Explanatory Masks for Interpretable Financial Analysis

Grant Holtes

April 9, 2024

1 Introduction

The popularity and effectiveness of "black box" machine learning models such as random forest and neural networks has driven the development the field of Explanatory AI (XAI), "a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms" (IBM, 2020). While true black box models can be less popular in quantitative finance, both the black box and traditional models often raise similar questions regarding how a result was determined. This note examines how one technique, data masking, can be borrowed from the field of computer vision and applied to financial use cases, using the simple illustrative example of correlation estimation.

2 Explanatory masks in computer vision

An "explanatory mask" aims to determine how each part of the input data influences the model's prediction. A simple method to accomplish this is to black out or "mask" areas in the image, and measure how much the prediction is impacted by taking the difference of the whole image prediction and the masked prediction. If done for every area within the image, the relative importance of each area in the image can be computed, which is often visualised as a gradient map of importance, as in Figure 1

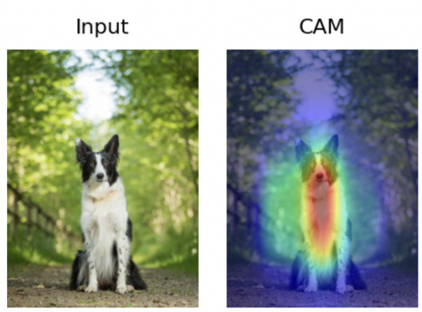

Figure 1: Example of an explanatory masks from Fernandez, 2020

3 Application to financial datasets

This approach can be quite cleanly shifted to use with financial data if the mask is applied to periods in the input time series rather than areas of an image. The contribution to a result from a particular period in time can be measured as the difference between the whole sample result and the result where the time period is excluded.

More formally, given a time series, X, and a model, $F(X)$, the impact of a time period, t, on the whole sample result can be given by Equation 1

$$\Delta F(X, t') = F(X) - F(X | t > t' + \epsilon \vee t < t' - \epsilon) \quad (1)$$

Where ϵ is the time window size, which determines the length of the mask applied. This could be set such that only one observation is masked, or larger so entire months or years are masked at once. Evaluating $\Delta F(X, t')$ for every t' in the time series allows for the visualisation of impact over the dataset.

4 Usage with correlation estimates

Correlation estimation provides a simple example. The graph below shoes the impact of each time period on the correlation between small and large cap stocks. It is clear that the correlation is stable with the exception of market shocks, with the dotcom crash decreasing the correlation by 6 percentage points, while the GFC and Covid pandemic increased the correlation by 2 percentage points.

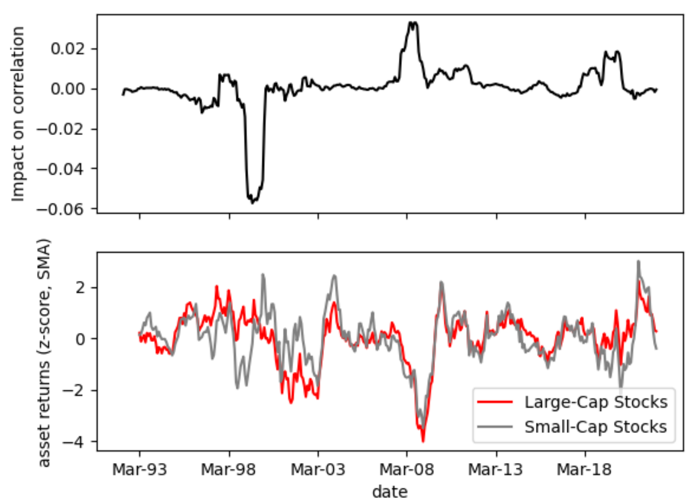

Figure 2: Evaluation of $\Delta F(X, t')$ on the correlation between large and small cap US equities, and normalised asset class returns. 12 month window used.

Selected further examples are provided in the appendix.

5 What about conditional estimates?

The obvious alternative to this method is to compute the analysis in question on a sliding window of time over the data - a estimate conditional on time. For correlation, this produces very similar results to the data masking approach, as in Figure 3.

