University of Pretoria
Department of Economics Working Paper Series

Forecasting Aggregate Retail Sales: The Case of South Africa

Goodness C. Aye
University of Pretoria
Mehmet Balcilar
Eastern Mediterranean University
Rangan Gupta
University of Pretoria
Anandamayee Majumdar
University of Pretoria
Working Paper: 2013-12
February 2013

Department of Economics
University of Pretoria
0002, Pretoria
South Africa
Tel: +27 12 420 2413

Forecasting Aggregate Retail Sales: The Case of South Africa

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

Abstract

Forecasting aggregate retail sales may improve portfolio investors' ability to predict movements in the stock prices of the retailing chains. Therefore, 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. We deviate from the uniform symmetric quadratic loss function typically used in forecast evaluation exercises. Hence, we consider loss functions that overweight forecast error in booms and recessions to check whether a specific model that appears to be a good choice on average is also preferable in times of economic stress. To this end, we use the weighted RMSE and weighted version of the Diebold-Mariano tests to evaluate the different forecasts. Focussing on the single models alone, results show that their performances differ greatly across forecast horizons and for different weighting schemes. However, the combination forecasts models in general produced better forecasts and are largely unaffected by business cycles and time horizons.

Key Words: seasonality, weighted loss, retail sales forecasting, combination forecasts, South Africa

JEL Classification: C32, C53, E32

1. Introduction

The retail industry in South Africa is classified under the tertiary sector and falls within the wholesale and retail sub-sector (also known as the trade sub-sector). In 2011, the tertiary sector contributed 69.1 percent to the country's economy. The wholesale and retail trade sub-sector contributed approximately 13.7 percent to the economy. The retail trade and repairs of goods made the largest contribution (45 percent) within the wholesale and retail trade sub-sector (IHS, Global Insight, 2012; Gauteng Province: Provincial Treasury Quarterly Bulletin, 2012). This indicates that the retail industry drives the performance of the trade sub-sector. The retail industry contributes about 5.7 percent of total GDP. The retail industry is among the top industries in the country in terms of the share of employed labour force. The industry's share of employment to 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. This placed the retail industry as the fifth largest employer in the country. At first place was, households at 10.5 percent, followed by other business activities at 10.1 percent. Third place was held by education, which accounted for 7.5 percent of total employment and in fourth place was public administration and defence which accounted for 7.2 percent (IHS, Global Insight, 2012; Gauteng Province: Provincial Treasury Quarterly Bulletin, 2012). The South Africa retail industry is one of largest retail industry in the Sub Saharan region that presents profitable investment

[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] Corresponding author. Department of Economics, University of Pretoria, Pretoria, 0002, South Africa, Email: rangan.gupta@up.ac.za.
[d] Soochow University Center for Advance Statistics and Econometric Research, Suzhou, China.

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 to a greater urgency for retailers to enter the specific country. The GRDI scores are based on country and business risk, market attractiveness, market saturation and time pressure variables. In 2011, South Africa was ranked 26th out of 30 developing countries with a score of 42.2, a deterioration from the 24th rank of 2010 (41.7). At the top of the rankings was Brazil with a GRDI score of 71.5. Focusing on the individual components of GRDI, South Africa scored 46.9 percent on market attractiveness, 89.3 on country and business risk, 15.2 on market saturation and 17.2 on time pressure. However, South Africa dropped out from the 2012 rankings because of market saturation of international retailers compared to other countries in the GRDI (Kearney, 2011, 2012). These statistics indicate the role of the retail industry in South Africa.

The management of retail sales is of paramount importance to retail organisations and retail policy makers. Due to competition and globalization, sales forecasting plays a prominent role as part of the commercial enterprise (Xiao and Qi, 2008). Most retailers are constantly struggling to reduce their cost and increase profits. An accurate sales forecasting system is an efficient way to achieve these goals as reliable prediction of sales can improve the quality of business strategy. Forecasting of the future demand is central to the planning and operation of retail business at both macro and micro levels. At the organizational level, forecasts of sales 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 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 labour force, as well as after sales services (Zhang, 2009). Therefore, the ability of retailing managers to estimate the probable sales quantity in the next period, can lead to improved customers' satisfaction, reduced destruction of products, increased sales revenue and more effective and efficient production plan (Chen and Ou, 2011a, 2011b). Finally, retail trade sales are believed to be a very close proxy for consumption expenditures in a country, which in turn, is the most dominant component of the GDP. For a country like South Africa, where consumption data is available only at the quarterly frequency, forecasts of the retail sales at monthly frequency can give the policy makers an idea about the tentative path of consumption (even before and without data being available for the same at that specific point in time), and hence the GDP of the economy.

