Advances in Manufacturing Technology XXX

The urgent need to keep pace with the accelerating globalization of manufacturing in the 21st century has produced rapid advancements in manufacturing technology, research, and expertise.

This book presents the proceedings of the 14th International Conference on Manufacturing Research (ICMR 2016), entitled Advances in Manufacturing Technology XXX. The conference also incorporated the 31st National Conference on Manufacturing Research, and was held at Loughborough University, Loughborough, UK, in September 2016. The ICMR conference is renowned as a friendly and inclusive environment which brings together a broad community of researchers who share the common goal of developing and managing the technologies and operations key to sustaining the success of manufacturing businesses.

The proceedings is divided into 14 sections, including: Manufacturing Processes; Additive Manufacturing; Manufacturing Materials; Advanced Manufacturing Technology; Product Design and Development, as well as many other aspects of manufacturing management and innovation. It contains 92 papers, which represents an acceptance rate of 75%.

With its comprehensive overview of current developments, this book will be of interest to all those involved in manufacturing today.
Advances in Transdisciplinary Engineering

Advances in Transdisciplinary Engineering (ATDE) is a peer-reviewed book series covering the developments in the key application areas of product quality, production efficiency and overall customer satisfaction. ATDE will focus on theoretical, experimental and case history-based research, and its application in engineering practice. The series will include proceedings and edited volumes of interest to researchers in academia, as well as professional engineers working in industry.

Editor-in-Chief
Josip Stjepandić, PROSTEP AG, Darmstadt, Germany

Co-Editor-in-Chief
Richard Curran, TU Delft, The Netherlands

Advisory Board
Cees Bil, RMIT University, Australia
Milton Borsato, Federal University of Technology, Paraná-Curitiba, Brazil
Shuo-Yan Chou, ‘Taiwan Tech, Taiwan, China
Fredrik Ehl, Linköping University
Parisa Ghodous, University of Lyon, France
Kazuo Hikata, University of Tokyo, Japan
John Mo, RMIT University, Australia
Essam Shahab, Cranfield University, UK
Mike Sobolevski, TTU, Texas, USA
Amy Trappey, NTUT, Taiwan, China
Wim J.C. Verhagen, TU Delft, The Netherlands
Wensheng Xu, Beijing, Huaqiang University, China

Volume 3
Recently published in this series:


ISSN 2352-751X (print)
ISSN 2352-7528 (online)
Contents

Preface and Acknowledgements
Yee Mey Goh and Keith Case v
Organising Committee ix

Part 1. Keynote
The Hidden Information and Its Potential in Manufacturing Data (abstract)
Tom Neat and Felix Ng 3

