Exploring the efficacy of higher moments in portfolio optimisation

Pat Burns @ LQG 2013 December

(The avocado talk)

Presented at the London Quant Group

The house rules are that the audience is to give the speaker a hard time. I'm happy to report that they performed their duty admirably.

There is one spot in particular in the presentation that is designed to be complained about.

What is

portfolio optimisation ?

Often a great place to start is with the question: What the hell are we talking about?

Traditional answer

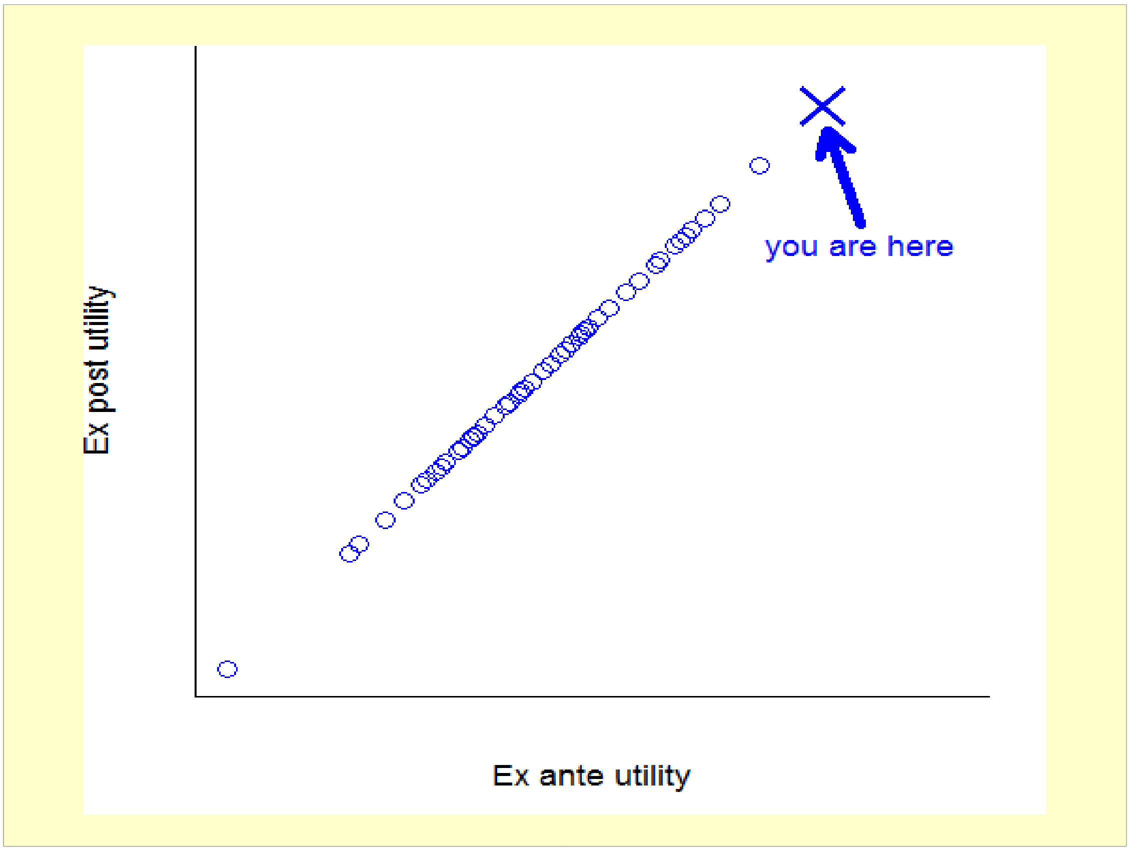
Maximise the utility

Subject to meeting

CONSTRAINTS

The traditional answer is not exactly wrong – if you got this answer on an academic test, you would have to give them points.

But it conjures up the wrong picture.



We picture actually getting the best answer.

It is n't gonna be that way

– Steve Forbert

Steve Forbert had some prophetic words about this.

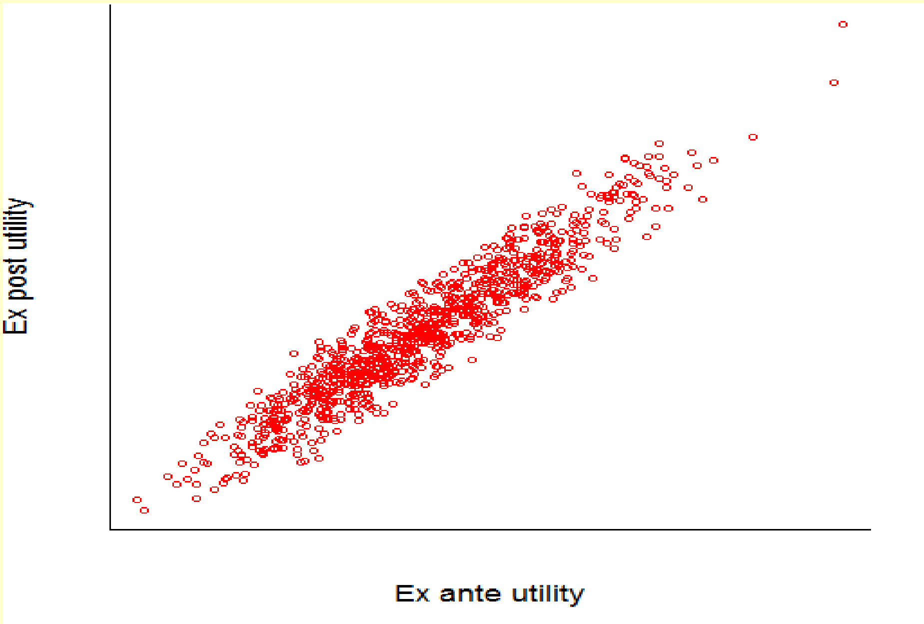
You can see more of his thoughts in this regard in the blog post that was the catalyst for this talk:

http://www.portfolioprobe.com/2013/10/14/four-moments-of-p

I can push that image out of my mind, but it takes work.

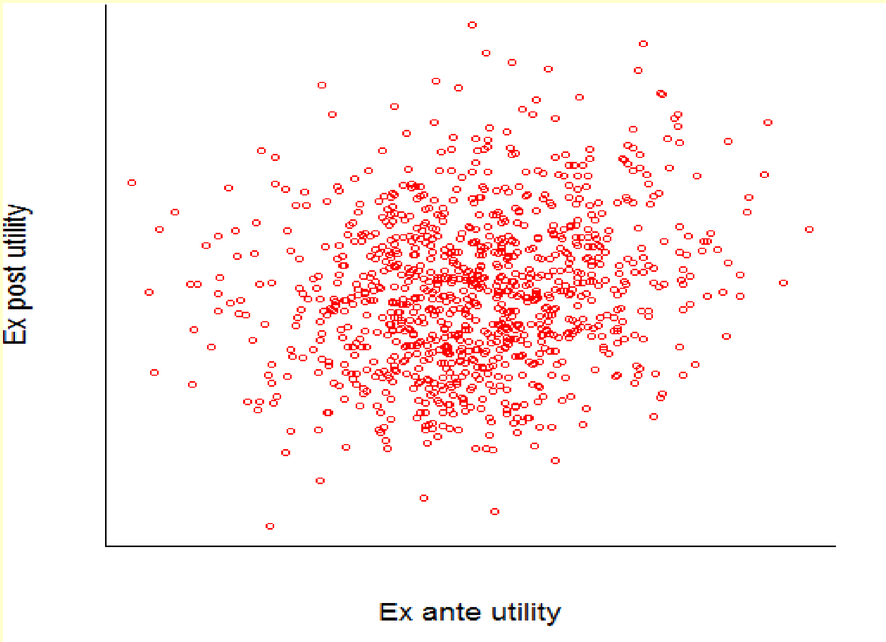
What image should we have?

As good as it gets



This is about the best we'll get. This is using volatility as the utility (the realised is for 20 trading days).

Still “good”



This is a more typical picture, and actually it is still quite good.

That this is “good” suggests that my wife was wrong to discourage me from being a professional basketball player – I'm at least as good at basketball as this.

optimisation

A key problem is this word. Even when it is spelled correctly – with a “z” – it is trouble.

It has an association with perfection that is just not warranted.

selection

A much better word is this one.

