EUROPEAN CENTRAL BANK

EUROSYSTEM

Occasional Paper Series

Lamia Allali, Nicolas Dierick, Alessandro Santoni

Earnings manipulation and probability of default: insights from AnaCredit and supervisory implications

No 385

Contents

Abstract

This article provides a novel insight into whether earnings manipulation signals are reflected in banks' internal credit risk estimates, as measured by the probability of default (PD) estimates, and whether such manipulation has an impact on credit risk (point in time or deferred). The hypothesis is that firms engaging in manipulation may be exposed to increased credit risk over time, which should be reflected in higher PD values.

Using AnaCredit – a granular dataset covering credit exposures from European banks between 2019 and 2022 – and financial statement data from Orbis, we constructed a sample of 4,649 publicly traded corporations, for which we computed the Beneish M-Scores that are used to detect potential earnings manipulation. This allowed us to determine the interrelation with PDs. Our results reveal a weak and negative correlation between M-Scores and PDs, suggesting that earnings manipulation signals are not fully absorbed by banks' internal models. Further analysis shows that these results are driven by the high prevalence of firms with no earnings manipulation signals. Firms for which the M-Score effectively indicates potential earnings manipulation (8.9% of the sample) are observed to have higher PDs, which also increase further as the M-Score worsens. These findings support the hypothesis that earnings manipulation signals are not fully reflected in credit risk estimates over time, indicating that their impact – when it occurs – is deferred instead of being captured immediately in internal models. Our results indicate that the relationship between potential earnings manipulation and banks' internal credit risk estimates is highly context-dependent and non-linear.

Cross-sectional analyses by country and industry show consistent patterns linking default risk to M-Scores in selected countries and sectors. Our results suggest there is a potential need for internal credit risk estimates to consider qualitative overrides and expert judgement, if earnings manipulation signals are not adequately factored into banks' internal credit risk estimates. To capture both default risk and the reliability of financial reporting, we introduced a manipulation-adjusted credit risk indicator, denoted as PD^+, which measures the extent to which manipulation risk (captured by probability of fraud (PF)) amplifies default risk.

JEL codes: G32, M41, M42, C33

Keywords: probability of default, earnings manipulation, Beneish M-Score, credit risk modelling, financial misreporting, AnaCredit

1 Executive summary and introduction

1.1 Executive Summary

This study analyses the extent to which banks recognise signs of earnings manipulation when estimating the default risk of publicly traded firms. The analysis compares an indicator of potentially unreliable financial reporting with bank's internal probability of default assessments. A distinct pattern emerges among the small subset that exceed the manipulation threshold. These firms are associated with higher estimated default risk, and their probabilities of default rise further as the indicator of earnings manipulation intensifies. Across countries and industries, elevated credit risk and stronger manipulation signals tend to cluster in specific markets and sectors. Overall, the evidence suggests that bank's internal credit risk models do respond to signs of earnings manipulation primarily among companies for which the M-Score already indicates potential earnings manipulation. These findings provide valuable insights for understanding how financial reporting quality interacts with credit risk assessment and highlight the relevance of monitoring manipulation indicators when evaluating firms' risk profile.

1.2 Introduction

Over the past few decades, a growing body of research has cultivated the fields of default risk estimation and financial statement manipulation. Earnings manipulation remains a critical concern, as it may distort reported financial performances and, consequently, banks' internal risk measures. The Beneish M-Score (Beneish, 1999; Matsumoto, 2002; Qiang & Terry, 2005) is widely used to detect potential earnings manipulation. Limited prior studies have examined the relationship between earnings manipulation and banks' internal credit risk models, highlighting potential challenges that banks face in capturing earnings manipulation signals within their internal credit risk models (Beaver et al., 2005).

This paper investigates whether banks' internal credit risk estimates, as represented by their PDs, are responsive to the potential risk of earnings manipulation. The hypothesis is that firms engaging in earnings manipulation may experience increased credit risk over time, which should be reflected in higher PD values. This assumption is based on the notion that earnings manipulation may mask underlying financial weaknesses that are not immediately detectable. Over time, these hidden deficiencies can impair a firm's ability to meet its financial obligations, thereby increasing the likelihood of default. Even so, stable PD values despite signs of earnings manipulation could highlight that internal credit risk estimates, which use quantitative or qualitative inputs from financial statements, are being determined on the basis of less reliable inputs.

Our analysis leverages on two complementary datasets: AnaCredit, which provides detailed credit exposure and granular credit data across the European Union, and

Orbis (Bureau Van Dijk), which offers detailed financial statements of listed companies. By interlinking these datasets, we constructed a sample of 4,659 publicly traded corporations for the period between 2019 and 2022.[1] We computed the Beneish M-Score directly from the raw financial data, rather than relying on pre-computed values. The M-Score computation includes the full set of ratios proposed by Beneish,[2] ensuring robust detection of potential earnings manipulation. Our methodology aligns with prior studies based on Compustat,[3] the most recurrent dataset used by previous studies. Importantly, the analysis places particular emphasis on firms whose financial statements exhibit potential earnings manipulation signals. It specifically focuses on the economic relevance of the relationship between earnings manipulation and credit risk, by concentrating on a restricted subsample of firms that exceed the Beneish manipulation threshold.

Our results indicate that the relationship between potential earnings manipulation and banks' internal credit risk estimates is context-dependent and non-linear. Within our dataset, only 8.9% of firms exceeded the Beneish M-Score. These firms exhibit higher default probabilities that rise further as manipulation indicators worsen beyond the threshold. In contrast, when considering the full sample, the M-Score displayed a negative association with banks' estimated probabilities of default, suggesting that higher manipulation indicators coincide with lower assessed credit risk.

