doi.org/10.26398/asaproc.0043

Exploiting consumer sentiment in volume sales forecasting: case studies

Domenica Fioredistella Iezzi [a], Roberto Monte [b]
[a] Department of Enterprise Engineering Mario Lucertini, Tor Vergata University of Rome, Italy
[b] Department of Civil Engineering and Computer Science Engineering, Tor Vergata University of Rome, Italy

1. Introduction

The impact of word of mouth on sales has been well known since before the arrival of the Internet and the explosion of the social media phenomenon. Paul Lazarsfeld's studies have highlighted that media messages can be further mediated by informal opinion leaders who intercept, interpret, and disseminate what they see and hear in the personal networks where they are inserted (Katz et al., 2006). The advent of the Internet with the introduction of digital word of mouth has increased and accelerated the assessment of messages about products and services. Social media, in particular, has dramatically amplified the echo of traditional word of mouth (Huete-Alcocer, 2017).

In the summer of 2023, we witnessed media phenomena linked to the publication of receipts on social media, highlighting the high prices of products or services in tourist locations. This information had a boomerang effect with large shares on the net and strongly negative comments. According to Yang (2017), a negative or positive attitude of customers towards a product or a service will influence customers' future purchase intentions.

This paper aims to identify the best forecast of sales of two goods using state-space models with signals produced by consumer feelings (Iezzi and Monte, 2022, 2023; Basili et al., 2023). We measure the consumer sentiment score to track attitudes expressed in tweets. We apply different approaches: from the Valence Aware Dictionary and sEntiment Reasoner (VADER) rules-based model, using the estimation of negative, positive, and compound feelings (Hutto and Gilbert, 2014), to machine learning-oriented techniques relying on logistic regression (Prabhat and Khullar 2017, to deep learning Word2Vec[1] (Mikolov et al., 2013). We want to investigate which approach might generate a better signal for predicting the sales of the goods. To forecast sales volumes, we use different models and compare the results: the ETS-AAA model[2] (Additive error, Additive slope component in the local trend, Additive seasonality), the ETS-ANA Model (Additive error, No additive slope component in the local trend, Additive seasonality), and Multivariate Autoregressive State-Space (MARSS) models (Harvey, 1990, Hyndman et al. 2002). In particular, to implement the MARSS models, which account for consumer feelings, we have collected two corpora from Twitter and built some sentiment signals with R and Python software.

Section 2 presents data and models; Section 3 presents the main findings; Section 4 concludes.

---

[1]Word2Vec is a group of models that are used to produce word embeddings. These models are two-layer neural networks trained to reconstruct words' linguistic contexts.

[2]The ETS models, by R. Hyndman and his coworkers (e.g., Hyndman et al., 2002), are a well-known family of time series models, consisting in the state-space form of the Holt-Winters models, having, in general, an error component (E), a local trend with an additional slope component (T), and a seasonal component (S).

2. Data and models

The two corpora collected from Twitter concern Toyota Camry US (165,000 tweets) and Gentilini Osvego biscuits (1,016 tweets). We consider Toyota Camry tweets from June 1, 2009, to December 31, 2022, and Gentilini Osvego biscuits tweets from January 1, 2013, to December 31, 2020. We use the Twitter Academic Research Product Track based on our academic profile (courtesy of Twitter Developer) to scrape Twitter time series [3]. Gentilini Osvego biscuits tweets are in Italian, but we translate them using the library "googleLanguageR", an R package allowing speech-to-text transcription, neural net translation, and natural language processing via the Google Cloud machine learning set. The choice to implement a supervised machine learning model has imposed the need to introduce a labeled dataset for training it. Between different possible options, we have decided to use Sentiment140, a dataset of 1.6 millions of positively and negatively labeled tweets. The reasons are essentially due to the following: the corpus of tweets collected, mainly concerning user opinions related to brands, products, or topics in general, was conceived precisely for classifying sentiment on Twitter.

