Comparing penalized dynamic regression with time series methods for modeling and forecasting retail product sales

Patrícia Ramos, INESC TEC and School of Accounting and Administration, Polytechnic Institute of Porto, Portugal, pramos@inesctec.pt
José Manuel Oliveira, INESC TEC and Faculty of Economics, University of Porto, Portugal, jmo@fep.up.pt
Nikolaos Kourentzes, Skovde Artificial Intelligence Lab, Skovde University, Sweden, nikolaos@kourentzes.com
Robert Fildes, Centre for Marketing Analytics and Forecasting, Lancaster University Management School, UK, r.fildes@lancaster.ac.uk

Retailers depend strongly on accurate sales forecasts to manage their supply chains and make decisions concerning purchasing, logistics, marketing, finance, human resources, etc. Inaccurate forecasts of product sales can lead to stock-outs which are indubitably very negative for the business. If the product is not available in shelf its potential sales are lost and there is the chance of customers to look for it in the competitors, making loyalty difficult to maintain. Ordering excess inventory, to reduce the risk of stock-outs and to improve customer's satisfaction, increases costs significantly (e.g. labor and storage) reducing the profit margin. Additionally, there is an increased awareness that food waste should be reduced. In 2012 the European Parliament called for urgent measures to halve food waste by 2025 and to improve access to food for needy EU citizens. Efficient inventory management can be achieved with accurate forecasts of SKU (Stock-Keeping Unit) sales at the store level, which enable the retailer to replenish in time and meet the customers' expectations. However, how to develop forecasting models that consider the main drivers that affect demand still constitutes a major challenge.

In this work we propose efficient and effective models to forecast retailer product sales at the SKU level. Working with a large dataset from a leading Portuguese retailer, we develop and test three different modeling approaches to forecast the demand during a promotion including routine against non-routine product categories. Integrated models that perform joint estimation of autoregressive components and promotional information outperform hybrid models that carry out this estimation on a two-step approach. On the other hand, this two-step approach performs best than univariate models which reveals the critical importance of external information. Models tested include ARIMA, TBATS, LASSO and Exponential Smoothing.