Part 2. Manufacturing Processes
Toolpath Geometry and High Speed Machine Axis Motion
L. Chanda and R.J. Cripps 7
Numerical Modelling and Simulation of Temperature Distribution in End Milling
Sunday Joshua Ojolo, Ibukunoluwa Aina and Gbeminiyi Musibau Sobamowo 13
Investigating the Sensitivities of Tool Wear When Applied to Monitoring Various Technologies for Micro Milling and Drilling
James M. Griffin and Hala Abdelaziz 19
Vibro-Impact Machining – An Application in Surface Grinding
Andre D.L. Batako, Heisum Ewad, Grzegorz Bechciniski and Witold Pawlowski 25
Structural Transformations of Polymers Using Laser Irradiation
Jimshner N. Aneli, Nana V. Bakradze, Teimuraz N. Dumbadze and Andre D.L. Batako 31
Finite Element Analysis of Single Point Incremental Forming
Spyridon Kotsis and Jonathan Corney 37
Injection Moulding – Properties Customisation by Varying Process Conditions
Tino Meyer, Paul Sherratt, Andy Harland, Barry Havorth, Chris Holmes and Tim Lucas 43
Simulation of Material Removal in Mould Polishing for Polymer Optic Replication
Rui Almeida, Rainer Börret, Wolfgang Rimkus, David K. Harrison and Anjali De Silva 49
Fabrication of Micro Components with MSZ Material Using Electrical-Field Activated Powder Sintering Technology
Hasan Hijji, Yi Qin, Kunlan Huang, Song Yang, Muhammad Bin Zulkifli and Jie Zhao 55
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Study of Linear Heat Transfer Gradients on 3D Printed Specimens</td>
<td>137</td>
</tr>
<tr>
<td>Robert Benham and Foyaz Rehman</td>
<td></td>
</tr>
<tr>
<td>Analysis and Impacts of Chips Formation on Hole Quality During</td>
<td>143</td>
</tr>
<tr>
<td>Fibre-Reinforced Plastic Composites Machining</td>
<td></td>
</tr>
<tr>
<td>Sihri Olorunisim Ismail, Hom Nath Dhakal, Ivan Popov</td>
<td></td>
</tr>
<tr>
<td>and Johnny Beangrand</td>
<td></td>
</tr>
<tr>
<td>Wear Behaviour of Laser Cladded Ni-Based WC Composite Coating</td>
<td>149</td>
</tr>
<tr>
<td>for Inconel Hot Extrusion: Practical Challenges and Effectiveness</td>
<td></td>
</tr>
<tr>
<td>Lynne McIntosh, Javad Falsafi, Stephen Fitzpatrick and Paul L. Blackwell</td>
<td></td>
</tr>
<tr>
<td>Effect of Powder Metallurgy Parameters on the Performance of EDM</td>
<td>155</td>
</tr>
<tr>
<td>Tool Electrodes</td>
<td></td>
</tr>
<tr>
<td>Amoljit Singh Gill and Sanjeev Kumar</td>
<td></td>
</tr>
<tr>
<td>A Novel Adaptation of the T-Peel Bimetal Bond Test Based on the Thin</td>
<td>161</td>
</tr>
<tr>
<td>Film Bonding Theory Using Cold Roll Bonded Al/Sn/Steel Bimetal Laminates</td>
<td></td>
</tr>
<tr>
<td>Laurie Da Silva, Mahmoud El-Sharif, Colin Chisholm and Stuart Laidlaw</td>
<td></td>
</tr>
<tr>
<td>Part 5. Advanced Manufacturing Technology</td>
<td></td>
</tr>
<tr>
<td>A Path Planning Algorithm for a Materials Handling Gantry Robot and</td>
<td>169</td>
</tr>
<tr>
<td>its Validation by Virtual Commissioning</td>
<td></td>
</tr>
<tr>
<td>Ruth Fleisch, Robert Schöch, Thorsten Prante and Reinhard Pfafferkorn</td>
<td></td>
</tr>
<tr>
<td>Integrating Optimisation with Simulation for Flexible Manufacturing</td>
<td>175</td>
</tr>
<tr>
<td>System Booyang Song, Winda Hutabarat, Ashutosh Tiwari and Shane Enticott</td>
<td></td>
</tr>
<tr>
<td>Automatic Generation of a Detailed Outfitting Planning and Determining</td>
<td>181</td>
</tr>
<tr>
<td>the Effect of Multi-Skilled Mounting Teams</td>
<td></td>
</tr>
<tr>
<td>Christopher Rose and Jenny Conen</td>
<td></td>
</tr>
<tr>
<td>Development of PID Controller for Vibratory Milling</td>
<td>187</td>
</tr>
<tr>
<td>Wael Najj Alharbi, Andre Batako and Barry Gemm</td>
<td></td>
</tr>
<tr>
<td>An Application of Autoregressive Hidden Markov Models for Identifying</td>
<td>193</td>
</tr>
<tr>
<td>Machine Operations</td>
<td></td>
</tr>
<tr>
<td>Dimitrios Pantazis, Adrian Ayastuy Rodriguez, Paul P. Conway</td>
<td></td>
</tr>
<tr>
<td>and Andrew A. West</td>
<td></td>
</tr>
<tr>
<td>Improving Data Accuracy in Simulation of Flexible Manufacturing</td>
<td>199</td>
</tr>
<tr>
<td>Systems Jastyna Rybiwicka, Ashutosh Tiwari and Shane Enticott</td>
<td></td>
</tr>
<tr>
<td>AFRC's Image Processing Platform: A High Speed User-Friendly</td>
<td>203</td>
</tr>
<tr>
<td>Architecture for Real Time Object Detection in Forging Processes</td>
<td></td>
</tr>
<tr>
<td>Danial Kahani and Remi Zante</td>
<td></td>
</tr>
<tr>
<td>A Multi Degree of Freedom Actuation System for Robot and Machine</td>
<td>211</td>
</tr>
<tr>
<td>Vision Industrial Applications</td>
<td></td>
</tr>
<tr>
<td>Mfortaw Elvis Asiu, Mahmoud Shafik and Ilia Oruifge</td>
<td></td>
</tr>
</tbody>
</table>
Part 6. Human Aspects of Manufacturing

