Deutsche Bank
Markets Research

Europe

Quantitative Strategy

Thematic Report - Europe

Date
16 February 2016

Mean Reversion II: Pairs Trading Strategies

Shuo Qu, Ph.D
shuo.qu@db.com

Spyros Mesomeris, Ph.D
spyros.mesomeris@db.com

Christian Davies
christian.davies@db.com

Caio Natividade
caio.natividade@db.com

Paul Ward, Ph.D
paul.ward@db.com

Jacopo Capra
jacopo.capra@db.com

James Osiol
james.osiol@db.com

Vivek Anand
vivek-v.anand@db.com

North America: +1 212 250 8983
Europe: +44 20 754 71684
Asia: +852 2203 6990

In a previous report, we discussed cross-sectional mean reversion strategies in equity markets. Pairs trading, which attempts to exploits a temporary mis-pricing between two securities with a stable relative price relationship, is another type of mean reversion strategy. In this report, we show how you can improve both the selection and trading aspects of a conventional pairs trading strategy.

Fundamental risk models help to identify profitable pairs

Pairs trading strategies typically look for co-integrated relationships between stocks belonging to the same country and sector/industry group. We believe there are superior means with which to capture the degree of "fundamental similarity" between stocks. For example, we show that utilizing a fundamental risk model to identify stock pairs significantly reduces divergence risk, and also improves the average return per pair.

News analytics overlay to further enhance pairs trading performance

Divergence risk increases in the proportion of idiosyncratic risk associated with a pair's constituent stocks. A news analytics overlay which helps to differentiate between price divergence due to news as opposed due to random price movements, significantly improves the performance of the trading strategy by reducing the number of non-convergent trades.

Beyond stock pairs

In looking for potential pairs candidates, we do not have to limit ourselves to stock pairs. We propose a novel method based on clustering and dynamic tree-cutting to systematically identify clusters of stocks as potential constituents for synthetic pairs trading strategies.

Source: Getty Images

Deutsche Bank AG/London

Table Of Contents

Introduction

The principle of pairs trading is remarkably simple. An investor finds assets whose prices have moved together historically, open a trade by shorting the winner and buying the loser when the spread between them widens, and close the trade when the spread converges. The strategy exploits temporary anomalies between prices of assets that have a relatively long-run equilibrium relationship. While methods may differ in their sophistication, all implementations rely on the use of statistical analysis of historical prices to identify pair candidates with a stable inter-relationship. As is often the case, it may sound simple, but the devil is in the details.

Since its invention in the 1980s[1], pairs trading has grown into one of the most popular statistical arbitrage strategies. The first extensive academic research on pairs trading was conducted by Gatev et al. (1999,2006), who examined the return and risk characteristics of a simple pairs trading strategy using CRSP data over the period 1962 to 2002. They documented economically and statistically significant profits of around 1% per month. They also found that these profits survived even with conservative transaction cost assumptions. Subsequent research, e.g. by Do and Faff (2009, 2010), has shown a dramatic decline in the profitability generated by trading stock pairs[2]. As is pointed by Chan (2013), it has become increasingly difficult to squeeze profit out such strategies as the market has become much more efficient over the past decade. Another specific reason for the decline in the profits of pairs trading is the decimalization of stock prices, which has caused bid-ask spreads to narrow dramatically. So pairs traders, who act as a type of market makers, find that their market -making profits decrease also (Serge, 2008).

The first step in a stock pairs trading program is finding pairs of related stocks with stable relative price relationships. However, it is often the case that the prior equilibrium relationship between two stocks does not persist in the subsequent out-of-sample period. Previously, researchers focused on pairs with common sector or industry affiliations. Gatev et al. (2006) restrict the selection of pairs within the same sector, and Do and Faff (2010) show that considerable benefit can be achieved by using finer industry classification schemes. More recently, research by Zhao et al. (2015) proposes using supply chain data to select pairs linked by a customer-supplier relationship. Ultimately, this body of research suggests that pairs with a fundamental relationship are better candidates for a pairs trading strategy. This result is intuitive as fundamentally similar assets are likely to be exposed to similar risks and opportunities. From an asset pricing perspective, this means that their cash flows and discount rates are highly correlated. Thus, they are more likely to have a true equilibrium relationship.

Once stocks are paired up in a sensible way, the next step is to initiate a long/short position when the spread between the pair diverges and unwind the position upon convergence. It is argued that profits and risks from trading stock pairs are very much related to the information event which creates

[1] Pioneered by Nunzio Tartaglia's quantitative group at Morgan Stanley.

[2] They extended the Gatev analysis dataset to June 2008 and had found a 60% decline in returns between 2002 and 2008.

divergence (Engelberg et. al. 2008). For example, the return from the strategy is often expected to be small if the divergence event is caused by negative idiosyncratic news to a constituent of the pair, as it is likely to cause long-term or even permanent differences in the prices of the pair. On the other hand, if the divergence is caused by common information which diffuses into the pair at differential rates causing prices to temporarily move apart, it is more likely that prices of the pair will converge and a profit will be realized.

Taken together, it is difficult to be consistently profitable in trading pairs of stocks without having a fundamental understanding of the companies forming the pair. In addition, the ability to differentiate between “good” divergence (which is likely to converge) from “bad” divergence, and to exit a position in time when the likelihood of further divergence increases is of paramount importance. The question is, how do we systematically achieve these aims?

In this report, we discuss a few methods that can boost the otherwise declining profit of pairs trading in the equities market. First, we propose the use of a fundamental risk model to systematically identify fundamentally similar companies as pairs candidates. Second, we show how news analytics can be used as an overlay to further improve the performance of a pairs strategy. This is done by inserting a step in the process that automatically checks for news and measures sentiment, which would indicate whether the divergence in the monitored price relationship had occurred as a result of a clear fundamental cause, or due to random price movements. If the latter we would expect the price relationship to revert to historic norms. Finally, in searching for pairs, one does not have to confine himself to a certain number of constituents. The constituent of a synthetic equity pair can be a single stock, a basket of stocks, an index or a combination of all three. We illustrate how the proposed fundamental risk model, combined with a smart clustering scheme, can be utilized to identify baskets of stocks as potential constituents for synthetic pairs.

Pairs Trading: The Basics

In this section we examine the three most commonly used methods to implement pairs trading, the distance approach, the stochastic spread approach and the co-integration approach. While more sophisticated models have been developed, they are generally derived from one of these three approaches.

Existing Pairs Trading Approaches

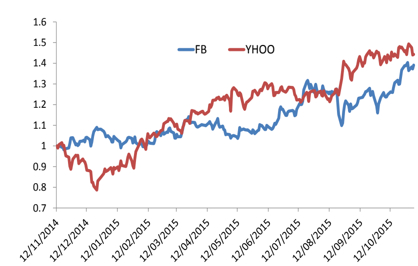
The Distance Approach

The distance approach looks to trade pairs whose prices closely match historically. This is taken as indication that the assets are fungible and by the Law of One Price (LOP) their prices should be nearly identical. The co-movement in a pair is measured by what is referred to as the distance. This is the sum of squared differences between the two normalized prices series over the formation period:

$$D = \sum_t (p_t^A - p_t^B)^2 \quad (1)$$

where p_t^A and p_t^B are the normalized prices[3] for stocks A and B respectively.

Figure 1: A pair example: normalized price paths of Facebook (FB) and YAHOO (YHOO)



Source: Deutsche Bank, FactSet, Axioma,RavenPack

Pairs are selected by choosing a matching partner that minimizes the historical distance over the formation period. Trading is triggered when the distance for each pair reaches a pre-defined threshold determined during the formation period. The most commonly cited research on this approach is that by Gatev et al. (1999).

This approach is model-free and consequently has the advantage of not being exposed to model mis-specification and mis-estimation. On the other hand, being non-parametric means that the strategy lacks forecasting ability regarding the convergence time or expected holding period. More fundamentally, it assumes that the price level distance is static through time, or the returns of the two stocks are in parity. Although such an assumption may be reasonable over short horizons, it is only for pairs whose risk-return profiles are close to identical.

The Stochastic Spread Approach

The stochastic spread approach as first outlined by Elliot et al. (2005) explicitly models the mean reverting behavior of the spread in a continuous time setting. The observed spread y_t is defined as the difference between the two stock prices. It is assumed that the observed spread is driven mainly by a state process plus some measurement error ω_t i.e.,

$$y_t = x_t + H\omega_t, \quad (2)$$

where $\omega_t \sim N(0,1)$ with $H > 0$ is the size measure of the error. The latent variable x_t is assumed to follow an Ornstein-Uhlenbeck process:

$$dx_t = \rho(\theta - x_t)dt + \sigma dB_t \quad (3)$$

[3] Both prices are normalized to begin at 1 over the formation period.

where dB_t is a standard Brownian motion. The state variable is known to revert to its mean θ at speed ρ.

The stochastic approach as specified by equations (2) and (3) offers some advantages from the empirical perspective. First, it captures the mean reversion which underlies pairs trading. Second, being a continuous time model, it is convenient for forecasting purposes. The model is completely tractable, with its parameters easily estimated by the Kalman filter within the framework of a state space setting. The expected time that the spread converges back to its long-term mean can be computed explicitly. This allows investors to have a more direct view on the expected holding period and expected return.

Despite the advantages, this approach has the fundamental limitation in that it restricts the long-run relationship between the two stocks to return parity. The stock pairs chosen must provide the same return such that any departure will be corrected in the future. This severely limits this model's generality as in practice it can only be applied in limited circumstances. For example, it may be well suited for companies that are dual-listed. In addition, the model can only be applied to determine trading decisions (e.g. when to enter and exit a position) once a pair has already been identified. It does not help identify profitable pairs in the first place.

The Co-integration Approach

The co-integration approach attempts to parameterize the pairs strategy by exploring the possibility of co-integration between two or more assets. Co-integrated price series possess a stationary long-run equilibrium relationship with the associated property of mean reversion. This is an important property as it means short-term departures from the equilibrium will eventually re-converge. The literature for this approach includes Alexander and Dimitriu(2002), Vidyamurthy(2004), Galenko et al.(2012) and Cartea et al. (2015).

A times series x_t is said to be integrated of order *d*, abbreviated as *I(d)*, if it needs to be differenced a minimum of *d* times to become a stationary process, denoted by *I(0).* Two or more integrated series are termed 'co-integrated' if a linear combination of these series is stationary.

There are a few popular tests for co-integration. Engle and Granger (1987) formulated one of the first tests of co-integration. The first step in the Engle-Granger's two-step procedure is to perform an ordinary least squares (OLS) regression on the integrated series, and the second step is to test the residuals for stationarity, for example, by applying the Augmented Dicky-Fuller(ADF) test.

However, the Engle-Granger test may suffer from bias when the number of variables is greater than two. It needs to explicitly define which series is to be used as the dependent variable in the regression. Finally, the OLS regression in the first step will lead to spurious estimators if the variables are not co-integrated. This makes the stationarity analysis on the residuals unreliable.

