Limits and Opportunities of Big Data For Macro-Prudential Modeling of Financial Systemic Risk

Willi Brammertz
Lecturer University of Zurich and ETH
Weiherallee 29
8610 Uster, Switzerland
+41-79-420-0750
willi@brammertz-consulting.ch

Allan I. Mendelowitz
Deloitte Consulting
1507 Blue Meadow Road
Rockville, MD 20854
+1-301-279-0744
amendelowitz@yahoo.com

ABSTRACT
We explore the use of "big data", i.e. large unstructured data sets, within financial risk analysis. We conclude it has value, but structured data remain critical. We show that forward-looking financial analysis on the systemic level needs a data structure that represents financial contracts as algorithms that produce state contingent cash flows. Currently the industry lacks such a standard, which precludes meaningful systemic forward-looking analysis. We introduce ACTUS as an emerging standard that will enable consistent analysis on all levels. This standard will also create an infrastructure for macro financial analysis[1].

General Terms
Algorithms, Economics, Finance, Standardization

Keywords
Data Standard for Financial Contracts, Financial Modeling, Systemic Level

1. INTRODUCTION
The failure of micro-prudential regulators and macroeconomic forecasters to predict the collapse of the financial system makes a strong case for new approaches for modeling financial risk at the macro level. Analysis of "big data", which we interpret as analysis of large unstructured data sets, has contributed to understanding of and providing insights into a range of topics from consumer behavior to threats to national security. However, an open question remains as to how it can contribute to an understanding of threats to financial stability. In this paper we look at the question of financial risk analysis, the data needed to understand threats to financial stability, which parts of the data might be gathered from the analysis of big data, and which parts need a complementary approach. This paper will explore insights that can be gleaned from big data and present a complementary approach that can extract additional information and increase understanding of threats to financial stability. Answers to such questions are closely related to the problem of risk aggregation.

2. FINANCIAL RISK ANALYTICS AND DATA REQUIREMENTS
Modern financial analysis is forward looking. Analysts want to know what happens to the financial sector when risk factors, such as market risk or credit risk, change. Such analyses can be focused on a single financial contract, a portfolio of contracts, a business entity such as a bank, or the entire financial system.

The starting point for such forward-looking analysis is the state contingent expected cash flows of individual financial contracts. For example, the analysis of the future fair value, income, or impact on liquidity of a financial contract can be represented by the following figure.

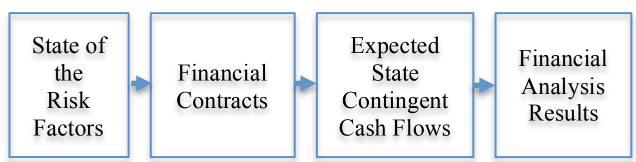

Figure 1. Logic Flow of the Elements of Forward Looking Financial Analysis

The risk factors are the stochastic inputs to financial analysis; they include market rates, counter-party ratings, and behavior elements. Financial contracts, the second box, are the deterministic elements in the financial system. They are the legally binding agreements, between two or more counterparties to exchange certain cash flows at certain times. The actual cash flows that will be exchanged under these agreements depend on both the terms of the contracts and the state of the risk factors. For example, a plain vanilla variable rate bond would have terms that reset the interest rate according to a specific cycle based on a spread over a specified market rate, such as Libor. Given the rules of the contract and given a state of the risk factors, the expected cash flows – the third box – can be derived. The derived cash flows are called "state contingent" due to their dependency

[1] Disclaimer: The views expressed herein represent the personal views of the authors and do not represent the views of any organization with which they might be affiliated.

DSMM'14, June 27 2014, Snowbird, UT, USA

ACM 978-1-4503-3012-1/14/06…$15.00.
http://dx.doi.org/10.1145/2630729.2630741

on the state of the risk factors. These expected cash flows are the critical input for analysis represented by the last box.

The sources of the risk factor data depend on the specific risk factor in question. Market data, such as interest rates and foreign exchange rates are available from commercial vendors such as Reuters and Bloomberg. On the other hand, individual counterparty risk data requires the unique identification of the counterparty and the associated data from which credit ratings, probability of default, and credit spreads are derived. Behavior risk, such as the probability of prepayment of a residential mortgager, must be modeled separately.

