Deutsche Bank
Markets Research

Global

Quantitative Strategy
Signal Processing

Date
8 September 2014

Seven Sins of Quantitative Investing

Sheng Wang
sheng.wang@db.com

Yin Luo, CFA
yin.luo@db.com

Miguel-A Alvarez
miguel-a.alvarez@db.com

Javed Jussa
javed.jussa@db.com

Allen Wang
allen-y.wang@db.com

Gaurav Rohal
gaurav.rohal@db.com

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

DB Quant Handbook, Part II

Quant investing is easy?

The rapid rise of computing power and wide availability of off-the-shelf backtesting software provided by many data vendors have given the impression that quant investing is easy, or is it?

Seven sins or biases in quantitative modeling

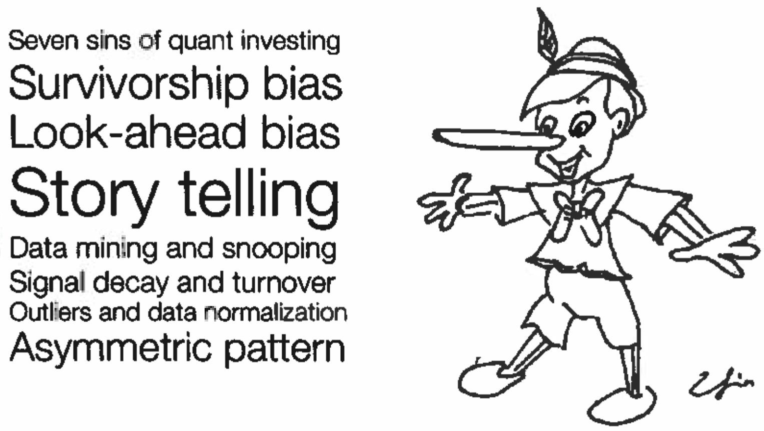
In this paper, we discuss the seven common mistakes investors tend to make when they perform backtesting and build quant models. Some of these may be familiar to our readers, but nonetheless, you may be surprised to see the impact of these biases. The other sins are so commonplace in both academia and practitioner's research that we usually take them for granted.

Unique features

There are a few unique features in this research that we have not seen in other places. We deliberate when to and when not to remove outliers; discuss various data normalization techniques; address the intricate issues of signal decay, turnover, and transaction costs; elaborate on the optimal rebalancing frequency; illustrate the asymmetric factor payoff patterns and the impact of short availability on portfolio performance; answer the question of "how many stocks should be held in the portfolio"; and review the tradeoffs of various factor weighting/portfolio construction techniques. Last but not least, we compare traditional active portfolio management via multi-factor models, with the new trend of smart beta/factor portfolio investing.

A hands-on tutorial on how to build a multi-factor model

Lastly, we use a real-life example to show how to avoid the seven sins when building multi-factor models and portfolios.



compare to survivor index?

Source: Yin Luo

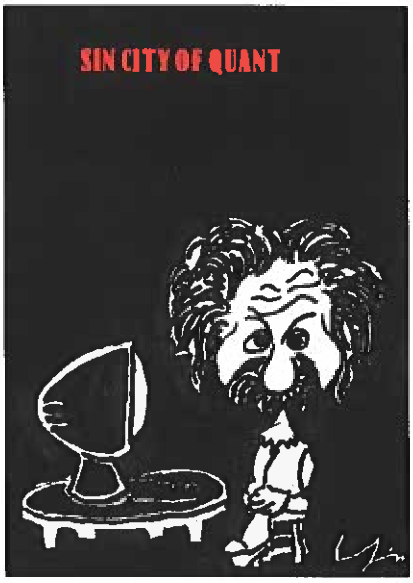
Table Of Contents

The sin city of quant

DB Quant Handbook, Part II

Building a fully automated investing system that makes money all the time has long been a dream of many investors, i.e., the Holy Grail of quant investing. With the rise and popularity of big data, computing power, and talent in recent years, it seems that it has become more and more a reality rather than just a dream.

Figure 1: Sin city of quant



Source: Yin Luo

Almost every major data vendor on the market (e.g., ClariFi, Capital IQ, Factset, Thomson Reuters, Bloomberg) has launched tools that make quantitative backtesting increasingly easier. Now, if you are willing to spend some modest amount, you can subscribe to anything from raw data (e.g., company financial statements, market data, etc.) to fancy news sentiment data feeds (which automatically quantify sentiment across all news stories and even web social media). Off-the-shelf software allows you to backtest every combination and permutation of possible investment strategies (or in quant jargon, factors), formulate multi-factor models, and construct investable portfolios. Everything from front office research, portfolio management, and trading to back office accounting and custodian services can be automated. Quant investing is no longer a game for a small group of elite mathematicians, computer scientists, and finance professors. Suddenly, building a quant model seems to be so easy that almost everybody can have it set up in no time. There is an also almost unlimited supply of talent – many universities are now offering MFE (master in financial engineering) degrees as well as other similar programs. The smartest young students are no longer joining physics departments; rather, they are pursuing a career in quantitative finance.

The reality, however, is not as rosy. Almost anyone who has actually managed assets would say that the real performance of a strategy is almost never the same as it is in backtesting. The summer of 2007 was one of the first warning signs. In a matter of two weeks, traditional factors like value and momentum lost so tremendously that investors called it a "six sigma" event (see Khandani and Lo [2007] and Cahan, et al [2012]). The ensuing 2008 global financial crisis, European sovereign debt crisis, and the frequent risk-on/risk-off regime shifts created more challenges for quant.

In this research, we discuss the seven most common mistakes investors tend to make when they build quant models or develop quantitative screening tools – collectively, we call these common biases in backtesting "the seven sins". This research paper forms Part II of our popular "DB Quant Handbook" series (see Luo, et al [2010a]), providing a general overview of quantitative modeling from the beginning to the end.

- Survivorship bias
- Look-ahead bias
- The sin of story telling
- Data mining and data snooping bias
- Signal decay and turnover
- Outliers – the story of spectacular successes and failures
- The asymmetric payoff pattern and shorting cost

Quants have traditionally believed their models were free from human behavioral biases and at times took advantage of these biases by trading against non-quant investors. What many quants do not realize, however, is that behavioral finance can be their friend or foe. Quants can equally suffer from the same behavior biases as any other investors.

Indeed, if we test enough factors and "torture" our data hard enough, we can almost always find something that works spectacularly in-sample. However, most of the time, these models would have no out-of-sample predictive power[1].

The problem in social science, like economics or finance, is that we only have one set of observations – unlike physical science, we typically cannot perform repeated experiments[2]. A strategy showing great performance could either be due to true skill or pure luck – the problem is that we can never really tell the difference with enough confidence. In other words, in economics and finance, we never have enough data.

In Luo, et al [2010a], we discuss the general principles of how to build a successful quantitative stock selection model. In this paper, we attempt to elaborate on some of the more intricate issues in outlier control, data normalization, signal decay/turnover/transaction costs, optimal rebalancing frequency, the asymmetric factor payoff patterns, short availability, and various factor weighting and portfolio construction techniques. We believe this discussion should be of great interest to quantitative investors, but also any investor who is interested in adding any systematic elements to his or her investment process.

You may not know, but many of your favorite data vendors and off-the-shelf backtesting software may suffer one or many of the sins discussed in this paper. You may believe these biases change the backtesting results somewhat, but would not fundamentally change your conclusion. You may be surprised. Please read on...

Regards,

Yin, Miguel, Javed, Sheng, Allen, and Gaurav
Deutsche Bank Quantitative Strategy

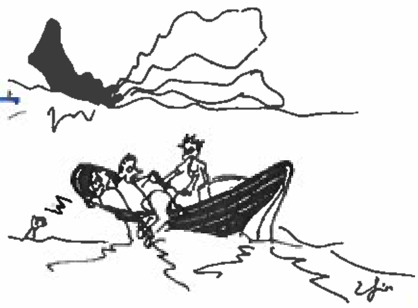
[1] The discrepancy between in-sample and out-of-sample performance is, of course, not all due to the biases in backtesting. As ideas become well known in the investment community, they are more likely to get arbitraged away (see McLean, et al [2014] on the impact of academic publications and subsequent alpha opportunities).

[2] There are some exceptions. For example, Forsythe, et al [1980], Plott and Sunder [1982], and Smith, et al [1988] lay out the foundation of so-called "experimental finance".

1. Survivorship bias

Survivorship bias is one of the common mistakes investors tend to make. Most people are aware of survivorship bias, but few understand its significance. It is widely discussed in academic literature, but remains ~~common~~ among practitioners. Mostly for convenience, practitioners tend to backtest certain investment strategies using only those companies that are currently in business, meaning stocks that have left the investment universe due to bankruptcy, delisting or being acquired are not included in the backtesting. Survivorship bias often leads to overly optimistic results and sometimes even draws the completely opposite conclusion.

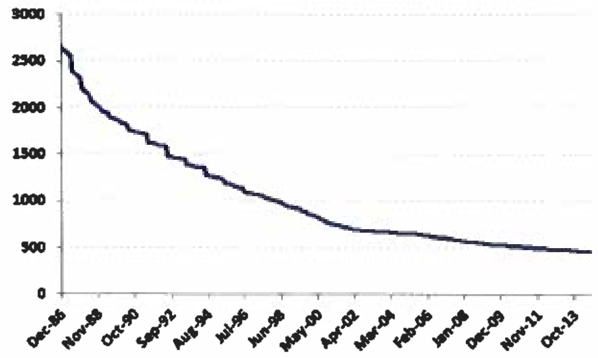
Figure 2: Survivorship bias



Source: Yin Luo

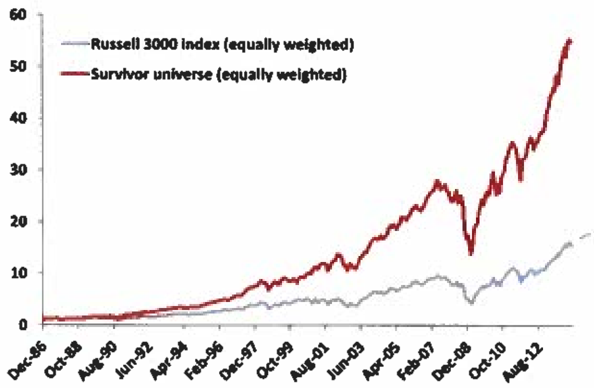
Now, let's show a simple example. If we use those companies that were in the Russell 3000 index on December 31, 1986 and have survived until today, i.e., we exclude those firms that were deleted from the index over the years (hereafter called "survivor universe"). As shown in Figure 3, it is obvious that the universe gets smaller and smaller over time – indeed, only less than 500 stocks (out of the 3,000 stocks in the index) have survived over the past 28 years. Then we track the performance of these stocks (equally weighted average) and compare that with an equally weighted Russell 3000 index. As shown in Figure 4, the companies that have survived outperform the index significantly, because stocks taken out of the index are mostly due to bankruptcy, delisting, or an extended period of underperformance (therefore, their market capitalizations drop below a certain threshold)[3].

Figure 3: Number of stocks in the Russell 3000 (as of 12/31/1986) that have survived until today



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 4: Stocks that have survived perform better than average



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Survivorship bias can also lead to completely opposite results. Figure 5 and Figure 6 illustrate the performance of the top and bottom quintile portfolios constructed on the Merton's distance to default factor[4], using the Russell 3000 universe[5] and the "survivor

[3] Companies that are taken out of the index could also be due to acquisition. In that case, stocks may have risen significantly. However, most index deletions are associated with underperformance.

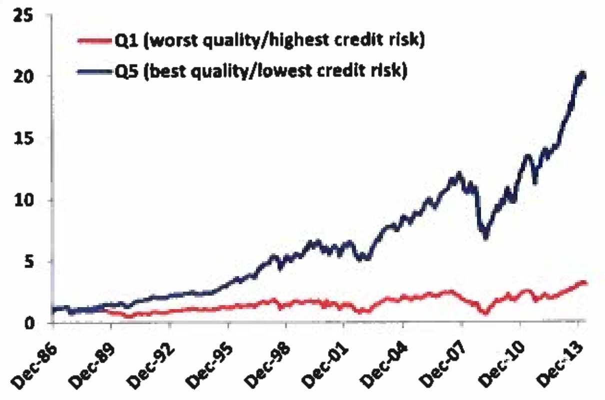
[4] The Merton's distance to default factor uses Merton's options pricing theory to measure the distress risk of a company. The larger the distance to default, the lower the implied credit risk.

universe"[6]. The proper backtesting using the correct investment universe suggests that companies with the best credit quality (i.e., largest distance to default) have considerably outperformed those distressed stocks (see Figure 5), i.e., a positive quality premium. Using the "survivor universe", however, we see something completely different (see Figure 6). It is astonishing that firms with the highest credit risk actually produce a 7x higher cumulative return compared to the best quality stocks.

As shown in Figure 7, using the proper investment universe, we can clearly see why low quality (highest credit risk) companies underperform – they tend to be companies that are more expensive and less profitable, have weaker growth and price momentum, with more negative earnings revision and higher volatility than those high quality firms.

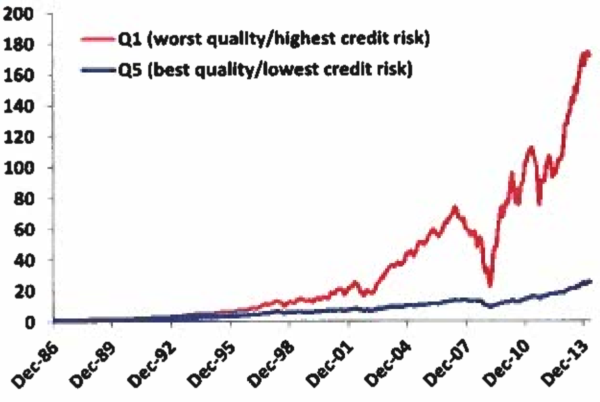
The close cousin of the survivorship bias is the look-ahead bias, i.e., using information that is not available as of the time of backtesting (a more detailed discussion about look-ahead bias will be presented in the next section). In this specific example, since most index deletions are due to share price underperformance, using the "survivor universe" essentially assumes that we know what companies underperform in the future. Figure 8 exhibits this look-ahead bias clearly. The worst credit risk companies in Q1 that have survived in the past 30 years no longer show negative characteristics across valuation, profitability, growth, sentiment, and momentum. It is not surprising that these companies have done exceptionally well (see Figure 6) – it is like going to a casino to play the most speculative game – if we win, we collect the money, but if we lose, we do not need to pay.

Figure 5: Merton distance of default factor on the Russell 3000 universe



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 6: Merton distance of default factor on the "survivor universe"

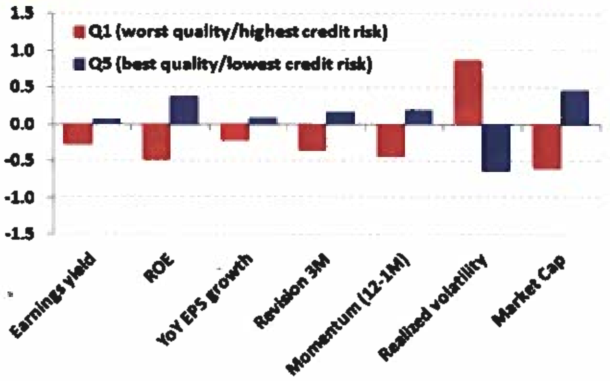


Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

[5] This is the correct universe without survivorship bias, where we use the point-in-time Russell 3000 index constituents at each month end to do the backtesting.

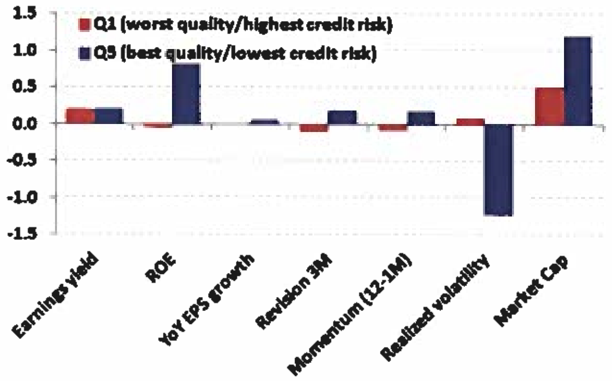
[6] This is the universe suffering from the survivorship bias, i.e., the universe of stocks that were in the Russell 3000 index on December 31, 1986 and have survived until present.

