Cambridge University Press
978-1-107-02343-7 - Handbook on Systemic Risk
Edited by Jean-Pierre Fouque and Joseph A. Langsam
Excerpt
More information

PART I

DATA: THE PREREQUISITE FOR MANAGING SYSTEMIC RISK

Edited by

H.V. Jagadish
University of Michigan

Cambridge University Press
978-1-107-02343-7 - Handbook on Systemic Risk
Edited by Jean-Pierre Fouque and Joseph A. Langsam
Excerpt
More information

Data for Systemic Risk

H. V. Jagadish

Whatever models may be constructed to assess systemic risk, they will require data to evaluate. This Part of this book explores the data needs for evaluating systemic risk, and describes the many challenges in meeting these data needs effectively.

Systemic risk models are particularly complicated because they are aggressively non-linear – an entity is either able to meet its contractual obligations or it is not, and which of the two scenarios we are in can affect the solvency of many other entities – and furthermore have heavy interlinkage between model entities so that independence assumptions are almost never possible. In consequence, the common mathematical simplifications, of linearity and independence, cannot be made. This leads to the need for models that are far more complex, potentially require Monte Carlo simulations to solve, and rely upon a large body of data and computation. In short, a large variety of detailed (granular) data is likely to be required, from a range of market participants. The size of this dataset makes computational assessment imperative: its complexity makes purely computational evaluation challenging. The five chapters in this Part together elucidate this challenge and also suggest directions toward a solution.

The first chapter, by Edward Hida (Deloitte and Touche), describes the typical financial information processing environment in several major types of financial services firms. Based on this, we can see what types of data are likely to be easy to obtain, and which may require more effort. The second chapter, by Roger Stein (Moody's Research), discusses the types of models one may wish to run to evaluate systemic risk, and points out the differing information needs for these. He also highlights some of the difficulties in obtaining fine-grained linked information, but points out that many models can do fine with less data. These two chapters together describe the current state of financial information as it exists, and the analyses that are possible within this context. In particular, the material in these chapters suggests the types of data that regulatory agencies may seek to be able to assess systemic risk.

Unfortunately, current practice leaves electronically unrecorded (or recorded electronically in a manner that makes it difficult to access) many important "details." It is likely that details not electronically captured are unlikely to be represented in valuation models, systemic risk models, or systemic risk analysis. This may be acceptable in every day operation because not all the details are required for any given function for which data is collected. For example, many corporate bond valuation and risk models do not require knowledge of the covenants. However, knowledge of these covenants may become important during a systemic crisis to support forensic financial research or to better understand credit freezes. More fundamentally, valuation models require information about cash flows under all states being modeled. Included in these states are those that are taken as parameters within the mathematical model as well as those whose stochastic evolution is being modeled. Capturing the state-specific cash flows is not sufficient. It is necessary to capture the contractual determinants of state changes. The descriptors for the cash flows and contributors to determining state changes are described in the contractual details. The contracts often will reference other contracts. Typically, there are important restrictions and covenants associated with any contract. Even for something as straightforward as a corporate bond offering, the prospectus usually runs into multiple pages. Even if most clauses in the contract rarely come into play, it is important to recognize that these details are in a contract for a purpose. Some details may only be referenced in extreme circumstances. These rare occurrences often coincide with systemic crises. Full knowledge of the contractual details is necessary for determining future cash flows under the states being modeled.

In short, current accounting practices can give us useful gross metrics such as concentration of exposure to a counterparty. However, they lack the information necessary to unravel contracts in atypical scenarios. They also may not capture transitive dependence on counterparties, through sequences of contracts. To be able to evaluate these more complex models that capture interconnections between parties, or even just to simulate possible scenarios, we require at least that contracts be captured electronically. The current state of the art in this regard is a language called FpML managed by the International Swaps and Derivatives Association (ISDA). The third chapter in this Part, by Andrew Jacobs (UBS) and Marc Gratacos (ISDA), describes the features and limitations of FpML.

