



Article

An Inductive Approach to Quantitative Methodology—Application of Novel Penalising Models in a Case Study of Target Debt Level in Swedish Listed Companies

Åsa Grek ^{1,*} , Fredrik Hartwig ² and Mark Dougherty ³

¹ School of Culture and Society, Dalarna University, 791 88 Falun, Sweden

² Department of Business and Economics Studies, University of Gävle, 801 76 Gävle, Sweden; fredrik.hartwig@hig.se

³ School of Information Technology, Halmstad University, 301 18 Halmstad, Sweden; mark.dougherty@hh.se

* Correspondence: age@du.se

Abstract: This paper proposes a method for conducting quantitative inductive research on survey data when the variable of interest follows an ordinal distribution. A methodology based on novel and traditional penalising models is described. The main aim of this study is to pedagogically present the method utilising the new penalising methods in a new application. A case was employed to outline the methodology. The case aims to select explanatory variables correlated with the target debt level in Swedish listed companies. The survey respondents were matched with accounting information from the companies' annual reports. However, missing data were present: to fully utilise penalising models, we employed classification and regression tree (CART)-based imputations by multiple imputations chained equations (MICEs) to address this problem. The imputed data were subjected to six penalising models: grouped multinomial lasso, ungrouped multinomial lasso, parallel element linked multinomial-ordinal (ELMO), semi-parallel ELMO, nonparallel ELMO, and cumulative generalised monotone incremental forward stagewise (GMIFS). While the older models yielded several explanatory variables for the hypothesis formation process, the new models (ELMO and GMIFS) identified only one quick asset ratio. Subsequent testing revealed that this variable was the only statistically significant variable that affected the target debt level.

Keywords: quantitative research; inductive research; survey data; register data; penalising models



Citation: Grek, Åsa, Fredrik Hartwig, and Mark Dougherty. 2024. An Inductive Approach to Quantitative Methodology—Application of Novel Penalising Models in a Case Study of Target Debt Level in Swedish Listed Companies. *Journal of Risk and Financial Management* 17: 207. <https://doi.org/10.3390/jrfm17050207>

Academic Editor: Thanasis Stengos

Received: 21 March 2024

Revised: 5 May 2024

Accepted: 8 May 2024

Published: 15 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Surveys are commonly employed in the social sciences for several compelling reasons: Survey data can address various research questions, each associated with multiple corresponding hypotheses. Typically, these hypotheses, aligned with deductive reasoning, are derived from existing theories and are subjected to various quantitative (primarily statistical) research methods. However, there are situations where an inductive approach could be more suitable for hypothesis development, such as a lack of established theory, limited empirical data on the subject, or when the phenomenon under investigation is intricate and involves multiple variables. When a deductive approach proves inadequate, researchers may need to adopt an inductive, exploratory approach for hypothesis generation. This inductive approach is often synonymous with qualitative research methods, relying on subjective data interpretation. However, Kell and Oliver (2004) argue that an inductive approach can also be founded on quantitative research methods when formulating hypotheses. Kell and Oliver (2004) contend that quantitative data can give rise to new hypotheses if it can speak for itself without preconceived notions or prior beliefs shaping the data.

Selecting a correctly specified model is essential to enabling an inductive approach to quantitative data. A correct model selection is imperative to free the data from preconceptions and allow them to reveal insights independently. In other words, a properly

chosen model minimises omitted variable bias. As [Clarke \(2005\)](#) points out, omitting a critical explanatory variable can increase the omitted variable bias. The risk is particularly high when numerous non-explanatory variables are included in the equation, especially if the explanatory variable is lost and left out. Therefore, a proper model selection should be conducted with care. There are two main approaches for achieving an appropriate model selection: (1) model selection through information criteria and (2) model selection through penalising models. Information criteria become impractical when dealing with many potential covariates ([Arnold 2010](#)). Therefore, penalising models are needed when the data have many potential covariates. The strength of penalising models is their ability to identify the most relevant explanatory variables in data with many covariates. Nevertheless, classic penalised models have their limitations. For instance, lasso and elastic net regression are inconsistent with ordinal data, commonly encountered in surveys, or only consistent under specific constraints ([Jia and Yu 2010](#)). There are several studies which have looked at different ways of selecting models ([Desboulets 2018](#); [Negrín et al. 2010](#); [Pacífico and Pilone 2024](#)).

Because surveys typically use ordinal data with many covariates, a penalising model that can use ordinal data is needed. Since the classic penalising models cannot consistently modulate ordinal data, new penalising statistical models have been developed to accommodate ordinal data with many covariates. Two such models are the cumulative generalised monotone incremental forward stagewise (GMIFS) method and the parallel element linked multinomial-ordinal (ELMO) model, both of which are adept at handling ordinal data ([Archer et al. 2014](#); [Wurm et al. 2021](#)). These novel models were developed for biostatistical application in cancer research. These methods have yet to be demonstrated to work on survey data.

What if there is a situation where there are numerous covariates but no established theory to inform hypothesis formulation? In this paper, we tackle this important concern. This paper proposes a novel method for conducting quantitative inductive research on survey data when the variable of interest follows an ordinal distribution. Thus, the primary contribution of this paper lies in presenting a new approach for conducting quantitative inductive research on survey data featuring an ordinal variable. To achieve this, we investigate the application of novel and classic penalising models to select explanatory variables affecting the target debt levels in Swedish listed firms. The primary aim of this study is the methodological description and application of the novel GMIFS and ELMO models on ordinal survey data to achieve quantitative inductive reasoning, in other words, to generate valid quantitatively based hypotheses. The employed case (the chosen target debt levels in Swedish listed firms) is used for the pedagogical purpose of outlining the methodological application. Through the case, this study aspires to transparently demonstrate this methodology for future application in other settings. As a methodological paper, the case's findings do not present this study's main purpose. However, it is necessary to discuss the case's outcome for transparency. Therefore, the secondary aim is used to differentiate from the main purpose of this study. Consequently, the secondary aim is generated from the application in this specific case: to discover variables that can predict a company's target debt level. Thus, establishing a hypothesis on the prediction of a company's target debt level will be referred to as the secondary aim of this study.

