To Appear in *Artificial Intelligence in the Capital Markets: State-of-the-Art Applications for Institutional Investors, Bankers and Traders*, Probus Publishing, Chicago 1995. pp. 315-326.

Market Analysis for Risk Management and Regulation: An Artificial Intelligence Approach

Roy S. Freedman & Joseph Mathai

Abstract

The key points in this chapter are:

1. Purely statistical market analysis techniques for determining risk are incomplete because, in the real world, risk is effected by non-statistical phenomena such as news. Quantitative models can ignore such "shocks," but analysts still need to explain their models in the face of such shocks.

2. Risk must include the effects of all relationships. This is only partially accounted by models using correlation statistics.

3. Market analysis techniques used for risk management have much in common with the techniques used in market surveillance for regulation.

In this paper, we provide a framework that can be applied either to risk management for trading or to the regulation of markets. Our framework lets analysts define patterns of market behavior and detect new or hidden relationships between subjects in order to evaluate risk and discover information flows.

Introduction

From the information technology perspective, Wall Street represents a complex and data intensive environment. This environment is discussed from an AI perspective in (Freedman 1991). Essentially, *products* are *traded* through different types of *orders* by *market participants*, who follow *market rules* and *comply* with *regulatory structures*. Market participants evaluate products and analyze news to determine when to place orders. The regulatory structures monitor news and market activity to determine when participants are not in compliance with market rules. Figure 1 shows this environment from an object-oriented perspective.

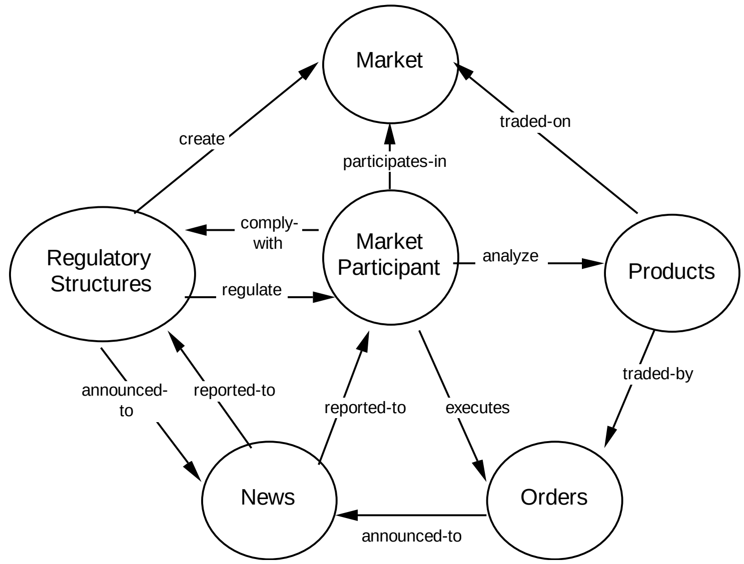

Figure 1. Object-Oriented Representation of Wall Street

A market is considered to be fair when all participants have access to the same information about the market and the products. When all participants have the same information, the value of a product will be determined by its supply and demand. In the real Wall Street environment, there are deviations between the true price and the market price because all participants *do not* have the same information. These deviations are further exaggerated by rumors, "good news," "bad news," and panics. The true price may also be viewed as a consequence of the efficient market hypothesis: it *is* the price observed when all participants act rationally and have perfect information. This *can* be true if all traders have access to identical telecommunication and information systems, and all information obtainable is known to all participants. In this rational and perfect information scenario, there is no profitable way to speculate *and returns will be proportional to the market risk in the trading strategy*.

Traders try to discover profitable opportunities. The discovery of these opportunities can take the form of a quantitative statistical price forecast, an analyst earnings estimate, or an econometric industry model. Note that these forecasts and estimates reflect *beliefs* about a company: they reflect a *model* of the *truth* about a company. Consequently, beliefs are *risky*: forecasts and estimates can be wrong. Trading management needs to control risk to assure consistent profits.

Advance true knowledge *not made public* about a new company product, a surprising earnings report, or a corporate takeover gives the possessor of the knowledge tremendous advantages in the markets where the securities of the company are traded. This advantage represents a profitable opportunity that is *riskless*. This advantage is sometimes used for personal gain or for market manipulation.

