Predicting Natural Gas Demand in Germany

Johann-Friedrich Salzmann
j.salzmann@students.hertie-school.org

Danial Riaz
d.riaz@students.hertie-school.org

Finn Krueger
f.krueger@students.hertie-school.org

Abstract

Predicting gas demand is one of the most crucial tasks for the German government today in order to limit unprecedented price hikes on consumers and create actionable policy to address the issue. Given the complexity of the dependence of gas on various segments within the consumer base i.e. residential, industrial, electricity, the consequent impact on prices and the economy and the reverse impact of prices and economic situation on demand itself, makes modelling gas demand particularly challenging. More over, social factors, such as 'responsible demand usage' by consumers is also unprecedented and one needs to introduce a model that is able to both learn from past patterns of annual seasonality as well as modify from recent day-to-day patters of consumption reduction. We have therefore made use of Machine Learning techniques in our paper to see the extent to which we can model the continuously changing daily gas demand in Germany.

We make use of various modelling techniques. We run Auto Regressive (AR), Seasonal Auto Regressive (SAR), Seasonal Autoregressive Integrated Moving Average (SARIMAX) as well as Linear Stochastic Gradient Descent (SGD) and Random Forest models and compare the performance of chronological and random test/training data splits. Our results indicate that a seasonal autoregressive random forest model with random split is the most effective in predicting gas demand with an R squared value of nearly 98%.

1. Introduction

The Russia-Ukraine war has lead Europe towards the largest energy crisis since the oil price shock of 1973. Since mid-2021, natural gas prices have been on a steep rise, with average wholesale prices at the TTF spot market well above 100 €/MWh between October 2021 and mid-2022 and prices of around 240 €/MWh in August. This is about ten times higher than the long-term pre-Covid price levels of 15–20 €/MWh (Ruhnau et al 2022).

The average price that industry paid for natural gas has quintupled compared to pre-crisis levels. Ammonia and aluminum industries were the first ones to reduce production followed by paper, brick, and steel industries in March 2022 as the spot prices increased above 200 €/MWh. The price related response to industrial consumption is complicated to predict given the complex relationship of the elasticities of these commodities as well as the competitiveness of the industries in which they operate in.

Average German residential retail prices have also increased, but with an even larger time lag. Initial survey reports, market data from the German regulator BnetzA as well as German gas market area manager Trading Hub Europe (THE) have shown a reduction of gas consumption by households. Price is of course not the only factor resulting in this reduced consumption, surveys have highlighted ethical considerations by consumers too in response to the Russian invasion of Ukraine.

Germany is an interesting case study to explore gas demand, as it is the largest export market for Russian natural gas. Furthermore, natural gas plays an essential role in Germany's industrial production as well as space heating. Reductions in Germany can therefore make a substantial contribution to solving the crisis at a European level. In order to effectively reduce the amount of gas demanded, it is crucial to have some foresight into how much gas will be demanded in the future. To that end, we will aim to predict daily German gas demand by implementing a variety of Machine Learning models.

2. Related Work

Existing research on energy demand prediction mostly makes use of classical, econometric autoregressive modeling techniqes and only sparsely incorporate machine learning methods.

Much of the literature concerning the forecast of gas demand applies Time Series (TS) models. They can be applied for a wide range of forecasting horizons (from annual to hourly) (Zhang, 2020).

One of the models often discussed is a Autoregressive Integrated Moving Average (ARIMA), which is widely applied in demand forecasting. ARIMA models use differencing to convert a non-stationary time series into a stationary one, and then predict future values from historical data. These models use “auto” correlations and moving averages over residual errors in the data to forecast future values. For short term predictions, ARIMA models require only prior data of a time series to generalize the forecast. They are however, poor long term forecasting models and not ideal for identifying unprecedented events. ARIMA model was applied to forecast annual or monthly gas demand of Turkey, with the consideration of GDP and price of gas (E. Erdogdu, 2010).

Seasonal Autoregressive Integrated Moving Average with exogenous factors, or SARIMAX, is an extension of the ARIMA class of models. Unlike the ARIMA, it is able to capture seasonality in the model. This enables it to provide qualified annual forecasting for demand of gas.

3. Proposed Method

We propose to use multiple classes of machine learning models in order to predict German national gas demand based on economic and weather data.

We assume that national aggregate gas demand can be explained by a simple decomposition:

$$D = H + I + E$$

whereas:

- D is the total gas demand.
- H is the household (and SME) demand, mostly used for residential heating, we can assume it mostly to be dependent on temperature and not much on gas prices or economics indicators,
- I is the industrial demand with regards to gas-intensive production; we assume industry market participants to react more sensitive to the overall economic situation, including short-term gas prices, and not much on weather conditions,
- E is the gas demand resulting from electricity generation of gas plants; despite this part of the demand has gone down following the scarcity of natural gas, it remains to be a relevant factor when electricity demand is peaking; we assume E to reflect an interaction of short-term gas and electricity prices.

However, we can alternatively assume that there is strong autocorrelation of daily gas demand in place, i.e., yesterday’s gas demand tends to be a strong predictor of today’s gas demand.

In our analysis, we base our demand prediction models on various models autoregressive models. Since the present value of national gas demand depends on its past value (i.e. auto-correlated) we have included seven day lagged predictor variables for all independent variables that aren’t time implicit. In comparison to the proposed ARIMA model, we have started off by adding the lagged predictor variables manually and have not applied differencing the lagged vs non lagged predictors. Manually adding the lagged variables instead of just applying ARIMA, allows us to get a better understanding of Times-Series regression models by comparing our manual regression model to ARIMA. This will provide us with first-hand insight into various effects such as differencing on our predictive capabilites and facilitate our learning experiences.

Based on the lagged variables we have build our first autoregressive model (AR) as well as an seasonal autoregressive model (SAR) that has included lagged dependent variables to capture all seasonal effects, since gas demand is highly seasonal. Our third autoregressive model is the Seasonal Autoregressive Integrated Moving Average (SARIMAX) model which is the benchmark in the current literature for autoregressive models (E. Erdogdu, 2010).

In order to use SARIMAX we have to investigate the pattern of our data to understand the seasonality.

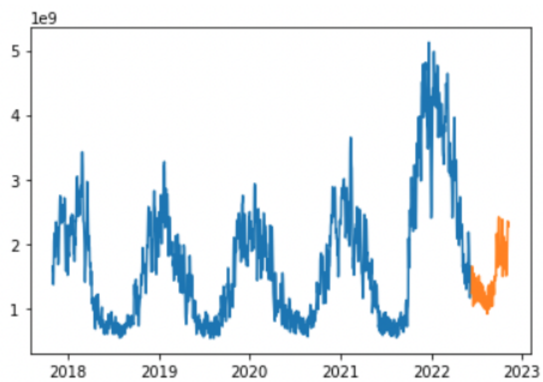

Figure 3.1. Descriptive analysis: Seasonal pattern of gas demand. Color indicates chronological train/test split

Plotting gas demand provides us with good understanding on the seasonality of our data. Sarimax models rely on seasonal differencing which is similar to regular differencing except it subtracts the value of the previous season instead of the consecutive term.

