Robustness of ARIMA models in long durée forecast of exports in post-systemic shock scenarios

Paulo Martins r2015469[1], Beatriz Fonseca r20201599[1], José Marçal r20201581[1], Luana Periquito r20191286[1]
[1] Nova Information Management School

Introduction

Exports are an important factor to consider when assessing a country's macro-economical and financial outlook.

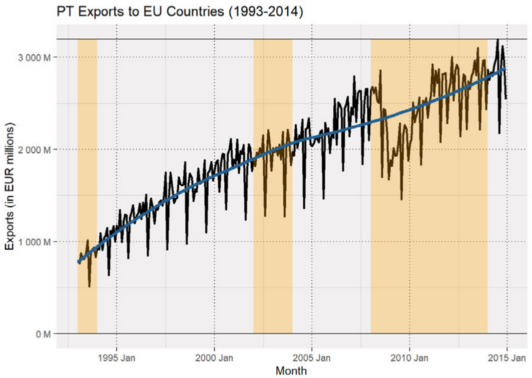

Figure 1: Figure 1. PT Exports to EU Countries (1993-2014). Time series of Portuguese exports to EU countries from 1993 to 2014 (black). Trending line (LOESS) of Portuguese exports from 1993 to 2014 (light blue). Recessionary periods (yellow).

At its core, this work seeks to demonstrate two-fold that; 1. ARIMA models are resilient to systemic shocks and thus should be at least considered with respect to forecasting exports; 2. backtested ARIMA model forecast accuracy might be indicative that a new normality might be on the horizon.

In total, 20+ years of data on Portuguese Exports was collected directly from the Instituto Nacional de Estatística website - Figure 1. Focusing exclusively on Portuguese exports to Intra-EU countries, the dataset was split into training and testing sets (backtesting) and the resulting models' accuracies were compared with established benchmark forecasting models.

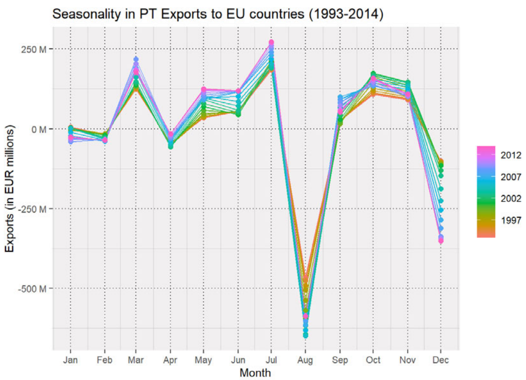

Figure 2: Figure 2. Seasonality in PT Exports to EU countries (1993-2014). Years in color label.

Methodology

Initially following the Box-Jenkins methodology, we tested the ARIMA models weak-stationarity of the response variable (Exports in EUR), through the Augmented Dickey-Fuller test. The response variable was log-transformed to stabilize its level variance, and seasonal differences were calculated.

Several candidate models were tested, with 1. Ljung-Box, 2.Shapiro-Wilks test; for auto-correlation of their residuals and their normality, respectively. Candidate models that met these requirements were compared on the basis of 1.lowest AICc score; 2. model complexity. And selected on a balance of the metrics: mean absolute proportional error (MAPE) and continuous ranking probability score (CRPS), along with a visual inspection of the forecasted timeseries plot.

Semi-final models were cross-validated using the stretch tilling method, with an initial seed of 120 months (12 years) and a periodic, these were culled using a balanced of performance metrics: root mean squared standard error (RMSSE), MAPE, CRPS, mean absolute error (MAE). Final model was chosen based on weighted average of MAPE "skill" and CRPS "skill", where skill is relative metric performance vis. benchmark (random walk - Naive model).

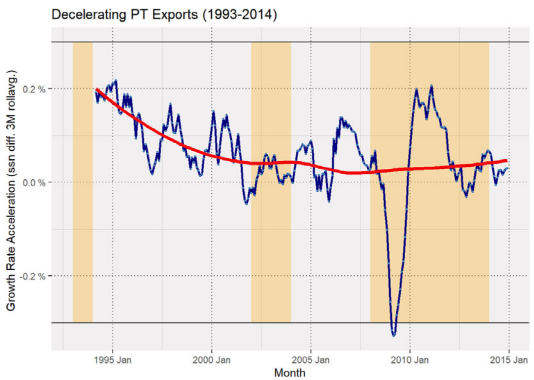

Figure 3: Figure 3. Decelerating PT Exports (1993-2014). Time series of log-transformed, seasonally differenced Portuguese exports to EU countries from 1993 to 2014, 3 month moving average (black). Trending line (LOESS) of Portuguese exports from 1993 to 2014 (red). Recessionary periods (yellow).