Figure 7: Factor exposure, the Russell 3000 universe



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 8: Factor exposure, the "survivor universe"



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

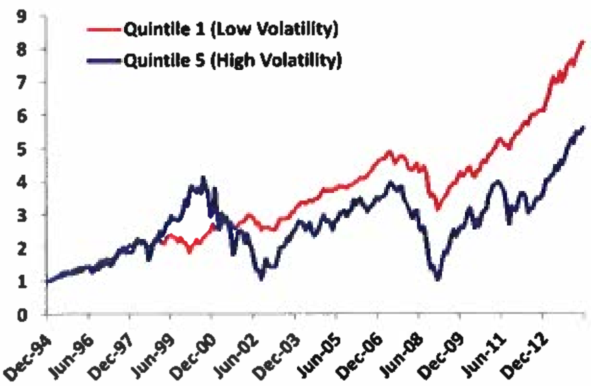
Backtesting with current index constituents

One particular kind of survivorship bias that is popular among non-quant investors (who want to incorporate quantitative screening into their investment process) is the current constituent bias. This happens when backtesting is conducted using only those stocks that are currently in an index. In this case, the backtesting universe comprises stocks that have survived over the years and companies that have been deleted from the index are excluded. Because maintaining historical index constituents is costly and painful, it is obviously much easier to use whatever companies are currently in the index to conduct backtesting.

Using current index constituents in a backtesting also suffers from look-ahead bias. Typically, underperforming stocks are deleted from the index and those outperforming ones are added. Using current index members back in time implies that we would have had the knowledge of companies that performed well in the future (i.e., index additions).

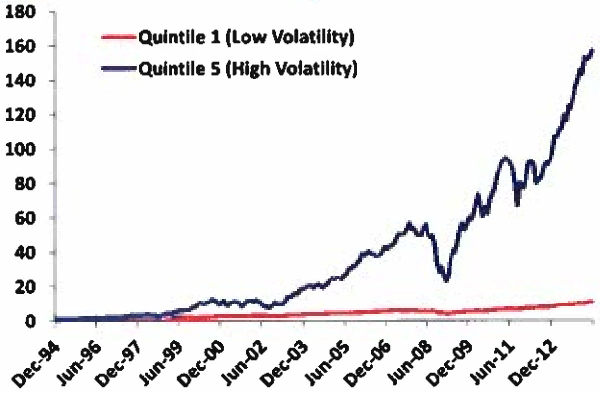
Let us take the realized volatility factor as an example. As we know, there is a low volatility anomaly in the equity world, i.e., companies with lower volatility actually tend to outperform highly volatile stocks (see Alvarez, et al [2011a], and Luo, et al [2013a]). Using the proper investment universe, e.g., stocks in the S&P 500 index point-in-time, as shown in Figure 9, low volatility stocks have clearly done better than high volatility stocks in the long term. Using the wrong universe, i.e., stocks currently in the S&P 500 index, we see exactly the opposite (see Figure 10) – high volatility stocks have outperformed low volatility stocks by 16x. This is very similar to (but even more extreme than) the Merton's distance to default example above. Stocks with the highest volatility also tend to show more extreme payoffs. Using current index constituents means that we capture all the positive extreme payoffs but do not have to pay for the negative ones.

Figure 9: Low volatility factor on the proper S&P 500 universe



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 10: Low volatility factor performance on the current S&P 500 index constituents



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

So, what should we do?

There are a few things that we should do to avoid the look-ahead bias:

- The backtesting universe should be point-in-time, including companies that went inactive (due to index deletion, bankruptcy, delisting, and acquisition).
- If you use a standard index (e.g., S&P 500 or MSCI World), you need to use those stocks in the index as of a given point-in-time and exclude companies that were added to the index at later dates.
- Take care of the missing value properly. If the factor scores for certain stocks are not available, we should keep value as NA (not available), instead of removing these stocks from the universe.

2. Look-ahead bias

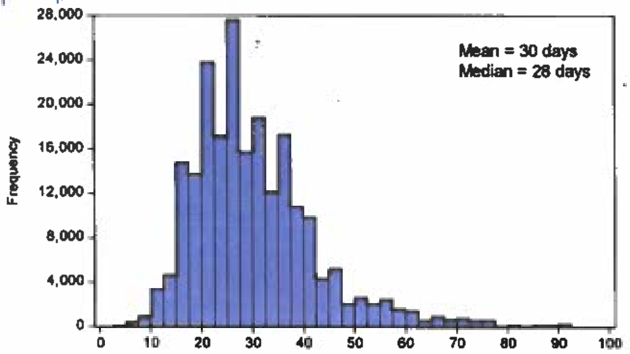
The second major mistake in backtesting is the look-ahead bias. It is the bias created by using information or data that were unknown or unavailable as of the time when the backtesting was conducted. It is probably the most common bias in the backtesting. To some extent, survivorship bias, discussed in the previous section, is a special case of the look-ahead bias, because the question whether a stock will survive (or be added in the index) in the future is unknown on the backtesting date.

An obvious example of look-ahead bias lies in companies' financial statement data. Companies report their financials on a regular basis[7] (quarterly, semi-annually, or annually). It takes, on average, one to two months for most companies to prepare and release their financial statements after the fiscal periods end (see Figure 12 and Figure 13), and in some occasions, it could take significantly longer. The general consensus is to apply reporting lags in the backtesting, i.e., we assume we do not know the fiscal period end financial statements until x months later. We typically set x equal to three months in standard backtesting.

Figure 11: Look-ahead bias

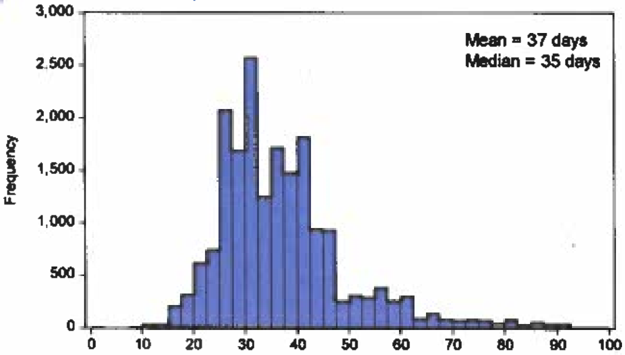
Source: Yin Luo

Figure 12: Number of days to file quarterly earnings – US companies



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 13: Number of days to file quarterly earnings – international companies



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

A more subtle form of look-ahead bias is restatement bias. Companies often restate their financial statements for various reasons. Many commonly used data sources only store the final revised data. In one of our previous research papers (see Luo, et al [2010b]), we show that economic data is also often restated and restatements are statistically significant. More importantly, models can be very different depending on whether the first reported or restated data is used.

The ideal solution is to use a PIT (point-in-time) database, where the vendor archives the originally reported data, along with the reporting dates, and keeps track of subsequent data revisions. For example, Compustat, Capital IQ, and Worldscope all provide PIT

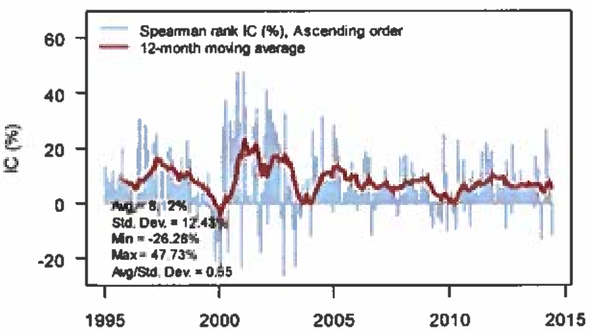
add? ask MAP

[7] See Luo, et al [2014] for a summary of reporting frequency by regions. In general, almost all public companies in the US and Canada report on a quarterly basis. Outside of the North American region, most companies report on lower frequencies (semi-annually or annually). However, we have seen more and more firms in Europe reporting quarterly financial statements in recent years.

company fundamental data for US and global companies. In our previous research, we have discussed the importance of using PIT data for backtesting (see Luo, et al [2010a] and Mesomeris, et al [2010]).

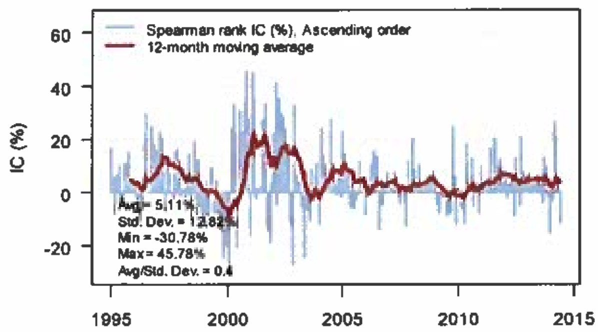
One of the common approaches that we use to assess factor performance is rank information coefficient (IC), i.e., the correlation between: 1) the stock ranking based on the factor scores as of a given date and 2) the stock ranking based on the returns over the subsequent period (often set as the next month's returns). Figure 14 shows the time series of performance of the earnings yield factor, using the non-PIT data. The non-PIT data assumes all financial statement data items become available on the days of fiscal period ends[8]. On the other hand, Figure 15 shows the earnings yield factor performance using proper PIT data. The biased non-PIT data inflates the performance of earnings yield by almost 60%.

Figure 14: The performance of the earnings yield factor, using non-PIT data



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 15: The performance of the earnings yield factor, using PIT data



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

The impact of reporting lag assumption

When PIT data is not available, investors typically apply some sort of reporting lags. There is not any fast or simple rule on how to set the exact lag. If we set the lag too short, we potentially suffer look-ahead bias. On the other hand, if we set the reporting lag too long, we could be using stale data.

Data coverage is another key parameter that we need to take into account. For example, Worldscope, as the *de facto* data vendor for international company fundamental data, introduced a point-in-time database in recent years. However, their PIT coverage only starts to match their traditional non-PIT data after 1999. Figure 16 shows the coverage of stocks in the S&P BMI UK universe, using ROE as an example.

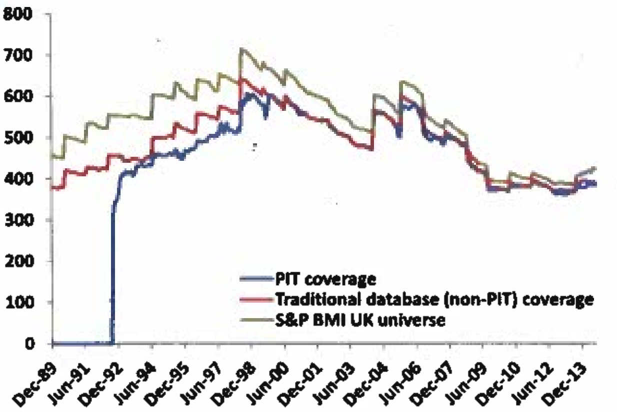
To make a fair comparison, we backtest our ROE factor in the UK universe only using data after 1999 to present, with the following setups:

[8] In this case, earnings yield is computed as EPS/price. For example, if a company has December 31, 2013 as the fiscal quarter end (Q4/2013), the non-PIT earnings yield is calculated using trailing four quarters EPS from Q1/2013, Q2/2013, Q3/2013, and Q4/2013. In reality, Q4/2013 EPS data is typically not available until Q1/2014; therefore, the proper EPS should be the sum of Q4/2012, Q1/2013, Q2/2013, and Q3/2013 data, assuming the company had reported Q3/2013 EPS by December 31, 2013.

- Using PIT data – this is the most realistic case
- Using non-PIT and making no reporting lag assumption[9] – this is the case with substantial look-ahead bias
- Using non-PIT and assuming a one-month reporting lag, i.e., we only know the ROE ratio one month after a company's fiscal period end
- Using non-PIT and assuming a two-month reporting lag
- Using non-PIT and assuming a three-month reporting lag

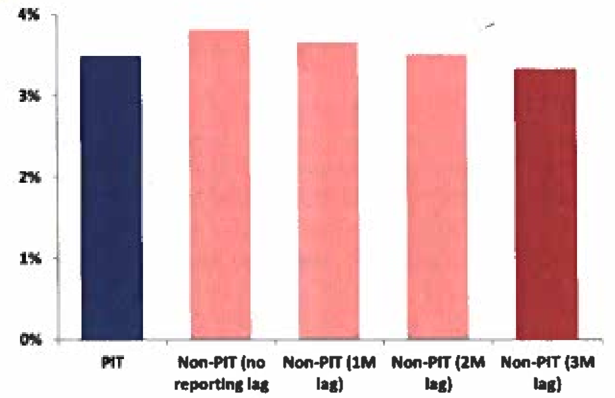
As shown in Figure 17, using non-PIT data and making no reporting lag assumption inflates ROE factor performance by 10%. If we do not have the PIT data and use non-PIT instead, as we extend our reporting lag assumption from one month to three months, ROE factor performance drops by 9%. The three-month reporting lag assumption appears to be somewhat overly conservative – as we use stale data, factor performance falls below that of using proper PIT data. The "ideal" reporting lag seems to be around two months. For most of our research, when we use non-PIT data, we typically use a three-month reporting lag to be conservative.

Figure 16: The coverage of UK stocks



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 17: The impact of reporting lag assumption, average monthly rank IC



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Look-ahead corporate action bias

Look-ahead bias can sometimes be quite intricate. For example, split adjustment factors can potentially produce look-ahead bias. From time to time, companies may decide to split their shares (or reverse split), to improve liquidity or attract certain types of clients. For most modeling purposes, we want everything to be split adjusted. For example, when we calculate earnings yield, EPS data typically comes from company financial statements with low reporting frequency (quarterly, semi-annually, or annually), while pricing information is from market data available at least daily. We need to make sure both EPS and price are split adjusted at the same time.

[9] For example, this essentially assumes that, for a company with a December 31, 2013 fiscal year end, we know the net income and shareholders' equity data of that company for the year 2013 on December 31, 2013.

Using split adjusted data can create a fairly subtle form of look-ahead bias. For example, one potential stock-selection factor is the share price itself. Investors are interested in whether the level of share price has predictive power of future stock returns. Company management and their investment banking advisors also want to know whether they should split (or reverse split) shares to keep the level of share price at a more reasonable range. In our backtesting, however, it is critical to use the original non-split adjusted share price. Many analysts either do not know the difference or do not believe it makes any difference[10]. We need to be aware of the fact that the adjusted price has future information.

Figure 18 shows the cumulative performance of the top 25 stocks with the lowest price in the S&P 500 index in the past 25 years. We can see that the adjusted price factor has done exceptionally well, beating the index by 2.5x, in terms of returns (see Figure 19) or 30% based on the Sharpe ratio (Figure 20). We may easily draw the wrong conclusion that buying low priced stocks is a good strategy, or recommend our corporate clients to split shares whenever possible.

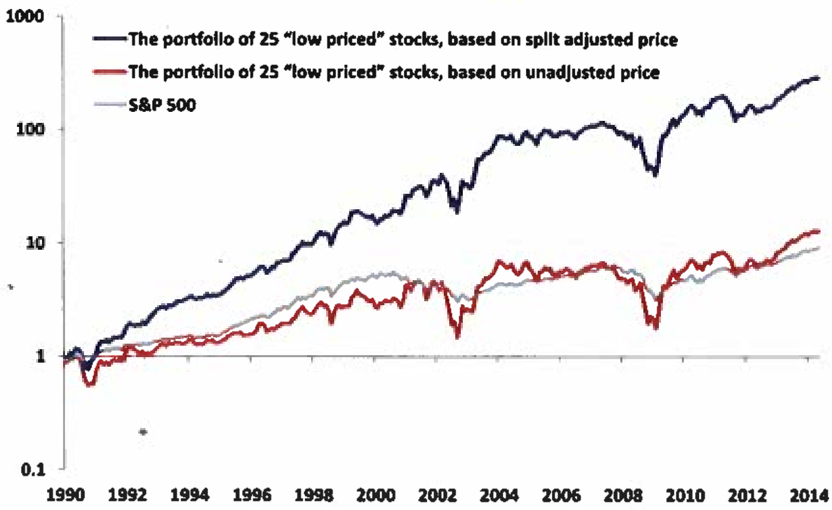
Using adjusted prices, unfortunately, suffers from look-ahead bias. For example, a stock that has had several share splits over the years, by definition, has very small adjusted prices going back in time. Companies generally choose to conduct a share split because their stock has risen to levels that are high enough to negatively impact its liquidity[11]. Therefore, going long these "low priced" stocks means buying the stocks that would have split over the next 25 years, which we obviously do not know at the time of investing.

The appropriate way to backtest this strategy is to use unadjusted prices, i.e., the "true" prices at that time of investing. The result is remarkably different (see Figure 19). The annualized return for those "low priced" stocks is roughly the same as the benchmark, but more than twice as volatile. As a result, the Sharpe ratio is half as that of the benchmark (see Figure 20). Therefore, buying the "low priced" stock may not be such a great strategy after all. This phenomenon confirms our intuition that extremely "low priced" or "penny" stocks are typically distress stocks that tend to underperform in the long term.

[10] That is partially because stock splits tend to be a fairly infrequent corporate event.

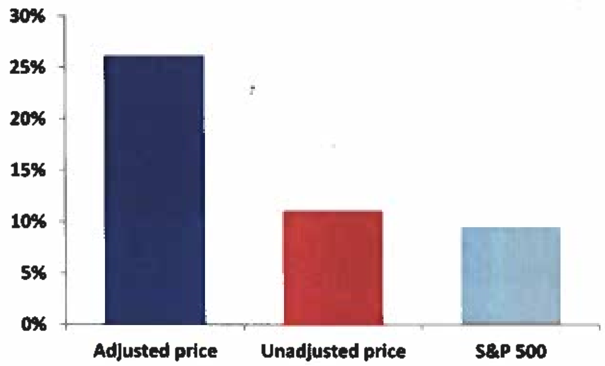
[11] In forthcoming research, we will address the reasons behind stock splits and other corporate actions. We will also perform an event study on these corporate actions.

