

SEPTEMBER 2015

BLACKROCK®

THE EVOLUTION OF ACTIVE INVESTING FINDING BIG ALPHA IN BIG DATA

MARKETS

INVESTMENT
ACTIONS

PORTFOLIO
DESIGN

RISK
MANAGEMENT

REGULATORY



Summary

A hallmark of today's digital age is the explosion in the quantity of data being produced and, increasingly, being put to work. How is this development reshaping the landscape for asset management? And what do investors need to know about 'big data'?

To help answer these questions we have produced a note summarising the insights discussed by our Scientific Active Equity team in recent webcasts and conferences in New York and San Francisco outlining what the concept of big data is, why it matters and how it can be used to generate alpha.

- ▶ We believe that in order to generate sustained alpha, investors should embrace acquiring, analysing and understanding the fast growing universe of data. Those who are unable to do so run the risk of falling behind in a rapidly changing investment landscape.
- ▶ Given the size of the datasets being created, it is necessary to use new tools and research methodologies, these include: machine learning, natural language processing, scientific data visualisation and distributed computing.
- ▶ These tools can help us to both understand what human indicators, such as analysts and employees, are trying to tell us about individual companies, as well as helping us construct better economic indicators that may offer clues for the outlook of entire industries.
- ▶ There is a fierce battle going on for talent. If investment managers hope to recruit the top computer scientists, statisticians and engineers, they will need to commit to a culture of innovation and empowerment that is less common within the world of finance.
- ▶ As with all investment signals, those generated using big data analysis will generally lose efficacy over time. However, superior data and superior methods of extracting information from that data have provided an advantage in investing, and we think that they always will.
- ▶ We firmly believe that a data-driven, scientifically based and technologically aware research culture can produce sustainable alpha.



RAFFAELE SAVI
Co-Head of Investments,
BlackRock's Scientific
Active Equity
(SAE) Group



JEFF SHEN
Co-Head of Investments,
BlackRock's SAE Group
and Head of Emerging
Markets, BlackRock



BRAD BETTS
Co-Head of Data
Science, BlackRock's
SAE Group



BILL MACCARTNEY
Co-Head of Data
Science, BlackRock's
SAE Group

Finding big alpha in big data

The world is awash in data. We are generating some 2.5 quintillion (2.5×10^{18}) bytes of it every day; IBM estimates that about 90% of the data in the world today has been created in the last two years. So it is no surprise that it seems that every day generates some new promise of how we can use this 'big data' to change the world. Proponents of big data analysis claim that it can do everything from helping companies sell more widgets to extending human life expectancy. However, what do we mean when we talk about big data? And can the concept possibly live up to the hype?

We are not big fans of the phrase 'big data'. Data science, while blander, would be a more accurate description, but the world seems to have settled on big data, so we'll use the terminology here. What really matters to us from an investment perspective is how data can be put to work to generate alpha. In practice, we are interested in all kinds of data, particularly as they relate to human economic behaviour. Most of these data are unstructured; making sense of them requires new approaches, a different set of tools and more powerful computing platforms.

As to the question of hype, with regard to investment management, the short answer is that big data are both overhyped and underappreciated. It is overhyped in the sense that, despite the fears of some and the wishes of others, computers are not going to make human fund managers obsolete anytime soon. Big data analysis is only as effective as the human expertise and wisdom that guides it.

Yet the role of big data is underappreciated because we believe their rise represents a watershed moment in the history of investment management. As the traditional line between passive and active has begun to blur with the rise of factor-based strategies, true alpha has proven increasingly elusive. In response, managers across the spectrum, from equity to fixed income are looking to big data for an edge.

However, that edge will not be easy to gain, or to maintain. We are in the midst of a race to build the tools and acquire the data and the talent that will lead to the next generation of valuable quantitative investing insights, while keeping in mind that today's best insights may become less effective as they are more widely adopted. Long-term success will therefore require not only a sizable commitment now, but also constant innovation.

Despite these challenges, we believe that in order to generate sustained alpha, investors must embrace the task of acquiring, analysing and understanding the ever growing data universe. Those who fail to do so run the risk of falling behind in a rapidly changing investment landscape.

