Twitter Thread by macrocephalopod



macrocephalopod
@macrocephalopod



A few things that I didn't cover yesterday when I talked about equity factor models (it's a huge area and it's impossible to more than scrape the surface)

> A few people in the DMs asking about equity factor models so here's a short explainer.
>
> Let's make it a concrete problem -- you are the risk manager at a big multi-manager hedge fund with ~100 sub-PMs each of whom has a portfolio of 10-50 stocks, long and short.
>
> — macrocephalopod (@macrocephalopod) February 2, 2021

1. How do you get the exposure matrix Xt?

There are different ways to estimate it, depending on the factor. Simplest is factors like industry or country exposure where the entries can be 0/1 depending on whether the stock is in that industry/country or not.

Some exposures can be estimated by linear regression on historical data, if you already have a time series which approximates the factor returns. E.g. exposure to the market factor (beta) is estimated this way, by regressing each stock against the S&P 500 (or some other index)

This also works for "macro" factors e.g. you can estimate exposures for each stock to commodity prices, exchange rates, interest rates, GDP or inflation surprises etc by regressing stock returns against the relevant historical time series.

Finally you can have exposures which are heuristically derived from other observable data about the stock, e.g. accounting data, analyst reports, past price movements etc. In this case you find some metric which measures the factor you care about (e.g. price to earnings) and

transform it so that it has a nice distribution in the cross-section - common approaches are z-scoring (subtract mean and divide by standard deviation) or ranking (the stock with the lowest metric gets exposure -1, highest gets +1 and others are linearly spaced between -1 and +1)

You want all the entries in the exposure matrix to have a similar scale (generally -3 <= X(i,j) <= +3 for all entries) as this makes it much easier to compare the factors with each other.

Implementation note -- with linearly dependent factors (e.g. each stock is in exactly one industry so sum of industry exposures equals market exposure) you can't use the normal equations below. You need a constraint on the factor returns. https://t.co/miAGIQPBV5

> Now you have a set of linear equations on each day, and you can solve the linear equations to get the vector of factor returns for each day using the normal equation - pic.twitter.com/YwVkUzSM69
>
> — macrocephalopod (@macrocephalopod) February 2, 2021

(Normally you would require the sum of all industry factor returns to be zero, sum of all country factor returns to be zero etc)

2. How many factors do you need?

It varies depending on the application. The simplest models would have just a few, maybe the market factor plus a couple of others that you care about (think about Fama-French 3 factor or 5 factor model) but it will normally be more.

A quant equity market neutral strategy might have a market factor, 20-40 industry factors, maybe ~10 country factors, 0-10 other risk factors (e.g. commodity exposure, currency exposure) and 10-50 alpha factors. So anywhere from 30-100 factors would be pretty common.

3. Wait, so you can literally make a factor out of anything?

Yes -- you hear a lot about the well known ones like value, momentum quality etc but there are hundreds of others which are widely known in academia and industry and thousands of proprietary in-house factors.

One way to tell if a factor is meaningful is to see how well it explains risk in the cross-section (equivalently what is the volatility of the factor return). For example the US market factor has ~20% annualized vol, a big factor like momentum will have 8-10% annualized vol, and

other factors that explain a meaningful component of returns might have 3-6% annualized vol. By comparison a random factor (literally generate random factor exposures between -1 and +1 each day) will have annualized vol of ~1% on the top 2,000 US stocks.

So if your factor return has only 1-2% annualized vol it is probably not explaining much risk. It may still have a positive expected return, but I would be skeptical whether that is real vs. over-fitting to past data.