Don't Get Caught with your Correlations Down (or Up): Capitalizing Risk in a Volatile World

September 17, 2019

David Odell, PhD.
Senior Statistician
Insureware

Don't Get Caught with your Correlations Down (or Up): Capitalizing Risk in a Volatile World

A discussion about Time-Risk and Correlation

Time Risk and Correlation

This presentation is in two parts. The first develops some simple ideas about risk into a useful distinction between two different risk situations. The respectively different quantifications of risk make it natural for risk transfer to occur. This is a structural approach to utility without explicit appeal to utility functions.
The second part is about the seemingly unrelated topic of how to properly measure and incorporate Line of Business correlations into a loss model. The main point is that you can't measure correlation until you've accounted for all the salient trends. If correlation is still measurable after de-trending then it is pure volatility correlation and needs to be incorporated into the model.
What connects the two parts is the emphasis on the time dimension. One of the key concepts in part one is the Time Average of outcomes. The cumulative effect of volatility over time is the hidden driver of our main example.
When more than one LOB is involved then it is not enough just to account for the volatility in each line. The correlation between lines is a sensitive determinant of the risk level and should be neither under- nor over-estimated.

We begin with a digression…

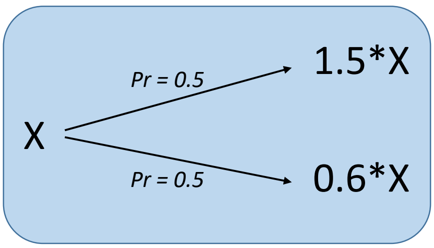
The following digression is based on some lecture notes from the London Mathematical Laboratory by Ole Peters and Alexander Adamou, entitled Ergodicity Economics.
They can be found here: https://ergodicityeconomics.com/lecture-notes/
All misunderstandings are my own.

The relevance to insurance will appear as we go.

A tempting offer

Imagine the following simple gamble:

You put in \$X and with probability exactly 0.5 you are repaid either \$1.5*X or \$0.6*X



Call the outcome of the game with stake X, G(X),
So, **E**(G(X)) = 0.05*X.

That is the expected profit you accrue from one round of the gamble.
The probabilities in successive rounds are i.i.d.

Is it worth playing? Certainly!

There are two different strategies you can follow in doing so…

Two strategies

#1 Accumulation

The first strategy is the ***Accumulation Strategy*** (AS). In this case you fix the value of X (say X = 100), and you wager exactly the same amount at each round.
After N rounds your expected funds are X*(1+0.05*N). So your expectation is that after the 20th round you will have doubled your stake, and at the 100th round your expected funds are 6*X.

#2 Compounding

The second strategy is the ***Compounding Strategy*** (CS). In this case you reinvest all your funds each time you play.
After N rounds your expected funds are X*(1.05)^N. So your expectation is that after the 20th round your funds will be 2.65*X, and at the 100th round your expected funds are 131.5*X.

I hope it's not obvious yet which is the better strategy. We need to do a bit more math to make up our minds.

In order to do this we'll make a distinction between the *Ensemble Average* and the *Time Average*.

Ensemble Average and Time Average

The **Ensemble Average** is the usual probability weighted sum of possible outcomes. In a large enough ensemble of independent instances of the game the proportion of these that produces a given outcome approximates as closely as we wish the probability of that outcome. This is a simple application of the CLT.

The **Time Average** is computed by averaging the outcome for a single trajectory over a long period of time.

In the case of the Accumulation Strategy (AS) these are exactly the same, because it makes no difference to the aggregate whether 100 players play one round each or one player plays 100 rounds.

In the case of the Compounding Strategy (CS), however they differ.
The outcome after N rounds is just the Nth power of the outcome after one round, and since we are multiplying here, the average factor per round is 1.05
To compute the Time Average, consider starting with X and playing N rounds, where N is large. Out of these say that k are wins and N-k are losses. The outcome is X*(1.5)^k * (0.6)^(N-k). Again, by the CLT k ≈ N/2.
So outcome is X*(1.5*0.6)^(N/2), and the per-round factor is ((0.9)^(N/2))^(1/N) = √0.9 ≈ 0.95

In Statistical Mechanics Ergodicity is defined as the agreement of the Ensemble Average and the Time Average. So the AS is ergodic but the CS is not.

Ensemble Average and Time Average

Here is a picture of some plays according to the CS which might make it clearer.

The blue column on the left we can treat as the player number out of a potentially very large pool.

The rounds of each players game are listed horizontally HTHT… etc.
Stopping provisionally after 46 rounds the orange column lists the funds of each player from a starting point at $1

So player #1527 has gone up to $3.46
Player #1532 is down to 22c
Player #1540 has amassed $845.24

	1	2	3	4	5	6	7	8	9	10	38	39	40	41	42	43	44	45	46	
1527	H	H	T	T	T	T	H	T	H	T	H	T	H	T	T	H	T	H	H	3.462085
1528	H	T	T	T	T	T	H	H	H	T	T	T	H	H	T	T	H	T	H	1.384834
1529	H	H	T	H	T	T	T	H	T	T	H	H	H	T	H	T	H	T	H	0.553934
1530	T	H	H	T	T	T	T	H	H	H	H	T	T	T	T	T	H	H	T	0.088629
1531	H	T	H	T	H	T	H	H	T	H	H	H	H	H	H	H	T	T	H	0.553934
1532	H	T	T	T	H	T	H	T	H	H	T	T	H	T	T	H	H	H	H	0.221573
1533	T	H	T	H	T	T	T	H	H	H	H	H	T	T	H	T	H	H	H	0.014181
1534	H	T	T	H	T	H	T	H	T	T	H	H	T	T	H	T	T	T	H	0.221573
1535	T	T	H	T	T	T	T	H	T	T	T	H	H	H	H	H	T	H	H	8.655213
1536	T	T	H	T	H	T	H	T	H	T	T	T	H	H	T	H	H	H	T	0.088629
1537	H	T	T	T	T	H	T	T	H	H	H	T	T	H	T	H	T	T	T	0.553934
1538	T	T	T	H	H	H	T	T	H	H	H	T	H	H	T	T	T	T	T	0.088629
1539	H	T	T	H	T	T	T	T	H	H	H	T	T	T	H	T	T	H	T	1.384834
1540	T	T	H	H	T	H	H	T	T	T	H	T	T	T	H	T	T	T	T	845.2356
1541	H	T	T	H	H	H	H	H	H	T	T	H	H	T	T	H	H	H	T	0.035452
1542	H	T	T	H	H	H	H	H	T	H	T	H	T	T	H	T	T	H	T	0.000908
1543	T	T	H	T	H	T	H	T	H	H	H	H	T	H	H	H	T	T	T	0.014181
1544	H	H	H	T	T	H	H	T	T	T	T	T	H	H	T	T	T	H	T	0.553934
1545	T	H	H	H	T	T	H	T	H	T	T	H	T	T	H	T	T	H	H	0.221573
1546	T	T	T	T	T	T	T	H	T	H	T	H	T	H	T	T	T	T	T	8.655213
1547	T	T	T	T	H	H	H	T	T	H	T	T	T	T	H	T	T	T	H	8.655213

