Time Series Forecasting with Hybrid ARIMA and exponential smoothing model:

A Study Based on Commodity Goods

Ng Qi Xuan

Introduction

Autoregressive integrated moving average (ARIMA) model and Holtz' Winter exponential smoothing model are popular time series models for forecasting [1]. ARIMA model combines autoregressive (AR), differencing (I), and moving average (MA) components. On the other hand, the exponential smoothing model assigns exponentially decreasing weights to past observations. While the robustness of these two models makes them suitable for non-stationary data with trend and/or seasonality, multiple research studies have shown that manually decomposing the time series first is still preferred [2,3]. This enables a more comprehensive understanding of the underlying patterns and provides flexibility for hybrid modelling [4]. In this study, autoregression is first performed to identify the non-stationary components in the time series before performing Seasonal and Trend Loess (STL) decomposition. The aim is to provide greater insights on their decomposed components and subsequently evaluate the performance of hybrid ML modelling.

Data Exploration

6 commodities with different trend and seasonality are chosen, namely olive oil, oranges, swine pork, sawn wood, shrimp and wool [5]. Their autocorrelation plots are shown in Fig. 1. A strong positive autocorrelation that gradually decrease suggests that a trend is present across all commodities. Moreover, a seasonality of m = 12 is also evident in Fig. 1 for oranges and swine as seen from the larger autocorrelation at its seasonal interval. The trend component and seasonality of m = 12 is first removed to further identify other possible features, including seasonality (where m ≠ 12), cyclic and noise, from the autocorrelation plot in Fig. 2.

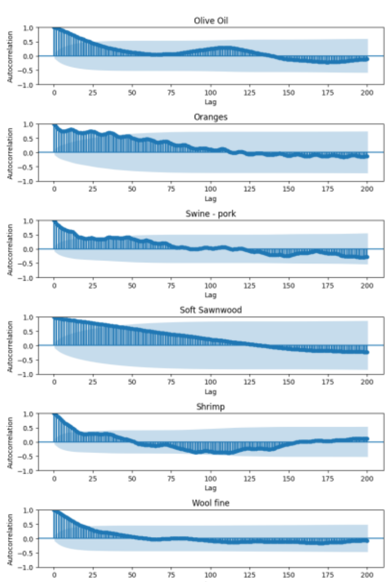

Fig. 1: Autocorrelation (original)

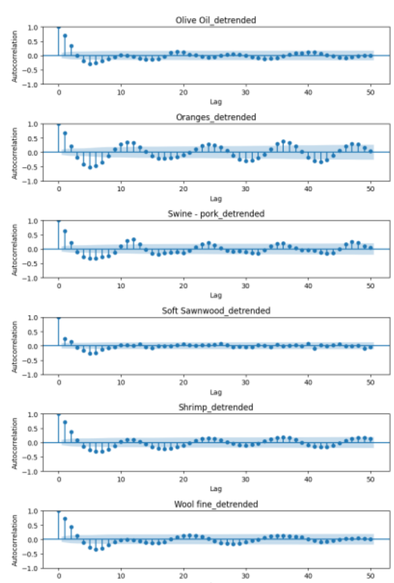

Fig. 2: Autocorrelation (de-trended)

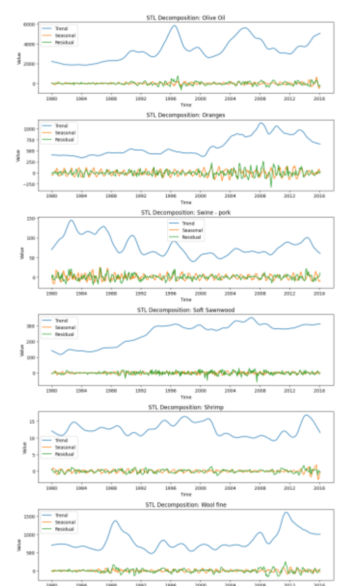

Fig. 3: Decomposed time series

This study employs the STL decomposition from statsmodels.tsa.seasonal to split each time series into trend, seasonal and residual (Fig. 3). Subsequently, autocorrelation is reperformed on the residual component (Fig. 4). This would reveal any other patterns in the time series that might be hidden from the more dominating trend and seasonal components. This process stops when at least one of the conditions below is fulfilled.

a) The time series becomes stationary and white noise is left (i.e. the residual component is indeed residual). Only statistically insignificant values at lag >1 remains on the autocorrelation plot.
b) No additional seasonal patterns could be identified and the residual component resembles a cyclic pattern. This is shown as multiple statistically significant peaks at no fixed interval on the autocorrelation plot.
c) The residual component is statistically insignificant (i.e. its 25th to 75th percentile < 5% of the trend component).

