0 Introduction

0.1 Goal

- 1 step-ahead electricity consumption prediction
- single house
- consider weahter and calendar effects

0.2 Data

- household in London
- January 2012 to December 2013
- 30 min resolution
- hourly temperature records
- date and time information

0.3 Recap of Visualization and Data Analysis

- Plots of Holiday vs. weekday vs. weekend: widely varrying patterns
- Time-series plots: data has sharp peaks
- Box plots vs. time and temperature: the occurance of these peaks has no visible pattern, and are diagnosed as outliers
- Auto-correlation and partial auto-correlation: lags 1, 2, 48, 49, 7*48, 7*48+1 are the peaks of the PACF (lag 48 is a full day)
- seasonal differences cannot be easily detected

0.4 Summary

- Need different models for weekday/weekend/holiday
- Predicting the peaks can be challenging or impossible
- AR models with lagged versions of the time-series as features are suitable for the task

0.5 In This Notebook

- One-step-ahead consumption prediction of a fixed household
- Feature selection problem: determining suitable auto-regressors
- Evaluating fit results
- Studying the effect of day part (morning/noon/...) on prediction error

0.6 General Setup

- Single fixed household

- Only Thursdays that were not a holiday (see section 1)
- Dependent variable (target): one-step-ahead consumption
- Regressors (features): constant / hour of day (x,y) / day of year (x,y) / temperature / auto-regressors(delayed versions)
- The same model is used for all parts of Tuesdays (morning/noon/...) and all seasons

Setup is the same in all sections unless specified

---

1 Preparing The Data for Regression

1.1 Cleaning

- interpolate missing values
- merge consumption, weather, and calendar data to a single dataframe
- no outlier removal method
- convert time of day and day of year to x,y
- min-max scaling to [-1, 1]

1.2 Filtering the data

- fix one day, e.g. Thursday
- remove all holidays

1.2 Constructing regression dataframe

- dependent variable: consumption at time interval t
- regressors: temperature, time of day (x,y), day of year (x,y), auto-regressors(delayed versions of the signal)
- dealing with holidays: if holidays appear among auto-regressors, replace them with the previous day. If the previous day is weekend, holiday, or already in the features, keep going 1 day back.

```
number of holidays
25.0

   ***   DATA IS READY FOR USE   ***


          LCLid stdorToU    Acorn Acorn_grouped      file
3871  MAC000068      Std  ACORN-L     Adversity  block_77
All groups:
['ACORN-' 'Affluent' 'Comfortable' 'Adversity' 'ACORN-U']
```

2 Feature Selection

2.0 Theory

- Question: which delays of the signal should be in the regressors?
- Feature selection problem: finding the most appropriate set of regressors, in this problem, delayed versions
- More regressors (features) maximize the explanatory power of our model, but could increase variance in the prediction
- The fit improves after feature reduction if MSE associated with parameter estimates will be smaller than the reduction in variance

Methods

- Regressors that are highly correlated with the dependent variable: univariate test, filtering is based on an arbitrary threshold or manually fixed number of regressors, or tuned by cross-validation
- Recursive feature elimination (RFE): start with all regressors, recursively remove less important regressors, repeat until a desired set of features remain.
- RFECV: RFE + tune number of features with CV
- Lasso: reduces several coefficients to zero leaving only features that are truly important, by adding L1 penalty to training objective (MSE)

Evaluation metrics

- Different number of regressors selection methods lead to different models, need to evaluate and compare them
- MSE
- R-squared: proportion of variation in the outcome explained by regressors (the same as the squared correlation between actual and predicted values for OLS)
- Adj. R-squared: adjusted with the number of features, adjusted R-squared is equal to 1 - (n-1)/(n–k-1)*(1-R2)
- AIC = -2log(likelihood) + 2*number of features

Notes

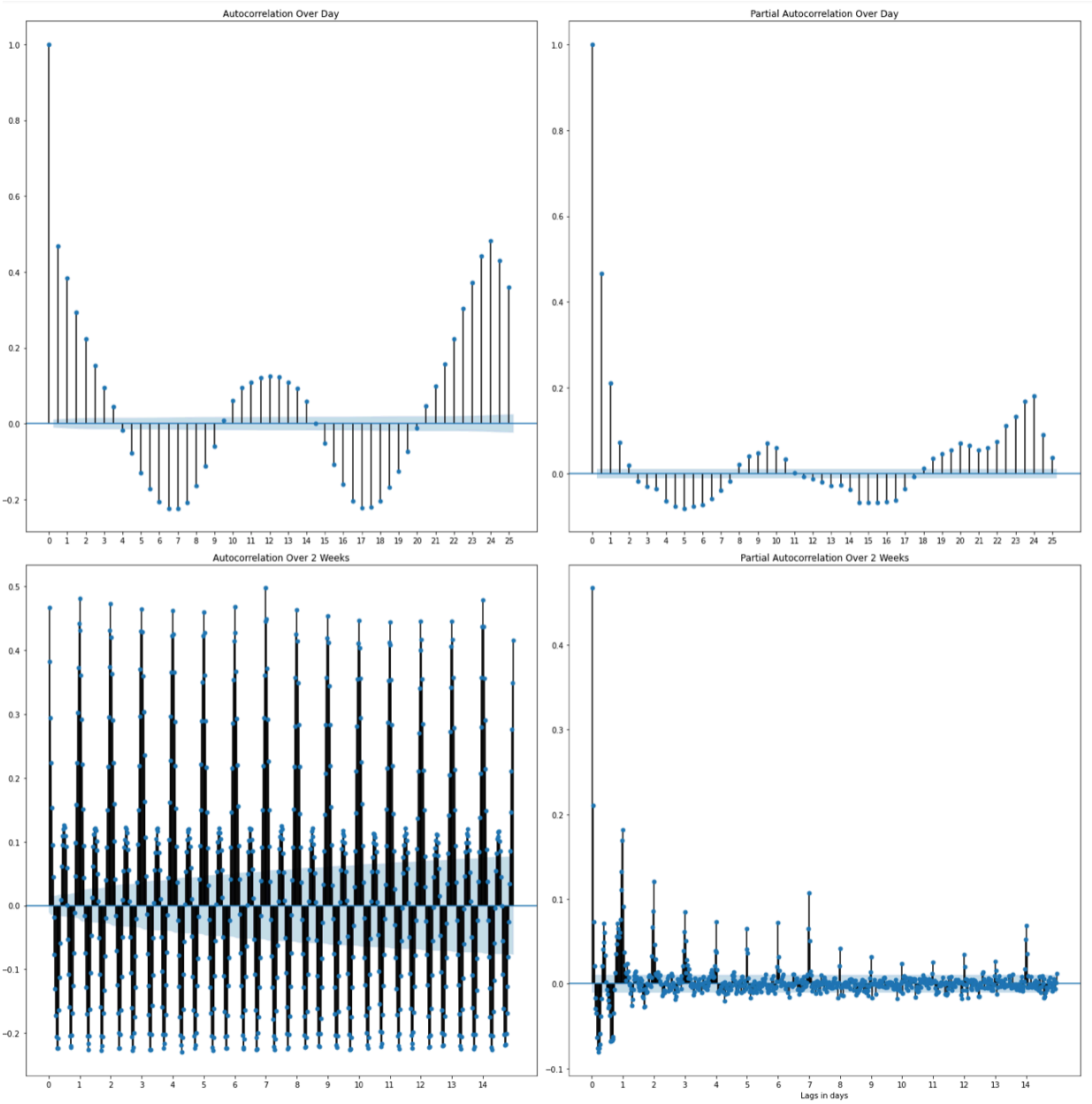
- For a fixed number of regressors, all methods obtain the same set of selected features
- Selection based on correlation + tuning number of features with CV is presented in

