

Determinants and dynamics of renewable energy consumption in the US

30 gennaio, 2022

Abstract

In this paper we forecast the dynamics of renewable energy consumption in the U.S. market. We considered monthly data on energy and fossil fuels from the US Energy Information Administration (EIA), macroeconomic variables from the FRED database (FED St. Louis) and monthly prices of crude oil and natural gas provided by the World Bank. We performed three variable selection methods in order to reduce the dimensionality of the data: hard thresholding, partial least squares and LASSO. Then, we focused on predictions using a recurrent neural network and a factor model. To estimate the factors we relied on two different approaches: PCA and EM algorithm. We all took part in the design of each step of this paper, however

The RNN forecasts on the test sample are accurate, as well as the results with the one-factor models.

1. Introduction

Due to increasing risk of irreparable environmental damages by pollution and overusage of resources, countries around the world try to establish and incentivize the usage of renewable ways of generating energy. Doing so, countries focus not only on shutting down resource intensive ways of energy production but also issue consumption targets for energy produced by renewable sources. Germany for example is about to shut down its last nuclear power plant at the end of this year (News 2022) and plans to be independent of energy produced by coal by 2038 (Government 2020). However, the German government also aims to increase the consumption share of renewable energy to 65% by 2030 (Government 2020). The plans of US president Joe Biden are even more ambitious, he wants all electricity to be from renewable sources by 2035 (Guardian 2021).

While the production share of renewable energy is heavily influenced by technological constraints and legal considerations, the demand side is much more of an economic problem and requires further investigation as an increase in consumption is supposed to put pressure on the producers and thereby accelerates the phasing out of old technologies. Hence, this study focuses on understanding the determinants as well as the dynamics of the share of renewable energy consumption in total energy consumption. Therefore, the analysis in this paper is twofold.

Firstly, it analyses the determinants of renewable energy consumption. Thus, variables are selected from a big dataset, which consists of macroeconomic indicators as well as consumption and production of alternative energy forms, using three different methods of variable selection. In line with Bai and Ng (2008), the paper compares the selection based on hard thresholding and LASSO. Additionally, it analyses which variables are selected by the Variables Importance in Projection (VIP) filtering method for partial least squares regressions (PLSR). Secondly, the above mentioned consumption share of renewable energy in the US is predicted and its development is evaluated. Two forecasting methods are compared, dynamic factor models and neural network, and their ability to predict the response variable is discussed.

Previous economic literature on the determinants of renewable energy consumption is dominated by investigating the relationship between energy consumption and economic growth. While output and employment were found to be key determinants of solar and wind energy (Sari R. 2008), a bi-directional causality between both economic growth and renewable energy consumption and between non-renewable and renewable energy consumption was investigated (Apergis N. 2012). Even though there is literature which investigates the influence of other macroeconomic and institutional measures on the consumption of renewable energy, variables are usually already pre-selected and, hence, heavily dependent on prior beliefs (see Ergun S. J. (2019) for literature review). Furthermore, by pre-selection of variables there was little need of employing machine learning methods in previous literature and only a few studies use comparable methodology (Sarkodie S. A. 2018). Thus, this study distinguishes from earlier literature by only imposing limited prior beliefs on what drives consumption of renewable energy and the methodology that arises from this approach.

While machine learning methodology with respect to forecasting is quite prevalent within energy economics in general (Ghoddusi H. 2019), it is more scarcely used with respect to renewable energy consumption in particular. Since most studies suffer from a comparably short time horizon, they either use completely different methodology (Tsai S. 2017) or only rely on neural networks (Azadeh A. 2013). Hence, the paper differs from previous literature in scale as of its long time horizon as well as in methodology.

The rest of the paper is organized as follows: Section 2 describes the data. Then, chapter 3 introduces the methodology for the variable selection and discusses the underlying results, while section 4 does the same for the forecasting part. Chapter 5 concludes.

2. Data description

The final dataset used for the underlying analysis consists of 145 variables describing the economic and the energy situation in the US ($N = 147$) from 1:1973 until 8:2021 ($T = 584$) on a monthly basis. As many of those series are non-stationary, they are transformed mostly using log-differences. In the following the composition and origin of the data will be discussed more detailed.

Firstly, monthly data on energy and fossil fuels is provided by the US Energy Information Administration (EIA) (Administration 2022). The raw file consists of 41 variables including production, consumption and exports time series of energy by source. Nine of those series are discarded as of too many missing values or collinearity. Furthermore, the dependent variable is constructed as the share of the two variables TotalRenewableEnergyConsumption over TotalPrimaryEnergyConsumption and, therefore, those two variables are also dropped from the analysis. Thus, one is left with 31 variables from this source.

Additionally, the Federal Reserve Bank of St. Louis maintains a large dataset, the FRED database, on macroeconomic performance in the US. 127 variables on topics like output and income, the labor market, money and credit, prices or interest and exchange rate are updated on a monthly basis (see McCracken M. W. (2016) for a detailed description of all included variables in the database). Time series are transformed according to the suggested transformation. As for some time series second log-differences are recommended, the first two observation points of all variables cannot be used in the following analysis ($T = 582$). While the transformation yields stationary time series for most of the series, the housing variables have to be excluded as of non-stationarity. Additionally, four series have to be dropped as of too many missings. Thus, 113 variables from this source are kept.

Moreover, monthly prices of crude oil and natural gas are included to allow for cross-elasticity of renewable energy consumption on price changes of its substitutes. The time series are provided by the Worldbank (Worldbank 2022). Data on the price of coal could not be used as of too many missings in the time series. All remaining time series are stationary and there are 8 missing values left in the data which have to be replaced as LASSO does not allow for missing observations. For simplicity and as of the comparably small number of missings, those are imputed by the reported value of the previous period. Furthermore, GeothermalEnergyConsumption and NuclearElectricPowerConsumption are found to be collinear and are, therefore, discarded ($N = 145$). An investigation for outliers yields 1912 extreme results out of 84390 elements. Outliers in this case are defined according to the definition of “far out” outliers in earlier literature on boxplots (W. 1977) and are, therefore, either three interquantile ranges ($IQR = Q3 - Q1$) above the third quantile ($Q3$) or three interquantile ranges below the first quantile ($Q1$). These outliers might drive some of the results in the forecasting section. Finally, the data is standardized to fluctuate around zero which is necessary for

the dynamic factor model and the PLSR. In the graphs below one can see the dynamics of all transformed, standardized and missings-corrected time series splitted up in three parts for better observability. As shown, the outliers are spread over the full period with the highest spikes in recent years.

