ArbMakerPlus

Arbitrage software

Quick guide to Principles & Practicalities

Principles

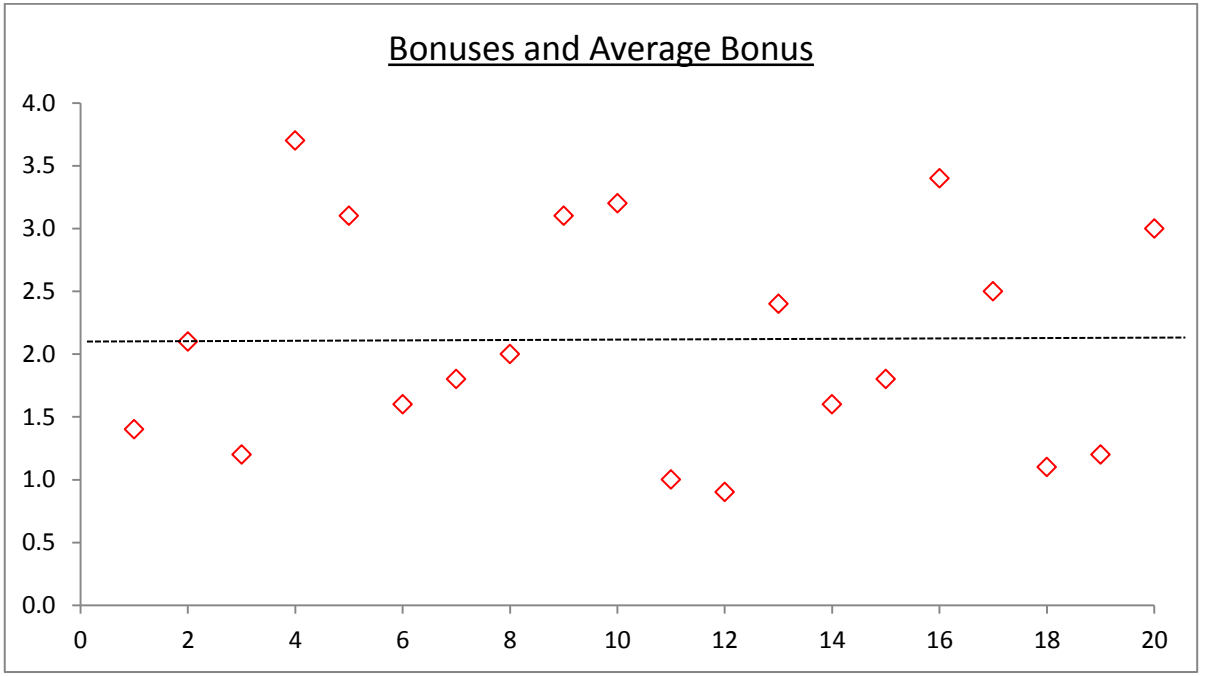
What is the key data users rely on to find potential trades with ArbMakerPlus cointegration software?

Something called ‘residuals’.

Picture the end-of-year bonus party at JP Morgan. Traders line up to get their bonuses. Maybe Jamie Dimon is handing them out. Each trader gets paid according to performance. Possibly, were they present, the CDS traders who lost $2bn (and counting) this year would be handing Mr Dimon cash to keep their jobs.