the following

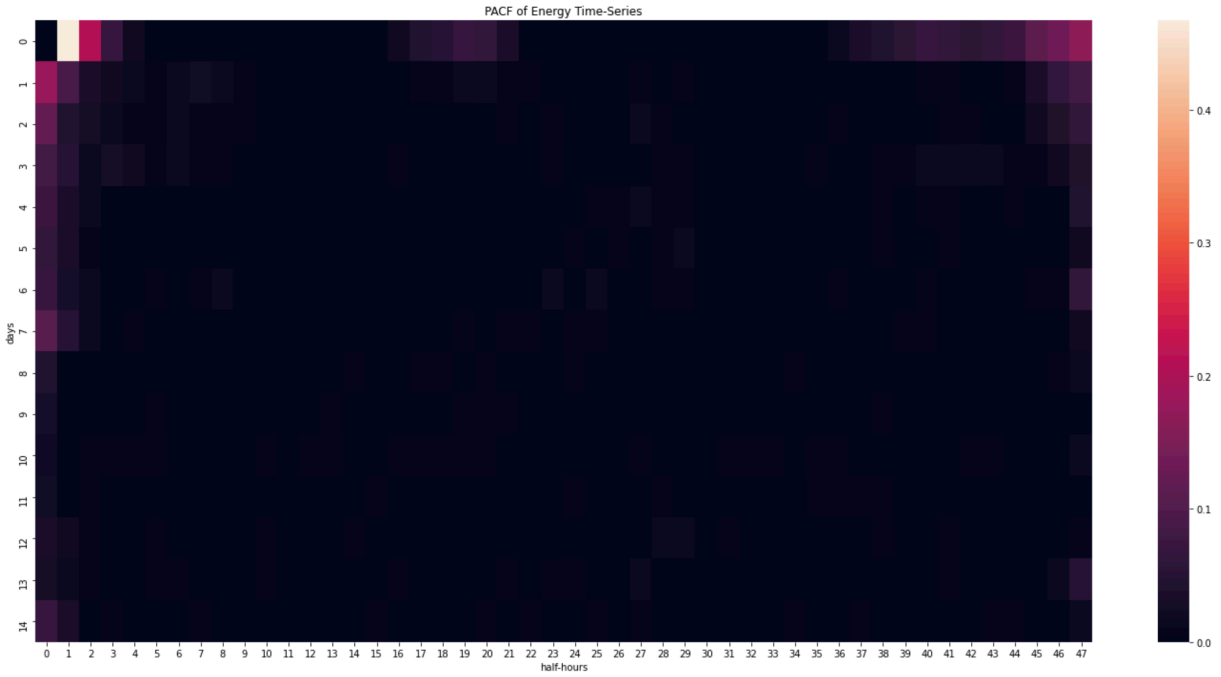
- In this problem, sum of squared errors is small and AIC is monotonically increasing with the number of features. Hence, AIC is not used for comparing models.
- Adj. R2 and MSE are used as evaluation metrics

2.1 Method 1: highly correlated features

- Plot auto-correlation and partial auto-correlation (PACF)
- Select highest picks in PACF
- Number of selected regressors is tuned by CV

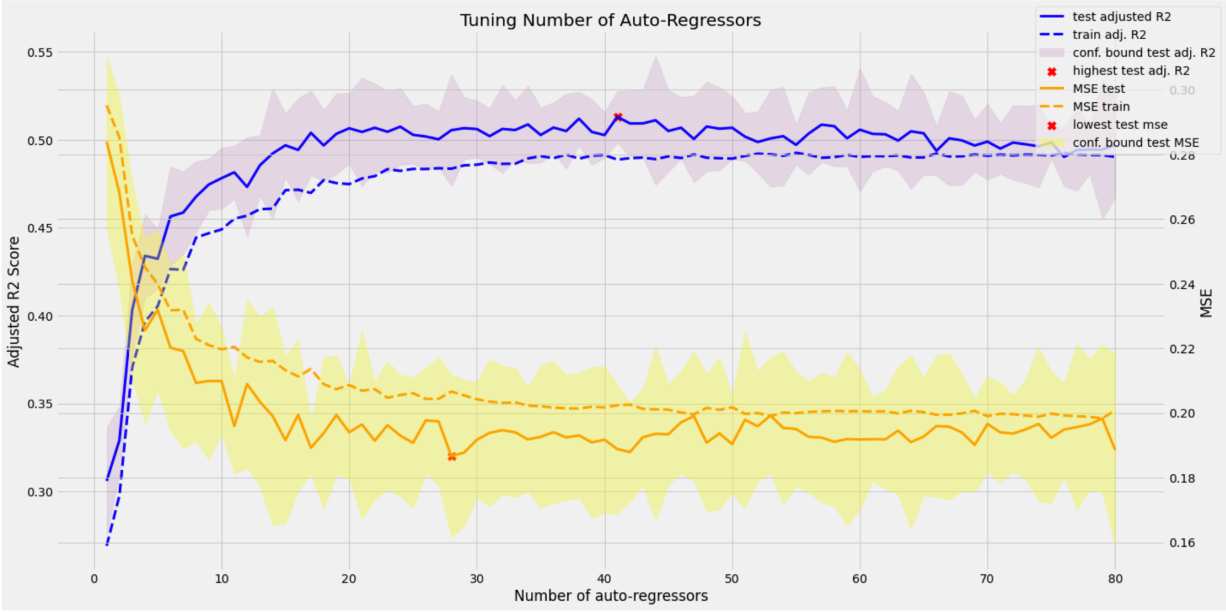


```
Top 20 lags in order of PACF:
[  1   2  48  47  46  96  45 336  49  95 144  44 192   3 288  19  40 672
  94  41]
```



Tune number of highly correlated auto-regressors

- Other features, hourofday(x,y)/dayofyear(x,y)/temperature, are included in all models
- Lags are sorted based on their PACF
- Number of auto-regressors is increased from 1 up to 80
- Error measures are calculated
- Selected number of features = first point with test MSE <= 1.1 *min test MSE and test adj. R2 >= 0.95* max test adj. R2