Figure 18: Performance of the top 25 names with the lowest share price



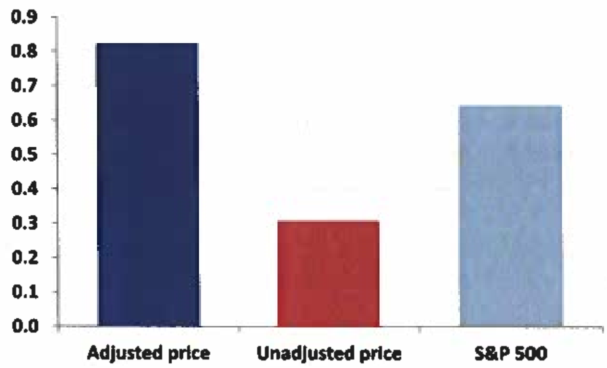
Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 19: Annualized return



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 20: Sharpe ratio



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

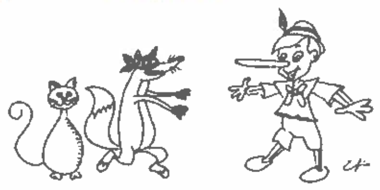
To avoid look-ahead bias in the backtesting, there are a few things that we could do:

- Use PIT (point in time) data whenever possible.
- If there is no PIT data available, lag the data conservatively.
- Be careful when to and when not to use split adjustment.
- Think twice when the backtesting results are counter intuitive (more on this later).
- Run the strategy live (on paper) to see if the performance matches with that in the backtesting.

3. The sin of storytelling

We all love stories. If you have ever attended trainings on presentation or public speaking, you probably remember that one of the most important ways to leave a deep impression with your audience is to tell stories rather than simply repeating facts and numbers. In the famous case in Langer, et al [1978], the authors set up an experiment where participants asked people in a queue if experimenters can squeeze in. In the cases where the participants did not give excuses, their success rate was only 42%. When the participants tell stories[12] on why they need to push in, 60% of those in line people granted the favor.

Figure 21: Storytelling



Source: Yin Luo

We all seem to be more comfortable with investing in something that we can understand and that we have an intuition behind. The problem is, once we find a pattern, we can almost always come up with a story to explain it. However, the fact that we can explain something has nothing to do with its out-of-sample performance.

The pattern about financial leverage is a classic example. Should companies with higher financial leverage (e.g., long-term debt/total assets) earn higher or lower subsequent returns? On the one side, you may argue that firms with higher leverage are inherently more risky; therefore, investors should demand higher returns (e.g., premia) for holding these stocks. On the other hand, you could easily explain the result using a behavior finance story either limited by arbitrage (e.g., we cannot get enough supply to short distress stocks) or overconfidence (e.g., investors love to invest in speculative stocks in the hope of making a killing); and therefore, highly leveraged firms should deliver lower subsequent returns. Well, if we backtest the simple debt/total assets (or debt/equity) ratio factor, over time, or in a different investment universe, the results can well be very different and contradictory; but which story should we believe?

More problematically, we all tend to suffer from the so-called confirmatory bias – looking for information that supports our prior beliefs and ignoring the information that does not.

How long is long enough?

One of the common critiques we hear from clients on new databases is "there is only a few years of data and the history is not long enough". In a sense, they are correct. Unlike physical science, in economics and finance, we have only one set of observations over time, i.e., we deal with time series data. Even if the data is stationary, we need a very long period of medium frequency (e.g., monthly) or a reasonable length of high frequency (e.g., daily or tick-by-tick) data. So, how long is long enough?

On one hand, we want the backtesting to go as far back as possible – at least over a few economic cycles to overcome the small sample bias. On the other hand, we know that economic data is often non-stationary; therefore, what happened 20 years (or something even a few years) ago may be irrelevant today. As we wrote in a number of our research papers (e.g., Luo, et al [2013b]), regime shifts are commonplace in finance. In particular, the Internet bubble in late 1990s, the 2007 quant crisis, and the more recent 2008 global financial crisis are often cited as break points into new regimes.

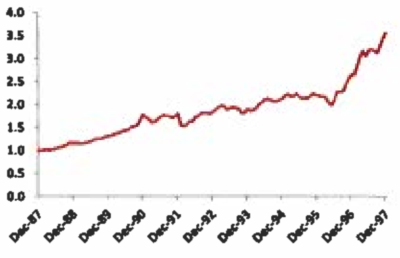
[12] These include even some obviously meaningless excuses like "I urgently need to make a copy".

For example, let us take value as an example and use earnings yield as a proxy for valuation. Value and momentum are probably the two most widely accepted factors in finance in both academia (see Fama and French [1992, 1993]) and by practitioners. Suppose we got into the investment business at the end of 1997 and backtested the earnings yield factor on US stocks. As shown in Figure 22, value is clearly a winning strategy – buying cheap and selling expensive stocks had worked out exceptionally well. We could easily come up with two stories. Value stocks are more like distressed companies; therefore we should demand higher returns (i.e., risk premia) to take on the extra risk. Alternatively, we could resort to the behavioral finance argument – investors tend to be overly confident about their skill and over invest in growth or "glamorous" stocks. Therefore, "glamorous" stocks are overly expensive and the future realized growth often turns out to be disappointing. Well, both stories sound quite convincing, so we would have probably invested in value stocks.

Now, let us see what happened to our value strategy in the technology sector in the subsequent two and a half years, from late 1997 to mid-2000. It turned out to be a disaster. In such a short period of time, we lost nearly 70% of our investment (Figure 23). By early 2000, we all had heard about the story that the old brick-and-mortar business model would soon be completely replaced by the so-called "new economy". It was more about who can come up with the next "Dot com" names and valuation seemed to be irrelevant. Unfortunately, the "new economy" story did not last very long either. In the next two and a half years from mid-2000 to late-2002, the old and boring value style recovered strongly (Figure 24). Looking through the entire 20 years of history (see Figure 25), valuation does not seem to matter much in the technology sector – it generates a decimal return with a huge volatility. The problem is that we never really know whether a factor is good or bad; nor do we even know the direction; i.e., whether being cheap would produce higher or lower subsequent returns, at a given point in time.

As shown in Figure 26, the Sharpe ratio of a typical value strategy in the technology sector can vary from significantly negative to considerably positive. So what story should we believe?

Figure 22: Wealth curve of earnings yield factor (the Russell 3000 index), 1987-1997

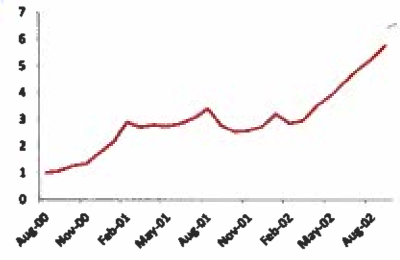


Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

Figure 23: Wealth curve of the earnings yield factor (technology sector), before tech bubble burst

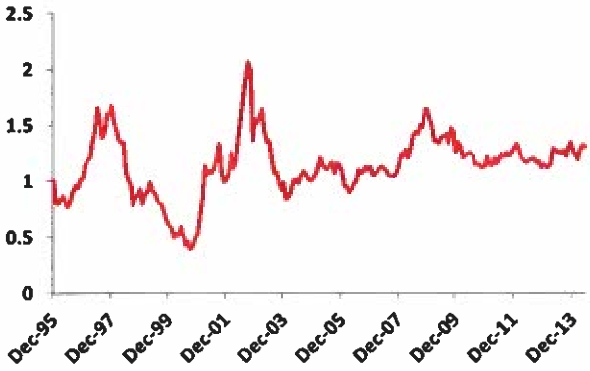
Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

Figure 24: Wealth curve of the earnings yield factor (technology sector), after tech bubble burst



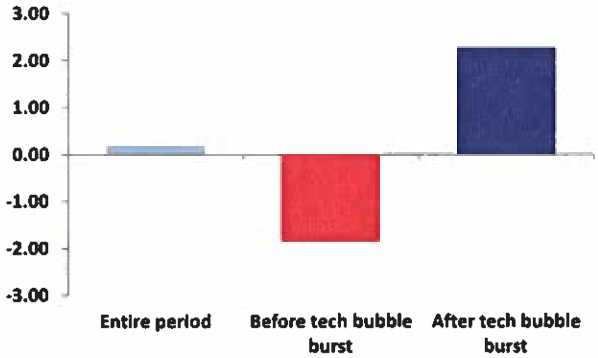
Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 25: Wealth curve of the earnings yield factor (technology sector), 1995 to 2014



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 26: Sharpe ratio



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

A bigger problem is that we all tend to publish only those factors that "work"[13] and disregard those ones that do not. A quick search on the archive of the *Journal of Portfolio Management* or SSRN seems to suggest that new and promising factors are being discovered every day.

To avoid time period bias, we can do the following:

- Use as long of a history as possible so that it covers as many economic and policy cycles as possible.
- Understand how performance varies in different economic cycles and examine whether government policy intervention creates potentially new regimes.
- Another philosophy is to not even care about explaining patterns. We build machine learning algorithms and let our models identify patterns, hoping those patterns are at least persistent in the short term. The so-called StatArb (or statistical arbitrage) strategy typically follows this argument. Many investors confuse machine learning with data mining. We would like to argue that they are very different concepts. We will discuss data mining in the next section.

[13] There are occasionally papers that discuss factors that do not predict returns or risk, but they are definitely the minority.

4. Data mining and data snooping bias

Figure 27: Data mining

Source: Yin Luo

Data mining is an interesting phrase. In the fields of computer science and statistics, it refers to the computational process of discovering patterns in large data sets, often involving sophisticated statistical techniques, computation algorithms, and large scale database systems. In finance, however, it often means "manipulating" data or models to find the desired pattern that an analyst wants to show.

We would rather refer to the finance definition of data mining as data snooping, i.e., the behavior of extensively searching for the patterns or rules that fit a model perfectly in-sample. Analysts often fine tune the parameters of their models and choose the ones that perform well in the backtesting.

With enough data manipulation, we can almost always find a model that performs very well in-sample. Let us show an example of data snooping in model construction. First, we start from the 72 common stock-selection factors (e.g., earnings yield, price momentum, return on equity, debt/equity ratio, etc.) in our factor library, using the S&P 500 as our investment universe. For our "in-sample model", we backtest the performance of these 72 factors from May 31, 2009 to June 30, 2014, pick the best factor from each one of the six style buckets (value, growth, momentum/reversal, sentiment, quality, and exotic), and equally weight the six factors, which forms our multi-factor alpha model. Then, we backtest the performance of this model ("in-sample model") over the same period from May 31, 2009 to June 30, 2014. Figure 28 shows the performance of our in-sample multi-factor model. It has done well with a Sharpe ratio of 0.7x in US large cap equities. See, we told you – quant is easy! The problem is, of course, as of May 31, 2009, we did not know which six factors would have done the best in the next five years; so obviously, the in-sample model suffers from the look-ahead bias.

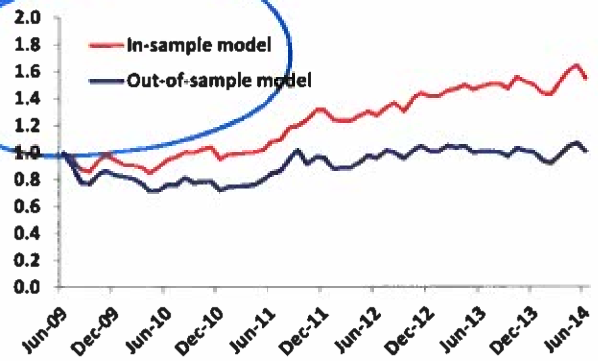
Next, we build an out-of-sample model. Starting from May 31, 2009, we use rolling 60 months of data to form our multi-factor model, following a similar methodology as the in-sample model. At the end of each month, we backtest all 72 factors, pick the best factor from each of the six style categories (but only using data available as of that time), and equally weight them, which becomes our "out-of-sample" multi-factor model. As shown in Figure 28, the out-of-sample model was essentially flat over the past five years – well, quant investing is not so easy after all – at least not in the US large cap equity space.

Equally weighting all factors is simple (and commonly used by practitioners), but not necessarily the optimal approach – it does not take into account either factor performance or risk (and correlation/diversification). Now, let us test another popular factor weighting algorithm – Grinold and Kahn's approach (see Grinold and Kahn [1999]). Grinold and Kahn's approach is essentially a mean-variance optimization at the factor level, by incorporating both factor returns and risk in the weighting decision. Figure 29 shows the performance of our in-sample versus out-of-sample models[14].

[14] In this setup, similar to the simulations above, we also backtest all 72 factors and select the best factor from each of the six style categories. The six factors are then optimized via the Grinold and Kahn algorithm to decide their weights in

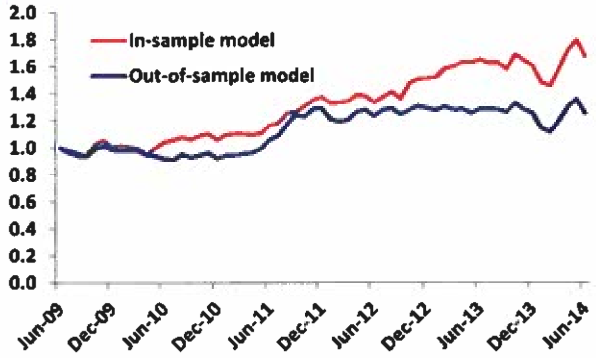
Comparing Figure 29 to Figure 28, using in-sample data, it appears that factor weighting decisions are not very important – equally weighting and MVO produce almost identical and decent returns. Using the correct out-of-sample data, we can clearly see the benefit of the mean-variance optimization in the factor weighting process. The Grinold and Kahn factor weighting algorithm produces some modest profit, while the equally weighted model was essentially flat.

Figure 28: Factor weighting – equally weighting algorithm



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 29: Factor weighting – Grinold and Kahn MVO algorithm



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Data snooping bias is probably the most difficult to deal with. In our opinion, we can probably never be able to completely avoid data snooping bias. However, some basic checking can help us stay honest:

- At the very minimum, we should avoid look-ahead bias. When we build models and backtest strategies, we need to make sure to only use data available to us as of the time, i.e., point-in-time.
- More importantly, we are strong believers of backtesting on a set of pre-defined rules rather than factors. If we always follow the same set of rules to backtest and select factors, weight factors, and build portfolios, and if the backtesting results are promising, we have higher confidence that the same methodology is likely to survive out-of-sample. Interested readers can find more details in Luo, et al [2010a, b] and Wang, et al [2012, 2013a].
- Ideally, we want to build our model using one set of data – for example, equity data in one country. Then, we apply the same methodology (without fine tuning the parameters) using different data sets, e.g., other countries/regions.
- If we deal with a single country (e.g., US equities or Japanese equities), we could reserve a set of true out-of-sample data (let us call it the "validation sample"). We should never touch the "validation sample" and use it only to check out final model performance.
- The real test is live performance.

the multi-factor model. We use rolling average returns as expected returns, and a rolling covariance matrix as our factor risk model.

5. Signal decay and turnover

Backtesting is often conducted in an ideal world: no transaction cost, no turnover constraint, and unlimited long and short availability. In reality, all investors are all limited by some constraints. In this section, we discuss one of the most important constraints – turnover and the associated transaction cost issue.

Figure 30: Signal decay

Source: Yin Luo

The impact of transaction cost

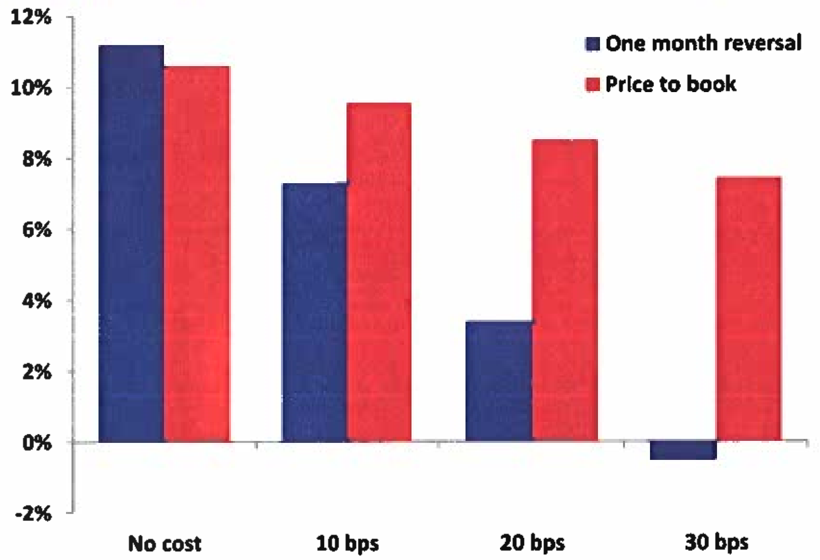
First of all, you may wonder why turnover or signal decay even matter. If you manage assets for a hedge fund, you probably have quite a bit of flexibility to turnover your portfolio anytime you like or as much as you see fit. The problem is that quick-decay factors require high turnover to extract the juicy returns. High turnover means more transactions, while transactions incur costs (commissions, bid-ask spreads, market impact, and for long/short portfolios, stock lending fees).