Our discussion thus far has assumed that the data we are working with is all organized in the same way. In practice, variations abound. The final two chapters in this Part address this challenge. The one by Arnon Rosenthal and Len Seligman (Mitre), considers many of the challenges to practical data integration at large scale. They eloquently point out that standards are a good thing, and should be pushed as far as we can. However, it is unrealistic to expect to create a universal standard for everything. Therefore we need techniques to deal with heterogeneity of representation and of values. For example, the current efforts to establish a universal

LEI (legal entity identifier) will move us forward significantly, and therefore is an important step to take; yet no one expects that the mere existence of LEI will solve all problems of financial data integration. Rather, having consistent use of LEI will mean there is one less problem to solve and one less source of error in analysis.

The final chapter by Mike Atkin and Mike Bennett (EDM Council), describes one way in which to deal with heterogeneity across information systems. The idea is to have each data entry represented as an instance of a concept that has been "properly defined" in terms of some ontology. By these means, at least there can be no misunderstanding of what the value represents. Then, by means of conceptual "cross-walking" across ontologies, we can merge information from multiple information systems created following different ontologies. While there are limits to this cross-walking, and possible information loss entailed, the first step, of having establishing well-grounded semantics, is well understood for many financial systems. In fact, the EDM Council has developed an ontology for this purpose that shows considerable promise.

In summary, managing and manipulating data is central to evaluating systemic risk. This Part considers how best to do this, going from the present to the future and going from the modeling needs to the technological possibilities. Along the way, it points to the great deal of research work required to build the data infrastructure necessary to get a handle on systemic risk. These topics include:

- Developing an effective semantic definition for all (or at least most) financial objects of interest. The Atkin–Bennett chapter sheds light on the current state of the art.
- Developing an efficient computer-readable specification for all (or at least most) financial objects of interest. The Jacobs–Gratacos chapter shows how FpML does this for many derivative agreements.
- Having the semantic definitions match the elements in the specification, so that a computer reading the specification can get the semantics right.
- Linking data across multiple repositories of interest. Some standards exist already, such as CUSIP. Some are being agreed upon at present, such as LEI. But many other entities remain without standards. Furthermore, linking data across databases often requires more than just identifying the linking entity – we need the organizational structure, we may need the semantics, and so forth. The Rosenthal–Seligman and Stein chapters have more to say on this issue.
- Once all the above piece-parts are there, developing automated techniques to reason through multiple chains of contracts and determining exposures or risks of interest.
- While there is a need to make technological progress along the lines of the bullets above, we must accept that there will be limits to this progress. We must also try

to get useful analyses performed meanwhile. Limitations of current practice are described in the Hida chapter. Trade-offs between what we would like and what we have are extensively considered by both Stein and Rosenthal–Seligman.

We note that there are many additional issues that are likely to arise in the course of creating and using the data infrastructure required. For lack of space, we do not cover these in detail: instead we mention them briefly here.

Information Extraction: Not all data of interest will be captured in structured reports being designed and the infrastructure being proposed here. There will remain data of importance that is available only in text documents, such as contract riders or collateral details unavailable in electronic form. We will need to apply natural language processing techniques to try to get computers to parse and "understand" these documents, or at least to use information extraction techniques to pull out facts of importance and represent these in structured form amenable to further analysis. Techniques currently exist for achieving at least partial success in such an endeavor. Further work is required to improve these techniques and to tune them to perform well for the financial and legal documents of interest.

The same information extraction problem also appears in other contexts related to financial information. For example, financial news services, such as Bloomberg and Reuters, are recognized as being of tremendous value. Much of this information is provided in textual form, for human consumption. Computerized processing requires information extraction. Similarly, human social networks, such as Twitter, comprise text messages. Analysis of these, as described in the last paragraph of this introduction, requires information extraction as a first step.