Section 2 provides an overview of prior research in the field. Section 3 outlines the methodology. Section 4 presents the case data. Section 5 reports the findings obtained from various penalised models. Finally, Section 6 discusses the findings and concludes the results obtained by different methods.

2. The Case

Studying the determinants and consequences shaping capital structure decisions is not new: In 1984, Myers asked, "How do firms choose their capital structures?" ([Myers 1984](#), p. 575). This query marked a starting point for what was to become the capital structure research field. Building on this inquiry, [Graham and Harvey \(2001\)](#) conducted a seminal

study that unveiled a significant finding: 81 per cent of the surveyed U.S. firms adhered to either a somewhat strict or flexible target debt level. Subsequently, a replication study spanning European countries—the UK, Netherlands, France, and Germany—by [Brounen et al. \(2004\)](#) reinforced this discovery, with 76 per cent of firms reporting a somewhat strict or flexible target debt level.

The existing literature offers valuable insights into the determinants for adopting target debt. [Mielcarz et al. \(2018\)](#) revealed a firm preference for debt-leverage-ratio targets. On the other hand, [Flannery and Rangan \(2006\)](#) observed a tendency among firms to align their debt with long-term capital structure targets. [Lemmon et al. \(2008\)](#) uncovered that leverage tends to exhibit constancy over time, persisting for over two decades. These findings suggest that determining target debt levels may depend on factors with enduring stability. [Miglo \(2020\)](#) discovered that firms that deviate further from their *target debt levels* are less inclined to adopt the zero-leverage policy compared to those closer to their targets. According to their model, this reluctance stems from the potential for a significant tax shield when moving towards the target, all else being equal. Similarly, [Gungoraydinoglu and Öztekin \(2021\)](#) revealed that shocks affecting firms may result in fluctuations in *target debt* ratios over time without necessarily causing observable changes in debt ratios. They also found that while leverage targets are influenced by observed leverage ratios, the degree to which cost and benefit considerations manifest in observed leverage versus leverage targets and/or target deviations may vary among firms and over time. Additionally, [Zhou et al. \(2016\)](#) unveiled a positive correlation between firms' target debt levels and the cost of their equity. In a related vein, [Hovakimian et al. \(2001\)](#) found the significance of median leverage within industry categories as a pivotal metric influencing the target debt levels of all companies operating within the same category.

Furthermore, [Harford et al. \(2009\)](#) uncovered the usage of capital structure targets among U.S. firms to facilitate substantial acquisitions. [Antoniou et al. \(2008\)](#) delved into the influence of solvency and firm size, revealing their positive effects on financial leverage, while increased profitability, growth prospects, and share prices negatively impacted financial leverage. This underscores the notion that the market environment within which the firms operate can either increase or decrease their leverage targets. [Campello \(2003\)](#) shed light on the impact of economic downturns, highlighting that firms with high debt burdens fare more poorly than their low-debt counterparts during recessions. This relationship particularly resonates in industry sectors with low debt exposure, whereas high-debt sectors exhibit greater resilience. Thus, maintaining a debt leverage target conforming with other firms within the same industry category emerges as a logical strategy during both economic downturns and upturns. [Marchica and Mura \(2010\)](#) concurred, emphasising the industry median leverage as a critical determinant of companies' target debt levels. [Memon et al. \(2021\)](#), [Touil and Mamoghli \(2020\)](#), and [Vo et al. \(2022\)](#) studied how fast firms adjusted their capital structures after a target debt ratio was set.

In addition, this case (secondary study) contributes by presenting the determinants of target debt levels within firms. Specifically, we delve into the intricate relationship between estimated target debt levels and accounting-based data—a subject to be explored in greater detail in Section 4.

3. Method

In line with the primary aim of this study, a presentation of the method used in the case is outlined below.

These questionnaires were dispatched in 2005 and 2008 to the CFOs of all companies listed on the Stockholm Stock Exchange. In cases where a CFO was absent, the questionnaire was directed to another senior executive responsible for financial management. Comprising 12 questions, with ten featuring subqueries, respondents were prompted to rank each query on a scale from zero (never/not important) to four (always/very important)¹. The consolidated dataset encompasses responses from both 2005 and 2008, featuring 292 companies. Of these, 42 remained active throughout both years. The overall adjusted

response rate for the two years stood at 39.1 per cent, with non-responses accounting for 60.9 per cent; see Table A1 for the descriptive statistics of the survey's main questions. This was an important reason why we utilised survey data in our article because the response rate was high, approximately 40 per cent (in contrast to similar studies, which have a response rate of around 10 per cent²). Another argument is, as mentioned, that survey data should only be viewed as a single case. The article's main aim is not to analyse the usage of target debt levels in Sweden today but rather to engage in a general discussion surrounding how to apply an inductive method using quantitative data.

The other variables were drawn from the Swedish Companies Registration Office (Bolagsverket), where all limited companies are mandated to submit annual reports. This database offers over 170 variables for all such companies. Matched with the Stockholm Stock Exchange market listings using corporate identification numbers, each company's potential auxiliary variables surpassed 170. Following the elimination of unusable variables, 167 variables were retained for further analysis.

3.1. Multiple Imputations via Classification and Regression Trees (CARTs)

If data are missing, utilising a penalising variable selection method without first addressing this issue would not be feasible; it would likely break down (Long and Johnson 2015). Therefore, it is imperative to employ a method for estimating the missing data before applying a penalising method.

In this study, we adopted multiple imputations by chained equations (MICEs) to estimate missing data, following the approach recommended by van Buuren and Groothuis-Oudshoorn (2011). However, using linear and logistic regressions for imputation presents challenges, such as multicollinearity in the linear regression model and separation in the logistic regression model (Albert and Anderson 1984). A solution to these problems was proposed by Burgette and Reiter (2010), who advocated for the use of Classification and Regression Trees (CARTs) (Breiman et al. 1984). Burgette and Reiter (2010) conducted a simulation study comparing the CART-based MICE algorithm to MICE using other algorithms. Their findings indicated that "quadratic and interaction terms, CART-based MICE results in notably lower mean squared errors and biases. Even the estimated main effects are somewhat closer to the truth. . . Across all β elements, approximately 70% of the intervals cover the truth when using CART-based MICE, compared with 53% for standard MICE" (Burgette and Reiter 2010, pp. 1072–73). Given that CART-based MICE outperformed other methods in the simulation study, we employed CART-based MICE to address non-response and missing values in the survey and the register data.

