Department of Statistics 2022-23

Estimating the Causal Effect of Germany's 2021 Nuclear Plant Closures on Electricity Generation Using Gaussian Process Regression

Candidate Number: 52187

LSE

Submitted for the Master of Science, London School of Economics, University of London

Table of contents

List of Figures

List of Tables

Summary

This study examines the repercussions of Germany's 2021 nuclear power plant closures on its fossil fuel electricity generation in 2022 using a comprehensive dataset of European electricity generation from 2015-2022. Four causal inference methodologies were employed: differences-in-differences, synthetic control, Bayesian structural time series, and Gaussian process regression. The goal was to determine the extent of potentially increased reliance on fossil fuels following the nuclear phase-out. The research integrates monthly electricity generation figures from various European countries and incorporates average temperature as a covariate.

The analysis reveals no evidence for an increase in Germany's fossil fuel electricity generation in 2022 as a result of the nuclear plant shutdowns. Among the models applied, Gaussian process regression stood out due to its nonparametric quality and Bayesian uncertainty quantification, complemented by a novel technique for automated feature selection. Contrary to popular anticipation as a result of past research focused on previous closures, the results challenge the preconceived notion of an inherent rise in fossil fuel utilization post-nuclear phase-out.

Methodologically, the study highlights the utility of combining flexible Bayesian modeling with more established causal inference techniques. In summary, this investigation not only offers valuable insights into the dynamics of energy transitions but also underscores the importance and potential of data-driven causal inference in understanding such shifts. The paper concludes with recommendations for more detailed exploratory studies incorporating richer datasets and enhanced modeling techniques in the future.

1 Introduction

Background and Context

Nuclear power has played a complex role in Germany's energy landscape over the past several decades. The country's first commercial nuclear reactor began construction in 1961, marking the beginnings of a nascent civilian nuclear industry. Through the 1960s and 1970s, nuclear energy capacity expanded as Germany, like other industrialized nations, pursued atomic power as a source of economic growth and energy security. By the 1980s, nuclear generated a substantial portion of Germany's electricity needs. However, even as the technology rapidly expanded, so did public skepticism and resistance. The roots of Germany's anti-nuclear movement emerged in the 1970s over concerns about radioactive waste disposal, safety, and local environmental impacts. Public opposition escalated sharply following the 1986 Chernobyl disaster, which directly contaminated parts of Germany and fueled anxiety about the risks of nuclear technology.

Against this backdrop, Germany's government decided in 2000 to phase out nuclear power completely by the early 2020s. The phase-out plan involved gradually shutting down all of the country's nuclear plants according to a decades-long unwinding scheme. However, in the late 2000s and early 2010s, efforts emerged to slow down the phase-out as part of climate change mitigation strategies focused on reducing fossil fuels. But the 2011 Fukushima nuclear accident in Japan abruptly reversed course. It prompted Chancellor Angela Merkel to return to the original phase-out policy, with eight nuclear plants closed immediately. Additional reactors ceased operations in 2015, 2017, 2019, in December 2021, and in 2023. The 2021 closures of three of the six remaining plants—seen by some to contribute to unusually high energy costs during 2022 after Russia cut off gas supplies to Europe—are the specific focus of this analysis.

The history of nuclear power in Germany has thus traversed changing political and public attitudes. The 2021 shutdowns mark the culmination of a decades-long societal debate over the merits and risks of atomic energy. Understanding the impacts of phasing out this major energy source remains an open question with implications for nuclear policy and climate change goals alike. This study aims to shed light on these complex dynamics.

Research Question

This analysis focuses on estimating the causal effect of Germany's 2021 nuclear plant closures on the country's electricity generation from fossil fuel sources during the year 2022. Specifically, it asks the question: How did the shutdown of Germany's last operating nuclear reactors in December 2021 impact fossil fuel electricity generation in 2022 compared to the counterfactual scenario without the closures?

