Short Term Power Burn Model

GAS Analytics

Rationale:

Dependent:

- gas demand

Covariates:

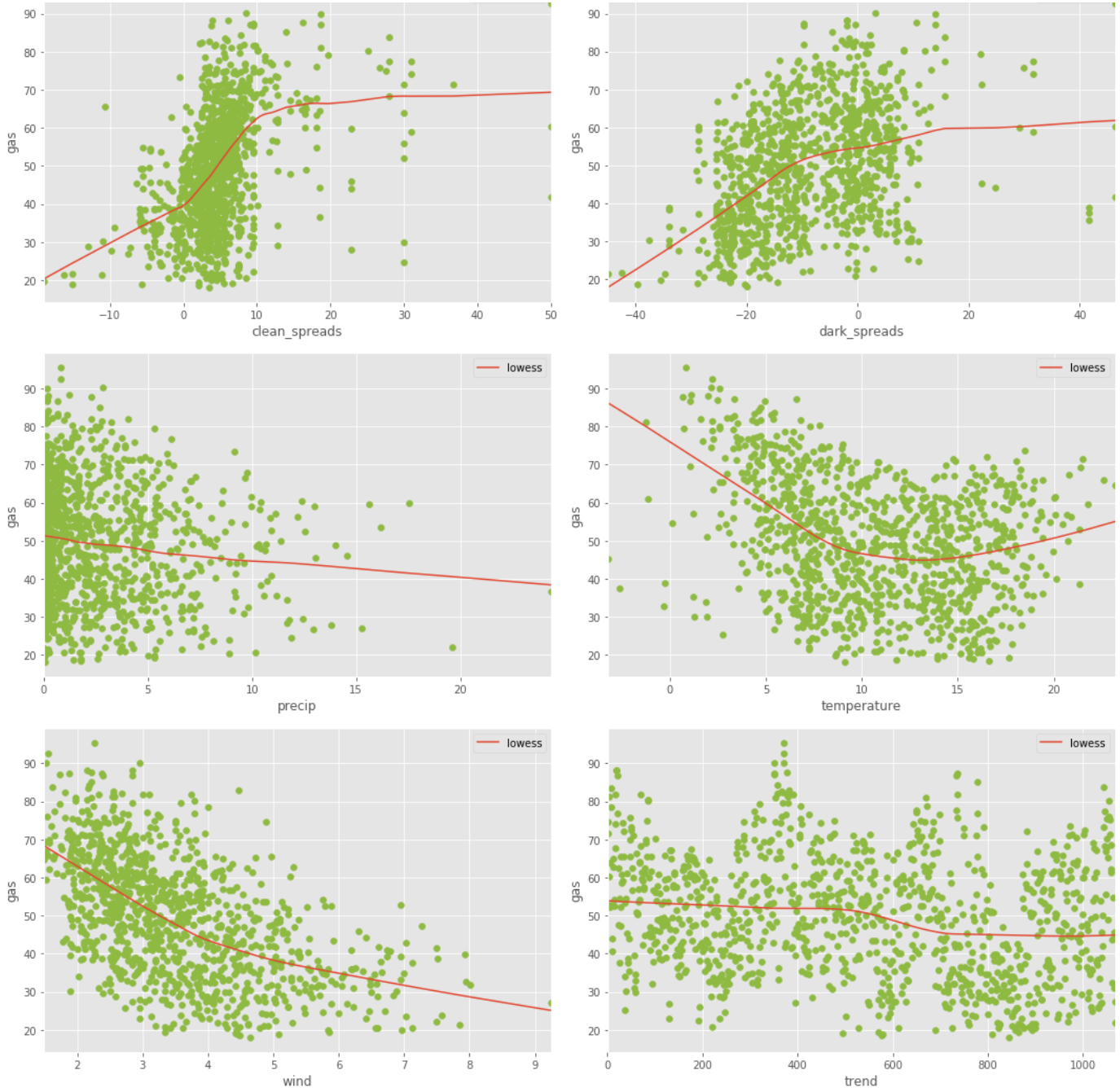
- clean spreads
- dark spreads

- temperature
- wind
- precipitation

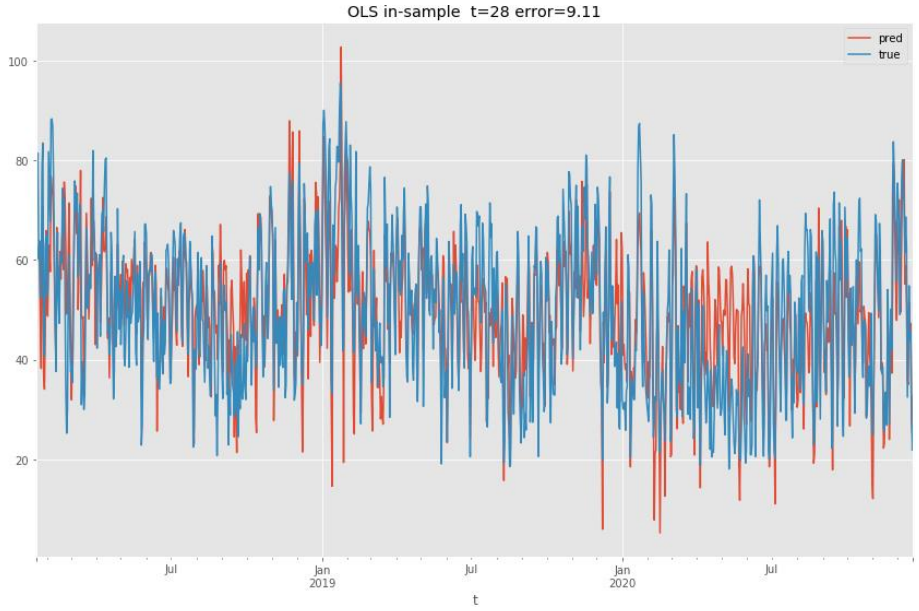
- monday_thursday flag
- Fourier series

Data:

