

*17th Ottawa group meeting
Rome, 7 – 10 June 2022*

Title: Modernising the measurement of clothing price indices using web-scraped data: classification and product grouping

Authors: Liam Greenhough, Hazel Martindale, Helen Sands

Abstract:

An increase in purchases from online-only clothing retailers has led to a decline in market coverage in the UK CPI. We have therefore been exploring the use of web-scraped data for clothing. While web-scraping clothing data allows us to increase coverage of the market, it brings about new challenges with regards to classification and representivity of our clothing indices. Web-scraped data alone do not resolve old challenges regarding the extremely dynamic nature of clothing markets due to constantly changing fashion and seasonal trends.

In this presentation we show how using data science methods (machine learning and natural language processing) to classify and group together similar products can reduce a downward bias in indices caused by the extremely dynamic nature of clothing markets.

REFERENCES

ONS, [Product grouping: measuring inflation in dynamic clothing markets](#), 2021

ONS, [Classification of new data in UK consumer price statistics](#), 2021

ONS, [Automated classification of web-scraped clothing data in consumer price statistics](#), 2020