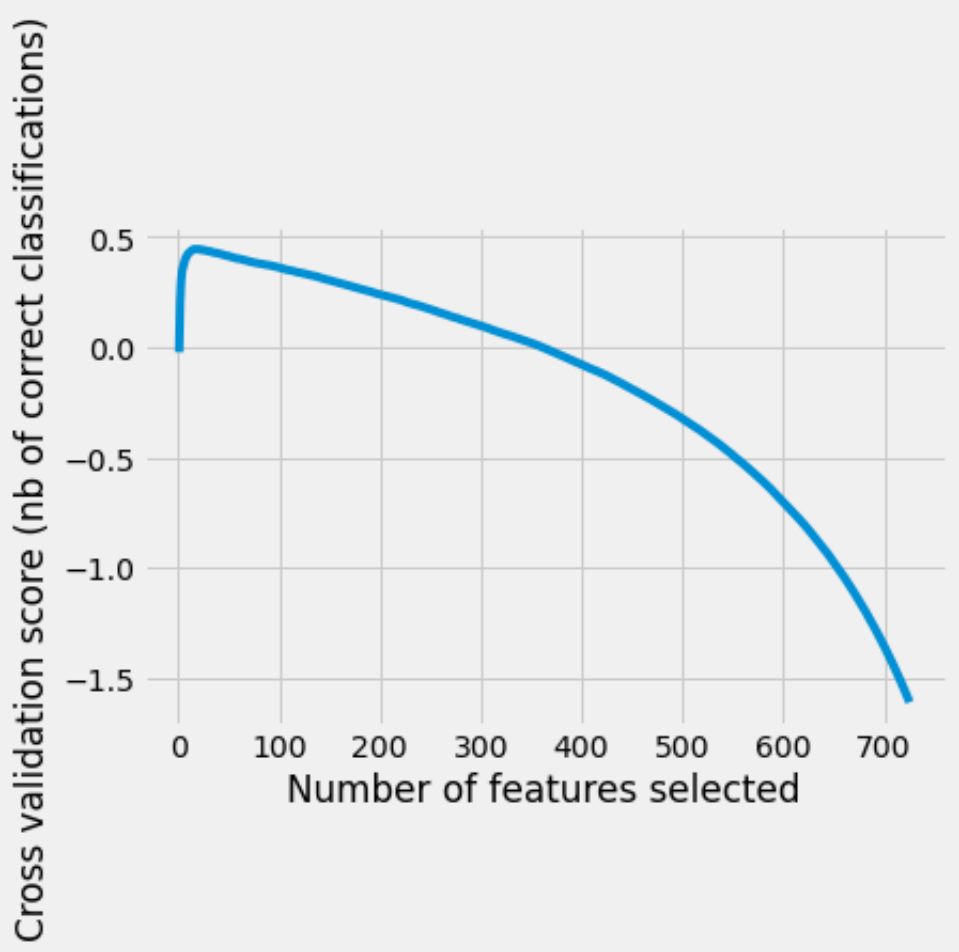

```
Results:
Criterion                    Num.auto-regressors    Adj. R2 test    AIC test    MSE test
-------------------------------------------------------------------------------------------
max adj. R2:                 41                     0.51            81.60       0.19
min MSE:                     28                     0.51            55.63       0.19
first in range:              14                     0.49            27.48       0.20
-------------------------------------------------------------------------------------------
```

Calculating error measures for the selected number of features

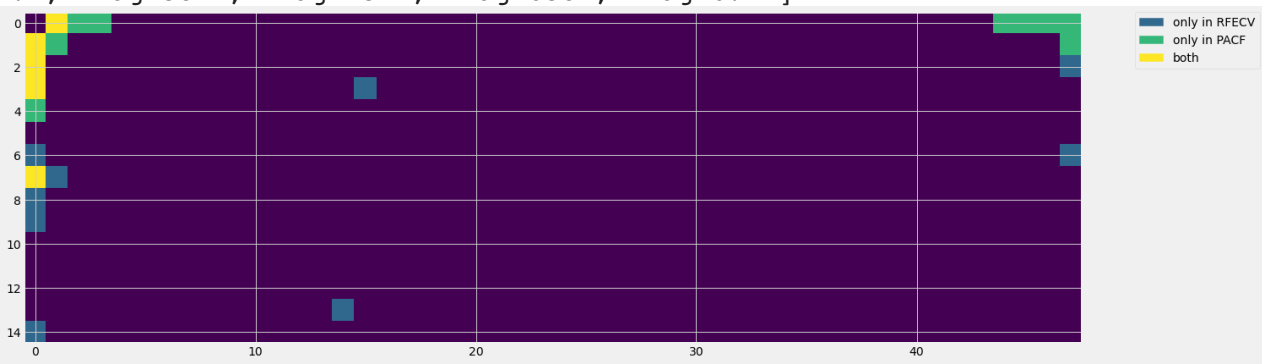
```
Using lags:
[  1   2   3  44  45  46  47  48  49  95  96 144 192 336]
Train R2 scores: 0.46956419514991155, 0.4692392758258581, 0.474517574455024
07, 0.47657609466632755, 0.46408008405380774
Test  R2 scores: 0.46960505299833977, 0.47184579589918574, 0.44681446970151
684, 0.44324314716579116, 0.4915172366842526
Mean absolute error: train 0.32, test 0.33
Mean squared error:  train 0.21, test 0.21
Explained Variance Score (best=1): train 0.46, test 0.49
Coefficient of determination (R2): train 0.46, test 0.49
Adjusted coeff. of determination:  train 0.46, test 0.49
AIC: train 24.57, test 27.34
```

2.2 Method 2: RFE+CV

- Decide on importance of regressors based on t-statistic
- Compare models using adjusted R2
- Search among 14 days ago (maximum 15x48-1 features)




```
Recurssive Feature Elimination + CV
Optimal number of features: 17
Selected features:
['constant', 'hourofd_y', 'temperature_hourly', 'lag 1', 'lag 48', 'lag 96'
, 'lag 143', 'lag 144', 'lag 159', 'lag 288', 'lag 335', 'lag 336', 'lag 33
7', 'lag 384', 'lag 432', 'lag 638', 'lag 672']
```




```
Train R2 scores: 0.49935356956717525, 0.4858271770864523, 0.485282124388776
6, 0.49135878427510526, 0.49812373195195303
Test  R2 scores: 0.45897744165502696, 0.5114642738558266, 0.516074581423753
3, 0.48354480324787064, 0.4614888062150112
Mean absolute error: train 0.32, test 0.30
Mean squared error:  train 0.21, test 0.18
Explained Variance Score (best=1): train 0.49, test 0.52
Coefficient of determination (R2): train 0.49, test 0.52
Adjusted coeff. of determination:  train 0.48, test 0.51
AIC: train 18.62, test 21.73
```

2.3 Comapring Feature Selection Methods

- APCF selects features that are more intuitive: only around the same hour, recent days
- MSE and adj. R2 of both methods are close, RFECV is slightly better
- RFECV selects fewer number of features
- PACF uses the whole time-series (all days) to ranks features
- RFECV can be easily performed on differnt parts of the time-series, e.g. different days, different parts of the day, ...

