Int. J. Production Economics 160 (2015) 66–79

Contents lists available at ScienceDirect

Int. J. Production Economics

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

Forecasting aggregate retail sales: The case of South Africa

Goodness C. Aye [a], Mehmet Balcilar [b], Rangan Gupta [a,*], Anandamayee Majumdar [c]

[a] Department of Economics, University of Pretoria, Pretoria, 0002, South Africa
[b] Department of Economics, Eastern Mediterranean University, Famagusta, Turkish Republic of Northern Cyprus, via Mersin 10, Turkey
[c] Soochow University Center for Advance Statistics and Econometric Research, Suzhou, China

A R T I C L E I N F O

Article history:
Received 27 February 2013
Accepted 24 September 2014
Available online 5 October 2014

Keywords:
Seasonality
Weighted loss
Retail sales forecasting
Combination forecasts
South Africa

A B S T R A C T

Forecasting aggregate retail sales may improve portfolio investors' ability to predict movements in the stock prices of retail chains. This paper uses 26 (23 single and 3 combination) forecasting models to forecast South Africa's aggregate seasonal retail sales. We use data from 1970:01–2012:05, with 1987:01–2012:05 as the out-of-sample period. Unlike the previous literature on retail sales forecasting, we not only look at a wide array of linear and nonlinear models, but also generate multi-step-ahead forecasts using a real-time recursive estimation scheme over the out-of-sample period, to better mimic the practical scenario faced by economic agents making retailing decisions. In addition, we deviate from the uniform symmetric quadratic loss function typically used in forecast evaluation exercises, by considering loss functions that overweight the forecast error in booms and recessions. Focusing on the results of single models alone shows that their performances differ greatly across forecast horizons and for different weighting schemes, with no unique model performing the best across various scenarios. However, combination forecast models, especially the discounted mean-square forecast error method, which weighs current information more than past, not only produced better forecasts, but were also largely unaffected by business cycles and time horizons. This result, along with individual nonlinear models performing better than linear models, led us to conclude that theoretical research on retail sales should look at developing dynamic stochastic general equilibrium models that not only incorporate learning behavior, but also allow the behavioral parameters of the model to be state dependent, to account for regime-switching behavior across alternative states of the economy.

1. Introduction

Forecasting, in general, is a difficult task, but is, perhaps, more challenging for emerging economies. This is primarily because emerging economies are subject to various structural changes more often than developed economies (Aye et al., forthcoming). This paper is the first to examine the ability of different linear and nonlinear models to forecast retail sales in an emerging economy, namely South Africa.

The management of retail sales is of paramount importance to retail organizations and retail policy makers. Due to competition and globalization, sales forecasting plays a prominent role in the commercial enterprise (Xiao and Qi, 2008). Many retailers struggle to reduce their costs and increase profits. An accurate sales forecasting system is an efficient way to achieve these goals, as the reliable prediction of sales can improve business strategy. Forecasting of future demand is central to the planning and operation of a retail business at both the macro and micro levels. At the organizational level, sales forecasts are essential inputs to many decision activities in various functional areas such as marketing, sales, and production/purchasing, as well as finance and accounting (Mentzer and Bienstock, 1998; Zhang, 2009). Sales forecasts also provide the basis for regional and national distribution and replenishment plans. For profitable retail operations, accurate demand forecasting is crucial in organizing and planning purchasing, production, transportation, and the labor force, as well as after-sales services (Zhang, 2009). Therefore, the ability of retail managers to estimate the probable sales quantity in the next period could lead to improved customer satisfaction, reduced waste, increased sales revenue and more effective and efficient production plans (Chen and Ou, 2011a, 2011b).

Forecasting in the retail industry has basically been done using either individual or aggregate retail sales. Industry forecasts are especially useful to big retailers who may have a greater market share (Alon et al., 2001). For the retail industry, Peterson (1993) shows that large retailers are more likely to use time-series methods and prepare industry forecasts, while small retailers emphasize judgmental methods and company forecasts. Better forecasts of aggregate retail sales may improve the forecasts of individual retailers, because changes in their sales levels are often

* Corresponding author.
E-mail address: rangan.gupta@up.ac.za (R. Gupta).

http://dx.doi.org/10.1016/j.ijpe.2014.09.033
0925-5273/

systematic (Peterson, 1993). More accurate forecasts of aggregate retail sales may improve portfolio investors' ability to predict movements in stock prices of retailing chains (Barksdale and Hilliard, 1975; Thall, 1992; Alon et al., 2001). Poor forecasting may result in redundant or insufficient stock, which will directly affect the revenue and competitive position (Agrawal and Schorling, 1996).

Given the critical role of retail sales and the importance of its forecasting, this study sets out to forecast South Africa's aggregate retail sales. South Africa is used as a representative region for emerging economies as monthly retail sales data are available over a long period (1970–2012). The period under study is long enough to accommodate major events, not only in South Africa but also in the dominant economies around the world, which in turn, had an impact on the emerging markets.

The retail industry in South Africa is classified under the tertiary sector and falls within the wholesale and retail subsector (also known as the trade subsector). In 2011, the tertiary sector contributed 69.1 percent to the country's economy. The wholesale and retail trade subsector contributed approximately 13.7 percent to the economy. The retail trade and repair of goods made the largest contribution (45 percent) within the wholesale and retail trade subsector (IHS Global Insight, 2012; Gauteng Province: Provincial Treasury Quarterly Bulletin, 2012). This indicates that the retail industry drives the performance of the trade subsector. The retail industry contributes about 5.7 percent of total GDP. It is among the top industries in the country in terms of share of the employed labor force. The industry's share of employment of the national total has been fluctuating around 7 percent. The highest contribution made by the retail industry to employment was in 2006, when it reached 7.9 percent. In 2010, 7.2 percent of employed people were in the retail industry. The South African retail industry is one of largest retail industries in the sub-Saharan region that presents profitable investment opportunities for new players (RNCOS, 2011). The Global Retail Development Index (GRDI) annual publication ranks the top developing countries for retail expansion internationally, where countries are ranked on a 100-point scale. A higher ranking translates into a greater urgency for retailers to enter the specific country. In 2011, South Africa was ranked 26th out of 30 developing countries, with a score of 42.2, a deterioration from its 24th ranking in 2010 (41.7). However, South Africa dropped in the 2012 rankings, because of the market saturation of international retailers compared with other countries in the GRDI (Kearney, 2011, 2012). These statistics indicate the important role of the retail industry in South Africa, and hence, justify the need to forecast retail sales. Furthermore, from a macroeconomic perspective, forecasting retail sales, with it being a good proxy for consumption (Garrett, et al., 2004; Case et al., 2005, 2012; Zhou and Carroll, 2012), is likely to provide an early indication (being at a monthly frequency) of where the general economy (Gross Domestic Product [GDP]) might be heading, given that consumption is the dominant component of GDP. Like most, if not all, emerging and developing economies, consumption data in South Africa is available only at a quarterly frequency.

Retail sales data show strong seasonal variations. How best to deal with seasonal time series and which seasonal model is the most appropriate for a given time series are still largely unsolved issues (Zhang and Kline, 2007). Some of the international studies on retail sales forecasting attempt to select the optimal forecasting model by comparing forecasts from single artificial neural network (ANN) models with one or two traditional methods such as the exponential smoothing, moving average (MA), autoregressive and integrated moving average (ARIMA), seasonal ARIMA (SARIMA) and generalized autoregressive conditional heteroscedastic (GARCH) models. The forecasting ability of the traditional models is limited by their assumption of linear behavior, and thus, they are not always satisfactory (Zhang, 2003). The ANN models – which offer an alternative – are also not without their limitations and criticisms (Moreno et al., 2011). The ANN models lack a theoretical foundation and a systematic procedure for the construction of the model, in comparison with the classical approximations such as the Box–Jenkins methodology (Box and Jenkins, 1976). According to Moreno et al. (2011), the most criticized aspect of using the ANN model is that the value of the parameters obtained by the network cannot be practically interpreted, unlike those of the classical statistical models. As a consequence, ANN models have been presented to the user as a 'black box'. However, attempts are being made at dealing with these criticisms (Hansen, et al., 1999; Montaño and Palmer, 2003; Palmer et al., 2008).

From the foregoing, it is clear that all commonly used single forecasting models (traditional or ANN models) have their own characteristics, strengths and weaknesses. Further, when employing a single model, only part of the effective information can be used, showing that the range of information is insufficient (Wan et al., 2012). A single model will also be affected by the model's set conditions and other factors. These factors may reduce the accuracy of individual forecasting methods and increase the size of errors. Combining different forecasts averages such errors (Makridakis, 1989). The origin of forecast combination dates back to the seminal work of Bates and Granger (1969). Empirical findings, in general, show that combining improves the accuracy of forecasting and reduces the variance among post-sample forecasting errors (Makridakis and Winkler, 1983), and this also holds true in statistical forecasting, when judging estimates and when averaging statistical and subjective predictions (Clemen, 1989). In this regard, most studies on retail sales forecasting have attempted to improve on forecasts from single models by combining forecasts from two or, at most, three ANN models (see Chang and Wang 2006; Doganis et al., 2006; Aburto and Weber, 2007; Chen and Ou, 2009, 2011a, 2011b; Ni and Fan, 2011).

Against this background, the academic contribution of this paper to the literature on forecasting retail sales involves considering a wider range of forecasting models (23). This set of models includes not only the commonly used ANN models and traditional linear models, whose weaknesses have been noted above, but also other nonlinear methods with good statistical and theoretical foundations. In addition, we use recent emerging Bayesian techniques, which permit prior distributions on the coefficients of the models, to quantify the uncertainty about parameter values, given the observed data. Using a large number of models gives us a greater chance of selecting the 'best' among the set with very little or no bias. Further, the study contributes by providing a combined forecast of these different models.

The theory of forecast combination suggests that methods that weight better-performing forecasts more heavily will perform better than simple combination forecasts, and that further gains might be obtained by introducing time variation in the weights or by discounting observations in the distant past (Stock and Watson, 2004). This study uses not only the simple methods, but also more complex and efficient forecast combination techniques, not previously used in the retail sales forecasting literature. Also, the studies relating to forecasting retail sales have used a fixed-parameter estimation method, whereby the parameters estimated over the in-sample period are kept fixed for the out-of-sample period, when generating forecasts. Given that the retail sales series, in general, arehighly volatile, to produce forecasts, the models should be estimated recursively over the out-of-sample horizon, by first identifying an in-sample where the series is likely to be more stable. We not only perform recursive estimations, but also look beyond one-step-ahead forecasting horizons by producing multi-step results for short, medium and longer forecasting horizons. Further, the performance of a particular model may be determined by the type of evaluation criteria employed: the use of improper criteria to evaluate forecasts may result in poor forecasting performance (Van Dijk and Franses, 2003). While the use of the standard root-mean square error (RMSE) and Harvey-modified Diebold-Mariano (MDM) tests (Harvey et al., 1997) are widely accepted in the forecasting literature, the recent recession has demonstrated that a good forecast of a rather extreme event might be of special interest beyond that of minimizing an average squared error. Hence, we extend previous studies on retail sales forecasting, by considering loss functions that overweight the forecast errors in either booms or recessions or both.

