of equity (book-to-market), minus the average return of three portfolios containing large size companies sorted by book-to-market. The book-to-market is separated into the categories Low, Medium and High or just Low and High while the size categories are Small and Big. The formula is presented below:

$$SMB = \frac{1}{3} * (Small\ Low + Small\ Medium + Small\ High) - \frac{1}{3} * (Big\ Low + Big\ Medium + Big\ High) \qquad (2)$$

The HML-factor is calculated by the average return on the two portfolios with high book-to-market (sorted by size), minus the average return of the low book-to-market portfolios (sorted by size) factor. The formula is presented below:

$$SMB = 1/2\ (Small\ High + Big\ High) - 1/2\ (Small\ Low + Big\ Low) \qquad (3)$$

The Fama-French Three-Factor Model provides an empirically sound framework for evaluating the M-score's ability to predict excess return in the stock market.

2.2 Empirical Studies

2.2.1 M-score Model

The M-score model, developed by Beneish (1999) is a quantitative model designed to detect earnings manipulation in listed companies by analysing data contained in their financial statements. It does so by aggregating common financial metrics that signal financial statement manipulation - so called forensic accounting principles - and factors that indicate a higher likelihood on the part of management to engage in manipulation, into a formula.

M-score is a function of eight variables, each with a different weighting which is commensurate to their correlation coefficients presented in Beneish (1999). The calculation of one of the factors, total accruals to total assets (TATA), was updated by Beneish, Lee and Nichols (2012) due to cash flow data becoming directly available via cash flow statements, which previously had to be derived from the balance sheet in Beneish (1999). No other change was made to the original model. The variables used in the M-score model are:

$DSRI_t$ = Days' Sales in Receivable Index at time t.
GMI_t = Gross margin index at time t.
AQI_t = Asset quality index at time t.
SGI_t = Sales growth index at time t.
$DEPI_t$ = Depreciation index at time t.
$SGAI_t$ = Sales general and administrative expenses index at time t.
$TATA_t$ = Total assets to total accruals at time t.
$LVGI_t$ = Leverage index at time t.

Descriptions of the accounting data required for the calculation of the variables, the calculations and rationale behind the individual M-score components are presented in Appendix A.

The complete formula of the M-score model is presented below:

$$mscore_t = -4{,}84 + 0{,}920 * DSRI_t + 0{,}528 * GMI_t + 0{,}404 * AQI_t + 0{,}892 * SGI_t + 0{,}115 * DEPI_t - 0{,}172 * SGAI_t + 4{,}679 * TATA_t - 0{,}327 * LVGI_t \quad (4)$$

The model assigns a numerical M-score value to a company at a specific point in time. If that value is greater or lesser than a predetermined limit-value, a company will be classified as a manipulator (greater than the limit-value) or a non-manipulator (lesser than the limit-value).

A limit-value is determined by the relative costs associated with model errors. The model is prone to two types of errors, it can classify a company as a non-manipulator when the company manipulates (Type I error), or it can classify a non-manipulator as engaging in manipulation (Type II error). Depending on the assumed relative misclassification costs associated with Type I and Type II errors, the cut-off probability which minimize the costs of errors will vary. Beneish computation of the expected costs of misclassification is presented in the formula below:

$$ECM = P(M) * PI * C1 + [1 - P(M)] * PII * CII \quad (5)$$

Where ECM denotes the expected costs of misclassification, P(M) the prior probability of encountering earnings manipulators, PI and PII the conditional probabilities of Type I and Type II errors and CI and CII are the costs of Type I and Type II errors. E.g. a relative error cost of 20:1 and 30:1 yielded the cut-off probability 0.0376. This corresponds to the M-score value -1.78, at which the expected costs of errors are minimized.

The M-score model was used to predict stock returns in Beneish, Lee and Nichols (2012). It was utilized to show how investment portfolios containing companies flagged as manipulators generated lower returns compared to portfolios of non-flagged companies. The test was used on stocks listed on NYSE, Nasdaq and AMEX, and performance was measured during the period of 1993-2009. The limit-value used to classify manipulators was -1.78. With this limit-value several known and discovered manipulators were flagged. However, the model also flagged a specific type of firm which, according to the authors, had experienced high growth but whose fundamentals had begun to deteriorate. These were referred to as "…look[ing] like manipulators".

The study concluded that the M-score model was an effective tool for predicting relative returns of non-flagged companies compared to that of flagged companies during the 1993-2009 period, showing that flagged companies yielded statistically significant lower stock returns. However, much of this predictive power derived from the variables which assessed a company's inclination to manipulate, i.e. variables which indicate deteriorating fundamentals rather than accrual-variables which measure the magnitude of accounting irregularities.

2.2.2 Other Relevant Empirical Research

One important study related to the subject of earnings manipulation is Dechow et al. (1996) which investigates the motives and consequences of earnings manipulation in American firms. Weak corporate governance structures were found to enable earnings manipulation. Further they discovered that the cost of capital for a given company increased materially after their overstatements of earnings were publicly revealed.

By researching companies which entered bankruptcy, Rosner (2003) found that failing firms, i.e. companies pre-bankruptcy, were significantly overstating earnings in bad years by using financial manipulation. Further she found that pre-bankruptcy firms mainly manipulated earnings via accruals and that analysis of actual cash flow displayed signs of deteriorating fundamentals. These findings on manipulation is consistent with research presented by Beneish (1999) as well as the construction and purpose of the M-score model.

2.3 Hypotheses

The previous literature and economic theories presented constitutes the underlying framework for this paper. It is on the basis of this literature that the thesis will present the hypotheses to be tested.

The M-score model is derived from research conducted by observing only US listed companies. Given the significant differences in size, liquidity and diversification between the US and Swedish market the first hypothesis will capture the model's transferability to the Swedish stock market and its explanatory power on said market. Thus, the first hypothesis of this thesis was formulated as follows:

Hypothesis 1: The M-score model for detecting earnings manipulation has explanatory power over future stock price returns on the Swedish stock market.

The intuition behind the second hypothesis draws from the evidence presented by Beneish et al. (2012) which found that the returns of portfolios consisting of manipulating firms were worse than portfolios consisting of non-manipulating firms. In conjunction with this the second hypothesis was formulated as:

Hypothesis 2: The companies flagged as earnings manipulators by the M-score model will have lower excess stock returns compared to companies which are not flagged as manipulators.

3. Data and Methodology

This section details the datasets used, how the sample companies were derived, what tests and regressions have been made and the variables used.

3.1 Notations

The different variables throughout the thesis are denoted as follows:

$mscore_t$: M-score in period t

R_{f_t}: Risk-free rate (annual average interest on Swedish 10-year government bonds) in period t

R_t: Annual total return (adjusted for any dividends) of stock in period t

$excret_t$: Excess return of stock in period t

ER_t: Adjusted excess return of stock in period t

R_{SMBt}: Fama-French small minus big portfolio returns in Sweden in period t

R_{HMLt}: Fama-French high minus low portfolio returns in Sweden in period t

R_{MKTt}: Fama-French market portfolio returns in Sweden in period t

ER_{nf_t}: Excess return of non-flagged companies portfolio in period t

ER_{f_t}: Excess return of flagged companies portfolio in period t

$adjER_{f_t}$: Adjusted excess return of flagged companies portfolio in period t

e_i: Error term

3.2 Dataset

The dataset consists of data collected from the databases of FactSet, AQR Library, the Swedish Riksbank and Wharton Research Data Services (WRDS). The dataset consists of accounting items necessary for calculating M-score as defined in Appendix A matched with total stock return data for each company during the same periods, as well as Fama-French return factors and risk-free rate data. The population the sample was drawn from consists of the 346 listed stocks traded on the Stockholm exchange as of the 7th of March 2018. The final sample consists of 126 listed companies and 1208 observations in total, spread out between the years of 2005-2017.

3.3 Choice of Data Sample

The sample used in this study was generated by removing entities from the population on the basis of sample criteria to ensure reliability and temporal compatibility with Fama-French factor data, return data and the M-score model.

The companies whose fiscal year did not follow a calendar year was excluded from the sample. This was to allow timing of the accounting data to be consistent with the timing of return data through all samples, to minimize the impact of seasonality, and to avoid misrepresentations in the calculation of Fama-French factors. In the cases where companies have dual share classes listed on the exchange e.g. A-shares and B-shares, the share class which was the least traded was excluded to avoid the possibility of illiquidity affecting return or Fama-French factor data.