Deployment of Assisted Living Technology Using Intelligent Body Sensors
Riyad Etsaadi and Mahmoud Shafik
Platform for Elderly People Health Monitoring
Ikeoma Chinazuekperi Iheji and Mahmoud Shafik

Motivation and Learning in Manufacturing Industries
Shafiei Mai, Keith Case, Shahril Mohamad and Yee Mey Goh

Cognitive Aspects Affecting Human Performance in Manual Assembly
Anna Brolin, Keith Case and Peter Thorvald

Ergonomic Risk Assessment – A Case Study of a Garment Manufacturing Industry
Amjad Hussain, Iqra Javed, Keith Case, Ashfaq Ahmad and Nadeem Safdar

Distributed Cognition in Manufacturing: Collaborative Assembly Work
Rebecca Androasson, Jessica Lindblom and Peter Thorvald

Agile Assembly Planning for Multi-Variant Production Based on 3D PDF
Felix Kuhl, Stefan Rathoff, Josip Sijepandic and Klaus Thaterkorst

The Impacts of Ageing on Manufacturing Sectors
Renuga Nagarajan and Patricia Ramos

Part 7. Product Design and Development

Design Optimisation of Passive Humidification Device for Intensive Care Medical Applications
Mahmod Shafik and Anne Lechevrel

Evaluation of 3D Renewable Micro Power Station for Smart Grid Applications
Moglo Komlavi, Mahmoud Shafik and Mfortaw Elvis Ashu

Design & Development of a Bicycle Cleat Adapter
Fayyaz Rehman and Santiago Diaz Montalvo

3D Alignment for Interactive Evolutionary Design
Theodora Retepsi, Ian J. Graham and Yee Mey Goh

Design and Manufacturing of a Miniature Insulin Administration Device for Non Compliant Diabetic Patients of Kingdom of Saudi Arabia (KSA)
Irfan Anjum Manarvi and Nader Maher Kamel Matta

Identifying Sequence Maps or Locus to Represent the Genetic Structure or Genome Standard of Styling DNA in Automotive Design
Shahrizan Zainal Abidin, Azlan Othman, Zafuddin Shamsuddin, Zaidi Samudin, Halim Hassan and Wan Asri Wan Mohamed

Design and Manufacturing of a Fire Protection Suit Through Reverse Engineering
Irfan Anjum Manarvi and Haris Rehman

Part 8. Digital Manufacturing (Industrie 4.0)

An Intelligent Anti-Collision System for Electric Vehicles Applications
Ikeoma Chinazuekperi Iheji and Mahmoud Shafik

A Case Study Analysis of the Application of Design for Manufacture Principles by Industrial Design Students
Russell Marshall and Tom Page

Industrie 4.0 Implementations in the Automotive Industry
Diana Segura-Velandia, Aaron Neal, Paul Goodall, Paul Conway and Andrew West

Component Detection with an On-Board UHF RFID Reader for Industrie 4.0
Aaron Neal, Diana Segura-Velandia, Paul Conway and Andrew West

Capability of Returnable Transit Items
Paul Goodall, Aaron Neal, Diana Segura-Velandia, Paul Conway and Andrew West

Factory Automation and Information Technology Convergence in Complex Manufacturing
Ip-Shing Fan and Leon Osuin

Optimising Mixed Model Assembly Lines for Mass Customisation: A Multi Agent Systems Approach
Olaimide Bajji, David Stewart and Maria Fasli

A Data Management System for Identifying the Traceability of Returnable Transit Items Using Radio Frequency Identification Portals
Paul Goodall, Aaron Neal, Diana Segura-Velandia, Paul Conway and Andrew West

Product Lifecycle Management Enabled by Industrie 4.0 Technology
Filipe Ferreira, José Faria, Américo Azevedo and Ana Luisa Marques

Part 9. Sustainable Manufacturing

Development of a Simulation-Based Approach to Smart Management of Energy Consumption in Buildings and Their Implementation
A.N.R. Pour and K. Cheng

A Decision Support Tool for Improving Value Chain Resilience to Critical Materials in Manufacturing
Liam Gardner and James Colwill

Lean & Green, How to Encourage Industries to Establish Pro-Environmental Behaviour
Elebista Pavlik, Dagmara Gutowska, Remigiusz Horbal, Zofia Matiuzszczak, Robert Miehe, Ivan Bogdanov and Ralph Schneider

A Manufacturing Approach to Reducing Consumer Food Waste
Aicha Jellali, Elliot Woolley, Guillermo Garcia-Garcia and Shakin Rahimifard
Part 10. Information and Knowledge Management

