# <u>Abstract</u>: The use of approximate expenditure weights for web scraped data in consumer price indices

Alternative data sources, such as web scraped data and point of sale scanner data, are becoming more commonly available and have potential to improve consumer price indices through more frequent data collection, increased coverage and larger sample sizes. However, one of the main drawbacks that have been identified with web scraped data is the lack of data to derive expenditure weights for individual price quotes.

As part of research into the use of these data, the United Kingdom's Office for National Statistics (ONS) has conducted some analysis to investigate the potential impact of this issue. Since all prices are scraped, calculating an unweighted index at the lowest level of aggregation means that more popular items may not have enough influence on the index. Unweighted indices therefore may not be representative of consumer spending if the prices of less popular products behave differently to those of more popular products.

This paper is an investigation into alternative weight allocation methods that could approximate expenditure weights for the web scraped products used in the calculation of an item's index. In the absence of product-level expenditure or quantity information, we have investigated various methods, including deriving weights from the product position on a web page or the market share. Geometric Laspeyres indices are calculated from these derived weights, such that their behaviour can be compared to the unweighted Jevons index calculated from the same data source. Further analysis looks at how the difference in market behaviour between items could also be considered; different distributional assumptions may be appropriate for those items where the market distribution indicates the existence of a market leader, compared with markets that exhibit signs of perfect competition.