INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

e-invoice time series nowcasting with R

Bruno Lima, Almiro Moreira, João Poças, Sofia Rodrigues

03-10-2022

Background

Portuguese Tax Administration implemented a mandatory system of reporting all invoices, by electronic way, issued for any commercial transaction (e-invoice):

- taxable amounts aggregated by issuer and acquirer VAT numbers;
- data received monthly;

When missing values are detected, for more relevant issuers, we need to *nowcast* those values.

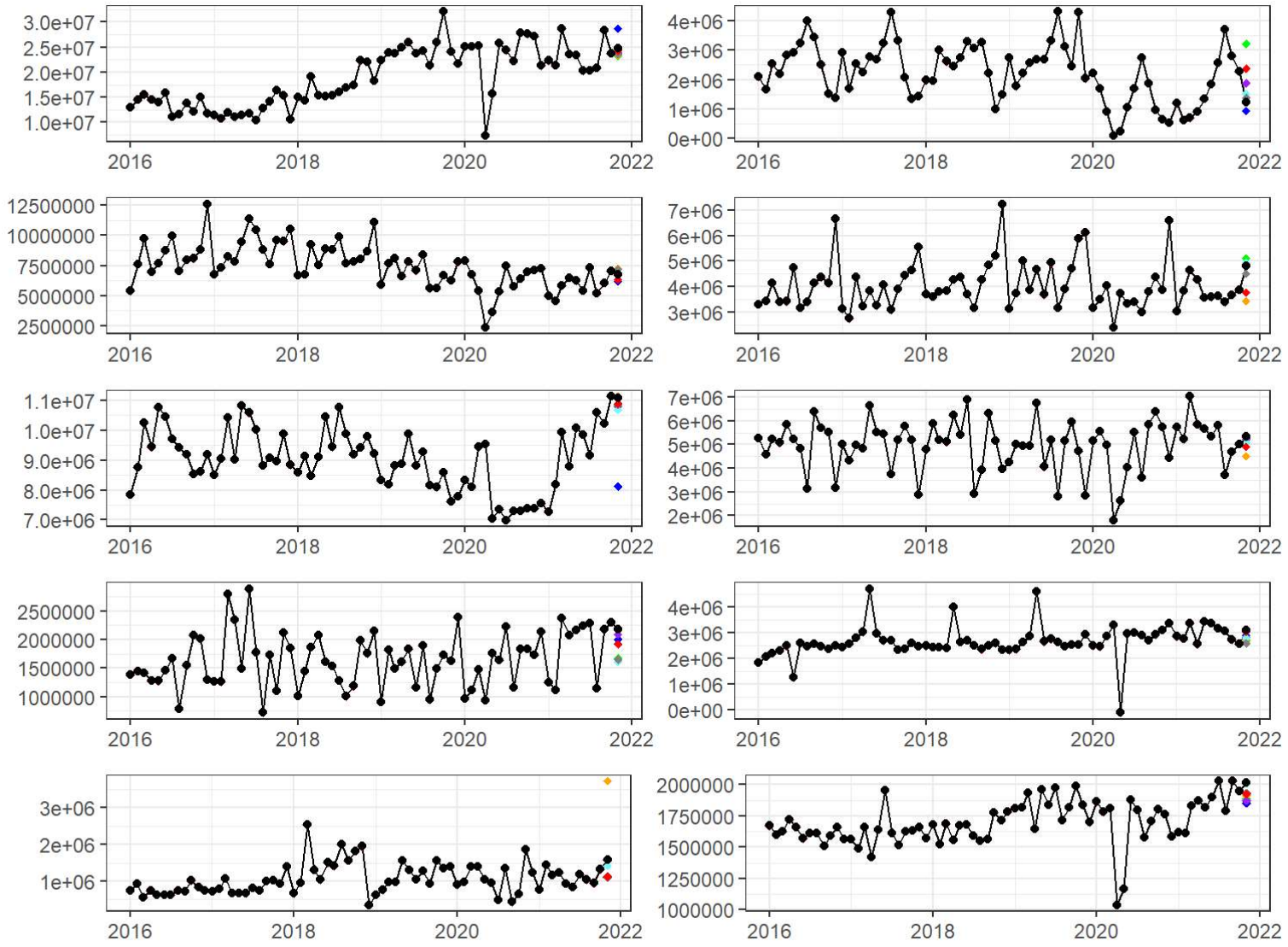
Using the historical univariate time series of the taxable values (VT) of each issuer, we need to do an immediate forecast and make all the data available to Statistical Portugal users.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Background (cont.)

e-invoice:

Time series are very heterogeneous and difficult to model.