An Investigation into the Management of Design for Manufacturing Knowledge in an Aerospace Company
Mohammed El Souri, James Gao, Oladele Owodunni, Clive Simmons' and Nick Martin

Networked Engineering Notebooks for Smart Manufacturing
Peter Denno, Charles Dickerson and Jennifer Harding

Investigation into the Design and Development of Knowledge-Based Innovators: Processes in Manufacturing Companies
Lakhvir Singh, Reza Ziarati, Martin Ziarati, Richard Gatward and Feng Chen

Intelligent Semantic Query for Manufacturing Supply Chain
Salmen Saadlou, Mozaffar Saadat and Ebrahim Amini Sharifi

Part 11. Organisation Management

Organizational Learning Capability: An Empirical Assessment of Innovative Supply Chain Development
Andrew Thomas, Peter Dorrington, Filipa Costa and Gareth Loudon

Manufacturing Supply Chain Demand Study – Profiling UK Manufacturing Performance
Andrew Thomas, Hefin Rowlands, Peter Dorrington, Filipa Costa, Mark Francis and Ron Fisher

An Investigation into Re-Shoring Decision: Case Study Approach
Hamid Moradiou and Chris Backhouse

Conceptual Understanding, Design and Management of Supply Chain in 21st Century – A Review
Sameh M. Saad and Adewale A. Ogunsanwo

A Review of Resilience Within the UK Food Manufacturing Sector
James Colwill, Stella Despouli and Ram Bhakma

Responsiveness Optimisation in the Fractal Supply Network
Sameh M. Saad and Ramin Bahadori

Part 12. Cost Engineering and Forecasting

Process Selection Using Variation and Cost Relations
Vincent McKenna, Yan Jin, Adrian Murphy, Michael Morgan, Caroline McClory, Colm Higgins and Rory Collins

Cost Modelling for Aircraft in a Multi-Disciplinary Design Context
Davide Di Pasquale, David Gore, Mark Savill, Timoleon Kipouros and Carren Holden

Management of Promotional Activity Supported by Forecasts Based on Assorted Information
Cátia Ribeiro, José Manuel Oliveira and Patricia Ramos

Sales Forecasting in Retail Industry Based on Dynamic Regression Models
José Manuel Pinho, José Manuel Oliveira and Patricia Ramos

Evaluating the Forecasting Accuracy of Pure Time Series Models on Retail Data
Patricia Ramos, José Manuel Oliveira and Rui Rebelo

Part 13. Lean and Quality Management

Development of Lean Six-Sigma Conceptual Implementation Model for Manufacturing Organisations
Sameh M. Saad and Mohamed Khumhak

Do Inspections Really Help?
Moshe Eben Chaime

The Quality Journey of Greek SMEs
G. Sains, G. Haritos, Th. Kriemadis and M. Fowler

Creating Industrially Relevant Environments for Teaching Lean Manufacturing at Karlstad University
Lasse Jacobsson, Anders Wickberg and Leo De Vin

Part 14. Decision Support and Optimisation

Critical Success Factors for In-House Production, Partial Production or Outsourcing in Garment Industry
C.S. Libânio, Fernando Gonçalves Amaral and Sérgio Almeida Migowski

Fuzzy Analytic Hierarchy Process for the Selection of Maintenance Policies Within Petroleum Industry
Abdel M. Mohamed and Sameh M. Saad

Application of a Multivariate Process Control Technique for Set-Up Dominated Low Volume Operations
Steven Cox, Scott Anderson, Neil Gray, Oliver Vogt and Apostolos Kotsiolas
Part 1

Keynote
When a comparison between a reference aircraft and a concept one is desired to understand if there is any benefit in terms of economic profitability, on the second run, using concept aircraft input, the ticket price is kept constant and the IRR left varying.

To estimate cost, the geometric features such as the fabrication areas of skin, spars and ribs, and the assembling perimeter of the wing are needed, as well as, mission and mass information. The extraction of these parameters has been implemented in an automatic fashion using again the Model Center integration/process building environment. The bottom left of Figure 2 shows the Model Center process for the cost suite input generation. The first component generates a password string to access the data management tool. The second component consists of a python script developed to extract those geometric features from the 3D CAD model of the wing structure to enhance the product structure and improve the validity of the cost evaluation. The third component is used to extract mass data to enable cost to reflect the actual design. The fourth one extracts info on the manufacturing complexity including process type, process time and raw material weight. The last one generates the input file for the cost suite. To summarise, the model is capable of calculating the Direct Operating Cost and provides the output in both net present value and the absolute cost. Different trade studies have been performed and cost results produced for different real case configurations.