Ensemble Average and Time Average

The Ensemble Average here is the average value of the orange column.

The Time Average for an individual player is the 46th root of the respective number in the orange column.

The limiting Time Average as the rounds go to infinity is the same for each player, (i.e. √0.9) so if we compute this first and then average over the players we'll get the overall Time Average for the game (√0.9).

	1	2	3	4	5	6	7	8	9	10	38	39	40	41	42	43	44	45	46	
1527	H	H	T	T	T	T	H	T	H	T	H	T	H	T	T	H	T	H	H	3.462085
1528	H	T	T	T	T	T	H	H	H	T	T	T	H	H	T	T	H	T	H	1.384834
1529	H	H	T	H	T	T	T	H	T	T	H	H	H	T	H	T	H	T	H	0.553934
1530	T	H	H	T	T	T	T	H	H	H	H	T	T	T	T	T	H	H	T	0.088629
1531	H	T	H	T	H	T	H	H	T	H	H	H	H	H	H	H	T	T	H	0.553934
1532	H	T	T	T	H	T	H	T	H	H	T	T	H	T	T	H	H	H	H	0.221573
1533	T	H	T	H	T	T	T	H	H	H	H	H	T	T	H	T	H	H	H	0.014181
1534	H	T	T	H	T	H	T	H	T	T	H	H	T	T	H	T	T	T	H	0.221573
1535	T	T	H	T	T	T	T	H	T	T	T	H	H	H	H	H	T	H	H	8.655213
1536	T	T	H	T	H	T	H	T	H	T	T	T	H	H	T	H	H	H	T	0.088629
1537	H	T	T	T	T	H	T	T	H	H	H	T	T	H	T	H	T	T	T	0.553934
1538	T	T	T	H	H	H	T	T	H	H	H	T	H	H	T	T	T	T	T	0.088629
1539	H	T	T	H	T	T	T	T	H	H	H	T	T	T	H	T	T	H	T	1.384834
1540	T	T	H	H	T	H	H	T	T	T	H	T	T	T	H	T	T	T	T	845.2356
1541	H	T	T	H	H	H	H	H	H	T	T	H	H	T	T	H	H	H	T	0.035452
1542	H	T	T	H	H	H	H	H	T	H	T	H	T	T	H	T	T	H	T	0.000908
1543	T	T	H	T	H	T	H	T	H	H	H	H	T	H	H	H	T	T	T	0.014181
1544	H	H	H	T	T	H	H	T	T	T	T	T	H	H	T	T	T	H	T	0.553934
1545	T	H	H	H	T	T	H	T	H	T	T	H	T	T	H	T	T	H	H	0.221573
1546	T	T	T	T	T	T	T	H	T	H	T	H	T	H	T	T	T	T	T	8.655213
1547	T	T	T	T	H	H	H	T	T	H	T	T	T	T	H	T	T	T	H	8.655213

Ensemble Average and Time Average

The Ensemble Average is the average over a large ensemble.

How large does it have to be?

Large enough for the empirical probability distribution to converge 'pretty closely' to the true distribution.

The Time Average is just the **typical experience of the typical player**. The average over a single sample path extended far out in time.

The Ensemble Average

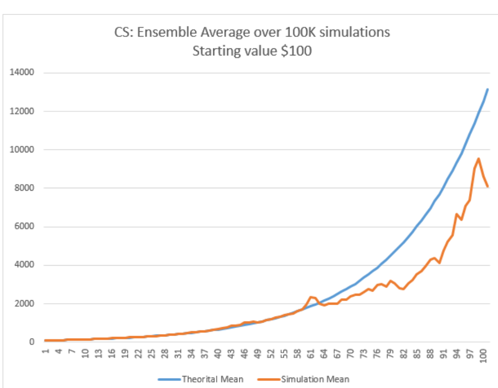
The Ensemble Average is the mathematical expectation but in the CS almost nobody ever comes up to it.

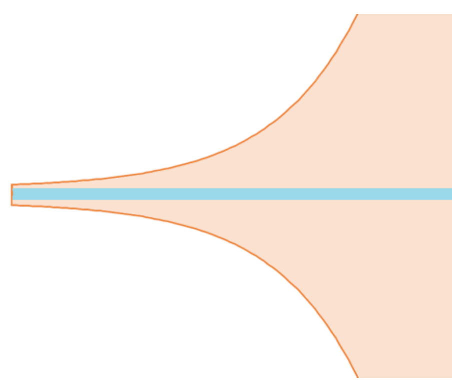
If I play the CS for 100 rounds, then the mean is at the 99th percentile, and just breaking even is at the 86th. These statistics make sense for ensembles of players.