Given the critical role of retail sales and the importance of its forecasting, this study is set out to forecast South Africa's retail sales. Specifically, we focus on forecasting aggregate seasonal retail sales. Industry forecasts are especially useful to big retailers who may have a greater market share (Alon et al., 2001). For the retailing industry, Peterson (1993) showed that large retails are more likely to use time-series methods and prepare industry forecasts, while small retails emphasize judgmental methods and company forecasts. Better forecasts of aggregate retail sales can improve the forecasts of individual retailers because changes in their sales levels are often systematic (Peterson, 1993). More accurate forecasts of aggregate retail sales may improve portfolio investors' ability to predict movements in the stock prices of retailing chains (Barksdale and Hilliard, 1975; Thall; 1992; Alon et al., 2001). However, poor forecasting would result in redundant or insufficient stock that will directly affect the revenue and competitive position (Agrawal and Schorling, 1996).

Improving the quality of forecasts is still an outstanding question (Granger, 1996). In particular, retail sales data present strong seasonal variations. Forecasting of time series that have seasonal variations remains an important problem for forecasters. How to best deal with seasonal time series and which seasonal model is the most appropriate for a given time series are still largely unsolved (Zhang and Cline, 2007). Historically, modelling and forecasting seasonal data is one of the major research efforts and many theoretical and heuristic methods have been developed in the last several decades. The available traditional quantitative approaches include heuristic methods such as time-series decomposition and exponential smoothing as well as time-series regression and autoregressive and integrated moving average (ARIMA) models that have formal statistical foundations (Chu and Zhang, 2003). Nevertheless, their forecasting ability is limited by their assumption of a linear behaviour and thus, it is not always satisfactory (Zhang, 2003). Soft computing methods such as fuzzy logic, artificial neural networks (ANNs), and genetic algorithms offer an alternative, taking into account both endogenous and exogenous variables and allowing arbitrary non-linear approximation functions derived (learned) directly from the data. ANN models are the most commonly used non-linear forecasting models. As pointed out by Moreno et al. (2011), ANNs are not without limitations and criticisms. ANNs lack a theoretical foundation and a systematic procedure for the construction of the model, comparable to the classical approximations such as the Box-Jenkins methodology (Box and Jenkins, 1976). As a result, the construction phase of the model involves the experimental selection of a wide number of parameters by trial and error. According to Moreno et al. (2011), the most criticised aspect in the use of ANN focuses on the study of the effect and significance of the input variables of the model, due to the fact that the value of the parameters obtained by the network does not possess a practical interpretation, unlike classical statistical models. As a consequence, ANN have been presented to the user as a 'black box' as it is not possible to analyze the role played by each of the input variables in the forecast carried out. However, attempts are being made to overcome these criticisms (Hansen, et al., 1999; Montaño and Palmer, 2003; Palmer et al., 2008).

From the foregoing, it is obvious that single forecasting models all have their own characteristics, strengths and weaknesses. Further, when employing a single model, only a certain point of the effective information can be used, showing that the range of information sources is insufficient (Wan et al., 2012). Single model will also be affected by the model's set conditions and other factors. These factors may deteriorate the accuracy of individual forecasting methods and increase the size of errors. Combining the 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 forecasting accuracy and reduces the variance of post-sample forecasting errors (Makridakis and Winkler, 1983) and this holds true in statistical forecasting, judgmental estimates and when averaging statistical and subjective predictions (Clemen, 1989). However, it is important to also note that, Bates and Granger (1969) and more recently Kapetanios et.al. (2008) observed that the combination forecasting does not always necessarily lead to a better forecasting performance. Banternghansa and McCracken (2010) also added that averaging approach should be used and interpreted with caution whereas "past model performance does not always ensure future model performance".

There are quite a number of forecast combining methods, some simple, others are sophisticated.[1] The theory of combination forecasting 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). However, empirical evidence has shown that simple forecast combining methods often outperform more complex methods.

Against this background, this study uses twenty three single and three combination forecasting approaches to forecast South Africa's aggregate seasonal retail sales. There exist quite a large number of studies on retail sales forecasting.[2] However, majority of these studies focus on individual retail sales forecasting. Despite the importance of retail sales and its forecasting and the large number of studies internationally, we are not aware of any study in South Africa on retail sales forecast. This study therefore contributes to the literature by forecasting South Africa's retail sales, given our knowledge of the economic structure of the economy. 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). In order to have a comprehensive picture of the forecasting performance of the twenty six different models, the forecast accuracy of each model over the 1987:01–2012:05 out-of-sample period is evaluated using the root mean square error (RMSE). However, 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 deviate from previous studies on retail sales forecasting by considering loss functions that overweight the forecast errors in either booms or recessions or both. We use van Dijk and Franses (2003) weighted mean square error to evaluate forecasts from different models. In addition, the Harvey et al. (1997) modified Diebold Mariano (MDM) test is used to evaluate whether the average loss differences between two models is significantly different from zero. For the MDM test, a weighted version proposed by van Dijk and Franses (2003) and adopted by Carstensen et al. (2010) is employed.