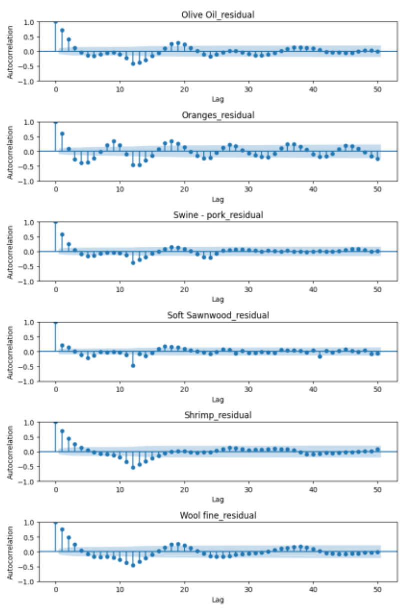

Fig. 4a: Autocorrelation (residual) after 1st STL decomposition iteration

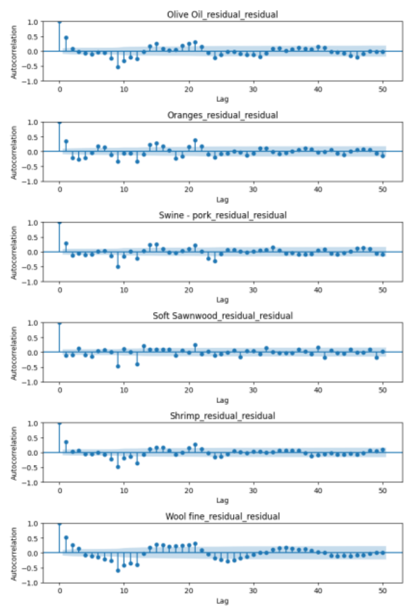

Fig. 4b: Autocorrelation (residual) after 2nd STL decomposition iteration

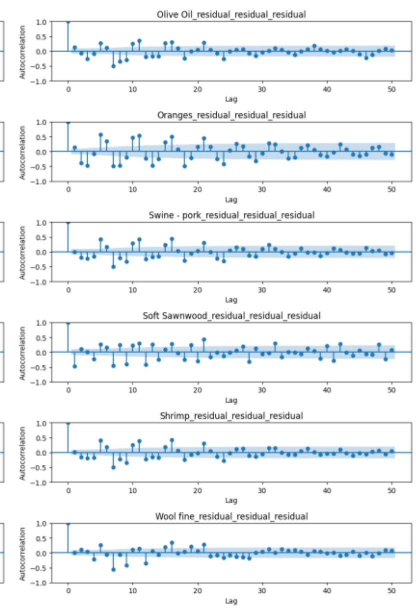

Figure 4c: Autocorrelation (residual) after 3rd STL decomposition iteration

While not all autocorrelation plots of the original time series reveal seasonal pattern of m = 12, all their residual components resemble closer to a stationary time series after the 1st round of STL decomposition. This suggests that there are other unidentified patterns present in their residual components. Else, the seasonal component should only be removed for commodities that shows an improvement in autocorrelation towards a stationary time series. Fig. 4b also reveals that oranges has an additional seasonal pattern of m = 9. After the 3rd round of STL decomposition, condition b) is fulfilled.

Commodity	residual_range/ trend_comp.mean()
Olive Oil	0.014192
Oranges	0.043089
Swine - pork	0.037377
Soft Sawnwood	0.021279
Shrimp	0.013548
Wool fine	0.018002

Table 1: Residual to trend ratio for each commodity

Results and Analysis

ARIMA model and exponential smoothing model are separately applied to the trend and seasonal of each commodity.