Real estate companies and enterprises operating in the financial industry e.g. insurance or banking were excluded, as the nature of these operations leads to the accounting data not being compatible with the M-score model. As the model is not able to reliably discover earnings manipulations in these industries, these firms were removed to avoid model failures.

Lastly, observations in periods where any of the M-score model's variables could not be calculated in a reliable way were also excluded. The choice of time period was based on the ability to maximize the number of observations while still being able to use reliable and comparable accounting data. As accounting standards change over time and companies' accounting data is often not readily available for years far back in time, going back much further would yield data of lower quality. The period of 2005-2017 was thus deemed to be an appropriate since this allowed the data to be of high quality and ensuring comparability while still allowing the period to contain a sufficient number of observations. During these years the Swedish stock market has had long periods of both up- and down-turns. This is optimal, as it is desirable to test the M-scores impact on excess returns in different stock market climates.

3.4 Regression

To test whether the M-score has explanatory power over future stock returns or not, a regression analysis was conducted using the dependent and independent variables described in 3.4.1 Variables. The complete regression model is given by:

$$ER_t = \beta_0 + \beta_{mscore} * mscore_t + \beta_{R_{SMB}} * R_{SMB_t} + \beta_{R_{HML}} * R_{HML_t} + \beta_{R_{MKT}} R_{MKT_t} + \varepsilon_i \quad (6)$$

3.4.1 Regression Variables

Dependent Variables

The dependent variable used in the primary regression is excess return. Excess return is defined as the stock's annual total return (adjusted for any dividends), subtracted by the risk-free rate. The risk free-rate is defined as the average interest rate on Swedish 10-year government bonds during period t. The formula for calculating excess return is thus given by:

$$excret_i = R_{t_i} - R_f \quad (7)$$

Due to extreme outliers in the excret variable it was decided to create a new dependent variable ER. This variable is equal to excess return adjusted to replace values outside the upper 1% bound with the lowest value of that percentile.

Independent variables

One of the independent variables used to approximate excess returns is mscore. The mscore variable is the value calculated using the M-score model previously described in 2.2.1 M-score model. In period t, the mscore variable is the companies' M-score value calculated at the beginning of the year using accounting data from the recently passed fiscal year. The mscore variable is calculated as follows for every period t:

$$mscore_t = -4{,}84 + 0{,}920 * DSRI_t + 0{,}528 * GMI_t + 0{,}404 * AQI_t + 0{,}892 * SGI_t + 0{,}115 * DEPI_t - 0{,}172 * SGAI_t + 4{,}679 * TATA_t - 0{,}327 * LVGI_t \quad (8)$$

Further information on the underlying components and calculations of these variables are presented in appendix A.

The three Fama-French variables R_{SMB} R_{HML} and R_{MKT}, more thoroughly described in 2.1.2 Fama-French Three-Factor Model, are also included as independent variables. This allows the regression to incorporate factors with strong explanatory power to avoid large error terms. I.e. by incorporating these factors into the regression model leads to an improved measure of M-score's impact on excess returns.

3.4.2 Regression Model Assumptions

The following assumptions have been made as part of the regression analysis:

1. In the population model the dependent variable, y, is related to the independent variable, x, and the error (or disturbance), u, as $y = \beta_0 + \beta_1 x + u$ where β_0, β_1 are the population intercept and slope parameters respectively.
2. There is a random sample of n, $\{(x_i,y_i) : i = 1,2,...,n\}$ following the population model in 1.
3. The sample outcomes on x, $\{x_i, i = 1,2..., n\}$, are not all the same value.
4. The error u has an expected value of zero given any value of the explanatory value
5. The population error u is independent of the explanatory variables $x_1, x_2, ..., x_k$ and is distributed $u \sim N(0; \sigma 2)$*

*Note: Results from a normality test presented in 4.2.1 Normality Test supports this assumption.

Generalized Least Squares (GLS) regression with panel sorted data is robust to heteroscedasticity, consequently the assumption of homoscedasticity is not required to hold for the regression.

3.5 t-test

To test the second hypothesis of this study, a Welch t-test was conducted to determine if the adjusted mean excess stock returns of a portfolio consisting of all flagged company stocks ($adjER_f$) was statistically significantly different from the mean excess stock returns of a portfolio of all non-flagged company stocks (ER_{nf}). The portfolio division of flagged and non-flagged is based on the annual M-score values where any value greater than -1.78 in a single period will lead to a company being classified as a manipulator (flagged) for that period.

Companies with an M-score value below -1.78 is classified as non-manipulators (non-flagged). According to previous research e.g. the efficient market hypothesis, securities generally have different variances. The t-test must accordingly allow for unequal variances. As a result, it was deemed necessary to conduct a Welch t-test instead of a Student's t-test.

3.5.1 t-test Variables

To measure and compare excess returns of the two portfolios in each period, two separate variables were created.

ER_{nf}

The variable ER_{nf} represents excess returns in the non-flagged company stocks portfolio. ER_{nf} is computed as follows:

$$ER_{{nf}_t} = excret_{{nf}_t} - (\beta_{R_{{SMB}_{{nf}_t}}} * R_{SMB_{nf_t}} + \beta_{R_{{HML}_{{nf}_t}}} * \ R_{HML_{nf_t}} + \ \beta_{R_{{MKT}_{nf}}} * R_{MKT_{nf_t}}) \tag{9}$$

The portfolio Betas in the Fama-French model is estimated for the period 2005-2017 using the following regression models:

$$\beta_{SMB_{nf}}: excret_{{nf}_t} = \ \beta_0 + \ \beta_{R_{SMB_{nf}}} * R_{SMB_{nf_t}} + \varepsilon_i \tag{10}$$

$$\beta_{MKT_{nf}}: excret_{{nf}_t} = \ \beta_0 + \ \beta_{R_{MKT_{nf}}} * R_{MKT_{nf_t}} + \varepsilon_i \tag{11}$$

$$\beta_{HML_{nf}}: excret_{{nf}_t} = \ \beta_0 + \ \beta_{R_{HML_{nf}}} * R_{HML_{nf_t}} + \varepsilon_i \tag{12}$$

Complete calculations of Beta values used in the model are presented in appendix B.

$adjER_f$

The variable $adjER_f$ computes excess returns in flagged company stocks. $adjER_f$ is computed as follows:

$$ER_{f_t} = excret_{f_t} - (\beta_{R_{{SMB}_{f_t}}} * R_{SMB_{f_t}} + \beta_{R_{{HML}_{f_t}}} * \ R_{HML_{f_t}} + \ \beta_{R_{{MKT}_f}} * R_{MKT_{f_t}}) \tag{13}$$

To adjust for extreme outliers, values in the highest 1% bound was replaced with the lowest value in those percentiles. Such extreme outliers were not present in ER_{nf}, which meant there was no need for adjustments to its values.

The portfolio Betas in the Fama-French model is estimated for the period 2005-2017 using the following regression models:

$$\beta_{SMB_f}: excret_{f_t} = \beta_0 + \beta_{R_{SMB_f}} * R_{SMB_{f_t}} + \varepsilon_i \quad (14)$$

$$\beta_{MKT_f}: excret_{f_t} = \beta_0 + \beta_{R_{MKT_f}} * R_{MKT_{f_t}} + \varepsilon_i \quad (15)$$

$$\beta_{HML_f}: excret_{f_t} = \beta_0 + \beta_{R_{HML_f}} * R_{HML_{f_t}} + \varepsilon_i \quad (16)$$

Complete calculations of Beta values used in the model are presented in appendix B.

3.5.2 t-test Assumptions

To conduct a t-test it must first be established that the residuals of the dependent variables $adjER_f$ and ER_{nf} are normally distributed. The variables ER_{nf} and $adjER_f$ were tested for normality using a skewness-kurtosis test, and the distribution of residuals were plotted against a kernel density estimation. This was done to determine if the normality assumption required for the t-test would hold. The other GLS assumptions are assumed to hold similarly to the same extent as for the primary regression. Using results presented in 4.3.2 Normality Test for t-test Assumptions, all necessary assumptions for conducting the t-test could be assumed to hold.

