Economic networks and corporate default prediction

Andreea Constantin

March 2017

Abstract

This paper investigates the role of industry-specific effects and structural properties of intersectoral customer-supplier relations on the corporate default prediction of individual firms. We focus on a large sample of US exchange-listed companies over the period 1997-2015 and show that default prediction models that account for input-output network effects have better in-sample and out-of-sample accuracy compared to benchmark models that focus only on firm-specific and macroeconomic attributes. We find that companies' default intensities are related to the aggregate financial health of the industry in which they operate and the competition level of customer/supplier industries. Moreover, the prediction accuracy of the model is improved when we account for companies' role as main commodity suppliers in the aggregate economy, as well as their position in the structural flow of commodities. Second-order effects, related to customers' and suppliers' position in the sectoral network, also prove to be relevant.

JEL Classification: C53; D57; E44; G33

Keywords: default, input-output tables, distance-to-default, accuracy ratio, customer-supplier, competition, network centrality

*Contact: Andreea Constantin, Université de Lausanne and the Swiss Finance Institute, Dept. of Banking and Finance, Batiment Extranef, Quartier de Dorigny, 1015 Lausanne, Switzerland. E-mail address: andreea.piloiu@gmail.com. We thank Prof. Michael Rockinger for helpful comments.

1 Introduction

Current credit risk models focus on companies' risk profiles and macroeconomic conditions as determinants of default, while the influence of default by one firm on the default risk of another is not specifically explored. For example, in corporate default prediction models that rely on doubly-stochastic settings, the default correlation between firms is only driven by common exposures to macroeconomic factors or correlations across firm-specific attributes. This means that firms' defaults are conditionally independent and the model does not allow for a direct impact of the default of one firm on the default intensity of another firm.

This paper is the first to account for interconnections between firms' economic activities, seen here as intersectoral input-output linkages, in reduced-form credit risk models. This is motivated by recent theoretical results in **?** and **?**, that study the microfoundations of aggregate fluctuations. The authors show that the "diversification argument", according to which microeconomic idiosyncratic shocks average out in the aggregate, is not valid in the presence of a fat-tailed distribution of firm sizes and asymmetric sectoral input-output linkages. These observations are also documented empirically.

Our empirical analysis is based on the multi-period corporate default prediction approaches presented in **?** and **?**, which estimate a term structure of firms' default probabilities using firm-specific and macroeconomic state variables. Thus, we use a decomposable pseudo-likelihood function for components related to exits due to default (called *default exits*) and exits due to reasons other than default (called *other exits*) in order to estimate firms' default intensities. Following the literature on credit risk, we define the financial health of firms in terms of a volatility-adjusted measure of leverage, called the distance-to-default, and other macroeconomic and firm-specific attributes generally used in corporate prediction models. However, until now there are no means to account for the influence of other firms on a company's conditional default probabilities.

The main contribution of our study is to address current limitations of the default prediction literature by including cross-sectional and structural effects at industry level, which capture companies' exposures to financial conditions arising from economic ties to the rest of the economy. In such way, we allow for firms' estimated default intensities to depend on the financial health of the industry in which they operate or that of the industries with which they have customer-supplier relations.

Firstly, we take an industry-level perspective and look at the effects on firms' estimated default intensity of industry-specific attributes, such as concentration and average distance-to-default, which **?** claim to be statistically insignificant. Secondly, we make use of the structure of economic relations on which the US economy operates and test whether incorporating the health condition and network connections formed by industries that act as customers or suppliers can improve the performance of the credit risk model.

We start by analyzing the structural properties of the commodity-by-industry annual direct requirements tables, disaggregated at summary level, and observe that the network characteristics that focus on the supply side (such as the outdegree) are particularly heavy-tailed. Industries' concentration index also seems to have have a similar behaviour. Following the results in **?** and **?**, these distribution-based features could have an impact on the formation of aggregate fluctuations from industry-specific shocks and could affect corporate default probabilities.

Further, we construct various industry-level attributes based on centrality measures coming from the intersectoral customer-supplier network as well as combinations with industry-specific characteristics. Given that the intersectoral network is based on input-output tables of commodities needed for industrial production, this approach allows us to introduce a reflection of the topology behind the production-based real economy in a credit risk model. The justification behind using centrality measures relies on their ability to capture different aspects about industries' role in the flow of inputs and outputs relative to the whole economy. The analysis is performed for both first-degree and second-degree relations[1].

The credit model estimation is based on the maximum likelihood procedure and uses a large sample of quarterly observations for listed US industrial firms together with annual input-output tables at summary level for the US economy. The period covered is 1997-2015. As benchmark variables we use balance-sheet data, stock prices, and monetary

[1]First-degree relations refer to an industry's supply and demand ties with other industries, whereas second-degree relations refer to the supply and demand ties of an industry's customers and suppliers. **?** show that the rate of decay in the aggregate volatility of the US economy is affected by the degree of fatness in the tails of both the first-order and second-order outdegree distributions. Their analysis is based on the detailed level of input-output tables, covering 474 industries in 1997.

policy variables. We find that companies' default intensities are strongly related to the aggregate financial health of the industry in which they operate as well as the competition level. Moreover, industry' role as main commodity suppliers in the aggregate economy is relevant for default prediction. The results also provide evidence that network effects go beyond first-order connections and that accounting for second-degree connections helps improve further the performance of the credit risk model. Out-of-sample evaluations confirm the hypotheses that exposures to customers credit risk and supply chain disruptions on the suppliers' side are relevant predictors of companies credit risk. Moreover, insufficient diversification of the customer and supplier base as well as exposures to variations in the flow of commodities and payments in the aggregate economy also have an impact on companies' default intensities.

The remainder of the paper is structured as follows. Section **??** offers a discussion of the related literature. Section **??** presents the credit risk model and forward-intensity approach. In Section **??**, we present the data and methods used to estimate the reduced-form credit model. The main results are discussed in Section **??**. Finally, Section **??** concludes.

2 Related Literature

This paper touches on two strands of literature. Firstly, it uses a credit risk modeling setting, where a reduced-form approach is used to model firms' default intensities. Second, it explores the importance of network effects, that arise from companies' customer-supplier relations at sector level, on the in-sample and out-of-sample performance of the credit model.

In the academic literature, credit risk has been studied through the lens of structural models and reduced-form models. From the structural point of view, the capital structure of firms is the main determinant of default and a firm defaults when its assets are insufficient relative to its liabilities. Typically, the asset process is modeled as a geometric Brownian motion and the distance-to-default[2], which is a volatility-adjusted measure of

[2]The distance-to-default is computed following the observation that a company's equity can be seen as a call option written on the firm's assets, where the strike price is based on its liabilities. Roughly speaking, it represents the number of standard deviations of asset growth by which the market value of

leverage, determines a firm's conditional default probability (see ?, ?, ?). Moodys KMV (?) further developed the Black-Scholes-Merton structural model for industry practice and is now the most important provider of estimates of default probabilities for listed firms.

Considering the performance of the Black-Scholes-Merton structural model, empirical studies show that the default probabilities generated by the model are significantly less than the empirically observed default rates (?), they perform only marginally better as a predictor in hazard models and in out-of-sample forecasts compared to a naive alternative with the same functional form (see ?), and the model generates very large estimation error when estimating credit default swaps (CDS) spreads (?). In our paper, we will use the theoretical foundations of the Merton model and include firms' distance-to-default as an attribute in the reduced-form credit model, besides other balance-sheet, stock market and monetary policy related variables.

Reduced-form models consider default risk in terms of exogenous variables, such as firm-specific, sector-specific, and macroeconomic state variables. They were first introduced by ? and ? and initially relied on a discriminant analysis of companies' accounting variables. Further, reduced-form models used binary response models where historically observed default events are modeled with logit and probit regressions (see ?, ?, ?, and ?). Altman's Z-score and Ohlson's O-score are now widely known as measures of financial distress. More recent papers acknowledge that the probability of default of a firm depends on the prediction horizon (see ?, ?, ?). In this paper we use the doubly stochastic Poisson intensity model proposed by ? and further adopted by ?, whose maximum likelihood estimation that has the advantage of being decomposable between default and other exits components. We conduct a similar exercise, using a reduced-form econometric approach that is extended to include network and industry-level effects.

Moreover, the paper is related to the literature on economic and financial networks. This is a rather recent research topic in economics and finance which yielded mixed results prior to the 2007-2009 financial crisis. Pioneering theoretical work on financial networks, contagion, and systemic risk was done by ? and ?, who offer an identification of the stylized elements of financial systems. However, their conclusions that system stability

a firm's assets exceeds that of its outstanding liabilities.

increases with the number of links have been challenged as being dependent on the type of flows that take place between the nodes and the resultant network structures (see **?**, **?**, and **?**). **?** use bilateral interbank data for German banks and show that there is a highly persistent tiered core-periphery structure which was ignored in previous theoretical research. Moreover, flow maps associated with financial activities of Indian banks,[3] show that the networks representing contingent claims exposures (e.g. derivatives) are much more concentrated in the core compared to networks associated with credit-based interbank lending and borrowing; the least concentrated networks, which display limited clustering, are the payment and settlement systems (see **?** for the Austrian payment and settlement systems, **?** for Fedwire). Our study is also related to **?**, who combine network linkages with distress prediction for European banks.

Concerning input-output economies, this paper is related to **?** who analyze systemic risk in the context of linear intersectoral exposures. Their main result is to show that the distributions of outdegrees and second-order connections[4] are determining the rate at which the impact of idiosyncratic shocks on the aggregate volatility is vanishing. They conclude that the law of large numbers fails for asymmetric networks, which have power-law tail structures for nodes' outdegrees. However, their analysis is performed using the detailed level of input-output tables, covering 474 industries in 1997, whereas our study is based on less disaggregated data (64 industries excluding the governmental sectors) that spans a longer period of time (yearly coverage for 1997-2014).

3 Corporate default prediction model

In this section we describe the econometric model for estimating firms' default intensities, starting from the approach in **?** and **?**. We present in Appendix **??** an overview of the theory behind the conditional default probabilities and forward intensity rates and limit this section to describing the econometric model.

For making predictions about the corporate default rates we need to introduce the

[3] Reserve Bank of India Financial Stability Report, December 2011.

[4] The outdegree of a node in the intersectoral network summarizes the industry's importance as supplier to the rest of the economy. Second-order connections look further in the network, at the importance as suppliers (customers) of the industries that are economically linked to a specific node.

combined exit intensity g_{it} and default intensity f_{it}, related to two independent Poisson processes used to model companies' exit and default times. In our study, the two intensities are defined by the conditional probabilities of exiting the sample or defaulting during over the coming year. The combined exit intensity covers exit events related both to default and other exit reasons, so it will naturally include the default intensity. Further, we will model f_{it} and g_{it} as functions of state variables $X_{it} = (U_{it}, Y_t)$, that are available at time t. We let U_{it} represent the firm-specific variables of firm i, that cease to be observable after its exit time, and Y_t represent the vector of macroeconomic variables that are observable at all times. Following **?** we choose the exponential function for the two exit intensities, as it is non-negative and yields $f_{it} \leq g_{it}$[5]:

$$f_{it} = \exp\left(\alpha_0 + \alpha_1 \cdot x_{it,1} + \alpha_2 \cdot x_{it,2} + \alpha_3 \cdot x_{it,3} + ... + \alpha_k \cdot x_{it,k}\right) \tag{1}$$

$$g_{it} = f_{it} + \exp\left(\beta_0 + \beta_1 \cdot x_{it,1} + \beta_2 \cdot x_{it,2} + \beta_3 \cdot x_{it,3} + ... + \beta_k \cdot x_{it,k}\right) \tag{2}$$

Predicting default and other exit probabilities boils down to estimating the parameters $\alpha = \{\alpha_1, \alpha_2, ..., \alpha_k\}$ and $\beta = \{\beta_1, \beta_2, ..., \beta_k\}$ of the model specified in Equations **??** and **??**. The model specified in **?** uses a maximum likelihood function to estimate the parameters $\hat{\alpha}$ and $\hat{\beta}$, and is based on the assumption that different firms are X-conditionally independent. This implies that the default of one firm has no direct impact on the default intensity of another firm nor on the dynamics of state variables, and any dependency arises from the correlation among firm-specific variables or firms' exposure to common factors. The pseudo-likelihood function used for estimating the default and combined exit intensities is constructed as :

$$\mathcal{L}_\tau(\alpha, \beta; \tau_C, \tau_D, X) = \prod_{i=1}^{N} \prod_{t=0}^{T-1} \mathcal{L}_{\tau,i,t}(\alpha, \beta) \tag{3}$$

The likelihood function above can be decomposed in pseudo-likelihood components by separating the terms involving α, related to defaults, and β, related to other types of

[5]However, other non-negative functions that ensure the default intensity is no greater than the combined exit intensity can be used as well.

exit:

$$\mathcal{L}(\alpha) = \prod_{i=1}^{N}\prod_{t=0}^{T-1}\mathcal{L}_{i,t}(\alpha),$$

$$\mathcal{L}(\beta) = \prod_{i=1}^{N}\prod_{t=0}^{T-1}\mathcal{L}_{i,t}(\beta).$$

where

$$\begin{aligned}
\mathcal{L}_{i,t}(\alpha) &= \underbrace{\mathbf{1}_{t_{0i}\leq t,\tau_{Ci}>t+\tau}}_{(1)} \cdot \exp(-f_{it}) \\
&+ \underbrace{\mathbf{1}_{t_{0i}\leq t,\tau_{Ci}\neq\tau_{Di},\tau_{Ci}\leq t+\tau}}_{(2)} \cdot \exp(-f_{it}) \\
&+ \underbrace{\mathbf{1}_{t_{0i}\leq t,\tau_{Ci}=\tau_{Di}\leq t+\tau}}_{(3)} \cdot[1-\exp(-f_{it})] \\
&+ \underbrace{\mathbf{1}_{t_{0i}>t}}_{(4)} + \underbrace{\mathbf{1}_{\tau_{Ci}\leq t}}_{(5)} \\
\mathcal{L}_{i,t}(\beta) &= \mathbf{1}_{t_{0i}\leq t,\tau_{Ci}>t+\tau}\exp[-(g_{it}-f_{it})] \\
&+ \mathbf{1}_{t_{0i}\leq t,\tau_{Di}\neq\tau_{Ci},\tau_{Ci}\leq t+\tau}\cdot\exp[1-\exp[-(g_{it}-f_{it})]] \\
&+ \mathbf{1}_{t_{0i}\leq t,\tau_{Di}=\tau_{Ci}\leq t+\tau} + \mathbf{1}_{t_{0i}>t} + \mathbf{1}_{\tau_{Ci}\leq t}
\end{aligned}$$

In the above equations τ_{Ci} represents the combined exit time and τ_{Di} represents the time of default of company i. These two moments are modeled as stopping times of independent Poisson processes with stochatic intensities (see Appendix **??** for more details). The likelihood for each firm is a sum of indicator functions multiplied by their respective probabilities [6]. Hence, the likelihoods above specify five independent indicator functions which define five independent cases that can occur during the following year: (1) the firm does not exit the sample and is considered as surviving, (2) the firm exits due to other reasons than default, (3) the firm exits the sample due to default, (4) the firm has not yet entered the sample, and (5) the firm has already exited the sample.