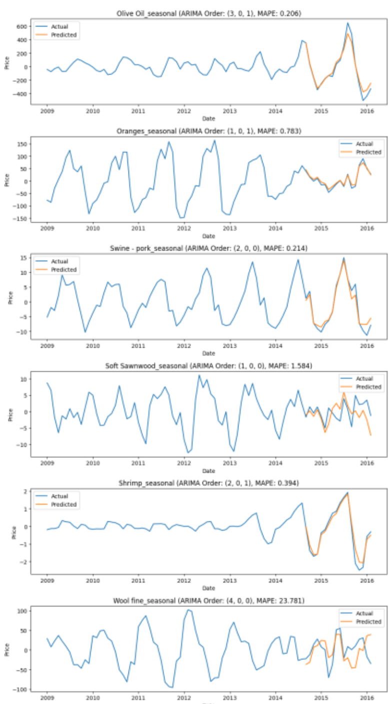

Fig. 5a: Seasonal ARIMA with 72 months of data

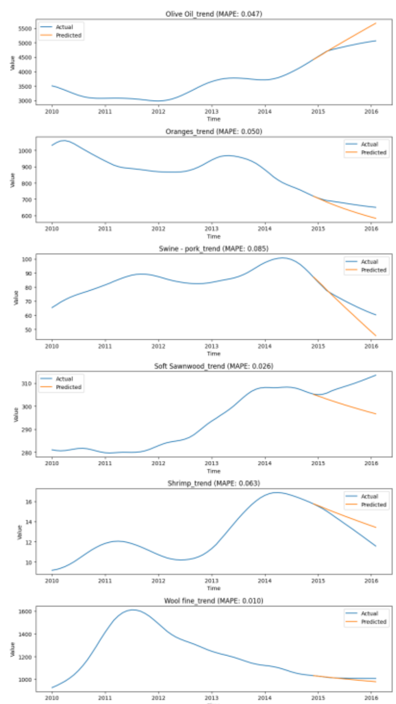

Fig. 5b: Exponential modelling with 84 months of data

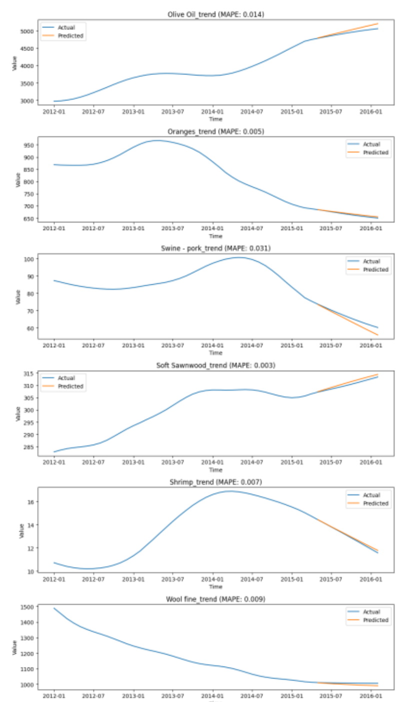

Figure 5c: Exponential modelling with 60 months of data

The models' performances are evaluated based on their mean average percentage error (MAPE). The best ARIMA models are selected using the pmdarima library. Rows prior to 2010 are dropped as the models ended up performing poorly due to overfitting.

As shown in Table 2, exponential smoothing model outperforms ARIMA model for the trend component. The MAPEs for the seasonal component is slightly higher due to the small data range. Nevertheless, Fig. 6 shows that the ARIMA model still manage to capture the overall pattern and further fine tuning could be performed.

Model	Commodity	Component	MAPE	
			2010 – 2016	2012 - 2016
Exponential smoothing	Olive oil	Trend	0.047	0.014
Exponential smoothing	Oranges	Trend	0.050	0.005
Exponential smoothing	Swine pork	Trend	0.085	0.031
Exponential smoothing	Sawn wood	Trend	0.026	0.003
Exponential smoothing	Shrimp	Trend	0.063	0.007
Exponential smoothing	Wool	Trend	0.010	0.009
ARIMA	Olive oil	Trend	0.062	0.188
ARIMA	Oranges	Trend	0.061	0.124
ARIMA	Swine pork	Trend	0.066	0.258
ARIMA	Sawn wood	Trend	0.016	0.031
ARIMA	Shrimp	Trend	0.055	0.126
ARIMA	Wool	Trend	0.023	0.198
ARIMA	Olive oil	Seasonal	0.355	0.762
ARIMA	Oranges	Seasonal	0.290	2.743
ARIMA	Swine pork	Seasonal	0.125	0.046
ARIMA	Sawn wood	Seasonal	1.499	0.464
ARIMA	Shrimp	Seasonal	0.269	0.035
ARIMA	Wool	Seasonal	2.143	0.735