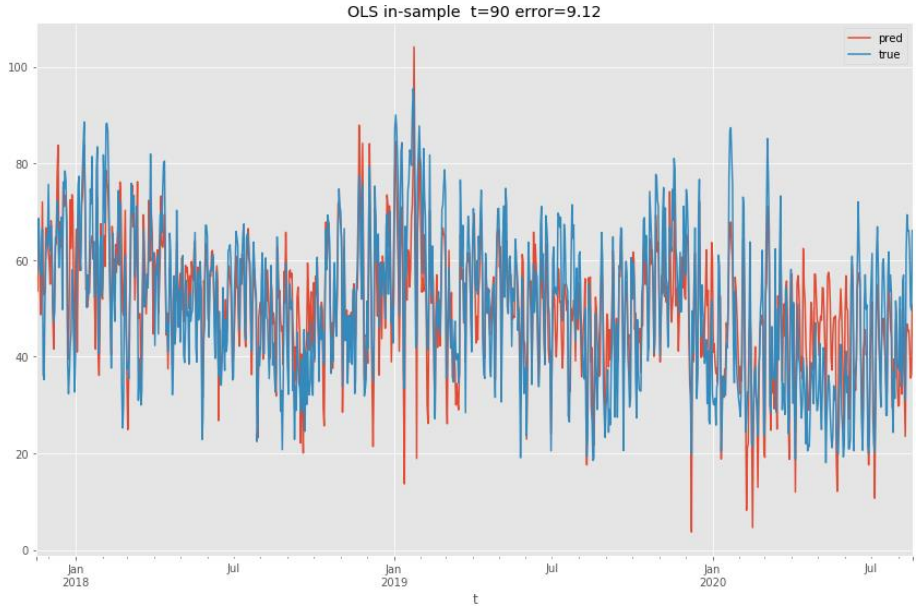
- 3 years daily sampling

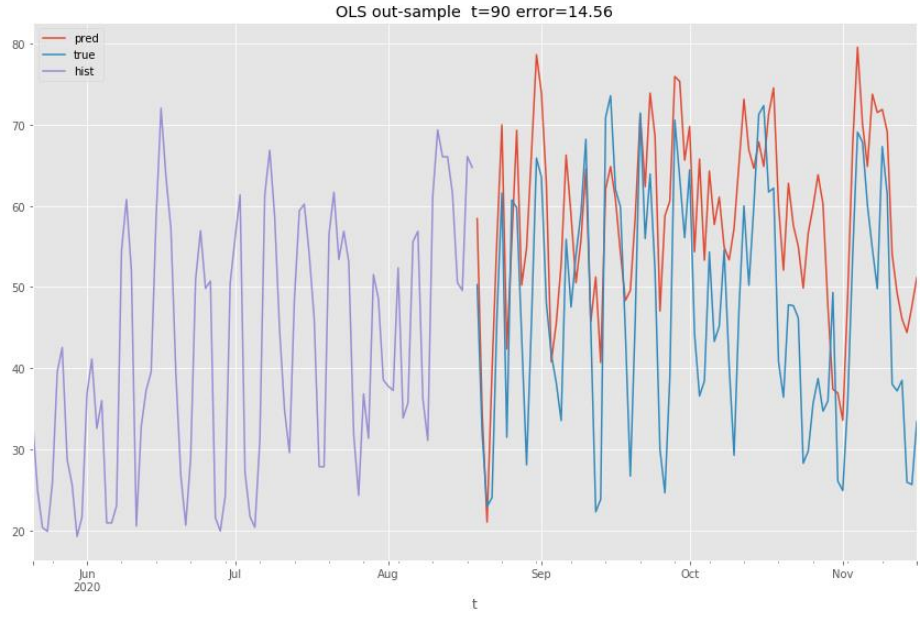


Calibrator (exog): Linear Regression



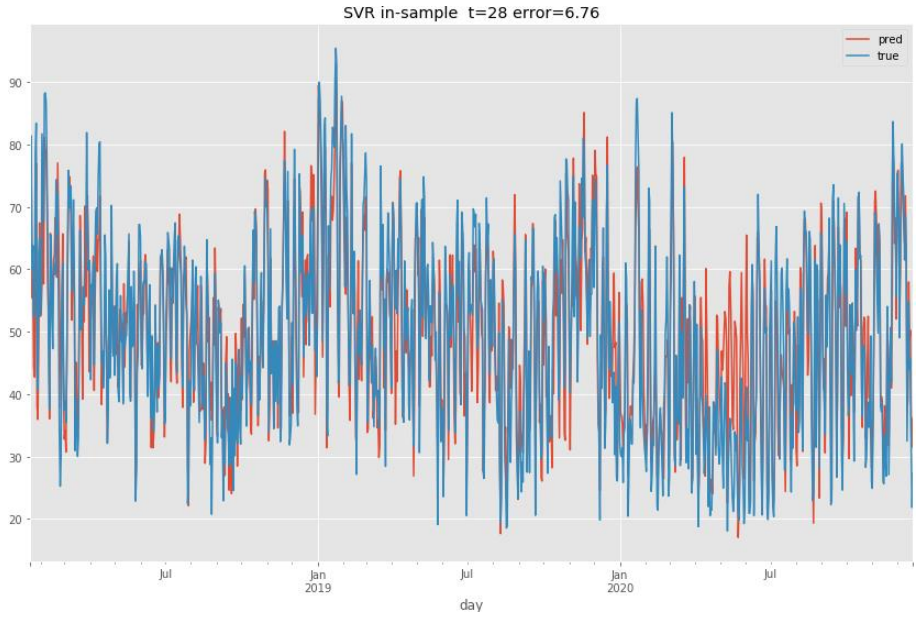
CV= 9.43 (2.08)