Methodological Overview

Germany's energy transition and nuclear phase-out present a complex setting with limited comparable case studies. Isolating the effect of the 2021 closures is challenging amidst contemporaneous shocks and the cumulative impacts of prior nuclear reductions. Controlling for all relevant factors using traditional regression techniques could prove inadequate. This study therefore implements multiple state-of-the-art causal inference methods, leveraging their complementary strengths and limitations. These include differences-in-differences, synthetic control modeling, Bayesian structural time series, and Gaussian process regression. The diversity of techniques provides analytical triangulation while exploring innovative applications of these approaches. Ultimately, the goal is two-fold: to produce substantively meaningful insights into the nuclear closures' impacts and to refine methodological practice in causal analysis of observational time series data.

The differences-in-differences approach compares changes in fossil fuel electricity generation in Germany before and after the closures relative to other control countries. DiD assumes parallel pre-treatment trends between groups. Violations of this can bias results, while matching on pre-treatment trends can help satisfy this assumption. However, DiD cannot capture complex temporal dynamics.

The synthetic control method constructs a weighted combination of control units that approximates the pre-treatment characteristics of the treated unit. The effect is estimated by comparing their post-treatment difference. This data-driven approach relaxes assumptions but lacks inherent uncertainty quantification.

Bayesian structural time series models decomposes time series into trends, seasonality, and regression components using a state space framework. This provides a flexible approach to prediction but relies on modeling assumptions. Bayesian inference naturally quantifies uncertainty.

Gaussian process regression directly models the outcome variable using a nonparametric Gaussian process prior distribution. This flexibility avoids restrictive assumptions while offering principled uncertainty estimates. Automatic relevance determination is utilized for input variable selection to prevent overfitting.

Applying these complementary methods together gives an additive sort of clarity towards the topic at hand while examining their comparative strengths and limitations in context.

Significance

This study makes both substantive and methodological contributions.

Substantively, quantifying the impacts of Germany's 2021 nuclear closures provides insights relevant to policymakers and analysts alike. Understanding the empirical effects on fossil fuel reliance informs debates around nuclear phase-outs, energy security, and climate change mitigation. The results also offer potential lessons for other countries considering reducing nuclear power.

Methodologically, this analysis demonstrates innovative applications of causal inference techniques using observational time series data. It elucidates the relative merits of differences-in-differences versus synthetic controls versus flexible Bayesian approaches. Implementing multiple methods enables helps uncover more of the truth about the the question at hand than a single approach would while expanding the toolkit for other applied researchers. In particular, the use of Gaussian process regression with principled input variable selection showcases an underexplored approach in this domain.

Overall, this study aims to generate new evidence and methodological learnings valuable from both an empirical and technical perspective. Rigorously estimating the causal effect of Germany's nuclear policy shifts on fossil fuel generation makes contributions on multiple fronts.

Outline

The remainder of the paper proceeds as follows:

The Literature Review situates this study in the context of prior research on electricity generation, Germany's energy transition, and methods for causal inference with time series data.

The Data section describes the panel dataset used covering fossil fuel electricity generation and relevant predictors across European countries over time.

The Methods section provides technical details on the implementation of the four causal inference approaches: differences-in-differences, synthetic control modeling, Bayesian structural time series, and Gaussian process regression.

The Results section presents findings from each method, including graphical representations of the counterfactual predictions and treatment effect estimates.

The Discussion compares and contrasts the results obtained across techniques, focusing on their complementary strengths and limitations.

Finally, the Conclusion summarizes the key takeaways, reflects on the substantive and methodological implications, acknowledges limitations, and suggests directions for future work.

2 Literature Review

2.1 Subject Review

Electricity Generation

Electricity generation is influenced by a myriad of factors, including resource availability, technological advancements, and policy directives. The portfolio of resources for power generation typically spans both nonrenewable sources such as coal, gas, and nuclear, and renewable ones like wind, solar, and hydropower. Governments, through their policy and regulatory frameworks, can either promote or deter specific energy forms. For instance, the global push towards renewable energy is driven by the imperative to reduce reliance on environmentally detrimental fossil fuels, which has direct implications for electricity generation from these sources ('Renewable energy', 2021).

