European Journal of Operational Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Bankruptcy prediction with fractional polynomial transformation of financial ratios

Zenon Taoushianis

Department of Banking and Finance, Southampton Business School, University of Southampton, SO17 1BJ, UK

ARTICLE INFO

Keywords:
Risk analysis
Fractional polynomials
Financial ratios
Bankruptcy prediction

ABSTRACT

We show that simple nonlinear transformations of financial ratios, within a multivariate fractional polynomial approach, yield substantial improvements in bankruptcy prediction. The approach selects optimal power functions balancing parsimony and complexity. Focusing on a dataset comprising of non-financial firms, we develop a parsimonious nonlinear logit model with minimal parameter specification and clear interpretability, outperforming linear logit models. The model improves the in-sample fit, while out-of-sample it significantly reduces costly misclassification errors and improves discriminatory power. Similar insights are obtained when applying fractional polynomials on a secondary dataset consisting of banking firms. Interestingly, the fractional polynomial model compares favourably with other nonlinear models. By simulating a competitive loan market, we demonstrate that the bank using the fractional polynomial model builds a higher-quality loan portfolio, resulting in superior risk-adjusted profitability compared to banks employing alternative models.

1. Introduction

The vast majority of academic studies employ logit models that assume a linear relationship between financial predictors and the risk of bankruptcy (see, for instance, Shumway, 2001; Chava & Jarrow, 2004; Altman & Sabato, 2007; Campbell et al., 2008; Tinoco & Wilson, 2013; Tian et al., 2015; Traczynski, 2017; Gupta & Chaudhry, 2019, among many others). However, linear logit models have several shortcomings when applied to bankruptcy prediction. First, in a stepwise logit framework used to select the optimal set of bankruptcy predictors, the exclusion of predictors may occur not because the predictors are truly insignificant, but because their relationship with bankruptcy is nonlinear, which is not captured within a linear stepwise selection process. This means potentially important variables might be missing in the final model, leading to an incomplete understanding of the true economic factors driving bankruptcy. As we show later, a nonlinear logit model constructed using the fractional polynomial approach does not include identical variables as those of a stepwise logit model. Second, linear logit models oversimplify the complexities inherent in financial data, where relationships between variables and outcomes are often nonlinear. Third, linear logit models can be poorly fitted to the data, resulting in lower predictive power. Finally, other techniques commonly used to model nonlinearities, such as splines or neural networks, are subject to well-known shortcomings, including difficulty in interpretability, prone to overfitting, complexity in their configuration and computation etc.

To overcome the limitations discussed above, we construct a nonlinear logit model using fractional polynomials—a family of integer and non-integer power functions, including quadradic, square root, logarithmic, cubic terms and other fractional functions, among others, that allow to transform predictors sequentially to model nonlinear relationships between inputs and outputs within a variable selection process (e.g., Royston & Altman, 1994; Sauerbrei & Royston, 1999). There are reasons to expect that financial ratios have a nonlinear association with the risk of bankruptcy. For instance, the impact of liquidity on bankruptcy risk may vary significantly at different levels, an effect that linear logit models cannot adequately capture. It is reasonable to expect that a reduction in liquidity would have different impact on bankruptcy risk depending on the initial liquidity position. A decrease in liquidity from an already high level is likely to have a minor effect on bankruptcy risk, as the firm may still have sufficient reserves to manage its obligations. Conversely, a similar-sized reduction from an already low liquidity position could significantly increase the risk of bankruptcy, as the firm may be left with limited resources to cover short-term liabilities. Indeed, as we show in subsequent univariate analysis, the bankruptcy frequency is significantly higher when liquidity is reduced from low to even lower levels, compared to higher levels, suggesting that a nonlinear relationship exists.

E-mail address: z.taoushianis@soton.ac.uk.

https://doi.org/10.1016/j.ejor.2025.07.036
Received 23 September 2024; Accepted 16 July 2025
Available online 19 July 2025

Unlike traditional stepwise logit regression, which assumes linear relationships, fractional polynomials adapt to the curvature of the data by selecting optimal power transformations for each predictor. In a multivariate fractional polynomial regression, all predictors start as linear functions in the logit model. The model then sequentially identifies the best transformation for each predictor, considering interactions with others. This process continues over multiple cycles, with each cycle refining the transformations based on the current best fit (i.e., while selecting the optimal transformation for each predictor, the logit regression includes other predictors with their optimal transformations determined up to that point). The selection process concludes when transformations stabilize across two consecutive cycles. The approach evaluates each predictor based on three criteria, using standard statistical tests in the following order:

Inclusion: Assessing whether the predictor adds predictive power to the model,

Linearity: Assessing whether a nonlinear transformation for the predictor is better than linearity,

Complexity: Assessing whether a higher-degree polynomial transformation for the predictor adds more predictive power than a lower-degree polynomial.

By applying this selection process iteratively across all predictors over multiple cycles, we obtain a model that not only includes the most important predictors but also embeds optimal nonlinear transformations. This iterative approach ensures that the final logistic model achieves the minimum deviance or, in other words, the best possible fit to the data which is the ultimate purpose. Effectively fractional polynomials generalize the stepwise logit regression because the linear relationship is just one of many possible transformations that can be applied to each predictor. Fractional polynomials explore a variety of power transformations, including both linear and nonlinear forms, to find the most appropriate functional relationship. As the model evaluates each predictor, it simultaneously tests for both its inclusion in the model and the optimal transformation that best fits the data, thus the approach is suitable for variable selection as well, allowing the construction of a parsimonious nonlinear logit model for bankruptcy prediction.

Earlier studies have considered splines (Giordani et al., 2014), or generalized additive models (GAMs) with splines (Djeundje and Crook, 2019b; Lohmann et al., 2023) or machine learning techniques such as artificial neural networks (Kumar & Ravi, 2007; Jones, 2017; Jones et al., 2017; Mai et al., 2019; Manthoulis et al., 2020; Petropoulos et al., 2020) to account for nonlinear effects in bankruptcy prediction. However, the proposed fractional polynomial approach has several important advantages. One of the key advantages of fractional polynomials is their interpretability. When a predictor is included in the final model, its optimal transformation is usually a one- or two-degree polynomial that is straightforward to understand. Additionally, each transformed variable is associated with a coefficient in the logit regression, making the relationship between the predictor and bankruptcy easy to interpret and communicate. In contrast, splines become difficult to interpret as the number of knots increases or when higher-order polynomials are used in each segment. The resulting piecewise functions can be complex and challenging to communicate, while segmentation in a fractional polynomial approach is not required. Neural networks, especially deep learning models, are often considered "black boxes" due to their complexity and the difficulty in interpreting the weights and interactions between variables. Another advantage of fractional polynomials is their simplicity and flexibility. Unlike splines and neural networks, which depend on critical parameters, fractional polynomials require minimal parameter specification. For splines, choosing the number and location of knots is crucial, while for neural networks, key parameters include the type of activation functions and the number of neurons in the hidden layers. As we demonstrate later, the performance of neural networks can vary significantly based on the number of neurons used. Moreover, splines and neural networks are prone to overfitting, particularly when using a large number of knots or neurons. Fractional polynomials, however, are less susceptible to overfitting due to their simpler structure. Finally, fractional polynomials offer the ability to account for interactions among predictors during the modelling process, which is particularly useful in a variable selection context. This is difficult to achieve with splines due to the complexity of specifying multiple knots. Similarly, the complex structure and black box nature of neural networks makes them less suitable for variable selection. In contrast, the fractional polynomial approach includes a step that tests the inclusion of each variable while the model includes the optimal transformations of the other predictors chosen up to that point. This approach effectively accounts for interactions without adding significant computational complexity or reducing efficiency.

To the best of our knowledge, our study is the first to introduce nonlinear logit models based on the multivariate fractional polynomial approach in the context of bankruptcy prediction, addressing the limitations of other nonlinear models discussed above. To demonstrate the advantages of this methodology, we conduct a large-scale and systematic empirical analysis using two extensive datasets consisting of financial and non-financial firms.

Overall, we document substantial improvements, in-sample and out-of-sample, when we apply multivariate fractional polynomials in bankruptcy prediction. We begin our analysis using a large dataset with quarterly non-financial firm observations, in the period December 1979 – September 2023, from which >2000 failed in the subsequent quarter. Applying fractional polynomials on a pool of financial ratios, we construct a parsimonious logit model with few predictors that is comparable to a stepwise logit model, however, all variables exhibit a nonlinear association with the risk of bankruptcy, as indicated by the optimal power functions selected by the fractional polynomial approach. Even more important is that the nonlinear logit model we construct, improves the in-sample fit and discriminatory power, measured by the Pseudo R^2 and the Area Under Curve (AUC), respectively, by approximately 60 % and 8 %, respectively. The improvement is striking considering that the stepwise and the nonlinear fractional polynomial models are comparable in terms of the variables they have.

Next, by considering a comprehensive out-of-sample analysis, we find that the nonlinear fractional polynomial logit model continues to outperform linear logit models substantially. Specifically, we find that the model concentrates 74 % of the failed firms in the highest risk deciles (compared to approximately 50 % achieved by the linear logit models), improves the AUC by >10 %, while reducing Type II errors (misclassifying failed firms) by nearly 50 % compared to linear logit models, highlighting the economic value of the nonlinear model in reducing costly misclassification errors.

We extend the above insights by applying the fractional polynomial approach on a large quarterly dataset consisting of FDIC-insured bank observations in the period June 1976 – September 2023, from which >2000 banks failed in the subsequent quarter. We document that standard CAMELS-based variables measuring capital adequacy, asset quality, earnings, and liquidity, exhibit a nonlinear relationship with bank risk. Importantly, depending on the forecasting horizon, the in-sample R^2 improves by up to 19 % while the in-sample AUC improves moderately. However, the out-of-sample AUC improves by up to 9 %.