It is important to note that the imputation of non-response in the survey data was carried out without using information obtained from the register data and vice versa. However, one downside of CART-based MICE is its hierarchical tree structure, which could exacerbate cross-correlations between different inputs. CART-based MICE was performed separately on the two datasets: one imputation of the survey data and another of the register data. We opted for ten imputations, as Schafer (1999) suggested, when using Rubin's formula for relative efficiency (Rubin 1987), five to ten imputations are typically sufficient. Three variables had more than 70 per cent missing values and were subsequently removed. Before CART-based MICE, all 162 register variables contained incomplete information; 147 variables had complete information after the imputation. If a variable still had some missing values after CART-based MICE, it was removed to facilitate lasso regression. In total, 15 variables were deleted. After imputation, the survey data retained 109 variables with complete information, including our primary variable of interest, the target debt level.

3.2. Multinomial Lasso

Standard methods for the selection of variables are the lasso (Tibshirani 1996), the ridge, and the elastic net (Zou and Hastie 2005) regressions: out of the three methods, the lasso regression is the most widely used for variable selection. The variable of interest in

this study has an ordinal distribution. However, the multinomial lasso model can fit the ordinal data. However, the performance could be better than a penalised model for ordinal independent variables (see Wurm et al. 2021).

Let n be the sample, $\{(x_{i,t-1}, y_{i,t})\}_{i=1}^n$, where $x_i = (x_{i1,t-1}, x_{i2,t-1}, \dots, x_{ip,t-1})$ is a p -dimensional vector of covariates, and $y_{i,t} \in R$ is the dependent variable following a multinomial distribution. The negative log-likelihood is maximised subject to the constraints ($|\beta_k|_1$)

$$-\frac{1}{n} \sum_{i=1}^n \log Pr \left(Y = y_{i,t} | x_{i,t-1}; \{\beta_{0k}, \beta_k\}_{k=1}^K \right) + \lambda \sum_{k=1}^K |\beta_k|_1$$

The constraint $|\beta_k|_1$ is the l_1 -norm (Hastie et al. 2015), where λ controls the shrinkage of the modes; if $\lambda = 0$, the models give the OLS estimates, and the shrinkage increases as λ increases (Archer et al. 2014). Some of the coefficients will shrink to be precisely zero when $\lambda > 0$; the solution for the lasso is unique.

The ungrouped multinomial lasso model above allows the lasso to select different variables for different outcomes. A different variety of the model described above is the grouped multinomial lasso: the grouping is then performed on the coefficients $\beta_j = (\beta_{1j}, \beta_{2j}, \dots, \beta_{Kj})$, and the likelihood is rewritten as

$$-\frac{1}{n} \sum_{i=1}^n \log Pr \left(Y = y_{i,t} | X = x_{i,t-1}; \{\beta_j\}_{j=1}^p \right) + \lambda \sum_{j=1}^p |\beta_j|_2$$

The constraint is using the l_2 -norm. The model penalises and selects the variables that should be included later in a regression model. Both lasso methods are fitted using a coordinate descent algorithm (Hastie et al. 2015).

3.3. Element Linked Multinomial-Ordinal (ELMO) Models

Wurm et al. (2021) proposed a class of models called the element linked multinomial-ordinal (ELMO). The ELMO is a subset of vector-generalised linear models and is generally fitted with a coordinate descent algorithm. Sometimes, it uses an algorithm for ordinal and multinomial regression models with an elastic net penalty. Simulations fit this elastic net and use accurate data to outperform the lasso and ridge regressions (Zou and Hastie 2005). Each of the three ELMO models operates with a link function consisting of two parts: The first part decides the model family³, and the second part is an ordinary link function⁴. Therefore, the ELMO class model has a procedure which makes it suitable for both ordinal data and unordered categorical data, i.e., the parallel and nonparallel form of the model. The nonparallel model can be shrunk towards the parallel model using an over-parameterising nonparallel model called the semi-parallel model (Wurm et al. 2021).

Let y_j be a vector of size $n \times 1$, where $y_{ij} = \{1$ if observation i belongs to class k , 0 otherwise with K classes. Let x be a matrix of size $n \times P$. The probability that observation i with covariates x_i belongs to class k will be denoted by $p_i = (p_{i1}, p_{i2}, \dots, p_{iK})$. Let β be a $P \times K$ matrix of regression coefficients and β_0 be a vector of K intercept values. The covariates' corresponding predictors are recorded into the vector $\eta_i = \beta_0 + \beta' x_i$. The class probabilities (η_i) are connected by $\eta_i = g(p_i)$, where g consists of two parts: a function over the distribution family and an elementwise link function (see Wurm et al. 2021).

The specification for the three ELMO models is defined below. The standard for all models is that the penalising parameter is defined by $\lambda \geq 0$. The elastic net penalty is defined by $\alpha \in [0, 1]$ and is a weighted average between the lasso and ridge regression's penalising parameters; furthermore, the penalty can either be manually set or automatically selected by the data (Wurm et al. 2021). The elastic net penalty behaves as the lasso: it shrinks coefficients to zero when no relationship to the independent variable can be found (Zou and Hastie 2005). Usually, the penalising parameter λ is set first, and then, the tuning parameter α is used to select the best value (Wurm et al. 2021).

The parallel model:

$$-\frac{1}{n_*}L(\beta_0, b) + \lambda \sum_{j=1}^p \left(\alpha |b_j| + \frac{1}{2}(1 - \alpha)b_j^2 \right)$$

In the parallel model, the columns of the β -matrix are restricted to being identical. Therefore, a new variable, b , is imposed, which stands for the common column vector, and will ensure that all cumulative class probabilities are moved in the same direction, where $n_* = \sum_{i=1}^n n_i$ is the sum of multinomial trials, and L is the log-likelihood (Wurm et al. 2021).

