Decoding Distress: The Behavior of Firms Preceding bankruptcy*

Martin Farias †
CEMFI

April 4, 2024

Abstract

This article offers evidence of the behavior and performance of Spanish firms preceding bankruptcy. I estimate a predictive model of bankruptcy which shows that the most important predictors of bankruptcy include measures of equity (net worth), profitability, size, and dividends, as well as the growth rate of aggregate credit. I complement the insights provided by the predictive model with an exploration of the dynamics of firms in the five years preceding bankruptcy, comparing it with the dynamics of firms before exiting the market without filing for bankruptcy. I find that firms arrive at the point of filing for bankruptcy in a very distressed financial situation, after experiencing greater decreases in earnings and equity than firms exiting the market without filing for bankruptcy. Despite this, other measures of the real and financial performance of these firms (investment rates, growth rates of employees, cash savings, and growth rate of account receivables) are mostly similar across the two groups of firms.

JEL codes: G30, G32, G33.
Keywords: Bankruptcy prediction, distressed firms, firm dynamics.

*I would like to thank Javier Suarez, Federico Kochen, Rafael Repullo, Vedant Agarwal, Utso Pal Mustafi, Tomás Budí-Ors, Pello Aspuru, Giorgio Pietrabissa, Francesco Chiocchio, Esteban Tisnés, and the participants of the Banking and Finance seminar at CEMFI for their helpful comments and suggestions. I acknowledge financial support from the Maria de Maeztu Unit of Excellence CEMFI MDM-2016-0684, funded by MCIN/AEI/10.13039/501100011033, the Santander Research Chair, Fundación Carolina and CEMFI.

†Contact: Email: martin.farias@cemfi.edu.es

1 Introduction

Bankruptcy prediction is a topic of great interest in the literature on corporate finance, with some of its most influential contributions dating back to Beaver (1966) and Altman (1968). This is because the ability to predict bankruptcy is of great importance to a variety of stakeholders, such as commercial banks that need to assess the credit risk of their borrowers, prudential supervisors who care about the implications for banks' solvency and asset quality, and policymakers who worry about the wider economic implications of firms' financial distress.

I contribute to this literature by offering further evidence of the behavior and performance of Spanish firms preceding bankruptcy. First, I use data on the quasi-universe of Spanish firms from 2000 to 2019 to estimate the one-year ahead probability that a firm files for bankruptcy. I find, in line with existing literature, that some of the most important predictors of bankruptcy are measures of equity (net worth), profitability, size, and dividends. Second, I offer a graphical description of the dynamics of firms preceding bankruptcy (the *path to bankruptcy*) across a list of variables that measure profitability, equity, and other measures of the real and financial performance of the firms. I find that firms arrive at the point of filing for bankruptcy in a very distressed financial situation, after experiencing greater decreases in earnings and equity than firms exiting the market without filing for bankruptcy. Despite this, other measures of the real and financial performance of these firms (investment rates, growth rates of employees, cash savings, and growth rate of account receivables) are mostly similar across the two groups of firms.

Bankruptcy proceedings in Spain are currently governed by a Bankruptcy Act that entered into force in 2004. The Act establishes a single in-court bankruptcy procedure, the *concurso de acreedores*, which may end in a restructuring agreement between the debtor and the creditors or in the liquidation of the company. The majority of the bankruptcy proceedings in Spain are initiated by the debtors (voluntary), reflecting the strong incentives for debtors to file for bankruptcy before their creditors force them to do so (Gómez and Sánchez, 2018). Two distinctive features of the Spanish bankruptcy system are that the majority of the bankruptcy proceedings end up in the liquidation of the debtor and that small firms rarely file for bankruptcy. Small firms tend to favor other mechanisms to deal with financial distress, such as carrying out debt enforcement via mortgage foreclosures (García-Posada and Mora-Sanguinetti, 2014; García-Posada, 2020).

I use administrative data of Spanish firms to estimate the conditional probability of bankruptcy for each firm-year observation and to describe the behavior of firms preceding bankruptcy. The data is obtained from the *Central de Balances Integrada* (CBI) dataset, which contains information from the balance sheet and income statements of the quasi-universe of Spanish firms. I verify that the evolution of the number of bankruptcies captured in the sample finally used in the analysis (after applying standard filters) tracks very closely the evolution of the aggregate number of bankruptcies reported by the judiciary statistics.

Then, I estimate a predictive model that uses a logistic regression linking a binary variable that indicates if a firm files for bankruptcy in a given year to a set of one-year lagged firm-level and aggregate variables. To construct this indicator, I use a variable included in the CBI dataset that registers the exact date on which a firm files for bankruptcy. I take an agnostic approach to select the covariates that are included in the predictive model. I begin with a large set of covariates that are typically used in the literature to predict bankruptcy, and I use LASSO (Least Absolute Shrinkage and Selection Operator), a popular regularization method, to select the most relevant covariates that maximize the predictive power of the model.

The results of the model show that the most important predictors of bankruptcy include (one-year lagged) measures of equity (net worth), profitability (as measured by EBITDA and operating cash flow), size (as measured by employees and total assets), and dividends. Interestingly, LASSO also keeps the lagged growth rate of aggregate credit as a relevant predictor of bankruptcy. I assess the predictive performance of the model using the Receiver Operating Characteristic (ROC) curve, the Area Under the ROC Curve (AUC), and the confusion matrix.[1] The AUC, which can be interpreted as the probability that the model ranks a randomly chosen positive occurrence as more likely positive than a randomly chosen negative occurrence, is 0.905. In turn, the confusion matrix shows that the model has high accuracy in predicting both bankruptcy and non-bankruptcy occurrences. The model correctly predicts 81.3% of the bankruptcy occurrences and 84.2% of the non-bankruptcy occurrences.

Then, I document six facts about the behavior of firms in the five years preceding bankruptcy. To highlight what is special about the dynamics of firms before bankruptcy, I compare them with the dynamics of other firms in the five years before exiting the market without filing for bankruptcy. Henceforth, I denote firms that eventually exit the market without filing for bankruptcy as *exiting* firms, and those that eventually file for bankruptcy as *bankrupt* firms. First, I find that bankrupt firms have lower profitability than exiting firms. Profitability decreases as the firms approach the year of either bankruptcy or exit, but the decrease is much more pronounced for bankrupt firms. Sec-

[1] See section 5.1.3 of Murphy (2022) for a review of these concepts.

ond, bankrupt firms have lower equity than exiting firms. Moreover, the equity of bankrupt firms tends to decrease as they approach the year of bankruptcy.

The remaining four facts are related to other measures of the real and financial performance of the firms. The third fact is that, despite the weaker financial condition of bankrupt firms, the investment rates of bankrupt and exiting firms do not differ significantly. The fourth fact is that bankrupt and exiting firms exhibit declining growth rates of employees as they approach the year of either bankruptcy or exit. The fifth and sixth facts talk about the dynamics of cash savings and the growth rate of account receivables. I find that bankrupt firms exhibit a declining trend in cash savings and the growth rate of account receivables as they approach the year of bankruptcy while exiting firms have more stable cash savings and growth rates of account receivables. One aspect that these last four facts have in common is that the dynamics of bankrupt and exiting firms are mostly similar across these measures. It is only the year before bankruptcy that eventually bankrupt firms start to exhibit a more pronounced decline in the growth rate of employees, cash holdings, and account receivables than firms that exit without filing for bankruptcy.

Finally, I explore whether the dynamics of bankrupt and exiting firms commented on above diverge across some relevant dimensions. Specifically, I examine whether these dynamics vary based on the age at which firms file for bankruptcy or exit the market, whether such decisions are made during recessionary or expansionary periods, and the size of the firms. I find that, in general, the dynamics of bankrupt and exiting firms are mostly similar across these dimensions.

Related literature. This article is directly related to three strands of the literature. First, it is related to the literature on bankruptcy prediction. This literature has a long tradition, with some of its most influential contributions dating back to Beaver (1966) and Altman (1968). The literature has used a variety of statistical methods to predict bankruptcy, including discriminant analysis (Altman, 1968), hazard models (Shumway, 2001), logit and probit models (Blanco, Ortiz, García-Posada, and Mayordomo, 2024), and other machine learning techniques. I follow one of the most typical approaches in the literature, which is to use a logistic regression model to predict the one-year ahead probability of bankruptcy, where a firm is considered to be bankrupt if it files for a bankruptcy procedure.

Second, I relate to the literature that assesses the properties of the bankruptcy procedure in Spain. Articles in this literature have studied what determines the low bankruptcy rates in Spain (Celentani, García-Posada, and Gómez, 2010; Garcia-Posada and Mora-Sanguinetti, 2012, text) and how the reforms implemented since the enactment of the Bankruptcy Act in 2004 have affected outcomes of the bankruptcy proceedings such as the likelihood of reorganization and the duration of the proceedings (Gómez and Sánchez, 2018), among other topics. Third, I relate to the literature that studies the effects of bankruptcy procedures on the dynamics of firm value and capital structure (Cooley and Quadrini, 2001; Bris, Welch, and Zhu, 2006; Hennessy and Whited, 2007; Corbae and D'Erasmo, 2021). My contribution to the last two strands of the literature is to complement the existing studies that document firm dynamics preceding bankruptcy, which have mostly used US data, with evidence from Spain.

The rest of the paper is organized as follows. Section 2 describes the institutional framework of corporate bankruptcy laws in Spain. Section 3 describes the data and the variables used in the analysis. Section 4 presents the predictive model and the results. Section 5 describes the behavior of firms preceding bankruptcy. Section 6 concludes.

2 Institutional framework: Corporate bankruptcy law in Spain

In Spain, bankruptcy proceedings are currently governed by a Bankruptcy Act that entered into force in 2004. There is only one in-court bankruptcy procedure, the *concurso de acreedores*. The bankruptcy procedure can be used both by companies that have already suspended payments and by those that foresee imminent payment difficulties, even if they are currently meeting their obligations to creditors. The application for bankruptcy can be made by any of the creditors or by the debtor company itself, which is obliged to file within two months of being in a situation of *insolvency*. For the Bankruptcy Act, a firm is assumed to be *insolvent* after three months of non-payment of taxes, social security contributions, or salaries.

Each application for bankruptcy is examined by a judge, who may accept or reject it. The degree of autonomy of the filing company differs depending on who files for bankruptcy. If it is the company itself, it continues to manage its assets and commercial activity, although its operations are supervised by an administrator. In turn, if it is requested by the creditors, the directors are relieved of their functions, and the management of the company is carried out by the administrator.

A proceeding may end in a restructuring agreement between the debtor and the creditors, which implies the survival of the company, or in the liquidation of the company. During the bankruptcy proceeding, there is an automatic stay on all unsecured claims, i.e., no action can be initiated to collect such debts. Creditors whose claims are guaranteed by assets involved in the production process of the firm are also affected by the suspension. Likewise, the accrual of interest to creditors is suspended during this period, except in the case of secured claims and wage claims.[2]

Some papers that focus on the examination of the bankruptcy proceedings in Spain, including García-Posada and Mora-Sanguinetti (2014) and García-Posada (2020), highlight the lengthy and costly nature of the proceedings. Moreover, some costs such as lawyers' salaries or judiciary fees associated with filing for bankruptcy are not proportional to firms' assets or debt liabilities. This is considered to be associated with two distinctive features of the Spanish bankruptcy system. First, the majority of the bankruptcy proceedings in Spain end up in the liquidation of the debtor. This may be because firms defer filing for bankruptcy in a gamble to avoid the associated legal and administrative costs. However, this delay leads, in many cases, to the deterioration of firms' financial health. Second, small firms rarely file for bankruptcy because they are disproportionately discouraged by the costs associated with the procedure. García-Posada and Mora-Sanguinetti (2014) argue that small firms prefer to carry out debt enforcement via mortgage foreclosures, which are cheaper procedures than bankruptcy, in case of financial distress. They provide evidence that the capital structure of small firms is indeed biased towards mortgage loans.

3 The data

Subsection 3.1 presents the data source used in this article and explains the cleaning process applied to the original data. Subsection 3.2 discusses the coverage of the data source for bankrupt firms. Finally, subsection 3.3 presents summary statistics of key variables used in the analysis.

3.1 Data source

My sample consists of balance sheet and income statements data of firms appearing in the *Central de Balances Integrada* (CBI) dataset of the Bank of Spain at any point between 2000 and 2019. The CBI contains information on the quasi-universe of Spanish firms, providing an accurate representation of the Spanish economic structure. Following common practice, I focus on for-profit, not government-owned corporations that do not belong to the financial industry, industries heavily influenced by the state (Education, Health, and International organizations), or industries where firms are a minority with respect to self-employed households.