Managing the electricity grid requires a balance between supply and demand. This entails methodologies like load forecasting, prioritizing power plants based on their marginal costs, and assimilating renewable energy inputs. The increasing integration of renewable energy introduces variability to grid management. To mitigate this, strategies encompassing energy storage, demand response, and power import/export mechanisms are deployed (Alquthami et al., 2022). The dynamics of the wholesale electricity market, which modulates electricity prices in response to supply-demand fluctuations, can also influence the extent to which fossil fuels are used for electricity generation (Michaelides, 2019).

Energy Mix in Germany

Germany's energy landscape offers insights relevant to the study's focus. The nation's "Energiewende" policy underscores its commitment to transitioning towards a more sustainable energy paradigm (Heering and Gustafson, 2021). Historically anchored in coal and nuclear energy, Germany's energy landscape has been undergoing a transformative shift with an eye towards climate change mitigation.

By the mid-2020s, renewable energy sources began to dominate Germany's electricity generation. This shift is attributed to substantial investments in wind and solar energy, catalyzed by state-backed incentives (Wehrle, 2022). However, despite the rise of renewables, natural gas and coal remain significant for ensuring a reliable electricity supply, especially during peak demand or renewable downtimes (Rode et al., 2023). This reliance on fossil fuels, in the absence of nuclear energy, is central to the research question of this paper.

The orchestration of Germany's electricity grid is steered by multiple transmission system operators (TSOs). Their mandate encompasses real-time demand forecasting, power plant scheduling, and renewable energy integration. With the renewable quotient on an upward trajectory, TSOs are channeling investments into grid modernization and pioneering technologies to manage the sporadic nature of wind and solar energy (Ernst et al., 2009). The

interconnectedness of Germany's energy infrastructure with neighboring nations also plays a role in its energy mix, influencing the extent to which fossil fuels are used for electricity generation (Rocholl and Bolton, 2016).

Nuclear Energy in Germany

The history of nuclear energy capacity in Germany dates back to the 1960s and has gone through several phases (Häfner, 2019). Initially, nuclear policy was central to German industrial policy and national energy policy, and later, to environmental policy (Mez, 2020). Following the development of civilian nuclear power in the US, USSR, UK, and France, German nuclear plants began coming online in the 1960s and increased in number and technical sophistication through the 1980s (International Atomic Energy Agency, 2023). However, the young industry began to fade as quickly as it had grown (Häfner, 2019).
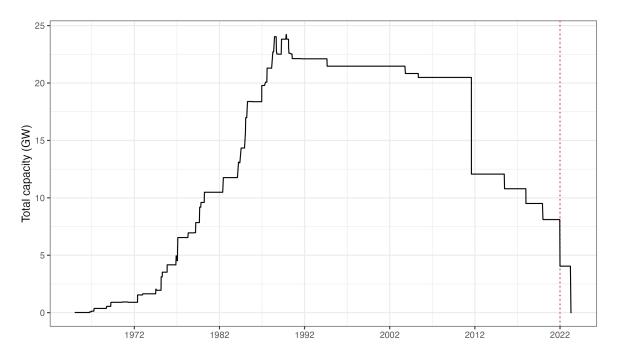
Figure 1: History of electricity generation capacity from nuclear power in Germany

Anti-nuclear sentiment started to grow around the same time as the first plans came online (Nelkin and Pollak, 1980). Activists were principally concerned with waste disposal but the cause quickly became associated with a larger political movement due to clashes with police and local governments. Public opinion was trending negative but was galvanized after much of Germany was exposed to radioactive fallout from the Chernobyl disaster in 1986 (Nelkin and Pollak, 1980). As told in Binder et al. (2001)'s Green Industrial Restructuring, the last new German nuclear plant was ordered in 1975 and activated in 1989; "since then no new nuclear power plants have been planned." After Chernobyl, the industry's focus pivoted from

5

expanding the number of reactors towards "improving the performance of existing reactors through upgrading, refurbishment, and managerial changes" Binder et al. (2001).