The Johansen test for co-integration is commonly regarded as superior to the Engle-Granger method. This is particularly true when the number of variables is greater than two. Johansen (1988) suggests a method for both determining how many co-integration vectors there are and also estimating all the distinct relationships. A short description of the Johansen method is included in Appendix A.

Based on the co-integration relationship, a portfolio comprising of $\beta = (\beta_1, \beta_2, ..., \beta_p)$ positions in stocks 1 to *p*, is by construction mean-reverting and can be modeled as a mean reverting process. This is again desirable from a forecasting perspective. The co-integration approach also has the advantage of offering a framework that can extend pairs trading to a basket setting, which allows investors to exploit the long-run equilibrium among multiple assets.

Apart from its obvious advantages, the co-integration approach has a number of limitations. Being a parametric approach, it is exposed to errors arising from the econometric techniques employed. The existing co-integration test statistics generally do not come from standard distributions and depend on the specifications for testing[4].Hence, higher co-integration test statistics for certain pairs of stocks don't necessarily imply that these candidate pairs are "superior" to others.

The Benchmark Strategy

Choice of Approach

We adopt a co-integration based approach in this report. This allows us to derive a dynamic definition of the long-rum equilibrium price spread that is implicitly mean reverting. The co-integration based approach also provides a framework for us to extend trading pairs into trading synthetic pairs (i.e. baskets), which we will discuss later in this report.

Data

We employ daily data from the MSCI Europe universe over the period January 2001 to December 2015. The chosen universe is rather liquid, which reduces the impact of microstructure factors such as bid-ask bounces, short-sale constraints and/or excessive short-selling costs and bankruptcy risk. The training period of the strategy is based on a rolling window of 12 months, and the trading period is based on a window of 3 months. We also screen out stocks with more than one missing price during the training period.

Methodology

To construct stock pairs, our benchmark strategy requires *country neutrality*, i.e., all stocks forming a pair must belong to the same country. This is to ensure that all stocks forming a pair are denominated in the same currency so as to avoid cross-market microstructure issues and currency risk.[5] Our benchmark strategy also requires *sector neutrality*, meaning that all stock pairs must also belong to the same GICS sector. This is a reasonable starting point as sector affiliation for a stock can explain a significant portion of its risk and return.

---

[4] Unit root test statistics have non-standard and non-normal asymptotic distributions under their respective null hypotheses. The limiting distributions of the test statistics are also affected by the inclusion of deterministic terms in the test regressions. These distributions are usually functions of stochastic process, so critical values must be tabulated by simulation techniques.

[5] There are indeed opportunities to profit from investing in cross-country pairs. However, to successfully execute pairs across markets, algorithms not only have to resolve market microstructure issues, but also manage currency risk. For example, if one leg of a pair must stop trading due to events like auctions, trading halts, lunch breaks, etc, the other leg must take action immediately to mitigate leg risk. Also, regulations and rules on lots, tick sizes and short selling in different markets require sophisticated order placement logic.

Pairs are matched exhaustively within the same country and GICS sector at the end of each month. For each pair, the Johansen test of co-integration is applied, based on the log prices of the pair from the 12-month training period. Pairs that have passed the co-integration test are eligible for trading for the next three months. We base our trading rules for opening and closing positions on a standard deviation metric. A long-short position is initiated when the spread of the pair diverges by two standard deviations[6], as measured over the prior12-month training period. All positions are closed out at the end of the three month trading period regardless of whether the prices of the pair have converged or not.

From a time series perspective, the speed at which the spread of a pair reverts back to its mean is different for different pairs of stocks. We take these differences into account when initiating a trade, as well as with formulating stop-loss strategies. First, pairs with a relatively long half-life[7] (for example, longer than the trading period itself), are unlikely to converge within the pre-defined trading period and will simply be closed before convergence, and therefore should be excluded from trading. Second, a position should not be initiated when the remaining time for trading is not long enough for the trade to converge (for example, when remaining time <2*half-life). In addition, it helps to provide an indication whether the position is valid during the trading period. If the spread has not reverted after multiple half-lives, we have reason to believe that probably a regime has changed, or our model may no longer be valid (broken relationship or new equilibrium).

The pairs position is therefore closed under any of the following conditions:

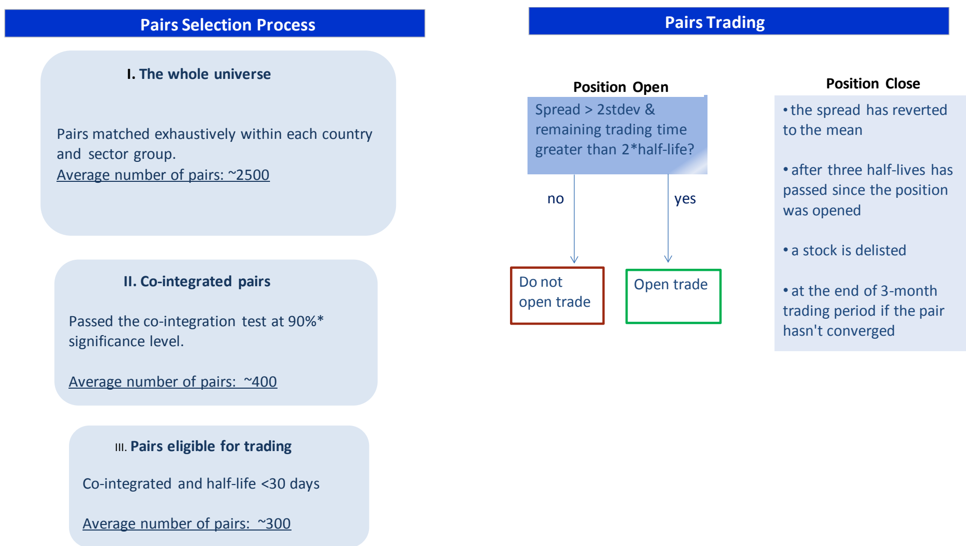
- The spread has reverted to the mean,
- After three half-lives has passed since the positions was opened,
- A stock in the pair is delisted from the universe,
- At the end of the 3-month trading period.

Details of the benchmark pairs selection and trading process is illustrated in Figure 2.

[6] In this report we do not seek to search for an optimal trading rule. The standard deviation based trading rule used here might not always cover transaction costs even when stock prices converge.

[7] The half-life of the spread gives us a rough estimate of how long we should expect the spread to remain far from its mean. The spread u_t can be modeled as a simple Ornstein-Uhlenbeck process $du_t = \rho(\theta - u_t)dt + \sigma dB_t$, and it can be shown that the half-life of this OU process equals $\ln(2)/\rho$.

Figure 2 The benchmark pairs selection and trading process



* In practice, we don't need perfect co-integration to implement a successful mean reversion strategy. In fact, research has shown that assets with the higest level of co-integration in-sample are usually the least robust out-of-sample(for example, Meucci 2010). In addition, in the case of co-integrated eigenseries, the volatility is the square root of the respective eigenvalue, which imples that the most mean-reverting series corresponds to a much lesser potential return.

Source: Deutsche Bank, FactSet, Axioma,RavenPack

Transaction Costs

Pairs trading is a cost-sensitive investment strategy. It involves frequent rebalancing, multiple openings and closings of trades and short-selling. In particular, the pairs strategy sells stocks that have done well relative to their "match" and buys those that have done poorly. Part of any observed price divergence is potentially due to price movements between bid and ask quotes. Therefore it is particularly important to consider transaction costs when evaluating the profitability of the strategy. Unless stated otherwise, we will report results for strategies that open (close) on the day following divergence (convergence) and include a one-way transaction cost of 10 basis points.

Note that the strategies discussed in this report have employed a simple standard deviation rule to open and close positions. The use of historical standard deviation to trigger the opening of pairs may open the pair too soon or at a point that would not provide returns over and above transaction costs (even if the pair subsequently converges). In practice, it may be possible to optimize the trading rules employed to maximize the profits to such strategies. This is an area that we would like to revisit in later work.

Benchmark Pairs Model Performance

We will briefly discuss the characteristics and performance of the benchmark pairs trading model.

Figure 3 shows the average number of pairs eligible for trading and the average number of open pairs each month. On average we have around 300 pairs eligible for trading each month and approximately 220 of them open for trading. In total, we trade over 40,000 pairs over the 15-year backtesting period.

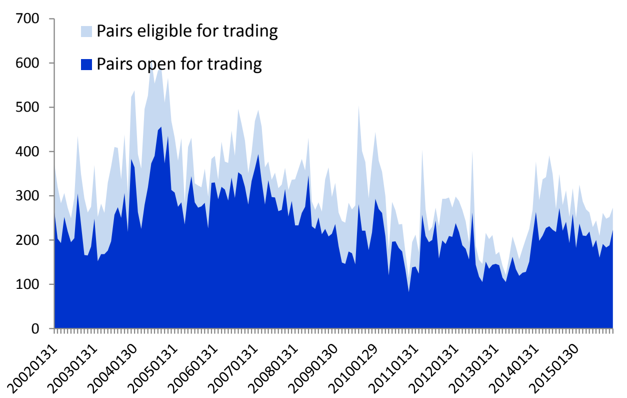
Figure 4 shows the percentage of traded pairs that fall into one of the three mutually exclusive groups: (1) pairs that do open but fail to converge during the designated trading period, (2) pairs that have one round trip trade and possibly another non-convergent trade, (3) pairs that have multiple round trip trades and possibly a final non-convergent trade. Group (1) represents the main risk in pairs trading: non-convergence risk and above 35% of the traded pairs fall into this category. Over half of the traded pairs have one roundtrip trade and another 10% have multiple roundtrip trades.

The profit distributions for different groups of pairs are shown in Figure 5. Overall, the benchmark strategy achieves an average profit of 0.5% per pair per three-month trading period, after transaction costs. Specifically, single and multiple roundtrip pairs (groups 2 and 3) achieve an average return of 1.4% and 3.2% respectively, whereas for non-converged pairs the average loss is over 1.5% (and the maximum loss 20%).

Figure 6 plots the distribution of time-to-convergence, conditional on a trade converging. The mode of the distribution is centered around 6 days, but there is a small percentage of trades that take more than 2 months to converge.

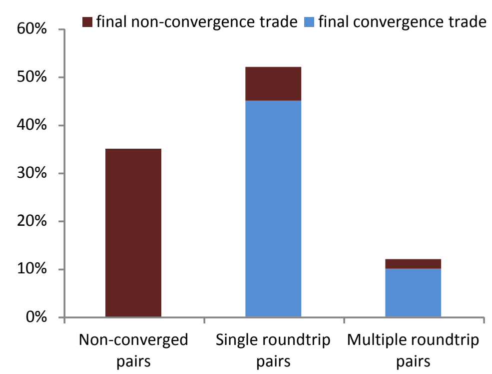
So why do some pairs manage to achieve higher returns than others? What happened on the divergence date and what pair characteristics contributed to the divergence? And what are the factors related to the speed of convergence? In the next section we make an attempt to answer these questions through an enhanced pairs trading model.

