CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
CERGE-EI

Electricity Load Forecasting: Impact of COVID-19 on the Czech Republic's Load Profile

Master's thesis

Author: Christopher Nyasha Mutama
Study program: MA Economic Research
Supervisor: Prof. Silvester van Koten
Year of defense: 2025

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.[1]

The author grants to Charles University permission to reproduce and distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, July 29, 2025

Christopher Nyasha Mutama

[1]During the preparation of this thesis, the author used Gemini and ChatGPT to assist with refining the translation of the abstract from English to Czech, LaTeX code formatting and PDF/A compliance of the final files. The author reviewed and edited the content after use of these tools and takes full responsibility for the final content.

Abstract

Accurate forecasting of electricity demand is critical for stable grid operation, energy policy formulation, and investment planning. Shocks threaten this stability, which in turn potentially introduces economic problems. This thesis investigates the impact of COVID-19 on the Czech Republic's electricity load profile. Unlike most forecasting methods which rely on historical consumption data, this study generates synthetic electricity load profiles based on weather variables. The annual peak load is the only load input for the model and is intended to condition the model to the consumption limits. The methodology is inspired by Behm et al. (2020), and the resulting counterfactual analysis lays the foundation for estimating the impact of future shocks on the Czech Republic's load profile in the medium and long term. Through an iterative and systematic evaluation of a LASSO regression model, an Artificial Neural Network (ANN), and an eXtreme Gradient Boosting (XGBoost) model, this study identifies XGBoost as the most robust and accurate method for this application. Using an XGBoost model with LASSO-selected features, I estimate that COVID-19 reduced electricity demand in the Czech Republic by 2.66 TWh between 2020 and 2021, equivalent to 4.0% of 2019's total annual load (66.15 TWh). This "COVID" effect was more pronounced in 2020 compared to 2021. A statistically significant difference was found between day and night impacts: the calculated average reductions in consumption were 145.68 MW during the day and 157.37 MW at night ($p = 0.03$). While unexpected, the pronounced night-time decline may reflect broader systemic changes in industrial and commercial consumption patterns that persisted, and may still persist beyond the typical working hours.

JEL Classification	C53, C58, Q40
Keywords	Electricity demand forecasting, COVID-19, Czech Republic, XGBoost, LASSO, synthetic load profile

Abstrakt

Přesné predikce poptávky po elektřině jsou klíčové pro stabilní provoz elektrizační soustavy, tvorbu energetické politiky a plánování investic. Různé šoky tuto stabilitu ohrožují a mohou vést k hospodářským problémům. Tato práce analyzuje dopad pandemie COVID-19 na odběrový profil elektřiny v České republice (dále jen ČR). Na rozdíl od většiny metod predikce, které vycházejí z historických dat o spotřebě, tato studie generuje syntetické odběrové profily na základě meteorologických proměnných. Roční špičkový odběr slouží jako jediný vstup modelu a pomáhá model upravit s ohledem na limity spotřeby. Metodologie je inspirována prací Behma a kol. (2020) a výsledná kontrafaktuální analýza poskytuje základ pro odhad dopadu budoucích šoků na odběrový profil ČR ve střednědobém a dlouhodobém horizontu. Na základě iterativního a systematického srovnání modelů LASSO regrese, umělé neuronové sítě (ANN) a metody eXtreme Gradient Boosting (XGBoost) identifikuje studie XGBoost jako nejrobustnější a nejpřesnější metodu pro tuto úlohu. Pomocí modelu XGBoost s proměnnými vybranými metodou LASSO odhaduji, že COVID-19 snížil poptávku po elektřině v ČR mezi lety 2020 a 2021 o 2,66 TWh, což odpovídá 4,0% celkové roční spotřeby v roce 2019 (66,15 TWh). Tento „COVID efekt“ byl výraznější v roce 2020 než v roce 2021. Statisticky významný rozdíl ($p = 0,03$) byl zjištěn mezi denními a nočními dopady: průměrné snížení spotřeby činilo 145,68 MW během dne a 157,37 MW v noci. Ačkoli je tento výsledek překvapivý, výraznější pokles spotřeby v nočních hodinách může odrážet hlubší strukturální změny v průmyslových a komerčních spotřebních vzorcích, které přetrvávaly a mohou nadále přetrvávat i mimo běžnou pracovní dobu.

JEL Classification	C53, C58, Q40
Klíčová slova	Predikce elektrické zátěže, COVID-19, Česká republika, XGBoost, LASSO, syntetický profil zátěže
Název	Predikce elektrické zátěže: Dopad COVID-19 na profil zátěže v České republice

Acknowledgments

First and foremost, I would like to sincerely thank my supervisor, doc. Silvester van Koten, Ph.D, for his expert guidance throughout the development of this thesis. His profound knowledge, insightful feedback, and unwavering encouragement have been instrumental in shaping my research. I am truly grateful for his mentorship and support, which have greatly enriched my academic journey.
I also extend my deepest gratitude to my professors, lecturers, tutors, and the entire CERGE-EI community. This thesis is the culmination of every interaction, discussion, and shared insight. Remove even one of these, and the road would have been far tougher.
To my mother, I owe boundless thanks for her unwavering encouragement and motivation. Her constant belief in my potential has been a driving force, reminding me that I can achieve whatever I set my mind to.
To my late father, your memory and values have been a constant source of strength throughout this endeavor.
My heartfelt appreciation goes to my uncle, a father figure and mentor, whose invaluable conversations and professional insights contributed significantly to this work.
Finally, I would like to express my gratitude to all my family, friends, and colleagues for their continual support, encouragement, and good wishes. Their collective presence has been a cornerstone of both my academic and personal growth.

Typeset in LaTeX using the IES Thesis Template.

Bibliographic Record
Mutama, Christopher Nyasha: *Electricity Load Forecasting: Impact of COVID-19 on the Czech Republic's Load Profile.* Master's thesis. Charles University, Faculty of Social Sciences, CERGE-EI, Prague. 2025, pages 71. Advisor: Prof. Silvester van Koten

Contents

List of Tables

List of Figures

Acronyms

AI Artificial Intelligence

ANN Artificial Neural Network

ARIMA Autoregressive Integrated Moving Average

ČEPS Czech Transmission System Operator; ČEPS, a.s.

CNN Convolutional Neural Network

ENTSO-E European Network of Transmission System Operators for Electricity

ES Exponential Smoothing

GBM Gradient Boosting Machine

GPR Gaussian Process Regression

LASSO Least Absolute Shrinkage and Selection Operator

LLM Large Language Model

LSTM Long Short-Term Memory

MLP Multi-layer Perceptron

MW Megawatt

MWh Megawatt-hours

ReLU Rectified linear unit

RNN Recurrent Neural Network

SARIMA Seasonal Autoregressive Integrated Moving Average

SARIMAX Seasonal ARIMA with Exogenous Variables

SVM Support Vector Machine

SVR Support Vector Regression

TWh Terawatt-hours

VAR Vector Autoregression

XGBoost eXtreme Gradient Boosting

Master's Thesis Proposal

Author	Christopher Nyasha Mutama
Supervisor	Prof. Silvester Van Koten
Proposed topic	Electricity Load Forecasting: Impact of COVID-19 on the Czech Republic's Load Profile

Motivation Accurate electricity demand forecasting is critical for stable grid operations, policy formulation, and investment decisions. The outbreak of the COVID-19 pandemic was met with government interventions which had an impact on how societies functioned. Electricity demand is driven by a complex and nonlinear set of factors, including weather conditions, economic activity, and behavioral changes. Shocks like the COVID-19 pandemic disrupt historical consumption patterns, and can alter the economic landscape of a country. The proposed work treats the pandemic as a unique chance to understand how the load profile's trajectory changed due to factors linked to the measures that followed the pandemic. This work aims to estimate the impact of the pandemic on electricity consumption without relying on historical load data. Instead it attempts to generate a purely synthetic load profile to capture the unprecedented changes induced by the pandemic. By relying on weather variables as the main predictors of demand, the proposed approach avoids the risk of encoding pandemic-era anomalies into future forecasts. For policymakers, the results of the study can serve as a foundation step for future assessments of various factors on the Czech Republic's electricity load profile, enabling more robust medium- to long-term forecasting and planning. In addition, the study contributes to an evolving body of knowledge seeking to understand how electricity demand response to prolonged non-economic shocks.

Research Question: To what extent did the COVID-19 pandemic alter the Czech Republic's electricity load profile for 2020 and 2021?

Methodology: This study proposes to generate synthetic electricity load profiles primarily based on weather variables, drawing inspiration from the method developed by Behm et al. (2020). While their method was originally conceptualized as the only candidate, the methodology will take an iterative and systematic approach to evaluate a LASSO regression model, an Artificial Neural Network (ANN), and an eXtreme Gradient Boosting (XGBoost) model. The basis of selecting these models is strong performance in prior energy forecasting literature, along with their ability to capture nonlinearities and complex interactions. XGBoost is anticipated to be identified as a robust and accurate method for this application. Data to be utilized will include hourly weather data from Visual Crossing Weather and hourly electricity load data from ČEPS. The COVID stringency index from OXGRT will be used to represent pandemic severity for analysis. The final XGBoost model, with LASSO-selected features, will be trained on pre-pandemic data to generate counterfactual "business-as-usual" load profiles for 2020 and 2021. The generated synthetic profiles will then be compared to the observed demand to measure the effect for disruptions related to the pandemic.

Model Justification: The study will involve a systematic evaluation of LASSO regression, Artificial Neural Networks (ANNs), and eXtreme Gradient Boosting (XGBoost) models. While ANNs are recognized for capturing complex non-linear relationships, XGBoost is expected to emerge as a superior model for this application due to its robustness and higher accuracy in predicting the Czech Republic's electricity load based on weather data. LASSO regression will also be crucial for initial model testing and for effective feature selection, which is anticipated to reduce dimensionality of the input features and improve the performance of both the ANN and XGBoost models during the evaluation. Trialing multiple models ensures that the forecasts are based on empirical performance and the systematic evaluation and selection process provides reliable counterfactual load profiles necessary for accurately quantifying the pandemic's impact.

Expected Contributions: Behm et al. (2020) position their study as the first to develop and validate ANN-generated synthetic profiles for European electricity load modeling, and present it as an alternative to regression-based methods. This study contributes to the literature on electricity load forecasting by demonstrating how synthetic counterfactual profiles can be used to isolate the impact of exogenous shocks. Furthermore, the proposed work offers one of the first empirical evaluations of pandemic-related demand changes in the Czech Republic using high-resolution data. The findings may inform future energy policy and electricity planning.

References

Behm, C., Nolting, L., & Praktiknjo, A. (2020). How to model European electricity load profiles using artificial neural networks. *Applied Energy, 277*, 115564. https://doi.org/10.1016/j.apenergy.2020.115564

ČEPS. (2016). *Mid-term adequacy forecast of the Czech Republic*. ČEPS.

Lai, J.-P., Chang, Y.-M., Chen, C.-H., & Pai, P.-F. (2020). A survey of machine learning models in renewable energy predictions. *Applied Sciences, 10*(17), 5975.

McKibbin, W., & Fernando, R. (2023). The global economic impacts of the COVID-19 pandemic. *Economic Modelling, 129*.

Nabavi, S. A., Mohammadi, S., Motlagh, N. H., Tarkoma, S., & Geyer, P. (2024). Deep learning modeling in electricity load forecasting: Improved accuracy by combining DWT and LSTM. *Energy Reports, 12*, 2873–2900. https://doi.org/10.1016/j.egyr.2024.08.070

Řanda, M. (2023). *Short-term electric load forecasting using Czech data* [Master's thesis, Charles University].

Chapter 1

Introduction

Today's economies have a strong connection with energy consumption and relies on a stable and predictable supply. Among all energy forms, electricity is particularly critical. It is a core input for industrial processes and an everyday need in households. Furthermore, electricity consumption has a complex and often bi-directional relationship with other economic indicators including industrial output and GDP. As such, it can serve as an indicator of the economic performance of a country, and this utility can be valuable for decision-making for economies that are coupled to electricity consumption. Electricity consumption can be represented by an electricity load profile, which is defined as a graphical representation of this consumption over a given period of time. Likewise, an energy profile is a broad representation of the consumption patterns across the entire energy portfolio. These patterns can reflect underlying structural changes and growth trends in an economy.

Consequently, unexpected changes to the dynamics of demand and supply can have significant implications on the economy. The impact horizon can differ depending on the nature of the change. Precise load forecasting is thus essential for both economic stability and operational efficiency of the power grid. As early as 1985, Bunn and Farmer showed that improving forecasts by just a 1% reduction in error could save the United Kingdom 10 million pounds annually in operating costs. Although precise figures for the Czech Republic's operating costs are not available, this logic is universal. Nabavi et al. (2024) explain that lower predictions can lead to shortages and blackouts, especially during peak times when power grids are overloaded. On the other hand, if the predictions are too high, input, storage, and maintenance costs rise. For the Czech Republic, the implications of forecasting accuracy extend beyond the country's borders. The Czech Republic is among the top four net exporters of electricity in the European Union. According to the Czech Statistical Office (2022), over 30,000 GWh of electricity were exported in 2022, an increase from just over 26,000 GWh the previous year. This is over 40% of the current domestic consumption, making electricity an important commodity. The primary export markets are Germany, Slovakia, and Poland, with the 2023 exports earning the country over 2 billion United States dollars (World Bank, 2023). Based on this, the economic implications of load forecasting precision are significant, and inaccurate predictions of domestic demand directly affect the Czech Republic's ability to maximize revenue from exports. Underestimating domestic demand reduces the surplus available for export, while overestimations could result in wasted resources and higher operational costs, making exports less profitable. Load forecasting is a challenging exercise even under normal conditions. When sudden economic and societal shocks are introduced, knock-on effects on the dynamics of electricity demand and supply can be expected, which in turn have economic implications.

1.0.1 Impact of Shocks

The paralyzing power outage in Prague, which occurred in the first week of July 2025, brings perspective to the impact of shocks on the forces of demand and supply. In a preliminary report, ENTSO-E (2025) discloses that about 1,500 MW of production and 2,700 MW of consumption were lost due to the event. While the estimated financial losses are yet to be determined, projected losses from the 2025 Spanish blackout paint a picture. Mestres Domènech & Martín Vilató (2025) reveal that on the day of the Spanish blackout, consumer spending declined by 34%, and the financial losses amounted to an estimated 8% of the daily GDP. While this interruption was a one-time shock in the short term, the disclosed financial losses are economically significant.

This thesis aims to understand how COVID-19 affected the Czech Republic's electricity load profile for strategic energy infrastructure planning and policy formulation. The impact horizon of the pandemic offers an opportunity to understand how shocks that result in similar societal restrictions can alter the load profile. To achieve this, a synthetic load profile of the Czech Republic is generated for the years 2020 and 2021. The research question is: **to what extent did the COVID-19 pandemic alter the Czech Republic's load profile?** While the blackouts in Spain and the Czech Republic are examples of sudden and direct shocks, the government measures accompanying the pandemic were sustained. COVID-19 represents indirect shocks that have the potential to change the dynamics of supply and demand in the medium and long term. The resulting measures led to widespread changes in mobility, work routines, and global economic activity in general. McKibbin and Fernando (2023) use a combination of a dynamic stochastic general equilibrium and computable general equilibrium model, which they term "the G-cubed model," to measure the global economic effects of the pandemic. Their results show that the economic impact of

COVID-19 was both severe and highly variable across sectors and countries. While precise figures depend on assumptions, their simulations suggest that global GDP losses ranged from significant short-term contractions in the most optimistic scenarios to multi-trillion-dollar declines under more severe pandemic trajectories. Even though the full economic impact is complex to quantify, McKibbin and Fernando (2023) note how trade and manufacturing were heavily impacted, with utilities experiencing shifts in demand patterns. Given the relationship between electricity consumption and the economy, it is likely that changes in the economy are mirrored by electricity consumption patterns. Several studies have explored the impact of the pandemic on the electricity load profile from different perspectives. One study aimed to compare the changes in sectoral demand, with increased residential and reduced industrial consumption noted in the case of Sweden and Chile (van Zoest et al. 2023). A detailed discussion of other studies related to the impact of the pandemic on electricity demand is presented in the literature review section. However, to the best of my knowledge, no study has investigated the impact of COVID-19 on the Czech Republic's load profile in the medium to long term.