SEEING THE BIGGER PICTURE

Much of the explosion in the quantity of data (illustrated in exhibit 1) is driven by new technologies, notably smart phones, internet searches and social media. Given the size of the datasets being created, sorting useful signals from the vast

EXHIBIT 1: UNSTRUCTURED EXPANSION

Data generated each minute by the following means

Facebook users like **4,166,667** posts

Tinder users swipe **590,278** times

Twitter users tweet **347,222** times

Instagram users like **1,736,111** new photos

Amazon receives **4,310** unique visitors

Apple users download **51,000** apps

Skype users make **110,040** Calls

Vine users play **1,041,666** videos

Netflix subscribers stream **77,160** hours of video

Pinterest users pin **9,722** Images

YouTube users upload **300** hours of new video

Uber passengers take **694** rides

Reddit users cast **18,327** votes

Snapchat users share **284,722** snaps

Buzz feed users view **34,130** videos

Source: DOMO, August 2015

amounts of noise generated has become challenging. This has led to the rise of techniques such as machine learning, natural language processing, scientific data visualisation and distributed computing, which can enable researchers to build a better understanding of extremely large datasets.

Before going through an example of how some of these research applications can be used to make sense of the proliferation of new forms of data, it is worth highlighting that some of the same techniques that can help us comb through the world's searches, tweets and 'likes' can also be put towards analysing the growing pile of analyst reports, SEC filings and key corporate developments that impact security prices.

Both fundamental and quantitative managers have, of course, long looked to analysts as a rich source of information. When it comes to analysing a single report – or any particular collection of data, for that matter – a human can almost certainly bring a more sophisticated judgment than an algorithm can, however humans cannot operate as fast. In a world that now sees an average of 4,000 brokerage reports a day comprising 36,000 pages in 53 languages, advanced text analysis could be seen as a necessity, not a luxury, for investment managers who wish to extract maximum utility from the growing mountain of analyst comments.

The traditional way of incorporating analysts' views into the investment process is for portfolio managers to read a small number of reports that relate to their areas of investment responsibility. Today's big data technology uses multilingual text mining to search the entire universe of analyst reports for new and potentially market-moving information.

In addition, conference calls also often contain valuable forward-looking information, which can now be transcribed in real time and machine-read.

Indeed, any written information that may contain relevant insights, including annual reports, regulatory filings, news articles, books and blog posts all offer opportunities to benefit from this approach.

Constructing an effective system requires a massive investment in a low-latency, distributed and resilient computer network. One that can identify different languages, understand context, discard irrelevant information and pick up on the difference between forward and backward looking statements. The potential benefit from this approach is too great to ignore: it offers the prospect of being able to take into account everything that analysts and other commentators are telling us about the future prospects for tens of thousands of publicly traded securities.

SEARCHING FAR AND WIDE

The advancement in the ability of algorithms to read and understand analyst reports represents the use of new perspective: if analysing 15 years of brokerage reports is characterised by scouring every square inch of a football stadium for a few nuggets of valuable information, analysing just one month's worth of Google search data can be characterised by combing an area the size of New York City.

With this much data, the possibilities for uncovering something of value are nearly endless. Yet so are the chances of getting lost in what is often a low signal-to-noise environment. In order to increase the likelihood of success, it can be helpful to return to some basic principles of the scientific method. For example, starting with a hypothesis and then using ancillary testing to confirm or disprove the results. During the testing phase, two concerns to be aware of are, the risks of overfitting and the challenges of dealing with non-stationarities. There are statistical techniques that can help guard against the former, while the latter often requires human judgment to identify changes in prevailing market or economic conditions and to adjust the algorithms accordingly.

One area of research that has proved rewarding is measuring the impact of employee sentiment on company profitability. Of course, the idea of examining the relationship between employee satisfaction and stock market performance is not new. In a 2010 paper published in the *Journal of Financial Economics*, associate professor Alex Edmans of the University of Pennsylvania's Wharton School of Business analysed the relationship between employee satisfaction and long-run stock returns and found that over a 25-year period, a value weighted portfolio of Fortune magazine's '100 Best Companies to Work For in America' outperformed industry benchmarks by 2.1%.

The results seemed to confirm a correlation many had long suspected. However, this was a backward-looking study of a dataset that is only updated once a year and is restricted to US companies. Today by using big data analysis and crowdsourcing research, we can not only measure employee sentiment on a much wider scale, but also in a more timely and accurate way.

