Global Factor Premiums notes

August 2020

1 Statistics stuff

1.1 P values / hypothesis testing

Given a prior and a posterior (prior associated with the "null hypothesis"), what can we conclude about a given scenario?

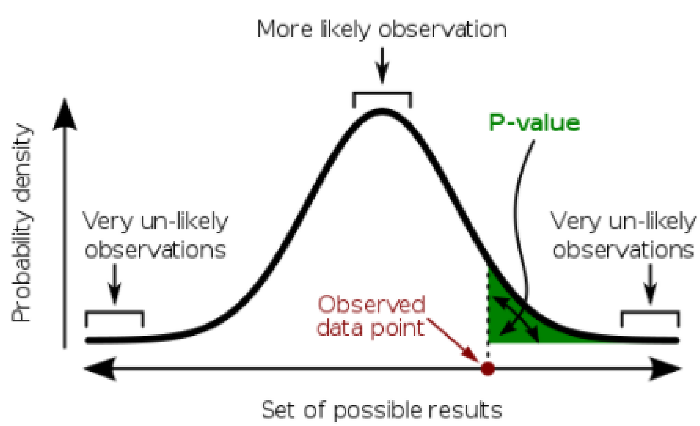
To quantify a measure of confidence that our experiment has uncovered something new (as opposed to no knowledge being gained - null hypothesis), we can compute a "z value"

$$z = \frac{\frac{\bar{x}-\mu}{\sigma}}{\sqrt{n}} \quad (1)$$

We assume our data is normally distributed. If our sample mean is dramatically different from the population mean, the z value will correspond (barring the $\sqrt{n}$ term) to how many standard deviations you are from the mean.

Therefore, a p-value is the probability that a test statistic in a reference distribution exceeds its value in the data.

$$p = Pr(\mathbf{T}(y_{test})) > \mathbf{T}(y))$$



1.2 Bayes factor testing

"Prior odds" are the odds we put on the null hypothesis (relative to an alternative hypothesis), using data external to the study. "Posterior odds" are the odds after seeing experimental data.

The ratio of the posterior odds : prior odds is called the *Bayes factor*, Intuitively, it measures how much our views change after observing new data. It's the ratio of the data's probability under two competing hypotheses.

Although it might make the math majors shutter, we can compute this directly from the same information used to calculate the p-value, if we make some (usually reasonable) assumptions. The strongest Bayes factor (minimum Bayes factor) against the null hypothesis is

$$MBF = e^{\frac{-Z^2}{2}}$$

Where Z is the number of standard errors from the null value. This is just called the "minimum Bayes factor". In the case where the prior probability distribution is symmetrical and descending around the null value, this reduces to a simpler equation.

For a Bayes factor of 1/10, i.e. the odds of the null hypothesis decrease by a factor of 10 after the study, we can say we have moved from our original position, but we cannot make an absolute claim. If the initial odds of the null hypothesis were 1 (probability of 50%), the odds after the study would be 1/10 (probability of 9%). If the initial odds of the null were already quite high, say 9 (probability of 90%), the odds after the study would be 9/10 (probability of 47%), which is still quite probable.

$$MBF_{SD} = \text{-}e\ p\ ln(p)$$

Where p is the p-value (since they are computed from the same information). Remember that in the earlier example of Bayes factors, that they only tell us how much we have learned, and we still need a frame of reference with which to compare it - namely the prior odds. To remedy this, we can then compute a Bayesian p-value as follows;

$$p_{Bayesian} = \frac{MBF_{SD} * odds_{prior}}{1 + MBF_{SD} * odds_{prior}}$$

Problem:[1] We do not know the prior odds. **Therefore**, what pick some "confidence value" to be the p-value, and solve for the prior odds.

$$odds_{break\ even} = \frac{\alpha}{(1-\alpha) * MBF_{sd}}$$

[1] All ideas past this point are original ideas presented in the paper

The authors argue against p-values because they require no context to interpret- which is a bit of a statistical no-no in my opinion. To compute an interpretability metric using the above, we require the context of confidence, which can be interpreted as "skepticism". In other words, given some level of skepticism, we can interpret the "break-even" prior odds as being the odds that would be required to just accept the alternative hypothesis.