Decomposing the likelihood function allows for a significant degree of tractability, as we will need to maximize separately the component related to the default intensity,

[6]The indicator function $\mathbf{1}_{A<B}$ is equal to one if $A < B$ or zero if $A \geq B$.

$\mathcal{L}(\alpha)$, from that related to other exit intensities, $\mathcal{L}(\beta)$[7]. Recall that for isolating the other exit intensities, which depend only on the β coefficients, we have to subtract the default intensities from the combined exit intensities.

The main contribution of this paper is to extended the econometric model to include new attributes for modeling f_{it} and g_{it} in order to capture network effects and the potential propagation of default risk across economically linked sectors. This approach, similar to the one used in **?**, allows firms' default intensities to respond to the financial health conditions in their sector and the ones with which they have customer-supplier relations. Firstly, we look at the effects of industry-average distance-to-default on firms' estimated default intensity and test the observation made by **?** that the industry-average of firms' distance-to-default is not a significant default predictor. Secondly, we make use of the structure of economic relations and test whether incorporating the health condition of industries that act as customers or suppliers can improve the performance of the credit risk model.

4 Data and Empirical Analysis

This section describes the datasets and the variables used in the default prediction model, a summary of their characteristics and analysis of their expected effects in the estimated model. The econometric model is estimated using quarterly data.

4.1 Data

The empirical implementation of the model is based on three types of data for the state variables used to estimate the intensities for default and other exits: firm-specific, macroeconomic, and sector level input-output data. We focus on US public firms, including financial institutions, over the period 1997-2015.

Firm-level data comes from Wharton Research Data Services (WRDS) using the CRSP/ Compustat merged database. Accounting, delisting, and industry classification data is taken from the Compustat quarterly file, while stock market data is taken from the CRSP daily files. The accounting items we collect are: total assets, cash and short-term

[7]See Proposition 2 in **?** for a similar approach.

investments, net income-loss, common shares outstanding, total long-term debt, and total current liabilities. We remove companies that have missing data for entire accounting variables, the delisting date and reason of deletion, or for the industry classification code. Also, we only use companies that have at least 8 quarterly observations and winsorize the accounting variables at 99.5% and 0.05%. In case an accounting observation is missing during the period a company appears in the database, we substitute it with the closest previous observation. The variable for the reason of deletion is coded from 1 to 10, out of which we separate code 2 (bankruptcy under Chapter 11) and code 3 (bankruptcy under Chapter 7) as being related to default, and consider the other codes as representing other types of exit. Companies with no deletion date are considered as surviving. Overall, our dataset consists of 8'886 companies that were active for at least 2 years during 1997-2015.

Table **??** shows for each year over 1997-2015 the number of companies in our sample that were active, as well as the number of bankruptcies and other types of exits. We can see that over time the number of listed, active companies in the US economy decreased continuously from about 6'000 in 1997 to 2'800 in 2015. The highest number of bankruptcies is observed between 1999-2002, where events such as the collapse of the dot-com bubble affected a large number of listed companies. Another increase in the number of bankruptcies takes place after the 2008-2009 financial crisis. Compared to **?**, who use numerous sources for the default and bankruptcy data (e.g. Bloomberg, Compustat, Moody's reports etc.), we can see that the number of defaults in our sample, which relies only on bankruptcies from Compustat, is smaller. This might affect the estimation of the corporate default model.

As macroeconomic variables we use the trailing 1-year return on the S&P500 index and the 3-month annualized US Treasury bill rate, which are taken from the WRDS CRSP files. The input-output data for US industries at summary level[8] is taken from the Bureau of Economic Analysis of the US Department of Commerce. We focus on the commodity-by-industry annual direct requirements table, denoted W, which indicate the commodity inputs required directly for a dollar of industry output. Each entry (I, J) in matrix W calculates the share of industry I's product in industry J's production

[8]The summary level input-output data is disaggregated over 64 industries, excluding governmental sectors.

technology. If we sum over the columns for row I, we obtain the outdegree of industry I, which corresponds to the industry's share in the input supply of the entire US economy.

4.2 Benchmark covariates

Firstly, we focus on a set of covariates that will be considered in the benchmark model and which is based on a selection of firm-specific and macroeconomic attributes similar to the ones used in **?** and **?**. Then we will introduce novel attributes based on the structure of the input-output industry-level tables, which will allow us to incorporate direct impacts from neighboring industries on the exit intensities of individual companies.

The covariates used in the benchmark model are listed below.

1. Trailing 1-year return on the S&500 index. We expect the estimated default intensity to decrease when the equity market performs well.

2. 3-month annualized US Treasury bill rate. We expect to have a negative relation between firms' default probabilities and the 3-month Tbill, due to the fact that lower rates are generally used during recessions in order to stimulate economic growth.

3. Trailing return: is the firms' stock return over the previous year. We expect to have a negative relation between the forward default intensity and firms' trailing return.

4. $Cash/TA$: is the ratio between cash and short-term investment to total assets; it characterizes the liquidity position of a firm, thus we expect to have a negative relation with the default intensity.

5. $NetInc/TA$: is the ratio between net income to total assets. A loss is registered as a negative net income. This ratio measures the profitability of a company, therefore we expect the default intensity to decrease when net income to total assets increases.

6. $Size$: is the logarithm of the ratio between a firm market equity value to the average market equity value of the whole S&P500. The market equity value is computed as the stock price multiplied by the number of shares outstanding at the end of each quarter. Firms' size is negative if it has a smaller market capitalization than the average market capitalization and positive otherwise. Given that large firms are

generally thought to have more financial flexibility and diversified operations, we expect to find a negative relation between size and default intensities[9].

7. *DtD*: the distance-to-default is based on companies' volatility-adjusted leverage and is used for assessing how far is the firm from the default point. It is measured in units of standard deviations of asset growth and is based on the structural model of Black-Scholes-Merton, which considers a firm's equity to be a call option written on the underlying assets and the strike being part of the debt level. We construct it similarly to **?** and **?**[10]. Previous studies[11] suggest that DtD is among the most important and significant predictors of credit risk; we expect to observe a negative relation between the default intensity and DtD.

8. *Mkt/Book*: market-to-book ratio is computed using the ratio between the market value and the book value of assets. The market value of assets comes from the construction of the DtD. The effect of the market-to-book ratio on the forward default intensity is uncertain, as this measure captures both the growth opportunities of companies and the extent of the market misvaluation. Depending on which of the two parts dominates, the relation with the default intensity should be negative for the former and negative for the latter.

Table **??** reports annualized descriptive statistics for the macroeconomic and firm-specific variables mentioned above and which represent the basis of the benchmark credit risk model. Table **??** presents correlations between the firm-specific variables. We can see that DtD is strongly correlated to $Size$ (46.5%) and also to $NetInc/TA$ (27.5%). Moreover, $Mkt/Book$ has an important correlation to $Cash/TA$ (31.5%) and DtD (27.3%). These results are in line with those reported by **?**, who also use Compustat balance sheet data for industrial firms. Compared to **?**, all results presented in our study use the current values of variables; we also tested the *level* and *trend* decomposition of firm-specific variables and found essentially the same results.

[9]However, **?** and **?** suggest that the relationship is not significant, while **?** finds that size is a significant determinant of default risk.

[10]More details are in Appendix **??**.

[11]See **?**, **?**, **?**, **?**.

4.3 Industry-specific covariates

In order to analyze the importance of possible cross-sectional effects on the default intensity of a company due to the financial health of its industry peers or due to the structure of economic ties arising from its operations, we test industry level attributes related to the distance-to-default, concentration, and network centrality. We will explain in this section how these attributes are constructed.

As the industry concentration and average distance-to-default rely on quarterly data, all variables based on these two attributes will also have a quarterly frequency. However, network centrality measures are built solely on the annual input-output tables, meaning that they will have only annual frequency. For the estimation of the econometric model, we transform all variables based on network centrality measures to quarterly frequency, by giving the same value to all quarters in a specific year. This implies that the variability of centrality-based attributes will be lower compared to the rest of the state variables, which could lead to an underestimation of their relevance in the respective model specifications.

Firstly, we assign each firm i to one of the 64 industries appearing in the input-output tables, based on its Nord American Industry Classification Code (NAICS). Table **??** details the number of companies that were part of each industry over the period 1997-2015, together with the defaults and other exits. Overall, we have 8'886 companies, that registered 341 defaults (3.84%) and 5415 other exits (61%). The industries with most listed companies are *Computer and electronic products* and *Chemical products*, which also register the highest number of defaults. However, the industries with high default frequencies are *General merchandise stores* and *Air transportation*, with about 16% of the companies in these industries defaulting over the period. Industries that performed well, with no defaults of listed companies, are *Utilities*, *Motor vehicles and part dealers*, several types of *Transportation*, *Legal services*, and *Amusements, gambling, and recreation industries* among others.

Even though the bankruptcy rate of 3.84% is in line with the one observed in the literature, not all defaults are reflected as balance sheet observations in our dataset, as some companies stop releasing quarterly balance sheet reports some time before they register default or go through bankruptcy proceedings. Overall, we only have about 25% of the bankrupted companies that filled quarterly information while being in bankruptcy proceedings. In order to better differentiate between healthy companies and distressed companies that end up bankrupt, we will consider companies that delisted due to bankruptcy to be in default status for one year prior to the default announcement. For simplicity, we will refer in the rest of the study to companies that exited the sample due to Chapter 11 or Chapter 7 bankruptcy as having defaulted.

For each of the industries we compute the average distance-to-default, its concentration index, and the centrality measures arising from the economic network in which it operates. The average distance-to-default for industry I in each quarter t is computed as in Equation **??**.

$$\overline{DtD}_{I,t} = \frac{1}{n_{I,t}} \sum_{i=1}^{n_{I,t}} DtD_{i,t} \tag{4}$$

$n_{I,t}$ is the number of companies in industry I and $DtD_{i,t}$ is the distance-to-default of company i for quarter t.

The concentration index of industry I is computed using the Herfindahl-Hirschman Index[12] and is shown in Equation **??**.

$$HHI_{I,t} = \sum_{i=1}^{n_{I,t}} (MS_{i,t})^2 \tag{5}$$

$MS_{i,t}$ represents the market share of firm i with respect to the whole industry. The index takes values in the interval $[0, 1]$. High values for $HHI_{I,t}$ indicate low competition and high market power of few companies in the industry. The concentration index and industry-average distance-to-default are computed at quarterly frequency.

Industries' centrality measures are based only on the annual input-output tables W_t, which can be represented by a directed weighted graph called the intersectoral network of the economy. Each node in the graph corresponds to an industry, and a directed link (I, J) with weight $w_{IJ} > 0$ exists from node I to node J if industry I is an input supplier to industry J. We will use the notions of intersectoral network and input-output matrices to refer to the structure of intersectoral flows.

[12]The index was developed independently by the economists A.O. Hirschman (in 1945) and O.C. Herfindahl (in 1950). Hirschman presented the index in his book, *National Power and the Structure of Foreign Trade* (Berkeley: University of California Press, 1945). Herfindahl's index was presented in his unpublished doctoral dissertation, *Concentration in the U.S. Steel Industry* (Columbia University, 1950).

The general concept of centrality is used for different aspects of the 'importance' or 'influence' of nodes within a network. The centrality measures we are interested in capture the opportunity of an industry to influence others as an important supplier (outdegree) or customer (indegree) and its exposures to risk (eigencentrality) or flow of resources (betweenness).

The weighted outdegree of sector I at time t is defined as the share of sector I's output in the input supply of the entire economy:

$$OutDeg_{I,t} = \sum_{J \neq I} w_{IJ,t} \tag{6}$$

where the sum considers the $N-1$ industries J, other than I, in the intersectoral network. The outdegree is related to sector I's role as supplier and is computed by summing the weights it directs to its customers (outgoing links).

The weighted indegree is defined as the share of inputs from the rest of the economy used by industry I with respect to its total input needs. It is computed by summing the incoming supply links and is related to sector I's role as customer:

$$InDeg_{I,t} = \sum_{J \neq I} w_{JI,t} \tag{7}$$

The outdegree and indegree can hardly be considered centrality measures, as we do not need to know the whole network structure in order to calculate them for each node.