To do this we use automated 'web-scraping' capabilities to look into job sites where employees provide feedback on thousands of employers across the world, as well as other sources of employee sentiment including social media, blogs and chat rooms. When all of the relevant data from these sources has been collected an overall sentiment score for a company is calculated and this score can be regularly updated and changes monitored. Exhibit 2 shows how a portfolio of the top ranked companies with respect to employee satisfaction has performed compared to a portfolio of the lowest ranked companies.

EXHIBIT 2: THE HAPPINESS DIVIDEND

Return of firms scoring in the top quintile for employee satisfaction versus those scoring in the bottom quintile



Source: BlackRock's Scientific Active Equity Group, February, 2015. These are cumulative specific returns for owning names in the top and bottom quintiles of employee sentiment in the MSCI ACWI IMI Index, as determined by SAE. The names in these quintiles are not static and change through time. This is a frictionless simulation (i.e., no transaction costs are applied).

CONNECTING THE DOTS

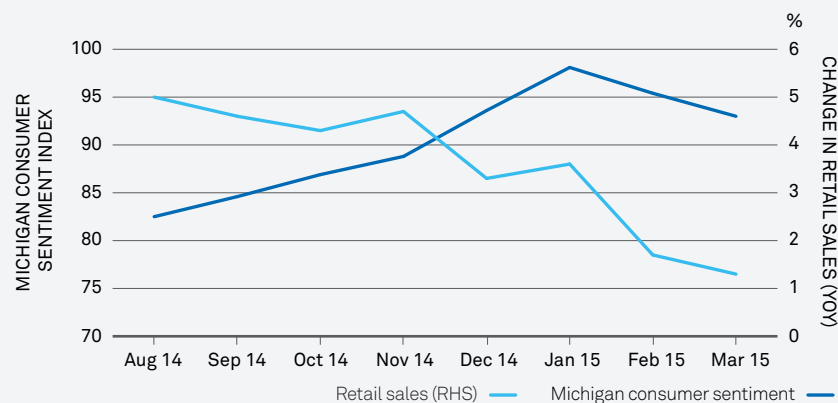
In addition to helping us understand what human indicators, such as analysts and employees, are trying to tell us about individual companies, big data analysis can also help us construct better economic signals that may offer clues about the outlook for entire industries.

A recent example can help illustrate this: it is generally believed that improving consumer confidence will positively impact retail sales. It should therefore come as no surprise that market participants focus on the University of Michigan's monthly Consumer Sentiment Index as a possible leading indicator of retail sales.

Unfortunately, there is at least one problem with this approach. People often say one thing when participating in a survey but then go on to do something completely different in the real world. As the price of oil plummeted late last year conventional wisdom suggested that consumers would take some of what they saved at the pump and spend it in shops. The Consumer Sentiment Index captured this optimism, rising by more than 10% over several months. However, as can be seen in exhibit 3, this optimism appears misguided, as retail sales momentum steadily declined over the same period.

EXHIBIT 3: SIGNALS CROSSED

University of Michigan consumer sentiment and retail sales growth



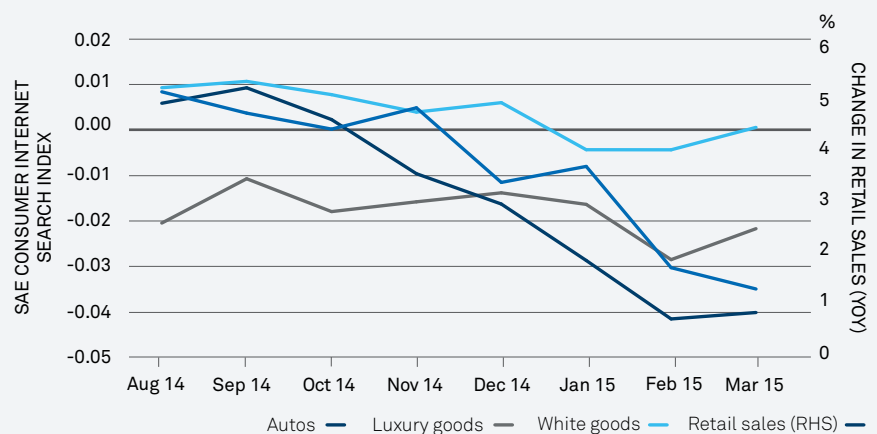
Source: Bloomberg, April 2015.

Rather than focusing on a 70-year-old monthly survey of 500 consumers, we think it is possible to achieve more accurate results with a more modern toolkit and a much larger sample size.