For hypothesis testing, the magnitude of the break even odds ratio is important. In the paper, a confidence level of 5%, and a prior odds ratio of 4:1, is rated as "perhaps". As the ratio becomes large, more skepticism is required to reject the alternative hypothesis.

2 Testing for significance

2.1 Robustness to researcher degrees of freedom

They tested the return factors in a variety of scenarios that they deem "common degrees of freedom" that could affect results;

- Rolling 10 year sample periods
- Period of portfolio rebalancing
- Trimming of extreme positive returns
- Accounting for lagged implementation (waiting to buy for some interval after a signal)

$$p = \int_{z}^{\text{inf}} \mathcal{N} dx \tag{2}$$

2.2 Large sample

They use data spanning 1800-1980. They also supplement this with a modern sample, 2012-2016. If their results were due to unintentional manipulation of confidence metrics, they would expect any significant factors to disappear in this new sample period.

3 Experiment

3.1 Four asset classes

They study the presence of each factor they test in four different asset classes;

1. International equity indices
2. 10-year government bond indices

3. Commodities
4. Currencies

The authors stress the point that they did not study factors across multiple asset classes to gain statistical confidence - rather, they expect certain factors to be present in some markets and not others. Nevertheless, they do compute an "aggregate" break-even odds for each factor spanning all 4 asset classes.

3.2 "Global"

The dataset they construct (beginning at 1800) consists of the following markets-

- U.S. and U.K. markets for Equities
- U.S, U.K, and French bond markets
- Wheat, Cotton, Cocoa, Copper, Silver
- (GBP/USD), (FRF/USD), (USD/USD) currency pairs.

Over the years, the number of markets gradually increased to a maximum of 68. They exclude markets that experienced hyperinflation over this time period.

3.3 The factors

Definitions are somewhat ambiguous, as they were in the original paper. When I reproduced some of the results, I found that the ambiguity did not really matter.

1. **Trend:**[3] 12 period trailing returns[2]
2. **Momentum:**[3] Top performers in a single period measured by return in that period [4]
3. **Value:** Depends on asset class
    - Dividend yield for equity indices
    - Real yield for bonds
    - 5-year reversal in spot price for commodities
    - Absolute and relative purchasing power parity for currencies

[2] When I reproduced this factor, I got roughly the same results doing a few different things - having a "buy" signal cutoff being positive average trailing 12 period returns, again with 10%, and again with positive *t - t-12* returns. I figured that in our models, we can use a learnable parameter for the cutoff.

[3] Jax code available on my Github

[4] Similar to the above, the top % did not really matter - again, this can be a learnable parameter in our models.

4. **Carry:** "Implied yield for each instrument"- Something about the difference between the yields of two assets - typically currency pairs or bonds of different durations. I am not sure what this would be for equities/commodities.

5. **Seasonality:**[3] Average returns on a certain period, over every occurrence of that period over some time period. Example: Average returns over every March 8th for the last 20 years to make trade decision on March 7.

6. **Betting Against Beta:**[3] Buy low beta assets. Beta is defined as follows;

$$\beta = \frac{Cov(r_{Asset}, r_{Index})}{Var(r_{Index})}$$

3.4 Experimental details

They first try replication on the period usually used for finance studies - between 1981 and 2011. They find that at best, most of the break-even prior odds were pretty low, despite the p-values being significant. Low break-even odds indicates that although there may be an anomaly present, one does not need to be very skeptical to disregard it.

Since Bayes factors intrinsically measure "how much have you learned from the study", we would expect more significance were there an actual trend if we have a larger sample period. Therefore they go on to test for factors on the period from 1800-1981.