[2]See García-Posada (2020) for a more detailed description of the bankruptcy procedure in Spain.

I also apply a variety of filters that exclude observations without valid and consistent information for the variables used in the analysis. The final sample consists of 7,837,901 firm-year observations from 1,414,577 unique firms. There are 44,142 firms that ever filed for bankruptcy, which accounts for 211,466 firm-year observations in the sample. A detailed explanation of the cleaning steps can be found in Appendix A.

3.2 Representativeness of the CBI dataset for bankrupt firms

Figure 1 shows the number of bankruptcies in my clean sample between 2004 and 2019. The black line represents the number of bankruptcies in the population, as reported by the National Institute of Statistics (INE), based on judiciary statistics, while the red line represents the number of bankruptcies contained in my clean sample from the CBI dataset. The evolution of the number of bankruptcies in Spain, which increased after 2007 and peaked in 2013, likely reflects the effects of several factors such as the impact of the business cycle, the various phases of the global financial crisis and the European sovereign debt crisis, and the effects of the reforms in the Bankruptcy Act that took place in 2009, 2011, and 2014.

Figure 1 speaks to the quality of my sample of bankrupt firms in Spain. My sample contains information about 57% of the bankruptcy proceedings initiated during the sample period. Additionally, the evolution over time of the number of bankruptcies in my sample mimics the evolution of the number of bankruptcies in the population.[3]

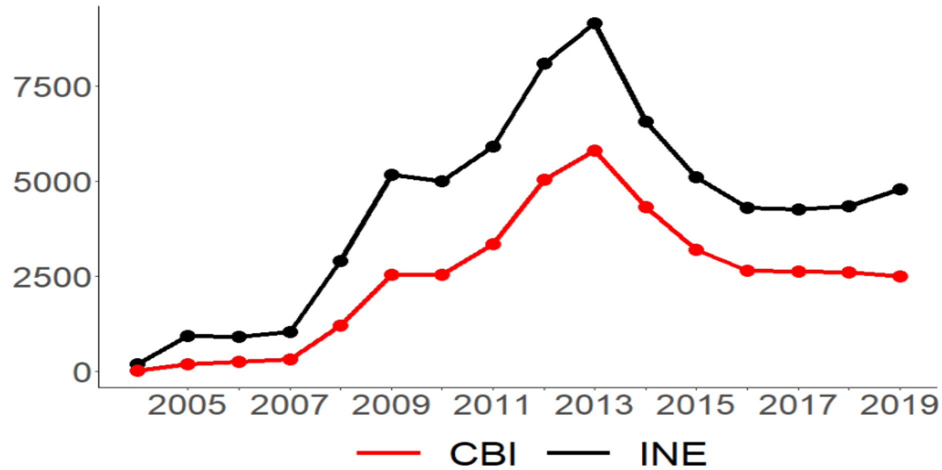
3.3 Summary statistics of key variables

Table 1 displays summary statistics that describe the differences between *in-bankruptcy* firms and the rest of the firms (*non-bankrupt*). Variables are defined in table A1. I consider a firm-year observation to be *in-bankruptcy* if the firm has filed for bankruptcy and (if applicable) has not completed the reorganization process yet. The rest of the firm-year observations are labeled as *non-bankrupt*. Variables are expressed as ratios to contemporaneous assets of each firm, except for *assets*, *employees*, *age*, and *Z-score*, which are the assets of the firm (in thousands of euros), the number of employees in the firm, the difference between the reporting year and the year of incorporation, and the Altman Z-score of the firms, respectively.

Several differences between in-bankruptcy and non-bankrupt firms are apparent from

[3] Figure B1 in Appendix B shows a figure similar to figure 1 but explaining in more detail how many bankruptcy events are captured by the CBI dataset in different steps of the cleaning process.

FIGURE 1. Number of bankruptcies in Spain (2005-2019) in the population (INE) and in the CBI dataset.



This figure shows data on the number of corporate bankruptcies in Spain between 2004 and 2019. Information about bankruptcy proceedings initiated by individuals is excluded. The figure shows the total number of bankruptcies reported by the National Institute of Statistics (INE) and the number of bankruptcies contained in the CBI dataset after the cleaning steps mentioned in section 3.1 are applied.

table 1. First, in-bankruptcy firms are less profitable, as captured by lower average earnings (*EBITDA*) or cash flow (*CF*). Accordingly, *in-bankruptcy* firms display lower average rates of cash savings (Δ*Cash*), real investment (*Real investment*), and financial investment (*Financial investment*). Second, in-bankruptcy firms tend to be larger (in terms of *assets* and *employees*) and older than non-bankrupt firms. This is in line with prior evidence in Spain (e.g., García-Posada and Mora-Sanguinetti (2014)), and it reflects the fact that smaller firms are discouraged from filing for bankruptcy to avoid the costs associated with the process, some of which are not proportional to firms' assets or debt liabilities. Third, in-bankruptcy firms, on average, exhibit higher leverage than non-bankrupt firms, as measured by total debt liabilities (*Liabilities*). In other words, given the relationship between assets, liabilities and equity, in-bankruptcy firms have lower equity than non-bankrupt firms. Finally, there are also significant differences in the average values of the Altman Z-scores of in-bankruptcy and non-bankrupt firms. The average Z-score of in-bankruptcy firms is 0.817, which falls in the area (Z-score < 1.23) where the score indicates that the firm is very likely to head towards bankruptcy in the next two years. In contrast, the average Z-score of non-bankrupt firms is 5.521, which falls in the area (Z-score > 2.99) where the score indicates that the firm is in solid financial health. The last two rows of table 1 show the percentage of in-bankruptcy and non-bankrupt firms that have a Z-score lower than 1.23 and higher than 2.99. We observe that 77.2% of in-bankruptcy firms have

a Z-score lower than 1.23 and only 5.6% have a Z-score higher than 2.99. In turn, 34.4% of non-bankrupt firms have a Z-score lower than 1.23 and 33.1% have a Z-score higher than 2.99.

4 Predicting bankruptcy

Subsection 4.1 specifies the model that I estimate to predict the probability of bankruptcy for each firm-year observation. In turn, subsection 4.2 describes the sample used and the covariates selection process, while subsection 4.3 presents the estimated predictive model and the assessment of its performance.

4.1 Predictive model

To assess a firm's bankruptcy risk, I estimate a predictive model that uses a logistic regression linking a binary variable $Y_{i,t}$ with a vector of covariates $X_{i,t-k}$ measured at year $t-k$. The binary variable $Y_{i,t}$ takes the value 1 if the firm i files for bankruptcy at year t and 0 otherwise. Specifically, the prediction is given by the following equation:

$$P\left(Y_{i,t}=1 \mid Y_{i,t-1}=0, X_{i,t-k}\right)=\frac{e^{\beta' X_{i,t-k}}}{1+e^{\beta' X_{i,t-k}}}, \tag{1}$$

where $X_{i,t-k}$ includes lagged values of firm-level and aggregate variables. The aggregate variables are the growth rate of GDP and the growth rate of aggregate credit. The firm-level variables include measures of firms' size, leverage, profitability, and the composition of assets and liabilities, among others. The complete set of covariates can be found in table B1. Following the standard practice in the literature, I use one-year lags of all the covariates, i.e., $k=1$, and these covariates are standardized before inputting them into the model.

4.2 Estimation: Sample and covariates selection

Existing papers in the literature have used a multitude of variables to predict firms' bankruptcy. As a result, an encompassing model may include a large number of covariates. However, the selection of covariates must balance the benefit of increasing the dimensionality of the vector $X_{i,t-1}$, which is the potential improvement of the predictive power of the model, with its drawbacks, mainly the risk of overfitting the model to the training sample. To strike a balance, I begin with the broad set of covariates that

TABLE 1. Summary statistics of key variables, and differences in means between in-bankruptcy and non-bankrupt firms.

Variable	**Mean**		**Difference**	**P-value**
	In-bankruptcy	**Non-bankrupt**		
EBITDA	-0.056	0.045	-0.101	<1e-4
CF	-0.054	0.025	-0.080	<1e-4
Δ *Debt*	-0.008	0.017	-0.024	<1e-4
Δ *Cash*	0.001	0.008	-0.009	<1e-4
Real investment	-0.032	0.032	-0.064	<1e-4
Financial investment	-0.001	0.002	-0.002	<1e-4
Δ *Account receivables*	-0.025	0.004	-0.029	<1e-4
Net Div	-0.003	-0.003	<1e-4	0.736
Assets	9,900.6	3,269.5	6,631.1	<1e-4
Cash	0.098	0.203	-0.105	<1e-4
Tangible assets	0.338	0.333	0.004	0.134
Intangible assets	0.023	0.021	0.003	0.006
Current assets	0.579	0.602	-0.023	<1e-4
Non-current assets	0.421	0.398	0.023	<1e-4
Current liabilities	0.718	0.493	0.225	<1e-4
Non-current liabilities	0.420	0.181	0.239	<1e-4
Liabilities	1.139	0.674	0.464	<1e-4
Employees	18.288	10.539	7.749	<1e-4
Age	20	13	7	<1e-4
Z-score	0.817	5.521	-4.704	<1e-4
Z-score < 1.23	0.772	0.344	0.428	<1e-4
Z-score > 2.99	0.056	0.331	-0.275	<1e-4

Note: Variables are defined in table A1. All the variables are measured as ratios to contemporaneous assets of each firm *(assets)*, except for *assets*, *employees*, *age*, and *Z-score*, which are the assets of the firm (in euros), the number of employees in the firm, the difference between the reporting year and the year of incorporation, and the Altman Z-score of the firms, respectively. The Altman Z-score is computed as follows: 0.717 * *working capital/assets* + 0.847 * *Retained earning/assets* + 3.107 *EBITDA/assets* + 0.420 *equity/Liabilities* + 0.998 * *Sales/assets*. A Z-score bigger than 2.99 indicates that the firm is in solid financial health; a Z-score lower than 1.23 indicates that the firm is very likely to head towards bankruptcy in the next two years; a Z-score between 1.23 and 2.99 indicates a moderate chance of bankruptcy in the next two years. *In-bankruptcy* firms are those that have filed for bankruptcy and (if applicable) have not completed the reorganization process yet. The date on which a firm files for bankruptcy is obtained from the *alta_sitc* variable reported in the CBI. The *Difference* column computes the difference in means between *in-bankruptcy* firms and *non-bankrupt* firms. The *P-value* column displays the p-value of the t-test of difference in means, whose null hypothesis is that the means are equal.

are presented in table B1 and I resort to LASSO (Least Absolute Shrinkage and Selection Operator), a popular regularization method that enhances the prediction accuracy of the model, to select the most relevant covariates.[4]

The LASSO estimator of the vector of coefficients β is defined as the solution to the following optimization problem:

$$\min_{\beta} \sum \left\{ -Y_{i,t} \left(\beta' X_{i,t-k} \right) + \log \left(1 + e^{\beta' X_{i,t-k}} \right) \right\} - \lambda \sum_{j=1}^{p} \left| \beta_j \right|, \tag{2}$$

where p is the number of covariates used in the model and λ is a tuning parameter that controls the strength of the penalty term. The LASSO regularization shrinks the magnitude of all the coefficients, and sets the smallest ones to zero, thus selecting the most relevant covariates. Additionally, LASSO handles very well perfect and imperfect colinearity between variables. In the particular case of having perfectly collinear covariates, the algorithm selects one of the perfectly correlated variables and sets the rest to zero.

The implementation of the LASSO regularization requires assigning a value to λ. To find the optimal value of λ, I use a 10-fold cross-validation procedure. This is done by splitting the sample into a training sample and a test sample. Then, the training sample is further split into 10 equally sized subsamples. Given one value for λ, a LASSO estimator is obtained using 9 out of the 10 subsamples of the training sample, and the out-of-sample prediction error is computed using the remaining subsample. This procedure is repeated 10 times, each time leaving out a different subsample, and finally, an average out-of-sample prediction error across the 10 repetitions is computed. The optimal value of λ is the one that minimizes the average out-of-sample prediction error across the 10 repetitions. This optimization process is done for a grid of 100 values of λ between λ_0 and λ_{max}, where λ_{max} is the smallest value for λ such that all the coefficients are zero and λ_0 is defined by multiplying λ_{max} by 10^{-4}.