In order to forecast gas demand with the autoregressive model SARIMAX, we would have to define varies parameters to define in out model. The model is represented as

$$\text{SARIMAX}(p, d, q)(P, D, Q)m$$

whereas:

- p is the order of the autoregressive (AR) term that is regressed on previous values from the same time series,
- q is the order of the moving average(MA) that is equal to the weighted noise of the consecutive term,
- d is the number of differencing that you need to apply to make the time series stationary,
- P is the seasonal autoregressive term,
- Q is the seasonal Moving Average term,
- D is the seasonal difference order to make the data stationary,
- M is the number of time steps for the seasonal period. As time steps in this case are days, we have $M = 365$.

To estimate all parameters, we make use of the `auto_arimax` search function, aiming to find the SARIMAX model configuration with the optimal Akaike Information Critera (AIC).

In figure B.11 in the Appendix, the output of the search function is depicted. The model configuration with optimal AIC, i.e. best parameters, is

$$\text{SARIMAX}(3, 0, 0)(2, 1, 0)12$$

Besides the implemented AR, SAR and SARIMAX models that solely rely on historic data on gas demand to make predictions, we have build various static models that use a variety of variables. This allows us to see whether we can make more precise predictions based on the factors that influence the amount of gas including weather and a variety of economic factors. It also indicates how well autoregressive models perform in comparison to models that are based on a variety of other variables.

We combine the strengths of static, AR and SAR models with two classes of well-researched machine learning models:

- *Stochastic Gradient Descent (SGD)* because of their ability to cope with large-scale datasets;
- *Random Forests (RF)* as an ensemble learning method, because it works particularly well when having a large number of correlated predictors.

We use AR, SAR and SARIMAX models as a benchmark of what prediction quality can be achieved by simple (or complex) timeseries analysis, as opposed to collecting many features and fitting big data sets with machine learning approaches.

4. Experiments

We have implemented the models with a tech stack of R (used for data merging, cleaning, imputation & feature engineering) and Python (used for actual ML modeling)[1].

Data: We compiled a dataset consisting of 1833 daily observations (roughly 5 years) and 510 predictor variables, including transformations. In order to predict the daily national gas demand in Germany we use weather data from various cities within Germany. We combine them with general economic and day-to-day market price data.

- Daily national gas demand (merged and transformed): national gas demand measured in kWh, retrieved from the THE (Trading Hub Europe, joint company of all German regional gas net 2 work providers). Before October 2021, NetConnect and Gaspool shared the market area, so THE has made available historic demand data from both markets. Gas demand is split into types (H-Gas vs. L-Gas, Psyn vs. Pana vs. RLMmT) that we combine to a summed total gas demand as predictor while adding daily gas type proportions of the total as additional predictors.
- Gas prices (merged and imputed): Daily gas spot price data was obtained through the Dutch TTF Natural Gas price data (Futures) available on investing.com and combined with Spot Market data retrieved from the European Energy Exchange (the European Energy Exchange (EEX) is an integration between Powernext and Gaspoint Nordic that publishes market data every business day after the conclusion of its settlement window).
- Weather data (cleaned and imputed): Weather data Average, minimum and maximum daily temperature measured in Celsius as well as other factors including precipitation, wind speed, and sun hours taken from Frankfurt, Berlin, Munich and Bonn retrieved from Meteostat (Meteostatis one of the largest vendors of open weather and climate data. The platform provides access to long-term time series of thousands of weather stations), allowing to feed the model with sufficient data on different weather conditions within the country.
- Co2 Prices (cleaned and imputed): Carbon Emissions Futures Historical Data over the last 5 years retrieved from investment.com , reflecting daily Emissions pricing in Europe.

[1] Code is available at `https://github.com/jfsalzmann/gasprices.git`

- Electricity Prices (cleaned and imputed): We obtained German electricity price data for the last 5 years through the DBEc1 Electricity futures data retrieved from investing.com.

- DAX (cleaned and imputed): The DAX, (Deutscher Aktienindex, German stock index) is a stock market index consisting of the 40 major German blue chip companies trading on the Frankfurt Stock Exchange. Data on DAX is publicly available on several platforms and reflects the general German economic performance and general fluctuations.. We obtained the daily DAX data over the past 5 years retrieved from from Yahoo Finance.

- Quarterly GDP (cleaned and imputed): German quarterly GDP data was obtained from data.worldbank.org for the last 5 years, reflecting the general German economic performance, too.

We use a variety of engineered features (in the following, $n \in N$ is the number of observations, whereas $i \in I$ are the model prediction variables).

- A number of numeric date based features is included (day, week, month, weekday, year), as well their cyclic sin and cos projections:

$$x_{i,\text{sin}}^n = \sin\left(2\pi \max_{n \in N}(x_i^n)\, x_i^n\right)$$

$$x_{i,\text{cos}}^n = \cos\left(2\pi \max_{n \in N}(x_i^n)\, x_i^n\right)$$

- For all variables, including date transformations, we add ln, square and elasticity transformations:

$$x_{i,\text{ln}}^n = \ln(x_i^n)$$

$$x_{i,\text{sq}}^n = (x_i^n)^2$$

$$x_{i,\text{el}}^n = (x_i^n)^{-2}$$

- Additionally, a 7-day lagged copy of all variables except date features is added.

$$x_{i,\text{lag}(7)}^n = x_i^{n-7}$$

Various imputation efforts are undertaken to address the issue that many of the variables, especially stock market data, contains missing observations. As a result of adding lagged variables,

- All transformations are ensured to produce valid results. Infinite values that arise as artifacts are replaced by very large numeric values, i.e.

$$\infty \;\mapsto\; 10^3 \max_{n \in N}(x_i^n)$$

- As we use time series data, last-value-put-forward (quarterly data), cyclic (gas type proportions), and linear (all other features) imputation is applied to features with missing observations based on the chronological order.

Not all features are included in every model; this applies particularly for gas demand based features used for autoregression, here referred to as y, and date features, here referred to as `dd`. Based on the findings of extensive feature importance analysis, we removed gas type proportions, here referred to as `gas`, as predictors in the *Full Combined Autoregressive* model and only made available gas type proportions with a 7-day lag. This constrains the model to learn only from data that realistically would be available or reasonably possible to impute/forecast for a gas demand prediction of the future.

See Table 5.1, 5.2 for the model specifications. See Appendix A for the respective specifications of the baseline models implemented.

Midterm Preliminary	y	`dd`	`gas`	...
x_i^n				
$x_{i,\text{ln}}^n$; $x_{i,\text{sq}}^n$; $x_{i,\text{el}}^n$				
x_i^{n-7}				
$x_{i,\text{ln}}^{n-7}$; $x_{i,\text{sq}}^{n-7}$; $x_{i,\text{el}}^{n-7}$				
$x_i^{n-d}\ \forall\, d \in \{1,\ldots,7\}$				

Table 4.1. Model configuration: predictors, original & transformed, included for respective models & data splits. Color indicates inclusion (green), exclusion (red), endogenous (orange).

Full Autoregressive	y	`dd`	`gas`	...
x_i^n				
$x_{i,\text{ln}}^n$; $x_{i,\text{sq}}^n$; $x_{i,\text{el}}^n$				
x_i^{n-7}				
$x_{i,\text{ln}}^{n-7}$; $x_{i,\text{sq}}^{n-7}$; $x_{i,\text{el}}^{n-7}$				
$x_i^{n-d}\ \forall\, d \in \{1,\ldots,7\}$				

Table 4.2. Model configuration: predictors, original & transformed, included for respective models & data splits. Color indicates inclusion (green), exclusion (red), endogenous (orange).