Betweenness centrality measures the importance of nodes in terms of the flow they control, in the sense that nodes with high betweenness are along many of the shortest paths between all possible pairs of nodes. It can also be interpreted as an index of frequency for commodities' variability reaching a node. It is computed based on the shares of shortest distances between pairs of nodes that pass through a specific node. For node I we have:

$$BTW_{I,t} = \sum_{J,K \neq I} \frac{g_{J,I,K}}{g_{J,K}} \tag{8}$$

where $g_{J,K}$ is the number of shortest paths between nodes J and K and the distance or length of a path is the sum of the weights of its edges.[13]

Eigencentrality, also called eigenvector centrality, is defined as the principal eigenvector of the input-output matrix. It can be seen as a weighted degree measure in which the centrality of a node is proportional to the sum of centralities of its suppliers or customers. Depending on whether we use matrix W, where entry (I, J) is from supplier I to customer J, or its transpose, there is the supplier-based eigencentrality (also called the in-eigenvector) or the customer-based one (out-eigenvector):

$$\begin{aligned} EigenCent^{C}_{I,t} \equiv \nu^{C}_{I,t} &= \frac{1}{\lambda^{C}_{t}} \sum_{J \neq I} w_{IJ,t} \cdot \nu^{C}_{J,t} \\ EigenCent^{S}_{I,t} \equiv \nu^{S}_{I,t} &= \frac{1}{\lambda^{S}_{t}} \sum_{J \neq I} w_{JI,t} \cdot \nu^{S}_{J,t} \end{aligned} \tag{9}$$

where λ^{S}_{t} is the biggest eigenvalue of the input-output matrix W at time t, and is associated with the eigenvector ν^{S}_{t}.[14] The relationship between the eigenvalue and corresponding eigenvector of matrix W can also be represented as $W \cdot \nu = \lambda \cdot \nu$.

Table **??** reports descriptive statistics for industries' concentration and centrality measures. We can see that the average concentration index for the 64 industries in our sample is about 23%, with a minimum of 2% and a maximum of almost 100%, meaning that over the years there are some industries with almost no competition at all and where the market power is controlled by few companies. However, given that the 75th percentile for the concentration index is 27.7%, most industries' are formed by many firms that share the market power. Concerning the centrality measures, the skewness and kurtosis indicate that outdegree and customer-based eigencentrality are the most asymmetric and display the heaviest tails. This means that in terms of out-going links, that proxy for the influence as commodities supplier to the rest of the network, the industries show important heterogeneity, where a small number of sectors have a disproportionately important part as input suppliers to others. These distribution-based observations, coupled with results in **?** and **?**, point to the possibility of aggregate fluctuations forming from industry-specific shocks, which could impact corporate default probabilities.

In network theory, the underlying degree distribution of a network is crucial in de-

[13] See the algorithm explained in **?**.

[14] Similarly, for λ^{C}_{t} and ν^{C}_{t} we use the transpose of W.

termining its resilience to shocks. Theoretical research on networks mainly focuses on two types of networks: Poisson (or random) networks and power-law (also called scale free) networks. The main difference between the two is that in the random network the links have the same probability to exist and the nodes have on average similar degrees, independent of network's history, whereas power-law networks typically have other characteristics like clustering and correlations between degrees. Moreover, depending on the types of flows depicted by the network, power-law degree distributions provide high resilience to random failure but a high sensitivity to attack strategies that focus on highly connected nodes, whereas Poisson degree distributions are similarly sensitive in both cases. Most real-world networks, such as financial networks, the world-wide-web, and input-output tables are documented to be power-law networks, where the degrees are usually well fitted by power-law distribution with an exponent between 2 and 3.[15] Figure **??** shows in detail the evolution over the period 1997 to 2014 of industries' concentration and customer-based centrality distributions. The distribution of the outdegree does not change much over time in terms of main shape, however the number of industries considered outliers tends to increase. Industries' distributions along customer-based eigencentrality and concentration are also quite stable over time.

Figure **??** analyzes further whether the outdegree of the nodes for the intersectoral network has a distribution that behaves as a power-law[16]. The figure illustrates the double logarithmic plot of the empirical distribution of outdegrees and their respective fitted power-law model for a selection of years spanning 1999-2014. As in the case of other real-world economic and financial networks (e.g. the Brazilian financial network analyzed by **?**, and the detailed input-output tables for the US economy by **?**), the tails of the outdegree distributions exhibit a linear decay in log-scale, with small changes over the years. This suggests that the networks are not random, but rather have Pareto tails for the degree distribution. This characteristic is specific for most real-world complex

[15] **?** shows that for a power-law (Pareto) distribution $x^{-\zeta}$, all moments $m \geq \zeta - 1$ diverge. This means that it has a well-defined mean over $x \in [1, \infty]$ only if $\zeta > 2$, and it has a finite variance only if $\zeta > 3$. Most identified power laws in nature have exponents such that the mean is well-defined but the variance is not, meaning that they are prone to black swan behavior, where unexpected tail events are consequential.

[16] **?** show that if the empirical distribution of outdegrees of the intersectoral network can be approximated by a Pareto (or power-law) distribution with shape parameter $\zeta \in (1, 2)$, the aggregate volatility in the economy decays at a rate slower than $\frac{\zeta - 1}{\zeta}$.

network, where the nodes display widely differing degrees.

Using the approach proposed by **?** we estimate, using maximum likelihood, the tail exponent ζ and the tail threshold x_m for power-law distributions fitted to outdegrees. The results in Figures **??** and **??** provide evidence for the Pareto tail hypothesis. Figure **??** shows the estimates and their respective standard errors of the power-law shape parameter for industries' outdegrees, over the period 1997 to 2014. We can see that in the case of output-input tables disaggregated at 64 industries, the shape parameter (or exponent) is always smaller than 2.5, and for about half of the years it could statistically be smaller than 2. This means that in all cases, the variance of the distribution is undefined and shocks to industries that play an important role as commodities suppliers will not diversify away and can lead to aggregate fluctuations. The differences with respect to **?**, whose analysis yields tail exponents smaller than 1.5 for detailed input-output tables, may come from the level of data disaggregation (more than 400 industries compared to 64 in our study), as it is well known that more aggregation leads to more uniform characteristics and smoothing of variability across industries. To a lesser extent, the different exponents may also be due to different parametrization of estimations.

Given that supply chains provide the channels of distress transmission across industries, we will further analyze the hypothesis that the importance of industries in the input-output structure of the economy has effects on companies' default intensities. Finally, the new covariates tested in the credit risk model are based on assigning firms to industries and analyzing the following structural and cross-sectional effects:

- Industry-specific variables (ISV): we include variables that capture the industry context in which firms' operate. We test the average distance-to-default $\overline{DtD}_{I,t}$ and the concentration index $HHI_{I,t}$.

- First-order customer-suppliers connections: based on the input-output matrices, we test the importance of industries' position in the intersectoral network as measured by $InDeg_{I,t}$ (commodity inputs coming from suppliers), $OutDeg_{I,t}$ (commodity outputs going to customers), betweenness, and customer and supplier-based eigencentrality. Moreover, we also construct attributes for each insdustry by combining their customers' input shares or suppliers' output shares with their concentration index and average distance-to-default:

$$\begin{aligned} Customer^{ISV}_{I,t} &= \sum_{J\neq I} w_{IJ,t} \cdot ISV_{J,t} \\ Supplier^{ISV}_{I,t} &= \sum_{J\neq I} w_{JI,t} \cdot ISV_{J,t} \end{aligned} \tag{10}$$

where ISV stands for the industry-specific variables mentioned above. We thus have $Customer^{\overline{DtD}}_{I,t}$, $Supplier^{\overline{DtD}}_{I,t}$, $Customer^{HHI}_{I,t}$, and $Supplier^{HHI}_{I,t}$.

- Second-order customer-supplier connections: for each industry, we take into account possible cascade effects by looking one step further in the network, at the customers and suppliers of the immediately neighboring industries. We compute the following attributes for industry specific variables:

$$\begin{aligned} Customer2nd^{ISV}_{I,t} &= \sum_{J\neq I} \sum_{K\neq J} w_{IJ,t} \cdot w_{JK,t} \cdot ISV_{K,t} \\ Supplier2nd^{ISV}_{I,t} &= \sum_{J\neq I} \sum_{K\neq J} w_{JI,t} \cdot w_{KJ,t} \cdot ISV_{K,t} \end{aligned} \tag{11}$$

where $Customer2nd_{I,t}$ stands for the customers of the customers of industry I and $Supplier2nd_{I,t}$ for the suppliers of the suppliers of industry I. For each industry I, we also consider its customers and suppliers' centrality measures, based on second-order connections:

$$\begin{aligned} Customer^{CM}_{I,t} &= \sum_{J\neq I} w_{IJ,t} \cdot CM_{J,t} \\ Supplier^{CM}_{I,t} &= \sum_{J\neq I} w_{JI,t} \cdot CM_{J,t} \end{aligned} \tag{12}$$

where CM denotes the centrality measures defined above. Therefore, we test the relevance of customers' and suppliers' indegree, outdegree, and betweenness. Given that supplier-based eigencentrality already takes into account the centrality of an industry's suppliers, and similarly for the customer-based one, we will only compute the reciprocal second-order connections, namely using the supplier-based eigencentrality of customers and the customer-based eigencentrality of suppliers. These second-order connections refer to the importance of suppliers in terms of the importance of the customers they serve, and vice-versa for customers. Computing weighted customers' out-degree is equivalent to the second-order connections studied by **?** for the detailed US input-output tables.

Regarding the economic interpretation, we expect to find effects related to industries' concentration index, which was shown by **?** to be empirically linked to lower average stock returns and implicitly lower risk. Moreover, we expect default intensities to be related to customers' indegree, which proxy for customers' payment obligations and can be seen as exposure to customers' credit risk, and suppliers' outdegree, which accounts for competition for input resources and business supply risk. On the other hand, customers' outdegree, which captures their role as main commodity providers to other industries, and suppliers' indegree, which captures their role as commodity buyers from other industries, do not make much economic sense and we would not expect them to provide relevant results. However, we include them in different model specifications in order to see whether arbitrarily constructed variables using customer-supplier connections could lead the model to deliver significant results, in which case we could not rule out spurious results in the other specifications.

5 Empirical Results

This section presents the estimation results for the default intensity parameters ($\{\alpha_1, \alpha_2, ..., \alpha_k\}$ in Equation **??**) related to different credit model specifications, using the maximum-likelihood (ML) procedure.[17] We group the results by the industry-specific attribute used to test the cross-sectional and structural effects related to industries' position in the customer-supplier network. The default horizon used in all estimations is one year.

We will follow the standard approach in the credit risk literature and use the Cumulative Accuracy Profile (CAP) and its summary statistic, the Accuracy Ratio (AR)[18] to evaluate the performance of different credit risk model specifications. We will differentiate between the in-sample AR, which uses the whole dataset both for estimating the

[17]We consider that the coefficients for other exit parameters are not of interest for default prediction and are not shown in this paper for reasons of space. However, they can be made available upon request.

[18]For more information about the CAP and AR, please see Appendix **??**.

model and assessing its prediction performance,[19] and the out-of-sample AR, which is the true performance measure as it splits the data sample in two subsamples, one for estimating the model - called the training sample - and one for evaluating the model - called the testing sample. The in-sample AR cannot be used to determine whether a model is superior to another, as it is susceptible to overfitting, and gives only a first indication about their performance. We also show results of likelihood ratio tests, which are used to compare the goodness of fit of the models with respect to the benchmark specification.

5.1 Estimation of default intensity parameters

The ML estimations for the benchmark default intensities are shown in column (0) of Tables **??**-**??**, which present results using the whole data sample from 1997q1 to 2014q4. As a reminder, the benchmark model includes firm-specific and macroeconomic variables taken from the standard credit risk literature, while it ignores cross-sectional industry effects and customer-supplier connections. We can see that most coefficients, both for macroeconomic and firm-specific variables, have the expected sign and are significant at 1% level. Moreover, the benchmark coefficients are stable across different model specifications: better past stock returns, both at company-level as well as on the aggregate market, reduce companies' default intensities; the same applies for increased company profitability, bigger relative size, and tighter monetary policy. Interestingly, firm's distance-to-default has the expected sign but does not appear to be significant in the benchmark model, even though it becomes significant in other model specifications. This may be because of its high correlation with $SizeComp$ and $NetIncTA$, which are both highly significant. The market-to-book ratio is not significant in any of the specifications.[20]

Table **??** shows the ML estimates for default intensity coefficients when we account for variables based on industries' average distance-to-default, on the left side, and industries' concentration, on the right side. We can see that for our sample, firms' default intensity is highly sensitive to the $\overline{DtD}$ of the industry in which it operates, with a strongly significant but positive coefficient, meaning that a decrease in the aggregate volatility-

[19]For in-sample AR, the model's fitted values are compared to the actual realizations using the same dataset. We report it together with the estimation results in Tables **??**-**??**.

[20]However, the results for other-exit estimations show that both firms' distance-to-default and market-to-book ratio are significant.

adjusted leverage of the industry increases individual companies' default intensity over the next year. Adding this variable alone to the estimations increases the in-sample accuracy, as measured by the AR, from 85.66% in the benchmark model, to 86.40%, even though part of the better performance may be due to firm level DtD becoming significant. This results contradicts the observation by **?** that the industry-average distance-to-default is not significant in the estimation of default intensities. Moreover, combining industries' $\overline{DtD}$ with their customer-supplier connections from the commodities' input-output tables does not seem to bring improvements in terms of in-sample fit. We have significant coefficients for firms' sensitivity to $\overline{DtD}$ of industries that act as customers, both as first and second connections, but without being translated in better in-sample AR.

Further, we can see that the concentration level of the industry where a firm operates has a negative association with companies' default intensity, though it is not statistically significant. This confirms the hypothesis that companies in industries with high competition, and low entry barriers, tend to have higher default intensities. The results of combining the customer-supplier connections and industries' concentration level show a positive and statistically significant association with companies' default intensities for the next year. In the case of customer industries, the in-sample accuracy also increases to 86.18%. This result could be explained by a limited diversification of the client base in the case of high aggregate customer concentration, which gives more exposure to clients' credit risk. Looking at second level connections gives significant coefficients, though they do not help to improve the model in-sample AR.

In the following part, we analyze the results for model specifications that include variables based on network centrality measures. As mentioned before, network centrality measures are built solely on the annual input-output tables, therefore they will have lower variability compared to the state variables of the benchmark model. This constraint makes it more difficult to obtain results where centrality measures prove relevant for corporate default prediction, making any significant results we find even stronger.