Instead of the common forecasting approaches that rely on historical values of the target variable, this thesis builds on the approach proposed by Behm et al. (2020) and generates synthetic load profiles from weather and calendrical data. The authors used the annual peak load as an input parameter to scale synthetic load profiles while preserving weather-dependent patterns. Similarly, I utilize the annual peak load for 2016, 2017, and 2018 to anchor the model in line with Behm et al. (2020). Using the pre-pandemic annual peak grounds the analysis in actual grid conditions and guards against grid overload (Nabavi et al. 2024). The hourly weather data was sourced from Visual Crossing Weather, while the hourly electricity load data was sourced from ČEPS, the Czech Transmission System Operator. To determine how the pandemic altered the load profile for the Czech Republic, I conduct a counterfactual analysis using a model trained on pre-pandemic data to predict consumption patterns, simulating a "business-as-usual" scenario using weather variables to generate a synthetic load profile. The severity of COVID-19 is represented by a COVID stringency index, whose values are sourced from the Oxford COVID-19 Government Response Tracker (OXGRT). Three different models will be trained and tested against 2019 data to select the most accurate. The selected model will be used to generate the synthetic load profiles for 2020 and 2021. In addition, a heuristic analysis is performed to provide a simple estimation of the effects of weaker and stronger stringency policies.

Consequently, this thesis contributes to the existing literature by offering a detailed case study to understand how COVID-19 and similar shocks may alter electricity consumption patterns in the Czech Republic. A counterfactual analysis is consistent with Perçuku et al.'s (2025) argument to future-proof electricity grids against unforeseen disruptions such as pandemics and weather-induced disasters. Finally, it validates Behm et al.'s (2020) findings that a purely synthetic load profile can be accurately generated without using past values of the dependent variable. Chapter 2 provides context of the Czech Republic's load profile since the year 2000. It discusses the energy profile in general, before focusing on the observed electricity load trends from 2016 to 2024. Chapter 3 reviews the existing literature and examines approaches to load forecasting and studies related to the impact of the pandemic on electricity consumption. Chapter 4 discusses the core models identified in the literature review in greater detail, while Chapter 5 describes the data and details the methodology. Chapter 6 presents the counterfactual analysis, including policy before the conclusion in Chapter 7 summarizing the findings is presented.

Chapter 2

Czech Republic Electricity Load Profile

This chapter contextualizes the electricity load profile of the Czech Republic, starting with the an overview of the energy profile. The observed load trends from 2016 to 2024 are discussed, focusing on the pre-COVID trends before discussing the post-COVID trends. Lastly, these observed trends are compared against earlier forecasts from ČEPS, highlighting any deviations along with implications for understanding the dynamics of Czech electricity demand.

2.0.1 Energy and Load Profile Overview

According to an ODYSSEE-MURE (2025) report, the general energy profile of the Czech Republic has shifted across economic sectors since the start of the millennium. From 2000 to 2022, the net energy consumption in Czechia increased by 0.5 million tonnes of oil equivalent (Mtoe), reaching 24.5 Mtoe by the end of 2022. However, the industrial and manufacturing base reduced their consumption by 26%, partly due to efficiency gains which averaged 2.8% annually. The services sector experienced a 17% reduction in consumption, and recorded average efficiency improvements of 1.4% per year.

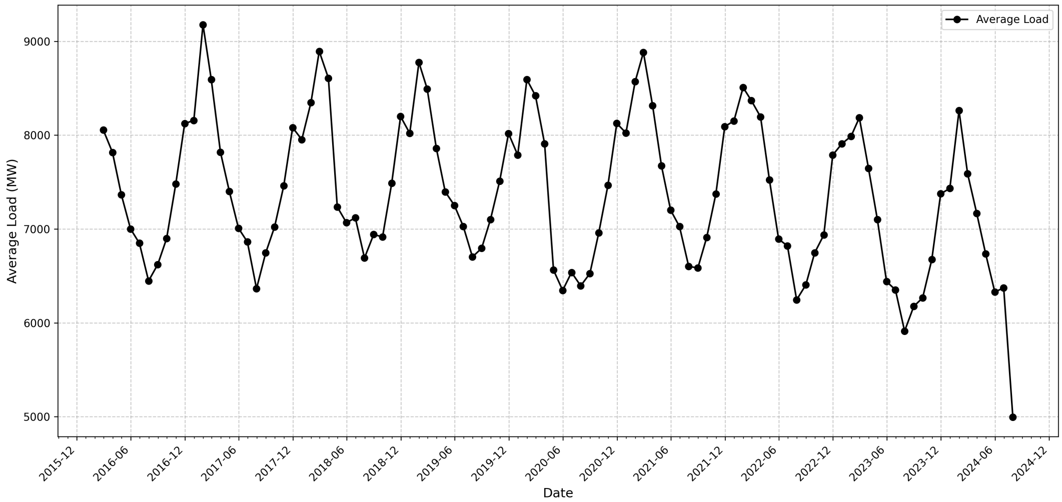
On the contrary, households increased their consumption over the same period. Energy consumption increased by 12%, despite a 1.1% annual efficiency gain. The report notes an increase in the number of residential buildings and appliance use as the driving forces. Similarly, the transport sector showed a significant increase in energy consumption by 65%, which was accompanied by a 2% decline in efficiency. Road transport dominated, accounting for 97% of its consumption. Overall, the net energy efficiency in the Czech Republic improved by an average of 1.5% per year from 2000 to 2022, resulting in a 28% cumulative gain. Despite saving of 7.8

Mtoe, the efficiency gains were inadequate to decouple general energy consumption from economic activity. Uğurlu & Jindřichovská (2024) find evidence of this coupling specific to electricity demand. In their analysis of the relationship between electricity consumption, trade, and GDP post COVID-19, the authors find a long run elasticity of 0.08, confirming that the Czech Republic's electricity demand remains structurally linked to macroeconomic performance.

ČEPS (2016) in their Mid-term Adequacy Forecast (MAF) detail the changes of the Czech Republic's electricity load profile. Starting in the second half of 2014 and becoming more pronounced in early 2015, the Czech electricity system experienced a significant rise in overall consumption. This increase coincided with the first measurable signs of economic recovery following the Great Recession of 2008, again suggesting an economy coupled with electricity consumption. A correlation between economic activity and electricity demand can be assumed (ČEPS, 2016). A more detailed examination of average weekly peak loads revealed an annual increase of around 1.6% between 2014 and 2015, pointing to a growth in baseline consumption as well as higher volatility in demand throughout the year (ČEPS, 2016).

Projections provided by ČEPS anticipated that the peak load would rise from 9,900 MW in 2016 to approximately 10,500 MW by 2025, marking a 6.1% expected increase. This relatively modest growth rate was among the lowest in the Central and Eastern European (CEE) region and reflects a conservative outlook on economic expansion and energy efficiency improvements (ČEPS, 2016). Furthermore, this increase in peak load was expected to occur alongside structural changes in the electricity energy mix, with higher penetration of variable renewable energy sources expected. At the same time, gradual reduction of fossil fuel generated electricity was expected. At the EU level, peak load was projected to grow faster after 2020, leading to concerns about load volatility. The report stressed the need to improve reserve capacity for the EU as an emerging risk likely to mirror the Czech electricity distribution system (ČEPS, 2016). Figure 1 depicts the average load profile of the Czech Republic from 2016 to 2024.

Figure 2.1: Czech Republic Average Load Profile (2016 - 2024)



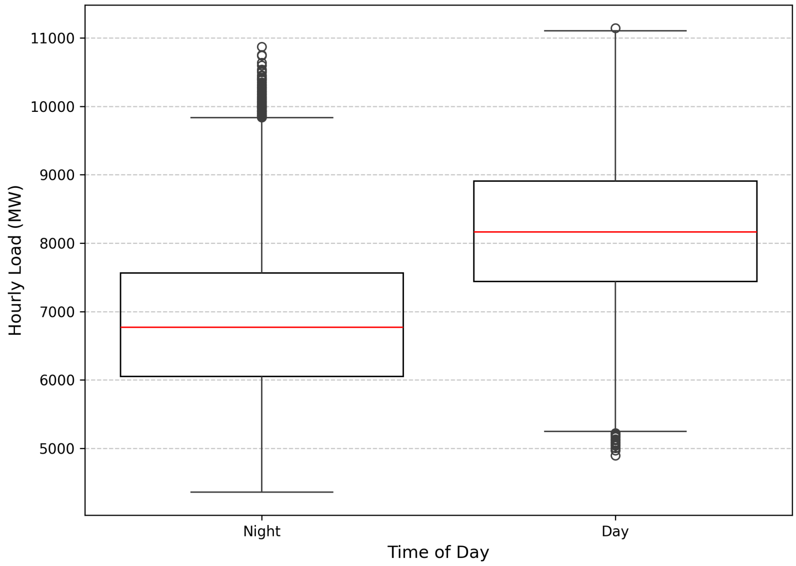
Source: This figure was compiled by the author.

2.0.2 Observed Load Trends (2016 to 2024)

An analysis of the Czech Republic's monthly average consumption patterns for the years 2016 to 2024 contextualizes the expected load profile in the absence of COVID-19. From 2016–2019, the load profile was characterized by consistent and predictable seasonality. The observed average monthly load was 7,523 MW, while the observed annual average loads were 7,343 MW in 2016, 7,534 MW in 2017, 7,620 MW in 2018, and 7,554 MW in 2019. The first three years recorded a marginal increase in the average annual load, followed by a slight decrease at the end of 2019. Monthly patterns consistently showed peak demand in winter, reaching 9,177 MW in January 2017, and low demand in summer, with a minimum of 6,365 MW recorded in July 2017.

In addition to the seasonal pattern, a temporal analysis shows that daytime loads were 8,158.38 MW compared to 6,874.56 MW at night, a statistically significant difference of 1,283.82 MW ($p < 0.001$). The hourly variation was characterized by standard deviations of 1105.13 MW for day and 1089.08 MW for night (difference of approximately 16 MW), suggesting that the hour-to-hour changes were very similar for day and night periods.

Figure 2.2: Pre-Pandemic Day and Night Variations



Source: This figure was compiled by the author.

The subsequent period, from 2020 to 2024, shows a slight but notable change in the load profile coinciding with the outbreak of the pandemic. The initial cases were confirmed in the Czech Republic in March 2020. The year 2020 observed an overall decline in the average monthly load, which fell to 7,247 MW. This reduction represents a 3.66% change compared to the 2016–2019 average. Furthermore, the Russia–Ukraine war broke out in late February 2022, which also coincided with the change in the load profile. On average, the annual loads observed were 7,319 MW in 2020, 7,607 MW in 2021, 7,355 MW in 2022, 6,955 MW in 2023, and 7,076 MW in 2024. During the 2020–2024 period, the annual average load initially increased from 2020 to 2021, then experienced a decline through 2023. This was followed by a slight increase in 2024, suggesting a recovery. The highest monthly average load observed after the pandemic was 8,882 MW in February 2021, and the lowest was 4,996 MW in July 2024. This fluctuating annual trend suggests a change in the country's electricity consumption patterns to which the response measures to the COVID-19 pandemic likely played a role.

The observed deviations from previous load trends are in contrast to the expectations outlined in the MAF published by ČEPS (2016). In their modeling, ČEPS assumed that net consumption would continue to rise in tandem with GDP, forecasting an increase to 65.5 TWh by 2020 and 67.0 TWh by 2025 (p. 8). The projections featured Scenario A and B. Scenario A reflected unadjusted correlations between GDP and electricity demand. Scenario B, on the other hand, incorporated additional factors such as energy efficiency improvements and the development of electric vehicles. Both scenarios were grounded in the assumption that electricity demand would remain closely coupled with macroeconomic activity, especially consumption for economic recovery following the 2008–2009 economic recession. Notably, ČEPS had already observed a rebound in gross consumption as early as 2014–2015, adjusted for temperature, and interpreted this as evidence of resumed coupling between energy use and economic growth (p. 9). This provides additional reasoning that the electricity load profile was expected to continue its upward trend or plateau at the very least. However, the data from 2020 to 2024 reveals a different pattern and electricity demand did not return to its pre-pandemic trajectory. Instead, average annual loads declined, particularly in 2022 and 2023, suggesting broader structural changes.

Chapter 3

Literature Review

The non-linear nature of the variables that influence demand makes load forecasting a complex task. For example, humidity amplifies temperature's effect, while precipitation conditions can change the ambient temperature. Coupling this with the categorical variable "day-of-the-week", interaction effects add to the non-linearity. For instance, the demand response to a specific high temperature on a workday can differ from the response on a public holiday or a weekend, even if the temperature is the same.

In general, weather variables are the most influential determinants of electricity demand, with temperature exhibiting a strong relationship with demand. In the Czech Republic, electricity spikes due to household air conditioning are minimal, with its share in household electricity consumption being only 1% in 2022, according to an ODYSSEE-MURE (2025) report. In cold weather, heating demand increases more gradually as temperatures fall below comfortable levels (Liu et al., 2021). Beyond temperature and humidity, other variables such as cloud cover and the type of precipitation play a role (Behm et al., 2020). Cloud cover can increase the need for artificial lighting during the day. On the other hand, the precipitation type impacts cooling needs. Snow conditions reduce temperature even further, which in turn increases heating needs. Depending on the forecast horizon, these non-linear features influence the appropriate forecast model. Short-term predictions typically focus on the next hour to a few days, usually 48 hours. Medium-term forecasts focus on up to 12 months, while long-term forecasts apply to anything 13 months and beyond (Behm et al., 2020). Traditionally, forecasting methods heavily relied on statistical techniques such as regression models and time-series methods before sophisticated machine learning tools rose to prominence. The scope of this literature review is to uncover the best medium- to long-term forecasting method for this counterfactual analysis and to understand how other studies measured the impact of COVID-19 on electricity demand in general.

3.1 Electricity Load Forecasting Approaches

McGrath and Jonker (2024) explain the process for load forecasting, and emphasize that the nature of the data and forecast horizon influences the model choice. For a given model, performance can be measured using either the root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and R^2 error (R^2). The RMSE captures the average magnitude of prediction errors, with larger errors penalized more heavily, making it sensitive to outliers. The MAE calculates the average absolute difference between predicted and actual values, while the MAPE expresses prediction error as a percentage of the actual values. This makes it especially useful when comparing performance across different scales. Lastly, the R^2 error measures how well a model captures the variance in the data.

A model is considered to perform well when error metrics such as MAE, RMSE, and MAPE are close to zero, indicating minimal difference between predicted and actual values, while the R^2 value is close to one. RMSE is best used to explain the average magnitude of prediction errors in the original data, while R^2 explains the overall explanatory power of the model and goodness of fit. A model is known to overfit when it learns the training data too well, capturing noise and specific details that do not generalize to new, unseen data.