We use the ETS-ANA and ETS-AAA models, introduced by Hyndman et al. (2008), as benchmarks. These can be written as follows:

$$\text{ETS-ANA}\begin{cases} y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t, \\ \ell_t = \ell_{t-1} + \alpha\varepsilon_t, \\ s_t = s_{t-m} + \gamma\varepsilon_t; \end{cases} \qquad \text{ETS-AAA}\begin{cases} y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t, \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t, \\ b_t = b_{t-1} + \beta\varepsilon_t, \\ s_t = s_{t-m} + \gamma\varepsilon_t, \end{cases}$$

where y_t is the value of the time series of interest at time t, the hidden variable ℓ_t [resp. b_t, resp. s_t] is the *local level* [resp. *slope*, resp. *seasonality*] of the Holt-Winters decomposition of y_t, and the variable ε_t represents the innovation term at time t with variance σ^2. The parameter σ^2 is determined together with the parameters λ, β, γ, and the initial states of the model in the estimation procedure. The only difference between the ETS-ANA and ETS-AAA models is the lack of the slope component in the former. Referring to the ETS-AAA model, our idea is to generalize the model in the context of State-Space models by attributing the role of consumer sentiment for the product to the slope variable b_t. Naturally, we cannot expect to observe this variable, which is kept as a hidden variable, while we assume to observe a proxy, which is the sentiment signal built from time to time. Accordingly, in this paper, we study a state space model in which the state and observation equations take the following forms:

$$\text{state equations}\begin{cases} \ell_t = \beta_{\ell,\ell}\ell_{t-1} + \beta_{\ell,b}b_{t-1} + \sigma_{\ell,\ell}w_t^{(l)}, \\ b_t = \beta_{b,b}b_{t-1} + \sigma_{b.b}w_t^{(b)}, \\ s_t = s_{t-m} + \sigma_{s,s}w_t^{(s)}, \end{cases} \tag{1}$$

$$\text{observation equations}\begin{cases} y_t = \beta_{y,\ell}\ell_t + \beta_{y,b}b_t + \beta_{y,s}s_t + \sigma_{y,y}w_t^{(y)}, \\ z_t = \beta_{z,b}b_t + \sigma_{z,z}w_t^{(z)}. \end{cases}$$

Here, y_t is still the observed sales volume time series of the product, and z_t represents the observed sentiment scores on the hidden sentiment variable b_t. To add more flexibility to our model compared to the ETS-AAA model, we introduce the additional parameters $\beta_{\ell,\ell}, \ldots, \beta_{z,b}$, Moreover, we introduce independent innovations $w_t^{(l)}, \ldots, w_t^{(z)}$, with variances $\sigma^2_{\ell,\ell} \ldots, \sigma^2_{z,z}$, respectively. All these parameters, together with the initial states of the models, are estimated in a recursive procedure using the functions of the *MARSS R* library (Holmes et al., 2003).

[3] see https://developer.twitter.com/en/blog/product-news/2021/enabling-the-future-of-academic-research-with-the-twitter-api

3. Discussion

To test our model, we consider the Toyota Camry US monthly sales volume time series from June 1, 2009, to December 31, 2022, and the Gentilini Osvego biscuits monthly sales volume from January 1, 2012, to December 31, 2020. We apply the ETS-AAA and ETS-ANA models to these time series and monthly sales volume logarithms as a benchmark and, as a further benchmark, we also use a univariate MARSS model, which is much closer to the ATS-AAA model since it does not account for the sentiment signal. The large difference in the scale of the monthly sales volume and sentiment signals has suggested the opportunity to use the monthly sales volume logarithm time series. To measure the forecasting accuracy in the models, the dataset was divided into two parts: training set (90%) and test set (10%). The results of our comparative analysis are summarized in Tables 1-2. We consider the negative VADER, logistic regression, and Word2Vec as the sentiment signals. Figures 1-2 show the Toyota Camry sales volume and the Gentilini Osvego monthly sales volume logarithm as forecasted by bivariate MARSS models accounting for the sentiment signal.