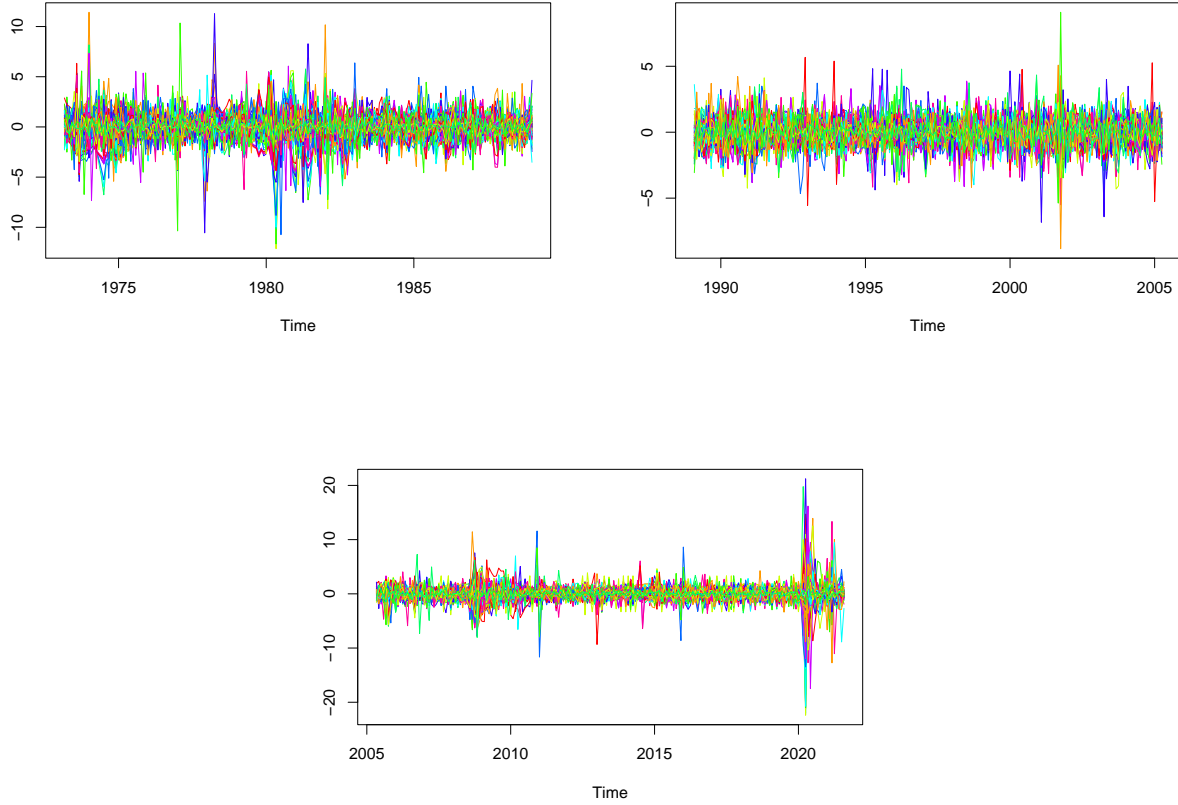


Figure 1: Standardized Time Series Plot

3. Variable selection

3.1 Methodology

Analyzing the determinants of renewable energy consumption with the underlying dataset requires variable selection. In line with earlier literature (Bai J. 2008), this study compares the results of three methodological approaches which differ in the reasoning why variables are chosen.

3.1.1 Hard Thresholding

Hard thresholding tests the relevance of a single predictor for the dependent variable y_t (Bai J. 2008). Hence, considering the autoregressive nature of the time series data, it performs one regression of the share of renewable energy consumption on each variable in the dataset $x_{i,t}$ plus its lags $x_{i,t-h}$ and the lags of the dependent variable. While there are always four lags of the dependent variable included in those regressions, the number of lags of the explanatory variables h vary between three different specifications ($h = 1, 3$ or 6). For the variable selection, the resulting p-values of $x_{i,t}$ are compared with a Bonferroni corrected 5% significance level (critical value = 0.00036). Thus, only variables with a p-value smaller than the critical value are kept as their predictive power on y_t is significant at 5% level.

##	Varname	Hard_Thresholding_lag1
## 1	dl_NaturalGasPlantLiquidsProduc	0.0001960176
## 2	dl_NuclearElectricPowerProductio	1.921052e-11
## 3	dl_TotalRenewableEnergyProductio	1.975522e-26
## 4	dl_NaturalGasImports	4.95853e-15
## 5	dl_CoalConsumption	2.216242e-24
## 6	dl_NaturalGasConsumptionExcludi	5.649158e-47
## 7	dl_PetroleumConsumptionExcluding	7.691579e-09
## 8	dl_TotalFossilFuelsConsumption	6.880973e-45
## 9	dl_indpro	9.369306e-05
## 10	dl_ipfpnss	5.721753e-05
## 11	dl_ipfinal	2.170328e-05
## 12	dl_ipcongd	1.476711e-06
## 13	dl_ipncongd	1.799365e-08
## 14	dl_ipb51222s	2.027636e-09
## 15	dl_natural_gas_us	5.108224e-05
## 16	dl_natural_gas_average	0.0003384302
##	Hard_Thresholding_lag3	Hard_Thresholding_lag6
## 1	NULL	NULL
## 2	7.296069e-12	1.712536e-11
## 3	3.091057e-30	1.176648e-63
## 4	5.206395e-14	1.178227e-09
## 5	4.730028e-25	8.260054e-25
## 6	8.087505e-32	1.404135e-21
## 7	3.238585e-11	1.767146e-15
## 8	5.096163e-39	6.481691e-30
## 9	7.250901e-05	0.0001003536
## 10	3.530604e-05	4.178134e-05
## 11	1.165203e-05	1.185811e-05
## 12	4.454579e-07	4.014271e-07
## 13	8.251865e-09	8.913939e-09
## 14	3.819317e-09	1.221317e-09
## 15	3.984908e-05	4.233306e-05
## 16	NULL	0.0003138963

As selected variables do barely vary between the different specifications, the relevance of the variables for the share of renewable energy consumption seems to be independent of the timing. Hence, there are no variables that only affect consumption of renewables in the middle term ($h = 6$) but not in the short term ($h = 1$). Focusing on immediate effects, the study, thus, proceeds with the one-lagged results.

