Balancing Complexity and Parsimony in Models

29 August 2024

Paul Sweeting

Thank you for joining…
…but why listen to me?

- I’ve been building asset models for a quarter of a century
- I’ve taught investment and enterprise risk management
- Much of my book covers the maths of asset models
- Many of my jobs have required me to build asset models
- I can be very cynical!

The challenge

- All models are wrong…
- …but some are useful
- This means they need to be useful enough…
- …and thus have sufficient complexity…
- …but how much is enough, and how much is too much?

Simple models

- Some approaches seek to reduce complexity…
- …as it can make models more tractable
- Main group known as dimensional reduction methods (principal component analysis, singular value decomposition)
- This can be really helpful, particularly for bonds
- However, it can remove/ignore important features…
- …so other approaches try to capture this complexity

Example: modelling bond yields with PCA

- Take your forward curve
- Take logarithms, strip out any trends
- Apply PCA to what's left
- Take the first three factors (which represent volatilities)…
- …and the weights for each forward term
- Model three random normal variables…
- …and weight to give projected (log) forward rates
- Recombine for the forward and spot curves

Example: modelling bond yields with PCA

- BUT
  - What if the log forward rates aren't normally distributed?
  - What if there's skew, or fat tails?
  - Or jointly fat tails?
  - What about negative rates?
  - Etc…
- There are potential solutions…
- (SVD and fitting non-normal distributions, but not trivial)
- …but excessive simplicity can understate risk

Complex models

- Two broad approaches to complexity
- First is econometric
  - Seeks to understand underlying relationships between variables…
  - …and to construct models accordingly
- The second is return-based
  - Doesn't care what processes led to the observed results…
  - …just seeks to model what is seen

Advantages of complexity

- Marketing
- Better reflect underlying distributions
  - Avoid “impossible” returns for bonds
  - Reflect fat tails for risk asset classes
- Capture more complex relationships
  - Build relationships explicitly (econometric “cascade” modelling)
  - Assume the mechanism is unknowable (return-based modelling)

Econometric approaches

- An example (made up, but probably exists somewhere…)
  - Start with price inflation…
  - Built as an autoregressive series driven by short-term rates…
  - …plus some noise
  - Short term rates themselves driven by inflation
  - Bond prices driven by bond yields…
  - …driven by long-term equilibrium, short rates and noise
  - Dividends driven by cash, inflation and noise
  - Equity prices driven by bond yields, dividends and noise
  - …and so on

Econometric approaches

- Extend this across asset classes, markets, currencies
- Use only relationships that show statistical significance
- (Though even here, some of these will be luck…
- …and are subject to random structural breaks…
- …which, of course, can also be modelled)
- Ideally, test significance of system as a whole…
- …rather than all components…
- …and with 240 observations (20 years monthly data)…
- …if you're lucky

Returns-based modelling

- Just care about returns and the relationship between them
- Several aspects to them
  - How are they distributed?
  - Are there any links between returns in subsequent periods?
  - Does the distribution change over time?
  - What is the extent of any relationship between asset classes?

How are returns distributed?

- Are there any fat tails (excess kurtosis), any skewness?
- (Both relative to a normal distribution)
- If so, might be looking at a non-normal distribution
- Student's t-distribution is the most commonly used one…
- …even though it was developed for hypothesis testing
- There also exists skew-t distributions (more than one)…
- …and many other distributions

Are there any links between returns in subsequent periods?

- The returns might not be fat at all…
- …it might just be serial correlation…
- …although this has an uncomfortable implication…
- …that past returns can be used to help predict future ones
- But can use autoregressive processes…
- (and moving averages, and integrated processes)
- …to explain observed returns better

Does the distribution change over time?

- For example, does the volatility change?
- It appears to in practice…
- …and this too can be modelled using a GARCH process
- (Generalised Autoregressive Conditional Heteroskedasticity)

What is the extent of any relationship between asset classes?

- Correlation is the starting point…
- …but what if this isn’t stable over time?
- Or if it is different for jointly extreme events?
- Might call for copulas…
- …which all for the “shape” of dependence to be modelled

Problems with complexity

- Not enough data to justify complex models
- Can only model previous (joint) distributions
- Tendency to spurious accuracy…
- …and risk of excessive sensitivity to inputs
- Is the data really non-normal? (And beware type I errors)
- Importance of non-normality falls as asset classes rise
    - Jointly extreme results less reliant on individually extreme events
- Excessive confidence in results
    - See "marketing"

Determinants of the right level of complexity

- How much data?
  - Need enough data to parameterise your model
  - Amount needed increases (rapidly) with number of asset classes
  - “Curse of dimensionality”
- Absolute or rank?
  - If you only need to know which is “better”…
  - …or an actual value?
- What degree of accuracy?
  - 1 in 200 year event? Or 2 significant figures?

Conclusion

- Complex financial modelling can be great…
- …if it really adds value
- I have a personal prejudice against overly complex models…
- …and complexity being used to market “cleverness”
- Start with the modelling objective, and work back from that