Our contributions are four folds. First, this is the first study on retail (aggregate or individual) sales forecasting in South Africa. Second, we are not aware of any study elsewhere that have applied large number (26) of seasonal forecasting models in the context of retail sales as done in this study. Third, this is the first study on retail forecasting that consider different weighting schemes for the standard loss function. Fourth, we employ three alternative forecast combination methods and we compare results from the 23 single models with the composite models.

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

[1]For forecast combining methods, see Chan et al., 1999, Stock and Watson (2004), Rapach and Strauss (2010).

[2] See the literature subsection for a review of available studies on forecasting of retail (aggregate and individual) sales.

2. Literature

In this section, we provide a review of empirical studies on retail (aggregate and individual) sales forecasting with a view to make clear the contribution of the current study. Alon (1997) found that the Winters' exponential smoothing model forecasts aggregate retail sales more accurately than the simple exponential and Holt's models. Alon et al. (2001) and Chu and Zhang (2003) investigated the forecasting properties of artificial neural networks (ANN), Winters exponential smoothing, ARIMA models and multivariate regression, applied to aggregate retail sales and their results suggested that the ANN methods produce the best results. Frank et al. (2003) forecast women's apparel sales using three different forecasting models namely single seasonal exponential smoothing, Winters' three parameter model, and ANNs. Their result indicates that ANN model outperform the other two models based on R^2 evaluation. Doganis et al. (2006) presented an evolutionary sales forecasting model which is a combination of two artificial intelligence technologies, namely the radial basis function and genetic algorithm. The methodology is applied successfully to sales data of fresh milk provided by a major manufacturing company of daily product. Chang and Wang (2006) integrated fuzzy logic and artificial neural network into the fuzzy back-propagation network (FBPN) for sales forecasting in Printed Circuit Board (PCB) industry. The results from FBPN are compared to those of Grey Forecasting (GF), Multiple Regression Analysis (MRA) and Back-propagation network (BPN). The experimental results indicate that the Fuzzy back-propagation approach outperforms other three different forecasting models in Mean Absolute Percentage Error (MAPE) measures.

Aburto and Weber (2007) presented a hybrid intelligent system combining ARIMA model and MLP neural networks for demand forecasting. It shows improvements in forecasting accuracy and a replenishment system for a Chilean supermarket, which leads simultaneously to fewer sales and lower inventory levels. Joseph et al. (2007) examine out-of-sample forecasts of aggregate sales using 3-month treasury bills interest rate in NeuroSolutions environment referenced against 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 out-perform the linear regression models. Au et al. (2008) illustrated 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 can converge faster and more accurate in forecasting for time series than the fully connected neuron network and traditional SARIMA model. Sun et al. (2008) also developed different sales forecasting models in fashion retailing. They applied ELM neural network model to investigate the relationship between sales amount and some significant factors which affect demand. The results demonstrate that the proposed methods outperform the back-propagation neural network model. Ali et al. (2009) explored forecasting accuracy versus data and model complexity trade-off in the grocery retailing sales forecasting problem, by considering a wide spectrum in data and technique complexity. The experiment results indicated 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 et al. (2009) developed the GMFLN forecasting model by integrating GRA and MFLN neural networks. The experimental results indicated that the proposed forecasting model outperforms the MA, ARIMA and GARCH forecasting models of the retail goods. Gil-Alana et al. (2010) examine whether retail sales forecasts can be better explained 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 find that retail sales forecasts are better explained in terms of a long memory model that incorporates both persistence and seasonal components. Chen and Ou (2011a, 2011b) developed the GELM model by integrating a Grey relation analysis (GRA) with extreme learning machine (ELM) to construct a forecasting model for fresh food retail industry. Results show that the GELM model outperforms the GARCH, GBPN, and GMFLN models. Ni and Fan (2011) proposed a two-stage dynamic forecasting model, which is a combination of the ART model and error forecasting model based on neural network to improve the accuracy of fashion retail forecasting. However, their results are not compared to other forecasting models.