In addition, another concern is data leakage. For statistical time series models, leakage occurs when future data is used to train a model, while for machine learning, data not available during real-world use is accidentally included in model training. A clear separation between training and testing datasets must be enforced and all data transformations, including scaling, imputation, feature selection, and hyperparameter tuning, must be applied exclusively to the training data and then propagated to the test set. Leakages result in models that show superior performance in tests, but have poor performance in real forecasting tasks. It is essential to prevent leakage to have accurate model evaluation metrics that reflect true forecasting performance. In this study, every step of the pipeline was designed to prevent leakage and mimic the conditions under which the model would be used in practice. Scaling is particularly important, and warrants a more detailed discussion which is presented.

3.1.1 Scaling and Hyperparameter Tuning

Scaling prevents feature dominance, encourages fair contribution and improves performance while enabling quicker model convergence. Scaling does this by transforming features to a common numerical range or distribution. This process ensures that features with larger numerical magnitudes do not disproportionately influence the model's learning process (Ahsan et al., 2021). There are several methods for data scaling, and the prominent ones are described below:

- **Standardization or Z-score Normalization:** Usually performed using a 'StandardScaler' method, this technique transforms features to have a mean of zero and a standard deviation of one. For each feature, the original value (x) is transformed into a new value (x') using its mean (μ) and standard deviation (σ). The mean and standard deviation are calculated from the training data:

$$x' = \frac{x - \mu}{\sigma}$$

This method works best for data that follows a Gaussian distribution (bell curve) or when algorithms assume zero-centered data (Ahsan et al., 2021).

- **Normalization or Min-Max Scaling:** This method scales features to a fixed range between 0 and 1. The formula for Min-Max scaling is:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

where $\min(x)$ and $\max(x)$ are the minimum and maximum values of the feature, respectively. This approach is beneficial when algorithms require input features to be within a specific bounded range (Ahsan et al., 2021).

- **Robust Scaling:** Contrary to the standardization and normalization scaling methods, robust scaling handles outliers more effectively by using the interquartile range (IQR) instead of the mean and standard deviation. It scales features using the median and the IQR, making it less susceptible to the influence of extreme values (Ahsan et al., 2021).

The choice of scaling method depends on the nature of the data and the requirements of the chosen machine learning algorithm. Scaling must be consistent scaling to ensure fair contribution of all features to the model's performance. Furthermore, the architecture of the model is critical for machine learning models. The architecture is defined by the parameters and hyperparameters Model parameters are internal variables that the model learns directly from the training data. They represent patterns discovered within the dataset. As an example, for an ANN, the weights and biases connecting neurons are the model parameters. These are adjusted during training. On the other hand, hyperparameters are external configuration settings not learned from data. Rather, they are set manually before training begins, although it is possible to optimize them automatically after the initial setting (Probst et al., 2019). The learning rate is an example of a hyperparameter specification, which influences how fast a model learns and its ultimate performance.

Hyperparameter tuning is applied to the external settings. This optimization process involves systematically testing different combinations of hyperparameter settings (e.g., the number of decision trees in XGBoost, or how quickly it learns from errors) to find the configuration that yields the most accurate and generalizable forecasts. Techniques such as randomized search and cross-validation are commonly employed to explore this vast space of possibilities efficiently and robustly evaluate performance on unseen data, ultimately minimizing errors like RMSE and maximizing explanatory power (R^2). This comprehensive approach to architecture design and hyperparameter optimization is the backbone of high-performing and reliable forecasting models.

3.1.2 Forecast Horizon

For short-term forecasts, the traditional time series models, particularly the Auto Regressive Integrated Moving Average with Exogenous (ARIMA), have shown strong performance in forecasting demand. Řanda (2023) combined Seasonal Auto Regressive Integrated Moving Average with Exogenous (SARIMAX) and Long Short-Term Memory (LSTM)-based Recurrent Neural Networks (RNNs) to forecast short-term electricity load for the Czech Republic. The hybrid model leveraged econometric time-series techniques to address seasonality and external factors, particularly weather and prices, while the LSTM component captured non-linear dependencies and long-term patterns. Compared to standalone SARIMAX, RNN, and bagged regression trees, the hybrid model was superior and outperformed official forecasts provided by ČEPS. Even though the study was conducted after the pandemic, the author did not perform a quantitative analysis of how the pandemic affected short-term demand.

For long-term predictions, mainly linear regression models have been in vogue, but the emergence of AI and machine learning models, particularly deep learning models, are changing the landscape (McGrath and Jonker, 2024). Support Vector Machines (SVM) and neural networks are useful when modelling complex non-linear relationships which are characteristic of load-weather variable data. Behm et al. (2020) present one of the first approaches to model weather-dependent electricity load profiles using only weather and calendrical variables. Their model was a deep neural network, which they trained on German weather and load data. The authors set up the model with 5 hidden layers and 1,024 hidden nodes per layer, making their architecture a deep neural network. To justify choosing this machine learning method, the authors assert that ENTSO-E relies on regression-based models, whose forecasts are presented in their Mid-Term Adequacy Forecasts (MAF). Their ANN model achieved a MAPE of 2.8% when predicting the 2016 German load profile, compared to the MAF forecast presented by ENTSO-E, which had a MAPE of 4.8%. To prove that the specified model could be used to forecast load for other European countries, they applied the same methodology to model the load for France, Spain, and Sweden using the respective countries' data. Essentially, the model learned the patterns from the German data to determine the internal architecture. The same ANN model featuring five hidden layers featuring 1,024 nodes each was then used to generate the profiles for France, Spain and Sweden. For validation, they compared predicted hourly load values for 2016 against actual data. Spain's MAPE was similar to Germany at 2.8% while the results for Sweden and France showed MAPEs of 3.4% and 3.2% respectively. These results led to their conclusion that their specific ANN model architecture generalizes well across different European climate zones based on improvements to ENTSO-E's MAF.

While Behm et al. (2020) offer a promising ANN-based alternative to non-machine learning approaches and machine learning approaches that rely on historical load values, their comparison with ENTSO-E's MAF is not based on access to ENTSO-E's internal models. Instead, it relies on a reconstructed proxy. To build the proxy, the authors extracted temperature-dependent cubic polynomial functions from published MAF datasets and followed the methodological steps outlined in ENTSO-E's publicly available documentation. Afterward, they validated their replication by comparing it to the original MAF load profiles for 1984 and 2007. Their replication matched the published data closely, with a reported MAPE of around 1.6% and R^2 above 0.98. This lends credibility to their approximation of ENTSO-E's approach, but is not a complete account of any unpublished procedures and parameter adjustments ENTSO-E could have applied. As such, while the ANN model demonstrates superior performance relative to this approximate benchmark, the strength of the comparison is limited by the absence of direct validation or collaboration with ENTSO-E.

Furthermore, Baur et al. (2024) point out that complex deep learning models such as the one used by Behm et al. (2020) can act as "black boxes" making their internal decision-making difficult to interpret. Traditional statistical models provide clear and interpretable relationships between input variables and outputs. On the other hand, deep neural networks learn complex feature representations internally. Even though there are methods such as feature importance or saliency maps to peek inside, the explanations are approximate and may not always fully demystify a model's decision-making process. This complexity makes it challenging to trace how specific inputs influence the final predictions. Consequently, the claimed improvements by Behm et al. (2020) need to be interpreted carefully.

Jain and Gupta (2024) tested SVMs, RNNs, and LSTMs on five years of hourly Indian load data. LSTM models were the most accurate. Although their study was short-term, the authors suggest that these architectures are scalable to longer horizons, depending on data quantity and tuning.

In addition to neural networks, tree-based ensemble methods like Random Forests

(RF) and XGBoost perform well. XGBoost is especially strong for complex forecasting. RF builds many independent trees and averages the results. XGBoost builds trees sequentially, correcting prior errors with gradient boosting (Fatima et al., 2023). It is faster, handles large datasets well, and tolerates missing data.

For counterfactual analysis, model interpretability and predictor importance are critical. Feature selection is essential to prioritize key variables. While stepwise or subset selection can be unstable, LASSO (Least Absolute Shrinkage and Selection Operator) selects variables and regularizes models simultaneously. It shrinks some coefficients to zero, reducing overfitting and enhancing interpretability (Tibshirani, 1996). As Freijeiro-González et al. (2021) note, LASSO is well-suited to high-dimensional problems. Its advantages and limitations are explored further in the next chapter.

3.1.3 Forecasting and Counterfactual Analyses Post COVID-19

Uğurlu & Jindřichovská (2024) investigated the relationship between electricity consumption, trade, and GDP, and the effect of the pandemic for the Czech Republic, Hungary, Poland, and Slovakia. Using a Pooled Mean Group-Autoregressive Distributed Lag (PMG-ARDL) approach designed to capture both short and long-run effects while allowing for heterogeneity, they measured the difference in the total amount of electricty consumed. A Pesaran CD test was used to confirm cross-sectional dependency and their results revealed that COVID-19 measures had a negative and statistically significant short-run effect on electricity consumption in the region. This impact was quantified using coefficients for the COVID-19 dummy variable. In the short run panel model, the coefficient for the COVID-19 dummy variable was -0.015, and varied for individual countries. The coefficients were -0.1076 for the Czech Republic, -0.1357 for Hungary, -0.0842 for Poland, and -0.1088 for Slovakia. These coefficients indicate a decrease in the electricity consumption due to pandemic-related measures. While this study provides valuable insights into the direct impact of COVID-19, it focuses on econometric modeling with dummy variables rather than employing forecasting or constructing a hypothetical "no-COVID" scenario.

Gulati et al. (2021) study the short-term impact of COVID-19 on Haryana's load profile in India. By analyzing daily load data from January to April 2020, the authors observed a significant drop in electricity demand following the lockdown. Industrial consumption was impacted the most compared to residential areas. As an example, the industrial hub of Faridabad reduced from around 900 MW to slightly over 500 MW. The study employed several conventional machine learning methods (linear regression, support vector regression, decision tree regression, and random forest) and an artificial neural network (ANN). Their results show that the ANN was more accurate as it achieved a lower MSE compared to the other models. As an example, the ANN's forecast error was 354.28 MSE versus 516.55 MSE for the textile region of Ambala. Based on this, the ANN captured non-linear demand patterns better than conventional machine learning methods.

Jinran Wu et al. (2023) conducted a study evaluating the impact of COVID-19 lockdowns on electricity demand in Victoria, Australia. Their analysis utilized a time series forecasting model incorporating predictors for pandemic impact and lockdown periods. Their results show that the total demand in Victoria decreased by 3.0% in 2020 compared to 2019, and by 1.81% from January to July 2021 relative to the same period in 2019. Specifically, the average half-hourly demand saw a reduction of 210.55 MW due to lockdowns. The study found that during lockdown periods, the lowest points of electricity demand (around 4:00 AM and 2:00 PM) became much lower. However, the highest demand times (around 9:30 AM and 6:30 PM) stayed about the same as non-lockdown periods, with the average demand being 224.92 MW higher at these peak times during lockdown. Weekend demand showed little difference, but weekdays, particularly Monday, Thursday, and Friday, had lower demand during lockdowns.

In another study, Abulibdeh et al. (2022) examined the impact of the COVID-19 pandemic on electricity consumption and forecasting accuracy in Qatar. They trained three different models — namely SVM, XGBoost, and RF — using historical electricity consumption data from 2010–2019. The trained models were then used to simulate what the electricity consumption would have been in 2020 and 2021 if the COVID-19 pandemic had not occurred. This simulation relied on historical load data and was based on the patterns learned from the actual pre-pandemic data. Their findings indicate that lockdowns led to decreased consumption in commercial and industrial sectors. Increased consumption was noted in the residential, governmental, and agricultural sectors, suggesting that the demand shifted to other sectors of the economy.

Feras Alasali et al. (2021) examined the impact of the COVID-19 pandemic on electricity demand and load forecasting accuracy in Jordan, utilizing five years of half-hourly smart meter data. Their study's primary tool was a proposed rolling stochastic ARIMAX model, selected for its ability to handle the non-smooth nature of demand during disruptions and to generate future demand scenarios through a probabilistic approach, thereby aiming to improve forecast performance. This model was benchmarked against a standard ARIMAX model and an Artificial Neural Network (ANN), with the proposed model demonstrating superior performance by reducing forecast error by up to 23.7%. To specifically quantify the pandemic's impact, the researchers performed a counterfactual analysis by forecasting electricity demand for 2020 as if the pandemic had not occurred, and then compared these forecasts with the actual consumption data.

This literature review shows how important advanced machine learning models like Artificial Neural Networks (ANNs), XGBoost, and LASSO regression are for dealing with the complex job of predicting electricity demand, especially because weather and calendar details have non-linear effects (Behm et al., 2020; Jain & Gupta, 2024; Fatima et al., 2023). ANNs have been good at understanding these complex relationships and working well even during unexpected events (Gulati et al., 2021). XGBoost is known for being accurate and efficient and can handle many different types of data. It builds trees sequentially, where each new tree is designed to learn from the mistakes made in the previous step, and corrects them thus reducing forecast errors (Fatima et al., 2023). LASSO selects the most important factors and simplifies models, which helps make predictions clear and easy to understand. This is key for knowing what caused changes in 'what if' situations (Baur et al., 2024; Tibshirani, 1996; Freijeiro-González et al., 2021).

Similar to the study by Abulibdeh et al. (2022) exploring the pandemic's effects on the load profile, this research also measures changes in electricity demand. However, a main difference in method is using Behm et al. (2020)'s approach, which creates electricity demand patterns based on observed hourly weather and calendric information, instead of using past hourly values of electricity demand. The next chapter explores the forecasting methods identified in the literature more closely.

Chapter 4

Understanding the Core Models

The preceding literature review identified the use of Artificial Neural Networks (ANNs), XGBoost, and LASSO regression for handling the complex factors that influence electricity demand. This chapter now explains in detail how each of the models works to better understand why they are well-suited for electricity load forecasting.

4.1 LASSO Regression

As discussed in the previous chapter, LASSO is a powerful and popular type of linear regression that can be applied to create models that are easier to understand and more efficient, especially when there are many possible input variables (Tibshirani, 1996). LASSO attempts to find the best straight line or hyperplane that minimizes the overall difference between the predicted and realized data points. LASSO adds a penalty to this process, based on the total absolute size of the model's coefficients (the numbers that show how much each input variable affects the output). This is known as L1 regularization. The objective of the LASSO method is to make the sum of squared errors small, while also adding this penalty related to the size of the coefficients. This penalty encourages the model to make the coefficients of some less important input variables exactly zero. When a variable's coefficient becomes zero, that variable is effectively removed from the model. This automatic way of selecting features is a major benefit of LASSO. It helps to simplify the model by getting rid of unnecessary variables, leading to a clearer and more concise model (Freijeiro-González et al., 2021).

With fewer variables, it becomes much easier to see which factors are truly driving the predictions, making the model more understandable. By reducing the number of variables and making coefficients smaller, LASSO also helps the model avoid being too sensitive to minor details in the training data. This ensures it performs better on new data it hasn't seen before, which helps prevent a problem called overfitting. LASSO's ability to select features automatically is very valuable in datasets with many features where only a few might actually be important. Its improved clarity is a big advantage, as simpler models are easier for people to understand. It also works well in situations where there are many possible input variables (Baur et al., 2024).

The objective function for LASSO regression is as follows:

$$\min_{\beta_0,\beta} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - (\beta_0 + \mathbf{x}_i^T \beta))^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \tag{4.1}$$

In this equation, n is the number of data points, and p is the number of features. y_i stands for the actual value for the i-th data point, and $\mathbf{x}_i$ is the set of features for that data point. β_0 is the intercept, and β is the set of coefficients for the features. The term λ (lambda, which is zero or greater) controls how strong the L1 penalty is. A larger λ means more coefficients will be set to zero.