In 100K simulations of 100 rounds the luckiest player turned $1 into $2.9M

By contrast the Accumulation Strategy produces a symmetric bell-shaped distribution. After 100 rounds mean and median are $5, and break-even is at the 18th percentile. The luckiest in 100K simulations had $23

The Ensemble Average





If we follow the mean of 100K simulations through 100 rounds, we see that it tracks the theoretical mean pretty well until a certain point when it falls away. If we kept going it would fall all the way to zero. This is because you need an ensemble size proportional to 2^N to get a fair sample. 100K soon becomes the smallest drop in the bucket. Early on the 100K contained many super-rich, but after a time it's just a collection of typical players. In my simulations, deviation from expected mean kicked in at round #58, 2^58 = 280,000Trillion, so we were doing well.

The Time Average and the Geometric Mean

In summary:
1) The Compounding Strategy has an exponentially increasing mean, but this is maintained in the ensemble by an ever dwindling proportion of super-players.
2) As time goes by all players become typical players.
3) The typical player's experience is determined by the Time Average, whose rate is the Geometric Mean of the two payouts.

Recall: If positive outcomes $X_1, X_2, X_3, \ldots X_n$ have respective probabilities $p_1, p_2, p_3, \ldots p_n$, where $\sum_1^n p_i = 1$, then the Geometric Mean is $\prod_1^n X_i^{p_i}$

The occasions when it is more appropriate to use this mean rather than the usual Ensemble Mean are yet to be precisely determined, but for our purposes the following rule of thumb will apply:

1. The typical experience is more relevant than the ensemble experience – especially when you only get one shot.
2. The outcome is in the form of a multiplicative factor over the initial state.

Rule of Thumb for when to use the Geometric Mean

This bears repeating.

Rule of Thumb

The occasions when it is more appropriate to use the Geometric Mean rather than the usual Ensemble Mean are yet to be precisely determined, but for our purposes the following rule of thumb will apply:

1. The typical experience is more relevant than the ensemble experience – especially when you only get one shot.
2. The outcome is in the form of a multiplicative factor over the initial state.

Looking ahead: Note that this rule does not apply to insurers (they follow the AS), but it does to the individuals who purchase insurance!

Two famous Expectation conundrums

Problems with mathematical Expectation, or Ensemble Average were noticed early on and expressed in the famous **St Petersburg Lottery** problem:

The following game is to be offered: A fair coin is tossed. If it comes out heads you get $1, if not it is tossed again. If it comes out heads you get $2, if not it is tossed again and so on. If the first heads appears at the Nth toss you get $\$2^{N-1}$. Game continues only until the first head.

It sounds like a good game. The question is, How much would you play for a ticket to play this game?

The expected return on the game is $ ∞ (=Infinity). Most people would pay a small price, say under $10 to play it, but no-one is going to give a large sum in return for infinite expectation.
Historically, this problem led to the concept of a Utility function. Personally, I think that even someone with a very concave Utility Function (=risk seeking – e.g. looking to fund a start-up) would bother investing more than small change in this.

The St Petersburg Lottery

Although this problem doesn't quite fit the proportionality condition of our Rule of Thumb, the idea of typicality and the associated Time Average does a good job of pricing.

What is the average waiting time to the first Head?
Ans. On average the first head occurs on the second toss.
Therefore a reasonable ticket price for a typical player = $2
A simple logarithmic calculation shows that this is exactly the Geometric Mean of the payout – actually it is exactly the same calculation as for the mean waiting time.

Does this mean that it would be a good business to offer players a St Petersburg payout for a ticket price of $5 (or even $10)? Emphatically no! You'd be bankrupt pretty quickly.

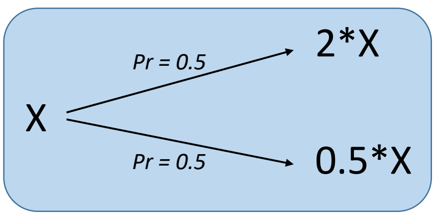
Does that mean that if tickets cost 'only' $5 you should play many rounds? Emphatically no! You'd be cleaned out pretty quickly.

The Two Envelopes Problem

This is another famous Expectation conundrum.

There are two indistinguishable envelopes, each containing a valid check for a certain sum. You know *only* that one amount is exactly twice the other amount.
You pick one envelope, and before opening it you are offered the opportunity to change your choice.

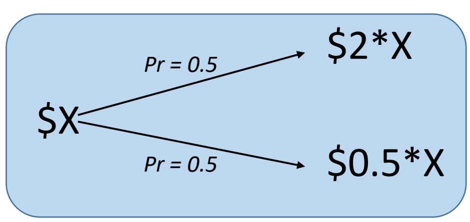
Say that the envelope you originally picked contains a check for $X, then with probability 0.5 the other envelope contains $2*X and with probability 0.5 it contains $0.5X



The expected result of swapping is $1.25*X

This is nonsensical! What is wrong?

The Two Envelopes Problem



E(Swap) = $1.25*X ???
So, you should keep swapping unopened envelopes forever?
This nonsense can be ‘resolved’ using some high-powered conditional probability theory (or so I’m told), but our Rule of Thumb applies perfectly here:

Geometric Mean(Swap) = $ sqrt(2*0.5)*X = $X.

This supports what is intuitively obvious, which is that it makes no difference at all to swap.

The Insurance Paradox

We are now ready to see how this applies to Insurance, in particular to resolving the Insurance Paradox, which is that on the basis of Expectation (= Ensemble Average) no-one should ever want to buy insurance.

Say that the potential loss is L, and the probability of that loss is p.
The insurance premium is $Prem = p*L + C$, where C is the Insurer's margin (= protection against 'Gambler's Ruin'.)

Expectations

	With Insurance	Without insurance
Insurer	$Prem - p * L = C$	0
Insured	$-Prem = -p * L - C$	$-p * L$

This makes no sense on the basis of Expectation. The usual solution is via risk reduction, equivalent to Utility function.

However note that the Insurer plays the AS, while the Insured being in a one-off and 'typical' player situation ought to play the CS.

The Insurance Paradox

Let's see how we can apply the Rule of Thumb to someone seeking insurance:

Recall, the **RoT**:

1. The typical experience is more relevant than the ensemble experience – especially when you only get one shot.
2. The outcome is in the form of a multiplicative factor over the initial state.