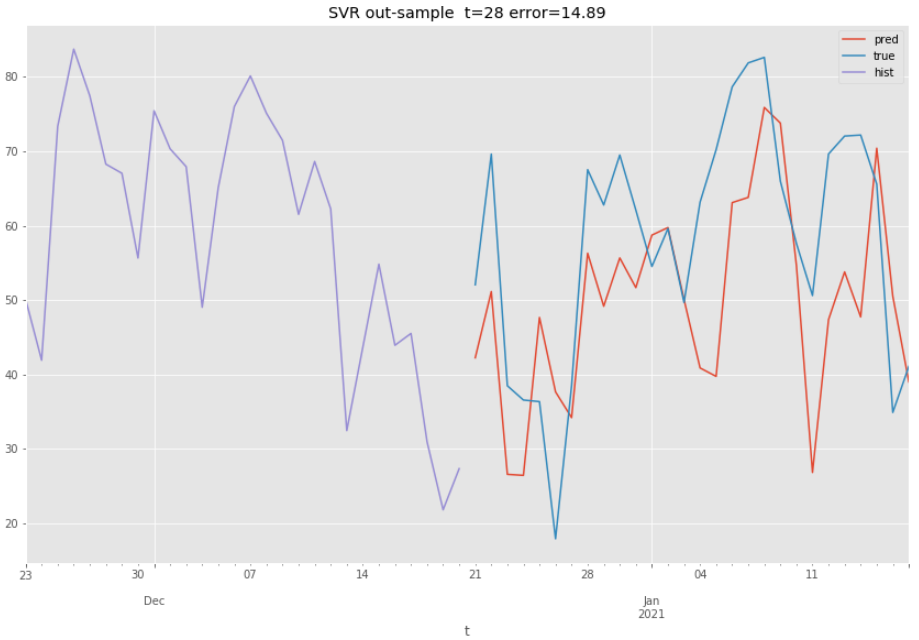


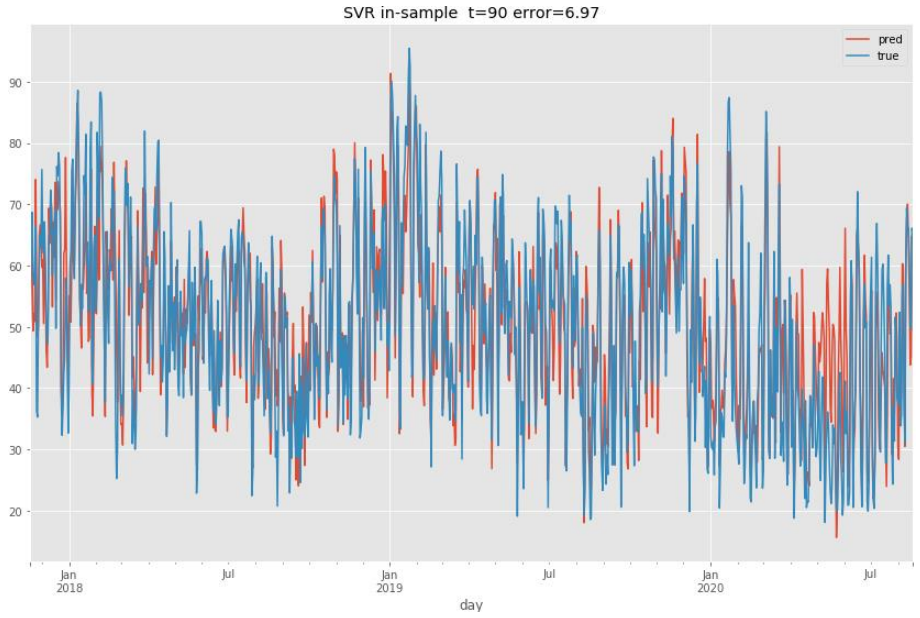


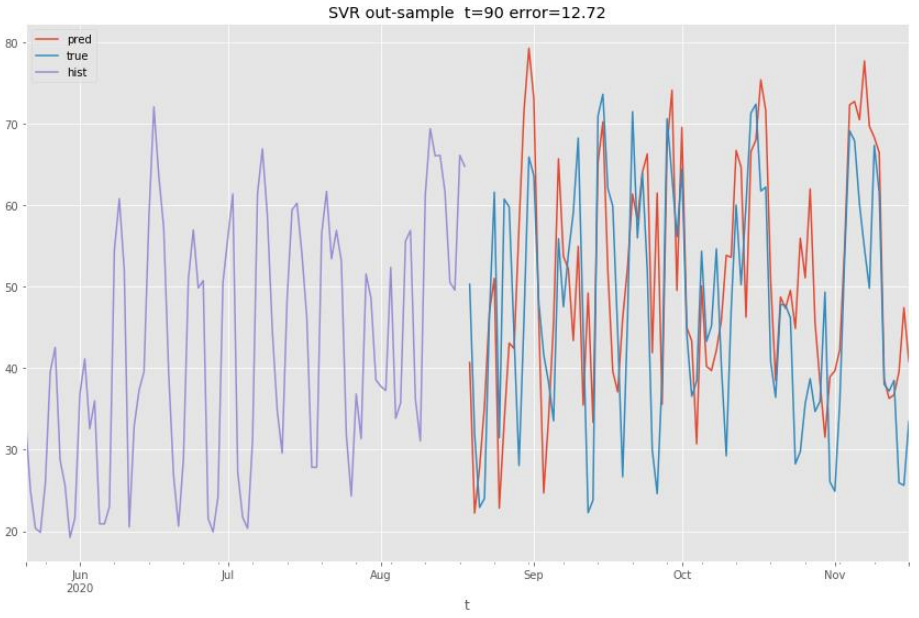
CV= 10.57 (1.79)