Even though LASSO streamlines machine learning models through feature selection well, it does have some drawbacks which are important to note when the goal is to perfectly identify every underlying feature. One example is when the dataset contains features that are too similar. LASSO has the tendency to choose only one from the feature group. In addition, Freijeiro-González et al. (2021) note that in very complex scenarios with numerous features, it might occasionally include a "noisy" feature that is not truly important. One way to mitigate this is a thresholded LASSO (Zhou, 2010). This method aims to select a sufficiently small set of important features without sacrificing the accuracy of the model. The result is close to ideal, essentially getting close to the ideal result one would achieve if they already knew which features were truly important.

4.2 Artificial Neural Networks (ANNs)

An ANN is made up of many connected processing units, often called "neurons" or "nodes," which are organized into distinct layers. A single neuron is the basic building block of an ANN. For any neuron j, its operation involves two main steps. First, it calculates a weighted sum of its inputs from the previous layer, which is called z_j.

$$z_j = \sum_{k=1}^{m} w_{jk} a_k + b_j \tag{4.2}$$

Where m is the number of inputs to neuron j, w_{jk} are the weights (numbers that are adjusted during training) that connect input k to neuron j, a_k is the activation (output) from input k in the previous layer, and b_j is the bias term for neuron j.

Second, this weighted sum z_j then goes through a non-linear activation function, $f(\cdot)$, to create the neuron's output, a_j.

The network starts with an input layer, which is a matrix of all variables that affect the load profile. Each neuron in this layer represents one input variable, such as temperature, humidity, hour of the day, or day of the week. These neurons simply pass the input values to the next layer.

After the input layer, there are one or more hidden layers. Although calculations can be made without hidden layers, hidden layers carry out more complex computations to learn intricate non-linear relationships in the data. If the ANN architecture includes hidden layers, each hidden layer neuron's input is the output from the previous layer. The inputs are multiplied by weights representing how strong the connection between the neurons is, before they are added together along with a bias value. The combined sum is subsequently processed by an activation function, to result in an output.

Rectified linear unit (ReLU) and the sigmoid function are common activation functions used to introduce non-linearity, mimicking the complex relationships found in load forecasting problems (Behm et al., 2020). Without activation functions, even a network with many layers would act like a single simple linear model affecting the quality of the final predictions produced by the output layer.

ANNs learn through training by adjusting weights and biases to minimize prediction errors. This involves a forward pass to generate predictions, error calculation to quantify the difference between predictions and actual values, backpropagation to distribute the error back through the network, and weight adjustment using optimization methods like gradient descent. Repeating this process over many epochs allows the ANN to learn and improve its accuracy in identifying patterns and relationships within the data.

4.3 Extreme Gradient Boosting (XGBoost)

Fatima *et al.* 2023 describe how the algorithm works. Essentially, XGBoost combines many simple decision trees. The framework is precise and highly adaptive to most types of data. The decision trees are sequentially created, and each new tree corrects the mistakes of the ones built before it. This stacking improves the predictive power of the model. Learning the patterns begins with an initial prediction for all data points, usually the mean. From there, the algorithm calculates the errors, which are the differences between the actual values and the current predictions. Importantly, instead of trying to predict the original target value directly, the next decision tree is specifically built to predict these errors, also known as residuals. This new tree's role is to fix the inaccuracies made by the previous model or prediction.

The predictions from this new error-predicting tree are then added to the previous overall prediction, gradually making the combined model more accurate. This process repeats: in each step, a new tree is added that works to correct the remaining errors from all the trees built so far. The term "gradient" refers to the use of mathematical optimization methods, similar to those for finding the lowest point on a graph, to determine the best way to reduce these errors.

The general objective function that XGBoost minimizes at each step t combines a function measuring how well the model fits the data with terms that control the model's complexity and help prevent it from overfitting:

$$\text{Obj}^{(t)} = \sum_{i=1}^{n} L(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)) + \sum_{k=1}^{t} \Omega(f_k)$$

L is a function that measures the difference between the actual value y_i and the prediction $\hat{y}_i^{(t)}$. $\hat{y}_i^{(t-1)}$ is the prediction from the model in the previous $t-1$ steps, and $f_t(\mathbf{x}_i)$ is the new tree (or simple model) added at step t for data point $\mathbf{x}_i$. $\Omega(f_k)$ is a regularization term for the k-th tree, which discourages the tree from becoming too complicated. A common form for $\Omega(f_k)$ is: The general objective function that XGBoost minimizes at each step t combines a function measuring how well the model fits the data with terms that control the model's complexity and help prevent it from overfitting:

$$\text{Obj}^{(t)} = \sum_{i=1}^{n} L(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)) + \sum_{k=1}^{t} \Omega(f_k) \tag{4.3}$$

L is a function that measures the difference between the actual value y_i and the prediction $\hat{y}_i^{(t)}$. $\hat{y}_i^{(t-1)}$ is the prediction from the model in the previous $t-1$ steps, and $f_t(\mathbf{x}_i)$ is the new tree (or simple model) added at step t for data point $\mathbf{x}_i$. $\Omega(f_k)$ is a regularization term for the k-th tree, which discourages the tree from becoming too complicated. A common form for $\Omega(f_k)$ is:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{4.4}$$

Chapter 5

Methodology

5.1 Theoretical Framework: Approaches to Counterfactual Analysis

This section reviews empirical strategies used to construct counterfactual electricity load profiles and discusses the assumptions and limitations. The Difference-in-Differences method is popular for counterfactual analysis, and would be used to compare load changes over time between a 'treated group and an untreated group. While reliable, the global nature of the pandemic makes it impossible to have the two groups, as the measures were applied to the whole country. Another method is the Synthetic Control Method (SCM), which would involve constructing a "synthetic Czech Republic" from a weighted combination of other countries' pre-pandemic load data to serve as the counterfactual. This is a useful method, but the major challenge is replicating granular hourly load values which can be challenging (Chen, 2023). Last but not least, Computable General Equilibrium (CGE) models have the ability to simulate economy-wide impacts of shocks and policies influencing energy demand at a macro level (Jia & Lin, 2022). However, one main challenge is that their aggregate nature means they are unable to directly model granular hourly electricity load. In this case, matching methods, which aim to balance covariates between treated and control groups would be impractical for widespread national shocks due to pervasive "treatment" and unobservable confounders. Given these limitations, this thesis naturally employs a generative, machine learning-based approach. This framework is supported by Kirilenko *et al.* 2024 who advocate for their suitability in complex problems, where other methods fail or require too much computation.

5.2 Data Description

Hourly electricity load data was sourced from the Czech Transmission System Operator, ČEPS. The data was available in sets, which had to be merged to create one complete dataset for the load. Weather data was sourced from Visual Crossing Weather and was available as a full dataset. To handle missing values, interpolation was applied to the datasets. Missing values for the variables `solarradiation`, `uvindex`, `sealevelpressure`, `snow`, `snowdepth`, and `load` were subsequently handled using a linear interpolation method. This approach was chosen as it is well-suited for environmental and time-series data, effectively filling gaps by estimating values based on a straight line between known data points. This method has been shown to be efficient for predicting missing values **?**.Merging of electricity load and weather datasets was performed manually using Excel, and subsequent pre-processing steps were applied to this merged dataset using Python 13.

[1].

5.2.1 Data Preprocessing Steps

The datetime column was converted to a datetime format. Public holidays in the Czech Republic (for the years 2016–2024) were identified, and their weekday value was changed to 6 (Sunday). Datetime components, specifically month, day, and hour, were extracted into separate columns. Data from January 2016 was removed from the dataset due to either missing weather values or load values for the entire month. The annual peak load was calculated and merged into the dataset. The column with the type of precipitation observed, `preciptype`, was split into separate binary columns, with 0 for no precipitation and 1 otherwise regardless of the type i.e. rain and snow. Following these transformations, some columns were manually dropped to remove non-feature identifiers and redundant variables from the dataset. Examples are `stations` (due to being a weather station identifier and thus not a feature), `severerisk` (weather risks e.g. flood), `windgust`, the original `preciptype` column (as it was replaced by its binary encodings), and `solarenergy` (since its information was incorporated into the `radiation` variable).

[1]The datasets and code (.ipynb files) are available in the GitHub repository: `https://github.com/NyxC33/load-forecasting-cz`. The workflow is split into seven notebooks that must be executed in chronological order: `data_processing`, `lasso_features_and_training`, `ann_with_lasso_features`, `gradient_boost`, `model_tuning`, `counterfactual_analysis`, and `Heristic_Counterfactual_Analysis`. These notebooks cover data preparation, modeling, evaluation, and both standard and heuristic counterfactual analyses.

5.2.2 Features and Target Definition

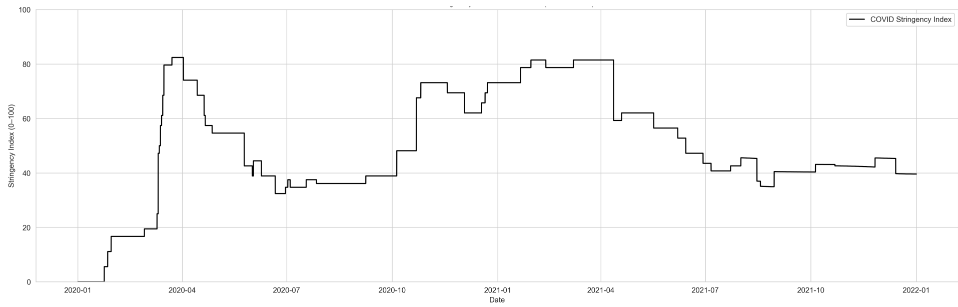
month	Month of the year. Unit: Integer (1–12).
day	Day of the month. Unit: Integer (1–31).
hour	Hour of the day. Unit: Integer (0–23).
weekday	Day of the week. Public holidays are marked as Sunday (6). Unit: Integer (0–6, where 0 is Monday).
annual_peak	Maximum load observed for a given year. Unit: MW.
temp	Average temperature, ranging from −16.7 to 37.3°C.
humidity	Relative humidity, ranging from 11.19% to 100%.
feels_like	Apparent temperature, ranging from −20.7 to 36.1°C.
dew	Dew point temperature, ranging from −19.9 to 23.3°C.
wind_speed	Wind speed. Unit: m/s.
wind_dir	Wind direction (0–360°).
solar_radiation	Solar radiation (unit not specified).
cloud_cover	Cloud cover (0–100%).
uv_index	UV index (0–10+).
year	Year of the observation.
sea_level_pressure	Atmospheric pressure at sea level.
snow	Amount of snow. Unit not specified.
visibility	Visibility (0 to 68.1 km).
snow_depth	Snow depth in meters.
preciptype_columns	Binary indicators for rain/snow. 1 = precipitation present, 0 = otherwise.
load	Target variable.

The dataset was split by year. The training data includes all processed observations from 2016 to 2018. The test data comprises observations from 2019 to prevent data leakage.

5.2.3 Stringency Index

The stringency index was used to represent the severity of COVID-19, and not directly used as an input feature in the model. It quantifies the strictness of government lockdowns implemented as a response to the pandemic. Sourced from the Oxford Covid-19 Government Response Tracker (OxCGRT), this composite index ranges from 0 to 100. The higher the value, the more stringent the measures. The index creators systematically tracked and scored several closure policies, particularly school and workplace closures, public event cancellations, restrictions on gatherings, public transport closures, stay-at-home requirements, general restrictions on internal movement, and international travel controls. The authors normalized and combined through a weighted average. In my thesis, I utilized the stringency index specifically over the broader Government Response Index (GRI) because its focused scope on restrictive measures allows for a more precise analysis of their direct impact on the Czech Republic's electricity load profile. This ensures that observed changes in electricity demand can be directly attributed to the enforcement of these specific restrictive policies.

Figure 5.1: Stringency Index (2020–2021)



Source: This figure was compiled by the author using data from Hale *et al.* (2021)

5.3 Model Development

Table 5.1 below summarizes the modeling approaches tested in this study. The performance metrics are presented along with the relevant implementation notes. These initial results guided the model selection as described in the discussion that follows.

Table 5.1: Overview of Model Methods and Performance

Framework	**R^2 / RMSE (MW)**	**Comment**
LASSO	0.453 / 926.1	Used LassoCV with 5-fold CV; ≥10 MW coefficient threshold for selection.
ANN	-4.204 / 2855.7	3-layer ANN (100-50-25); negative test R^2 suggests indicates that the model is worse than just predicting mean. Can be poor model set up. Overfitting likely happening.
ANN_LASSO	0.746 / 631.2	Same architecture as above; trained on LASSO-selected features; alpha tuned via GridSearchCV.
XGB	0.924 / 344.4	GridSearchCV tuning; used all available features.
XGB_LASSO	0.929 / 334.3	Final model; trained on LASSO-selected features; optimized via RandomizedSearchCV.

Behm et al. (2020) justify selecting an ANN to forecast load due to its ability to represent the same “input-output” relationships as common regression models such as the one the authors purport ENSTOE use. Moreover, the authors were motivated by a need to transfer promising machine learning methods to electricity load forecasting, partly influencing their model choice. Instead of selecting a single model directly, in this study, I use an iterative approach to select the best model from LASSO regression, ANN, and XGBoost based on the review of the literature.

LASSO played a central dual role in my modeling approach, and was first used as a standalone model before being used for feature selection for both the ANN and XGBoost models. Leveraging LASSO to identify and use the most influential variables as features reduces model dimensionality to enhance performance. Results by Řanda (2023) validate the systematic review by Nti et al. (2020) that a hybrid model enhances accuracy when forecasting load. This is justification for my approach to iteratively test a combination of LASSO with ANN and LASSO with XGBoost. For a hybrid model, LASSO functions only to select the features that are important to forecast load. Each iteration featured a central tracker to identify the overall best-performing model across the entire workflow. This ensured a structured and reproducible method for selecting the final candidate model for forecasting.

To select the best performing model, I will score them by how well they can predict 2019, the year before COVID-19. Once the best-performing model has been identified, I use the best-scoring model to create the counterfactual scenario.

5.3.1 LASSO Model: Technical Specifications

The LASSO feature selection process began by loading training data from 2016 to 2018. All potential features were identified and then scaled using a `StandardScaler`. This was necessary to ensure equal contributions to the L1 penalty by each feature when training the model. As a result, variables with larger scales were prevented from dominating the regularization process for the model and for feature selection in the subsequent steps.

The model was trained on the scaled training data, employing 5-fold cross-validation to select the optimal alpha parameter from a logarithmically spaced range of 100 values between 0.000001 and 1000. Five-fold cross-validation was used to balance bias and variance in estimating the optimal regularization parameter without excessive computational cost (Tibshirani, 1996). A wide alpha range specification allowed the model to explore both light and strong penalization to accommodate possible variations in feature relevance and multicollinearity. Features were selected if the absolute value of their coefficient in the trained LASSO model exceeded a minimum impact threshold of 10 MW, representing roughly 0.001% of the annual peak load. This additional thresholding step filtered out features with negligible real-world influence, promoting interpretability and focusing the model on drivers with operational significance. To interpret the coefficient, a rescaler was applied and the coefficients were grouped into temporal and non-temporal categories to better contextualize their effects.