Finally, we evaluate the performance of the fractional polynomial model, developed for both non-financial and financial firms, against other nonlinear benchmarks, specifically to a feedforward artificial neural network, a recurrent neural network, splines, and a generalized partially linear single index model. Our results indicate that the fractional polynomial logit models perform better than the benchmarks with the results being substantially more pronounced using the non-financial firms sample. In addition, unlike neural networks which show volatile performance depending on neuron configuration, the fractional polynomial model offers stable and reliable predictions. By simulating a competitive loan market, we show that the bank using the fractional polynomial model in its decision-making has economic gains as the bank

manages a higher-quality loan portfolio with higher risk-adjusted profitability compared to banks using alternative models. The results, overall, are consistent when considering longer-term horizons.

The rest of the paper is organized as follows. Section 2 discusses the data, Section 3 describes the fractional polynomial methodology, Section 4 discusses the results and Section 5 concludes the paper.

2. Data

2.1. Sample

Quarterly financial data for non-financial firms are downloaded from Compustat in the period December 1979 – September 2023 (accounting also for a three-month reporting delay) and they are matched with a failure indicator that we construct, that takes values of 1 for firms that failed in the subsequent quarter and 0 otherwise.

Failure information, such as the name, date, and failure reason, are downloaded from BankruptcyData of New Generation Research Inc, which is a database that contains detailed information for distressed companies in the U.S.[1] Failed firms are defined as firms that filed for Chapter 7 (liquidation) or Chapter 11 (administration). On the other hand, non-failed firms are those that have survived until the end of the sample period or exited the sample for other reasons, such as merger, acquisition etc.

Overall, we have a sample with 1,262,355 firm-quarter observations in the period December 1979 – September 2023 and failure information for the next quarter. In the dataset, we have identified 2298 failed firms. In Table IA.1 of the Internet Appendix, we report information for the number of observations in the sample period.

To get a full understanding of the failure sample and its variation over time, in Fig. 1 below, we plot the number of failed firms in each quarter over the sample period.

As can be seen, the number of failed firms spike around the 1990 – 1991 U.S. recession period, in the early 2000 – 2001 period during the dot.com bubble, in the 2008 – 2009 period during the global financial crisis and in the 2020 due the arrival COVID-19 pandemic. Overall, our sample with failed firms is representative of the prevailing market conditions. Additionally, in Table IA.2 of the Internet Appendix we show the distribution of failed firms per industry.

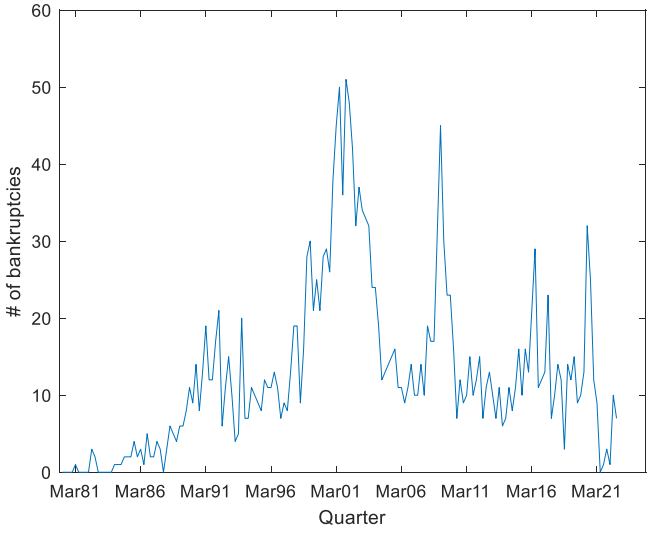

Fig. 1. This figure shows the fluctuation of the number of bankruptcies each quarter in the years 1980 – 2023.

For profitability, we construct the ratio of net income to total assets (NI/TA). For liquidity, we construct three financial ratios: cash and short-term equivalents to current liabilities (CH/CL), current assets to current liabilities (CA/CL) and working capital to total assets (WC/TA). For leverage, we construct the ratios of total liabilities to total assets (TL/TA) and short-term debt to total assets (STD/TA). For coverage, we construct the ratio of net income to interest expense (NI/IE), for capital we construct the ratio of equity to total assets (EQ/TA) and finally, the natural logarithm of total assets is used for size (SIZE). The set of financial predictors that we construct are commonly used in the literature and are expected to capture the risk of failure quite accurately (e.g., Shumway, 2001; Chava & Jarrow, 2004; Altman & Sabato, 2007; Tian et al., 2015, and many others). Finally, the short list of candidate financial predictors that we maintain allows us to build a parsimonious model for prediction that focuses on the most critical variables. This approach enhances the model's robustness and reliability, minimizes the risk of overfitting, and simplifies practical application as these variables are available for almost all firms.

The summary statistics reported in Table IA.3 of the Internet Appendix, reveal distinct differences between failed and non-failed firms. Failed firms exhibit lower capital levels, as indicated by a lower EQ/TA ratio, lower liquidity, evidenced by lower CA/CL, CH/CL, and WC/TA ratios. Additionally, failed firms are less profitable, with a lower NI/TA ratio, and have higher leverage, shown by higher TL/TA and STD/TA ratios.

3. Methodology

3.1. Variable transformations

The logit function (i.e. log-odds), $G(\beta, x)$, used in standard logistic regression, is a linear function of the predictor variables and has the following form:

$$\ln\left(\frac{p}{1-p}\right) = G(\beta, x) = \beta_0 + \sum_{i=1}^{k} \beta_i x_k \tag{1}$$

where p is the probability of failure, x is the vector of predictors and β the vector of coefficients. A more general and flexible approach is when allowing the relationship between the logit function and the predictors to be nonlinear, without of course excluding the possibility of linearity. Fractional polynomials fit, for the i *th* predictor variable, polynomial functions, F, up to a certain degree, J, to transform the predictor variable nonlinearly; $F_1(x_i), F_2(x_i), \ldots, F_J(x_i)$, leading to the following form of the logit function (with other predictors included but for the simplicity of exposition, we do not include them in the following equation):

$$G(\beta, x) = \beta_0 + \sum_{j=1}^{J} \beta_{i,j} F_j(x_i) + \ldots \tag{2}$$

In this context, we consider fitting one-degree polynomial functions ($J = 1$) and two-degree polynomial functions ($J = 2$). When $J = 1$, the function F is a power function from the set $p = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$, where 0 denotes the natural logarithm of the variable. The nonlinear transformation is then defined as:

$$\mathrm{F_1(x_i)} = \mathrm{x_i^{p_1}} \tag{3}$$

When $J = 2$, two polynomial functions are fitted: $F_1(x_i)$ as in Eq. (3) and $F_2(x_i)$ is defined as follows:

$$\mathrm{F_2(x_i)} \begin{cases} \mathrm{x_i^{p_2}} & \mathrm{p_2 \neq p_1} \\ \mathrm{F_1(x_i)ln(x_i)} & \mathrm{p_2 = p_1} \end{cases} \tag{4}$$

For example, when $p_1 = -1$ and $p_2 = 2$, $F_1(x_i) = \frac{1}{x_i}$ and $F_2(x_i) = x_i^2$ thus the logit function that includes the i *th* transformed predictor is:

[1] https://newgenerationresearch.com/.

$$G(\beta, x) = \beta_0 + \beta_{i,1}\frac{1}{x_i} + \beta_{i,2}x_i^2 + \ldots \quad (5)$$

In another example, suppose that $p_1 = p_2 = 3$. In this case, $F_1(x_i) = x_i^3$ and $F_2(x_i) = x_i^3\ln(x_i)$. Hence, the logit function is:

$$G(\beta, x) = \beta_0 + \beta_{i,1}x_i^3 + \beta_{i,2}x_i^3\ln(x_i) + \ldots \quad (6)$$

These transformations allow the logistic regression model to capture nonlinear relationships between predictor variables and the logit function, thus providing a more flexible and more accurate representation of the data, potentially yielding a better-fitted model.

Finally, we do not consider more than two-degree polynomials. The multivariate fractional polynomial approach is designed to balance model flexibility with parsimony. Higher-degree polynomials introduce excessive complexity, increasing the risk of overfitting while making the resulting model difficult to interpret. The principles of inclusion, linearity, and complexity (discussed in Section 3.2 subsequently) guide the selection process, ensuring that the chosen functional forms provide an optimal trade-off between goodness-of-fit and interpretability. In addition, considering up to two-degree polynomial functions is reasonable because we capture non-linearities while offering some degree of generalization that is important in out-of-sample predictions, which turn out to be true in our empirical analysis.

3.2. Optimal polynomial function selection

Optimality is defined as minimizing the deviance (D) of the final logistic model (−2*Log-Likelihood), by selecting for each predictor the best nonlinear transformation function as discussed in Section 3.1, aiming to include the most important predictors (e.g. parsimonious model) and maintaining the complexity of the model at low levels. To achieve optimality, we establish an iteration process for each predictor on a given cycle to identify the transformation that significantly reduces the Deviance of the model, following the principles of Inclusion, Linearity and Complexity as shown in Fig. 2.

More specifically, the following process is applied to each predictor during a given cycle:

Initial Step: The best one-degree polynomial function corresponds to the model with the lowest deviance, $D(J = 1)$. The best two-degree polynomial function corresponds to the model with the lowest deviance, $D(J = 2)$. Then, the optimal polynomial function for each predictor is chosen as follows:

Inclusion: We first assess whether including a given predictor improves the model fit. This is determined by comparing the null model (without the predictor) to the best two-degree polynomial model. The gain statistic, G(null, $J = 2$) = D(null) – D($J = 2$), follows a Chi-square distribution with 4 degrees of freedom. If statistically significant at the 5 % level, the predictor is retained; excluded otherwise.

Linearity: If included, we test whether the predictor exhibits a nonlinear relationship with the logit function. This is done by comparing the linear model with the best two-degree polynomial model. The gain statistic, G(1, $J = 2$) = $D(1)$ – D($J = 2$), follows a Chi-square distribution with 3 degrees of freedom. If significant at the 5 % level, we conclude that a nonlinear transformation is necessary; otherwise, the linear transformation is retained.