3 Visualizing Predictions

Goal: Produce plots to evaluate the performance of the linear model

3.1 Residual plots: Plot residual (predicted-actual) vs. predicted

Ideal:

- Points are symmetrically distributed
- Points are clustered near the y-axis
- No clear patterns

Observations:

- (-) Cone-shaped plots, error variance increases with consumption
- (-) Error magnitude is large w.r.t predicted value; e.g., residual of 1 at noon when consumption is around 1 => 100% error

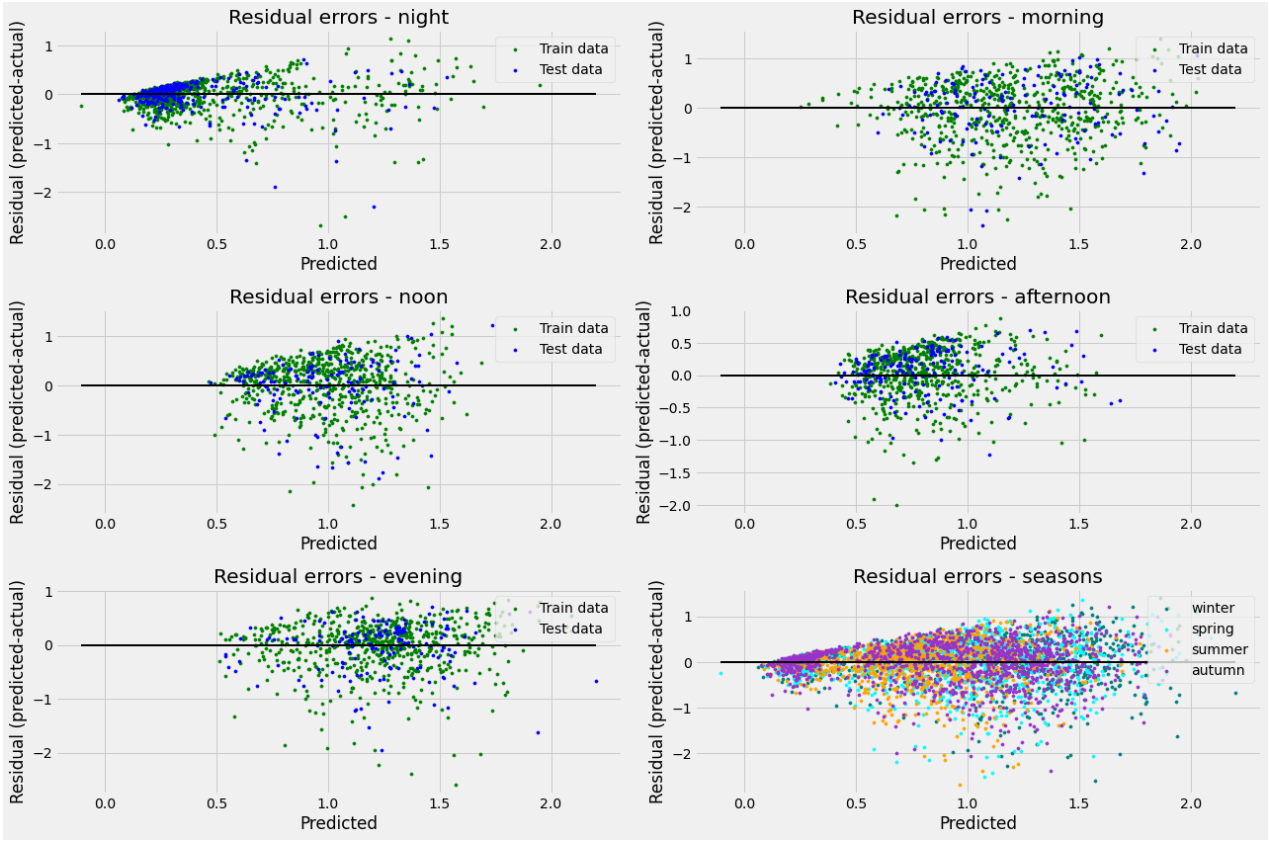
Interpretation:

- If cone-shaped, data is heteroscedastic (i.e. the variance of the errors is not constant)

Consequences of heteroscedasticity:

- OLS will not give you the estimator with the smallest variance (i.e. your estimators will not be useful).
- Significance tests will run either too high or too low.
- Standard errors will be biased, along with their corresponding test statistics and confidence intervals.

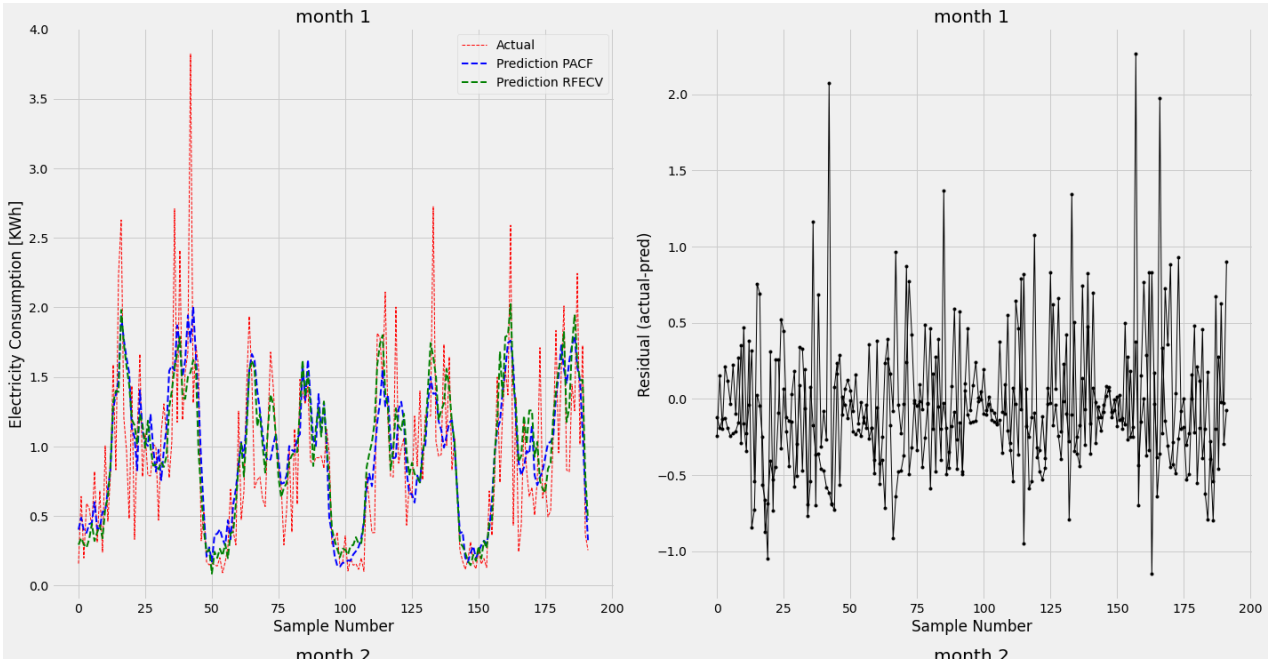
ref: https://www.qualtrics.com/support/stats-iq/analyses/regression-guides/interpreting-residual-plots-improve-regression/