The first condition is straightforward. You only get one shot; you can't apply the Law of Averages.*

To meet the second condition, that is, to make it look as if you are in a CS game, just express the possible loss as a proportion of something.
Proportion of what? Of your Net Worth? We will limit ourselves to cases where you are insuring an object or a venture whose loss could be considered to reduce your 'net worth' by a roughly determinable proportion, less than 100%. If this looks like a Utility Function that is no coincidence, but in many cases it can be estimated without any introspection. This limitation will prove not to be as limiting as it looks.

*The cases where the Insured are able to find a way of profiting from the Law of Averages (or the AS,) they generally form a Captive or join a Mutual.

The Insurance Paradox

Let X is the Net Worth of the one seeking insurance, we will let

rX = amount insured, and

p = probability of a loss event, which for now will mean a loss of exactly rX.

The Insurer charges a premium that is greater than prX and to keep the proportionality we will write the
Premium = $prX + cX = (pr+c)X$

We can now factor X out of the calculations, or equivalently let X =1.

Insured's Geometric Mean = $1-(pr+c)$, with insurance; = $(1-r)^p$, without insurance.

So insurance is worth getting if $1-(pr+c) \geq (1-r)^p$.

This is equivalent to $c \leq 1-pr-(1-r)^p$.

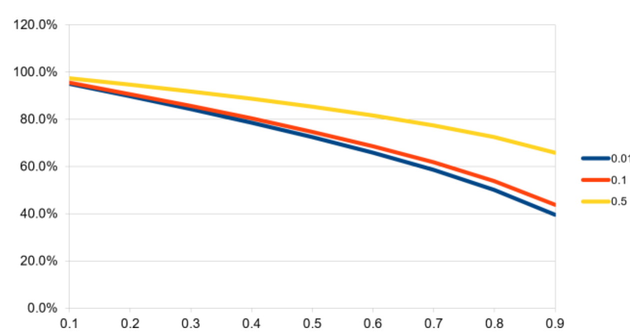
For the Insurer to offer insurance c must be positive, so $1-pr-(1-r)^p$ is the threshold value.

Assuming this to be positive the Insurer's decision on whether to offer an insurance contract will depend on the ELR for the policy =

Expected(Loss)/Premium = $pr/(pr+c) = pr/(1-(1-r)^p)$.

The Insurance Paradox Resolved

We can see that insurance is worth purchasing at all charted conditions, but in some cases the threshold value of ELR is so high that you'd be lucky to find an insurer who would offer it.
For example, if you are insuring 20% of net worth against a risk that is 50% likely, you'll need to find an insurer who'll agree to an ELR on the policy of 94.7% or more – which is unlikely.
But if you are insuring 50% of net worth against a 1% risk, then any policy with an ELR above 72.4% is good economics, and you might find a seller.
At 90% of net worth and a 1% risk, the threshold is 39.5%. That's probably more than you can afford, but the market will take care of it and you'll likely find insurers who will take on the risk at a higher ELR.



Threshold Loss Ratio for the value of insurance, by proportion of 'net worth' insured, at different risk rates.

In the case of extreme possible losses, such as in Liability and MedMal, this model says that insurance is essential and probably very expensive, but if you can find coverage at an affordable ELR you should take it.

End of Digression and segue

This digression has really been about risk and it has aimed at making a distinction between two types of risk situation: 1. those that can benefit from the law of averages, and 2. those that cannot because they are essentially 'one-off'. This is a distinctions that runs through many areas, e.g. frequentist vs. Bayesian probability, the group and the individual, destiny and free-will, ways of thinking about time….

Situation 1: Large numbers of independent instances; Accumulation Strategy; Mathematics based on theory of Expectation works well as model; Utility function is linear, ie, not needed.	⟷	Situation 2: One-shot situation; Player as 'typical' individual; Better understood as a version of the Cumulative Strategy; need to look at Time Averages, a mathematically different approach; Utility function is relevant, and is non-linear, usually logarithmic.

In the insurance market-place sellers are in 1. and buyers are in 2., and this is what makes transactions possible. Insurance companies, however, are in 2. as far as their own solvency is concerned. (Hence reinsurance!) In this respect they need to think in terms of Time Averages, of outcomes projected long into the future. Calendar trends are of first importance here. If your model accounts for all salient trends you are still not done. The correlation between LOBs is the next obstacle that emerges in the time-perspective.

Correlations for Loss Forecasts

Context: We are thinking about correlations in time, correlations of time-series. More specifically in this context, correlations of cash-flows over time.

Correlations are important only in respect of the variability in the forecast, and hence in any calculations of risk capital.

$$Forecast \sim Mean + Variability$$

Mean and Variability are supplied by the model, and both are time dependent.
Variability is expressed as a distribution, whose form may vary with time.
Simplifying:

$$Frc(T) = Forecast(T) = Mean(T) + \varepsilon(T)$$

Where $\varepsilon(T) \sim Dist(0,\ \sigma_T^2)$

If there are two lines:

$$Frc^A(T) = E^A(T) + \varepsilon^A(T); \quad Frc^B(T) = E^B(T) + \varepsilon^B(T)$$

Where $\varepsilon^A(T) \sim Dist(0,\ \sigma(A)_T^2)$ and $\varepsilon^B(T) \sim Dist(0,\ \sigma(B)_T^2)$.

Correlations for Loss Forecasts

$$Frc^{A+B}(T) = E^A(T) + E^B(T) + \varepsilon^{A+B}(T)$$

Where $\varepsilon^{A+B}(T) \sim Dist(0,\ \sigma(A+B)_T^2)$

And $Dist(0,\ \sigma(A+B)_T^2) = Dist(0,\ \sigma(A)_T^2 + 2 * correlation * \sigma(A)_T * \sigma(B)_T\ + \sigma(B)_T^2)$

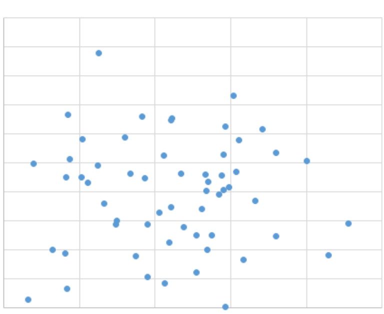
If $correlation = 1$, this is $Dist(0,\ (\sigma(A)_T + \sigma(B)_T)^2)$ and all diversification benefit is lost.