Evaluation method: We utilise and compare two different models: A variety of autoregressive (AR) models, the seasonal extension (SAR), and SARIMAX as well as an SGD Regressor model, and a Random Forest model with the number of estimators set to 20.

All models are run with a chronological train/test split of the data, and a random split, each using 100 observations for testing and 1733 observations for training/validation purposes. For the chronological split, the last 100 observations get separated (i.e., data from the last couple of months), whereas in the random split a random sample is drawn. While the latter may break seasonal patterns for autoregression and lagged features, the former puts models to stress as Germany's gas demand pattern deviated significantly from the previous years following public efforts to save as much gas as possible. However, using such a model in production would obviously result in the requirement to predict on out-of-sample data more similar to the most recent observations.

We therefore systematically compare the performance of all model types in both split configurations over a total of 18 models. It turned out to be challenging to train models in a way that they perform especially good over those last observations, while focusing less on prediction excellence of the historic time series.

To normalise features and ensure comparability across models, we use a `StandardScaler` and a `QuantileTransformer` with the number of quantiles set to 100.

We compare model results based on a visual, qualitative general impression of fit, with particular focus on prediction performance conditional on its date; as well as we compare quantitative indicators such as $RMSE$ and R^2. We also analyse and report feature importance for Linear SGD and Random Forest Models. While the latter model class provides feature importance by design, for the former we use the absolute estimated parameters β_i under the assumption of data to be normalised and centered:

$$\text{varimp}(\beta_i^{LinearSGD}) = \left|\beta_i^{LinearSGD}\right|$$

Experimental details: With the model configurations introduced above, each model is trained and fit using their specific data sets generated in the `R` based pre-processing environment. In a standardised manner, performance indicators, variable importance and graphs are automatically generated in `Python` Jupyter Notebooks[2].

Most of the data gets pulled using web APIs or scraping approaches. In a production environment, this would allow the models to be run automatically every day to predict next days' gas demand based on latest data. It would allow to run models over a period of several days in order to observe and enhance performance robustness. The architecture of the modeling environment is designed to be modular and production-ready in every aspect, being agnostic to data sources and predictors used. This is not only relevant for reproduction and allows for flexible recycling of the models implemented, but it also matters with regards to the constraint that high-quality data that is required to run high-quality models in a production environment would be based on commercial market and weather data, most likely.

[2]Reproducible model runs are accessible in the notebooks at `https://github.com/jfsalzmann/gasprices/tree/main/learning`

Results: In our initial report, we proposed a proof-of-concept baseline model with only a few key variables that was generally uncleaned, unimputed and only for a short time frame of about a year, without any lagged features. It performed well enough to conduct further research and strive for refinement. However, systematic model biases were visible and shaped our further research and modeling work.

In our midterm report, we had reported premilinary modeling results based on static (i.e., non-autoregressive) features with strong focus on feature engineering. We reported the model performance to be overall good based on qualitative assessment of the model plots and with regards to quantitative measures. Note that we report corrected performance results for the preliminary models as we found several modeling issues after submitting the previous report.

Most literature on energy demand prediction uses autoregressive models. We therefore implemented a variety of autoregressive baseline models, including SARIMAX. We improved adjusted the configuration of our full combined final models as we strived for even higher model performance.

Table 4.3 shows the results of all respective models.

Figures 4.1, 4.2 visualise the fit and performance of the baseline SARIMAX model.

Figures 4.3, 4.4, 4.5, 4.6 visualise the performance of the best performing model in each model class, highlighted as grey in Table 4.3. Table 4.4 shows the variable importance of the best performing final model.

See Appendix B & C for the respective figures of all other models.

#	Quality	Type	Model	Split	$RMSE$	R^2
1	P.O.C.	Static	Linear SGD	Chron.	3.5e8	0.9039
2	Baseline	AR	Linear SGD	Rand.	2.4e8	0.8553
3	Baseline	AR	Linear SGD	Chron.	3.6e8	0.6532
4	Baseline	AR	RF Trees	Rand.	2.1e8	0.9446
5	Baseline	AR	RF Trees	Chron.	2.0e8	0.7517
6	Baseline	SAR	Linear SGD	Rand.	3.4e8	0.8680
7	Baseline	SAR	Linear SGD	Chron.	3.7e8	0.1642
8	Baseline	SAR	RF Trees	Rand.	1.4e8	0.9644
9	Baseline	SAR	RF Trees	Chron.	1.4e8	0.8764
10	Baseline	SARIMAX	Linear AR	Chron.	1.6e8	0.8150
11	Preliminary	Static	Linear SGD	Rand.	2.7e8	0.9308
12	Preliminary	Static	Linear SGD	Chron.	3.4e8	0.3025
13	Preliminary	Static	RF Trees	Rand.	1.4e8	0.9860
14	Preliminary	Static	RF Trees	Chron.	4.4e8	−0.150
15	Full Comb.	SAR	Linear SGD	Rand.	2.2e8	0.9434
16	Full Comb.	SAR	Linear SGD	Chron.	3.4e8	0.3018
17	Full Comb.	SAR	RF Trees	Rand.	1.3e8	0.9794
18	Full Comb.	SAR	RF Trees	Chron.	1.5e8	0.8680

Table 4.3. Model performance: quantitative measures for the respective models & data splits. Color indicates inclusion (green), exclusion (red), endogenous (orange).

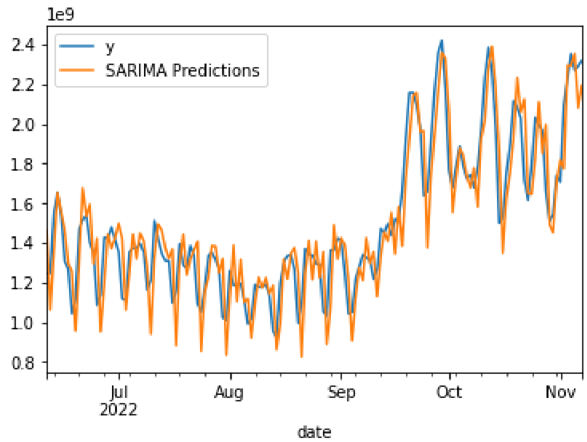

Figure 4.1. Baseline SARIMAX model, chronological split: gas demand, June-November 2022. Color indicates predicted/actual gas demand dates

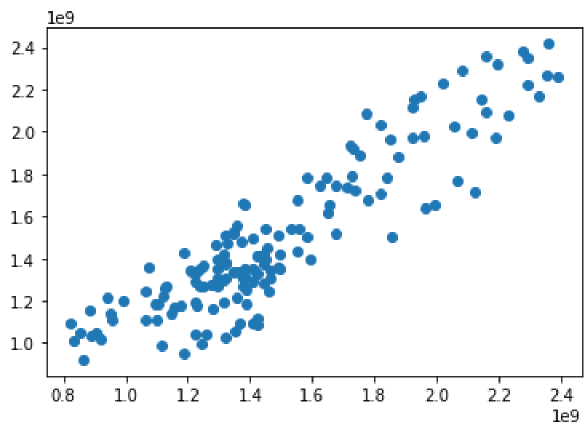

Figure 4.2. Baseline SARIMAX Model, chronological split: predicted vs. actual gas demand, test data. Color indicates prediction dates