Results[1]

Surprisingly, the model with best forecasting power - measured in terms of skill vis. a Naive model on the held out data - was the SARIMA(100011). This model failed the LJung-Box test, yet passed the Shapiro-Wilks. We posit that this was due to Exports in Portugal being a superset of seasonal sub-series, where small but distinct movements require their own components to net the series of correlations; the best found runner-up was SARIMA(700316) for which not enough data is available to compute the cross-validation. Moreover, in general, the AICc proved to be a deceiving metric, as the series included periods of recession and economic growth, for which cyclical and not periodic movements are relevant and may perturb the calculation, as most models with high AICc had either very big forecast intervals, or awful point forecasts.

SARIMA100011

$$(1 - 0.0827B)(1 - B^{12})y_t = (1 + (-0.6907)B^{12})\varepsilon + 0.0093$$

model	MAPE	MAPE skill	CRPS	CRPS skill	Final Decision
ETS_MAM	7.87	0.62	271161986	0.53	0.57
SARIMA100011	8.60	0.58	223570437	0.61	0.60
SARIMA700116	7.74	0.62	291879986	0.49	0.56
SNAIVE_DRIFT	9.46	0.54	265080246	0.54	0.54

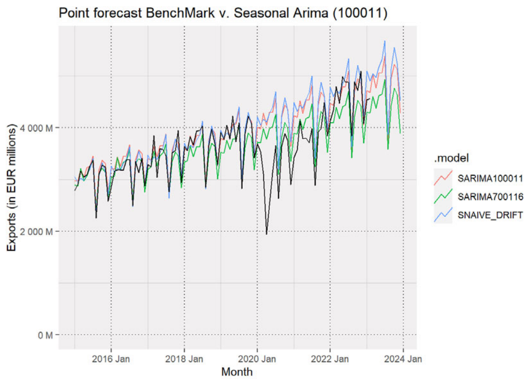

Figure 4: Figure 4. Point forecast BenchMark v. Seasonal Arima (100011, 700116). Model in color label.

Conclusions

Figure 2 detailed that growth rate had been decelerating over the period 1995-2014. This trend was apparently being inverted in the years proceeding and up to COVID-19 - causing Portuguese exports to plummet. Predicatively, our ARIMA models showed no forecasting power during the supply shock that the pandemic ensued, as they use the past to predict the future strictum sensum. However, our backtests have shown that ARIMA models can recover predictive power as the drivers of exports return to their previous levels - Figures 3, 4.

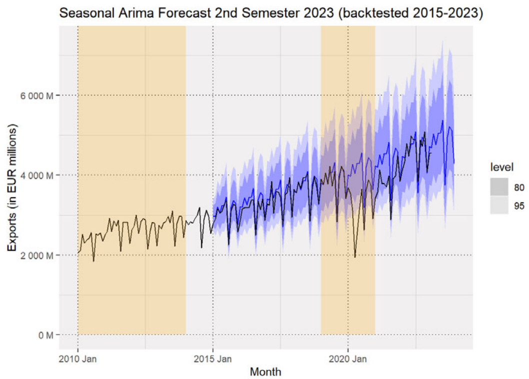

Figure 5: Figure 5. Seasonal Arima Forecast 2nd Semester 2023 (backtested 2015-2023). Time series of Portuguese exports 2010-2023[February] (black). SARIMA 100011 backtest, forecast and confidence intervals (blue line and shaded areas). Recessionary periods (yellow).

References

https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&contecto=pi&indOcorrCod=0005720&selTab=tab0

https://www.ffms.pt/pt-pt/estudos/crises-na-economia-portuguesa

Hyndman, R., & Athanasopoulos, G. (2021). Forecasting: Principles and Practice. (3rd ed.) OTexts. https://otexts.com/fpp3/

SHAPIRO, S. S., & WILK, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3–4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Footnotes

[1]We initially follow Box-Jenkins methodology, due to poor performance of models obtained and the reasons claimed in Results about possible auto-correlations in subsets of the dataset (or other unknown phenomena), every model's skill was calculated, regardless of failing on any one arbitrary test that would have disqualified it, if the model demonstrated good metrics otherwise.