Figure 3: Average number eligible pairs and traded pairs



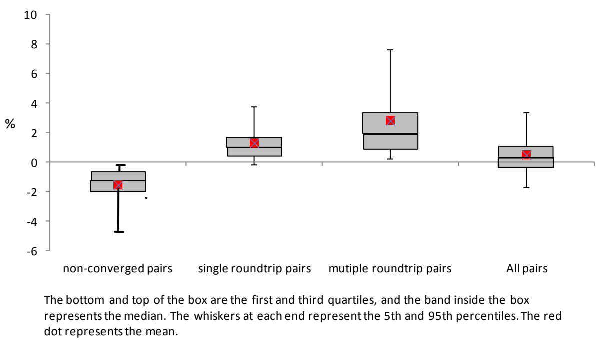
Source: Deutsche Bank, FactSet, Axioma,RavenPack

Figure 4: A breakdown of traded pairs



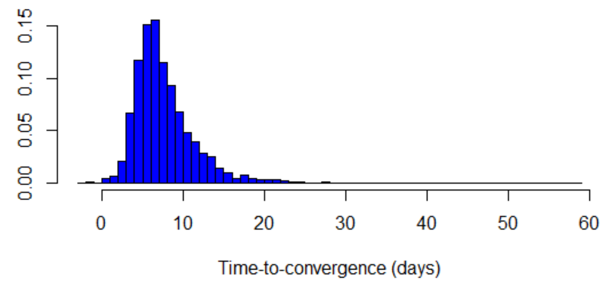
Source: Deutsche Bank, FactSet, Axioma,RavenPack

Figure 5: Benchmark pairs trading profit distribution (per pair)



Source: Deutsche Bank, FactSet, Axioma,RavenPack

Figure 6: Time-to-convergence



Source: Deutsche Bank, FactSet, Axioma,RavenPack

An Enhanced Pairs Trading Model

Identifying Pairs Using a Fundamental Risk Model

In the previous section, we described a benchmark pairs trading strategy based on the principles of co-integration. As with any other time-series methodology, there is a tendency for time-series modeling to be led by the technique, rather than any real relationship. Quite often, co-integration between two assets breaks down out-of-sample and trading the pair is a losing proposition. This is evident as over a third of traded pairs under the benchmark strategy failed to converge. However, it is often quite difficult to detect the breakdown of co-integration, except in hindsight. The benchmark strategy accounts for this by forming pairs within the same country and sector group in an attempt to minimize divergence risk.

While both country and sector affiliations are significant factors, risk is multi-dimensional. Considering these additional dimensions can help to better understand the relationships between assets. We believe a more robust approach to identifying and utilizing the common fundamental drivers between assets is using a fundamental risk model.

Fundamental risk models provide a consistent and interpretable framework for performance and risk attribution. They also separate stock-specific and market-level events, and track the evolution of market and factor correlations. Changes in an issuer's exposures or operations are more likely to be reflected on a timely basis, and compared to times series models fundamental models are less likely to confuse noise for signal. Stocks that are similar in factor space are more likely to move in tandem and less likely to diverge significantly. Therefore, they make for very promising candidates for pairs trading.

We employ Axioma's Europe short horizon fundamental risk model[8] to identify stocks that are closely related. We define "closeness" of two stocks as their pair-wise correlation implied by the risk model. The *NxN* asset correlation matrix of a fundamental factor model can be calculated as:

$$COR = \frac{\hat{C}}{VV'} \qquad (4)$$

where $\hat{C} = \hat{X}FX$ is a NxN asset covariance matrix and $V = \sqrt{diag(\hat{C})}$, F is a KxK factor covariance matrix, X is a KxN matrix of asset exposures to factors. The list of *K* risk factors used by the model can be found in the table in Figure 7. Apart from market, country and industry factors the model employs 9 style factors. They are value, leverage, size, growth, short term and medium term momentum, volatility, liquidity and exchange rate sensitivity.

[8] The short-horizon fundamental risk model uses a 125-day half life in calculating factor correlations and volatilities and a half-life of 60 days in calculating specific risk.

Figure 7: Axioma's Europe short horizon fundamental risk model factor list

Factor Class	Factors
Market	European Market
Country	Austria, Belgium, Czech Republic, Denmark,Finland, France, Germany, Greece, Hungary, Ireland, Italy, Luxembourg, Netherlands, Norway, Poland, Portugal, Russia, Spain, Sweden, Switzerland, Turkey
Industry	GICS II industries
Style	Value Leverage Growth Size Short-Term Momentum Medium-Term Momentum Volatility Liquidity Exchange Rate Sensitity

Source: Deutsche Bank, FactSet, Axioma,RavenPack

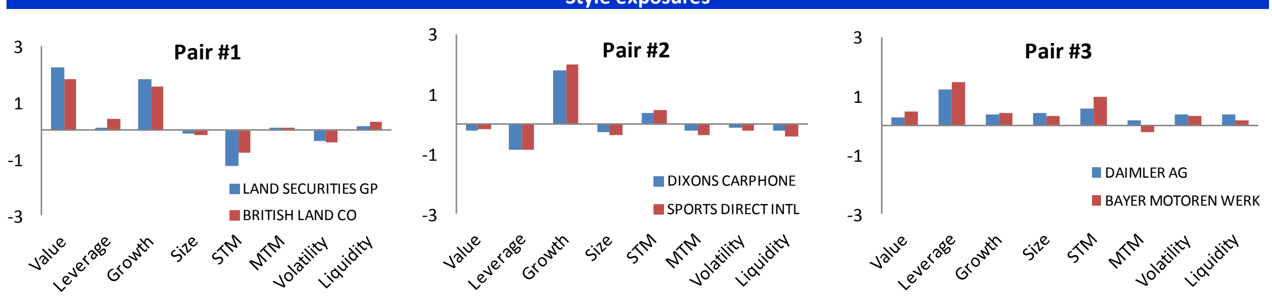
High correlation implies two stocks are exposed to similar risk factors, and are therefore fundamentally "close". As a result, their prices are more likely to move together in the near future in the absence of idiosyncratic events. Figure 8 shows a few examples of high correlation pairs as of 30-Nov-2015. Note that these pairs not only have the same country and industry affiliations, but also have similar style exposures. For example, pair Dixon Carphone and Sports Direct Intl are both UK retail stocks. In addition, they also have similar levels of valuation, leverage, volatility and liquidity exposures. They are both growth stocks with similar market capitalizations and also have similar short term and medium term momentum. We will come back to this pair example later.

Figure 8: Examples of high correlation pairs

Country & Industry exposures

	Stock name	Country	Sector	Industry
	LAND SECURITIES GP	United Kingdom	Financials	Real Estate
Pair#1	BRITISH LAND CO	United Kingdom	Financials	Real Estate
	DIXONS CARPHONE	United Kingdom	consumer discretionary	Retailing
Pair#2	SPORTS DIRECT INTL	United Kingdom	consumer discretionary	Retailing
	DAIMLER AG	Germany	consumer discretionary	Automobiles & Components
Pair#3	BAYER MOTOREN WERK	Germany	consumer discretionary	Automobiles & Components

Style exposures



Source: Deutsche Bank, FactSet, Axioma,RavenPack

So how do we make use of the factor correlation to choose better pairs? The benchmark strategy requires country and sector neutrality, i.e. stocks forming a pair must come from the same country and sector group. Here we still keep the country neutrality requirement to avoid currency and cross-market microstructure issues. After an exhaustive match of pairs within the same country, we further refine the search universe by selecting those pairs that are highly correlated (top 5%) as implied by the risk model. This filtering not only ensures that the chosen pairs will share commonality of risk factor exposures more broadly, but also reduces our search universe significantly.

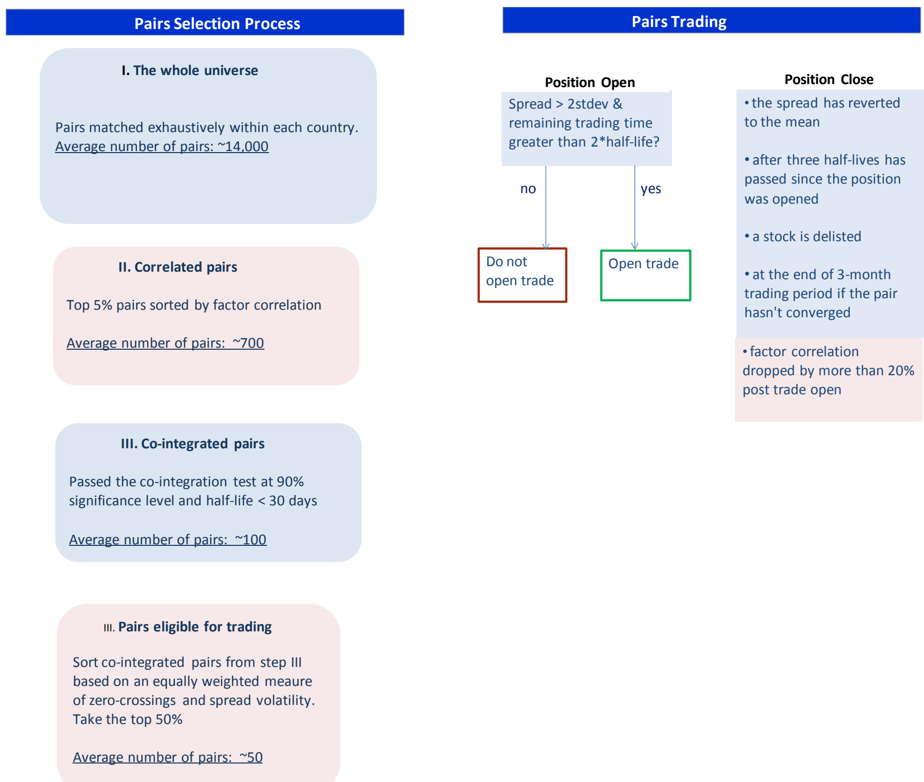
Now we have a method that matches pairs using fundamental similarities and co-integration properties. There are however other considerations when searching for profitable pairs. The performance of a given pairs trading strategy is a function of various factors which includes not only the degree of fungibility between the paired securities, but also the magnitude and frequency of mis-pricings. Profitable pairs trading strategies need to balance these considerations. For example, to be profitable, the strategy requires frequent reversal in the price spread, implying the need for paired stocks to oscillate around each other. More importantly, the expected return of a trade has the same order of magnitude as its absolute spread. This means that profitability is ultimately related to the volatility of the spread. The spread needs to be sizable enough for the strategy to make a profit, otherwise transaction costs could make the pair that frequently diverge and converge by small amounts unviable. We therefore further distinguish co-integrated pairs based on an equally weighted metric of the following to measures:

- The number of zero crossings of the spread (descending order)
- Volatility of the spread adjusted by trading cost (descending order)

We can further improve the stop-loss strategy by taking into account the change in factor correlations between a pair. If the divergence in prices between a pair is supported by a substantial decrease in their correlation, it shows that the pair is no longer exposed to similar risk factors. It is then more likely that the co-integrating relationship has broken. Therefore we add one more condition to the trading rules: close a position when the correlation between the pair has dropped by more than 20%.