Riskless trading advantages that result from acting on information acquired from *non-public* information sources is a concern to participants in markets that are supposed to operate fairly and honestly (Lynch & Mathews 1989). The use of such proprietary or non-public "inside" information for financial gain – *insider trading* – is not legal in the U.S. financial markets. The Insider Trading and Securities Fraud Enforcement Act of 1988 (H.R. 5133), mandates the Securities and Exchange Commission (SEC) with the authority to impose civil penalties on any person purchasing or selling securities through the facilities of a national securities exchange “while in possession of material, non-public information,” or “by communicating such information in connection with a transaction.” According to the Insider Trading Proscriptions Act of 1987 (S.1380), such wrongful use of this information impairs the “fairness, honesty, and integrity of the Nation's securities markets.” The phrase “material, non-public information” is generally interpreted to refer to knowledge that would be important to an investor in making a decision to buy or sell a security that is not available to the general public.

Regulatory organizations try to prevent illegal information advantage in the market so that trading remains fair and orderly. Thus, as Figure 1 indicates, a market can be considered to be defined by its regulatory structure. Regulatory organizations monitor trading and take action when they determine that the market is not fair or orderly.

When a regulatory organization determines that unfair information advantage exists in its market, actions can be taken that include the temporary halting of trading in the market or the suspension of a market participant. Regulatory organizations can also levy financial fines and penalties for rule violations. Here are a few examples:

> In 1991, a former executive vice president and chief financial officer at Crazy Eddie Inc. was charged with insider trading and participating in a multi-million dollar fraud scheme. He was ordered to pay $160,000 in fines.
>
> In 1990, a former money manager at Drexel Burnham agreed to pay $8 million to settle an SEC lawsuit alleging insider trading and fraud.
>
> In 1989, Drexel Burnham agreed to a $300 million fine to the Justice Department and established a $350 million fund for civil suits resulting from their inability to properly monitor the effectiveness of their insider trading compliance procedures.

A comparison of the risk management characteristics of trading versus regulation that shows that they are in fact dual aspects of each other with respect to a trading event, is shown in Figure 2.