Further analysis at country, industry and enterprise size level also identified clusters of countries and corporations associated with elevated PDs and M-Scores. Portugal, Luxembourg and Greece, industries such as construction and real estate, financial and insurance, as well as small and micro enterprises, showed the highest estimated default risk within the sample. These were also observed to be associated with a higher likelihood of earnings manipulation.

[1] This timeframe aligns with the implementation of AnaCredit, thus reflecting the availability of data from that year onwards.

[2] Days' Sales in Receivables Index (DSRI), Gross Margin Index (GMI), Asset Quality Index (AQI), Sales Growth Index (SGI), Depreciation Index (DEPI), Sales, general and Administrative Expenses Index (SGAI), Leverage Index (LEVI), and Total Accruals to Total Assets (TATA)

[3] In his analysis, Beneish considered this database, which includes detailed financial information on publicly traded firms in the United States for which data were available for the 1982-1992 period.

2 Literature

The present article contributes to existing theories, by relying on banks' PD estimates derived from the AnaCredit dataset to explore the potential relationship between earnings manipulation and internal credit risk assessments. The study bridges two traditionally distinct research domains; credit risk modelling and detection of earnings manipulation.

Credit risk modelling has evolved substantially, with foundational approaches such as structural default models and the introduction of key concepts like distance to default (DtD) (Vasicek, 1984; Merton, 1974), which remain central to risk quantification. Altman's (1968) discriminant model was among the earliest attempts to link accounting ratios to default predictions, providing the groundwork for later credit scoring methodologies. Survival analysis techniques, including the Weibull distribution (Weibull, 1951), are widely applied in hazard rate models to predict time-to-default. In parallel, market-based indicators such as credit default swap (CDS) spreads offer real-time insights into default risk. This is in addition to their capacity to capture market expectations of credit events, thus offering a dynamic complement to static balance sheet metrics (Duffie & Singleton, 1999; Cont, 2001). Moreover, regression-based approaches, particularly logistic regression, are commonly employed in PD modelling due to their interpretability and robustness (Altman & Saunders, 1998; Beaver, 1966).

Earnings manipulation and its implication for credit risk, poses a substantial challenge. Since PD models typically rely on historical financial data to estimate default probabilities, any distortion in reported earnings could lead to misclassification of a firm's risk profile. The Beneish M-Score (1999) – a probabilistic model designed to detect earnings manipulation among firms exhibiting inflated financial performance – identified 76% of earnings manipulation cases in the period between 1982 and 1992. Later studies, such as Vladu, et al. (2016), confirmed its effectiveness across jurisdictions, although adaptations were needed to account for regulatory differences. This assumption is based on the idea that earnings manipulation can hide real financial problems that are not visible in standard financial reports. These hidden issues can weaken a firm's financial position over time, making it harder to meet debt obligations and increasing the risk of default (Rosner, 2003; Dechow, et al. 2011; Penman, 2001).

Further research (Beneish, et al. 2013) reveals that firms flagged by the M-Score tend to underperform. This suggests that manipulation indicators also hold valuable information about expected returns and risk pricing. Nonetheless, research on how such qualitative fraud signals are integrated into banks' internal credit risk models remains limited. Beaver et al. (2005) emphasise the disconnection between financial statement analysis and credit risk modelling, partly explaining why manipulation signals are often underrepresented in PD estimates. This study seeks to explore the extent to which banks' PD models reflect the risks associated with earnings manipulation, offering new insights into the intersection between financial reporting integrity and risk assessment.

3 Data

Our study uses accounting data from Orbis to construct the M-Score, which serves as our measure of potential earnings manipulation. In particular, we leverage on Orbis' detailed accounting statements for listed companies, covering balance sheet, P&L and cash flow statistics. These provide us with an available universe of 24,875 companies.[4] We took accounting data from Orbis for the period between 2018 and 2022 to construct the indices required for the M-Score, which relies on accounting figures across two financial years. This reduces the overall universe of publicly traded companies from 24,875 to 8,298. Subsequently, we identified the detailed list of metrics needed to compute the M-Score (see Annex 1). This reduced our sample size to 4,659 publicly traded companies with sufficient accounting data.

We leverage on AnaCredit data to measure the interrelation between banks' probability of default and earnings manipulation. AnaCredit is an analytic credit dataset containing detailed information on individual bank loans in the euro area, harmonised across all Member States. Most importantly, this database includes information on PDs for different companies that have been granted loans by the banks in question. This metric serves as our main measure for estimating bank default risk and is defined as the counterparty's probability of default over one year, determined in accordance with Articles 160, 163, 179 and 180 of the Capital Requirements Regulation (EU) No 575/2013.

Our sample period covers the period from 2019 to 2022. This allows us to account for the data collection process in AnaCredit, which did not start until September 2018, as well as the second edition of the AnaCredit manual, which was published in May 2019. The latter changes to the manual incorporated various additional explanations provided through Q&As conducted between 2017 and 2018.

The Orbis and AnaCredit datasets are interlinked via each publicly traded corporation's legal entity identifier (LEI) code and reference period. Among the 4,659 companies with sufficient accounting data, 1,489 are also observed within AnaCredit,[5] of which 1,349 include information on the banks' PD for those firms. The final dataset includes 9,681 observations across 1,349 firms, four financial years and 51 creditors with loans and PDs attributable to those firms. This paragraph can be summarised in Table 2 below.

Table 1
Data handling

Category	Number of observations
Listed corporation universe with detailed balance sheets, P&L statements and LEI codes	24,875
Of which: has a year-end closing date between 2019 and 2022	8,298

[4] Orbis provides data in dollars or euro. In our case preference was given to the euro currency.

[5] Regarding the AnaCredit dataset, only data on loans to publicly traded corporations are reported and only on loans with an exposure of more than €25 thousand. See also ECB Regulation on AnaCredit.