Complexity: If nonlinearity is confirmed, we determine whether a two-degree polynomial is necessary by comparing it to the best one-degree polynomial transformation. The gain statistic, G($J = 1$, $J = 2$) = $D(J = 1)$ – D($J = 2$), follows a Chi-square distribution with 2 degrees of freedom. If significant at the 5 % level, we retain the two-degree polynomial; otherwise, we use the simpler one-degree polynomial.

A cycle ends when the above process is applied on all predictors. Hence, performing the above process across all predictors over multiple cycles, yields a globally optimized structure with minimum Deviance, while controlling for parsimony and complexity.

It is important to note here that, the process of selecting the optimal polynomial function for each predictor in a multivariate fractional polynomial approach is designed to account for interactions between the predictors. The approach selects the best polynomial function for each predictor one at a time. During this process, all other predictors are included in the regression using the best polynomial functions identified so far. Hence, the choice of the polynomial function for any given predictor is conditional on the polynomial functions already chosen for the other predictors. In the first cycle, each predictor is transformed sequentially, beginning with the predictor that has the lowest p-value when all predictors are treated as linear. We determine the optimal polynomial function for each predictor in turn, while the others are included in the model with their best functions identified up to that point. In the second cycle we re-fit the polynomial functions for all predictors one at a time, considering the updated optimal functions for all variables, ensuring that the selection of functions for each predictor is adjusted based on the most current choices for the other predictors. The process continues, with each cycle refining the polynomial functions according to the principles of inclusion, linearity and complexity as explained above, until two consecutive cycles result in the same set of functions for all predictors. At this point, the fractional polynomial selection process is completed, minimizing the overall deviance of the model (e.g., maximizing the fit). That is, optimality is achieved in terms of minimum deviance when, over multiple cycles, the approach selects the optimal transformation for each predictor. Typically, it only takes two to three cycles to reach this point.

Another point is that, within the multivariate fractional polynomial framework, the core statistical properties of logistic regression are retained, including consistency, asymptotic normality, and likelihood-based inference (based on deviance), ensuring that the fundamental assumptions of generalized linear models remain intact. This is because MFP essentially estimates logit regressions at each step, thus the models in all steps share the robust properties of logistic regression. Moreover, MFP mitigates potential biases through a structured and systematic model selection process. The approach relies on deviance tests (in the same spirit to likelihood ratio tests) at each step to determine whether higher-order polynomials significantly improve model fit. Additionally, the stepwise procedure ensures that selected transformations are iteratively re-evaluated in the presence of other variables, reducing the risk of overfitting.

Outliers could influence the estimation and performance results, thus, in the subsequent analysis, we winsorize both the in-sample and out-of-sample data at the 5 % level.[2] Moreover, as the functions include the logarithm ($p = 0$) and square root ($p = -0.5$ or $p = 0.5$), the predictors, x, must be greater than zero. Before implementing the fractional polynomial process, we use the min-max linear transformation method such that each predictor takes values between 0.1 and 1. The min-max transformation does not alter the properties of the predictor neither changes the fundamental nonlinear relationship with the logit function. For consistency, we use the min-max predictors in the other models as well.

4. Results

4.1. Univariate

First, we explore the optimal nonlinear relationships of the predictors with the logit function univariately, using the full sample period (December 1979 – September 2023) with 2298 failed firms and 1,260,057 healthy firm-quarter observations. Each row in Table 1 shows the results of a univariate regression.

[2] Hence, where we refer to full sample results, we merge the winsorized in-sample and out-of-sample datasets. The winsorization does not affect the results but rather it is made to ensure that the results and the nonlinear relationship are both not affected by outliers.

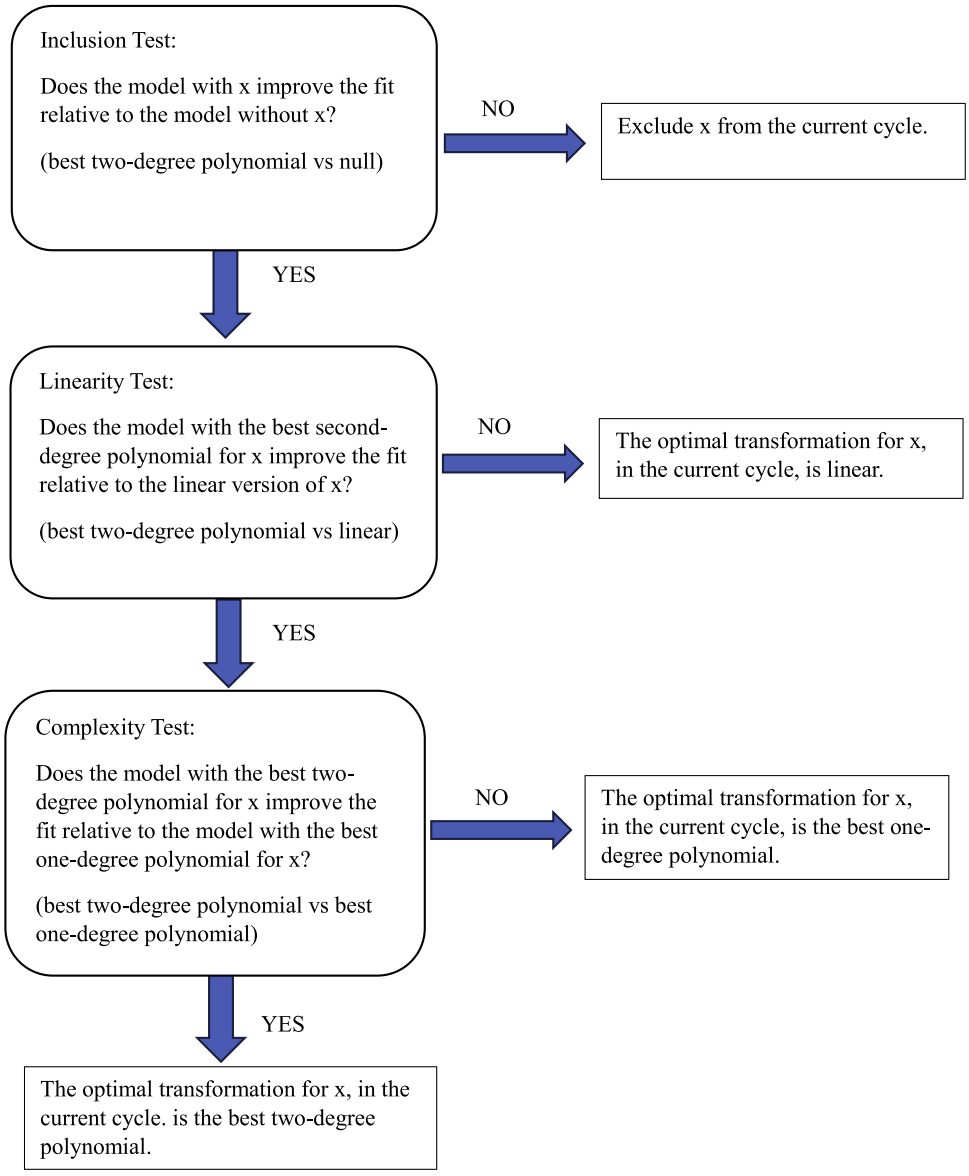

Fig. 2. This figure shows the three steps that the fractional polynomial approach follows for predictor x in the current cycle.

Table 1
Univariate results.

x	Power(s)	$F_1(x)$	$F_2(x)$	R^2 (Nonlinear)	R^2 (Linear)
EQ/TA	$(-0.5, 0.5)$	$1/\sqrt{x}$	$\sqrt{x}$	0.0594	0.0526
CA/CL	$(-2, -2)$	$1/x^2$	$1/x^2 \ln(x)$	0.0523	0.0444
NI/TA	$(3, 3)$	x^3	$x^3 \ln(x)$	0.0438	0.0245
CH/CL	$(-2, 3)$	$1/x^2$	x^3	0.0361	0.0267
TL/TA	$(2, 3)$	x^2	x^3	0.0590	0.0513
STD/TA	$(-2, 1)$	$1/x^2$	x	0.0573	0.0564
NI/IE	$(3, 3)$	x^3	$x^3 \ln(x)$	0.0029	0.0004
WC/TA	$(-1, 0)$	$1/x$	$\ln(x)$	0.0443	0.0439
SIZE	$(2, 3)$	x^2	x^3	0.0080	0.0000

This table shows univariate fractional polynomial estimation results in the full sample period. Each row is a univariate logit regression applying the fractional polynomial approach, described in Section 3, on the financial ratio shown in the first column. The second column reports the optimal power functions and the corresponding transformation, F, is shown in the third and fourth columns. The last two columns report the pseudo R^2 for the fractional polynomial (e.g., nonlinear) model and the linear model. In each regression, we use observations in the period December 1979 – September 2023 matched with a failure indicator in the subsequent quarter.

In particular, columns 3 and 4 show the optimal functions selected by the fractional polynomial process when the regression includes only one predictor at a time. As can be seen, the optimal relationship in all cases is nonlinear. Taking as an example the liquidity variable CH/CL, the optimal transformation chosen is a two-degree polynomial with powers $(-2, 3)$, yielding the two functional forms $F_1(x) = 1/x^2$ and $F_2(x) = x^3$. That is, the predictor passed the inclusion test (e.g., this two-degree polynomial transformation is better than the null), the linearity test (e.g., this two-degree polynomial transformation is better than the linear), and the complexity test (e.g., this two-degree polynomial transformation is better than the best one-degree transformation). As can be seen in the last two columns, the transformation improves the fit. The pseudo-R^2 increases from 2.67 % to 3.61 % (relative percentage change is 35 %). Similar is the case of other predictors, experiencing substantial improvements in many cases.