The overall political landscape in the 1990s, especially with the Green Party's ascent to the federal government post-reunification, bolstered the anti-nuclear stance (Rossnagel and Hentschel, 2012). Succumbing to the mounting political and public pressure, the German government legislated a nuclear phase-out by 2021 in 2000 (Rossnagel and Hentschel, 2012), leading to the decommissioning of the first plant in 2003 (International Atomic Energy Agency, 2023). However, the phase-out trajectory faced opposition in the late 2000s, primarily from a coalition spearheaded by the Christian Democrat party (Rossnagel and Hentschel, 2012). They underscored the role of nuclear energy in achieving Germany's climate objectives (Norman and Sweeney, 2010). Amidst significant public dissent, the phase-out deadline was extended to 2036 (Scherwath et al., 2020).

At the time of the Fukushima earthquake, nuclear power constituted one-quarter of Germany's electricity supply. However, the disaster was nonetheless a death blow to the industry. The events in Japan reignited public protests on an unprecedented scale and compelled the German government to reconsider its stance (Schreurs, 2013). A mere three days post-Fukushima, Prime Minister Angela Merkel announced a renewed commitment to the original phase-out plan.

The sequence of nuclear plant closures since 2011 (International Atomic Energy Agency, 2023) is as follows:

1. 2011: Immediate shutdown of eight nuclear power plants post-Fukushima, including Neckarwestheim 1, Biblis A and B, Brunsbüttel, Isar 1, Unterweser, Philippsburg 1, and Krümmel.
2. 2015: Grafenrheinfeld was decommissioned.
3. 2017: Gundremmingen B was decommissioned.
4. 2019: Philippsburg 2 was decommissioned.
5. 2021: Grohnde, Gundremmingen C, and Brokdorf were decommissioned.
6. 2023: The final three plants, Isar 2, Emsland, and Neckarwestheim 2 are decommissioned following a delay due to the 2022 Russian invasion of Ukraine (Alkousaa, 2023)

The 2021 closures are the focus of this paper.

A recent study by Jarvis et al. (2019) analyzed the impacts of Germany's nuclear phase-out following the 2011 Fukushima disaster. Using a machine learning approach on plant-level data, they estimated that the lost nuclear generation was replaced primarily by increased output from coal plants and electricity imports. This resulted in higher emissions of CO_2 and local air pollutants, increasing mortality risk from exposure. The study contrasts with this one in that it focuses on the 2011 closures and uses a different methodology for its period of consideration, which takes place prior to when ENTSO-E data is available. It is the single most useful point of comparison for the results obtained here.

6

2.2 Methods Review

The question I hope to answer can be characterized as a causal inference problem for time-series cross-sectional data (as usefully described in Carlson and Brown, 2020). To contextualize my methodological choices, I will review relevant literature on select approaches to questions following this general framework.

Differences-in-Differences

Differences-in-differences is a quasi-experimental approach commonly used to estimate the causal effect of a policy intervention or treatment (Doudchenko and Imbens, 2016). It involves comparing the change over time in an outcome for a treatment group to the change over time for a control group. The key assumption is that in the absence of treatment, the average outcomes for the treated and control groups would follow parallel trends over time.

DID has been applied extensively in economics and policy evaluation research. While a simple and intuitive technique, it relies on several strong assumptions, including no spillover effects between treatment and control groups, and balance between groups such that the parallel trends assumption holds (Doudchenko and Imbens, 2016). Violations of these assumptions can lead to biased effect estimates. Existing generalized forms help relax assumptions, but require more complex modeling and data to credibly identify effects (Xu, 2016).

Synthetic Control Method

The synthetic control method, introduced by Abadie and Gardeazabal (2003), is a comparative case study technique for causal inference with aggregate panel data. It constructs a synthetic control unit as a weighted average of control units that best approximates the pretreatment characteristics of the treated unit. The effect is estimated by comparing the posttreatment outcomes of the treated unit to its synthetic control (Abadie et al., 2010).