Of which: has sufficient accounting data to calculate the M-Score	4,659
Of which: has observations in AnaCredit based on LEI code matching	1,489
Of which: has observations in AnaCredit on PD	1,349

Notes: Table 1 summarises the progressive filtering of the listed corporation universe based on data availability and matching criteria, from initial balance sheet and P&L coverage to AnaCredit observations on PD.

Table 2
Final dataset

Category	Number of observations
Number of observations	9,681
Number of unique debtor LEI codes	1,349
Number of unique creditor LEI codes	51
Number of unique closing dates	4

Notes: Table 2 provides an overview of the final dataset composition, including the total number of observations and of unique identifiers for debtors, creditors and closing dates.

Our main focus will be on European company data, as our final dataset primarily comprises observations across Italian (28.8%), German (26.8%), Spanish (14.1%) and French companies (11.6%). With regard to their size, 6,829 observations concern large enterprises (70.5%), with medium (13.9%), small (9.7%) and micro enterprises (4.30%) representing a less significant part of the sample population. Their business operations are primarily oriented towards professional, scientific and technical (33.7%), manufacturing (31.2%), information and communication (10.0%), and financial and insurance (6.04%) activities. Finally, our dataset mainly covers loans granted by significant institutions (SIs)[6] (98%) rather than less significant institutions (LSIs) (2%). Based on these descriptive statistics, our sample exhibits firm size and sectoral characteristics comparable to those typically observed in the Compustat database, which focuses predominantly on large, publicly traded firms in the manufacturing and service industries (Beneish, 1999; Dechow et al, 2011).

Listed companies within our dataset can generally be considered large firms, with median total assets of €1,115 million, median sales of €727 million and a median market value of €631 million. From a liquidity perspective, these firms show a moderate capacity to meet short-term obligations, as indicated by a median working capital-to-total-assets ratio of 0.09 and a median current ratio of 1.31. Their median total-debt-to-total-assets ratio of 0.63 suggests a relatively high reliance on debt financing. In terms of profitability, firms generate on average a 3% return on assets (ROA), while the median sales growth rate of 1.07 reflects modest but positive growth performance (see Table 3).

[6] Currently, the significance criteria is based on: size (i.e. total value of assets higher than €30 billion or considered as one of the three largest banks in a country), economic importance (i.e. either for a specific country or the EU economy as a whole), cross-border activities (i.e. total value of assets higher than €5 billion and ratio of cross-border assets/liabilities, in more than one country, to total assets/liabilities higher than 20%) and public financial assistance.

Table 3

Summary statistics

Characteristics	P1	P25	P50	P75	P99
Size (expressed in EUR millions)					
Total assets	9	209	1,115	6,257	264,917
Sales	2	137	727	3,789	149,419
Market value	4	105	631	3,600	76,428
Liquidity / leverage					
Working capital to total assets	-0.36	0.01	0.09	0.22	0.58
Current ratio	0.37	1.02	1.31	1.80	5.06
Total debt to total assets	0.22	0.52	0.63	0.75	1.18
Profitability / growth					
Return on assets	-32%	0%	3%	6%	23%
Sales growth	0.40	0.97	1.07	1.21	2.69

Notes: Table 3 reports descriptive statistics on listed companies' characteristics, while the percentiles analysis is useful in determining thresholds for identifying outliers within the dataset.

4 Methodology

To investigate the contemporary relationship between a firm's likelihood of engaging in earnings manipulation and the bank's PD estimate, we specified the following empirical model:

$$PD_{i,j,t} = \alpha + \beta \cdot MScore_{i,t} + x'_{i,j,t} \cdot \zeta + \varepsilon_{i,j,t} \tag{1}$$

In equation (1), $PD_{i,j,t}$ denotes the bank's PD for bank j in year t. Our primary explanatory variable of interest is the $MScore_{i,t}$. It captures a firm's earnings manipulation risk using the Beneish (1999) $Mscore_{i,t}$ framework, which is measured on the basis of eight different indices using financial statement data (see Annex 2):

$$MScore_{i,t} = -4.84 + 0.92 \cdot DSRI_{i,t} + 0.528 \cdot GMI_{i,t} + 0.404 \cdot AQI_{i,t} + 0.892 \cdot SGI_{i,t} + 0.115 \cdot DEPI_{i,t} - 0.172 \cdot SGAI_{i,t} + 4.679 \cdot TATA_{i,t} - 0.327 \cdot LVGI_{i,t} \tag{2}$$

The value of the M-Score is then associated with a potential likelihood for earnings manipulation. If firm i were to receive a score greater than -1.78, it would signal a higher likelihood that it may be manipulating its earnings.

The vector $x'_{i,j,t}$ reflects an additional list of control variables that may influence the estimated dependency between the $MScore_{i,t}$ and the $IRB\ PD_{i,j,t}$. These controls include:

- bank fixed effects, to account for differences in business model, portfolio composition or internal rating methodologies
- industry fixed effects, to capture sector-specific risk dynamics
- year fixed effects, to control for macroeconomic and regulatory changes over time

We use ordinary least squares (OLS) to estimate the model and account for a one-way clustering of standard errors at the firm level. This is to allow for potential autocorrelation and heteroskedasticity in repeated observations.

5 Empirical results

5.1 Contemporary analysis

Following on from the descriptive statistics outlined in Table 3 from the previous section, we present summary statistics on the Beneish M-Score and PD estimates in Table 4. In line with previous studies, we observe that the proportion of firms with an M-Score exceeding the -1.78 threshold amounts to 8.9% of all observations. This shows that potential earnings manipulation are an uncommon event within the broader dataset. Beneish's (1999) model identified approximately 13% of the holdout sample as manipulators, with later studies confirming the M-Score's effectiveness across jurisdictions (Vladu, et al. 2016). By comparison, Özkán and Alfarhan (2025) also found that approximately 13% of firm-year observations across G7 countries are classified as potential manipulators using the Beneish M-Score.

