Big Data in Credit Risk Management

L. Bottazzi and C. Tebaldi
University of Bologna, Bocconi University and IGIER

April 5, 2019

Abstract

Using information on more than 100K corporate customers extracted from the data-lake of a major italian lender, we explore the possibility to use big data analytics to identify channels of direct distress contagion to define optimal lending policies in the presence of network externalities among borrowers.

Traditional reduced form and structural credit risk pricing models are implicitly based on the assumption that only variations in default intensities, driven by systematic factors, steer the risk of aggregate portfolios. See for example Jarrow, Lando and Yu (2005). This is equivalent to the assumption that fluctuations driven by firm specific default events can be completely diversified out. Recent empirical evidence, see for example Azizpour, Giesecke and Schwenkler (2018), highlights that the impact of direct contagion driven by credit-events is the major source of default clustering. We may expect that contagion takes place through equilibrium firm-to-firm linkages that arise endogenously both for production, like supply-chains, and/or financial reasons, that is bank-firm financing relationships, or both, like trade-credit relationships. These linkages form a network that plays a major role in propagating distress shocks.

Recently, Jacobson and Von Schedvin (2015) quantify the importance of trade credit chains for the propagation of corporate bankruptcy using an exhaustive data set on claims held by Swedish trade creditors (suppliers) on failed trade debtors (customers). Controlling for creditor characteristics, they find that trade debtor failures are associated with substantially enhanced bankruptcy risks for the involved trade creditors. The estimated average marginal effect implies an increase in annual creditor failure risk by around 53% when exposed to a debtor failure. Losses by the creditors are due both to lack of repayments and a drop of demand. The propagation mechanism explains a significant part of the aggregate bankruptcy frequency, especially during economic downturns. A similar quantification has been confirmed by Arata (2018) using high performance computing facilties to exploit information from a comprehensive Japanese firm-level dataset. These authors, however, find that small-world structure of the network has a considerable dampening effect in bankruptcy propagation.

The above results support a more intensive use of the data-lakes where the banks store private information on corporate customers. These data may play an important role to improve credit risk management at the institutional level and may be useful to improve informational content of early warning indicators, credit risk scoring and pricing, used by financial intermediaries, above and beyond traditional approaches.

The present research project challenges this view, relying on the analysis of a propietary data set provided by a major italian bank. The data include high-frequency, high quality information

on 100K corporate customers. The dataset has been collected devoting specific attention to the information that may be useful to detect and track the network of productive and financial relations between customers, on top of the direct relationship that the bank has with them. The peculiar structure of the italian industrial system, populated by a high number of small and medium inter-related enterprises working as suppliers of big manufacturers seems particularly relevant to assess the informativeness of proprietary data. The dataset has been organized following a 'network based' architecture. Each information is classified as 'node' (i.e. single firm) or 'edge' (firm-to-firm linkage).[1]. Preliminary descriptive analysis based on the definition of new network based measures (e.g. firm centralities) and network extensions of traditional credit risk monitoring indicators (e.g. a network extension of the Z-score of Altman (1968)) confirm the significance of the information carried by observed network relations improving out-of-sample predictability.

It is fair to say that, looking at this dataset through the lens of traditional applied microeconometric analysis a number of critical issues have to be overcome to produce effective big-data, network based credit policy tools, relying only on the partial information available to the bank. We briefly list below the one that we find most relevant and that we plan to address in the next future.

First of all, an issue of endogeneity arises. In fact, it is important to observe that the quality of information on corporate creditors available to the bank is heterogeneous and endogenously determined by the type of relationship established by the customer with the bank. Secondly, one important problem is to establish the quality of information that can be retrieved from the bank dataset.

In fact a typical trait of italian firms and individuals is that they establish multiple lending relationships with different banks. Hence, having data only from one bank gives us only a partial vision of the bank customer relation as the full net of linkages are only partially observed. Partial information from other lenders can be retrieved from the national credit registry that we can access and that we plan to merge with the bank data. Therefore a screening and an a priori assessment of the lending relationship is needed to infer the quality of the information available to the bank that depends on the type of bank-customer undergoing relationship.

From a methdological point of view this analysis requires the use of relevant automated diagnostic tools to profile each customer (a node) on the basis of the individual financial information and its relation to the network. We aim to find efficient identification strategies that may be useful to pin down robust information, on persistent channels of propagation of distress risk that may arise as an effect of two type of interaction: i) the bank-firm and ii) firm-to-firm linkages relying on recent innovative approaches to the analysis of credit demand-supply shocks, see for example Amiti and Weinstein (2018).

A viable approach to risk mitigation consists in analyzing counterparty risk contagion adopting a disease spreading model. We can think at the bank as an entity whose goal is to establish the best structure of the 'vaccination' campaign to reduce the possibility of contagion while keeping into account the most persistent connections that characterize the portfolio of bank customers, see e.g. Trajanovski et al. (2015). In this respect,the traditional credit risk measures should be complemented with contagion risk measures similar to those adopted in the analysis of systemic risk analysis.

Having properly identified interfirm relationships and the impact on credit risk, the second

[1]This type of dataset is known in big-data science under the name: relational datasets modeled as graph. Techniques to handle this type of dataset and get reliable measures of connectivity are discussed inAvella-Medina et al. (2018)

step of our work is the definition of a machine learning algorithm to assess the optimal response of the bank management to the individual demand of credit products from lenders, in the presence of endogeneous and exogeneous shocks impacting the portfolio. In this case, the bank plays the role of a central planner that sets prices and quantities of loans and financial services optimizing the global tradeoff between revenue maximization and risk mitigation while keeping into account network externalities of each intervention.

References

Altman, Edward I. 1968. "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy." The journal of finance, 23(4): 589–609.

Amiti, Mary, and David E Weinstein. 2018. "How much do idiosyncratic bank shocks affect investment? Evidence from matched bank-firm loan data." Journal of Political Economy, 126(2): 525–587.

Arata, Yoshiyuki. 2018. "An Empirical Analysis of the Propagation of Corporate Bankruptcy." Research Institute of Economy, Trade and Industry (RIETI).

Avella-Medina, Marco, Francesca Parise, Michael Schaub, and Santiago Segarra. 2018. "Centrality measures for graphons: Accounting for uncertainty in networks." IEEE Transactions on Network Science and Engineering.

Azizpour, Shahriar, Kay Giesecke, and Gustavo Schwenkler. 2018. "Exploring the sources of default clustering." Journal of Financial Economics, 129(1): 154–183.

Jacobson, Tor, and Erik Von Schedvin. 2015. "Trade credit and the propagation of corporate failure: an empirical analysis." Econometrica, 83(4): 1315–1371.

Jarrow, Robert A, David Lando, and Fan Yu. 2005. "Default risk and diversification: Theory and empirical implications." Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics, 15(1): 1–26.

Trajanovski, Stojan, Yezekael Hayel, Eitan Altman, Huijuan Wang, and Piet Van Mieghem. 2015. "Decentralized protection strategies against SIS epidemics in networks." IEEE transactions on control of network systems, 2(4): 406–419.