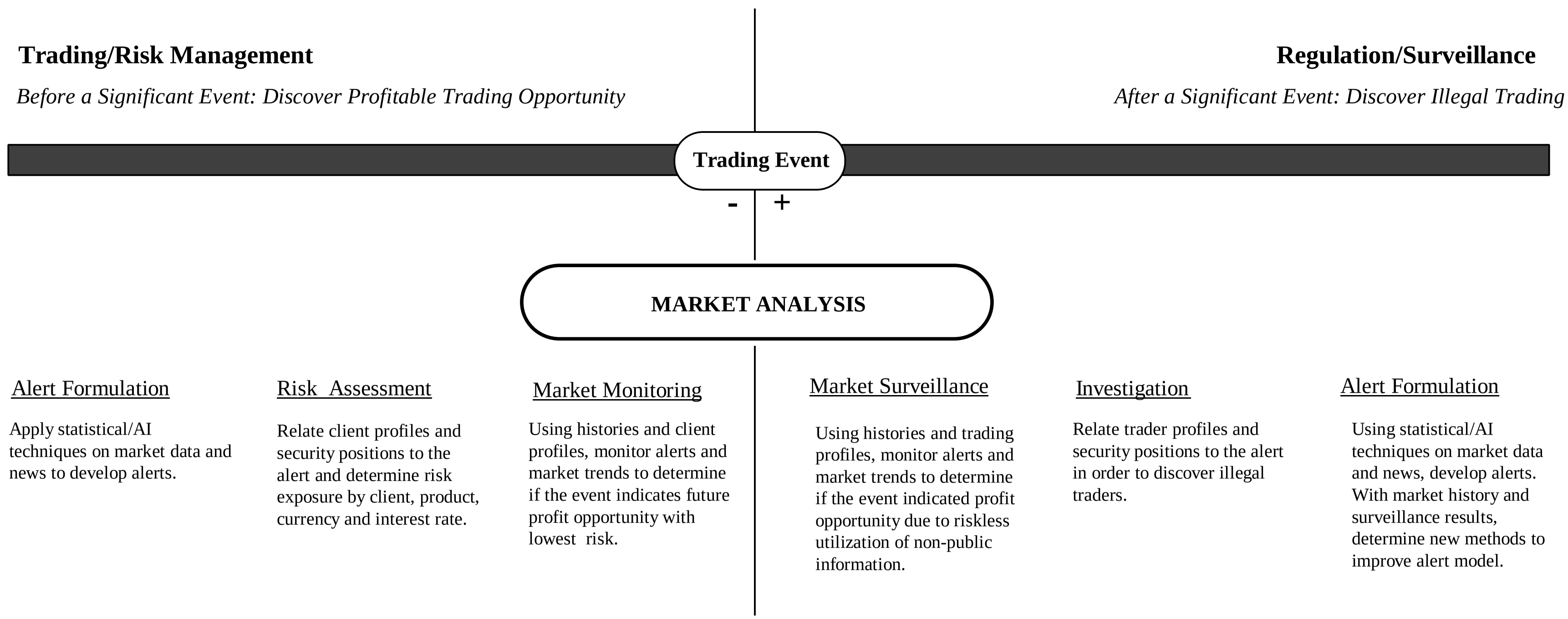

Figure 2. Trading vs. Regulation: An Inverse Relationship

The risk management issues may be summed up as follows:

From the Trading Perspective. From research or a model, identify trades that the trader believes are potentially profitable, with a risk profile believed acceptable to the investor. The trader or investor cannot know the return and risk profiles with certainty because past performance is no guarantee of future performance.

From the Compliance/Regulatory Perspective. From research or a model, identify trades that the compliance analyst believes are possibly illegal: trades that probably relied on true knowledge that was not publicly available. The return and risk profiles of potentially illegal trades are known (the ideal ones are profitable and riskless) — it is the identity of the actual trades that must be discovered.

Market analysis techniques used for regulation are almost identical to the market analysis techniques used for trading. They both rely on identifying *unusual* market behavior in order to indicate abuses in the market place. These techniques range from using simple changes in prices and volume in a stock (or finding an unusually large trading presence by a broker-dealer), to deriving statistical measures of unusual market patterns using cluster analysis using economic factors. However, these techniques are at best alerting systems for market aberrations – they do little in identifying trading on inside information or true riskless arbitrage opportunities.

Tips help in market regulation, *viz.* the Caracas tip on a Merrill Lynch broker that led to the unraveling of the 1986 Levine-Boesky-Milken case (Stewart 1991). None of the defendants in this insider trading scandal were investigated by the exchanges where they completed their trades. Nonetheless, once the scandal became public, several surveillance organizations set up task forces to scrutinize their internal procedures and techniques to determine why they missed all the culprits in this case. What is required is a methodology to detect insider trading even in the absence of tips.

Automating Market Analysis

In general, automated data analysis techniques (sometimes called "dataveillance" (Clarke 1988) depend on data that identifies subjects and their relationships. These techniques have been used for a number of years to draw an analyst's attention to a particular subject or group of subject that are “unusual.” In automated analysis for either regulation or trading, cases are typically organized into data collections that contain computer printouts, results of database queries, notes, and intermediate lists.

Traditional dataveillance procedures rely on matching, merging, sorting, and statistical operations available in many database management systems. Here, evidence for or against unusual patterns is gathered as a result of a specific set of operations.

To be effective, dataveillance requires database access from a variety of public and proprietary sources. Many of these databases are not normalized, and are corrupted by various types of "noise." Regulatory data collected for market surveillance typically has spelling errors, spelling variants, abbreviations, and other context dependencies that result in high proportions of bad matches or undetected matches. Indeed, (Clarke 1988) quotes a study in which half the matches in a surveillance query were erroneous. Determining whether this corruption is due to aberrant activity or from erroneous data is an important aspect of the analysis. One solution to this "dirty data" problem is to use knowledge-based parsing techniques (Frail & Freedman 1991a, 1991b). A parsing algorithm determines the meaning of a sentence in a particular language by inducing and identifying the fundamental language structures. For example, a computer can be programmed to "understand" natural language if it knows about noun phrases, articles, and verbs, and how these structures fit together. In the "dirty data " problem, it turns out that a good part of market data can be considered to be a language in its own right, and that rules associated with these "fragmented language domains" can be derived. This implies that many spelling errors, spelling variants, abbreviations, and other context dependencies can be recognized and corrected. From the trading perspective, this "dirty data" problem is seen in building automated systems that evaluate the impact of news.