Consumption and production of alternative energy sources including production of renewable energy have a very high predictive power on y_t . Additionally, various time series from the output and income group of the FRED dataset have a significant impact on the share of renewable energy consumption. Therefore, the result are in line with earlier literature exploring the relationship between output and consumption of renewables and consumption of non-renewables and y_t (Apergis N. 2012). Lastly, some of the variability in the dependent variable is explained by the price of natural gas. This is not very surprising as an increase in prices of natural gas is expected to decrease consumption of natural gas and consumers, therefore, would have to substitute. However, it is puzzling that crude oil price as another included commodity price in the analysis is not part of the selected variables.

3.1.2 Partial Least Squares (PLS)

Partial Least Squares (PLS) is a *supervised learning* method for dimension reduction, thus dealing with the so-called large p small n problem. To form a relationship between the response variable Y , which can be both univariate or multivariate, and the set of explanatory variables X_1, \dots, X_p , PLS constructs new explanatory variables called *components* which are linear combinations of the original predictors. Differently from Principal Component Regression (PCR), components in PLS are sequentially constructed also considering the response variable of interest. In other words, PLS forms components that capture most of the information in the input variables that is useful for predicting Y , while reducing the dimensionality by using fewer components than the number of input variables (Garthwaite 1994).

The original aim of PLS is to find the relevant linear subspace of the explanatory variables, nevertheless many methods for variable selection using this method have been proposed in order to improve model interpretation. Variable selection methods based on PLS regression can be divided into three main categories: filter-, wrapper-, and embedded methods. This paper focuses on the selection criteria based on *Variable Importance in Projection* (VIP) which belongs to the filter methods class. The VIP score of each predictor j is computed using the formula below. First, the optimal number of factors has to be determined, which here is done through cross-validation considering the root mean squared error of prediction (RMSEP), as shown in Figure 2, and indicates use of 14 factors.

$$v_j = \sqrt{p \cdot \frac{\sum_{a=1}^A SS_a (w_{aj} / \|w_a\|^2)}{\sum_{a=1}^A (SS_a)}}$$

where SS_a is the sum of squares explained by the a th component and $(w_{aj} / \|w_a\|^2)$ is the normalized weight for j variable and a component, which represents the importance of the j th variable (Mehmood T. 2012).

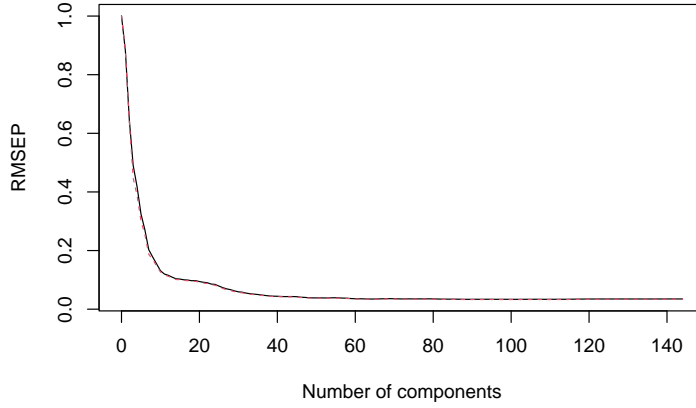


Figure 2: Validation Plot

Selection of more influential variables occurs by comparing each VIP score v_j with a positive threshold value defined by the researcher. Since the average of the squared VIP scores equals 1, generally a ‘greater than one rule’ is used and variables with $v_j > 1$ are selected. This is not a statistically justified threshold point and it has been found that a cutoff point between 0.83 and 1.21 might be more appropriate based on certain characteristics of the data such as proportion of relevant predictors and magnitude of correlation (Chong I. 2005).

The cutoff point = 1 is considered. Compared to the hard thresholding method described previously, the number of selected variables based on VIP score is three times larger but all variables selected with the first method are also selected here. Moreover, with very few exceptions, matching variables are the ones with the highest VIP score and from the plot below these are the variables between index 10 and 40 which correspond to the energy related predictors.

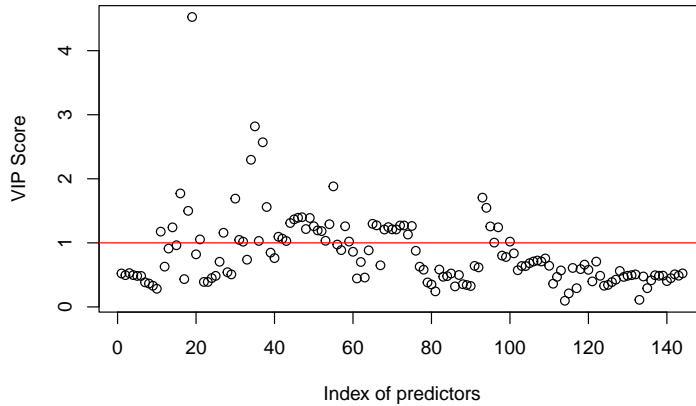


Figure 3: VIP Scores

3.1.3 LASSO

The LASSO, firstly introduced by Tibshirani (1996), is a *supervised learning* technique to implement variable selection. It shrinks some coefficients, associated to some of the the p predictors, to zero. Aside from variable selection, another advantage of this process is to improve the accuracy of the predictions - with respect to the ones of the linear model - through a reduction of the variance which outweighs the increase in the bias.

Theoretically, the LASSO coefficient is defined as:

$$\left(\hat{\beta}_{0,\lambda}^L \dots \hat{\beta}_{p,\lambda}^L\right)' = \arg \min_{b_0 \dots b_p} \sum_{i=1}^n \left(y_i - b_0 - \sum_{j=1}^p x_{ij} b_j\right)^2 + \lambda \sum_{j=1}^p |b_j|.$$

The LASSO coefficients minimizes a quantity equal to the residual sum of squares plus a penalty term, called l_1 penalty. If the coefficient λ (*tuning parameter*) is large enough, some coefficients are shrunk exactly to zero, allowing for variable selection. Therefore, the higher is λ , the more the penalty matters, the more coefficients are shrunk to zero.