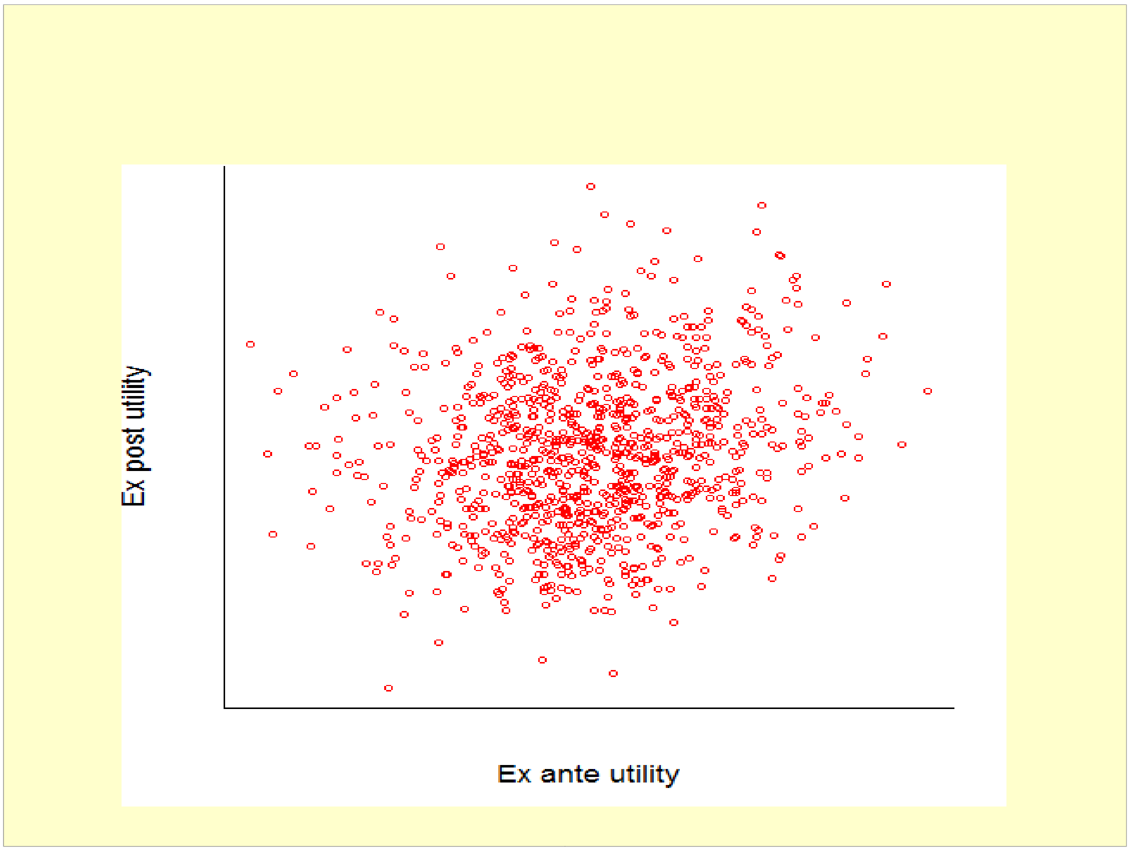
Among the portfolios

that obey the constraints

guess good ones

So let's try again with our question of what the hell are we talking about. Now we are asking: What is portfolio selection?

In this new version constraints get top billing and utility is relegated to a footnote.

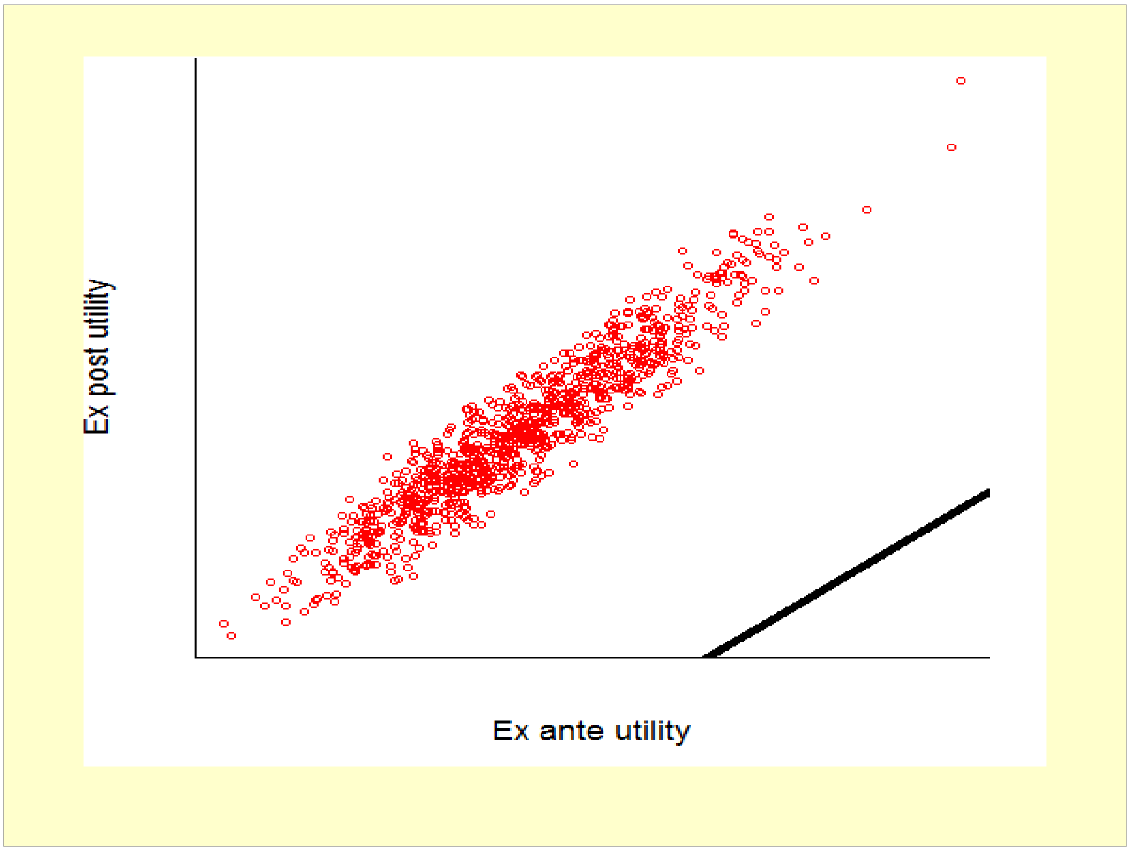


We've seen this picture before. It was a minor opportunity to complain. You should complain if it wasn't obvious to you what is being plotted.

Each point in the plot is a portfolio that obeys the constraints of the problem.

The collection of points is a random sample of portfolios that obey the constraints. That is called “random portfolios” and you can learn more about them at:

http://www.portfolioprobe.com/about/random-portfolios-in-finance/



We've almost seen this picture before as well. There is an addition: the y = x line. Now we can see that in this case we actually got more utility than what was predicted.

Does that matter? For some things it does. For what we are doing it doesn't matter one little bit.

What matters is: if you have a portfolio in each hand, how well you know which will be better.

Let's forget about optimising silly portfolios for a while and optimise guacamole.

At this point I pulled an avocado out of my pocket. Ed asked if this was planned or did I just happen to have an avocado in my pocket. I replied that it was planned and that I had told my wife that she couldn't use the avocados yet because they were for my talk. Her reply was: “This is a switch, you're buying fruit to throw at the audience.”

Photo by Brybs via stock.xchng

When I'm optimising guacamole, I need to select an avocado at the store. What do I want? I want an avocado that is flavourful and not rotten. But I only get to know what I've got once I take it home, cut it open and have a taste.

At the store I have to use characteristics that seem to correspond to good results. What is my selection process? I hold an avocado in each hand. I look at their colour – darker is better, more likely to be ripe.

I gently squeeze them – hard means it isn't ripe, too soft means it could be rotten.

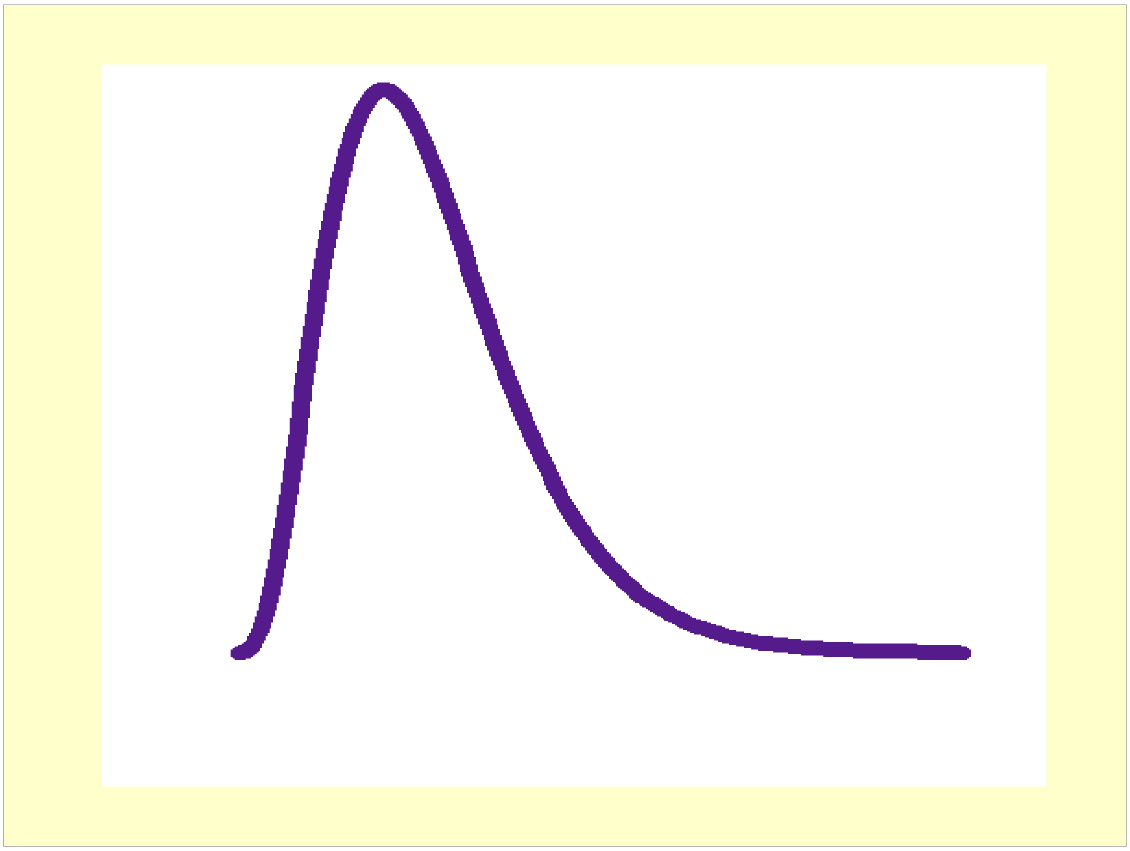
I don't try to compare their density. That works for citrus, but I doubt it works for avocados.