Relevant financial contract data consist of the contract terms that are necessary to determine cash flow obligations. These include such terms as the notional amount of the contract, how the interest rate is determined (including interest reset cycles if applicable), an interest payment cycle, amortization patterns (if applicable), and so on.

Expected cash flows are the result of the interaction between the risk factors and the contract terms. However, because financial contracts are written in text, i.e. words, the actual contracts thus represented are not suitable for the analysis of large volumes of contracts. Forward looking analysis of large volumes of contract data requires a machine-readable mathematical representation of the contracts, i.e. algorithms that represent mathematically the contract agreements and which generate the expected cash flows that depend on the both the contract terms and the state of the risk factors.

3. ACCOUNTING STANDARDS ARE NOT A RISK METRIC

The bulk of financial data that is currently available for analysis is reported under one or another Financial Accounting Standards. The current regulatory approach to monitoring financial risk also relies heavily on such data. For example, the annual capital stress tests to which Systemically Important Financial Institutions (SIFIs) are subjected rely on results reported according to accounting standards. In such exercises the regulators specify the risk scenarios and several months later the SIFI's report back the impact of the stress scenarios on their capital levels. These exercises are very costly and time consuming; nevertheless, they are of limited value. Each SIFI uses its own internal models and simplifying assumptions to determine the impact of the stress scenarios on their balance sheets. Consequently, the results of the stress tests at one SIFI are not comparable to the results reported by another SIFI. Furthermore, the results of the stress tests do not give the regulators insight into how the conditions of the SIFIs might change under different risk scenarios.

4. ACCESS TO BIG DATA BY ITSELF IS NOT ENOUGH

The Dodd-Frank Act in an attempt to ameliorate the risks of derivatives trading pushed the clearing and settlement of swaps onto clearing houses. In order to better understand the swaps market the legislation also required the reporting of granular swaps trade data to Swaps Data Repositories[2] (SDRs). The Commodities Future Trading Commission (CFTC) implemented this requirement in regulation starting with Interest Rate Swaps (IRS) and Credit Default Swaps (CDS). Since the passage of the legislation and the adoption of the implementing regulations millions of swaps trades have been reported to the SDRs. However, there was nothing in the legislation or implementing regulations that required that the reported swaps data be in a data standard that would make possible meaningful analysis of the reported data. In fact, the data standard used varies from SDR to SDR. On February 10, 2014 the CFTC's Technical Advisory Committee held a public hearing at which the senior career officials of the CFTC testified that the reported data was "all over the place" and they could not do anything with it. They could not analyze the data to understand even notional value, they could not represent cash flows, and they simply could not analyze the data to the point at which it could be turned into useful information.

What lesson can we learn from these events? Some data may be usefully analyzed without having it represented in a data standard that supports analysis. However, other data is of less value without such a data standard. Mining of "Big data" must have as a starting point a clear understanding of what data requires a data standard to be usefully analyzed and what data can yield useful information without being represented in a data standard that supports analysis. In the case of financial contracts, an algorithmic representation of granular financial contract data is needed to generate state contingent cash flow and produce the desired analytical results.

5. STRUCTURED AND UNSTRUCTURED DATA

Where can "Big Data" provide helpful insights with respect to forward-looking financial analysis? Access to big data serves this purpose best in the area of new and alternative information on the state of risk factors. For example, it could be possible to find important insights into market risk data on the Internet. It might be possible to find some hints about future evolutions of market risks by mining the information contained in the vast amount of Tweets and Facebook posts and messages. Counter-party risk is another risk factor for which insights may be gained from mining big data. Credit risk was at the center of the 2008 financial crisis. Mortgage originators essentially dispensed with any meaningful credit underwriting in granting hundreds of billions of dollars of home mortgages. Mining big data could have yielded valuable information on the deterioration of the credit quality of home mortgage borrowers. That insight could have been used to model counter-party risk and its implications with respect to value, income and liquidity for the investors in those home mortgage.

However, while mining big data may provide important insights in to the state of the risk factors, more information is needed to precisely calculate expected state contingent cash flows and the analysis that starts with those cash flows. For example, understanding the implications of market interest rates on financial risk requires specific details on contract terms such as day count method, payment cycles, interest rate spreads over market bench marks, etc. Such information is generally private and not systematically retrievable along with interest rate data from the unstructured data contained in large databases.