It is worth highlighting at this point that the specific setting of my analysis features a severe class imbalance since the majority of the observations are associated with firms that did not file for bankruptcy (see table 2). As a result, a predictive model that uses such a sample is likely to be biased towards predicting that firms do not file for bankruptcy ($Y_{i,t} = 0$), resulting in a model with high accuracy (as measured by the percent of correctly classified observations) but with a low power to predict filing for bankruptcy ($Y_{i,t} = 1$). To address this issue, I use a technique called *undersampling*, which consists of creating a perfectly balanced sample by randomly selecting a subset of observations from the ma-

[4]See section 11.4 of Murphy (2022) for a review of regression analysis using LASSO regularization.

jority class ($Y_{i,t} = 0$) to match the number of observations in the minority class ($Y_{i,t} = 1$).[5]

Table 2 reports the sample sizes in the *original*, *undersampling*, *training*, and *test* samples. First, note that the original sample contains fewer observations than the sample reported in subsection 3.1. This is because estimating the predictive model requires having lagged covariates for each observation of the dependent variable used in the estimation, which adds an additional requirement to the sample. Also, this sample does not contain observations of firms in the years after they file for bankruptcy. The original sample contains 9,412 firm-year observations for the class $Y_{i,t} = 1$ and 5,449,535 firm-year observations for the class $Y_{i,t} = 0$. After undersampling, the sample consists of a perfectly balanced sample of 18,824 firm-year observations. The undersampling sample is split into a training sample and a test sample, containing 2/3 and 1/3 of the observations in the undersampling sample, respectively. The training sample is used to select the optimal value of λ and estimate the model. The test sample is used to compute the out-of-sample prediction error of the selected model.

TABLE 2. Sample sizes in the original, undersampling, training, and test samples.

Sample	**Observations where**	
	$\boldsymbol{Y_{i,t} = 0}$	$\boldsymbol{Y_{i,t} = 1}$
Original	9,412	5,449,535
Undersampling	9,412	9,412
Training	6,299	6,250
Test	3,113	3,162

This table reports the number of firm-year observations in the original, undersampling, training, and test sample. The original sample contains fewer observations than those reported in subsection 3.1 because estimating the model requires having lagged covariates for each observation of the dependent variable used in the estimation, which adds an additional requirement to the sample. Also, the original sample does not contain observations of firms in the years after they file for bankruptcy. The undersampling sample is a subset of the original sample, and it was obtained by randomly selecting a subset of observations from the majority class ($Y_{i,t} = 0$) to match the number of observations in the minority class ($Y_{i,t} = 1$). The training and test samples contain 2/3 and 1/3 of the observations in the undersampling sample, respectively.

[5]See sections 5.1 and 10.3 of Murphy (2022) for a review of undersampling and other techniques to deal with class imbalance that involve changing the performance metric to assign a larger weight to the prediction error associated with the minority class. Another tradition in the literature (which started with Chawla, Bowyer, Hall, and Kegelmeyer (2002)) advocates for oversampling the minority class by creating synthetic observations. From this menu of different techniques to deal with class imbalance, I prefer undersampling for its simplicity.

4.3 Results

Figure 2 shows in the horizontal axis the values of λ that were considered (in log scale) and in the vertical axis the average and the standard deviation of the out-of-sample prediction error associated with each λ.[6] The two vertical dotted lines indicate, from left to right, the values of λ that minimize the out-of-sample prediction error (called λ_{min}) and the largest value of λ such that out-of-sample prediction error is within one standard deviation of the error associated with λ_{min} (called λ_1). Additionally, the numbers on top of the figure indicate how many non-zero covariates are left by each value of λ. Figure 2 shows that the average out-of-sample prediction error is minimized at $log(\lambda_{min}) = -7.67$, where the model contains 28 covariates and the binomial deviance equals 0.79.

In turn, figure 3 shows the variable importance of the 15 most important covariates in the model associated with λ_{min}. The measure of variable importance is the absolute value of the coefficients in the selected model. Note that this is a valid measure of variable importance because, following the standard practice in prediction exercises, the covariates have been standardized before inputting them into the model. The symbols on the right side of the figure indicate the sign of the coefficients. The figure shows that some of the most important predictors are measures of equity ($\Delta Equity_{i,t-1}$), profitability ($EBITDA_{i,t-1}$, $CF_{i,t-1}$), size ($Employees_{i,t-1}$, $Log(Assets)_{i,t-1}$), and dividends ($Net\ Dividends_{i,t-1}$). Notably, the growth rate of aggregate credit ($Creditgrowth_{t-1}$) is also among the most important predictors of filing for bankruptcy. Table 3 shows all the coefficients of the estimated model, and figure B2 in Appendix B plots all the parameters contained in table 3. Note that table 3 does not report the standard errors of the estimated coefficients because there is no straightforward way to compute them when using LASSO regularization. Nevertheless, in the context of this exercise, the predictive performance of the model is of more interest than are confidence intervals for the individual coefficients.

In the final step, I evaluate the model's predictive performance on the test sample. I compute the probability of bankruptcy for each observation within the test sample. Subsequently, any observation with a predicted probability exceeding 0.5 is predicted to belong to the class $Y_{i,t} = 1$, while the remaining observations are predicted to belong to the class $Y_{i,t} = 0$.

The predictive power of the model is assessed using the Receiver Operating Characteristic (ROC) curve and its corresponding Area Under the ROC Curve (AUC), which are shown in figure 4.[7] The ROC curve plots the true positive rate (TPR), which is the propor-

[6] Deviance is the standard prediction error measure in binary classification models. This measure is defined through the difference of the log-likelihoods between the fitted model and the actual binary outcomes.

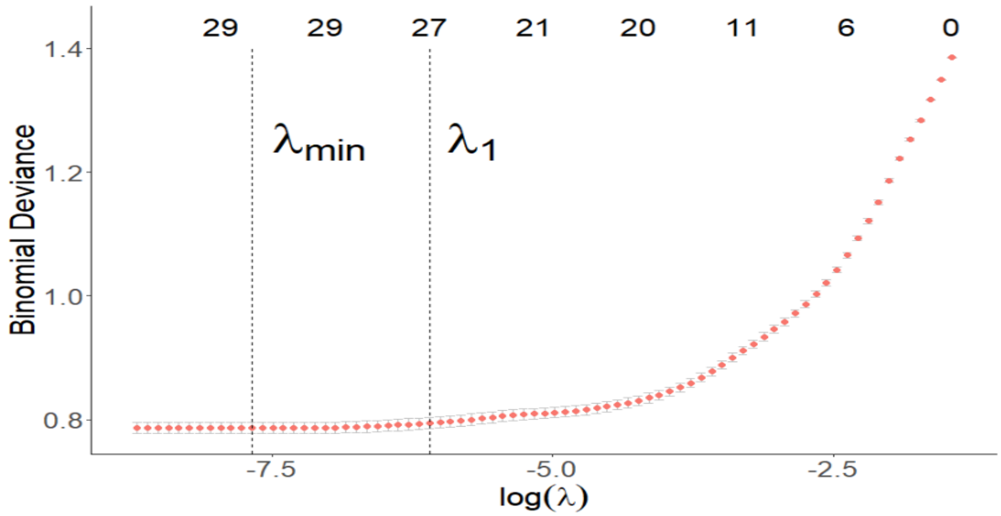
[7] See section 5.1.3 of Murphy (2022) for a review of these concepts.

TABLE 3. Estimated coefficients of the LASSO probit model associated with λ_{min}.

Covariate	Coefficient	Covariate	Coefficient
$\Delta Equity_{i,t-1}$	-1.923	$EBITDA_{i,t-1}$	1.397
$Log(Employees)_{i,t-1}$	-0.934	$Log(Assets)_{i,t-1}$	0.910
$Net\ Dividends_{i,t-1}$	-0.727	$CF_{i,t-1}$	-0.655
$Creditgrowth_{i,t-1}$	-0.534	$EBITDA_{i,t-1} < 0$	0.505
$\Delta CF_{i,t-1}$	0.491	$Equity_{i,t-1}$	-0.476
$\Delta EBITDA_{i,t-1}$	-0.433	$Cash_{i,t-1}$	-0.373
$Non\text{-}current\ liabilities_{i,t-1}$	0.310	$Tangible\ assets_{i,t-1}$	-0.298
$Account\ receivables_{i,t-1}$	0.292	$\Delta Current\ assets_{i,t-1}$	-0.178
$Sales\ growth_{i,t-1}$	-0.168	$Intangible\ assets_{i,t-1}$	0.166
$Equity_{i,t-1} < 0$	-0.133	$\Delta Capital_{i,t-1}$	-0.105
$\Delta Assets\ other\ than\ cash_{i,t-1}$	-0.099	$Age_{i,t-1}$	0.099
$\Delta Non\text{-}current\ liabilities_{i,t-1}$	-0.097	$\Delta Cash_{i,t-1}$	0.080
$\Delta Employees_{i,t-1}$	-0.075	$Current\ assets_{i,t-1}$	0.072
$Financial\ investment_{i,t-1}$	-0.014	$\Delta Tangibleassets_{i,t-1}$	-0.003
$GDPgrowth_{i,t-1}$	.	$\Delta Account\ receivables_{i,t-1}$	.
$\Delta Liabilities_{i,t-1}$	.	$Non\text{-}current\ assets_{i,t-1}$	.
$Capital_{i,t-1}$	.	$Liabilities_{i,t-1}$	.
$Current\ liabilities_{i,t-1}$	.	$Net\ debt_{i,t-1}$	.
$Working\ capital_{i,t-1}$	.	$\Delta Current\ liabilities_{i,t-1}$	.

This table reports the estimated parameters in the predictive model associated with λ_{min}. The variables without a reported estimate are those that were set to zero by the LASSO regularization. The variables are defined in table A1. All the variables are measured as ratios to average assets *(assets)* of each firm, except for *GDP growth*, *credit growth*, *EBITDA* < 0, *sales growth*, *Ln(assets)*, *age*, *employees*, *Δemployees*, and *Equity* < 0. These variables are measured as the growth rate of *GDP*, the growth rate of *aggregate credit*, a dummy variable that takes a value of 1 if the firm has negative *EBITDA*, the annual log difference in *sales*, the logarithm of *assets*, the difference between the reporting year and the year of incorporation, the number of *employees* in the firm, the annual difference in the number of *employees* in the firm, and a dummy variable that takes a value of 1 if the firm has negative *equity*, respectively.

FIGURE 2. Goodness of fit associated with each value of the tuning parameter λ.

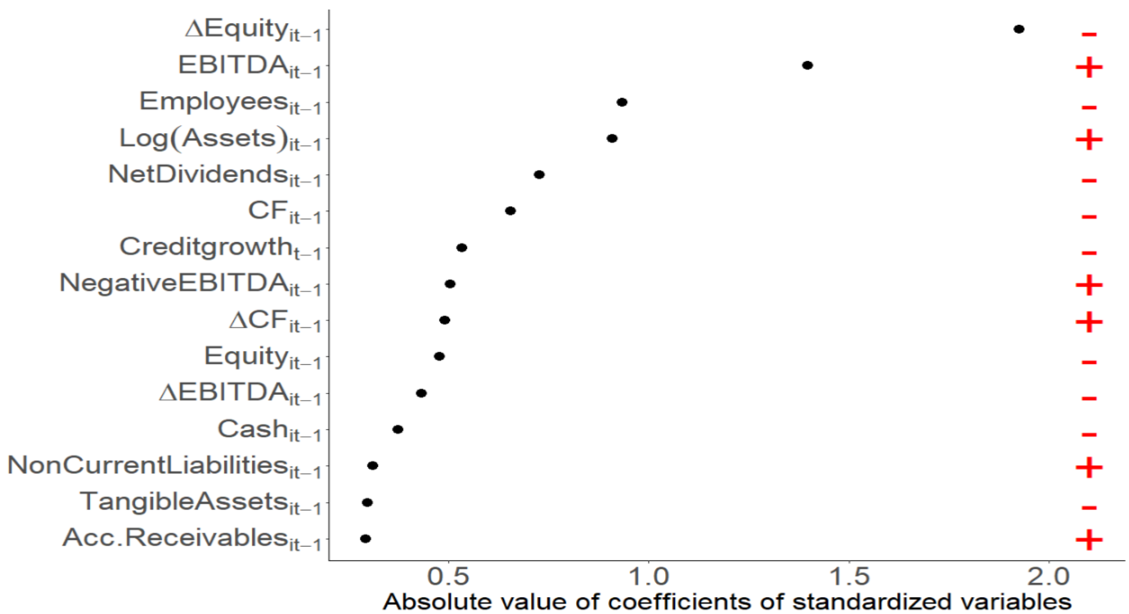


This figure reports the average binomial deviance associated with each value of the tuning parameter λ. Also, the standard deviations of the binomial deviance are reported. The two vertical lines indicate, from left to right, the values of λ that minimize the out-of-sample prediction error (λ_{min}) and the largest value of λ such that error is within one standard deviation of the error associated with λ_{min} (λ_1). The numbers on top of the figure indicate how many non-zero covariates are left for each value of λ.

tion of true positives that are correctly identified (that is, observations that are observed to belong to the class $Y_{i,t} = 1$ and are also predicted to belong to the class $Y_{i,t} = 1$), in the vertical axis, and the false positive rate (FPR), which is the proportion of false positives that are incorrectly identified (that is, observations that are observed to belong to the class $Y_{i,t} = 0$ but are predicted to belong to the class $Y_{i,t} = 1$), in the horizontal axis. The AUC measures the area under the ROC curve. A probabilistic interpretation of the AUC is that it measures the probability that the model ranks a randomly chosen positive occurrence as more likely positive than a randomly chosen negative occurrence. For reference, figure 4 also reports the 45° line, which would correspond to an uninformative model with an AUC equal to 0.5. In this case, the AUC is 0.905, which indicates that the model has a high power to predict bankruptcy one year ahead.