Calibrator (exog): Support Vector (kernel) Regression

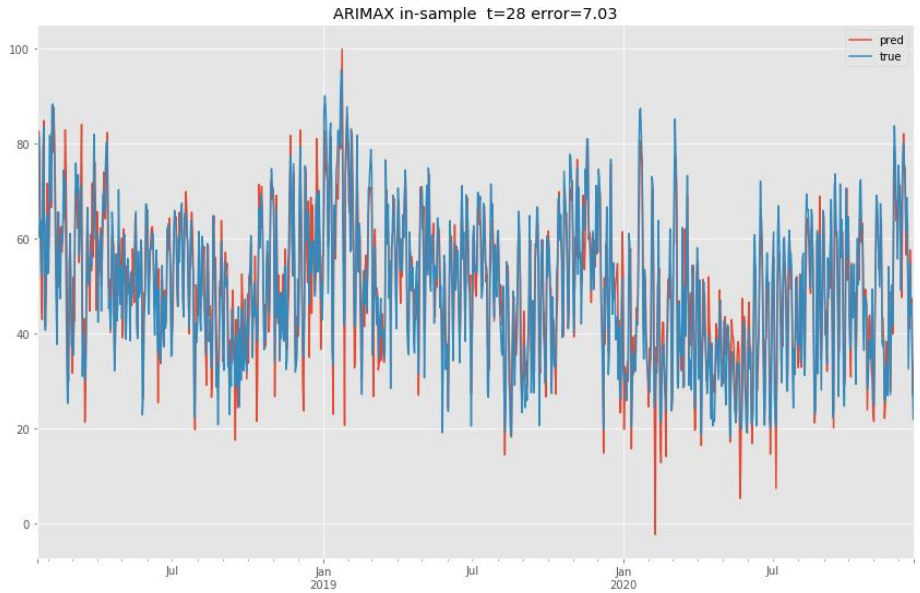


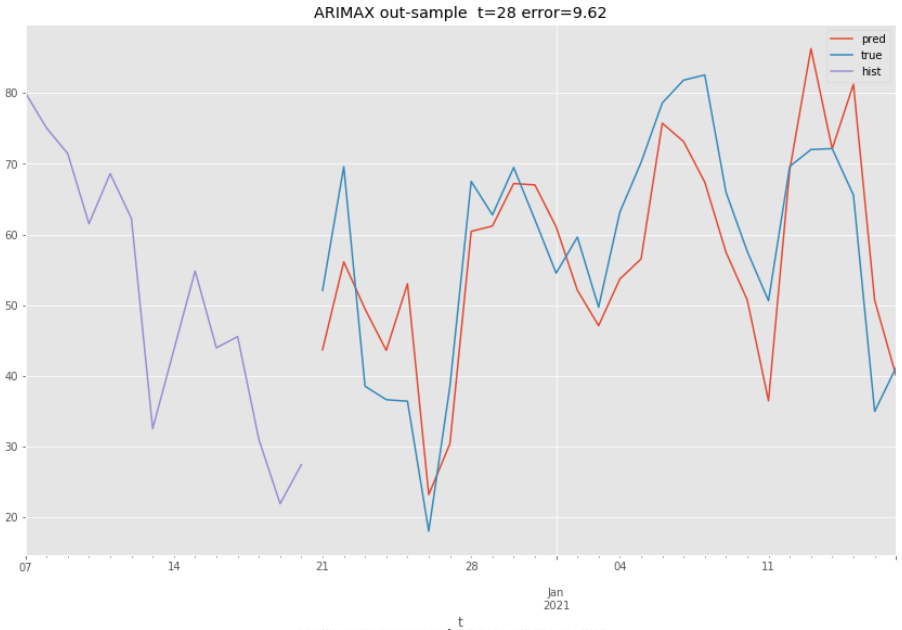