The amount of risk capital we need to hold depends on the size of the variance term.

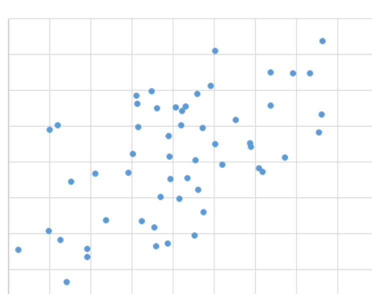
So, we desire the correlation to be as small as possible. And if negative correlations are considered to be too good to be true, then we want them as close to zero as we can get.

Correlations

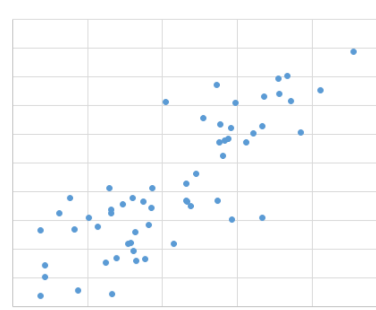
There's a simple formula we all know by which you can compute the correlation between two equal length time series. This boils it all down to a single measure, the Pearson correlation coefficient. There's no need to write this down because it corresponds so well to a picture.



Correlation = 0.0



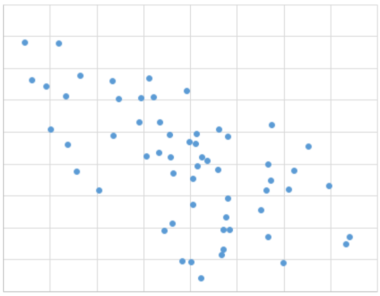
Correlation = 0.5



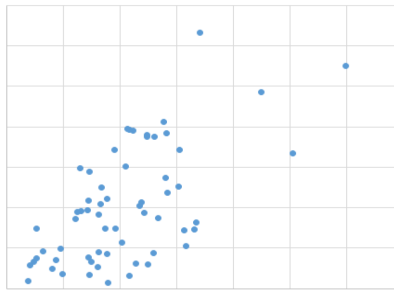
Correlation = 0.8

Correlations

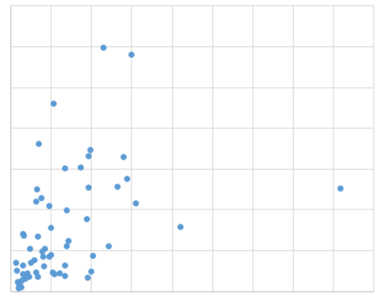
And of course you can also have negative correlations and non-linear correlations. The latter can involve special forms of correlation such as ‘tail-correlation’ – often reducible to the linear form via a transformation.



Correlation = -0.6



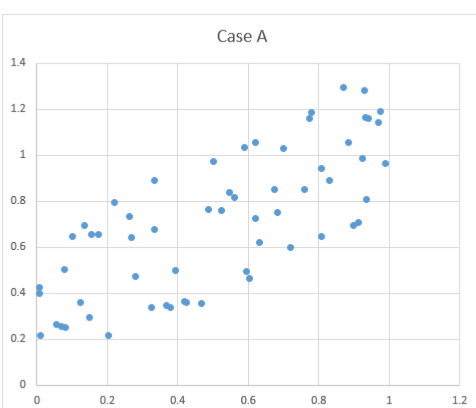
Correlation = 0.6
But clearly non-linear.



Correlation = 0.45
Log transform would reveal an underlying linear correlation of 0.6

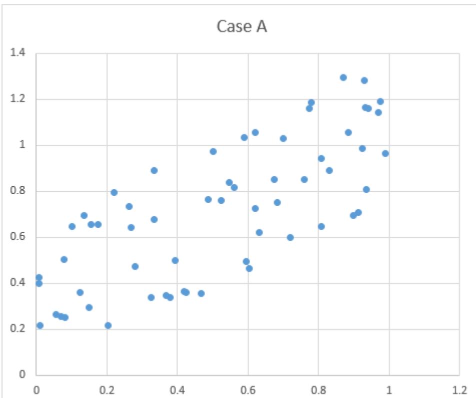
Correlations

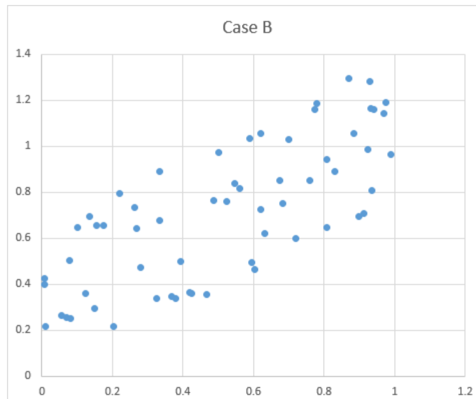
In these graphs I’ve plotted Series X against Series Y and so have obscured the time dependence.
Here are two different cases which both come up with a correlation of around 0.75
We’ll see that the two cases aren’t the same and that the correlation in one case is largely illusory.



Correlations

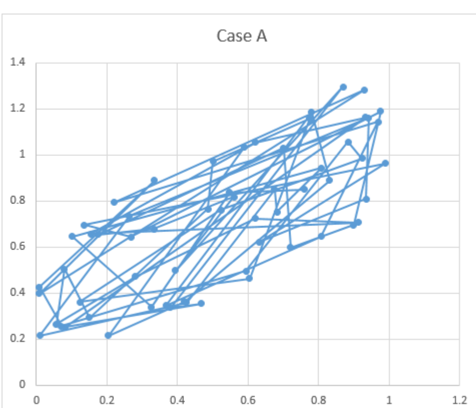
In these graphs I've plotted Series X against Series Y and so have obscured the time dependence.
Here are two different cases which both come up with a correlation of around 0.75
We'll see that the two cases aren't the same and that the correlation in one case is largely illusory.