Figure 1 highlights the existence of some shared peaks and troughs between the observations of each year. March, May, and August have the highest sales over the years, while February and November have the lowest. Possible anomalous values are also evident, such as those recorded in April 2020 and August 2009, the first of which is undoubtedly due to the restrictive measures imposed worldwide after the outbreak of the COVID-19 pandemic.

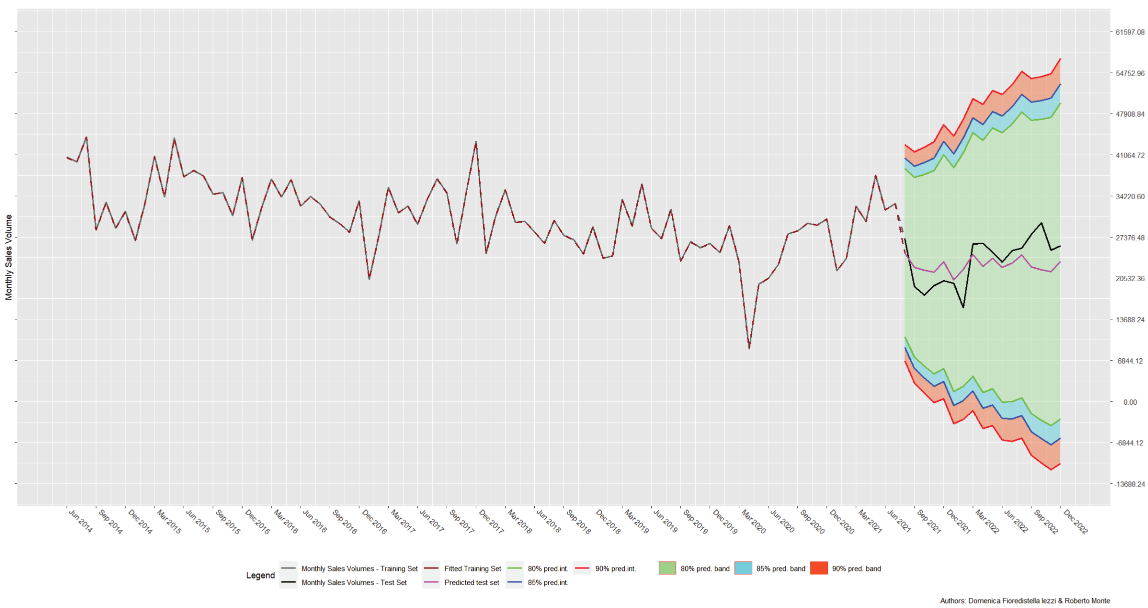

Figure 1: **Toyota Camry US - MARSS bivariate model with monthly sales volumes and Word2Vec sentiment observed signals**

From Tables 1-2, the accuracy measures[4] MAPE, SMAPE%, MASE, and RMSSE show that MARSS bivariate models outperform all univariate benchmark models. In this context, RMSE and MAE are less reliable. The comparison of the information measures is less indicative because the structures and the numbers of parameters in the bivariate and univariate models differ.

[4]Accuracy measures: logLik=log-likelihood, AIC=Akaike information criterion, AICc=Akaike information criterion corrected, BIC=Bayes Information Criterion, RMSE=Root Mean Square Error, MAE=Mean Absolute Error, MAPE=Mean Absolute Percentage Error, SMAPE=Symmetric mean absolute percentage error, MASE=Mean Absolute Scaled Error, RMSSE= Root Mean Square Error.