Fig. 3 illustrates the nonlinear relationship between selected predictors (NI/TA, CH/CL, TL/TA, SIZE) and the logit function using fractional polynomial functions. To create each plot in the figure, we sorted each predictor in ascending order and divided its values into 100 equal-sized groups. For each group, we calculated the mean value and the failure rate, defined as the number of failures in the group divided by the

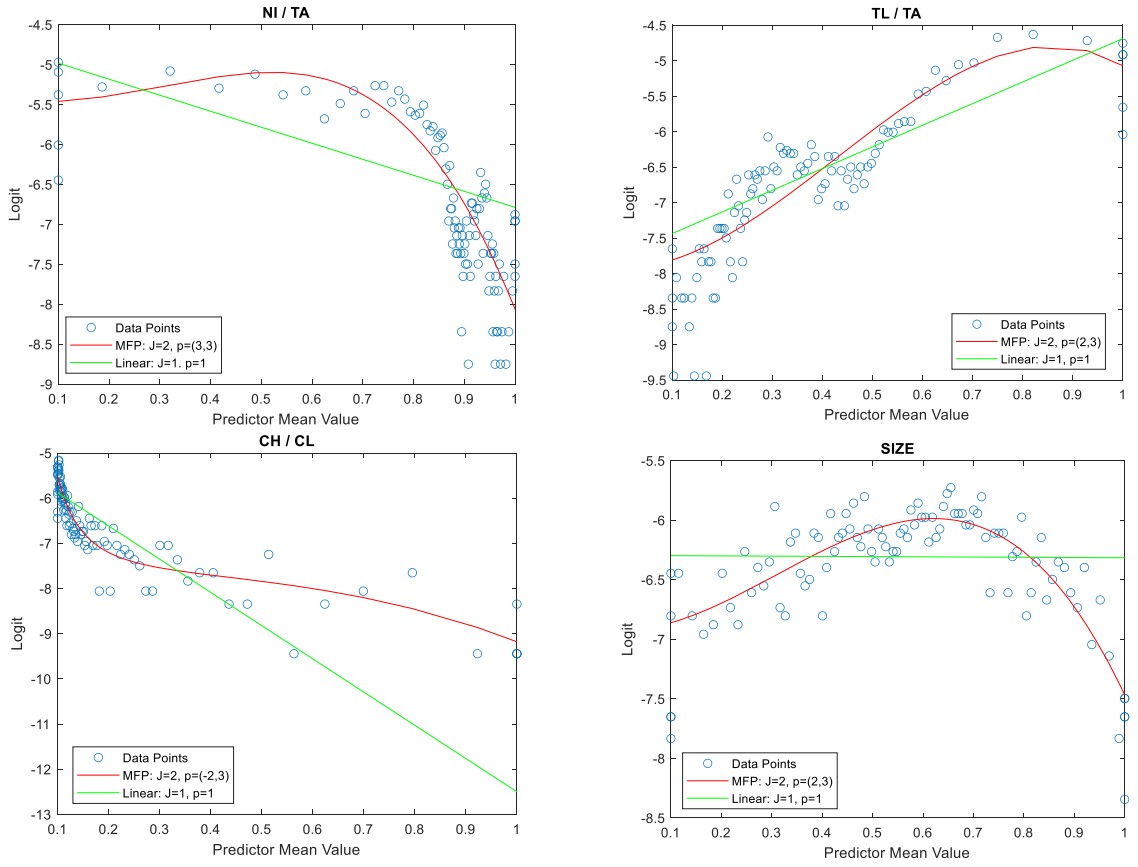

Fig. 3. This figure plots univariate relationships for selected financial ratios with the logit function in the full sample period. Financial ratios are sorted and divided in percentiles, and we compute the predictor mean value for each percentile. For each percentile, the failure rate is the number of failures in the percentile divided by the number of observations in the percentile. The logit is the log-odds = ln [failure rate / (1-failure rate)]. Then we plot the mean predictor value (x) and the logit (y), represented by the blue circles. The green line represents fitted values (at the predictor mean value) from a univariate linear logit regression. The red line represents fitted values (at the mean predictor value) from a univariate nonlinear logit regression, using the fractional polynomial approach described in Section 3. J is the polynomial degree, and p is the selected power function. The corresponding transformation, F, is shown for each predictor in Table 1. The financial ratios are constructed in the period December 1979 – September 2023 matched with the failure indicator in the subsequent quarter.

total number of observations in the group. We then computed the logit function (log-odds) for each group as $\ln\left(\frac{failure\ rate}{1-failure\ rate}\right)$ and plotted these values, represented by the blue circles. The scatter plot clearly shows that the relationship between the predictors and the logit function is nonlinear. The green line represents fitted values (estimated at the mean predictor values) from a linear logistic regression while the red curve represents fitted values (estimated at the mean predictor values) using the fractional polynomial functions shown in Table 1. According to the plots, fractional polynomial transformation of the predictors fits the data smoothly and more accurately, demonstrating the ability of fractional polynomials to capture the nonlinear relationship between predictors and the logit function much better than the linear approach.

An interesting case is that of the SIZE variable. While many studies report a negative association between firm size and distress (e.g., Bharath & Shumway, 2008; Campbell et al., 2008; Hadlock & Pierce, 2010, among others) our results suggest that this is the case for relatively very large firms.

In particular, as smaller firms grow to a certain point, they can face higher distress as they mainly finance their growth through leverage. However, above that point, firms that grow even more enjoy larger economies of scale and easier access to capital, eventually facing less distress.

Another case showing indeed that financial ratios have a nonlinear relationship with the likelihood of failure is that of the liquidity variable (CH/CL). As can be seen, even a small increase in liquidity reduces substantially the likelihood of failure, however, at very high levels the effect is marginal.

4.2. A multivariate fractional polynomial model

4.2.1. In-sample results

In this section, we provide multivariate fractional polynomial results where the best function(s) for each predictor is estimated in conjunction with other predictors in the regression. While selecting the optimal functional form for each predictor, the logit regression includes other predictors as well but transformed with the best function(s) chosen up to that point, thus accounting for interactions between the predictors.

Table 2 reports multivariate results for a full linear logit model that includes all predictors (FULL), for a stepwise linear logit model that includes predictors selected from a stepwise logit regression at the 5 % significance level (STEPWISE) and finally, for a nonlinear logit model which incorporates the optimal functions of those predictors chosen using multivariate fractional polynomials (MFP). The results are based on the in-sample period December 1979 – December 2008, with 1619 failed firms and 849,499 non-failed firm observations, which accounts for approximately 70 % of the full sample.

Interestingly, the fractional polynomial logit model selects the variables also selected by the stepwise logit model but also accounting for nonlinearities, showing its flexibility to accommodate nonlinear functions while maintaining a parsimonious set of predictors. The WC/TA variable is selected by fractional polynomials but not by the stepwise

Table 2
Estimation results.

Variables	Power (s)	F (x)	MFP (β)	dp / dx	STEPWISE (β)	FULL (β)
EQ/TA						0.40
						(0.75)
CA/CL						0.38
						(0.72)
NI/TA	(3, 3)	x^3	−3.82***	−0.011	−1.97***	−1.94***
			(−24.63)		(−14.79)	(−13.93)
		x^3 ln (x)	−6.70***			
			(−8.17)			
CH/CL	(−2, 3)	x^{-2}	0.01***	−0.022	−1.69***	−1.75***
			(6.62)		(−4.38)	(−3.40)
		x^3	−1.91***			
			(−3.73)			
TL/TA	(3, 3)	x^3	1.81***	0.003	2.36***	2.63***
			(16.00)		(13.70)	(4.91)
		x^3 ln (x)	−9.05***			
			(−12.47)			
STD/TA	(1, -)	x	1.03***	0.002	1.95***	1.88***
			(10.14)		(16.92)	(16.11)
NI/IE						0.12
						(0.54)
WC/TA	(2, 3)	x^2	−4.58***	−0.002		−0.44*
			(−6.12)			(−1.65)
		x^3	4.25***			
			(5.20)			
SIZE	(1, 2)	x	15.24***	0.006	2.94***	2.94***
			(27.50)		(25.11)	(24.59)
		x^2	−10.95***			
			(−22.47)			
Constant			−10.80***		−8.24***	−8.55***
			(−50.85)		(−38.43)	(−13.73)
$Y = 1$			1619		1619	1619
$Y = 0$			849,499		849,499	849,499
Deviance			19,147.37		20,766.17	20,762.40
Pseudo R^2			0.1859		0.1171	0.1173
AUC			0.8805		0.8194	0.8189

This table shows logit regression estimation results, using in-sample observations, for three models. FULL is a linear logit model that includes all predictors. STEPWISE is a linear logit model constructed from backward elimination of the insignificant predictors at the 5 % level. MFP is a nonlinear logit model constructed using the multivariate fractional polynomial approach described in Section 3. The "Power(s)" column reports the optimal set of powers selected for each predictor while the corresponding transformation, *F*, is also shown along with the corresponding coefficient in the next columns. The column "dp/dx" reports marginal effects, computed as the partial derivative of the probability of failure with respect to a financial predictor.
The financial predictors are constructed for the in-sample period December 1979 – December 2008 matched with a failure indicator in the subsequent quarter.
The last rows present fitness (Deviance and Pseudo-R^2) and discriminatory power (AUC) statistics for the models. The standard errors are adjusted for clusters at the firm level. *, **,*** denote statistical significance at the 10 %, 5 % and 1 % levels.

logit model, highlighting the underlying limitation of linear logit models; Linear logit models may exclude variables not because they are insignificant but because they have nonlinear relationships and thus poorly fitted by linear models. In some cases, the transformations differ from those chosen in the univariate analysis (Table 1) since the optimal functions are chosen conditional on the presence of other predictors, plus the results are based on a different time period (e.g., the in-sample period). Hence, slight differences are expected. For example, with powers (2, 3), the optimal transformations for WC/TA conditional on other predictors are $F_1(x) = x^2$ and $F_2(x) = x^3$ while in univariate regressions the optimal functions, with powers (−1, 0) are $F_1(x) = 1/x$ and $F_2(x) = \ln(x)$, however, the transformation in other cases is similar (NI/TA, CH/CL). Moreover, the coefficient for each function F is highly statistically significant at the 1 % level.

To examine whether the fractional polynomial model is economically intuitive as the stepwise logit model, we estimate marginal effects (dp / dx) for each predictor, reported in Table 2. According to the results, the sign of the marginal effects is consistent with the coefficient signs of the stepwise logit model, suggesting that higher profitability and liquidity reduces the likelihood of failure while higher leverage increases the likelihood of failure. Overall, the fractional polynomial model is consistent with economic intuition.