3.6 Selection Bias and Other Sample Issues

Survivorship bias is a concern because the sample was derived from companies currently traded on the Swedish stock exchange. Implicitly, companies that were delisted before March 6th 2017 is not included in the original population from which the sample was developed. The possibility of manipulating companies to enter bankruptcy as outlined by Rosner (2003) or become delisted after the manipulation is discovered, leads to a potential sample issue. When a company is no longer publicly traded, the accounting information needed generally becomes difficult to obtain because it is removed from databases which only include data of contemporary listed firms.

This is a particularly problematic sample issue for this study since a delisting or bankruptcy due to fraud generally corresponds to large negative returns. Because these observations would be missing from the dataset there is a risk that the average stock returns of the subgroup flagged as manipulators skew more positive compared to what the actual average would be if these observations were could be included in the dataset. Related to this is a concern about a single observation in the dataset, a very extreme outlier in the excess return variable in the form of Fingerprint Cards with a total stock return of 1 598% percent during 2015. The same year it was flagged as a manipulator which means it dramatically distorts the average excess return data of the flagged companies.

Another significant problem is the exclusion of firms as a consequence of them not reporting certain accounting items required to complete the computation of M-score. The model requires two sets of twelve unique accounting items, one from the current year and one from the previous, to calculate an M-score value. Most of these items are mandatory to disclose, but they also include some which a company could conceal if they wished. Consequently, it is possible that manipulating firms which are aware of the M-score model's existence could design the structure of their financial statement to conceal some specific accounting items required by M-score. For example, by reporting their costs by type e.g. personnel costs and raw material costs, instead of by corporate function e.g. selling costs manufacturing costs, a business can conceal their selling general and administrative expenses, which is one of the line items used in the model.

Another problem with the sample is that in total only 118 observations out of the 1 208 are occasions when companies are flagged as manipulators. Using a relatively small sample size means the results presented runs a higher risk of not being applicable on the entire population. A larger sample would be preferable to ensure reliability, but the study has been restricted by the amount of readily accessible data. It is important to note that the sample is limited by the Swedish stock market becoming a developed market later than the U.S. stock market, that it consists of far less firms and consequently it is not as rich in historical data. In contrast it is an innate aspect of the study's hypothesis to test the model's efficacy on a small stock market which in itself necessitates a relatively small sample.

4. Empirical results

This section will present the aggregate summary statistics for the variables representing M-score and excess returns. After that the results from the primary regression testing the explanatory power of M-score is presented. Subsequently the results from a Welch t-test for unequal variances is exhibited to determine if the average returns of non-flagged and flagged firms significantly differ from each other. Lastly a robustness check of the primary regression was conducted where the independent M-score values were lagged to see if the predictive power was enhanced when regressed to returns the years after the model flagged a specific company.

4.1 Summarized Statistics

This segment shows the descriptive statistics of the sample used in the various tests.

Table 1: Descriptive summary statistics for excess returns.

Percentiles				
	Smallest		Obs	1208
1%	-.76484	-.9509574	Sum of Wgt.	1208
5%	-.5338275	-.8951218	Mean	.1809682
10%	-.3968927	-.8713795	Std. Dev.	.5398571
25%	-.1492434	-.8588538	Variance	.2914457
50%	.11325	-5.527717	Skewness	1.540389
	Largest		Kurtosis	7.611767
75%	.3990877	2.603858		
90%	.7798992	2.603858		
95%	1.118693	2.603858		
99%	2.603858	2.603858		

Note: The sample consists of 126 listed Swedish firms that operate outside the real estate and financial sectors. The observations are collected from the time period 2005-2017. The number of observations vary each year as a consequence of some firms not being listed certain years while others are listed throughout the whole time period. ER (excess return) is the annual stock returns of the individual sample firms adjusted by the risk-free rate and the Fama-French factor returns. Values in the upper 1% bound are replaced with the lowest value in that percentile.

Table 2: Descriptive summary statistics for mscore.

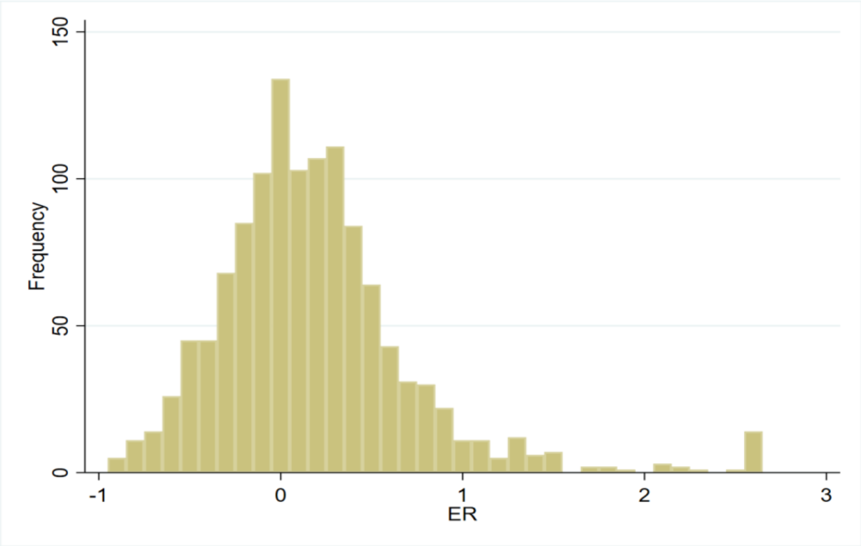
Percentiles					
		Smallest		Obs	1208
1%	-5.527716	-5.527716		Sum of Wgt.	1208
5%	-3.315179	-5.527716		Mean	-2.334961
10%	-3.001319	-5.527716		Std. Dev.	1.048498
25%	-2.667397	-5.527716		Variance	1.099348
50%	-2.427924			Skewness	3.259616
		Largest		Kurtosis	23.7588
75%	-2.190317	4.4			
90%	-1.789418	4.4			
95%	-1.116863	4.4			
99%	4.4	4.4			

Note: The sample consists of 126 listed Swedish firms that operate outside the real estate and financial sectors. The observations are collected from the time period 2005-2017. The number of observations vary each year as a consequence of some firms not being listed certain years while others are listed throughout the whole time period. mscore is the M-score value the model assigned to each company at the beginning of each fiscal year adjusted to replace values outside the upper and lower 1% bounds with the lowest value in those percentiles. mscore had large outliers outside of that boundary which could upset the regression result.

Figure 1: Excess returns by M-score values

Note: This figure presents a scatter diagram where the value of *ER* for each observation measured on the y-axis is plotted against the corresponding mscore value on the x-axis. mscore is the M-score value (see equation 4 for computation) for an individual company coinciding with the excess return of a given year adjusted to replace values outside the upper and lower 1% bounds with the lowest and highest values in those percentiles respectively. ER (excess return) is the annual stock returns of the individual sample firms adjusted by the risk-free rate.

Figure 2: Excess returns

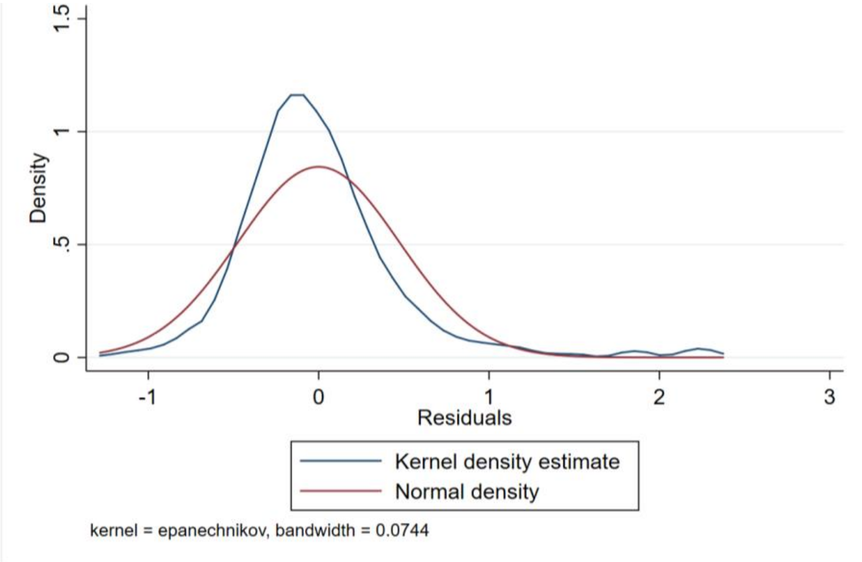


Note: This figure presents a histogram over *ER*. *ER* (excess return) is the annual stock returns of the individual sample firms adjusted by the risk-free rate. Values in the upper 1% bound are replaced with the lowest value of that percentile.