The nonparallel model:

$$-\frac{1}{n_*}L(\beta_0, \beta) + \lambda \sum_{j=1}^p \sum_{k=1}^K \left(\alpha |\beta_{jk}| + \frac{1}{2}(1 - \alpha)\beta_{jk}^2 \right)$$

In the nonparallel model, there are no restrictions on β , and as a result, not all the cumulative class probabilities will be compelled to move in the same direction. Due to the properties of the nonparallel model, it is more suitable for unordered multinomial data, but it can still be used on ordinal data (Wurm et al. 2021).

The semi-parallel model:

$$-\frac{1}{n_*}L(\beta_0, b, \beta) + \lambda \left(\rho \sum_{j=1}^p \left(\alpha |b_j| + \frac{1}{2}(1 - \alpha)b_j^2 \right) + \sum_{j=1}^p \sum_{k=1}^K \left(\alpha |\beta_{jk}| + \frac{1}{2}(1 - \alpha)\beta_{jk}^2 \right) \right)$$

The semi-parallel model can be used on both ordinal response data and unordered multinomial data: this is because it includes an over-parameterised nonparallel model, which has both parallel and nonparallel coefficients. Depending on x_i , the semi-parallel model may contain only the parallel or nonparallel coefficients, where $\rho \geq 0$ is a third tuning parameter, especially for parallel terms (Wurm et al. 2021).

The semi-parallel model has an additional restriction $\eta_i = \beta_0 + \beta' x_i + (b' x_i) \times \ll$, compared to the non-parallel $\eta_i = \beta_0 + (b' x_i)$ and the parallel model $\eta_i = \beta_0 + (b' x_i) \times \ll$, where \ll represents a vector of length K of ones. All three ELMO models are optimised using a coordinate descent algorithm. The semi-parallel model combines both parallel and nonparallel coefficients, making it somewhat over parameterised compared to a purely nonparallel model. By employing an elastic net penalty, the penalised likelihood generally converges to a unique solution in most cases. Some covariates in the penalised semi-parallel model may exhibit only parallel coefficients, effectively setting nonparallel coefficients to zero. In some cases, the semi-parallel model incorporates both parallel and nonparallel coefficients (Wurm et al. 2021).

3.4. Cumulative Generalised Monotone Incremental Forward Stagewise (GMIFS) Method

The cumulative generalised monotone incremental forward stagewise (GMIFS) method was developed by Archer et al. (2014). The GMIFS method aims to fit a penalised method on ordinal data. This method is built on the incremental forward stagewise (IFS), which gives a penalised solution on non-ordinal data. The IFS method for linear regression has a process comparable to forward stepwise regression, which has a greedy procedure. The difference is that, compared to the forward stepwise regression, the coefficient updates are smaller and made more carefully for the IFS.

Let y_j be a vector of size $n \times 1$, where $y_{ik} = \{1 \text{ if observation } i \text{ is class } k, 0 \text{ otherwise}\}$, and the classes have K levels. Let x be a matrix of size $n \times P$. The probability that observa-

tion i with covariates x_i belongs to class k is denoted by $\pi_k(x_i)$ (Archer et al. 2014). Hence, the likelihood of an ordinal response model can be written as

$$L = \prod_{i=1}^n \prod_{k=1}^K \pi_k(x_i)^{y_{ik}}$$

Consequently, the log-likelihood can be expressed as

$$\log L = \sum_{i=1}^n \sum_{k=1}^K y_{ik} \log (\pi_k(x_i))$$

For the cumulative logit model, the corresponding log-likelihood with respect to β_p is written as

$$\frac{\delta \log L}{\delta \beta_p} = \mathbf{x}_p^T \left(\frac{y_1}{1 + \exp(\alpha_1 + \mathbf{x}\beta)} - \sum_{k=2}^{K-1} \frac{(\exp(\alpha_k + \alpha_{k-1} + 2\mathbf{x}\beta) - 1) y_k}{(1 + \exp(\alpha_k + \mathbf{x}\beta))(1 + \exp(\alpha_{k-1} + \mathbf{x}\beta))} - \frac{\exp(\alpha_{K-1} + \mathbf{x}\beta) y_K}{1 + \exp(\alpha_{K-1} + \mathbf{x}\beta)} \right)$$

One of the advantages of the GMIFS method is the estimation of a cross-correlation matrix that addresses the problem of cross-correlations between the different input variables. As GMIFS estimates a cross-correlation matrix, the model takes several detailed steps during its calculations and makes many computationally expensive iterations. Consequently, the model has the longest computational time (Archer et al. 2014).

4. Data

In line with the secondary aim of the current study, we investigate companies' target debt levels. An inductive approach requires utilising a wide array of variables. This study's survey data was combined with the vast database from The Swedish Companies Registration Office (Bolagsverket). This study included all variables accessible through the companies' annual reports.

4.1. Survey Data

The survey was administered to Chief Financial Officers (CFOs) of companies with primary listings on the Stockholm Stock Exchange in 2005 and 2008. In cases where no individual held the title of CFO, the survey was directed to another senior executive responsible for the company's financial management. This questionnaire closely replicated the survey initially developed by Graham and Harvey (2001). It comprised 12 main questions, resulting in 112 variables, with one question focusing on the target debt level. For an English translation of the survey questionnaire, please refer to Daunfeldt and Hartwig's (2014) Appendix 1. The dependent variable under examination corresponded to one of the survey questions, which was as follows:

Does your company have a target range for the solvency (or the debt-to-equity) ratio? Please, choose one of the alternatives.

1. No target debt level.
2. Yes, a flexible target range (=the aim is that the solvency/debt-to-equity ratio should be within a wide range).
3. Yes, a somewhat tight target range (=the aim is that the solvency/debt-to-equity ratio should be within a relatively narrow range).
4. Yes, a strict target range (=the aim is that the solvency/debt-to-equity ratio should be at, or very close to, a certain percentage figure.).