Given the above forecasting setup, the results obtained (discussed in detail further on) have important implications for the retail sales industry. Clearly, given the complexity involved in forecasting, the industry is likely to benefit from a wider array of linear and nonlinear models than just the few models currently used. This is primarily because different models capture different aspects of the series (be they nonlinearity or seasonality) at different points of the forecasting horizon. In addition, forecasting during periods of boom and recession also affect the success of the forecasting models, since predictability during periods of calm may be very different from that during periods of turmoil. Finally, since forecast combination methods reduce forecast variability across horizons, as well as during booms or recessions, it is not only important for the retail sector to have a wide array of models, but also for it to have knowledge of forecast combination methods. This is of pivotal importance to the retail industry, since combination approaches are likely to provide more accurate forecasts, as well as less variable forecasts across extreme periods, thus allowing the sector to make better planning and organizing input decisions concerning the production process, transportation, inventory management and after sales services.

The rest of the paper is organized as follows. The next section discusses the literature on retail sales forecasting. The data and econometric methodology are discussed in Section 3. The empirical results are reported in Sections 4 and 5.

2. Literature

In this section, we provide a review of the empirical studies on retail (aggregate and individual) sales forecasting, with a view to confirming the contribution of this study, as described in the Introduction.

Alon (1997) found that the Winters' exponential smoothing model forecasts aggregate retail sales more accurately than the simple exponential and Holt's models and that it accurately forecasts individual product sales, company sales, income statement items, and aggregate retail sales. Alon et al. (2001) compared the performance of artificial neural networks (ANNs) to the traditional time-series models, namely, the Winters' exponential smoothing, ARIMA models and multivariate regression, using monthly aggregate retail sales data for the U.S. Their results, based on a mean absolute percentage error (MAPE) suggested that the ANN methods produced the best results, as they were able to "capture the dynamic nonlinear trend and seasonal patterns, as well as the interactions between them."

Chu and Zhang (2003) compared the out-of-sample forecasting performance of linear (ARIMA with time series, regression with dummy variables, and regression with trigonometric variables) and nonlinear (neural networks) seasonal forecasting models for the U.S.'s monthly aggregate retail sales from January 1985 to December 1999. They found that the neural network estimated using deseasonalized data outperformed the rest of the models, based on three performance measures (the root-mean squared error [RMSE], the mean absolute error [MAE] and the [MAPE]). They also found that although seasonal dummy variables could be useful for predicting retail sales, their performance may not be robust, and that trigonometric models are not useful in aggregate retail sales forecasting.

Frank et al. (2003), using U.S. annual data from 1997 to 2000 on women's apparel sales, evaluated the forecasting performance of three different forecasting models, namely single seasonal exponential smoothing, Winters' three parameter model, and ANN model. Their result indicated that the ANN model outperformed the other two models, based on an R^2 evaluation. Doganis et al. (2006) presented an evolutionary sales forecasting model that is a combination of two artificial intelligence technologies, namely the radial basis function and a genetic algorithm (GA-RBF). This methodology was applied to sales data of fresh milk provided by a major manufacturing company of daily produce in Greece, and the findings from different formulations of the model were compared with linear (AR, ARMA, RLS, Holt-Winters) models. Their findings showed that the adaptive formulation of the combined neural network model had the least MAPE, indicating that models that allow correction of themselves as new information becomes available are able to forecast sales more accurately. Chang and Wang (2006) integrated fuzzy logic and an artificial neural network into the fuzzy back-propagation network (FBPN) for sales forecasting in the printed circuit board (PCB) industry in Taiwan. The results from FBPN were compared with those of Grey forecasting (GF), multiple regression analysis (MRA) and back-propagation networks (BPNs). The experimental results indicate that the fuzzy back-propagation approach outperforms the other three different forecasting models in Mean Absolute Percentage Error (MAPE) measures.

Aburto and Weber (2007) presented a hybrid intelligent system, combining an ARIMA model and MLP neural networks for demand forecasting, and found that the MLP model outperformed the ARIMA model, while the hybrid model outperformed the individual models, based on a MAPE and a normalized MSE. They also showed that a replenishment system for a Chilean supermarket, based on improved forecasting accuracy, simultaneously led to fewer sales failures and lower inventory levels. Joseph et al. (2007) examined the out-of-sample forecasts of aggregate sales using the 3-month Treasury bill interest rate in a NeuroSolutions environment, referenced against the forecasts of linear regression models. Two types of dynamic neural network models trained with the Levenberg–Marquardt back propagation algorithm under supervised learning were used. The neural network models outperform the linear regression models. Au et al. (2008) illustrated an evolutionary neuron network for sales forecasting, and showed that when guided with the BIC and the pre-search approach, the non-fully connected neuron network could converge faster and more accurately in forecasting time series than the fully connected neuron network and traditional SARIMA model based on an MSE criterion. Sun et al. (2008) also developed different sales forecasting models for fashion retailing in Hong Kong. They applied an ELM neural network model to investigate the relationship between sales amount and some significant factors that affect demand. The results demonstrate that the proposed methods outperform the back-propagation neural network model. Ali et al. (2009) explored the trade-off between forecasting accuracy and data and model complexity in the sales forecasting problem facing grocery retailing, by considering a wide spectrum in data and technique complexity. The results of the experiment indicate that simple time-series techniques perform very well for periods without promotions. However, for periods with promotions, regression trees with explicit features improve accuracy substantially.

Chen and Ou (2009) developed the GMFLN forecasting model by integrating GRA and MFLN neural networks. The experimental results indicate that the proposed forecasting model outperforms the MA, ARIMA and GARCH forecasting models for retail goods.

Gil-Alana et al. (2010) examined whether retail sales forecasts could be explained better in terms of a model that incorporates both long-run persistence and seasonal components in a fractional differencing framework than models that use integer degrees of differentiation. They found that retail sales forecasts are explained better in terms of a long memory model that incorporates both persistence and seasonal components. Chen and Ou (2011a, 2011b) developed a Grey relation analysis with an extreme learning machine (GELM) model for forecasting future daily sales in the fresh food retail industry in Taiwan. Using the MSE and MAD statistics, they showed that the GELM model outperforms the standard statistical time-series model, GARCH, as well as two other artificial neural network (GBPN, and GMFLN) models. Ni and Fan (2011) proposed a two-stage dynamic forecasting model, which is a combination of the ART model and an error forecasting model based on a neural network, to improve the accuracy of retail fashion forecasting. However, their results are not compared with other forecasting models.

As can be seen from the above, most of the studies reviewed emphasized the importance of different forms of neural network models (hence, nonlinearity in general), and compared the forecasts with a few linear forecasting models. These studies evaluate the forecasts from different models using the standard loss function, which essentially minimizes an average squared error, to show that ANN models, in general, tend to outperform standard linear models. However, in this study, we consider 26 seasonal forecasting (23 single and 3 combined) models for aggregate retail sales, and we employ forecast evaluation techniques with different weighting schemes to see how each model performs in times of booms and recessions. Broadly, we extend the literature on retail sales studies by considering a wide array of linear and nonlinear models beyond neural network methods. This is important, given the criticisms of neural network models, with the primary concern being that it is not theoretically founded. Moreover, not only do we perform recursive estimations to capture the volatilities inherent in retail sales, but we also, contrary to the literature, look beyond one-step-ahead forecasting horizons by producing multi-step results for short, medium and longer forecasting horizons.

At this stage, it is important further to emphasize the role of a recursive forecasting scheme, conducting multi-step-ahead forecasts, and forecast combination. The recursive approach mimics the real-time forecasting scenario of a retailer when making any decision pertaining to the production chain, in the sense that a retailer can use only available (in-sample) information on retail sales. In addition, forecasting for retail sales is complicated by the problem that a retailer, in real time, must reach decisions regarding various aspects under uncertainty concerning the "optimal" forecasting model. The real-time forecasting approach resolves this problem by assuming that a retailer uses a search-and-updating technique to predict retail sales. The search part requires that a retailer, in every period of time when a decision must be reached, estimates a large number of forecasting models and then identifies an "optimal" model by means of some model selection criteria based on available data, after identifying in and out-of-sample periods. The updating part, in turn, requires that a retailer re-estimates the forecasting models whenever new information on retail sales becomes available. The real-time forecasting approach, thereby, accounts not only for model uncertainty but also for the possibility that the optimal forecasting model may change over time.

As discussed previously, given a series as volatile as retail sales, changes in the optimal forecasting model are very likely due to structural breaks and regime shifts, something that we see across forecasting horizons and also periods of extremes in the discussion of our results, below. Previous studies, which relied on a once-off estimation of a model or a small set of models over an identified in-sample, are thus likely to be misspecified, since these studies not only reduce the possibility of choosing a better model, but also fail to account for parameter instability, and hence, model uncertainty. In light of this, it is highly possible that the observation made in the literature that a specific model always performs better than a benchmark over the entire out-of-sample horizon is spurious, since economic conditions change, and it is impossible for a single forecasting model to capture all the dynamics over the out-of-sample horizon. As far as multi-step-ahead forecasts are concerned, retailing decisions clearly are made not only over the short-run, but also over the medium- to long-run. Hence, multi-step-ahead forecasts based on a recursive approach (unlike the constant parameter method), which accounts for parameter and model uncertainty, are more realistic in nature and are also likely to be more precise.

Finally, forecast combination approaches allow us to reduce the problem of model uncertainty, since this provides us a statistical approach to combine the best features of the various models used, given model uncertainty over the out-of-sample horizon. However, it is also important to emphasize that forecast combination is not likely to perform well if the array of models used is limited in number. In other words, unlike the literature, our approach of using a large number of individual models, recursive estimation, multi-step-ahead forecasts and forecast combination is a more realistic and better way of conducting a forecasting exercise for a volatile series like retail sales.

3. Data and methodology

We used monthly aggregate sales data for South Africa covering the period 1970:01–2012:05, making a total of 509 observations. The period covers a number of economic events, thereby capturing

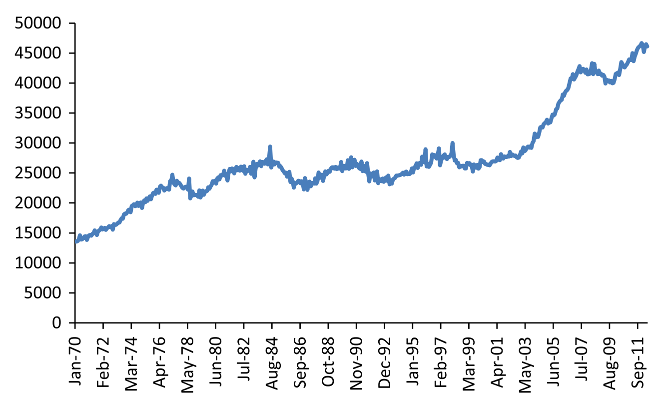


Fig. 1. Aggregate seasonal retail sales series in million rands.