4.2.1 Normality Test for Regression Assumptions

To determine if the normality assumption of the GLS regression holds, the regression residuals were plotted against a kernel density estimate as shown below in Figure 3. After analysing the distribution of residuals compared to the Kernel density estimate, it is deemed reasonable to conclude that the normality assumption is indeed satisfied.

Figure 3: Kernel density test for excess returns



Note: This figure shows the Kernel density estimate of the main regression (see equation 6) and the normal density of the residuals in the main regression. The plot indicates the residuals are normally distributed

4.2.2 Primary Regression Results

Examining the regression results presented in Table 3, the regression model is shown to have weak explanatory power over the dependent variable. The overall R^2 value is 0.2338, meaning that 23.38% of the variance in excess returns is explained by the regression model.

The Fama-French independent variables R_{SMB} and R_{MKT} explain the majority of the variance in excess returns and are both statistically significant at the 0.1%-level with p-values below 0.001. mscore is found to have a p-value of 0.018 and is statistically significant at the 5%-level. Further it has a Beta value of -0.0563764 signifying a negative correlation between M-score and excess returns.

Table 3: Regression results

Random-effects GLS regression	Number of obs = 1208
Group variable:	
companyid	Number of groups = 126
R-sq:	Obs per group:
within = 0.2602	min = 5
between = 0.0001	avg = 9.6
overall = 0.2338	max = 13
corr(u_i, X) = 0 (assumed)	Wald chi2(4) = 378.44
	Prob > chi2 = 0.0000

ER	Coefficient	Std. Err.	z	P>z	[95% Conf. Interval]	
mscore	-.0308553*	.0130212	-2.37	0.018	-.0563764	-.0053342
R_{SMB}	.0131651***	.0012793	10.29	0.000	.0106576	.0156726
R_{MKT}	.0067709***	.0005769	11.74	0.000	.0056403	.0079015
R_{HML}	-.001722	.0015993	-1.08	0.282	-.0048566	.0014126
Constant	.0428412	.0345242	1.24	0.215	-.024825	.1105074

sigma_u .08451943

sigma_e .46707755

rho .03170606

(fraction of variance due to u_i)

Note: This table presents the regression outlined in equation 6 using excess returns as an independent variable. Excess return (ER) is defined as the total annual stock return of the individual sample firms minus the risk-free return. mscore is the M-score value (see equation 4 for computation) for an individual company coinciding with the excess return of any given year. A company's Mscore value (mscore) explain excess returns with significance at the 5% level. ***, ** and * represent statistical significance at the 0,1 %, 1% and 5 % levels, respectively.

4.3 Welch t-test

In this section the study will present the result of the Welch t-test, testing whether the companies flagged as earnings manipulators will show statistically significant lower excess stock returns compared to companies not flagged as manipulators.

4.3.1 Normality Test for t-test Assumptions

Reviewing the skewness/kurtosis test results for the two variables presented in Table 4, it is found that the p-value of skewness and kurtosis combined for the variable ERnf is below 0.01. This implies that the hypothesis of non-normality cannot be rejected for the excess returns of non-manipulator portfolio. However, when interpreting Figure 4 which depicts the residuals of ER_{nf} plotted against a Kernel density estimate, they appear to cluster somewhat according to a normal distribution. There are some outliers in the right tail of the distribution which likely explain the result of the skewness-kurtosis test. It is also known that this test is sensitive when the number of observations is relatively large, which is the case with ER_{nf}. The test result combined with the graphic examination results in the assessment that ER_{nf} is to be considered normally distributed. The p-value 0.1928 for the combined skewness-kurtosis test of $adjER_f$ is above 0.05 and thus the variable appears to follow a normal distribution. This is further evidenced by examining Figure 5 where the observations contained in the $adjER_f$ variable is plotted against a Kernel density estimate. They appear to follow a normal distribution. This might be a result of the disparity in number of observations between the two variables.

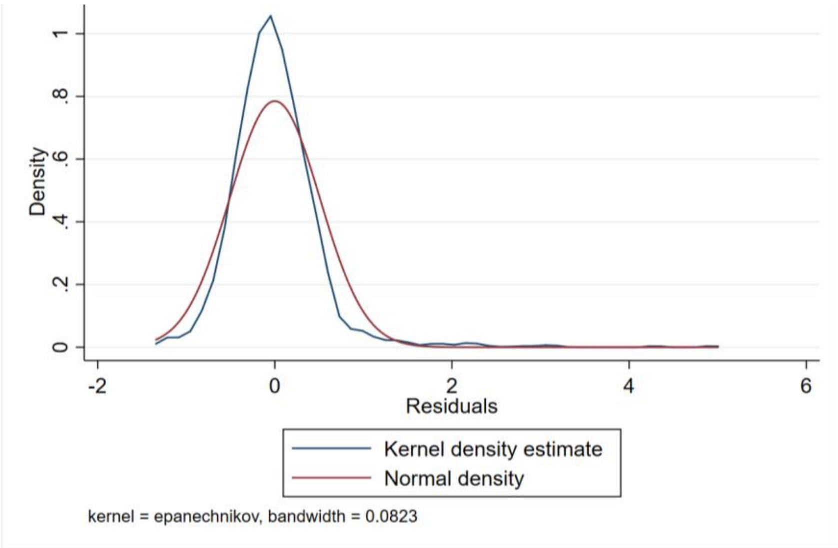
The variables are found to not meaningfully diverge from the estimated normal distributions. It is therefore concluded that the variables satisfy the condition of normality necessary to ensure the reliability of a t-test.

Table 4: Skewness/Kurtosis tests for Normality

Variable	Obs	Pr(Skewness)	Pr(Kurtosis) adj	chi2(2)	Prob>chi2
ER_{nf}	1090	0.0000	0.0000	.0	.0000
$adjER_f$	118	0.0755	0.8063	3.29	0.1928

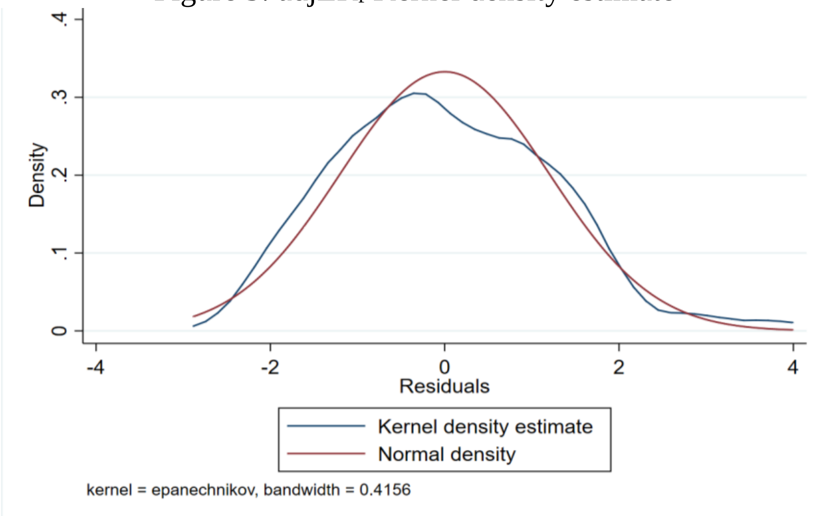
Note: This table displays the result of the Skewness/Kurtosis tests for the variables ER_{nf} and $adjER_f$ (annual excess returns of the equal weighted portfolios separated into firms with M-score values above -1.78 and firms with M-score values below -1.78). The test rejects the hypothesis that the residuals of ER_{nf} follow a normal distribution. Further, the test fail to reject the hypothesis that the residuals of $adjER_f$ follow a normal distributing.

Figure 4: ER_{nf} Kernel density estimate



Note: This figure shows the Kernel density estimate of the regression of ER_{nf} and the normal density of the residuals in the same regression. The plot indicates the residuals are normally distributed.

Figure 5: adjER_f Kernel density estimate



Note: This figure shows the Kernel density estimate of the regression of adjER_f and the normal density of the residuals in the same regression. The plot indicates the residuals are normally distributed.