(Daunfeldt and Hartwig 2014, p. 111)

In 2005, the survey was distributed to all listed companies on the Stockholm Stock Exchange, a total of 244 companies. The survey was mailed on three occasions: the 8th of January, the 14th of March, and the 23rd of May. For those companies that did not respond

in the initial round, follow-up contact was made via phone, with a polite encouragement to participate in the survey. Of the 244 companies surveyed, 112 returned the completed survey. Seven of these responses were deemed unusable, resulting in an adjusted response rate of 43.0 per cent.

In 2008, the survey was again dispatched to all 249 listed companies on the Stockholm Stock Exchange. The survey was mailed on four occasions: the 18th of February, the 10th of March, the 3rd of April, and the 16th of June. As in the previous survey, non-respondents in the initial round were contacted via phone. Out of the 249 initial surveys, 92 were returned. However, four of these responses could not be utilised, resulting in an adjusted response rate of 35.3 per cent. When combining the response rates from both surveys, the overall adjusted response rate amounted to 39.1 per cent. It is worth noting that these survey data have been previously used in studies conducted by [Hartwig \(2012\)](#) and [Daunfeldt and Hartwig \(2014\)](#).

4.2. Register Data

In Sweden, all limited companies are legally obliged to submit their annual reports to the Swedish Companies Registration Office (Bolagsverket). This information is subsequently cross-referenced with each company's unique registration number. The data utilised in this study were sourced from PAR, a private consultancy agency. PAR has acquired accounting data from the Swedish Companies Registration Office and organised it into a comprehensive database. The PAR database comprises 162 accounting variables for all limited companies in Sweden. However, it is important to note that some values may be missing for certain companies.

This study used register data from the years preceding each survey, namely, 2004 and 2007. This choice was made to establish causality in alignment with the survey responses. Subsequently, following multiple imputations, these register data were matched with the questionnaire responses using each company's unique registration number. This matching process allowed us to identify the variables that may influence the adoption of a target debt level.

5. Results

The various statistical methods employed in this study, as detailed in Section 3, were used to identify determinants affecting the target debt level in Swedish listed companies. Once the variables were collected from the surveys and accounting data were obtained from the Swedish Companies Registration Office (Bolagsverket), missing variables were estimated using CART-based MICE. Each of the methods used in this study was executed using R. The results below reveal that the quick asset ratio is the sole variable significantly influencing the target debt level.

5.1. Results of the Multinomial Lasso

The multinomial lasso was executed in two different versions: the grouped and the ungrouped multinomial lasso.

Table 1 shows the variable coefficients for λ that were not shrunk to zero. The selected variables included company age, machinery, bank overdraft facility utilisation, operating profit (loss) per employee, equity ratio, and quick asset ratio. These selected variables were used in an unrestricted ordered logit regression. According to [Hastie et al. \(2001, p. 91\)](#), non-zero coefficients identified by the lasso regression can be incorporated into an unrestricted regression model (see Tables A2 and A3). The outcomes of this unrestricted ordered regression model can be seen in Table A2, which demonstrates that the quick asset ratio is the sole statistically significant variable.

Table 1. Results of the grouped multinomial lasso.

Potential Explanatory Variables	No Target Debt Level	Flexible Target Debt Level	Tight Target Debt Level	Strict Target Debt Level
Intercept	-1.94×10^{-2}	3.69×10^{-1}	-4.41×10^{-2}	-3.05×10^{-1}
Company age	-1.73×10^{-5}	-9.39×10^{-4}	3.21×10^{-4}	6.35×10^{-4}
Machinery	-5.28×10^{-9}	-7.61×10^{-10}	9.75×10^{-9}	-3.71×10^{-9}
Bank overdraft facility utilised	-6.33×10^{-9}	-1.21×10^{-11}	1.01×10^{-8}	-3.75×10^{-9}
Operating profit (loss) per employee	3.45×10^{-8}	-4.06×10^{-8}	3.56×10^{-8}	-2.94×10^{-8}
Equity ratio	1.18×10^{-4}	1.58×10^{-4}	-2.83×10^{-4}	8.54×10^{-6}
Quick asset ratio	6.46×10^{-5}	5.99×10^{-5}	-9.33×10^{-5}	-3.13×10^{-5}

Table 2 summarises the results for the ungrouped multinomial lasso. The output is represented as target debt levels 1 to 4, corresponding to the survey question about the target debt level. Furthermore, the table presents the variables that did not shrink to zero: company age, machinery, bank overdraft facility utilised, and equity ratio.

Table 2. Results of the ungrouped multinomial lasso.

Potential Explanatory Variables	No Target Debt Level	Flexible Target Debt Level	Tight Target Debt Level	Strict Target Debt Level
Intercept	-0.0325	0.3719	-9.33×10^{-3}	-0.3301
Company age		-0.0013		
Machinery			1.96×10^{-8}	
Bank overdraft facility utilised			2.22×10^{-8}	
Equity ratio			-1.75×10^{-3}	

We conducted an unrestricted ordered logistic regression model using company age, machinery, bank overdraft facility utilised, and equity ratio. The results of this analysis are available in Tables A4 and A5. None of these variables demonstrated a significant impact on the target debt level. However, it is important to note a difference in the variables selected by the grouped and ungrouped multinomial lasso regressions. While the ungrouped lasso identified the same variables, the grouped lasso also selected the operating profit (loss) per employee and quick asset ratio variables. This difference in variable selection between the two models arises from their distinct estimation methods (Wurm et al. 2021; Archer et al. 2014).

5.2. Results of the ELMO and the GMIFS

Out of all the 108 possible explanatory variables, the ELMO and the GMIFS selected the quick asset ratio as the only potentially explanatory variable (Tables 3 and 4). The difference in the estimation of the quick asset ratio affected the target debt level between ELMO and GMIFS, which their different estimation methods can explain. When the results from ELMO and GMIFS were incorporated into the unrestricted model, following Hastie et al. (2001, p. 91), a statistically significant negative correlation between the quick asset ratio and the target debt level was observed. Further details can be found in Tables A6 and A7.

Table 3. Results from the ELMO.