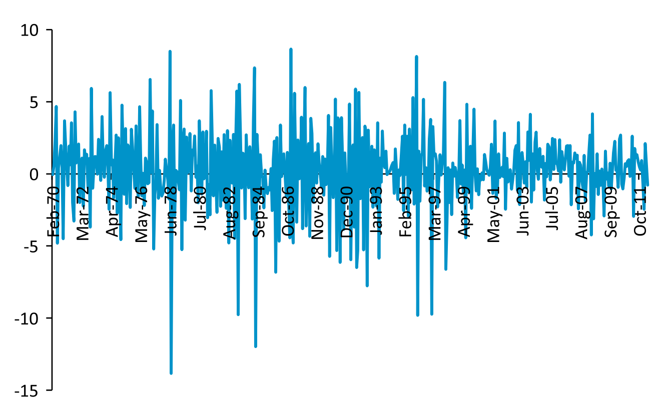


Fig. 2. Growth rate of aggregate seasonal retail sales.

Table 1
Model description and specification.

S/N	Code	Description and specification
A. Models with seasonal dummy variables		
1.	RW	Random walk, equivalent to ARIMA(0,1,0)
2.	ARIMA	Autoregressive integrated moving average; estimated model is ARIMA(2,1,0)
3.	ARFIMA	Autoregressive fractionally integrated moving average; estimated model is ARFIMA(2,1 + d,0)
4.	BARIMA	Bayesian ARIMA model parameters are estimated to minimize the 24-step MSE once over the out-of-sample period. We start with a long model with ARIMA (p,1,q), where $p,q \leq 12$. Estimates arising from minimizing 24-step MSE are used as informative priors in the recursive estimation
5.	BCAR	Bias corrected AR model. Estimated model is AR(2) with first differencing. The method we use is described in Stine and Shaman (1989)
6.	MSAR	Markov switching autoregressive model, Estimated model is MS-AR(2) with two regimes and first differencing.
7.	SETAR	Self-exiting threshold autoregressive model. Estimated model is SETAR (k,p,d), with $k=2$ (# of regimes), $p=2$ (autoregression order) and $d=1$ (delay order)
8.	LSTAR	Logistic smooth transition autoregressive model. Estimated model is LSTAR (k,p,d), with $k=2$ (# of regimes), $p=2$ (autoregression order) and $d=1$ (delay order)
9.	ARANN	Autoregressive artificial neural network. Autoregressive order is 2. We use three hidden layers. The ANN is a multi-layer perceptron (MLP) feed-forward network, with a hyperbolic-tangent (tansig) activation function for the hidden layers and a linear activation function for the output layer
10.	NPAR	Fully non-parametric (auto) regression. An autoregressive model with lag order equal to 2
11.	SPAR	Semi-parametric (auto) regression. An autoregressive model with lag order equal to 2
12.	GARCH	Generalized autoregressive conditional heteroscedasticity model. We use an ARIMA(2,1,0)–EGARCH(1,1) model
13.	GA	This is genetic algorithm-based forecasting. Two lags are used as inputs (see Szpiro, 1997 for the approach we used). Function approximation is terminated at a maximum step of 3000
14.	FUZZY	Evolutionary fuzzy modeling. The approach is taken from Peña-Reyes (2004). Fuzzy fitting uses a population of 200 and 60000 generations
15.	DISC	Discounted forecast combination. The discount factor we used is 0.50
16.	PC	Principal components forecast combination. We used a maximum of 4 principal components, based on the Bai and Ng (2002) method
17.	MEAN	Simple mean forecast combination
B. Full seasonal models		
1.	SRW	Seasonal random walk, ARIMA(0,1,0)(0,1,0), so both seasonal and regular random walk components exist
2.	HW	Holt-Winters methods; tree smoothing parameters are estimated
3.	TBATS	State-space exponential smoothing model with trigonometric seasonal component (See Hyndman et al. 2002)
4.	SARANN	Seasonal autoregressive ANN. The ANN is multi-layer perceptron (MLP) feed-forward network with a hyperbolic-tangent (tansig) activation function for the hidden layers and a linear activation function for the output layer. We use the approach in Taskaya-Temizel and Casey (2005) to set the number of delays (AR order). A total of 9 hidden layers are used
5.	SUTSEA	Seemingly unrelated structural time-series model with local trend and additive seasonal component (see Harvey, 2006)
6.	SUTSET	Seemingly unrelated structural time-series model with local trend and trigonometric seasonal component (see Harvey, 2006)
7.	SARIMA	Seasonal ARIMA. The estimated model is SARIMA(2,1,0)(2,0,0)
8.	BSARIMA	Bayesian SARIMA model parameters are estimated to minimize the 24-step MSE once over the out-of-sample period. We start with a long model with ARIMA (p,1,q)(P,0,Q), and where $p,q,P,Q \leq 4$. Estimates arising from minimizing the 24-step MSE are used as informative priors in the recursive estimation
9.	SARFIMA	Autoregressive fractionally integrated moving average. The estimated model is ARFIMA (2,1 + d,0)(2,0,0)

both the boom and recession periods in South Africa. The data was sourced from Statistics South Africa. The full data set is split into two. We use data from 1970M1 to 1986M6 (204 observations) for the in-sample. Data from 1987:01 to 2012:05 (305 observations) was used for the out-of-sample period.

The plot of the seasonally adjusted aggregate retail sales series is shown in Fig. 1, while its growth rate is plotted in Fig. 2. There is a noticeable seasonal variation in the data. Fig. 1 shows that retail trade sales follow a particular pattern annually. Every December, retail sales figures spike upward and in January, contractions occur. This trend is explained by the tendencies of households to shop more during the month of December, since most people are on holiday or have received bonuses during this period. In the month of January, consumer spending reduces, as people prepare to go back to work or school, and also to pay off short-term debts incurred in December. The overall trend is an increase in retail trade sales. Fig. 2 also depicts strong volatility, with the highest peak in January 1987 (8.6%), thus justifying our choice of 1987:01–2012:05 as the out-of-sample period.

3.1. *Forecasting Models*

A model was identified using the in-sample data; then the same model was recursively re-estimated, and 1 to 24 step-ahead forecasts were obtained recursively over the out-of-sample period. Only the parameters were re-estimated in the recursive forecasting, but the identified model structure was kept constant.

We have two classes of models. The acronyms and brief description of the models we used are presented in Table 1.[1] The first class consists of 17 models with seasonal dummy variables. This is equivalent to deterministic seasonal adjustment. For instance, for the ARIMA model we estimate

$$\phi(L)\Delta^d y_t = \mu + \sum_{s=1}^{11} \gamma_s d_{s,t} + \theta(L)\varepsilon_t, \tag{1}$$

where y is the log of the aggregate retail sales and $d_{s,t}$ is a dummy variable taking the value of one for month s. At each recursive estimation, step dummy are included in the regression and forecasts of seasonal component are easily obtained from $\mu + \sum_{s=1}^{11}\gamma_s d_{s,t}$. The models are presented in panel A of Table 1. When joint estimation of the seasonal component and non-seasonal component is not feasible, for the Genetic Algorithm (GA) method, for instance, the seasonal component is pre-estimated using linear regression, and the non-seasonal component is forecast in a second step; final forecasts are recovered by adding $\mu + \sum_{s=1}^{11}\gamma_s d_{s,t}$. The second class consists of nine full seasonal models. The models are presented in panel B of Table 1. In all models, data is the log of first differences, since there is a unit root. Level forecasts are recovered from the forecasts of the growth rates. All model order is selected using BIC. In each case, forecasts were made at four horizons: 1, 4, 12 and 24 months.

[1] However, given the pivotal role of forecast combination in this paper, detailed descriptions of the forecast combination models are given in the next subsection.

Each model we use in our study has one or more features. Full seasonal models attempt to capture seasonal variation using parameterization suitable for stochastic and complicated seasonality. Models with seasonal dummies assume that seasonal variation is deterministic. RW and SRW models serve as benchmark models. We can split these models into two basic groups. The first group is a class of linear time-series model, and these include ARIMA, ARFIMA, BARIMA, BCAR, HW, TBATS, SUTSEA, SUTSET, SARIMA, BSARIMA, and SARFIMA. These models are the most commonly used class of models for modeling linear short- and long-memory time series. The ARFIMA and SARFIMA models assume a fractionally integrated time series and capture long memory. The BARIMA and BSARIMA models are based on Bayesian estimation of the parameters; also the order of the models is chosen based on informative priors to improve the out-of-sample forecasting errors. The BCAR corrects the bias in autoregressive parameter estimations, which may improve the forecasting performance in cases where parameter estimates may be biased due to the small sample size and deviations from assumptions. The second group of models can be classified broadly as nonlinear models. The nonlinear models include the following: MSAR, SETAR, LSTAR, ARANN, NPAR, SPAR, GARCH, GA, FUZZY, and SARANN. These models capture various types of nonlinearities and may have better forecasting performance if the underlying time series has nonlinear dynamics. The regime switching models MSAR, SETAR, and LSTAR are well known in the literature and best fit cases where a time series follows asymmetric dynamics, like recessions and booms. The MSAR models the regime switching based on a latent regime variable that follows a first-order Markov process, and therefore has unexplained switching. The SETAR and LSTAR models have explained switching, and therefore regime switching follows a known structure. SETAR models assume a swift switching that is completed in one period, while the LSTAR model assumes a smooth switching in and out of a regime that spreads to more than one period. The GARCH model assumes an autoregressive conditionally heteroscedastic error term and suitable time-varying variance. The ARANN, NPAR, SPAR, GA, FUZZY, and SARNN models do not assume any known parametric functional form and successfully approximate quite general nonlinear functions. Moreover, we also form a combination of the different models using three combination methods not previously used in any of the retail sales forecasting papers, to ensure that the combination model that best fits the data is selected

3.1.1. Forecast combination methods

Three forecast combination methods are considered: the simple forecasts (MEAN), the discounted MSFE (DISC) and the principal component (PC) methods. Our selection of these three is based on their good performance as reported in previous studies. The forecast combination methods differ in the way they use historical information to compute the combination forecast and in the extent to which the weight given an individual forecast is allowed to change over time. Some of the combining methods require a holdout period to calculate the weights used to combine the individual model forecasts, and we use the first P_0 observations from the out-of-sample period as the initial holdout period, following Rapach and Strauss (2010). The combination forecasts of y^h_{t+h} made at time t, $\hat{y}^h_{CB,t+h|t}$, are typically a linear combination of the individual model forecasts

$$\hat{y}^h_{CB,t+ht} = \sum_{i=1}^{n} w_{i,t}\hat{y}^h_{i,t+h|t} \tag{2}$$

where $\sum_{i=1}^{n} w_{i.t} = 1$. When the weights, $\{w_{i,t}\}_{i=1}^{n}$, are estimated, we use the individual out-of-sample forecasts and y^h_{t+h} observations available from the start of the holdout out-of-sample period to time t. For each of the combining methods, we compute combination forecasts over the post-holdout out-of-sample period. This leaves us with a total of $P_h = P-(h-1)-P_0$combination forecasts, $\left\{\hat{y}^h_{CB,t+h|t}\right\}_{t=R+P_0}^{T-h}$, available for evaluation.[2]

3.1.1.1. Simple combination forecasts. The simple combination forecasts compute the combination forecasts without regard to the historical performance of the individual forecasts. Stock and Watson (1999, 2003, 2004) find that simple combining methods work well in forecasting inflation and output growth using a large number of potential predictors. They (Stock and Watson, 2004) note that there seems to be little difference between the performance of the mean and the trimmed mean forecasts, while the median forecast typically has a somewhat higher relative MSFE than either the mean or trimmed mean forecasts. Therefore, we consider the mean combination forecast (MEAN). The mean combination forecast simply involves setting $w_{i,t} = 1/n$ $(i = 1, \ldots, n)$ in (2). Thus, the simple combining methods do not require a holdout out-of-sample period.