As in every other machine learning method, the estimation of the LASSO is performed considering two key-passages: splitting the data into training and test set, and then consider the training set to estimate the model and tune the parameters (in this case λ , using cross validation). The first step is trivial, as we consider the first 291 observations as training set (roughly the 50%) and the remaining as test set. We fit the LASSO model using the glmnet method in R, specifying alpha=1.

instead, cross validation with time series data can be tricky. We cannot perform the standard k-fold procedure, which is based on picking randomly samples out of the training set and split it again in different folds, one of which will be the test for the validation procedure. It is evident that the random choice of the folds cannot be applied in this context, as there is a temporal dependency across observations that must be preserved. Therefore, we opted for a procedure suitable for time series, based on a *expanding window*: first we train the LASSO on 100 observations and evaluate its prediction on the next 50 of the sequence. Subsequently, we keep iterating this procedure by incorporating the previous test set into a new training set and testing the accuracy on the following observations. The λ chosen with cross validation is:

```
model1$bestTune
```

```
##      alpha      lambda
## 130      1 0.003847427
```

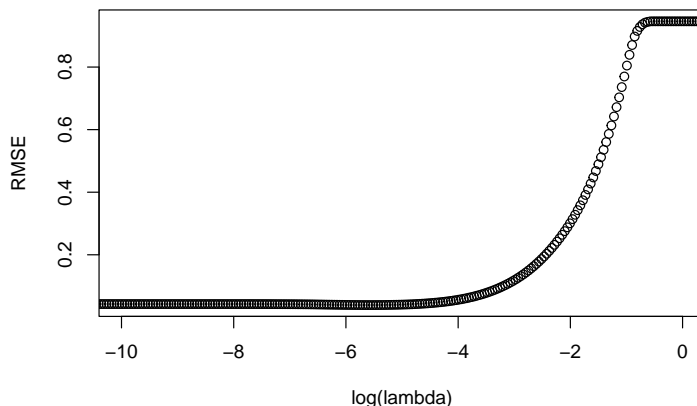


Figure 4: RMSE of different $\log(\lambda)$, $\text{minimumat} = -5.56$

We consider this value of the hyperparameter to perform variable selection with LASSO. The selected variable consists of small subset (8 variables) of our initial starting features. To have a glimpse of which variables contributes the most, we consider this variable importance plot, where we consider only the variables associated to non-zero coefficients.

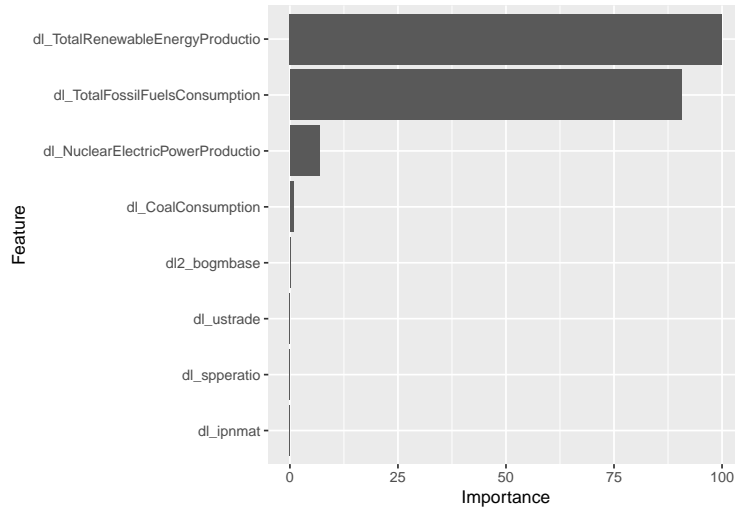


Figure 5: LASSO Variable importance

As we have argued before, the LASSO model could be used to perform variable selection by dropping the coefficients that have been shrunk to zero. However, we should be careful to proceed in this way, because, as in any other model, LASSO results comes with an estimation error. If we perform variable selection only based on LASSO, we are neglecting this error.

3.2 Compare results

Now we compare the results of the different variable selection methods. However, for space reasons the table has been omitted. The production of renewable and nuclear energy, as well as the consumption of coal and fossil fuel are selected by all three methods, suggesting that they are indeed important feature in predicting the share of renewable energy consumption in the US market. Between the models, the LASSO selects fewer features, and some are not selected by the other two methods. Moreover, as we argued before, we can't consider LASSO to perform variable selection simply neglecting its estimation error.

However, considering the other two procedure, there is some consistency: even if PLS selects more variables, all the ones selected with hard thresholding selected are also selected with PLS. To perform predictive analysis using an artificial neural network in the next section, we adopt the subset of features selected with the hard thresholding.

CHAPTER 4: Forecast Analysis

4.1 Neural Networks

The first model we consider to predict the share of renewable energy consumption in the U.S. is a *Recurrent Neural Network* (RNN). In general, the family of artificial neural networks is composed by a wide range of models that can be applied in various contexts, from image recognition to documents classification. All neural networks are composed by several layers: an input layer, composed by p contained variables in the dataset, and multiple hidden layers, each of which contains several hidden units (or neurons), that takes as inputs the initial features or the activations of previous layers. These layers build a non-linear function $f(X)$ to predict the response Y .

Recurrent Neural Networks are designed to take as input a sequence, therefore they are useful to build predictive models with time series data. Our dataset consist of one time series for each variable, and we want to predict the share of renewables $Y_l = y_t$ using a mini-sequence of vectors of lagged predictors (containing also the target variable) $X_l = \{X_{t-1}, X_{t-2}, \dots, X_{t-L}\}$, where L is the number of lags considered. Since $T = 582$ and we opted for $L = 5$, we created $l = 1, \dots, 577$ (X_l, Y_l) pairs.

The structure of the RNN is different from the standard feed-forward neural network. The model considers the 577 mini-sequence of inputs sequentially: each activation vector A_l (composed by K units) inside the hidden layer takes as input the corresponding mini sequence X_l and the previous activation vector A_{l-1} . In particular, each unit is defined as:

$$A_{lk} = g \left(w_{k0} + \sum_{j=1}^p w_{kj} X_{lj} + \sum_{s=1}^K u_{ks} A_{l-1,s} \right)$$

where w_{kj} for $j = 0, \dots, p$ and u_{ks} for $s = 1, \dots, K$ are the parameters to be estimated and $g(z)$ is a non-linear activation function. In our model, we opted for the hyperbolic tangent:

$$g(z) = \frac{(e^z - e^{-z})}{(e^z + e^{-z})}.$$

Each activation vector feeds the output layer and produces a prediction O_l :

$$O_l = \beta_0 + \sum_{k=1}^K \beta_k A_{lk}$$

β_k for $k = 0, \dots, K$ are the parameters to be estimated.