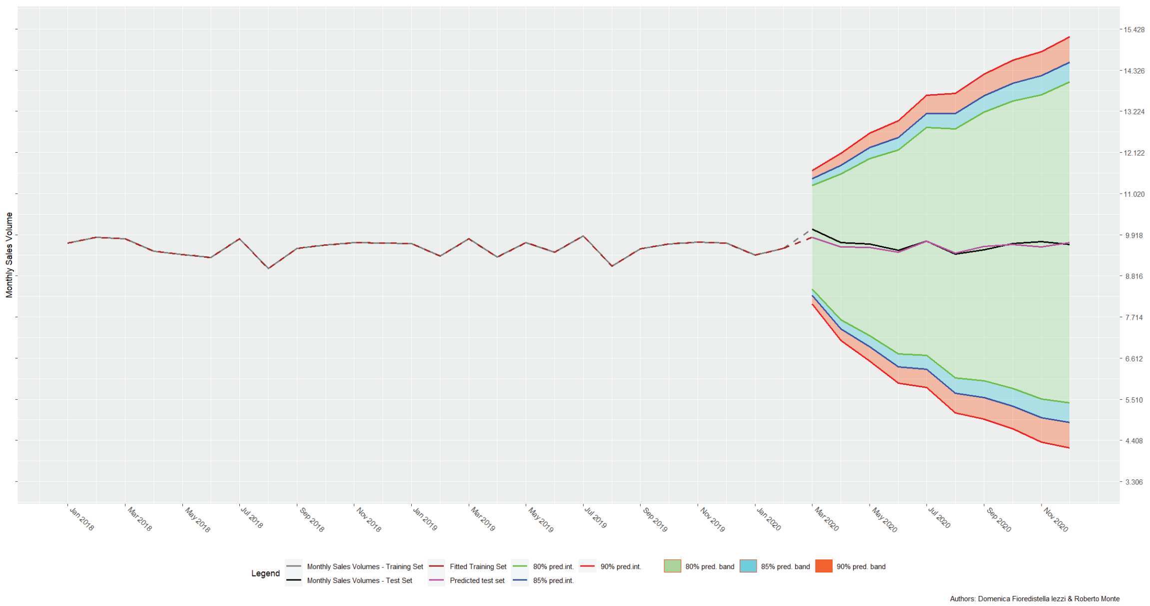

Figure 2: **Gentilini Osvego biscuits - MARSS bivariate model with monthly sales volumes logarithm and negative VADER sentiment observed signals**

Models	logLik	AIC	BIC	AICc	RMSE	MAE	MAPE	SMAPE%	MASE	RMSSE
ETS-ANA (y)	-1591.434	3212.869	3257.623	3216.561	9440.055	8610.602	39.914	15.902	1.669	1.389
ETS-AAA (y)	-1592.126	3218.252	3268.973	3223.033	10809.330	10241.120	46.714	18.321	1.985	1.590
MARSS (y) - No Sent.	-1567.278	3154.555	3184.319	3156.185	7291.802	6087.463	28.754	11.831	1.180	1.073
MARSS (y) - Neg. VADER	-1562.234	3148.469	3184.272	3150.814	4599.050	3608.202	17.271	7.811	0.699	0.676
MARSS (z) - Neg. VADER	-1562.234	3148.469	3184.272	3150.814	0.026	0.018	59.089	24.169	1.156	0.971
MARSS (y) - LR	-1789.955	3621.910	3653.713	3624.255	3947.773	3119.797	14.460	6.925	0.605	0.580
MARSS (z) - LR	-1789.955	3621.910	3653.713	3624.255	0.118	0.062	9.790	5.319	0.818	0.990
MARSS (y) - Word2Vec	-1435.063	2894.125	2929.928	2896.471	3625.778	3064.401	13.633	6.741	0.594	0.533
MARSS (z) - Word2Vec	-1435.063	2894.125	2929.928	2896.471	0.076	0.036	7.971	3.235	0.977	0.992

Table 1: **Information parameters and accuracy measures for Toyota Camry US monthly sales volumes: ETS-ANA model, ETS-AAA model, MARSS univariate, and MARSS bivariate with negative VADER, logistic regression (L.R.), and Word2Vec sentiment signals**.