Aim

Evaluate nowcasting methods, available in R packages:

- applying them to different and heterogeneous time series;
- comparing the results with the next data received.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Methods

package	method	description
{RJDemetra}	X-13ARIMA-SEATS (**X13**)	the seasonal adjustment software officially recommended to the members of the European Statistical System (ESS)
{RJDemetra}	TRAMO/SEATS (**TS**)	Time series Regression with ARIMA noise, Missing values and Outliers"(TRAMO) and "Signal Extraction in ARIMA Time Series"(SEATS)
{forecast}	**ETS**	Exponential smoothing state space model
{forecast}	Auto ARIMA (**AA**)	Fit best ARIMA model to univariate time series

package	method	description
{imputeTS}	Kalman Smoothing (**KS**)	Missing Value Imputation by Kalman Smoothing and State Space Models
{prophet}	Facebook's forecasting procedure	Additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects

Methods (cont.)

Nowcasting methods were applied to raw data and normalized data ($\frac{VT-min(VT)}{max(VT)-min(VT)}$).

Measuring errors:

- Root Mean Square Deviation:

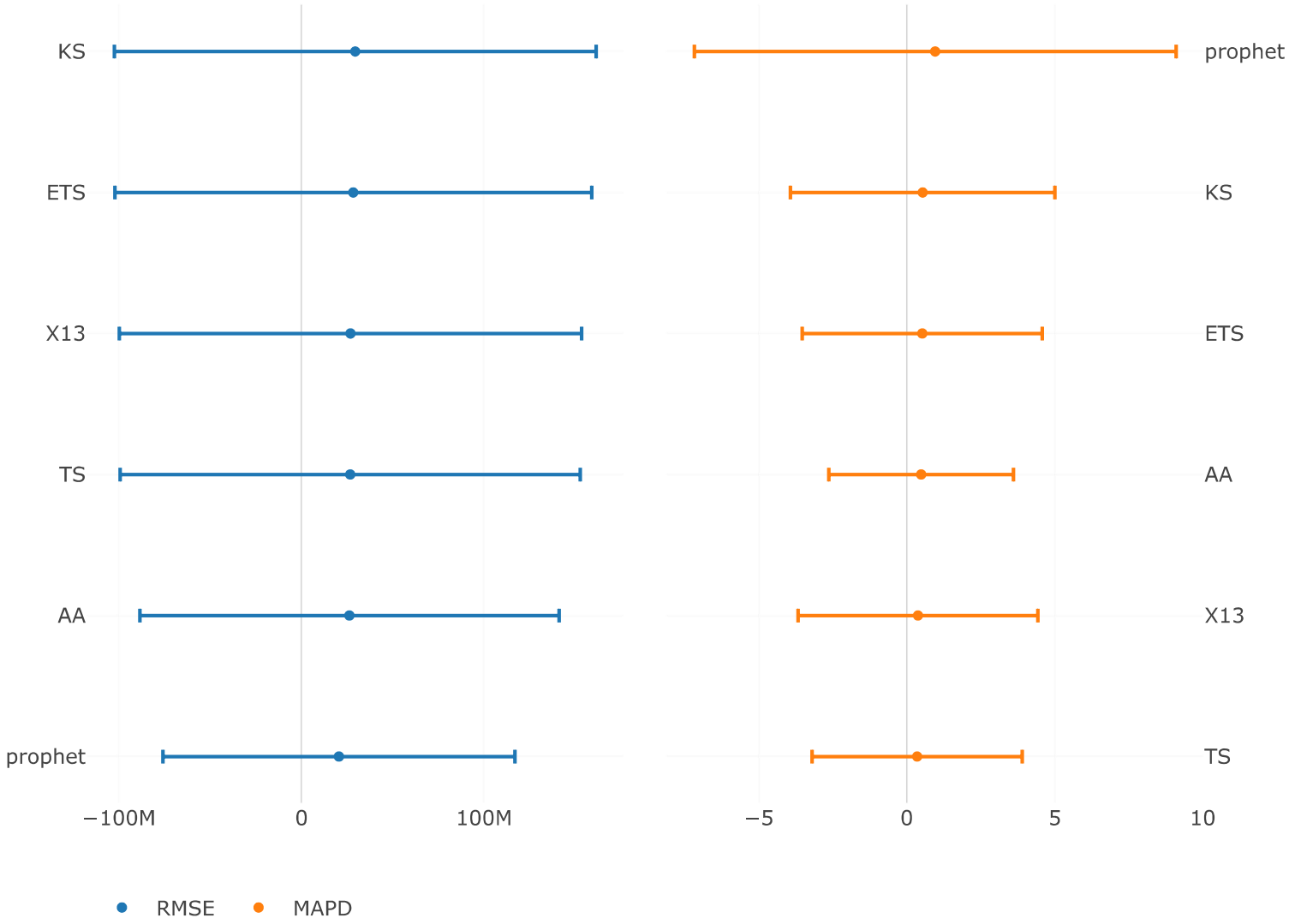
$$RMSE = \sqrt{\frac{1}{N}\sum(f-o)^2};$$

