Using New Methods and Data Sources to Improve Economic Statistics

David Curran
UK Office of National Statistics
david.curran@ext.ons.gov.uk

Tony Zemaitis
The Turing Institute

We would be delighted to showcase the collaboration and the three projects which explore how to use new methods and data sources to improve economic statistics in practice.

There are three distinct projects within the scope of a collaboration that began in 2021 between the Economic, Social and Environment Group (ESEG) within ONS, the Data Science Campus (DSC) within ONS and The Alan Turing Institute which is the United Kingdom's national institute for data science and artificial intelligence (https://www.turing.ac.uk/news/ons-and-turing-join-forces-produce-better-and-faster-estimates-economic-changes).

The collaboration directly supports the UK Statistics Authority's new five-year strategy of "Statistics for the public good: Informing the UK. Improving lives. Building the future.", by aiming to deliver key information faster, at a more granular level, and through more efficient and sophisticated data modelling.

Further, the projects closely align with the strategic objectives of other key government department stakeholders.

The first project is named Understanding Economic Networks, and aims to build a greater understanding of the economic linkages between industries and regions through a combination of data processing for firm-to-firm transactions data held by ONS, economic modelling and using data from the whole economic system, and the development of methods for efficiently reducing the dimensionality and complexity of network-generated data.

Initial stages of the Economic Networks project included the acquisition of high frequency UK firm-to-firm payments and the secondment of ONS analysts to Financial Institutions to recreate a monetary flows database based on payment systems data. Additional work has also developed validation exercises on cuts of aggregated payment systems data to be compared with sector level series at ONS as well as designing linkage algorithms to match firm level data to Companies House and other business registers.

The second project is named Economic Nowcasting, which seeks to build methods to allow rapid assimilation and combination of disparate data to produce estimates of key economic variables, such as household expenditure, household income and consumer prices. The aim is to produce quicker aggregate modelled estimates about these variables and generate estimates at a more detailed local level. As the quantity of data available for ONS to process increases, there needs to be comparable development in the methods of exploiting this data. A key concern is that, while more data is available, it is often not sampled in a convenient manner, and the incorporation of multiple streams of data is needed to provide accurate understanding.

In the initial stage of the project, we've explored the use of the signature methodology as an alternative to traditional time-series approaches to producing economic nowcasts, with a focus on household income. We compare these with a range of comparable benchmark nowcasting models used in the existing economic literature. We've also considered a range of indicator variables at different levels of granularity, with the aim of improved timeliness. Our approaches address the mixed frequency data issue that is a well-known problem in most empirical nowcasting exercises. The third project is named Synthetic Data and Privacy Preservation, which aims to develop tools and examples to allow sharing of private datasets with a wider range of stakeholders, while preserving privacy. The intention is to use synthetic data generators (SDGs), which offer a naturally private way to generate high-quality data that preserves the statistical features of the original data set. Applying this methodology to sensitive data will allow greater future flexibility for collaboration with researchers in the wider community and government.

The team has evaluated a range of methods for synthetic data generation and benchmarked them by generating synthetic data from the UK Census Microdata Teaching File. This has been achieved by using the synthetic data evaluation framework, SynthGauge, developed to facilitate a rapid-cycle evaluation of the data utility preservation in the synthetic data.

Chaired by: TBC (ONS)

1. Economic Networks and Transaction Data – Presented by TBC (ONS) / TBC (Turing)

2. Economic Nowcasting - Presented by TBC (ONS) / TBC (ONS) / TBC (Turing)

3. Synthetic Data and Privacy Preservation – Presented by TBC (ONS) / TBC (ONS) / TBC (Turing)