Most consumers now precede big purchases with some research on the internet. By tracking search activity, around the world, for big-ticket items, we can gain a more complete picture of consumers' purchasing intentions, which also includes granular data on the types of items that are likely to be purchased in different geographic regions. The results from this type of analysis appear promising. As can be seen in exhibit 4, our measures of consumer sentiment have been better recent predictors of retail sales.

EXHIBIT 4: PREDICTIVE POWERS

Internet search sentiment and retail sales growth



Sources: BlackRock's Scientific Active Equity Group and Bloomberg, April 2015.

FIXING A HOLE

These techniques can be used to generate signals that are valuable to both equity and fixed income managers in analysing individual companies and forecasting economic data. Furthermore, fixed income markets may involve complex non-standardised issues, idiosyncratic structured products and opaque secondary markets, which can present additional hurdles when conducting security-level analysis.

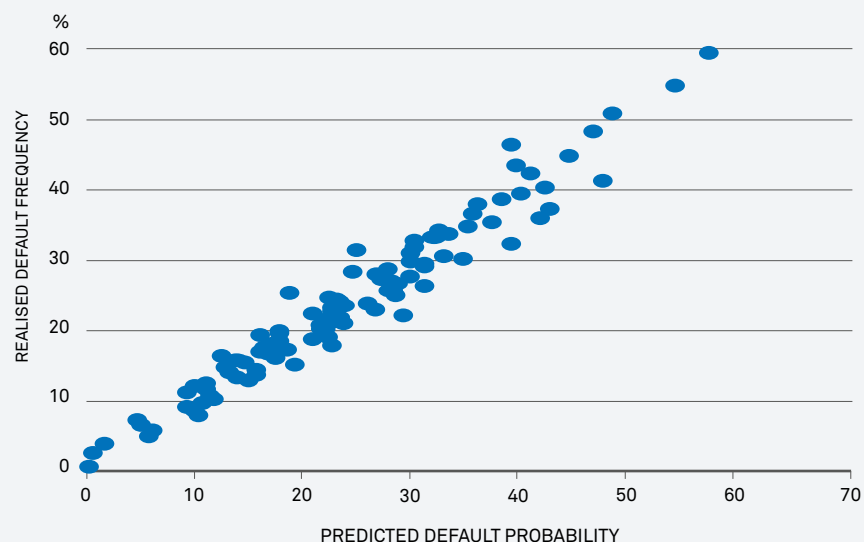
These markets can provide fertile ground for managers who can systematically analyse a large array of complex individual securities. For example, on the surface, the universe of commercial mortgage backed securities (CMBS) appears to be a rather unruly dataset – but beneath the surface are rich veins of data at the individual loan or pool level. By combining machine-learning techniques and portfolio manager expertise a deeper understanding of the underlying data can be gained.

In order to extract value from data on tens of thousands of CMBS loans, it is necessary to write programmes that can read important information that is generally contained in an annex to the prospectus supplement of each security. This is a complicated process with a high potential for error that requires strict quality control standards. In addition to the top line data that a well designed algorithm can extract, CMBS can also contain subordinate financing details that are buried within sub paragraphs or footnotes of loan documentation. These can significantly alter the top line data.

The details of this kind of subordinate financing need to be assessed manually by experts with product specific knowledge in order to gain a more complete understanding of the risk profile of individual loans. If managers are able to combine the necessary human market expertise with a powerful computing platform, it is possible to build a model that can help to better predict loan default. As shown in exhibit 5.

EXHIBIT 5: FOREWARNED

Predicted and realised CMBS defaults



Source: BlackRock Solutions data on CMBS loans securitized between 2004 and 2007. Default information up until September 2014.

THE BATTLE FOR TALENT

It isn't just the construction of fixed income models that necessitates collaboration between man and machine. Many solutions to the vexing puzzles that impact 21st-century investors requires the best that data, technology and people have to offer. At the same time, we need to acknowledge that human beings are responsible for building, using and interpreting the advanced models that are making sense of the world's information.

The need to hire high calibre personnel is nothing new for investment managers; the industry has long recognised and rewarded top talent in the fields of portfolio management and trading. From now on in order to gain an edge in today's markets, organisations may need to spend the same kind of money, time and care to cultivate data science talent with expertise in fields such as machine learning, natural language processing, distributed computing and scientific data visualisation.