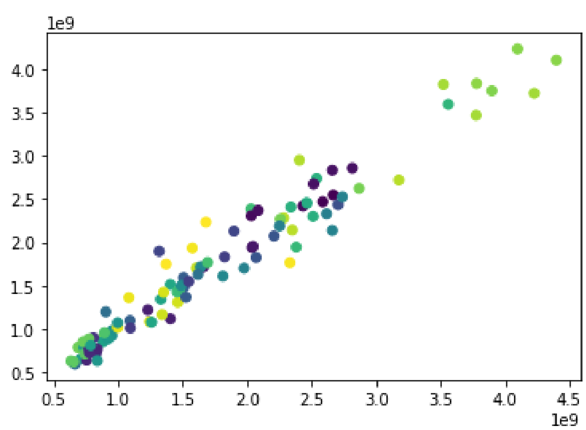

Figure 4.3. Baseline AR Model, Random Forest, chronological split: predicted vs. actual gas demand, test data. Color indicates prediction dates

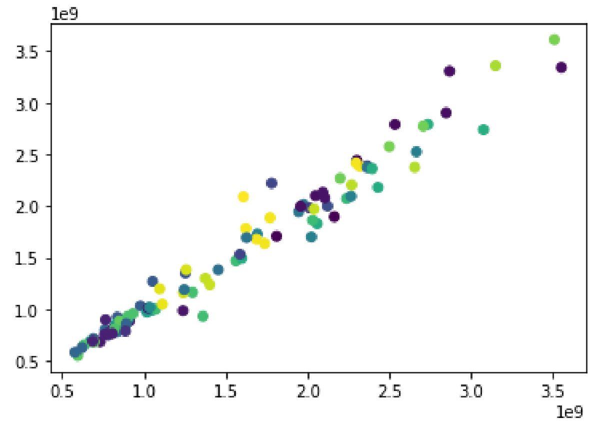

Figure 4.4. Baseline SAR Model, Random Forest, random split: predicted vs. actual gas demand, test data. Color indicates prediction dates

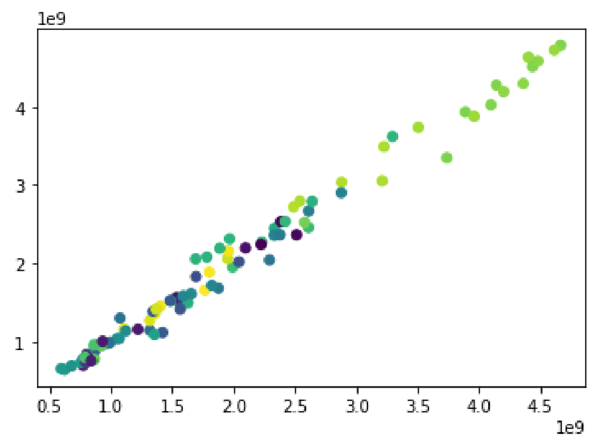

Figure 4.5. Preliminary Static Model, Random Forest, random split: predicted vs. actual gas demand, test data. Color indicates prediction dates

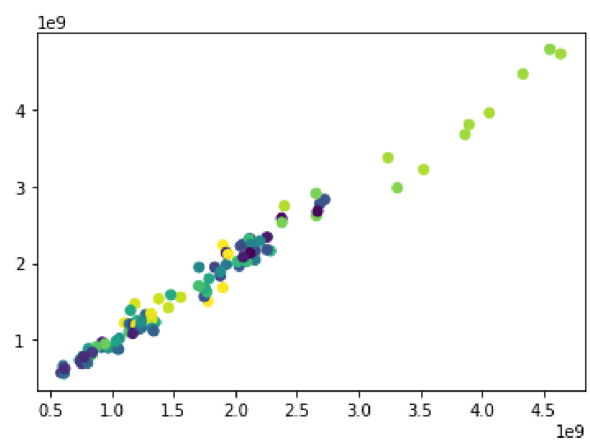

Figure 4.6. Fully Combined Autoregressive Model, Random Forest, random split: predicted vs. actual gas demand, test data. Color indicates prediction dates

5. Analysis

From our various baseline models, we find that:

- Gas demand can be predicted using weather and economic data; the proof-of-concept Linear SGD Model with chronological split performs with $R^2 = 0.9039$,
- Gas demand is strongly autocorrelated; when building simple AR models, the best model is a Random Forest with random split, performing with $R^2 = 0.9446$,
- Seasonal patterns matter; when extending to SAR models, the best model is a Random Forest with random split, performing with $R^2 = 0.9644$,
- Machine Learning models win over classical econometrical models; the SARIMAX model with chronological split performs with $R^2 = 0.8150$, confirming that there is strong autocorrelation in place in any case.

These models provide a benchmark for evaluating any production model candidate. Our propsed models are able to pass the threshold, even though not all of them perform better than the benchmark. Still, we are able to present two models that do:

- As for our midterm report, we presented preliminary static (i.e., non-autoregressive) models; the best model is a Random Forest with random split, performing with $R^2 = 0.9860$; however, as noted above, we changed some of the model configurations to avoid overfitting and therefore propose
- Final models with SAR; the best model is a Random Forest with random split, performing with $R^2 = 0.9794$. When comparing $RMSE$ values, it outperforms the preliminary model.

More generally, we find:

- In almost all model configurations, Random Forest models outperform Linear SGD models,
- Most Random Forest models, but almost no Linear SGD model outperform the baseline benchmark results, implying that classical econometric modeling approaches would be superior,
- Our combination strategy works: pairing autoregressive and machine learning modeling approaches increases the predictive strength. The SAR ML models with full variable configuration do the trick and are able to catch the underlying principles of gas demand. With regards to variable importance, we can observe that the best model combines the lagged autoregressrive predictor with a cyclic feature, weekday, as well as transformed and original temperature time series.

Variable	Importance
`y____lag7`	9.613811e-01
`dd_weekday`	2.816362e-03
`M_tavg__sqrd`	2.419733e-03
`M_tavg__sqrd`	2.403892e-03
`M_tavg`	1.443557e-03

Table 4.4. Fully Combined Autoregressive Model, Random Forest, random split: variable importance of the five most important variables

Besides the general success of the proposed approach, we recognise the complex dependencies of our identified independent variables on predicting gas demand. This is in line with our expectations as the relationship between our target and model features is not a very direct one and many times we have made use of either variable approximations or variable transformations. Random Forests are particularly good at being agnostic towards the nature and number of features one adds to their model.

Our findings are in line with that of Zhang (2019), though we do not make use of Deep-RNN models for natural gas demand prediction, we do see the effectiveness of Non-linear Autoregressive Models (NAR) as highlighted by Zhang and colleagues. There are some challenges in making comparisons of the fit of our model with models studied in other studies in the past; namely that of the difference in data used. The events we have seen unfolded in the last few months and subsequently the market dynamics, particularly in Germany and Europe at large, have really been an anomaly over the last decade and thus any model comparison from other studies needs to be covering the same period of events.

Interestingly, we observe that in most cases, random splits outperform chronological splits in our models. This is not something we had anticipated as generally with time-series data where auto-correlation is present, splitting train-test data on time is preferable to splitting it randomly in order to improve accuracy for predicting fresh data. In a future analysis we would like to explore this observation further.