4.3.2 t-test Results

As can be seen in Table 5, the result of the Welch t-test yields a p-value of 0.9373 for the two-tailed distribution testing the hypothesis of the difference of the two portfolio's excess returns are statistically significantly different from zero. This p-value is much higher than the 0.05 value required to reject the null hypothesis at the 5% significance level.

It follows that the test does not reject the null hypothesis and thus the mean excess returns cannot be said to be statistically significantly different. The p-values for the two one-tailed t-distributions produce the same conclusions. Neither of the p-values are below 0.05, consequently there is no empirical evidence supporting the hypothesis that the mean of excess return of flagged firms is statistically significantly higher of lower than the mean excess return of non-flagged firms.

Table 5: Two-sample t test with unequal variances

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Er_{nf}	1090	.098686	.0154164	.5089747	.0684368	.1289352
$adjER_f$	118	.0898148	.1113979	1.210091	-.1308028	.3104325
combined	1208	.0978194	.017635	.6129268	.0632208	.1324181
diff		.0088712	.1124596		-.2137611	.2315035

diff = mean(ER_{nf}) - mean($adjER_f$) t = 0.0789

Ho: diff = 0 Welch's degrees of freedom = 121.597

Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(T < t) = 0.5314	Pr(T > t) = 0.9373	Pr(T > t) = 0.4686

Note: The table shows the results of the Welch t-test with unequal variances. The p-values indicate that the null hypothesis stating ER_f is statistically significantly equal to $adjER_f$ cannot be rejected. ER_f is the annual stock returns of the sample firms flagged as manipulators minus the risk-free rate and minus the portfolio Fama-French factor returns. $adjER_f$ is the annual stock returns of the sample firms flagged as manipulators minus the risk-free rate and minus the Fama-French factor returns further adjusted by replacing values outside the upper 1% bound with the lowest value within the same percentile.

4.4 Robustness Tests

In the primary regression, the correlation coefficient of M-score is found to be significant at the 5%-level, however, the overall R^2-value is very poor. To test the soundness of these results a robustness test was conducted, the result of which is presented in this section. The test consisted of lagging the M-score variable and consequently pairing each value with excess returns of a later time period, rather than the testing for excess return in the same year. This was done to determine whether the M-score model's predictive power over returns during the years following its detection of potential earnings manipulation were stronger i.e. if it takes more than 1 year for the model's prediction to be actualized. The intuition behind lagging M-score values to test robustness stem from the reasoning that it might take time for the market participants to discover and react to earnings manipulation. Consequently, the negative returns associated with discovered earnings manipulation might appear in one of the years subsequent to the model labelling of a company as a manipulator.

It is not intuitive that investors could know whether the model will flag a firm or not in the future. Due to this fact, the robustness test has not incorporated any effects of leading M-score values i.e. transferring them backwards to pair them with earlier return values. Any result from such a test would not be meaningful to investors neither would it be based on any logically sound assumptions.

Presented in appendix C are Tables 9, 10 and 11 containing the results of these three regressions. Comparing the results to that of the results of the primary regression in Table 3 it is noted that the R^2 values when M-score is lagged becomes lower. In the regression where M-score is lagged one year the R^2 is 0.1981 which is lower than 0.2338 for the primary regression. The same is true when the lagging effect was extended to two years which yielded and R^2 of 0.2085 which is also lower than 0.2338. However, this regression displays a higher R^2 compared to the regression where M-score was lagged with only one year. When lagging M-score with a third year the results yield an R^2 value of 0.2044. This is slightly lower than 0.2085 for the two-year-lagged values but slightly higher than the R^2-value 0.1981 of the regression with the one-year-lag. It is notable that none of the M-score coefficients are significant at, or below, the 5%-level in these regressions.

The results of this robustness test compared to the results of the main regression indicate that the explanatory power of M-score is comparatively greater when regressed to the returns generated by the flagged companies' stocks for the same year.

5. Discussion

In this section the results of the empirical study are discussed through the framework of previous literature and the hypotheses previously outlined in this thesis.

5.1 Discussion: Hypothesis 1

Hypothesis 1: The M-score model has explanatory power over future stock price returns on the Swedish stock market.

The results of this paper find no strong empirical evidence that the M-score model has explanatory power when tested against excess returns on the Swedish stock exchange. With the primary regression model explaining merely 23% of the variance in excess returns and the mscore variable only being significant at the 5%-level, the M-score appears not to be a reliable predictor of excess returns on the Swedish stock exchange. The results did not find evidence to strengthen the first hypothesis of this thesis.

The beta value (-0.0308553) of the mscore variable displayed in the primary regression result is negative and thus consistent with the hypothesis. A higher M-score indicates a greater probability of manipulation which, in accordance with previous research, means the stock should generate lower excess returns. This negative correlation between M-score and future stock returns were however not strong enough to support the hypothesis.

There are several possible explanations to why the model has not performed as strongly as would be suggested by previous research. Firstly, the model could be country specific to the United States in the same manner as the Fama-French model. Or it may require the firm population examined to have properties similar to those of the US stock market population. The inconsistency of the results could stem from the characteristics of the Swedish stock market and its differences compared to the United States' market.

Another potential motivation for the poor fitness of the regression model can be found in Beneish et al. (2012) where the authors argue that broader knowledge of the M-score model can diminish its effectiveness. If a manipulating firm is aware of the model it could restrict the extent of manipulation to make sure the model will mischaracterize it as a non-manipulator. On this basis the authors argue M-score is likely to lose some of its predictive power in the future as it becomes more widely spread.

Lastly an explanation can be drawn from Kama and Melumad (2011) which found that new methods appeared to have been adopted by U.S. firms to conceal the effects of earnings manipulations. Of the methods introduced, factoring of receivables stands out as particularly detrimental to the functionality of the M-score model and it heavy reliance on identifying inflated earnings via accruals such as receivables. This highlights the perennial flaw of a static model eventually becoming outdated as it tries to operate in tandem with fluid and constantly developing accounting standards.

5.2 Discussion: Hypothesis 2

Hypothesis 2: The companies flagged as earnings manipulators by the M-score model will have lower excess stock returns compared to companies which are not flagged as manipulators.

The results of the study do not provide evidence that the excess stock return of companies classified as manipulators are statistically significantly lower compared to companies classified as non-manipulators. Results from the Welch t-test concludes that the average annual returns of portfolios containing all flagged manipulators is not statistically different from the annual average returns of portfolios consisting of non-manipulators.

This evidence does not strengthen the second hypothesis outlined in this paper. Even though one extreme outlier in excess return of the flagged portfolio was replaced, the t-test results did not support the hypothesis. This would appear to contradict previous research, however there are potential explanations which can be drawn directly from that previous research.

Beneish et al. (2012) concluded that one should expect the M-score model to suffer reduced efficacy over time. In the same study a decrease of the M-score model’s effectiveness in separating stocks generating low excess returns from those that generate high excess return is visible post 2002. The observed trend matches the results found in this thesis. It should also be noted that the sample issues as discussed in 3.6 Selection Bias and Other Sample Issues could have had meaningful impact on the t-test results. Without data restrictions it is possible the test would have yielded a different result.

Further, it is possible that the Beneish model does a poor job of correctly classifying businesses with extreme growth. If a business is growing very fast, several of the M-score factors will attain a value resulting in a high M-score value and by extension implying a high probability of earnings manipulation. We argue that the modern economic landscape lends itself more naturally to extreme growth firms compared to that same landscape before the digital developments of the 21st century. New industries and business models requiring relatively greater working capital investments or emerging industries where competitive forces are still in the process of pushing down profitability to margin equilibrium could also represent an adequate explanation of the poor efficacy.

7. Conclusion

The aim of this thesis was to study the relationship between earnings manipulation, as defined and identified by the M-score model, and excess returns of stocks listed on the Swedish stock exchange. The data used covered 126 companies between the years of 2005 and 2017.

The results of this study show that the M-score model for detecting earnings manipulation does not have significant explanatory power over excess returns on the Swedish stock exchange in the period 2005-2017. These results occur both when testing for stock returns in the same year as companies being flagged, and when testing for potential lagged effects in subsequent years. These results stand contrary to previous research by Beneish et al. (2012) concluding M-score had explanatory power over excess returns on American stocks. There are several possible explanations for this contradiction, but as noted by Beneish et al. (2012) the likely explanation stems from the ongoing evolution of more complicated financial manipulation schemes, and knowledge of the model becoming more common amongst manipulating firms. It is thus reasonable to assume that the using the M-score model to predict excess stock returns will become an even less efficient method in the future.