Table 5.2: Quantitative Impact of Selected Features (Above 10 MW)

Feature	**Impact (MW per unit)**
temp	-4.68 per °C
feelslike	-67.26 per °C
dew	-11.93 per °C
windspeed	+1.80 per km/h
winddir	-0.56 per °
solarradiation	+2.55 per W/m^2
cloudcover	+3.69 per %
snow	+193.11 per cm
visibility	-1.87 per km
snowdepth	+73.14 per cm
preciptype_rain	+13.18 when raining

[1] LASSO selected 16 out of 21 available features, including both meteorological and temporal indicators. Temporal variables like `month`, `day`, `hour`, `weekday`, and `year` were also included but are not shown here due to their categorical nature. The same selected features were identified and used in the downstream ANN and XGBoost models.

Subsequently, the test set from 2019 was prepared using the same features and scaling transformation. The model's performance was evaluated on this test set using only the selected features, calculating the Root Mean Squared Error and R-squared. Finally, the list of selected features and the trained model, including the fitted scaler, selected coefficients, and performance metrics were saved for later comparisons provided that the test R-squared exceeded 0.6 with at least one feature selected.

5.3.2 The Artificial Neural Network

The first ANN was trained using all available features, excluding the manually dropped variables which were not relevant such as weather stations' identification numbers. After splitting the data into training (2016–2018) and testing (2019) sets, all features were scaled using a `StandardScaler`. This transformation of the features set the mean to 0 and the standard deviation to 1 to help with convergence and improve stability of the model. ANNs can adjust for varying feature scales internally, but external scaling accelerates convergence and improves stability during training.

The ANN was configured as a Multi-Layer Perceptron Regressor (MLP), similar to Behm et al. (2020). While inspired by Behm et al.'s deep MLP architecture for national load forecasting, a more computationally feasible design of three hidden layers with 100, 50, and 25 neurons was used to effectively capture complex non-linear relationships. This architecture resembles a pyramid (larger base) and was intended to progressively condense representations to balance model capacity with generalization.

The ReLU activation function was selected over the sigmoid function to mitigate the vanishing gradient problem. Vanishing gradient occurs when gradients become too small to effectively update network weights in the hidden layers, leading to poor learning. In addition, an Adam solver was used to automatically adjust the learning rate. Dynamic learning rates respond to training performance and reduce the need for manual tuning should the ANN be selected as the best performing model. A batch size of 32 was selected to balance computational efficiency and gradient estimation quality. Training was set to a maximum of 1000 iterations and incorporated early stopping, with 20% of the training data reserved for validation and a patience of 50 iterations without improvement. These specifications aimed to prevent overfitting while allowing sufficient learning time.

The model's performance was evaluated by calculating RMSE and R-squared (R^2) on both training and test sets, and an overfitting check was performed. The test R^2 error was negative, suggesting that irrelevant features were included, or poor model architecture. In addition, internal feature importance was assessed based on the average magnitude of the first layer's weights. While not definitive, this internal check can provide a rough idea of feature influence in the model's attempt to improve the test R^2 error.

A second independent training was conducted using input features selected by LASSO regression. The same regression from the LASSO model was applied to the dataset to identify a subset of features based on a minimum impact threshold of 10 MW. These selected features were then independently scaled using a new `StandardScaler`. This second ANN had the same architecture as in the first and was trained on this reduced set of features. The model's R^2 score improved to 0.76 suggesting that the initial model was affected by noisy features which swamped the signal. Keeping the architecture the same allowed for a direct comparison between models trained on the full feature set versus the model trained on LASSO-selected features.

In addition, the regularization parameter, alpha, for this ANN was optimized using a three-fold `GridSearchCV`, searching across a logarithmic range of alpha values. `GridSearchCV` works by defining a grid of hyperparameter values to test and then systematically tries every single possible combination of these values. The regularization aimed to reduce overfitting, especially with smaller feature sets, while tuning alpha ensured the right balance between model complexity and generalization. The performance of this optimized ANN, using the LASSO-selected features, was then evaluated with RMSE and R^2 on both training and test sets. Its results were compared against the previously trained standalone LASSO model's performance and the ANN trained in the first stage.

5.3.3 XGBoost

The XGBoost implementation also followed a two-part approach, similar to the ANN, focusing on training with all features first before training a model with LASSO-selected features.

For the first XGBoost model, all features were scaled using a `StandardScaler` after splitting the data. Although XGBoost is less sensitive to scaling, this was necessary to maintain pipeline uniformity across the models and allow a fair comparison. A basic XGBoost Regressor was created, which is a powerful algorithm for predicting continuous numbers, similar to how an MLP works. This initial setup became the foundation for systematically identifying the best configuration.

The hyperparameters were optimized using a three-fold cross-validation with `GridSearchCV`. This means the dataset designated for hyperparameter tuning was divided into three equal subsets with the model being iteratively trained on two subsets and validated on the last subset. The results from these three iterations were then averaged to provide a robust estimate of performance and reduce overfitting.

The parameters tuned were the number of estimators, learning rate, maximum tree depth, subsample ratio, and column subsample ratio. These parameters were selected because they control model capacity, regularization strength, and overfitting tendencies which are key drivers of performance for gradient-boosted trees. After training the optimal model, its performance was evaluated using RMSE and R^2 on both the training and test sets. Feature importance was also calculated using the model's inherent feature importance scores, which offer insight into which variables contribute most to reducing prediction error across the individual decision trees.

Similar to the ANN, the second training of the XGBoost used features selected by the LASSO model. The selected features were loaded, and new training and test datasets were created using only these features, which were then scaled independently with a `StandardScaler`. The same XGBoost Regressor was again used, and its hyperparameters were optimized with `GridSearchCV`, similar to the first part. However, the hyperparameter ranges were adjusted to account for the smaller input set, leading to a less complex model. The optimal model's performance was then evaluated using RMSE and R^2 on the training and test sets, and feature importance was determined for the selected features.

The two-stage implementation of the XGBoost algorithm was designed to enable fair comparisons while keeping accuracy and interpretability consistent when using all versus LASSO-selected inputs.

5.3.4 Theoretical Framework: Approaches to Counterfactual Analysis

This section reviews empirical strategies used to construct counterfactual electricity load profiles and discusses the assumptions and limitations. The first method is the Difference-in-Differences (DiD) method, which compares load changes over time between a 'treated group and an untreated group. While reliable, the global nature of the pandemic makes it impossible to have the two groups, as the measures were applied to the whole country. Another method is the Synthetic Control Method (SCM), which would involve constructing a "synthetic Czech Republic" from a weighted combination of other countries' pre-pandemic load data to serve as the counterfactual. This is a useful method, but the major challenge is replicating granular hourly load values which can be challenging (Chen, 2023). Last but not least, Computable General Equilibrium (CGE) models have the ability to simulate economy-wide impacts of shocks and policies influencing energy demand at a macro level (Jia & Lin, 2022). However, the challenge is that their aggregate nature means they are unable to directly model granular hourly electricity load. In this case, matching methods, which aim to balance covariates between treated and control groups would be impractical for widespread national shocks due to pervasive "treatment" and unobservable confounders. Naturally, this thesis utilizes a generative, machine learning-based approach.

5.3.5 Model Selection and Tuning

The systematic and iterative model selection process ultimately identified the XGBoost model with LASSO-selected features as the top performer for load forecasting. Before settling on the model, tuning was also performed. Initially, each model type (LASSO, Artificial Neural Network, and XGBoost) underwent separate hyperparameter optimization. For instance, the LASSO model utilized `LassoCV` to find its optimal alpha through 5-fold cross-validation across a wide range of values. Both ANN and XGBoost models, when trained with all features and then with LASSO-selected features, underwent hyperparameter tuning using `GridSearchCV` function. This facilitated a systematic search for the best parameters such as the number of estimators, learning rate, and tree depth for XGBoost, or network architecture and regularization in the case of the ANN. Throughout these training phases, the performance of each model on the test set (measured by R^2 and RMSE) was continuously compared against a running record of the overall best model. This ensured that the best model identified at any point was always tracked.

Finally, the optimized XGBoost model with LASSO-selected features was further subjected to hyperparameter tuning using a `scikit-learn` pipeline. This pipeline integrated the `StandardScaler` with the XGBoost model. `RandomizedSearchCV` was used for this final optimization, exploring a broader set of hyper-parameters than in the initial tuning phases. Table 5.3 below presents the final parameters, while table 5.4 illustrates the hyper-parameter optimization process.

Table 5.3: Final XGBoost Hyper-parameters after Tuning

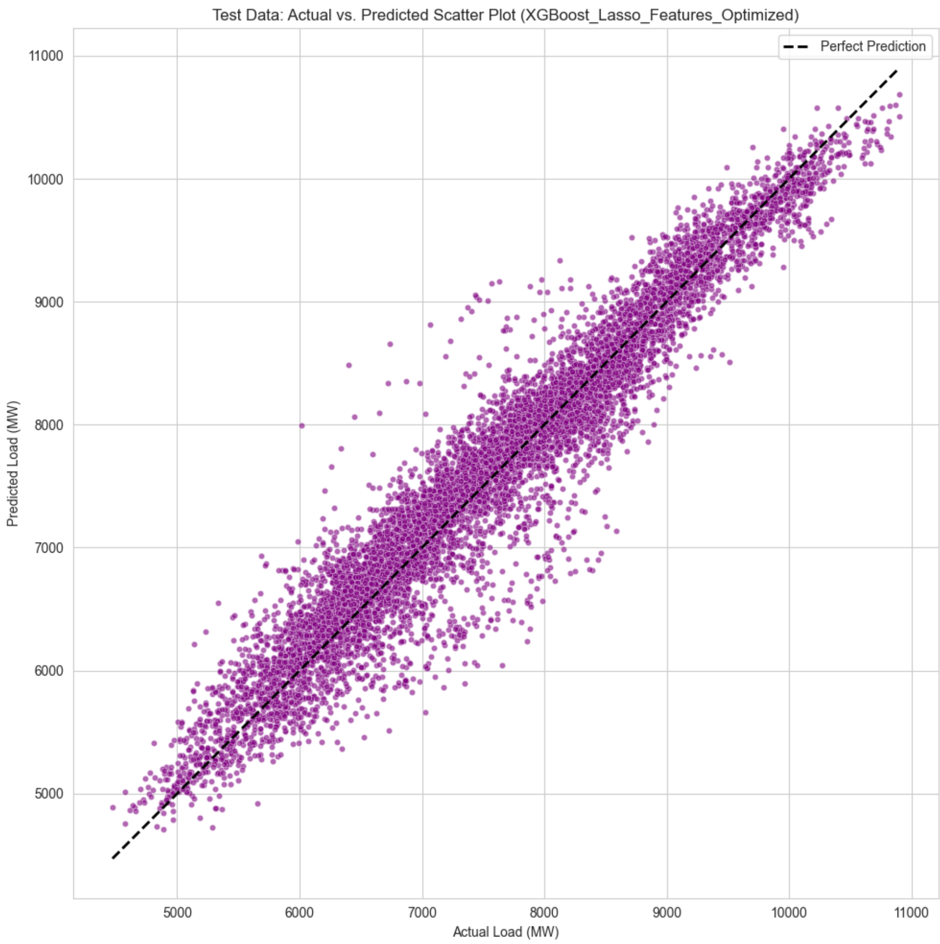
`model__subsample`	0.8
`model__reg_alpha`	0.001
`model__n_estimators`	500
`model__min_child_weight`	5
`model__max_depth`	4
`model__learning_rate`	0.1
`model__gamma`	0.2
`model__colsample_bytree`	0.9

Table 5.4: Python Code for Model Tuning and Evaluation

```
# Performing hyperparameter tuning
tuner = RandomizedSearchCV(
  estimator=pipeline,
  param_distributions=param_distributions,
  n_iter=n_iterations_tuning,
  cv=cv_folds_tuning,
  scoring='neg_mean_squared_error',
  n_jobs=-1,
  verbose=2,
  random_state=42 )
tuner.fit(X_train_full, y_train)
final_pipeline = tuner.best_estimator_

# Evaluating final model performance
train_pred = final_pipeline.predict(X_train_full)
test_pred = final_pipeline.predict(X_test_full)
train_rmse = np.sqrt(mean_squared_error(y_train, train_pred))
test_rmse = np.sqrt(mean_squared_error(y_test, test_pred))
train_r2 = r2_score(y_train, train_pred)
test_r2 = r2_score(y_test, test_pred)
```

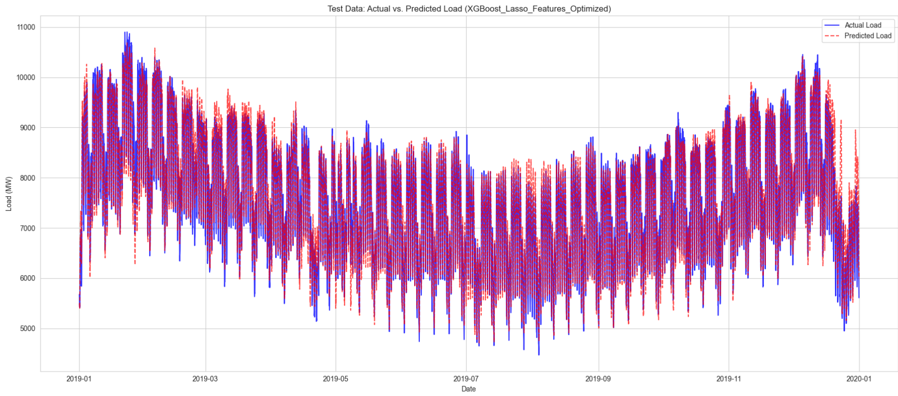
Optimizing `subsample` and `reg_alpha` (L1 regularization) was performed to encourage simpler models by penalizing larger weights and control data sampling, while `n_estimators` and `learning_rate` were tuned to improve the model's overall learning capacity. The parameters `min_child_weight`, `max_depth`, and `gamma` were adjusted to manage individual tree complexity and prevent overfitting. Further generalization was achieved by tuning `colsample_bytree` to control feature sampling. Model tuning re-confirmed the XGBoost model with 16 LASSO selected features as the most robust and accurate model. The final performance metrics showed an R^2 of 0.929 and a RMSE of 334.5 MW on the test set. Despite the extensive final tuning, this performance was consistent with the previous best, indicating the stability and optimal configuration achieved earlier.

Figure 5.2: Actual vs Predicted Scatter Plot



Source: This figure was compiled by the author.

Figure 5.3: Model Performance: Actual Load vs Predicted Load



Source: This figure was compiled by the author.

This chapter outlined the methodology for constructing a counterfactual electricity load profile, including the pre-processing steps, iterative model development, and model tuning. The next chapter presents the results of the counterfactual analysis, showing the estimated COVID-19 impact on the Czech Republic's electricity load. It also provides insights into possible scenarios along with recommendations for energy economics policy.

Chapter 6

Counterfactual Analysis

The counterfactual analysis was performed using the optimized XGBoost model to project the counterfactual load profile for 2020 and 2021. The model with LASSO selected features was provided with the actual observed values for the weather variables. Since this model was trained solely on pre-COVID-19 data, its predictions for 2020 and 2021 directly represent the electricity demand that would have been realized without the pandemic. The code block below illustrates the key steps in the counterfactual analysis. Table 6.1 details the key steps in the counterfactual analysis, while the resulting counterfactual load profile illustrated in Figure 6.1.

Formally, for each hour t in the years 2020 and 2021, let:

- Y_t^{actual}: the observed electricity load,
- X_t: the vector of weather and calendar features used by the model (excluding any COVID-related indicators),
- $f_{\text{XGB}}(\cdot)$: the trained XGBoost model from the final pipeline.