To run our RNN we installed the Keras and Tensorflow packages, which are essential to perform deep learning. This procedure is *particularly* challenging for Windows users. Thankfully, we were able to install them relying on a (Radziwill 2017). We considered the subset of variables selected with the hard thresholding. Our sample is roughly divided 70% – 30% between training and test set. We run a simple AR model with 5 lags to compare its performance with the RNN predictions. In order to run both models, a function which generates five lagged datasets is defined. The R-squared for the AR(5) is:

```
1 - mean((arpred - arframe[!istrain, "dl_ShareRenCons"])^2) / V0
```

```
## [1] 0.2520075
```

To estimate the RNN we must specify the values assumed by some parameters. First, within each activation vector there are 12 hidden units. We are aware that in the literature there are many rule-of-thumb methods for determining the correct number of neurons, and we relied on the one which suggests to use a number equal to 2/3 the size of the input layer.

The batch size and the number of epochs are two crucial hyperparameters. The first defines the number of samples (i.e. how many mini-sequence of input vectors X_t) to work through before updating the internal model parameters, while the latter defines the number times that the learning algorithm will work through the entire training dataset Brownlee (2018). We chose both values relying on the trends of the literature. The batch size is equal to 64 (typically under 32 there is a tendency to overfit), whereas the number of epochs is 200. To be sure that we are not overfitting the training data, we plot the history of the mean squared error of the model (labelled as 'loss') for each epoch i.e the learning curve, both for the training and the validation set. We can see that our RNN presents comparable performance on both sets. If, instead, these parallel plots start to depart consistently, it might be a sign that we are overfitting the training data.

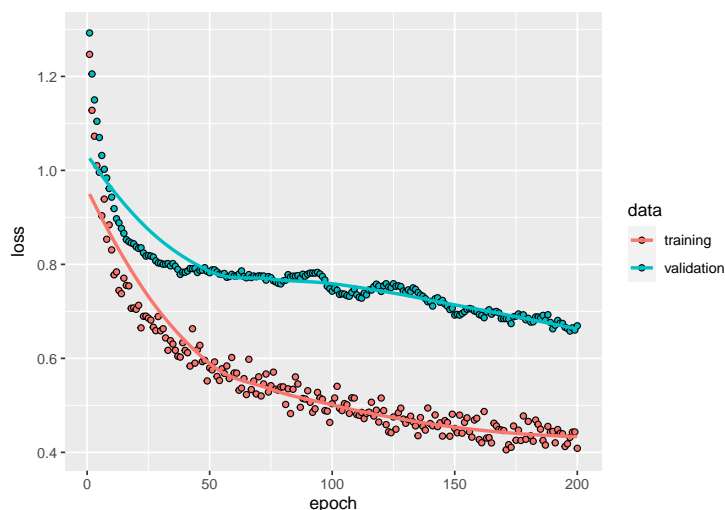


Figure 6: Learning curve

As expected, this model performs considerably better than the AR(5), because it captures complex non-linear relations between our features.

```
1 - mean((kpred - arframe[!istrain, "dl_ShareRenCons"])^2) / V0
```

```
## [1] 0.4389399
```

This goodness of fit is even more evident when we compare graphically the time series of our predicted share of renewable energy consumption (in red) with the actual one (in black).

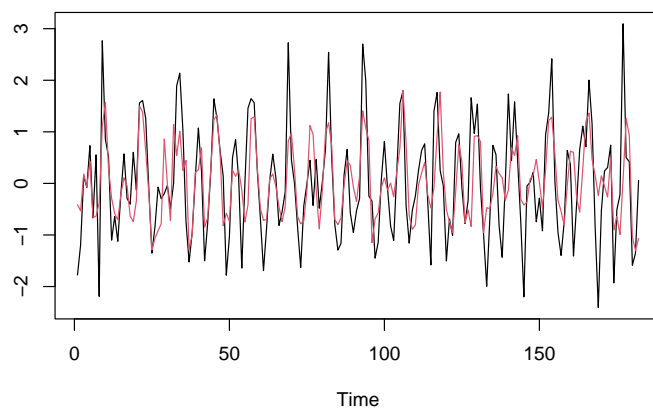


Figure 7: predicted y vs true y

4.2 Factor Models

The second approach implemented for the prediction of the share of renewable energy consumption in the U.S. is *factor analysis*, which allows to work with the entire dataset and it is particularly suited for working with high-dimensional data. In this section an estimation of the factors using two different methods will be provided: (i) Principal Component Analysis; and (ii) Expectation Maximization algorithm. Based on the factors estimated two separate direct forecasts will be performed, more precisely computing two one-step ahead forecasts of the response variable, and the results of both the estimated factors and the estimated predictions will be compared.

Factor analysis is a technique of *unsupervised learning* that in the last few decades, thanks to the increasing availability of large datasets, has been more and more applied in the field of macroeconometrics. Factor models, and in particular Dynamic Factor Model (DFM), are a dimension reduction technique that have been vastly applied in forecasting exercises. The underlying idea is that co-movement of many observed series can be summarized by few unobserved factors, therefore simplifying the high-dimensional problem (& M. M. Bańbura M. 2010). Forecasting can be then be carried out as a two-step process: first, estimate the factors from the observed predictors; second, estimate by a linear regression the relationship between the variable of interest and the few selected factors (Stock J. H. 2002).

This study focuses on a dynamic factor model, where the dynamics of each series is split in two orthogonal components: the first captures the co-movements and is driven by few common factors and the second is treated as an idiosyncratic residual representing individual features (M. M. Bańbura M. Giannone D. 2013). Moreover, working with a large dataset (i.e. large n), the more realistic approximate DFM can be considered, which allows the idiosyncratic component to be weakly correlated with all other idiosyncratic components. The approximate DFM in state-space form is therefore written as follow:

$$x_{it} = \lambda_i' F_t + \xi_{it} \quad i = 1, \dots, n, \quad t = 1, \dots, T \quad (1)$$

$$F_t = A F_{t-1} + v_t, \quad t = 1, \dots, T \quad (2)$$

where $\lambda_i' F_t$ is the *common* component and ξ_{it} the *idiosyncratic* component. We assume the vector of factors F_t to follow a VAR(1) process with $\{v_t\}$ being a zero-mean white noise process.