Most studies emphasized different forms of neural network models and compared the forecast with few other forecasting models. These studies evaluate the forecasts from different models using the standard loss function which is essentially minimizing an average squared error. However, in this study, we consider twenty (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. None of these studies is conducted for South Africa. Hence, we focus on South Africa's aggregate retail sales.

3. Data and Methodology

We use monthly aggregate sales data for South Africa covering 1970:01 to 2012:05 making a total of 509 observations. The period covers a number of economic events thereby capturing both the boom and the recession periods in South Africa. The data is sourced from Statistics South Africa. The full data set is split into two. We use data from 1970M1-1986M6 (204 observations) for in-sample. Data from 1987:01-2012:05 (305 observations) is used for the out-of-sample period. The plot of the seasonally adjusted aggregate retail sales series is shown in Figure 1 while its growth rate is plotted in Figure 2. There is a noticeable seasonal variation in the data. Figure 1 shows that retail trade sales follow a particular pattern annually. Every December, retail sales figures spiked-upward and in January, a contraction occurred. This trend is explained by the tendencies of households to shop more during the December month since most people are on holiday or have received bonuses. In the month of January, consumer spending reduces as people prepare to go back to work or school and also pay off short-term debts incurred in December. The overall trend is an increase in retail trade sales. Figure 2 also depicts strong volatility with the highest peak in January 1987 (8.6%); thus justifying our choice of 1987:01-2012:05 out-of-sample period.

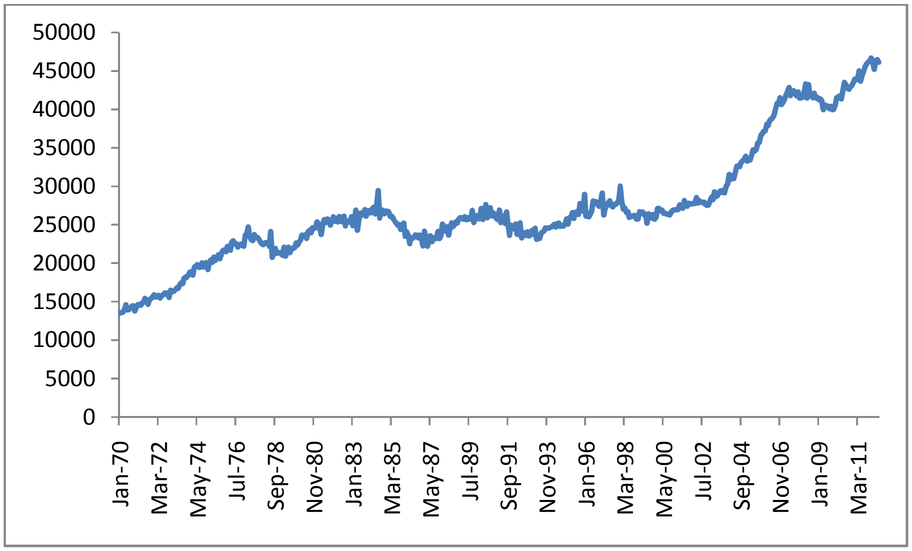

Figure 1: Aggregate seasonal retail sales series in million rand

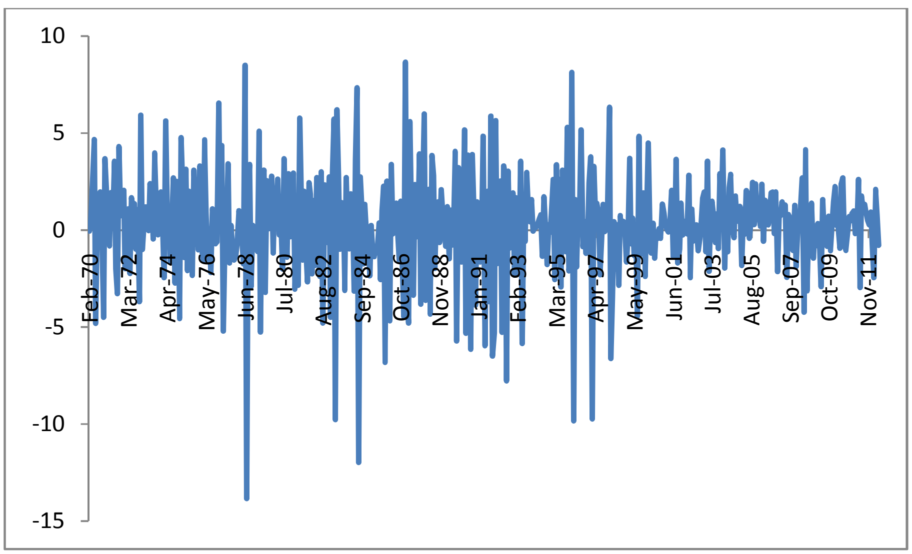

Figure 2: Growth rate of aggregate seasonal retail sales

3.1 Forecasting Models

A model is identified using the in-sample data and then the same model recursively re-estimated and 1 to 24 step ahead forecasts are obtained recursively over the out-of-sample period. Only the parameters are re-estimated in the recursive forecasting, but identified model structure is kept constant. We have two classes of models. The acronyms and brief description of the models we used are presented in Table 1.[3] The first class consists of 17 models with seasonal dummy