References


Management of Promotional Activity Supported by Forecasts Based on Assorted Information

Cátiia RIBEIROa, José Manuel OLIVEIRAa,b and Patricia RAMOSa,b,c,

Abstract. Aggressive marketing causes rapid changes in consumer behavior and some significant impact in the retail business. In this context, sales forecasting at the SKU level can help retailers to become more competitive by reducing inventory investment and distribution costs. Sales forecasts are often obtained combining basic univariate forecasting models with empirical judgment. However, more effective forecasting methods can be obtained by incorporating promotional information, including price, percentage of discount (direct discount or loyalty card discount), calendar events and weekend indicators not only from the focal product but also from its competitors. To deal with the high dimensionality of the variable space, we propose a two-stage LASSO regression to select optimal predictors and estimate the model parameters. At the first stage, only focal SKUs promotional explanatory variables are included in the Autoregressive Distributed Lag model. At the second stage, the in-sample forecast errors from the first stage are regressed on the explanatory variables from the other SKUs in the same category with the focal SKU, and to use that information more effectively three different approaches were considered: select the five top sales SKUs, include all raw promotional information, and preprocess raw information using Principal Component Analysis. The empirical results obtained using daily data from a Portuguese retailer show that the inclusion of promotional information from SKUs in the same category may improve the forecast accuracy and that better overall forecasting results may be obtained if the best model for each SKU is selected.

Keywords. Business analytics, demand forecasting, competitive information, autoregressive distributed lag models, LASSO, principal component analysis.

1. Introduction

The new paradigm of mass customization is forcing manufacturers to redesign and change products constantly [1]. As a consequence, products life cycles have been decreasing making sales at the SKU (Stock Keeping Unit) level in a particular store difficult to forecast, as times series for this products tend to be short. Moreover, retailers are increasing marketing activities such as price reductions and promotions due to more intense competition and recent economic recession [2]. Products are typically on promotion for a limited period of time, e.g. one week, during which
demand is usually substantially higher occurring many stock-outs due to inaccurate forecasts. Stock-outs can be very negative to the business because they lead to dissatisfied customers [3]. How to balance the loss due to stock-outs and the cost of safety stocks is clearly an important issue for today’s retailers. It is well accepted that SKU level retail store sales are affected not only by own promotional information but also by the influence of products of the same category [4, 5]. When building a product sales forecasting model, an obvious obstacle to use all this information is the high dimensionality of the variable space that may turn the model easily over-fitted or impossible to estimate [6]. Traditional methods used to solve problems of dimensionality reduction include the stepwise selection, the Least Absolute Shrinkage and Selection Operator (LASSO) and the information summary approach. The stepwise selection method has been criticized for being likely to select some unimportant predictors and miss important ones. LASSO has received great attention from the scientific community due to its convexity and sparsity solutions. It estimates the regression model in the usual way with an additional constraint for the sum of the absolute values of the parameter coefficients. The information summary approach condenses the information of all variables into a small set of predictors at an acceptable cost of information loss. In the retail business it is common to promote similar products simultaneously (e.g. different SKUs under the same brand) which makes promotional variables of different products highly correlated to each other. A popular information summary technique usually used to solve problems of multi-collinearity is Principal Components Analysis (PCA). In the forecasting context, an important issue associated with PC is that the new predictors are estimated without taking into account the dependent variable. Thus, when only a few components are selected to represent all variable space, they might not have enough prediction ability to forecast the dependent variable because useful components were rejected. In this paper, we focus on developing an effective and automatic modelling approach to forecast sales at the SKU level in a particular store that overcomes the issues discussed. The performance of the forecasting system is demonstrated through a case study of daily data of many hundreds of retail SKUs from a Portuguese retailer in order to produce short term forecast. The rest of the paper is organized as follows. In Section 2, we discuss the methodology involved in the work and present the models used. In Section 3, we describe the data, present the empirical study details and discuss the results obtained. Finally, in Section 4, we point the main conclusions of the work.

