

Modern data analysis tools in personal financial services: a quantitative revolution?

David J. Hand

Imperial College London
d.j.hand@imperial.ac.uk

Abstract. I argue that dramatic advances in methods and tools of data analysis are having an equally dramatic effect on how decisions are made in the personal financial services sector. A series of examples are given, and the sorts of tools needed are described.

1 Introduction

The mathematical revolution in the financial market sector is now well established. Over the past several decades it has become more and more advanced, employing legions of researchers, predominantly mathematicians and physicists, who might otherwise have gone into the academic world, and who have knowledge and skills of the types of mathematics needed for the financial sector.

This revolution is rather unusual because it represents a direct link between advanced mathematics and practical application: most applications of mathematics involve fairly straightforward and well established mathematics, even mathematics which is centuries old. In contrast, the mathematics of financial markets, of options, futures, swaps, and other complex products is cutting edge. Several journals catering for these advances have sprung up, and many books, ranging from the introductory to the advanced have been written about this new area of application of mathematics.

Broadly speaking, this area is concerned with managing risk at a macroscopic level. The risk that a corporation's share price will drop, that cold weather will damage the harvest, that a hurricane will hit the Eastern seaboard of the USA, and so on. Of course, now that tools which essentially allow one to bet on the future outcome of uncertain events have been developed they have begun to be applied to a wide range of outcomes of other events. They are, after all, essentially gambling instruments.

What has attracted less media interest, however, are the parallel changes which are beginning to impact the personal financial services sector. Perhaps it is because most of these are concerned with the micro level, the level of individual people, that they do not attract such interest. Perhaps it is because the revolution is still at an early stage, and has not built up the momentum of the derivatives market. Perhaps it is because there have been no vastly expensive disasters, such as the LTCM fiasco, which have captured the public attention. Whatever the cause, the

changes which are permeating the retail sector generally only occasionally make the newspapers. One will find the occasional article about credit scoring instruments, describing how financial institutions make their decisions on whom they should lend to, the odd article about someone who has fallen deeper and deeper into debt, or articles in the personal financial pages advising people on how to switch mortgage suppliers or credit cards. Often, these articles are not complimentary to the financial services sector. In general, however, the personal sector has not attracted as much media attention, or inspired popular interest (and awe?), as the financial market sector.

I speculate that one reason is because of the concentration of the large sums in areas such as financial markets and hedge funds. The tens, hundreds, or thousands of millions occur in one place. In contrast, although the sums are just as large in the retail sector (for example, the total amount lent to individual customers exceeds the total amount lent to corporations - perhaps not so surprising when one considers mortgages and car finance), they arise as a large number of relatively small amounts. Of course, it is this very 'large number' which makes the area so perfectly suited to statistical, machine learning, and data mining approaches. Similarly, in the investment banking sector, because large sums are concentrated in particular places, the potential for disaster is more obvious: if a bank, which characteristically manipulates many millions of euros, collapses, those many millions are immediately there to hit the headlines. In contrast, for the consumer market to lead to disaster something must simultaneously hit many small borrowers, which accumulate to give overall large sums. The result, perhaps equally large, is not so immediately obvious. Of course, this does not mean that the media are not aware of the possibility - and are not averse to the occasional scare story: 'One million credit card holders are in serious danger of having their debt spiral out of control,' (MORI Market Dynamics, 2003); 'More than 300,000 homebuyers could be plunged into negative equity if property prices fall as fast as experts predict,' (Mail on Sunday, 2003).

The quantitative tools in the financial markets sector are of a particular kind, typically involving such innovations as stochastic calculus and the Black-Scholes model and its extensions. The quantitative tools used in the retail sector are of a very different kind. They owe more to statistics and data analysis than to mathematics.

In one sense this difference is fundamental to the difference between the two areas. The mathematical finance sector is creating normative tools, telling bankers and others how they should behave to manipulate and control their risks. The retail sector is creating descriptive tools, describing how customers actually behave with their credit cards, loans, or mortgages, and then of course using these tools to guide decision making. The mathematical finance sector is mapping from theory to practice. The personal finance sector is mapping from data to theory (and thence to practice). This distinction, even opposition, mirrors the distinction between probability and statistics: the former starts with the model and describes what one will observe (two sixes will come up around 1/36th of the time), while the latter starts with the observations and infers what led to them (that each face of the dice had an equal probability of occurring).