I keep comparing avocados in each hand, always keeping the one that seems better. I continue until either I've examined all the choices within the constraints of the avocado display, or I get bored.

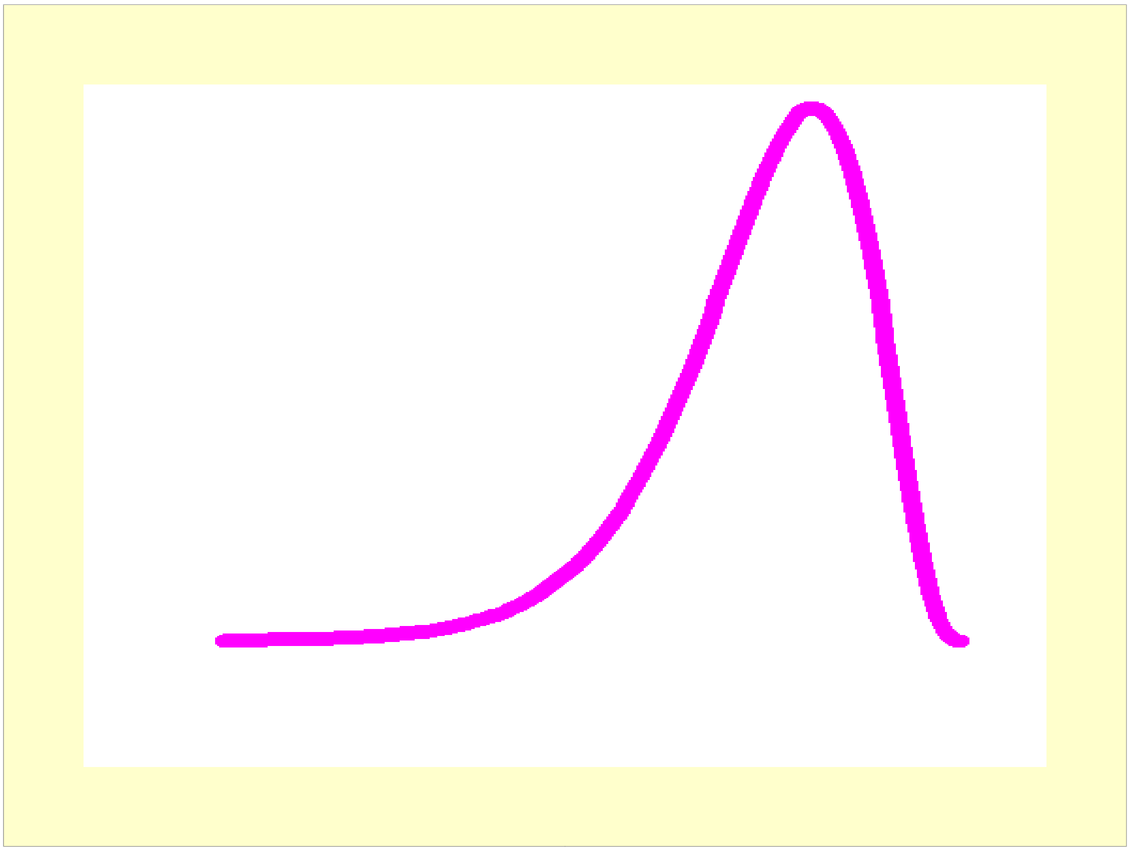
Skewness & Kurtosis

But we are supposed to be talking about silly portfolios, and actually about higher moments in portfolios. In particular skewness and kurtosis.

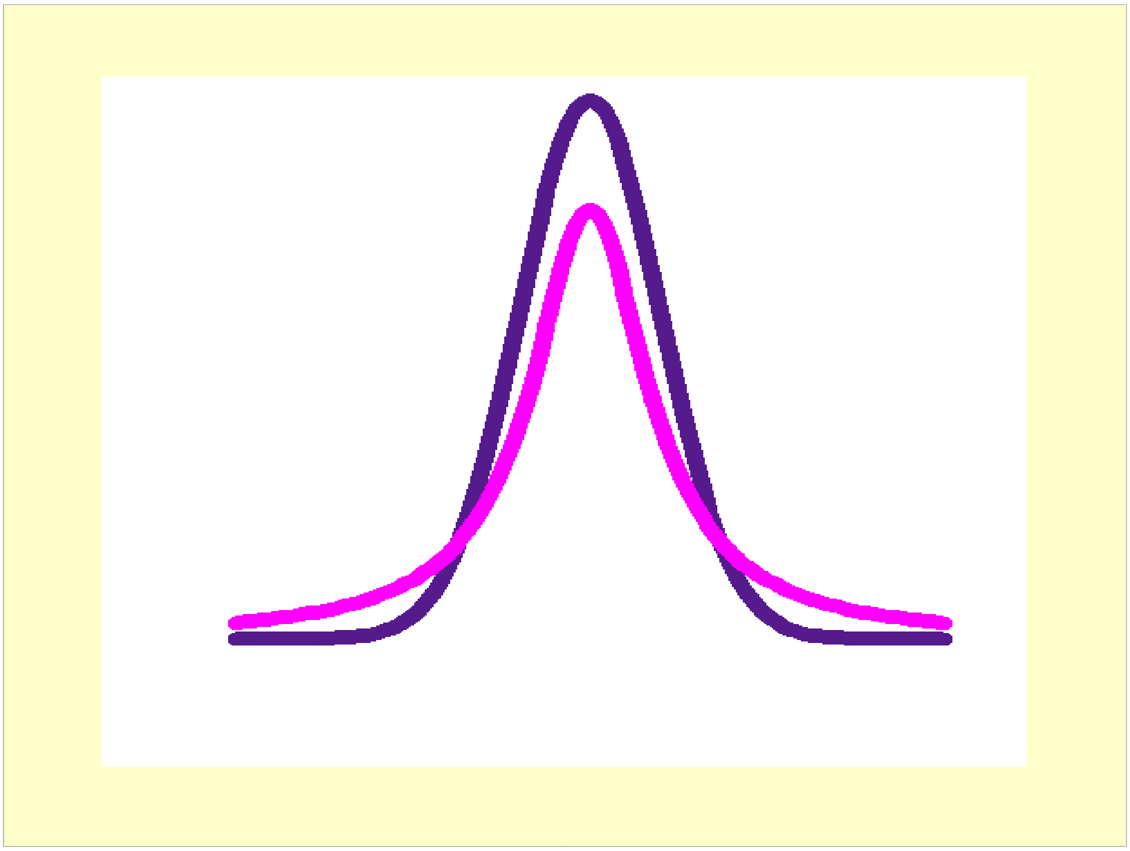
The one thing where there is universal agreement is in how to relate to these.



We like positive skewness.



We hate negative skewness.



We dislike high kurtosis.

There are at least two ways to think about kurtosis.

One way is to think that our dislike of the negative tail outweighs our like for the positive tail.

Another way is to think that we very much like certainty, and high kurtosis is all about uncertainty.

The experiment (1)

- 400+ US large cap **equities**
- Portfolios with exactly 200 names
- As of the start of 2004
- 0.1% < weight < 3% at inception
- Hold through 2012
- Sample of 1000 portfolios

Here is the experiment that started off this sordid mess. It was reported in the blog post:

http://www.portfolioprobe.com/2013/10/14/four-moments-of-p

We will create random portfolios with some constraints and hold them (with no trading) for 9 years. The actual number of portfolios created doesn't much matter, but we need to remember how many there are.

We are going to over-generalise the results of this experiment – that is what humans do. However, we should be especially careful when trying to generalise beyond equities. Different assets can be expected to be quite different in this regard.

The experiment (2)

- 250-day rolling estimates of:
  - Mean
  - Variance
  - Skewness
  - Kurtosis

We want estimates of the first four moments for each of the 1000 portfolios over rolling windows of 250 trading days.

We start with a matrix of returns for the portfolios. It has 2267 rows (days over the 9 years) and 1000 columns (portfolios).

We create 4 new matrices that are each 2018 by 1000. Each holds the estimates of one of the moments.

The experiment (2)

- 250-day rolling estimates of:
  - Mean
  - Variance
  - Skewness
  - Kurtosis
- At each time: rank across portfolios

Now we create 4 additional matrices (and we'll actually care about these). For each of the moments we get the rank of the portfolios within each day.

We look at one row (day) at a time. We replace the smallest statistic with 1, the second smallest statistic with 2, …, the largest statistic with 1000.