2. Methodology

As mentioned in the previous section, in the retail business it is common to promote products in the same category simultaneously. This results in the promotional variables from different SKUs of the same category being highly correlated with each other. When a regression model includes a group of explanatory variables with high pairwise correlations, LASSO tends to select only one of the variables in the group. If relevant predictors are highly correlated with irrelevant ones, LASSO may reject important variables by selecting an unimportant one. From previous studies [4, 6], it is known that an SKU’s own promotional variable is more relevant than that of other SKU’s. If we include simultaneously in the regression model promotional explanatory variables from the focal SKU and promotional explanatory variables from the SKUs in the same category with the focal SKU, LASSO may reject the focal SKU’s own predictors. To deal with this problem, we propose a two-stage LASSO regression to select optimal predictors and estimate the model parameters. At the first stage, only focal SKUs promotional explanatory variables are included in the Autoregressive Distributed Lag (ADL) model, namely sales lags, price and its lags, calendar events and weekend indicators, which is given by

\[
\ln(Y_{it}) = \beta_0 + \sum_{t=1}^{T} \delta_t \ln(Y_{i,t-1}) + \gamma_t \ln(P_{i,t-1}) + \phi_{j,t} \text{PDD}_{j,t} + \psi_{k,t} \text{PDC}_{j,t} + \theta_{t,E} + \mu_{t,LW} + \epsilon_{i,t}, \quad t = 1, \ldots, T
\]

where \(\ln(Y_{it}), \ln(P_{i,t-1}), \text{PDD}_{j,t}, \text{PDC}_{j,t}\) are respectively, the log sales, the log price, the percentage of direct discount and the percentage of discount in card of focal SKU \(i\) in category \(k\) in day \(t\); \(\epsilon_{i,t}\) is an indicator variable for a calendar event (New year’s Day, Carnival, Good Friday, Easter, Freedom Day, Labour Day, Corpus Christi, Portugal Day, Assumption Day, Republic Day, All Saints Day, Restoration of Independence, Immaculate Conception Day, Christmas Eve and Christmas Day); \(j\) if day \(t\) is a calendar event, 0 otherwise; \(LW\) is an indicator variable for the last weekend of the month; \(1\) if \(t\) is a day from the last weekend of the month, 0 otherwise; \(\nu_t, \delta_t, \gamma_t, \phi_t, \psi_t, \theta_t, \mu_{t,LW}\) are the model parameters; \(\epsilon_{i,t}\) is the disturbance term and \(L\) is the order of the lags included which was assumed to be 2. A 10-fold cross-validation was used to obtain the optimal value of the penalty parameter in LASSO regression. After variable selection and parameters estimation by LASSO regression, the estimated ADL model is used to compute the in-sample forecast errors \(\epsilon_{i,t}\) and to forecast on the out-of-sample period. At the second stage, the in-sample forecast errors from the first stage \(\epsilon_{i,t}\) are regressed on the explanatory variables from the other SKUs in the same category with the focal SKU, being the ADL model given by

\[
\epsilon_{i,t} = \beta_{0,t} + \sum_{j=1}^{J} \delta_{j,t} \ln(Y_{j,t}) + \gamma_{j,t} \ln(P_{j,t}) + \phi_{j,t} \text{PDD}_{j,t} + \psi_{j,t} \text{PDC}_{j,t} + \mu_{j,t, LW} + \epsilon_{i,t}
\]

where \(\ln(Y_{j,t}), \ln(P_{j,t}), \text{PDD}_{j,t}, \text{PDC}_{j,t}\) are respectively, the log sales, the log price, the percentage of direct discount and the percentage of discount in card of SKU \(j\) in category \(k\) in day \(t\); \(\delta_{j,t}, \gamma_{j,t}, \phi_{j,t}, \psi_{j,t}\) are the model parameters; \(\epsilon_{i,t}\) is the disturbance term. As in the first stage, variable selection and parameters estimation is done by LASSO regression. Then, the estimated ADL model is used to forecast on the out-of-sample period. The final out-of-sample forecasts are the sum of the forecasts in the two stages. In the second stage, to deal with the high dimensionality of the potential explanatory variables from the SKUs in the same category with the focal SKU, and to use that information more effectively, three different approaches were considered: (1) identify the five top sales SKUs and use them as representatives of the whole category; (2) include the information from all SKUS in the category directly into LASSO regression; (3) preprocess the information from all SKUs in the category to lower the dimensionality using PCA. A PCA was conducted for each type of variable (sales, price, percentage of discount and percentage of discount in card) across SKUs in the same category. So, in model (2), if the inputs are PC obtained by PCA, then the variables \(Y_{i,P}, \text{PDD} \text{ and } \text{PDC}\) are the corresponding PC, not necessarily extracted in the same number; if the inputs are from the five top sales SKUs then \(n_{i,p}\) is 5; if the
inputs are from all SKUs in the category, then \( n_c \) is the number of SKUs in category \( k \).