Table III: Statistical perspectives on global return factors: replicating sample

The table summarizes various statistical perspectives on the historical performance of the global return factors. Shown per factor per asset class are the historical frequentist p-value ("p-value"), Bayesianized p-value using a 4-to-1 prior odds ratio ("Bayesian-p") and break-even prior odds at a 5% confidence level ("BE-odds") of its performance. The sample starts in January 1981 and ends December 2011 and is at the monthly frequency. Covered are equity indices ("Equities"), 10-year maturity government bond indices ("Bonds"), commodities ("Commodities"), currencies ("FX"), and their equally weighted combination across the four asset classes ("Multi Asset").

		Trend	**Momentum**	**Value**	**Carry**	**Seasonality**	**BAB**
Equities	p-value	0.00	0.00	0.06	0.00	0.05	0.04
	Bayesian-p	0.04	0.05	0.64	0.00	0.60	0.58
	BE-odds	4.03	3.44	0.10	947.75	0.12	0.13
Bonds	p-value	0.00	0.11	0.11	0.00	0.62	0.05
	Bayesian-p	0.17	0.72	0.72	0.01	0.76	0.60
	BE-odds	0.93	0.07	0.07	22.78	0.06	0.12
Commodities	p-value	0.00	0.00	0.22	0.01	0.01	0.95
	Bayesian-p	0.00	0.04	0.78	0.40	0.29	0.34
	BE-odds	1,255.10	4.03	0.05	0.28	0.47	0.37
FX	p-value	0.00	0.20	0.08	0.00	1.00	0.47
	Bayesian-p	0.17	0.78	0.68	0.17	1.00	0.79
	BE-odds	0.93	0.05	0.09	0.93	0.00	0.05
Multi Asset	p-value	0.00	0.00	0.00	0.00	0.03	0.03
	Bayesian-p	0.00	0.00	0.13	0.00	0.51	0.53
	BE-odds	>9,999	105.45	1.26	>9,999	0.18	0.17

4 Results

Table V: Statistical perspectives on global return factors: 1800 - 2016

The table summarizes various statistical perspectives on the historical performance of the global return factors. Shown per factor per asset class are the historical frequentist p-value ("p-value"), Bayesian p-value using a 4-to-1 prior odds ratio ("Bayesian-p") and break-even prior odds at a 5% confidence level ("BE-odds") of its performance. The sample starts in January 1800 and ends December 2016 and is at the monthly frequency. Panel A shows the results for the 'new sample' periods (1800-1980 and 2012-2016), Panel B shows the full sample results (1800-2016). Covered are equity indices ("Equities"), 10-year maturity government bond indices ("Bonds"), commodities ("Commodities"), currencies ("FX"), and their equally weighted combination across the four asset classes ("Multi Asset").

Panel A: 1800-1980 and 2012-2016

		Trend	**Momentum**	**Value**	**Carry**	**Seasonality**	**BAB**
Equities	p-value	0.00	0.00	0.00	0.00	0.00	0.00
	Bayesian-p	0.00	0.00	0.05	0.00	0.00	0.00
	BE-odds	>9,999	>9,999	3.67	>9,999	>9,999	2,021.00
Bonds	p-value	0.00	0.00	0.00	0.00	0.00	0.77
	Bayesian-p	0.00	0.00	0.06	0.00	0.00	0.68
	BE-odds	>9,999	280.39	3.23	>9,999	>9,999	0.09
Commodities	p-value	0.00	0.32	0.00	0.01	0.00	0.40
	Bayesian-p	0.05	0.80	0.00	0.36	0.00	0.80
	BE-odds	3.33	0.05	>9,999	0.33	>9,999	0.05
FX	p-value	0.00	0.00	0.72	0.00	0.00	0.56
	Bayesian-p	0.00	0.00	0.72	0.00	0.00	0.78
	BE-odds	>9,999	546.48	0.07	>9,999	147.25	0.05
Multi Asset	p-value	0.00	0.00	0.00	0.00	0.00	0.06
	Bayesian-p	0.00	0.00	0.00	0.00	0.00	0.64
	BE-odds	>9,999	>9,999	>9,999	>9,999	>9,999	0.11