3.1.1.2. Discounted MSFE combination forecasts. Following Stock and Watson (2004) and Rapach and Strauss (2010), we consider a combining method, where the weights in (2) are a function of the recent historical forecasting performance of the individual models. The discounted MSFE combination (*DISC*) h-step-ahead forecast method has the form (2) where the weights are as follows:

$$w_{i,t} = m_{i,t}^{-1} / \sum_{j=1}^{n} m_{j,t}^{-1}, \tag{3}$$

where

$$m_{i,t} = \sum_{s=R}^{t-h} \gamma^{t-h-s}(y^h_{s+h} - \hat{y}^h_{i,s+h|s})^2 \tag{4}$$

and γ is a discount factor. When $\gamma = 1$, there is no discounting, and (3) produces the optimal combination forecast derived by Bates and Granger (1969) for the case where the individual forecasts are uncorrelated. When $\gamma < 1$, greater importance is attached to the recent forecasting accuracy of the individual models. We consider γ value of 0.5. The results are the same with a γ value of 0.70. Although, this seems to be a low discount factor, it however, may be due to the seasonal time series we are forecasting, and as a recent past is the most important, weights given to the past forecast were required to decline very fast in our case.

3.1.1.3. Principal component forecast combination. Principal component forecast combination (PC) requires (i) recursively computing the first few principal components of the estimated common factors of the panel of forecasts, (ii) estimating a regression of $y^h_{s+h|s}$ onto these principal components, and (iii) forming the forecast based on this regression (Stock and Watson, 2004). The reduction of many forecasts to a few principal components provides a convenient method for allowing some estimation of factor weights, yet reduces the number of weights that must be estimated. This method has been used by Chan et al. (1999), Stock and Watson (2004) and Rapach and Strauss (2010), among others. One motivation for using PC is that recent work on large forecasting models suggests that large macroeconomic data sets are well described by a few common dynamic factors, which are useful for forecasting, and that the common factors can be estimated by principal components (Forni, et al., 2000, 2003; Stock and Watson, 1999, 2002, 2004).

[2] We use 1987:01–1996:12 as the initial hold-out out-of sample period.

The principal component forecasts are constructed as follows. Let $\hat{F}^h_{1,s+h|s},...,\hat{F}^h_{m,s+h|s}$ for $s=R,...,t$denote the first m principal components of the uncentred second-moment matrix of the individual model forecasts, $\hat{y}^h_{i,s+h|s}$ $(i=1,...,n; s=R,...,t)$. To form a combination forecast of y^h_{t+h} at time t based on the fitted principal components, we estimate the following regression model

$$y^h_{s+h}=\theta_1\hat{F}^h_{1,s+h|s}+\cdots+\theta_m\hat{F}^h_{m,s+h|s}+v^h_{s+h}, \tag{5}$$

where $s=R,...,t-h$. The combination forecast is given by $\theta_1\hat{F}^h_{1,s+h|s}+\cdots+\theta_m\hat{F}^h_{m,s+h|s}$, where $\hat{\theta}_1,...,\hat{\theta}_m$are OLS estimates of $\theta_1,...,\theta_m$, respectively, in (5). We use the IC_{p3} information criterion developed by Bai and Ng (2002) to select m (considering a maximum value of 4) when calculating combination forecasts using the *PC* method. Bai and Ng (2002) show that familiar information criteria such as the Akaike information criterion (AIC) and the Schwarz Bayesian information criterion (SIC) do not always consistently estimate the true number of factors, and they develop alternative criteria that consistently estimate the true number of factors under more general conditions. In extensive Monte Carlo simulations, and using a large sample size, as in our study, Bai and Ng (2002) find that the IC_{p3} criterion performs well in selecting the correct number of factors.

3.2. Forecast evaluation using weighted loss functions

The standard period-t loss function used in most of the forecast evaluation literature is the squared forecast error

$$L_{i,t}=e^2_{i,t}, \tag{6}$$

where $e_{i,t}=y_t-y^f_{i,t}$ is the forecast error of model i, y_t is the realization of the target variable, y, and aggregate retail sales in our case, $y^f_{i,t}$, is the value predicted by model i. Comparing the average loss difference of the two competing models 1 and 2 implies computing their mean squared forecast errors

$$\text{MSFE}_i=\frac{1}{P}\sum_{t=T+1}^{T+P}e^2_{i,t}, \quad i=1,2, \tag{7}$$

over the forecast period $T+1$ to $T+P$, and choosing the model with the smaller MSFE.

However, according to Carstensen et al. (2010), there are many occasions in which different loss functions can make more sense for the applied forecaster but also for the user of a forecast, such as a politician or the CEO of a company. For instance, the case of the recent recession, which demonstrated that a good forecast of a rather extreme event might be of special interest beyond that of minimizing an average squared error. Consequently, banks could have taken earlier measures to shelter against the turmoil, governments could have started stimulus packages in time, and firms might have circumvented their strong increase in inventories. This is in line with Van Dijk and Franses (2003) argument that a weighted squared forecast error can be used to place more weight on unusual events when evaluating forecast models. Following Van Dijk and Franses (2003) and Carstensen et al. (2010), we use a weighted squared forecast error. Hence, the loss function in (6) can be respecified as follows:

$$L^w_{i,t}=w_te^2_{i,t}, \tag{8}$$

where the weight w_t is specified as

1. $w_{left,t}=1-\hat{F}(y_t)$, where $F(\cdot)$is the cumulative distribution function of y_t, to overweight the left tail of the distribution. This gives rise to a "recession" loss function.
2. $w_{right,t}=\hat{F}(y_t)$, to overweight the right tail of the distribution. This gives rise to a "boom" loss function.
3. $w_{tail,t}=1-\hat{f}(y_t)/\max(\hat{f}(y_t))$, where $F(\cdot)$ is the density function of y_t, which allows a focus on both tails of the distribution giving rise to both the recession and boom loss functions.

When equal weights, $w_t=1$, are imposed, the weighted loss function (8) collapses to the standard loss function (6), giving rise to the conventional "uniform" loss function.

To evaluate a forecast model i over a forecast period $T+1$to $T+P$ using the weighted loss function simply requires calculating the weighted mean squared forecast error,

$$\text{MSFE}_i=\frac{1}{P}\sum_{t=T+1}^{T+P}w_te^2_{i,t}, \tag{9}$$

In order to compare, say, model i to a benchmark model 0, one calculates the weighted loss difference

$$d_{i,t}=L^w_{0,t}-L^w_{i,t}=w_te^2_{0,t}-w_te^2_{i,t}, \tag{10}$$

and averages over the forecast period

$$\overline{d}_i=\frac{1}{P}\sum_{t=T+1}^{T+P}d_{i,t}=\frac{1}{P}\sum_{t=T+1}^{T+P}w_te^2_{0,t}-\frac{1}{P}\sum_{t=T+1}^{T+P}w_te^2_{i,t} \tag{11}$$

We use this weighted loss and analyze the forecast accuracy of different models with respect to the different weighting schemes introduced above. There are a large number of tests proposed in the literature to analyze whether empirical loss differences between two or more competing models are statistically significant. The most influential and widely used is the pairwise test introduced by Diebold and Mariano (1995). In this study, we employ the modified Diebold-Mariano (*MDM*) test proposed by Harvey et al. (1997), which corrects for a small sample bias. The *MDM* test is a pairwise test designed to compare two models at a time, say, model *i* with benchmark model 0. The null hypothesis of the *MDM* test is that of equal forecast performance,

$$E[d_{i,t}]=E\left[L^w_{0,t}-L^w_{i,t}\right]=0 \tag{12}$$

Following Harvey et al. (1997), we use the modified Diebold-Mariano test statistic

$$\text{MDM}=\sqrt{\frac{P+1-2h+h(h-1)/P}{P}}\frac{\overline{d}_i}{\sqrt{\hat{V}(\overline{d}_i)}}, \tag{13}$$

where h is the forecast horizon and $\hat{V}(\overline{d}_i)$is the estimated variance of series $d_{i,t}$. The *MDM* test statistic is compared with a critical value from the t-distribution with $P-1$ degrees of freedom.

The forecasting performance of a candidate forecast is also evaluated by comparing its out-of-sample RMSE to the benchmark forecast, following Chan et al. (1999), Stock and Watson (2004) and Rapach and Strauss (2010). The benchmark forecast used here is from the random walk (RW) model. If the candidate forecast has a relative RMSE of less than one, then it outperformed the RW benchmark over the forecast period.

RMSE is simply the square root of the mean square error (MSE), which is the most frequently used forecast error measure by the academicians and practitioners (Carbone and Armstrong, 1982). The RMSE is equivalent to the root mean square percentage error when the forecasted series is in logs, which is the case in this study. RMSE is scale dependent and would not be recommended for comparing methods for a group of series. Chatfield (1992) points out that it is perfectly reasonable to evaluate forecasts from different models by using the RMSE for single series. We compare the forecasting performance of a group of models for a single series, and using the RMSE does not have any disadvantage over the other forecast error measures. Further, Zellner (1986) points out that using the mean of the predictive probability density function for a series is optimal relative to a squared error loss function and the MSE or RMSE, and hence, the RMSE is an appropriate measure to evaluate the performance of forecasts when the mean of the predictive probability density function is used – as it happens to be in our case – when estimating the nonlinear and Bayesian models.

4. Empirical results

In this section, we report the results from all the 26 aggregate retail forecasting models. We first present the uniform, boom, recession, and boom and recession weighted RMSEs and their corresponding ranks. These results are presented in Tables 2–5 for horizons of 1, 4, 12 and 24 months, respectively. The rankings in most – but by far not in all – cases differ greatly between boom and recession periods, and even at different forecast horizons. In general, models with seasonal dummy variables seem to have a smaller RMSE than full seasonal models. Also as a general result, the average forecast errors based on the uniform weighting scheme are strongly driven by the forecast errors made during booms, which are substantially higher than during recessions. This holds true for all models and forecast horizons. It implies that improvements in terms of model building should aim at better predictions of boom periods.

Interestingly, the combination forecasts, especially the DISC and PC models, outperform the single or individual forecast models. The outstanding performance of the DISC appears to be robust to both the weighting scheme and forecast horizons, taking 1st rank in 12 cases out of 16, and 2nd for the remaining 4 cases. This implies that, in general, the DISC model has the smallest RMSE. Following closely behind the DISC is the PC model. However, we observe that at medium- and longer-term horizons ($h=12$ and $h=24$), the PC model's performance for either the recession forecasts or tail forecasts is not quite as impressive, as it ranks between 6th and 18th in these cases. Another interesting finding in this study with respect to the RMSE evaluation criterion is that the more sophisticated forecast combination methods outperformed the simple mean combination method, unlike the other studies cited previously.

Table 2
Root-mean squared forecast errors ($h=1$).