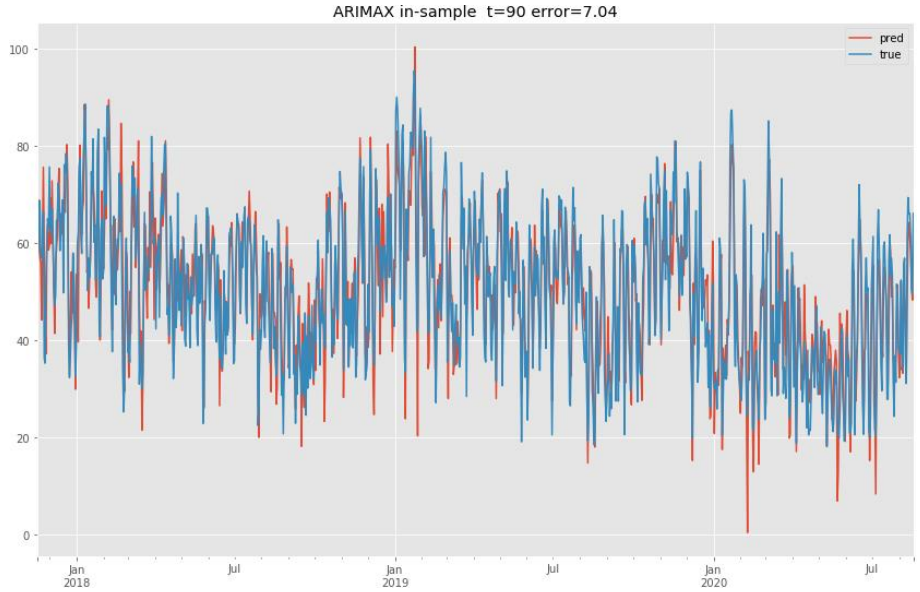


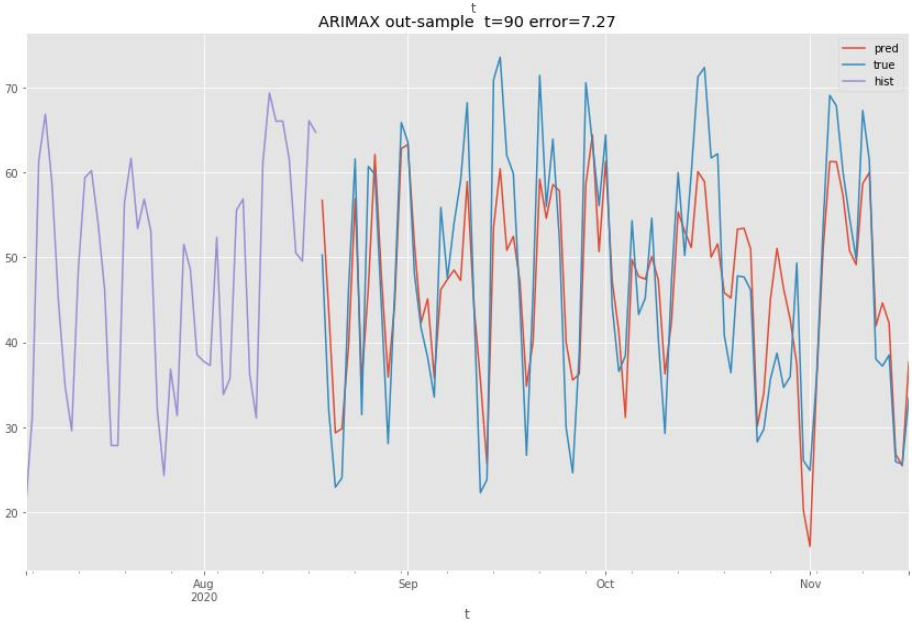
Calibrator (exog): SARIMAX





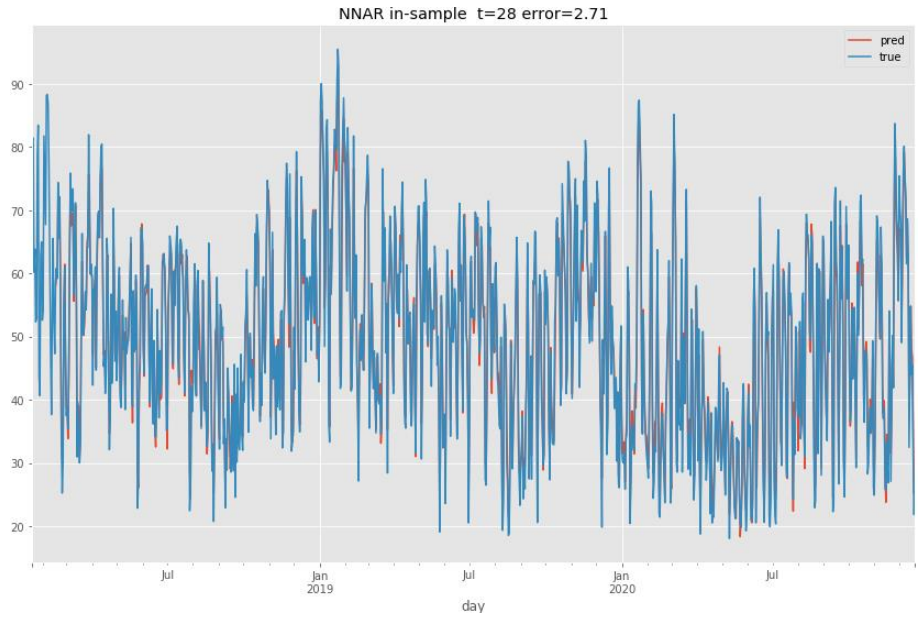
CV= 8.44 (1.71)