Data Cleaning: There will be errors in the data. These may be errors in the original data record, or could be errors introduced by the information integration or information extraction processes applied computationally, both of which will be less than 100% accurate. Models run on erroneous data will produce erroneous results. The more complex the model, the more likely that the erroneous results will be accepted at face value without questioning their basis and without finding errors in the input that caused the wrong conclusions. For these reasons, it is crucial that data be thoroughly cleaned prior to analysis. There is work on data cleaning technologies, but this is hard to do in general, without knowledge of domain or error model. In the context of the financial information we are gathering, we need to develop data cleaning methods particularly designed for such data.

In financial reporting, current practice often has a restatement of numbers as errors are corrected and initial estimates are rendered more accurate. In other words, there is a versioning of some financial data. As we build a complex chain of derivations, it is important to know which versions of the base data were used in the derivation, and also to update derived data as the base data is updated. The former issue is addressed by means of provenance technology, discussed next; the latter is addressed by database view maintenance technology.

Data Provenance: We may often care from where a piece of data was obtained. There are several reasons why this can matter. We just discussed that many financial data are restated. We must keep track of the version used in such instances. We know that there are sometimes errors in reporting, whether due to fraudulent intention or not. If we need to reconcile data from two or more entities, we need to keep track of which piece of data came from which entity. Not all entities define financial terms in the same manner. While standardization and the use of appropriate ontologies can alleviate this problem, we should expect that their application will remain imperfect in the foreseeable future. It becomes important to note whose definition of financial term is used (based on who reported the data) if this is ever to be unraveled. Finally, even if we start with the same base data, there can be a variety of models applied to it, with a range of possible assumptions and parameter values. Any derived results must clearly indicate how they were derived.

The field of data provenance has developed in recent years. It involves techniques for the capture, storage, manipulation, and querying of provenance information to go along with data. This provenance is "meta-data" (data about the data) that can address questions such as those listed in the preceding paragraph. Challenges in doing this effectively have to do with capturing provenance without adding undue overhead to data collection and processing workflows, storing provenance without hugely increasing the amount of data storage required, and using provenance effectively to answer user questions in a meaningful way.

Data Visualization: The result of a model execution may sometimes be a single number (or even a single bit – e.g. there is a bubble/there is not a bubble). More often, an analyst, whether a regulator or a risk manager at a financial institution, will need more detail. Given the large volume of the data, and the complexity of the models used, it is easy to overwhelm the human. Computer systems must be designed to be appropriately informative. A decision-maker needs to know not just a final number, but also an appropriate amount of information surrounding it: this could be model details, parameter values assumed (which she may want to tweak), various intermediate results (needed so the decision-maker can trust the final result produced). Data visualization techniques are required to be able to present the decision-maker with such information in a maximally usable form.

Security and Sharing: To be able to conduct systemic analyses, information is required from multiple interconnected firms (or assumptions have to be made about the missing information). Regulatory agencies may be able to acquire the needed proprietary data. However, that leaves everyone else in the dark. If this data, suitably sanitized, were to be made public, we could have a much broader range of people analyze it, leading to greater intellectual ferment, better models, better prediction, and hence better regulation. For these reasons, regulatory agencies should consider what data they could make public, after suitably removing proprietary information, either through suppression of selected specifics, or through aggregation. (After a suitable delay, most financial data can be made public since it no longer has proprietary value. This should be done, and does not pose a technical challenge). Financial institutions too should want such managed disclosure because this disclosed data could help them create better risk models for their own use as well.

There has been considerable work done on how to disclose data in a manner suitable for analysis without disclosing information that must be kept private, of under the label of privacy-preserving data mining. A common scenario in which to consider this problem is a medical record, where individual patient information should be kept private, but there is a great deal of valuable information that can and should be derived in the aggregate. This type of analysis techniques should also be developed for financial information.

Human Choice: Many contracts have options that may or may not be exercised. For example, a homeowner has an option to prepay a mortgage at any time, closing out the contract (usually done by selling the house or by refinancing with a new loan, possibly from a different lender). Whereas stock options are always exercised if they expire in the money, other options may require an understanding of human psychology to model. Going beyond options, there is a growing understanding of how many people trade based on sentiment or momentum. These aspects of human choice are not normally considered in financial models. As we build systemic models, we will probably need to understand better the humans in the loop, and to model them better as well. There is work on sentiment analysis, based on blog posts, review sites, or tweets. The idea is to have a computer monitor such sources, and to point to trending changes in sentiment.