Table 4 further highlights the observed distribution for the M-Score to be non-normal with high skewness, kurtosis and presence of outliers. To avoid a potential bias in subsequent statistical testing procedures, we exclude outliers identified as observations with an M-Score below the 1st and above the 99th percentile. Finally, within Table 4 we also observe a segment of the sample population to already be classified in default, as reflected by the PD equalling 100%. These observations will be excluded in subsequent analyses, due to such obligors no longer being effectively assessed with banks' rating models.

Table 4

Summary statistics on M-Score and PD

Characteristics	No.	P1	P25	P50	P75	P99
M-Score	9,681	-4.62	-2.87	-2.60	-2.28	2.46
2019	2,379	-5.06	-2.93	-2.71	-2.47	2.90
2020	2,497	-5.17	-3.09	-2.80	-2.53	3.13
2021	2,438	-4.09	-2.71	-2.46	-2.14	1.72
2022	2,367	-4.30	-2.68	-2.41	-2.12	1.88
PD	9,681	0.00%	0.16%	0.35%	1.08%	100.00%
2019	2,379	0.00%	0.13%	0.33%	1.10%	100.00%
2020	2,497	0.00%	0.16%	0.38%	1.14%	100.00%
2021	2,438	0.03%	0.17%	0.37%	1.10%	100.00%
2022	2,367	0.00%	0.15%	0.33%	0.88%	100.00%

Notes: Table 4 reports descriptive statistics on the M-Score and banks' PD. The percentile analysis supports the identification of outliers by establishing thresholds values.

In Table 5, we present the outcome of pairwise correlation tests between the Beneish M-Score and the PDs within the entire sample and by year of observation. This analysis reveals the correlation to be low and negative, at -2.33% across the entire sample data, which is statistically significant at the 5% significance level. This implies that, in practice, firms with a higher likelihood of earnings manipulation, as represented by an increase in the M-Score, are associated with lower PD estimates.

Table 5
Pairwise correlation between M-Score and PD

M-Score	Correlation with PD	No.
Full sample	-2.33%**	9,188
2019 sample	-0.00%	2,232
2020 sample	-2.23%	2,345
2021 sample	-1.32%	2,336
2022 sample	-6.45%***	2,275

Notes: Table 5 presents the pairwise correlation between the M-Score and the PD for the reference sample, which excludes defaulted obligors and outliers. Coefficients marked with *** and ** indicate significance at the 1% and 5% level, respectively.

These preliminary results are further corroborated by the outcome of our regression model shown below in Table 6. The model estimates the contemporary relationship between earnings manipulation and banks' default risk estimation, as outlined in equation (**1**). The results indicate that firms with a higher likelihood of earnings manipulation (i.e. higher M-Score) tend to have lower PD estimates. Conceptually, this suggests that banks' internal models may not fully capture the potential risk of earnings manipulation within financial statements. In some cases, internal models may even inadvertently interpret manipulated financial statements as a signal of stronger credit quality. Economically, a one-unit increase in the M-Score (i.e. higher likelihood of earnings manipulation) corresponds to a 31 basis point decrease in the PD, while a one standard deviation change in the M-Score (0.65) is associated with a 20 basis point decrease in the PD.

Table 6
Baseline model

	(1)	(2)	(3)	(4)
M-Score	-0.16%	-0.20%*	-0.26%**	-0.31%***
Control variables				
Bank fixed effects	No	Yes	Yes	Yes
Sector fixed effects	No	No	Yes	Yes
Year fixed effects	No	No	No	Yes
No.	9,188	9,188	9,188	9,188
R2	0.05%	3.24%	5.11%	5.18%

Notes: Table 6 reports the baseline OLS regression results, indicating the coefficients associated with the predictor variable M-Score and its respective impacts on the PD of bank j for firm i in year t. Coefficients marked with ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Nevertheless, our baseline results may be affected by the fact that most of the sample data is not characterised by any risk of potential earnings manipulation (i.e. 91.1% of all observations have an M-score below or equal to -1.78). Consequently, we explore in Table 7 the dependency of the pairwise correlation between M-Score and PDs, depending on the distance to the M-Score threshold signifying potential earnings manipulation. These results reveal a substantive increase in the correlation from -2.33% to 12.78% within the subsample of firms with likely earnings manipulation (i.e. an M-Score value above -1.78). We also extend our baseline model specification in Table 8. Introducing a dummy variable that equals 1 if the M-Score exceeds -1.78, and an interaction effect therewith, allows for a different intercept and slope coefficient among likely earnings manipulators. This shows that firms for which financial statements indicate potential earnings manipulation are associated with a statistically significant higher PD. Further increases are also depicted in the PD as the M-Score worsens beyond the threshold value of -1.78. In economic terms, a one-unit increase in the M-Score corresponds to a 43 basis point increase in the PD, whereas a one standard deviation change in the M-Score among manipulators (0.85) is associated with a 37 basis point increase in the PD. This pattern indicates that the relationship between the likelihood of earnings manipulation and banks' PDs is non-linear, emerging only among firms displaying stronger manipulation signals. Banks' internal credit risk estimates can therefore be considered responsive to the potential risk of earnings manipulation, rather than being subject to the risk of misleading financial statements where PDs signal better creditworthiness.