Further Research

The M-score model is based on a study conducted on U.S. companies. One possible are of further research is to replicate the study performed in Beneish (1999) on regional stock markets or on the global stock market to investigate whether there are model variations depending on the independent characteristics of the investigated markets.

Sweden has four smaller exchanges where equities are traded not classified as part of the official stock market, First North, Aktietorget, NGM and Nordic MTF. These exchanges are not as heavily regulated as the main exchange and therefore the ability of firms listed on these marketplaces to commit earnings manipulation is arguably greater due to less individual scrutiny. This creates room for a broadening of the research conducted in this this study to include these exchanges and contrasting the resulting efficacy of M-score with the results found in this paper.

Another area of interest is the whether any of the individual components of the M-score model, or any combination thereof, has greater explanatory power than the combined model. Because different components can offset each other resulting in a lower total M-score value, transmuting the model composition has the potential to better predict excess returns.

This study mainly focuses on quantitatively investigating the linkage between excess returns and earnings manipulation. Through the lens of behavioural finance, a potential area of research is to look at individual and institutional investor behaviour and reactions in connection to the revelation of earnings manipulation.

Lastly a potential area of interest concerns the fundamentals of the M-score model, the relative costs to investors of investing in companies publicly ousted as earnings manipulators. To determine the actual average relative costs could lead to the development of more precise earnings manipulation models useful in probabilistic approaches to quantitative investment strategies.

References

AQR library. Datasets.
Available at: https://www.aqr.com/Insights/Datasets
[Retrieved 26-03-2018]

Beneish, Messod D. 1999, “The Detection of Earnings Manipulation”
Financial Analysts Journal, Vol. 55, No. 5 (Sep. - Oct., 1999), pp. 24-36
URL: http://www.jstor.org/stable/4480190
[Retrieved 21-02-2018]

Beneish, Messod D., Lee, Charles M.C and Nichols D. Craig. 2012, “Fraud detection and expected returns”
URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1998387
[Retrieved 07-03-2018]

Dechow, P., Sloan, R., & Sweeney, A. (1995). “Detecting Earnings Management”. *The Accounting Review*, 70(2), 193-225.
URL: http://www.jstor.org/stable/248303
[Retrieved 27-02-2018]

Eugene F. Fama and Kenneth R. French, “The Cross-Section of Expected Stock Returns”
The Journal of Finance, Vol. 47, No. 2 (Jun., 1992), pp. 427-465
URL: http://www.jstor.org/stable/2329112
[Retrieved 19-02-2018]

Fama, Eugene F. “Efficient Capital Markets: A Review of Theory and Empirical Work.” *The Journal of Finance*, vol. 25, no. 2, 1970, pp. 383–417. JSTOR, JSTOR
ULR: http://www.jstor.org/stable/2325486
[Retrieved 19-02-2018]

Kama, Itay, Melumad, Nahum. (2011). “Camouflaged Earnings Management”
URL:http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.221.6429&rep=rep1&type=pdf
[Retrieved 25-04-2018]

Rosner, R. L. (2003). “Earnings manipulation in failing firms” *Contemporary Accounting Research*, 20(2), 361.
ULR: https://search.proquest.com/docview/194210575?accountid=39039
[Retrieved 27-03-2018]

The Riksbank. Swedish 10 year government bonds, annual data, date 2005-2017.
URL: https://www.riksbank.se/en-gb/statistics/search-interest--exchange-rates/
[Retrieved 27-02-2018]

Appendix

Appendix A - The calculation of, and rationale behind, the M-score components

I. $$DSRI = \left(\frac{(Net\ Receivables_t)}{Sales_t}\right) / \left(\frac{(Net\ Receivables_{t-1})}{Sales_t - 1}\right) \quad (17)$$

Days' sales in receivables index is an M-score component designed to measure if the level of receivables relative to sales are balanced between two years. A meaningful increase in days' sales in receivables could be the result to a firm trying to inflate sales thereby exaggerating earnings.

II. $$GMI = \left(\frac{(Sales_{t-1} - Cost\ of\ Goods\ Sold_{t-1})}{Sales_{t-1}}\right) / \left(\frac{(Sales_t - Cost\ of\ Goods\ Sold_t)}{Sales_t}\right) \quad (18)$$

Gross margin index is a variable that captures the change in gross margin between two years. Then the value is greater than 1 the margin is weakening signalling that the firm's prospects are poor. Firms with deteriorating fundamentals are assumed to be more prone to manipulation thus this metric captures the inclination to engage in earnings manipulation.

III. $$AQI = \left(1 - \left(\frac{Current\ Assets_t + PP\&E_t}{Sales_t}\right)\right) / \left(1 - \left(\frac{Current\ Assets_{t-1} + PP\&E_{t-1}}{Sales_{t-1}}\right)\right) \quad (19)$$

Asset quality index is a variable built to measure the proportion of assets on the balance sheet whose realization is uncertain to provide actual future benefits to the company. This measure is used to capture changes in cost deferrals. A value greater than one implies an increased tendency of capitalizing costs thus inflating current earnings.

IV. $$SGI = \frac{Sales_t}{Sales_{t-1}} \quad (20)$$

Sales growth index is a variable designed to capture sales growth between two years. Companies experiencing high growth are more likely and have greater incentives to engage in manipulation to reach earnings target.

V. $DEPI = \left(\frac{(Depreciation_{t-1})}{Depreciation_{t-1}+PPE_{t-1}}\right) / \left(\frac{(Depreciation_t)}{Depreciation_t+PPE_t}\right)$ (21)

Depreciation index is a ratio describing the changes in nominal depreciation between two years and a value above one could indicate the depreciation speed has been slowed. This can create artificial increases in income if the useful life of the assets is in fact shorter than the amortization necessitating future write downs.

VI. $SGAI = \left(\frac{SG\&A\ Expenses_t}{Sales_t}\right) / \left(\frac{SG\&A\ Expenses_{t-1}}{Sales_{t-1}}\right)$ (22)

Selling, general and administrative expenses index is a ratio of said expenses to sales between two years. Like the gross margin index its primary function in the model is capturing worsening fundamentals of businesses resulting in increased likelihood of manipulation.

VII. $LVGI = \left(\frac{(Long\ Term\ Debt_t+Current\ Liabilities_t)}{Total\ Assets_t}\right) / \left(\frac{(Long\ Term\ Debt_t+Current\ Liabilities_t)}{Total\ Assets_t}\right)$

(23)

The Leverage index is designed to capture the effect of increasing leverage. When a company becomes increasingly levered constraints such as covenants tightening leading firms to become more predisposed to manipulate earnings to improve their metrics and avoid breaking covenants.

VIII. $TATA = \frac{Income\ Befoe\ Extraordinary\ Items_t - Cash\ Flow\ From\ Operations_t}{Total\ Assets_t}$ (24)

Total accruals to total assets measures how well reported earnings coincide with real cash inflows. The rationale behind this variable is that firms with large or inflated accruals i.e. poor conversion of profits into actual cash has a higher probability of committing earnings manipulation.

Appendix B – Portfolio beta regression results for ER_{nf} and $adjER_f$

The dataset of Fama-French factor returns for the Swedish stock market was organized into monthly time intervals. To make this return data compliant with the annual total return data and annual M-score model data the following computations were used to transform the monthly return data into annual return data:

$$R_{SMB_{yt_i}} = \left(1 + R_{SMB_{mt_i}}\right) * \left(1 + R_{SMB_{mt-1_i}}\right)\left(1 + R_{SMB_{mt-2_i}}\right) \dots * (1 + R_{SMB_{mt-11_i}}) - 1 \quad (25)$$

$$R_{HML_{yt_i}} = \left(1 + R_{HML_{mt_i}}\right) * \left(1 + R_{HML_{mt-1_i}}\right) * \left(1 + R_{HML_{mt-2_i}}\right) \dots * (1 + R_{HML_{mt-11_i}}) - 1 \quad (26)$$

$$R_{MKT_{yt_i}} = (1 + R_{MKT_{mt_i}}) * (1 + R_{MKT_{mt-1_i}})(1 + R_{MKT_{mt-2_i}}) \dots * (1 + R_{MKT_{mt-11_i}}) - 1 \quad (27)$$