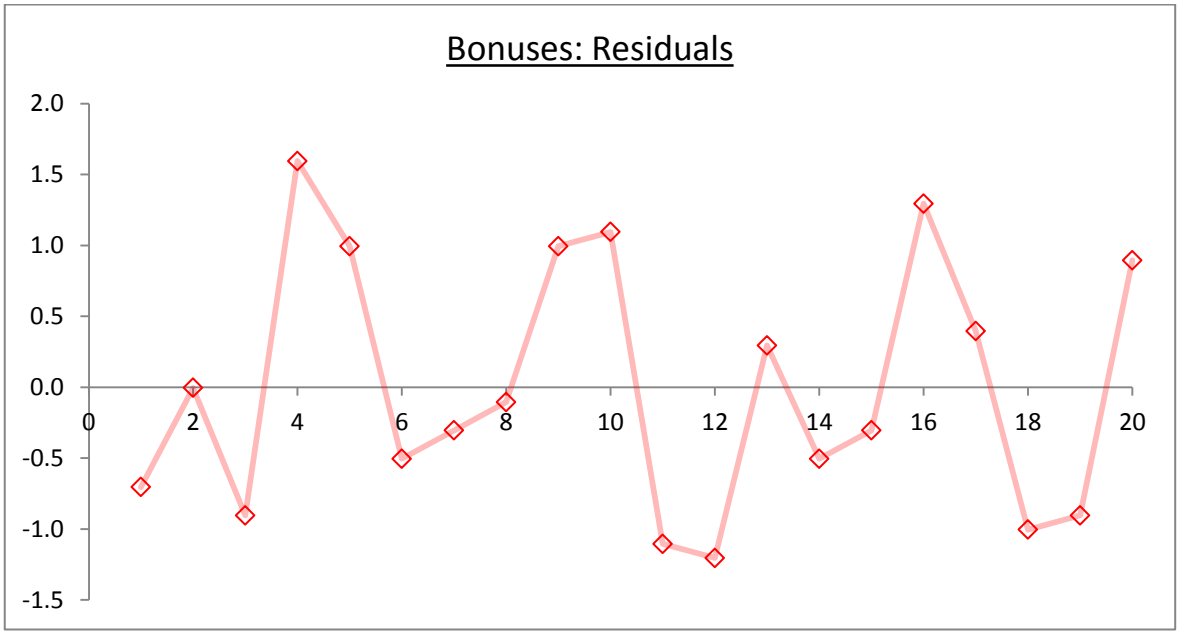
But at the end of the process let’s say the average bonus is $2.1m. Some traders got more than the average and some less. All bonuses could be plotted in graph form using this average of $2.1 as a line of best fit. Some bonuses would fall above the average, some below.

These *differences of each bonus from the average bonus are called residuals*. Statisticians also refer to them as ‘fitting errors’ because they do not fit the average line.



Why are residuals useful?

In short, because they always must sum to zero – that is, they *revert* to the average. Take a look below at the residuals derived from the bonus graph above:



Bonus in $m:	1.4	2.1	1.2	3.7	3.1	1.6	1.8	2	3.1	3.2	1	0.9	2.4	1.6	1.8	3.4	2.5	1.1	1.2	3	**2.1**	**Average**
Residuals:	-0.705	-0.005	-0.905	1.595	0.995	-0.505	-0.305	-0.105	0.995	1.095	-1.105	-1.205	0.295	-0.505	-0.305	1.295	0.395	-1.005	-0.905	0.895	**0.000**	**Sum of residuals**

It is bonus levels that drive the value of each residual. The analogy to financial prices is that when these rise (fall) and push residuals *away* from their average trading opportunities arise to sell (buy) the underlying instruments. Why? Because of the mean-reverting properties of residuals.