Table 7

Pairwise correlation between M-Score and PD by subsample

M-Score	Correlation with PD	Average PD	Average M-Score	No.
Full sample	-2.33%**	1.30%	-2.53	9,188
> -3.00	4.06%***	1.17%	-2.37	7,712
> -2.75	5.60%***	1.17%	-2.24	6,016
> -2.50	5.91%***	1.24%	-2.00	3,716
> -2.25	5.64%***	1.36%	-1.70	2,103
> -2.00	6.99%**	1.44%	-1.34	1,153
> -1.78	12.78%***	1.45%	-1.01	724

Notes: Table 7 presents the pairwise correlation between the M-Score and the PD for the reference sample, as well as subsamples for which the M-Score exceeds the specified threshold value. Coefficients marked with ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table 8

Extended baseline model

	(1)	(2)	(3)	(4)
M-Score	-0.72%***	-0.72%***	-0.75%***	-0.90%***
Manipulator	2.70%***	2.52%***	2.35%***	2.68%***
M-Score x manipulator	1.33%***	1.29%***	1.19%***	1.33%***
Control variables				
Bank fixed effects	No	Yes	Yes	Yes
Sector fixed effects	No	No	Yes	Yes
Year fixed effects	No	No	No	Yes
No.	9,188	9,188	9,188	9,188
R2	0.59%	4.22%	5.51%	5.68%

Notes: Table 8 reports the extended baseline OLS regression results, indicating the coefficients associated with the predictor variable M-Score and its respective impacts on the PD of bank j for firm i in year t. The "Manipulator" variable represents a dummy variable that equals 1 if the M-score exceeds -1.78. Coefficients marked with ∗∗∗, ∗∗ and ∗ indicate significance at the 1%, 5% and 10% level, respectively.

As a robustness check, we also estimated an alternative specification using the M-Score from the previous period to account for a delay in banks' ability to incorporate financial statement data only upon the effective receipt thereof. The results, which remained broadly consistent, are presented in Table 9,.

Table 9
Extended baseline model with lagged M-Score

	(1)	(2)	(3)	(4)
M-Score t-1	-0.7%	-0.7%	-0.8%	-0.9%
Manipulator	2.5%	2.5%	2.4%	2.7%
M-Score t-1 x manipulator	0.7%	1.3%	1.2%	1.3%
Control variables				
Bank fixed effects	No	Yes	Yes	Yes
Sector fixed effects	No	No	Yes	Yes
Year fixed effects	No	No	No	Yes
No.	5,483	5,483	5,483	5,483
R2	0.59%	3.71%	5.51%	5.68%

Notes: Table 8 reports the extended baseline OLS regression results, indicating the coefficients associated with the predictor variable M-Score and its respective impact on the PD of bank j for firm i in year t. The variable M-score variable has been included with a time lag of one year. Coefficients marked with ∗∗∗, ∗∗ and ∗ indicate significance at the 1%, 5% and 10% level, respectively.

5.2 Predictive analysis

We also explore the ability of the M-Score to predict default events within the next 12 months by estimating a linear probability model of the following form:

$$\begin{aligned} Def_{i,j,t} &= \alpha + \beta \cdot M\ Score_{i,t-1} + \gamma \cdot Manipulator_{t-1} \\ &+ \delta \cdot M\ Score_{i,t-1} \cdot Manipulator_{t-1} + \mathrm{x}'_{i,j,t} \cdot \zeta + \varepsilon_{i,j,t} \end{aligned} \quad \textbf{(3)}$$

where the dependent variable reflects a binary indicator variable set equal to one if firm i is classified as defaulted in year t by bank j, and 0 if otherwise. This predictive analysis is motivated by the need to evaluate whether firms exhibiting higher levels of earnings manipulation, as captured by the Beneish M-Score, actually display higher default risk in practice. By using the lagged M-Score, we also continue to reflect the scenario in which banks assign ratings based on the most recently available financial statements, rather than contemporaneous data. Estimation results are presented in Table 9.

Table 10

Default prediction

	(1)	(2)	(3)	(4)
M-Score t-1	-1.8%***	-1.8%***	-1.9%***	-2.1%***
Manipulator	4.6%***	4.4%***	4.3%***	4.9%***
M-Score t-1 x manipulator	1.7%***	1.6%***	1.6%***	1.8%***
Control variables				
Bank fixed effects	No	Yes	Yes	Yes
Sector fixed effects	No	No	Yes	Yes
Year fixed effects	No	No	No	Yes
No.	5,484	5,484	5,484	5,484
R2	0.77%	1.61%	2.27%	2.51%

Notes: Table 9 displays the results of the default prediction model where the target variable is probability of default of bank j for firm i in year t. The predictors in the model include M-Score along with control variables. The table provides the estimated coefficients for each predictor accompanied by their statistical significance levels. Coefficients marked with ∗∗∗, ∗∗ and ∗ indicate significance at the 1%, 5% and 10% level, respectively.

The results reinforce the previous evidence of a relationship between the Beneish M-Score and the likelihood of a firm defaulting within the next 12 months. Across all four model specifications, the lagged M-Score coefficient is consistently negative and significant at the 1% level, ranging from -1.8% to -2.1%. This indicates that, in our predictive framework, firms with a higher M-Score in the preceding year are associated with a lower observed default occurrence in the subsequent year. However, the inclusion of a differential slope coefficient for likely earnings manipulators highlights this observed relationship to be predominantly driven by firms with an M-Score below -1.78. Instead, for likely earnings manipulators the partial effect through the M-score nets out, implying likely earnings manipulators to be associated with a higher likelihood of default, due to the positive differential intercept. Nevertheless, it should be noted that the estimation results in this scenario should be considered sensitive to the number of observed default occurrences. More specifically, we recorded only 38 defaults to have occurred within our analysis (0.70%) across 5,484 firm-bank-year observations. By including bank, sector and year fixed effects, this progressively improves the model’s explanatory power (R^2 from 0.36% to 2.30%), even though the overall fit remains modest.

5.3 Country analysis

The country-level breakdown in Table 10 reveals three distinct clusters based on exposure-weighted average PD and M-Score values. These offer insights into countries’ economic characteristics and potential earnings manipulation.