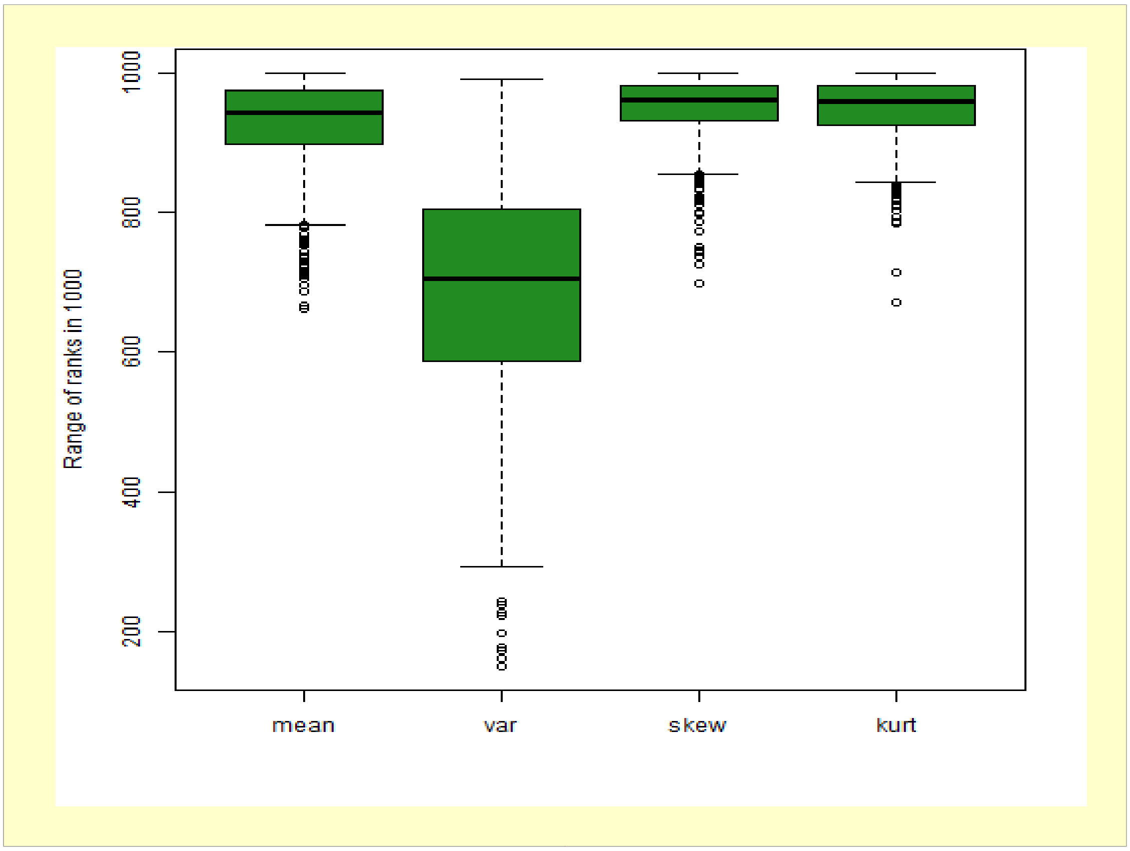
Why ranks? Because what we care about is if we can tell if the portfolio in our right hand is better than the one in our left. The actual value of the moment doesn't matter much.

The experiment (3)

- **Look at the range of ranks for each portfolio**

If a moment is useful for us to predict, then the rank of portfolios needs to be stable for it. If the ranks of the portfolios were constant through time, then we could do perfect optimisation.

Our statistic for stability is the range of ranks that each portfolio had over the 8 years of estimates that we have. That is, the maximum rank it achieved minus the minimum rank.



So for each of the 4 moments we have a distribution (over 1000 portfolios) of ranges of ranks.

The maximum range is 999. Ideal is 0.

The mean is quite unstable. The variance is much more stable. Skewness and kurtosis are even more unstable than the mean.

This is a boxplot. If you don't understand it, there is a blog post that explains them:

http://www.portfolioprobe.com/2012/12/24/miles-of-iles/

that was

STUPIDD

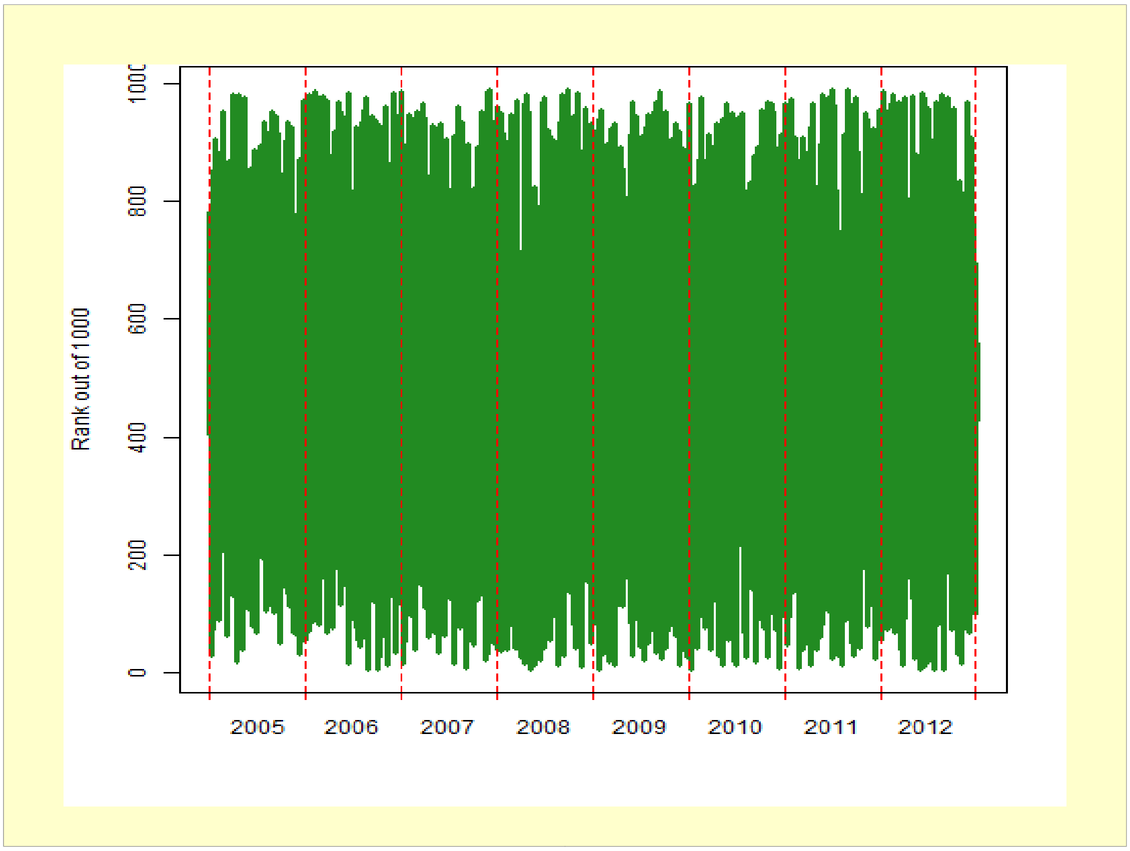
The range of ranks is not a very good measure of stability.

For one thing as we increase the length of time that we hold portfolios, it probably becomes a very good estimator of 999 – we can expect more and more portfolios to run the full range of ranks over longer time frames.

But it is even worse than that ...

Here is a time plot of the rank for some moment of a hypothetical portfolio.

The range of ranks for this is 999 – the maximum.



Here is a similar plot for another hypothetical portfolio. The range of ranks for this is 985.

So by the range of ranks statistic this portfolio is more stable than the previous one. Which of course is nonsense.

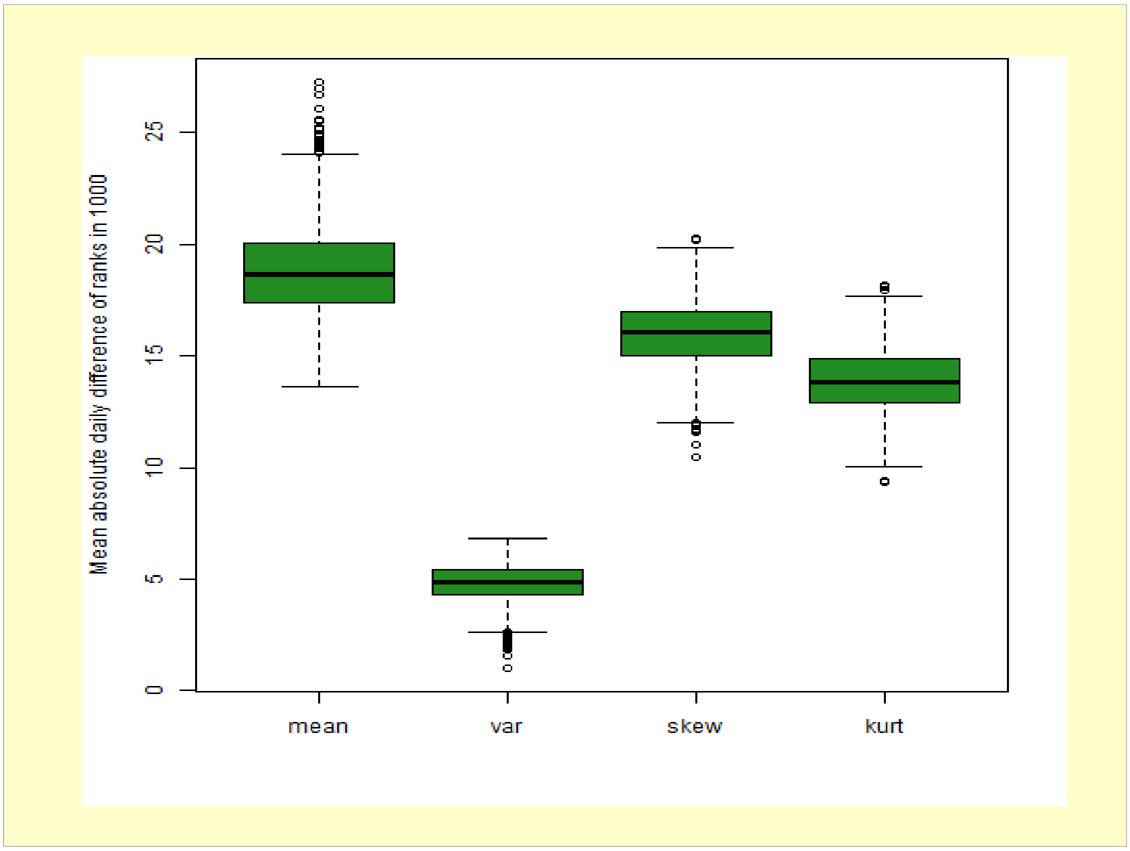
Maybe we can do better.