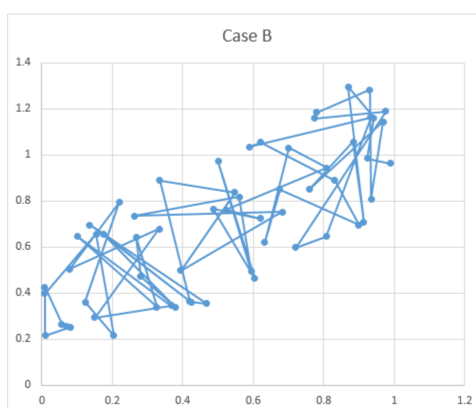




Correlations

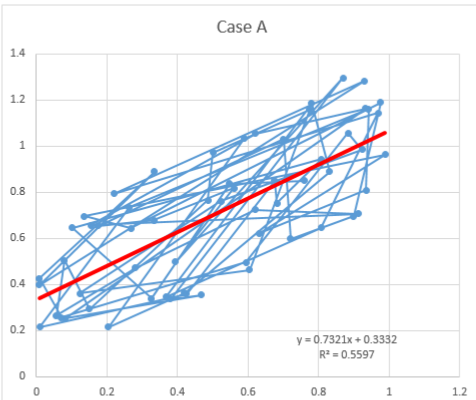
I've added a line that goes in the order in which the data points arose, that is, I've put the timing back in the picture.

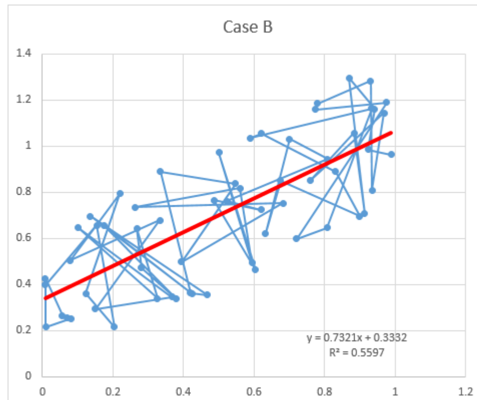




Correlations

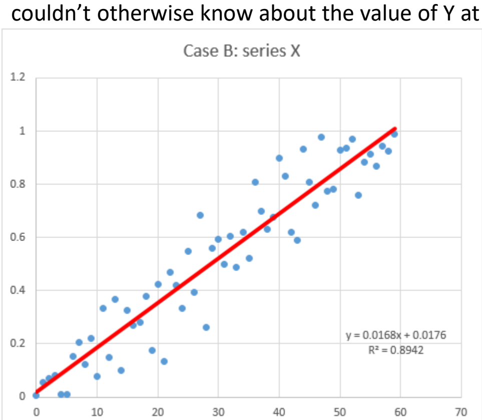
The trend lines are, of course, identical, but some extra structure is apparent in case B. Something that can be modelled independently of the relationship of X and Y. Both X and Y in this case have an increasing trend in time, even if it currently looks a bit drunken. Let's model those.

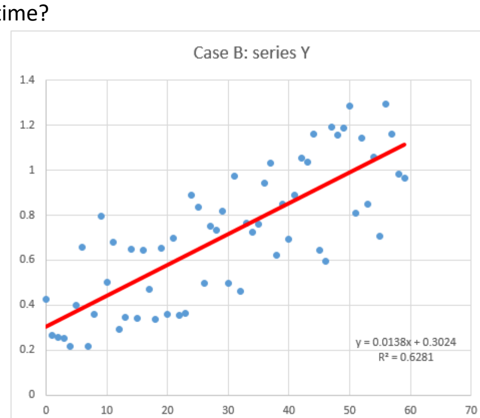




Correlations

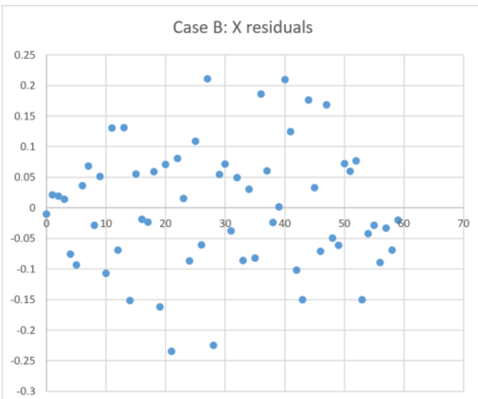
In case B, both the X series and the Y series have linear trends in time. If these were independent measures of interest in a financial context, this relationship would not be missed and any forecasts of X or Y would be based on it. So, if we find out the value of X at some known future time, does this tell us something we couldn't otherwise know about the value of Y at that time?

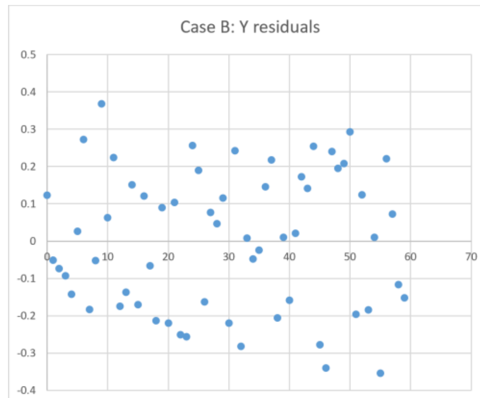




Correlation?

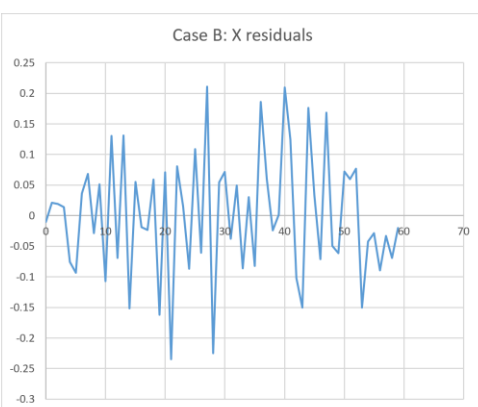
Using the regression to detrend the data we get the respective residuals from our simple linear trend models.