Table **??** presents the effects of industries' role as main commodity customers or suppliers to the rest of the economy firms' default intensities. The model specifications use industries' indegree, outdegree, and their combination with customer-supplier connections. Including only industries' indegree does not yield significant coefficients; however, including both the industry-specific indegree together with that of customer industries gives a good performance both in terms of in-sample AR (86.37%) as well as significance of coefficients. A positive indegree coefficient means that the more an industry depends on inputs from the rest of the economy, the higher companies' default intensities; the same effect happens if an industry's customers are among important consumers of inputs. This result is consistent with the hypothesis of exposure to customers' credit risk. On the other hand, exposure to suppliers' indegree does not yield significant coefficients, nor improvements in the in-sample fit.

The coefficient related to industry-specific outdegree is positive and significant at 1%, meaning that the more important an industries' role as commodity supplier, the higher the default intensities of companies operating in that industry. Including only the outdegree helps improve the in-sample fit by 0.3 percentage points. This result is expected, as the outdegree captures the general exposure toward customers, and confirms the previous results related to customers' indegree. The role played by an industry's commodity suppliers as major suppliers towards the rest of the economy (suppliers' oudegree) is also statistically and economically relevant, as it captures companies' exposure to the risk of supply chain disruptions[21]. On the other hand, the role played by a customer as important supplier in the economy (customers' outdegree) does not prove to be useful for predicting corporate default.

The relevance of an industry's exposure to the flow of commodities in the intersectoral network, as captured by the betweenness centrality, is analyzed in the left side of Table **??**. Firstly, we can see that there is a statistically significant and positive association between a company's default intensity over the next year and its industry role as bridge in the supply chain. This observations is also valid for second-order connections: having economic ties to a supplier with high betweenness increases the default intensity and yields the best in-sample fit with respect to the benchmark specification (86.49% vs. 85.66%.) This confirms the hypothesis that an industry's structural position in the general flow of commodities, which exposes it to variations in the availability of commodities as well as payment flows, is having an impact on companies credit and business risk. The right

[21] Operational accidents and natural disasters are a good example of risks that can cause extensive supply chain disruption, as it was the case of Toyota, Sony and other Japanese companies that needed to suspend production in the wake of the 2011 earthquake and tsunami. See **?**

side of table **??** analyzes the performance of the default prediction model when variables based on industries' eigencentrality are introduced. At industry-level, we show results for in-eigencentrality, where the centrality of an industry is high if its suppliers have high centrality. We can see that while the coefficient is positive and statistically significant, the in-sample AR of the model is not greatly improved. The same is valid for the second-order connections, where we consider the supplier-based eigencentrality of customers and the customer-based eigencentrality of suppliers.

Figure **??** shows how the in-sample estimated default intensities compare to the actual number of defaults (depicted as bars) over the sample period. The exponential model seems to fit reasonably well the empirical defaults; moreover, introducing industry-level variables in the benchmark specification does not change the shape of fit, rather leads to local adjustments. For example, including industry-average distance-to-default lowers the estimated credit risk during 1998-2000 (which the benchmark model seems to overstate) and increases it during 2004-2007 and 2013-2014. On the other hand, including supplier betweenness does not change much the benchmark fit, except for a short overstatement of credit risk in mid-2001.

Further, we look at *Exide Technologies*, a company in the *Electrical equipment, appliances, and components* industry that went through two Chapter 11 bankruptcy protection filings, in 2002 and 2013. The company emerged with debt reductions from both cases, in 2004 and 2015. Even though there were a couple of critical financial distress episodes, in our sample the company is registered with one exit event due to default in April 2015. In Figure **??**, which shows the in-sample estimated default intensities for *Exide* over 1997q1-2014q4, we can see that the corporate default model is able to pick spikes in credit risk for several quarters preceding both bankruptcy filings. As before, different specifications including industry effects appear as variants to the benchmark model. For this specific example, customer-based centrality seems to provide the highest estimated default intensities before and during the bankruptcy periods. Other increases in credit risk were also registered during 2005 and 2007-2009.

5.2 Out-of-sample performance

For assessing the ability of a model to predict corporate default, we need to split the data sample into two parts, such that the data used to estimate the model (called the training sample) is different from the data used to evaluate it (called the validation or testing sample). This procedure ensures that the model performance, compared to other alternatives, is not overstated. There are two ways to split the dataset in training and testing samples: using time separation for two consecutive periods or creating a cross-sectional separation for estimation and prediction.

In the case of over-time out-of-sample (OT-OOS) performance, we use an iterative procedure in which for each out-of-sample quarter over 2007q1-2014q4 we estimate the model specification, based on all data available up to that point in time, and use it to determine firms' default intensities for the upcoming year. For each iteration, we collect the estimated default intensities and use them in the end to calculate the out-of-sample AR.

The out-of-sample results using over-time sample separation are shown on the left side of Table **??** and evaluate the accuracy ratio for a selection of model specifications that include industry-specific effects and customer-supplier relations. It appears that the main channel of distress transmission, which has a quantitative impact on firms' default probability, is through supplier industries. Compared to the in-sample fit, the OT-OOS AR is more strict in assessing models' prediction power, with the benchmark model already scoring very high with an AR of 94.54%. The highest AR, of about 95.08%, is obtained for the model specification including suppliers' outdegree,[22] meaning that the importance of an industry's suppliers as main commodity providers to the rest of the economy is increasing companies' default intensities. This observation highlights the possible impact of business risk, and more specifically supplier risk, related especially to episodes of supply-chain disruptions.

Considering the possible impact of other firms' financial health on companies' default probabilities, we can see in that an industry's average distance-to-default helps predict the default of individual companies (OT-OOS AR of 94.91%). This specification was also among the best performing specifications in terms of in-sample fit and contradicts the

[22] See second-order connections in Equation **??**.

claims in ? regarding the ineffective use of industry-average distance-to-default. Other attributes that increase model's prediction power are the concentration level inside supplier industries and suppliers' customer-based eigencentrality, which indicates the importance of suppliers in terms of the importance of the customers they serve.

For the cross-sectional out-of-sample evaluation (CS-OOS), we employ the k-fold cross-validation method, where the original sample is randomly partitioned in k equal sized subsamples, or block. Then, for each of the k subsample, we retain one subsample for testing and use the remaining $k-1$ subsamples to train the model. This method insures that each observation is used for validation exactly once and that all observations are used both for training and testing. In our case, we use $k = 100$. The CS-OOS AR is computed after all k subsamples have been used for validation.

The right side of Table ?? shows the prediction performance of different credit model specifications using cross-validation. The AR in this case is significantly different from the OT-OOS AR, as the latter uses only the years 2007-2014 as out-of-sample data, whereas for cross-validation all observations are considered for out-of-sample. From this point of view, CS-OOS should be compared to in-sample AR, allowing to account for models' over-fitting. We can see that, with few exceptions, models including industry-specific and customer-supplier effects tend to out-perform the benchmark model, which has an AR of 84.50%. The specification including suppliers' betweenness proves to be the best performing one, with an AR of 85.51%. This result reinforces the importance of exposures to variations in the availability of commodities and payment flows through suppliers. Customer concentration is also among the best performers, confirming the in-sample observation that a lack of customer base diversification negatively affects companies' credit risk. Overall, the ranking of model prediction performance obtained with cross-validation matches the initial results based on in-sample AR.

6 Conclusion

This paper tests the relevance of attributes based on the structure of intersectoral customer-supplier networks for the prediction of corporate default. Using industry-specific variables and customer-supplier relations allows us to incorporate direct impacts from neighboring industries on the default intensities of individual companies. We employ a reduced-form corporate default model used in **?** and **?**, who rely on a set of macroeconomic and firm-specific variables to estimate the intensities of default and other types of exits. The model is based on two independent doubly-stochastic Poisson processes that enable a separable likelihood function for default and other-type of exits. One of the limitations of the original model is the conditional independence of companies' default hazard rates, meaning that the model does not allow for the default or other exit of a company to have a direct influence on the default or other exit intensity of another firm. The only type of default intensity dependence between companies stems from their common dependence on macrovariables or from the correlation across firms of firm-specific covariates. In our study, we specifically test whether including the customer/supplier structural role of industries, and various combinations with industry-specific attributes such as average distance-to-default and concentration level, influence firms' default intensity.

The use of commodity-to-industry input-output tables is motivated by the observation in **?** that the law of large numbers fails for asymmetric networks, which have power-law tail structures for nodes outdegrees, meaning that the impact of idiosyncratic shocks on the aggregate volatility is not vanishing as fast as previously thought. Moreover, **?** also studies the microfoundations of aggregate fluctuations and shows that the "diversification argument" is not valid in the presence of a fat-tailed distribution of firm sizes, because shocks to firms with highest market power will propagate at aggregate level. Our analysis shows that both the distribution of industries' concentration indices, which is based on firms' size, and the distribution of network measures based on out-going links, that relate to industries' role as commodity suppliers, are heavy tailed and asymmetric. These initial observations point to the possibility of aggregate fluctuations forming from industry-specific shocks, which could impact corporate default prediction.

The empirical method used in this study is based on the maximum likelihood estimation of the default intensity function, where we set the default horizon at one year. We find empirically that industry-level attributes such as distance-to-default and competition are relevant for the prediction of corporate default. Moreover, the results provide evidence that economic ties to customer/supplier industries that have low competition relate positively to companies' default intensities. The best in-sample fit, as measured by the accuracy ratio, is obtained when accounting for industry-average distance-to-default and industries' exposure to the flow of commodities in the economy, as captured by betweenness. We also observe that arbitrarily built network-based variables do not yield significant results, which helps confirm that the results obtained for economically meaningful variables are not artefacts.

The over-time out-of-sample prediction performance points to the superiority of the supplier-based model specifications but is generally more pessimistic than the in-sample accuracy when assessing improvements over the benchmark. Overall, the financial health context of an industry, as measured by industry-average distance-to-default, helps predict individual corporate defaults in out-of-sample tests. This result contradicts previous observations in the literature. Concerning the customer-supplier relations, cross-sectional out-of-sample results confirm the hypotheses that exposures to customers' credit risk, assessed by customers' indegree, and supply chain disruptions, measured by suppliers' outdegree, are relevant predictors of companies' credit risk. Moreover, insufficient diversification of the customer and supplier base as well as exposures to variations in the flow of commodities and payments in the aggregate economy also have an impact on companies' default intensities.

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80:1977–2016.

Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2010). Cascades in networks and aggregate volatility. Technical report, National Bureau of Economic Research.

Allen, F. and Gale, D. (2000). Financial contagion. *Journal of Political Economy*, 108(1):1–33.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4):589–609.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4:71–111.

Bharath, S. T. and Shumway, T. (2008). Forecasting default with the Merton distance to default model. *Review of Financial Studies*, 21(3):1339–1369.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–54.

Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2):163–177.

Campbell, J. Y., Hilscher, J., and Szilagyi, J. (2008). In search of distress risk. *The Journal of Finance*, 63(6):2899–2939.

Chava, S. and Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review of Finance*, 8(4):537–569.

Clauset, A., Shalizi, C., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*.

Constantin, A., Peltonen, T. A., and Sarlin, P. (2017). Network linkages to predict bank distress. *Journal of Financial Stability*.

Cont, R., Moussa, A., and Santos, E. B. (2012). Network structure and systemic risk in banking systems. *Handbook of Systemic Risk, Cambridge University Press.*

Craig, B. and Von Peter, G. (2014). Interbank tiering and money center banks. *Journal of Financial Intermediation*, 23(3):322–347.

Crosbie, P. and Bohn, J. (2003). Modeling default risk. *Research Report, Moody's—KMV Corp.*

Duan, J.-C., Sun, J., and Wang, T. (2012). Multiperiod corporate default prediction: A forward intensity approach. *Journal of Econometrics*, 170:191–209.

Duffie, D., Saita, L., and Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3):635–665.

Eisenberg, L. and Noe, T. H. (2001). Systemic risk in financial systems. *Management Science*, 47(2):236–249.

Engelmann, B., Hayden, E., Tasche, D., et al. (2003). Measuring the discriminative power of rating systems. *Deutsche Bundesbank Frankfurt.*

Eom, Y. H., Helwege, J., and Huang, J.-z. (2004). Structural models of corporate bond pricing: An empirical analysis. *Review of Financial Studies*, 17(2):499–544.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica*, 79:733–772.

Hillegeist, S. A., Keating, E. K., Cram, D. P., and Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9(1):5–34.

Hou, K. and Robinson, D. T. (2006). Industry concentration and average stock returns. *The Journal of Finance*, 61(4):1927–1956.

Kyriakopoulos, F., Thurner, S., Puhr, C., and Schmitz, S. W. (2009). Network and eigenvalue analysis of financial transaction networks. *The European Physical Journal B*, 71(4):523–531.

Leland, H. E. (1994). Corporate debt value, bond covenants, and optimal capital structure. *The Journal of Finance*, 49(4):1213–1252.

Markose, S., Giansante, S., and Shaghaghi, A. R. (2012). 'Too interconnected to fail' financial network of US CDS market: Topological fragility and systemic risk. *Journal of Economic Behavior & Organization*, 83(3):627–646.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29(2):449–470.

Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5):323–351.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, pages 109–131.

Park, Y., Hong, P., and Roh, J. J. (2013). Supply chain lessons from the catastrophic natural disaster in Japan. *Business Horizons*, 56(1):75–85.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*, 74(1):101–124.

Sobehart, J. R., Keenan, S. C., and Stein, R. (2000). Benchmarking quantitative default risk models: a validation methodology. *Moody's Investors Service.*

Soramäki, K., Bech, M. L., Arnold, J., Glass, R. J., and Beyeler, W. E. (2007). The topology of interbank payment flows. *Physica A: Statistical Mechanics and its Applications*, 379(1):317–333.

Tarashev, N. A. (2005). An empirical evaluation of structural credit risk models. *BIS Working Paper.*

Vassalou, M. and Xing, Y. (2004). Default risk in equity returns. *The Journal of Finance*, 59(2):831–868.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, pages 59–82.