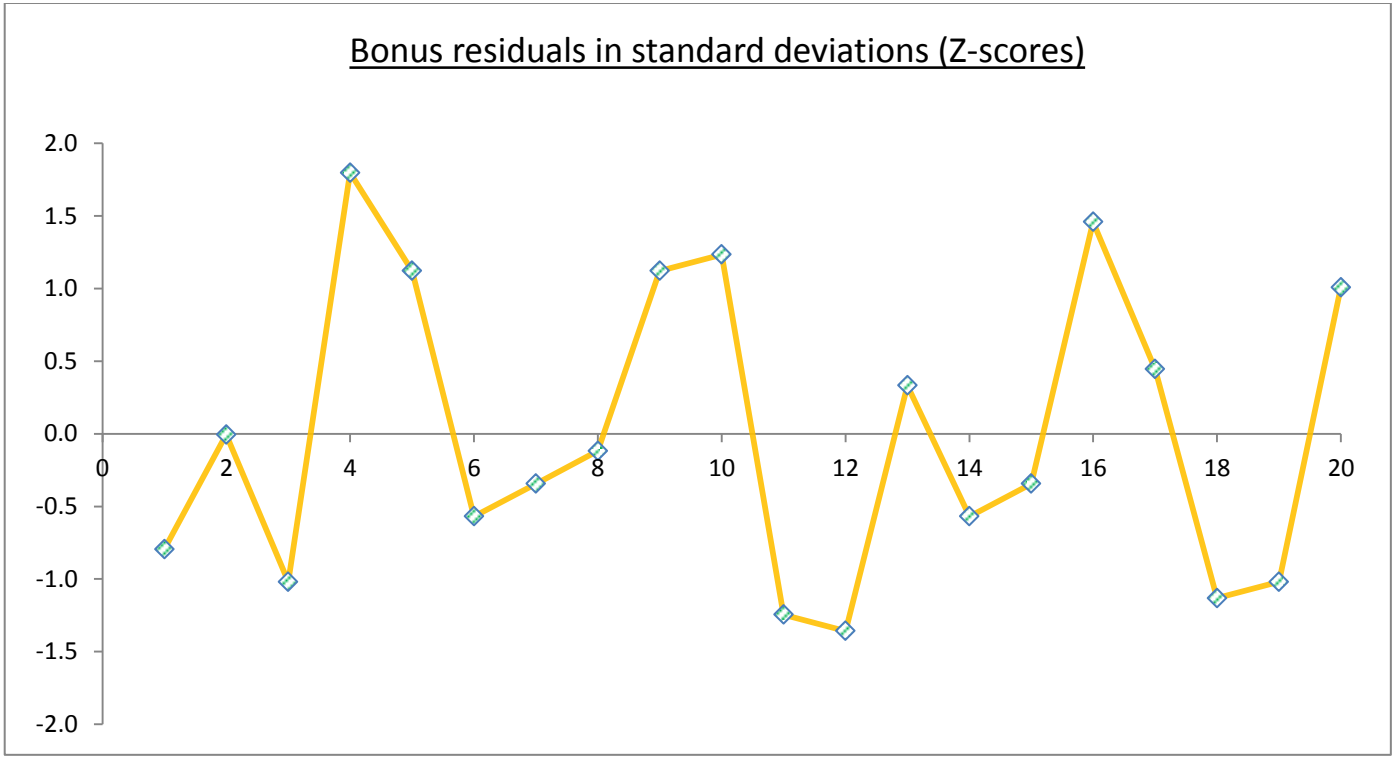
In short, it is starting to look like residuals at the extremes of the observed values permit the direction of the subsequent observations to be assigned probabilities and traded upon.

Two crucial caveats

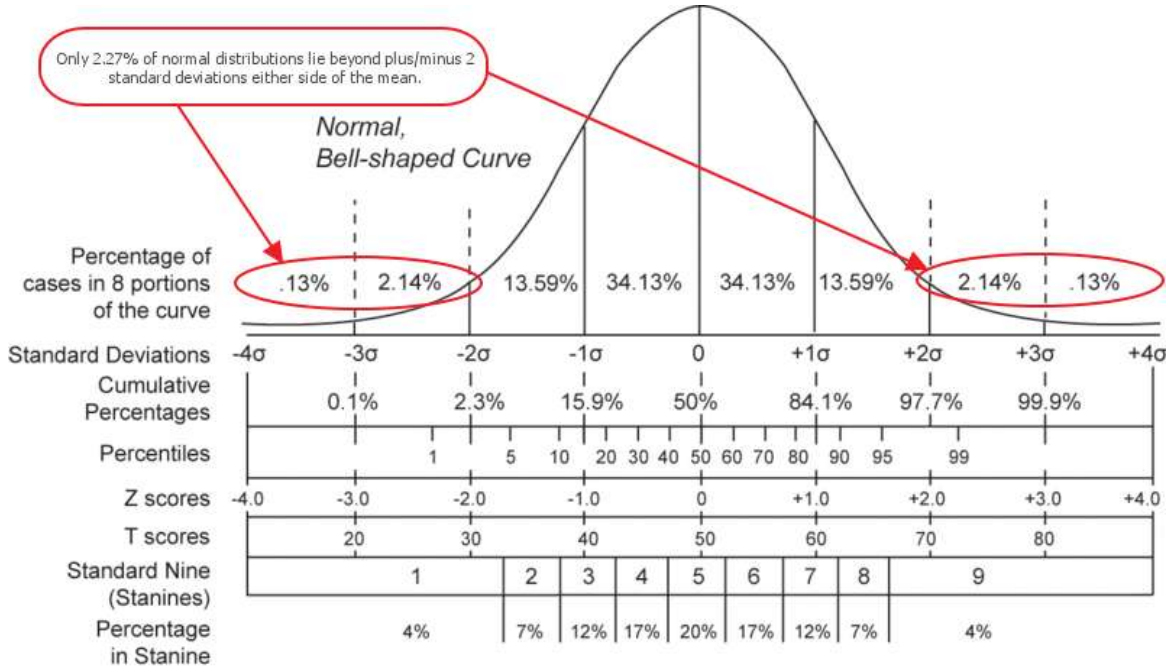
Before taking any position based on residual data it is crucial that:

1. The residuals should be normally, or near normally, distributed. This is because normal (or Gaussian) distributions have cast-iron probability qualities essential for trading; and
2. The residuals should be mutually independent and what statisticians call “independent and identically distributed” – or random.

Assume that this is the case in our bonus example. To take advantage of normal residual data it is helpful to transform each observation to a ‘Z-score’. This is done by dividing each observation by the standard deviation of the population. What you get, measured in standard deviations, is how far each data point is from the mean of the population. This distance is also called the Z-score. In graph form:



Here is where normality is crucial. Under a normal, or Gaussian, distribution the probability of an observation being, say, 2 standard deviations (ie Z-scores) away from the mean is less than 5%. The next graphic sets out the entire probability distribution for normal data, permits specific probabilities to be assigned to observed data and thus allows assessment of the likely direction of future observations:



ArbMakerPlus analyzes the residuals from linear equations as a part of its cointegration process and not fictitious bonus rewards. But the principles above apply in the same way. Thus if the incoming data continue to be normal and cointegrated then probabilities about the trend direction of those incoming observations can be determined – they will revert back to the mean.

It is a simple but powerful method of modeling uncertainty.

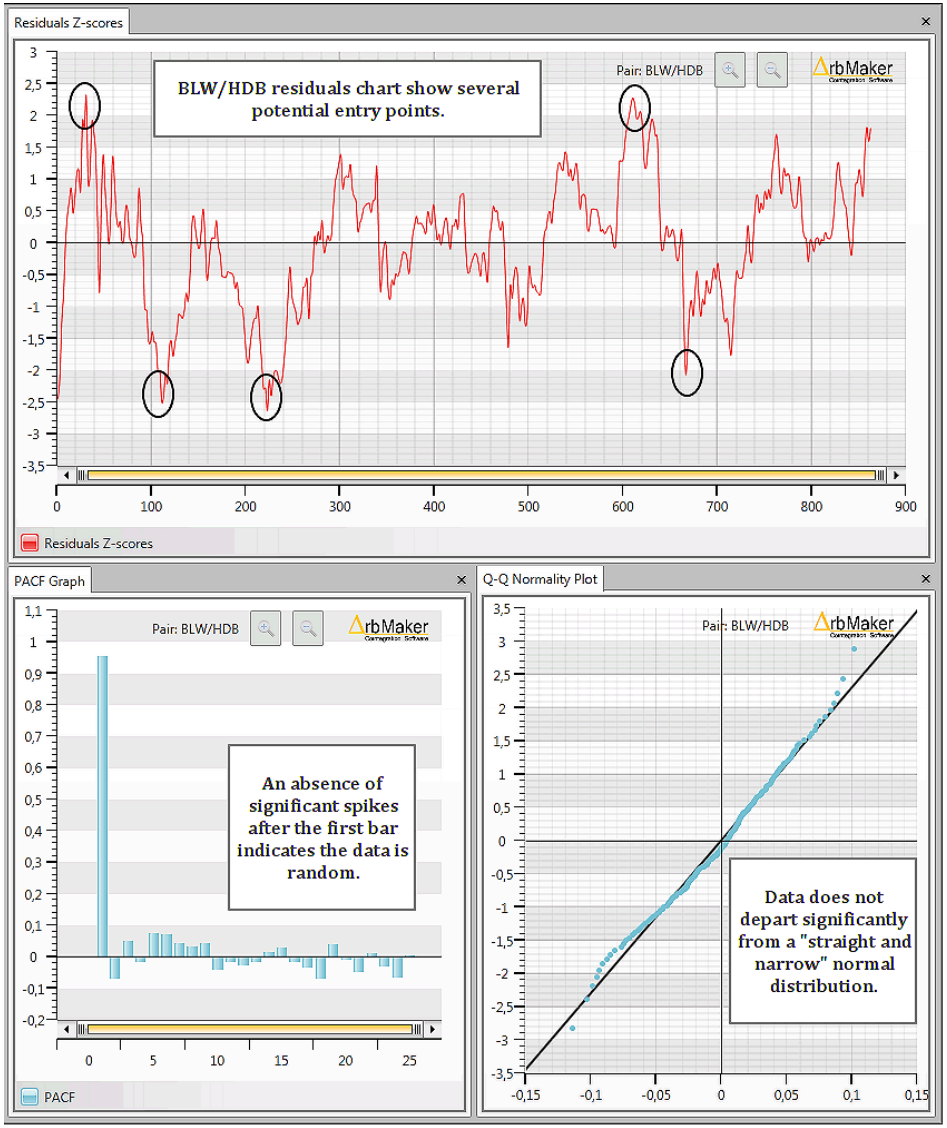
The main documentation here covers how ArbMakerPlus establishes cointegration and sifts the data in order to identify the normality and randomness of data. It should be used to cross-reference this Quick Guide.