Potential Explanatory Variables	The Parallel Model		
	logit(P[Y <= 1])	logit(P[Y <= 2])	logit(P[Y <= 3])
Intercept	−1.1774	0.3353	1.4698
Quick asset ratio	0.0002	0.0002	0.0002
The semi-parallel model			
	logit(P[Y <= 1])	logit(P[Y <= 2])	logit(P[Y <= 3])
Intercept	−1.1324	0.3772	1.5096
The non-parallel model			
	logit(P[Y <= 1])	logit(P[Y <= 2])	logit(P[Y <= 3])
Intercept	−1.1324	0.3772	1.5096

Table 4. Results from the cumulative GMIFS.

Potential Explanatory Variables	Output
Intercept: 1	−1.1351
Intercept: 2	0.3783
Intercept: 3	1.5131
Quick asset ratio	0.0560

6. Discussion and Conclusions

The primary aim of this study was to outline a new inductive approach based on a quantitative research method. This inductive approach could lay the foundation for further exploration of quantitative hypothesis formation. The outcomes of our case study support this assertion. Data with over 170 different variables were analysed through the novel and traditional penalising methods. Subsequently, a set of potential explanatory variables emerged from the penalising models, each explanatory variable giving rise to a hypothesis that was subsequently tested in the regression phase. The results of this regression can be found in the Appendix A and are discussed below.

The six penalising models identified the following variables: company age, machinery, bank overdraft facility utilised, operating profit (loss) per employee, equity ratio, and quick asset ratio. Consequently, a key question emerged: Which among the different models should be considered the most valid? Wurm et al. (2021) investigated which penalty model performed the best on accurate ordinal data. They tested seven different models, including three versions of ELMO (parallel, nonparallel, and semi-parallel), two versions of multinomial logistic regression (ungrouped and grouped), GMIFS, and a cumulative logit model with forward stepwise variable selection by using AIC. The findings of Wurm et al. (2021) indicated that GMIFS outperformed all the other models, achieving a mean misclassification rate of 0.073. The parallel and semi-parallel ELMO were ranked second-best, with a mean misclassification of 0.091. The ungrouped multinomial logistic lasso had a mean misclassification of 0.108, and the ungrouped multinomial logistic lasso’s mean misclassification was 0.158. The worst performing was the nonparallel ELMO, with a mean misclassification rate of 0.373.

In this study, the novel penalising models ELMO and GMIFS significantly improved selectivity in identifying potential explanatory variables. They yielded one explanatory variable: the quick asset ratio (see Tables 3 and 4). The traditional models generated multiple hypotheses, which, in the regression phase, were not statistically significant. Therefore, the novel ELMO and GMIFS outperformed the traditional models, aligning with the findings by Wurm et al. (2021). The selectivity of statistically significant hypotheses is of great value in large-scale applications in future studies. In cases with a large quantity of possible variables or a combination of variables for deductive testing, the manual inductive reasoning of subsequent testing is not feasible. Hence, a selective penalising model could be vital for efficient quantitative hypothesis formation for deductive testing.

Hence, it is evident that an inductive approach can be grounded in quantitative methods, provided that the data can autonomously offer insights, as emphasised by [Kell and Oliver \(2004\)](#). Nevertheless, to enable the data to articulate themselves effectively, it is imperative to employ an accurately specified model, thereby reducing the risk of omitted variable bias interfering with the hypothesis formulation process, according to [Clarke's assertion \(Clarke 2005\)](#). Considering this, the authors of this study advocate for the advantages of using the new penalising models, ELMO and GMIFS, especially when dealing with ordinal data. These models prove invaluable for selecting optimal explanatory variables, a critical step in hypothesis development. However, the novelty of this study is also a limitation. To the authors' knowledge, this methodology paper is the first to apply the ELMO and GMIFS models inductively to survey data. Consequently, this study utilises penalising models, which have never previously been proven to work on survey data, thereby creating uncertainties relating to instances where the novel ELMO and GMIFS might provide unexpected errors or incorrect results, especially when they are applied in a new setting. In concluding the primary aim, this paper proves that these novel penalising models work on this dataset. However, there is a need for future studies to further explore quantitative inductive research based on novel penalising models.

Aligning with the secondary aim of this study, the generated hypothesis is that the quick asset ratio was the sole variable that significantly impacted the target debt level. This hypothesis was later tested, and within the boundaries of the case, it was proven true. These conclusions are further supported by the cumulative ordinal logit model, which reveals a significant negative effect of the quick asset ratio on the target debt level (for additional details, refer to [Tables A2, A3, A6 and A7](#)).

Because the quick asset ratio variable was later confirmed to be the only significant variable statistically affecting the target debt level, the novel ELMO and GMIFS were superior to the older penalising models. The older penalising models generated a larger list of potential explanatory variables, including non-significant explanatory variables. Furthermore, the ungrouped multinomial lasso did not include the significant explanatory variable: quick asset ratio. Therefore, this method included several non-explanatory variables while excluding the sole explanatory variable. This error leads to increased omitted variable bias, as [Clarke \(2005\)](#) highlighted.

There are several limitations of our paper. The first limitation of the case is that the survey data are old. Arguably, the age of survey data does not impact the main purpose of this study, as the case is mainly employed to outline the method pedagogically. Therefore, its age is of less importance for this study's primary aim. Nevertheless, the survey data age impairs the certainty of drawing real-world conclusions regarding the generated hypothesis of quick asset ratio and target debt levels. However, there is also an important strength of the case: the high response rate in the survey and the amount of possible explanatory variables. Consequently, the high response rate is relevant for the case application and outlining the method, aligning with the primary aim of this study, which proves the novel penalising model's inductive application on survey data. However, future research with contemporary data must verify the real-world association between quick-asset ratio and target debt levels.

Another limitation concerns this study's primary aim. While the novel penalising models have proved effective in this dataset, their application to survey data is new and not without uncertainties. Unexpected errors or misclassifications could occur, particularly when these models are employed in new settings. Thus, future research must investigate these models' robustness across different datasets and conditions.

Furthermore, several variables identified by traditional models were not statistically significant, which invites reflection on potential influential factors not considered in this analysis. The iterative nature of model building, especially in an inductive approach, suggests that the inclusion of additional variables could provide deeper insights.