[3] However, given the pivotal role of forecast combination in this paper, detailed descriptions of the forecast combination models are given in the next sub-section.

variables. This is equivalent to deterministic seasonal adjustment. For instance for the ARIMA model we estimate,

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

where y is the log of the aggregate retail sales and $d_{s,t}$ is a dummy variable taking value of one for month s. At each recursive estimation, step dummy are included in the regression and forecasts of seasonal component is 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 Genetic Algorithm (GA) method for instance, seasonal component is pre-estimated using linear regression and non-seasonal component is forecasted in a second step; and final forecasts are recovered by adding $\mu + \sum_{s=1}^{11} \gamma_s d_{s,t}$. The second class consists of 9 full seasonal models. The models are presented in panel B of Table 1. In all models data is log of first differences, since there is a unit root. Level forecast are recovered from the forecasts of the growth rates. All model order are selected using BIC. In each case, forecasts were made at four horizons: 1, 4, 12 and 24 months.

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*) and where $p, q \le 12$. Estimates arising from minimizing 24-step MSE are used as informative priors in the recursive estimation.
5.	BCAR	Bias corrected AR model, the estimated model is AR(2) with first differencing. The method we used is described in Stine and Shaman (1989).
6.	MSAR	Markov Switching autoregressive model, estimated model is MS-AR(2) with 2 regimes and first differencing.
7.	SETAR	Self-exciting 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 3 hidden layers. The ANN is multi-layer perceptron (MLP) feed-forward network with 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, it is an autoregressive model with lag order equal to 2.
11.	SPAR	Semi-parametric (auto)regression, it is an autoregressive model with lag order equal to 2.
12.	GARCH	Generalized Autorregressive Conditional Heteroskedasticity model. We use ARIMA(2,1,0)-EGARCH(1,1) model.
13.	GA	This is the Genetic Algorithm based forecasting. Two lags are used as inputs (see Szpiro, 1997 for the approach we used). Function appriximation is terminated at a maximum step of 3000.

14.	FUZZY	Evolutionary Fuzzy Modeling. The approach is taken from Reyes (2002). Fuzzy fitting uses 200 population and 60000 generations.
15.	DISC	Discounted forecast combination. The discount factor we used is 0.50.
16.	PC	Principal components forecast combination. We used maximum of 4 principal components based on 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 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 unrealted 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 $p, q, P, \mathrm{Q} \leq 4$. Estimates arising from minimizing 24-step MSE are used as informative priors in the recursive estimation.
9.	SARFIMA	Autoregressive fractionally integrated moving average, estimated model is ARFIMA (2,1+*d*,0)(2,0,0).

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 the 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_{t+h}^h made at time t, $\hat{y}^h{}_{CB,t+h|t}$, typically are a linear combination of the individual model forecasts

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

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_{t+h}^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[4].

[4] We use 1987:01-1996:12 as the initial holdou out-ouf sample period.

Simple Combination Forecasts

The simple combination forecasts compute the combination forecast 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. Stock and Watson (2004) noted that there seems to be little difference between the mean and the trimmed mean forecast performance while the median typically has somewhat higher relative MSFE than either the mean or trimmed mean. Therefore, we consider the mean combination forecast (MEAN). The mean combination forecast simply involves setting $w_{i,t} = 1/n \quad (i = 1,...,n)$ in (2). Thus, the simple combining methods do not require a holdout out-of-sample period.

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:

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

where $m_{i,t} = \sum_{s=R}^{t-h} \gamma^{t-h-s} (y_{s+h}^{h} - \hat{y}_{i,s+h|s}^{h})^2$ (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, however, it may due to the seasonal time series we are forecasting and recent past is the most important, weights given to past forecast required to decline very fast in our case.

Principal component forecast combination

Principal component forecast combination (*PC*) requires (i) recursively computing the first few principal components of estimated common factors of the panel of forecasts, (ii) estimating a regression of $y_{s+h|s}^{h}$ onto these principal components, and (iii) forming the forecast based on this regression (Stock and Watson, 2004). Reduction of the 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 use of *PC* is that, recent work on large forecasting models suggests that large macroeconomic data sets are well described by a few common dynamic factors that are useful for forecasting, and that the common factors can be estimated by principal components (Forni, et al., 2000, 2001; Stock and Watson, 1999, 2002, 2004).