Finally, I also report the confusion matrix in table 4. The confusion matrix shows the number of true positives, true negatives, false positives, and false negatives. The confusion matrix allows me to compute an overall measure of the accuracy of the model, which is the ratio of the sum of true positives and true negatives over the total number of observations. This measure equals 82.8% ((2,631+2,563)/6,275), which indicates that the model has a high power to predict bankruptcy one year ahead. The confusion matrix also highlights that the model has high accuracy in predicting both $Y_{i,t} = 1$ and $Y_{i,t} = 0$ instances. The model correctly predicts 81.3% (2,563/3,152) of the $Y_{i,t} = 1$ instances and

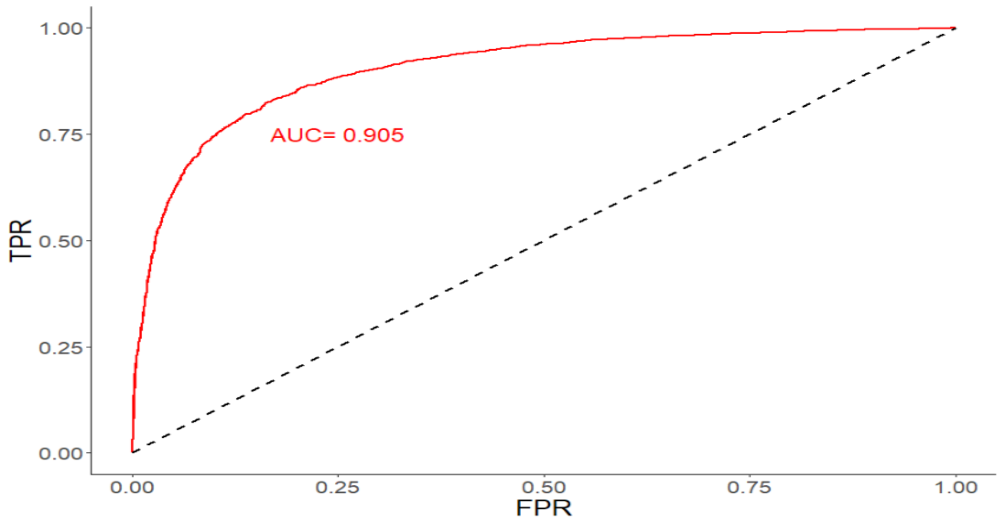
FIGURE 3. Importance of each of the 15 most important covariates in the predictive model associated with λ_{min}.



This figure plots the variable importance of the top 15 most important covariates in the predictive model associated with λ_{min}. The variable importance is computed as the absolute value of the coefficients of the selected model. Note that this is a valid measure of variable importance because, following the standard practice in prediction exercises, the covariates have been standardized before inputting them into the model. The symbols on the right side of the figure indicate the sign of the coefficients. The variables are defined in table A1. All the variables are measured as ratios to average assets *(assets)* of each firm, except for *Credit growth*, *Ln(Assets)*, *age*, *employees*, and *Equity* < 0. These variables are measured as the growth rate of *aggregate credit*, the logarithm of *assets*, the difference between the reporting year and the year of incorporation, the number of *employees* in the firm, and a dummy variable that takes a value of one if the firm has negative *equity*, respectively.

84.2% (2,631/3,123) of the $Y_{i,t} = 0$ instances.

FIGURE 4. ROC curve of the predictive model associated with λ_{min}.



This figure reports the receiver operating curve (ROC) curve of the predictive model associated with λ_{min}. The ROC curve plots the true positive rate (TPR), which is the proportion of true positives that are correctly identified, in the vertical axis, and the false positive rate (FPR), which is the proportion of false positives that are incorrectly identified, in the horizontal axis. The area under the curve (AUC) is also reported. The AUC is a measure of the predictive power of the model. A probabilistic interpretation of the AUC is that it measures the probability that the model ranks a randomly chosen positive occurrence as more likely positive than a randomly chosen negative occurrence.

TABLE 4. Confusion matrix of the predictive model associated with λ_{min}.

		Observed		
		0	1	Total
Predicted	0	2,631	589	3,230
	1	492	2,563	3,055
	Total	3,123	3,152	6,275

This table reports the confusion matrix of the predictive model associated with λ_{min}. The model is used to predict the observations in the test sample. The confusion matrix shows the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

5 Dynamics prior to bankrputcy

In this section, I present six facts about the dynamics of firms in Spain in the years leading up to filing for bankruptcy. To highlight what is special about the dynamics of these firms, I contrast them with the dynamics of other firms before exiting the market without filing for bankruptcy. Consistent with other articles studying firm exit using the CBI dataset, such as Budí-Ors (2024), I identify the year of exit as the last year in which the firm is present in the dataset. To ensure that the right censoring of the data does not incorrectly identify firms as exiting the market in the last years of my sample, I extend the sample to include the years 2020-2022 only to measure firm exit. Henceforth, I denote firms that eventually exit the market without filing for bankruptcy as *exiting* firms, and those that eventually file for bankruptcy as *bankrupt* firms.

Subsection 5.1 describes the empirical specification used to describe the dynamics of bankrupt and exiting firms. Subsection 5.2 presents the main results of the analysis and subsection 5.3 discusses additional results.

5.1 Empirical specification

I describe the dynamics of *bankrupt* and *exiting* firms using the following specification:

$$y_{i,t} = \sum_{k=0}^{\bar{k}} \left(\alpha_k^b \, Bankruptcy_{i,t}^k + \alpha_k^e Exit_{i,t}^k \right) + \lambda_s + \lambda_t + \epsilon_{it}, \tag{3}$$

where $y_{i,t}$ is the outcome variable of interest for firm i at year t. $Bankruptcy_{i,t}^k$ is an indicator variable taking the value of 1 if firm i files for bankruptcy k years ahead of year t, and 0 otherwise. Similarly, $Exit_{i,t}^k$ is an indicator variable with a value of 1 if firm i exits the market without filing for bankruptcy k years ahead of year t, and 0 otherwise. Finally, λ_s and λ_t are 2-digit industry fixed effects and year fixed effects, respectively.

The parameters of interest are α_k^b and α_k^e, which describe the dynamics of the outcome variable as firms approach the year of either bankruptcy or exit, respectively, after controlling for industry and year fixed effects. Note that these parameters measure differences with respect to the excluded category, which in this case are the firms that neither file for bankruptcy nor exit the market without filing for bankruptcy during the sample period.

To be concrete, equation (3) is estimated using a sample (derived from the one described in section 3.1) which contains observations from three groups of firms: First, firms that file for bankruptcy at some point during the sample period. Second, firms that exit the market without filing for bankruptcy at some point during the sample period. Third, firms that neither file for bankruptcy nor exit the market without filing for bankruptcy during the sample period. For the first two groups of firms, the sample is restricted to include at most five years before filing for bankruptcy or exiting the market, which means that $\bar{k} = 5$ in equation (3). The observations of firms after filing for bankruptcy or exiting the market are also excluded from the sample. Finally, I exclude observations of firms that report multiple bankruptcy dates. Although this might lead to a loss of information about firms that file for bankruptcy more than once, I do this on the grounds of preventing what could be a misreporting of the bankruptcy date.[8] The resulting sample contains 6,981,532 firm-year observations from 1,375,015 unique firms.

5.2 Main results

The results of estimating equation (3) are reported in figure 5. Panel (a) shows the evolution of earnings for bankrupt and exiting firms. Earnings are measured as the ratio of *EBITDA* to *output*, and the regression weights observations by contemporaneous *output*. The panel shows that bankrupt and exiting firms tend to have lower earnings than their industry peers and that this difference increases as they approach the year of either bankruptcy or exit. Moreover, bankrupt firms exhibit significantly lower earnings than exiting firms in the years leading up to bankruptcy.

Panel (b) shows the evolution of the equity of bankrupt and exiting firms. Equity is measured as the ratio of *equity* to *assets*, and the regression weights observations by contemporaneous *assets*. The panel shows that bankrupt and exiting firms tend to have lower equity than their industry peers. This is more patent for bankrupt firms, whose equity also tends to decrease in the years leading up to bankruptcy, in comparison with their industry peers. Given that equity is measured as the difference between assets and liabilities, an alternative reading of panel (b) is that bankrupt firms are significantly more leveraged (as measured by the ratio of liabilities over assets) than their industry peers and that such a gap increases as they approach the year of bankruptcy.

Panel (c) shows the evolution of the real investment of bankrupt and exiting firms. Real investment is measured as the growth rate of *capital*, and the regression weights observations by lagged *capital*. Panel (c) shows that the real investment rates of bankrupt and exiting firms seem to be slightly lower than the industry average, although the statistical significance of these differences is not clear. What is more, there are no significant differences in the real investment rates of bankrupt and exiting firms.

[8] The deletion of observations associated with firms that report multiple bankruptcy dates is unlikely to affect the results of the analysis, as these firms represent a negligible fraction of the sample.

Panel (d) shows the evolution of the growth rate of *employees* of bankrupt and exiting firms. In this regression, observations are weighted by the lagged number of employees. The panel shows that, compared to their industry peers, both types of firms exhibit significantly lower growth rates of employees in the years leading up to bankruptcy or exit. The growth rates of employees seem to be similar for bankrupt and exiting firms up to the year before bankruptcy when bankrupt firms begin to exhibit a significantly lower growth rate of employees.