Model	Uniform		Boom		Recession		Tail	
	RMSE	Rank	RMSE	Rank	RMSE	Rank	RMSE	Rank
RW	0.0252	15	0.0193	15	0.0074	13	0.0112	14
DISC	0.0139	1	0.0109	1	0.0039	2	0.0057	2
PC	0.0180	3	0.0136	3	0.0058	3	0.0086	3
MEAN	0.0209	4	0.0154	8	0.0067	4	0.0101	5
ARIMA	0.0217	12	0.0166	14	0.0068	5	0.0100	4
ARFIMA	0.0213	9	0.0159	12	0.0070	7	0.0103	6
BARIMA	0.0264	16	0.0205	16	0.0074	14	0.0114	16
BCAR	0.0213	8	0.0157	9	0.0072	10	0.0106	9
MSAR	0.0150	2	0.0121	2	0.0038	1	0.0057	1
SETAR	0.0212	7	0.0153	7	0.0070	8	0.0105	8
LSTAR	0.0209	5	0.0153	6	0.0069	6	0.0103	7
ARANN	0.0215	11	0.0158	11	0.0072	11	0.0107	10
NPAR	0.0210	6	0.0148	4	0.0073	12	0.0109	12
SPAR	0.0218	13	0.0159	13	0.0071	9	0.0108	11
GARCH	0.0374	24	0.0282	24	0.0117	22	0.0177	22
GA	0.0215	10	0.0152	5	0.0078	16	0.0113	15
FUZZY	0.0219	14	0.0157	10	0.0076	15	0.0111	13
SRW	0.0325	21	0.0235	22	0.0109	21	0.0161	21
HW	0.0285	18	0.0221	19	0.0095	17	0.0131	17
TBATS	0.0367	23	0.0273	23	0.0124	24	0.0180	24
SARANN	0.0329	22	0.0222	20	0.0118	23	0.0180	23
SUTSEA	0.0292	19	0.0228	21	0.0096	18	0.0133	18
SUTSET	0.1031	26	0.0738	26	0.0378	26	0.0537	26
SARIMA	0.0278	17	0.0208	17	0.0098	19	0.0136	19
BSARIMA	0.0471	25	0.0338	25	0.0179	25	0.0248	25
SARFIMA	0.0297	20	0.0214	18	0.0105	20	0.0150	20

Notes: This table reports the root MSFEs and their corresponding ranking for each forecasting horizon and weighting scheme.

Table 3
Root-mean squared forecast errors ($h=4$).

Model	Uniform		Boom		Recession		Tail	
	RMSE	Rank	RMSE	Rank	RMSE	Rank	RMSE	Rank
RW	0.0334	15	0.0239	15	0.0127	18	0.0176	16
DISC	0.0227	1	0.0174	1	0.0071	1	0.0104	1
PC	0.0276	2	0.0201	2	0.0102	4	0.0143	3
MEAN	0.0298	11	0.0220	11	0.0105	6	0.0150	6
ARIMA	0.0292	5	0.0228	14	0.0094	2	0.0133	2
ARFIMA	0.0291	4	0.0215	8	0.0105	7	0.0148	5
BARIMA	0.0348	16	0.0264	16	0.0111	13	0.0161	14
BCAR	0.0295	8	0.0213	6	0.0112	15	0.0157	13
MSAR	0.0300	12	0.0225	13	0.0099	3	0.0144	4
SETAR	0.0295	7	0.0214	7	0.0104	5	0.0151	7
LSTAR	0.0292	6	0.0212	4	0.0106	11	0.0151	9
ARANN	0.0297	9	0.0216	9	0.0110	12	0.0155	12
NPAR	0.0289	3	0.0206	3	0.0105	9	0.0152	11
SPAR	0.0297	10	0.0217	10	0.0106	10	0.0152	10
GARCH	0.0415	22	0.0324	23	0.0111	14	0.0177	17
GA	0.0310	14	0.0212	5	0.0131	19	0.0178	18
FUZZY	0.0301	13	0.0223	12	0.0105	8	0.0151	8
SRW	0.0414	21	0.0311	22	0.0136	21	0.0194	21
HW	0.0381	20	0.0294	20	0.0135	20	0.0181	20
TBATS	0.0471	23	0.0341	24	0.0162	23	0.0240	23
SARANN	0.0477	24	0.0289	19	0.0178	24	0.0290	24
SUTSEA	0.0378	19	0.0300	21	0.0120	16	0.0166	15
SUTSET	0.1087	26	0.0798	26	0.0389	26	0.0549	26
SARIMA	0.0371	17	0.0277	18	0.0126	17	0.0180	19
BSARIMA	0.0698	25	0.0487	25	0.0278	25	0.0388	25
SARFIMA	0.0376	18	0.0272	17	0.0136	22	0.0195	22

Notes: See notes to Table 2.

Table 4
Root-mean squared forecast errors ($h=12$).

Model	Uniform		Boom		Recession		Tail	
	RMSE	Rank	RMSE	Rank	RMSE	Rank	RMSE	Rank
RW	0.0577	18	0.0335	7	0.0297	23	0.0393	23
DISC	0.0400	1	0.0278	1	0.0159	1	0.0221	1
PC	0.0469	2	0.0299	2	0.0209	10	0.0289	6
MEAN	0.0483	4	0.0315	3	0.0205	7	0.0289	5
ARIMA	0.0478	3	0.0325	5	0.0189	3	0.0271	2
ARFIMA	0.0506	6	0.0322	4	0.0228	18	0.0313	18
BARIMA	0.0512	9	0.0337	9	0.0218	15	0.0304	10
BCAR	0.0521	10	0.0325	6	0.0240	21	0.0328	20
MSAR	0.0509	7	0.0344	13	0.0209	11	0.0293	7
SETAR	0.0524	12	0.0355	17	0.0209	9	0.0298	9
LSTAR	0.0522	11	0.0341	11	0.0222	17	0.0310	17
ARANN	0.0529	15	0.0337	10	0.0236	19	0.0326	19
NPAR	0.0526	14	0.0348	14	0.0215	13	0.0306	12
SPAR	0.0525	13	0.0349	16	0.0219	16	0.0307	13
GARCH	0.0598	24	0.0443	25	0.0179	2	0.0283	4
GA	0.0595	21	0.0348	15	0.0305	25	0.0402	24
FUZZY	0.0541	16	0.0366	20	0.0218	14	0.0309	14
SRW	0.0590	20	0.0417	22	0.0203	5	0.0305	11
HW	0.0587	19	0.0388	21	0.0238	20	0.0339	22
TBATS	0.0612	25	0.0363	19	0.0299	24	0.0402	25
SARANN	0.0509	8	0.0342	12	0.0212	12	0.0298	8
SUTSEA	0.0597	23	0.0421	24	0.0204	6	0.0310	16
SUTSET	0.1122	26	0.0775	26	0.0430	26	0.0616	26
SARIMA	0.0595	22	0.0417	23	0.0206	8	0.0309	15
BSARIMA	0.0554	17	0.0356	18	0.0243	22	0.0338	21
SARFIMA	0.0495	5	0.0335	8	0.0196	4	0.0281	3

Notes: See notes to Table 2.

Table 5
Root-mean squared forecast errors (h=24).

Model	Uniform		Boom		Recession		Tail	
	RMSE	Rank	RMSE	Rank	RMSE	Rank	RMSE	Rank
RW	0.0997	19	0.0544	9	0.0550	24	0.0714	23
DISC	0.0692	1	0.0454	2	0.0326	2	0.0416	1
PC	0.0823	2	0.0451	1	0.0451	18	0.0581	17
MEAN	0.0857	6	0.0536	8	0.0407	11	0.0544	11
ARIMA	0.0851	5	0.0533	6	0.0406	10	0.0539	10
ARFIMA	0.0873	7	0.0520	3	0.0426	15	0.0571	15
BARIMA	0.0848	4	0.0535	7	0.0391	9	0.0530	7
BCAR	0.0877	8	0.0520	4	0.0424	14	0.0572	16
MSAR	0.0831	3	0.0528	5	0.0374	5	0.0510	3
SETAR	0.0899	15	0.0597	19	0.0359	3	0.0514	4
LSTAR	0.0881	11	0.0559	12	0.0387	8	0.0536	8
ARANN	0.0895	14	0.0552	11	0.0408	12	0.0561	13
NPAR	0.0880	10	0.0573	15	0.0369	4	0.0520	5
SPAR	0.0879	9	0.0567	13	0.038	6	0.0529	6
GARCH	0.0947	18	0.0675	21	0.032	1	0.0481	2
GA	0.1029	20	0.0583	16	0.0549	23	0.0716	24
FUZZY	0.0926	16	0.0617	20	0.0382	7	0.0538	9
SRW	0.1202	24	0.0854	24	0.0471	20	0.0659	20
HW	0.1159	22	0.0800	22	0.0496	22	0.0669	21
TBATS	0.1064	21	0.0567	14	0.0566	25	0.0754	25
SARANN	0.0885	12	0.0548	10	0.0431	16	0.0569	14
SUTSEA	0.1211	25	0.0855	25	0.0475	21	0.0670	22
SUTSET	0.1504	26	0.1024	26	0.0610	26	0.0862	26
SARIMA	0.1166	23	0.083	23	0.0462	19	0.0644	19
BSARIMA	0.0939	17	0.0586	18	0.0435	17	0.0590	18
SARFIMA	0.0894	13	0.0586	17	0.0409	13	0.0545	12

Notes: See notes to Table 2.

We can generally infer that the relative performance of the DISC model is unaffected by specific economic conditions. Another model that seems to perform fairly well is the MSAR. This is particularly so for the shortest- (ranking 1st for recession and tail forecasts and 2nd for boom and uniform forecasts) and longest-term forecasts (with a ranking of 3rd for both uniform and tail forecasts and 5th for both boom and recession forecasts). However, for the rest of the models, the rankings in most cases differ greatly between boom and recession periods, and even at different forecast horizons. Take the GARCH model, for instance: while it seems to be the most useful model for recession forecasts, with a rank of 1 at h=24, it ranks 21st for the boom forecasts. The same model ranks 22nd and 24th for the recession and boom forecasts respectively, at h=1, 23rd and 14th at h=4 and 25th and 2nd at h=12.

If we focus on different horizons, we can easily pick out the three best models for recession or boom forecasts. For example, at the shortest-term horizon (h=1), the top three models for booms are DISC, MSAR and the PC models, in that order, while the top three models for recessions are the MSAR, DISC and PC models. At the four-month horizon, the top three models for booms are DISC, PC and NPAR, while the top three models for recessions are DISC, ARIMA and MSAR. At the 12-month horizon, the top three models for booms are DISC, PC and MEAN, while the top three models for recessions are DISC, GARCH and ARIMA. At the longest-term horizon (h=24), the top three models for booms are PC, DISC and ARFIMA, while the top three models for recessions are GARCH, DISC and SETAR. In practice, the choice of an appropriate model may depend on both the forecast horizon and on the specific loss function, since – as is shown by our results – some single models are more suited for short-term horizons, while other models perform better in the longer run. Moreover, given that the performance of the single models differs depending on the economic condition that the country is in, forecasters who particularly dislike forecast errors during recessions should use a slightly different set of models than forecasters who are more interested in a correct prediction during booming markets. This is consistent with the findings in Carstensen et al. (2010).