After the annual Fama-French factor return data was acquired the calculation of their beta values were conducted by regressing the three individual Fama-French factors to the returns of the portfolios minus the risk-free rate. Fama-French Beta regression formulas are presented below:

$$ER_{t_i} = \beta R_{SMB_{t_i}} * R_{SMB_{t_i}} + \varepsilon_i \quad (27)$$

$$ER_{t_i} = \beta R_{MKT_{t_i}} * R_{MKT_{t_i}} + \varepsilon_i \quad (27)$$

$$ER_{t_i} = \beta R_{HML_{t_i}} * R_{HML_{t_i}} + \varepsilon_i \quad (27)$$

Table 6: Fama-French Beta R_{MKT} Regression

Random-effects GLS regression	Number of obs = 1208
Group variable: companyid	Number of groups = 126
R-sq:	Obs per group:
within = 0.1013	min = 5
between = 0.0061	avg = 9.6
overall = 0.0869	max = 13
corr(u_i, X) = 0 (assumed)	Wald chi2(4) = 120.81
	Prob > chi2 = 0.0000

ER	Coefficient	Std. Err.	z	P>z	[95% Conf. Interval]	
R_{MKT}	.0089858	.0008175	10.99	0.000	.0073835	.0105881
Constant	.1105088	.0269538	4.10	0.215	.0576803	.1633373

sigma_u .1708313

sigma_e .70541979

rho .05539733

(fraction of variance due to u_i)

Note: ***, ** and * represent statistical significance at the 0,1 %, 1% and 5 % levels, respectively.

Table 7: Fama-French Beta R_{SMB} Regression

Random-effects GLS regression	Number of obs = 1208
Group variable: companyid	Number of groups = 126
R-sq:	Obs per group:
within = 0.1063	min = 5
between = 0.0200	avg = 9.6
overall = 0.0964	max = 13
corr(u_i, X) = 0 (assumed)	Wald chi2(4) = 132.07
	Prob > chi2 = 0.0000

ER	Coefficient	Std. Err.	z	P>z	[95% Conf. Interval]	
R_{SMB}	.0201723	.0017512	11.52	0.000	.0167401	.0236046
Constant	.2148841	.0252648	8.51	0.000	.1653659	.2644022

sigma_u .16822423

sigma_e .70345089

rho .05409499

(fraction of variance due to u_i)

Note: ***, ** and * represent statistical significance at the 0,1 %, 1% and 5 % levels, respectively.

Table 8: Fama-French Beta R_{HML} Regression

Random-effects GLS regression	Number of obs = 1208
Group variable: companyid	Number of groups = 126
R-sq:	Obs per group:
within = 0.0013	min = 5
between = 0.0282	avg = 9.6
overall = 0.0016	max = 13
corr(u_i, X) = 0 (assumed)	Wald chi2(4) = 1.82
	Prob > chi2 = 0.1779

ER	Coefficient	Std. Err.	z	P>z	[95% Conf. Interval]	
R_{HML}	.0033003	.0024497	1.35	0.178	-.001501	.0081017
Constant	.2065979	.0250620	8.24	0.000	.1574772	.2557186

sigma_u .14512025

sigma_e .74362431

rho .03668733

(fraction of variance due to u_i)

Note: ***, ** and * represent statistical significance at the 0,1 %, 1% and 5 % levels, respectively.

Appendix C - Regression tables detailing robustness test results

Table 9: Lagged effects one year

Random-effects GLS regression	Number of obs = 1082
Group variable: companyid	Number of groups = 126
R-sq:	Obs per group:
within = 0.2178	min = 4
between = 0.0559	avg = 8.6
overall = 0.1981	max = 12
corr(u_i, X) = 0 (assumed)	Wald chi2(4) = 273.24
	Prob > chi2 = 0.0000

ER	Coefficient	Std. Err.	z	P>z	[95% Conf. Interval]	
mscorelag1	-.0179186	.0137688	-1.30	0.193	-.0449049	.0090677
R_{SMB}	.0121304***	.00135	8.99	0.000	.0094845	.0147763
R_{MKT}	.0066226***	.0006136	10.79	0.000	.00542	.0078251
R_{HML}	-.001722	-.0016159	-0.99	0.325	-.0048303	.0015986
Constant	.0656936	.0372964	1.76	0.078	-.0074061	.1387932

sigma_u .09240779

sigma_e .46836571

rho .03746813

(fraction of variance due to u_i)

Note: The table show the results from the lagged effect regressions in the robustness test. Excess return (ER) is defined as the total annual stock return of the individual sample firms minus the risk-free return. mscorelag1 is the M-score value (see equation 4 for computation) for an individual company one year prior to when the excess return is generated ($mscore_{t-1}$). ***, ** and * represent statistical significance at the 0,1 %, 1% and 5 % levels, respectively.

Table 10: Lagged effects two years

Random-effects GLS regression	Number of obs = 956
Group variable: companyid	Number of groups = 126
R-sq:	Obs per group:
within = 0.2210	min = 3
between = 0.1471	avg = 7.6
overall = 0.2085	max = 11
corr(u_i, X) = 0 (assumed)	Wald chi2(4) = 254.05
	Prob > chi2 = 0.0000

ER	Coefficient	Std. Err.	z	P>z	[95% Conf. Interval]	
mscorelag2	-.0053243	.0143042	0.37	0.710	-.03336	.0227114
R_{SMB}	.0139912***	.0015068	9.29	0.000	.011038	.0169444
R_{MKT}	.0060964***	.000697	8.75	0.000	.0047302	.0074626
R_{HML}	-.0023199	.0017452	1.33	0.184	-.0057405	.0011007
Constant	.0860473*	.0383639	2.24	0.025	.0108553	.1612392

sigma_u .09757278

sigma_e .46967026

rho .04137335

(fraction of variance due to u_i)

Note: Excess return (ER) is defined as the total annual stock return of the individual sample firms minus the risk-free return. mscorelag2 is the M-score value (see equation 4 for computation) for an individual company two years prior to when the excess return is generated ($mscore_{t-2}$). ***, ** and * represent statistical significance at the 0,1 %, 1% and 5 % levels, respectively.

Table 11: Lagged effects three years

Random-effects GLS regression	Number of obs = 830
Group variable: companyid	Number of groups = 126
R-sq:	Obs per group:
within = 0.2397	min = 2
between = 0.0271	avg = 6.6
overall = 0.2044	max = 10
corr(u_i, X) = 0 (assumed)	Wald chi2(4) = 226.03
	Prob > chi2 = 0.0000

ER	Coefficient	Std. Err.	z	P>z	[95% Conf. Interval]	
mscorelag3	.0111217	.0155433	0.72	0.474	-.0193427	.041586
R_{SMB}	.0146661***	.0016128	9.09	0.000	.011505	.0178271
R_{MKT}	.0061246***	.000731	8.38	0.000	.0046919	.0075573
R_{HML}	-.0027664	.0018609	1.49	0.137	-.0064137	.000881
Constant	.1248666	.042385	2.95	0.003	.0417935	.2079397

sigma_u .14447709

sigma_e .46384877

rho .0884366

(fraction of variance due to u_i)

Note: Excess return (ER) is defined as the total annual stock return of the individual sample firms minus the risk-free return. mscorelag3 is the M-score value (see equation 4 for computation) for an individual company three years prior to when the excess return is generated ($mscore_{t-3}$). ***, ** and * represent statistical significance at the 0,1 %, 1% and 5 % levels, respectively.

Appendix D: Descriptive statistics for t-test variables ***ER_{nf}, ER_f*** and ***$adjER_f$***

Table 12: Descriptive summary statistics for ER_{nf}

Percentiles				
	Smallest		Obs	1090
1%	-.870675	-1.166042	Sum of Wgt.	1090
5%	-.5421775	-1.159658	Mean	.098686
10%	-.4014579	-1.060311	Std. Dev.	.5089747
25%	-.1873749	-1.051862	Variance	.2590553
50%	.0507254		Skewness	2.566787
	Largest		Kurtosis	19.89165
75%	.3219036	3.150784		
90%	.5791472	3.271222		
95%	.7698932	4.365471		
99%	2.047764	4.997435		

Note: The table presents descriptive statistics for ER_{nf} computed as annual stock returns of the sample firms not flagged as manipulators minus the risk-free rate and minus the Fama-French factor returns. The sample consists of 126 listed Swedish firms that operate outside the real estate and financial sectors and the observations are collected from the time period 2005-2017.