In conclusion, this study has demonstrated the potential of novel penalising models like ELMO and GMIFS in a quantitative inductive framework. These models have shown

superior performance in identifying key variables that influence the target debt level in Swedish listed companies. However, the research’s inductive nature and the application’s novelty also suggest a cautious approach, advocating for further empirical testing and refinement of these models.

By outlining these potential variables and methodological enhancements for future research, this paper provides a roadmap for subsequent empirical inquiries. The groundwork laid in this study encourages an iterative quantitative inductive research strategy.

Author Contributions: Conceptualization, Å.G. and F.H.; methodology, Å.G. and M.D.; software, Å.G.; analysis, Å.G. and M.D.; data curation, F.H. and Å.G.; writing—original draft preparation, Å.G.; writing—review and editing, Å.G., F.H. and M.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data available on request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Descriptive statistics of the variables from the survey.

Variable from the Survey	Number of Observations	Min	Max	Median	Mean	Standard Deviation
Net Present Value	188	0	4	3	2.51	1.37
Internal Rate of Return	188	0	4	1	1.44	1.52
Annuity Method	188	0	4	0	0.378	0.795
Earnings Multiple Approach	188	0	4	0	1.40	1.60
Adjusted Present Value	188	0	4	0	0.527	1.047
Payback Period	188	0	4	3	2.29	1.43
Discounted Payback Period	188	0	4	0	0.888	1.354
Profitability Index	188	0	4	0	0.681	1.217
Accounting Rate of Return	188	0	4	0	1.08	1.49
Sensitivity Analysis	188	0	4	2	1.95	1.57
Value at Risk	188	0	4	0	0.468	0.983
Target Debt	183	1	4	2	2.22	1.03
Management Own	179	0	4	1	1.39	0.94
CEO Education	173	1	5	2	1.827	0.845
CEO Age	187	1	4	2	2.422	0.724
CEO Tenure	187	1	3	1	1.615	0.756
Foreign Sales	185	1	4	4	3.308	0.971

Table A2. Ordered logit regression model based on results for the grouped multinomial lasso.

Potential Explanatory Variables	Coefficient
Age	0.0046375 (0.002915)
Machinery	1.81×10^{-8} (3.55×10^{-8})
Bank overdraft facility utilised	8.27×10^{-8} (1.88×10^{-7})
Operating profit (loss) per employee	-1.57×10^{-6} (1.23×10^{-6})
Equity ratio	-0.0006198 (0.0037459)
Quick asset ratio	-0.0005948 * (0.000299)

Standard error in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A3. Ordered logit regression model based on results for the grouped multinomial lasso.

Potential Explanatory Variables	Odds Ratio
Age	1.004648 (0.0029286)
Machinery	1 (3.55×10^{-8})
Bank overdraft facility utilised	1 (1.88×10^{-7})
Operating profit (loss) per employee	0.9999984 (1.23×10^{-6})
Equity ratio	0.9993804 (0.0037436)
Quick asset ratio	0.9994054 * (0.0002988)

Standard error in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A4. Ordered logit regression model based on results for the ungrouped multinomial lasso.

Potential Explanatory Variables	Coefficient
Age	0.0045153 (0.0028872)
Machinery	2.05×10^{-8} (3.56×10^{-8})
Bank overdraft facility utilised	8.08×10^{-8} (1.88×10^{-7})
Equity ratio	-0.0046035 (0.0031584)

Standard error in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A5. Ordered logit regression model based on results for the ungrouped multinomial lasso.

Potential Explanatory Variables	Odds Ratio
Age	1.004526 (0.0029003)
Machinery	1 (3.56×10^{-8})
Bank overdraft facility utilised	1 (1.88×10^{-7})
Equity ratio	0.995407 (0.0031439)

Standard error in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A6. Ordered logit regression model based on results for the cumulative lasso.

Potential Explanatory Variables	Coefficient
Quick asset ratio	-0.0007021 ** (0.000248)

Standard error in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A7. Ordered logit regression model based on results for the cumulative lasso.

Potential Explanatory Variables	Odds Ratio
Quick asset ratio	0.9992982 ** (0.0002478)

Standard error in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes

- ¹ See Daunfeldt and Hartwig's (2014) Appendix 1 for the translated questionnaire.
- ² Brounen et al. (2004) had a response rate of 5 per cent, while Graham and Harvey obtained a 9 per cent response rate.
- ³ For example: adjacent category, continuation ratio, cumulative probability, and stopping ratio.
- ⁴ For example: complementary, log-log, logit, or probit.