The last rows of Table 2 report fitness (Pseudo R^2 and Deviance) and discriminatory power (AUC) statistics for the three logit models. The nonlinear logit model, with fractional polynomial transformations of the predictors, substantially improves the fit and discriminatory power as the relative percentage change in Pseudo-R^2 and AUC is approximately 60 % and 8 % respectively compared to linear logit models. In addition, significant reduction in Deviance is achieved when going from a linear model to a multivariate fractional polynomial model. The full model, which includes all predictors in linear form, can be seen as the initial state with a Deviance equal to 20,762.40. In contrast, our proposed MFP model iteratively selects the best transformations, reducing the Deviance to 19,147.37, demonstrating a substantial improvement in model fit. The findings confirm the notion that simple nonlinear transformations of the predictors improve the model performance. The improvement is substantial, considering that the models include similar set of variables.

4.2.2. *Out-of-sample results*

To truly understand the benefits of using the nonlinear MFP logit model, we conduct an out-of-sample analysis employing an array of tests. For the tests, we use the out-of-sample observations spanning the period March 2009 – September 2023.

First, using the models in Table 2, we compute the failure probability for the out-of-sample observations and rank these observations in deciles, where the first decile includes the observations with the lowest probabilities and the tenth decile includes the observations with the highest probabilities. In each row of Table 3, we report the percentage of failed firms in each decile, defined as the number of failures in the decile divided by the total number of failures (in the out-of-sample period).

A good model would concentrate more failures towards the last deciles. According to the results, the MFP model concentrates 74 % of failed firms in the ninth and tenth decile, while the stepwise and the full logit models concentrate approximately half of the failures in those deciles. In the last row, the AUC statistics suggest that the relative percentage increase in discriminatory power achieved by the nonlinear

Table 3
Out-of-sample results: decile rankings.

Decile	MFP	STEPWISE	FULL
1	0.007	0.009	0.007
2	0.012	0.019	0.021
3	0.022	0.027	0.025
4	0.013	0.040	0.043
5	0.021	0.069	0.072
6	0.028	0.066	0.071
7	0.055	0.115	0.111
8	0.099	0.140	0.137
9	0.296	0.113	0.116
10	0.448	0.402	0.398
AUC	0.8299	0.7508	0.7483

This table reports the out-of-sample decile rankings for the models estimated in Table 2. Using out-of-sample observations in the period March 2009 – September 2023, we compute the probability of failure using each model and sort the probabilities in ascending order. We form deciles, where the first decile includes observations with the lowest risk and the tenth decile includes observations with the highest risk. For each decile, the concentration of failures is defined as the number of failures in the decile divided by the total number of failures. This number is reported in each cell. The last row reports the discriminatory power measured by the AUC.

logit model is >10 %.

From an economic perspective, misclassifying a failed firm (Type II error) is going to have far more damaging effects than misclassifying a healthy firm (Type I error). In the first case, the bank may possibly lose up to the total value of the loan outstanding whereas in the second case, it is just a lost opportunity to gain the interest from the loan. In Table 4 we report Type I and Type II errors when classifying firms into failed or healthy using various cut-off percentiles.

The results show that the nonlinear (e.g., MFP) model produces substantially lower Type II errors. For instance, the median cut-off point where firms higher than the median probability are classified as failed and non-failed otherwise, produces a Type II error equal to 7.5 % for the nonlinear MFP logit model, whereas for the stepwise and full logit models the errors are 16.3 % and 16.8 %, respectively. In many other cases the difference in Type II errors between the models is quite substantial as well.

Next, we assess the predictive ability of the various logit models in longer horizons. We run logit regressions equivalent to Table 2 but lagging the predictors by four, eight and twelve quarters.[3] Table 5 reports in-sample and out-of-sample results.

Overall, we continue to document substantial improvements, in-sample and out-of-sample, when using fractional polynomials to transform the predictors. The in-sample R^2 more than doubles in the four, eight and twelve quarter horizons while the in-sample AUC improves by up to 12 %. Moreover, the out-of-sample AUC improves by up to 15 %. Finally, in Table IA.4 of the Internet Appendix, we report results per industry, and we continue to document substantial improvements when applying the fractional polynomial approach, consistent with our main results.

Table 4
Out-of-sample results: misclassification errors.

	MFP		STEPWISE		FULL	
Cut-off (c)	Type I	Type II	Type I	Type II	Type I	Type II
5	0.950	0.006	0.950	0.004	0.950	0.004
10	0.900	0.007	0.900	0.009	0.900	0.007
15	0.850	0.010	0.850	0.018	0.850	0.019
20	0.800	0.019	0.800	0.028	0.800	0.028
25	0.750	0.034	0.750	0.037	0.750	0.037
30	0.700	0.041	0.700	0.054	0.700	0.053
35	0.649	0.043	0.650	0.074	0.650	0.075
40	0.599	0.054	0.599	0.094	0.599	0.096
45	0.549	0.063	0.549	0.110	0.549	0.122
50	0.499	0.075	0.499	0.163	0.499	0.168
55	0.449	0.090	0.449	0.197	0.449	0.202
60	0.399	0.103	0.399	0.230	0.399	0.239
65	0.349	0.125	0.349	0.281	0.349	0.283
70	0.299	0.158	0.299	0.345	0.299	0.349
75	0.249	0.193	0.249	0.423	0.249	0.426

This table reports the out-of-sample misclassification errors for the models estimated in Table 2. Using out-of-sample observations from March 2009 to September 2023, we calculate the probability of failure for each model. For classification, we use a threshold probability c, defined as the c-th percentile of the estimated failure probabilities. Observations are classified as "failed" if their estimated probability of failure exceeds the threshold; otherwise, they are classified as "non-failed." For each threshold shown in the first column, we assess two types of errors: Type I error, which is the percentage of non-failed firms incorrectly classified as failed, and Type II error, which is the percentage of failed firms incorrectly classified as non-failed.

[3] Regressions are not reported for brevity, but they are available upon request.

Table 5
Out-of-sample results: predictions in longer-term horizons.

	MFP	STEPWISE	FULL
Panel A: $t+1$			
In-Sample R^2	0.1859	0.1171	0.1173
In-Sample AUC	0.8805	0.8194	0.8189
Out-of-Sample AUC	0.8299	0.7508	0.7483
Panel B: $t+4$			
In-Sample R^2	0.1061	0.0493	0.0495
In-Sample AUC	0.8191	0.7373	0.7379
Out-of-Sample AUC	0.7527	0.6571	0.6575
Panel C: $t+8$			
In-Sample R^2	0.0620	0.0264	0.0265
In-Sample AUC	0.7584	0.6792	0.6797
Out-of-Sample AUC	0.6881	0.6123	0.6146
Panel D: $t+12$			
In-Sample R^2	0.0402	0.0157	0.0159
In-Sample AUC	0.7104	0.6465	0.6470
Out-of-Sample AUC	0.6238	0.6190	0.6133

This table reports the in-sample and out-of-sample performance for the multivariate fractional polynomial logit model (MFP), the stepwise logit model (STEPWISE), and the full logit model (FULL) in one, four, eight and twelve quarter forecasting horizons. Panel A reports the results also reported in the last rows of Tables 2 and 3. For the results reported in Panels B – D we re-run the analysis but lagging the financial predictors accordingly, to derive in-sample statistics (pseudo R^2 and AUC) and the out-of-sample AUC for the various forecasting horizons.

4.3. Fractional polynomials for banks

In this section, we assess the effectiveness of fractional polynomials in predicting bank failures. Our analysis is based on a large dataset of FDIC bank-quarter observations from Call Reports sourced from the Federal Reserve Bank of Chicago, spanning from June 1976 to September 2023 (accounting for a three-month reporting delay). The dataset comprises 1,934,396 bank-quarter observations matched with a failure indicator for the subsequent quarter. This indicator takes values of 1 for the quarter in which a bank either terminated its operations or received an assistance transaction, and 0 otherwise. Within this dataset, 2217 banks were identified as having failed, according to the Federal Deposit Insurance Corporation (FDIC) database of failed banks.

To maintain a sufficiently large number of failures, we construct financial ratios that are consistently reported throughout the sample period following the CAMELS framework widely used in the literature (see for instance Cole & White, 2012; Betz et al., 2014; Cleary & Hebb, 2016; Audrino et al., 2019, among others). We construct capital adequacy as the ratio of Equity to Total Assets (EQ/TA), asset quality as the ratio of Loan Loss Provision to Total Assets (LLP/TA), earnings quality as the ratio of Net Income to Total Assets (NI/TA), liquidity as the ratio of Cash to Total Assets (CH/TA) and SIZE as the natural logarithm of total assets.

Table 6 shows the estimation results for the in-sample period spanning the period June 1976 – December 2002 (approximately 70 % of the dataset). Applying the stepwise logit regression, all predictors are highly statistically significant, with the expected sign, confirming that measures of capital, asset quality, liquidity, and profitability, are key variables in predicting bank failures. Bank size has a negative coefficient consistent with the “too big to fail” feature of large banks. However, fractional polynomial transformation results, reported under the MFP column, confirm that there exists a nonlinear relationship between the predictors and the probability of bank failure that improves both the model fit (R^2 improves by 10 %) and the model discriminatory power improves slightly as the discriminatory power is already high.

Table 7 reports in-sample and out-of-sample results for various forecasting horizons. In all cases, the optimal nonlinear transformation

Table 6
Estimation results: bank dataset.

