Computers and Operations Research 000 (2017) 1–12

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

Coping with demand volatility in retail pharmacies with the aid of big data exploration

Christos I. Papanagnou*, Omeiza Matthews-Amune

Salford Business School, University of Salford, Manchester, M5 4WT, UK

ARTICLE INFO

Article history:
Received 30 October 2016
Revised 30 June 2017
Accepted 13 August 2017
Available online xxx

Keywords:
Retail pharmacy
Data mining
Time series
Forecasting
Big data
Demand uncertainty

ABSTRACT

Data management tools and analytics have provided managers with the opportunity to contemplate inventory performance as an ongoing activity by no longer examining only data agglomerated from ERP systems, but also, considering internet information derived from customers' online buying behaviour. The realisation of this complex relationship has increased interest in business intelligence through data and text mining of structured, semi-structured and unstructured data, commonly referred to as "big data" to uncover underlying patterns which might explain customer behaviour and improve the response to demand volatility. This paper explores how sales structured data can be used in conjunction with non-structured customer data to improve inventory management either in terms of forecasting or treating some inventory as "top-selling" based on specific customer tendency to acquire more information through the internet. A medical condition is considered - namely pain - by examining 129 weeks of sales data regarding analgesics and information seeking data by customers through Google, online newspapers and YouTube. In order to facilitate our study we consider a VARX model with non-structured data as exogenous to obtain the best estimation and we perform tests against several univariate models in terms of best fit performance and forecasting.

1. Introduction

Due to the increasing prominence of concomitant complex factors regarding consumer choices, retail industry has been challenging with demand volatility and uncertainty, which affects demand planning and inventory holding (Azadeh et al., 2015). The detrimental consequences of this uncertainty in demand are often amplified from downstream echelons (retail stores) to upstream echelons (manufacturing) in a phenomenon referred to as bullwhip effect (Lee et al., 1997; Papanagnou and Halikias, 2008). Another major consequence of demand volatility is the increasing inaccuracy of forecasts which have resulted in excessive stocking leading to expiries and losses especially when considering products with a predetermined shelf life (Betts, 2014). Literature suggests a considerable amount of various univariate time series models and common estimation techniques in drug retail to address demand volatility. However, these cannot tackle limitations related to their applicability on historical data from short past windows, which are often characterised by linearity (Buzia et al., 2016; Anusha et al., 2014, see). Demand planning and inventory control in the retail industry have long been adversely affected by the combination of a complex set of factors determining consumer choices (Chao, 2015). Consumer behaviour related factors ranging from social, cultural and economic to psychological and personal have been implicated by researchers as complicit in promoting demand volatility, which can bring negative consequences on inventory control, consumer confidence and shareholder profits in the long run (Ferreira et al., 2015). Considering the high stakes involved with retail demand prediction, it is a vital problem for every retail company to address fluctuations on medicine inventories. The pharmaceutical retail business presents a unique challenge as profits are increasingly threatened by short expiry dates, increasing government regulation in the sale of medicines and fierce (sometimes monopolistic) competition by rivals (Gibson, 2012).

Cadeaux and Dubelaar (2012) attributed the challenge of demand volatility faced by retail pharmacies to the lack of a reliable inventory management system which should provide useful forecasting information on relationships between sales volatility and stock at the level of the specific product item. This has resulted in overstocking in some cases leading to excessive inventories, while understocking on the other hand resulting in poor customer satisfaction levels (Mahar et al., 2012). Yadav (2015) argue that retail pharmacies are a popular choice in low-income countries for individuals seeking healthcare for minor ailments as a re-

* Corresponding author.
E-mail addresses: c.papanagnou@salford.ac.uk, chrobo@yahoo.com (C.I. Papanagnou).

http://dx.doi.org/10.1016/j.cor.2017.08.009
0305-0548/

sult of the ease of access as compared to the bureaucratic processes, cost and time involved in hospital visitations. Also, in many smaller towns where hospitals are unavailable or reside in bigger cities, retail pharmacies are the first point of call for treatment and advice. Anusha et al. (2014) identify stock outs due to poor demand planning whereas the high cost of medicines arising from increased length of the supply chain. Considering the importance of retail pharmacies in the health care supply chain, accurate demand planning is critical to balancing demand and supply, which ensures product availability, waste reduction, customer satisfaction, improved inventory management, minimisation of over and under stocking and increased profits.

Linoff and Berry (2011) elucidate that accurate demand estimation is critical to minimising over and under stocking thereby minimising losses and most importantly maximises sales and customer satisfaction. Liu et al. (2013) attempt to address this challenge by both proposing a simulation optimisation system for pharmacy inventory management using empirical distributions to model demand. In a simulation-based approach, Kroger Company - which operates 1950 in-store pharmacies around the United States - managed to reduce out-of-stocks but no consideration was taken in terms of demand seasonality, trends and "disturbances" in demand pattern (Zhang et al., 2014).