Models	logLik	AIC	BIC	AICc	RMSE	MAE	MAPE	SMAPE%	MASE	RMSSE
ETS-ANA (y)	-34.444	98.887	135.703	105.744	0.209	0.168	1.724	0.873	0.743	0.755
ETS-AAA (y)	-36.170	106.341	148.065	115.341	0.227	0.189	1.939	0.983	0.836	0.818
MARSS (y) - No Sent.	-128.888	277.777	302.320	280.710	0.122	0.085	0.065	0.437	0.376	0.438
MARSS (y) - Neg. VADER	48.060	-72.121	-42.669	-67.847	0.101	0.081	0.829	0.416	0.358	0.364
MARSS (z) - Neg. VADER	48.060	-72.121	-42.669	-67.847	0.043	0.029	-	43.208	0.788	0.891
MARSS (y) - LR	-5.606	35.212	64.664	39.486	0.132	0.086	0.877	0.442	0.381	0.477
MARSS (z) - LR	-5.606	35.212	64.664	39.486	0.045	0.040	6.984	3.596	0.630	0.525
MARSS (y) - Word2Vec	-55.288	134.576	164.028	138.850	0.105	0.068	0.6901	0.347	0.300	0.380
MARSS (z) - Word2Vec	-55.288	134.576	164.028	138.850	0.120	0.088	15.913	7.276	0.825	0.816

Table 2: **Information parameters and accuracy measures for Gentilini Osvego biscuits monthly sales volumes logarithm: ETS-ANA model, ETS-AAA model, MARSS univariate, and MARSS bivariate with negative VADER, logistic regression (L.R.), and Word2Vec sentiment signals.**

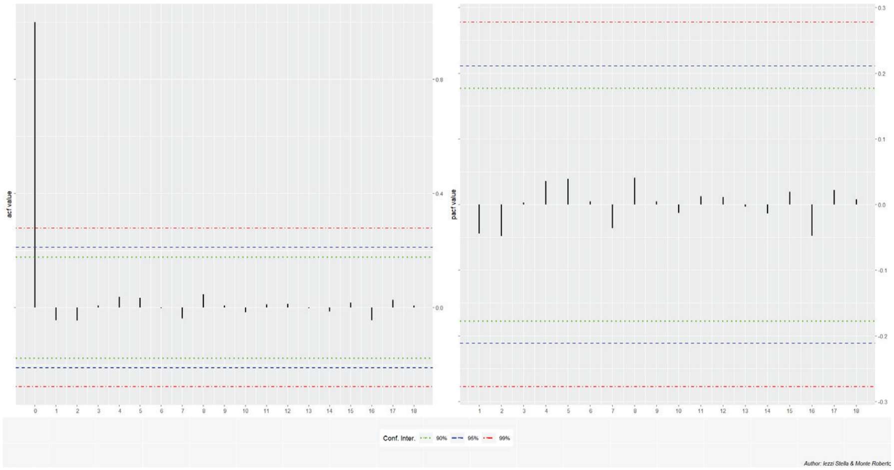

Figure 3: **Autocorrelogram and partial autocorrelogram of the residuals of the bivariate MARSS model for the Osvego Gentilini monthly sales volumes logarithm and negative Vader sentiment**

Figure 1 illustrates the MARSS bivariate model with monthly sales volumes and Word2Vec sentiment observed signals for the Toyota Camry US. So far, this appears to be the best model for our case study. The fitted monthly sales volumes are so close to the states in the training set (it is almost impossible to distinguish the training and fitted set in the plot) that one could think of a typical case of overfitting. However, the predicted monthly sales volumes are so close to the test set that overfitting is unlikely. Figure 2 describes the MARSS bivariate model with the monthly sales volumes logarithm and negative VADER sentiment observed signals for the Gentilini Osvego biscuits. As observed in other research (Iezzi and Monte, 2023; Xiaolin et al., 2019), negative sentiment provides a better signal than the others (positive and compound) possible with VADER. The same considerations as in Figure 1 apply, with even stronger emphasis due to the logarithm transformation, which corrects the difference in scale of the observed signals. In this case, we also provide the plot of the autocorrelogram and partial autocorrelogram of the residuals (see Figure 3), which renders the suspect of overfitting unlikely. The logarithmic transformation of sales volumes generally improves the performance of the models because it tends to reduce the disproportion between the signals. It also manages to reduce sensitivity to extreme deviations by reducing the effect of extreme values (outliers), thus making the data distribution less influenced by them.