2 Why now?

Why am I suggesting that we are at the beginning of a quantitative revolution in the personal financial services sector? After all, scorecards have been around for many years (one can trace their origins back to the early part of the twentieth century). Also, if there is such a revolution, what is causing it?

The answer to the first question is simply the evidence all around us. The complexity of the personal financial services space has increased dramatically in the past decade or two. New financial products, and variants of old ones, are being produced all the time. Telephone banking is by now a familiar example, internet purchasing has become an everyday activity (although there is still some hesitation because of the reports of crime). Internet banking is growing. Paper cheques will soon become obsolete. Other new kinds of products include current account mortgages, which integrate the two financial products, customer specific interest rates (rather than a single universal rate, applied regardless of the perceived value or risk of the customer), sweeper accounts, debt consolidation operations, and a gradual trend into new market niches. This last is particularly interesting: I have observed clear signs of quantitative scorecard technology moving downstream, to the so-called subprime sector, the high risk customers. This can only be a good move, both for the financial institutions and for the individual customers.

Technological advances, and advances in the range and type of financial products, have been accompanied by changes in the relationship between the financial institutions and the customers. Customers have become more impatient. They want a loan decision now: if you take a month to decide, you will have lost them as they go elsewhere.

In a sense, there is really a single underling cause of these changes: progress in computer technology. And the result is very much a leapfrog act. New data manipulation technology leads to the possibility of new financial products, which leads to a need for new ways to handle the data describing customers. Modern customer relationship management, and lifetime value calculations are examples of this. Data mining, in particular, is facilitating progress in the former area, in ways which would have been quite inconceivable not many years ago. Huge amounts of data on customer characteristics and behaviour is captured and retained, and can later be probed in detail to extract useful and perhaps previously unsuspected features.

These changes and advances lead to their own problems, requiring novel, typically quantitative and data analytic, solutions. An obvious and by now familiar example is cardholder-not-present credit card fraud. This arises when purchases are made over the telephone or internet by someone who is using stolen credit card details, and therefore is not physically present to show the card.

Another cause of change (though, of course, they are all related) is legislation. This is constantly evolving, perhaps usually to protect the customer. The Basle II requirements, in particular, insist on quantitative innovations.

3 Some examples

Whenever a decision based on evidence or information must be made, then there is a scope for quantitative information and formal decision-making procedures. Just a few of these situations, as they occur in the retail banking sector, are described in this section.

3.1 The classic problem: application default scoring

The oldest established usage of statistical methods in the retail banking sector is probably that of predicting which applicants are bad risks and which are good risks: who is likely to default on a loan or other product, and who not? Statistical methods have been used to build such predictive models in the retail banking sector for at least 30 years, using standard tools such as linear regression, linear discriminant analysis, and logistic regression. The last of these has become something of an industry standard, though the way it is used in the industry gives it greater flexibility than one might expect. In particular, all variables are categorised, and the categories coded with indicator variables. This is equivalent to using a nonlinear transformation on the raw variables, yielding a model equivalent to a generalised additive model.

Inevitably, since the problem here is the standard one of supervised classification, many of the large number of other tools which have been developed for supervised classification have also been applied in this context. These include classification trees, neural networks, support vector machines, and nearest neighbour methods. Despite their claimed superiority of performance (though Hand, 2005, has reservations about this), such methods have not been widely adopted. One reason must be that, in front end roles (those in which the decision is communicated to the customer) there is a premium on interpretability. Neural networks may yield slightly better classifications, but they are hardly transparent. In contrast, in back end roles, in which the decisions do not have to be 'justified' to the customer (such as fraud detection), more sophisticated and less interpretable methods can be used.

Although application scoring is well-established as a role for quantitative methods in the retail finance sector, many open questions remain. A key one is how best to handle the design sample selection which results from the fact that the set of customers granted a loan is not a random sample from the entire population of applicants. This means that care has to be used in constructing a classifier for new cases on the base of the observed outcomes of those given a loan. This sample selection problem is known as 'reject inference' in the industry, and some poor and potentially misleading methods have been adopted to tackle it.

3.2 Behavioural scoring

Application scoring is a one-time decision for any given customer: the accept/reject decision is made and that is that. Behavioural scoring, a more recent development, involves following up the accepted customers so that interventions can be made if problems appear to be arising. Such methods have a role in loan situations, but are perhaps more important in revolving credit operations, such as credit cards.