There is a fierce battle going on for talent. Not just within the field of finance, but increasingly across nearly every sector of the economy, led, of course, by the technology industry. If investment managers hope to interrupt the flow taking many of the top computer scientists, statisticians and engineers to tech firms, they will need to commit to a culture of innovation and empowerment that is less common within the world of finance.

It is also important to keep in mind that some things have not changed. Market insights still matter, and people still need to perform intelligent analysis of the available data in order to arrive at sensible investment recommendations, just as they did in the era of Graham and Dodd. It's easy to forget today, but back in the 1930s and 1940s, it required a visionary mind-set and dedication to collect and analyse many pages of accounting statements and pricing data. Institutions that build investment teams that comprise the best of both technological expertise and investing acumen stand a greater chance of success.

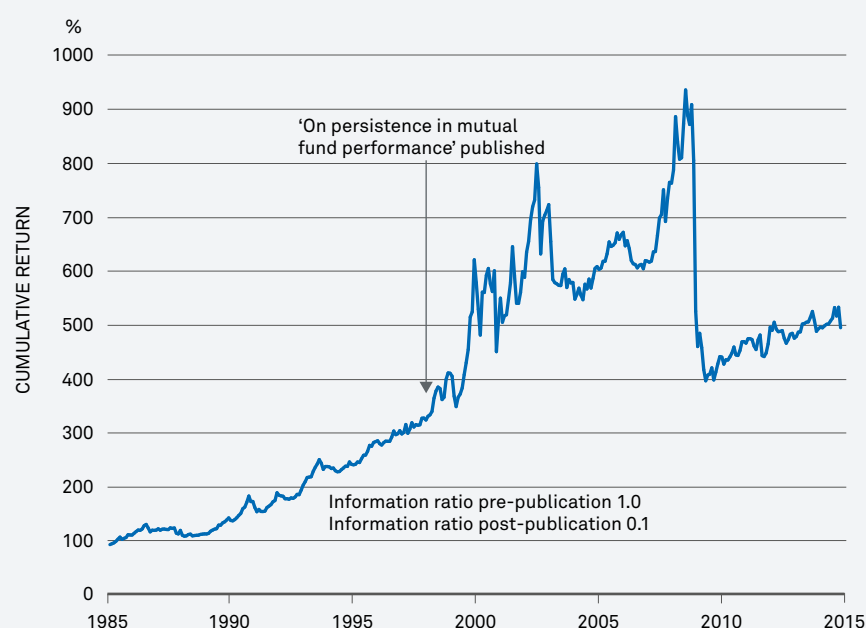
ONE STEP BACK, MANY PATHS FORWARD

Even if investment managers assemble the brightest minds and equip them with the best tools, there is no guarantee that the insights they uncover will have staying power, in the pursuit of alpha.

The widely followed momentum factor offers an instructive case of how once reliable signals can decay over time. As we can see in exhibit 6, momentum delivered enviable results to managers that were early to identify it as a reliable predictive tool and incorporate it into their investment strategies. Over the 12-year period from 1985 to 1997, the Fama-French momentum factor delivered over 300% cumulative return in a steady fashion, with an information ratio of 1.0.

EXHIBIT 6: LOSING MOMENTUM

Long-term returns to the Fama-French momentum factor



Source: Fama French. as of 30 April 2015.

Then the word got out. In March 1997 Mark Carhart published his seminal Journal of Finance paper 'On Persistence in Mutual Fund Performance', which detailed the strong historical performance of a momentum-based strategy, the returns from following such a strategy changed dramatically. Instead of realising slow, consistent alpha with few large drawdowns, momentum investors from 1997 on have experienced significantly lower cumulative returns, higher volatility, a 90% reduction in the information ratio and a genuine crash that shook much of the quant and fundamental establishment to its core.

Of course, there are alternative explanations for the performance of momentum in the post-Carhart era, but we think investors would be remiss if they did not at least consider the possibility that the change in factor return patterns was at least partially due to increased awareness of the factor itself.

Before entrusting money to the current generation of quantitative managers, investors are right to ask whether the strategies that are now successfully harnessing big data to derive alpha may similarly decay or become commoditised. The truth is that any particular signal will generally lose efficacy over time, for a variety of reasons including increased awareness of the pattern, regulatory changes and shifts in market and macroeconomic climates.

