

Big data-based inflation calculation for Hungary

Topic 1 – Bringing in information from where we can get it

Keywords: Big data, inflation, data comparison, pricing techniques

Introduction

Since official inflation (or CPI) calculation algorithms have not significantly changed in the last 30-40 years, many research have been done recently to find alternative ways of calculation. These methods focus partly on precision of inflation data, and partly on time period of processing, getting closer to real-time calculations.

Our approach and goal is defining alternative inflation calculation methods for Hungary, which also consider specific attributes of the local economic conditions and legal regulations. We determine those 21th century challenges that statisticians have to face, e.g., difficulties of efficient data collection and consideration of latest pricing practice of retailers, which include dynamic pricing and zillions of different marketing campaigns to provide indirect price decrease.

Currently, our approach is in concept development stage. Implementation is expected after pilots, based on the experience of tests and refined algorithms.

Methods / Problem statement

Regarding raw data collection for inflation calculations, we apply a completely new method to current manual practices: automatic bots will browse the internet to identify valid price information. Our method applies big data algorithms. This means that even millions of data can be processed daily. However, in this type of data gathering we face new challenges that human manual work could easily avoid: redundancy and unreliability of prices, outdated offers, imprecise description of products and ambiguity.

In addition, we present the pricing practice of retailers. We plan to analyze data of price comparison sites. These sites provide price information real time, so price changes can be identified easily. Here the challenges include product categorization, appropriate selection of consumer basket.

Results / Proposed solution

Our research concentrates on three tasks:

1. Billion Prices Project adaptation to Hungary: The most well-known approach is the Billion Prices Project (BPP) by MIT to gather daily prices and calculate practically real-time inflation. We analyze all aspects that must be considered to adaptation of these algorithms to the local environment.
2. The current practice of price data collection is not efficient: data recording is completely manual: agents record prices in stores all around the country and make aggregations using weighting that were defined one or two years before. We suggest automated price gathering: software bots constantly browse the internet for prices. Filtering algorithms must be applied to screen the useful price data. Based on this approach, millions of price data can be gathered every day, this is why we face a big data problem.
3. Considering real price data: list prices (Suggested/Recommended Retail Prices) are always higher than street price. In many areas (tourism, uber, etc.) dynamic prices are part of everyday life. Different types of

e-commerce pricing cannot be caught by traditional price collection. Weekly or monthly data recording of the old school is inaccurate. We must concentrate on “exit” prices, the actual “buyer price” of a product instead of street prices or advertised price offers.

Conclusions

Based on the first steps of our research we concluded the following:

- We suggest big data approach for price collection of inflation calculation. Millions of price data can be gathered daily efficiently. Screening and analysis of these data can only be done using big data calculation algorithms to provide frequent (daily or every two/three days) inflation figures.

- We propose new data cleaning and validation algorithm that can be considered as an independent check (“second opinion”) of the calculations. This part of the process becomes a completely new step and the most crucial in automated price data gathering.

- New algorithms are needed to tackle the modern pricing practices. Dynamic and individual pricing, coupons, sale campaigns of many types can only be measured if we concentrate on “exit” prices, that is the actual cost, and not the offers.