Figure 9 illustrates the enhanced pairs selection and trading process. Any difference from the benchmark strategy is shaded in light pink.

Figure 9: The enhanced pairs selection and trading process

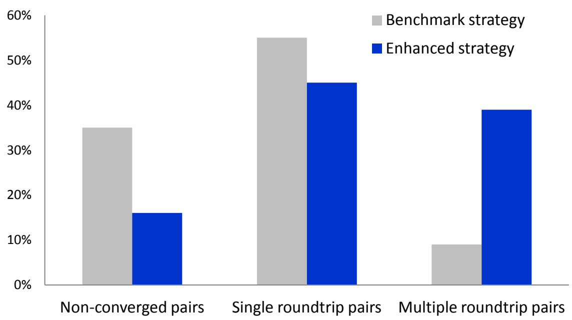


Source: Deutsche Bank, FactSet, Axioma,RavenPack

On average, the enhanced strategy based on the fundamental risk model trades 40 pairs per month. Figures 10 and 11 compare the performance of the enhanced strategy to the benchmark strategy. As is shown in Figure 10, the enhanced strategy significantly reduces the percentage of non-converged pairs, from over a third to 15%. In addition, the percentage of multiple roundtrip pairs also increased sharply, from 10% to 40%. There is also a substantial increase in the average return per pair, from 0.5% to over 2.3%, confirmed by significant p-value[9] from the one-sided t-test.

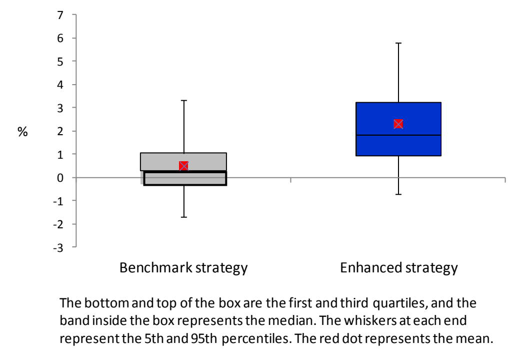
[9] <0.05

Figure 10: Breakdown of traded pairs



Source: Deutsche Bank, FactSet, Axioma,RavenPack

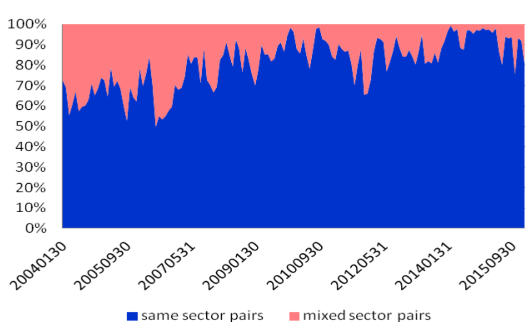
Figure 11: Profit distributions



Source: Deutsche Bank, FactSet, Axioma,RavenPack

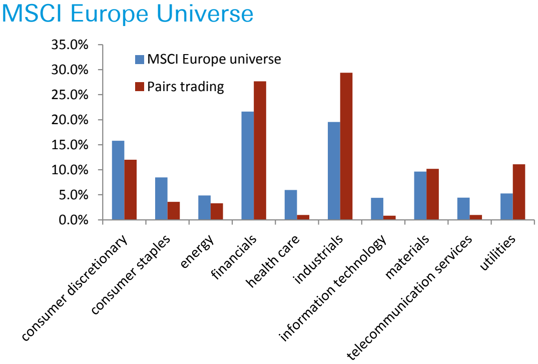
Figure 12 shows the decomposition of pairs selected by the enhanced strategy. On average, the majority (>80%) of the chosen pairs are same sector pairs, and the rest are mixed sector pairs. This confirms that sector affiliation is not always sufficient when searching for profitable pairs. The percentage of mixed sector pairs has been declining over time, from more than 30% in the early 2000 to less than 10% in recent years. Figure 13 shows the percentage difference in coverage between the MSCI Europe universe and the universe of pairs selected by the enhanced strategy. Industrials, financials and utilities sectors appear to contribute to a disproportionate percentage of the chosen pairs. This is not surprising as these sectors involve lines of business activities that are fairly uniform and the cross-sectional differences in factor exposures in these sectors are relatively small.

Figure 12: Pairs decomposition



Source: Deutsche Bank, FactSet, Axioma,RavenPack

Figure 13: Sector coverage of the new pairs strategy and the MSCI Europe Universe



Source: Deutsche Bank, FactSet, Axioma,RavenPack

Results suggest that imposing a fundamental risk model on top of a purely technical co-integration model improves the overall return and risk characteristics of a pairs trading strategy considerably. The fundamental risk model can be employed to facilitate decisions not only on pairs selection, but also about trade exit. It also makes the strategy operationally simpler by tracking much fewer pairs each month. Some pair-traders perform this fundamental filter manually, one way or another, based on their understanding of the constituent companies' economic situation.

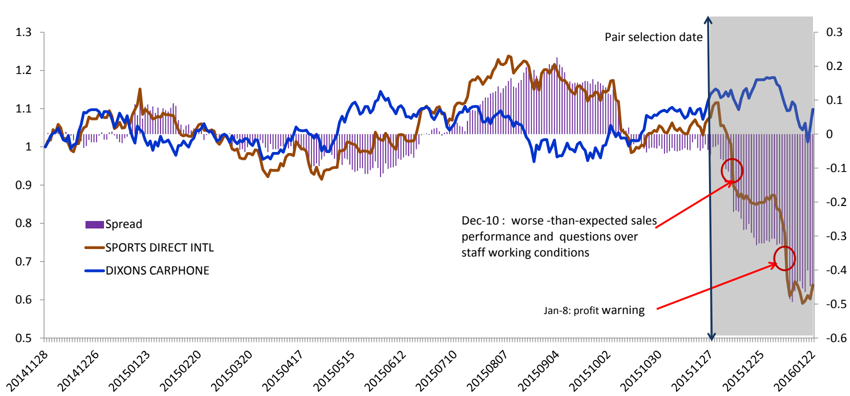
In the next section, we show how news analytics can be used as an overlay to the strategy to differentiate between "good" divergence and "bad" divergence, so as to improve the performance even further.

Applying A News Analytics Overlay to Pairs Trading

The enhanced strategy has clearly boosted performance by choosing more "sensible" pairs. However, this is only one part of the jigsaw for a profitable pairs trading strategy. A pair of stocks with similar fundamental exposures is more likely to move in tandem in the near future, but there is no guarantee for such behavior as for any single stock a large proportion of the price movement is driven by idiosyncratic risk. This can be best illustrated by the Dixon Carphone and Sports Direct Intl pair example mentioned earlier.

The pair was selected by the enhanced strategy on 30-Nov-2015. As is shown in Figure 8, the pair of UK retail stocks is one of the top correlated pairs in the universe. The Johansen test indicates the pair is co-integrated with 95% probability. In addition, the spread of the pair has moderate levels of zero-crossings and volatility over the formation period (Figure 14). If history was going to repeat itself, trading the pair in the subsequent months would have generated a profit. On 10-Dec-2015 Sports Direct Intl slumped by 11%, triggering an entry signal for a pairs trade. This drop in price was the largest the company had suffered for nearly two years, and the sell-off was due to a double whammy of worse-than-expected sales performance and revelations over its pay and working conditions[10]. The share price of Sports Direct Intl dropped another 14% on 8-Jan-2016, after the retailer issued a profit warning. Overall, the Sports Direct Intl price dropped by more than 40% over a period of two months. Clearly, trading the pair would have realized a loss.

Figure 14: A pairs example: Sport Direct Intl and Dixons Carphone



Source: Deutsche Bank, FactSet, Axioma, RavenPack

[10] It was revealed that workers at the company are subject to an extraordinary regime of searches and surveillance and the company also pays below the minimum wage. The company was branded a "scar on British business" by the Institute of Directors, was rounded on by its own shareholders and opposition MPs demanded that the company be investigated by HMRC. [Source: Guardian].

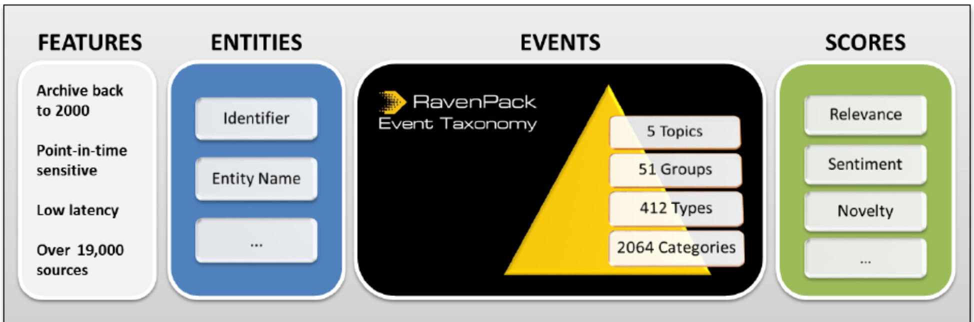
Pairs trading has two key ingredients: divergence and convergence. The profits and risks from trading stock pairs are very much related to the type of information event which creates divergence. As we saw in the Dixons and Sports Direct Intl pair example, there is a good chance that prices will diverge further if divergence is caused by a piece of news related specifically to one constituent of the pair. On the other hand, if divergence is caused by random price moments or a differential reaction rate to common information[11], convergence is more likely to follow after the initial divergence.

In this section, we show how news analytics can be used as an overlay to the strategy to differentiate "good" divergence from "bad" divergence, so as to improve the performance further. For this purpose we utilize the news and sentiment data from RavenPack News Analytics. RavenPack systematically tracks and analyzes information on over 40,000 companies from all major real-time newswires, online media, and other sources to produce real-time news analytics[12]. Its event taxonomy systematically tags news with high level topic codes down to granular event categories. It covers corporate events relating to both scheduled and unscheduled news about companies such as layoffs, mergers and acquisitions, product releases, analyst guidance, and earnings. For any detected news event, RavenPack generates a set of scores, rating different aspects of the event in relation to the entity in a matter of milliseconds. Among them we highlight Relevance, Event Novelty Scores (ENS) and Event Sentiment Scores (ESS). A Relevance score takes values between 0 and100 and indicates how strongly related an entity is to the underlying news story, with higher values indicating greater relevance. An ESS is a granular score between 0-100 that represents the news sentiment for a given entity by measuring various proxies sampled from the news. An ENS is a score between 0-100 that represents how novel a news story is within a 24-hour time window across all news stories. (see Appendix B for detailed definitions). Finally, Figure 15 presents a schematic view of RavenPack's News Analytics data.