The motivation of trading is to achieve profitable outcomes at *lowest* risk. The motivation of insider trading is to achieve profitable outcomes at *no* risk, based on the use of non-public knowledge. The fundamental problem in market regulation and compliance — the discovery of the particular *relationships* between traders and the actual possessors of this proprietary knowledge — can be modeled as an information flow between subjects, where the subjects denote account holders, brokers, dealers, registered representatives, and so on. Similarly, the fundamental problem in risk management — the discovery of the particular *relationships* between assets and their hedging instruments — can be modeled as an information flow between subjects, where the subjects denote assets and their derivative securities that are either already held or are available in the market. Difficulties arise because in general, (i) the unusual subjects are not known — they must be discovered or inferred from the data; (ii) the definition of "unusual pattern of behavior" is subjective and possibly changes with every analysis and over time; and (iii) the quantity of the data in an analysis is overwhelming.

Artificial Intelligence (AI) models are designed to emphasize the representation and integration of qualitative (subjective and judgmental) information within a decision-making environment. Our framework is based on AI models for a subject calculus and relationship calculus where —

> the first models the accumulation of evidence for or against a particular hypothesis for individual subjects;
>
> the second models the certainty of relationships between subjects.

In the following section, we show how to formalize the concepts of a subject calculus and relationship calculus in terms of an uncertainty calculus using probabilistic logic.

The Subject and Relationship Calculus

We define a subject calculus and relationship calculus in terms of an extension of predicate logic with an uncertainty calculus. An uncertainty calculus attaches a belief qualifier to any predicate. These belief quantifiers are typically between 0 and 1, with 1 denoting certainty and 0 denoting impossibility. If these beliefs are consistent with statistics and probability theory, then the resultant structure is termed probabilistic logic; otherwise, the resultant structure is termed fuzzy logic. See Nilsson 1986, Guggenheimer and Freedman 1987, and Guggenheimer 1990 for more information about probabilistic logic.

We use a Prolog notation for predicate logic. Let X denote a subject, u(X) denote the predicate "X is unusual", and e1(X), ..., en(X) denote evidence predicates that describe the behavior of the subject X in the market. The subject calculus is a first order rule that accumulates the certainty of evidence for X.

Let p be an uncertainty calculus — a probability function that is induced from a set of propositions about X that is consistent with probabilistic logic, so that the evidence e1(X), ..., en(X) is qualified by p(e1(X)),..., p(en(X)). The simplest subject calculus rule takes the form

```
clause(u(X), R) :-  clause(e1(X), p(e1(X))..., clause(en(X), p(en(X))),
                    R = f(p(e1(X)),...,  p(en(X))).
```

Here, the subject calculus is realized by the function *f* (a conjunctive uncertainty calculus) that is chosen to be consistent with probabilistic logic. To illustrate, consider the case n=2. If we observe *e1(X)* and *e2(X)* with certainty *p(e1(X))* and *p(e2(X))* and if the rule *u(X) if e1(X) and e2(X)* has certainty *R*, then *u(X)* is true with certainty *f(p(e1(x)), p(e2(x)))*. Note that here the predicate for "unusual" can be precisely defined by procedures for any kind of market analysis. To complete the subject calculus framework, uncertainty calculi for disjunction and negation (as well as conjunction) need to be specified as well.

Note that the predicate "X is unusual" is an indirect measure of risk. From the trading perspective, this can be seen if we a construct a simple rule, "X is profitable with risk *R* if X is unusual with certainty *R*." Defining and measuring risk in trading strategies becomes more complicated when non-stationarity and non-normality of the evidences used to develop these rules are taken into account.

One problem with the traditional analytic framework for evaluating risk is that it is difficult to accommodate *subjective* risk classifications and yet be consistent across risk classes and product categories (see Mark 1993). Most risk measures are purely quantitative: they depend on a statistical model and on statistical parameter estimation.