The classic example is short-term reversal, i.e., stocks that have performed well recently (say, the last month) are more likely to revert (underperform) in the subsequent month. The reversal factor is a great stock selection signal in the Japanese equity market (before transaction costs anyway). In one of our recent research papers (see Wang, et al [2014]), we find that reversal has even stronger predictive power among dividend paying stocks in Japan.

As shown in Figure 31, in a theoretical world with no transaction costs, a simple long/short strategy (buying the top 20% dividend paying stocks in Japan with the worst performance in the previous month; and shorting the bottom 20% stocks with the highest returns in the prior month) has generated an annual return of 12%, beating the classic value factor of price-to-book. As we increase our transaction cost assumption from 0bps to 30bps per trade, the alpha of the reversal strategy drops to negative, while the excess return of the price-to-book value strategy drops only modestly.

For now, the solution seems to be easy, for fast decay/strong performing factors, maybe we can simply add a tight turnover constraint, then we can capture their predictive power while limiting transaction costs. The answer, of course, is never that easy. As we will show in the next section, it is really a balance between predictive power, decay/turnover, and cost. For models with exceptional predictive power, we actually want to trade more – even after transaction cost, we may still generate higher returns.

Figure 31: Annualized return with different transaction cost assumptions



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Turnover constraint matters

In reality, not every manager can trade as much as they want. Most institutional portfolio managers face some level of turnover constraints. Turnover constraint can have a considerable impact on model performance. Now, let us show the impact of the turnover constraint using our N-LASR[15] global stock-selection model as an example. The N-LASR uses a sophisticated machine learning technique called AdaBoost to build a highly adaptive model. The N-LASR model has shown great out-of-sample and live performance. However, due to its adaptive nature, it tends to have fairly high information decay, which requires a high level of turnover to capture the benefit.

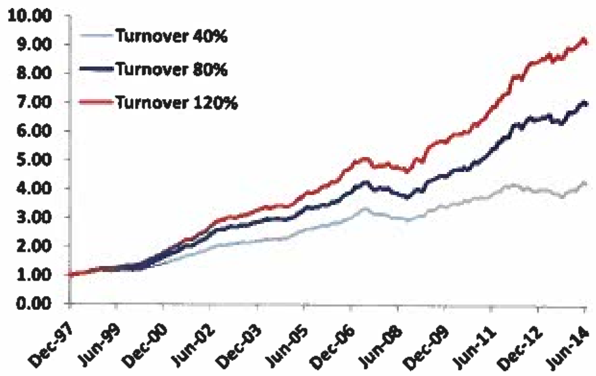
Figure 32 and Figure 33 show the performance of a few optimized portfolios based on the N-LASR model on the US equity market, under different levels of turnover constraints. We also impose other common constraints, but keep them the same for comparison purposes:

- Long/short market neutral strategy
- 2x leverage, i.e., for $1 capital, the strategy invests in $1 long and $1 short
- Target annualized volatility of 4%
- Maximum single stock weight of 1.5%
- Beta neutral (maximum 0.1 net beta exposure)
- Sector neutral (maximum 10% absolute sector exposure)
- Transaction cost at 20bps per trade

[15] N-LASR is our proprietary global stock selection model based on a set of around 70 factors, using a machine learning technique called AdaBoost. Details can be found in Wang, et al [2012, 2013].

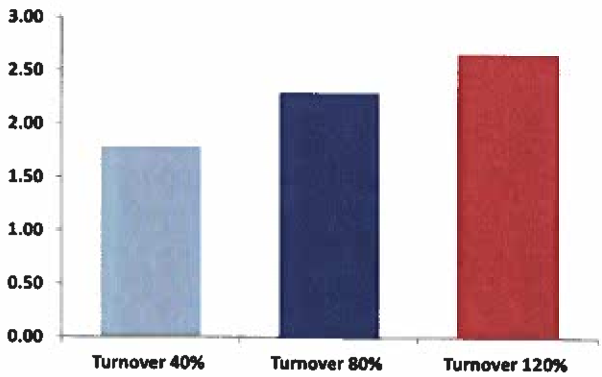
As we tighten the turnover constraint (two-way per month) from 120% to 40%, the Sharpe ratio plunges from 2.7x to 1.8x (almost by one third). In this case, because the N-LASR model has such a strong predictive power for near-term stock returns, a tighter turnover constraint, even after transaction cost, actually hurts model performance.

Figure 32: Wealth curve for N-LASR portfolio with different turnover constraints (after transaction cost)



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 33: Sharpe ratio with different turnover constraints (after transaction cost)



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Optimal rebalancing frequency

Having a tight turnover constraint, however, does not necessarily mean that we should have a very low rebalance frequency. In many instances, we have heard comments such as "we are long-term value investors; we hold stocks for three to five years; and therefore, we rebalance once a year". New information comes in constantly and we should adjust our models and beliefs accordingly. Even if we have a tight turnover constraint, we may still want to frequently adjust our positions – albeit modestly each time.

To illustrate the benefit of frequent rebalancing, let us use a simple example. Assume we manage a long-only value-tilt mutual fund portfolio, with a fairly tight turnover constraint of no more than 36% per year. We constructed two long-only portfolios based on the earnings yield factor[16], tracking the Russell 3000 index, with a 4% target annual tracking error.

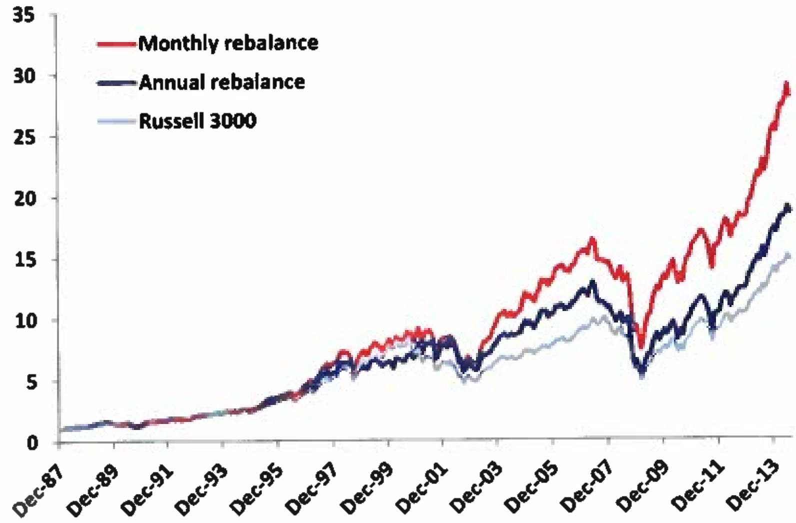
- **Annual rebalance:** we rebalance our portfolio once a year on December 31 and keep the turnover at 36%.
- **Monthly rebalance:** we rebalance our portfolio once a month, at each month end, and constrain the monthly turnover to 3%; therefore, the total annual turnover is the same as the strategy above, at 36%.

Sixth have bank do

As shown in Figure 34, although both strategies outperform our benchmark, the more frequently rebalanced strategy ("monthly rebalance") beats the annual rebalance strategy by 1.5x over the past 26 years.

[16] As a value factor, the earnings yield has slow information decay and does not require much turnover.

Figure 34: Annual versus monthly rebalance for a low turnover value portfolio



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Signal decay

Factors with faster decay require higher turnover, while higher turnover incurs greater transaction cost, which has to be balanced with the predictive power of the models. Some signals decay so quickly that alpha opportunity disappears within hours, minutes or even seconds. For example, a simple backtesting of the one-day reversal factor (i.e., buying stocks that have fallen the most on the previous day) seems to suggest that short-term reversal is a great strategy. The only problem is that the factor itself can only be computed after the market closes; therefore, the earliest time we can trade on the signal is at the next day's open.

Figure 35 shows the performance of our one-day reversal factor, assuming[17]:

- We compute the factor using returns from the previous day's close to today's close; then trade at the same day's close (only possible in theory) – hereafter called "trading at the same day's close".
- We compute the return of each stock from the previous day's close to today's close, but trade at the next day's open (possible in practice) – hereafter called "trading at the next day's open".

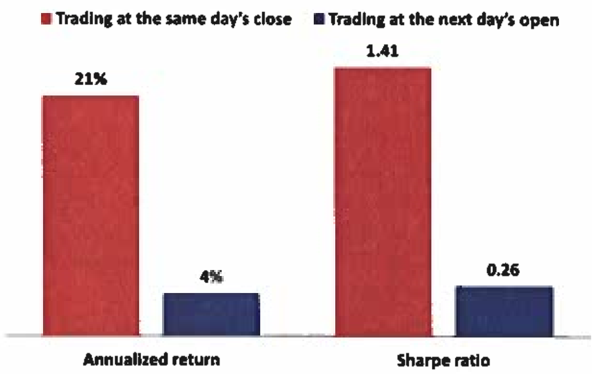
If we can calculate the one-day reversal factor and trade on the same day's closing price, we can generate a Sharpe ratio of 1.4x – pretty good for a single factor model. However, in reality, we can only trade at the second day's open, in which case the Sharpe ratio plummets to merely 0.3x (down almost 80%).

[17] The performance is based on a simple long/short quintile portfolio, daily rebalanced.

Figure 35: Performance of one day reversal

Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 36: Annualized return and Sharpe ratio



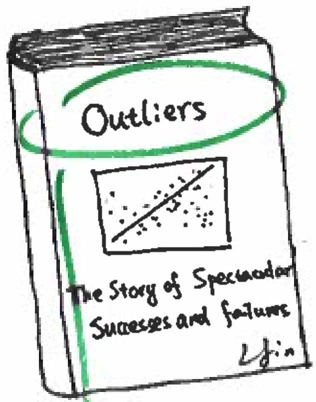
Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

How to optimally weight in signal decay, turnover, rebalance frequency, and transaction cost is partially science and partially art. Interested readers can find more details in Alvarez, et al [2011b] and more practically, an application in emerging markets, where transaction cost tends to be prohibitively high (see Wang, et al [2013b]).

6. Outliers – spectacular successes and failures

As discussed in Luo, et al [2013b], outliers are commonplace and as expected in financial data. They can be friend or foe – it is all about how to treat them. Outliers are not necessarily bad. Indeed, outliers referred to in Malcolm Gladwell's book, "Outliers: The Story of Success" [18], are mostly star hockey players or mathematicians.

Figure 37: Outliers



Source: Yin Luo

LHS outliers

Controlling outliers via winsorization

The easiest way to deal with outliers is to control them. More often than not, outliers are caused by data errors or specific events that are unlikely to be repeated in the future. There is little information on outliers and they usually cause more harm than good. We could either remove them from our sample completely (called trimming, truncation, or censoring, in econometrics jargons) or impose some restrictions by winsorization[19].

Unlike data truncation[20], winsorization replaces data greater or less than certain extreme x percentiles (often set as 5%, 1%, or 0.1%) with the xth percentile values.

Taking earnings yield as an example, Figure 38 and Figure 39 show the distribution of the earnings yield factor for US firms and Asian ex Japan stocks, respectively. Outliers are clearly present in both regions, but appear to be more dominating in AxJ. Financial data quality is generally better for US companies than for international firms. A naïve way to winsorize data is to replace those earnings yields below -100% with -100% (and similarly for yields above 100%).

When we aggregate company fundamental data to the market index level, outliers could have an even bigger impact.

Figure 40 shows the aggregate trailing 12-month earnings yield for the Russell 3000 index. It is calculated using the total earnings[21] of all the Russell 3000 stocks divided by the total market cap of the same stocks. There are a few spikes in the index aggregate earnings yield. Diving into the data reveals that the spike in 1994 was caused by Welbilt Corp.'s EPS being over $60,000, while the share was priced at just $20[22]. Similarly, the spike in 2000 was caused by Crown Media Holdings Inc's EPS being hugely negative. Figure 41 shows the aggregated index earnings yield after winsorizing at the 1% level – the smoothed data series appears to makes more sense.

[18] By the way, if you have not read the book, we would highly recommend you pick up a copy at http://www.amazon.com/Outliers-Story-Success-Malcolm-Gladwell/dp/0316017930/ref=sr_1_1?s=books&ie=UTF8&qid=1406826384&sr=1-1&keywords=outlier

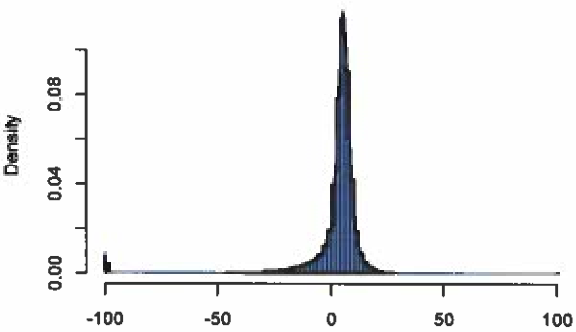
[19] Winsorization is named after the famous statistician Charles P. Winsor.

[20] Data truncation typically eliminates outliers from the analysis, which will be discussed in the next section.

[21] Total earnings are calculated as $\sum EPS_i \times NoSharesOS_i$

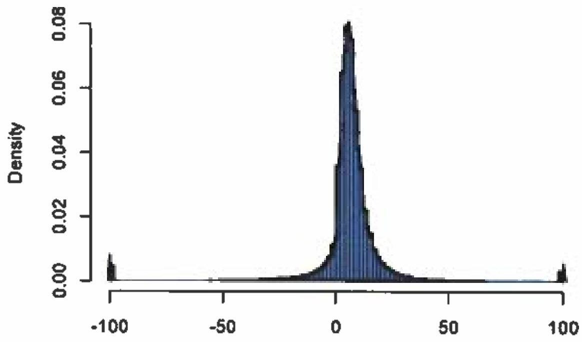
[22] In this case, it is probably better to use the net income data directly from a company's financial statement. However, as discussed in Luo, et al [2012], it is not that easy because the net income figure may not be consistent with how market cap is computed, partially due to the dilution effect from employee stock options, etc.

Figure 38: US earnings yield distribution (percentile)



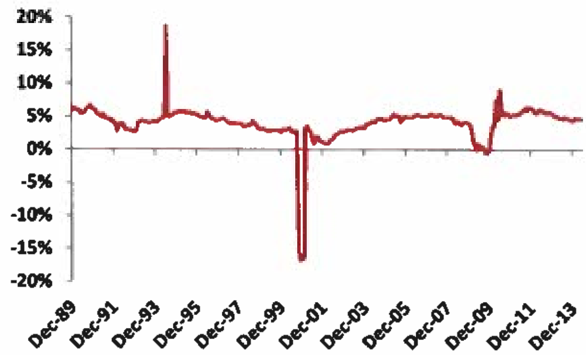
Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 39: Asia ex Japan earnings yield distribution (percentile)



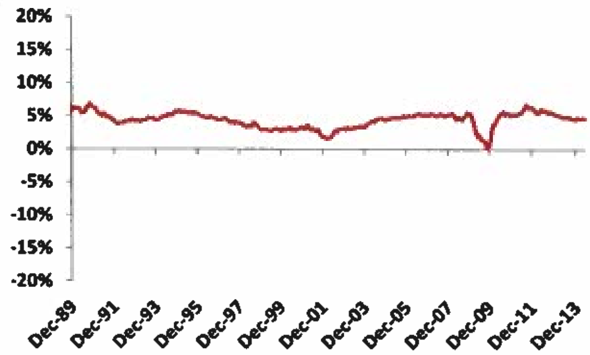
Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 40: Aggregate earnings yield, using raw data



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 41: Aggregate earnings yield, using winsorized data

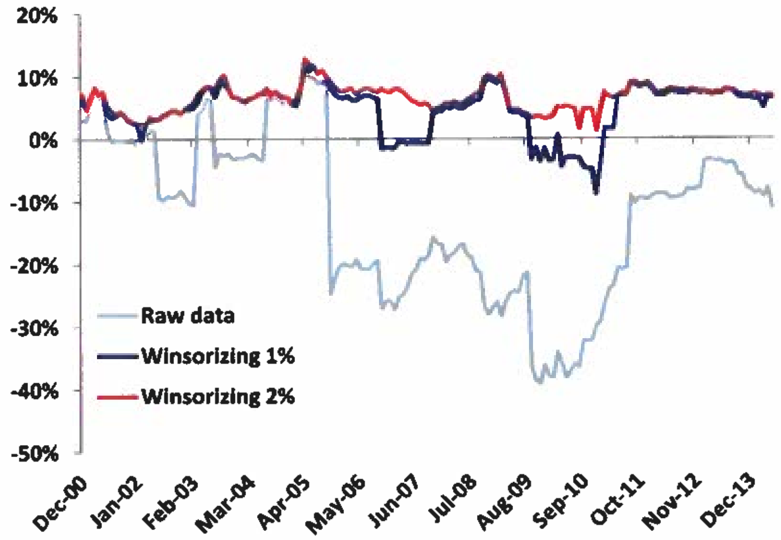


Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

If we calculate the same earnings yield for the S&P BMI Korean universe, which has about 350 stocks, as shown in Figure 42, the raw index earnings yield was hugely volatile and stays deeply negative throughout most of its history. If we winsorize the underlying data at the 1% level, much of the ups and downs disappear. If we set the winsorization threshold at 2%, most outliers are gone and the index yield becomes a much smoother series, but we may inevitably also have removed some useful observations. Depending on the purpose, we may want to use either the original data series, the 1% winsorized, or clipping[23] data at 2%.