Table 11

Country analysis

Country	No.	Average M-Score	Weighted PD
Italy	2,613	-2.47	0.55%
Germany	2,537	-2.56	0.42%

Spain	1,228	-2.43	0.75%
France	1,064	-2.66	1.05%
Finland	268	-2.60	0.56%
Austria	207	-2.54	0.25%
Netherlands (the)	193	-2.58	0.36%
Other	191	-2.51	0.98%
Switzerland	135	-2.61	0.12%
Belgium	116	-2.45	1.09%
Portugal	105	-2.56	3.81%
Luxembourg	79	-2.46	3.97%
Sweden	57	-2.63	0.54%
Greece	50	-2.50	2.92%
Lithuania	50	-2.66	0.49%
Denmark	47	-2.43	0.19%
Estonia	46	-2.81	0.66%
Norway	45	-2.49	0.67%
Poland	40	-2.49	0.94%
United Kingdom of Great Britain and Northern Ireland (the)	39	-2.59	1.71%
Russian Federation (the)	37	-2.44	1.13%
Canada	21	-2.43	1.53%
Peru	20	-2.31	1.64%

Notes: Table 10 reports the averages of the M-Score and PD, grouped by our dataset's country, over the period 2019 to 2022.

First, several countries, such as Portugal (3.81%), Luxembourg (3.97%) and Greece (2.92%), show higher estimated default risk among the loans in our sample, accompanied by M-Scores (around -2.46 to -2.56). This pattern suggests weaker financial health among borrowers in these markets. In contrast, countries such as Italy (0.55%), Germany (0.42%), Finland (0.56%), Austria (0.25%) and Sweden (0.54%) exhibit lower estimated default risk, consistent with more stable borrower profiles and limited signals of potential manipulation. The aggregated "Other" category also displays a relatively elevated PD (0.98%) together with a moderately negative M-Score (-2.51). This group consists of all countries in the dataset with fewer than 20 observations, which were combined into a single category to ensure statistical robustness and avoid the reporting of potentially unstable country-level estimates based on very small samples.

5.4 Industry analysis

The industry-level breakdown in Table 11 illustrates three distinct clusters based on exposure-weighted average PD and M-Score values. This offers insights into sector-specific profile and potential earnings manipulation.

Table 12
Industry analysis

Industry	No.	Average M-Score	Weighted PD

Professional, scientific and technical activities	3,106	-2.53	0.58%
Manufacturing	,2895	-2.55	0.57%
Information and communication	904	-2.60	0.53%
Financial and insurance activities	567	-2.62	1.52%
Wholesale and retail trade, motor vehicle and motorcycle repair	487	-2.46	0.59%
Transportation and storage	279	-2.56	0.56%
Construction	210	-2.19	1.37%
Electricity, gas, steam and air conditioning supply	204	-2.27	0.46%
Administrative and support service activities	112	-2.37	0.37%
Human health and social work activities	94	-2.48	0.92%
Mining and quarrying	91	-2.38	0.95%
Real estate activities	77	-2.15	1.91%
Accommodation and food service activities	65	-2.81	1.72%
Other	39	-2.45	2.12%
Water supply, sewerage, waste management and remediation activities	38	-1.97	0.21%
Other service activities	20	-3.01	0.66%

Notes: Table 11 summarises the averages of the M-Score and PDs, grouped by our dataset's industry, over the period 2019 to 2022

Several industries, such as real estate (1.91%), financial and insurance (1.52%), and construction (1.37%), display the highest exposure-weighted PDs in our sample, accompanied by M-Scores between -2.15 and -2.62. This pattern suggests weaker financial conditions among firms operating in these sectors and, in some cases, a higher likelihood of earnings manipulation. In contrast, sectors such as transportation and storage (0.56%), manufacturing (0.57%), information and communication (0.53%), and water supply, sewerage, waste management and remediation activities (0.21%) exhibit lower estimated default risk together with M-Scores ranging from approximately -2.0 to -2.5, consistent with more stable borrower profiles and similar signals of potential manipulation. The aggregated Other category displays a relatively elevated PD (2.12%) alongside a moderately negative M-Score (-2.45). This group consists of industries with fewer than 20 observations, which were combined into a single category to ensure statistical robustness and to avoid reporting potentially unstable industry-level estimates based on very small samples.

5.5 Enterprise size analysis

The analysis of enterprise size in Table 12 highlights notable differences in both default risk and potential earnings manipulation.

Table 13

Enterprise size analysis

Enterprise size	No.	Average M-Score	Weighted PD
Large enterprises	6,564	-2.53	0.57%
Medium enterprise	1,262	-2.50	0.75%
Small enterprise	847	-2.52	1.51%
Micro enterprise	376	-2.57	1.61%
Other	139	-2.60	1.67%

Notes: Table 12 summarises the averages of the M-Score and PD, grouped by our dataset's enterprise size over the period 2019 to 2022

Small and micro enterprises exhibit the highest estimated default risk in the sample, with exposure-weighted PDs of 1.51% and 1.61%, respectively. They are accompanied by average M-Scores between -2.52 for small enterprises and -2.57 for micro enterprises. This pattern suggests relatively weaker financial conditions among smaller firms and a limited susceptibility to earnings manipulation. Medium enterprises display a moderate level of estimated default risk, with a weighted PD of 0.75% and an average M-Score of -2.50. Large enterprises have the lowest estimated default risk of 0.57%, which together with a similarly negative M-Score of -2.53 is consistent with more stable borrower profiles. The aggregated "Other" category exhibits a weighted PD of 1.67% alongside a strongly negative average M-Score (-2.60). This group includes firms that could not be clearly classified in the standard enterprise size categories and is retained as a separate category to ensure completeness of the dataset while avoiding a potential misclassification bias.

