Conditional quantile estimation using high-dimensional time series data

Motivation

Stock market returns are subject to large drawdowns, which can take years to recover. The global financial crisis led to a fall of $7.9 trillion in the market capitalisation of the S&P500. It took investors over 3 years to recover their capital, assuming they stayed fully invested: many do not. The fallout from the dot-com bubble took even longer to recover. Avoiding large drawdowns is crucial for savers and pensioners. It can have a dramatic effect on the ending wealth of their portfolios. Abstract movements in stock markets are life-changing events.

As asset allocators, the view of future returns is a crucial input in securing solid performance and avoiding large drawdowns. The challenge is not just to understand the median or mean return but the distribution of returns and the size and probability of tail risk. The problem we outline below is a significant part of this challenge.
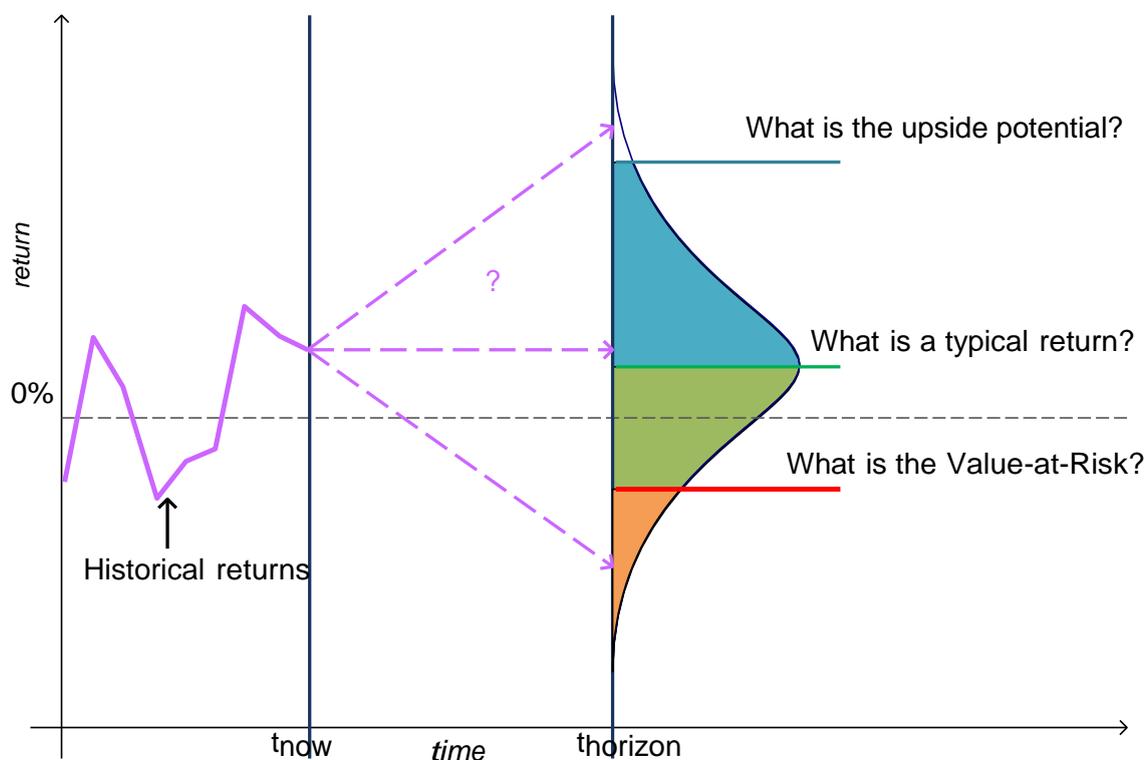

Figure 1 Conditional quantile estimation of asset returns

The problem

An important problem faced in our long-term strategy work is to deal with estimating the broad future outcomes of time series, with a slightly greater emphasis on extremes. In this instance, the preference is to focus on estimating conditional quantiles of the series, given other historical data, rather than estimating a generative parametric model for the data.

We have available data consisting of monthly and quarterly observations of around 50 continuous economic and financial variables covering the period since 1970. We also have available a further time series over the same period to be treated as a response. Our goal is to estimate, for each time point, the conditional quantiles of the response series given historical observations of the remaining series.

Our problem may differ from classical forecasting problems in the following aspects:

- We are interested in estimating conditional quantiles rather than conditional means.
- We have a large number (around 50) of covariate series we would like to condition on instead of conditioning on the history of the response series itself. Moreover, we have relatively few observations (around 600 for the monthly series) making our problem high- dimensional.
- We are interested in exploring methods that do not make strong parametric assumptions on the time series.

Questions

Q1: Given the nature of the problem described above, which methods are particularly suitable to attack this problem?

Q2: What would be a good validation framework allowing us to compare performance of various conditional quantile estimation methods (existing versus new methods)?

Q3: How robust / stable are the proposed methods in terms of their sensitivity to assumptions, and can one provide guarantees for the estimated conditional quantiles in terms of confidence intervals?

Q4: So far, the problem has focused on the conditional distribution of a single response variable. How could one extend the methods to cater for joint conditional distribution estimation when we have a vector of time series?

The available dataset

Whilst some of the economic time series were designed with stationarity in mind, other time series will have to be transformed to be approximately stationary.

An important aspect to consider is avoiding look-ahead bias: when estimating the conditional quantile of interest, only sufficiently lagged observations can be used (for instance, end-of-quarter data are typically released few months later). Look-ahead bias must be also carefully reflected in cross-validation (CV) set construction/generalisation error estimation.

Note: datasets are provided to the working group under a Confidentiality Agreement signed on behalf of the University of Cambridge.