The counterfactual prediction is then:

$$Y_t^{\text{cf}} = f_{\text{XGB}}(X_t) \tag{6.1}$$

The estimated impact of the COVID-19 pandemic on load is calculated as:

$$\Delta_t = Y_t^{\text{actual}} - Y_t^{\text{cf}} \tag{6.2}$$

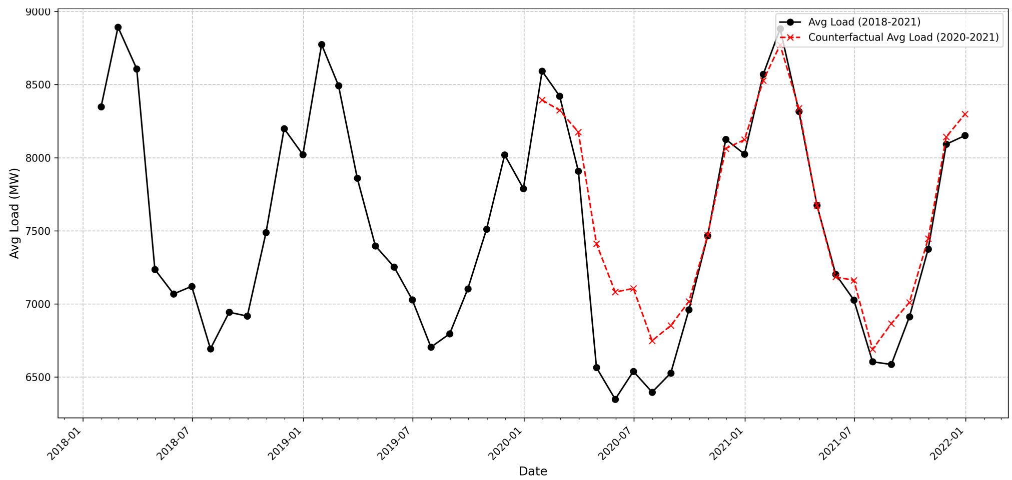
The total and average hourly impact over the analysis window $T \subseteq \{t : \text{year}(t) \in \{2020, 2021\}\}$ are:

$$\text{Total Impact} = \sum_{t \in T} \Delta_t, \qquad \text{Average Hourly Impact} = \frac{1}{|T|} \sum_{t \in T} \Delta_t \tag{6.3}$$

Table 6.1: Python Code for Counterfactual Impact Calculation

```
counterfactual_data = counterfactual_data[
  (counterfactual_data['datetime'].dt.year >= 2020) &
  (counterfactual_data['datetime'].dt.year <= 2021)
  ].copy()
# Prepare features for the counterfactual period
X_counterfactual_full = counterfactual_data[loaded_features]
y_counterfactual_actual = counterfactual_data['load']


# Generate counterfactual predictions (what load would have been
without COVID)
y_counterfactual_pred = final_pipeline.predict(X_counterfactual_full)
# Calculate estimated COVID_19 impact
estimated_covid_impact = y_counterfactual_actual -
y_counterfactual_pred
# Calculate total and average impact
total_impact_mwh = estimated_covid_impact.sum()
average_hourly_impact_mw = estimated_covid_impact.mean()
```

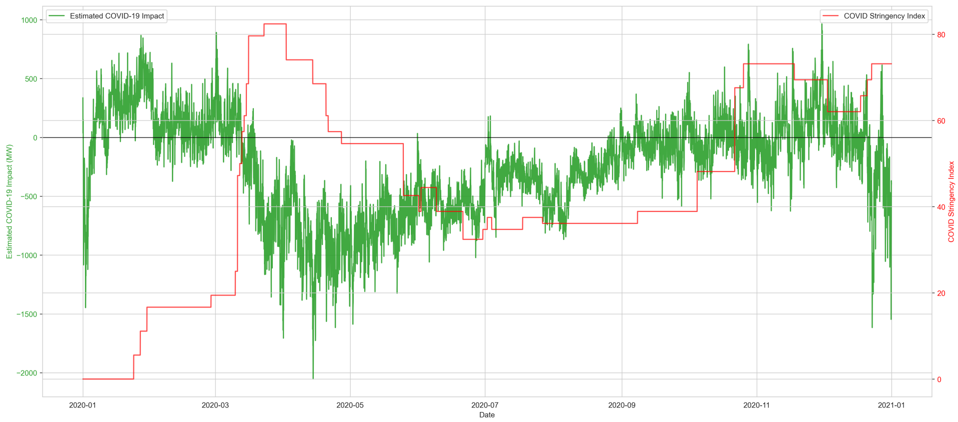
Figure 6.1: Actual and Counterfactual Load Profile



Source: This figure was compiled by the author.

While the stringency index was instrumental to link the observed difference to the severity of the measures, the model did not learn from this index since pre-Covid data was used to train the model. During the study period, the maximum observed stringency was 82.41, while the minimum observed value was 5.56. By the end of 2021, the stringency value had fallen to 39.59. To calculate the "COVID impact" on the load profile, the difference between the actual observed load during the pandemic and the generated counterfactual load was used. The results show that for the entire period (2020 - 2021), electricity demand fell by 2,658,411 MWh (or about 2.66 TWh representing about 270 million euro if the electricity is valued at €100/MWh), corresponding to an average hourly reduction of 151.53 MW. For the year 2020, the total consumption was reduced by 2,123,687.71 MWh with an average hourly reduction of 241.77 MW. On the other hand, for the year 2021 the impact of COVID-19 was less pronounced with a total reduction of 534,723.31 MWh compared to the baseline indicating a recovery. The calculated average hourly reduction for 2021 is approximately 61 MW. The figure below depicts the relationship between stringency and the corresponding changes to the load profile.

Figure 6.2: Load Reductions vs Stringency (2020)

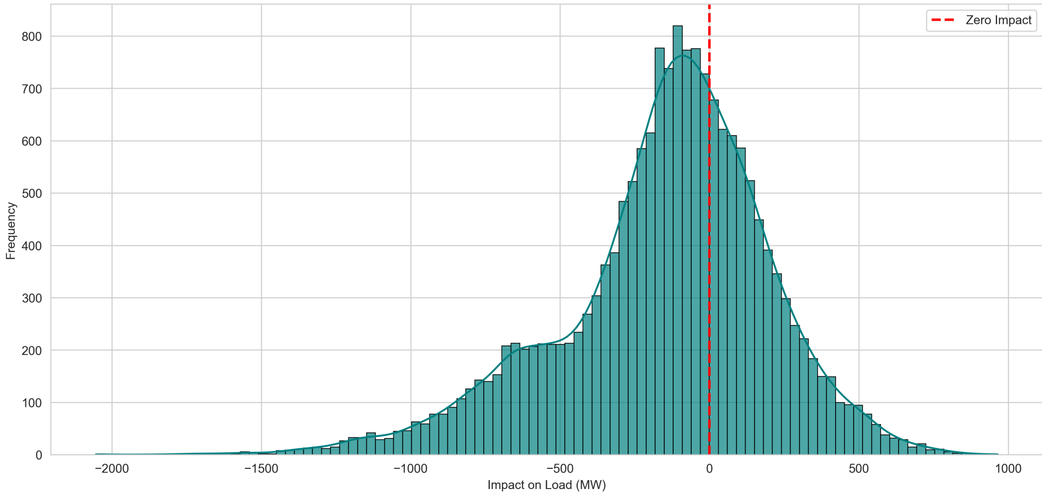


Source: This figure was compiled by the author.

Figure 6.3: Load Reductions vs Stringency (2021)

Source: This figure was compiled by the author.

Figure 6.4: Distribution of Estimated COVID-19 Impact on Electricity Load



Source: This figure was compiled by the author.

6.0.1 Rationale

In the first quarter of 2020, the stringency index rose sharply indicating the increased lockdown measures. This was accompanied by a sharp decline, which peaked at around 850 MW in May 2020 when the stringency was around 0.65. This suggests a direct association between the initial, rapid implementation of strict government interventions and a significant decrease in electricity consumption. Following the initial peak, the stringency declined and then rose again from August 2020 into early 2021. During this period, the estimated load impact exhibited a more consistent but visually inverse relationship. As stringency measures tightened towards late 2020 and early 2021, the estimated load impact generally became more negative, dipping below zero again after a period of less negative impact. Conversely, when stringency eased, the load impact became less negative or moved towards recovery, demonstrating a developing correspondence between the intensity of interventions and the direction of load reduction. Throughout 2021, as the lockdown measures eased to just over 0.4, the estimated load impact reduced to under 200 MW.

Further examination into the temporal dynamics of the change revealed distinct patterns between day and night periods. The average hourly COVID-19 impact during the day was a reduction of 145.68 MW with demand dropping even further during the night by 157.37 MW (T-statistic: 2.1758, P-value: 0.0296). The COVID-19 impact was more pronounced during night hours. A more granular analysis showed average negative impacts of approximately 112 MW for mornings, 179 MW for afternoons, 175 MW for evenings and 149 MW for night time.

6.0.2 A Proposed Heuristic for Stringency Impact Simulation

To illustrate the model's capabilities to simulate how the load profile could have looked under different stringency levels without retraining the model, a heuristic approach has been implemented. This method allows an estimation of the potential impact on electricity load if COVID-19 stringency measures had been consistently lower or higher than observed. The data shows that the highest observed stringency index was 82.41. A factor of 1.21 would represent a stringency of 99.71 which is close to the maximum level of 100. Likewise, a more relaxed scenario is calculated by dividing the observed stringency values by a factor of 1.21 for a balanced analysis. The heuristic operates on the assumption that the observed deviation in electricity load due to COVID-19 is proportionally related to the average stringency level during the period. This is a strong assumption, but is sufficient to illustrate this potential application.

The Heuristic Estimated Impact ($I_{\text{heuristic}}(t)$) at any given time t for a hypothetical, sustained target stringency level (S_{target}) is calculated as:

$$I_{\text{heuristic}}(t) = I_{\text{observed}}(t) \times \frac{S_{\text{target}}}{\bar{S}_{\text{actual}}} \tag{6.4}$$

Where:

- $I_{\text{observed}}(t) = L_{\text{actual}}(t) - L_{\text{no\_covid}}(t)$: The observed deviation of actual load ($L_{\text{actual}}(t)$) from the "no COVID" baseline predicted load ($L_{\text{no\_covid}}(t)$) at time t.
- S_{target}: The specific, sustained target stringency index value for the hypothetical scenario.
- $\bar{S}_{\text{actual}}$: The average actual stringency index observed over the entire counterfactual period (2020–2021).

Once $I_{\text{heuristic}}(t)$ is determined, the Heuristic Hypothetical Load ($L_{\text{heuristic}}(t)$) for that scenario is:

$$L_{\text{heuristic}}(t) = L_{\text{no\_covid}}(t) + I_{\text{heuristic}}(t) \tag{6.5}$$

The actual stringency profile correlates to a load reduction of approximately 2.66 million MWh over the 2020–2021 period, averaging -151.53 MW hourly. After simulating a low stringency scenario the estimated total impact is reduced to approximately -2.13 million MWh, with an average hourly impact of -121.22 MW. Simulating a high stringency scenario shows a higher impact of approximately -3.32 million MWh over 2020–2021, corresponding to an average hourly reduction of -189.41 MW. These simulations demonstrate that, under the assumed proportional relationship, stricter measures would have led to a more pronounced suppression of electricity demand, while less stringent measures would have resulted in a comparatively smaller reduction.

It is important to acknowledge that this approach is a simplification. Another key assumption is the relaxation or strengthening of the measures would have occurred at the same time as the government implemented the observed measures. Even though a model explicitly trained on stringency data as one of the inputs is developed, the results would still rely on the same assumption. Furthermore, the simulation requires one to account for the Russia–Ukraine war which followed the end of the lockdown measures. Nevertheless, for the current scope of this thesis, a simple heuristic approach offers valuable insights.

Table 6.2: Hypothetical Impact of Different Stringency Levels on Electricity Consumption (2020–2021).

Scenario	**Year**	**Total Impact (MWh)**	**Avg. Hourly Impact (MW)**
Actual	2020	-2,123,687.71	-241.77
	2021	-534,723.31	-61.04
	Total	-2,658,411.02	-151.53
Low	2020	-1,698,950.17	-193.41
	2021	-427,778.65	-48.83
	Total	-2,126,728.82	-121.22
High	2020	-2,654,609.64	-302.21
	2021	-668,404.13	-76.30
	Total	-3,323,013.78	-189.41

6.1 Discussion of Results

The results of the counterfactual analysis provide strong evidence that the COVID-19 pandemic significantly altered electricity consumption patterns in the Czech Republic, both in magnitude and distribution across time periods. Using the most accurate forecasting model — XGBoost with LASSO-selected features — this study estimated a total reduction of 2.66 TWh in electricity demand across 2020 and 2021. This represents a meaningful 4.0% decrease compared to 2019's baseline consumption of 66.15 TWh.

These findings reinforce the assertion by McKibbin & Fernando (2023) that the pandemic introduced non-linear shocks to economic activity, which were clearly reflected in electricity load profiles. In addition to measuring the total change, the analysis also measured the variations between day and night time. The average load dropped by 145.68 MW during daytime hours and by 157.37 MW during the night. These differences were statistically significant ($p = 0.03$). While unexpected, this trend may point to a deeper structural shift, where industrial and commercial baseloads were more heavily curtailed than residential consumption. Furthermore, this finding aligns with similar patterns noted in the Victoria, Australia study by (Wu *et al.* 2023), where industrial zones demonstrated flatter and more depressed load curves during lockdowns.

Importantly, the stringency index analysis confirms a close correlation between governmental restrictions and demand reductions. The second quarter of 2020 experienced the most severe stringency, which coincided with the most substantial drops in load. This correlation further validates the use of non-demand-side predictors (weather and calendar variables) in generating accurate counterfactuals, as first proposed by Behm *et al.* (2020). It also supports Nabavi *et al.*'s (2024) argument that building models capable of anticipating the effects of non-linear societal disruptions is essential to load forecasting.

When compared with other studies, particularly Gulati *et al.* (2021) for India and Abulibdeh *et al.* (2022) for Qatar, the Czech case mirrors falls in line with their findings of decreased industrial demand and somewhat stable residential consumption. However, this study adds to the understanding of how the pandemic changed the load profile for the Czech Republic by synthetically generating a load profile based on weather variables only. This methodological distinction avoids problems of post-shock model distortion and validates Behm *et al.*'s (2020) core idea of using weather related features without relying on historical load values.

Going to the specifications, a final consideration is the performance of the modeling framework itself. The robustness of the XGBoost model with LASSO-selected features ($R^2 = 0.929$, RMSE = 334.3 MW) validates both the model choice and feature reduction strategy. Compared to standalone LASSO ($R^2 = 0.453$) and unfiltered ANN ($R^2 = -4.204$), the hybrid approach achieved a superior balance between predictive accuracy and interpretability — crucial for policy-relevant insights.

6.2 Economic Implications

The results of this study show that the measures to contain the pandemic changed the load profile. The findings contrast ČEPS's expectations presented in their 2016 Mid-term Adequacy Forecast for the Czech Republic. At that time, ČEPS projected an increase in the Czech Republic's net consumption, expecting it to reach 65.5 TWh in 2020 and 67.0 TWh in 2025. While the analysis quantifies the immediate impact of the pandemic on the load profile, it is possible that these changes may be part of broader underlying economic transformations. One possibility is the potential decoupling between economic growth and electricity consumption.