The experiment (4)

- **Look at the mean absolute daily change in rank**

A less stupid statistic is to look at the change in rank from day to day, take its absolute value, and then take the mean of that over all the days for the portfolio.

We still have a distribution over 1000 portfolios for each of the moments.

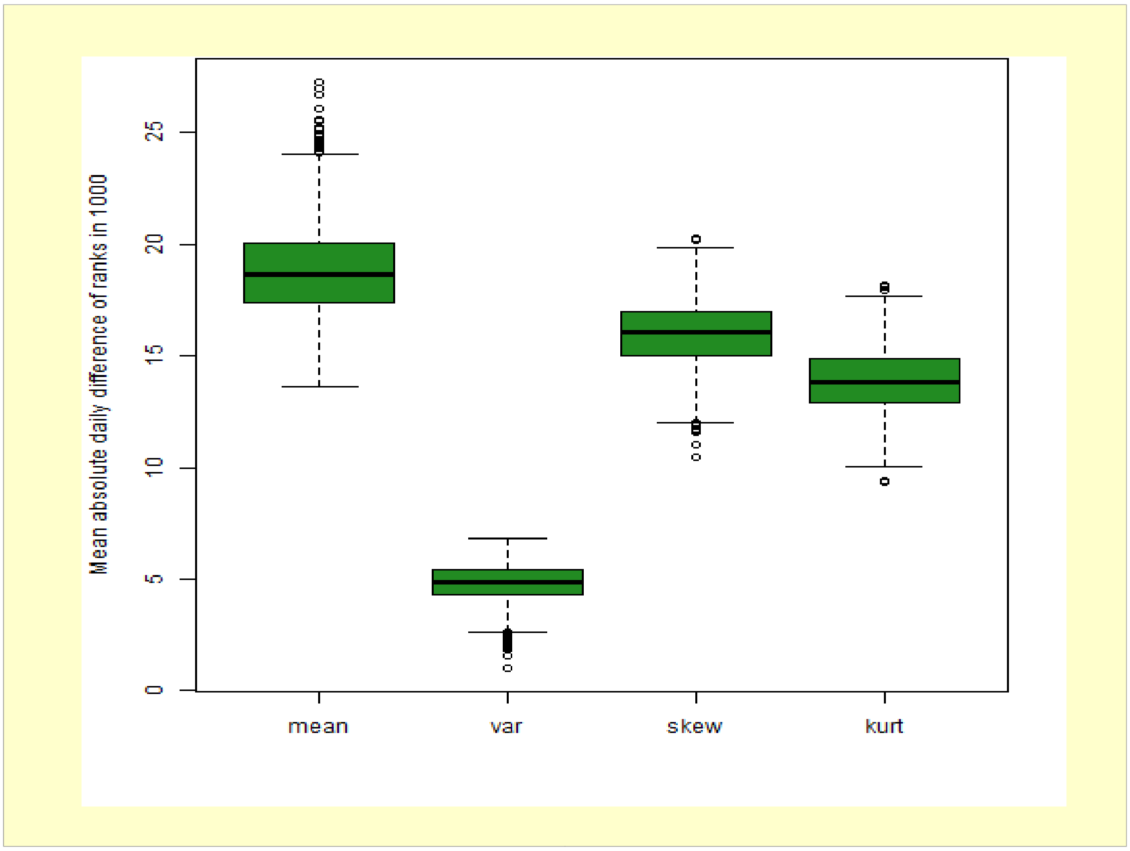


It is still the case that up is bad (less useful) and down is good.

We know that the historic mean is useless. I'm okay if people want to put their own money at risk with it, but I hope that no one is using it with other people's money.

We know that the variance is useful. So somewhere between where those two land is the line between useful and useless.

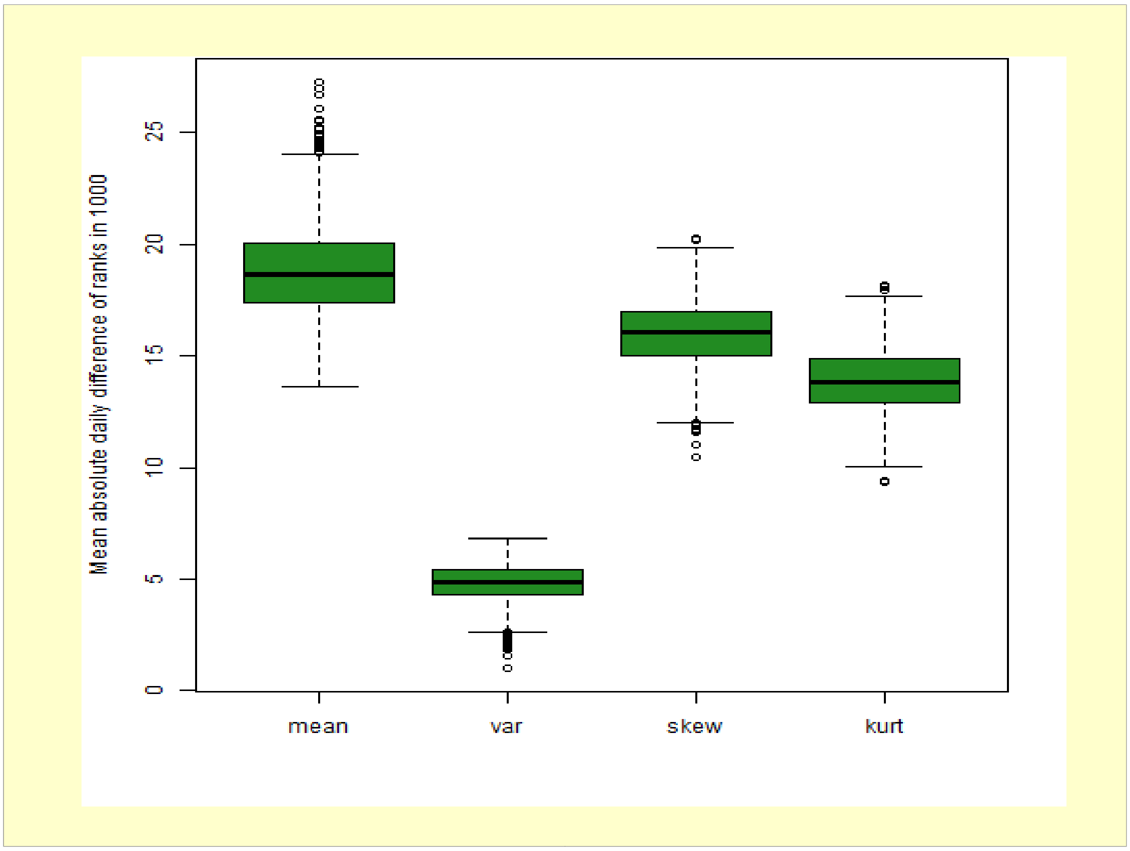
Robert Macrae suggested that the line might slope downward.



There is another way that the useful-useless line could be illusory: changes through time might render the line fuzzy or wiggly.

There was a question about why the outliers are all on the upside for the mean but the downside for the variance. Antonia Lim hypothesised that it was autocorrelation.

The original blog post shows a similar plot for portfolios that have 20 instead of 200 assets – it is very similar, including the same outlier pattern.



I was surprised by the location of both skewness and kurtosis in this plot.