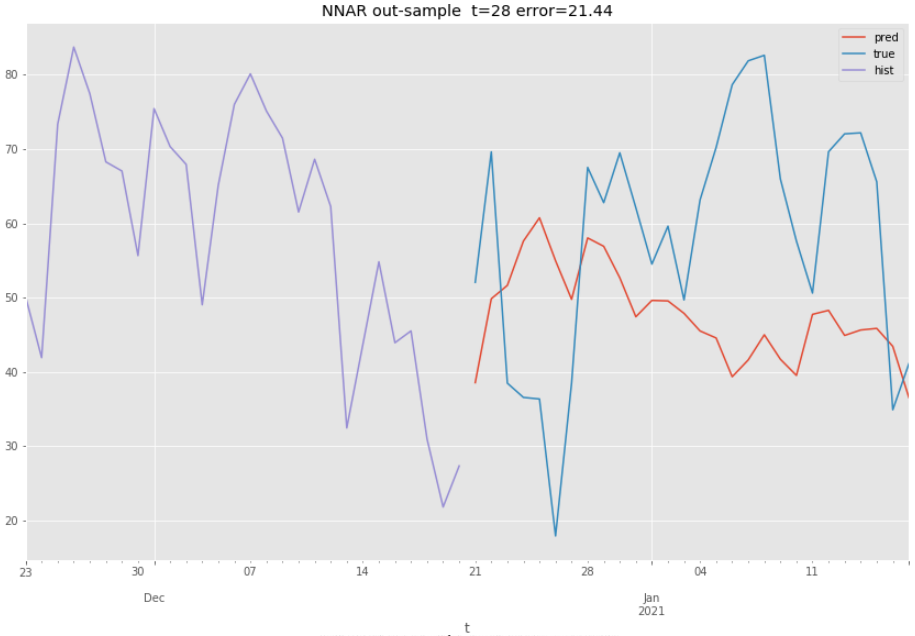


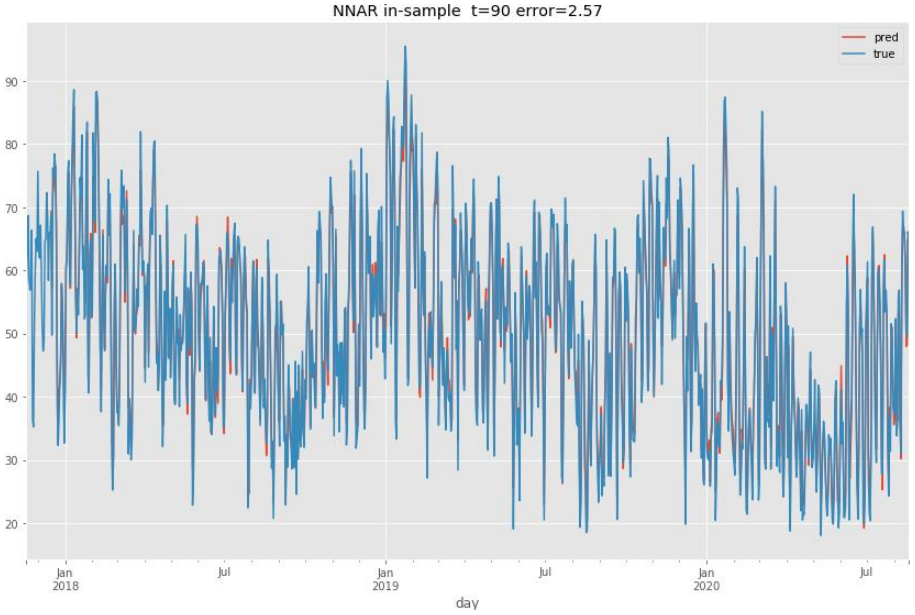


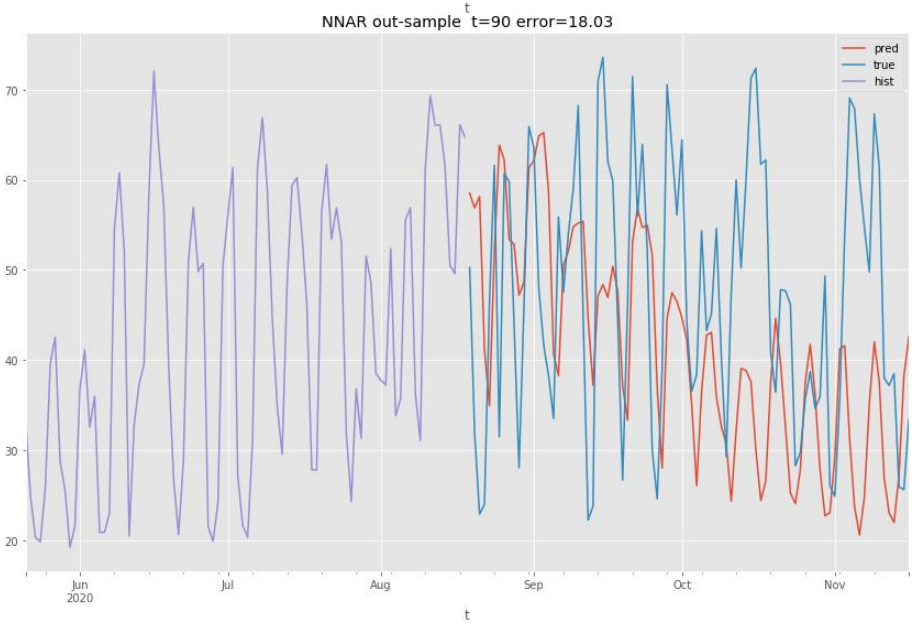
CV= 10.37 (2.41)