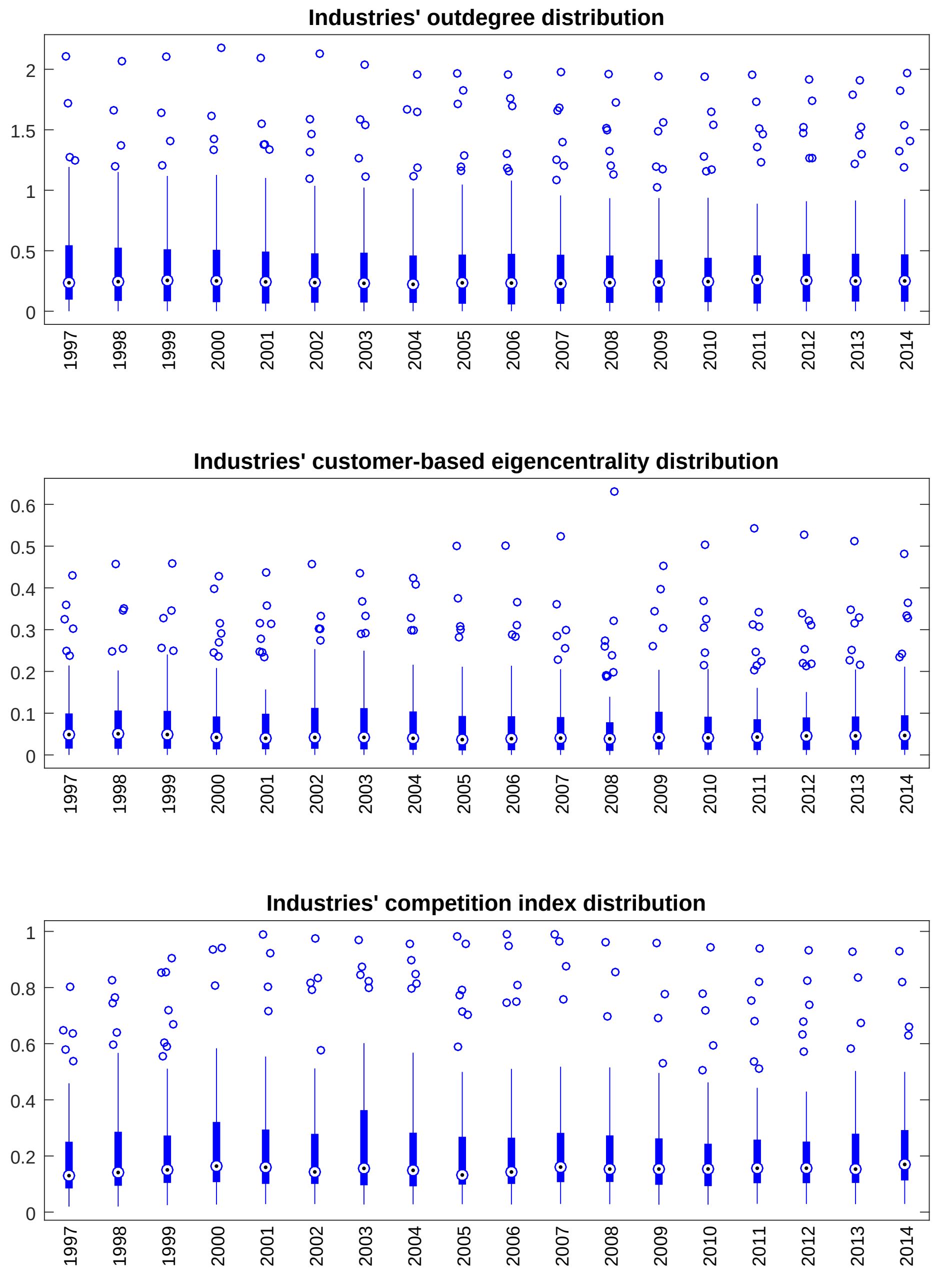

Figure 1: The figure shows the evolution over the period 1997 to 2014 of industries' concentration and centrality distributions. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the 'o' symbol.

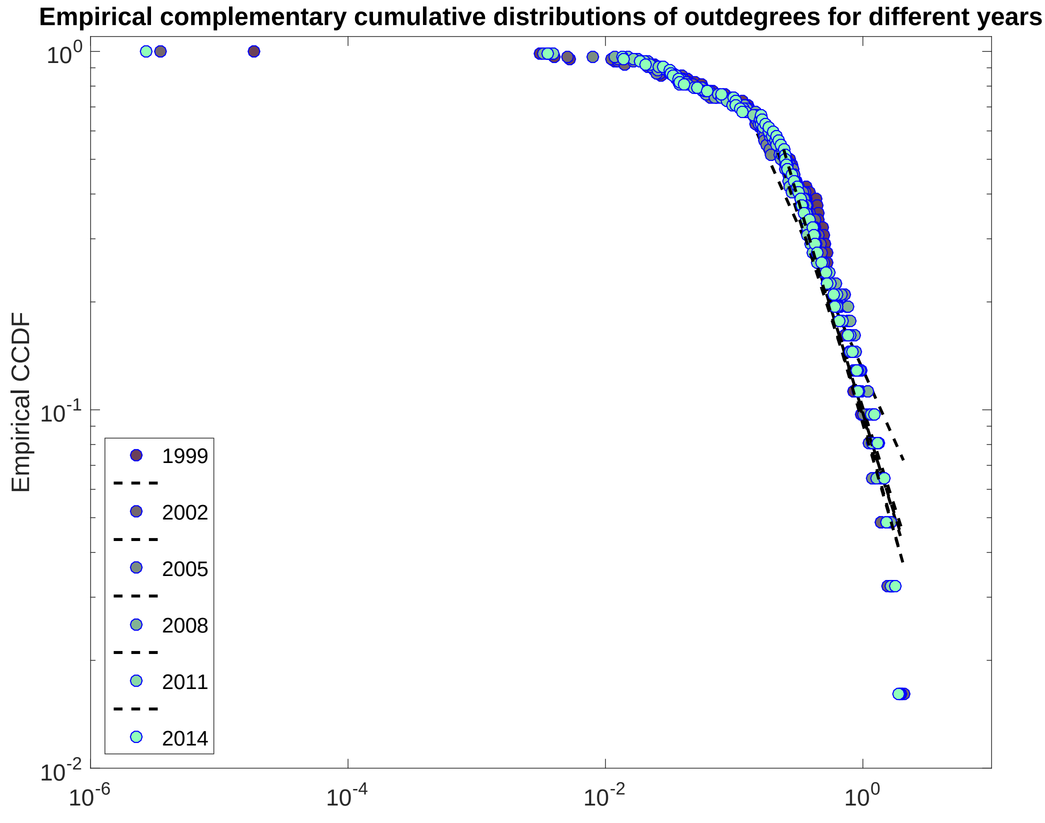

Figure 2: The figure shows the empirical complementary cumulative distributions, also called tail distributions, of outdegrees (marked with colored discs) and their respective fitted power-law model (marked with dashed lines) for a selection of years spanning 1999-2014.

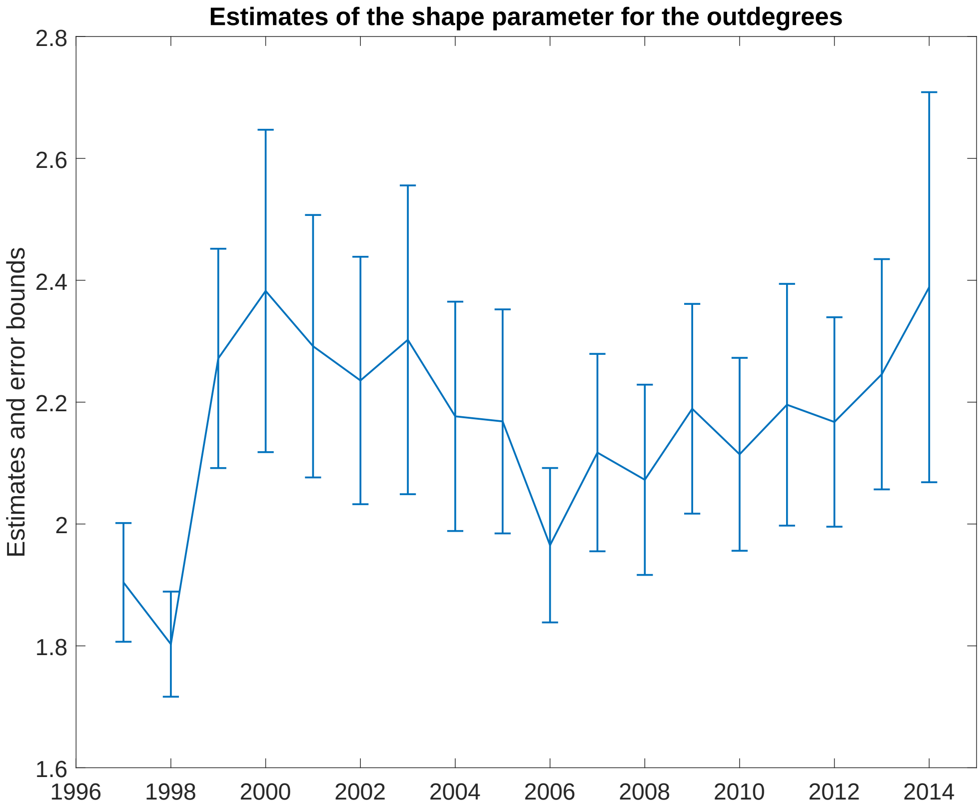

Figure 3: The figure shows the estimates and their respective standard errors of the power-law shape parameter for industries' outdegrees, over the period 1997 to 2014.

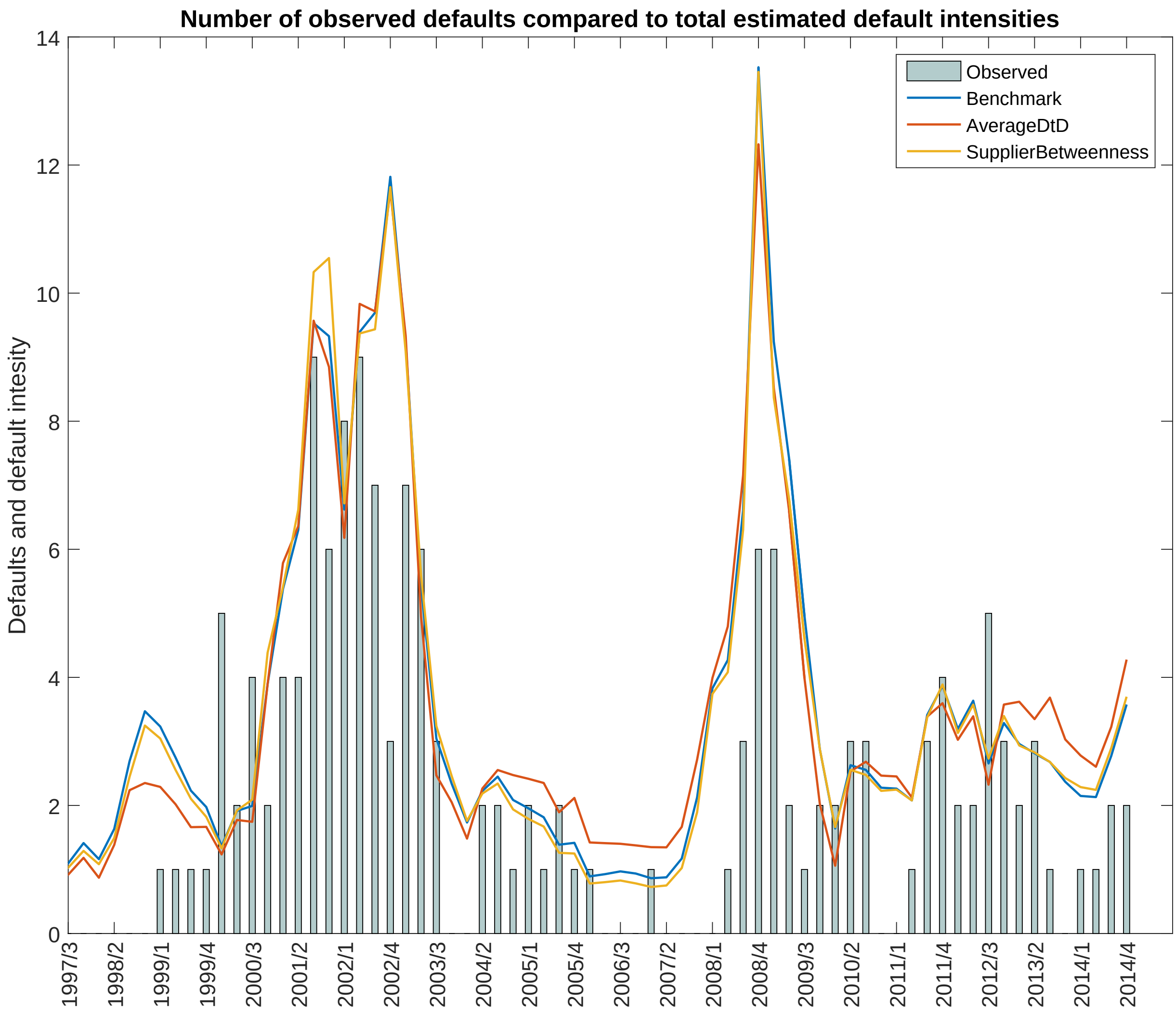

Figure 4: The figure shows for each quarter the observed (bars) and in-sample predicted aggregate number of defaults for different corporate default model specifications, over the period 1997 to 2014.