[23] The term clipping is often used in the signal processing field.

Figure 42: Aggregate earnings yield for the Korean market



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

It turns out that winsorization may not be the ideal approach in macro data aggregation. In Luo, et al [2012a], we discuss in detail the correct way of aggregating bottom-up data to the top-down level. It essentially involves a process of creating a set of index level financial statements, using index adjustment factors and taking into account missing values.

Here are a few issues with winsorization:

- By definition, this method will remove extreme values in our data sample, regardless if they are data errors or important observations. Outliers may indeed contain useful and important information (more on this later).
- The threshold percentile is often chosen subjectively and arbitrarily. There is no one-size-fits-for-all solution. More importantly, even for a single factor, the percentile of outliers may also change over time.
- Backtesting results may be sensitive to the cutoff value we choose.

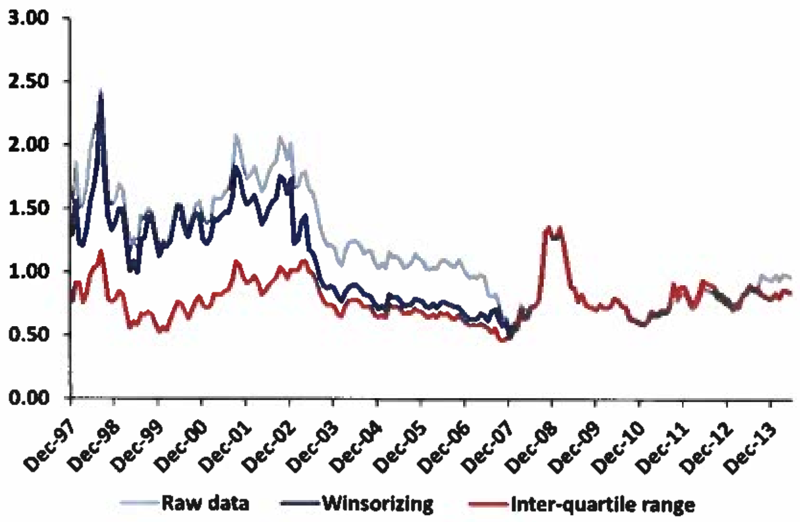
Truncating outliers via Inter-quartile range (IQR)

Another easy and effective way to identify/remove outliers is based on a measure called inter-quartile range (see Upton and Gram [1996], Zwillinger and Kokoska [2000]). It is a measure equal to the difference between the upper and lower quartiles. Assume $Q1$ and $Q3$ are the lower and upper quartile, respectively, Q3-Q1 is called the inter-quartile range (IQR).

An outlier is defined to be any observations outside the range: $[Q1-k(Q3-Q1), Q3+k(Q3-Q1)]$, for some nonnegative constant value k. This method is intuitive and robust to the number of outliers. As long as the data is not highly skewed, it can exclude as many as 50% (25% on each side) of the observations as outliers, or keep the entire sample if there is no outlier.

Let us use the aggregate book-to-price ratio for the Hong Kong[24] market as an example. As shown in Figure 43, the IQR approach produces the smoothest line, followed by winsorization, while the book yield using raw data exhibits more extreme values. Examining the numbers under the hood reveals a few large outliers with book-to-price over 25x still present prior to 2002, even after winsorization. Without outlier control or using the winsorization technique, the Hong Kong market was even cheaper in the early 2000s than it was during the 2008 global financial crisis. It appears that IQR is more robust in this case, but again, the "best" technique depends on the purpose of our analysis.

Figure 43: Aggregate book-to-market for the Hong Kong market



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

The IQR method would not work properly if the data is highly skewed or if the distribution is far from normal. In which case, we need to transform the data first. For example, market cap is highly skewed. We could perform a log transformation, then apply the IQR truncation. Therefore, the downside of using IQR is that we need to manually examine the distributional property of each factor first[25].

Z-score data normalization versus ranking normalization

Of course, the most popular way to normalize data is probably via z-scores, i.e., we subtract the mean from the raw data and then divide by standard deviation.

$$z_{i,t} = \frac{x_{i,t} - \mu_t}{\sigma_t}$$

Where,

$z_{i,t}$ is the raw factor score for stock i at time t,

μ_t is the cross-sectional mean of all stocks in the universe as of time t, and

[24] We use the S&P BMI Hong Kong index as our universe.

[25] We could, of course, automate the process with Jacque-Bera or other normality statistical tests.

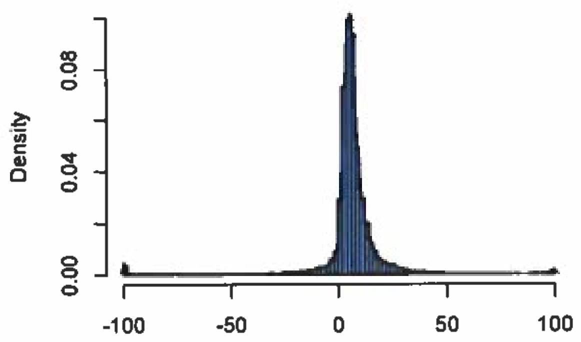
σ_t is the cross-sectional standard deviation (i.e., dispersion) of all stocks in the universe as of time t.

For most factors, even after the z-score transformation, there are still significant outliers and the distribution can still be far from normal. Therefore, many analysts would apply further outlier controls (e.g., via winsorization or IQR truncation).

Our preferred approach of data normalization is via a ranking transformation. We rank all stocks based on their factor scores. The ranking transformation automatically transforms whatever distribution to a uniform distribution. All outliers are also transformed into a reasonable data range. In our proprietary models, we further convert the data via an inverse normal transformation; therefore, our factors always follow exact standard normal distribution with zero means and standard deviations equal to one.

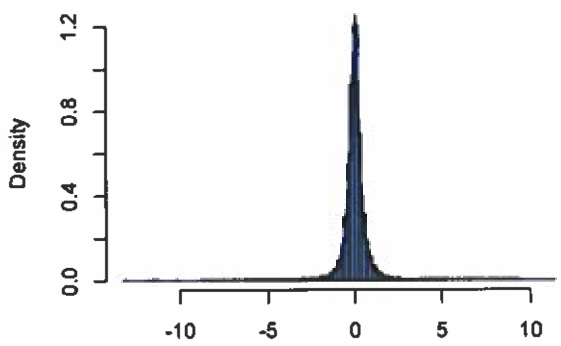
Figure 44 shows the distribution of the original factor data – in this case, we use the earnings yield for Indonesia as an example. We can clearly see that data does not follow a normal distribution and has plenty of outliers. Figure 45 illustrates the distribution after the z-score transformation. Outliers are still present, but the distribution is less skewed. Figure 46 exhibits the distribution after ranking transformation – it clearly follows a uniform distribution. Finally, you can see our proprietary algorithm converts data into a standard normal distribution (see Figure 47).

Figure 44: The distribution of Indonesia earnings yield – raw data



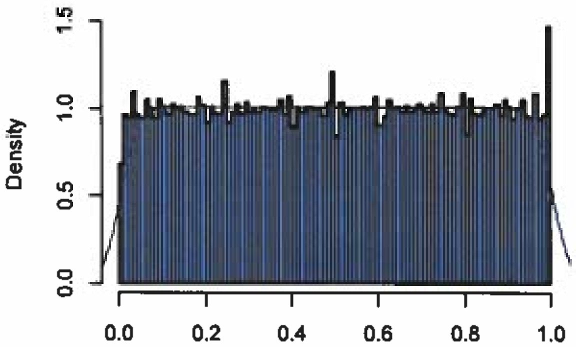
Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 45: The distribution of Indonesia earnings yield – z-score transformation



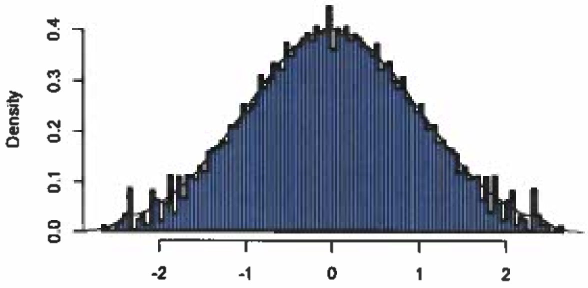
Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 46: The distribution of Indonesia earnings yield – the ranking transformation



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 47: The distribution of Indonesia earnings yield – our proprietary transformation



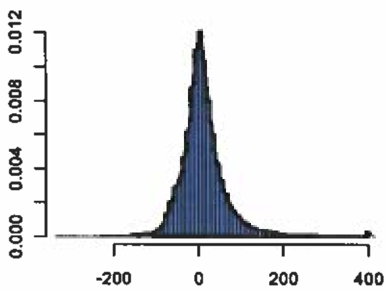
Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Now, let us see the impact of data normalization on model performance. We build an equally weighted multi-factor model on four simple factors: value (trailing 12-month earnings yield), momentum (12-1M total return), sentiment (three-month earning revision), and quality (ROE), using three different data normalization techniques:

- Use raw factor scores with no normalization
- Z-score normalization
- Ranking, then inverse normal transformation

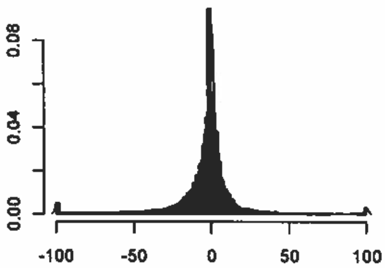
Figure 48 to Figure 50 show the distribution of raw factor scores for Indonesia, we can see that different factors have different distributional properties.

Figure 48: Momentum distribution



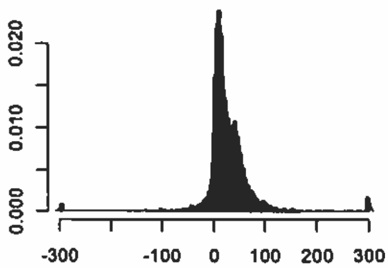
Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 49: Sentiment distribution



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 50: ROE distribution



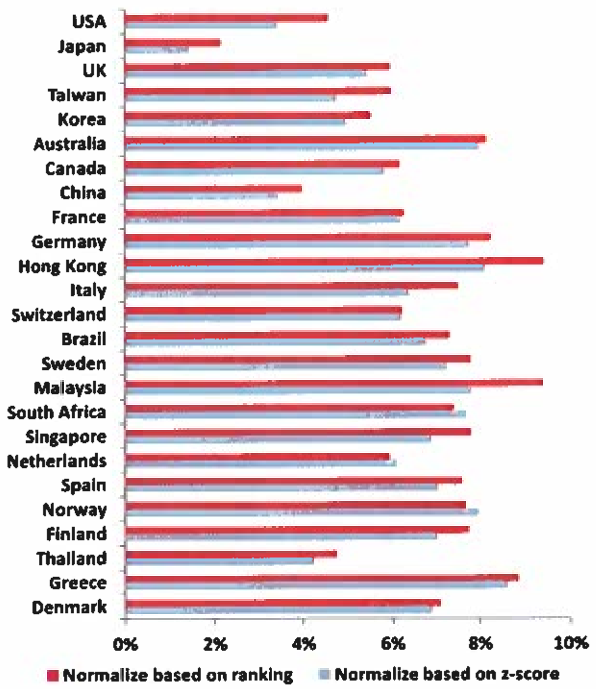
Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 51 compares the performance of our multi-factor models with three different data normalization techniques, for the top 25 countries with the most number of stocks in the MSCI ACWI. It is interesting to see that:

- Ranking transformation outperforms z-score transformation in most countries, which seems to suggest that information is more about rankings – the distance between factor scores does not add incremental information (rather, it may just add noise).

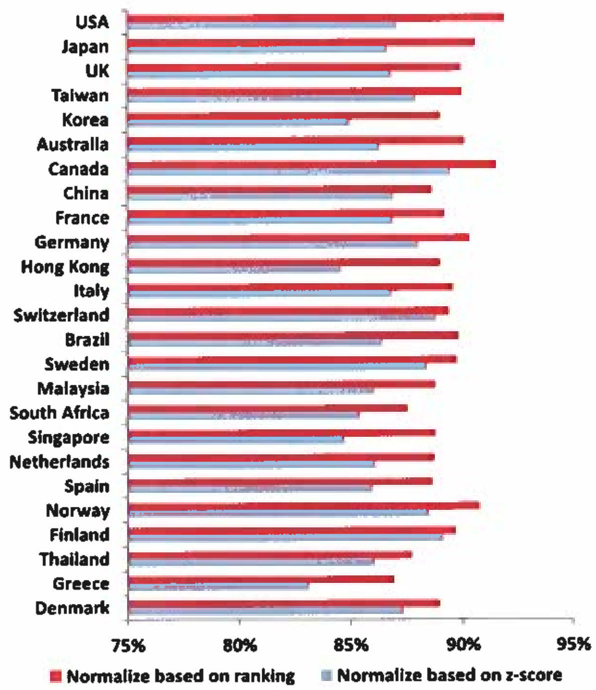
- In addition, the average model signal serial correlation based on the ranking transformation is also higher (see Figure 52), which implies lower turnover.

Figure 51: Average model performance (rank IC), using different data normalization techniques



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 52: Average signal serial correlation, using different normalization techniques



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Do outliers contain any useful information?

The issue with ranking transformation is that we lose the information contained in the distance between factor scores – so long as the factor score of stock A is higher than that of stock B, it does not matter if A is 1% higher or 200% higher than B. For most factors, ranking is all that matters, while distance adds nothing more than noise.

Price momentum, however, may prove to be an exception. In Alvarez, et al. [2011c], we suggest that the distance between momentum scores contains useful information about stocks' future returns. In the same paper, we also propose a new momentum factor called "neutralized momentum", by adjusting each stock's exposure to market beta. We find both the original price momentum and neutralized momentum factors perform better without performing the ranking normalization.

Figure 53 and Figure 54 compare the performance of our neutralized price momentum factor: 1) using the original price momentum[26] scores without the ranking normalization;

Most quant models control for outliers - we are explicitly trying to use them

[26] Here we actually use our neutralized momentum, but the results would be qualitatively similar if we use the traditional 12-1M momentum.

and 2) using the ranking transformed momentum scores. We lose almost 35% of predictive power by normalizing the price momentum factor first.

More intuitively, let us show the impact of normalization using portfolios. We can build a simple portfolio in that each stock's weight is proportional to its momentum score, i.e., signal weighted portfolio.

$$\omega_i \sim \lambda f_i$$

Subject to:

$$\sum_i |\omega_i| = 1$$ [27]

Where,

ω_i is the weight of stock i,

λ is a scaling factor, and

f_i is the price momentum score for stock i

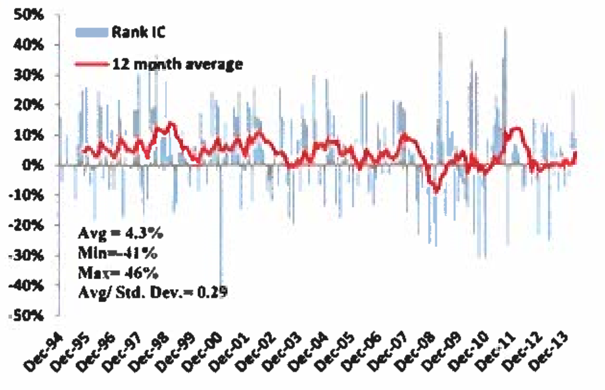
Figure 55 shows the signal weighted portfolio based on the two momentum signals: normalized and non-normalized. Price momentum without normalization beats the ranking transformed factor by 205% in 20 years.

Figure 56 shows the aggregate net exposures for the two momentum portfolios. By construction, half of the stocks rank above the median and the other half fall below. Therefore, the net exposure of our portfolio should be close to zero, i.e., dollar neutral. The portfolio constructed on the raw score, however, takes on significant net long or short positions. Interestingly, it tends to take on net short exposure during bear markets and net long exposure during bull markets, which suggests that the distance between price momentum scores seems to have some market timing ability.

Unlike most other factors, the momentum signal, by definition, is in the return space, which has similar magnitude and distributional properties as expected returns. Therefore, normalization may not be as necessary as for other factors.