Variables	Power(s)	F(x)	MFP (β)	dp / dx	STEPWISE (β)
EQ/TA	(−2, 2)	x^{-2}	0.04***	−0.078	−16.01***
			(31.90)		(−19.07)
		x^{2}	−2.36***		
			(−4.83)		
LLP/TA	(−2, 0)	x^{-2}	0.01***	−0.002	1.04***
			(6.30)		(9.39)
		ln(x)	0.82***		
			(9.05)		
NI/TA	(2, 3)	x^{2}	−17.74***	−0.004	−5.13***
			(−23.24)		(−21.90)
		x^{3}	16.15		
			(20.54)		
CH/TA	(0, -)	ln(x)	−0.37***	−0.002	−0.82***
			(−9.02)		(−7.60)
SIZE	(0, -)	ln(x)	−0.50***	−0.002	−1.16***
			(−12.18)		(−10.37)
Constant			−7.10***		−0.77***
			(−47.59)		(−4.33)
$Y = 1$			1656		1656
$Y = 0$			1,352,174		1,352,174
Deviance			15,813.91		16,622.35
Pseudo R^2			0.3804		0.3487
AUC			0.9656		0.9562

This table shows logit regression estimation results on the bank dataset consisting of FDIC-insured banks, using in-sample observations, for two models. MFP is a nonlinear logit model constructed using the multivariate fractional polynomial approach described in Section 3. STEPWISE is a linear logit model constructed from backward elimination of the insignificant predictors at the 5 % level. The "Power(s)" column reports the optimal set of powers selected for each predictor while the corresponding transformation, *F*, is also shown along with the corresponding coefficient in the next columns. dp / dx are marginal effects, calculated as the partial derivative of the probability of failure with respect to a financial predictor.
The financial predictors are constructed for the in-sample period June 1976 – December 2002 matched with a failure indicator in the subsequent quarter.
The last rows present fitness (Deviance and Pseudo-R^2) and discriminatory power (AUC) statistics for the models. The standard errors are adjusted for clusters at the bank level. *, **,*** denote statistical significance at the 10 %, 5 % and 1 % levels.

of the predictors, selected by the fractional polynomial process, improves the model performance. Specifically, the in-sample R^2 improves by up to 19 % while the in-sample AUC improves moderately. However, the out-of-sample AUC improves by up to 9 %.

Overall, the results suggest that fractional polynomials improve the prediction of bank failures compared to a linear logit model.

4.4. Comparison with alternative nonlinear approaches

A natural question that arises is how the nonlinear logit model, constructed using fractional polynomial transformations, compares to other approaches that also account for nonlinearities in their structure. We consider four nonlinear models. The first one, is a spline model (SPLINES) in the spirit of Giordani et al. (2014) using B-splines for basis function and a penalty parameter for overfitting the data[4] (see also Luo et al., 2016; Djeundje & Crook, 2019a). The second one, is a Generalized Partially Linear Single Index Model (GPLSIM) which provides the option to treat some predictors linearly and the rest nonlinearly via a single index and an unknown smooth function.[5] A potential advantage of this approach is that, instead of modelling a high-dimensional nonlinear relationship directly, it reduces dimensionality by creating a single index, making it computationally efficient (Carroll et al., 1997; Li et al., 2023). Next, a widely used alternative that we consider in our study as a nonlinear benchmark is a feedforward Artificial Neural Network (ANN). To this end, many studies indeed showed that ANN outperform linear logit models as ANN can capture complex relationships in the data (see for instance Kumar & Ravi, 2007; Tinoco & Wilson, 2013; Jones et al., 2017; Mai et al., 2019). In the hidden layer we use the tan-sigmoid transfer function, while in the output layer we use the soft-max function, suitable for binary or multiclass classification. We run four different ANN configurations,[6] varying the number of neurons in the hidden layer to 3, 5, 10, and 15. Finally, we run a recurrent neural network (RNN) with a long short-term memory layer with 50 hidden units which allows to learn the long-term dependencies of the data.[7]

Table 8 reports the out-of-sample performance of the various models when predicting firm and bank failures in various forecasting horizons. We focus on the AUC performance as it is a key statistic of discriminatory power widely used in the literature (Fitzpatrick & Mues, 2016; Jabeur et al., 2020, Ma et al., 2025, among others).

Overall, we conclude that the MFP approach has higher discriminatory power compared to other nonlinear approaches, with the results being substantially more pronounced in the case of predicting firm failures. These results are not surprising as many of the alternative approaches, for example splines, neural networks etc., have critical parameters to configure or more complex structure which may compromise performance as opposed to fractional polynomials, consistent with our conjectures discussed earlier. For instance, in Table IA.5 of the Internet Appendix, we document that the performance of ANN is very volatile for the various number of neurons used, which hinders a serious limitation for ANN. This volatility introduces model risk, as selecting the optimal ANN configuration becomes challenging and can lead to unstable predictive performance.

Next, we re-run the analysis but changing the out-of-sample period.

Table 7
Performance in longer-term forecasting horizons: bank dataset.

	MFP	STEPWISE
Panel A: $t + 1$		
In-Sample R^2	0.3804	0.3487
In-Sample AUC	0.9656	0.9562
Out-of-Sample AUC	0.9686	0.9119
Panel B: $t + 4$		
In-Sample R^2	0.2399	0.2132
In-Sample AUC	0.9127	0.8940
Out-of-Sample AUC	0.9163	0.8683
Panel C: $t + 8$		
In-Sample R^2	0.1148	0.0998
In-Sample AUC	0.8225	0.7958
Out-of-Sample AUC	0.7598	0.7166
Panel D: $t + 12$		
In-Sample R^2	0.0545	0.0457
In-Sample AUC	0.7484	0.7201
Out-of-Sample AUC	0.6533	0.5971

This table reports the in-sample and out-of-sample performance on the bank dataset consisting of FDIC-insured banks, for the multivariate fractional polynomial logit model (MFP) and the linear stepwise logit model (STEPWISE) in one, four, eight and twelve quarter forecasting horizons. The estimation of the models is shown in Table 6. Panel A reports the results also reported in the last rows of Tables 6 (plus the out-of-sample AUC). For the results reported in Panels B – D we re-run the analysis but lagging the financial predictors accordingly, to derive in-sample statistics (pseudo R^2 and AUC) and the out-of-sample AUC for the various forecasting horizons.

[4] We use the 'gam' function from the 'mgcv' package in R. We use 10 knots.
[5] We estimate the model using the 'gplsim' package in R. We consider all predictors to be nonlinear, thus all predictors enter in the single index. The smooth function is a spline with 10 knots.
[6] We train ANN using MATLAB's fitcnet command.
[7] We use the 'lstmLayer' command in MATLAB to configure the RNN.

Table 8
Out-of-sample AUC of various models: main results.

	Non-Financial Firms				Financial (FDIC) Institutions			
	AUC ($t+1$)	AUC ($t+4$)	AUC ($t+8$)	AUC ($t+12$)	AUC ($t+1$)	AUC ($t+4$)	AUC ($t+8$)	AUC ($t+12$)
MFP	0.8299	0.7527	0.6881	0.6238	0.9686	0.9163	0.7598	0.6533
SPLINES	0.7851	0.6939	0.6478	0.6005	0.9673	0.9136	0.7566	0.6673
GPLSIM	0.7867	0.7203	0.6098	0.5833	0.9318	0.8702	0.7294	0.6008
ANN	0.7833	0.7237	0.6322	0.6165	0.9556	0.9111	0.7462	0.6441
RNN	0.7652	0.6796	0.6337	0.6223	0.8765	0.8662	0.7290	0.5921
STEPWISE	0.7508	0.6571	0.6123	0.6190	0.9119	0.8683	0.7166	0.5971
FULL	0.7483	0.6575	0.6146	0.6133	0.9119	0.8683	0.7166	0.5971

This table reports the out-of-sample AUC performance of the various models when predicting firm (e.g. non-financial) and bank failures. Performance for the firm dataset is based on observations in the period March 2009 – September 2023 whereas for the bank dataset performance is based on observations in the period March 2003 – September 2023. Firm and bank observations are matched with a failure indicator in the subsequent quarter. For the ANN, we report average performance when considering 3, 5, 10, 15 neurons in the hidden layer. Full results for the ANN are reported in Table IA.5 of the Internet Appendix.

In particular, we measure the AUC performance in the period 2000 – 2023 and 2012 – 2023 (while the models are trained accordingly in the previous periods).

Overall, results reported in Table 9 confirm that MFP has higher discriminating ability compared to the alternative approaches, suggesting that changing the out-of-sample period does not alter the main findings.

Finally, we conduct a simulation analysis by generating artificial firms with financial predictors that exhibit the same statistical properties as the original dataset by maintaining the covariance structure of the original financial variables. This allows us to assess how well the models generalize when trained on synthetic data that mimics real-world statistical properties. First, we create a simulated training dataset by generating synthetic firms whose financial predictors follow a multivariate normal distribution. The mean and covariance matrix of this distribution are derived from the original training dataset to ensure statistical similarity. The size of this simulated training dataset is set to 10 % of the original training sample to enhance computational efficiency. Once the simulated training sample is generated, we use it to train the models, just as we would with real data. Next, we generate 500 separate simulated testing samples, each constructed in the same way as the training sample. Specifically, we draw random numbers from a multivariate normal distribution whose covariance matrix matches that of the original testing dataset. Each of these 500 simulated test sets has a size equal to 10 % of the original testing sample to ensure computational feasibility. Once the 500 simulated test samples are created, we apply the models (trained on the simulated training data) to each of these test sets. For every simulation run, we compute the AUC and average the AUC values across all 500 test samples to obtain a robust estimate of the models' performance.

Table 9
Out-of-sample AUC of various models: alternative testing samples.

	Non-Financial Firms		Financial (FDIC) Institutions	
	2000 - 2023 AUC	2012 - 2023 AUC	2000 - 2023 AUC	2012 - 2023 AUC
MFP	0.8476	0.8351	0.9632	0.9785
SPLINES	0.8143	0.7809	0.9621	0.9778
GPLSIM	0.7935	0.7803	0.9243	0.9649
ANN	0.8014	0.8216	0.9496	0.9707
RNN	0.7873	0.7593	0.8680	0.8880
STEPWISE	0.7668	0.7514	0.9065	0.9419
FULL	0.7653	0.7476	0.9065	0.9425

This table reports the out-of-sample AUC performance of the various models when predicting firm (e.g. non-financial) and bank failures in different testing samples. Performance for both datasets is based on observations in the period March 2000 – September 2023 and March 2012 – September 2023. Firm and bank observations are matched with a failure indicator in the subsequent quarter. For the ANN, we report average performance when considering 3, 5, 10, 15 neurons in the hidden layer.