- Mean Absolute Percentage Deviation:

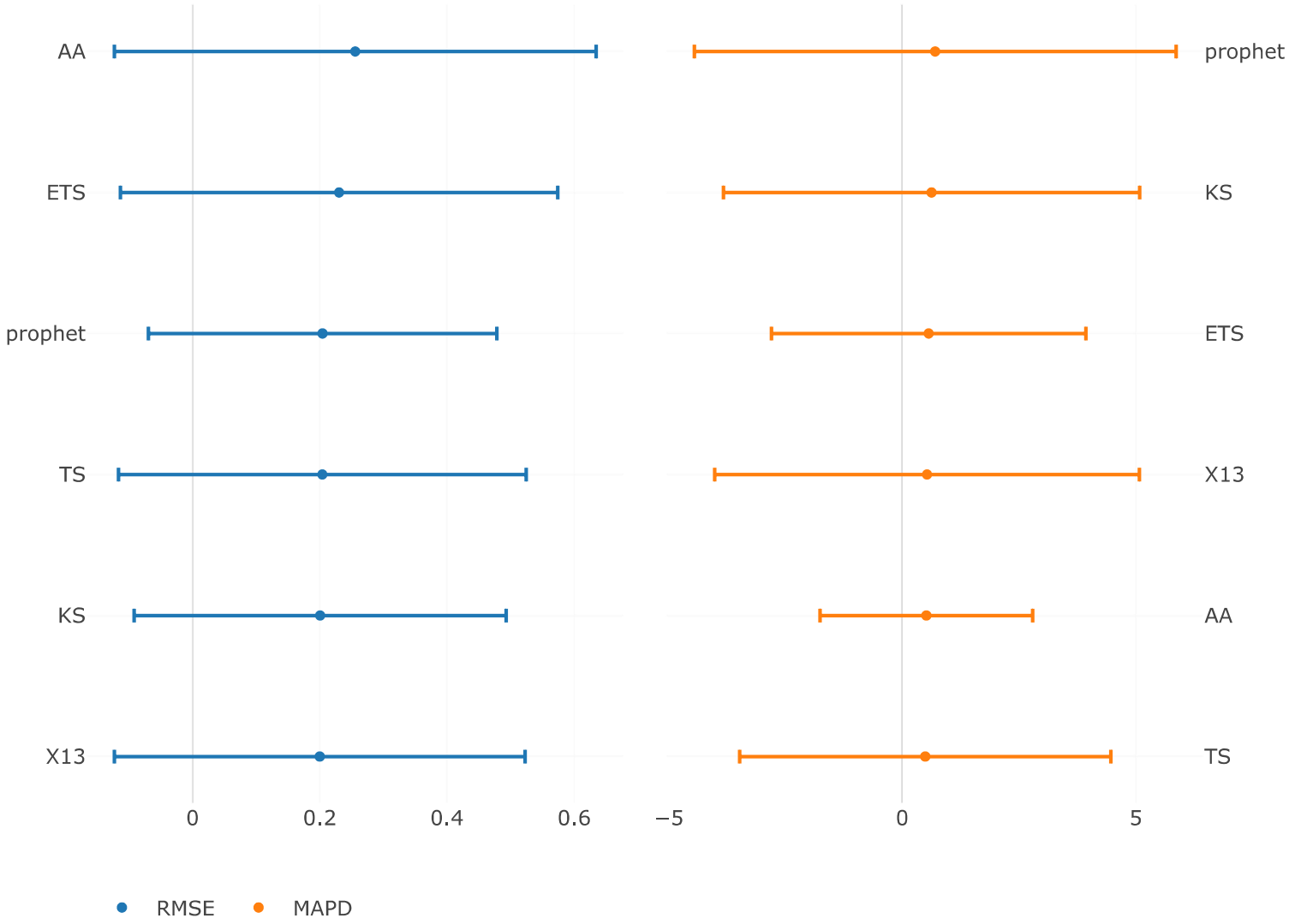
$$MPAD = \frac{100\%}{N}\sum\left|\frac{f-o}{o}\right|;$$

(f *forecast*, o *observed*)

nowcasting performance (raw data)



nowcasting performance (normalized data)



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Conclusions

- times series are too heterogenous to find out a '*one fits all*' method;
- MAPD is a better metric to compare nowcasting performance methods, in such heterogeneous time series;
- to compare nowcasting results, we should use normalized data;
- {prophet} performes poorly (both for raw and normalize data) according to MAPD;
- {JDmetra} methods (TRAMO/SEATS and X-13ARIMA-SEATS) showed the best results according to MAPD.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Cheers from Portugal

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Thank you for your time.