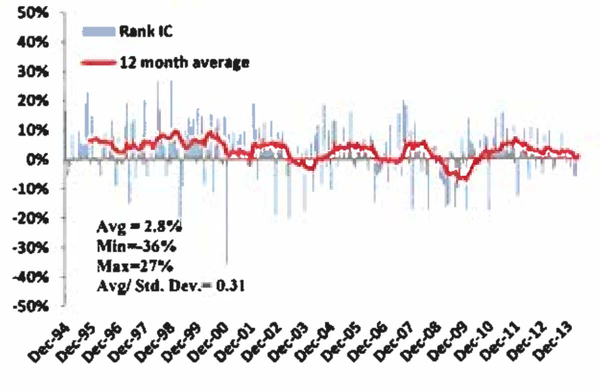
[27] This means that our portfolio can take long and short positions and we do not restrict on net exposures. However, we restrict the gross exposure to be always one, i.e., we do not take any leverage.

Figure 53: Neutralized momentum factor, without the ranking normalization



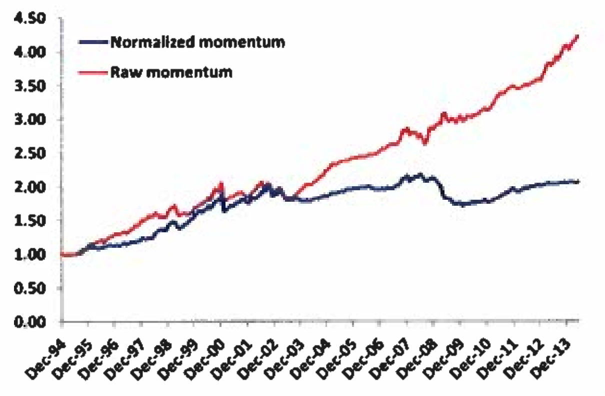
Source: Deutsche Bank

Figure 54: Neutralized momentum factor, with the ranking normalization



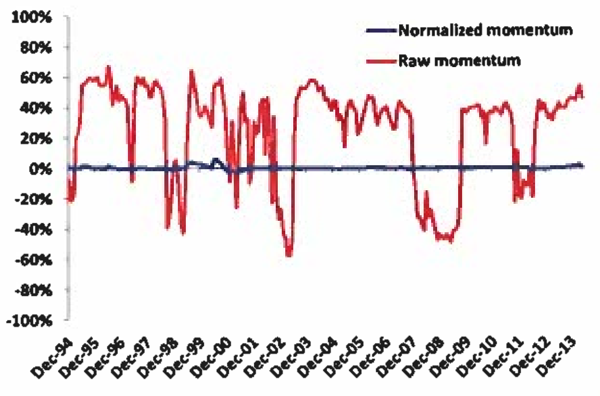
Source: Deutsche Bank

Figure 55: Momentum portfolio performance



Source: Deutsche Bank

Figure 56: Aggregate net exposure



Source: Deutsche Bank

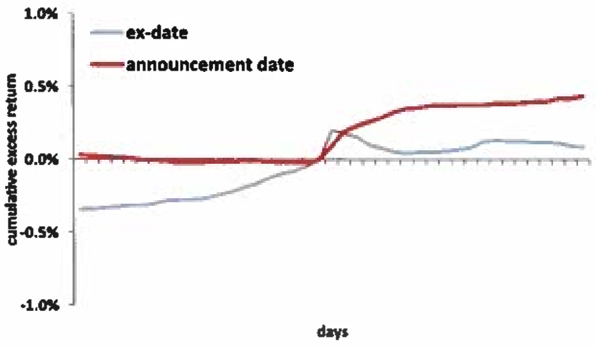
How to calculate excess returns?

When we normalize our data, we have to compute our factors relative to certain universes or benchmarks. Interestingly, the results can be day-and-night depending on which benchmark or universe we use. We use an event study to show the impact of benchmark selection bias.

In an event study, we need to compute abnormal returns. For example, suppose a company announces earnings and the stock rises by 5%, but the market also goes up 3% on the same day. We need to understand how much of the 5% should be attributed to the earnings announcement event. One of the common approaches is to subtract market return from stock return, adjusting or not adjusting for the stock's beta. The market return is dominated by large cap stock; therefore, we may want to use the simple average (or median) return in our investment universe.

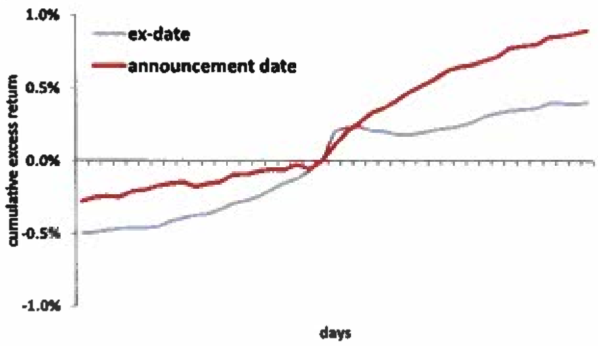
In our global dividend research (see Wang, et al [2014]), we conduct an event study on dividend announcement. We normalize each stock's return by subtracting the average return of all dividend-paying stocks on the same day. As shown in Figure 57, on average, there is no price movement prior to the event date, i.e., there is probably no leakage of dividend announcement information. However, if we choose the wrong benchmark – where we use the broad equity market, e.g., the S&P 500 index, we see stocks actually tend to go up before the dividend announcement (see Figure 58). The reason is possibly due to the fact that dividend-paying stocks tend to earn higher returns than the broad market. Using the wrong benchmark makes it impossible to tell whether the price drift before the dividend announcement is due to the dividend premium or the dividend announcement.

Figure 57: Excess return over the equally weighted average of dividend-paying stocks



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 58: Excess return over the S&P 500 index



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

7. The asymmetric pattern and shorting cost

When we measure factor performance, we typically rely on two metrics:

- **Long/short quantile return spread.** We typically form a long/short hedged portfolio by buying the top ranked stocks in the first quantile, based on the factor of interest; and at the same time, shorting those stocks in the bottom quantile. Stocks in the long (and short) baskets are either equally weighted or capitalization weighted. Typical choices of quantile are tercile, quartile, quintile, decile, and occasionally, percentile, depending on the size of our data sample. The long/short portfolio is rebalanced periodically.
- **Rank information coefficient.** As explained in the previous section, rank IC is the correlation coefficient between: 1) the ranking of stocks based on the scores of a given factor as of a given period end and 2) the ranking of the same universe of stocks based on their forward returns in the next period. Then we repeat the same calculation each period.

Figure 59: Asymmetric pattern

Source: Yin Luo

If a factor's payoff pattern is symmetric, linear, and monotonic, then both approaches should draw a similar conclusion. In this section, we address the issue of asymmetric payoff patterns. In order to measure the efficacy on the long and short sides separately, we calculate the excess return of the following two portfolios:

- **Long portfolio excess return:** long the top quartile stocks (equally weighted) against the average (or median) return of our investment universe (which is equivalent to shorting a basket of all stocks in our universe, equally weighted)
- **Short portfolio excess return:** short the bottom quartile stocks (equally weighted) against the average (or median) return of our investment universe (which is equivalent to using the proceeds from our short positions to fund a long portfolio of all stocks in our universe, equally weighted)

Figure 60 and Figure 61 show the long and the short excess returns for earnings yield and price momentum, respectively. It is interesting to note that the alpha of earnings yield is concentrated on the long side, while the excess return of price momentum is dominated by the short side. Therefore, not only the value factors tend to have slower information decay/lower turnover, but it is also easier to capture the alpha compared to price momentum, because we can collect most of the value premium from the long side, without much of the need of shorting.

Figure 62 shows the long and the short excess returns for 28 common factors in the US. These factors are sorted based on the spread between "short excess return" and "long excess return". The higher up on the list, the more difficult it is to capture the alpha, due to heavier demand for shorting and likely higher shorting cost (shorting cost will be discussed in the next section). Value factors generally collect their premia from the long side, while price momentum/reversal and quality factors generate more alpha from the short side. Analyst revision factors tend to show more symmetric payoff patterns.

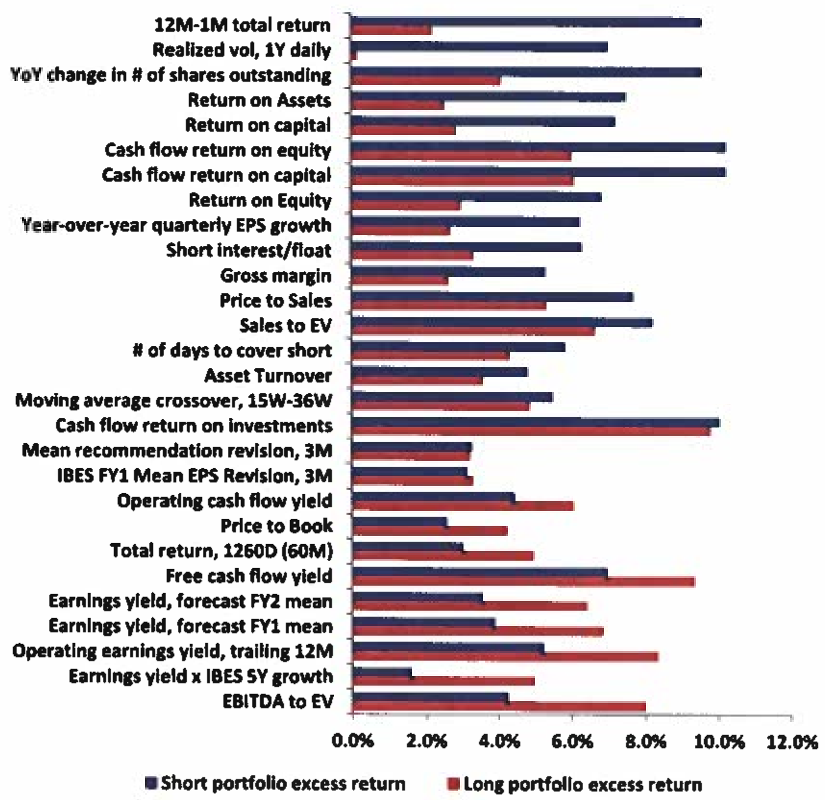
Figure 60: Earnings yield

Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 61: Price momentum

Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 62: The asymmetric payoff pattern



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Understanding the asymmetric nature of factors returns is critical, especially if our mandate is long only or dedicated short.

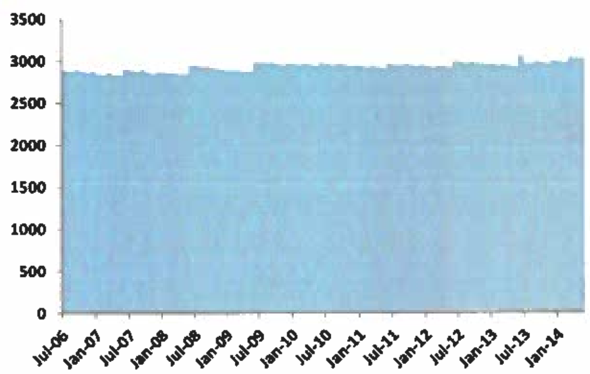
Accounting for short availability and shorting costs

In typical backtesting, analysts generally assume they can short any stocks at no cost or at the same level of cost. However, we need to be aware that borrowing cost can be prohibitively high for some stocks, while on other occasions, it could be impossible to locate the borrowing. For certain stocks or industries or countries, there could also be government or exchange imposed rules that prohibit any shorting at all.

In a previous research paper (see Cahan, et al [2011]), we show how to use the securities lending data from a vendor called Data Explorers[28] in stock selection strategies. A nice feature about the database is that they provide information about short availability and shorting costs for a fairly wide range of stocks globally, almost in real time. They also have point-in-time historical data. One of the interesting data items from the vendor is a cost of borrowing score called DCBS – a discrete number from one to 10, for each stock, on a daily basis. DCBS is indicative of the fee charged by the agent lender, where one stands for the cheapest and 10 stands for the most expensive to borrow.

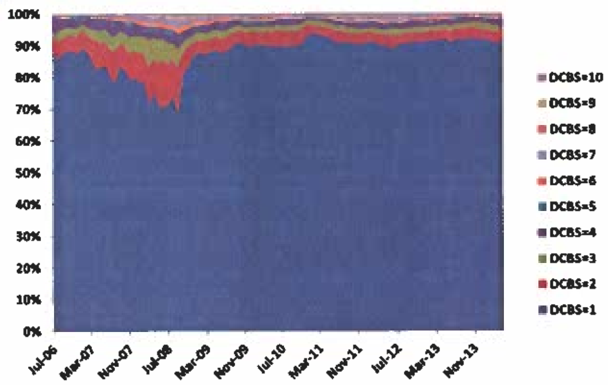
Figure 63 shows the coverage for this factor in the US over time, which matches well the Russell 3000 universe. Figure 64 shows the percentage of stocks across each of the 10 DCBS scores. The data starts in 2006 and almost 90% of US stocks are classified as cheap/easy to borrow (labeled one). During the 2008 global financial crisis period, borrowing costs skyrocketed – some financials stocks were even banned from being shorted, reflected by higher percentages of expensive-to-borrow stocks during this episode.

Figure 63: Coverage for the cost of borrow score (DCBS) from Data Explorers



Source: Bloomberg Finance LLP, Compustat, IBES, Markit, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 64: Cost-of-borrow score composition



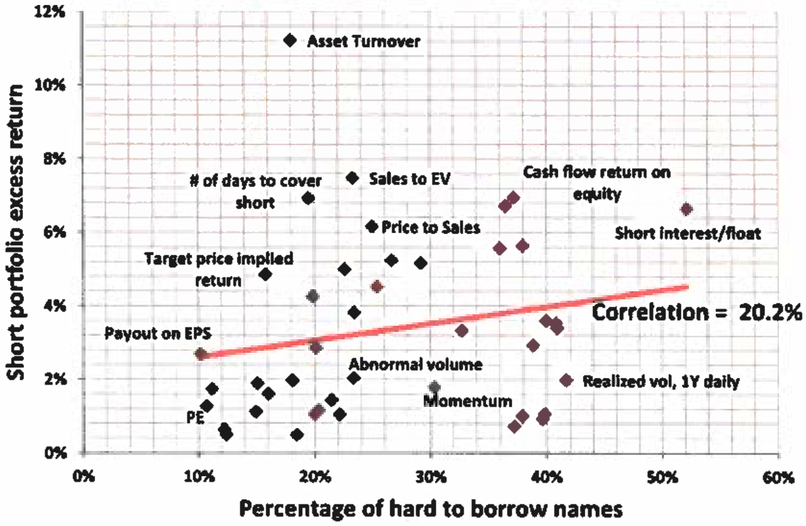
Source: Bloomberg Finance LLP, Compustat, IBES, Markit, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Since there are only around 10% of stocks classified as hard-to-borrow, do we really need to account for short availability in our backtesting? One of the reasons (if not the most important reason) that the poorly ranked stocks by many factors tend to underperform is due to the so-called "limited arbitrage" argument. These stocks are difficult and expensive to short, which prevents arbitrageurs from immediately forcing prices to fair values.

[28] The company, Data Explorers was subsequently acquired by Markit in 2012.

Figure 65 shows the percentage of expensive-to-short (i.e., with DCBS scores above one) stocks in the bottom decile in each of the common factors (the x-axis). In Figure 65, we also display the efficacy of each factor on the short side (the y-axis). Interestingly, those factors that produce greater alphas on the short side often contain more hard-to-borrow names (e.g., short interest/number of shares outstanding, cash flow return on equity, etc.). In addition, almost all the short portfolios on average have over 10% of hard-to-borrow stocks – more than the average 10% of hard-to-borrow names in the overall universe (see Figure 64).

Figure 65: Percentage of hard-to-borrow names in the short portfolio versus short portfolio performance



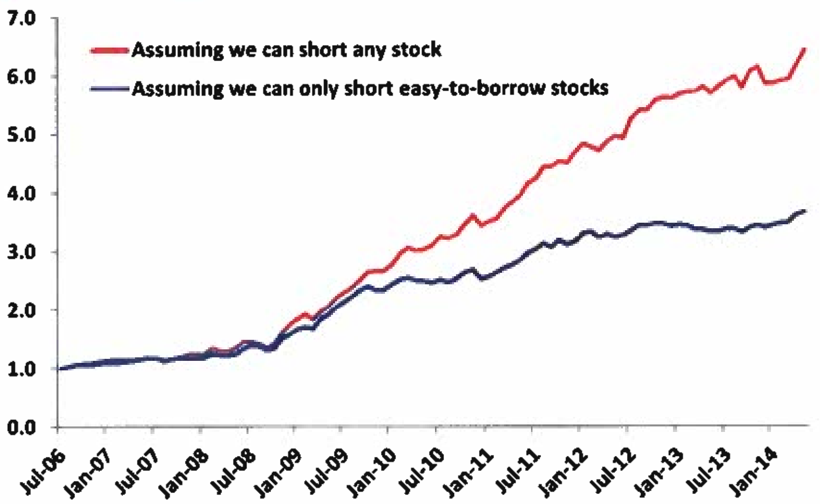
Source: Bloomberg Finance LLP, Compustat, IBES, Markit, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

How much difference would it make if we cannot short those hard-to-borrow stocks? Figure 66 shows the performance of two long/short portfolios based on our N-LASR global stock selection model:

- **Unrealistic portfolio:** we long the top 100 stocks and short the bottom 100 stocks, ranked by the N-LASR model, assuming we can short any stocks.
- **Realistic portfolio:** we long the top 100 stocks and short the bottom 100 stocks, ranked by the N-LASR model. On the short side, we assume that we can only short those easy-to-borrow stocks with DCBS score of one.