Calibrator (endog): Neural Network Autoregression

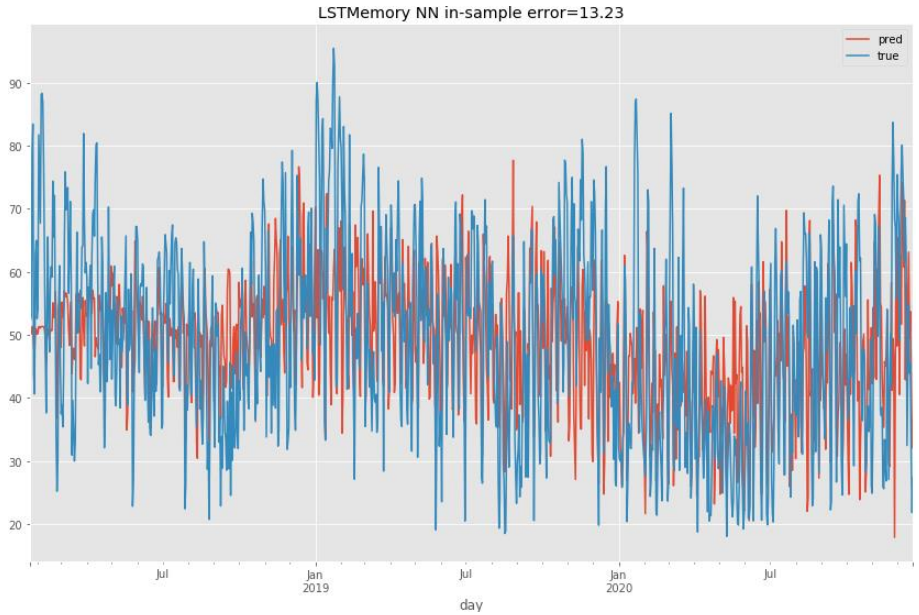


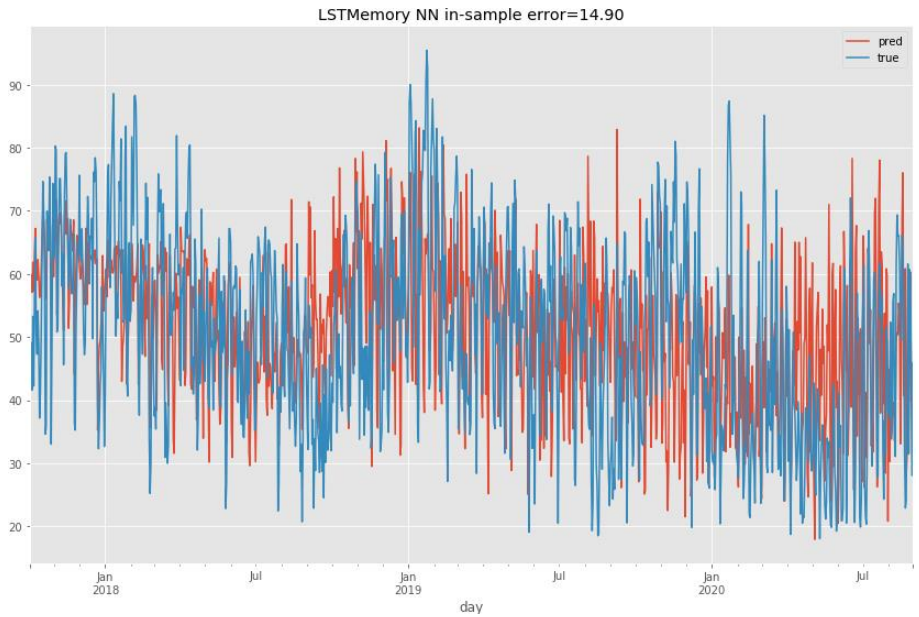


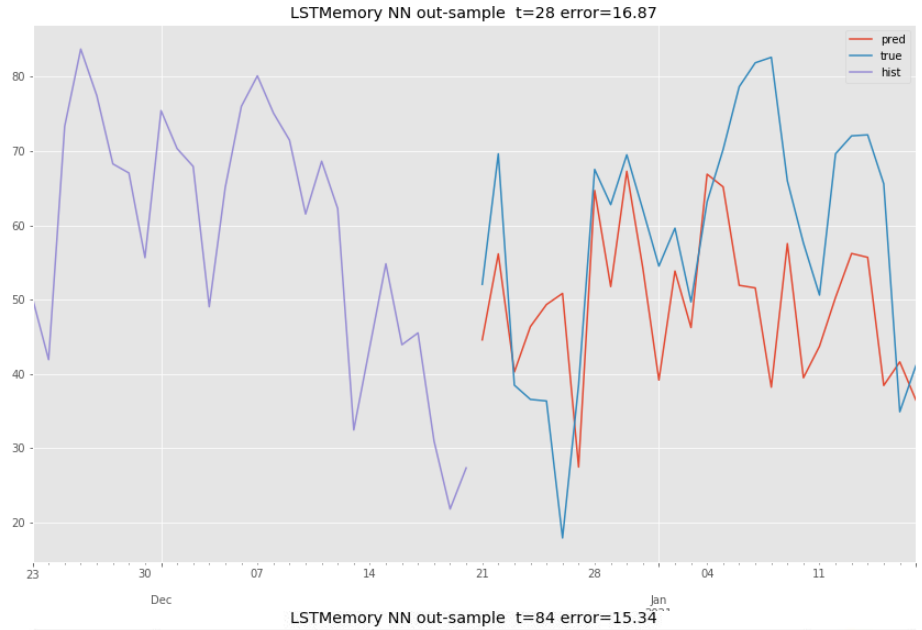


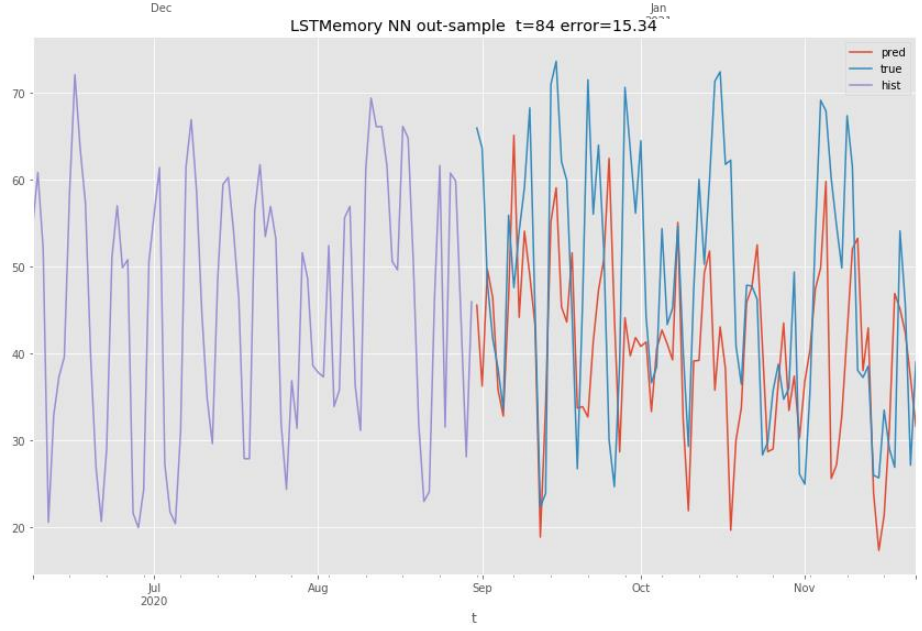


Calibrator (endog): Neural Network Long Short Term Memory









Appendix A:

Model	Parameters	Description
EXOGENEOUS RELATIONSHIP:		
Linear Regression	const, trend, clean_spreads, precipitation, temperature, wind, S1-7, C1-7, S2-7, C2-7, S3-7, C3-7, mo_th_yes	Modelling in levels Exog variables: grid search with BIC criterion 3 terms of Fourier series at weekly frequency
SARIMAX	SARIMAX(1, 0, 1)x(1, 0, 1, 7) const, dark_spreads, precipitation, temperature, wind,	Modelling in levels Specification: grid search with BIC criterion Exog variables: grid search with BIC criterion
Support Vector Regression	type="eps-regression" kernel='radial' • cost= 8 • gamma= 0.0625 • epsilon= 0.3	Dependent variable and features scaling: standardization Specification: grid search with 10-fold CV
ENDOGENEOUS RELATIONSHIP:		
Neural Network Autoregression	• Model: NNAR(29,1,15)[7] • Average of 20 networks, each of which is a 29-15-1 network with 466 weights options were - linear output units	Dependent variable scaling: standardization
Long Short Term Memory Neural Network		

Appendix B:

ALL DATA: IN TIME : ACTUALS 4y								OUT OF TIME: FUTURE
IN SAMPLE 3y								OUT SAMPLE 28days	
CROSS VALIDATION 3y:									
FOLD1 1y		FOLD2 1.5y		FOLD3 2.5y		FOLD4 3y			
train sample 0.8y	test sample 28days	train sample 1.3y	test sample 28days	train sample 2.3y	test sample 28days	train sample 2.8y	test sample28 days		
	error 1		error 2		error 3		error 4		