Estimation of the factors

Significant dimension reduction occurs when the number of selected factors is much smaller than the number of original predictors, i.e. $r \ll n$. Determination of the number of r static factors to estimate in the model can be achieved by a combination of a priori knowledge, visual inspection of a scree plot and the use of information criteria. From the cumulative variance of the first 8 and 14 factors it can be observed that already with only 8 almost 50% of the variability of \mathbf{X} is explained. Moreover, the scree plot below shows that after 13 factors the additional explanatory value of an extra factor becomes very small.

```
sum(var_explained[c(1:8)])
```

```
## [1] 0.4956183
```

```
sum(var_explained[c(1:14)])
```

```
## [1] 0.6065449
```

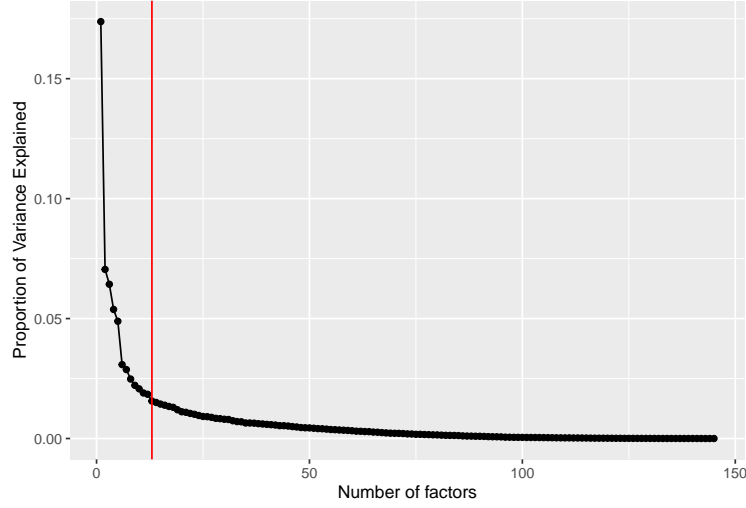


Figure 8: Scree plot

Furthermore, the information criteria proposed by Bai J. (2008) is considered, which is expressed as $IC(r) = \ln V_r(\hat{\Lambda}, \hat{F}) + rg(N, T)$, where $V_r(\hat{\Lambda}, \hat{F})$ is the sum of squared residuals when r factors are estimated using PCA and $g(N, T)$ is a penalty factor. This criteria allows to consistently estimate the number of factors from the observed data in approximate factor models and in this case indicates 10 as the optimal number, which will be considered for the estimation that follows.

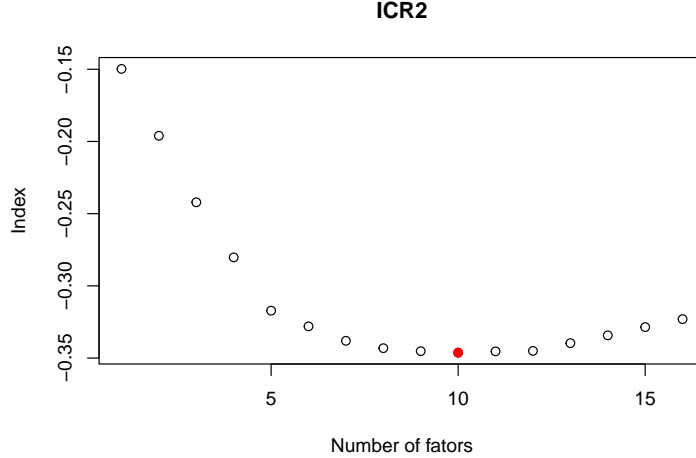


Figure 9: IC plot

Once the number of factors is determined we can proceed with the estimation of the factors. For the approximate DFM in state-space form, in the case of large n , there are two main classes of consistent estimators, which will be implemented hereafter:

1. Principal Component Analysis (PCA);
2. techniques based on Kalman Filter (KF) and Expectation Maximization (EM) algorithm

Principal Component Analysis PCA is a non-parametric technique which is simple to implement, does not need any distributional assumption and give consistent estimates of the factors in an approximate factor model for large n (Stock J. H. 2002). Factors are build as linear combinations of the observed variables in such a way that the variance of the idiosyncratic component is minimum and the variance of the common component is maximum. The first 10 factors are estimated through PCA, which account for 53.8% of the total variance. In Figure 10 the first 4 factors are plotted and show some extreme values which could be caused by the presence of outliers in the data.

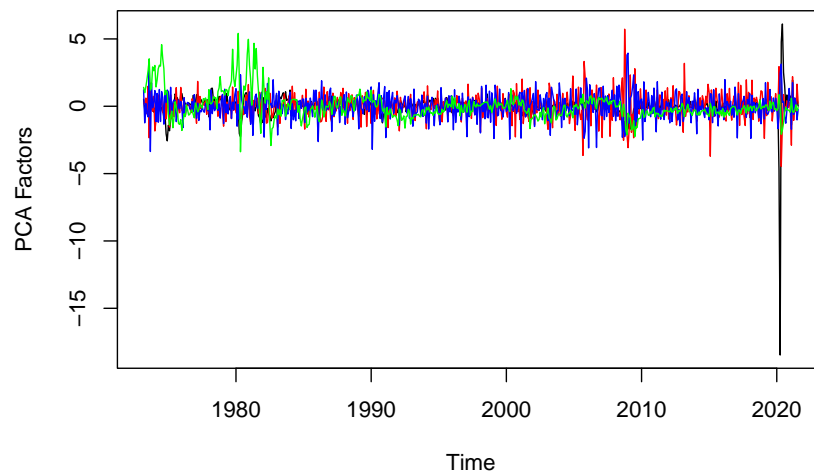


Figure 10: PCA Factors

Once factors are estimated, the one-step ahead direct forecast of the response variable can be computed. As a benchmark estimation a prediction considering only the first factor is conducted. In direct forecast the dependent variable y_{t+1} is projected onto the factor at time t . The results are showed in Figure 11, where it can be observe that the predicted response (red) under-estimate the observed one (black) throughout the series and the variance of the prediction is much smaller. The limited prediction accuracy of the one-factor model is in part determined by the fact that the first factor accounts only for 17.4% of the variability of the data, while the remaining variance is explained by individual features of the series.

```
##      Beta Coefficient PCA
## [1,]      -0.007264982

##      Y observed   Y predicted
## [575,] -0.013335495 -0.0034009139
## [576,]  0.205900952 -0.0055264587
## [577,]  0.034561403  0.0099351339
## [578,]  0.029044667 -0.0141079262
## [579,] -0.103553243 -0.0009043704
## [580,] -0.088168137 -0.0044156156
## [581,]  0.005143368 -0.0032186252
## [582,]  0.042635307 -0.0043166838
```