Table 13: Descriptive summary statistics for ER_f

Percentiles				
	Smallest		Obs	118
1%	-2.219806	-2.28625	Sum of Wgt.	118
5%	-1.75425	-2.219806	Mean	.1835842
10%	-1.453818	-2.065271	Std. Dev.	1.784103
25%	-.7570702	-1.793027	Variance	3.183024
50%	-.0403906		Skewness	4.661135
	Largest		Kurtosis	38.84963
75%	1.09352	2.601733		
90%	1.624055	2.65097		
95%	1.910959	3.691167		
99%	3.691167	14.75596		

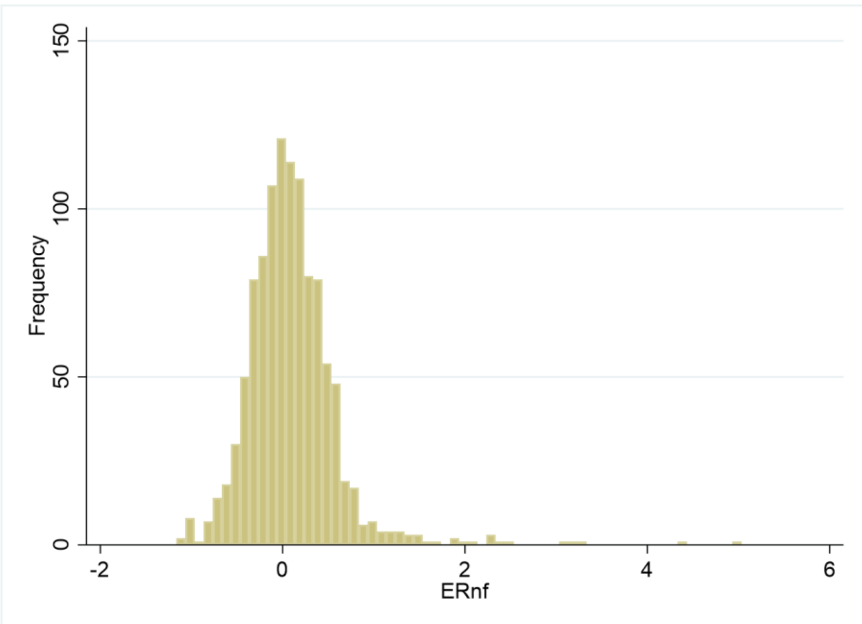
Note: The table presents descriptive statistics for ER_{nf} computed as annual stock returns of the sample firms not flagged as manipulators minus the risk-free rate and minus the Fama-French factor returns. The sample consists of 126 listed Swedish firms that operate outside the real estate and financial sectors and the observations are collected from the time period 2005-2017.

Table 14: Descriptive summary statistics for *adjER$_f$*

Percentiles					
	Smallest			Obs	118
1%	-2.219806	-2.28625		Sum of Wgt.	118
5%	-1.75425	-2.219806		Mean	.0898148
10%	-1.453818	-2.065271		Std. Dev.	1.210091
25%	-.7570702	-1.793027		Variance	1.464319
50%	-.0403906			Skewness	.3912367
	Largest			Kurtosis	2.977014
75%	1.09352	2.601733			
90%	1.624055	2.65097			
95%	1.910959	3.691167			
99%	3.691167	3.691167			

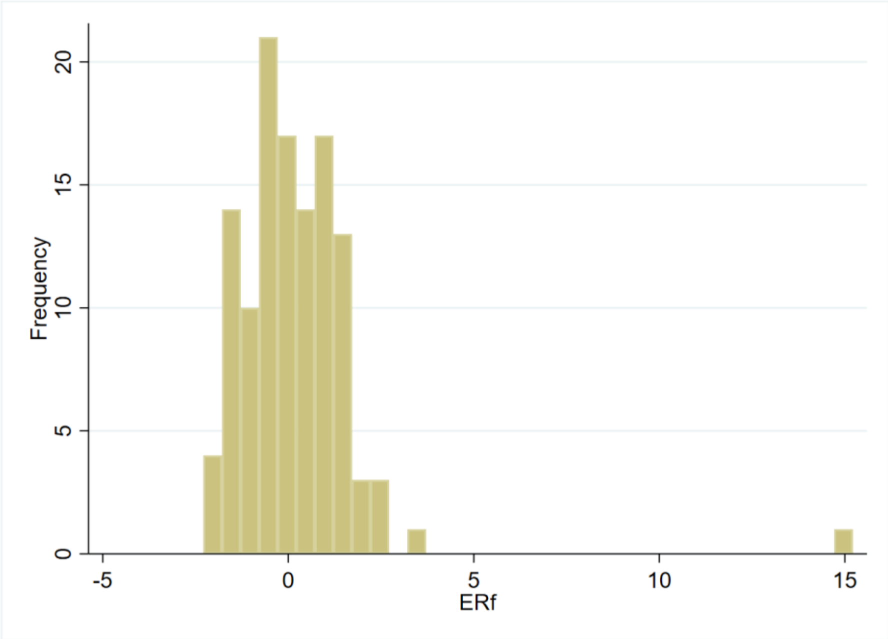
Note: The table presents descriptive statistics for *ER$_{nf}$* computed as annual stock returns of the sample firms not flagged as manipulators minus the risk-free rate and minus the Fama-French factor returns. The sample consists of 126 listed Swedish firms that operate outside the real estate and financial sectors and the observations are collected from the time period 2005-2017. Values in the upper 1% bound are replaced with the lowest value in that percentile.

Figure 6: Excess returns of stocks in the non-flagged portfolio



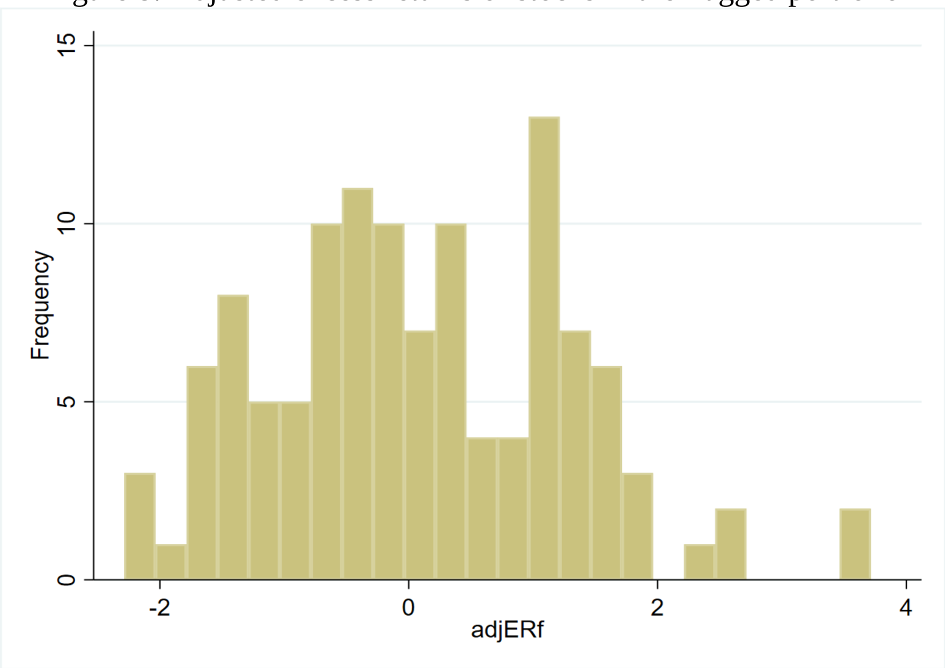
Note: This figure presents a histogram over the variable *ER$_{nf}$*. *ER$_{nf}$* is computed as annual stock returns of the sample firms not flagged as manipulators minus the risk-free rate and minus the Fama-French factor returns.

Figure 7: Excess returns of stocks in the flagged portfolio



Note: This figure presents a histogram over the variable ER_f. ER_f is computed as annual stock returns of the sample firms flagged as manipulators minus the risk-free rate and minus the Fama-French factor returns.

Figure 8: Adjusted excess returns of stocks in the flagged portfolio



Note: This figure presents a histogram over the variable $adjER_f$. $adjER_f$ is computed as annual stock returns of the sample firms flagged as manipulators minus the risk-free rate and minus the Fama-French factor returns. The variable has been adjusted by replacing values outside the upper 1% bound with the lowest value within the same percentile.