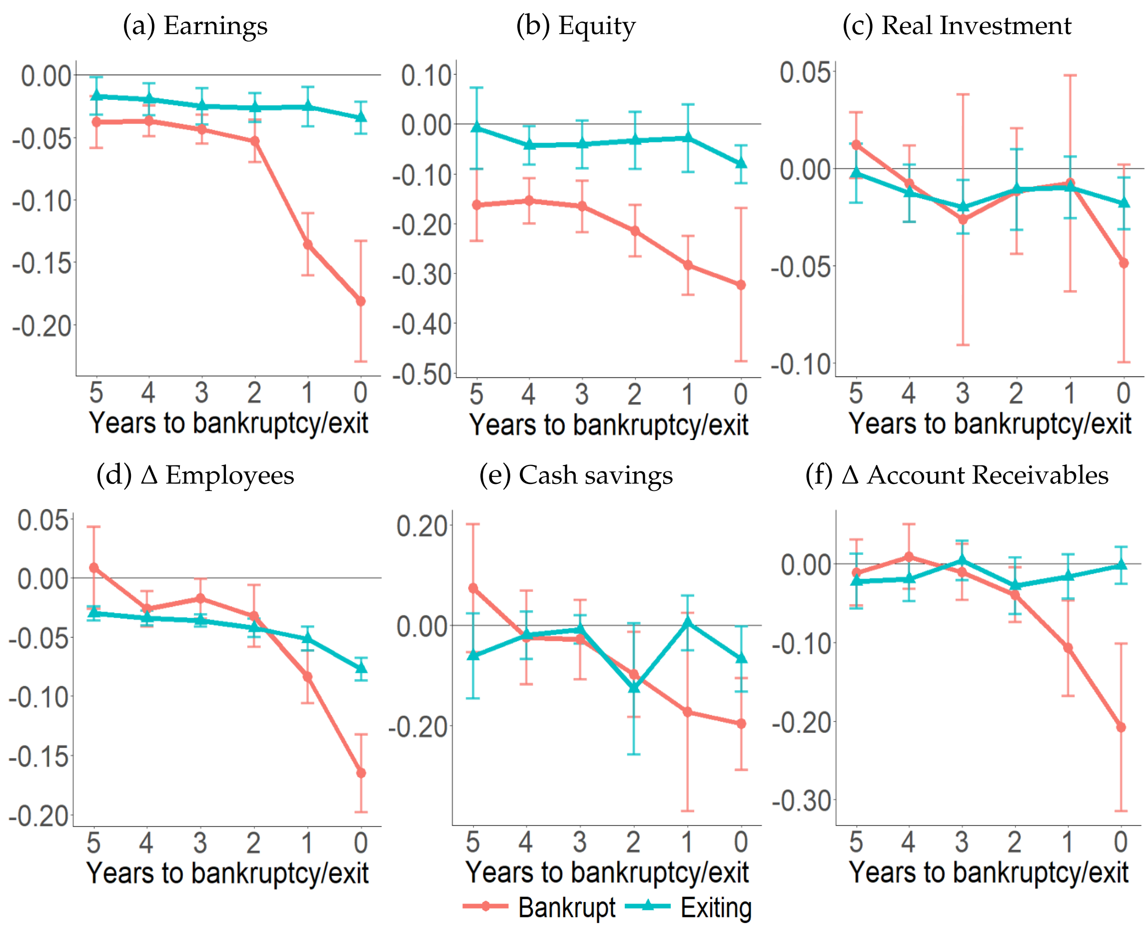
Panel (e) shows the evolution of cash savings of bankrupt and exiting firms. Cash savings are measured as the growth rate of cash holdings (*cash*), and the regression weights observations by lagged cash holdings. The panel shows that exiting firms do not seem to have significantly different cash savings than their industry peers in the years before exiting. In turn, bankrupt firms, in comparison with their industry peers, seem to exhibit a declining trend in their cash savings, and they start to have lower cash savings rates than their industry peers two years before filing for bankruptcy.

Panel (f) shows the evolution of the growth rate of *account receivables* of bankrupt and exiting firms. In this regression, observations are weighted by the lagged account receivables. This panel shows that exiting firms have similar growth rates of account receivables to their industry peers. Bankrupt firms, in turn, start to exhibit significantly lower growth rates of account receivables than their industry peers two years before filing for bankruptcy.

Discussion. Panels (a) and (b) together suggest that bankrupt firms arrive at the point of filing for bankruptcy in a very distressed financial situation. After controlling for industry and year fixed effects, bankrupt firms exhibit lower earnings and equity than exiting firms in the years before filing for bankruptcy. In particular, bankrupt firms experience substantial decreases in earnings and equity starting from the year before filing for bankruptcy. These findings are consistent with the widespread opinion in the literature that Spanish firms use bankruptcy as a measure of last resort (García-Posada, 2020).

The deterioration in the financial situation of bankrupt firms seems to be reflected in the other measures of real and financial performance, as bankrupt firms seem to exhibit declining investment rates, growth rates of employees, cash savings, and growth rates of account receivables in the years leading up to bankruptcy. However, panels (c)-(f) show that the measures of bankrupt firms are mostly similar to those of exiting firms. It is only in the year of filing for bankruptcy, or at most in the year before, that the weaker financial situation of bankrupt firms is accompanied by significantly lower growth rates of employees, cash holdings, and growth rates of account receivables than those of exiting firms.

FIGURE 5. Earnings, equity, and other measures of the real and financial performance of *bankrupt* and *exiting* firms.



This figure reports the estimated values of the parameters α_k^b and α_k^e in regression (3) for the outcome variables earnings, equity, real investment, Δ Employees, cash savings, and Δ Account Receivables. Year and industry fixed effects are controlled for by demeaning the outcome variables. Earnings are measured as *EBITDA* over *output*, equity as *equity* over *assets*, real investment as the growth rate of *capital*, Δ Employees as the growth rate of the number of employees (*employees*), cash savings as the growth rate of *cash*, and Δ Account Receivables as the growth rate of *account receivables*. All variables are defined in table A1. The bars centered around each coefficient represent the 95% confidence interval. Standard errors are clustered by industry. The regression weights observations by contemporaneous *output* for earnings, contemporaneous *assets* for equity, lagged *capital* for real investment, lagged *employees* for Δ Employees, lagged *cash* for cash savings, and lagged *account receivables* for Δ Account receivables.

5.3 Additional results

I explore whether the dynamics of bankrupt and exiting firms depicted in figure 5 diverge across some relevant dimensions. Specifically, I examine whether these dynamics vary based on the age at which firms file for bankruptcy or exit the market, whether such decisions are made during recessionary or expansionary periods, and the size of the firms.

Age. I extend the specification in equation (3) by interacting the indicators $Bankruptcy^k_{i,t}$ and $Exit^k_{i,t}$ with a dummy variable that takes a value of 1 if the firm files for bankruptcy or exits at an age above the median age of the sample, which is 20 years, and 0 otherwise. The estimated parameters measure differences with respect to the excluded category, which are the firms that neither file for bankruptcy nor exit the market during the sample period. The results are reported in figure 6. Column (a) shows the dynamics of bankrupt and exiting firms for firms filing for bankruptcy or exiting at an age between 0-20, while column (b) shows the dynamics of bankrupt and exiting firms for firms filing for bankruptcy or exiting at an age higher than 20.

Figure 6 shows that, after controlling for year and industry fixed effects, the dynamics of bankrupt and exiting firms are mostly similar across firms filing for bankruptcy or exiting the market with an age above and below the median age of the sample.

Recessions vs expansions. I extend the specification in equation (3) by interacting the indicators $Bankruptcy^k_{i,t}$ and $Exit^k_{i,t}$ with a dummy variable that takes a value of 1 if the firm files for bankruptcy or exits during a recession, and 0 otherwise. The estimated parameters measure differences with respect to the excluded category, which are the firms that neither file for bankruptcy nor exit the market during the sample period. The results are reported in figure 7. Column (a) shows the dynamics of bankrupt and exiting firms for firms filing for bankruptcy or exiting between the years 2008-2013, while column (b) shows the dynamics of bankrupt and exiting firms for firms filing for bankruptcy or exiting at any other time.

Figure 7 shows that, after controlling for year and industry fixed effects, the dynamics of bankrupt and exiting firms are mostly similar across firms that file for bankruptcy or exit the market during recessions and expansions. One aspect to highlight is that, compared with their industry peers, bankrupt firms seem to experience a more severe decrease in earnings and equity during expansions than during recessions, especially starting from the year before filing for bankruptcy. Consistent with this, in the year before filing for bankruptcy, firms tend to exhibit lower investment rates, growth rates of employees, and cash savings rates when they file for bankruptcy during expansions than during recessions.

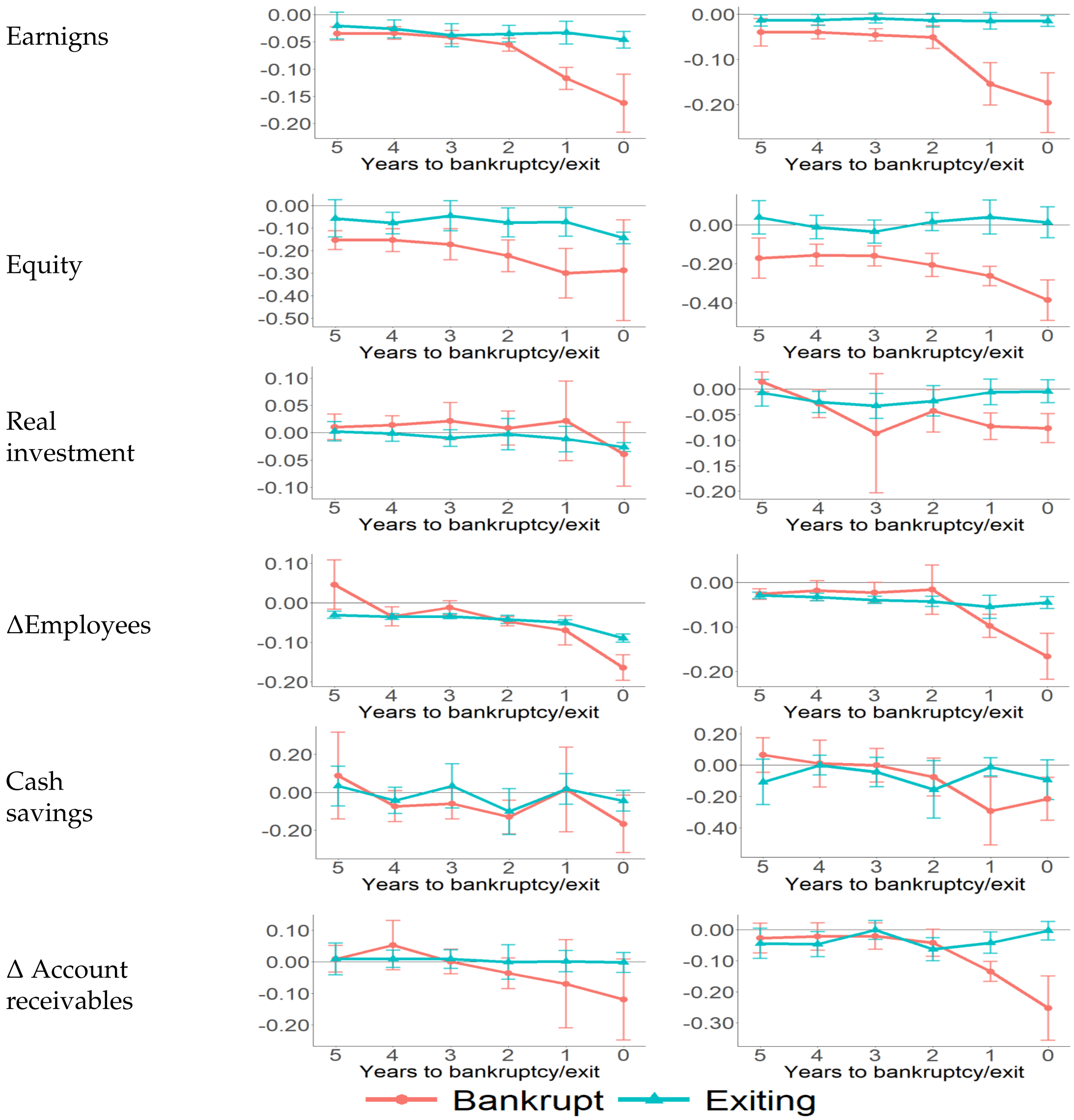
SMEs vs large firms. I extend the specification in equation (3) by interacting the indicators $Bankruptcy_{i,t}^{k}$ and $Exit_{i,t}^{k}$ with a dummy variable that takes a value of 1 if the firm is a small and medium enterprise (SME), and 0 otherwise (large firms).[9] For each firm, I use the first available observation in the last five years before filing for bankruptcy or exiting to categorize it as SME or large. The estimated parameters measure differences with respect to the excluded category, which are the firms that neither file for bankruptcy nor exit the market during the sample period. Column (a) shows the dynamics of bankrupt and exiting firms for SMEs, while column (b) shows the dynamics of bankrupt and exiting firms for large firms.

Figure 8 shows that, after controlling for year and industry fixed effects, the dynamics of bankrupt and exiting firms are mostly similar across SMEs and large firms. One notable difference is that large firms that exit the market do not exhibit lower earnings or equity than their industry peers in the years leading up to exit, while SMEs that exit the market do.

[9] Firms are categorized according to the criteria of European Recommendation 2003/361/EC.