References

- Albert, Adelin, and John. A. Anderson. 1984. On the Existence of Maximum Likelihood Estimates in Logistic Regression Models. *Biometrika* 71: 1–10. [[CrossRef](#)]
- Antoniou, Antonios, Yilmaz Guney, and Krishna Paudyal. 2008. The Determinants of Capital Structure: Capital Market-Oriented versus Bank-Oriented Institutions. *Journal of Financial and Quantitative Analysis* 43: 59–92. [[CrossRef](#)]
- Archer, Kellie J., Jiayi Hou, Qing Zhou, Kyle Ferber, John G. Layne, and Amanda E. Gentry. 2014. ordinalgmifs: An R package for ordinal regression in high-dimensional data settings. *Cancer Informatics* 13: 187–95. [[CrossRef](#)] [[PubMed](#)]
- Arnold, Todd W. 2010. Uninformative parameters and model selection using Akaike's Information Criterion. *The Journal of Wildlife Management* 74: 1175–78. [[CrossRef](#)]
- Breiman, Leo, Jerome Friedman, R. A. Olshen, and Charles J. Stone. 1984. *Classification and Regression Trees*. Belmont: Wadsworth.
- Brounen, Dirk, Abe de Jong, and Kees C. G. Koedijk. 2004. Corporate finance in Europe: Confronting theory with practice. *Financial Management* 33: 71–101. [[CrossRef](#)]
- Burgette, Lane F., and Jerome P. Reiter. 2010. Multiple imputation for missing data via sequential regression trees. *American Journal of Epidemiology* 172: 1070–76. [[CrossRef](#)] [[PubMed](#)]
- Campello, Murillo. 2003. Capital structure and product markets interactions: Evidence from business cycles. *Journal of Financial Economics* 68: 353–78. [[CrossRef](#)]
- Clarke, A. Kevin. 2005. The Phantom Menace: Omitted Variable Bias in Econometric Research. *Conflict Management and Peace Science* 22: 341–52. [[CrossRef](#)]
- Daunfeldt, Sven-Olof, and Fredrik Hartwig. 2014. What Determines the Use of Capital Budgeting Methods? Evidence from Swedish Listed Companies. *Journal of Finance and Economics* 2: 101–12. [[CrossRef](#)]
- Desboulets, Loann D. D. 2018. A review on variable selection in regression analysis. *Econometrics* 6: 45. [[CrossRef](#)]
- Flannery, Mark J., and Kasturi. P. Rangan. 2006. Partial adjustment toward target capital structures. *Journal of Financial Economics* 79: 469–506. [[CrossRef](#)]
- Graham, John R., and Campbell R. Harvey. 2001. The theory and practice of corporate finance: Evidence from the field. *Journal of Financial Economics* 60: 187–243. [[CrossRef](#)]
- Gungoraydinoglu, Ali, and Özde Öztekin. 2021. Financial leverage and debt maturity targeting: International evidence. *Journal of Risk and Financial Management* 14: 437. [[CrossRef](#)]
- Harford, Jarrad, Sandy Klasa, and Nathan Walcott. 2009. Do firms have leverage targets? Evidence from acquisitions. *Journal of Financial Economics* 93: 1–14. [[CrossRef](#)]
- Hartwig, Fredrik. 2012. The Use of Capital Budgeting and Cost of Capital Estimation Methods in Swedish-Listed Companies. *The Journal of Applied Business Research* 28: 1451–76. [[CrossRef](#)]
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2001. *The Elements of Statistical Learning*, 10th ed. Springer Series in Statistics; New York: Springer.
- Hastie, Trevor, Robert Tibshirani, and Martin Wainwright. 2015. *Statistical Learning with Sparsity*. Boca Raton: CRC Press.
- Hovakimian, Armen, Tim Opler, and Sheridan Titman. 2001. The Debt-Equity Choice. *Journal of Financial and Quantitative Analysis* 36: 1–24. [[CrossRef](#)]
- Jia, Jinzhu, and Bin Yu. 2010. On model selection consistency of the elastic net when $p \gg n$. *Statistica Sinica* 20: 595–611.
- Kell, Douglas B., and Stephen G. Oliver. 2004. Here is the evidence; now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays* 26: 99–105. [[CrossRef](#)]
- Lemmon, Michael L., Michael R. Roberts, and Jaime F. Zender. 2008. Back to the beginning: Persistence and the cross-section of corporate capital structure. *The Journal of Finance* 63: 1575–608. [[CrossRef](#)]
- Long, Qi, and Brent A. Johnson. 2015. Variable selection in the presence of missing data: Resampling and imputation. *Biostatistics* 16: 596–610. [[CrossRef](#)]
- Marchica, Maria-Teresa, and Roberto Mura. 2010. Financial flexibility, investment ability, and firm value: Evidence from firms with spare debt capacity. *Financial Management* 39: 1339–65. [[CrossRef](#)]
- Memon, Pervaiz A., Rohani Md-Rus, and Zahiruddin B. Ghazali. 2021. Adjustment speed towards target capital structure and its determinants. *Economic Research-Ekonomska Istraživanja* 34: 1966–84. [[CrossRef](#)]
- Mielcarz, Paweł, Dmytro Osichuk, and Ryszard Owczarkowski. 2018. Financial restructuring and target capital structure: An iterative algorithm for shareholder value maximization. *Review of Accounting and Finance* 17: 280–94. [[CrossRef](#)]
- Miglo, Anton. 2020. Zero-debt policy under asymmetric information, flexibility and free cash flow considerations. *Journal of Risk and Financial Management* 13: 296. [[CrossRef](#)]
- Myers, Stewart C. 1984. The capital structure puzzle. *The Journal of Finance* 39: 574–92. [[CrossRef](#)]

- Negrín, Miguel A., Francisco J. Vázquez-Polo, María Martel, Elías Moreno, and Francisco J. Girón. 2010. Bayesian variable selection in cost-effectiveness analysis. *International Journal of Environmental Research and Public Health* 7: 1577–96. [[CrossRef](#)] [[PubMed](#)]
- Pacifico, Antonio, and Daniela Pilone. 2024. Penalized Bayesian Approach-Based Variable Selection for Economic Forecasting. *Journal of Risk and Financial Management* 17: 84. [[CrossRef](#)]
- Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Schafer, Joseph L. 1999. Multiple imputation: A primer. *Statistical Methods in Medical Research* 8: 3–15. [[CrossRef](#)] [[PubMed](#)]
- Tibshirani, Robert. 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58: 267–88. [[CrossRef](#)]
- Touil, Marwa, and Chokri Mamoghli. 2020. Institutional environment and determinants of adjustment speed to the target capital structure in the MENA region. *Borsa Istanbul Review* 20: 121–43. [[CrossRef](#)]
- van Buuren, Stef, and Karin Groothuis-Oudshoorn. 2011. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 45: 1–67. [[CrossRef](#)]
- Vo, Thuy A., Mieszko Mazur, and An Thai. 2022. The impact of COVID-19 economic crisis on the speed of adjustment toward target leverage ratio: An international analysis. *Finance Research Letters* 45: 102157. [[CrossRef](#)] [[PubMed](#)]
- Wurm, Michael J., Paul J. Rathouz, and Bret M. Hanlon. 2021. Regularized ordinal regression and the ordinalNet R package. *Journal of Statistical Software* 99: 1–42. [[CrossRef](#)] [[PubMed](#)]
- Zhou, Qing, Kelvin J. Tan, Robert Faff, and Yushu Zhu. 2016. Deviation from target capital structure, cost of equity and speed of adjustment. *Journal of Corporate Finance* 39: 99–120. [[CrossRef](#)]
- Zou, Hui, and Trevor Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series* 67: 301–20. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.