Next we also evaluate the forecasting models based on their RMSE relative to the benchmark RW forecast.[3] If the relative RMSE of any model is less than 1, then it outperformed the RW model. Almost all the models with seasonal dummy variables outperformed the benchmark RW model, whereas the RW model outperformed all the full seasonal models at the 1-month and 4-month horizons. This is robust to different weighting schemes. However, at 12- and 24-month horizons, both full seasonal models and models with seasonal dummy variables outperformed the RW model, especially for the recession and tail forecasts. It is also observed that the DISC-combined forecast has substantial gains over both the benchmark RW and the rest individual models. For instance, the RMSE for the DISC model is lower than the RMSE for the RW model by about 43% and 48% respectively for the boom and recession forecasts at horizon one. However, this gain reduces as one progresses to longer horizons. Looking at horizon 24, the gain relative to the RW model reduces to 26% and 41% respectively for the boom and recession forecasts. The MSAR is the best individual model at horizon one, with an improvement of 37% and 48% respectively for the boom and recession forecasts. At the 24-month horizon, the best individually performing model for the recession forecasts is GARCH, with an improvement of 42% over the RW model, whereas for the boom period, the RMSE of the former is 24% higher than that of the latter. The best performing individual model (ARFIMA) for the boom forecasts improves upon the RW model by only 4% at h=24. Overall, the performance of the models relative to the RW model differs both in terms of forecast horizon and different weighting schemes.

To evaluate whether the above findings are statistically significant, we employ the weighted version of the modified Diebold-Mariano pairwise test. The null hypothesis of the MDM test is that of equal forecast performance. The result is reported in Tables 6–7.[4] The columns with the heading "+" show the number of times a specific model significantly outperforms its competitors. The columns with heading "−" show the number of times a specific model is outperformed by its competitors. Recalling that we have 26 forecasting models, a rank of 25 is therefore the maximum a specific model can either outperform other models by, or be outperformed by other models. At the 1-month horizon, the DISC and the MSAR models significantly outperform the rest of the competing models by 24 times, and thus they are not significantly dominated by any other model. This simply implies that these two models yield significantly smaller losses than their competitors. The next well performing model is the PC model. These results are robust to the different weighting schemes. At horizon 4, the DISC model significantly outperforms the remaining 25 models, and is not outperformed by any other model, irrespective of the weighting scheme used. Following the DISC model is the PC model. A similar result holds at the 12-month horizon, with the exception being that the PC model does not perform equally well for the recession and tail forecasts. At horizon 24, the DISC model again is the best performing model. It is followed by the PC model for the boom forecasts and by the BARIMA model for the uniform and recession forecasts. The worst performing model at all horizons and weighting schemes is the SUTSET, as it never outperforms any model significantly, but is rather significantly dominated by other models.

Overall, there appears to be no single model that performs relatively better than other single models at all forecast horizons and for all weighting schemes. It is the MSAR model at horizon

[3] The relative values are essentially the ratio of each model to the RW model. We do not present the results here, but they are available upon request.

[4] We report only the rankings and show the best model in bold. The MW-DM statistics with the p-values are available from authors on request.

Table 6
Summary of modified Diebold-Mariano forecast accuracy tests ($h=1$ and $h=4$).

Model	$h=1$								$h=4$							
	Uniform		Boom		Recession		Tail		Uniform		Boom		Recession		Tail	
	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−
RW	10	14	8	14	11	12	10	9	10	14	11	12	7	13	4	13
DISC	24	0	24	0	24	0	24	0	25	0	25	0	25	0	25	0
PC	23	2	23	2	23	2	23	2	21	1	18	1	21	1	20	1
MEAN	12	3	12	3	13	3	13	3	12	2	12	2	13	3	12	3
ARIMA	12	3	12	6	12	3	15	3	12	1	12	2	18	1	22	1
ARFIMA	12	3	13	3	14	3	15	3	13	2	12	1	15	2	15	2
BARIMA	6	15	5	15	9	15	7	12	6	14	5	15	8	13	6	13
BCAR	12	3	13	3	13	4	13	3	13	2	12	1	13	5	12	4
MSAR	24	0	24	0	24	0	24	0	12	2	11	2	13	1	12	1
SETAR	12	3	12	3	12	3	12	3	12	2	12	1	13	1	12	2
LSTAR	12	3	13	3	16	3	14	3	14	1	14	1	15	2	12	3
ARANN	12	3	12	3	12	5	12	6	12	2	12	1	13	5	12	4
NPAR	12	3	13	3	12	3	11	3	13	1	12	1	13	2	12	3
SPAR	12	3	12	4	12	3	11	3	12	2	12	3	13	3	12	3
GARCH	2	21	2	22	2	21	2	20	3	16	2	18	3	14	3	13
GA	12	3	12	3	11	7	10	9	12	6	12	2	7	13	5	13
FUZZY	12	3	12	3	11	4	10	5	12	3	11	3	13	3	12	3
SRW	3	19	3	18	3	20	2	20	3	19	2	18	3	18	3	16
HW	5	15	4	14	6	16	7	15	3	15	3	15	3	16	4	13
TBATS	2	20	2	21	2	20	2	20	2	22	2	20	2	22	2	22
SARANN	2	17	4	16	2	19	2	19	2	16	2	15	2	18	2	19
SUTSEA	4	15	4	16	6	16	7	15	4	15	3	15	5	13	5	4
SUTSET	0	25	0	25	0	25	0	25	0	25	0	25	0	25	0	25
SARIMA	6	14	7	14	7	15	7	15	4	15	5	15	5	13	5	13
BSARIMA	1	24	1	24	1	24	1	24	1	24	1	24	1	24	1	24
SARFIMA	5	15	5	14	5	17	5	19	4	15	5	15	3	16	3	15

Notes: The columns headed by "+" indicate the number of times a specific model significantly outperforms its competitors. The columns headed by "−" indicate the number of times a specific model is outperformed by its competitors.

Table 7
Summary of modified Diebold-Mariano forecast accuracy tests ($h=12$ and $h=24$).

Model	$h=12$								$h=24$							
	Uniform		Boom		Recession		Tail		Uniform		Boom		Recession		Tail	
	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−
RW	1	9	6	2	1	17	1	16	1	3	5	2	1	18	1	18
DISC	25	0	25	0	25	0	25	0	25	0	24	0	25	0	25	0
PC	22	1	22	1	10	1	9	1	12	1	24	0	8	1	5	1
MEAN	21	1	19	1	15	1	13	1	11	1	12	2	10	1	10	1
ARIMA	15	1	10	1	19	1	17	1	10	1	11	2	10	1	10	1
ARFIMA	13	3	15	2	8	5	7	3	8	1	16	2	7	3	5	4
BARIMA	10	3	8	3	8	3	7	1	13	1	15	2	13	1	10	1
BCAR	7	4	14	2	4	8	4	9	8	2	15	2	5	3	5	4
MSAR	7	3	6	3	8	2	8	1	11	1	10	2	12	1	9	1
SETAR	5	3	6	6	8	1	9	1	8	1	6	12	10	1	6	1
LSTAR	6	3	9	5	6	3	6	3	9	1	8	4	10	1	7	2
ARANN	4	5	8	5	4	13	4	12	7	3	8	5	5	6	5	6
NPAR	5	4	5	5	7	3	5	4	9	1	8	5	11	1	8	1
SPAR	5	4	7	5	7	3	6	3	10	1	8	5	10	1	10	1
GARCH	1	16	1	18	3	1	3	1	5	6	5	16	8	1	8	1
GA	1	15	1	5	1	18	1	18	1	14	5	6	1	18	0	18
FUZZY	3	8	2	11	4	5	4	4	6	10	6	13	8	3	5	2
SRW	1	9	1	17	2	3	1	2	1	18	1	21	1	15	1	5
HW	1	13	1	11	2	12	1	8	1	18	1	21	1	18	1	17
TBATS	1	17	1	14	1	21	1	18	1	17	5	9	0	18	0	18
SARANN	10	4	7	3	7	4	8	3	8	3	10	2	8	4	5	4
SUTSEA	1	8	1	17	1	2	1	2	1	18	1	21	1	14	1	5
SUTSET	0	25	0	25	0	25	0	25	0	25	0	25	0	24	0	23
SARIMA	1	8	1	17	2	4	1	4	1	18	1	21	1	15	1	5
BSARIMA	2	10	5	7	4	13	3	13	6	11	5	9	5	9	4	9
SARFIMA	11	2	8	3	13	1	10	1	7	4	6	7	8	1	8	1

Notes: The columns headed by "+" indicate the number of times a specific model significantly outperforms its competitors. The columns headed by "−" indicate the number of times a specific model is outperformed by its competitors.

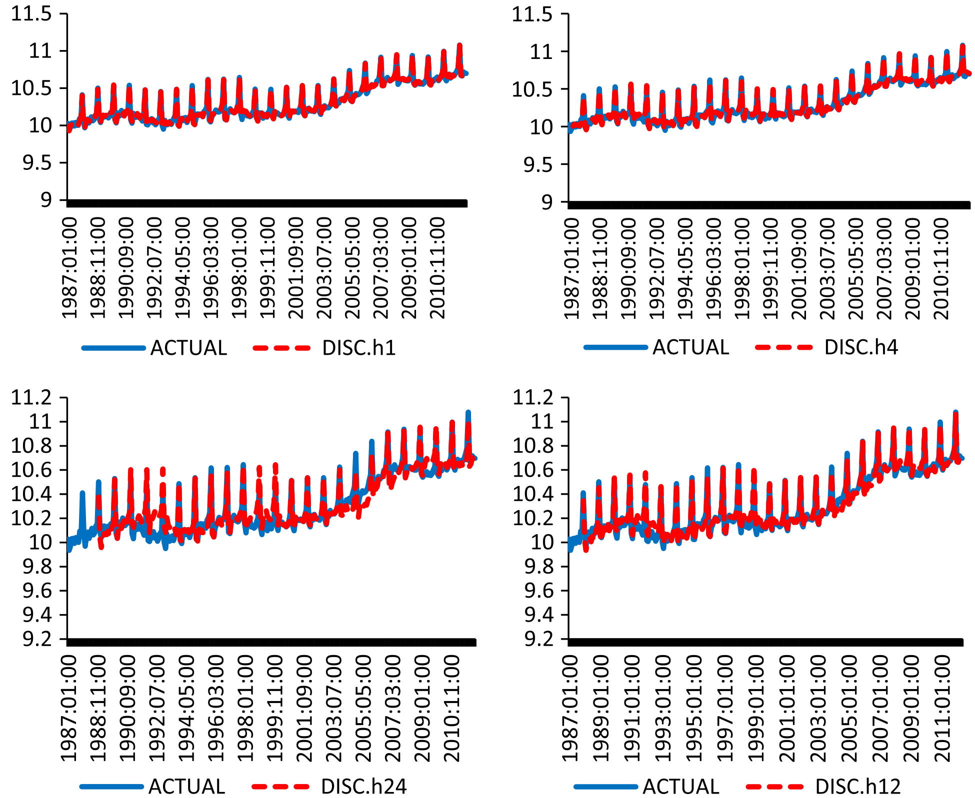


Fig. 3. Actual and DISC forecast of retail sales at different horizons for the uniform weighting scheme. *Note*: DISC.h1, DISC.h4, DISC.h12 and DISC.h24 represents forecasts from the discounted forecast combination model at horizons 1, 4, 12 and 24, respectively.