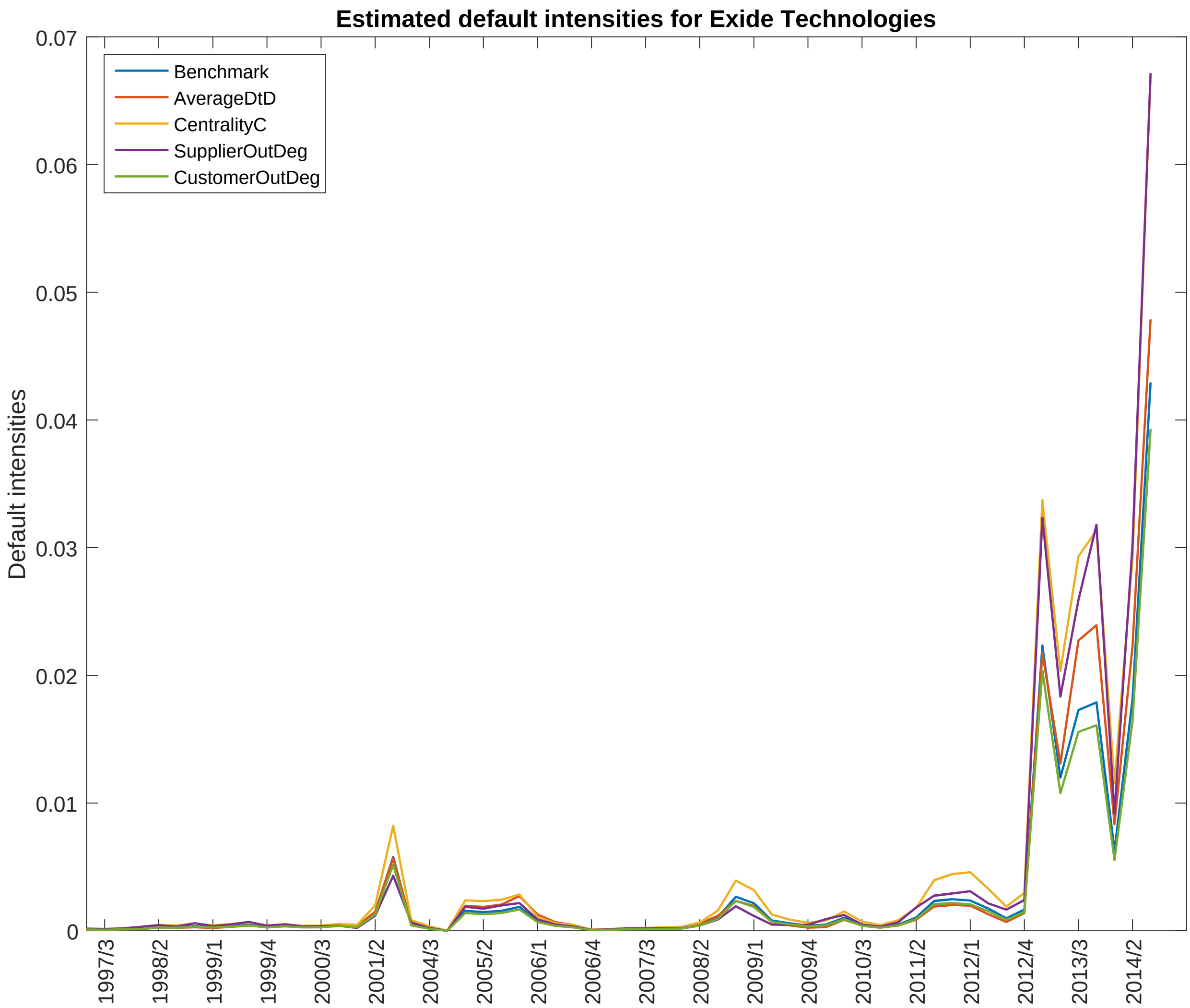

Figure 5: The figure shows the in-sample estimated default intensities for Exide Technologies, a multi-billion dollar company producing electrical machinery, equipment, and supplies that exited the Compustat sample in April 2015 due to Chapter 11 bankruptcy filing.

Table I: Number of companies, defaults/bankruptcies, and other exits for 1997-2015

Year	Active firms	Defaults	Other exits
1997	6034	0	190
1998	5904	10	331
1999	5595	30	432
2000	5388	56	487
2001	4914	42	365
2002	4499	27	248
2003	4156	16	314
2004	4022	16	442
2005	3969	16	254
2006	3884	11	290
2007	3799	5	321
2008	3598	7	210
2009	3356	17	189
2010	3259	15	283
2011	3127	20	209
2012	3066	15	203
2013	3095	16	186
2014	2965	8	186
2015	2828	10	185

The table reports the number of active firms, defaults/bankruptcies, and other exits for each year over the period 1997-2015.

Table II: Descriptive statistics for macroeconomic and firm-specific attributes

	Mean	Std. dev.	Min	.25pctl	Median	.75pctl	Max	Skewness	Kurtosis
				Macroeconomic variables					
$Trailing\ S\&P500$	0.083	0.189	-0.397	-0.011	0.114	0.211	0.466	-0.489	2.839
$3m\ T-bill$	2.570	2.126	0.010	0.150	2.190	4.790	6.000	0.098	1.365
				Firm-specific variables					
$Trailing\ return$	0.224	1.561	-0.972	-0.306	0.000	0.333	28.521	10.841	164.882
$Cash/TA$	0.202	0.236	0.000	0.025	0.101	0.302	0.990	1.438	4.254
$NetInc/TA$	-0.017	0.124	-1.693	-0.016	0.007	0.020	0.988	-4.964	67.170
$Size$	-4.314	2.140	-9.903	-5.881	-4.345	-2.859	2.227	0.166	2.720
$Market/Book$	2.085	3.093	0.262	0.908	1.337	2.211	56.899	10.017	148.765
DtD	4.024	3.197	-2.174	1.811	3.533	5.566	26.011	1.540	8.114

The table reports annualized descriptive statistics for the attributes used in the benchmark credit model. For macroeconomic variables, $Trailing\ S\&P500$ is the return of the S&P500 over the previous year and $3m\ T-bill$ is the three-month Treasury rate. For firm-specific variables, $Trailing\ return$ is firm's stock return over the previous year, DtD is the distance-to-default, $Cash/TA$ is cash and short-term investments over the total assets, $NetInc/TA$ is the net income over the total assets, $Size$ is log of firm's market equity value over the average market equity value of an S&P500 company, and $Market/Book$ is the market to book ratio using the total asset values.

Table III: Correlation matrix for firms-specific variables

	$Trailing\ return$	$Cash/TA$	$NetInc/TA$	$Size$	$Market/Book$	DtD
$Trailing\ return$	1.000	0.059	0.028	0.026	0.120	0.123
$Cash/TA$	0.059	1.000	-0.171	-0.107	0.315	0.068
$NetInc/TA$	0.028	-0.171	1.000	0.240	-0.037	0.275
$Size$	0.026	-0.107	0.240	1.000	0.082	0.465
$Market/Book$	0.120	0.315	-0.037	0.082	1.000	0.273
DtD	0.123	0.068	0.275	0.465	0.273	1.000

Table IV: Number of companies by industry: Part 1 of 2.

	Firms	Defaults	(%)	Other exits	(%)
Farms	27	1	3.70	16	59.26
Forestry, fishing, and related activities	6	0	0.00	4	66.67
Oil and gas extraction	302	9	2.98	168	55.63
Mining, except oil and gas	100	4	4.00	44	44.00
Support activities for mining	58	1	1.72	29	50.00
Utilities	209	0	0.00	112	53.59
Construction	80	4	5.00	43	53.75
Wood products	35	3	8.57	17	48.57
Nonmetallic mineral products	49	3	6.12	31	63.27
Primary metals	96	4	4.17	60	62.50
Fabricated metal products	113	3	2.65	57	50.44
Machinery	329	11	3.34	204	62.01
Computer and electronic products	1150	41	3.57	739	64.26
Electrical equipment, appliances, and components	135	2	1.48	78	57.78
Motor vehicles, bodies and trailers, and parts	110	6	5.45	55	50.00
Other transportation equipment	71	3	4.23	33	46.48
Furniture and related products	44	4	9.09	21	47.73
Miscellaneous manufacturing	297	12	4.04	201	67.68
Food and beverage and tobacco products	195	4	2.05	117	60.00
Textile mills and textile product mills	39	2	5.13	28	71.79
Apparel and leather and allied products	123	7	5.69	74	60.16
Paper products	75	2	2.67	52	69.33
Printing and related support activities	43	1	2.33	32	74.42
Petroleum and coal products	55	2	3.64	27	49.09
Chemical products	912	30	3.29	465	50.99
Plastics and rubber products	85	4	4.71	57	67.06
Wholesale trade	367	18	4.90	229	62.40
Motor vehicle and parts dealers	32	0	0.00	17	53.13
Food and beverage stores	44	3	6.82	33	75.00
General merchandise stores	44	7	15.91	20	45.45
Other retail	353	27	7.65	205	58.07
Air transportation	42	7	16.67	17	40.48

The table reports the number of companies that were active for each industry during 1997-2015 as well as the number of defaults/bankruptcies and other exits.

Table IV (Continued): Number of companies by industry: Part 2 of 2.

	Firms	Defaults	(%)	Other exits	(%)
Rail transportation	16	0	0.00	9	56.25
Water transportation	24	0	0.00	11	45.83
Truck transportation	51	5	9.80	28	54.90
Transit and ground passenger transportation	5	0	0.00	4	80.00
Pipeline transportation	55	0	0.00	25	45.45
Other transportation and support activities	40	2	5.00	23	57.50
Warehousing and storage	4	0	0.00	4	100.00
Publishing industries, except internet (includes software)	579	19	3.28	451	77.89
Motion picture and sound recording industries	53	3	5.66	38	71.70
Broadcasting and telecommunications	355	13	3.66	242	68.17
Data processing, internet publishing, and other	301	9	2.99	159	52.82
Federal Reserve banks, credit intermediation, and other	49	3	6.12	27	55.10
Securities, commodity contracts, and investments	90	5	5.56	39	43.33
Insurance carriers and related activities	83	2	2.41	61	73.49
Funds, trusts, and other financial vehicles	16	0	0.00	14	87.50
Other real estate	62	1	1.61	27	43.55
Rental and leasing services and lessors of intangible assets	105	7	6.67	47	44.76
Legal services	5	0	0.00	4	80.00
Computer systems design and related services	277	7	2.53	199	71.84
Miscellaneous professional, scientific, and technical services	232	8	3.45	151	65.09
Administrative and support services	186	6	3.23	124	66.67
Waste management and remediation services	61	4	6.56	41	67.21
Educational services	42	1	2.38	23	54.76
Ambulatory health care services	160	12	7.50	100	62.50
Hospitals	28	0	0.00	16	57.14
Nursing and residential care facilities	47	2	4.26	36	76.60
Social assistance	8	0	0.00	7	87.50
Performing arts, spectator sports, museums, and other	19	1	5.26	13	68.42
Amusements, gambling, and recreation industries	50	0	0.00	29	58.00
Accommodation	63	1	1.59	48	76.19
Food services and drinking places	164	5	3.05	102	62.20
Other services, except government	36	0	0.00	28	77.78
Total	8886	341	3.84	5415	60.94

The table reports the number of companies that were active for each industry during 1997-2015 as well as the number of defaults/bankruptcies and other exits.

Table V: Descriptive statistics for industries' competition and centrality measures

	Competition	Indegree	Outdegree	EigenCent^C	EigenCent^S	Betweenness
Mean	0.230	0.374	0.373	0.079	0.117	9.407
Std. dev.	0.206	0.124	0.425	0.097	0.041	11.924
Min	0.020	0.099	0.000	0.000	0.032	0.000
.25pctl	0.100	0.278	0.057	0.013	0.084	3.146
Median	0.152	0.364	0.243	0.045	0.116	6.416
.75pctl	0.277	0.461	0.494	0.096	0.149	8.531
Max	0.990	0.791	2.177	0.631	0.257	70.540
Skewness	1.819	0.430	1.849	2.180	0.284	2.944
Kurtosis	5.936	3.116	6.501	8.286	2.502	13.304

The table reports descriptive statistics for industries' competition and centrality measures. Competition is computed as the sum of squared market shares of companies in each industry (the HHI index), indegree is the share of commodity inputs used from the rest of the economy, outdegree is the share of commodity outputs delivered by an industry towards the rest of the economy, eigencentrality is the relative importance of industries in the sectoral network, and betweenness is industries' exposure to the flow of supply-chain commodities in the economy. See Section detailed in Section **??** for more information.

Table VI: Default intensity coefficients when considering industry-specific variables and customer-supplier effects.