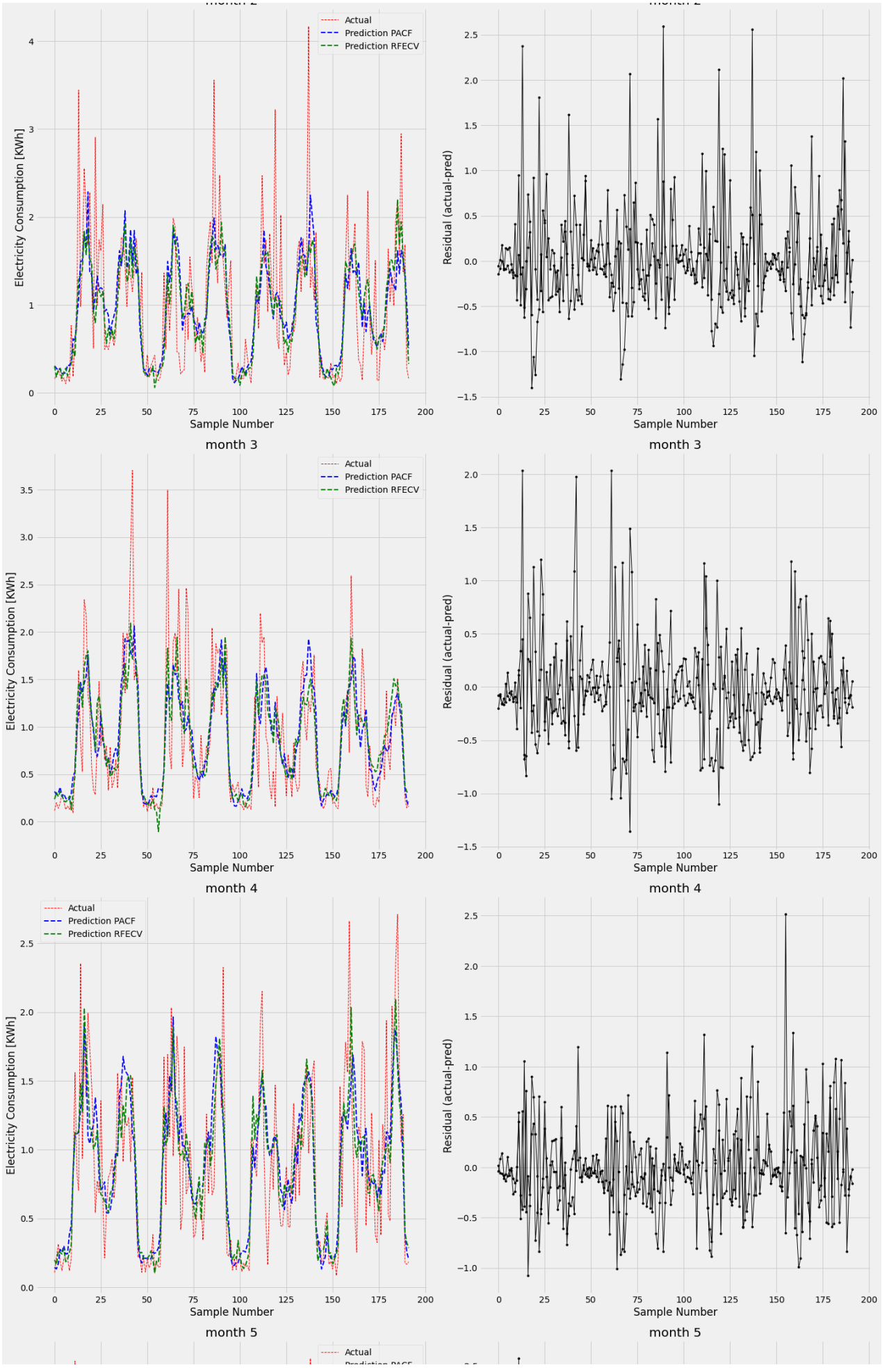


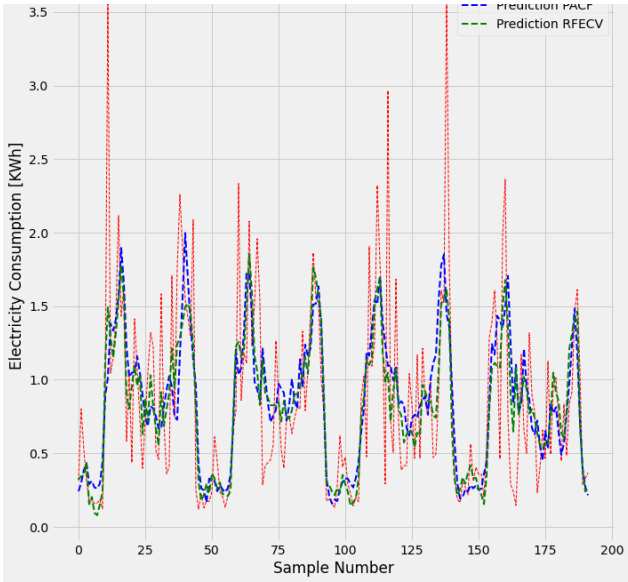
3.2 Predictions in different months

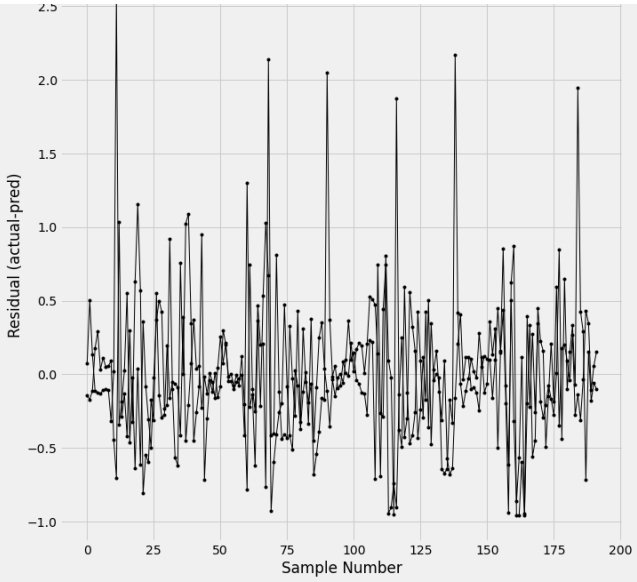
Observations:

- (+) The general trend is captured
- (+) Residuals show no clear trend with time
- (-) Consumption peaks were poorly modeled that results in considerable peaks in residuals
- No clear distinction between model performance in different months



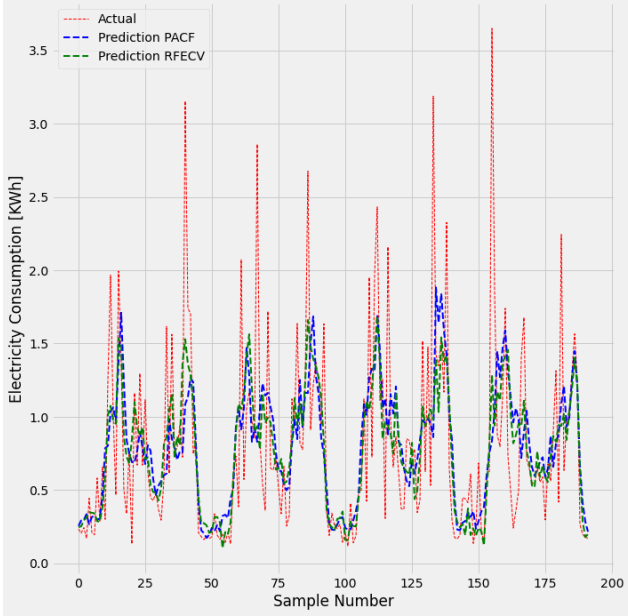

Actual
Prediction PACF
Prediction RFECV
Electricity Consumption [KWh]
Residual (actual-pred)
Sample Number
month 3
month 4
month 5