Figure 66 shows that adding a realistic constraint on the short side cuts down the cumulative performance by almost half.

Figure 66: Performance with and without short constraints



Source: Bloomberg Finance LLP, Compustat, IBES, Markit, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

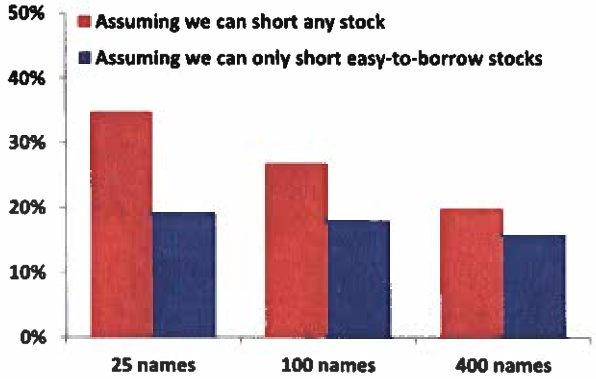
High conviction or diversification

One popular view in the investment world, especially a view shared by many fundamental investors, is that we should fully take advantage of our "high conviction" ideas; therefore, a more concentrated portfolio is more desirable than a portfolio holding hundreds of stocks. On the other hand, some managers (more likely a quant) believe in diversification and typically hold fairly diversified portfolios.

Given the strong out-of-sample and live performance of our N-LASR model, we use it as a proxy to generate high conviction ideas. As shown in Figure 67, in an ideal setup in which we can short any stock, adding more stocks to the portfolio tends to lower performance – as we move from a 25-name portfolio (long top 25 stocks/short bottom 25 stocks) to a 400-name portfolio, active annual return goes down from 35% to 20%. However, once we remove those hard-to-borrow stocks from the backtest, the benefit of a concentrated portfolio shrinks significantly. More importantly, while a highly concentrated portfolio may produce great returns, it comes with the cost of being more volatile as well. From a Sharpe ratio perspective, as shown in Figure 68, if we assume we can short any stock, a more diversified portfolio actually shows slightly stronger performance. Once we adjust for those hard-to-borrow names, the diversification benefit clearly outweighs higher conviction. A more diversified portfolio with 800 stocks (400 long/400 short) improves the Sharpe ratio from 1.5x (based on a portfolio of 25 long/25 short) to 2.2x (almost 50%).

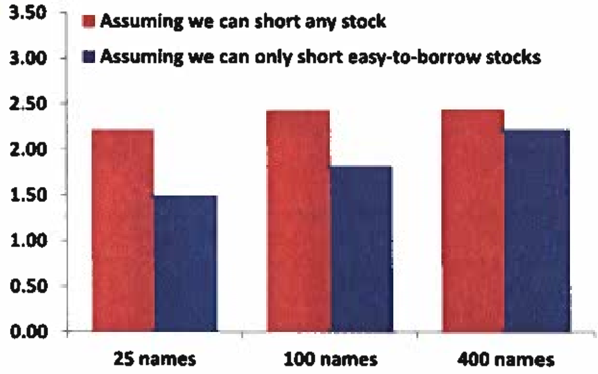
We also note an interesting pattern that the constraint on hard-to-borrow stocks has a much stronger impact on more concentrated portfolios. As shown in Figure 68, the short constraint reduces the Sharpe ratio of the concentrated portfolio (25 long/25 short) by over 30%, but trims down the performance by less than 10% for a more diversified portfolio (400 long/400 short), which suggests that the worst ranked stocks by the N-LASR model are more likely to have relatively high borrowing cost.

Figure 67: Annualized return, long/short N-LASR portfolios



Source: Bloomberg Finance LLP, Compustat, IBES, Markit, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 68: Sharpe Ratio, long/short N-LASR portfolios



Source: Bloomberg Finance LLP, Compustat, IBES, Markit, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

A hands-on tutorial of how to build a realistic model

In this section, we show a hands-on tutorial of how to build a multi-factor model, based on five commonly used factors:

- Value (trailing 12 month earnings yield)
- Growth (year-over-year quarterly EPS growth)
- Quality (ROE)
- Momentum (12-1M total return), and
- Sentiment (IBES three-month earnings revision)

The purpose of this exercise is not meant to show the best possible model – rather, we intend to use a real-life example to show our readers how to avoid the seven sins discussed in the previous section. The ultimate goal is to construct a model that hopefully will not only work well in-sample, but also survive out-of-sample.

How to avoid the seven sins

We set up our backtesting framework to avoid each of the seven sins discussed above:

- **Survivorship bias.** We perform our backtesting on the Russell 3000 index universe, using those companies in the index as of a given point in time.
- **Look-ahead bias.** We use point-in-time data to calculate all of our factors. Company fundamental data is sourced from Compustat point-in-time database, which reflects whatever was available at each month end.
- **Story telling and data history.** We follow the convention for the direction of each factor: buying stocks that are cheaper, that enjoy higher growth, that are more profitable, that have stronger price momentum, and that have more positive analyst sentiment. Our backtesting is conducted over the past 20 years, from 1994 to 2014, covering multiple economic cycles.
- **Data mining and data snooping bias.** The four factor weighting algorithms are extensively tested across multiple countries/regions and asset classes.
- **Signal decay and turnover.** We avoid fast decay factors in this exercise. Portfolio performance is computed after transaction costs.
- **Outlier control.** We use our proprietary data normalization technique to transform each factor to a standard normal distribution, before we combine them together into multi-factor models.
- **The asymmetric payoff pattern and shorting cost.** We study the impact of short availability in detail in this section.

An essay of various factor weighting algorithms

As we move from single factor backtesting to multi-factor models, one of the most common questions that we hear from clients is how to weight various factors in a multi-factor model. Alternatively, clients may invest in factor portfolios (more on this topic later), which also requires portfolio construction algorithms to put these factor portfolios (or smart beta portfolios) together.

We have published extensively in this space:

- In "Quant GEM" (see Wang, et al [2013b]), we compare a few factor weighting algorithms in the global emerging market space: equally weighting, alpha weighting, risk parity weighting, Grinold & Kahn (i.e., mean-variance optimization) weighting, and alpha risk parity weighting.
- Factor performance is regime dependent. In a different macroeconomic environment, factor performance can be drastically different. In "DB Quant Handbook" (see Luo, et al [2010a]), "Style Rotation" (see Luo, et al [2010b]), and "From Macro to Micro" (see Luo, et al [2012b]), we show how to incorporate macroeconomic regimes in our factor weighting decisions.
- Mean-variance optimization is the dominate approach for portfolio construction. In "Robust Factor Models" (see Luo, et al [2011a]), we study how a Bayesian shrinkage estimator of the factor covariance matrix, conditional on the VIX index can improve the performance of mean-variance optimized factor weighting algorithms, e.g., the Grinold & Kahn approach.
- In "Tail Risk in Optimal Signal Weighting" (see Luo, et al [2011b]), we demonstrate the benefit of incorporating higher moments (e.g., coskewness and cokurtosis matrices) in factor weighting.
- In "Risk Parity and Risk-based Allocation" (see Alvarez, et al [2011d]), we introduce our proprietary "alpha risk parity" weighting algorithm. The alpha risk parity approach takes into account the factor performance expectation, but is less sensitive to alpha prediction than the traditional mean-variance optimization algorithms.
- The most comprehensive review of portfolio construction techniques is offered in "DB Handbook of Portfolio Construction, Part 1" (see Luo, et al [2013a]). In this research, we also propose a suite of new factor and portfolio construction techniques. In particular, we show the benefit of our proprietary "minimum tail dependence", "minimum variance-tail dependence", and "robust minimum conditional value-at-risk" portfolios.
- In "DB Handbook of Portfolio Construction, Part II" (see Luo, et al [2013b]), we further illustrate how to improve the risk estimation side of portfolio construction. We find that, regardless of the portfolio construction technique, portfolio performance almost always improves with better risk estimates.

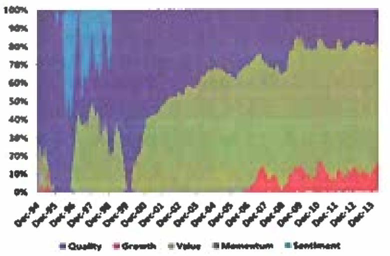
In this research, we compared four factor weighting algorithms:

- **Equally weighting:** each of the five factors is allocated a fixed 20% weight.
- **Grinold & Kahn:** we perform a simple mean-variance optimization at each month end. Both expected factor returns and covariance matrix are computed using an expending window.
- **Alpha risk parity:** we perform the alpha risk parity optimization at each month end. Both expected factor returns and covariance matrix are computed using an expanding window. The main idea of alpha risk parity is to allocate the same amount of risk budget per unit of alpha. In this scheme, a factor with a higher expected return, and/or a lower volatility, and/or lower correlations with other factors, gets a higher weight.

- **Minimum tail dependence:** we perform the minimum tail dependence optimization at each month end. The tail dependence matrix is computed using an expanding window. The tail dependence method was introduced in our previous research as one of the most effective portfolio optimization techniques; see Luo, et al [2013a, 2013b] for details. It attempts to find assets that are less dependent to other assets at the tail level to avoid the crowded trade. Those factors that are less dependent to other factors get higher weights.

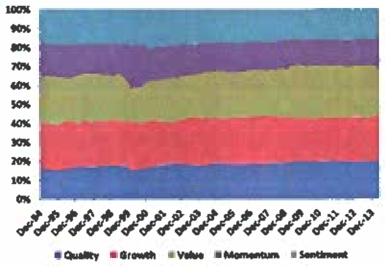
Figure 69 to Figure 71 compare the factor weighs over time, for each of the three dynamic factor weighting algorithms, respectively. Grinold & Kahn's approach, due to the nature of mean-variance optimization, tends to be quite volatile and concentrated – the model is mostly dominated by the value factor in recent years (see Figure 69). At the other extreme, the alpha risk parity approach delivers the most stable and balanced allocation among the five styles (see Figure 70). The downside is that it may not be adaptive enough to the changes in the market environment. Our minimum tail dependence algorithm seems to strike a good balance between stability and adaptation (see Figure 71).

Figure 69: Grinold & Kahn factor weights



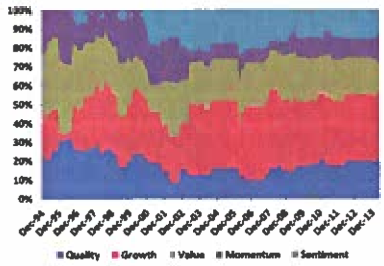
Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 70: Alpha risk parity factor weights



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 71: Minimum tail dependence factor weights

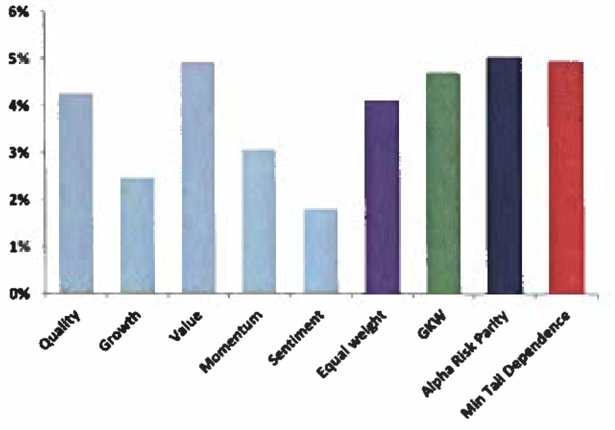


Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Performance wise (see Figure 72 and Figure 73), we find:

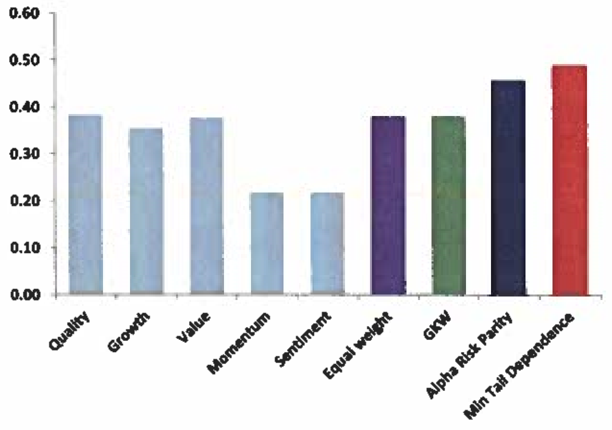
- Almost all multi-factor models outperform any of the five single factors on a risk-adjusted basis (see Figure 73).
- The more sophisticated factor weighting schemes outperform the more naïve equally weighted algorithm, especially on a risk adjusted basis (see Figure 73).
- Comparing the alpha risk parity and minimum tail dependence approaches, the average return is about the same (see Figure 72), but the minimum tail dependence algorithm clearly dominates after adjusting for risk (see Figure 73), which suggests that the minimum tail dependence algorithm has better risk control.

Figure 72: Average rank IC



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 73: Risk adjusted rank IC



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Risk control at the model building stage – country and sector neutralization

Traditionally, analysts focus on model building, while portfolio managers are responsible for risk control at the portfolio level. In this section, we show the benefit of adding some risk control at the alpha model construction stage.

As we know, company characteristics (e.g., valuation, growth profile, and profitability) vary greatly from country to country, and from industry to industry. For example, technology companies are generally more expensive than utilities stocks, but enjoy greater growth potential. A model that ranks stocks regardless of their country/sector (and other contextual differences) essentially engages not only in stock selection, but also country/sector rotation. Because the breadth in stock selection is generally much larger than that in country/industry rotation[29], it is typically better to use our risk budget in stock selection rather than country/sector rotation.

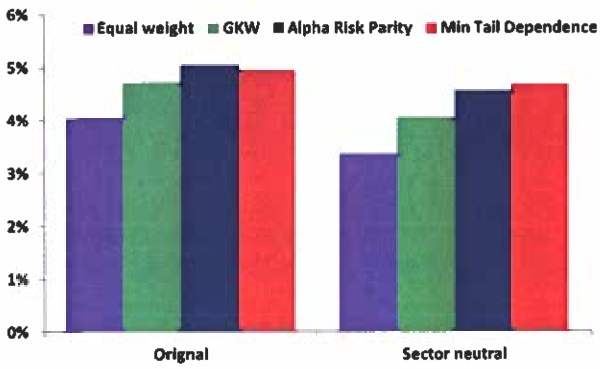
One way to make our stock selection model more robust and less volatile is to control for country/sector differences via a technique called neutralization. We have applied country/sector neutralization in our N-LASR global stock selection model (see Wang, et al [2013a]) and our global emerging market models (see Wang, et al [2013b]).

In this section, we use sector neutralization as an example, but the benefit is equally applicable to country neutralization for regions with more than one country. The method of sector (or country) neutralization is quite simple. We essentially normalize each factor within each of the sectors (or countries), by mapping the factor scores into a standard normal distribution (as we discussed in the previous section); therefore, stocks from different sectors (or countries) are more comparable.

After sector neutralization, although the average model performance drops modestly (see Figure 74), the risk adjusted performance improves considerably (see Figure 75), regardless of the factor weighting algorithms used.

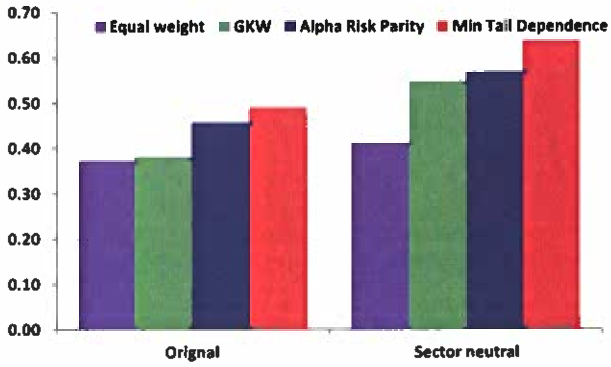
[29] There are typically hundreds or thousands of stocks, but only a handful of countries and/or industries.

Figure 74: Average rank IC



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 75: Risk adjusted rank IC



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Smart beta investing via factor portfolios

In most of our previous research (see Wang, et al [2013b, 2014]), we construct multi-factor models by combining a suite of factors and then construct portfolios of stocks using these multi-factor models – hereafter called active portfolio management via "multi-factor models". An alternative approach is to build a suite of factor portfolios first (each factor portfolio comprises a number of stocks), and then conduct an asset allocation exercise by combining these factor portfolios – hereafter called smart beta investing via "multi factor-portfolios".

This first approach is often used by active equity portfolio managers, while the second approach has gained great popularity among asset allocators in recent years as factor investing becomes more main stream. The multi-factor models give portfolio managers the flexibility of constructing their proprietary models, by selecting more unique factors. On the other hand, the "multi factor-portfolios" approach shifts the power from asset managers to asset owners. Asset owners now have more flexibility to construct their portfolios by mixing a suite of beta and smart beta (i.e., factor portfolios) strategies.