The overall energy efficiency gains reported in the ODYSSEE-MURE (2025) energy profile report suggest an economy that is on track to decouple in the medium term. Net energy gains became 28% more efficient from 2000 to 2022. Although these efficiency gains and structural shifts have not yet fully offset the growth in energy consumption driven by overall economic activity, they suggest a weakening link between economic expansion and increased electricity demand. Therefore, the measured impact of COVID-19 might partially overlap with an accelerated manifestation of existing or new economic trends. For policy makers interested in exploring the economic link, this implies that future load profile predictions must account for these complex interactions, moving beyond single-factor explanations such as the impact of COVID-19 explored in this thesis. It is prudent to account for the dynamic relationship between energy efficiency, sectoral shifts, and overall economic activity even in the context of unforeseen shocks.

6.3 Future Work

The counterfactual analysis quantifies the change in the electricity load during the 2020–2021 period that coincided with the COVID-19 pandemic and associated measures. By comparing the observed load with a synthetic load profile, this study estimates how the load profile deviated from expectations. While the observed impact clearly demonstrates the influence of the pandemic, the load forecasting model was not designed to directly incorporate the stringency index as a predictive feature. The scope of the thesis was confined to measuring the extent to which the COVID-19 pandemic changed the load profile. Consequently, the pre-COVID data would feature stringency index values of zero before 2020. The feature selection would have eliminated this variable.

However, integrating the COVID stringency index into the model's feature set is an opportunity to understand how different stringency levels affected the load profile. By extending the training set to the end of 2021 and retraining the model, the model could learn the relationship between policy measures and electricity demand. This can facilitate more sophisticated counterfactual simulations. From there it will be possible to determine how a different level of stringency could have affected the load profile. In the event of a future shock that requires the government to implement similar measures, it will be possible to simulate different scenarios.

Furthermore, the model can be expanded to incorporate a broader set of other economic indicators as feature. Examples are macroeconomic variables such as GDP, sectoral output, or demographic trends like population growth. This study has demonstrated the effectiveness of LASSO as a feature selection tool. It is possible use LASSO to select the appropriate features from a richer set of candidate features Z_t, which would include not only traditional economic and demographic variables but also variables capturing societal shocks and behavioral responses. The electricity load can then be modeled as:

$$L_t = f(Z_t) + \epsilon_t,$$

where $f(\cdot)$ is a flexible non-linear function such as XGBoost, trained on LASSO-selected features from Z_t.

A broad model with other macroeconomic variables will likely lead to more nuanced forecasts. As a policy recommendation, planners should adopt synthetic forecasting models as part of scenario planning for future crises. For shocks that exhibit structural similarities to COVID-19, this approach could support rapid contingency planning and enhance the resilience of electricity systems.

6.3.1 Opportunities to Incorporate Artificial Intelligence

Beyond enhancing the model to incorporate a broader range of economic indicators, another opportunity lies in Artificial Intelligence (AI). It is generally expected that AI will be used more in a wide range of tasks. Large Language Models (LLMs) and conversational AI in particular, could be potentially be used to analyze extensive datasets to identify patterns and predict future outcomes. Su *et al.* (2024), in their systematic literature review highlight that LLMs are inherently well-suited for timestamped data. They also have the ability to recognize deviations from the norm. Interactive AI could be used to draw insights from the core forecasting model to quickly deliver insights. For example, a stakeholder could pose questions in natural language, such as: "What is the projected electricity demand for next winter under a specific scenario?" or "How would a given percentage increase in renewable energy sources impact grid stability?" This interactive capability would enable real-time exploration, allowing input for hypothetical questions for policy and related use cases.

Chapter 7

Conclusion

This thesis sought to understand how the electricity load profile for the Czech Republic would have been had the COVID-19 pandemic not occurred. Given the non-linear nature of variables that affect electricity demand, it was essential to specify a model capable of dealing with the complex interaction between variables. By generating synthetic load profiles based solely on weather and calendrical variables, the study's finding is that the Czech Republic's load profile was significantly altered during the pandemic. Using a modeling approach inspired by Behm et al. (2020), selecting the machine learning algorithm for the model took an iterative approach.

Three models types were evaluated, namely LASSO regression model, an artificial neural network model, and an extreme gradient boosting model. LASSO was also used to perform feature selection, and each model was trained on pre-pandemic data from 2016—2018 and tested on 2019 data. The XGBoost model with LASSO selected features was the best performing model with an R^2 of 0.929 and RMSE of 334.3 MW. This model was used to produce the counterfactual load profiles for 2020 and 2021, representing what electricity demand would have been without the pandemic.

A comparison between the observed and counterfactual load profiles showed that COVID-19 reduced electricity consumption by approximately 2.66 TWh over the two years. This change is equivalent to 4.0% of 2019's total demand, with the biggest change observed in 2020. The changes were more significant during night hours. This suggests that the effects of societal restrictions extended beyond typical working hours, perhaps due to prolonged reductions in commercial and industrial activity.

The results also confirm that synthetic load profiles generated from weather and calendar features can serve as reliable estimates of baseline demand. Although this study intentionally limited model inputs to weather and calendrical variables for interpretability and transferability, future work could improve this framework by incorporating additional predictors such as population data among other factors related to electricity consumption. Further research could also apply this method to regional or sector-specific data, or test in other countries.

In closing, this study establishes an adaptable approach to counterfactual modeling framework that can help energy planners and policymakers better understand and prepare for medium to long-term changes in electricity demand following unexpected events.

Bibliography

Abulibdeh, A., E. Zaidan, & R. Jabbar (2022): "The impact of COVID-19 pandemic on electricity consumption and electricity demand forecasting accuracy: Empirical evidence from the state of Qatar." *Energy Strategy Reviews* **44**: p. 100980.

Ahsan, M. M., M. P. Mahmud, P. K. Saha, K. D. Gupta, & Z. Siddique (2021): "Effect of data scaling methods on machine learning algorithms and model performance." *Technologies* **9(3)**: p. 52.

Baur, L., K. Ditschuneit, M. Schambach, C. Kaymakci, T. Wollmann, & A. Sauer (2024): "Explainability and interpretability in electric load forecasting using machine learning techniques - a review." *Energy and AI* **16**: p. 100358.

Behm, C., L. Nolting, & A. Praktiknjo (2020): "How to model european electricity load profiles using artificial neural networks." *Applied Energy* **277**: p. 115564.

Chen, S.-S. (2023): "A direct approach to kilian–lewis style counterfactual analysis in vector autoregression models." *Journal of Applied Econometrics* **38(7)**: pp. 1068–1076.

Czech Statistical Office (2022): "Czech republic electricity exports (1993–2022)." [Data set]. CEIC.

Elamin, N. & M. Fukushige (2018): "Modeling and forecasting hourly electricity demand by sarimax with interactions." *Energy* **165**: pp. 257–268.

ENTSO-E (2025): "Power system incident in czech republic." `https://www.entsoe.eu/news/2025/07/04/power-system-incident-in-czech-republic/`. Accessed: 2025-07-25.

Fatima, S., A. Hussain, S. B. Amir, S. H. Ahmed, & S. M. H. Aslam (2023): "Xgboost and random forest algorithms: An in-depth analysis." *Pakistan Journal of Scientific Research, PJOSR* **3(1)**: pp. 26–31.

Freijeiro-González, L., M. Febrero-Bande, & W. González-Manteiga (2022): "A critical review of lasso and its derivatives for variable selection under dependence among covariates." *International Statistical Review* **90(1)**: pp. 118–145.

García-Puente, B., A. Rodríguez-Hurtado, M. Santos, & J. E. Sierra-García (2023): "Evaluation of xgboost vs. other machine learning models for wind parameters identification." *Renewable Energy and Power Quality Journal* **21**: pp. 388–393.

Goodfellow, I., Y. Bengio, & A. Courville (2016): *Deep Learning*. Cambridge, MA: MIT Press.

Gulati, P., A. Kumar, & R. Bhardwaj (2021): "Impact of Covid19 on electricity load in Haryana (India)." *International Journal of Energy Research* **45**: pp. 3397–3409.

Hale, T., N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hallas, S. Majumdar, & H. Tatlow (2021): "A global panel database of pandemic policies (oxford covid-19 government response tracker)." *Nature Human Behaviour* .

Harvey, D., S. Leybourne, & P. Newbold (1997): "Testing the equality of prediction mean squared errors." *International Journal of Forecasting* **13(2)**: pp. 281–291.

Jain, A. & S. C. Gupta (2024): "Evaluation of electrical load demand forecasting using various machine learning algorithms." *Frontiers in Energy Research* **12**: p. 1408119.

Jia, Z. & B. Lin (2022): "Ceeea2.0 model: A dynamic cge model for energy–environment–economy analysis with available data and code." *Energy Economics* **112**: p. 106117.

Kirilenko, D., P. Barbiero, M. Gjoreski, M. Luštrek, & M. Langheinrich (2024): "Generative models for counterfactual explanations." In "HI-AI@KDD, Human-Interpretable AI Workshop at the KDD 2024," .

Lai, J.-P., Y.-M. Chang, C.-H. Chen, & P.-F. Pai (2020): "A survey of machine learning models in renewable energy predictions." *Applied Sciences* **10(17)**: p. 5975.

Lee, J. & Y. Cho (2022): "National-scale electricity peak load forecasting: Traditional, machine learning, or hybrid model?" *Energy* **239**: p. 122366.

LEPOT, M., J.-B. AUBIN, & F. H. L. R. CLEMENS (2017): "Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment." *Water* **9(10)**: p. 796.

MCKIBBIN, W. & R. FERNANDO (2023): "The global economic impacts of the COVID-19 pandemic." *Economic Modelling* **129(C)**.

MESTRES DOMÈNECH, J. & Z. MARTÍN VILATÓ (2025): "The economic impact of the blackout in detail: A cross-sector and regional analysis based on caixabank data." *Analysis report*, CaixaBank Research.

NABAVI, S. A., S. MOHAMMADI, N. H. MOTLAGH, S. TARKOMA, & P. GEYER (2024): "Deep learning modeling in electricity load forecasting: Improved accuracy by combining dwt and lstm." *Energy Reports* **12**: pp. 2873–2900.

NTI, I. K., M. TEIMEH, O. NYARKO-BOATENG, & ET AL. (2020): "Electricity load forecasting: A systematic review." *Journal of Electrical Systems and Information Technology* **7(1)**: p. 13.

ODYSSEE-MURE (2025): "Czechia: Energy profile."

PAPAIOANNOU, G. P., C. DIKAIAKOS, A. DRAMOUNTANIS, & P. G. PAPAIOANNOU (2016): "Analysis and modeling for short- to medium-term load forecasting using a hybrid manifold learning principal component model and comparison with classical statistical models (sarimax, exponential smoothing) and artificial intelligence models (ann, svm): The case of greek electricity market." *Energies* **9(8)**: p. 635.

PERÇUKU, A., D. MINKOVSKA, & N. HINOV (2025): "Enhancing electricity load forecasting with machine learning and deep learning." *Technologies* **13(2)**: p. 59.

PROBST, P., A.-L. BOULESTEIX, & B. BISCHL (2019): "Tunability: Importance of hyperparameters of machine learning algorithms." *Journal of Machine Learning Research* **20**: pp. 1–32.

SU, J., C. JIANG, X. JIN, Y. QIAO, T. XIAO, H. MA, & J. LIN (2024): "Large language models for forecasting and anomaly detection: A systematic literature review." *arXiv preprint arXiv:2402.10350* .

UĞURLU, E. & I. JINDŘICHOVSKÁ (2024): "The relationship between electricity consumption, trade, and gdp and the effect of covid-19: a panel ardl approach on the visegrad countries." *Frontiers in Energy Research* **11**: 1141847.

VISUAL CROSSING CORPORATION (2025): "Visual crossing weather (2016 - 2021)." [data service].

Wang, P., B. Liu, & T. Hong (2016): "Electric load forecasting with recency effect: A big data approach." *International Journal of Forecasting* **32(3)**: pp. 585–597.

World Bank (2023): "Czech republic electrical energy exports by country in 2023." World Integrated Trade Solution (WITS).

Wu, J., N. Levi, R. Araujo, & Y.-G. Wang (2023): "An evaluation of the impact of COVID-19 lockdowns on electricity demand." *Electric Power Systems Research* **216**: p. 109015.

Zhou, S. (2010): "Thresholded Lasso for high dimensional variable selection and statistical estimation."

ČEPS (2016): "Mid-term adequacy forecast of the czech republic." *Technical report*, ČEPS, Prague, Czech Republic.

ČEPS (2025): "All data." `https://www.ceps.cz/en/all-data`. [Accessed 2025-03-02].

Řanda, M. (2023): *Short-term Electric Load Forecasting Using Czech Data.* Master's thesis, Charles University.

Appendix A

Pre-COVID Load Analysis

Table A.1: Pre-COVID (2016-2019) Day vs Night Load Analysis Summary

Metric	**Value**
Average Day Load (MW)	8,158.38
Average Night Load (MW)	6,874.56
Difference (Day - Night, MW)	1,283.82
Standard Deviation Day Load (MW)	1,105.13
Standard Deviation Night Load (MW)	1,089.08
T-statistic	108.39
P-value	$< .001$

Appendix B

Iterative Modelling

Figure B.1: LASSO-selected features and their impacts used in downstream models.[1]

```
Base directory set to: C:\Users\cnmut\OneDrive\Desktop\CERGE\Thesis\Data
Alpha range tested: 0.00000100 to 1000.0
Best alpha: 5.33669923
Alpha position: 75/100 in tested range

--- Selected Features Impact ---
Quantitative impact (MW per unit change, unless specified):
- **temp**: -4.68 MW per 1°C change
- **feelslike**: -67.26 MW per 1°C change
- **dew**: -11.93 MW per 1°C change
- **windspeed**: 1.80 MW per 1 km/h change
- **winddir**: -0.56 MW per 1° direction change
- **solarradiation**: 2.55 MW per 1 W/m² change
- **cloudcover**: 3.69 MW per 1% change
- **snow**: 193.11 MW per 1 cm change
- **visibility**: -1.87 MW per 1 km change
- **snowdepth**: 73.14 MW per 1 cm change
- **preciptype_rain**: 13.18 MW when this condition occurs (binary feature)

Temporal features capturing load patterns: month, day, hour, weekday, year

Total selected features: 16 out of 21 total

Selected features list saved to 'selected_features_for_ml.json'.

❌ Model rejected - insufficient performance
Test R²: 0.453, RMSE: 926.1 MW
```

Figure B.2: ANN model training (No LASSO).

```
=== TRAINING COMPLETED ===
ANN converged in 504 epochs
Final training loss: 13729.5775
Early stopping: Yes

=== ANN PERFORMANCE (ALL FEATURES) ===
Training RMSE: 163.5 MW
Training R²: 0.984
Test RMSE: 2855.7 MW
Test R²: -4.204

=== OVERFITTING CHECK ===
R² gap (train - test): 5.187
⚠  Potential overfitting detected

=== TOP 10 FEATURES BY AVERAGE WEIGHT MAGNITUDE ===
 1. preciptype_snow     : 0.8254
 2. month               : 0.6796
 3. hour                : 0.5813
 4. preciptype_rain     : 0.5568
 5. year                : 0.4905
 6. weekday             : 0.4541
 7. annual_peak         : 0.4354
 8. day                 : 0.3977
 9. temp                : 0.3964
10. dew                 : 0.3297

Results for ANN with all features saved to: ann_all_features_results.json
Temporary ANN model (all features) and scaler saved to: C:\Users\cnmut\OneDrive\Desktop\CERGE\Thesis\Data\_temp_ann_all_features_model.pkl
```