[11] For example, two energy companies might react differently to a piece of news regarding the energy sector.

[12] It covers over 98% of the investable global market. Coverage by region: Americas 45%, Asia 31%, Europe 19%, Oceania 4%, Africa 1%. [Source:Ravenpack].

Figure 15: A schematic view of RavenPack's News Analytics



Timestamp	RavenPack ID	Entity Name	Topic	Group	Type	Category	Relevance	ESS	ENS	G_ENS
8/24/2015 11:41	147C38	Southern Co.	business	acquisitions-mergers	acquisition	acquisition-acquirer	100	49	24	1
8/24/2015 11:41	87EAB7	AGL Resources Inc.	business	acquisitions-mergers	acquisition	acquisition-acquiree	100	76	24	1
8/25/2015 17:10	5F2FF7	Delta Air Lines Inc.	business	products-services	product-release	product-release	100	64	100	100
8/27/2015 11:02	E08AF3	CenturyLink Inc.	business	labor-issues	layoffs	layoffs	100	43	24	24

Source: RavenPack

To test the effects of news on a pairs trading strategy, we use two aggregated indicators derived from the RavenPack news analytics data that measure sentiment and media attention. Sentiment provides directional opinion, and media attention focuses on whether a company receives more news than expected. These are useful not only for the prediction of future volatility or trading liquidity, but also for sentiment interpretation (Hafez et. al. 2015).

The average daily sentiment indicator (*AS_1D*) is a numeric value between -1 and +1 representing the average sentiment strength of a company over the previous 24 hours. A value of -1 is highly negative and a value of +1 is highly positive, whereas a value of 0 is neutral. Only novel and highly relevant news items that have non-neutral sentiment score are included in the computation[13]:

$$AS\_1D_t = \sum_{i=1}^{n} \frac{sent_i}{n}, \qquad (5)$$

where $sent_i = (ESS_i - 50)/50, i \in U$, and $U = \{1,\ldots,n\}$ is the number of news events with Relevance =100 , ENS=100 and ESS !=50.

Abnormal news volume (*ANV_1D*) captures how different the actual news volume over the past 24 hours is when compared to the normal news volume for the company over the last 365 days:

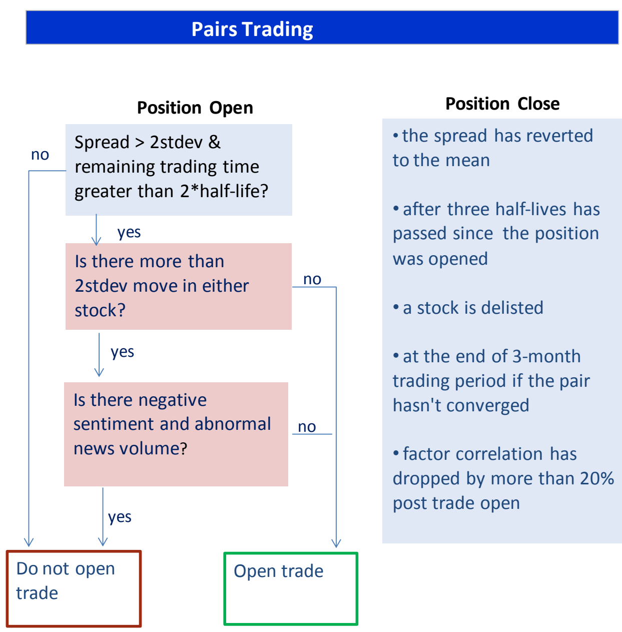
$$ANV\_1D_t = \frac{News\_Volume\_1D - avg\_365D(News\_Volume\_1D)}{stdev\_365D(News\_Volume\_1D)} \qquad (6)$$

A positive number represents the number of standard deviations the volume is above average, whereas a negative number represents the number of standard deviations the volume is below average. While a number of 0 means the news volume is the same as average.

[13] Certain type of news stories categorized as "order imbalance", "insider trading" and "technical analysis" are excluded as they tend to add noise given their lack of sentiment, high volume and frequency.

One of the prominent asymmetries in news analytics is the price impact of positive v.s. negative news. Several studies from Ravenpack have shown that market reaction to negative news is generally stronger than the reaction to positive news, and this phenomenon is consistent over time, across different types of signals. In addition, negative news also tends to decay more slowly than positive news (Hafez 2015). Therefore, we insert a step in the pairs trading process that checks if divergence of the pair can be attributed to negative news and sentiment. Specifically, upon divergence we first check to see if there is "abnormal" return[14] in either of the stocks. If either stock in the pair has an abnormal return, we check if there is negative sentiment (AS_1D <0) together with abnormal news volume (ANV_1D>2) supporting the price shock. If there is negative sentiment and abnormal news volume supporting the divergence, we do not initiate the trade. Figure 16 illustrates the pairs trading process with news overlay. Again any difference from the enhanced strategy is shaded in light pink.

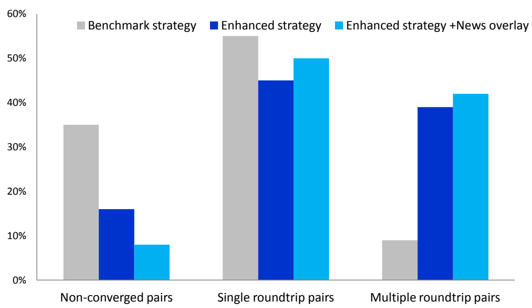
Figure 16: The new trading process with news overlay



Source: Deutsche Bank, Factset, Axioma, Ravenpack

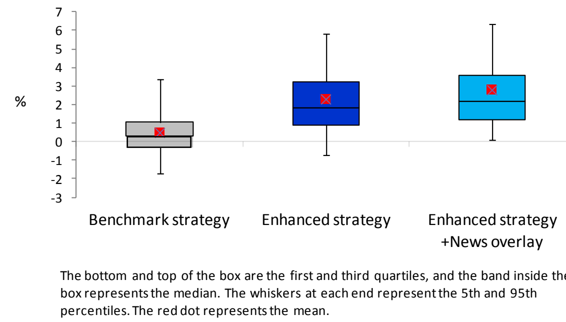
[14] We define "abnormal" return as the absolute return of the stock on the "open" day greater than two standard deviations.

Figure 17: Breakdown of traded pairs: MSCI Europe



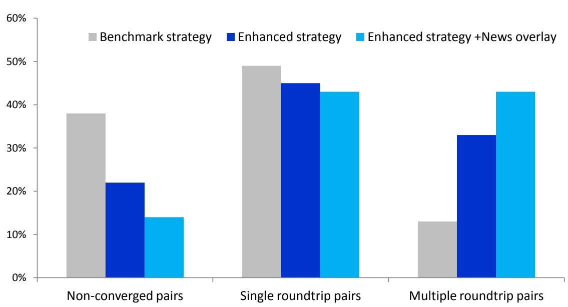
Source: Deutsche Bank, FactSet, Axioma,RavenPack

Figure 18: Profit Distributions: MSCI Europe



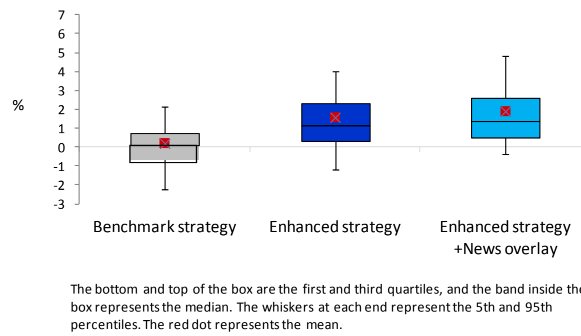
Source: Deutsche Bank, FactSet, Axioma,RavenPack

Figure 19: Breakdown of traded pairs: MSCI U.S.



Source: Deutsche Bank, FactSet, Axioma,RavenPack

Figure 20: Profit Distributions: MSCI U.S.



Source: Deutsche Bank, FactSet, Axioma,RavenPack

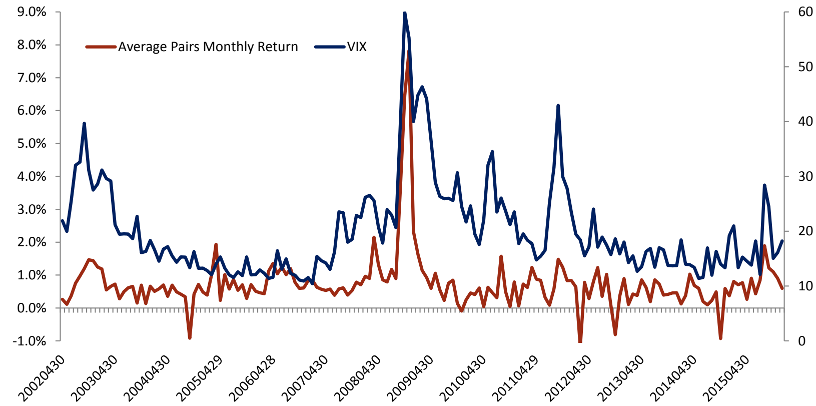
As can be seen from Figures 17 and 18, using the Ravenpack new analytics overlay further improves the strategy's performance significantly. First, the strategy with news overlay suffers much lower divergence risk: the percentage of non-converged pairs dropped by over a half from 15% to 7%. In addition, average profit per pair also increased from 2.3% to 2.8%, and the return distribution becomes more positively skewed. The increase in average returns is confirmed by significant p-values (<0.05) from the one-sided pair-wise t-test.

Figures 19 and 20 show the results of the pairs strategies applied on the MSCI U.S. universe. As can be seen from the graphs, the same conclusions can be reached, albeit the strategies have relatively lower returns in the U.S.. The average return per pair under the benchmark strategy, the enhanced strategy using the risk model, and the final strategy with both risk model and news overlay are 0.2%, 1.6% and 1.9% respectively. The improvement in returns is again supported by significant p-values from the one-sided pair-wise t-test.

Risk Exposures of the Final Strategy

Figure 21 shows the relationship between the average monthly return of the final pairs strategy versus the VIX. It is clear from the graph that there's a positive relationship between pairs trading profitability and market volatility. In line with the properties of mean-reversion strategies, pairs trading tend to work better in periods of high volatility.

Figure 21: Average pairs return versus market volatility



Source: Deutsche Bank, FactSet, Axioma,RavenPack

Like other mean-reversion strategies, pairs trading works better under higher market volatilities..

Selecting pairs using a fundamental risk model further reduces the strategy's exposure to common risk factors.

Finally, the table in Figure 22 shows the systematic risk exposures of various pairs trading strategies discussed in this report. We regress the average monthly returns of the benchmark and the final pairs strategy against the three Fama-French factors together with momentum and reversal factors. Because pairs strategies are market-neutral, the exposures to the market are small and statistically insignificant. Exposures to the size, value and momentum factors are also insignificant. As expected, some of the winner stocks that a pairs strategy shorts are short-term winners, and some of the loser stocks that a pairs strategy buys are short-term losers, and therefore a pairs strategy is positively exposed to the reversal factor.