1 for all weighting schemes, the ARFIMA model for recession and tails forecasts and the LSTAR model for uniform and boom and also recession forecasts at horizon 4. At horizon 12, it is the ARIMA model for recession and the ARFIMA model for boom forecasts. At horizon 24, it is the ARFIMA model for the boom and the BARIMA model for the uniform and recession forecasts. However, the combination forecast, especially the DISC model forecast, is the best at all horizons, no matter which weighting scheme is employed.

These findings confirm the superiority of combined forecasts over individual forecasts for forecasting South Africa's aggregate retail sales. This is confirmed graphically in Fig. 3, where we plot the actuals and the forecasts at different horizons from the discounted MSFE combination method under the uniform loss weighting scheme.[5] As can be seen, the forecasts closely track the actuals for all the horizons considered.

Thus, academics and practitioners should not depend on a single forecasting model, as no single model uniformly dominates others, but at the same time, they should have a wide array of models, rather than just a few for the sake of efficiency gains, since for a series as volatile as retail sales, different models tend to capture different aspects of the unknown data-generating process that defines the structure of retail sales. This fact is highlighted by the better performance of the principal component forecast combination approach relative to that of the simple mean-based forecast combination method. The principal component approach uses a statistical procedure to provide more weight to the forecasts of those models for which the forecasts tend to move closely. This is likely to be the case for models which produce better forecasts, since otherwise this approach would not have performed better than the mean-based combination approach.

However, the outstanding performance of the discounted MSFE combination method highlights the importance of using forecast information sets that are more recent. That we needed to discount the past quite strongly (using a discount factor of 0.50), from an empirical point of view, is probably because the series is characterized by volatility and strong seasonal patterns, to the extent that forecast information from the distant past is of little value. However, from a theoretical perspective, realizing that the retail sales series that we are working with comprises realized values (hence are market clearing or equilibrium retail sales) is possibly an indication of learning by retailers. In other words, retailers tend to update their information sets, valuing current information more than that of the recent past, to better predict the future, realizing that economic conditions and demand are continuously changing over time, and that supply decisions need to change accordingly to accommodate such dynamism. At this stage, it is important to point out that – unlike previous studies (see Rapach and Strauss, 2010 for a detailed discussion) that used combined forecasts when forecasting various variables of interest, and that found the simple mean-based forecast combination method often to perform better than more sophisticated combination methods, like the principal components and discounted MSFE approaches – we find the relatively more statistically rigorous discounted MSFE to be the standout forecast combination approach. This, we believe, could emanate from the theoretical reasons involving learning, discussed above.

The nonlinearity, especially of a parametric nature, seems to be important and should always be taken into account. While the MSAR model tends to perform better relative to the other forms of nonlinear models over short horizons, the LSTAR model tends to stand out over longer horizons. While, as discussed earlier, for the

[5] See Appendix A for similar plots for the best single models.

MSAR model, the regime-switching is based on a latent regime variable that follows a first-order Markov process, therefore having unexplained switching, the LSTAR model has explained switching; hence, regime-switching follows a known structure, with a smooth transition in and out of a regime that spreads to more than one period. Theoretically this result is important for us, as it more than likely suggests the development of dynamic stochastic general equilibrium models to capture the nonlinear dynamics of equilibrium retail sales, with the parameters of the model defining consumer and producer choices being state-dependent, to account for regime-switching behavior across alternative states of the economy. Lastly, from an empirical perspective, attempting to model the seasonal variation with models that have complicated seasonal features does not seem to be worth the effort, as a lots of the nonlinearity due to the seasonal behavior already seems to have been captured by the nonlinear, but non-seasonal, models.

5. Conclusion

In this paper, we assess the performance of 26 models in forecasting South Africa's aggregate seasonal retail sales over 1987:01–2012:05, the out-of-sample period. The recent recession has demonstrated that a good forecast of a rather extreme event might be of special interest beyond that of minimizing an average squared error. Hence, we allowed for departures from the uniform symmetric quadratic loss function, typically used in forecast evaluation exercises. We overweighted forecast errors during periods of high or low growth rates to check how the indicators perform during booms and recessions, i.e., in times of particularly high demand for good forecasts. Specifically, we use Van Dijk and Franses (2003) weighted MSFE and weighted modified MDM tests to evaluate the forecasts of the 26 different forecasting models. We estimated two broad classes of modes: 17 models with seasonal dummy variables and 9 full seasonal models.

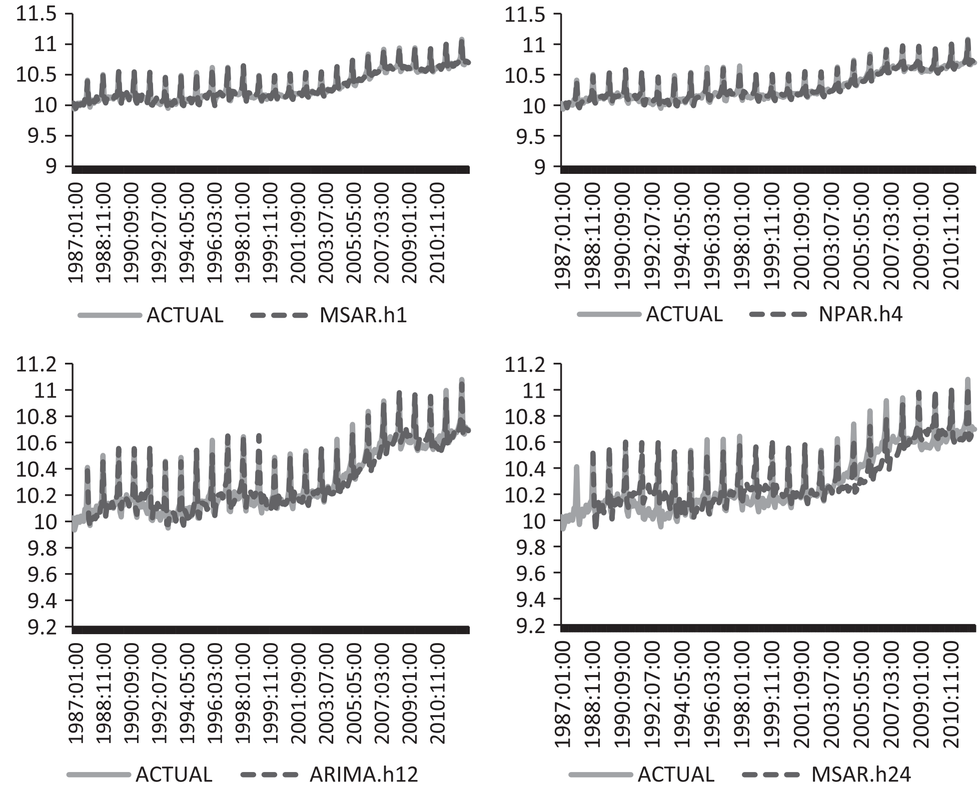
In general, the models with seasonal dummy variables produce better forecasts than the full seasonal models. Most of the models performed better than the random walk benchmark. The most widely used nonlinear model in the retail sales forecasting literature, namely, the ANN model, is consistently outperformed by other types of nonlinear models that are more strongly theoretically founded. From the analysis, it is difficult to identify a specific individual model as the best for forecasting South Africa's aggregate retail sales. Some single models are well suited for booms, while others are well suited for recessions, and this differs across forecast horizons. However, the combination forecasts offer ways of incorporating and culling information from a larger number of forecasting models. This group of models ultimately outperforms the individual models, in general. Specifically, the discounted combination forecast model (DISC) outperforms all the single models, and the other two combination forecast (simple mean and principal component) models, and its performance is largely unaffected by specific economic or business cycle situations or forecast horizons.

The findings of this study demonstrate the need to include a wider array of linear and nonlinear models than just few models when forecasting retails sales, since each model captures different aspects of the series at different forecasting horizons. Moreover, it is also important to consider forecasting during both periods of booms and recessions, since the predictability of the single models for periods of calm is vastly different than for periods of turmoil. Moreover, identifying the best model for forecasting retail sales – which happens to be a combined model, in this case, based on a discounted MSFE – helps not only to provide accurate forecasts over short to long horizons but also less variable forecasts across extreme periods. This should, in turn, enable businesses and investors to manage their inventories efficiently, to make proper business plans and strategies, and also to make optimal portfolio resource allocation decisions, all of which are likely to affect profitability. Further, government officials would require such accurate aggregate retail sales forecasts in designing and implementing optimal public policy for the retail industry, which will subsequently benefit both consumers and businesses. The fact that better retail sales forecasting gives an indication of the path of consumption and also is a vital pre-inflationary indicator has implications for policy makers and investors. For instance, if the growth of retail sales is stalled or slowing, it could signal a recession, because of the significant role personal consumption plays in the growth of the economy. More so, a sudden rise in retail sales in the midst of a business cycle, would cause the Reserve Bank to increase the interest rate at least in the short term, to curtail any possible inflation. This might negatively affect financial assets such as stocks and bonds, as well as investors' future cash flows. Therefore, accurate forecasts of, and timely policy intervention in the retail industry, based on the best econometric model, is crucial for economic growth and stability.

This paper primarily adds empirically to the literature on forecasting retail sales and suggests possible theoretical models that need to be developed to capture appropriately the dynamics of equilibrium retail sales, based on the results obtained. From an applied perspective, contrary to the existing literature on retail sales forecasting, we not only look at a wider array of linear and non-linear models, but also conduct multi-step-ahead forecasts based on a real-time (recursive) scenario, as is likely to be faced by the agents. In addition, we look at the performance of the empirical models across different phases of the retail sales growth equilibrium. Our results indicate the outstanding performance of the discounted MSFE combination methods and the importance of nonlinearity. They also highlight the need for developing dynamic stochastic general equilibrium models for retail sales, which incorporate not only learning behavior but also provides frameworks that allow the behavioral parameters of the model to be state-dependent, to account for regime-switching behavior across alternative states of the economy. Nevertheless, the current paper adopts a univariate approach, with retail sales being predicted by their past values only, either in a linear or nonlinear fashion. However, as indicated by Dias et al. (2010), retail sales are likely to be affected by a large number of economic variables. Hence, future research could aim at using linear and nonlinear models for forecasting retail sales. This could involve using macroeconomic and financial variables as possible predictors, and in turn, comparing the results with the various univariate models discussed in this paper. The findings would then allow us to strengthen our theoretical conclusions, which – to some extent at this stage – are quite conjectural in nature.

Appendix A. Actual and forecast of retail sales at different horizons for the uniform weighting scheme for the best single models

See Fig. Appendix A.



Appendix A. Actual and forecast of retail sales at different horizons for the uniform weighting scheme for the best single models.

References

Aburto, L., Weber, R., 2007. Improved supply chain management based on hybrid demand forecasts. Appl. Soft Comput. 7 (1), 126–144.

Agrawal, D., Schorling, C., 1996. Market share forecasting: an empirical comparison of artificial neural networks and multinomial logic model. J. Retail. 72 (4), 383–408.