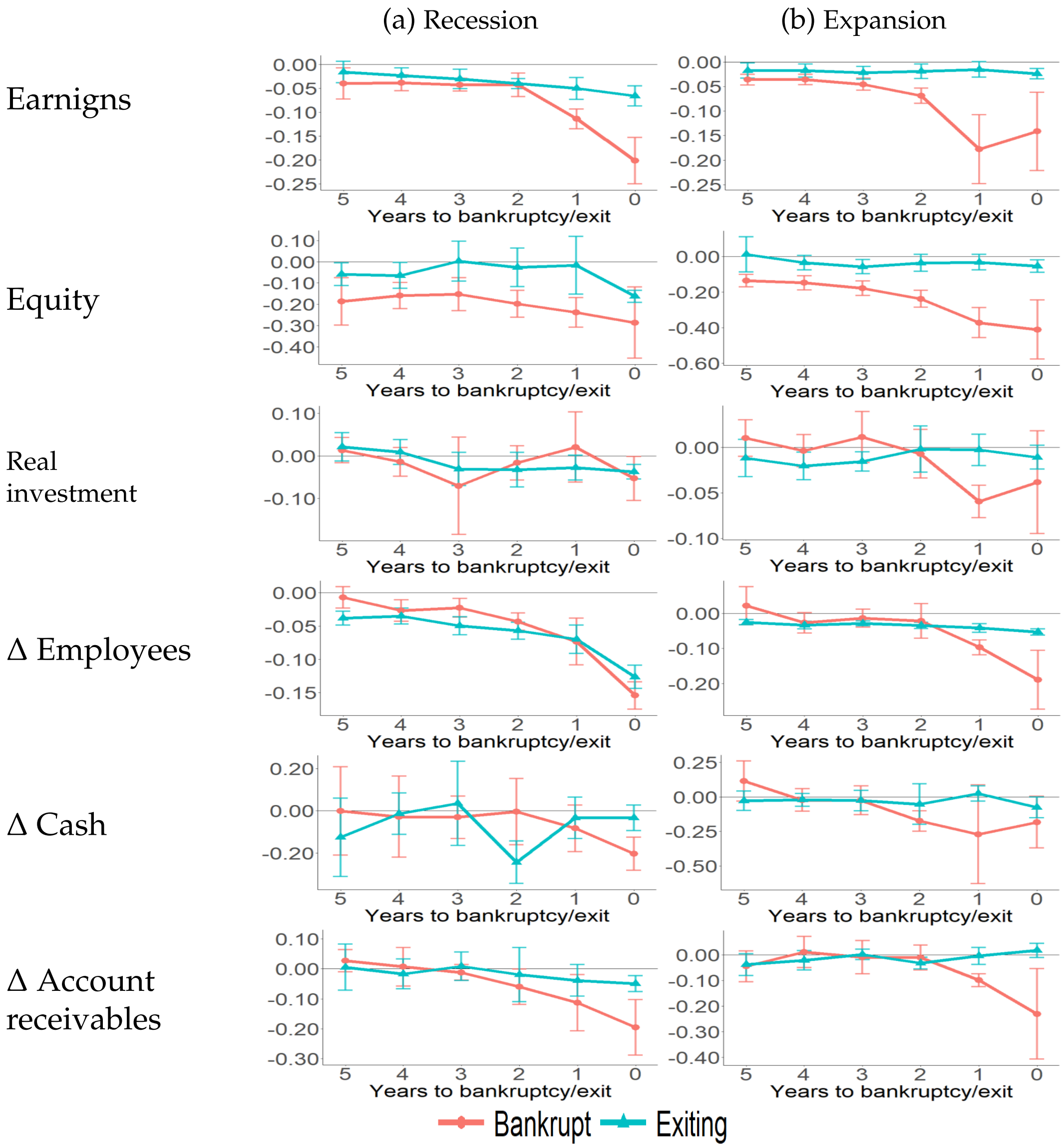
FIGURE 6. Earnings, equity, and other measures of the performance of *bankrupt* and *exiting* firms, for firms filing for bankruptcy or exiting at ages between 0-20, and at ages higher than 20.

(a) Firms filing for bankruptcy or exiting between ages 0-20 (b) Firms filing for bankruptcy or exiting with an age more than 20



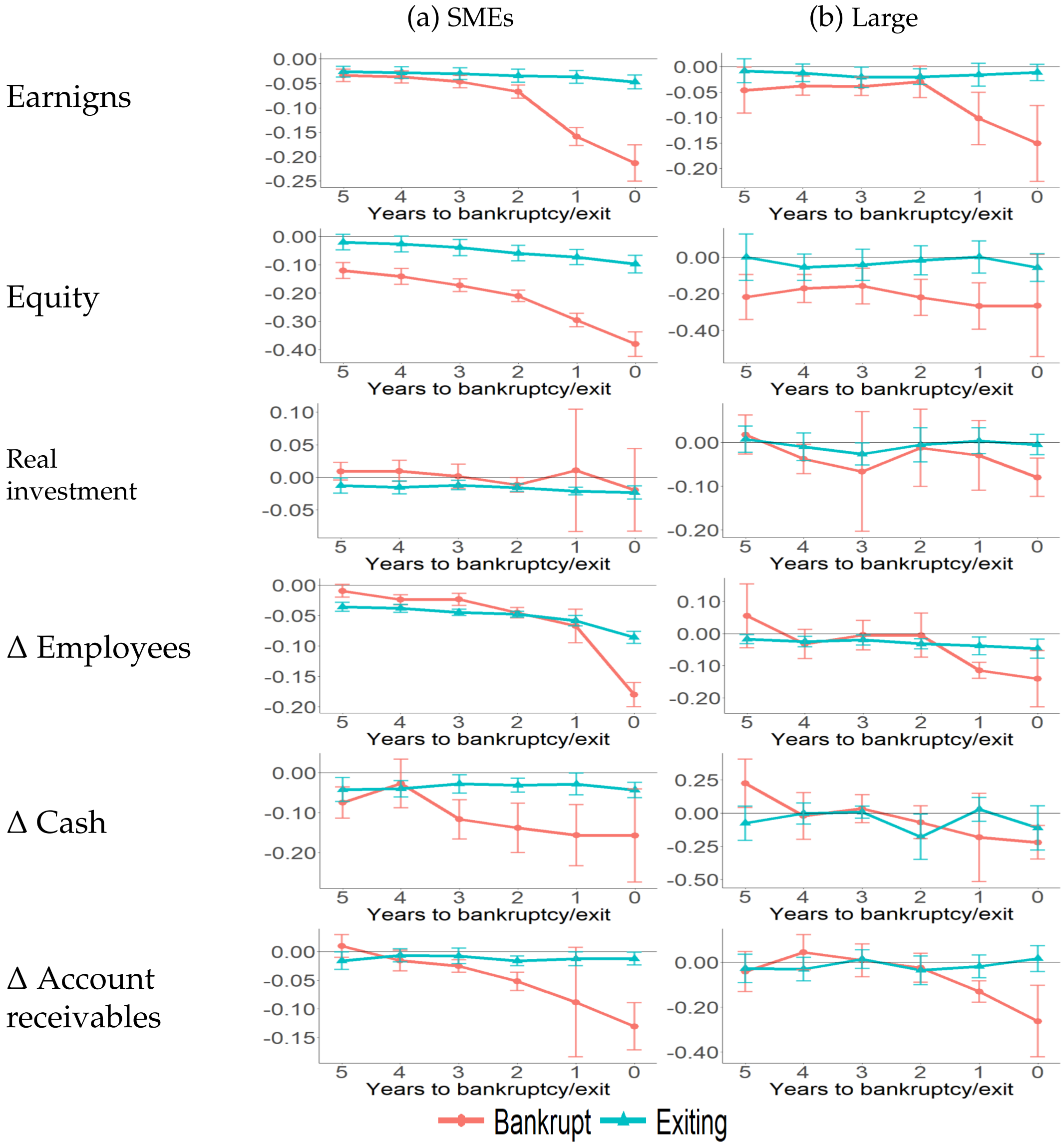
This figure reports the estimated parameters in a regression that results from extending equation (3). This is done by interacting the indicators of bankruptcy and exit with a dummy variable that takes a value of 1 if the firm files for bankruptcy or exits at an age above the median age of the sample, which is 20 years, and 0 otherwise. The resulting specification is estimated for the outcome variables earnings, equity, real investment, Δ Employees, cash savings, and Δ Account Receivables. Earnings are measured as *EBITDA* over *output*, equity as *equity* over *assets*, real investment as the growth rate of *capital*, Δ Employees as the growth rate of the number of employees (*employees*), cash savings as the growth rate of *cash*, and Δ Account Receivables as the growth rate of *account receivables*. All variables are defined in table A1. The bars centered around each coefficient represent the 95% confidence interval. Standard errors are clustered by industry. The regression weights observations by contemporaneous *output* for earnings, contemporaneous *assets* for equity, lagged *capital* for real investment, lagged *employees* for Δ Employees, lagged *cash* for cash savings, and lagged *account receivables* for Δ Account receivables.

FIGURE 7. Earnings, equity, and other measures of the performance of *bankrupt* and *exiting* firms, for firms filing for bankruptcy or exiting in recession and expansion periods.



This figure reports the estimated parameters in a regression that results from extending equation (3). This is done by interacting the indicators of bankruptcy and exit with a dummy variable that takes a value of 1 if the firm files for bankruptcy or exits during a recession, and 0 otherwise. The resulting specification is estimated for the outcome variables earnings, equity, real investment, Δ Employees, cash savings, and Δ Account Receivables. Year and industry fixed effects are controlled for by demeaning the outcome variables. Earnings are measured as *EBITDA* over *output*, equity as *equity* over *assets*, real investment as the growth rate of *capital*, Δ Employees as the growth rate of the number of employees (*employees*), cash savings as the growth rate of *cash*, and Δ Account Receivables as the growth rate of *account receivables*. All variables are defined in table A1. The bars centered around each coefficient represent the 95% confidence interval. Standard errors are clustered by industry. The regression weights observations by contemporaneous *output* for earnings, contemporaneous *assets* for equity, lagged *capital* for real investment, lagged *employees* for Δ Employees, lagged *cash* for cash savings, and lagged *account receivables* for Δ Account receivables.

FIGURE 8. Earnings, equity, and other measures of the performance of *bankrupt* and *exiting* firms, for SMEs and large firms.



This figure reports the estimated parameters in a regression that results from extending equation (3). This is done by interacting the indicators of bankruptcy and exit with a dummy variable that takes a value of 1 if the firm is a small and medium enterprise (SME), and 0 if it is a large firm. For each firm, I use the first available observation in the last five years before filing for bankruptcy or exiting to categorize it as SME or large. The resulting specification is estimated for the outcome variables earnings, equity, real investment, Δ Employees, cash savings, and Δ Account Receivables. Year and industry fixed effects are controlled for by demeaning the outcome variables. Earnings are measured as *EBITDA* over *output*, equity as *equity* over *assets*, real investment as the growth rate of *capital*, Δ Employees as the growth rate of the number of employees (*employees*), cash savings as the growth rate of *cash*, and Δ Account Receivables as the growth rate of *account receivables*. All variables are defined in table A1. The bars centered around each coefficient represent the 95% confidence interval. Standard errors are clustered by industry. The regression weights observations by contemporaneous *output* for earnings, contemporaneous *assets* for equity, lagged *capital* for real investment, lagged *employees* for Δ Employees, lagged *cash* for cash savings, and lagged *account receivables* for Δ Account receivables.

6 Conclusions

This article offers evidence of the behavior of Spanish firms preceding bankruptcy. I estimate a predictive model of bankruptcy which shows that the most important predictors of bankruptcy include (one-year lagged) measures of equity, profitability, size, and dividends. Interestingly, the growth rate of aggregate credit is also an important predictor of bankruptcy. I complement the insights provided by the predictive model with an exploration of the behavior of firms in the five years preceding bankruptcy (henceforth, *bankrupt* firms). To highlight what is special about the dynamics of these firms, I contrast them with the dynamics of other firms before exiting the market without filing for bankruptcy (henceforth, *exiting* firms).

I find that bankrupt firms arrive at the point of filing for bankruptcy in a very distressed financial situation. After controlling for industry and year fixed effects, bankrupt firms exhibit lower earnings and equity than exiting firms in the years before filing for bankruptcy. In particular, bankrupt firms experience substantial decreases in earnings and equity starting from the year before filing for bankruptcy.

The deterioration in the financial situation of bankrupt firms seems to be reflected in the other measures of real and financial performance, as bankrupt firms seem to exhibit declining investment rates, growth rates of employees, cash savings, and growth rates of account receivables in the years leading up to bankruptcy. However, panels (c)-(f) show that the measures of bankrupt firms are mostly similar to those of exiting firms. It is only in the year of filing for bankruptcy, or at most in the year before, that the weaker financial situation of bankrupt firms is accompanied by significantly lower growth rates of employees, cash savings, and growth rates of account receivables than those of exiting firms.

I further explore whether the dynamics of bankrupt and exiting firms commented on above diverge across some relevant dimensions. Specifically, I examine whether these dynamics vary based on the age at which firms file for bankruptcy or exit the market, whether such decisions are made during recessionary or expansionary periods, and the size of the firms. I find that, in general, the dynamics of bankrupt and exiting firms are mostly similar across these dimensions.