6 Main findings & conclusion

In this study, we investigated whether signals of earnings manipulation, as captured by the Beneish M-Score, are reflected in banks' internal credit risk estimates, as measured by banks' PD estimates, and whether such signals influence short-term credit risk. For this purpose, we relied on accounting data from Orbis over the period from 2019 to 2022 to calculate the M-Scores directly. We took this approach rather than using pre-existing or harmonised datasets to differentiate our study. Subsequently, we combined the Orbis dataset with AnaCredit data to retrieve information on banks' internal default estimates.

Our results indicate that the relationship between potential earnings manipulation and banks' internal credit risk estimates is highly context-dependent and non-linear. Across the full sample, we observe a negative correlation between the M-Score and banks' PDs, suggesting a higher indicator for potential earnings manipulation to be paradoxically associated with a lower estimated default risk. Only a small share of firms (8.9%) within our dataset effectively breaches the Beneish threshold, thereby confirming signs of potential earnings manipulation to be uncommon, such that the results are driven by firms not characterised by any risk of potential earnings manipulation. Further subsample analyses reveal banks' internal credit risk estimates to be effectively responsive. More specifically, we observe higher PDs among companies for which the M-Score indicates potential earnings manipulation. At the same time, we also see further increases in the PDs when the M-Score worsens beyond the threshold value. Based on our empirical results, we see that a one standard deviation change in the M-Score among manipulators is associated with a 37 basis point increase in the PD. Furthermore, these results are shown to be robust when accounting for a time lag in the availability of financial statements for usage within internal rating models and corroborated by an analysis of the effective default incidence within the following 12 months.

At the aggregate country and industry levels, these dynamics are reflected in patterns of financial health and potential earnings manipulation. Countries such as Portugal, Luxembourg and Greece show higher PDs, while Germany, Finland, Austria and Sweden, on the other hand, display lower PDs and similar negative M-Scores, which is consistent with more robust borrower profiles. Across industries, sectors such as real estate, financial and insurance activities, and construction exhibit both elevated PDs and similar M-Scores. Conversely, sectors such as transportation and manufacturing show more stable credit risk profiles with similar M-Scores and lower PDs.

In summary, the overarching negative relationship in the full sample reflects the dominance of baseline firms, with manipulation risk effects emerging only within a smaller subset of likely manipulators. Our findings emphasise that banks' PD models, which often rely on financial ratios and quantitative indicators, are responsive to signals of earnings manipulation indices. In this context, the Beneish M-Score may be interpreted as a proxy for PF, reflecting the likelihood that a firm's reported earnings may be affected by manipulation. Our results also imply that the value of the Beneish M-Score can serve as a potentially useful indicator for banks to identify possible earnings manipulation within firms' financial statements. Finally, our results further suggest the potential need for banks' internal credit risk estimates to consider qualitative overrides and expert judgement, when signals of earnings manipulation may not be adequately factored into these estimates.

References

Altman, E. I., & Saunders, A. (1998), “Credit risk measurement: Developments over the last 20 years”, *Journal of Banking and Finance*, 21(11-12), 1721-1742.

Beaver, W. H., McNichols, M. F. & Rhie, J. W. (2005), “Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy”, *Review of Accounting Studies*, 10(1), 93-122.

Beneish, M. D. (1997), “Detecting GAAP violation: implications for assessing earnings management among firms with extreme financial performance”, *Journal of Accounting and Public Policy*, 16(3), 271-309.

Beneish, M. D., (1999), “Incentives and Penalties Related to Earnings Overstatements That Violate GAAP”, *The Accounting Review*, 74(4), 425-457.

Beneish, M. D. (1999, “The detection of earnings manipulation”, *Financial Analysts* Journal, 55(5), 24-36.

Beneish, M. D. (2001), “Earnings management: A perspective”, *Managerial Finance*, 27(12), 3-17.

Beneish, M. D., Lee, C. M. & Nichols, D. C. (2013), “Earnings manipulation and expected returns”, *Financial Analysts Journal*, 69(2), 57-82.

Bhattacharya, U., Daouk, H. & Welker, M. (2003), “The world price of earnings opacity”, *The Accounting Review*, 78(3), 641-678. Cont, R. (2001), “Empirical properties of asset returns: stylized facts and statistical issues’’, Quantitative finance, 1(2), 223.

Dechow, P. M., Sloan, R. G., & Sweeney, A. P. (1995), “Detecting earnings management”, *The Accounting review*, 193-225.

Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R.G. (2011), “Predicting Material Accounting Misstatements”, *Contemporary Accounting Research*, 28(1), 17-82.

Dechow, P., Hutton, A., Kim, J. H., & Sloan, R. G. (2012), “Detecting earnings management: A new approach”, *Journal of Accounting Research*, 50(2), 275-334.

Duffie, D., & Singleton, K. J. (1999), “Modeling term structures of defaultable bonds”, *The Review of Financial Studies*, 12(4), 687-720.

European Central Bank. (2018), AnaCredit - Analytical Credit Dataset.

Healy, P.M., & Palepu, K.G. (2003), “The Fall of Enron”, *Journal of Economic Perspectives*, 17(2), 3-26.

Jones, J.J. (1991), “Earnings Management During Import Relief Investigations”, *Journal of Accounting Research*, 29(2), 193-228.

Matsumoto, D. A. (2002), “Management's incentives to avoid negative earnings surprises”, *The Accounting Review*, 77(3), 483-514.

Merton, R. C. (1974), “On the Pricing of Corporate Debt: The Risk Structure of Interest Rates”, *The Journal of Finance*, 29(2), 449-470.

Nelson, M.W., Elliott, J.A. & Tarpley, R.L. (2002), “Evidence from Auditors about Managers’ and Auditors’ Earnings Management Decisions”, *The Accounting Review*, 77(s-1), 175-202.