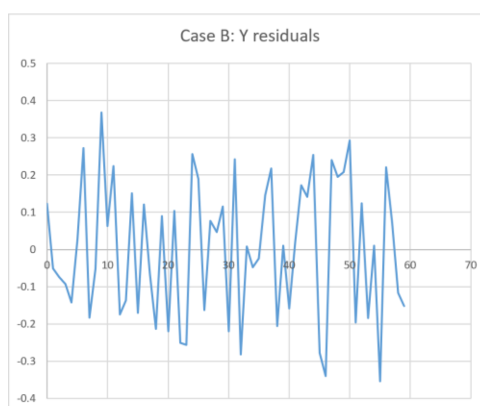




Correlation?

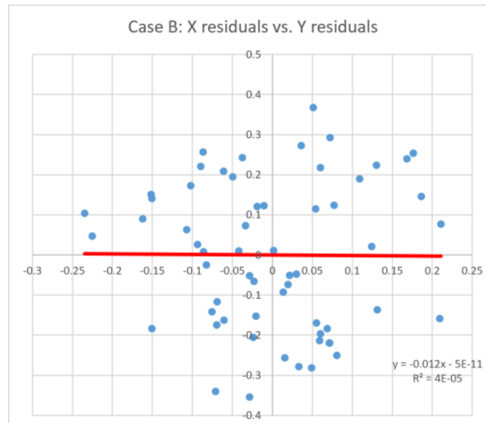
Using the regression to detrend the data we get the respective residuals from our simple linear trend models.





No Correlation!

Plotting the two series against each other no further structure is discernible.
The Pearson correlation is -0.0065.

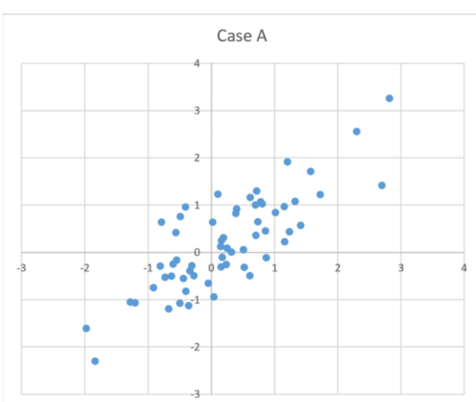


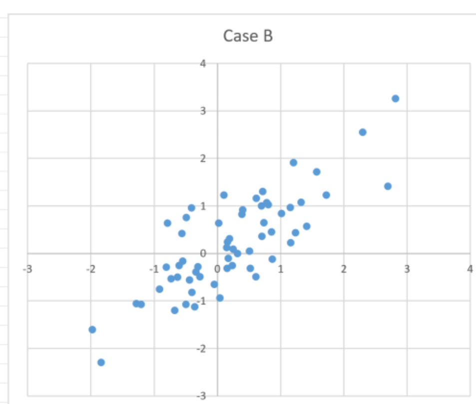
This is good. We can model Series A as linear in time, and Series B as linear in time. Two independent models; no correlation between the variability distributions.
So when we forecast for A + B we get full zero-correlation diversification benefit.

But…
Things need not have come out so nicely….

Correlations

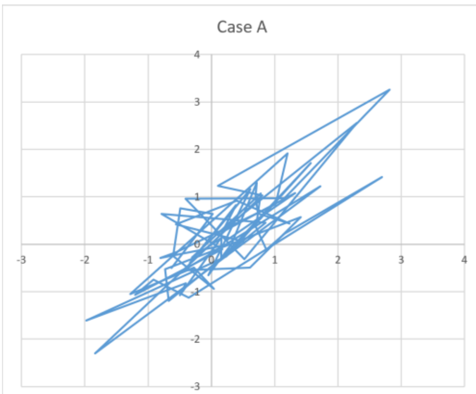
Consider these two cases, which both show a correlation of 0.8

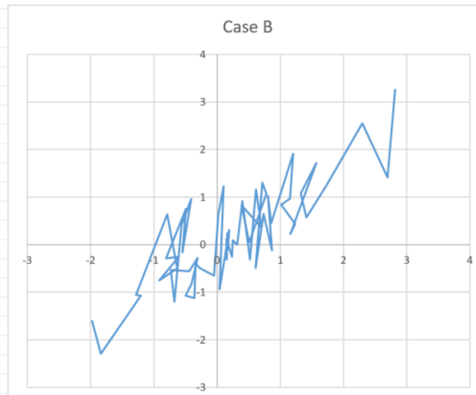




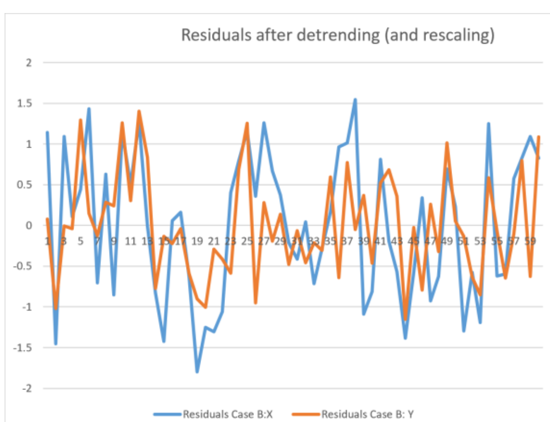
Correlations

Add in the time flow. Case A again is chaotic; there's not much more to be done. Case B again has unmodeled trends. So again we model the (linear) trends in Case B, Series X and Series Y and compute the residuals.





Correlation



Now, even after detrending the residuals have a correlation of 0.55. Separate modelling of the two datasets has reduced the measured correlation from 0.8 to 0.55, but this is significant enough that it really ought to be a part of our model.