But let's now look at some real data examples in the following 'Practicalities' section.

Practicalities

Example 1: BLW/HDB

The charts below came from an end-of-day scan of the US financial sector:



It appears to be a promising pair. However, common sense demands the checking of a few other items.

Potential trades: have a check-list

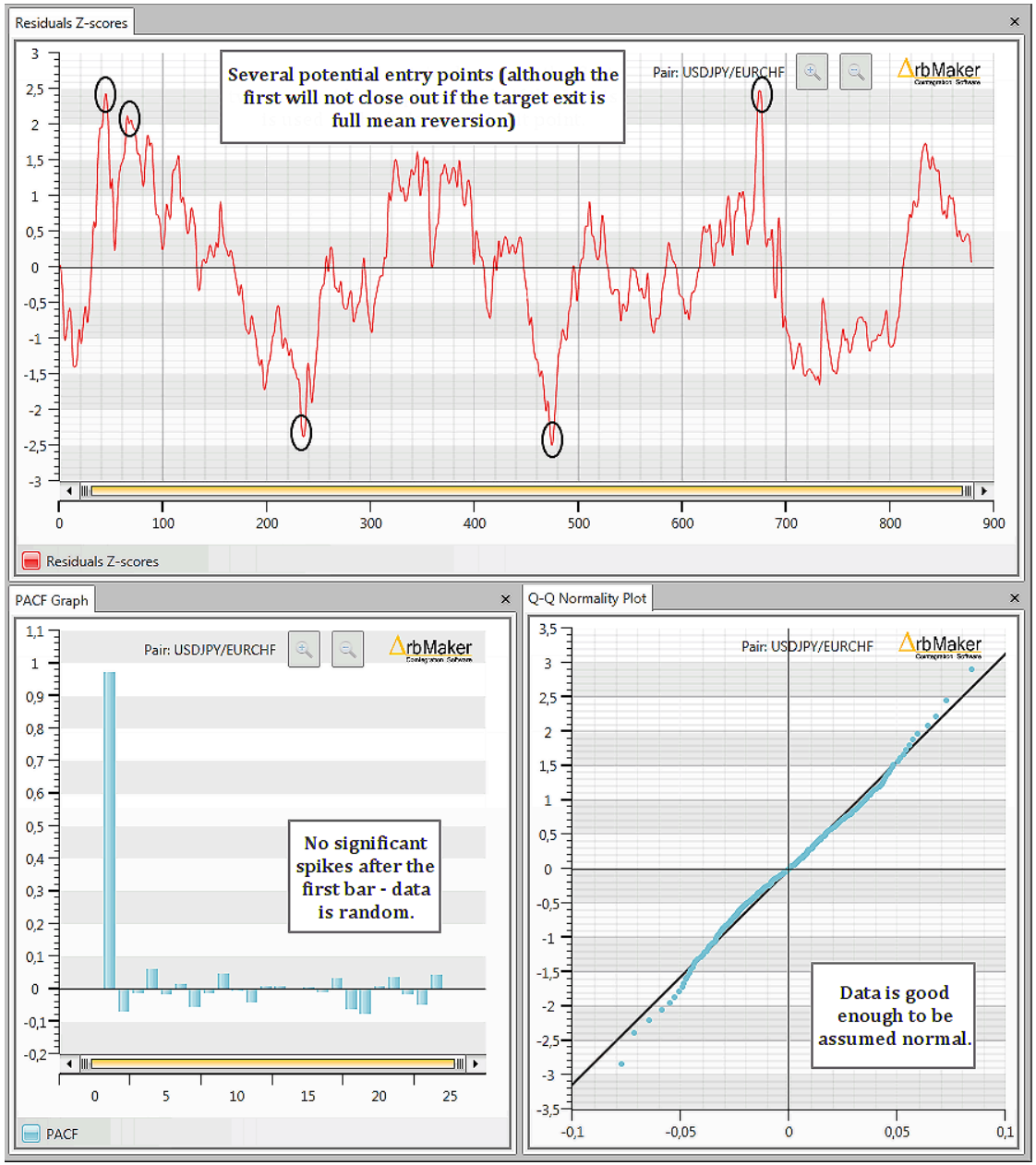
There is more to consider than the charts. Some examples of additional factors to consider include:

1. **Strength of cointegration**: a higher degree of cointegration, indicating a stronger relationship between the underlying instruments, may be generally more desirable than less. So a pair cointegrated at 99% (like this one) may be preferable to one at 95%.
2. **R^2:** a high R^2 validates the explicative power of the relationship between BLW and HDB (or any X/Y pair) more than a low one. This pair has an R^2 of 0.87 or 87%.
3. **Beta**: in the case of BLW/HDB the beta is 0.354. This suggests a dollar move in HDB will see a $0.354 move in BLW. The immediate consequence for a pair trade is that balancing the position will require $1/$0.354 (or 2.83) units of BLW for each unit of HDB. It is not 'dollar neutral'. This may not detract from the trade itself but it may raise questions of capital efficiency for the trader's portfolio.
4. **Fundamentals**: BLW is Blackrock Limited Duration Income Trust, an investment company mainly focused on the United States and Europe. HDB is HDFC bank Limited, a large Indian bank (with a NYSE listing) essentially focused on its domestic retail and mortgage lending markets. While there may be overlap in the two companies' portfolios the businesses are some distance apart structurally and geographically. Whether the two offset even sector risk ought to be a concern.
5. **Reporting dates**: it is not always comfortable to hold positions in firms about to report earnings. Certainly company risk rises greatly (both positive and negative) approaching and during earnings season.

The list is non-exhaustive.

Example 2: USDJPY/EURCHF

The chart below is from an end-of-day scan of our currencies database:

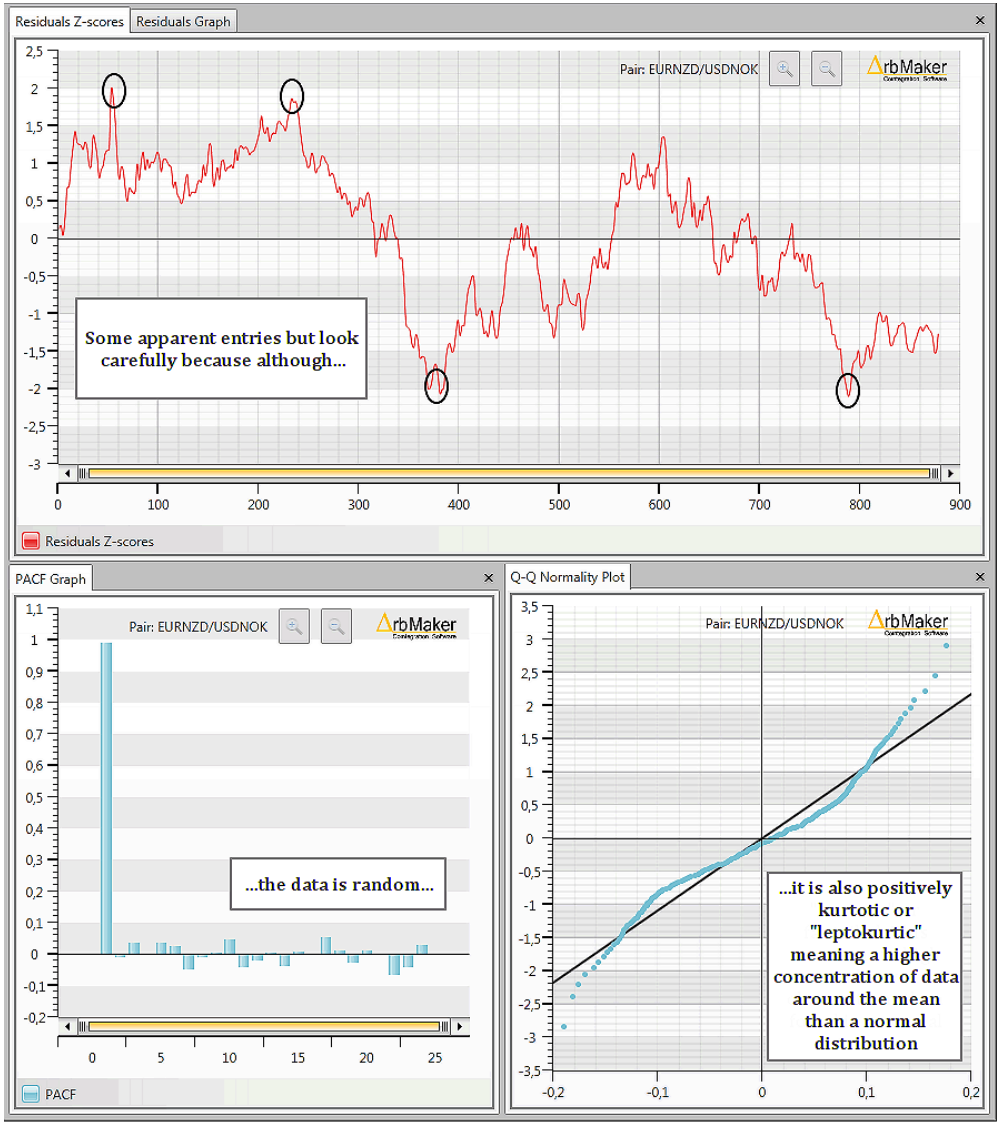


Once again the chart is decent but it is worth digging deeper with a checklist. Some (still non-exhaustive) items might be:

1. **Strength of cointegration**: unlike the first example this pair is cointegrated at the 95% confidence interval. Good, but not as strong a relationship as at 99%. At the intra-day level this can be an important consideration as the statistical relationship can change (and decay) relatively swiftly.
2. **R^2:** it is about 83% for this FX pair. To put this in context a useful way of thinking about these numbers is to consider that 17% of the actual data is not explained by the model. Given that FX is habitually leveraged at 40% (and much higher) the margin for error is magnified by lower R^2s.
3. **Economic data & policies**: there is a very large body of academic work that considers the different and varying impacts on exchange rates of central bank interventions, economic announcements and data releases. Some actions and data are considered particularly influential (US payrolls announcements, for example) and may be hard to trade around. So it is generally a good idea to have an economic calendar relevant to the currencies in line for potential trades handy.

