Cameron Hume

Fixed Income Specialist

Good data, not "Big Data"

Data, and its incorporation into the investment process, is often a core preoccupation for investment managers and there is a lot of excitement and discussion around "Big Data". According to Gartner: "Big Data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation"[1].

Our view is that institutional investment managers ("investors") do not use Big Data; in fact they struggle with the amount and quality of conventional data, and how to incorporate this into the investment process.

We believe that investors need to deal with the issues surrounding large amounts of data, rather than try to bring Big Data solutions to bear on their requirements. Our preoccupation is with solving quality, integrity and accessibility issues, and not those associated with the challenges of Big Data. The solutions we use are intended expressly to avoid the problems we have seen elsewhere.

In our experience the root causes of data related issues are:

1. multiple, and often conflicting, sources of data;
2. proliferation of data definitions and technical standards.
3. use of spreadsheets for analysis;
4. failure to integrate structured analytical data into the investment process;
5. inability to change the assumptions involved in analysis quickly; and
6. trying to solve data quality or integrity problems through reconciliation rather than use of data.

As a specialist fixed income investor, we believe the combination of large volumes of data coupled with sophisticated modelling tools and massive increases in processing power presents a powerful opportunity to address these issues, to the benefit of our clients.

[1] Gartner IT Glossary © 2017 Gartner, Inc

Cameron Hume
Exchange Place 1, 1 Semple Street
Edinburgh EH3 8BL
www.cameronhume.com

CAMERON HUME

Data is not easy to incorporate into the investment process

Investors typically use financial and economic data. Key inputs for the analysis of fixed income instruments are financial data (for example bond, future and swap prices) and the important characteristics of these instruments. Unlike economic data (see below), financial data remains relatively expensive and opaque, and is typically only available through commercial data providers (a model which doesn't appear likely to change any time soon). New categories of financial data are also becoming available, for example information on environmental, social and governance (ESG) factors or the introduction of the Legal Entity Identifier (LEI), which provides a unique standardised reference for issuers.

ESG data is increasingly a requirement for our clients, however is not easy to incorporate into the investment process, particularly in fixed income. Our approach to data makes this possible, sourcing extensive ESG data from the MSCI ESG Research service, which we integrate with investment portfolios in our data management systems so that we can monitor these exposures alongside more traditional fixed income factor exposures. By doing so, we believe that we can establish a basis for measuring compliance with the policies and ESG objectives of our clients.

There are other categories of data that, historically, have been neglected by investors. However, when combined with advances in models and processing power, these provide further opportunities for fixed income investors. This data includes the vast amount of economic data being released by central banks, the IMF and government statistical agencies. The amount, quality and structure of economic data has improved in recent years, with more information on the structure and links between data, helped by the development of standards such as SDMX[2].

These large amounts of financial and economic data need to be incorporated into the investment process. There are several hurdles that need to be overcome to allow this to happen.

1. the data feeds need to be automated, checked and available in a consistent form for use across all investment teams;

2. tools are required to find and chart the data for use as the input for further analysis;

3. models and analysis of the data need to be understood and used by the investment team; and

4. key portfolio attributes such as exposure, risk and performance require appropriate and consistent treatment.

Historically a portfolio manager would produce their own analysis using a spreadsheet. This is error-prone, and difficult to update regularly or share across the investment team. This method is unlikely to cope with the volume or complexity of data that might be required. Open source tools such as R (a statistical programming language) allow more sophisticated analysis on large sets of data and can be combined to produce powerful but flexible models. However it is important these models are understood and used in a consistent way across the investment team.

Packaged analytical tools can be used, but often fail to offer sufficient breadth of analysis, timeliness or competitive advantage and do not cater well for fixed income investors. For example, the analysis and reporting of active positions for equities is usually inappropriate and uninformative in fixed income and is not consistent with the investment process and internal dialogue on active positions.

[2] SDMX (Statistical Data and Metadata eXchange) is an international initiative that aims at standardising the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries.

Flexible data architectures can be scaled very quickly

The challenges for the investor are growing exponentially: today there are vast volumes of data that are potentially relevant to the investment process and the different categories of data have the potential to grow rapidly. For example, we did not envisage two years ago that we would build ESG data into our investment process and portfolio position and compliance monitoring, and in five years' time it is highly likely that we will be dealing with other data types.

Our approach to this is to use proven, currently available technology to build an architecture that is flexible, highly scalable and can be deployed rapidly. We gather and retain data in its native form quickly and effectively. Before we use that data, we then process it, for example, to identify and rationalise conflicting data, or match data from a single issuer. We further separate our data layer from our analytical layer to enable the development of the analytical layer in such a way that it can consume new data as it becomes available.

We have developed a set of tools which provides the investment team with financial, economic and ESG data and the tools to analyse this data. We see this tool set as augmenting the investment process – allowing ideas and common assumptions to be tested, and providing the catalyst for potential investment ideas for further investigation.

The result is that everyone understands the models and the algorithms used, as opposed to these being embedded in the spreadsheets of individual investment managers. Analysis can be altered in investment committee meetings in the event that there is a challenge to a particular set of assumptions. New, more subjective factors like ESG can be easily integrated into the analysis and investment process and we get a single view of data within analysis, order management, and reporting.

The emergence of cloud-based infrastructure has enabled the rapid scaling of computing resource that can be required to cope with increasing types and volumes of data.

Preparing for the future

It is likely that Big Data will become more relevant to institutional investment managers. Investors are considering new types of data that may form a meaningful part of the investment process. With a proper structure and process for dealing with data, the incorporation of Big Data into the investment process offers some exciting possibilities. For example, web-scraping technology can provide real-time estimates of inflation, while satellite imagery can infer crop growth or customer volumes ahead of public statistics.

There is much speculation on the role computers will have in the future, a topic discussed in detail in the "Second Machine Age[3]". The authors make the case that humans must adapt to collaborate with machines, and when that collaboration happens, the end result is stronger. They provide a fascinating anecdote using chess as an example – although it is well known that a computer can beat the best chess players, what is less well known is that a team of humans and computers working together is superior to both. At Cameron Hume we agree that such collaboration in fixed income improves the investment process, augmenting our long term fundamental analysis with data and models.

[3] The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies by Erik Brynjolfsson and Andrew Mcafee.

Contact

Keith Logan
keith.logan@cameronhume.com
T +44 (0) 131 603 6988
M +44 (0) 7860 925 131

Author

Caspar Cook is Investment Analyst
at Cameron Hume Limited