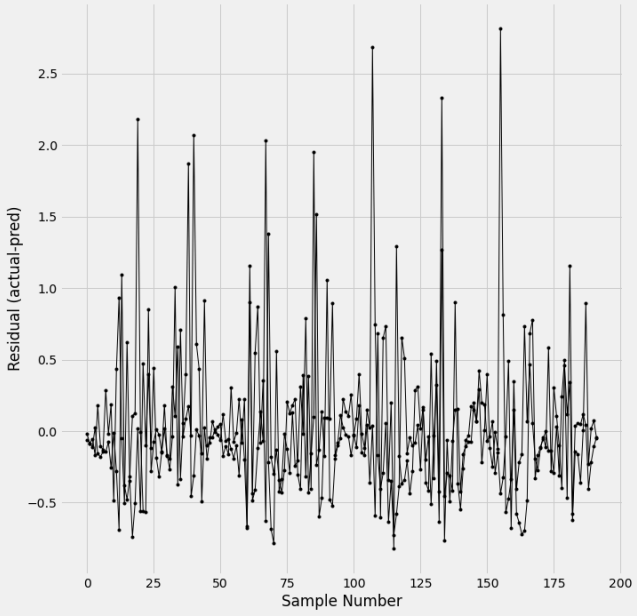




month 6

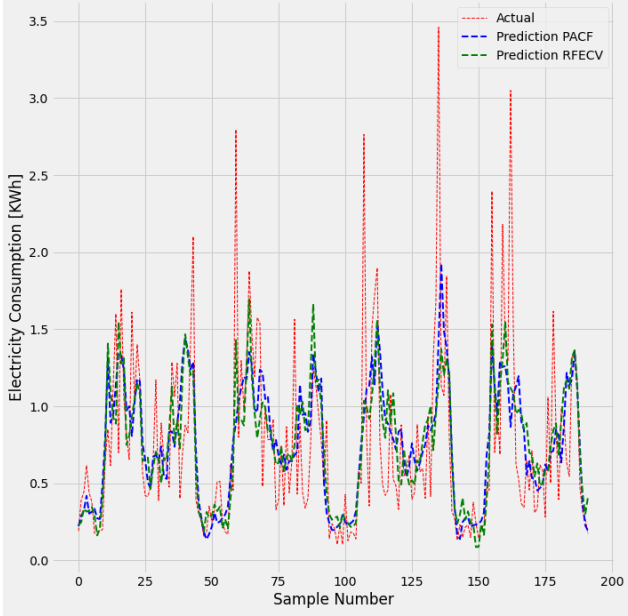
month 6

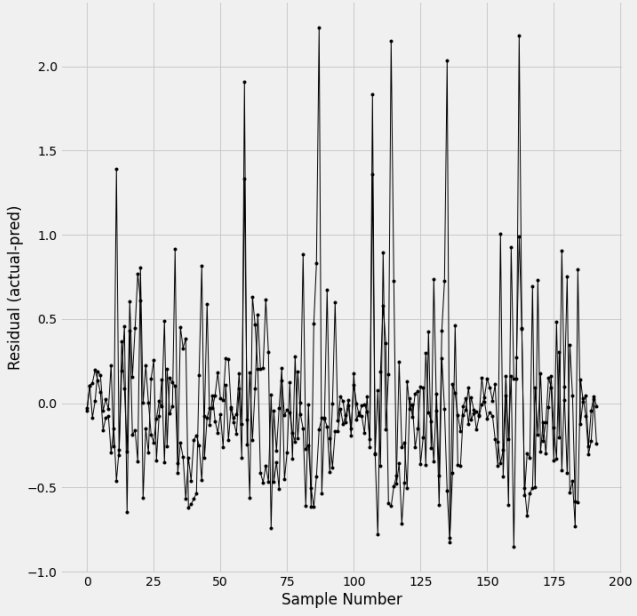




month 7

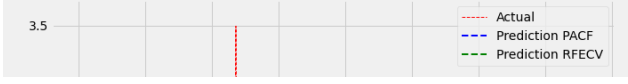
month 7

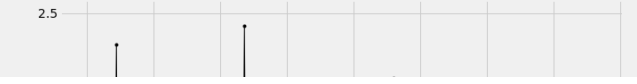