3. Empirical study

The data used in this study came from a Portuguese retailer leader on food distribution and product manufacturing. For the purpose of the research, one of the largest stores in sales volume of the company’s retail chain was selected. Only the departments of Food and Freshers of this store which have the greatest impact on sales were considered. A diverse set of 11 product categories was selected for the empirical study. The data comprise daily product information at the SKU level, including unit sales, price and promotions between January 7th 2014 and April 27th 2015 (476 days). There are two types of promotions: “direct discount” which is a simple price reduction and “discount in card” which saves the discount in a loyalty card to be used in next purchases. We focused our study on continuous demand products since most of the SKUs with intermittent demand usually are not promoted. Thus only the SKUs with sales on at least 95% of the days were selected. The forecasting performance of the proposed models was then evaluated with 692 SKUs from which 109 are focal. Table 1 presents the total number of SKUs and the number of focal SKUs in each category and the percentages of weeks concerning promotional activities ordered by percentage of weeks with direct discount. It can be seen that a quite diverse set of product categories was selected and that the direct discount is more frequent than the card discount across all categories (with exception to pasteurized milk). Figure 1 shows daily sales, price (in euros) and promotional periods of a SKU chosen randomly from the dataset. It can be seen that promotions may increase sales considerably, however their effects can be quite different. For each focal SKU in the dataset, 6 different models were evaluated: 1- ADL-only based on equation (1) using the focal SKUs own predictors; 2- ADL-intra-top5 based on equations (1) and (2) using the focal SKUs own predictors and the predictors from the five top sales SKUs in the same category; 3- ADL-intra-all based on equations (1) and (2) using the focal SKUs own predictors and the predictors from all the SKUs in the same category; 4-6- ADL-intra-PCA(k) based on equations (1) and (2) using the focal SKUs own predictors and the predictors extracted from PCA whose cumulative percentage of total variation is k. Our criterion was to consider k equal to 70%, 80% and 90%.

Table 1. Descriptive statistics of the dataset.

<table>
<thead>
<tr>
<th>Category</th>
<th>N° of focal SKUs</th>
<th>Mean percentage of weeks with promotional activities</th>
<th>Total n° of SKUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pork</td>
<td>3</td>
<td>49.31</td>
<td>109</td>
</tr>
<tr>
<td>Chicken</td>
<td>11</td>
<td>28.47</td>
<td>61</td>
</tr>
<tr>
<td>Apples</td>
<td>5</td>
<td>25.88</td>
<td>37</td>
</tr>
<tr>
<td>Canned fish</td>
<td>11</td>
<td>13.23</td>
<td>37</td>
</tr>
<tr>
<td>Cooking oil</td>
<td>4</td>
<td>12.13</td>
<td>25</td>
</tr>
<tr>
<td>Toothbrush</td>
<td>20</td>
<td>9.34</td>
<td>118</td>
</tr>
<tr>
<td>UHT milk</td>
<td>21</td>
<td>6.23</td>
<td>61</td>
</tr>
<tr>
<td>Coke</td>
<td>16</td>
<td>4.80</td>
<td>64</td>
</tr>
<tr>
<td>Manual laundry detergent</td>
<td>5</td>
<td>3.23</td>
<td>14</td>
</tr>
<tr>
<td>Sugar</td>
<td>8</td>
<td>0.92</td>
<td>44</td>
</tr>
<tr>
<td>Pasteurized milk</td>
<td>5</td>
<td>0.29</td>
<td>27</td>
</tr>
</tbody>
</table>

In order to evaluate the forecasting performance of these models, each SKU’s time series (sales, price, percentage of direct discount and percentage of discount in card) were split into in-sample (the first 355 days) and out-of-sample (the last 119). For each focal SKU, all models estimated by LASSO using the in-sample were used to forecast on the out-of-sample. The forecasting performance of the models was compared using the most traditional and popular error measures: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Scaled Error (MASE) defined in [5]. The results obtained (Table 2) show that the overall performance of the models is quite similar with ADL-only model performing a little better than the others with respect to RMSE, MAE and MASE. ADL-intra-top5 performed better concerning MAE which indicates that the inclusion of promotional information from SKUs in the same category may improve the forecast accuracy. Figures 2 and 3 show the MAE and MASE values of all models by category. Not much difference between the models is observed in MAE values. However, MAPE results show that none of the models outperforms all the others in all categories, indicating that better overall forecasting results may be obtained if the best model for each SKU is selected.