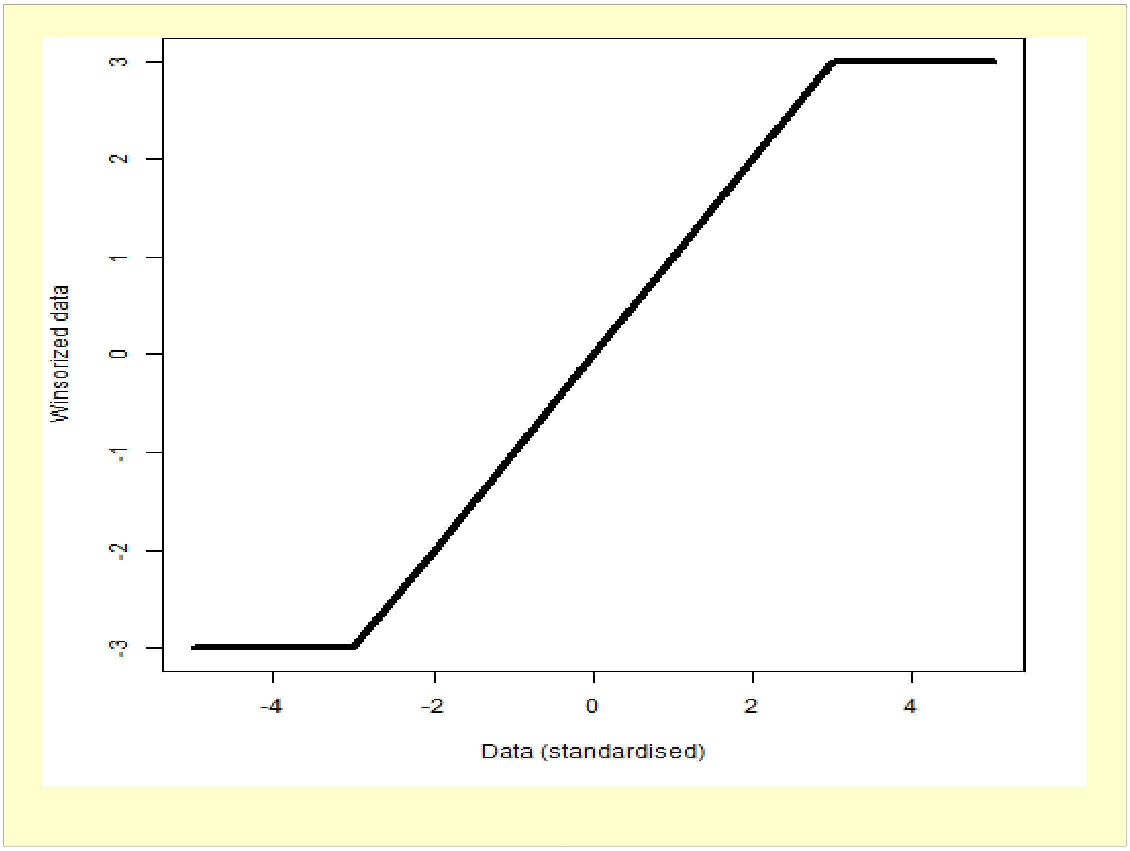
I expected skewness to be at least as unstable as the mean. Now that I've seen it more stable, I can make up a story for that – but I don't believe myself.

I expected kurtosis to be down lower, closer to the variance. As Steve Satchell pointed out, the variance of the kurtosis estimator depends on the 8^{th} moment.

statistical robustness

Perhaps we have a statistical robustness problem.

When we are calculating the kurtosis, we are raising the returns to the 4th power. So the extreme returns have a big, big effect on the result.

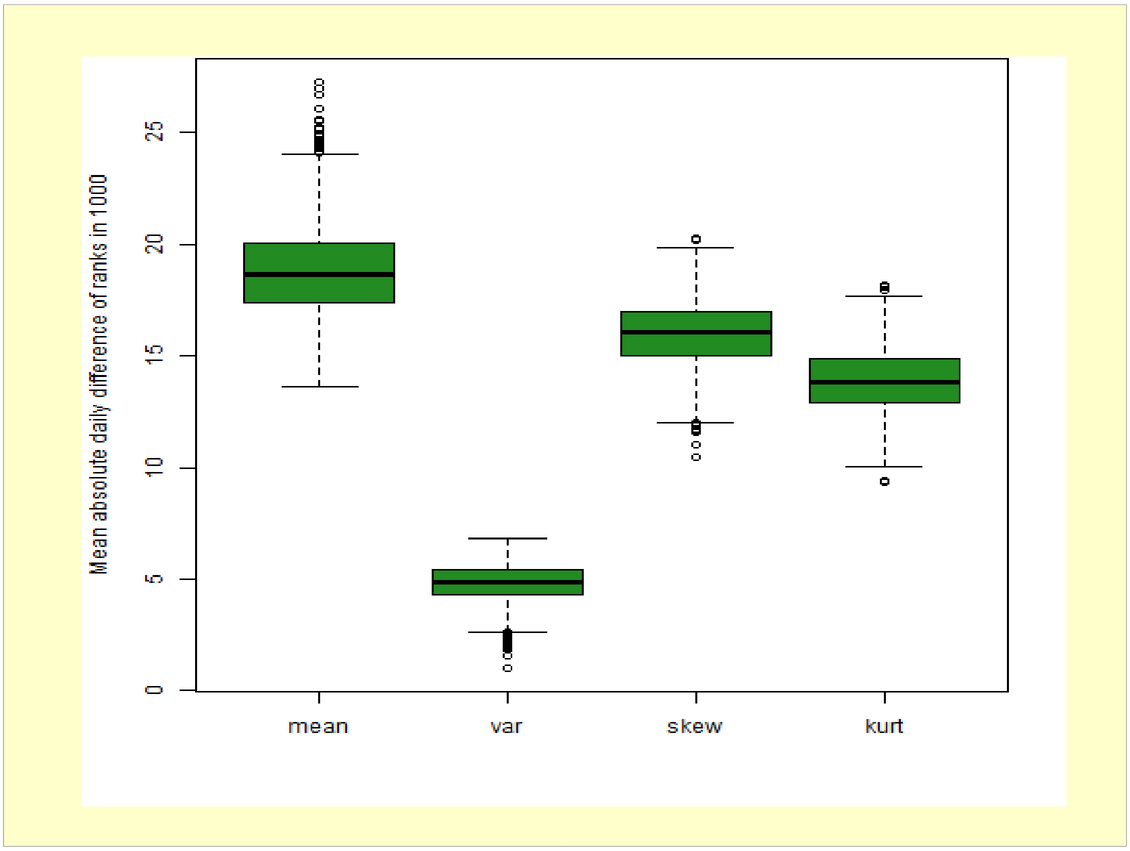


Winsorisation is a handy way of getting robust estimates. It is nice in that you can get robustness without creating new robust estimation techniques. You can use the same worn out machine you've always used and robustify the data instead.

If Winsorisation is new to you, then there is a blog post about it:

http://www.portfolioprobe.com/2011/06/30/winsorization/

The form of Winsorisation used was the second type in that post, but the function actually used was 'winsorize' in the 'robustHD' R package.



We're going to draw another picture like this, but with the returns Winsorised at 3 standard deviations (where standard deviation is estimated robustly via the median absolute deviation from the median).

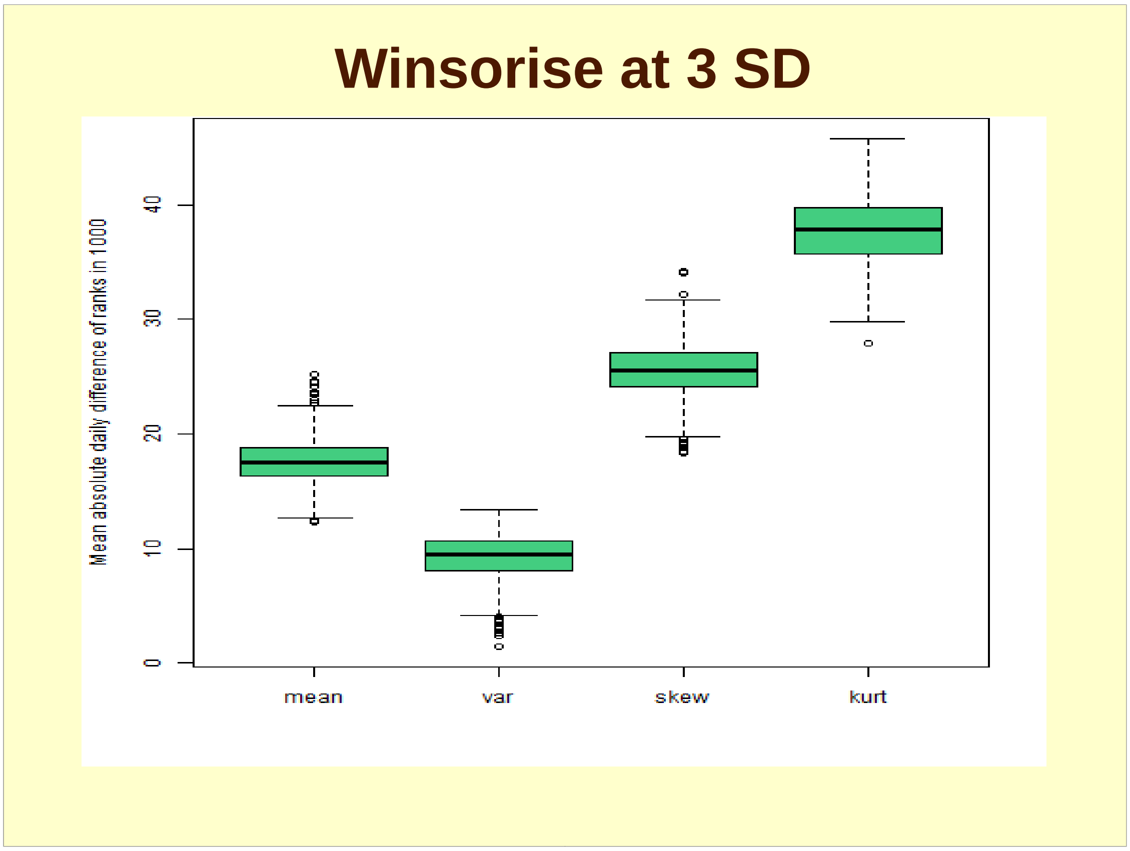
Let's vote

- **Skewness & kurtosis both better**
- **Both about the same**
- **Skewness better, kurtosis worse**
- **Kurtosis better, skewness worse**
- **Skewness & kurtosis both worse**

The vote was held on the effect that Winsorisation would have.