		Average DtD							Concentration						
	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	-9.612***	-10.437***	-9.824***	-9.764***	-9.886***	-9.793***	-10.455***	-10.309***	-9.554***	-10.008***	-9.892***	-10.077***	-10.232***	-9.837***	-9.658***
	(0.492)	(0.554)	(0.506)	(0.503)	(0.548)	(0.537)	(0.559)	(0.571)	(0.502)	(0.507)	(0.507)	(0.513)	(0.527)	(0.512)	(0.524)
TrailingRet	-0.571***	-0.579***	-0.581***	-0.597***	-0.577***	-0.586***	-0.586***	-0.604***	-0.584***	-0.564***	-0.596***	-0.588***	-0.626***	-0.605***	-0.582***
	(0.183)	(0.186)	(0.187)	(0.188)	(0.185)	(0.187)	(0.189)	(0.191)	(0.187)	(0.185)	(0.189)	(0.186)	(0.192)	(0.189)	(0.186)
SP500	-0.768**	-2.053***	-1.202***	-1.029**	-1.171**	-0.956**	-2.161***	-1.800***	-0.754**	-0.726**	-0.766**	-0.595	-0.562	-0.626	-0.835**
	(0.420)	(0.593)	(0.461)	(0.454)	(0.545)	(0.511)	(0.608)	(0.617)	(0.422)	(0.421)	(0.420)	(0.422)	(0.422)	(0.419)	(0.426)
T3m	-0.291***	-0.223***	-0.268***	-0.275***	-0.278***	-0.282***	-0.220***	-0.225***	-0.293***	-0.268***	-0.274***	-0.295***	-0.289***	-0.287***	-0.289***
	(0.047)	(0.051)	(0.048)	(0.048)	(0.049)	(0.048)	(0.052)	(0.051)	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)	(0.048)	(0.048)
CashTA	0.750***	0.736**	0.668**	0.663**	0.746***	0.763***	0.740**	0.790***	0.746***	0.580**	0.630**	0.864***	0.892***	0.750***	0.797***
	(0.288)	(0.286)	(0.290)	(0.292)	(0.288)	(0.289)	(0.294)	(0.286)	(0.289)	(0.289)	(0.290)	(0.290)	(0.287)	(0.291)	(0.294)
NetIncTA	-1.675***	-1.645***	-1.634***	-1.651***	-1.663***	-1.672***	-1.652***	-1.693***	-1.674***	-1.568***	-1.580***	-1.708***	-1.400***	-1.773***	-1.812***
	(0.470)	(0.471)	(0.473)	(0.472)	(0.472)	(0.473)	(0.475)	(0.469)	(0.472)	(0.476)	(0.473)	(0.471)	(0.482)	(0.471)	(0.467)
SizeComp	-0.685***	-0.670***	-0.680***	-0.683***	-0.681***	-0.685***	-0.661***	-0.667***	-0.685***	-0.682***	-0.689***	-0.684***	-0.688***	-0.662***	-0.698***
	(0.061)	(0.062)	(0.062)	(0.061)	(0.061)	(0.062)	(0.062)	(0.062)	(0.062)	(0.062)	(0.062)	(0.061)	(0.061)	(0.061)	(0.061)
Mkt2Book	0.033	0.029	0.025	0.028	0.032	0.031	0.028	0.030	0.031	0.021	0.027	0.030	0.047	0.018	0.027
	(0.034)	(0.034)	(0.035)	(0.034)	(0.034)	(0.034)	(0.035)	(0.034)	(0.034)	(0.035)	(0.034)	(0.034)	(0.033)	(0.034)	(0.034)
DtD	-0.075	-0.114**	-0.079	-0.073	-0.082	-0.080	-0.117**	-0.120**	-0.077	-0.075	-0.070	-0.084	-0.089**	-0.085	-0.062
	(0.054)	(0.056)	(0.055)	(0.054)	(0.055)	(0.055)	(0.057)	(0.057)	(0.054)	(0.054)	(0.054)	(0.053)	(0.054)	(0.054)	(0.052)
ISV		0.235***					0.194***	0.253***	-0.456					-0.088	-1.774**
		(0.071)					(0.074)	(0.080)	(0.743)					(0.762)	(0.865)
CustomerISV			0.085***				0.249**			2.562***				7.372***	
			(0.033)				(0.107)			(0.518)				(1.204)	
Customer2ndISV				0.095**			-0.304**				2.574***			-8.723***	
				(0.051)			(0.172)				(0.840)			(2.180)	
SupplierISV					0.165			-0.286				6.561***			15.228***
					(0.136)			(0.491)				(2.005)			(4.705)
Supplier2ndISV						0.177		0.324					14.578***		-24.413**
						(0.223)		(0.763)					(4.211)		(9.751)
Likelihood ratio (χ^2)		10.622	6.277	2.953	1.512	0.767	17.337	10.042	0.432	22.804	8.963	6.543	-4.902	38.014	19.609
(p-value)		(0.001)	(0.007)	(0.053)	(0.152)	(0.310)	(0.000)	(0.008)	(0.489)	(0.000)	(0.002)	(0.006)		(0.000)	(0.000)
In-sample AR (%)	85.66	86.40	85.64	85.64	85.56	85.66	86.24	86.41	85.66	86.18	85.80	85.68	85.41	86.04	85.73

The table shows results for maximum likelihood estimates of the default intensity function. Column (0) covers the benchmark model, where only macroeconomic and firm-specific variables are used. The following columns show results for specifications with industry-specific variables (ISV) based on industries' average distance-to-default and concentration. Customer-supplier effects, as detailed in section **??**, are also included. The likelihood ratio test assesses the models' goodness of fit with respect to the benchmark. In-sample fit of the model is evaluated in the last line, using the accuracy ratio (AR). In parentheses are the standard errors for the estimated coefficients. Statistical significance at the 10%, 5% and 1% level is indicated by *, **, and ***.

Table VII: Default intensity coefficients when considering industry indegree/outdegree and customer-supplier effects.

		InDegree				OutDegree			
	(0)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Intercept	-9.612***	-9.794***	-9.910***	-9.659***	-10.970***	-9.843***	-9.816***	-10.124***	-10.185***
	(0.492)	(0.551)	(0.505)	(0.544)	(0.608)	(0.502)	(0.506)	(0.569)	(0.573)
TrailingRet	-0.571***	-0.585***	-0.579***	-0.568***	-0.660***	-0.607***	-0.581***	-0.552***	-0.562***
	(0.183)	(0.187)	(0.185)	(0.183)	(0.198)	(0.190)	(0.186)	(0.181)	(0.183)
SP500	-0.768**	-0.782**	-0.781**	-0.751**	-0.662	-0.782**	-0.763**	-0.763**	-0.760**
	(0.420)	(0.423)	(0.419)	(0.419)	(0.417)	(0.420)	(0.420)	(0.421)	(0.419)
T3m	-0.291***	-0.288***	-0.285***	-0.293***	-0.309***	-0.287***	-0.289***	-0.296***	-0.299***
	(0.047)	(0.048)	(0.047)	(0.048)	(0.047)	(0.047)	(0.047)	(0.047)	(0.048)
CashTA	0.750***	0.819***	0.614**	0.747***	0.755***	0.680**	0.643**	0.602**	0.663**
	(0.288)	(0.293)	(0.289)	(0.287)	(0.292)	(0.288)	(0.292)	(0.298)	(0.302)
NetIncTA	-1.675***	-1.801***	-1.612***	-1.722***	-1.748	-1.649***	-1.602***	-1.570***	-1.621***
	(0.470)	(0.469)	(0.471)	(0.467)	(0.466)	(0.472)	(0.474)	(0.477)	(0.474)
SizeComp	-0.685***	-0.662***	-0.688***	-0.681***	-0.684***	-0.683***	-0.692***	-0.689***	-0.680***
	(0.061)	(0.061)	(0.061)	(0.061)	(0.061)	(0.061)	(0.062)	(0.062)	(0.062)
Mkt2Book	0.033	0.029	0.024	0.031	0.020	0.025	0.028	0.033	0.031
	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)	(0.035)	(0.034)	(0.034)	(0.034)
DtD	-0.075	-0.076	-0.072	-0.078	-0.068	-0.068	-0.071	-0.078	-0.083
	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)
∼Degree		0.904			2.222**	0.391***			0.770**
		(0.773)			(1.260)	(0.138)			(0.376)
Customer∼Degree			1.040***		1.456***		0.583**		-0.976
			(0.322)		(0.378)		(0.267)		(0.787)
Supplier∼Degree				0.612	1.124			1.424**	1.278
				(1.552)	(2.298)			(0.714)	(0.811)
Likelihood ratio (χ^2)		1.357	10.117	0.147	15.352	8.048	4.808	4.171	11.366
(p-value)		(0.174)	(0.001)	(0.968)	(0.001)	(0.003)	(0.016)	(0.024)	(0.005)
In-sample AR (%)	85.66	85.72	85.96	85.61	86.37	85.97	85.71	85.78	86.13

The table shows results for maximum likelihood estimates of the default intensity function ($\{\alpha_1, \alpha_2, ..., \alpha_k\}$ in Equation **??**). Column (0) covers the benchmark model, where only macroeconomic and firm-specific variables are used to assess companies' default probabilities. The following columns show results for specifications including variables based on industries' indegrees/outdegrees ($InDeg$ in Equation **??**, $OutDeg$ in Equation **??**), which capture industries' role as commodity customers/providers in the economy. Customer-supplier effects, as detailed in section **??**, are also included. The likelihood ratio test assesses the models' goodness of fit with respect to the benchmark. In-sample fit is evaluated in the last line, using the accuracy ratio (AR). In parentheses are the standard errors for the estimated coefficients. Statistical significance at the 10%, 5% and 1% level is indicated by *, **, and ***.

Table VIII: Default intensity coefficients when considering industry centrality and customer-supplier effects.

		Betweenness				Eigencentrality			
	(0)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Intercept	-9.612***	-9.825***	-9.787***	-10.789***	-10.774***	-9.938***	-9.847***	-10.688***	-10.524***
	(0.492)	(0.497)	(0.499)	(0.519)	(0.516)	(0.507)	(0.503)	(0.569)	(0.622)
TrailingRet	-0.571***	-0.592***	-0.576***	-0.585***	-0.579***	-0.550***	-0.577***	-0.508***	-0.593***
	(0.183)	(0.187)	(0.186)	(0.186)	(0.185)	(0.181)	(0.185)	(0.174)	(0.191)
SP500	-0.768**	-0.769**	-0.824**	-0.637	-0.681	-0.797**	-0.761**	-0.777**	-0.824**
	(0.420)	(0.420)	(0.421)	(0.422)	(0.420)	(0.421)	(0.419)	(0.419)	(0.426)
T3m	-0.291***	-0.290***	-0.282***	-0.332***	-0.314***	-0.279***	-0.286***	-0.317***	-0.294***
	(0.047)	(0.047)	(0.047)	(0.048)	(0.048)	(0.047)	(0.047)	(0.048)	(0.048)
CashTA	0.750***	0.852***	0.723**	0.734**	0.767***	0.751***	0.631**	0.578**	0.753**
	(0.288)	(0.291)	(0.288)	(0.289)	(0.290)	(0.288)	(0.289)	(0.287)	(0.312)
NetIncTA	-1.675***	-1.770***	-1.633***	-1.669***	-1.669***	-1.569***	-1.686***	-1.412***	-1.604***
	(0.470)	(0.466)	(0.470)	(0.463)	(0.463)	(0.475)	(0.469)	(0.480)	(0.480)
SizeComp	-0.685***	-0.681***	-0.678***	-0.682***	-0.677***	-0.695***	-0.683***	-0.686***	-0.685***
	(0.061)	(0.061)	(0.061)	(0.060)	(0.060)	(0.062)	(0.061)	(0.061)	(0.062)
Mkt2Book	0.033	0.033	0.029	0.036	0.033	0.025	0.023	0.034	0.028
	(0.034)	(0.033)	(0.034)	(0.033)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)
DtD	-0.075	-0.074	-0.077	-0.080	-0.080	-0.076	-0.072	-0.091**	-0.072
	(0.054)	(0.054)	(0.054)	(0.052)	(0.053)	(0.054)	(0.054)	(0.055)	(0.055)
Centrality		0.016***			0.009	2.025***			2.385**
		(0.004)			(0.006)	(0.615)			(1.064)
CustomerCentrality			0.039***		0.011		2.962***		-0.380
			(0.013)		(0.014)		(1.058)		(1.866)
SupplierCentrality				0.237***	0.206***			15.682***	4.818
				(0.039)	(0.042)			(3.311)	(4.559)
Likelihood ratio (χ^2)		10.609	8.190	30.677	34.545	11.001	7.216	2.116	14.791
(p-value)		(0.001)	(0.002)	(0.000)	(0.000)	(0.000)	(0.004)	(0.095)	(0.002)
In-sample AR (%)	85.66	85.69	85.78	86.49	86.45	85.81	85.89	85.31	85.94

The table shows results for maximum likelihood estimates of the default intensity function ($\{\alpha_1, \alpha_2, ..., \alpha_k\}$ in Equation **??**). Column (0) covers the benchmark model, where only macroeconomic and firm-specific variables are used to assess companies' default probabilities. The following columns show results for specifications with variables based on industries' centrality in the input-output economy, measured by betweenness and (supplier-based) eigencentrality. We also include customer-supplier effects, as detailed in section **??**. The likelihood ratio test assesses the models' goodness of fit with respect to the benchmark. In-sample fit of the model is evaluated in the last line, using the accuracy ratio (AR). In parentheses are the standard errors for the estimated coefficients. Statistical significance at the 10%, 5% and 1% level is indicated by *, **, and ***.

Table IX: Accuracy ratios (%) for over-time and cross-sectional out-of-sample tests.

	Over-time separation			Cross validation		
	ISV	Supplier	Customer	ISV	Supplier	Customer
$\overline{DtD}$	**94.91**	**94.61**	93.88	**84.69**	84.27	**84.54**
Concentration	94.46	**94.75**	93.34	84.43	**84.62**	**84.94**
InDeg	**94.64**	93.97	93.84	**84.53**	84.26	**84.85**
OutDeg	94.05	**95.08**	94.05	**84.84**	**84.80**	**84.78**
Betweenness	94.18	94.43	94.07	**84.81**	**85.52**	**84.76**
EigenCent^S	93.83	**94.60**	93.61	**84.65**	84.23	**84.84**

The table shows results for the prediction performance, as measured by out-of-sample accuracy ratios, for a selection of relevant model specifications that include industry-specific effects and customer-supplier relations. The bolded values represent an improvement compared to the benchmark specification, which has an AR of 94.54% for over-time out-of-sample and 84.50% for cross-validation. *ISV* represent specifications where the attributes used were either based on an industry's internal context (average distance-to-default and concentration) or its position in the customer-supplier network. The columns *Supplier* and *Customer* refer to attributes where we account for customers and suppliers' internal context and industry-position in the commodities' supply-chain. See Section **??** for more details.

Appendices

A Conditional default probabilities and forward intensity rates

In this section we describe the probabilistic model for firms' default starting from **?**, who a use doubly stochastic Poisson intensity model to estimate default probabilities over multiple periods. In a doubly stochastic setting, the time of default (τ_{iD}) of a firm i is modeled as the first jump of a Poisson process, whose intensity λ_{it} is itself random. The stochastic intensity is a function of some state variables whose dynamics is not affected by default. Besides default, public companies are delisted from trading on the stock exchange as a result of mergers and acquisitions, going back to private company, and other reasons. When studying default, it is important to differentiate whether companies leave the sample as a result of bankruptcy or because of other reasons. This is done by introducing a second independent Poisson process governing other types of exit, whose intensity ϕ_{it} is also stochastic.

For example, let us introduce the counting processes N and M as independent Poisson processes with conditionally deterministic time-varying intensities. We denote by N_{it} the default counting process of firm i with intensity $\lambda_{i,t}$ and denote M_{it} to be the other exit counting process with intensity $\phi_{i,t}$. We define the default time as stopping time $\tau_{Di} = \inf\{t \in \mathbb{R}_+ | N_{it} > 0, M_{it} = 0\}$ and the exit time due to other reasons as $\tau_{Oi} = \inf\{t \in \mathbb{R}_+ | M_{it} > 0, N_{it} = 0\}$. The survival probability for the interval $[t, t+\tau]$ is given by the probability of having no jump in the combined counting process between time t and $t+\tau$, meaning that the combined exit time $\tau_{Ci} = \inf\{t \in \mathbb{R}_+ | N_{it} + M_{it} > 0\}$ is not yet reached.