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 uncentered 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

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

where $s = R,...,t-h$. The combination forecast is given by $\theta_1 \hat{F}^h_{1,s+h|s} + ... + \theta_m \hat{F}^h_{m,s+h|s}$, 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 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*, aggregate retail sales in our case, $y^f_{i,t}$ is the value predicted by model i. Comparing the average loss difference of two competing models 1 and 2 implies computing their mean squared forecast errors

$$MSFE_i = \frac{1}{P}\sum_{t=T+1}^{T+P} e^2_{i,t}, \qquad i = 1,2, \tag{7}$$

over the forecast period $T+1$ to $T+P$ and choosing the model with the smaller MSFE.[5]

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

[5] RMSE is simply the square root of MSFE

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 re-specified as:

$$L_{i,t}^{w} = w_t e_{i,t}^2 \tag{8}$$

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 to focus on both tails of the distribution given rise to both recession and boom loss function.

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

$$MSFE_i = \frac{1}{P} \sum_{t=T+1}^{T+P} w_t e_{i,t}^2, \tag{9}$$

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

$$d_{i,t} = L_{0,t}^{w} - L_{i,t}^{w} = w_t e_{0,t}^2 - w_t e_{i,t}^2 \tag{10}$$

and averages over the forecast period

$$\bar{d}_i = \frac{1}{P} \sum_{t=T+1}^{T+P} d_{i,t} = \frac{1}{P} \sum_{t=T+1}^{T+P} w_t e_{0,t}^2 - \frac{1}{P} \sum_{t=T+1}^{T+P} w_t e_{i,t}^2 \tag{11}$$

We use this weighted loss and analyse the forecast accuracy of different models with respect to the different weighting schemes introduced above. There is large number of tests proposed in the literature to analyse whether empirical loss differences between two or more competing models are statistically significant. The most influential and most 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 small sample bias. *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,

$$E[d_{i,t}] = E[L_{0,t}^{w} - L_{i,t}^{w}] = 0 \qquad (12)$$

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

$$MDM = \sqrt{\frac{P+1-2h+h(h-1)/P}{P}} \frac{\bar{d}_i}{\sqrt{\hat{V}(\bar{d}_i)}} \qquad (13)$$

where h is the forecast horizon and $\hat{V}(\bar{d}_i)$ 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 less than one, then it outperformed the RW benchmark over the forecast period.

4. Empirical Results

In this section, we report the results from all the 26 aggregate retails forecasting models. We first present the uniform, boom, recession, and boom and recession weighted RMSE and their corresponding ranks. These results are presented in Tables 2, 3, 4 and 5 for horizons of 1, 4, 12 and 24, 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 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.

Table 2: Root Mean Squared Forecast Errors (h=1)

	Uniform		**Boom**		**Recession**		**Tail**	
Model	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 the corresponding ranking for each forecasting horizon and weighting scheme.

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 the first rank in 12 cases out of 16 and 2nd for the remaining 4 cases. This implies that the DISC model has the smallest RMSE in general. Following closely to 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 impressive as it takes the rank of between 6th and 18th for these cases. Another interesting finding in this study with respect to RMSE evaluation criterion is that the more sophisticated forecast combination methods outperformed the simple mean combination method unlike other studies cited previously.

Table 3: Root Mean Squared Forecast Errors (h=4)

	Uniform		**Boom**		**Recession**		**Tail**	
Model	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.

We can generally infer that the relative performance of the DISC model is unaffected by the 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 rank of 3 for both uniform and tail forecasts and 5 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.

Table 4: Root Mean Squared Forecast Errors (h=12)

	Uniform		**Boom**		**Recession**		**Tail**	
Model	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.

If we focus on different horizons, we can easily pick out the best three 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 PC models in that order while the top three models for recessions are MSAR, DISC and PC models. At the 4-month horizon, the top three models for booms are DISC, PC and NPAR models while the top three models for recessions are DISC, ARIMA and MSAR models. At the 12-month horizon, the top three models for booms are DISC, PC and MEAN models while the top three models for recessions are DISC, GARCH and ARIMA models. At the longest term horizon (*h*=24), the top three models for booms are PC, DISC and ARFIMA models while the top three models for recessions are GARCH, DISC and SETAR models. In practice, the choice of an appropriate model may depend on both the forecast horizon and on the specific loss function. Forecasters who particularly dislike forecast errors during recessions should use a slightly different set of models than forecasters who are more interested in correct boom prediction. This is consistent with the findings in Carstensen et al., 2010).