Watson et al. (2014) compared the efficacy of various six sigma analytical models in improving the inventory replenishment policies and minimising out of stock rates in a large volume pharmacy setting. While out of stock rates were reduced, the research was based on process improvement and undesired "shortage" events with a lack of consideration of historical demand data. Ribeiro et al. (2016) adopted Pegels method a multiplicative trend exponential smoothing based technique - in order to guarantee sufficient stock levels and perform better short-term forecasting horizons for specific drug types in a pharmaceutical distribution company by investing solely on demand patterns. However, the presence of demand fluctuations constitutes a constraint for this approach as the proposed model can become very sensitive to volatility. In addition, the authors argue that the validity of the results is subject to additional data such as data types considered in the present paper.

Other studies suggest that the pharmaceutical retail demand may be challenged from a different angle by considering temporal and economical features that are associated with consumers and may affect demand in an extreme gradient boosting model (Sawon et al., 2016). However, the main limitation of this approach revolves around the lack of use of customer related data in favour to store related data. One solution for this could be given by merging customer related data in order to minimise drug inventory levels in a large medical distribution organisation. Neural networks computing systems were used to merge very basic customer related information (e.g., gender, unique customer number), which was obtained from already existing transactional (but limited) data in the store (Bansal et al., 1998). In another study, a combined ARIMA methodology with artificial neural networks (ANNs) was advanced in order to capture with accuracy both linear and nonlinear patterns of sales in a drug retail store (Khalil Zadeh et al., 2014). In particular, an explorative network based analysis was conducted on medicine sales by analysing linear covariations due to the lack of enough past sales records for each drug and customer related data.

The importance of considering customer related data by focusing on the impact of new channels of information sourced from the internet on consumer purchasing habits has been explicated in pharmaceutical industry. With the aid of internet-based platforms - like social media applications which encourage consumer interactions - the internet has become a primary health-related information source allowing data propagation through user-generated content and sharing capabilities (Greene and Kesselheim, 2010; Shankar and Li, 2014; Tyrawski and DeAndrea, 2015). Wosinska (2002) also documents the rising influence of internet-based direct-to-consumer advertising on supporting patients choices for several medicines and treatment. This has resulted in an increase in the total market demand for a therapeutic group as well as increased purchase frequency by means of greater therapy compliance.

Considering the rising importance and the vital role that consumer generated data from the internet plays on drug purchasing preferences, Kim et al. (2015) underscore the shortage of relevant existing demand planning research as the majority of available studies focus on simple data analysis. For the impact of such complex internet data to be evaluated, Thomassey and Fiordaliso (2006) suggest that data mining and machine learning have been shown to provide better results than statistical models in nonlinear data structures or when complex relationships exist. Hamuro et al. (1998) also support the notion that harnessing consumer information from data mining provides retail pharmacies with a competitive advantage in the demand estimation process.

To clearly assess the interdependencies of complex relationships obtained by data mining of the internet consumer-generated data, several VAR models have been suggested (Khalil Zadeh et al., 2014). The advantages and applicability of VAR models rest on the fact that they allow all key variables in the model to be considered symmetrically by identifying an equation for each variable, which contains lags and delays of all other variables in the model. Despite their numerous benefits, VAR models have been applied only scantily for demand estimation in pharmaceutical industry. Watson et al. (2014) favour the application of advanced VAR models in minimising out of stock rates in pharmacies by emphasising that the assumptions of conventional inventory theories (e.g., EOQ policies) were insufficient and unrealistic. Kim et al. (2015) attempted to evaluate the impact of user generated data from social media blogs by employing the VARX model in considering the impact of exogenous variables on upstream demand forecasting on pharmaceutical supply chains. They showed that the strengths of the VARX model reinforced by its ability to simultaneously analyse the impact of all the variables in the system on each other - making it highly adaptable to structural changes.

This paper proposes the use of the VAR in the analysis of the impact of exogenous variables obtained from a variety of internet based sources of endogenous historic sales data. According to Kenny (2011), exogenous variables are determined by external factors outside the model, which have an impact on the endogenous variables. This study suggests that Google search intensity, online newspaper keywords and YouTube video duration, which are treated as external variables, can form the exogenous variables. These variables can be explanatory as information seeking from patients may influence demand for medicines (Suziedelyte, 2012). The inclusion of external variables in estimation and forecasting models has increased in recent times as a result of the realisation that external explanatory factors - which have an impact on other variables existing within the model - may affect accuracy. Employing this model is expected to address the deficiencies of other "orthodox" models discussed by evaluating the complex nature of relationships and interdependencies in systems were situations like nonlinearity and seasonality coexist with a critical emphasis on the impact of consumer generated data on real business estimating situations.

1.1. Big data analytics and demand estimation techniques

The increase in quantity and speed of universal data generation from various sources over the last few decades has resulted in the emergence of big data term. Sources of such data have been

Please cite this article as: C.I. Papanagnou, O. Matthews-Amune, Coping with demand volatility in retail pharmacies with the aid of big data exploration, Computers and Operations Research (2017), http://dx.doi.org/10.1016/j.cor.2017.08.009