Compared to DID, SCM makes weaker assumptions and can yield improved effect estimation when there are no adequate existing control units or when the parallel trends assumption is violated. However, it still assumes no spillovers or anticipatory effects. The transparency of the technique has contributed to its rising popularity across the social sciences (Doudchenko and Imbens, 2016).

A key limitation is that the SCM does not naturally quantify uncertainty, hampering statistical inference. Extensions using placebo tests have been proposed to assess significance, but can suffer from multiple testing problems. Bayesian implementations are an active area of methodological research aimed at producing well-calibrated measures of uncertainty with the SCM (Kim et al., 2020).

Bayesian Structural Time Series

Bayesian structural time series models are a flexible approach for time series modeling and forecasting. They decompose a time series into interpretable components like trends, seasonality, and regression effects using latent state variables within a Bayesian state space modeling framework (Brodersen et al., 2015).

For causal inference, BSTS is proposed as an alternative to SCM that naturally quantifies uncertainty and accounts for anticipation effects using the regression component (Giudice et al., 2022). It does not require pre-specification of control units. However, it relies on modeling assumptions and places priors on components that can influence results. BSTS also cannot incorporate time-varying treatments. Extensions to the multivariate case have been developed, but can increase complexity (Menchetti and Bojinov, 2022).

Gaussian Process Regression

Gaussian process regression is a nonparametric Bayesian approach that specifies a prior directly over the space of functions describing the data, making it "a very attractive model in many application domains" (Rasmussen and Williams, 2006). This makes for a flexible model that lets the data speak for themselves without imposing restrictive assumptions upfront, making it "arguably the best default alternative for analyzing [time-series cross-sectional] data." (Carlson, 2018).

For panel data analysis and causal inference, GPR avoids limitations of SCM and BSTS. It can flexibly model trends and capture spillovers between units while quantifying uncertainty. It also does not require pre-specifying control units or components. These advantages have motivated applications of GPR for causal inference, though it is not yet as widely adopted as alternatives (e.g., Carlson and Brown, 2020). Challenges include computational scaling and choices involved in specifying the GP prior and likelihood. Extensions to multi-output GPs and deep GPs are promising avenues for future research (Eli Ben-Michael, 2023).

3 Methods

3.1 Data

The electricity generation data used in this analysis is sourced from the European Network of Transmission System Operators (ENTSO-E) Transparency Platform. This platform provides hourly data on electricity demand, supply, transmission, and generation reported by participating transmission system operators across Europe.

The raw data extracted from ENTSO-E covers the hourly domestic electricity generation from January 2015 to December 2022 for the 27 member countries of the European Union along with Serbia, Bosnia and Herzegovina, and the United Kingdom. Only generation from fossil fuel sources ("Fossil Gas," "Fossil Hard coal," "Fossil Oil," "Fossil Brown coal/Lignite," "Fossil Coal-derived gas," "Fossil Oil shale," and "Fossil Peat") is retained while renewables and nuclear are omitted.

The data is aggregated to monthly resolution due to the noisiness of hourly fluctuations. Some light preprocessing is applied, including identifying relevant generation sources as "fossil fuel" and filtering out countries with substantial missing observations.

Temperature data is sourced from OpenWeatherMap as an important covariate. Maximum and minimum temperatures are queried for every day between January 1st, 2015 and December 31st, 2022 for the weather station nearest the capital city of each country. The average of the maximum and minimum temperatures is used as the temperature value for each day. Average temperatures are then aggregated to the month level and incorporated into the main dataset based on the country and month corresponding to each electricity generation observation.

Dummy variables are constructed indicating the treatment group (Germany) and treatment period (post-nuclear closure).

The final time-series cross-sectional dataset is a panel of 22 countries over 103 months, resulting in 2266 total observations. Each observation includes the date (month), country name, total domestic fossil fuel electricity generation in MWh, and temperature.