Another problem is in mixing and evaluating dissimilar quantitative and qualitative risk measures. For example, an unusual change in the statistical price pattern of a financial asset represents a significant change in the risk profile of the asset when the change can not be explained by any news in the market: here, the absence of a relationship is significant. From the regulatory perspective, someone trading in a statistically unusual way does not indicate anything in itself; however, it becomes significant if this trader is also a corporate insider. Here, the presence of a relationship is significant.

These problems are addressed by the relationship calculus. An example of a simple rule in the relationship calculus is the recursive rule that "X is unusual if X is related to Y and Y is unusual." This rule takes the form

```
clause(u(X), R) :-  related_to(X, Y, s), clause(u(Y), y), clause(u(X), x),
                                R = g(x,y, s).
```

Here, the relationship calculus is realized by the function *g* (a conjunctive uncertainty calculus) that is chosen to be consistent with probabilistic logic. The predicate *related_to(X,Y,s)* defines the strength *s* of the relationship between X and Y. This relationship must also be discovered by the evidence. In trading, building the predicate *related_to(X,Y,s)* can be as simple as constructing a table of industry affiliations or statistical correlations; here, this probabilistic rule models the knowledge that companies in the same industry move together. Our framework also allows analysts to associate news events to subjects by defining such explicit relationships, and by having the model determine implicit relationships using logical inference.

Further complications can be introduced by requiring both the the subject and relationship calculus to be dynamic. A dynamic uncertainty calculus (*viz*., the uncertainty functions for negation, conjunction, and disjunction) can change across different types of market analyses.

Reasoning with incomplete information is a paradox. How can we build a model that uses incomplete information? If we know what the information is, then we incorporate these factors in the model, and collect the data; if we do not know what the information is, then how could we incorporate the factors into a model? Again, such information is typical of judgments and other non-quantitative information. Excluding these so-called "exogenous factors" — those factors completely determined outside a quantitative model — may make mathematical sense but limits the analysis. Even economists "tune" their econometric models to bring them "in line" with their subjective opinions.

Note that the issue of reasoning with incomplete information is also addressed in many AI models. In our framework, this is addressed by requiring that the subject calculus and relationship calculus be dynamic.

Implementation Considerations

We have outlined a framework for specifying probabilistic rules in trading risk management and regulation. These procedures can be defined and implemented as probabilistic rules in Horn clauses or other rule-based languages. However, to be successful, the implementation must be able to work when confronted with large quantities of data.

One approach around this difficulty is to build a *deductive database*. Here, a database management system is loosely coupled to a rule-based interpreter. The database queries are incorporated in a set of stored procedures which are invoked by the interpreter. This approach is similar to that discussed in (Imielinski 1987). However, it is difficult to implement recursive rules and it is not clear how to couple the deductive database consistently to the subject and relationship calculus.

Another approach that we recommend is the *inductive database* approach. Here, the subject and relationship calculus are more closely coupled — they are implemented directly within a database management system. Here, the functionality of the subject and relationship calculus is modeled within the relational calculus constructs of the database.

In some sense, reasoning with an inductive database is similar to case-based reasoning (Slade 1991). A case-based reasoning approach looks for and ranks similarities between cases. For example, a case-based reasoning system for risk management would rank trades with respect to a case database of unusual trades. A case-based expert system has been built for market surveillance at the Toronto Stock Exchange (Buta and Barletta 1991) for alert monitoring. Our inductive database approach is analogous to the case-based approach, as long as case similarities are measured in terms of the subject calculus. However, our approach is much more powerful since it supports linkages between subjects in terms of the relationship calculus.

Conclusion

Our approach, based on the concepts of a subject calculus and relationship calculus, provides a rigorous foundation for sophisticated regulatory and trading risk management systems. It leads to a simple and efficient representation of the subjective and intuitive notions of accumulating evidence for detecting an unusual event, and for associating unusual behavior of subjects. Our approach handles uncertain and incomplete data, and does not require subjects to be identified *a priori*. The framework directly supports the quick prioritization of tasks, by enabling simple and efficient sorting representations. This framework also demonstrates scalability, as demonstrated by:

Syntactic extensibility. Knowledge can be added to improve accuracy. New evidence about subjects can be dynamically defined and added by analysts. Linkages between subjects can be identified from any source and added to the analysis at anytime. The subject and relationship calculus will automatically revise rankings with the new information. For example, from the trading perspective, if a new database of corporate suppliers is discovered, it can be added to improve the *related_to* predicate.