Even where privacy is not an issue, big data with detailed information about granular financial contracts is rare. The only

[2] These Data Repositories are also referred to as Swaps Trade Repositories (TRs).

information that is generally available are the notional values and, in a subset of cases, the market values of individual contracts. Such information gives useful insights into historical and current measures of market size and activity. However, it does not support forward-looking analysis.

Fully fleshed-out granular financial contract data is generally not publically available. However, as the case of the CFTC demonstrates, even the regulatory agencies that have access to detailed proprietary data face a major challenge. Even granular data that includes contract level specifics can resist interpretation. The challenge of financial analysis is that it requires specific data on contract terms **and** algorithms that convert the contract terms and the state of the risk factors into state contingent cash flows. The two requirements cannot be separated.

This can be represented by slightly modifying Figure 1 above:

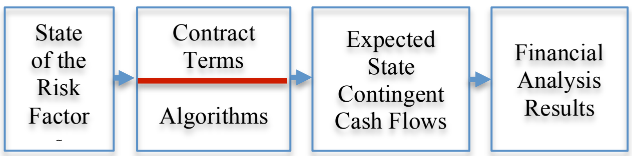

Figure 2. Contract terms and algorithmic representation of financial contracts

"Contract Terms" represent the typical contract terms such as interest rate, time to maturity, notional etc. By "algorithms" we mean the mathematical representation of the legal obligations to exchange cash flows that are included in the contracts. A classic example is the interest payment obligation contained in a contract. An interest payment obligation has to be described by a rate, an interest payment cycle and a notional amount on which the interest will be paid. The calculation also requires information on additional mechanisms, such as the day count convention, an indication whether the interest is paid or capitalized, and so on. The algorithm combines all this information and calculates the date and the size of the interest payment. Such information cannot simply be stored on a database and aggregated in a meaningful way. In order to get meaning, we need computer algorithms that are able to combine the contract terms with all necessary logic statements and the state of the risk factors. A more detailed example is provided in the Appendix.

Typically proprietary databases include the contract terms. However, they do not generally include the algorithmic representations of the legal obligations of the contracts that would support further analysis. What these algorithms represent is buried in the text of legal documents.

Once risk factors, contract terms, and the contract algorithms that represent the legal obligations of the contract are available, the calculation of the expected state contingent cash flows is a mechanical process, i.e. the simple execution of the computer code. From the contract terms, the state of the risk factors, and the algorithms we get a vector of expected dates/cash-flows. Finally, with standard analytical approaches, it is possible to derive analytic results from the expected cash flows such as fair value, liquidity, and income.

The problems the CFTC experienced are a direct consequence of the lack of standard algorithms representing the agreements of financial contracts. Lacking these algorithms we lack a common basis – or language – that supports forward-looking analysis. This lack of a standard plagues not only external oversight bodies, such as the CFTC, but banks themselves in carrying out their management obligations, including risk management. The lack of such a standard limits how the insights that can be gleaned from big data can be used to better monitor threats to financial stability. What we need is an industry standard that represents financial contracts algorithmically in order to support analysis. This will solve many of today's data problems. Furthermore, it will provide the computational infrastructure for monitoring threats to financial stability that can put the insights gleaned from big data to good use.

6. PROJECT ACTUS: BUILDING THE INFRASTURUTRE FOR MACRO-FINANCIAL ANLYSIS

Project ACTUS[3] is an undertaking to create an open source algorithmically-based data standard for finance that will support forward-looking analysis. In this project we clearly distinguish between the deterministic elements of finance and the stochastic elements. The deterministic elements are the legal obligations contained in financial contracts that represent the agreements of the contracting parties to exchange cash flows (the second box of Figure 2). The stochastic elements are the risk factors: market risk, counter-party risk, and behavior risk (the left box in Figure 2). The expected cash flows that are exchanged pursuant to any financial contract are determined by the legal obligations and the state of the risk factors. Value, income, and liquidity can be directly and consistently derived from the expected cash flows.