Figure B.3: ANN model training (LASSO selected features).

```
Alpha range tested for LASSO: 0.00000100 to 1000.0
Best alpha for LASSO: 5.33669923

Selected 16 features out of 21 total:
- month: 15.77 MW (standardized)
- day: -10.71 MW (standardized)
- hour: 232.83 MW (standardized)
- weekday: -484.64 MW (standardized)
- temp: -40.83 MW (standardized)
- feelslike: -663.35 MW (standardized)
- dew: -79.69 MW (standardized)
- windspeed: 11.88 MW (standardized)
- winddir: -55.60 MW (standardized)
- solarradiation: 521.80 MW (standardized)
- cloudcover: 123.53 MW (standardized)
- year: 118.66 MW (standardized)
- snow: 15.91 MW (standardized)
- visibility: -25.14 MW (standardized)
- snowdepth: 83.36 MW (standardized)
- preciptype_rain: 13.18 MW (standardized)

=== ANN TRAINING WITH LASSO SELECTED FEATURES (OPTIMIZING ALPHA) ===
Starting ANN alpha optimization with GridSearchCV...
Fitting 3 folds for each of 7 candidates, totalling 21 fits

=== ANN OPTIMIZATION COMPLETED ===
Best alpha found for ANN: 0.00100000
Best cross-validation score (negative MSE): -575597.03
Optimal ANN converged in 428 iterations

=== OPTIMAL ANN PERFORMANCE (LASSO FEATURES) ===
Training RMSE: 166.4 MW
Training R²: 0.983
Test RMSE: 631.2 MW
Test R²: 0.746

=== LASSO PERFORMANCE (for comparison) ===
Training RMSE: 844.8 MW
Training R²: 0.563
Test RMSE: 926.1 MW
Test R²: 0.453

=== MODEL COMPARISON ===
Optimal ANN (LASSO Features) Test R²: 0.746, RMSE: 631.2 MW
LASSO Regression Test R²: 0.453, RMSE: 926.1 MW

Results saved to ann_lasso_features_optimized_results.json
```

Figure B.4: XGBoost model training (No LASSO).

```
=== XGBOOST TRAINING WITH ALL FEATURES ===
Training data shape: (25560, 22)
Test data shape: (8760, 22)
Number of features: 21
Training samples: 25560
Test samples: 8760

=== FEATURE SCALING COMPLETED ===
Feature means (should be ~0): [-1.24539572e-16  4.67023394e-17  3.05789127e-18  3.11348930e-17
 -4.26992818e-16]...
Feature stds (should be ~1): [1. 1. 1. 1. 1.]...

=== XGBOOST TRAINING (ALL FEATURES - OPTIMIZING HYPERPARAMETERS) ===
Starting XGBoost hyperparameter optimization...
Fitting 3 folds for each of 108 candidates, totalling 324 fits

=== XGBOOST OPTIMIZATION COMPLETED ===
Best hyperparameters found for XGBoost: {'colsample_bytree': 1.0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 300, 'random_state': 42, 'subsam
ple': 0.8}
Best cross-validation score (negative MSE): -96504.16

=== XGBOOST PERFORMANCE (ALL FEATURES) ===
Training RMSE: 86.3 MW
Training R²: 0.995
Test RMSE: 344.4 MW
Test R²: 0.924

=== TOP 10 FEATURES BY XGBOOST IMPORTANCE ===
 1. feelslike          : 0.3059
 2. hour               : 0.2412
 3. weekday            : 0.2136
 4. temp               : 0.0984
 5. month              : 0.0340
 6. solarradiation     : 0.0253
 7. dew                : 0.0215
 8. annual_peak        : 0.0162
 9. snowdepth          : 0.0144
10. day                : 0.0101

Results for XGBoost with all features saved to: xgb_all_features_results.json

Current XGBoost (all features) (Test R²: 0.924) did not beat the overall best (Test R²: 0.929).

Overall workflow best model tracker updated in: workflow_best_model_tracker.json
```

Figure B.5: XGBoost model training (LASSO selected features).

```
=== XGBOOST TRAINING WITH LASSO SELECTED FEATURES ===
Training data shape: (25560, 22)
Test data shape: (8760, 22)

Successfully loaded 16 features selected by LASSO.
Number of LASSO selected features: 16
Training samples with selected features: 25560
Test samples with selected features: 8760

=== FEATURE SCALING COMPLETED FOR SELECTED FEATURES ===
Selected feature means (should be ~0): [-1.24539572e-16  4.67023394e-17  3.05789127e-18  3.11348930e-17
 -7.56133115e-17]...
Selected feature stds (should be ~1): [1. 1. 1. 1. 1.]...

=== XGBOOST TRAINING (LASSO FEATURES - OPTIMIZING HYPERPARAMETERS) ===
Starting XGBoost hyperparameter optimization for LASSO features...
Fitting 3 folds for each of 72 candidates, totalling 216 fits

=== XGBOOST OPTIMIZATION COMPLETED (LASSO FEATURES) ===
Best hyperparameters found for XGBoost (LASSO Features): {'colsample_bytree': 0.7, 'learning_rate': 0.07, 'max_depth': 6, 'n_estimators': 350, 'random_st
ate': 42, 'subsample': 0.7}
Best cross-validation score (negative MSE): -91928.53

=== XGBOOST PERFORMANCE (LASSO SELECTED FEATURES) ===
Training RMSE: 98.9 MW
Training R²: 0.994
Test RMSE: 334.3 MW
Test R²: 0.929

=== TOP 10 FEATURES BY XGBOOST IMPORTANCE (LASSO FEATURES) ===
 1. weekday            : 0.2291
 2. hour               : 0.2148
 3. feelslike          : 0.1783
 4. dew                : 0.0917
 5. solarradiation     : 0.0897
 6. temp               : 0.0640
 7. month              : 0.0466
 8. snow               : 0.0178
 9. snowdepth          : 0.0159
10. year               : 0.0145

Results for XGBoost with LASSO selected features saved to: xgb_lasso_features_optimized_results.json
```

Appendix C

Model Tuning

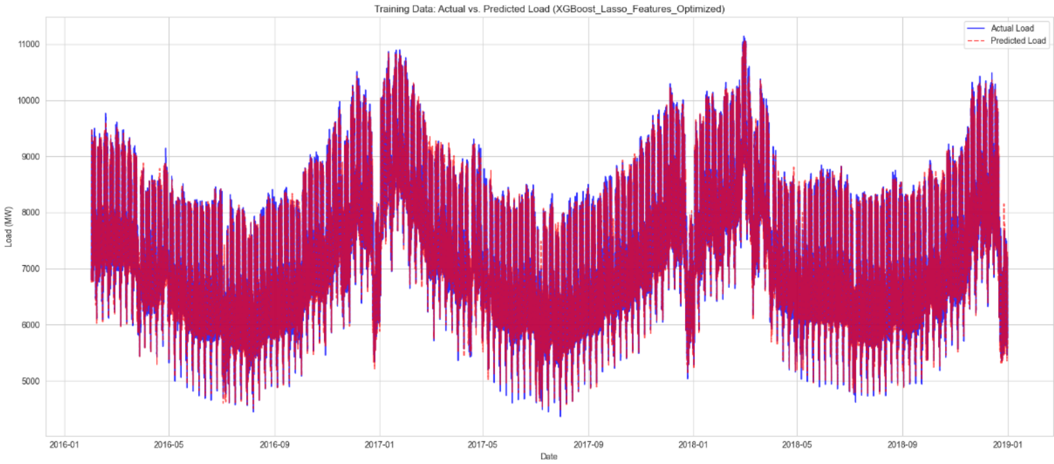
Figure C.1: Model Tuning Results (part 1).

```
=== CONSOLIDATING OVERALL BEST MODEL TRACKER ===
This step ensures 'workflow_best_model_tracker.json' points to the absolute best model from all previous experiments.
Loaded existing tracker. Current best: XGBoost_Lasso_Features_Optimized (R²: 0.929)
 - ANN_lasso_features_optimized (Test R²: 0.746) is not better than current best.
Warning: lasso_results.json not found. Lasso-only performance not considered.
 - ANN_all_features (Test R²: -4.204) is not better than current best.
 - XGBoost_all_features_optimized (Test R²: 0.924) is not better than current best.

Finalized workflow best model tracker updated in: workflow_best_model_tracker.json
Overall Best Model identified: **XGBoost_Lasso_Features_Optimized** (Test R²: 0.929, RMSE: 334.3 MW)
```

Figure C.2: Model Tuning Results (part 2).

```
Base directory set to: C:\Users\cnmut\OneDrive\Desktop\CERGE\Thesis\Data
=== FINAL MODEL TUNING WITH SCIKIT-LEARN PIPELINE ===

1. Identifying and Loading the Current Best Model...
Current best model: XGBoost_Lasso_Features_Optimized
Loading model from: C:\Users\cnmut\OneDrive\Desktop\CERGE\Thesis\Data\_temp_xgb_lasso_optimal_model.pkl
Model and features loaded successfully.
Features selected: 16
Training samples: 25560, Test samples: 8760

2. Building the Scikit-learn Pipeline...
Best model is XGBoost. Initializing XGBRegressor.
Pipeline created with: XGBRegressor

3. Performing hyperparameter tuning...
Fitting 5 folds for each of 75 candidates, totalling 375 fits

=== PIPELINE OPTIMIZATION COMPLETED ===
Best hyperparameters: {'model__subsample': 0.8, 'model__reg_alpha': 0.001, 'model__n_estimators': 500, 'model__min_child_weight': 5, 'model__max_depth': 
4, 'model__learning_rate': 0.1, 'model__gamma': 0.2, 'model__colsample_bytree': 0.9}
Best CV score (neg MSE): -86242.39

4. Evaluating and saving the final pipeline...

Training RMSE: 133.2 MW
Training R²: 0.989
Test RMSE: 334.5 MW
Test R²: 0.929

Previous best model Test R²: 0.929, RMSE: 334.3 MW
Saved final pipeline to: C:\Users\cnmut\OneDrive\Desktop\CERGE\Thesis\Data\final_production_model_pipeline_xgboost_lasso_features_optimized.pkl

Final tuned pipeline did NOT beat previous best model.

Updated best model tracker saved: C:\Users\cnmut\OneDrive\Desktop\CERGE\Thesis\Data\workflow_best_model_tracker.json
Final best model in workflow: XGBoost_Lasso_Features_Optimized (Test R²: 0.929)
```

Figure C.3: Tuned Model Performance.



Appendix D

Counterfactual Analysis

Figure D.1: Counterfactual Analysis Summary.

```
6. Summarizing the estimated impact...

Total estimated COVID-19 impact on load (2020-2021): -2658411.02 MWh
Average hourly COVID-19 impact on load (2020-2021): -151.53 MW

--- Yearly COVID-19 Impact Summary ---
Year 2020: Total Impact = -2123687.71 MWh, Average Hourly Impact = -241.77 MW
Year 2021: Total Impact = -534723.31 MWh, Average Hourly Impact = -61.04 MW

Counterfactual analysis completed successfully!
```

```
C:\Users\cnmut\AppData\Local\Temp\ipykernel_16264\4103525637.py:234: FutureWarning: 'Y' is deprecated and will be removed in a future version, please use
'YE' instead.
  yearly_impact_mwh = df_impact['estimated_impact'].resample('Y').sum()
C:\Users\cnmut\AppData\Local\Temp\ipykernel_16264\4103525637.py:235: FutureWarning: 'Y' is deprecated and will be removed in a future version, please use
'YE' instead.
  yearly_avg_impact_mw = df_impact['estimated_impact'].resample('Y').mean()
```

```python
# This cell should be run AFTER the full preceding code block has executed,
# ensuring 'counterfactual_data' and 'covid_stringency_actual' are defined.

if 'covid_stringency' in counterfactual_data.columns and not counterfactual_data['covid_stringency'].isnull().all():
    highest_observed_stringency = counterfactual_data['covid_stringency'].max()
    print(f"The highest observed COVID Stringency Index for CZ in your data is: {highest_observed_stringency:.2f}")
else:
    print("The 'covid_stringency' column was not found or is entirely null in 'counterfactual_data'.")
```

```
The highest observed COVID Stringency Index for CZ in your data is: 82.41
```

Figure D.2: Heuristic Counterfactual Analysis.

```
=== HEURISTIC COUNTERFACTUAL ANALYSIS: VARYING COVID STRINGENCY ===
NOTE: This analysis uses a heuristic approach to simulate stringency impact,
      by directly scaling the observed impact, without retraining the primary load forecasting model.

1. Loading the final pipeline and data...
Features loaded from tracker (these are what the model was trained on): ['month', 'day', 'hour', 'weekday', 'temp', 'feelslike', 'dew', 'windspeed', 'win
ddir', 'solarradiation', 'cloudcover', 'year', 'snow', 'visibility', 'snowdepth', 'preciptype_rain']
Final optimized model/pipeline loaded from: C:\Users\cnmut\OneDrive\Desktop\CERGE\Thesis\Data\_temp_xgb_lasso_optimal_model.pkl

2. Loading data and retrieving actual stringency...
  'covid_stringency' column successfully loaded.
  Actual Stringency (2020-2021) - Min: 0.00, Max: 82.41, Mean: 49.97

3. Generating 'No COVID' predictions and observed impact...
Scaler loaded from: C:\Users\cnmut\OneDrive\Desktop\CERGE\Thesis\Data\_temp_xgb_lasso_optimal_scaler.pkl
  'No COVID' predictions generated. This represents the load that would have occurred
  if COVID-19 related behavioral changes (beyond what the model learned from other features)
  had not taken place.
  Observed COVID-19 impact calculated. This is the actual deviation from the 'No COVID' baseline.

4. Heuristically scaling observed impact for low and high stringency scenarios (Direct Scaling)...
  Justification for this revised heuristic method:
  Given the primary load forecasting model was not trained on 'covid_stringency',
  and to directly reflect the user's desired scaling, this method directly scales
  the *observed COVID-19 impact* by fixed factors (1/1.25 for low, 1.25 for high).
  This ensures that higher stringency leads to a larger magnitude of negative impact,
  and lower stringency leads to a smaller magnitude of negative impact, providing
  a clear illustrative simulation for thesis purposes.
  Hypothetical impacts for low and high stringency calculated based on direct scaling factors.

5. Summarizing the heuristic estimated impact for each scenario...

--- Heuristic Estimated COVID-19 Impact Summary by Stringency Scenario ---

Scenario: Observed Impact (Actual Stringency)
  Total Impact (2020-2021): -2658411.02 MWh
  Average Hourly Impact (2020-2021): -151.53 MW
  Yearly Breakdown for Observed Impact (Actual Stringency):
    Year 2020: Total Impact = -2123687.71 MWh, Average Hourly Impact = -241.77 MW
    Year 2021: Total Impact = -534723.31 MWh, Average Hourly Impact = -61.04 MW

Scenario: Heuristic Impact (Low Stringency)
  Total Impact (2020-2021): -2126728.82 MWh
  Average Hourly Impact (2020-2021): -121.22 MW
  Yearly Breakdown for Heuristic Impact (Low Stringency):
    Year 2020: Total Impact = -1698950.17 MWh, Average Hourly Impact = -193.41 MW
    Year 2021: Total Impact = -427778.65 MWh, Average Hourly Impact = -48.83 MW

Scenario: Heuristic Impact (High Stringency)
  Total Impact (2020-2021): -3323013.78 MWh
  Average Hourly Impact (2020-2021): -189.41 MW
  Yearly Breakdown for Heuristic Impact (High Stringency):
    Year 2020: Total Impact = -2654609.64 MWh, Average Hourly Impact = -302.21 MW
    Year 2021: Total Impact = -668404.13 MWh, Average Hourly Impact = -76.30 MW
```