6. Conclusions

In this project, we built a natural gas demand predicting model completely from scratch using intuitive assumptions about model requirements as well as constraints. We sourced open source data on a variety of variables of interest obtaining information on a daily level.

We demonstrated the effectiveness of Random Forest models over Linear SDG models, particularly those with a random split rather than a chronological split in making these predictions.

Additionally, we have shown how Random Forest models are better than conventional Auto Regressive models such as Sarimax, due to their advantage to disentangle complex information from highly correlated independent variables.

In a future study, we would like to explore how our model performs with unexpected demand side fluctuations (voluntary or through regulatory means). We would also like to test out the ability of deep RNN models for natural gas demand forecasting on longer time horizons, e.g., days or months. Lastly, it would be interesting to see whether a hybrid method of Sarimax together with a Random Forest model would be able to combine the quality of both and lead to even more precise prediction.

7. Acknowledgements

We would like to acknowledge support from Hertie School PhD candidate Oliver Ruhnau who has conducted previous research in EU gas market modelling. Oliver clarified a number of concepts for us as well as indicated potential sources where reliable and publicly available data can be obtained for gas markets.

8. Contributions

We worked on a collaborative basis on most aspects of this assignment. FK had the lead in identifying popular approaches in academia for modelling gas demand, supported by DR and JFS. DR had the leaed in identifying data sourcing, integration and analysis of the findings, supported by JFS and FK. JFS had the lead in the model architecture, specification and troubleshooting, supported by FK and DR.

9. References

1. Erdogdu, E. (2010). Natural gas demand in Turkey.Applied Energy,87(1), 211-219.

2. J. D. Hunter, "Matplotlib: A 2D Graphics Environment" (2007), Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95

3. Pandas development team, T., 2020. pandas-dev/pandas: Pandas. URL: https://doi.org/10.5281/zenodo.3509134, doi:10.5281/zenodo. 3509134.

4. Pedregosa, F., Varoquaux, Gael, Gramfort, A., Michel, V., Thirion, B., Grisel, O., … others. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825–2830.

   , E. H. A., Fawzy, H., Fattah, A. M. A. (2021). Time series forecasting using tree based methods. J. Stat. Appl. Probab, 10, 229-244.

5. Rose, B., Floehr, E. (2017). Analysis of high-temperature forecast accuracy of consumer weather forecasts from 2005–2016. Forecast Watch Report.(Dublin, OH, USA).

6. Ruhnau, Oliver; Stiewe, Clemens; Muessel, Jarusch; Hirth, Lion (2022) : Gas demand in times of crisis: energy savings by consumer group in Germany, ZBW – Leibniz Information Centre for Economics, Kiel, Hamburg

7. Sadamori, IEA Press Release, 3rd Oct 2022 https://www.iea.org/news/natural-gas-markets-expected-to-remain-tight-into-2023-as-russia-further-reduces-supplies-to-europe

8. Su, H., Zio, E., Zhang, J., Xu, M., Li, X., Zhang, Z. (2019). A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced Deep-RNN model.Energy,178, 585-597."

9. Vinayak Sharma, Ümit Cali, Bhav Sardana, Murat Kuzlu, Dishant Banga, Manisa Pipattanasomporn (2021). Data-driven short-term natural gas demand forecasting with machine learning techniques. Journal of Petroleum Science and Engineering, (), –. doi:10.1016/j.petrol.2021.108979

10. van der Walt, S., Colbert, S.C., Varoquaux, G., 2011. The numpy array: A structure for efficient numerical computation. Computing in Science Engineering 13, 22–30.

Appendix A. Model Configuration

This is an extension of the model configurations presented in Section 5.

AR Baseline	y	`dd`	`gas`	...
x_i^n	green	red	red	red
$x_{i,\text{ln}}^n$; $x_{i,\text{sq}}^n$; $x_{i,\text{el}}^n$	red	red	red	red
x_i^{n-7}	red	red	red	red
$x_{i,\text{ln}}^{n-7}$; $x_{i,\text{sq}}^{n-7}$; $x_{i,\text{el}}^{n-7}$	red	red	red	red
$x_i^{n-d}\ \forall\, d{\in}\{1,\ldots,7\}$	green	red	red	red

Table A.1. Model configuration: predictors, original & transformed, included for respective models & data splits. Color indicates inclusion (green), exclusion (red), endogenous (orange).

SAR Baseline	y	`dd`	`gas`	...
x_i^n	green	green	red	red
$x_{i,\text{ln}}^n$; $x_{i,\text{sq}}^n$; $x_{i,\text{el}}^n$	red	red	red	red
x_i^{n-7}	red	red	red	red
$x_{i,\text{ln}}^{n-7}$; $x_{i,\text{sq}}^{n-7}$; $x_{i,\text{el}}^{n-7}$	red	red	red	red
$x_i^{n-d}\ \forall\, d{\in}\{1,\ldots,7\}$	green	red	red	red

Table A.2. Model configuration: predictors, original & transformed, included for respective models & data splits. Color indicates inclusion (green), exclusion (red), endogenous (orange).

SARIMAX Baseline	y	`dd`	`gas`	...
x_i^n	green	orange	red	red
$x_{i,\text{ln}}^n$; $x_{i,\text{sq}}^n$; $x_{i,\text{el}}^n$	red	red	red	red
x_i^{n-7}	red	red	red	red
$x_{i,\text{ln}}^{n-7}$; $x_{i,\text{sq}}^{n-7}$; $x_{i,\text{el}}^{n-7}$	red	red	red	red
$x_i^{n-d}\ \forall\, d{\in}\{1,\ldots,7\}$	orange	red	red	red

Table A.3. Model configuration: predictors, original & transformed, included for respective models & data splits. Color indicates inclusion (green), exclusion (red), endogenous (orange).

Midterm Preliminary	y	`dd`	`gas`	...
x_i^n	red	green	green	green
$x_{i,\text{ln}}^n$; $x_{i,\text{sq}}^n$; $x_{i,\text{el}}^n$	red	green	green	green
x_i^{n-7}	red	red	green	green
$x_{i,\text{ln}}^{n-7}$; $x_{i,\text{sq}}^{n-7}$; $x_{i,\text{el}}^{n-7}$	red	red	green	green
$x_i^{n-d}\ \forall\, d{\in}\{1,\ldots,7\}$	red	red	red	red

Table A.4. Model configuration: predictors, original & transformed, included for respective models & data splits. Color indicates inclusion (green), exclusion (red), endogenous (orange).

Full Autoregressive	y	`dd`	`gas`	...
x_i^n	green	green	red	green
$x_{i,\text{ln}}^n$; $x_{i,\text{sq}}^n$; $x_{i,\text{el}}^n$	red	green	red	green
x_i^{n-7}	green	red	green	green
$x_{i,\text{ln}}^{n-7}$; $x_{i,\text{sq}}^{n-7}$; $x_{i,\text{el}}^{n-7}$	red	red	green	green
$x_i^{n-d}\ \forall\, d{\in}\{1,\ldots,7\}$	red	red	red	red

Table A.5. Model configuration: predictors, original & transformed, included for respective models & data splits. Color indicates inclusion (green), exclusion (red), endogenous (orange).

Appendix B. Model Performance

This is an extension of the model performance graphs presented in Section 5.