Project ACTUS is building a set of 30 algorithmic Contract Types (CTs) that will be capable of representing the cash flow obligations of almost all financial obligations with a high level of precision.[4]

ACTUS can directly solve the problems the CFTC faced with the SDR data. The SWAPS CT algorithm to be used for IRS contracts has been programmed and tested. The data dictionary that defines the contract terms is complete. And, public assess to try out the CT is available on the ACTUS website. The precision and the wide range of the resulting expected state contingent cash flows is made possible by the algorithms themselves and the variety of contract terms that are accommodated by the algorithms. The appendix explains the plain vanilla IRS in some detail. The power of the ACTUS approach is that the same CT that is used to model a plain vanilla interest rate swap is also used for the full range of ever more complex interest rate swap contracts.

---

[3] Project ACTUS is the implementation of the concepts developed in *Unified Financial Analysis: The Missing Links in Finance*, Willi Brammertz, et.al. The project has benefited from the financial support of the Alfred P. Sloan Foundation, Deloitte Consulting, Zurich University of Applied Sciences, and the Stevens Institute of Technology.

[4] See: www.projectACTUS.org

7. RISK AGGREGATION

Macro-level financial analysis has often been characterized as an aggregation problem [5]. While aggregation is certainly an important step, the entire process needs a more nuanced understanding.

TR´s typically contain what we described in Figure 2 as "contract terms". However, these terms are not typically candidates for aggregation. Interest payment cycles or principal payment patterns are not data that can be meaningfully aggregated. The most likely candidate for aggregation is the "notional" value of transactions. The second most likely candidate – however less often available – is the "market value". Both values are "additive" in the mathematical sense. The value of a portfolio is simply the sum of the contracts that make up the portfolio. These measures carry useful, but limited, analytic meaning. A notional value might be a good indication of credit exposure in a mortgage portfolio. However, for derivatives this value alone carries less meaning. The market value at a point in time indicates how much value has been gained or lost since inception. However, it does not indicate future cash flows and does not support forward-looking analysis.

The situation could be improved if, for example, the notional of an IRS is enhanced with data indicating the timing of its interest payments tied to the fixed and variable legs of the swap. If the IRS is a plain vanilla interest rate swap, this additional information provides some forward-looking insights. However, additional information is needed if the swaps contracts are more complicated; for example, if the contracts were amortizing swaps or constant maturity swaps. Each additional rule calls for an additional mechanism. If we follow this approach to its logical conclusion, we arrive at exactly what is proposed by ACTUS, a generalized data standard that supports financial analysis.

8. CONCLUSION

Project ACTUS is creating a computational infrastructure that makes possible the analysis and monitoring of threats to stability of the financial system. It guarantees consistent representation of financial contracts and risk factor scenarios. The analytical results supported by this infrastructure are fully consistent and can be aggregated to any level: the level of a portfolio, a financial institution, or the financial system as a whole. Thus ACTUS provides an empirical base for the financial sector currently missing in standard macro-economic models.

With the ACTUS infrastructure, analysts can make good use of new insights into the state of the risk factors derived from the analysis of Big Data

9. ACKNOWLEDGEMENTS

The authors wish to thank the reviewers of our initial draft for their insights and helpful suggestions. The authors remain solely responsible for any errors or omissions.

5 See for example "Feasibility study on approaches to aggregate OTC derivatives trade repository data" released by the FSB (Financial Stability Board) on 4 February 2014.

10. REFERENCES

[1] Brammertz, W., Akkizidis, I., Breymann, W., Entin, R., Rustmann, M. 2009. *Unified Financial Analysis: The Missing Links of Finance*, John Wiley and Sons..

[2} Financial Stability Board, 2014, "Consultation Paper: Feasibility study on approaches to aggregate OTC derivatives data" 4 February, 2014: http://www.financialstabilityboard.org/publications/r_140204.htm

Appendix: Plain Vanilla Interest Rate Swap (IRS) Example

The purpose of the appendix is to demonstrate how "Contract Terms" (see Figure 2 above) interact with a Contract Type's algorithms to generate expected state contingent cash flows. For this demonstration we use a plain vanilla interest rate swap. An IRS can be viewed as the simultaneous exchange of the cash flow obligations of two bonds: one a fixed-rate bond and the other a variable-rate bond. The fixed side of a plain vanilla IRS can be described by the following contract terms:

Table 1. Contract terms for the fixed side of the IRS

Term	Our example	Explanation
Initial exchange date*	1/1/2013	Date when the first principal flow is exchanged
Maturity date*	12/31/2015	Date when the principal flow is returned
Principal*	$1,000	Principal amount
Interest rate	5%	Per annum rate
Day count method*	30/360	Measurement of time
Business day convention*	Modified following	Rule if payment falls on non-working day
Pay or receive	Receive	Corresponds to an asset

These terms are the input for the Expected State Contingent Cash Flows shown in Figures 1 and 2. Principal payments are shown in red arrows and interest payments in green arrows. Dashed lines in red and green indicate notional values and interest accruals.

