Data snooping and other model specification problems

Bernt Arne Ødegaard

24 November 2021

1 Data snooping

At the end of any empirical class is a good time to reiterate that any test we choose to run is subject to the fact that we have specified the hypothesis/model we apply it to.

It is a common way of doing empirical research that one modifies the model after having subjected it to the data. While this is of course legitimate, it always run the danger that the model is built to fit the data so well that while the model is accepted by the test that are run, the model will not survive a meeting with new data. The question is when the modeling and specification is colored by the data to an unacceptable degree.

Data snooping biases is a common name for this.

For example, in the finance context, consider the fact that we after a number of decades of research has ended up with using portfolios sorted on size, book/market and momentum to explain the crossection of stock returns.

A worry is that these characteristics of the data has emerged as a result of doing analysis on the same set of data over and over, these characteristics are just relevant by chance in the current dataset?

This kind of worry motivates the analysis of Lo and MacKinlay (1990).

We will not have time to cover the analysis of the paper, but anybody ending up doing empirical asset pricing research should look at this.

Let me just quote the abstract of the paper.

> *Tests of financial asset pricing models may yield misleading inferences when properties of the data are used to construct the test statistics. In particular, such tests are often based on returns to portfolios constructed by sorting on some empirically motivated characteristics of the securities, such as market value of equity. Analytical calculations, Monte Carlo simulations, and two empirical examples show that the effects of this type of data snooping can be substantial.*

They have also a very pertinent conclusion:

> .. nonexperimental inference may never be completely free from data-snooping biases since the attention given to empirical anomalies, incongruities, and unusual correlation is also the modus operandi fore genuine discivery and progress in the social sciences. Formal statistical analyses such as ours may serve as primitive guides to a better understanding of economic phenomena, but the ability to distinguish between the spurious and the substantive is likely to remain a cherished art.

2 Does Academic Research Destroy Stock Return Predictability?

The paper Mclean and Pontiff (2016) is an interesting way of asking to what degree results are due to data mining.

Consider an academic paper that identifies a new variable that seem to predict stock returns.

What are the alternatives?

- The effect is in reality not there, there is actually no predictability
    - The effect is due to data mining)
- The effect is there, one can predict stock returns.

So, what should happen if one returns to a predictive variable *after* the publication of the academic paper that first identifies it.

One can here have several hypotheses;

- The effect is due to data mining
    - $\rightarrow$ The effect should disappear
- The effect is due to risk differences, or other rational reasons.
    - If the predictability is because one identifies the high risk stocks - those *should* have a higher return.
    - Similarly with non-risk, such as transaction costs
    - $\rightarrow$ The predictabity should remain
- The effect is due to misprising
    - Investors not realizing that the predictive variable should be used in their trading decisions.
    - $\rightarrow$ The effect should disappear
    - $\rightarrow$ Alternatively, if there are impediments to trading allowing mispricing to persist, the predictability should be less.

Specifics

Of interest: Cross-sectional predictability (ability to rank returns on same date/time period)

Findings

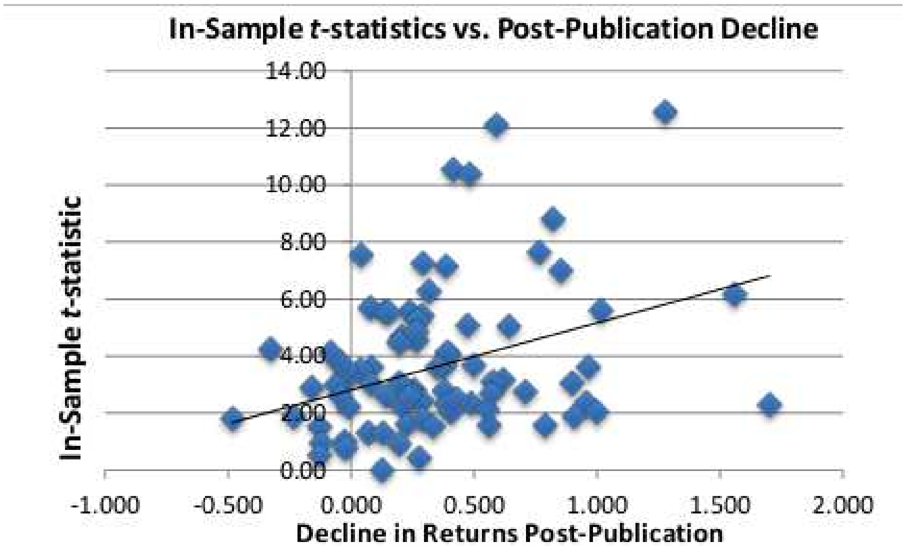
Measuring returns of long-short portfolios sorted on the predictive characteristic.

- Declines by 26% out of sample
- Declines by 58% post-publication.

Hypoteses

- Stock return predictability disappears entirely – Rejected
- Stock return predictability remains unchanged – Rejected
- Findings consistent with mispricing accounting for some or all of the original return predictability, and investors learning about this mispricing.

Cute picture



3 Modelling of asset prices

While data snooping as a concept has always been with us, it has recently seen some important applications in crossectional asset pricing, a major topic of this course.

After the specific example of the Mclean and Pontiff (2016), let look at a more wide-ranging discussion, that put the cats among the pigeons.

3.1 The Harvey 2017 Presidential address

Harvey (2017) is a Presidential Adress to the American Finance Association. Some of these addresses have had impact on how research is being done, others are quickly forgotten. This is among the more cited addresses.