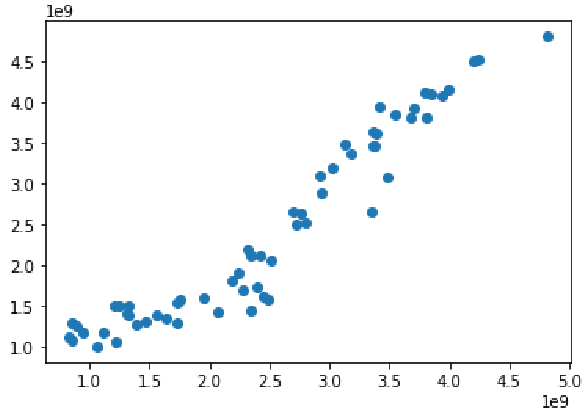

Figure B.1. Baseline Proof of Concept Model, Linear SGD, random split: predicted vs. actual gas demand, test data. Color indicates prediction dates

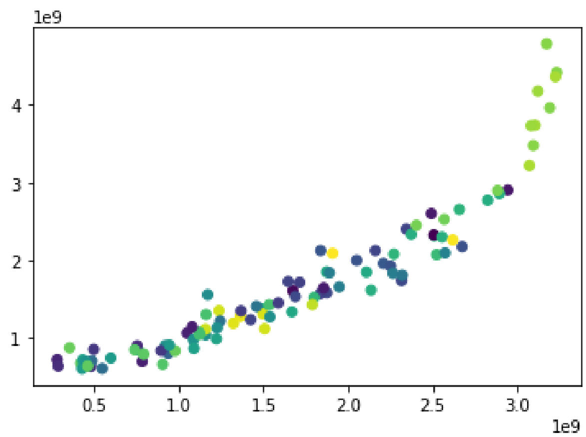

Figure B.2. Baseline AR Model, Linear SGD, random split: predicted vs. actual gas demand, test data. Color indicates prediction dates

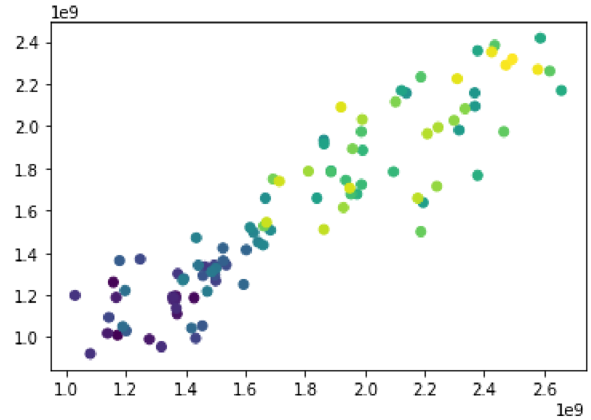

Figure B.3. Baseline AR Model, Linear SGD, chronological split: predicted vs. actual gas demand, test data. Color indicates prediction dates

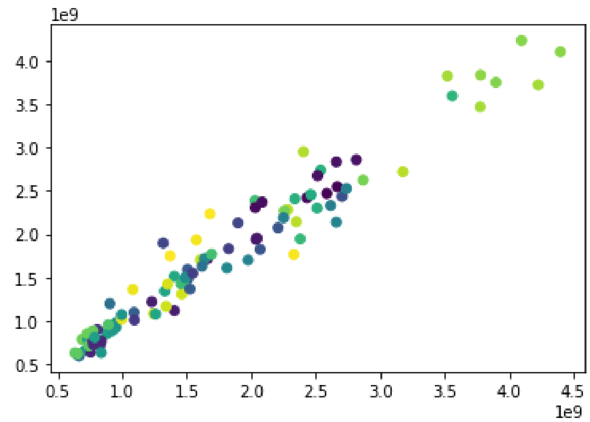

Figure B.4. Baseline AR Model, Random Forest, random split: predicted vs. actual gas demand, test data. Color indicates prediction dates

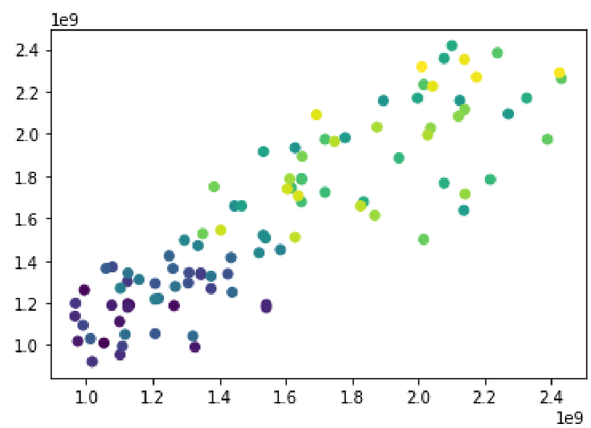

Figure B.5. Baseline AR Model, Random Forest, chronological split: predicted vs. actual gas demand, test data. Color indicates prediction dates

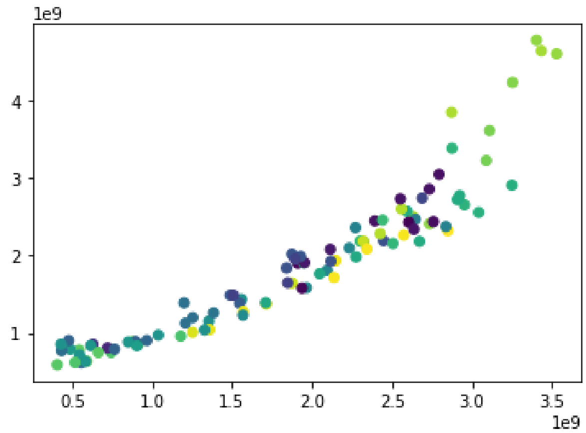

Figure B.6. Baseline SAR Model, Linear SGD, random split: predicted vs. actual gas demand, test data. Color indicates prediction dates

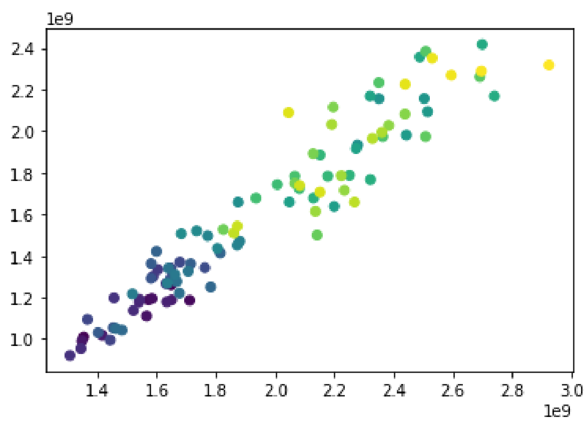

Figure B.7. Baseline SAR Model, Linear SGD, chronological split: predicted vs. actual gas demand, test data. Color indicates prediction dates