Table 5: Root Mean Squared Forecast Errors (*h*=24)

	Uniform		**Boom**		**Recession**		**Tail**	
Model	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.

Next we also evaluate the forecasting models based on their RMSE relative to the benchmark RW forecast.[6] 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-month 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 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 progress to longer horizons. Looking at horizon 24, the gain relative to RW model reduces to 26% and 41%, respectively, for the boom and recession forecasts. 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 individual 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 the later. 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 by both forecast horizon and different weighting schemes.

[6] 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.

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

Table 6: Summary of Modified Diebold-Mariano Forecast Accuracy Tests(*h*=1 and *h*=4)

	h=1								*h*=4							
	Uniform		**Boom**		**Recession**		**Tail**		**Uniform**		**Boom**		**Recession**		**Tail**	
Model	+	–	+	–	+	–	+	–	+	–	+	–	+	–	+	–
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

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

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

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 MSAR at horizon 1 for all weighting schemes, ARFIMA for recession and tails forecasts and LSTAR for uniform and boom and also recession forecasts at horizon 4. At horizon 12, it is ARIMA for recession and ARFIMA for boom forecasts. At horizon 24, it is ARFIMA for the boom and BARIMA model for the uniform and recession forecasts. However, the combination forecasts, 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.

Table 7: Summary of Modified Diebold-Mariano Forecast Accuracy Tests (h=12 and h=24)

	h=12								h=24							
	Uniform		**Boom**		**Recession**		**Tail**		**Uniform**		**Boom**		**Recession**		**Tail**	
Model	+	–	+	–	+	–	+	–	+	–	+	–	+	–	+	–
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 "+" indicate the number of times a specific model significantly outperforms its competitors. The columns "-" indicate the number of times a specific model is outperformed by its competitors.

5. Conclusion

In this paper, we assess the forecasting performance of 26 models of South Africa's aggregate seasonal retail sales over 1987:01 – 2012:05 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 overweighed 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 forecasts from 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 full seasonal models. Most of the models performed better than the random work benchmark. 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 differ 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 turned out to outperform the individual models in general. Specifically, the discounted combination forecast model (DISC) outperform all the single models and the other two combination forecasts (simple mean and principal component) models and the performance is largely unaffected by specific economic or business cycle situation.

References

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

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

Ali, Ö. G., Sayin, S., Woensel, T. V. and Fransoo, J. (2009). SKU demand forecasting in the presence of promotions. *Expert Systems with Application*, 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*, 234–236.

Alon, I., Qi, M. and Sadowski, R. J. (2001). Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods. *Journal of Retailing and Consumer Services*, 8(3), 147–156.

Au, K. F., Choi, T. M. and Yu, Y. (2008). Fashion retail forecasting by evolutionary neural networks. *International Journal of Production Economics*, 114 (2), 615–630.

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

Banternghansa, C. and McCracken, M.W. (2010) Real-time forecast averaging with ALFRED. *The Federal Reserve Bank of St. Louis Working Paper* 2010-033A, 49-66.

Bates, J. M. and Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly*, 20, 451–468.

Barksdale, H.C. and Hilliard, J.E. (1975). A cross-spectral analysis of retail inventories and sales. *Journal of Business*, 48(3), 365-82.

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

Chan, Y. L., Stock, J. H. and Watson, M. W. (1999). A dynamic factor model framework for forecast combination. *Spanish Economic Review*, 1, 91–121.

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

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

Chen, F.L. and Ou, T.Y. (2011a). Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry. *Expert Systems with Applications*, 38, 1336–1345.

Chen, F.L. and Ou, T.Y. (2011b). Constructing a sales forecasting model by integrating GRA and ELM: A case study for retail industry. *International Journal of Electronic Business Management*, 9 (2), 107-121.

Chu, C. W., and Zhang, P. G. Q. (2003). A comparative study of linear and nonlinear models for aggregate retail sales forecasting. *International Journal of Production Economics*, 86, 217–231.

Clemen, R.T (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5, 559-583.

Diebold, F. X. and Mariano, R.S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13(3), 253–263.

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

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The generalized factor model: identification and estimation. *Review of Economics and Statistics*, 82, 540–554.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2001). Do financial variables help forecasting inflation and real activity in the EURO area? CEPR Working Paper

Frank, C., Garg, A., Sztandera, L. and Raheja, A. (2003). Forecasting women's apparel sales using mathematical modeling. *International Journal of Clothing Science and Technology*, 15(2), 107–125.

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

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

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

Harvey, David I., Stephen J. Leybourne, and Paul Newbold (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13, 281–291.

