

A systematic approach to the linkage of hedge fund data sources

Università della Svizzera Italiana

MSc Digital Finance – Data Analytics for Finance

Matteo Bottacini – matteo.bottacini@usi.ch

October 2, 2020

The problem and its relevance

The purpose is to provide a systematic approach for synthesizing information from multiple heterogeneous hedge fund data sources.

This synthesis is necessitated by two cardinal features of hedge fund data: (i) the lack of a single comprehensive data source and (ii) the manifestation of identical asset-allocation algorithms along different share classes.

The new methodology and its contribution

The methodology adopted consists of the following steps:

- i. *Data pre-processing*: extraction, cleaning, standardization, bursting and enhancement;
- ii. *Indexing*: blocking to reduce the complexity splitting the population in smaller sets;
- iii. *Linkage method*: Fellegi-Sunter method to decide whether two reference are equivalent or not;
- iv. *Matching results*: evaluation of the linking quality (i) following what is discuss in “Quality and complexity measures for data linkage and deduplication” by Peter Christen and Karl Goiser and (ii) through a Monte Carlo simulation.

Its contribution has been confirmed by the results in “The Effect of Investment Constraints on Hedge Fund Investor Returns” by Juha Joenväärä, Robert L. Kosowski, P. Tolonen.

The data set

Six different sources: BarclayHedge, EurekaHedge, HedgeFund Intelligence, HedgeFundNet, Morningstar and Tremont Advisory Shareholders Services (TASS).

Morningstar database is set as the “golden source” benchmark testing sample due to its data quality. Each database can potentially be affected by several types of biases.

Conclusion – Summary of the results

Results indicate that the proposed matching algorithm produces a proper matching performance.

The precision of the process can be further increased, for instance either by applying an additional clerical review on large matching groups or by further tightening the conditions for blocking.

Further empirical research can benefit from adopting this data synthesis process for hedge funds.