CV error = 1/4 * (error1 + error2 + error 3 + error4) **CV std = Standard Deviation (error1, error2, error3, error4)**									
IN SAMPLE ERROR								OUT SAMPLE ERROR	OUT OF TIME ERROR

error = Root Mean Square Error

Appendix C:

Modelling Framework

Design Matrix:

- Imputations
- Design Matrix:

$$t \mid y_t \xrightarrow{f} v_t, \mid x_t^1,\ x_t^2,\ \dots\ x_t^k \xrightarrow{g} g_1 x_t^1,\ g_2 x_t^2, \dots, g_k x_t^k$$

$$\mathrm{DM}_t = \mathrm{DM}_t(y_t, f, f^{-1}, \{x_t^i\}_{i=1,k}, \{g_t^i\}_{i=1,k})$$

Exploratory analysis

- Autocorrelation:
  - $\mathrm{ACF}(x_t)$
  - $\mathrm{PACF}(x_t)$
- Scatter Plots:
  - y_t next to v_t and x_t^k next to $g_k x_t^k$
  - y_t vs. x_t^j for $j = 1, k$ with LOWESS for dependency shape analysis
  - x_t vs. x_{t-h} with LOWESS for autocorrelation analysis
  - y_t vs. x_{t-h}^k for given k with LOWESS for lagged-leading relationship

Calibrator:

$$\mathrm{C}(\boldsymbol{HyperParams}) \rightarrow \mathrm{C}$$

Model:

$$\mathrm{M} = \mathbb{M}(\mathrm{C},\ \mathrm{DM}) \rightarrow \{\hat{\theta}_l\}_{l=1,m}$$

Model Specification

$$\{\hat{\theta}_l\} \xrightarrow[\text{GridSearch}]{I(\theta):\ AIC,\ AICc,\ BIC} \{\hat{\theta}_l^*\}$$

Model Selection:

Cross Validation

$$\mathrm{CV} = \mathbb{CV}(\mathrm{M}, \mathrm{Partitioning},\ \mathrm{Performance\ Metric})$$

$$\mathrm{C} \xrightarrow{\varepsilon_{CV}} \mathrm{C}^*$$

Residuals Diagnostics:

$$\hat{\varepsilon}_t = v_t - \hat{v}_t;\ \hat{\varepsilon}_t^{std};\ \hat{\varepsilon}_t^{stu}$$

$$\mathrm{RD} = \mathcal{RD}(\hat{\varepsilon}_t, \hat{\varepsilon}_t^{std}, \hat{\varepsilon}_t^{stu})$$

Forecast:

$$\mathrm{C}^*(\boldsymbol{DM}_{t+1} = g_i(x_{t+1}^i)_{i=1,k} \mid \{\hat{\theta}^*{}_l\}_{l=1,m}) = \hat{v}_{t+1} \xrightarrow{f^{-1}, y_t} \hat{y}_{t+1}$$