Results reported in Table 10 demonstrate that the MFP model possesses superior discriminatory power compared to alternatives, consistent with previous findings. The results are more pronounced for firm (e.g., non-financial firms).

4.5. Economic performance

In this section, we assess whether greater discriminatory power of bankruptcy models translates into higher economic benefits for banks. Following the approach of Agarwal and Taffler (2008) and Bauer and Agarwal (2014), we consider a scenario where banks operate within a competitive credit market valued at $100 billion. Each bank applies a distinct bankruptcy prediction model to assess the creditworthiness of potential borrowers, and a credit spread is charged.

To estimate credit spreads, we use the training data (December 1979 – December 2008). Firm-year observations of the training data are ranked based on bankruptcy risk and divided into ten deciles of equal size. The first decile (lowest risk) is assigned a fixed credit spread denoted by k. For the remaining deciles (2 through 10), credit spreads (CS_i) are calculated based on the methodology of Blochlinger and Leippold (2006), using the formula:

$$CS_i = \frac{p(Y=1|S=i)}{p(Y=0|S=i)} LGD + k \tag{7}$$

In this expression, $p(Y=1|S=i)$ is the bankruptcy probability and $p(Y=0|S=i)$ is the survival probability within the i th decile. LGD refers to the loss given default, and its value is set at 45 %. The constant spread k is set at 0.3 %. The bankruptcy probability for each group is computed as the actual bankruptcy rate defined as the number of bankrupt firms divided by the total number of firms in that group (thus survival probability is one minus the bankruptcy probability).

Table 10
Out-of-sample AUC of various models: simulation.

	Non-Financial Firms AUC	Financial (FDIC) Institutions AUC
MFP	0.8718	0.9391
SPLINES	0.8628	0.9383
GPLSIM	0.6925	0.8942
ANN	0.8387	0.9222
RNN	0.5857	0.5450
STEPWISE	0.7070	0.8839
FULL	0.6874	0.8839

This table reports the out-of-sample AUC performance of various predictive models in forecasting firm (non-financial) and bank failures. The AUC values reported are the average across 500 simulated testing samples, each drawn from a multivariate normal distribution with a covariance matrix matching that of the original testing dataset. The models are trained on a simulated training sample, also generated from a multivariate normal distribution with a covariance matrix matching that of the original training dataset. Each simulated sample (both training and testing) is 10 % of the size of the original dataset.

We use observations from the testing dataset (March 2009 – September 2023) e.g. customers, to simulate a competitive loan market to evaluate the economic performance of each bankruptcy model. Banks compete to issue loans to firms (represented by firm-quarter data points of the testing dataset). Each institution uses its own bankruptcy model to rank firms by risk and excludes the riskiest 5 % from receiving loans. The remaining borrowers are sorted into ten deciles, and each is assigned a credit spread based on the estimates derived from the training sample. Customers are assumed to be fully rational and price-driven: they always select the bank offering them the lowest credit spread. This assumption removes any influence of factors such as bank relationships, service quality, or marketing, and ensures that observed allocation depends entirely on the model's ability to rank customers by risk. That is, a more accurate model would assign "good" customers to low-risk deciles where they are offered competitive, lower credit spreads to attract them, and "bad" customers to high-risk deciles where they are charged higher spreads or potentially denied credit altogether. Eventually, an accurate model would have a loan portfolio with better quality thus higher risk-adjusted profitability. In contrast, banks using less accurate models may misclassify riskier customers and offer them lower spreads, leading to poorer loan portfolio quality and lower profitability. This setup isolates the impact of predictive accuracy by removing other potential influences on customer choice, ensuring that observed differences in performance are attributable purely to the power of the model.

Bank profitability is evaluated using two performance metrics. The first is **Return on Assets (ROA)**, which is defined as total profits divided by the value of assets lent. The second is **Return on Risk-Weighted Assets (RRWA)**, which adjusts for credit risk by dividing profits by risk-weighted assets. Risk weights are computed following the regulatory guidelines established in the Basel Committee on Banking Supervision (Basel Committee, 2011).

It is important to note that, to capture the economic value generated by using a particular bankruptcy model, it is crucial that banks differ only in the model they adopt. Accordingly, we assume that all banks are identical in terms of size, lending capacity, and other characteristics. This simplification is necessary for the simulation: if banks varied in other ways such as having greater market power or a stronger ability to attract customers, it would be difficult to disentangle whether superior economic performance was due to the predictive model or to those competitive advantages. By keeping all other factors constant, we ensure that any differences in outcomes can be directly attributed to the effectiveness of the bankruptcy model itself.

Table 11 shows the economic results of seven banks, each one using a distinct bankruptcy model in their decision making (indicated in the second row). As can be seen, Bank 1 that uses the MFP model achieves superior economic performance from its competitors.[8] First, and most important, Bank 1 manages a higher quality loan portfolio, as only 0.04 % of firms eventually failed to repay the loan, while for Banks 2 - 7 the bankruptcy rate ranges from 0.10 % to 0.37 %. Second, Bank 1 is more profitable than the competitors. Based on the simple return on assets (profit divided by market value of loans outstanding), ROA is 0.30 % for Bank 1 compared to a range from 0.19 % to 0.28 % for the competing banks. The simple ROA does not account for the inherent risk in the loan portfolio. To account for that, we use the risk-adjusted ROA (profit divided by risk-weighted assets) using standard formulas provided by the Basel Accord. Adjusting for risk, Bank 1 delivers an ROA equal to 3.03 % while the competing banks deliver an ROA ranging from 0.84 % to 2.88 %.

Overall, the results suggest there are economic benefits by using the MFP model in credit decisions since the bank employing the MFP model manages a better loan portfolio translating to better economic performance in terms of return on risk-adjusted assets.

4.6. Discriminatory power, goodness of fit, and statistical significance

To fully assess whether the higher performing ability of the MFP model is truly meaningful for banks, we analyze the extent to which there are statistically significant differences in the performance of the MFP model against the other competing models.

The first columns in Table 12 report the AUC values of the various models in predicting firm failures in the out-of-sample period (March 2009 – September 2023) and z-statistics, following DeLong et al. (1988) to test whether the difference between the AUC of the MFP model and the AUC of competing models are statistically significant.[9] As can be seen from the table below, the discriminatory power of the MFP is statistically significant at the 1 % level.

Next, we measure the out-of-sample goodness of fit of the various models based on McFadden's pseudo R^2, reported on the following columns in Table 12. As can be seen, the R^2 of the MFP, although not particularly high, is substantially higher compared to other models. Finally, we assess whether the goodness of fit of the MFP model is significantly better than the competing models. Specifically, the last two columns report the log-likelihoods and z-statistics following Vuong (1989) to test whether the differences between the log-likelihood of the MFP model and the log-likelihoods of the competing models are statistically significant. As can be seen, the results indicate that the MFP model fits the data significantly better than the majority of the competing models.

5. Conclusion

A common problem arising when using linear logit models in bankruptcy prediction is that financial ratios exhibit nonlinear relationships with the risk of bankruptcy. Splines and machine learning techniques, on the other hand, are subject to well-known shortcomings. We overcome those limitations by constructing a parsimonious nonlinear logit model using a multivariate fractional polynomial approach with minimal parameter specification and clear interpretability. The approach evaluates, for each predictor, whether it should be included in the model, whether the relationship is indeed nonlinear and what the optimal function is from all possible one-degree and two-degree polynomial functions. Importantly, the method accounts for interactions among predictors, effectively generalizing the stepwise logistic regression used for variable selection.

Using a large dataset consisting of non-financial firms, we show that the fractional polynomial logit model improves the out-of-sample discriminatory power, reduces significantly costly misclassification errors but also enhances the economic performance for banks. Specifically, banks adopting the model are better positioned to manage portfolios with higher credit quality, leading to increased risk-adjusted profitability. This highlights the practical value of incorporating nonlinear dynamics into bankruptcy prediction models within a multivariate fractional polynomial approach.

The robustness of our findings is further validated by applying the fractional polynomial approach on financial ratios of FDIC banks. Specifically, the fractional polynomial model has higher predictive power, in-sample and out-of-sample, in various horizons, highlighting the generalizability and effectiveness of the fractional polynomial methodology across different types of entities.

Perhaps more important is that the fractional polynomial model, overall, performs comparably or even better against other nonlinear approaches in a variety of out-of-sample tests, including tests in different

[8] For Bank 4 that uses ANN, we average the bankruptcy probabilities of four neural networks estimated using 3, 5, 10 and 15 neurons. Their performance is shown in Table IA.5 in the Internet Appendix.

[9] For the neural network, we average the bankruptcy probabilities predicted by four distinct neural networks using 3, 5, 10 and 15 neurons. The averaged bankruptcy probabilities are used as inputs for the DeLong and Vuong tests.

Table 11
Bank economic performance.

	Bank 1 MFP	Bank 2 SPLINE	Bank 3 GPLSIM	Bank 4 ANN	Bank 5 RNN	Bank 6 STEPWISE	Bank 7 FULL
Market Share (%)	16.69	30.19	14.00	19.41	5.65	5.38	6.13
Bankruptcies	28	221	69	82	81	63	94
Bankruptcy Rate (%)	0.04	0.18	0.12	0.10	0.35	0.28	0.37
Average Spread (%)	0.32	0.32	0.34	0.31	0.35	0.34	0.44
Revenues ($M)	53.83	95.78	48.05	60.95	19.74	18.10	26.91
Loss($M)	3.53	27.86	8.90	10.34	10.21	7.94	11.85
Profit($M)	50.30	67.92	39.15	50.61	9.53	10.16	15.06
Return on Assets (%)	0.30	0.22	0.28	0.26	0.17	0.19	0.28
Return on RWA (%)	3.03	2.25	2.10	2.88	1.09	1.22	0.84

This table reports economic results for seven banks in a competitive loan market worth $100 billion. Bank 1 uses a logit model constructed using the multivariate fractional polynomial approach. Bank 2 uses splines, Bank 3 uses a generalized partially linear single index model, Bank 4 uses an artificial neural network, Bank 5 uses a recurrent neural network, Bank 6 uses a stepwise logit model and Bank 7 uses a full logit model.
The banks sort prospective customers (e.g. out-of-sample observations in the period March 2009 – September 2023) and reject the 5 % of customers with the highest risk. The remaining firms are classified in 10 groups and for each group, a credit spread is calculated as explained in the text using the training dataset in December 1979 – December 2008. The bank that grants the loan is the one offering the lowest credit spread. Market share is the number of loans given divided by the number of observations, Revenues = (market size)*(market share)*(average spread), Loss=(market size)*(prior probability of bankruptcy)*(share of bankruptcies)*(loss given default). Profit=Revenues-Loss. Return on Assets is profits divided by market size*market share and Return on Risk-Weighted-Assets is profits divided by Risk-Weighted Assets, obtained from formulas provided by the Basel Accord (2011). The prior probability of bankruptcy is the bankruptcy rate for firms in the in-sample period December 1979 – December 2008 and equals 0.19 %. Loss given default is 45 %.