To conclude this introduction, there are many important topics on which further work is required as we improve our ability to manage data for systemic risk. The five chapters in this Part lay out the ground for some of the more fundamental topics that must be addressed.

1

Systemic Risk Information Requirements: Current Environment, Needs, and Approaches for Development

Edward T. Hida II

Abstract This chapter begins by describing the current financial information processing environment in a typical financial services firm. Then it points to the many places in this environment where data of value for systemic risk assessment can be found. This leads to an assessment of the data elements that are likely to be found with moderate ease and those that will likely be more difficult to obtain. These will be compared with the data collection mandates being undertaken by the systemic risk regulator and financial market utilities. This assessment of disparate data will set the stage for a discussion of data integration in later chapters of this Part.

1.1 Introduction

The global financial crisis highlighted the need for greater availability and transparency of standardized information to assess and monitor systemic risk. During the crisis, regulators and others needed timely information to monitor the health of financial firms, understand their exposures, and assess concentrations and interconnections between firms and within markets. This information was not always readily available. As a result, there has been an intensifying focus on information related to systemic risk oversight.

Systemic risk oversight looks at the risks to the overall financial system and the interactions between financial institutions and between markets. Such oversight has the potential to broaden the regulatory view from the traditional "microprudential" focus on individual institutions to a broader "macroprudential" focus on the financial system and on the potential for contagion in the financial system.

Given this, regulators performing systemic risk oversight will be tasked with receiving and monitoring types of information which they have typically not received before. Measures and information which enable assessment of network effects and interactions and concentrations will be needed. These include measures of network strength and resiliency and information about network nodes and interconnections[1]. Information long used in risk management may be viewed and used in new ways. A focus on counterparty risk exposures, with more information about the interactions between counterparties will likely be needed. In addition, detailed transaction and position information may be used by the systemic risk regulator in the analysis of overall exposures and concentrations across the financial system. Other information which can identify linkages in transactions and exposures will also be required.

In general, systemic risk oversight approaches are expected to require more and different information from financial firms, markets and others. Thus, the quality of systemic risk oversight will largely depend on the accessibility and quality of this information and the ability of financial firms to provide it.

1.2 Purpose

In this chapter we will provide an overview of the types of information required for systemic risk monitoring. We will then summarize some elements of the financial information processing environment in financial services firms. We will point to the many different places in this environment where important pieces of data of value for systemic risk assessment can be found. This leads to an assessment of the data elements that are likely to be found with moderate ease and the data elements that will be more difficult to obtain. These will be compared with the data collection mandates being undertaken by the systemic risk regulator and financial market utilities. This assessment of disparate data will set the stage for a discussion of data integration in later chapters of this Part.

1.3 Overview of types of systemic risk information required

Effective systemic risk monitoring will require a wide variety of information. In an Office of Financial Research ("OFR") working paper, Bisias et al. report on a survey[2] of a wide range of systemic risk analytics developed by academics and others. They also provide taxonomies of systemic risk measures organized by data requirements, supervisory scope, event/decision time horizon, and research method.

In this chapter we will focus primarily on information obtained or derived from transaction activities of firms. Position and transaction information will be required to monitor the exposures of individual firms as well as to assess concentrations in

[1] Andrew G. Haldane, "Rethinking the Financial Network," Speech delivered at the Financial Student Association, Amsterdam, April 2009, accessed March 10, 2012, `http://www.bankofengland.co.uk/publications/speeches/2009/speech386.pdf`.

[2] Dimitrios Bisias, et al, Working Paper #0001, "A Survey of Systemic Risk Analytics," Office of Financial Research, January 5, 2012, accessed March 10, 2012, `http://www.treasury.gov/initiatives/wsr/ofr/Documents/OFRwp0001_BisiasFloodLoValavanis_ASurveyOfSystemicRiskAnalytics.pdf`.