Table 2. The models’ forecasting accuracy.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>MAE</th>
<th>MPE</th>
<th>MAPE</th>
<th>MASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADL-only</td>
<td>84.92</td>
<td>45.94</td>
<td>42.07</td>
<td>74.45</td>
<td>0.876</td>
</tr>
<tr>
<td>ADL-intra-top5</td>
<td>85.31</td>
<td>45.22</td>
<td>41.70</td>
<td>75.72</td>
<td>0.891</td>
</tr>
<tr>
<td>ADL-intra-all</td>
<td>85.35</td>
<td>45.78</td>
<td>41.65</td>
<td>77.81</td>
<td>0.899</td>
</tr>
<tr>
<td>ADL-intra-PCA(70%)</td>
<td>86.25</td>
<td>46.26</td>
<td>41.13</td>
<td>76.21</td>
<td>0.896</td>
</tr>
<tr>
<td>ADL-intra-PCA(80%)</td>
<td>86.31</td>
<td>46.34</td>
<td>41.80</td>
<td>75.31</td>
<td>0.894</td>
</tr>
<tr>
<td>ADL-intra-PCA(90%)</td>
<td>86.79</td>
<td>46.30</td>
<td>42.04</td>
<td>75.41</td>
<td>0.898</td>
</tr>
</tbody>
</table>

Note: Bold values show the best result in the column.

4. Conclusions

A two-stage LASSO regression is proposed to deal with the problem of high dimensionality in sales forecasting at the SKU level. At the first stage, only focal SKUs promotional explanatory variables are included in the ADL model, namely sales lags, price and its lags, calendar events and weekend indicators. At the second stage, the in-sample forecast errors from the first stage are regressed on the explanatory variables from the other SKUs in the same category with the focal SKU. In order to evaluate the forecasting performance of the proposed models, daily information at the SKU level from a Portuguese retailer leader on food distribution and product manufacturing was used. The empirical results obtained with 692 SKUs from 11 categories show that the inclusion of promotional information from SKUs in the same category may improve the forecast accuracy and that better overall forecasting results may be obtained if the best model for each SKU is selected.
Sales Forecasting in Retail Industry
Based on Dynamic Regression Models

José Manuel PINHO©, José Manuel OLIVEIRA© and Patrícia RAMOS©
© Faculty of Economics, University of Porto, 4200-464 Porto, PORTUGAL.
© INESC Technology and Science, Campus da FEUP, 4200-465 Porto, PORTUGAL.
© School of Accounting and Administration of Porto, Polytechnic Institute of Porto,
4465-004 S. Mamede de Infesta PORTUGAL.

Abstract. Sales forecasts gained more importance in the retail industry with the increasing of promotional activity, not only because of the considerable portion of products under promotion but also due to the existence of promotional activities, which boost product sales and make forecasts more difficult to obtain. This study is performed with real data from a Portuguese consumer goods retail company, from January 2012 until April 2015. To achieve the purpose of the study, dynamic regression is used based on information of the focal product and its competitors, with seasonality modelled using Fourier terms. The selection of variables to be included in the model is done based on the lowest value of AIC in the train period. The forecasts are obtained for a test period of 30 weeks. The forecasting models overall performance is analyzed for the full period and for the periods with and without promotions. The results show that our proposed dynamic regression models with price and promotional information of the focal product generate substantially more accurate forecasts than pure time series models for all periods studied.

Keywords. Retailing, machine learning, forecasting, time series, promotions, dynamic regression.

1. Introduction

The effectiveness of sales forecasting is gaining increasingly importance in the retail sector. With retailers continuously trying to minimize stock and increase customer satisfaction it helps to reduce inventory investment costs and to improve logistics operations. A bad sales forecast may cause losses to the retailer, either by rupture or by excess of stock. Recent studies [1] indicate that in the case of rupture of stock of a product consumers decide to change to another store, not purchasing a replacement product, as initially thought. The promotional activity has increased sharply in recent years leading usually to a considerable increase of sales in the periods in which products are under promotional actions [2]. The efficacy of simple forecast methods, often used in the retail sector, is reduced when applied to periods when there are promotions [3]. The aim of this work is to incorporate promotions on econometric models to improve sales forecasting, especially in periods of promotional campaigns, and measure the impact of promotions on sales. This analysis is of particular importance because in recent literature models that integrate the promotional activity of