Overall, these findings are consistent with the widespread opinion in the literature that Spanish firms use bankruptcy as a measure of last resort (García-Posada, 2020). In other words, the financial distress of bankrupt firms is not generally the result of a sudden shock, but rather the result of a gradual process of deterioration in the financial situation of the firm. The similarities in the measures of real and financial performance of bankrupt and exiting firms, even though the former are in a worse financial situation, may be of interest to policymakers who want to understand how the properties of the Spanish bankruptcy law affect the incentives of firms to file for bankruptcy at a late stage of financial distress, and how these incentives may be different for firms with different characteristics or in different economic environments.

References

ALMUNIA, M., D. LOPEZ RODRIGUEZ, AND E. MORAL-BENITO (2018): "Evaluating the macro-representativeness of a firm-level database: an application for the Spanish economy," *Banco de Espana Ocassional Paper*.

ALTMAN, E. I. (1968): "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The journal of finance*, 23, 589–609.

——— (2013): "Predicting financial distress of companies: revisiting the Z-score and ZETA® models," in *Handbook of research methods and applications in empirical finance*, Edward Elgar Publishing, 428–456.

BEAVER, W. H. (1966): "Financial ratios as predictors of failure," *Journal of accounting research*, 71–111.

BLANCO, R., E. F. ORTIZ, M. GARCÍA-POSADA, AND S. MAYORDOMO (2024): "A New Estimation of Default Probabilities Based on Non-Performing Loans," *Finance Research Letters*, 105149.

BRIS, A., I. WELCH, AND N. ZHU (2006): "The costs of bankruptcy: Chapter 7 liquidation versus Chapter 11 reorganization," *The journal of finance*, 61, 1253–1303.

BUDÍ-ORS, T. (2024): "The Life-Cycle of Firms and the Productivity Advantages of Large Cities," *Mimeo*.

CELENTANI, M., M. GARCÍA-POSADA, AND F. GÓMEZ (2010): "The Spanish business bankruptcy puzzle and the crisis," *FEDEA Documento de Trabajo*, 11, 1–52.

CHAWLA, N. V., K. W. BOWYER, L. O. HALL, AND W. P. KEGELMEYER (2002): "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, 16, 321–357.

COOLEY, T. F. AND V. QUADRINI (2001): "Financial markets and firm dynamics," *American economic review*, 91, 1286–1310.

CORBAE, D. AND P. D'ERASMO (2021): "Reorganization or liquidation: Bankruptcy choice and firm dynamics," *The Review of Economic Studies*, 88, 2239–2274.

GARCÍA-POSADA, M. (2020): "Análisis de los procedimientos de insolvencia en españa en el contexto de la crisis del Covid-19: los concursos de acreedores, los preconcursos y la moratoria concursal," *Documentos Ocasionales/Banco de España, 2029*.

Garcia-Posada, M. and J. S. Mora-Sanguinetti (2012): “Why do Spanish firms rarely use the bankruptcy system? The role of the mortgage institution,” .

García-Posada, M. and J. S. Mora-Sanguinetti (2014): “Are there alternatives to bankruptcy? A study of small business distress in Spain,” *SERIEs*, 5, 287–332.

Gómez, M. G.-P. and R. V. Sánchez (2018): “Bankruptcy reforms in the midst of the Great Recession: The Spanish experience,” *International Review of Law and Economics*, 55, 71–95.

Hennessy, C. A. and T. M. Whited (2007): “How costly is external financing? Evidence from a structural estimation,” *The Journal of Finance*, 62, 1705–1745.

Murphy, K. P. (2022): *Probabilistic Machine Learning: An introduction*, MIT Press.

Shumway, T. (2001): “Forecasting bankruptcy more accurately: A simple hazard model,” *The journal of business*, 74, 101–124.

Appendix A: Data Appendix

This appendix outlines the steps undertaken in the cleaning process of the *Central de Balances Integrada* (CBI) dataset. These steps were executed to yield the sample of observations that form the basis of analysis in this study. The definitions of the variables mentioned below can be found in table A1.

The CBI contains the balance sheet and profit and loss account of Spanish corporations since 1995. Following Almunia et al. (2018), I exclude data from 1995 to 1999 because in these years the quality of the data was relatively poor and its coverage was limited. To prevent the events triggered by the pandemic from affecting my estimates, the data included in this study only extends until 2019.

Step 1. I delete observations that do not pass the quality tests applied by the provider, which are reflected in the variable *calidad* that takes a value of 1 if the observation passes the tests and 0 otherwise. The forms received from the companies are subject to a company-by-company filtering process to guarantee the quality and consistency of the information incorporated into the database. This implies that the raw data received by the managers of the CBI is not integrated into the database until it has passed numerous tests, both logical and arithmetical, which are aimed at guaranteeing internal and external consistency. The details of this filtering process can be found in the supplementary material that appends the annual publication of the main results in the CBI dataset by the Bank of Spain. This supplementary material is only available in Spanish (see, for example, *suplemento metodológico 2020*).

Step 2. Further to the quality controls applied by the data provider, I implement three filters to focus on private, for-profit, non-financial corporations. First, I use the first letter of the tax identifier (*cif*) to exclude entities that are arguably not-for-profit enterprises. Table A2 lists what type and how many observations were deleted in this step. Second, I use the *gsec09* variable, which represents the section codes of the National Classification of Economic Activities (CNAE-2009), to delete observations for firms belonging to the financial industry (K), the public administration (O), industries heavily influenced by the government (P, Q, U), or industries where firms are a minority compared to self-employed households (T). Table A3 provides details on what industries and how many observations were deleted in this step. Third, I use the *grup* variable, which is an identifier of government or non-government ownership of firms, to delete any remaining entities controlled by the government. Panel A of table A4 provides details on how many observations were deleted in this step.

Step 3. I remove all observations corresponding to firms that report dubious informa-

TABLE A1. Definition of variables.

Name	Definition
Sales	*Net turnover* (c200001) + *Other operating income* (c200006).
Inputs	*Net purchases* (c200010) + *Other operating costs*. (c200012).
Personnel Costs	*Personnel Costs* (c200025).
EBITDA	*Sales - Inputs - Personnel Costs*.
Net Financial Revenue	*Net Financial Revenue* (c290042).
Corporate income taxes	*Corporate income taxes* (c200069).
CF	*Sales - Inputs - Personnel Costs + Revenue - Corporate income taxes*.
Retained earnings	Retained earnings (c290088)
Provisions	*Provisions* (c200177).
Assets	*Assets* (c200135).
Cash	*Cash and equivalents* (c200129) + *Short-term financial investment* (c200128).
Adjustment	*Depreciation* (c200043) - *Gains on disposal* (c290059) - *Changes in fair value* (c290068) + Annual change in *Provisions* (c200177) - *Variation in stock of final goods* (c200003) - *Variation in stock of raw materials* (c200011) - *Tasks performed for asset* (c200005).
P&L	*Profit (loss) for the year* (c290070).
Equity	*Equity* (c200145).
Net dividends	*P&L* - annual change in *Equity*.
alta_sitc	Represents the date on which the firm files for bankruptcy (*alta_sitc*).
Year	Year associated with the information reported by the firm (any).
Year of incorporation	Year in which the firm was incorporated (anyconst).
Age	Year - Year of incorporation.
Calidad	Indicator if the firm complies or not with quality standards (calidad).
Cif	Tax identification number associated with the firm (cif).
Gsec09	CNAE 2009 section code. It has a length of one alphanumeric position (gsec09).
grup	Identifier of government or non-government ownership of firms (grup).
Employees	Average number of employees (units) (c200084).
Output	Value of output (c200075).
Current assets	Current assets (c200134).
Non-current assets	Non-current assets (c200115).
Assets	*Current assets* + *Current assets*.
Tangible assets	*Tangible assets and Property* (c200098).
Intangible assets	*Intangible assets* (c200089).
Capital	*Intangible assets* + *Tangible assets*.
Assets other than cash	*Assets - Cash*.
Sales growth	Annual change in log of *Sales*.
Current liabilities	*Current liabilities* (c200180).
Non-current liabilities	*Non-current liabilities* (c200158).
Liabilities	*Non-current Liabilities* + *Current Liabilities*.
Net debt	*Liabilities - Cash*.
Account receivables	*Trade and other receivables* (c200121).
Financial investment	*Long-term financial investment* (c200103).
Real investment	Annual change in *Capital*
Working capital	*Current assets* (c200134) - *Current liabilities*
GDP	Gross domestic product.
Aggregate credit	Aggregate amount of bank loans to non-financial sector.
Z-score	$0.717 * \frac{Working\ capital}{Assets} + 0.847 * \frac{Retained\ earning}{Assets} + 3.107\ \frac{EBITDA}{Assets} + 0.420\ \frac{Equity}{Liabilities} + 0.998 * \frac{Sales}{Assets}$.

This table reports the variables in the CBI dataset that were used to construct the variables used in the empirical analysis. The names under the "Name" column refer to the names used in this paper, while the names and the codes under the "Definition" column refer to the names and the codes of the variables in the CBI dataset. The codes are reported in parentheses. The definition of the Altman Z-score was taken from Altman (2013) and it constitutes a revised version of the original index that is adapted for the analysis of private firms.

tion on their *Year of incorporation*. Specifically, I exclude firms that report a *Year of incorporation* that is either negative or greater than any of the years in which the firm reports data.

Step 4. I apply a set of filters to remove observations that involve apparent reporting inconsistencies. To begin with, I only keep observations if the firm was observed in the year before. This is to ensure that all variables are well-defined since some of them are computed using differences over two consecutive years. Second, I delete observations with negative *output*, *sales*, *assets*, or *capital*. Third, I delete observations where *cash* is negative or larger than *assets*. Fourth, I eliminate observations that exhibit negative values for either *personnel costs* or *inputs*. Fifth, I delete observations where *current liabilities* or *non-current liabilities* is negative or larger than *liabilities*, and observations where *liabilities* are negative. Similarly, I delete observations where *current assets*, *non-current assets*, or *tangible assets* are negative or larger than *assets*, and observations where *assets* are negative. Sixth, I delete observations for which the sum of *equity*, *liabilities*, and *provisions* differs substantially from the value of *assets*. Concretely, I delete observations for which the ratio of the sum of equity, liabilities, and provisions to assets is larger (lower) than the percentile 99.9 (0.1) in the sample prior to this deletion. Additionally, I delete observations for which the ratio of the sum of the uses of cash flow to cash flow is larger (lower) than 1.01 (0.99).[10]

Step 5. I delete observations where any of the following variables (divided by contemporaneous *assets*) take values smaller (larger) than the 0.01 (0.99) percentile of the corresponding variable in the sample prior to this deletion: *CF*, *Real investment*, *Financial investment*, Δ*Account receivables*, Δ *Liabilities*, Δ *Cash*, *Net dividends*, *Liabilities*, *Current liabilities*, and *Non-current liabilities*. This last step aims to eliminate any remaining outliers that could potentially result from measurement errors in the reported data.

[10]The quality filter applied by the data provider admits arithmetic errors that are not substantial in magnitude. More details on this can be found in the supplementary material (*suplemento metodológico*). In the same spirit, I do not require the accounting identities to hold exactly in the data.

TABLE A2. Description of preserved and deleted observations based on tax identifiers.

	Code	**Firm-year Observations**	**Description**	**Description in Spanish**
Preserved entities	A	1,236,381	Corporation	*Sociedades anónimas*
	B	12,855,089	Limited liability company	*Sociedades de responsabilidad limitada*
	C	2,395	Business partnership	*Sociedades colectivas*
	D	1,242	Limited partnership	*Sociedades comanditarias*
	J	4,976	Civil society	*Sociedades civiles*
	U	1,758	Joint venture	*Uniones Temporales de Empresas*
	N	569	Foreign entity	*Entidades extranjeras*
	W	1,336	Branch entity	*Establecimientos permanentes de entidades no residentes en territorio español*
Deleted entities	E	966	Joint ownership, inheritance in abeyance, or other entity	*Comunidades de bienes, herencias yacentes y demás entidades carentes de personalidad jurídica no incluidas expresamente en otras claves*
	F	46,490	Cooperative society	*Sociedades cooperativas*
	G	14,070	Association	*Asociaciones*
	H	88	Residents' association under a horizontal property regime	*Comunidades de propietarios en régimen de propiedad horizontal*
	P	8	Local corporation	*Corporaciones Locales*
	Q	1,079	Public institution	*Organismos públicos*
	R	257	Religious institution	*Congregaciones e instituciones religiosas*
	S	13	State or Autonomous Community Institution	*Órganos de la Administración del Estado y de las Comunidades Autónomas*
	V	11,178	Other type undefined in another code	*Otros tipos no definidos en el resto de claves*
	Invalid	961	Observations having a cif with an invalid structure	

TABLE A3. Description of preserved and deleted industries.