Therefore, given that the company survived until time t, the conditional survival and default probabilities to time $t+\tau$ can be expressed as [23]:

$$\mathbb{P}\left(\tau_{Ci} > t+\tau\right) = \mathbb{E}_t\left[\exp\left(-\int_t^{t+\tau}(\lambda_{is}+\phi_{is})ds\right)\right]$$

$$\mathbb{P}\left(\tau_{Di} < t+\tau\right) = \mathbb{E}_t\left[\int_t^{t+\tau}\exp\left(-\int_t^{s}(\lambda_{iu}+\phi_{iu})du\right)\cdot\lambda_{is}ds\right]$$

As the instantaneous intensities λ_{it} and ϕ_{it} are only known at or after time t, **?** propose to use the approach of *forward intensity rates*. They introduce the quantity $\psi_{it}(\tau)$ to be the spot combined exit intensity for default and other exits together and denote by $F_{it}(\tau)$ the conditional distribution function of the combined exit time evaluated at $t+\tau$. Therefore, $1-F_{it}(\tau)$ is the probability of surviving over the period $[t, t+\tau]$, which is also given in equation **??**, and:

$$\psi_{it}(\tau) \equiv -\frac{\ln(1-F_{it}(\tau))}{\tau} = -\frac{\ln\ \mathbb{E}\left[\exp\left(-\int_t^{t+\tau}(\lambda_{is}+\phi_{is})ds\right)\right]}{\tau}$$

Thus, the surviving probability over the period $[t, t+\tau]$ can be also expressed as $\exp(-\psi_{it}(\tau)\cdot\tau)$. Moreover, **?** makes the assumption that ψ_{it} is differentiable and define the forward

[23] Please refer to proof of Proposition 1 in **?**

combined exit intensity as:

$$g_{it}(\tau) \equiv \frac{F'_{it}}{1-F_{it}} = \psi_{it}(\tau) + \psi'_{it}(\tau).$$

which gives $\psi_{it}(\tau)\tau = \int_0^\tau g_{it}(s)ds$. In order to define the forward default intensity, we need to separate the combined exit time of firm i, τ_{Ci}, from the default time, τ_{Di}. If the firm exits due to default we have $\tau_{Ci} = \tau_{Di}$, otherwise it exits due to other reasons while not being in default and we have $\tau_{Ci} < \tau_{Di}$. The forward default intensity $f_{it}(\tau)$ for $[t+\tau, t+\tau+\Delta t]$ is defined by the conditional probability of defaulting in the time interval given that the firm has survived until $t+\tau$:

$$\begin{aligned} f_{it}(\tau) &\equiv e^{\psi_{it}(\tau)\cdot\tau} \cdot \lim_{\Delta t \to 0} \frac{\mathbb{P}_t(t+\tau < \tau_{Di} = \tau_{Ci} \leq t+\tau+\Delta t)}{\Delta t} \\ &= e^{\psi_{it}(\tau)\cdot\tau} \cdot \lim_{\Delta t \to 0} \frac{\mathbb{E}_t\left[\int_{t+\tau}^{t+\tau+\Delta t} \exp\left(-\int_t^s (\lambda_{iu} + \phi_{iu})du\right) \lambda_{is} ds\right]}{\Delta t}. \end{aligned}$$

The default probability over the period $[t, t+\tau]$ can now be written as $\int_0^\tau e^{-\psi_{it}(s)s} f_{it}(s)ds$.

The approach in **?** models $f_{it}(\tau)$ and $g_{it}(\tau)$ as exponential functions of some state variables and uses a maximum likelihood function to estimate the two exit intensities. The likelihood for each firm is :

$$\begin{aligned} \mathcal{L}_{\tau,i,t}(\alpha,\beta) = &\ \mathbf{1}_{t_{0i}\leq t, \tau_{Ci} > t+\tau} \cdot P_t(\tau_{Ci} > t+\tau) \\ &+ \mathbf{1}_{t_{0i}\leq t, \tau_{Ci} = \tau_{Di} \leq t+\tau} \cdot P_t(\tau_{Di} = \tau_{Ci} \leq t+\tau) \\ &+ \mathbf{1}_{t_{0i}\leq t, \tau_{Ci} \neq \tau_{Di}, \tau_{Ci} \leq t+\tau} \cdot P_t(\tau_{Di} \neq \tau_{Ci}, \tau_{Ci} \leq t+\tau) \\ &+ \mathbf{1}_{t_{0i} > t} + \mathbf{1}_{\tau_{Ci} \leq t} \end{aligned}$$

The likelihood for each firm is a sum of indicator functions multiplied by their respective probabilities, covering five independent cases that can occur during time interval $[t, t+\tau]$. Discretizing the model allows us to express the forward default and combined exit probabilities in terms of f_{it} and g_{it}, which we will estimate. The discretized version of the probabilities expressed in the likelihood function is:

$$P_t(\tau_{Ci} > t+\tau) = e^{-\sum_{s=0}^{\tau-1} g_{it}(s)\Delta t}$$

$$P_t(\tau_{Di} = \tau_{Ci} \leq t+\tau) =$$

$$= \begin{cases} 1 - e^{-f_{it}(0)\Delta t} & \text{if } \tau_{Ci} = t+1, \\ e^{-\sum_{s=0}^{\tau_{Ci}-t-2} g_{it}(s)\Delta t} \cdot \left(1 - e^{-f_{it}(\tau_{Ci}-t-1)\Delta t}\right) & \text{if } t+1 < \tau_{Ci} \leq t+\tau \end{cases}$$

$$P_t(\tau_{Di} \neq \tau_{Ci}, \tau_{Ci} \leq t+\tau) =$$

$$= \begin{cases} e^{-f_{it}(0)\Delta t} - e^{-g_{it}(0)\Delta t} & \text{if } \tau_{Ci} = t+1, \\ e^{-\sum_{s=0}^{\tau_{Ci}-t-2} g_{it}(s)\Delta t} \times \left(e^{-f_{it}(\tau_{Ci}-t-1)\Delta t} - e^{-g_{it}(\tau_{Ci}-t-1)\Delta t)}\right) & \text{if } t+1 < \tau_{Ci} \leq t+\tau \end{cases}$$

B Construction of distance-to-default

One measure relevant to estimating the default risk of a firm is its distance-to-default. **?** describe the distance-to-default as the number of standard deviations of annual asset growth by which the asset level exceeds the firm's liabilities. Here we follow **?**, **?**, and **?** and define the distance-to-default of a given company at time t, over the period T, as:

$$DtD_{i,t} = \frac{\ln\left(V_{A,t}/L_t\right) + \left(\mu_A - \frac{1}{2}\sigma_A^2\right)T}{\sigma_A\sqrt{T}}$$

Typically, the chosen time horizon T is of four quarters. We can see that for computing DtD_i we need the debt value L_t with time to maturity T, the market value of assets V_A, the mean rate of asset growth μ_A, and asset volatility σ_A. We will explain next how to solve iteratively for asset market value and asset volatility.

We use the **?** formula and state the market value of equity V_E, seen as a call option on V_A with time to expiration T and strike price L_t, and its volatility (derived from Ito's Lemma and the Geometric Brownian Motion assumption of the model):

$$V_E = V_A\Phi(d) - Le^{-rT}\Phi(d - \sigma_A\sqrt{T}) \tag{13}$$

$$V_E \sigma_E = V_A \sigma_A \frac{\delta V_E}{\delta V_A} \tag{14}$$

where

$$d_1 = \frac{\ln\left(V_A/L\right) + \left(r + \frac{1}{2}\sigma_A^2\right) T}{\sigma_A \sqrt{T}}$$

and r is the risk-free rate and Φ is the cumulative density function of the standard normal distribution.

We calculate σ_A together with V_A by solving equations **??** and **??** in an iterative procedure. For each quarter, we look at daily stock price information over the previous year and use the estimate of the volatility of equity, σ_E, as initial value for σ_A. We then use the Black-Scholes formula and compute V_A for each trading day over the last year, using the daily price information and the number of shares outstanding for $V_{E,t}$ and the sum of total liabilities and half of total long term debt for L_t. As risk-free rate, we use the one-year T-bill rate at daily frequency. We compute the asset volatility σ_A as the standard deviation of daily V_A and use the new σ_A for the next iteration. We repeat this procedure until the value σ_A converges up to a tolerance level of 10E-4. Once σ_A converged, we use its value in the Black-Scholes formula to compute the final market asset values. Finally, we estimate the drift μ_A with the mean of changes in $\ln(V_A)$. We repeat the above procedure for each end of the quarter, with an estimation window always kept at one year, in such way that we obtain V_A, μ_A, and σ_A needed to compute $DtD_{i,t}$.

C Computing the accuracy ratio

There are several ways to evaluate the ability of a model to discriminate between different classes of objects, which in our case represent defaulted and non-defaulted firms. Two of the most commonly used statistics are the Cumulative Accuracy Profile (CAP) and its summary statistic, the Accuracy Ratio (AR), and the Receiver Operating Characteristic (ROC) and its summary statistic, the area below the ROC curve (AUC) (see **?**). The CAP (which is explained in details in **?**) is the most popular technique currently used in economic and financial practice, whereas the ROC curves are mostly used in medicine and psychology.[24] In this paper, we follow the standard in the credit risk literature and use the CAP and the accuracy ratio to evaluate the performance of different credit risk model specifications. In this way, we can compare our results with the previous ones

In order to compute the accuracy ratio, we start by ranking the companies according to their estimated default intensity as predicted by the mode, from highest to lowest. Then, for every integer $\lambda \in 1..100$, we check how many companies actually defaulted (M_λ) among the companies within the $\lambda\%$ of firms with the highest default risk and record the corresponding number of defaulted companies as a percentage of total number of defaulted companies (M) in the sample over the time horizon:

$$f(\lambda) = \frac{M_\lambda}{M}$$

where $f(\lambda)$ takes values between 0 and 1, and is an increasing function of λ. Moreover, $f(0) = 0$ and $f(100) = 1$. In case of a perfectly performing default risk model, which would capture all defaults for each integer λ, $f(\lambda)$ would be given by:

$$f(\lambda) = \frac{\lambda \cdot N}{M} \quad \text{for} \quad \lambda < \frac{M}{N}, \qquad \textit{and} \qquad f(\lambda) = 1 \quad \text{for} \quad \lambda \geq \frac{M}{N}$$

In case we would have no information about the likelihood of default, the companies would be ranked randomly and we would eventually obtain $f(\lambda) = \lambda$.

For testing the performance of the credit risk model, we need to compute $f(\lambda)$ for each quarter, while always keeping one-year horizon, and take the average $f(\lambda)$ over all quarters covered by the testing sample. In the case of an uninformative model, the average $f(\lambda)$ function would correspond to a 45° line. The performance of a model is thus measured by the area between the graph of the average $f(\lambda)$ function and the 45° line, where a further lying average $f(\lambda)$ function from the 45° line means a better performance.

The accuracy ratio is defined as the ratio between the area associated with the model's average $f(\lambda)$ function and the one associated with a perfectly performing model (see Figure **??**, where $AR = \alpha_R/\alpha_P$.)

[24]The relationship between the accuracy ratio and the area under the ROC curve is $AUC = 1/2 \times (AR + 1)$

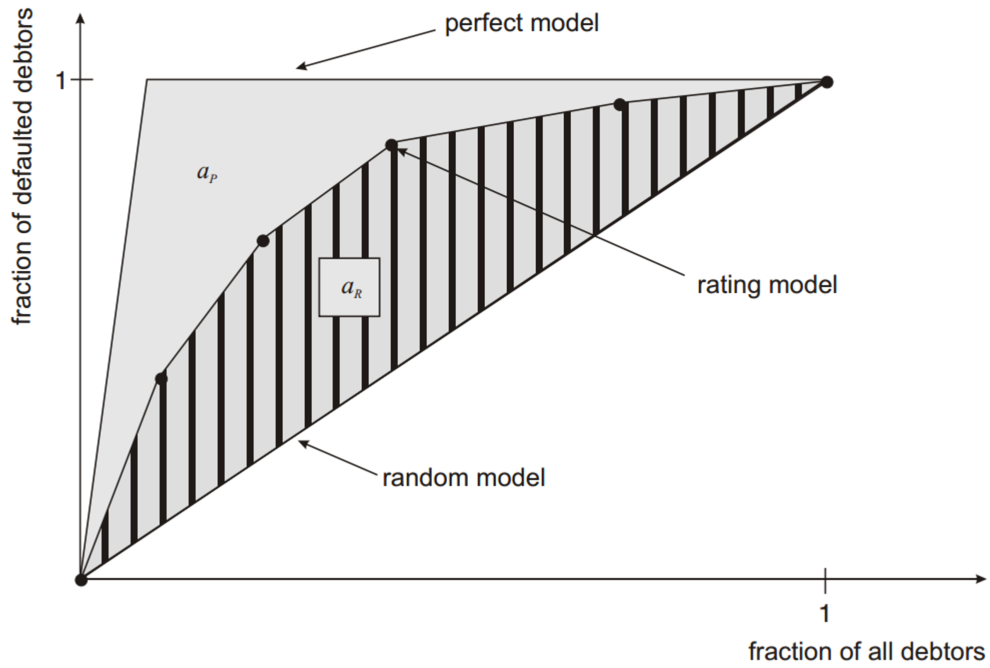

Figure 6: *Source:* **?**. This figure illustrates the concept of a Cumulative Accuracy Profile. For each percentage λ on the horizontal axis, the polygon shows the fraction of companies that defaulted within one year that were ranked in the λ% of firms with the highest estimated default probability at the beginning of the period. The upper line represents the case of perfect information, where all defaults are assigned to the lowest rating scores. The straight line below represents the naive case of zero information or random assignment of rating scores. The Accuracy Ratio is the ratio of the performance improvement of the model being evaluated over the naive model (α_R) to the performance improvement of the perfect model over the naive model (α_P).