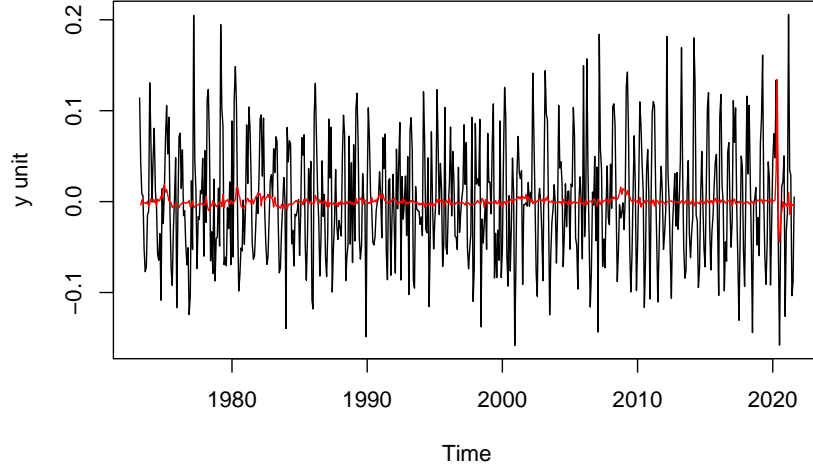


Figure 11: PCA Prediction

One of the main limitations of PCA estimation of the common component is that it does not take into account the dynamics of the factors themselves that are described in equation (2). Moreover, in the estimation of factors and loadings only the sample covariance is considered and therefore possible cross-sectional correlation of the idiosyncratic terms is not being accounted for. For this reason other techniques based on the computation and maximization of the likelihood in general are able to deliver more efficient estimates. In particular, in the following part, the Expectation Maximization algorithm is implemented.

Expectation Maximization algorithm The most efficient estimator in a parametric model such as the one we defined above is the Maximum Likelihood estimator, which requires the assumption of Gaussianity of the whole model for \mathbf{x}_t . The EM estimator is fully parametric and is computed as if the idiosyncratic components were serially and cross-sectionally uncorrelated. Nevertheless, Barigozzi & Luciani (2019) showed that the loss in efficiency in the estimation of the factor, due to this potential misspecification of the idiosyncratic covariance matrix, has a negligible impact on the estimates of the common components as n and T increase. The EM algorithm is an iterative procedure which can be employed even when n is large and is based on two sequential steps: an expectation step (E-step) and a maximization step (M-step). The EM estimator converges up to an approximation error, to the maximum likelihood estimator of the parameters of the model. Unlike PCA, the EM procedure considers also the dynamics of the factors in the estimation, delivering in general a more accurate result.

The factors are estimated considering the common shocks q equal the static factors and the EM algorithm is initialized with the PCA-VAR estimator of the parameters. Estimation of a one-factor model is performed, which, by assuming $q = r = 1$, implies that factors follow a AR(1) process and that there are no dynamics in the loadings. From the plot of the single factor in Figure 12 (left) it can be observed that also in this case there are some extreme values throughout the period. As in the PCA estimation, a one-step ahead direct forecast is conducted and again the resulting prediction of the response variable (in red) is under-estimating the observed one throughout the series and shows much lower variance. Same consideration can be made: the single first factor capture the trend of the response variable but the majority of its variability is due to the idiosyncratic component.

```
##      Beta Coefficient EM
## [1,]      0.001468433
```

```
##          Y observed   Y predicted
## [575,] -0.013335495 -0.0035692572
## [576,]  0.205900952 -0.0018173424
## [577,]  0.034561403  0.0041422131
## [578,]  0.029044667 -0.0083848352
## [579,] -0.103553243 -0.0002709016
## [580,] -0.088168137 -0.0033132509
## [581,]  0.005143368 -0.0027568005
## [582,]  0.042635307 -0.0044487695
```

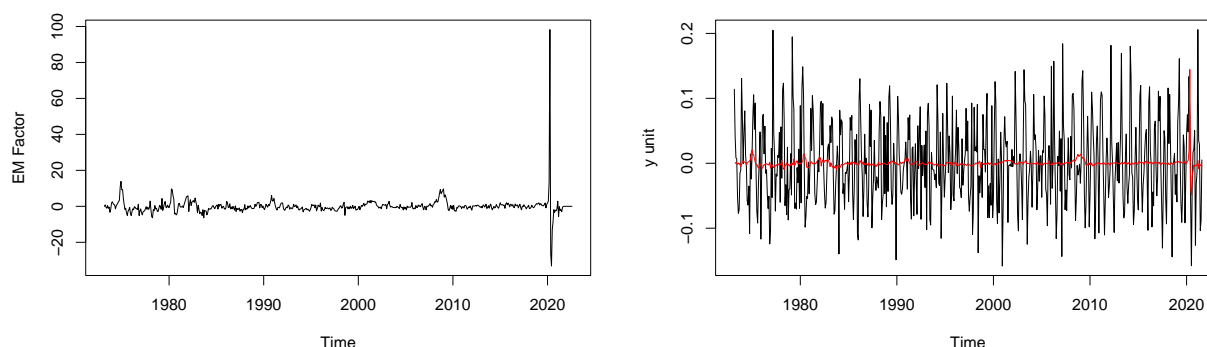


Figure 12: EM Factors and Prediction

Factor estimation by PCA and EM algorithm delivers similar results regarding the factors estimates, this is mainly due to the fact that both methods do not consider the possible cross-sectional correlation of the idiosyncratic component and the EM algorithm is initialized with the PCA estimation of the model parameters. The main difference between the two methods is that the EM algorithm is able to consider the dynamics of the factors for parameter estimation, which in our one-factor model are described by a AR(1) process. In the one-factor model case, the magnitude of the coefficients and the variance of the first factor estimated by PCA are much smaller compared to the one from the EM algorithm (Figure 13 (left)). The prediction accuracy of both models is in any case almost the same, as the computed Root Mean Squared Error (RMSE) and Figure 13 (right) demonstrate. A direct forecast considering more than one factor was also performed but the prediction results were completely imprecise and worsened with the number of factor increasing. One possible reason for this is the presence of outliers in the data which can distort the results of PCA estimation, being based on the sample covariance matrix. Moreover, the accuracy of the one-factor model results might indicate that this model specification is adequate in predicting our response variable and that the unexplained variability, being driven by outliers, is idiosyncratic.

```
##          RMSE PCA    RMSE EM
## [1,]  0.06554941  0.06552961
```