No special class of data analytic tools dominates behavioural scoring, but obviously stochastic process models, such as Markov chains, are useful. With these one can explore the probabilities of customers moving between different states. Examples of such models are given in Frydman et al (1985) and Till and Hand (2003).

3.3 Survival analysis

It is the essence of retail finance data that it is dynamic: things progress over time. Behavioural scoring captures this, in a way that application scoring does not, but behavioural scoring is typically used to answer questions such as 'has this happened?' Another class of questions is 'when is this likely to happen?' The survival analysis class of models are appropriate for answering such questions - hence the title of the paper by Thomas and others (Thomas et al, 1999): Not if, but when will borrowers default? Such models are also beginning to find use in many other applications. For example, in Hand and Kelly (2001), they are used to try to overcome the problem that in a loan situation a good outcome (i.e. not default) cannot be definitely known until the loan term has passed. This can be a long time (e.g. 25 years for a mortgage) and it is often not realistic to wait that long before one can build a model for new customers (the mortgage example clearly illustrates this). I expect survival analysis models to find many more retail finance applications in the future.

3.4 Models based on relationships between multiple variables

People, of course, are not unidimensional entities, and many aspects of their behaviour may be of interest as far as personal finance goes. Various classes of models have been applied to try to capture relationships between multiple variables in this context, including Bayesian belief networks and LISREL models. Examples are given in Hand et al (1997), Stanghellini et al (1999), and Sewart and Whittaker (1998). Since customers often have more than one financial product, such approaches have an obvious application in combining everything into a single model. However, they are more generally useful than this, since they allow one to handle partial and incomplete data (a common occurrence with retail banking data) with ease. Whether such models yield more accurate prediction than models specifically built to predict a single outcome variable is a different question.

Before formal methods of making credit granting decisions were developed, decisions were sometimes said to hinge around the three Cs. These stood for character, collateral, and capacity, and were thought to capture different aspects of the quality of a loan. From a modern perspective, these might be regarded as latent variables - and that raises the possibility that more sophisticated modern latent variable methods could be used. Hand and Crowder (2005) describe just such a model, in which the observed customer characteristics are separated into two classes: the primary variables, which are regarded as possible determinants of 'customer quality', and behavioural variables, which are consequences of customer quality. It is not always straightforward to decide into which class a variable should be put, but to give two obvious examples, age would be a primary variable (it is certainly not influenced by how trustworthy a customer is, for example) while a binary indicator of whether someone has or has not previously defaulted on a loan would be a behavioural variable. Intermediate between these two classes is a single unobservable 'customer quality' measure, the value of which can be

estimated from the observed values of the other variables.

Yet another class of models of this general form is described in Hand et al (2001). Often, the 'good' and 'bad' classes are described in terms of quite complex interrelationships between *simple variables (e.g. how long an account has been open, whether it is active, how large an overdraft has been, etc.). This suggests two alternative strategies for predicting class membership. One can go directly for it, predicting the class in the standard manner of supervised classification, or one can predict each of the variables used in the class definition separately (or even jointly) and then combine the predictions deterministically to yield a predicted class. These sorts of issues are explored further in Benton (2002) and Benton and Hand (2002).

3.5 Data mining

Perhaps one of the most exciting, and certainly one of the most recent, areas of development in the retail financial services sector is the possibility of mining the large transaction databases owned by the banks and other financial institutions. Data mining offers the promise of finding previously unsuspected patterns and relationships in these large bodies of data. It goes without saying that it is data storage and computer technology which has permitted the accumulation of such databases, and the possibility of searching them for patterns.

The scope here is effectively unlimited, with the opportunities for novel developments in customer relationship management particularly great. On the other hand, one must beware of the hype which has accompanied data mining. While it is almost certainly true that there is valuable information in the large transaction databases, it is even more certainly true that finding this information will be difficult. The situation is confounded by problems of multiplicity (with large enough data sets unusual configurations are almost certain to arise, purely by chance), of poor quality data (discovering persistent configurations of missing values may not be very useful in determining creditworthiness), and of discovering the obvious (it is not of great interest to discover that the number of married men is almost equal to the number of married women).