A key limitation is that the ENTSO-E data only covers domestic generation within each country. It does not account for electricity imports and exports across borders (although the two outcomes are certainly related). This precludes examining potential spillover effects of Germany's nuclear phase-out on its neighboring countries.

3.2 Notation

This section outlines the notation employed throughout the dissertation, offering a structured and consistent approach to representing various elements of the study. The chosen notation draws inspiration from Doudchenko and Imbens (2016) and Giudice et al. (2022).

Consider units represented by i with $i \in 1,\dots,m$, observed over time periods t, with $t \in 1,\dots,T$. The variable of interest is the quantity of electricity generated (measured in MWh), represented as $y_{i,t}$. Using the ENTSO-E panel, there are m country units and T week periods. Specifically, Germany, a key focus of this study, is denoted by $i = 1$.

The treatment assignment indicator $w_{i,t}$ defines each observation $y_{i,t}$ with $w_{i,t} \in \{0,1\}$. The value '1' indicates treatment (e.g., Germany during the treatment period) and '0' represents controls (e.g. Germany outside the treatment period and all other countries at all times).

The closure of nuclear power plants in Germany on December 31st, 2021, serves as the primary treatment event. Therefore, T_0 represents the last period before this event, and $T_0 + 1$ the period including this date. Germany received treatment during $t = T_0 + 1$ to $t = T$.

The primary focus is on treatment effects for Germany during its treatment period, $t = T_0 + 1$ to T. The relationship between treatment $w_{i,t}$ and the realized outcome $y_{i,t}^{obs}$ can be illustrated as follows:

$$
y_{i,t}^{obs} = \begin{cases} y_{i,t}(0) & \text{if } w_{i,t} = 0, \\ y_{i,t}(1) & \text{if } w_{i,t} = 1. \end{cases}
$$

Note that some $y_{i,t}^{obs}$ values, especially for Germany post t_0, are missing since electricity from fossil fuels is observed once post-treatment.

The difference between two potential outcomes, $y_{i,t}(1)$ and $y_{i,t}(0)$, provides the causal effects at both the subject and time levels.

For this study, there is a single country- and time-varying predictor observed at each time period denoted with $x_{i,t}$. This predictor is used to control for the effects of exogenous variables on the outcome of interest.

Combining several of the above terms for convenience in applying some of the methods described below, I let $\mathbf{y} = \{y_{1,1},\dots,y_{1,T}\} \in \mathbb{R}^T$ be the time series of the treated country (Germany). Then $X = \{x_{i,1},\dots,x_{i,T}\}' \in \mathbb{R}^T$ be the vector of the associated covariate. Then with $\mathbf{z}_1'$ as the vector containing $\{y_{2,1},\dots,y_{m,1}\}$, the variable of interest observed for all control units at $t = 1$, I can define $Z = \{\mathbf{z}_1',\dots,\mathbf{z}_T'\}' \in \mathbb{R}^{T \times m-1}$ the matrix containing the time series of the relevant units for the synthetic control (other European countries). Then $X^* = \{X', Z'\}'$ is the combined matrix of the single covariate and all control units—each with sample size T.

3.3 Implementation

3.3.1 Differences-in-Differences

The Differences-in-Differences approach is a widely-used econometric technique designed to infer causal effects by leveraging temporal and cross-sectional variations in treatment expo-

sure. Given the panel structure of my dataset, where countries are observed over time and treatment is applied starting at a specific time, DiD serves as a valuable baseline.

3.3.1.1 Model Specification

The treatment of interest is the closure of nuclear power plants in Germany starting from $t = T_0 + 1$. For each country i and time t, the outcome $y_{i,t}^{obs}$ is said to vary based on its treatment assignment. To capture the causal effect of the treatment, the DiD method contrasts the change in outcomes of Germany (from pre-treatment to post-treatment) against the analogous change in outcomes of control countries.