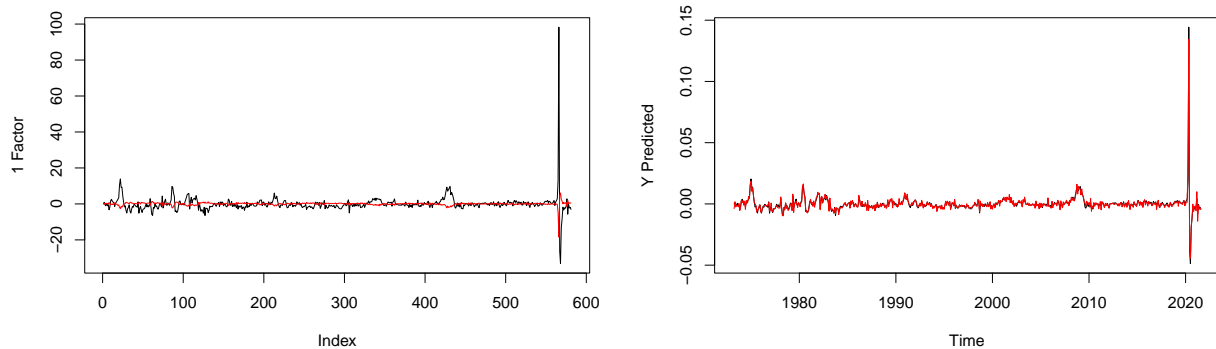



Figure 13: PCA - EM Comparison

CHAPTER 5: Conclusions

The aim of this paper was to develop machine learning algorithms to forecast the dynamics of renewable energy consumption in the US. First, we performed variable selection recurring hard thresholding, PLS and LASSO. Among these models, the LASSO selects fewer features, whereas PLS the most (three times the variables selected with hard thresholding). There is consistency between PLS and hard thresholding, as the latter selects variables that are selected also by the former. Then, we estimated an AR(5) model and trained a RNN algorithm considering the subset of predictors selected by hard thresholding. The RNN outperforms remarkably the AR(5) model, as the r-squared roughly doubles. Then, we focused on the factor model, estimating the factors with PCA and EM algorithm. Even though slightly different, the predictions with the one-factor model are accurate, whereas when we consider two factors the results are less precise, probably because of the presence of outliers in the data.

Bibliography

- Administration, US Energy Information. 2022. "Monthly Energy Review." <https://www.eia.gov/totalenergy/data/monthly/>.
- Apergis N., Payne J. 2012. "Renewable and Non-Renewable Energy Consumption-Growth Nexus: Evidence from a Panel Error Correction Model." *Energy Economics* 34: 733–38.
- Azadeh A., & Asadzadeh S. M., Babazadeh R. 2013. "Optimum Estimation and Forecasting of Renewable Energy Consumption by Artificial Neural Networks." *Renewable and Sustainable Energy Reviews* 27: 605–12.
- Bai J., Ng S. 2008. "Forecasting Economic Time Series Using Targeted Predictors." *Journal of Econometrics* 146: 304–17.
- Bañbura M., & Modugno M. 2010. "Maximum Likelihood Estimation of Factor Models on Data Sets with Arbitrary Pattern of Missing Data." *ECB Working Paper Series*, no. 1189.
- Bañbura M., Modugno M., Giannone D. 2013. "Now-Casting and the Real-Time Data Flow." *ECB Working Paper Series*, no. 1564.
- Barigozzi M., & Luciani M. 2019. "Quasi Maximum Likelihood Estimation and Inference of Large Approximate Dynamic Factor Models via the EM Algorithm" arxiv.1910.03821.
- Brownlee, Jason. 2018. "What Is the Difference Between a Batch and an Epoch in a Neural Network?" *Machine Learning Mastery* 20.
- Chong I., Jun C-H. 2005. "Performance of Some Variable Selection Methods When Multicollinearity Is Present." *Chemometrics and Intelligent Laboratory System* 78: 103–12.
- Ergun S. J., & Rivas M. F., Owusu P. A. 2019. "Determinants of Renewable Energy Consumption in Africa." *Environmental Science and Pollution Research* 26: 15390–405.
- Garthwaite, Paul H. 1994. "An Interpretation of Partial Least Squares." *Journal of the American Statistical Association* 89 (425): 122–27.
- Ghoddusi H., & Rafizadeh N., Creamer G. G. 2019. "Machine Learning in Energy Economics and Finance: A Review." *Energy Economics* 81 (May): 709–27.
- Government, German Federal. 2020. "More Energy from Renewable Sources." <https://www.bundesregierung.de/breg-en/issues/climate-action/electricity-generation-without-coal-1797118>.
- Guardian, The. 2021. "Biden's Clean Energy Plan Would Cut Emissions and Save 317,000 Lives." <https://www.theguardian.com/us-news/2021/jul/11/biden-administration-clean-energy-climate-crisis>.
- McCracken M. W., Ng S. 2016. "FRED-MD: A Monthly Database for Macroeconomic Research." *Journal of Business & Economic Statistics* 34: 574–89.
- Mehmood T., Snipen L., Hovde L. K. 2012. "A Review of Variable Selection Methods in Partial Least Squares Regression." *Chemometrics and Intelligent Laboratory Systems* 118: 62–69.
- News, abc. 2022. "Germany Shuts down Half of Its 6 Remaining Nuclear Plants." <https://abcnews.go.com/International/wireStory/correction-germany-nuclear-shutdown-story-82051054#:~:text=Germany%27s%20remaining%20three%20nuclear%20plants,by%20the%20end%20of%202022.&text=%E2%80%9C9CNuclear%20power%20plants%20remain%20high,Funke%20media%20group%20this%20week>.
- Radziwill, Nicole. 2017. "A Newbie's Install of Keras & Tensorflow on Windows 10 with r." *R-Bloggers*. <https://www.r-bloggers.com/2017/10/a-newbies-install-of-keras-tensorflow-on-windows-10-with-r/>.
- Sari R., & Soytaş U., Ewing B. T. 2008. "The Relationship Between Disaggregate Energy Consumption and Industrial Production in the United States: An ARDL Approach." *Energy Economics* 30: 2302–13.
- Sarkodie S. A., Adom P. K. 2018. "Determinants of Energy Consumption in Kenya: A NIPALS Approach." *Energy* 159: 696–705.
- Stock J. H., & Watson M. W. 2002. "Forecasting Using Principal Components from a Large Number of Predictors." *Journal of the American Statistical Association* 97 (460).
- Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1): 267–88.
- Tsai S., Zhang J., Xue Y. 2017. "Models for Forecasting Growth Trends in Renewable Energy." *Renewable and Sustainable Energy Reviews* 77: 1169–78.
- W., Tukey J. 1977. "Exploratory Data Analysis" 2: 44.
- Worldbank. 2022. "Commodity Markets." <https://www.worldbank.org/en/research/commodity-markets>.