Abstract:

> Given the competition for top journal space, there is an incentive to produce "sig- nificant" results. With the combination of unreported tests, lack of adjustment for multiple tests, and direct and indirect p-hacking, many of the results being published will fail to hold up in the future. In addition, there are basic issues with the interpre- tation of statistical significance. Increasing thresholds may be necessary, but still may not be sufficient: if the effect being studied is rare, even t > 3 will produce a large num- ber of false positives. Here I explore the meaning and limitations of a p-value. I offer a simple alternative (the minimum Bayes factor). I present guidelines for a robust, transparent research culture in financial economics. Finally, I offer some thoughts on the importance of risk-taking (from the perspective of authors and editors) to advance our field."

Harvey further summarizes his points as follows

- "Empirical research in financial economics relies too much on p-values, which are poorly understood in the first place.
- Journals want to publish papers with positive results and this incentivizes researchers to engage in data mining and "p-hacking."
- The outcome will likely be an embarrassing number of false positives—effects that will not be repeated in the future.

- The minimum Bayes factor (which is a function of the p-value) combined with prior odds provides a simple solution that can be reported alongside the usual p-value.
- The Bayesianized p-value answers the question: What is the probability that the null is true?
- The same technique can be used to answer: What threshold of t-statistic do I need so that there is only a 5% chance that the null is true? The threshold depends on the economic plausibility of the hypothesis."

The address makes a number of points, will not discuss all

Publication bias This refers to the fact that journal editors is more likely to reward (with publication) *new* discoveries. To do that it is necessary to reject a null (of no effect).

For example,

distribution of t-stats for studies looking at "factors" in the crossection

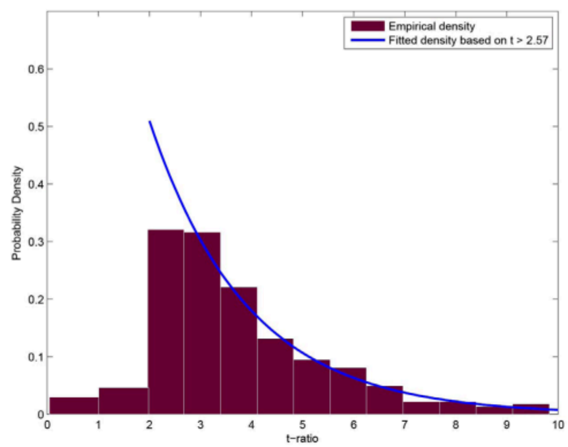


Figure 1. Distribution of reported *t*-statistics from factor studies, 1963 to 2012. Evidence from Harvey, Liu, and Zhu (2016).

p-hacking Making much of the same point as **?** and Mclean and Pontiff (2016). Since everybody has access to the same CRSP data, one can keep trying new ways of sorting the data until the p-value goes above a desired significance-level (p-hacking).

Could this be fixed by increasing the magnitude of p-level necessary? (e.g. to 3) – no, according to Harvey.

And how should you interpret p-values?

Are we all Bayesians? Claim: The more unlikely the theory, the easier it is to by chance confirm it.

→ that the plausibility of a particular theory should be a factor in deciding how much evidence we need.

If you accept the previous statement: You are a Bayesian, because you bring in prior beliefs.

Can you adapt Bayesian methods to more formally ask how much evidence is necessary?

Agency problems – publications Observed behaviour result of incentive to get published.

Possible solutions:

- Registered reports – once the investigation is approved by the journal, the editor promises publication no matter the results.
- Increase the cost of p-hacking: Replications repudiating earlier results – lower prestige of earlier results.

3.2 What happened to the search for factors?

The search for factors in the crossection is one of the examples used in the above articles, and the literature is still evolving here.

Some of the later contributions even introduces the characterization of the “Factor Zoo” Feng, Giglio, and Xiu (2020).

In terms of methods, an interesting counterpoint to the p-hacking arguments of Harvey is a recent JF paper by Andrew Chen.

He arguest that not all factors identified can be spurious.

Ask the question: What is the probability of identifying *this* many factors in the same dataset? Turns out the probability of that is infinitesimal.

To quote Chen’s abstract:

> Suppose that the 300+ published asset pricing factors are all spurious. How much p-hacking is required to produce these factors? If 10,000 researchers generate eight factors every day, it takes hundreds of years. This is because dozens of published t- statistics exceed 6.0, while the corresponding p-value is infinitesimal, implying an astronomical amount of p-hacking in a general model. More structure implies that p-hacking cannot address about 100 published t-statistics that exceed 4.0, as they require an implausibly nonlinear preference for t-statistics or even more p-hacking. These results imply that mispricing, risk, and/or frictions have a key role in stock returns.

So, there must be*some* structure in the crossection of stock returns.

4 Conclusion

So where are we?

- The finance research profession is much more aware of the potential for data snooping.
- Reproducibility – include data/code with research.
- Revisit old studies. Do they hold up to new data?
- Also be interested in non-results.

References

Guanhao Feng, Stefano Giglio, and Dacheng Xiu. Taming the factor zoo: A test of new factors. *Journal of Finance*, 75(3): 1327–1370, 2020. doi: 10.1111/jofi.12883.

Campbell R Harvey. Presidential address: The scientific outlook in financial economics. *Journal of Finance*, 72(4):1399–1440, 2017.

Andrew W Lo and A Craig MacKinlay. Data snooping biases in tests of financial asset pricing models. *Review of Financial Studies*, 3:431–467, 1990.

R David Mclean and Jeffrey Pontiff. Does academic research destroy stock return predictability? *Journal of Finance*, LXXI (1):5–30, 2016.