The most popular choice was the first: both skewness and kurtosis more useful with Winsorisation. The other choices were all about equally popular.

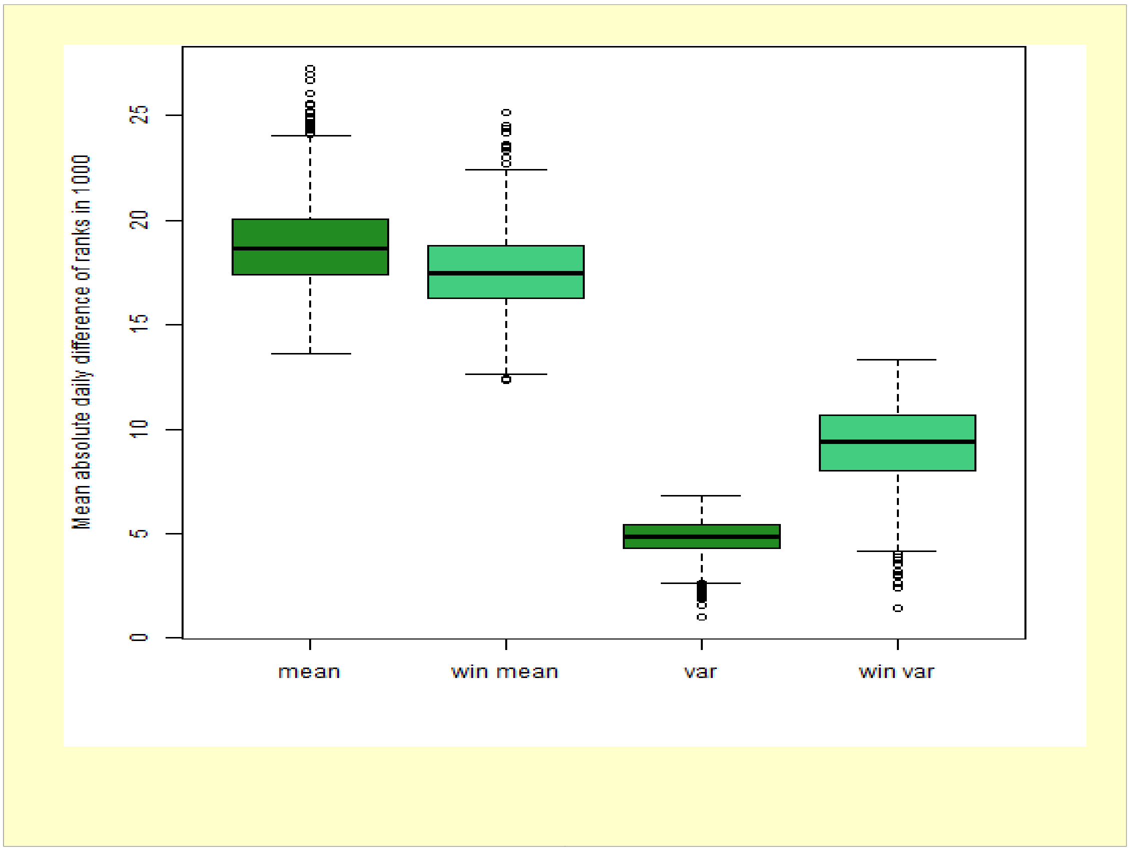
The truly most popular choice was abstention.



The result is utter disaster. Now we have skewness and kurtosis being much worse than useless.

A criticism that seems quite valid to me is that there could be an effect from the overlap within the 250-day window.

Perhaps a better measure is the mean change in absolute rank between days that are 250 days apart. It is not out of the realm of possibility that there will be a future Portfolio Probe blog post exploring that.



We can compare mean and variance results with and without Winsorisation.

Winsorisation improves the mean marginally.

Winsorisation is not at all an enhancement for variance. I think this might be the most intriguing thing in the talk.

Robert Macrae explains this (and the result for skewness and kurtosis) as the Winsorisation effectively destroying the tails which is the very thing we are trying to estimate (Robert's phrasing was much better).

Ed Fishwick's immediate intuition was to believe Robert's explanation for variance but not for skewness and kurtosis.

My intuition is just the opposite: that is the explanation I had come up with for the skewness and kurtosis before the talk. But I have seen a similar effect for the variance by other means:

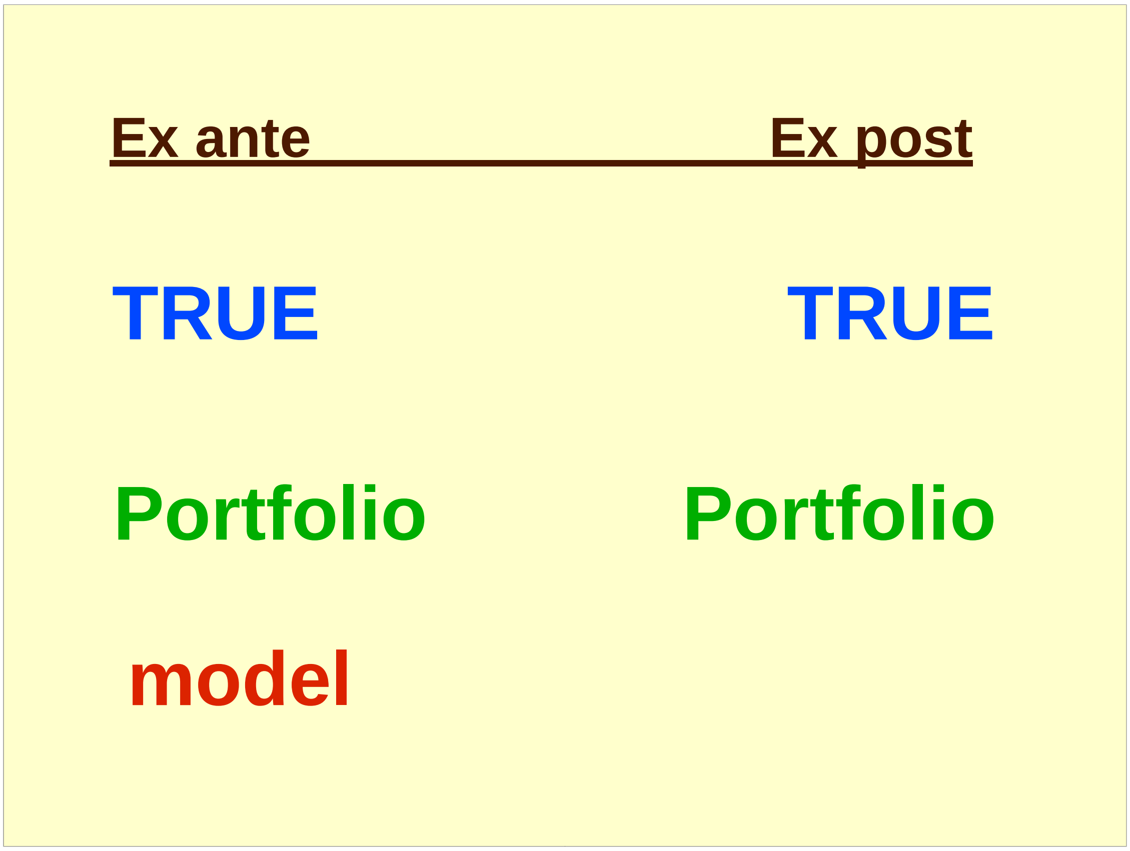
I estimated variance assuming a t distribution. So a more professional statistical analysis that in theory should give a better estimate of variance, but the result was worse prediction.

That suggests to me that the extreme returns are more informative than they “should” be.

Let's now think about actually implementing the use of skewness and kurtosis.

When we do that, there is likely to be some sleight of hand.

Picture by miamiamia via stock.xchng



This slide identifies the sleight of hand, and it also unmasks the colour coding that is used in the presentation.

The skewness and kurtosis that we've looked at so far has all been at the portfolio level – realised values. If we use skewness and kurtosis we will probably model them at the asset level and weight that up to the portfolio level. We don't have to, but it is likely.

We never get to see true blue – the truth is beyond our powers of perception.

# assets	Var
10	55
100	5050
1000	500,500

Let's inspect what we have for a few sizes of universe of assets.

We start with something that we think we know about.

The variance is an n by n matrix. But some of those numbers are redundant. A variance matrix for 10 assets has 100 total numbers but only 55 unique numbers.

# assets	Var	Skew
10	55	220
100	5050	171,700
1000	500,500	167,167,000

Skewness is an n by n by n block of numbers.

In mathematics that is a tensor.

In R jargon it is a 3-dimensional array.

