Applications of Information Theory to Econometrics

Abstract

The purpose of this research is to apply concepts from information theory to study economic and financial data. The availability of ultra-high frequency financial data (tick-by-tick data) has provided more in-depth information. However, researchers also faced increased levels of market micro structure noise. The question arises, how does one differentiate information from noise with minimum assumptions on the data generating process? It is the aim of this study to develop appropriate econometric techniques to answer this question.

One of the fundamental concepts in information theory is entropy which can be viewed as a measure of uncertainty associated with a random variable. This study aims to examine the behaviour of stock returns by modelling their probability distributions using Maximum Entropy Density (MED). Entropy is a functional of probability distributions which summarizes the characteristics of the random variables under investigation (stock returns). The estimated MED maximizes the entropy over all probability distributions given a set of constraints. These constraints contain the empirical characteristics of the return data and as such the resulting return distribution along with its parameters can be estimated directly from the data itself using Maximum Likelihood Estimation (MLE) or the Hellinger distance estimator.

This study provides two useful applications of MED for stock returns namely, testing for intra-daily seasonality and Value-at-Risk (VaR) forecasts. Intra-daily seasonality refers to cyclical patterns that are present in return data through the hours (time slots) of a given trading day. Given the abundance of stock data, one can estimate a MED for a given time slot and compare it with resulting densities from other time slots. Any significant differences in the parameters of the resulting distributions indicate that the return distribution changes between the time slots. This pattern repeated over a period of time corresponds to intra-daily cyclicality.

From a risk management perspective, the computed MED can also be used to estimate tail probabilities which provides a novel way to estimate Value-at-Risk (VaR) as well as Conditional VaR (CVaR). It also provides an interesting alternate for density forecast which contains richer information than a point or interval forecast. Furthermore, the asymptotic properties of the maximum likelihood estimator for time varying MED will be examined in detail.

In addition to this, the study examines the behaviour of long memory in financial time series which is often captured by fractionally integrated process. This would imply that the level of dependence decays at a much slower rate compared to an autoregressive moving average (ARMA) process. Another novel contribution of this study is to derive n-step ahead forecasts for fractionally integrated financial data and investigates the behaviour of n-step ahead forecasts as $n \to \infty$.

Finally, this study aims to investigate the relationship between long memory and entropy as well as CVaR. The connection between the two concepts has not been examined in details despite their relative importance.