Nigrini, M. (2012), “Benford’s Law: Applications for Forensic Accounting, Auditing, and Fraud Detection”, Wiley.

Özkán, F., & Alfarhan, A. (2025), “Earnings manipulation and cash holdings: a Beneish M-score analysis in G7 nations”, *Cogent Business & Management*, 12(1), 2502542.

Penman, S.H. (2001), “Financial Statement Analysis and Security Valuation”, McGraw-Hill/Irwin.

Cheng, Q., Warfield, T.D. (2005), “Equity Incentives and Earnings Management”, *The Accounting Review*, 80(2), 441-476.

Rosner, R. L. (2003), “Earnings manipulation in failing firms”, *Contemporary Accounting Research*, 20(2), 361-408.

Vasicek, O. (1984), Credit Valuation, KMV Corporation.

Vladu, A. B., Amat, O., & Cuzdriorean, D. D., (2017), “Truthfulness in Accounting: How to Discriminate Accounting Manipulators from Non-Manipulators”, *Journal of Business Ethics*, 140(4), 633-648.

Weibull, W. (1951), “A Statistical Distribution Function of Wide Applicability”, *Journal of Applied Mechanics*.

Annexes

Annex 1

Beneish M-Score ratio analyses

Formula	Rationale	Symptoms of potential accounting manipulation
DSRI = (net receivables t / sales t) / (net receivables t-1 / sales t-1)	DSRI measures how quickly a company collects its accounts receivable.	A high DSRI may indicate aggressive revenue recognition, where sales are recorded prematurely to inflate reported earnings.
GMI = [(sales t-1 – cost of goods sold t-1) / Sales t-1] / [(sales t – cost of goods sold t) / sales t]	GMI measures a company's gross margin at two different time points to detect any abnormal changes that might indicate manipulation.	A low GMI may suggest declining profitability, potentially due to aggressive revenue recognition or cost manipulation.
AQI = [1 – (current assets t + PPE t / total assets t)] / [1 – (current assets t-1 + PPE t-1 / total assets-1)]	AQI assesses the quality of a company's assets, focusing on the proportion of non-current assets.	A high AQI may indicate aggressive capitalisation of expenses or overvaluation of assets, inflating reported earnings.
SGI = sales t / sales t-1	SGI measures sales growth compared to the previous year.	A high SGI combined with low or negative cash flow may suggest aggressive revenue recognition to mask deteriorating financial performance.
DEPI = [depreciation t-1 / depreciationt-1 + PPE t-1] / [depreciation t / depreciation t + PPE t]	DEPI measures the ratio of the rate of depreciation in year t-1 to the corresponding rate in year t.	A low DEPI may indicate slower-than-expected depreciation, potentially inflating reported earnings.
SGAI = [sales, general and administrative expenses t / sales t] / [sales, general and administrative expenses t-1 / sales t-1]	SGAI compares the growth rate of selling, general and administrative expenses to sales growth.	Rapid SGAI growth relative to sales growth may indicate cost manipulation or inefficient expense management.
TATA = total accruals / total assets t	TATA ratio measures the proportion of accruals to total assets, providing insights into the quality of reported earnings.	It may suggest a larger portion of earnings being derived from accruals rather than cash transactions. It indicates likelihood of earnings manipulation or aggressive revenue recognition.
LEVI = [LTD t + current liabilities t / total assets t] / [LTDt-1 + current liabilities t-1 / total assets t-1]	LVGI assesses the company's leverage (debt) relative to its assets.	High leverage combined with poor profitability may signal financial distress, potentially leading to earnings manipulation to maintain appearances.

Source: Beneish (1997, 1999)

Annex 2

M-Score's index definitions

Variable	Definition
DSRI	(netaccountsreceivable[t] / netsales[t]) / (netaccountsreceivable[t-1] / netsales[t-1])
GMI	(1 - (-1 * costofgoodssold[t-1]) / netsales[t-1]) / (1 - (-1 * costofgoodssold[t]) / netsales[t])
AQI	(1 - (netpropertyplantequipment[t]+ totalcurrentassets[t]) / totalassets[t]) / (1 - (netpropertyplantequipment[t-1] + totalcurrentassets[t-1]) / totalassets[t-1])
SGI	(netsales[t]) / (netsales[t-1])
DEPI	((-1 * depreciation[t-1]) / ((-1 * depreciation[t-1]) + netpropertyplantequipment[t-1])) / ((-1 * depreciation[t]) / ((-1 * depreciation[t]) + netpropertyplantequipment[t]))
SGAI	((-1 * researchdevelopmentexpenses[t]) / netsales[t]) / ((-1 * researchdevelopmentexpenses[t-1]) / netsales[t-1])
TATA	(earningsaftertax[t] - netcashfromoperatingactivities[t]) / (totalassets[t])
LEVI	((noncurrentliabilities[t] + totalcurrentliabilities[t]) / totalassets[t]) / ((noncurrentliabilities[t-1] + totalcurrentliabilities[t-1]) / totalassets[t-1])

Source: Beneish (1997, 1999)

Acknowledgements

Alessandro Santoni
European Central Bank, Frankfurt am Main, Germany; email: alessandro.santoni@ecb.europa.eu

Nicolas Dierick
European Central Bank, Frankfurt am Main, Germany; email: nicolas.dierick@ecb.europa.eu

Lamia Allali
European Central Bank, Frankfurt am Main, Germany; email: lamia.allali@ecb.europa.eu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

This paper can be downloaded without charge from the ECB website, from the Social Science Research Network electronic library or from RePEc: Research Papers in Economics. Information on all of the papers published in the ECB Occasional Paper Series can be found on the ECB's website.

PDF ISBN 978-92-899-7637-4, ISSN 1725-6534, doi:10.2866/3145034, QB-01-26-025-EN-N