# assets	Var	Skew	Kurt
10	55	220	715
100	5050	171,700	4,421,275
1000	500,500	167,167,000	41,917,125,250

Kurtosis is a 4-dimensional array: n by n by n by n.

If we are thinking about equities, then a 1000 asset universe is at most medium sized.

Meanwhile, 42 billion is getting to be a non-trivial number of parameters.

# assets	Var	Skew	Kurt
10	55.0%	22.0%	7.2%
100	50.5%	17.2%	4.4%
1000	50.0%	16.7%	4.2%

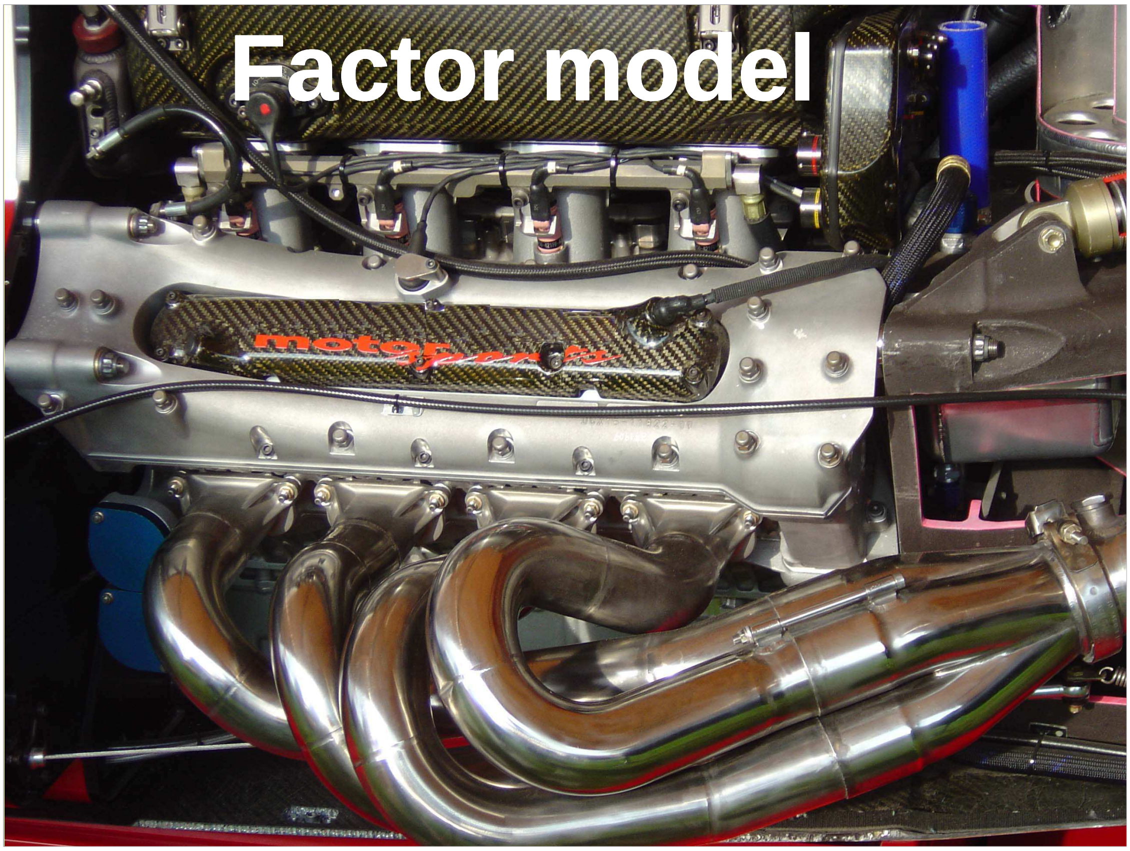
Here is the fraction of the parameters to the total size of the object.

It seems safe to assume that these asymptote to ½, 1/6 and 1/24.

Modelling strategies

How might we model skewness and kurtosis?

There are two main strategies that we use with variance.



With factor models we hypothesise that there is a finite number of pistons that drive the engine, and that we can identify the pistons well enough.

Photo by LeoSynapse via stock.xchng

Shrinkage

The other common strategy with variance is shrinkage.

sample

The first step with shrinkage is to get the sample estimate.

Let's think about a 1000 asset universe where we use 2 years of daily data. So we are bringing half a million numbers to the table.

Now we want to estimate our 42 billion kurtosis parameters. Something magic happens under quite reasonable assumptions: each of those 42 billion estimates is unbiased. Even in finance where we are unreasonable, we can phrase the question so that it will be about right.

But whatever the equivalent of positive definite is for kurtosis tensors, I don't think we are there.

w * sample + (1-w) * simple

The second part of shrinkage is to have a simple model.

Then to have a linear combination of the sample estimate and the simple model.

It is easy to get the sample estimate. It should be easy to think up a simple model and to get estimates for it.

It might be a bit of work to find an appropriate way to calculate the weights.

0 * sample +   1 * simple

A special form of shrinkage is to put all the weight on the simple model.

Some people are foolish enough to think that this isn't really shrinkage.

Shrinkage: Variance

Two types of values

- **Variances (diagonal)**
- **Covariances (off-diagonal)**

Shrink correlations towards a common value

In a variance matrix there are two types of parameter.

A common thing to shrink to is equal correlation. The most likely value for that correlation is the realised mean correlation. Note that it is not equal covariance that is being used.

The variances could be shrunk towards a central value as well. That is done sometimes, but not often.

Shrinkage: Skewness

Three types of values

- **iii (diagonal)**
- **iij**
- **ijk**

There are three types of parameter in a skewness tensor. They involve 1, 2 or 3 assets.

Skewness target

0

What might we shrink towards with skewness? If we are thinking about equities, then all zeros might be an okay target.

If you are in a setting where you expect skewness (some of the audience expect some equities to have skewness of known sign), then shrinking towards zero is not an especially good thing to do.

One possibility is shrinking towards correlation-like quantities. In skewness and kurtosis there are multiple correlation type values, not just one as in variance.

Martellini & Ziemann (2010)

An especially good place to learn about this sort of thing is this paper, which is entitled:

Improved Estimates of Higher-Order Comoments and Implications for Portfolio Selection

Thanks to Robert Macrae for pointing me to the paper.

Martellini & Ziemann (2010)

It has estimates for factor models, equal correlation type models, and shrinkage models.

It also has a section on using real data. Given that they did all that work to derive estimators for skewness and kurtosis, it is in their interests to show that they are useful. If I were trying to show that, then using equities as my example wouldn't be my first choice. But they do use equities – they're going for gold.

They start with what I think is mostly nonsense where they show skewness and kurtosis having a positive impact on results. Finally in the last table they ask a question that might really be of practical interest. There they get no usefulness from monthly data and a tiny bit of usefulness from weekly and daily using a five year window.

Martellini & Ziemann (2010)

Some of what they do make the higher moments look too useful, and some things make them not look useful enough.

They show that skewness and kurtosis have more effect in portfolios with larger numbers of assets (they had portfolios of size 10, 20 and 30).

That makes sense to me. If there are more assets, then there is more room for wider ranges of skewness and kurtosis.

Their portfolios have no constraints. Robert Macrae thinks that no constraints is appropriate as the test since there is no confusion about the effect. However, it seems to me that any usefulness of the moments will be reduced from imposing reasonable constraints.

Martellini & Ziemann (2010)

They are selecting the assets (generally 20, I think) and then using the moment estimates to find weights. But the real problem is to select the assets as well as the weights. Skewness and kurtosis could potentially be more valuable for that, but there is a war between added signal and added noise.

I'm not at all sure which wins out.

Where are we ?

As in: where are we with skewness and kurtosis? In the case of equities the audience – with one notable exception – seemed to share my opinion that skewness and kurtosis are unlikely to be of practical value.

There also seemed to be general agreement that using the higher moments in settings where we have views about the skewness and/or kurtosis of assets is going to be useful. My example of such a setting was in creating a fund of hedge funds.

No one offered any wisdom about how to really go about doing that though. The tensor route with small universes should be workable. But is that better than scenario optimisation with something like the omega ratio?

confession

Let's step back from skewness and kurtosis to the wider picture. Where I have a confession.

that was

STUPIDD

The avocado model has a major weakness.

portfolio optimisation

We started off talking about portfolio optimisation.

We decided that “optimisation” wasn't such a good word.

portfolio

selection

“selection” seems better.

~~**portfolio**~~

selection

It turns out that “portfolio” isn't such a good word either.

trade selection

“trade” is a better word.

In the avocado model, we were deciding between a portfolio and a completely different portfolio.

But almost always we have a portfolio in our hand and we want to transform it a little bit. It is more like adjusting the spices in a mixture.

When we say “portfolio” rather than “trade”, we usually get the problem wrong. In particular, trading costs often magically disappear.

Giles Heywood asked how volatility clustering, which has a definite effect on kurtosis and perhaps on skewness, should be brought into the picture. I didn't have an answer for him then, and I still don't. I'm hoping that one day I will.

One of the highlights for me was a Steve Satchell comment on a Robert Macrae comment, reported by Ed Fishwick. Steve said: “That's a good answer … but I'm not sure it's right.”

Steely Dan on trade selection

Throw back the little ones
And pan-fry the big ones
Use tact, poise and reason
And gently squeeze them

I end with a quote from the highest authority on the subject that I know.