To construct factor portfolios, we perform the following optimization at each month end, for each of the five factors separately:

- Long/short market neutral strategy
- 2x leverage, i.e., for $1 capital, the strategy invests in $1 long and $1 short
- Target annualized volatility of 6%
- Maximum single stock weight 1.5%
- Beta neutral (maximum 0.1 net beta exposure)
- Sector neutral (maximum 10% absolute sector exposure)
- Turnover constrained at 30% one-way per month (60% two-way turnover)
- Transaction cost 20 bps per trade

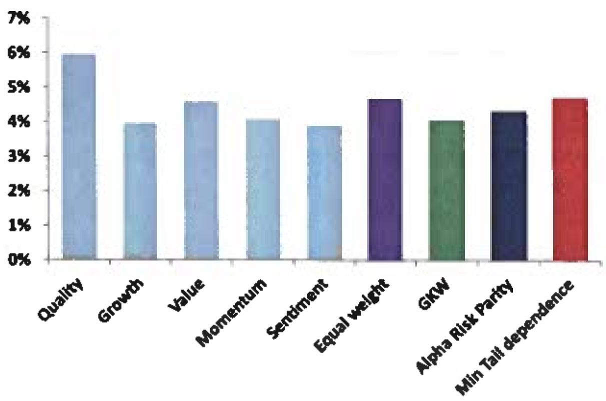
As shown in Figure 77, the realized volatilities of all five factor portfolios are higher than our target risk of 6% per annum, which should not be a surprise, as risk models on average tend to under-estimate risk in optimized portfolios (see Alvarez, et al [2012]).

After constructing each of the five factor portfolios, we then build an overall asset allocation strategy by mixing the five factor portfolios, using the same four portfolio construction techniques:

- Equally weighting
- Mean-variance optimization
- Alpha risk parity
- Minimum tail dependence

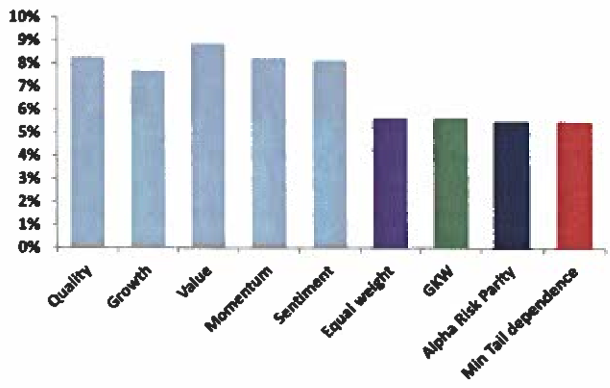
The overall asset allocation strategies tend to have comparable returns as the single factor portfolios (see Figure 76), but much lower volatilities (see Figure 77); and therefore dominate single factor portfolios on a risk adjusted basis (see Figure 78). The strength of our minimum tail dependence portfolio construction technique is clear – it has the second highest return (see Figure 76) and the highest Sharpe ratio (see Figure 78), but more importantly, it demonstrates a far superior ability of downside risk control (see Figure 79).

Figure 76: Annualized return



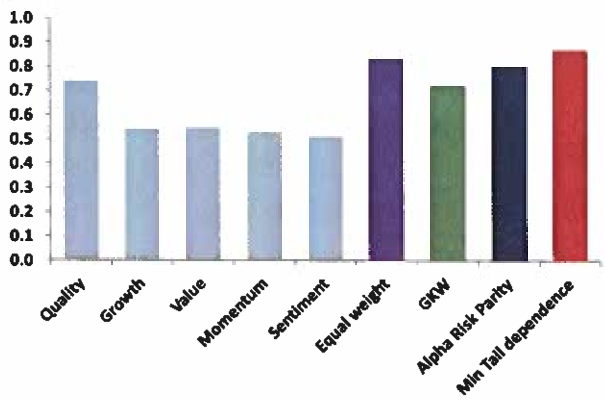
Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 77: Annualized volatility



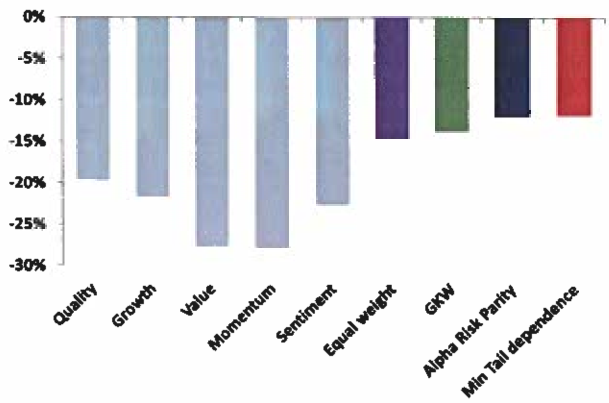
Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 78: Sharpe ratio



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 79: Maximum drawdown



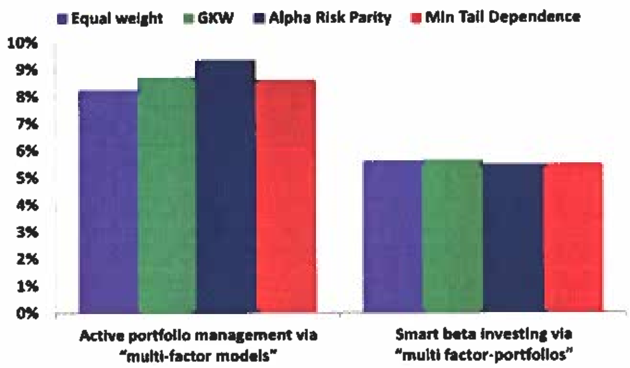
Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Active portfolio management versus smart beta investing

In this section, we show an interesting comparison of the two approaches: active portfolio management via multi-factor models versus smart beta investing via multi factor-portfolios. There are pros and cons for each of the two approaches:

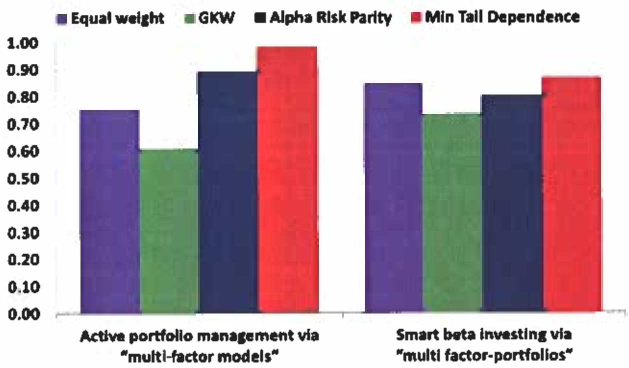
- Active portfolio management via multi-factor models tend to have higher realized risk than smart beta portfolios (see Figure 80). The second stage optimization in multi factor-portfolios further reduces risk.
- Active portfolio management via multi-factor models tends to produce higher Sharpe ratios (see Figure 81) – especially with more sophisticated portfolio construction techniques like alpha risk parity and minimum tail dependence, as these models are more efficient than multi factor-portfolios. For example, one stock might be held as a long position in one factor portfolio, while it could be shorted in another; therefore, smart beta investing via multi factor-portfolios does not allow netting and tends to be less efficient.
- The biggest benefit of smart beta via multi factor-portfolios is that it empowers asset owners (or GTAA portfolio managers) by providing additional investment instruments to their asset allocation strategies. Managers can not only invest in traditional asset classes (e.g., equities, bonds, currencies), but also across factor portfolios.
- On the surface, the fact that active portfolio management only marginally outperforms smart beta investing might be a little surprising. However, that is because we only use the same five simple and naïve factors in both active management and smart beta simulations. In practice, active managers are more likely to have more unique and proprietary factors in their multi-factor models[30].

Figure 80: Realized portfolio risk



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 81: Sharpe ratio



Source: Bloomberg Finance LLP, Compustat, IBES, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

30 It once again highlights the importance of having unique, less crowded, and unconventional factors in a multi-factor models for an active portfolio manager.

Shorting? Is it really worth the effort?

Many research papers would argue that there are more market inefficiencies on the short side than on the long side. As we stated in earlier sections, the problem is that, in practice, we cannot short every single stock and also shorting is not free, not to mention the additional administrating costs involved in long/short strategies.

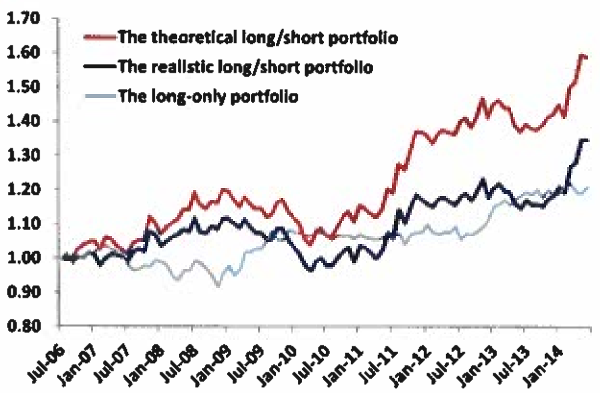
In this section, we show a real life example of three portfolios:

- **The theoretical long/short portfolio:** we can short any stocks without any extra cost
- **The realistic long/short portfolio:** we can only short those easy-to-borrow stocks
- **The long-only portfolio:** we build a long-only optimized portfolio and short the S&P 500 futures to hedge

From Figure 82 and Figure 83, we find:

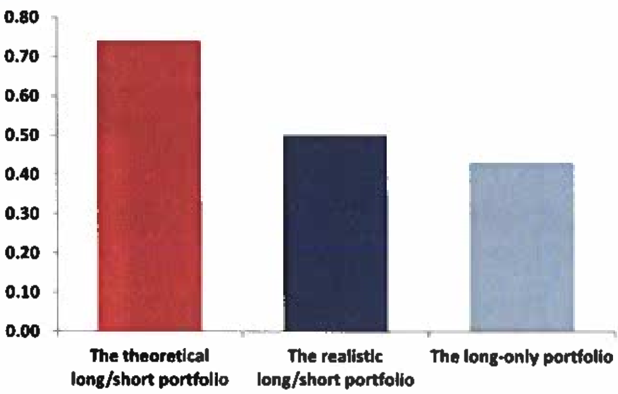
- The theoretical backtesting that ignores short availability inflates the Sharpe ratio by 48%.
- Even adjusting for hard-to-borrow names, our long/short portfolio still outperforms the long-only strategy by 16%.

Figure 82: Cumulative performance



Source: Bloomberg Finance LLP, Compustat, IBES, Markit, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Figure 83: Sharpe ratio



Source: Bloomberg Finance LLP, Compustat, IBES, Markit, Russell, S&P, Thomson Reuters, Worldscope, Deutsche Bank Quantitative Strategy

Bibliography

Alvarez, M, Y. Luo, R. Cahan, J. Jussa, and Z. Chen [2011a]. "Minimum Variance: Exposing the 'Magic'", *Deutsche Bank Quantitative Strategy*, 9 February 2011

Alvarez, M, Y. Luo, R. Cahan, J. Jussa, and Z. Chen [2011b]. "Learning to Drive in the Fast Lane", *Deutsche Bank Quantitative Strategy*, 26 April 2011

Alvarez, M, Y. Luo, R. Cahan, J. Jussa, and Z. Chen [2011c]. "Reviving Momentum: Mission Impossible?", *Deutsche Bank Quantitative Strategy*, 6 July 2011

Alvarez, M, Y. Luo, R. Cahan, J. Jussa, and Z. Chen [2011d]. "Risk Parity and Risk Based Allocation", *Deutsche Bank Quantitative Strategy*, 13 October 2011

Alvarez, M, Y. Luo, R. Cahan, J. Jussa, Z. Chen, and S. Wang [2012]. "Risk & Alpha Alignment", *Deutsche Bank Quantitative Strategy*, 22 August 2012

Cahan, R., Y. Luo, M. Alvarez, J. Jussa, and Z. Chen [2011]. "The Long and the Short of it", *Deutsche Bank Quantitative Strategy*, 18 January 2011

Cahan, R., Y. Luo, M. Alvarez, J. Jussa, Z. Chen, and Wang, S. [2012]. "Standing out from the Crowd", *Deutsche Bank Quantitative Strategy*, 31 January 2012

Fama, E. F.; and French, K. R. [1992]. "The Cross-Section of Expected Stock Returns". The Journal of Finance 47 (2): 427

Fama, E. F.; and French, K. R. [1993]. "Common risk factors in the returns on stocks and bonds". *Journal of Financial Economics* 33: 3

Forsythe, R., Palfrey, T., and Plott, C.R. [1982]. "Asset Valuation in an Experimental Market", Econometrica, 50(3), 537-568

Grinold, R., and Kahn, R. [1999]. *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk*, McGraw-Hill, 2nd Edition, October 26, 1999

Langer, E., Blank, A., and Chanowitz, B. [1978]. "The Mindlessness of Ostensibly Thoughtful Action: The Role of 'Placebic' Information in Interpersonal Interaction", *Journal of Personality and Social Psychology*, 36, 635-642

Luo, Y., R. Cahan, J. Jussa, and M. Alvarez [2010a]. "DB Quant Handbook", *Deutsche Bank Quantitative Strategy*, 22 July 2010.

Luo, Y., R. Cahan, J. Jussa, and M. Alvarez [2010b]. "Style Rotation", *Deutsche Bank Quantitative Strategy*, 7 September 2010.

Luo, Y., R. Cahan, M. Alvarez, J. Jussa, and Z. Chen [2011a]. "Robust Factor Models", *Deutsche Bank Quantitative Strategy*, 24 January 2011

Luo, Y., R. Cahan, M. Alvarez, J. Jussa, and Z. Chen [2011b]. "Tail Risk in Optimal Signal Weighting", *Deutsche Bank Quantitative Strategy*, 7 June 2011

Luo, Y., R. Cahan, M. Alvarez, J. Jussa, and Z. Chen [2011c]. "Quant Tactical Asset Allocation (QTAA)", *Deutsche Bank Quantitative Strategy*, 19 September 2011

Luo, Y., S. Wang, R. Cahan, M. Alvarez, J. Jussa, and Z. Chen [2012a]. "New Insights in Country Rotation", *Deutsche Bank Quantitative Strategy*, 9 February 2012

Luo, Y., S. Wang, R. Cahan, M. Alvarez, J. Jussa, and Z. Chen [2012b]. "From Macro to Micro", *Deutsche Bank Quantitative Strategy*, 2 May 2012

Luo, Y., S. Wang, R. Cahan, M. Alvarez, J. Jussa, and Z. Chen [2013a]. "DB Handbook of Portfolio Construction: Part 1", *Deutsche Bank Quantitative Strategy*, 30 May 2013

Luo, Y., S. Wang, J. Jussa, Z. Chen and M. Alvarez [2013b]. "Dynamic Allocation: Extremes, Tail dependence, and Regimes", *Deutsche Bank Quantitative Strategy*, 30 May 2013

Luo, Y., Chen, Z., Wang, S., Alvarez, M., Jussa, J., and Wang, A. [2014]. "Surprise! DB Handbook of Earnings (and beyond Earnings) Surprise", *Deutsche Bank Quantitative Strategy*, March 10, 2014

Mesomeris, S., M. Salvini, J. Avettand-Fenoel, A. Kassam [2010]. "Point-In-Time Data: Backtesting without Bias", *Deutsche Bank Quantitative Strategy*, 8 December 2010

McLean, R., J. Pontiff [2014] "Does Academic Research Destroy Stock Return Predictability?", SSRN Working Paper: http://ssrn.com/abstract=2156623

Plott, C.R., and Sunder, S. [1982]. "Efficiency of Experimental Security Markets with Insider Information: An Application of Rational Expectations Models", Journal of Political Economy, 90(4), 663-698

Smith, V.L., Suchanek, G., and Williams, A. [1988]. "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets", Econometrica, 56(5), 1119-1151

Upton, G., and Cook, I. [1996]. *Understanding Statistics*, Oxford University Press

Wang, S., Y. Luo, R. Cahan, M. Alvarez, J. Jussa, and Z. Chen [2012]. "The Rise of the Machines", *Deutsche Bank Quantitative Strategy*, 5 June 2012

Wang, S., Y. Luo, R. Cahan, M. Alvarez, J. Jussa, Z. Chen [2013a]. "The Rise of the Machines II", *Deutsche Bank Quantitative Strategy*, 24 February 2013

Wang, S., Y. Luo, M. Alvarez, J. Jussa, Z. Chen [2013b]. " Quant GEM", *Deutsche Bank Quantitative Strategy*, 2 October 2013

Wang, S., Y. Luo, M. Alvarez, J. Jussa, A. Wang [2014]. "Global Dividend Investing", *Deutsche Bank Quantitative Strategy*, 2 October 2013

Zwillinger, D., and Kokoska, S. [2000]. *CRC Standard Probability and Statistics Tables and Formulae*, CRC Press

Appendix 1

Important Disclosures

Additional information available upon request

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. Sheng Wang/Yin Luo/Miguel-A Alvarez/Javed Jussa/Allen Wang/Gaurav Rohal

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.