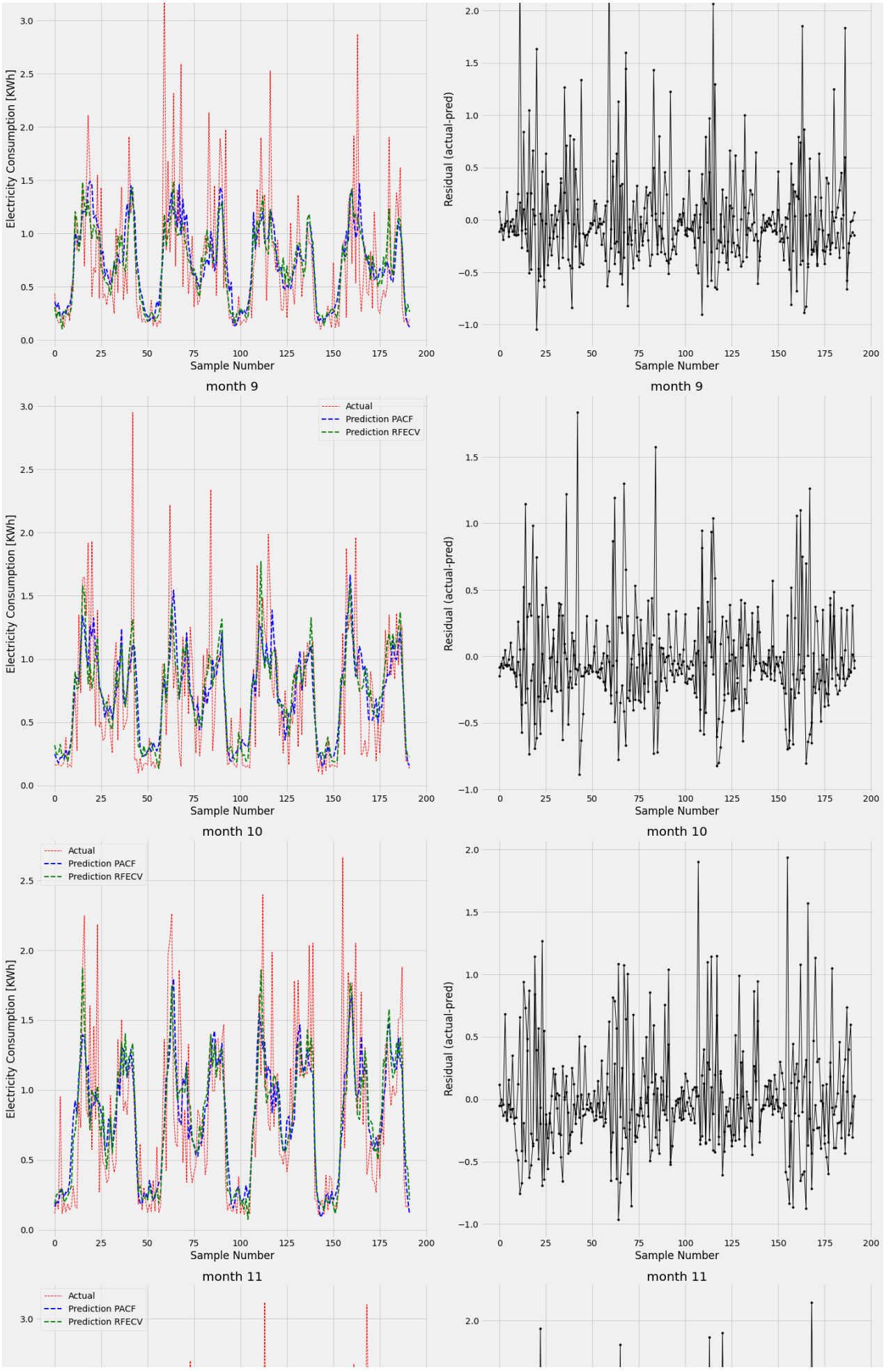


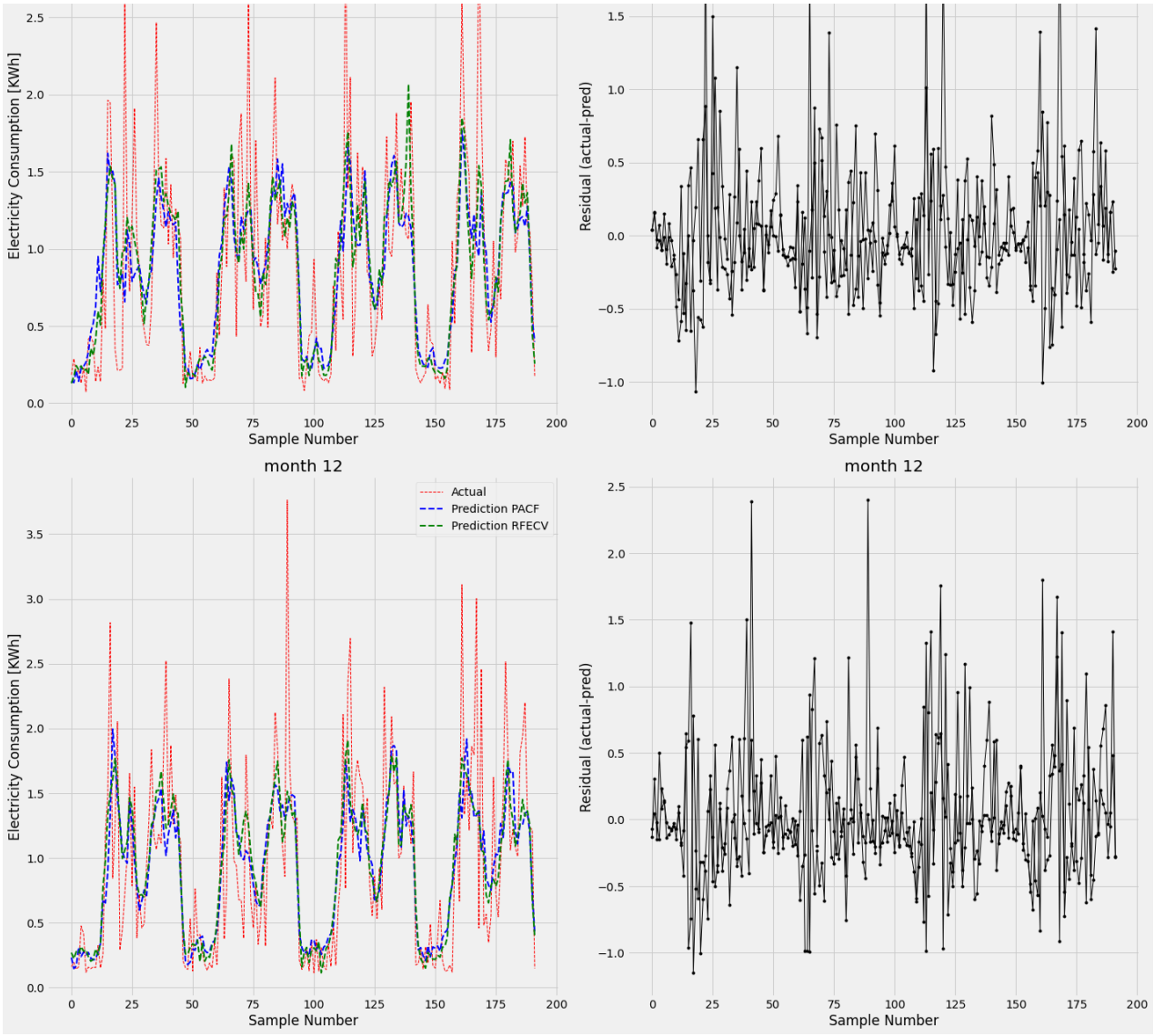
month 8

month 8





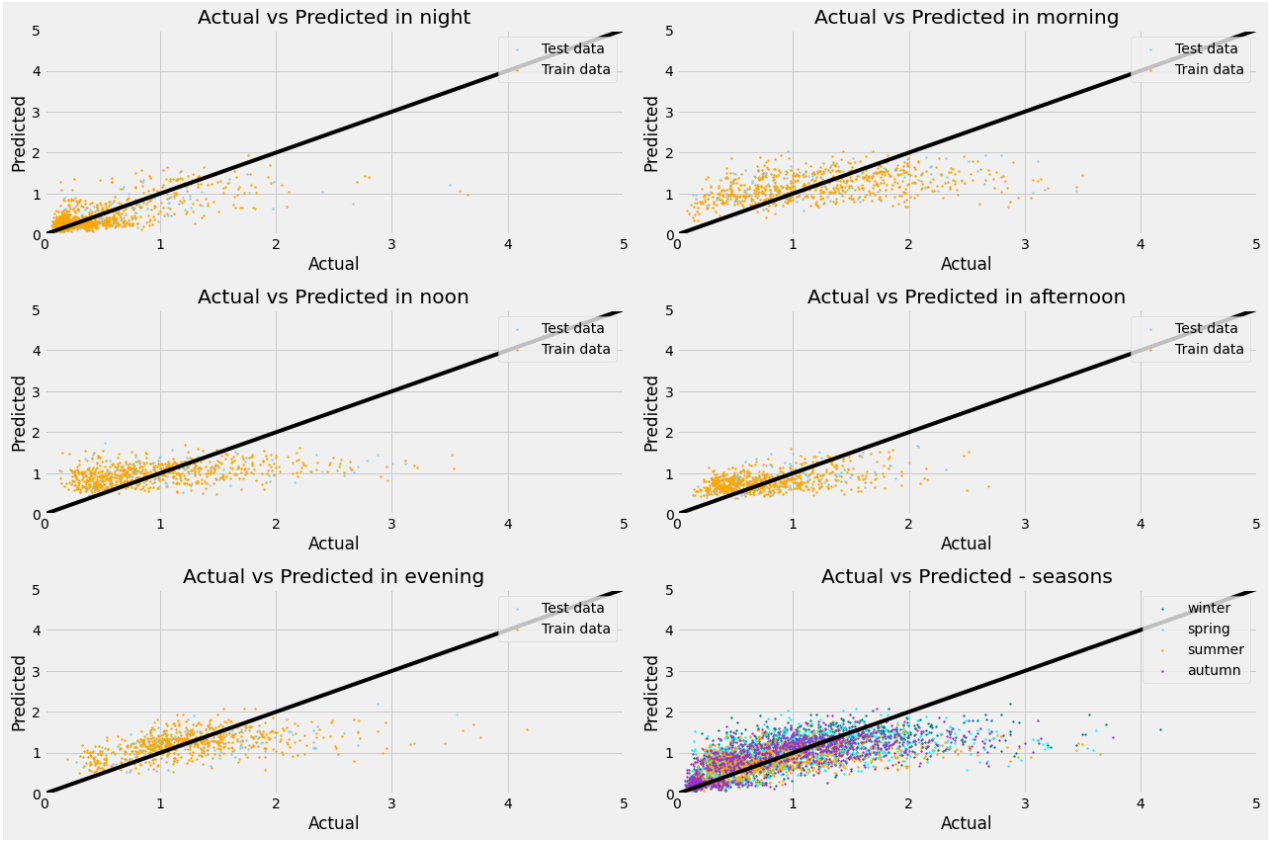

Electricity Consumption [KWh]
Sample Number
Residual (actual-pred)
month 9
Actual
Prediction PACF
Prediction RFECV
month 10
month 11