Overall only a small portion of the returns from pairs trading strategies can be attributed to the five risk factors. The five-factors together explain 11% of the returns from the benchmark pairs strategy, and less than 2% for the final pairs strategy. This is understandable as stocks are matched based on their fundamental exposures under the final strategy. So essentially the strategy is buying and selling two stocks with similar exposures, bringing the net exposure close to zero. The results indicate that pairs trading offers an uncorrelated source of alpha.

Figure 22: Systematic risk of pairs trading strategies

	Benchmark Strategy	Final Strategy
Intercept	0.001	0.009
	(2.95)	(4.66)
Market	-0.03	-0.01
	(-0.92)	(-0.57)
Size	0.06	0.02
	(1.1)	(0.72)
Value	0.05	0.02
	(1.2)	(0.65)
Momentum	-0.06	-0.02
	(-1.38)	(-0.63)
Reversal	0.07	0.03
	(2.53)	(1.25)
R-sq	11%	2%

Market: Equally-weighted MSCI Europe
Momentum: First 11 months
Reversal: 1 month

Source: Deutsche Bank, FactSet, Axioma,RavenPack

An Addendum to Stock Pairs

So far, we have discussed the classical formulation of the pairs strategy which looks at the mispricing of two stocks. However, there is a much wider universe of possibilities present outside stock pairs. For example, one could also exploit the relative mis-pricings of a basket of co-integrated stocks. Trading baskets has the benefit of capturing additional mean-reverting relationships not captured by pairs. For example, stocks A, B and C may be co-integrated as a triplet with a mean-reverting spread, but there is a chance that none of the pair-wise combinations (A,B), (AC) or (BC) is co-integrated. Furthermore, the constituents of a basket do not have to be confined to single stocks. Abundant opportunities can be found in trading a basket of indices or trading indices against their component stocks.

However, when extending the pairs framework to baskets it is not clear how many and which securities to include in a basket. Should we trade triplets, quadruplets or quintuplets, etc.? One naive approach would be to conduct an exhaustive search within some defined universe and run co-integration tests on all possible combinations. The problem with this approach is two-fold. First, the size of search grows roughly at the magnitude of O(N^K), where K is the size of the basket and N is the number of assets, which gets computationally infeasible as the number of stocks grows. Second, as we saw in the pairs case, a co-integration relationship often breaks down out-of-sample if there are no fundamental links between the paired assets.

In this section, we propose a clustering-based approach to systematically identify stocks as potential constituents of a basket. Specifically, we utilize the correlation matrix from the fundamental risk model as a measure of distance between different assets. We then apply hierarchical clustering and employ a dynamic tree-cut algorithm to automatically detect clusters of stocks that have similar risk exposures.

Hierarchical clustering starts with each data point assigned to its own cluster and iteratively merges the two closest clusters together until all the data belong to a single cluster. In order to decide which clusters should be combined, a measure of dissimilarity between sets of observations is required. For the purpose of our research we define dissimilarity as one minus the implied correlation from a fundamental risk model

$$dissimilarity = 1 - correlation, \qquad (7)$$

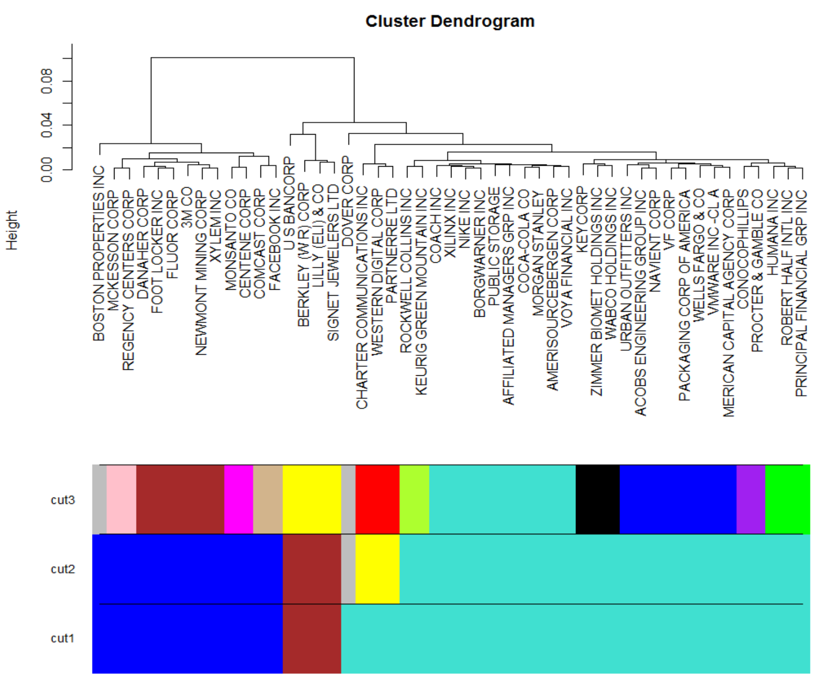
where correlation in equation (7) refers to the correlation matrix calculated from equation (4). We also adopt an average linkage method, in which the dissimilarity of two clusters A and B is the average pair-wise dissimilarities between all pairs of objects (a,b) such that $a \in A, b \in B$.

Hierarchical clustering methods produce a dendrogram which is a data structure containing information on which objects were merged at each step.

Clusters in a dendrogram are usually easy to tell visually. However, identifying the clusters in a systematic fashion can be difficult in practice[15]. In this report, we apply a dynamic branch cutting approach for detecting clusters in a dendrogram based on their shape. The technique has the following advantages: (1) it is capable of identifying nested clusters; (2) it is flexible: cluster shape parameters can be tuned to suit the application at hand; (3) it is suitable for automation. For technical details of this algorithm refer to Langfelder et. al (2009).

Figure 23 shows the clustering analysis performed on the constituents of energy sector ETF XLE. The ETF XLE is composed of some 40 stocks as of December 2014. The constituent stocks are clustered based on how fundamentally related they are, as implied by the fundamental risk model[16]. Clusters detected are shown by the rows of colors below the dendrogram, where each cluster is assigned a colour and unassigned objects are colored grey. Clusters of different sizes can be generated by the dynamic tree cut algorithm[17]. For example, under a loose tree cut algorithm ("cut1") we obtain two large clusters (blue and turquoise) whereas under a more stringent tree cut algorithm ("cut3") smaller clusters can be generated.

Figure 23: Cluster analysis on the constituent stocks of sector ETF XLE



Source: Deutsche Bank, FactSet, Axioma,RavenPack

[15] The most widely used method to identify clusters in a dendrogram is the fixed height branch cut. The user chooses a fixed height on the dendrogram and each contiguous branch of objects below that height is considered a separate cluster. However, the fixed tree cut technique is not ideal in situations where one expects a complicated dendrogram structure. Often dendograms exhibit distinct branches corresponding to the desired modules, but no single fixed cut height can identify them correctly.

[16] The dissimilarity matrix used in hierarchical clustering is calculated by using Axioma's US short horizon fundamental risk model.

[17] The size of each basket can also be controlled by tuning the shape parameters in the dynamic tree cut algorithm.

For each identified cluster the Johansen test of co-integration can be applied. When the size of a cluster is greater than two, more than one co-integration relationship may be found, in which case one could take the most significant co-integration relationship. Clusters that have passed the co-integration test can be traded as baskets. The hedging factors and the basket spread can be calculated from equations (A3) and (A5). The same trading rules can be followed as in the pairs case.

Though trading stock baskets has the benefit of capturing additional mean-reverting opportunities, there are a number of issues. First we find the average return of a basket generally decreases as the size of the basket increases. This may be due to the fact that specific risk is more diversified away by holding a larger number of assets, and therefore basket spreads are less volatile, leading to lower realized returns. There is also added operational difficulty in tracking the fundamental and stock-specific changes on the constituents of a basket, as well as in trade execution. As a result, it may be more preferable to trade a basket of indices than single stocks. The fundamental economics of an index change much more slowly than that of a single company, and the co-integration relationship between multiple indices is likely to last longer.

In addition to trading stock baskets, one could also trade an index against its constituent stocks. This falls into the category of index arbitrage strategies. An index arbitrage strategy trades on the difference in value between a portfolio of stocks constituting an index and the futures on that index. If we include all constituent stocks in the portfolio, then the market value of the portfolio will co-integrate very tightly with the index future contract, leaving little room for arbitrage opportunities. In order to increase the profit margin, one could pick a smaller subset of stocks from the constituents of an index to construct a basket to trade against the index itself.

We demonstrate the method by using the same energy sector EFT XLE example. The goal is to pick a smaller subset of these stocks to form a long basket to trade against XLE. The larger the basket, the better it would co-integrate with the XLE index, but the smaller the profit. Depending on the risk-reward preference, baskets of different sizes can be generated using the same clustering and dynamic tree cut method described earlier. Once baskets have been identified (Figure 23), we can from long-only portfolios[18] for each cluster and test if the portfolio is co-integrated with XLE.

This is done by a linear regression using the log price of XLE as the regressand and the log prices of stocks as the regressors. The coefficients of the linear regression, i.e., the hedge ratios can be estimated using maximum likelihood estimation with constraints that all hedge ratios must be positive. We then test if the residuals from the regression are stationary. As an example, we formed a portfolio of stocks from the blue cluster generated from ‘cut1’ as shown in Figure 23. The basket contains 13 stocks and the spread between the long-only portfolio and the XLE index is found to be stationary with 95% probability. Trading the portfolio against the XLE index generated a return of 4.5% and a Sharpe ratio of 1.3, from December 2014 to December 2015.

[18] If we have short positions in the stock portfolio and a short index position simultaneously, we would be double-shorting some stocks even when we are long the stock portfolio, which increases specific risks.

Conclusions

Existing pairs trading schemes almost exclusively rely on time series analysis of historical prices in combination with sector affiliation to identify tradable pairs. We argue that sector grouping is only one dimension of risk. A desirable pairs trading scheme should start by defining fundamentally homogeneous asset groups, within which statistical measures derived from historical prices can then be used to rank and select pairs. We believe the best way of defining such groups is by using a comprehensive fundamental risk model. Additionally, the profitability and risk of a pairs strategy is related to the type of information event driving the divergence. We propose a new pairs trading model by utilizing a fundamental risk model and with a news analytics overlay, which not only reduces the divergence risk significantly, but also boosts the average returns per pair.

A classical pairs strategy profit from the mean reversion properties of the spread between two stocks. In reality, one could exploit the co-integration properties of a basket of stocks. However, identifying such a basket is not a trivial task. We propose a novel way to identify potential baskets, with the help of the fundamental risk model and a smart clustering scheme.