Semantic extensibility. Different applications can be performed within the same framework by changing the evidence functions. Rules for assigning certainty can be changed at any time. For example, from the regulatory perspective, customizable evidence functions can be specified for market manipulation, front-running, churning, and other market abuses.

The common risk management problem between trading and regulation can be summarized as follows: both the trading analyst and the regulatory analyst are looking for a riskless trade. From the perspective of the trader, a riskless trade represents a profit opportunity. However, this opportunity is illegitimate if is derived from the use of proprietary information or illegal trading strategies. On the other hand, the regulatory analyst needs to identify these same riskless trades that result from the misuse of information or the application of illegal strategies.

References

Buta, P. , Barletta, R., 1991. Case-Based Reasoning for Market Surveillance. *Proceedings of the First International Conference on Artificial Intelligence Applications on Wall Street, 116-121*. IEEE Computer Society Press.

Clarke, Roger A., 1988. Information Technology and Dataveillance. *Communications of the Association for Computing Machinery*, 31: 498-512.

Frail, Robert P., Freedman, Roy S., 1991a. Name Classification in Noisy Domains. *Proceedings of the First International Conference on Artificial Intelligence Applications on Wall Street*, 16-22, IEEE Computer Society Press.

Frail, R., Freedman, R.S., 1991b. Text Classification in Fragmented Sublanguage Domains. *Proceedings of the Seventh IEEE Conference on Artificial Intelligence Applications*, 33-36, IEEE Computer Society Press.

Freedman, R.S., 1991. AI on Wall Street. *IEEE Expert*, Vol. 6, No. 2 : 2-7.

Guggenheimer, H., Freedman, R.S., 1987. Foundations of Probabilistic Logic. In *Proceedings of the Sixth International Conference on Artificial Intelligence*, 939-941. Menlo Park, California:International Joint Conference on Artificial Intelligence, Inc.

Guggenheimer, H.W., 1990. Limit Theorems for Dempster Composition in Probabilistic Logic. *Computers and Artificial Intelligence* 9 (4), 313-324.

Imielinski, Tomasz, 1987. Intelligent Query Answering in Rule-Based Systems. In *Foundations of Deductive Databases and Logic Programming* (Minker, J., ed.) 275-312. Morgan Kaufmann Publishers, Inc., Los Altos, California.

Lynch, G., Mathews, A., 1989. *Insider Trading and Securities Fraud Enforcement Act of 1988*. Clifton, New Jersey: Prentice-Hall Law & Business.

Nilsson, N., 1986. Probabilistic Logic. *Artificial Intelligence*, *28* : 71-87.

Mark, Robert, 1993. New Directions in Risk Management. Presentation given at the *Second International Conference on Artificial Intelligence Applications on Wall Street*.

Slade, S., 1991. Case-Based Reasoning for Financial Decision Making. *Proceedings of the First International Conference on Artificial Intelligence Applications on Wall Street, 232-237*. IEEE Computer Society Press.

Stewart, James, 1991. *Den of Thieves*. Simon & Schuster.

Author Biosketch

Roy S. Freedman

Roy S. Freedman has designed asset allocation, forecasting, and market surveillance systems that are in routine use at the world's largest stock exchanges, investment banks, and trading firms. He has published extensively on quantitative modeling, artificial intelligence, and software engineering. One of Mr. Freedman's strongest contributions was in the design and deployment of *ICASS* — the New York Stock Exchange market surveillance system.

Joseph Mathai

Joseph Mathai is currently in charge of Equity Systems at Fidelity Investments. He was the Director of Research at Instinet Corporation, where he developed the order management system for executing market neutral program strategies on Instinet and the New York Stock Exchange. Previously he was the Managing Director for Market Surveillance Systems at the New York Stock Exchange, where he led the development of the Integrated Computer Assisted Surveillance System for insider trading investigations and specialist surveillance.