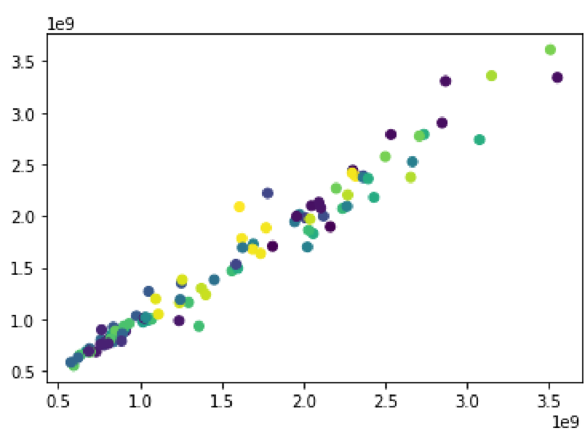

Figure B.8. Baseline SAR Model, Random Forest, random split: predicted vs. actual gas demand, test data. Color indicates prediction dates

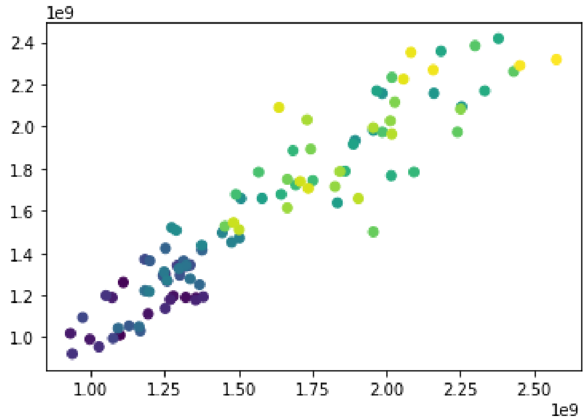

Figure B.9. Baseline SAR Model, Random Forest, chronological split: predicted vs. actual gas demand, test data. Color indicates prediction dates

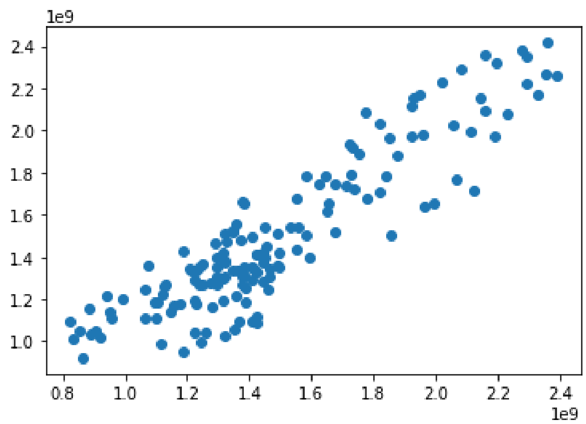

Figure B.10. Baseline SARIMAX Model, chronological split: predicted vs. actual gas demand, test data. Color indicates prediction dates

SARIMAX Results

Dep. Variable:	y	**No. Observations:**	1833
Model:	SARIMAX(3, 0, 0)x(2, 1, 0, 12)	**Log Likelihood**	-37319.423
Date:	Thu, 01 Dec 2022	**AIC**	74650.846
Time:	10:25:37	**BIC**	74683.889
Sample:	11-01-2017	**HQIC**	74663.036
	- 11-07-2022		
Covariance Type:	opg		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
ar.L1	1.3567	0.029	46.873	0.000	1.300	1.413
ar.L2	-0.6554	0.046	-14.385	0.000	-0.745	-0.566
ar.L3	0.2164	0.029	7.460	0.000	0.160	0.273
ar.S.L12	-0.7794	0.028	-28.007	0.000	-0.834	-0.725
ar.S.L24	-0.3290	0.027	-12.063	0.000	-0.382	-0.276
sigma2	5.429e+16	nan	nan	nan	nan	nan

Ljung-Box (L1) (Q):	0.00	**Jarque-Bera (JB):**	242.47
Prob(Q):	0.99	**Prob(JB):**	0.00
Heteroskedasticity (H):	1.66	**Skew:**	-0.02
Prob(H) (two-sided):	0.00	**Kurtosis:**	4.79

Figure B.11. Baseline SARIMAX model, chronological split: parameters estimated by the `Auto_ARIMAX` function `dates`

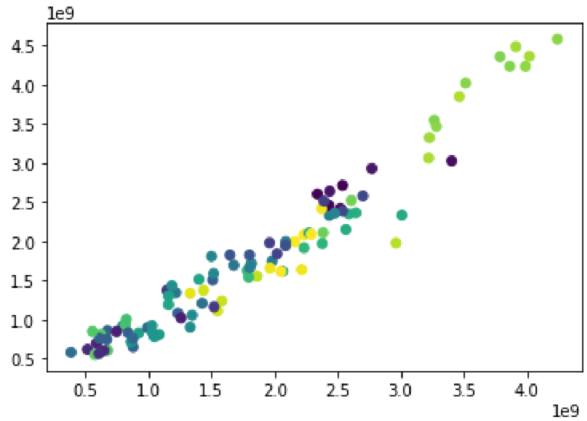

Figure B.12. Preliminary Static Model, Linear SGD, random split: predicted vs. actual gas demand, test data. Color indicates prediction dates

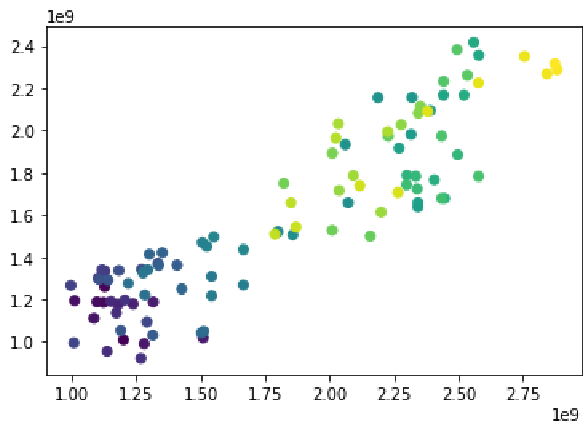

Figure B.13. Preliminary Static Model, Linear SGD, chronological split: predicted vs. actual gas demand, test data. Color indicates prediction dates

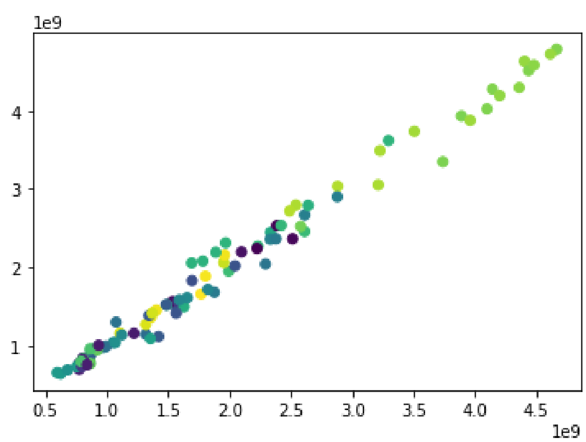

Figure B.14. Preliminary Static Model, Random Forest, random split: predicted vs. actual gas demand, test data. Color indicates prediction dates

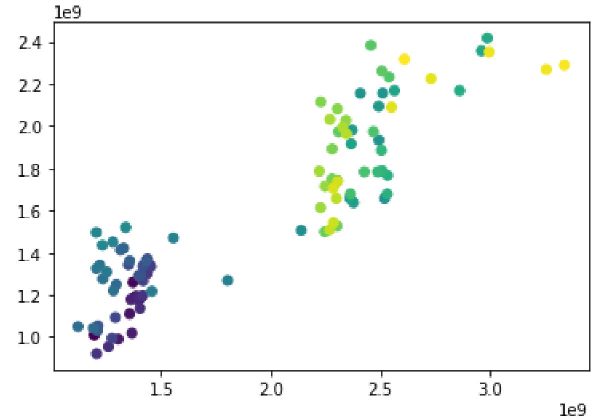

Figure B.15. Preliminary Static Model, Random Forest, chronological split: predicted vs. actual gas demand, test data. Color indicates prediction dates