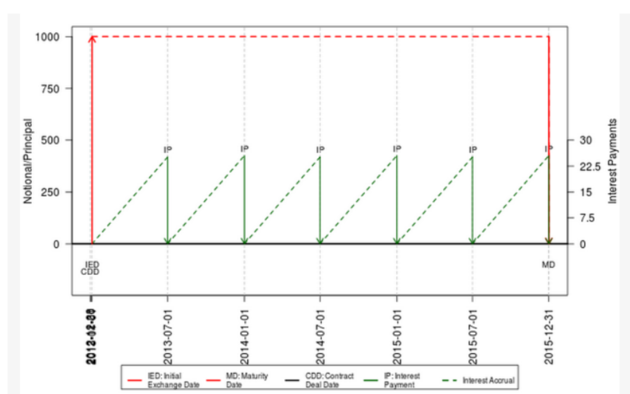

Figure 3. Expected state contingent cash flow: Fixed-interest leg of a plain vanilla IRS

Obviously there are algorithms between the data in Table 1 and the output of Figure 3. Describing these algorithms in detail would take several pages, therefore a few examples will have to suffice. Cycles, for example, play an important role in financial contracts. Interest payments, principal payments, options etc. all follow cycles. A cycle might look simple, however, cycle algorithms that determine exact dates require some hundred lines of code. It needs a start date, the notion of the cycle itself (in days, weeks, months etc.), it must be able to handle non-work days, the challenge that the length of a month varies between 28 and 31 days, irregular stubs at the beginning and the end of the cycle, and so on. A second example is the measurement of time, which in finance is an extremely complex issue. The 30/360 method, which considers every months to consist of thirty days and a year to consist of 360 days when computing the amount of interest payments, has to solve the challenge of months with 28, 29 or 31 days when scheduling interest payment dates. A large number of possible solutions exist. Even using the actual/actual day-count method, a seemingly straightforward alternative, poses problems with respect to leap years. Combining the necessary algorithms for all possible event types of even a plain vanilla bond needs more than a thousand lines of code.

We continue our example with the variable interest rate side of the IRS. Since we describe a plain vanilla IRS, all terms marked with an * in Table 1, must be the same for the variable side and "Pay or Receive" must be set to pay. In addition, at a minimum, the following attributes must be defined:

Table 2. Contract terms for the variable rate side of the IRS

Term	Our example	Explanation
Interest rate	3%	Per annum rate
Rate reset cycle	1Year	Resetting interest rate to market conditions
Reference rate	Libor USD	
Spread	2%	Rate added to the reference rate
Pay or receive	Pay	Corresponds to a liability

The following figure shows the combined results for the fixed and variable sides of this IRS example. The green arcs represent the rate resetting cycles. Note the opposing directions of the arrows which represents inflows and outflows of cash payments. As a plain vanilla swaps swap, the principal cash flows net out.

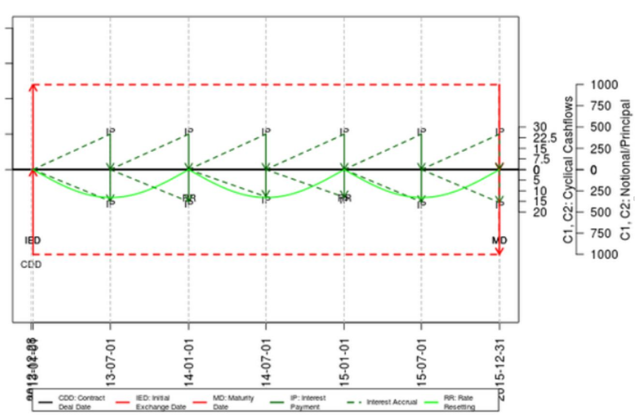

Figure 4. Contract events for a plain vanilla IRS

The more general term "contract events", rather than "expected state contingent cash flow", is used in the title of Figure 4 because the graph includes rate resets on the variable interest-rate leg of the swap. These rate resets are critical to cash flow determination. Rate resetting requires additional algorithms. These algorithms include a mechanism to point to the correct yield curve that is an external market condition. Interest rates have to be calculated which – under arbitrage free conditions – must follow implied forward rates. A spread must be added which determines the rate applied for the next cycle. These mechanisms taken together determine the next interest payment.