Harvey, A. (2006). Forecasting with Unobserved Components Time Series Models. Faculty of Economics, University of Cambridge. In Elliott,G., Granger, C.W.J and Timmermann, A. (eds), *Handbook of Economic Forecasting*, *1, 327–412*. Elsevier B.V.

Hyndman, R.J., Koehler, A.B., Snyder, R.D. and Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3), 439-454.

IHS Global Insight Regional Explorer-Encyclopedia (Ver 2.04) (2012).

Joseph, A., Singh, E. and 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*, 17.

Kapetanios, G., Labhard, V. and Price, S. (2008). Forecast combination and the Bank of England's suite of statistical forecasting models. *Economic Modelling*, 25(4), 772-792.

Kearney, A.T. (2011). Retail global expansion: a portfolio of opportunities-Global Retail Development Index. http://www.atkearney.com/documents/10192/3903b4b7-265c-484e-932c-50169e5aa2f3.

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

Makridakis, S. (1989). Why combining works? *International Journal of Forecasting*, 5, (4), 601-603.

Makridakis, S.. and Winkler, R.L. (1983). Averages of forecasts: Some empirical results. *Management Science,* 29, 987-996.

Mentzer, J. T. and Bienstock, C. C. (1998). *Sales forecasting management: understanding the techniques, systems and management of the sales forecasting process*. Thousand Oaks, CA: Sage publications.

Montaño, J.J., and Palmer, A. (2003). Numeric sensitivity analysis applied to feed-forward neural networks. *Neural Computing and Applications*, 12, 119-125.

Moreno, J.J.M, Alfonso Palmer Pol, A. P. and Gracia, P.M. (2011). Artificial neural networks applied to forecasting time series. *Psicothema*, 23 (2), 322-329.

Ni, Y. and Fan, F. (2011). A two-stage dynamic sales forecasting model for the fashion retail. *Expert Systems with Applications*, 1529–1536

Palmer, A., Montaño, J.J. and Franconetti, F.J. (2008). Sensitivity analysis applied to artificial neural networks for forecasting time series. *Methodology*, 4, 80-84.

Peterson, R. T. (1993). Forecasting practices in the retail industry. *Journal of Business Forecasting*, 12, 11–14.

Rapach, D.E. and Strauss, J.K. (2010). Bagging or combining (or both)? An analysis based on forecasting U.S. Employment growth. *Econometric Reviews*, 29(5–6), 511–533.

Reyes, C.A.P. (2002). Coevolutionary Fuzzy Modeling. Unpublished PhD Thesis, Faculté Informatique Et Communications, École Polytechnique Fédérale De Lausanne.

RNCOS (2011). South African retail industry forecast to 2013. September 1, 2011. http://www.rncos.com/Report/IM275.htm.

Szpiro, G.G. (1997). Forecasting chaotic time series with genetic algorithms. *Physical Review E*, 55(3), 2557-2568.

Stine, R. A., and Shaman, P. (1989). A fixed point characterization for bias of autoregressive estimators. *The Annals of Statistics*, 17, 1275-1284.

Stock, J.H. and Watson, M.W. (1999). Forecasting inflation. *Journal of Monetary Economics*, 44, 293–335.

Stock, J.H. and Watson, M.W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20, 147–162.

Stock, J. H. and Watson, M. W. (2003). Forecasting output growth and inflation: The role of asset prices. *Journal of Economic Literature*, 41:788–829.

Stock, J. H. and Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23:405–430.

Sun, Z. L., Choi, T. M., Au, K. F. and Yu, Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. *Decision Support Systems*, 46, (1), 411-419.

Thall, N. (1992). Neural forecasts: A retail sales booster. *Discount Merchandiser*, 23 (10), 41–42.

Taskaya-Temizel, T. and M.C. Casey (2005). Configuration of neural networks for the analysis of seasonal time series. Proceedings of the Third international conference on Advances in Pattern Recognition - Volume Part I Springer-Verlag.

van Dijk, D. and Franses, P.H. (2003). Selecting a Nonlinear Time Series Model using Weighted Tests of Equal Forecast Accuracy. Oxford Bulletin of Economics and Statistics, 65, 727–744.

Wan, L., Luo, B., Ji, H-M. and Yan, W-W (2012). A type of combination forecasting method based on time series method and PLS. *American Journal of Operations Research*, 2012, 2, 467-472.

Xiao, T. J. and 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.

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/

Zhang, G.P. and Kline, D.M. (2007). Quarterly time-series forecasting with neural networks. *IEEE Transactions on Neural Networks*, 18(6), 1800–1814.