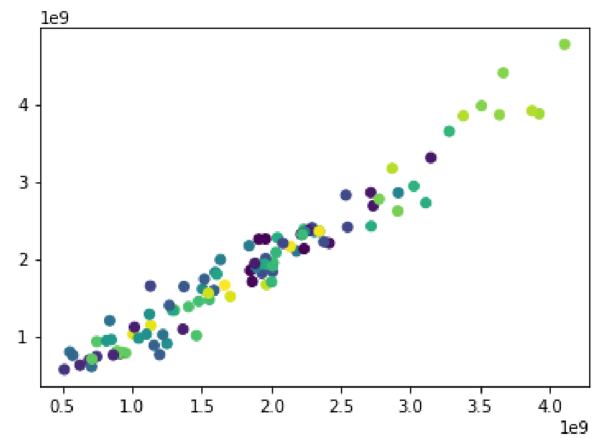

Figure B.16. Fully Combined Autoregressive Model, Linear SGD, random split: predicted vs. actual gas demand, test data. Color indicates prediction dates

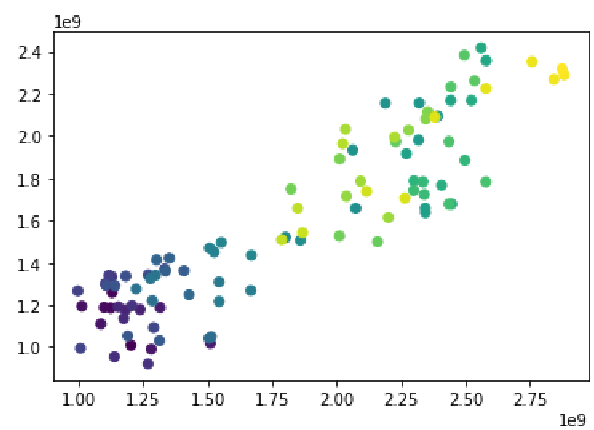

Figure B.17. Fully Combined Autoregressive Model, Linear SGD, chronological split: predicted vs. actual gas demand, test data. Color indicates prediction dates

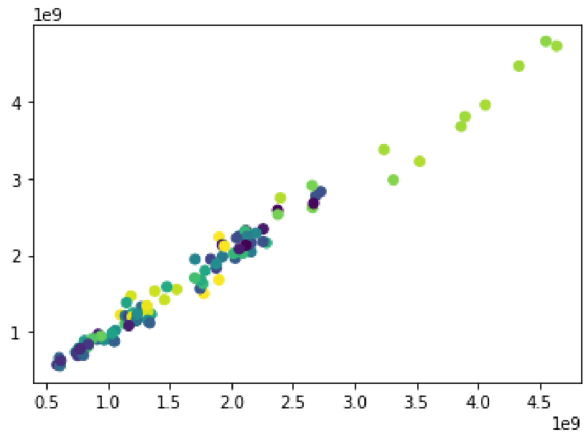

Figure B.18. Fully Combined Autoregressive Model, Random Forest, random split: predicted vs. actual gas demand, test data. Color indicates prediction dates

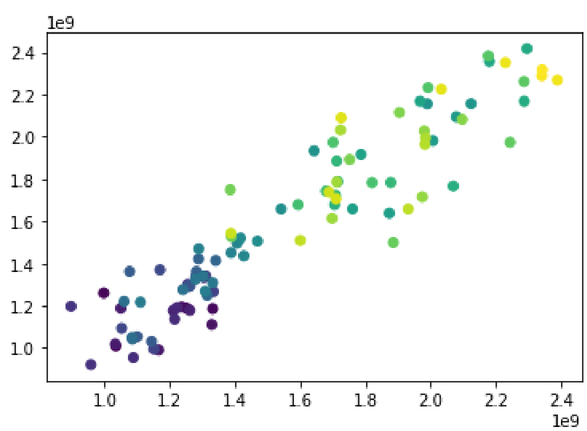

Figure B.19. Fully Combined Autoregressive Model, Random Forest, chronological split: predicted vs. actual gas demand, test data. Color indicates prediction dates

Appendix C. Variable Importance

This is to document variable importance as mentioned in Section 5. We report the 5 most important variables according to the model specific ordered importances.

Variable	Importance
y____lag7	9.590452e-01
M_tavg	2.882924e-03
dd_weekday	2.471446e-03
M_tavg__sqrd	1.995944e-03
FF_tavg__sqrd	1.556226e-03

Table C.1. Variable Importance of the five most important variables, Model 1

Variable	Importance
y____lag7	9.613811e-01
dd_weekday	2.816362e-03
M_tavg__sqrd	2.419733e-03
M_tavg__sqrd	2.403892e-03
M_tavg	1.443557e-03

Table C.2. Variable Importance of the five most important variables, Model 2

Variable	Importance
M_tavg	1.713032e-01
M_tavg__ln	1.409057e-01
FF_tavg__sqrd	8.723867e-02
FF_tavg_ln	6.749186e-02
FF_tavg__elas	5.959719e-02

Table C.3. Variable Importance of the five most important variables, Model 3

Variable	Importance
FF_tavg__sqrd	1.408371e-01
M_tavg__ln	1.112041e-01
FF_tavg	9.357799e-02
FF_tavg_ln	7.166409e-02
M_tavg	5.911612e-02

Table C.4. Variable Importance of the five most important variables, Model 4

Variable	Importance
M_tavg	1.713032e-01
M_tavg__ln	1.409057e-01
FF_tavg__sqrd	8.723867e-02
FF_tavg_ln	6.749186e-02
FF_tavg__elas	5.959719e-02

Table C.5. Variable Importance of the five most important variables, Model 5

Variable	Importance
FF_tavg__sqrd	1.408371e-01
M_tavg__ln	1.112041e-01
FF_tavg	9.357799e-02
FF_tavg_ln	7.166409e-02
M_tavg	5.911612e-02

Table C.6. Variable Importance of the five most important variables, Model 6

Variable	Importance
y____lag6	2.169415e-01
y____lag3	1.748242e-01
y____lag2	1.542732e-01
y____lag7	1.375416e-01
y____lag4	1.025686e-01

Table C.7. Variable Importance of the five most important variables, Model 7

Variable	Importance
y____lag6	2.169415e-01
y____lag3	1.748242e-01
y____lag2	1.542732e-01
y____lag7	1.375416e-01
y____lag4	1.025686e-01

Table C.8. Variable Importance of the five most important variables, Model 8

Variable	Importance
y____lag7	4.362982e+08
y____lag5	4.362982e+08
y____lag4	4.362982e+08
y____lag3	4.362982e+08
y____lag2	4.362982e+08

Table C.9. Variable Importance of the five most important variables, Model 9

Variable	Importance
y____lag7	4.068534e+08
y____lag6	4.068534e+08
y____lag5	4.068534e+08
y____lag4	4.068534e+08
y____lag3	4.068534e+08

Table C.10. Variable Importance of the five most important variables, Model 11