This example includes only the most basic algorithms. Amortizing swaps, constant maturity swaps and other variations all call for additional algorithms, some of which involve considerable complexity.

The events of our example can also be shown in tabular form (see Table 3 below). The first three columns are of primary interest. "Event Date" indicates the date of the event and "Event Type" is self-explanatory. In our simple IRS example the following events occur:

- AD0: Analysis date
- IED: Initial exchange date, the date of the first principal cash flow exchange. In a simple IRS two offsetting cash flows happen at the same point in time, netting to zero
- IP: Interest payment. The last column "Level" indicates the leg of the swap to which the payment belongs.
- RR: Rate reset. This points to the forward rate of the variable side. The implied rate is shown in the column "Rate"
- MD: Maturity date, date of final principal cash flow exchange. In a simple IRS two offsetting cash flows happen at the same point in time, netting to zero

Table 3. Contract events in tabular form

Event Date	Event Type	Event Value	Notional	Rate	IPAC	MVF	NPV	Level
2012-12-31T00	AD0	0.00	0.00	0.0000	0.00	0.00	0.00	P
2013-01-01T00	IED	-1000.00	1000.00	0.0500	0.00	1128.85	1128.85	C1
2013-01-01T00	IED	1000.00	-1000.00	0.0300	0.00	-1066.01	-1066.01	C2
2013-07-01T00	IP	25.00	1000.00	0.0500	0.00	1105.52	1105.52	C1
2013-07-01T00	IP	-14.88	-1000.00	0.0300	0.00	-1052.70	-1052.70	C2
2014-01-01T00	IP	25.00	1000.00	0.0500	0.00	1082.65	1082.65	C1
2014-01-01T00	IP	-15.12	-1000.00	0.0300	0.00	-1039.61	-1039.61	C2
2014-01-01T00	RR	0.03	-1000.00	0.0258	0.00	-1039.61	-1039.61	C2
2014-07-01T00	IP	25.00	1000.00	0.0500	0.00	1060.49	1060.49	C1
2014-07-01T00	IP	-12.79	-1000.00	0.0258	0.00	-1029.55	-1029.55	C2
2015-01-01T00	IP	25.00	1000.00	0.0500	0.00	1038.87	1038.87	C1
2015-01-01T00	IP	-13.01	-1000.00	0.0258	0.00	-1019.82	-1019.82	C2
2015-01-01T00	RR	0.03	-1000.00	0.0308	0.00	-1019.82	-1019.82	C2
2015-07-01T00	IP	25.00	1000.00	0.0500	0.00	1019.02	1019.02	C1
2015-07-01T00	IP	-15.27	-1000.00	0.0308	0.00	-1009.60	-1009.60	C2
2015-12-31T24	IP	25.00	1000.00	0.0500	0.00	0.00	0.00	C1
2015-12-31T24	MD	1000.00	0.00	0.0500	0.00	0.00	0.00	C1
2015-12-31T24	IP	-15.53	-1000.00	0.0308	0.00	0.00	0.00	C2
2015-12-31T24	MD	-1000.00	0.00	0.0308	0.00	0.00	0.00	C2

More complex instruments have more event types.

With respect to Figure 1 and Figure 2, these events represent the third box from the left. From these events it is possible to derive any financial analytics of interest, which are represented by the box on the far right in Figure 1 and Figure 2. It is possible, for example, to establish the value of the IRS at any point in time under any accounting standard. If mark-to- market is used as the accounting standard, discount factors derived from current market conditions must be used. It is also possible to calculate income and, of course, cash flows, which are directly visible in the table. Sensitivity can be derived from fair values. Risk measures combine risk factor shocks or distributions with changes in either value, income or liquidity. Since all types of analysis derive from the same underlying mechanisms, consistency between all of the results is a natural consequence.