As financial markets continue to become more integrated, opportunities abound for multiple-asset pairs in the realms of cash equities, FX, futures, options, and other derivative instruments. Also, the demand for multiple-leg pairs has been increasing as more complex investment models are being developed. Last but not least, due to decreasing profit margin and the challenges present in trading stock pairs, there has been a large shift for the strategy to be applied intraday in order to capture the best prices and avoid changes in fundamental and specific exposures which plague longer-term positions. Trading pairs intraday is subject to market microstructure issues and requires much smarter execution algorithms. These emerging trends represent the new frontier of the next generation pairs algorithms.

References

Alexandar, C. and A. Dimitriu "The Co-integration Alpha: Enhanced Index Tracking and Long-Short Equity Market Neutral Strategies", *Discussion paper, ISMA center* (2002).

Cartea A. and S. Jaimungal "Algorithmic Trading of Co-integrated Assets" *Working Paper* (2015)

Chan E. *Algorithmic Trading: Winning Strategies and their Rationale* (2013), (Wiley: New Jersey).

Chan L.K.C., J. Lakonishok and B. Swaminathan "Industry Classification and Return Co-movement". *Financial Analysts Journal* (2007), 63, 56-70.

Do,B., R. Faff "Does Simple Pairs Trading Still Work?", *Financial Analysts Journal* (2010), 66, 83-95.

Do,B., R. Faff "Are Pairs Trading Profits Robust to Trading Costs?", *Journal of Financial Research* (2012), 35, 261-287.

Elliott R., J. V. D. Hoek and Malcolms W. "Pairs Trading" *Quantitative Finance* (2005), 5(3), 271-276

Engelberg, J., P. Gao and R. Jagannathan "An Anatomy of Pairs Trading: the Role of Idiosyncratic News, Common Information and Liquidity", *Working Paper* (2008)

Engel R. F. and C.W.J. Granger "Co-integration and Error Correction: Representation, Estimation, and Testing" *Econometrica* (1987) 55(2), 251-276

Hafez P. and J Xie "Enhancing Short-Term Stock Reversal Strategies with News Analytics" *RavenPack Research Paper* (2013)

Hafez P., J. A. Guerrero-Colon and S. Duprey "Thematic Alpha Streams Improve Performance of Equity Portfolios" *RavenPack Research Paper* (2015)

Hafez P., J. A. Guerrero-Colon and S. Duprey "A Guide to Trading and Investment Applications Using News Analytics" *RavenPack Research Paper* (2015)

Galenko A. and E. Popova "Trading in the Presence of Co-integration" *The Journal of Alternative Investments* (2012) 15(1), 85-97

Gatev, E., W. Goetzmann and K. Rouwenhorst "Pairs Trading: Performance of a Relative-Value Arbitrage Rule", *NBER Working Paper* (1999).

Gatev, E., W. Goetzmann and K. Rouwenhorst "Pairs Trading: Performance of a Relative-Value Arbitrage Rule", *The Review of Financial Studies* (2006), 19, 798- 827.

Kittrell J. "Behavioral Trends and Market Neutrality", *RavenPack Research Paper* (2012).

Langfelder, P., B. Zhang and S. Horvath "Defining Clusters from a Hierarchical Cluster Tree: the Dynamic Tree Cut Package for R", *Bioinformatics* (2008), 24(5), 719-720.

Meucci A. "Review of Statistical Arbitrage, Co-integration, and Multivariate Ornstein-Uhlenbeck" *Working Paper* (2009)

Nath, P. "High Frequency Pairs Trading with U.S. Treasury Securities: Risks and Rewards for Hedge Funds" *LBS Working Paper* (2003)

Serge A., "Where Have All the Stat Arb Profits Gone?"*Columbia University Working Paper* (2008)

Vidyamurthy, G., *Pairs Trading, Quantitative Methods and Analysis* (2004), (Wiley: New York).

Zhao G. and K. Webster "Opportunistic Strategies during Turbulent Times", *Deutsche Bank Market Research report* (2015).

Appendix

Appendix A: Johansen Co-Integration Test

The Johansen method is the maximum likelihood estimator of the so-called reduced rank model. Consider a vector autoregressive (VAR) process for a *p*-dimensional vector

$$X_t = \sum_{i=1}^{k} A_i X_{t-i} + u_t \qquad \text{(A1)}$$

By using the difference operator Δ, this *VAR(k)* process can be transformed into a vector error correction model (VECM) as follows:

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Delta X_{t-i} + u_t \qquad \text{(A2)}$$

where the multiplier matrix Π can be decomposed into two matrices such that

$$\Pi = \alpha\beta' . \qquad \text{(A3)}$$

The vector or matrix β represents the co-integration vectors, and α is the matrix of error-correction coefficients which measure the rate each variable adjusts to the long-run equilibrium. The number of co-integrating vectors are identical to the number of stationary relationships in the Π -matrix. Mathematically, the rank of Π determines the number of independent rows in Π, and therefore also the number of co-integrating vectors .The rank of Π is given by the number of significant eigenvalues found in Π. Each significant eigenvalue represent a stationary relationship.

For example, if the number of variables p is three and there are two co-integrating vectors, then the matrix of co-integration vector β is 3x2 and the matrix of error-correction coefficient α is also 3x2. The coefficients in the co-integration vector β' multiplied by the variables X gives a set of linear combinations that are stationary, that is:

$$X'\beta = \begin{pmatrix} X_1 & X_2 & X_3 \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \\ \beta_{13} & \beta_{23} \end{pmatrix} = \left(\sum_{i=1}^{3} X_i \beta_{1i}, \quad \sum_{i=1}^{3} X_i \beta_{2i} \right) \qquad \text{(A4)}$$

Originally Johansen derived two tests, the maximum eigenvalue test and the trace test[19]. The maximum eigenvalue statistic tests the null hypothesis of *r* co-integrating vectors against the alternative of *r+1* vectors. The trace test on the other hand tests the null hypothesis of at most *r* co-integrating relations against a general alternative. Critical values of these tests have been calculated by Johansen et. al.(1990). Once the co-integration vectors are identified, the spread from a particular co-integrated relationship can be calculated as

$$u_j = X'\beta_j \qquad \text{(A5)}$$

---

[19] It has been found that the race test is a better test, since it appears to be more robust to skewness and excess kurtosis. In addition, the trace test can be adjusted for degrees of freedom, which can be of importance in small samples.

Appendix B: Ravenpack News Analytics Definition

A *Relevance* score is between 0-100 that indicates how strongly related the entity is to the underlying news story, with higher values indicating greater relevance. A score of 0 means the entity was passively mentioned while a score of 100 indicates that the entity identified plays a key role in the news story and is considered highly relevant.

An *Event Sentiment Score* (ESS) is a granular score between 0 and 100 that represents the news sentiment for a given entity by measuring various proxies sampled from the news. The score is determined by systematically matching stories typically categorized by financial experts as having short-term positive or negative financial or economic impact. The strength of the score is derived from a collection of surveys where financial experts rated entity-specific events as conveying positive or negative sentiment and to what degree. Their ratings are encapsulated in an algorithm that generates a score ranging from 0-100 where 50 indicates neutral sentiment, values above 50 indicate positive sentiment and values below 50 show negative sentiment.

An *Event Novelty score* (ENS) is a score between 0 and 100 that represents how "new" or novel a news story is within a 24-hour time window across all news stories . Any two stories that match the same event for the same companies will be considered similar according to ENS. The first story reporting a categorized event about one or more entities is considered to be the most novel and receives a score of 100. Subsequent stories about the same event for the same companies receive scores following a decay function based on the number of stories in the past 24- hour window.

Appendix 1

Important Disclosures

Additional information available upon request

*Prices are current as of the end of the previous trading session unless otherwise indicated and are sourced from local exchanges via Reuters, Bloomberg and other vendors . Other information is sourced from Deutsche Bank, subject companies, and other sources. For disclosures pertaining to recommendations or estimates made on securities other than the primary subject of this research, please see the most recently published company report or visit our global disclosure look-up page on our website at http://gm.db.com/ger/disclosure/DisclosureDirectory.eqsr

Analyst Certification

The views expressed in this report accurately reflect the personal views of the undersigned lead analyst(s). In addition, the undersigned lead analyst(s) has not and will not receive any compensation for providing a specific recommendation or view in this report. Shuo Qu

Hypothetical Disclaimer

Backtested, hypothetical or simulated performance results have inherent limitations. Unlike an actual performance record based on trading actual client portfolios, simulated results are achieved by means of the retroactive application of a backtested model itself designed with the benefit of hindsight. Taking into account historical events the backtesting of performance also differs from actual account performance because an actual investment strategy may be adjusted any time, for any reason, including a response to material, economic or market factors. The backtested performance includes hypothetical results that do not reflect the reinvestment of dividends and other earnings or the deduction of advisory fees, brokerage or other commissions, and any other expenses that a client would have paid or actually paid. No representation is made that any trading strategy or account will or is likely to achieve profits or losses similar to those shown. Alternative modeling techniques or assumptions might produce significantly different results and prove to be more appropriate. Past hypothetical backtest results are neither an indicator nor guarantee of future returns. Actual results will vary, perhaps materially, from the analysis.

Regulatory Disclosures

1.Important Additional Conflict Disclosures

Aside from within this report, important conflict disclosures can also be found at https://gm.db.com/equities under the "Disclosures Lookup" and "Legal" tabs. Investors are strongly encouraged to review this information before investing.

2.Short-Term Trade Ideas

Deutsche Bank equity research analysts sometimes have shorter-term trade ideas (known as SOLAR ideas) that are consistent or inconsistent with Deutsche Bank's existing longer term ratings. These trade ideas can be found at the SOLAR link at http://gm.db.com.

Additional Information

The information and opinions in this report were prepared by Deutsche Bank AG or one of its affiliates (collectively "Deutsche Bank"). Though the information herein is believed to be reliable and has been obtained from public sources believed to be reliable, Deutsche Bank makes no representation as to its accuracy or completeness.

If you use the services of Deutsche Bank in connection with a purchase or sale of a security that is discussed in this report, or is included or discussed in another communication (oral or written) from a Deutsche Bank analyst, Deutsche Bank may act as principal for its own account or as agent for another person.

Deutsche Bank may consider this report in deciding to trade as principal. It may also engage in transactions, for its own account or with customers, in a manner inconsistent with the views taken in this research report. Others within Deutsche Bank, including strategists, sales staff and other analysts, may take views that are inconsistent with those taken in this research report. Deutsche Bank issues a variety of research products, including fundamental analysis, equity-linked analysis, quantitative analysis and trade ideas. Recommendations contained in one type of communication may differ from recommendations contained in others, whether as a result of differing time horizons, methodologies or otherwise. Deutsche Bank and/or its affiliates may also be holding debt securities of the issuers it writes on.

Analysts are paid in part based on the profitability of Deutsche Bank AG and its affiliates, which includes investment banking revenues.