4. Conclusions

The results highlight that using a sentiment signal may improve the models' predictive capabilities. This advocates using data from customers' word of mouth. The different models tested for sentiment quantification (VADER, Logistic Regression, and Word2Vec) all show good results with a slight advantage for VADER and Word2Vec. The reason might be that VADER, a model-based approach built on social media, also allows the quantification of emoticons and symbols, capturing the nuances of sentiment in tweets well. Word2Vec is a neural network model that converts words into vectors of real numbers so that semantically similar words are represented by neighboring vectors in vector space. This approach helps to analyze and understand the meaning of words in a specific context. However, it is essential to note that Word2Vec may not fully capture the irony or sarcasm present in tweets, and its effectiveness depends on the quantity and quality of training data. Given those results in future tests, other more advanced deep learning models such as LSTM (Long Short-Term Memory) and BERT (Bidirectional Encoder Representations from Transformers) will be applied. The recurring structure of LSTMs makes them suitable for capturing temporal relationships and dependencies in text, which can be crucial in sentiment analysis. BERT captures bidirectional contexts of words, considering both directions of the context in which a word appears. This allows for a better understanding of the context of words, improving sentiment analysis.

References

Basili, R., Croce, D., Iezzi, D.F., Monte, R. (2023). The role of BERT in neural network sentiment scoring for time series forecast. In *Proceedings of the Statistics and Data Science Conference*, P. Cerchiello, A. Agosto, S. Osmetti, A. Spelta (eds.) pp. 55-60. PAVIA, Pavia University Press.

Harvey, A.C. (1990). *The Econometric Analysis of Time Series*, 2nd Edition, MIT Press Books, The MIT Press, Cambridge, Massachusetts.

Holmes, E.E, Ward, E.J., Scheuerell, M.D., Wills, K. (2023). MARSS: Multivariate Autoregressive State-Space Modeling. *R Package Version 3.11.8*.

Huete-Alcocer, N. (2017). A Literature review of word of mouth and electronic word of mouth: Implications for consumer behavior. *Frontiers in Psychology* **8** pp. 1-4.

Hutto, C., Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*. **8**(1) (May 2014), pp. 216-225.

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

Hyndman, R.J., Akram, M., Archibald, B.C. (2008). The admissible parameter space for exponential smoothing models. *Annals of the Institute of Statistical Mathematics* **60**(2), pp. 407-426.

Iezzi, D.F., Monte, R. (2022). Sales forecast and electronic word of mouth: the power of feelings. In *Proceedings of the 16th International Conference on statistical analysis of textual data (JADT22)* **2**, pp. 489-494. M. Misuraca, G. Scepi, M. Spano (eds.)

Iezzi, D.F., Monte, R. (2023 - in press). Social media effects on sales: consumer sentiment in a state-space model. In *New Frontiers in Textual Data Analysis* (Book Series: Studies in Classification, Data Analysis, and Knowledge Organization) M. Misuraca, G. Giordano (eds.), Springer.

Katz, E., Lazarsfeld, P.F., Roper, E., (2006). *Personal Influence: The Part Played by People in the Flow of Mass Communications* (1st ed.). Routledge.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems 26 (NIPS 2013)*.

Prabhat, A., Khullar, V., (2017). Sentiment classification on big data using Naïve Bayes and logistic regression, *2017 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, pp. 1-5.

Xiaolin, Li, Chaojiang, Wu, Fen, M. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis, *Information & Management*, **56**(2), pp. 172-184.

Yang, F.X. (2017). Effects of Restaurant Satisfaction and Knowledge Sharing Motivation on eWOM Intentions: the Moderating Role of Technology Acceptance Factors. *Journal of Hospitality & Tourism Research*. **41**, pp. 93-127.