Modeling with Correlation

The new bi-variate model ought to look like this, at least to start with.

$$X(t) = \alpha^X * t + \beta^X + \varepsilon_t^X$$

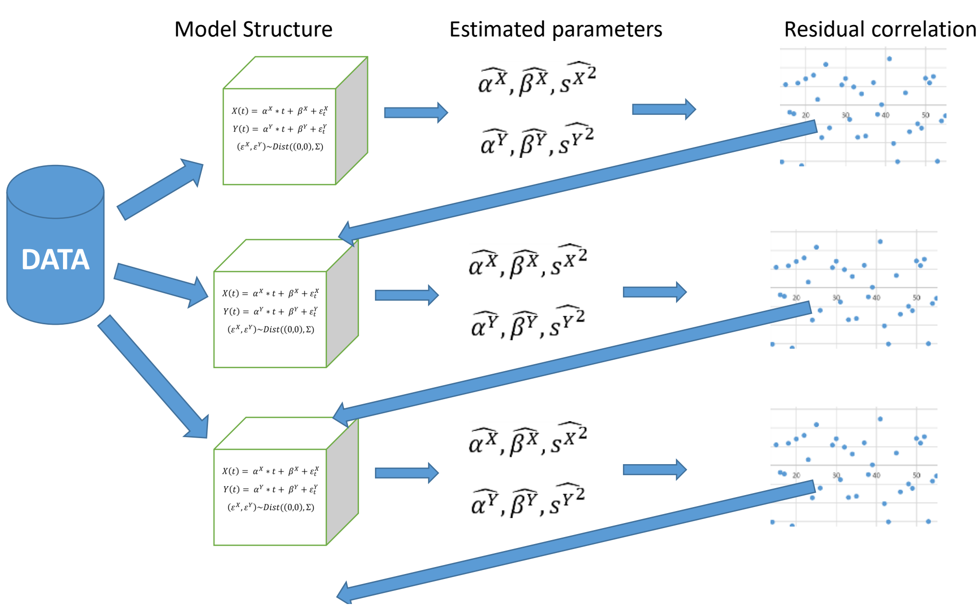
$$Y(t) = \alpha^Y * t + \beta^Y + \varepsilon_t^Y$$

$$(\varepsilon^X, \varepsilon^Y) \sim Dist((0,0), \Sigma)$$

Where $\Sigma = \begin{bmatrix} \sigma^X * \sigma^X & 0.55 * \sigma^X * \sigma^Y \\ 0.55 * \sigma^X * \sigma^Y & \sigma^Y * \sigma^Y \end{bmatrix}$

Here we've actually set the value for the residual correlation in the model, instead of treating like another parameter to be estimated. Actually the correlation is a parameter and will be estimated iteratively. The measured value from the independent modeling case is just the initial value.
It functions as a starting point; successive iterations will close in on the best estimate.

Modeling with Correlation



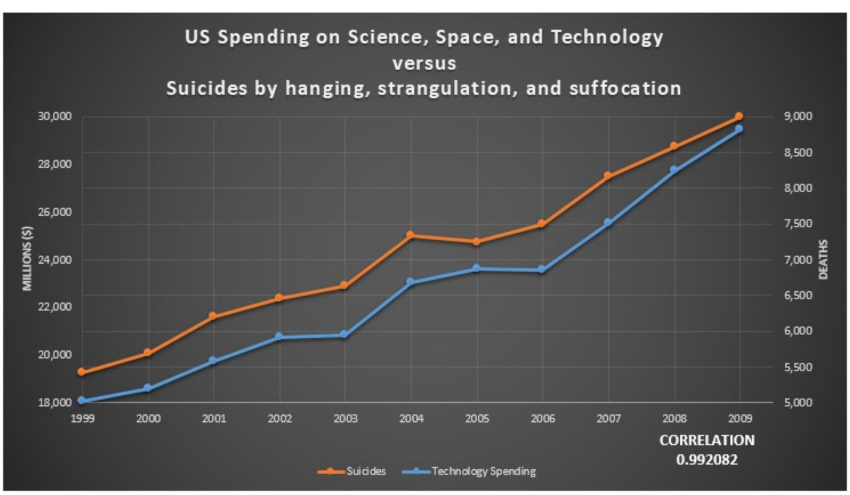
Modeling with Correlation

What happens to the volatility correlation when we do this?

It decreases, but some of the slack is taken up by parameter correlations between the two models. These need to be taken into account in simultaneously forecasting the two series.

The net effect, however, should be a further lowering of the variance of A+B in the forecast region, and hence represents a further lowering of the level of risk capital.

The old saw about correlation not being causation



You've probably seen many examples like this which are taken to show obviously spurious correlations.
But to immediately declare that all the correlation is false still shows some confusion between causation and correlation.
There's no evident causal connection between the two series, and they have not been detrended so the measured correlation of 0.992 is meaningless.

However, note that the changes in rate of increase track pretty well. A simple back-of-the-envelope detrending yields a residual correlation of 0.58. In the immortal words of Nassim Nicholas Taleb in a recent tweet…..

The one about correlation not being correlation

“Sometimes a correlation is not a correlation!”

There is a truth in this, but I think the same point is made in a more precise way by saying: A correlation is only a correlation; whether it is advisable to consider it in making a forecast is another matter. When there is no obvious causal chain connecting the two series, as in the previous slide, the common-sense answer is no, it’s a coincidence and has no place in a forecast model.

As against common-sense, we could argue:

1. Common sense is often doomed to play ‘catch-up’. Data-mining software would typically include every measured correlation, suitably weighted, and these programs generally out-perform educated experts. They are like a way of systematizing ‘beginner’s luck’.
2. It depends on your risk appetite. If your priority is minimizing risk, and the possible correlation has a negative effect, then you are bound to count it in your model. Risk management is about dealing with the unforeseen, and an unforeseen correlation in the model is a perfect example of this. Again, in reference to the first part of this talk, you are in Situation 2: you only get one go, so anything that might improve your odds should be counted.

Casualty Actuarial Society
4350 North Fairfax Drive, Suite 250
Arlington, Virginia 22203

www.casact.org