Apart from general issues of searching for patterns which could be commercially valuable (e.g. Hand and Blunt, 2001) another important application of data mining methods in this area is in fraud detection. Fraud can be perpetrated in many ways, but each way gives rise to a characteristic behavioural trace (e.g. in the way a credit card is used; in the way money is split up for laundering) and these can be detected using mining techniques. On the other hand, note that retrospective detection is all very well, but what is really needed is prospective detection: ideally, we need to know that fraud is being perpetrated as it happens, and not some months after the event. Fraud detection is particularly challenging because it is an evolutionary area: once one avenue has been blocked off (e.g. the 1970s regulation in the US that all monetary deposits of more than 10,000 must be reported) others are developed.

3.6 Customer value

A relatively new area of work is that of customer value models. This is difficult because, of course, it depends on unknown future developments. Nevertheless, models can be built which take account of economic climate variables, which can themselves be predicted from other,

econometric, models. Customer value models need to take account of the activity and lifetime of the customer, and also of the lifestage. For example, credit card users have a particular usage pattern, beginning with tentative and occasional use, and graduating to more general and frequent use (and then perhaps reverting to a more cautious use, as awareness of fraud grows). Of course, once again the ultimate aim is not merely to model and describe a customer's value, but to use this information to intervene and make decisions about what actions to take.

4 Conclusions

In the above I have described only a few of the many ways in which statistical, machine learning, and data mining models are having an important impact on understanding and decisions in the retail financial services sector. There are many other areas. Other examples of which I have experience include:

- models for stocking ATM machines, so that the probability that they will run out of money is optimally balanced against the cost of borrowing the money to stock them;
- models for deciding whether the cocaine contamination of a stash of banknotes was greater than expected from the general background of contamination;
- market segmentation, dividing customers into financial behaviour classes;
- tools for effective validation of predictive models in this sector. Indeed, in Hand (2005b), I argue that many of the tools which are in very common use are entirely inappropriate, and could be very misleading;
- churn models, to predict who is likely to move to another supplier (e.g. of mortgages);
- customer specific risk models.

These are just a few of the ways in which progress in computer systems and data analytic tools is impacting on the retail financial services sector. There are many, many others. The area is a natural one for inductive models, moving from the data to descriptions and inferences about how people behave - above all, the area is characterised by large numbers. It also provides challenges, and opportunities for the development of novel tools. I think that it is entirely appropriate to describe what we are witnessing as the beginning of a revolution in the application of quantitative methods.

References

- Benton T. (2002) Theoretical and empirical models. PhD thesis, Department of Mathematics, Imperial College, London.
- Benton T.C. and Hand D.J. (2002) Segmentation into predictable classes. *IMA Journal of Management Mathematics*, 13, 245-259.

Frydman H., Kallberg J.G., and Kao D.-L. (1985) Testing the adequacy of Markov chain and mover-stayer models as representations of credit behaviour. *Operations Research*, 33, 1203-1214.

Hand D.J. (2005) Classifier technology and the illusion of progress. Technical Report, Imperial College, Department of Mathematics.

Hand D.J. (2005b) Good practice in retail credit scorecard assessment. To appear in *Journal of the Operational Research Society*.

Hand D.J. and Blunt G. (2001) Prospecting for gems in credit card data. *IMA Journal of Management Mathematics*, 12, 173-200.

Hand D.J. and Crowder M.J. (2005) Measuring customer quality in retail banking. *Statistical Modelling*, 5, 1-14.

Hand D.J. and Kelly M.G. (2001) Lookahead scorecards for new fixed term credit products. *Journal of the Operational Research Society*, 52, 989-996.

Hand D.J., Li H.G., and Adams N.M. (2001) Supervised classification with structured class definitions. *Computational Statistics and Data Analysis*, 36, 209-225.

Hand D.J., McConway K.J., and Stanghellini E. (1997) Graphical models of applicants for credit. *IMA Journal of Mathematics Applied in Business and Industry*, 8, 143-155.

Sewart P. and Whittaker J. (1998) Fitting graphical models to credit scoring data. *IMA Journal of Mathematics Applied in Business and Industry*, 9, 241-266.

Stanghellini E., McConway K.J., and Hand D.J. (1999) A discrete variable chain graph for applicants for bank credit. *Journal of the Royal Statistical Society, Series C, Applied Statistics*, 48, 239-251.

Thomas L.C., Banasik J., and Crook J.N. (1999) Not if, but when will borrowers default? *Journal of the Operational Research Society*, 50, 1185-1190.

Till R.J. and Hand D.J. (2003) Behavioural models of credit card usage. *Journal of Applied Statistics*, 30, 1201-1220.