	Code	Firm-year observations	Description
Preserved industries	A	351,080	Agriculture, livestock, forestry and fisheries
	B	32,676	Extractive industries
	C	1,454,482	Manufacturing industry
	D	217,692	Supply of electric energy, gas, steam and air conditioning
	E	37,238	Water supply, sanitation activities, waste management and decontamination
	F	2,390,040	Construction
	G	3,137,984	Wholesale and retail; repair of motor vehicles and motorcycles
	H	497,163	Transportation and storage
	I	821,118	Hospitality
	J	398,036	Information and communications
	L	1,443,678	Real estate activities
	M	1,465,583	Professional, scientific and technical activities
	N	562,090	Administrative activities and auxiliary services
	R	244,527	Artistic, recreational and entertainment activities
	S	256,332	Other services
Deleted industries	K	48,857	Financial and insurance activities
	O	0	Public administration and defense; compulsory social security
	P	206,896	Education
	Q	302,770	Health and social services activities
	T	0	Household activities as employers of domestic personnel; household activities as producers of goods and services for their own use
	U	0	Activities of extraterritorial organizations and organizations
	-	235,504	Missing data

TABLE A4. Selecting the final sample.

Panel A: Details on the number of observations deleted using the *grup* variable.

	Code	Firm-year observations	Description
Preserved firms	1	13,296,771	Private
Deleted firms	0	0	Not reported
	2	884	Government-owned (unclassified)
	3	1,922	Central government-owned
	4	1,992	Regional government-owned
	5	8,150	Local government-owned

Panel B: Details on the number of observations deleted in Step 4.

	Remaining firm-year Observations
4.1 Well-defined variables	9,907,701
4.2 Negative *output*, *sales*, *assets*, or *capital*	9,886,049
4.3 Negative *cash* or *cash* larger than *assets*	9,720,875
4.4 Negative *personnel costs* or *inputs*	9,709,901
4.5 Discrepancies in asset and liability structures	9,428,745
4.6 Discrepancies in main balance sheet components or cash flow identity	8,732,756

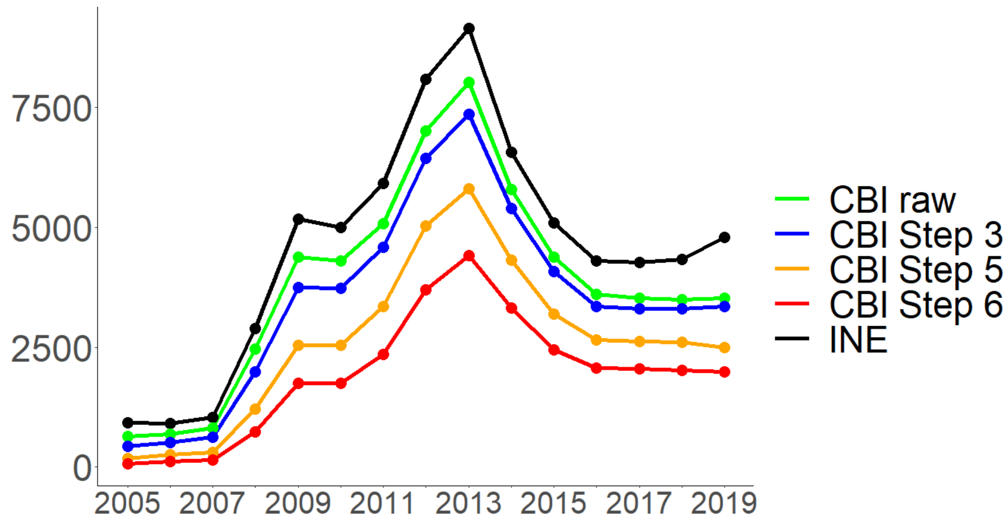
Panel C: Summary of cleaning steps.

Sequentially applied filters	Remaining firm-year observations
0. Initial sample	17,767,139
1. Preserving observations passing the quality filter (*calidad*)	14,178,981
2. Preserving for-profit firms in relevant industries	13,296,771
3. Deleting firms with dubious data on year of incorporation	12,952,141
4. Deleting observations with apparent reporting inconsistencies	8,732,756
5. Deleting outliers	7,837,901

This table contains details about the cleaning steps that are applied to the CBI dataset. Please, see the text of Appendix A for additional explanations.

Appendix B: Additional tables and figures

FIGURE B1. Number of bankruptcies in Spain (2005-2019) in the population (INE) and in several steps of the cleaning process of the CBI dataset.



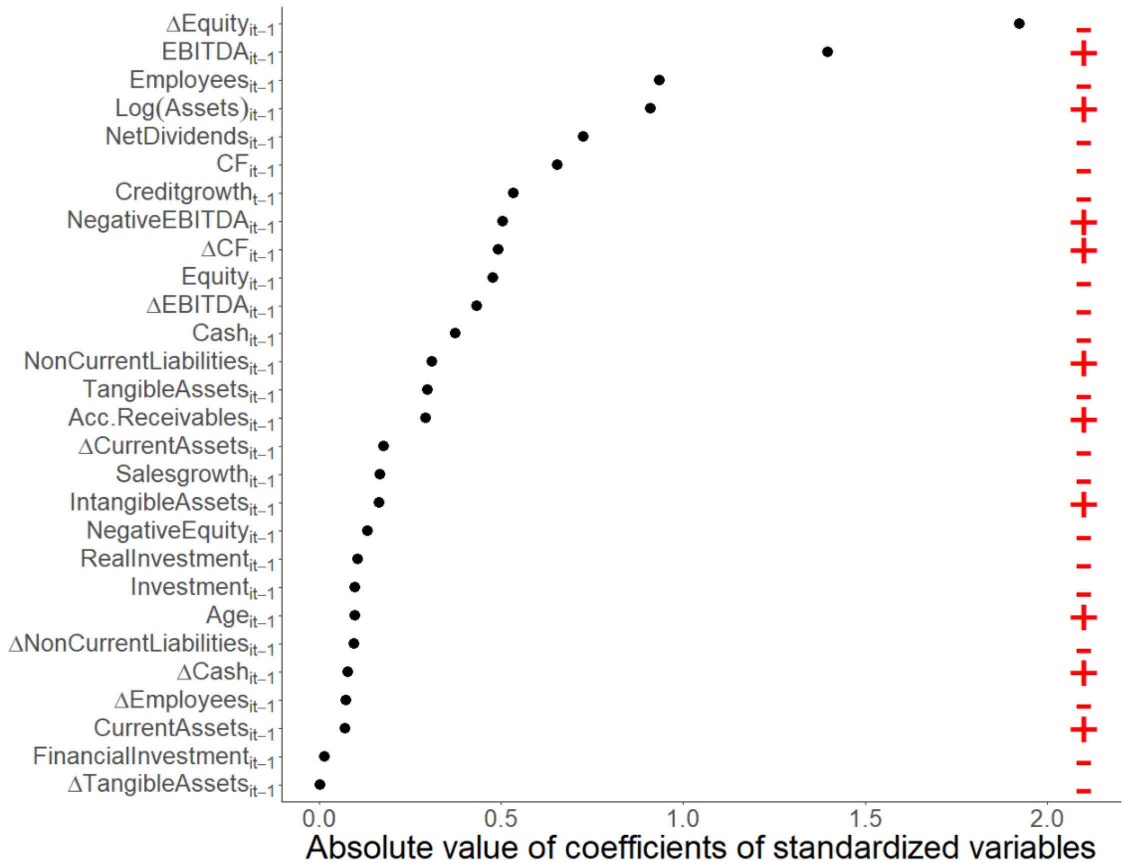
This figure shows information on the number of corporate bankruptcies in Spain between 2004 and 2019. Information about bankruptcy proceedings associated with individuals is excluded. The figure shows the total number of bankruptcies reported by the National Institute of Statistics (INE) and the number of bankruptcies contained in the CBI dataset at some of the cleaning steps mentioned in section 3.1. The *INE* line represents the total number of bankruptcies in Spain, while the lines *CBI raw*, *CBI Step 3*, and *CBI Step 5* represent the number of bankruptcies in the CBI dataset before cleaning, and after applying the steps 3 and 5 of the cleaning process explained in section 6, respectively. The line *CBI Step 6* represents the number of bankruptcies in the CBI dataset after keeping only the observations that are the observations used to estimate the predictive model of section 4.

TABLE B1. Variables used in the predictive model (1).

$\textit{GDP growth}_{t-1}$	$\textit{Tangible assets}_{i,t-1}$
$\textit{Credit growth}_{t-1}$	$\Delta\textit{Tangible assets}_{i,t-1}$
$\textit{EBITDA}_{i,t-1}$	$\textit{Liabilities}_{i,t-1}$
$\textit{EBITDA}_{i,t-1} < 0$	$\Delta\textit{Liabilities}_{i,t-1}$
$\Delta\textit{EBITDA}_{i,t-1}$	$\textit{Current liabilities}_{i,t-1}$
$\textit{CF}_{i,t-1}$	$\Delta\textit{Current liabilities}_{i,t-1}$
$\Delta\textit{CF}_{i,t-1}$	$\textit{Non-current liabilities}_{i,t-1}$
$\textit{Sales growth}_{i,t-1}$	$\Delta\textit{Non-current liabilities}_{i,t-1}$
$\log(\textit{assets})_{i,t-1}$	$\textit{Net debt}_{i,t-1}$
$\textit{Current assets}_{i,t-1}$	$\textit{Equity}_{i,t-1}$
$\Delta\textit{Current assets}_{i,t-1}$	$\Delta\textit{Equity}_{i,t-1}$
$\textit{Non-current assets}_{i,t-1}$	$\textit{Equity}_{i,t-1} < 0$
$\textit{Cash}_{i,t-1}$	$\Delta\textit{Assets other than cash}_{i,t-1}$
$\Delta\textit{Cash}_{i,t-1}$	$\textit{Financial investment}_{i,t-1}$
$\textit{Capital}_{i,t-1}$	$\textit{Account receivables}_{i,t-1}$
$\textit{Real investment}_{i,t-1}$	$\Delta\textit{Account receivables}_{i,t-1}$
$\textit{Intangible assets}_{i,t-1}$	$\textit{Net Dividends}_{i,t-1}$
$\textit{Age}_{i,t-1}$	$\textit{Employees}_{i,t-1}$
$\textit{Working capital}_{i,t-1}$	$\Delta\textit{Employees}_{i,t-1}$

This table reports the variables included in the vector $X_{i,t-k}$ in the predictive model (1). Δ denotes annual differences. The variables are defined in table A1. All the variables are measured as ratios to average assets *(assets)* of each firm, except for *GDP growth*, *Credit growth*, *EBITDA* < 0, *Sales growth*, *Ln(Assets)*, *age*, *employees*, Δ*employees*, and *Equity* < 0. These variables are measured as the growth rate of *GDP*, the growth rate of *aggregate credit*, a dummy variable that takes a value of 1 if the firm has negative *EBITDA*, the annual log difference in *sales*, the logarithm of *assets*, the difference between the reporting year and the year of incorporation, the number of *employees* in the firm, the annual difference in the number of *employees* in the firm, and a dummy variable that takes a value of 1 if the firm has negative *equity*, respectively. The variables have been standardized before inputting them into the predictive model. The variables are lagged by one year, i.e., $k = 1$.

FIGURE B2. Importance of each covariate in the predictive model associated with λ_{min}.



This figure plots the variable importance of each covariate in the predictive model associated with λ_{min}. The variable importance is computed as the absolute value of the coefficients of the selected model. Note that the absolute value of the coefficients is a valid measure of variable importance because, following the standard practice in prediction exercises, the covariates have been standardized before inputting them into the model. The variables are defined in table A1. All the variables are measured as ratios to average assets *(assets)* of each firm, except for *GDP growth*, *Credit growth*, *EBITDA* < 0, *Sales growth*, *Ln(Assets)*, *age*, *employees*, Δ*employees*, and *Equity* < 0. These variables are measured as the growth rate of *GDP*, the growth rate of *aggregate credit*, a dummy variable that takes a value of 1 if the firm has negative *EBITDA*, the annual log difference in *sales*, the logarithm of *assets*, the difference between the reporting year and the year of incorporation, the number of *employees* in the firm, the annual difference in the number of *employees* in the firm, and a dummy variable that takes a value of 1 if the firm has negative *equity*, respectively.