Having superior data and superior methods of extracting information from that data have provided an advantage in investing, and we believe they always will. This is not unique to the class of signals that draw on machine learning, natural language processing or any of the other techniques that we have discussed here. We will use these tools for as long as they give us alpha. If they stop doing so, we will use other tools. That's the nature of good science. Our allegiance is to finding alpha, not a particular set of tools or ideas.

Nonetheless, we firmly believe that a data-driven, scientifically based, technologically aware research culture can produce sustainable alpha. Furthermore, we are just beginning to scratch the surface with regard to understanding everything that both data and humans can tell us about markets. We expect that all of the techniques described in this paper will come to be seen as routine in the future. We also expect that as investment managers understand more about how the power of artificial intelligence, machine learning and conditional thinking can be used, important new insights will be revealed. We find these prospects very exciting. They inspire our continued commitment to the scientific process, in pursuit of a greater understanding of human beings, financial markets and the world's data.

Want to know more?



+44 (0)20 7743 3300



insights@blackrock.com



blackrock.com

The following notes should be read in conjunction with the attached document: This material is for distribution to Professional Clients (as defined by the FCA Rules) and should not be relied upon by any other persons. Issued by BlackRock Investment Management (UK) Limited, authorised and regulated by the Financial Conduct Authority. Registered office: 12 Throgmorton Avenue, London, EC2N 2DL. Tel: 020 7743 3000. Registered in England No. 2020394. For your protection telephone calls are usually recorded. BlackRock is a trading name of BlackRock Investment Management (UK) Limited. Past performance is not a guide to future performance. The value of investments and the income from them can fall as well as rise and is not guaranteed. You may not get back the amount originally invested. Changes in the rates of exchange between currencies may cause the value of investments to diminish or increase. Fluctuation may be particularly marked in the case of a higher volatility fund and the value of an investment may fall suddenly and substantially. Levels and basis of taxation may change from time to time. Reference to the names of each company mentioned in this communications is merely for explaining the investment strategy, and should not be construed as investment advice or investment recommendation of those companies. Issued in Switzerland by the representative office BlackRock Asset Management Schweiz AG, Bahnhofstrasse 39, CH-8001 Zürich. Issued in the Netherlands by the Amsterdam branch office of BlackRock Investment Management (UK) Limited: Amstelplein 1, 1096 HA Amsterdam, Tel: 020 - 549 5200. Please be advised that BlackRock Investment Management (UK) Limited is an authorised Financial Services provider with the South African Financial Services Board, FSP No. 43288. This information can be distributed in and from the Dubai International Financial Centre (DIFC) by BlackRock Advisors (UK) Limited - Dubai Branch which is regulated by the Dubai Financial Services Authority ("DFSA") and is only directed at 'Professional Clients' and no other person should rely upon the information contained within it. Neither the DFSA or any other authority or regulator located in the GCC or MENA region has approved this information. This information and associated materials have been provided to you at your express request, and for your exclusive use. This document is not intended for distribution to, or use by, any person or entity in any jurisdiction or country where such distribution would be unlawful under the securities laws of such. Any distribution, by whatever means, of this document and related material to persons other than those referred to above is strictly prohibited. Any research in this document has been procured and may have been acted on by BlackRock for its own purpose. The results of such research are being made available only incidentally. The views expressed do not constitute investment or any other advice and are subject to change. They do not necessarily reflect the views of any company in the BlackRock Group or any part thereof and no assurances are made as to their accuracy. This document is for information purposes only and does not constitute an offer or invitation to anyone to invest in any BlackRock funds and has not been prepared in connection with any such offer. © 2015 BlackRock, Inc. All Rights reserved. BLACKROCK, BLACKROCK SOLUTIONS, iSHARES, BUILD ON BLACKROCK, SO WHAT DO I DO WITH MY MONEY and the stylized i logo are registered and unregistered trademarks of BlackRock, Inc. or its subsidiaries in the United States and elsewhere. All other trademarks are those of their respective owners.

UNLESS OTHERWISE SPECIFIED, ALL INFORMATION CONTAINED IN THIS DOCUMENT IS CURRENT AS OF SEPTEMBER 2015, SOURCE BLACKROCK. 005065b-INST-EMEA-SEPT15-EN-CARS ID IMPEG 0258-COMPLIANCE END DATE: 9 SEPT 2016

BLACKROCK®