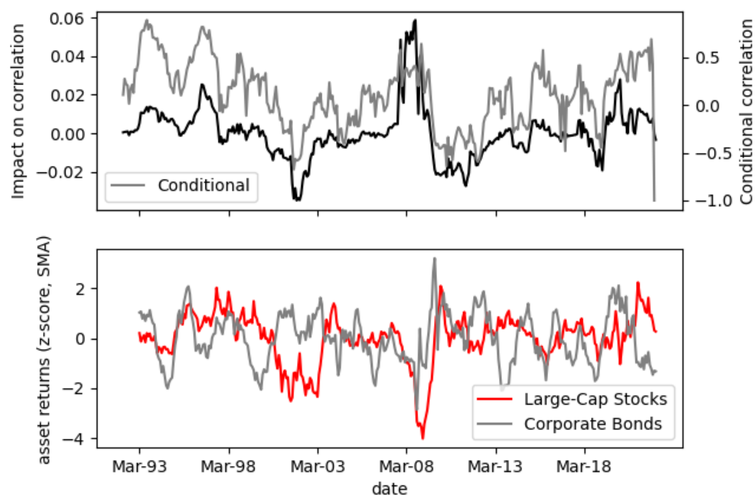

Figure 3: Evaluation of $\Delta F(X, t')$ on the correlation between large and small cap US equities, and normalised asset class returns. 12 month window used.

The data masking approach has a few advantages:

1. Small time windows, ϵ, can be used, even single time step size. To directly compute the results on such small samples sizes may not be reliable or possible depending on the properties of the estimator.

2. The larger sample sizes used produce less noise.

3. The method explicitly answers the question of "how does this time period impact the result".

Note:

References

Fernandez, F.-G. (2020). *Torchcam: Class activation explorer* [accessed at github.com/frgfm/torch-cam].

IBM. (2020). *What is explainable ai?* [accessed at www.ibm.com/topics/explainable-ai].

6 Appendix

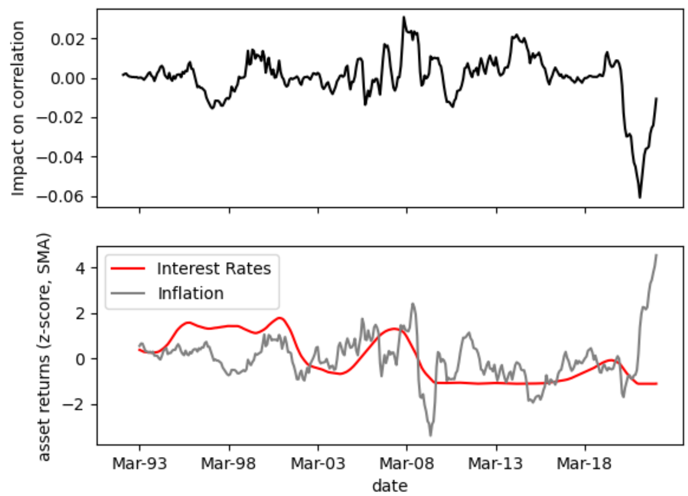

Figure 4: Evaluation of $\Delta F(X, t')$ on the correlation between interest rates and inflation, and normalised series values. 12 month window used.

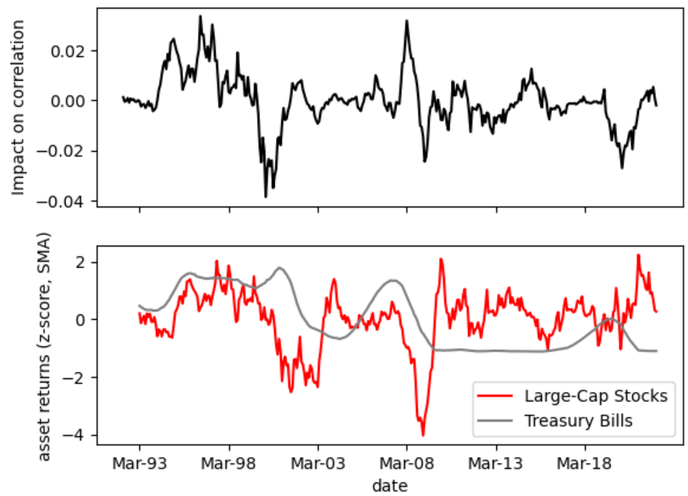

Figure 5: Evaluation of $\Delta F(X, t')$ on the correlation between large cap stock and treasury bill returns, and normalised series values. 12 month window used.