Clustering of common goods and commodities based on the time-series characteristics of their wholesale price index

Abstract

This case study provides a basis to obtain homogeneous clusters of common goods and commodities based on the time-series characteristics of their wholesale price index to obtain insights regarding the underlying similarities between them. These similarities enabled us to identify trends associated with the groups of commodities. The time-series data used in this case study was taken from the **Wholesale Price Index** catalogue of the **"data.gov.in"** website. We took monthly data from 2011 to 2020 with 2011-12 as the base year for calculating WPI. Before clustering, a kernel estimator was used to remove noise from the raw data; then, hierarchical clustering was performed over the smooth version of the original data. The obtained clusters contained items that appeared to be non-homogeneous. Hence, we performed ARIMA modeling over the noise that was previously removed by the kernel estimator. From this procedure, we obtained residuals and the squared sum of these residuals was calculated, over which hierarchical clustering was performed to obtain subclusters. The result of the analysis provided homogeneous clusters of the common goods and commodities based on the shape of the associated time-series data.