Ali, Ö.G., Sayin, S., Woensel, T.V., Fransoo, J., 2009. SKU demand forecasting in the presence of promotions. Expert Syst. Appl. 36 (10), 12340–12348.

Alon, I., 1997. Forecasting aggregate retail sales: the Winters' model revisited. In: Goodale, J.C. (Ed.), The 1997 Annual Proceedings. Midwest Decision Science Institute, pp. 234–236.

Alon, I., Qi, M., Sadowski, R.J., 2001. Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods. J. Retail. Consum Serv 8 (3), 147–156.

Au, K.F., Choi, T.M., Yu, Y., 2008. Fashion retail forecasting by evolutionary neural networks. Int. J. Prod. Econ. 114 (2), 615–630.

Aye, G.C., Dua, P., Gupta, R., 2014. Forecasting Indian macroeconomic variables using medium-scale var models. In: Dey, D.K., Loganathan, A., Singh, U., Upadhyay, S.K. (Eds.), Current Trends in Bayesian Methodology with Applications (forthcoming).

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Bates, J.M., Granger, C.W.J., 1969. The combination of forecasts. Oper. Res. Q. 20, 451–468.

Barksdale, H.C., Hilliard, J.E., 1975. A cross-spectral analysis of retail inventories and sales. J. Bus. 48 (3), 365–382.

Box, G.E.P., Jenkins, G.M., 1976. Time series analysis: Forecasting and Control. Holden Day, San Francisco.

Carbone, R., Armstrong, J.S., 1982. Evaluation of extrapolative forecasting methods: results of a survey of academicians and practitioners. J. Forecast. 1, 215–217.

Carstensen, K., Wohlrabe, K., Ziegler, C., Predictive ability of business cycle indicators under test: A case study for the Euro area industrial production. CESifo working paper Empirical and Theoretical Methods, No. 3158. http://www.econstor.eu/bitstream/10419/39840/1/634266772.pdf (accessed 08.06.12).

Case, K.E., Quigley, J.M., Shiller, R.J., 2005. Comparing wealth effects: the stock market versus the housing market. Adv. Macroecon. 5 (1), 1–34.

Case, K.E., Quigley, J.M., Shiller, R.J., 2012. Wealth effects revisited 1975–2012. Crit. Finance Rev. 2, 101–128.

Chan, Y.L., Stock, J.H., Watson, M.W., 1999. A dynamic factor model framework for forecast combination. Span. Econ. Rev. 1, 91–121.

Chang, P.C., Wang, Y.W., 2006. Fuzzy Delphi and back-propagation model for sales forecasting in PCB industry. Expert Syst. Appl. 30 (4), 715–726.

Chatfield, C., 1992. A commentary on error measures. Int. J. Forecast. 8, 100–102.

Chen, F.L., Ou, T.Y., 2009. Grey relation analysis and multilayer function link network sales forecasting model for perishable food in convenience store. Expert Syst. Appl. 36 (3), 7054–7063.

Chen, F.L., Ou, T.Y., 2011a. Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry. Expert Syst. Appl. 38, 1336–1345.

Chen, F.L., Ou, T.Y., 2011b. Constructing a sales forecasting model by integrating GRA and ELM: a case study for retail industry. Int. J. Electron. Bus. Manag. 9 (2), 107–121.

Chu, C.W., Zhang, P.G.Q., 2003. A comparative study of linear and nonlinear models for aggregate retail sales forecasting. Int. J. Prod. Econ. 86, 217–231.

Clemen, R.T., 1989. Combining forecasts: a review and annotated bibliography. Int. J. Forecast. 5, 559–583.

Dias, F., Pinheiro, M., Rua, A., 2010. Forecasting using targeted diffusion indexes. J. Forecast. 29, 341–352.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econ. Statist. 13 (3), 253–263.

Doganis, P., Alexandridis, A., Patrinos, P., Sarimveis, H., 2006. Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. J. Food Eng. 75 (2), 196–204.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized factor model: identification and estimation. Rev. Econ. Statist. 82, 540–554.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2003. Do financial variables help forecasting inflation and real activity in the euro area? J. Monet. Econ. 50, 1243–1255.

Frank, C., Garg, A., Sztandera, L., Raheja, A., 2003. Forecasting women's apparel sales using mathematical modeling. Int. J. Cloth. Sci. Technol. 15 (2), 107–125.

Garrett, T.A., Hernàndez-Murillo, R., Owyang, M.T., 2004. Does consumer sentiment predict regional consumption? Fed. Reserve Bank St. Louis Rev. 87 (2), 123–135.

Gauteng Province: Provincial Treasury Quarterly Bulletin; Gauteng Province 2012. The Retail Industry on the Rise in South Africa. Economic Analysis Unit, April–June, 2012. ⟨http://www.treasury.gpg.gov.za/Document/Documents/QB1%20The%20Retail%20Industry%20on%20the%20Rise.pdf⟩ (accessed 15.10.12).

Gil-Alana, L.A., Barros, C.P., Assaf, A., 2010. Retail sales persistence in the short term and long-term dynamics. Working Paper No. 03/10. January 2010, Faculty of Economics, Universidad de Navarra, Spain (accessed 10.06.12).

Hansen, J.V., McDonald, J.B., Nelson, R.D., 1999. Time series prediction with genetic-algorithm designed neural networks: An empirical comparison with modern statistical models. Comput. Intell. 15 (3), 171–184.

Harvey, David I., Leybourne, Stephen J., Newbold, Paul, 1997. Testing the equality of prediction mean squared errors. Int. J. Forecast. 13, 281–291.

Harvey, A., 2006. Forecasting with Unobserved Components Time Series Models. In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, 1. Faculty of Economics, University of Cambridge, pp. 327–412.

Hyndman, R.J., Koehler, A.B., Snyder, R.D., Grose, S., 2002. A state space framework for automatic forecasting using exponential smoothing methods. Int. J. Forecast. 18 (3), 439–454.
IHS Global Insight Regional Explorer-Encyclopedia (Ver 2.04), 2012.
Joseph, A., Singh, E., Larrain, M., 2007. Forecasting aggregate sales with interest rates using multiple neural network architectures. In: Dagli, C.H. (Ed.), Part IV: Financial Engineering and Data Mining from: Intelligent Engineering Systems Through Artificial Neural Networks, p. 17.
Kearney, A.T., 2011. Retail global expansion: a portfolio of opportunities-NSGlobal Retail Development Index. ⟨http://www.atkearney.com/documents/10192/3903b4b7-265c-484e-932c-50169e5aa2f3⟩ (accessed 10.06.12).
Kearney, A.T. (2012). Global retail expansion- keep on moving-Global Retail Development Index.⟨http://www.atkearney.com/documents/10192/302703/Global+Retail+Expansion+Keeps+On+Moving.pdf/4799f4e6-b20b-4605-9aa8-3ef451098f8a⟩ (accessed 10.06.12).
Makridakis, S., 1989. Why combining works? Int. J. Forecast. 5 (4), 601–603.
Makridakis, S., Winkler, R.L., 1983. Averages of forecasts: Some empirical results. Manag. Sci. 29, 987–996.
Mentzer, J.T., Bienstock, C.C., 1998. Sales Forecasting Management: Understanding the Techniques, Systems and Management of the Sales Forecasting Process. Sage Publications, Thousand Oaks, CA.
Montaño, J.J., Palmer, A., 2003. Numeric sensitivity analysis applied to feed-forward neural networks. Neural Comput. Appl. 12, 119–125.
Moreno, J.J.M., Alfonso Palmer Pol, A.P., Gracia, P.M., 2011. Artificial neural networks applied to forecasting time series. Psicothema 23 (2), 322–329.
Ni, Y., Fan, F., 2011. A two-stage dynamic sales forecasting model for the fashion retail. Expert Syst. Appl., 1529–1536.
Palmer, A., Montaño, J.J., Franconetti, F.J., 2008. Sensitivity analysis applied to artificial neural networks for forecasting time series. Methodology 4, 80–84.
Peña-Reyes, C.A., 2004. Coevolutionary Fuzzy Modeling. In: Peña-Reyes, C.A. (Ed.), Lecture Notes in Computer Science, 3204; 2004, p. XIV.
Peterson, R.T., 1993. Forecasting practices in the retail industry. J. Bus. Forecast. 12, 11–14.
Rapach, D.E., Strauss, J.K., 2010. Bagging or combining (or both)? An analysis based on forecasting U.S. Employment growth. Econom. Rev. 29 (5–6), 511–533.
RNCOS, 2011. South African retail industry forecast to 2013. September 1, 2011. ⟨http://www.rncos.com/Report/IM275.htm⟩ (accessed 20.06.12).
Szpiro, G.G., 1997. Forecasting chaotic time series with genetic algorithms. Phys. Rev. E 55 (3), 2557–2568.
Stine, R.A., Shaman, P., 1989. A fixed point characterization for bias of autoregressive estimators. Ann. Stat. 17, 1275–1284.
Stock, J.H., Watson, M.W., 1999. Forecasting inflation. J. Monet. Econ. 44, 293–335.
Stock, J.H., Watson, M.W., 2002. Macroeconomic forecasting using diffusion indexes. J. Bus. Econ. Stat. 20, 147–162.
Stock, J.H., Watson, M.W., 2003. Forecasting output growth and inflation: the role of asset prices. J. Econ. Lit. 41, 788–829.
Stock, J.H., Watson, M.W., 2004. Combination forecasts of output growth in a seven-country data set. J. Forecast. 23, 405–430.
Sun, Z.L., Choi, T.M., Au, K.F., Yu, Y., 2008. Sales forecasting using extreme learning machine with applications in fashion retailing. Decis. Support Syst. 46 (1), 411–419.
Thall, N., 1992. Neural forecasts: a retail sales booster. Discount Merch. 23 (10), 41–42.
Taskaya-Temizel, T., Casey, M.C., 2005. Configuration of neural networks for the analysis of seasonal time series. Pattern Recognit. Data Min., Lect Notes Comput. Sci. 3686, 297–304.
Van Dijk, D., Franses, P.H., 2003. Selecting a nonlinear time series model using weighted tests of equal forecast accuracy. Oxf. Bull. Econ. Stat. 65, 727–744.
Wan, L., Luo, B., Ji, H.-M., Yan, W.-W., 2012. A type of combination forecasting method based on time series method and PLS. Am. J. Oper. Res. 2012 (2), 467–472.
Xiao, T.J., Qi, X.T., 2008. Price competition, cost, and demand disruption and coordination of a supply chain with one manufacturer and two competing retails. Omega 36, 741–753.
Zellner, A., 1986. A tale of forecasting 1001 series: the Bayesian knight strikes again. Int. J. Forecast. 2, 491–494.
Zhang, G.P., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159–175.
Zhang, G.P., 2009. Neural networks for retail sales forecasting. category: business information systems, 2806–2810, IGI Global. ⟨http://www.irma-international.org/viewtitle/13986/⟩ (accessed 20.06.12).
Zhang, G.P., Kline, D.M., 2007. Quarterly time-series forecasting with neural networks. IEEE Trans.Neural Netw. 18 (6), 1800–1814.
Zhou, X., Carroll, C.D., 2012. Dynamics of wealth and consumption: new and improved measures for U.S. States. B.E. J. Macroecon. 12 (2), 1–42.