Opinions, estimates and projections constitute the current judgment of the author as of the date of this report. They do not necessarily reflect the opinions of Deutsche Bank and are subject to change without notice. Deutsche Bank has no obligation to update, modify or amend this report or to otherwise notify a recipient thereof if any opinion, forecast or estimate contained herein changes or subsequently becomes inaccurate. This report is provided for informational purposes only. It is not an offer or a solicitation of an offer to buy or sell any financial instruments or to participate in any particular trading strategy. Target prices are inherently imprecise and a product of the analyst's judgment. The financial instruments discussed in this report may not be suitable for all investors and investors must make their own informed investment decisions. Prices and availability of financial instruments are subject to change without notice and investment transactions can lead to losses as a result of price fluctuations and other factors. If a financial instrument is denominated in a currency other than an investor's currency, a change in exchange rates may adversely affect the investment. Past performance is not necessarily indicative of future results. Unless otherwise indicated, prices are current as of the end of the previous trading session, and are sourced from local exchanges via Reuters, Bloomberg and other vendors. Data is sourced from Deutsche Bank, subject companies, and in some cases, other parties.

Macroeconomic fluctuations often account for most of the risks associated with exposures to instruments that promise to pay fixed or variable interest rates. For an investor who is long fixed rate instruments (thus receiving these cash flows), increases in interest rates naturally lift the discount factors applied to the expected cash flows and thus cause a loss. The longer the maturity of a certain cash flow and the higher the move in the discount factor, the higher will be the loss. Upside surprises in inflation, fiscal funding needs, and FX depreciation rates are among the most common adverse macroeconomic shocks to receivers. But counterparty exposure, issuer creditworthiness, client segmentation, regulation (including changes in assets holding limits for different types of investors), changes in tax policies, currency convertibility (which may constrain currency conversion, repatriation of profits and/or the liquidation of positions), and settlement issues related to local clearing houses are also important risk factors to be considered. The sensitivity of fixed income instruments to macroeconomic shocks may be mitigated by indexing the contracted cash flows to inflation, to FX depreciation, or to specified interest rates – these are common in emerging markets. It is important to note that the index fixings may -- by construction -- lag or mis-measure the actual move in the underlying variables they are intended to track. The choice of the proper fixing (or metric) is particularly important in swaps markets, where floating coupon rates (i.e., coupons indexed to a typically short-dated interest rate reference index) are exchanged for fixed coupons. It is also important to acknowledge that funding in a currency that differs from the currency in which coupons are denominated carries FX risk. Naturally, options on swaps (swaptions) also bear the risks typical to options in addition to the risks related to rates movements.

Derivative transactions involve numerous risks including, among others, market, counterparty default and illiquidity risk. The appropriateness or otherwise of these products for use by investors is dependent on the investors' own circumstances including their tax position, their regulatory environment and the nature of their other assets and liabilities, and as such, investors should take expert legal and financial advice before entering into any transaction similar to or inspired by the contents of this publication. The risk of loss in futures trading and options, foreign or domestic, can be substantial. As a result of the high degree of leverage obtainable in futures and options trading, losses may be incurred that are greater than the amount of funds initially deposited. Trading in options involves risk and is not suitable for all investors. Prior to buying or selling an option investors must review the "Characteristics and Risks of Standardized Options", at http://www.optionsclearing.com/about/publications/character-risks.jsp. If you are unable to access the website please contact your Deutsche Bank representative for a copy of this important document.

Participants in foreign exchange transactions may incur risks arising from several factors, including the following: (i) exchange rates can be volatile and are subject to large fluctuations; (ii) the value of currencies may be affected by numerous market factors, including world and national economic, political and regulatory events, events in equity and debt markets and changes in interest rates; and (iii) currencies may be subject to devaluation or government imposed exchange controls which could affect the value of the currency. Investors in securities such as ADRs, whose values are affected by the currency of an underlying security, effectively assume currency risk.
Unless governing law provides otherwise, all transactions should be executed through the Deutsche Bank entity in the investor's home jurisdiction.

United States: Approved and/or distributed by Deutsche Bank Securities Incorporated, a member of FINRA, NFA and SIPC. Analysts employed by non-US affiliates may not be associated persons of Deutsche Bank Securities Incorporated and therefore not subject to FINRA regulations concerning communications with subject companies, public appearances and securities held by analysts.

Germany: Approved and/or distributed by Deutsche Bank AG, a joint stock corporation with limited liability incorporated in the Federal Republic of Germany with its principal office in Frankfurt am Main. Deutsche Bank AG is authorized under German Banking Law (competent authority: European Central Bank) and is subject to supervision by the European Central Bank and by BaFin, Germany's Federal Financial Supervisory Authority.

United Kingdom: Approved and/or distributed by Deutsche Bank AG acting through its London Branch at Winchester House, 1 Great Winchester Street, London EC2N 2DB. Deutsche Bank AG in the United Kingdom is authorised by the Prudential Regulation Authority and is subject to limited regulation by the Prudential Regulation Authority and Financial Conduct Authority. Details about the extent of our authorisation and regulation are available on request.

Hong Kong: Distributed by Deutsche Bank AG, Hong Kong Branch.

India: Prepared by Deutsche Equities Private Ltd, which is registered by the Securities and Exchange Board of India (SEBI) as a stock broker. Research Analyst SEBI Registration Number is INH000001741. DEIPL may have received administrative warnings from the SEBI for breaches of Indian regulations.

Japan: Approved and/or distributed by Deutsche Securities Inc.(DSI). Registration number - Registered as a financial instruments dealer by the Head of the Kanto Local Finance Bureau (Kinsho) No. 117. Member of associations: JSDA, Type II Financial Instruments Firms Association and The Financial Futures Association of Japan. Commissions and risks involved in stock transactions - for stock transactions, we charge stock commissions and consumption tax by multiplying the transaction amount by the commission rate agreed with each customer. Stock transactions can lead to losses as a result of share price fluctuations and other factors. Transactions in foreign stocks can lead to additional losses stemming from foreign exchange fluctuations. We may also charge commissions and fees for certain categories of investment advice, products and services. Recommended investment strategies, products and services carry the risk of losses to principal and other losses as a result of changes in market and/or economic trends, and/or fluctuations in market value. Before deciding on the purchase of financial products and/or services, customers should carefully read the relevant disclosures, prospectuses and other documentation. "Moody's", "Standard & Poor's", and "Fitch" mentioned in this report are not registered credit rating agencies in Japan unless Japan or "Nippon" is specifically designated in the name of the entity. Reports on Japanese listed companies not written by analysts of DSI are written by Deutsche Bank Group's analysts with the coverage companies specified by DSI. Some of the foreign securities stated on this report are not disclosed according to the Financial Instruments and Exchange Law of Japan.

Korea: Distributed by Deutsche Securities Korea Co.

South Africa: Deutsche Bank AG Johannesburg is incorporated in the Federal Republic of Germany (Branch Register Number in South Africa: 1998/003298/10).

Singapore: by Deutsche Bank AG, Singapore Branch or Deutsche Securities Asia Limited, Singapore Branch (One Raffles Quay #18-00 South Tower Singapore 048583, +65 6423 8001), which may be contacted in respect of any matters arising from, or in connection with, this report. Where this report is issued or promulgated in Singapore to a person who is not an accredited investor, expert investor or institutional investor (as defined in the applicable Singapore laws and regulations), they accept legal responsibility to such person for its contents.

Qatar: Deutsche Bank AG in the Qatar Financial Centre (registered no. 00032) is regulated by the Qatar Financial Centre Regulatory Authority. Deutsche Bank AG - QFC Branch may only undertake the financial services activities that fall within the scope of its existing QFCRA license. Principal place of business in the QFC: Qatar Financial Centre, Tower, West Bay, Level 5, PO Box 14928, Doha, Qatar. This information has been distributed by Deutsche Bank AG. Related financial products or services are only available to Business Customers, as defined by the Qatar Financial Centre Regulatory Authority.

Russia: This information, interpretation and opinions submitted herein are not in the context of, and do not constitute, any appraisal or evaluation activity requiring a license in the Russian Federation.

Kingdom of Saudi Arabia: Deutsche Securities Saudi Arabia LLC Company, (registered no. 07073-37) is regulated by the Capital Market Authority. Deutsche Securities Saudi Arabia may only undertake the financial services activities that fall within the scope of its existing CMA license. Principal place of business in Saudi Arabia: King Fahad Road, Al Olaya District, P.O. Box 301809, Faisaliah Tower - 17th Floor, 11372 Riyadh, Saudi Arabia.

United Arab Emirates: Deutsche Bank AG in the Dubai International Financial Centre (registered no. 00045) is regulated by the Dubai Financial Services Authority. Deutsche Bank AG - DIFC Branch may only undertake the financial services activities that fall within the scope of its existing DFSA license. Principal place of business in the DIFC: Dubai International Financial Centre, The Gate Village, Building 5, PO Box 504902, Dubai, U.A.E. This information has been distributed by Deutsche Bank AG. Related financial products or services are only available to Professional Clients, as defined by the Dubai Financial Services Authority.

Australia: Retail clients should obtain a copy of a Product Disclosure Statement (PDS) relating to any financial product referred to in this report and consider the PDS before making any decision about whether to acquire the product. Please refer to Australian specific research disclosures and related information at https://australia.db.com/australia/content/research-information.html

Australia and New Zealand: This research, and any access to it, is intended only for "wholesale clients" within the meaning of the Australian Corporations Act and New Zealand Financial Advisors Act respectively.
Additional information relative to securities, other financial products or issuers discussed in this report is available upon request. This report may not be reproduced, distributed or published by any person for any purpose without Deutsche Bank's prior written consent. Please cite source when quoting.

David Folkerts-Landau
Chief Economist and Global Head of Research

Raj Hindocha Global Chief Operating Officer Research	Marcel Cassard Global Head FICC Research & Global Macro Economics	Steve Pollard Global Head Equity Research
Michael Spencer Regional Head Asia Pacific Research	Ralf Hoffmann Regional Head Deutsche Bank Research, Germany	Andreas Neubauer Regional Head Equity Research, Germany

International Locations

Deutsche Bank AG
Deutsche Bank Place
Level 16
Corner of Hunter & Phillip Streets
Sydney, NSW 2000
Australia
Tel: (61) 2 8258 1234

Deutsche Bank AG
Große Gallusstraße 10-14
60272 Frankfurt am Main
Germany
Tel: (49) 69 910 00

Deutsche Bank AG
Filiale Hongkong
International Commerce Centre,
1 Austin Road West,Kowloon,
Hong Kong
Tel: (852) 2203 8888

Deutsche Securities Inc.
2-11-1 Nagatacho
Sanno Park Tower
Chiyoda-ku, Tokyo 100-6171
Japan
Tel: (81) 3 5156 6770

Deutsche Bank AG London
1 Great Winchester Street
London EC2N 2EQ
United Kingdom
Tel: (44) 20 7545 8000

Deutsche Bank Securities Inc.
60 Wall Street
New York, NY 10005
United States of America
Tel: (1) 212 250 2500