Table 2: Model evaluation for the trend and seasonal components

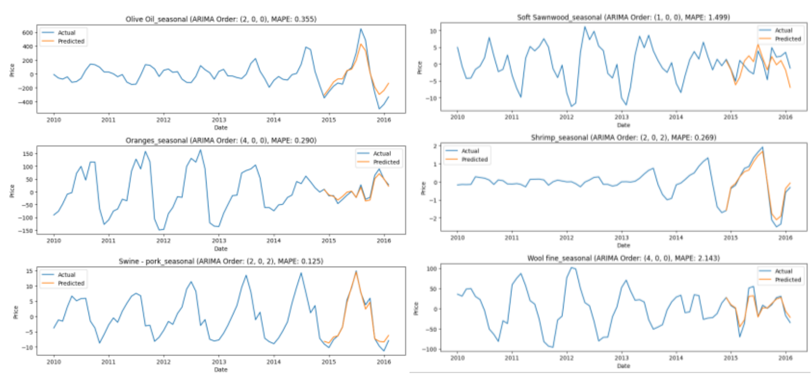

Fig. 6: Seasonal ARIMA modelling

This study also evaluates how much the models' performance differ when the residual component is analyze with the trend component and when the residual component is analyze with the seasonal component, known as the de-seasoned and de-trended time series respectively.

Model	Commodity	Component	MAPE	
			2010 – 2016	2012 - 2016

Exponential smoothing	Olive oil	De-seasoned	0.100	0.011
Exponential smoothing	Oranges	De-seasoned	0.098	0.062
Exponential smoothing	Swine pork	De-seasoned	1.552	0.485
Exponential smoothing	Sawn wood	De-seasoned	0.092	0.155
Exponential smoothing	Shrimp	De-seasoned	1.546	1.793
Exponential smoothing	Wool	De-seasoned	0.039	0.052
ARIMA	Olive oil	De-seasoned	0.085	0.006
ARIMA	Oranges	De-seasoned	0.083	0.043
ARIMA	Swine pork	De-seasoned	1.215	0.708
ARIMA	Sawn wood	De-seasoned	0.094	0.082
ARIMA	Shrimp	De-seasoned	1.103	1.535
ARIMA	Wool	De-seasoned	0.153	0.028
ARIMA	Olive oil	De-trended	1.137	1.078
ARIMA	Oranges	De-trended	3.574	2.643
ARIMA	Swine pork	De-trended	0.787	1.416
ARIMA	Sawn wood	De-trended	2.004	1.779
ARIMA	Shrimp	De-trended	2.165	0.981
ARIMA	Wool	De-trended	2.295	115.188

Table 3: Model evaluation for the de-seasoned and de-trended components

From the results shown above, ARIMA model outperforms exponential smoothing in all de-seasoned components except for swine pork. Even though the residual component only contributed to minor fluctuations in the time series, it has greatly impaired the models from capturing the overall pattern, resulting in significantly higher MAPE compared to Table 2.

Conclusion

Adopting a hybrid approach in modeling the individual components of the decomposed time series results in improved forecasting accuracy [6]. While some ML models can automatically identify and capture the seasonal component, the presence of multiple seasonal periods poses a challenge in accurately capturing all these patterns simultaneously. Hence, manual analysis of the autocorrelation of each dataset and iteratively removing non-stationary components remains necessary until the residual component exhibits true white noise characteristics.

References

[1] A. Kotillová, "Very short-term load forecasting using exponential smoothing and ARIMA models," *Journal of Information, Control and Management Systems*, vol. 9, no. 2, pp. 85-91, 2011

[2] A. M. De Livera, R. J. Hyndman, and M. Otto, "Forecasting time series with complex seasonal patterns using exponential smoothing," Journal of the American Statistical Association, vol. 106, no. 496, pp. 1513-1527, 2011.

[3] M. Theodosiou, "Forecasting monthly and quarterly time series using STL decomposition," International Journal of Forecasting, vol. 27, no. 3, pp. 695-711, 2011.

[4] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," Neurocomputing, vol. 50, pp. 159-175, 2003.

[5] Vagifa, "USA Commodity Prices," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/vagifa/usa-commodity-prices

[6] E. M. de Oliveira and F. L. C. Oliveira, "Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods," Energy, vol. 2018.