3.3 Actual vs. Prediction

Observations:

- Performance is different over morning/noon/...
- No clear distinction between seasons
- (-) Over-predicting at high consumptions
- (-) Under-predicting at low consumptions



4 Statistical Analysis of Ordinary Least Squares with statsmodels.api

4.0 Theory:

- Ordinary Least Squares model
- trained k=5 times, each round using (k-1)/k samples as the training dataset
- adjusted R2 criterion is calculated on the other 1/k of the data (validation dataset)
- the model with the highest adj. R2 on validation dataset is returned

Evaluation measures:

- R-squared: proportion of variation in the outcome explained by regressors (the same as the squared correlation between actual and predicted values for OLS)
- Adj. R-squared: adjusted with the number of features, adjusted R-squared is equal to 1 - (n-1)/(n–k-1)*(1-R2)
- F-statistic: whether a complex model is better than a simpler version of the same model in explaining the variance in the dependent variable
- Prob (F-statistic): t-test for how large F is
- Log-Likelihood:
- AIC: in general, AIC=2*num_of_regressors*-2log(likelihood). For linear regression, use log(sum of squared errors) instead of log(likelihood). Smaller AIC is preferred.

- BIC:

Interpreting the coefficients:

- std err: estimate of the standard deviation of the coefficient, the amount it varies across cases
- t: coefficient divided by its standard error. If a coefficient is large compared to its standard error, then it is probably different from 0.
- P>|t|

More statistics:

- Omnibus & Skew: a test of the skewness of the residual, ideally 0 => errors are Gaussian => linear regression approach would probably be better than random guessing but likely not as good as a nonlinear
- Prob Omnibus: statistical test indicating the probability that the residuals are normally distributed,ideally 1
- Kurtosis: a measure of "peakiness", greater Kurtosis can be interpreted as a tighter clustering of residuals around zero, implying a better model with few outliers
- Durbin-Watson: tests for homoscedasticity, ideal value between 1 and 2
- Jarque-Bera (JB)/Prob(JB): like the Omnibus test in that it tests both skew and kurtosis
- Condition Number: sensitivity of a function's output as compared to its input. When we have multicollinearity, we can expect much higher fluctuations to small changes in the data, hence, we hope to see a relatively small number, something below 30.

Ref: https://www.accelebrate.com/blog/interpreting-results-from-linear-regression-is-the-data-appropriate

Package: https://www.statsmodels.org/stable/generated/statsmodels.regression.linear_model.Regressi

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.
464
Model:                            OLS   Adj. R-squared:                  0.
461
Method:                 Least Squares   F-statistic:                     17
5.9
Date:                Thu, 22 Apr 2021   Prob (F-statistic):               0
.00
Time:                        07:47:03   Log-Likelihood:                -250
1.0
```

```
No. Observations:                3879   AIC:                             50
42.
Df Residuals:                    3859   BIC:                             51
67.
Df Model:                          19
Covariance Type:            nonrobust
==============================================================================
===========
                         coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
-----------
constant                0.1796      0.040      4.487      0.000       0.101
0.258
hourofd_x               0.0736      0.013      5.741      0.000       0.048
0.099
hourofd_y              -0.0751      0.013     -5.893      0.000      -0.100
-0.050
dayofy_x               -0.0168      0.012     -1.394      0.163      -0.040
0.007
dayofy_y               -0.0376      0.017     -2.157      0.031      -0.072
-0.003
temperature_hourly     -0.2279      0.078     -2.916      0.004      -0.381
-0.075
lag 1                   0.3675      0.065      5.693      0.000       0.241
0.494
lag 2                   0.2625      0.061      4.283      0.000       0.142
0.383
lag 3                   0.2420      0.059      4.120      0.000       0.127
0.357
lag 44                  0.0387      0.061      0.637      0.524      -0.080
0.158
lag 45                  0.0368      0.064      0.576      0.564      -0.088
0.162
lag 46                  0.2148      0.065      3.311      0.001       0.088
0.342
lag 47                  0.4003      0.066      6.036      0.000       0.270
0.530
lag 48                  0.5300      0.070      7.586      0.000       0.393
0.667
lag 49                  0.1853      0.067      2.763      0.006       0.054
0.317
lag 95                  0.2351      0.080      2.926      0.003       0.078
0.393
lag 96                  0.4878      0.080      6.073      0.000       0.330
0.645
lag 144                 0.5278      0.079      6.696      0.000       0.373
0.682
lag 192                 0.3687      0.075      4.947      0.000       0.223
0.515
lag 336                 0.5204      0.064      8.075      0.000       0.394
0.647
==============================================================================
===
Omnibus:                     1106.315   Durbin-Watson:                   2.
093
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3945.
013
Skew:                           1.397   Prob(JB):                        0
.00
Kurtosis:                       7.075   Cond. No.                        1
```

```
5.1
================================================================================
===

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is corr
ectly specified.
```

Significance of regressors:

Above results show that all features weren't significant. We remove these features and train another model on the reduced features set.

Result: AIC improves but other measures do not change significantly.

```
Significant regressors:
['constant', 'hourofd_x', 'hourofd_y', 'dayofy_y', 'temperature_hourly', 'l
ag 1', 'lag 2', 'lag 3', 'lag 46', 'lag 47', 'lag 48', 'lag 49', 'lag 95',
'lag 96', 'lag 144', 'lag 192', 'lag 336']
Removed regressors:
['dayofy_x' 'lag 44' 'lag 45']
Mean absolute error: train 0.33, test 0.32
Mean squared error:  train 0.21, test 0.21
Explained Variance Score (best=1): train 0.46, test 0.50
Coefficient of determination (R2): train 0.46, test 0.50
Adjusted coeff. of determination:  train 0.46, test 0.49
AIC: train 18.55, test 21.41
```

5 Summary

Overview:

- Goal: 1-step ahead consumption prediction in 30 min intervals
- Scope: single household
- Regressors: temperature, time of the day, day of the year, auto-regressors with different lags
- Methodology: linear regression

What has been done so far:

- Data cleaning
- Feature selection with RFECV and PACF+CV methods
- Significance analysis
- Visulaization

Results:

- The general trend can be modeled
- Sudden high peaks are underestimated
- Both feature selection methods obtain similar error measures
- The number of selected features is less than 20
- Important features are among 48k +/- 2, k between 0 and 7

To do next: Improving the fit

- Are outliers the only source of error?
- 1 hour resolution
- Modeling the residuals

Robustenss analysis:

- Effect of weekday
- Effect of time of the day
- Comparing error between different households

Federation:

- Is this problem a good instance for federated learning?

```
  File "<ipython-input-72-ca20abd091f5>", line 1
    jupyter nbconvert --to html --TemplateExporter.exclude_input=True Lin_R
eg_Part1.ipynb
            ^
SyntaxError: invalid syntax
```