Table 12
Statistical significance.

Model	Discriminatory Power Test		Goodness of Fit Test		
	AUC	DeLong Stat.	Out-of-Sample R^2	Log-Likelihood	Vuong Stat.
MFP	0.8299	–	0.0207	−4931.4	–
SPLINE	0.7851	9.20***	0.0037	−5016.7	5.51***
GPLSIM	0.7867	7.33***	< 0	−5266.3	15.67***
ANN	0.7833	8.98***	0.0137	−4966.6	2.21**
RNN	0.7652	9.49***	< 0	−5051.2	4.64***
STEPWISE	0.7508	11.02***	0.0165	−4952.4	0.79
FULL	0.7483	11.22***	0.0104	−4983.3	1.96**

This table reports results on discriminatory power, measured by the Area Under Curve (AUC), and on goodness of fit (McFadden's pseudo R^2 and Log-Likelihood) for the various models in predicting firm failures in the out-of-sample period March 2009 – September 2023. The models are estimated using the training sample December 1979 – December 2008. DeLong Stat. refers to z-statistics following DeLong et.al. (1988) to evaluate whether the differences in AUCs between the MFP model and the competing models are statistically significant. Vuong Stat. refers to z-statistics following Vuong (1989) to evaluate whether the differences in Log-Likelihoods between the MFP model and the competing models are statistically significant. *** and ** indicate statistical significance at the 1 % and 5 % levels, respectively.

periods and considering a simulation analysis by generating artificial firms and banks that mimic the structure of the original data. For example, while ANN can offer high predictive accuracy, their performance tends to be volatile and highly dependent on the configuration of neurons and layers. In contrast, the fractional polynomial models provide more stable and reliable predictions as they require minimal parameter specification, making them a practical and efficient tool for bankruptcy prediction.

To sum up, our study highlights the advantages of using a multivariate fractional polynomial approach in bankruptcy prediction. The method is easy to implement but also significantly improves model performance. Moreover, it delivers practical economic benefits, making it a valuable tool for both researchers and practitioners in accounting and finance.

Future work may consider the application of fractional polynomials on various distress phenomena such as predicting credit ratings, or financial distress. These two are states that precede bankruptcy which we consider in this study. Next, this study focuses on transforming financial ratios, which can be constructed for nearly all companies, ensuring broad applicability. Future research could explore nonlinear transformations of other predictors, such as stock market variables (e.g., equity returns, volatility) or macroeconomic indicators based on fractional polynomials, to further enhance predictive modelling. Another promising avenue for future research, is the integration of multivariate fractional polynomials with machine learning models. For instance, one could employ such polynomials as a pre-processing step to transform financial predictors and then use them as inputs to machine learning models to improve predictive performance. Another machine learning-related application would be to combine the outcomes of the various distinct models to develop an ensemble model to enhance predictive accuracy, or, to develop a weighting scheme that aggregates their predictions into a single, potentially more robust, outcome. Finally, while our study applies the multivariate fractional polynomial approach within a logistic regression framework, future research could explore its application in asset pricing models by incorporating nonlinear transformations of risk factors into traditional pricing frameworks, such as the Fama-French models or the Intertemporal Capital Asset Pricing Model (ICAPM). This could provide deeper insights into the relationship between systematic risk factors and asset returns, particularly in the presence of nonlinearities that standard linear models may fail to capture.

CRediT authorship contribution statement

Zenon Taoushianis: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2025.07.036.

References

Agarwal, V., & Taffler, R. (2008). Comparing the Performance of Market-Based and Accounting-Based Bankruptcy Prediction Models. *Journal of Banking and Finance, 32*, 1541–1551.
Altman, E., & Sabato, G. (2007). Modeling credit risk for SMEs: Evidence from the US market. *Abacus, 43*, 332–357.

Audrino, F., Kostrov, A., & Ortega, J.-P. (2019). Predicting U.S. bank failures with MIDAS logit models. *Journal of Financial & Quantitative Analysis, 54*, 2575–2603.
Basel Committee. (2011). Basel III: A global regulatory framework for more resilient banks and banking systems. *Basel Committee on Banking Supervision*.
Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking and Finance, 40*, 432–442.
Betz, F., Oprica, S., Peltonen, T. A., & Sarlin, P. (2014). Predicting distress in European banks. *Journal of Banking & Finance, 45*, 225–241.
Bharath, S. T., & Shumway, T. (2008). Forecasting default with the Merton distance to default model. *The Review of Financial Studies, 21*, 1339–1369.
Blochlinger, A., & Leippold, M. (2006). Economic benefit of powerful credit scoring. *Journal of Banking and Finance, 30*, 851–873.
Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In search of distress risk. *The Journal of Finance, 63*, 2899–2939.
Carroll, R. J., Fan, J., Gijbels, I., & Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association, 92*, 477–489.
Chava, S., & Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review of Finance, 8*, 537–569.
Cleary, S., & Hebb, G. (2016). An efficient and functional model for predicting bank distress: In and out of sample evidence. *Journal of Banking & Finance, 64*, 101–111.
Cole, R. A., & White, L. J. (2012). Deja Vu all over again: The causes of US commercial bank failures this time around. *Journal of Financial Services Research, 42*, 5–29.
DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristics curves: A nonparametric approach. *Biometrics, 44*, 837–845.
Djeundje, V. B., & Crook, J. (2019a). Dynamic survival models with varying coefficients for credit risks. *European Journal of Operational Research, 275*, 319–333.
Djeundje, V. B., & Crook, J. (2019b). Identifying hidden patterns in credit risk survival data using generalised additive models. *European Journal of Operational Research, 277*, 366–376.
Fitzpatrick, T., & Mues, C. (2016). An empirical comparison of classification algorithms for mortgage default prediction: Evidence from distressed mortgage market. *European Journal of Operational Research, 249*, 427–439.
Giordani, P., Jacobson, T., von Schedvin, E., & Villani, M. (2014). Taking the twists into account: Predicting firm bankruptcy risk with splines of financial ratios. *Journal of Financial & Quantitative Analysis, 49*, 1071–1099.
Gupta, J., & Chaudhry, S. (2019). Mind the tail, or risk to fail. *Journal of Business Research, 99*, 167–185.
Hadlock, C. J., & Pierce, J. R. (2010). New evidence on measuring financial constraints: Moving beyond the KZ index. *The Review of Financial Studies, 23*, 1909–1940.
Jabeur, S. B., Sadaaoui, A., Sghaier, A., & Aloui, R. (2020). Machine learning models and cost-sensitive decision trees for bond rating prediction. *Journal of the Operational Research Society, 71*, 1161–1179.
Jones, S. (2017). Corporate bankruptcy prediction: A high dimensional analysis. *Review of Accounting Studies, 22*, 1366–1422.
Jones, S., Johnstone, D., & Wilson, R. (2017). Predicting corporate bankruptcy: An evaluation of alternative statistical frameworks. *Journal of Business Finance & Accounting, 44*, 3–34.
Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques-A review. *European Journal of Operational Research, 180*, 1–28.
Li, S., Tian, S., Yu, Y., Zhu, X., & Lian, H. (2023). Corporate probability of default; A single-index hazard model approach. *Journal of Business & Economic Statistics, 41*, 1288–1299.
Lohmann, C., Mollenhoff, S., & Ohliger, T. (2023). Nonlinear relationships in bankruptcy prediction and their effect on the profitability of bankruptcy prediction models. *Journal of Business Economics, 93*, 1661–1690.
Luo, S., Kong, X., & Nie, T. (2016). Spline based survival model for credit risk modeling. *European Journal of Operational Research, 253*, 869–879.
Ma, X., Che, T., & Jiang, Q. (2025). A three-stage prediction model for firm default risk: An integration of text sentiment analysis. *Omega*.
Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research, 274*, 743–758.
Manthoulis, G., Doumpos, M., Zopounidis, C., & Galariotis, E. (2020). An ordinal classification framework for bank failure prediction: Methodology and empirical evidence for US banks. *European Journal of Operational Research*, 786–801.
Petropoulos, A., Siakoulis, V., Stavroulakis, E., & Vlachogiannakis, N. E. (2020). Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting, 36*, 1092–1113.
Royston, P., & Altman, D. G. (1994). Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Journal of the Rotal Statistical Society, Series C, 43*, 429–453.
Sauerbrei, W., & Royston, P. (1999). Building multivariable prognostic and diagnostic models: Transformation of the predictors using fractional polynomials. *Journal of the Royal Statistical Society, Series A, 162*, 71–94.
Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business, 74*, 101–124.
Tian, S., Yu, Y., & Guo, H. (2015). Variable selection and corporate bankruptcy forecasts. *Journal of Banking & Finance, 52*, 89–100.
Tinoco, M. H., & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis, 30*, 394–419.
Traczynski, J. (2017). Firm default prediction: A bayesian model-averaging approach. *Journal of Financial & Quantitative Analysis, 52*, 1211–1245.
Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica : Journal of the Econometric Society, 57*(2), 307–333.