This document proposes avenues to identify causation patterns for variables with limited historical data, for instance for assets in Private Equity or Private Debt. This paper specifically focuses on causation, and not correlation, as correlation doesn't necessarily imply causation, and that the correlation coefficient should be computed on large datasets to be relevant.

1. Introduction

The purpose of this document is suggesting a line of thought to assess the causation between variables with limited historical datasets, for instance for Private Equity or Private Debt assets. The concept of causation is chosen over the notion of correlation as correlation does not necessarily imply causation (and *vice versa*).

The first Section of this document is dedicated to highlight the key differences between correlation and causation, the second one will propose avenues to derive causation patterns between series with limited historical data.

2. Causation *versus* Correlation

If both notions tend to be very similar, causation and correlation are not substitutable. Causation between two variables exists when an action from a variable undeniably impacts the other variable; correlation means that a relationship exists between two variables, but the action from one doesn't necessarily impact the other one.

Causation may happen with no correlation, for instance because of too limited historical datapoints or manipulated datasets, on the other hand, highly correlated assets may present low causation because of distribution specificities (*i.e.* non-linear relationship between variables, outliers in the distribution, data clustering, or heteroskedasticity).

3. Inferring Causation

3.1. From Systemic Shocks

A systemic (aka macro or system-wide) shock is defined as an exogeneous event implying a collapse of an industry, company, or even of the economy as a whole. If the shock first impacts a company or an industry and then spreads to the economy as a whole, one talks about a contagion effect.

Inferring causation from a systemic event offers as main advantage to simplify the modelling by reducing the number of pairwise correlation to assess; one does not assess how much one variable can be explained by another variable, but well the sensitivity of each variable to a common factor, usually the state of the economy.

This approach is frequently used to model joint-default correlations, rather than assess joint-correlations between every pair of variables, the default distribution is represented as a function of a latent variable capturing a macro and idiosyncratic component. In short, rather than assuming that the value of a one variable directly depends on the value of other ones, the causation is described as the dependence of the value of the variable on the general state of the economy, and all variables are linked to each other by a single risk factor.

The idiosyncratic risk that could unpredictably be captured into the sensitivity factor is expected to be reduced by diversification. One expects for a large portfolio of companies that the sensitivity factor only captures the risk common to companies sharing observable characteristics, and not specific to every company. If this observation holds for large portfolios, macro shocks may generate idiosyncratic patterns difficult to identify and diversify. As a consequence, measuring the causation from cyclical patterns could offer a simpler and more robust alternative.

3.2. From a Cyclicality Pattern

A cyclical variable is sensitive to general macroeconomic conditions; a cyclical company or industry tends to perform poorly during a recession, while a non-cyclical company or industry tends to perform pretty well during economic downturn as they propose products still in demand.

Cyclical causation can be inferred by the dependence and sensitivity of industries or companies to the economic cycles. Again, the idiosyncratic component won't be captured into the sensitivity factor as portfolios are expected to be sufficiently large and diversified.

Modelling causation from cyclical patterns or from systemic shocks may appear pretty similar at first glance, but those two approaches should be analysed separately, nonetheless. Cycles are recurrent patterns, with known - or at least expected - amplitudes, systemic shocks are random by nature, with amplitudes difficult to predict; as a consequence, systemic shocks may better capture the unexpected loss component of a portfolio than a cyclical pattern analysis.

3.3. From A/B Testing

Causation can be tested by running experiments, for instance by running A/B testing. The test consists in having multiple versions of a single variable and check how the other variables respond against those different versions. This method is frequently used in Marketing, for instance, to assess how final users react to various versions of a same webpage.

Kind of similar approach could be applied to the Private Equity or Private Debt world, where the testing would rather take the form of a stress-test. The test could consist in checking how valuations of companies are impacted by stressing a common balance-sheet or income statement component and causation patterns between variable can potentially be drawn from the result of the stress-testing.

3.4. From a Scatter Plot Analysis

The approach consists in analysing a scatter plot and identifying potential causation patterns. The scatterplot shows the relationship between two numerical variables and suggests correlations between them.

As defined above, a high correlation coefficient doesn't necessarily mean cause and effect, but analysing the chart may confirm that a causation exists and confirm if the correlation coefficient - if computed - tends to be relevant or bias.

Patterns can be identified by analysing the direction (positive or negative relationship), the form (linear, non-linear, or cluster), and the intensity of the relationship (how quick a variable change *per* unit of change of another variable).