Example 3: One to be wary of

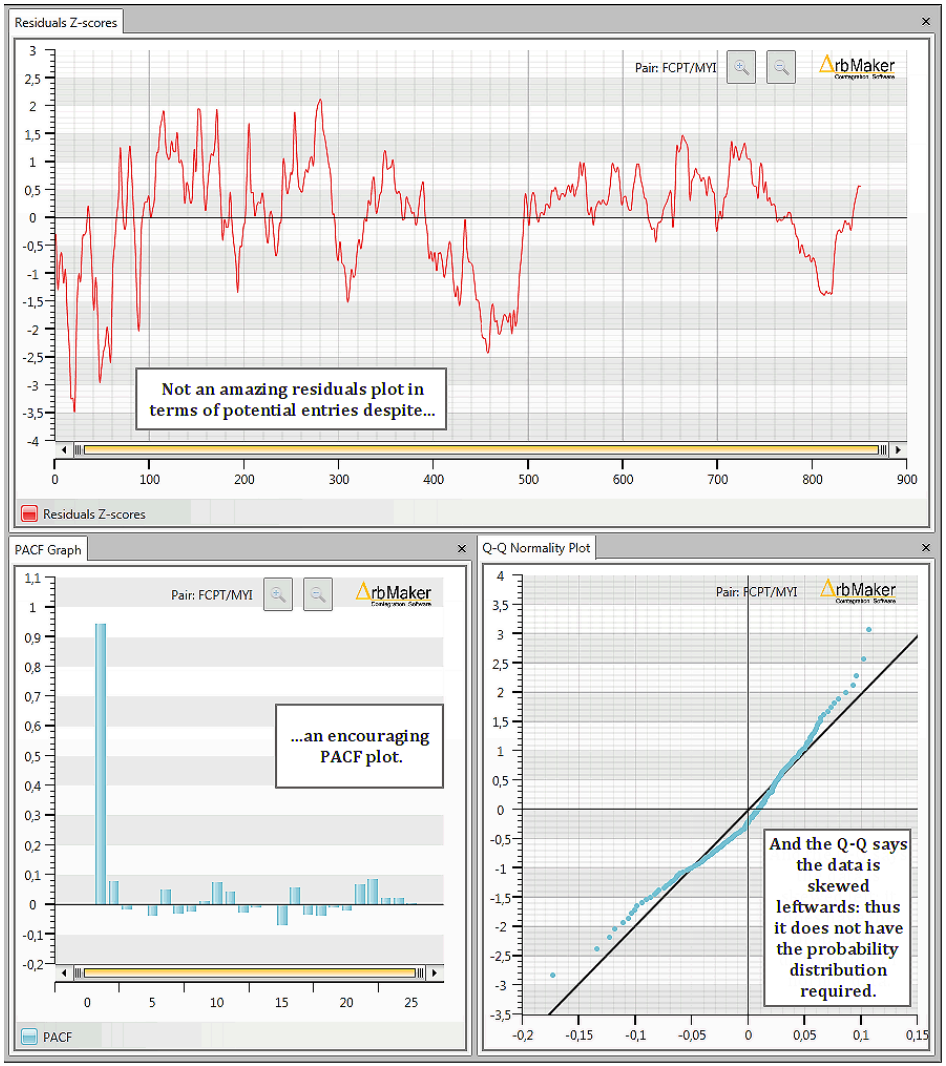
Sometimes the residual distribution is not normal. One of the two main manifestations of this is when it is kurtotic. Kurtosis comes in 2 forms: positive (leptokurtic) which implies data concentrated around its mean; and negative (platykurtic) which suggests a wider distribution of returns away from the mean (and thus greater volatility) than a normal distribution. Here's an example using end-of-day data of the leptokurtic forex pair EURNZD/USDNOK. The probability benefits of a normal distribution will not apply:



A platykurtic distribution would print the mirror image of the Q-Q plot shown above.

Example 4: And another to watch for

The other main departure from normality is a skewed distribution. Consider the shot below using end-of-day data from the London Stock Exchange for the pair FCPT/MYI (Foreign & Colonial Commercial Property Trust Ltd and Murray International Trust PLC) illustrating left, or negative, skew:



Right skew would show up as the mirror image to that shown in the Q-Q plot above. Overall, what counts is recognizing the non-normal patterns, not whether the pattern is lepto this or platy that, skewed and so forth. No matter how promising the Z-score charts may appear the implied tradability will be false if the data is not normal or near normal.