To capture the causal effect of the treatment, the DiD method contrasts the change in outcomes of Germany (from pre-treatment to post-treatment) against the analogous change in outcomes of control countries.

To account for potential confounding factors, my model incorporates exogenous predictors, denoted by $x_{i,t}$, to control for their effects. The DiD model can be represented as:

$$
\begin{aligned}
y_{i,t} =& \alpha + \beta_1 \text{treatment\_period}_t + \beta_2 \text{treatment\_group}_i + \\
& \delta \text{treatment\_period}_t \times \text{treatment\_group}_i + \\
& \sum_{k=1}^{m} \gamma_k \times I(i = k) + \theta x_{i,t} + \sum_{k=1}^{m} \lambda_k \times x_{i,t} \times I(i = k) + \epsilon_{i,t}
\end{aligned}
$$

In this equation:

- $y_{i,t}$ is the quantity of electricity generated for country i at time t.
- α is the intercept term.
- β_1 captures the effect of the post-treatment period, with $\text{treatment\_period}_t$ indicating whether the observation is from the post-treatment period.
- β_2 captures the effect of being in the treatment group (Germany), with treatment_group_i indicating whether the observation is from Germany.
- δ captures the DiD estimator of the treatment effect, with $\text{treatment\_period}_t \times \text{treatment\_group}_i$ representing the interaction between the treatment period and the treatment group.
- $\sum_{k=1}^{m} \gamma_k \times I(i = k)$ represents the country fixed effects, where $I(i = k)$ is an indicator function that equals 1 if country i is the k-th country and 0 otherwise.
- $\theta x_{i,t}$ represents the controlled impact of the exogenous predictor on electricity generation for country i at time t
- $\sum_{k=1}^{m} \lambda_k \times x_{i,t} \times I(i = k)$ accounts for the controlled impacts of the exogenous predictor, temperature, interacted with country fixed effects. As above, $I(i = k)$ is an indicator function that equals 1 if country i is the k-th country and 0 otherwise, and λ_k is the coefficient for the interaction between temperature and the k-th country.

- $\epsilon_{i,t}$ is the error term.

An essential characteristic of the DiD method is the "parallel trends assumption." This presumes that, in the absence of treatment, the treated and untreated units would have shown parallel trends in outcomes. The strength of the DiD method lies in its simplicity and its ability to control for time-invariant unobserved heterogeneities. However, it's essential to be aware of its limitations. The method's reliance on the parallel trends assumption can be a significant source of bias if this assumption is violated. Additionally, any post-treatment events or shocks that differentially affect the treated and control groups can introduce bias. Lastly, the DiD approach assumes that the treatment effect is consistent across time, which might not hold in all contexts.

To estimate the causal effect of the nuclear plant closures on electricity generation, I employ a two-step strategy:

1. Pre-treatment matching: Before applying the DiD method, I match Germany with potential control countries based on pre-treatment characteristics and trends. This ensures that my treated and control units are as similar as possible before the treatment. Specifically, I normalize the outcome variable (electricity generation) within each country to lie between 0 and 1. I then compute the difference in trends between Germany and each potential control country during the pre-treatment period. Countries with statistically insignificant differences in trends, as determined by linear regression, are selected as controls.

2. DiD estimation: After matching, I apply the DiD method to the matched dataset according to the model given above.

3.3.1.2 Checks

As described, there are a number of potential pitfalls in applying the DiD method. To address these, I address several standard robustness checks.

As the parallel trends assumption is crucial for the validity of DiD estimates, I check for violations by graphically inspecting the pre-treatment trends of the outcome variable for Germany and potential control countries. This visualization, combined with the statistical tests mentioned earlier, helps ensure that the parallel trends assumption holds.

If other countries in my control group were indirectly affected by Germany's nuclear plant closures, my estimates could be biased in what is known as a "spill-over effect." To address this, I could conduct sensitivity analyses excluding countries that might have experienced significant indirect effects. However, I determine this concern to be less relevant in this context since the treatment is reasonably modest in scale and the necessary logical path from plant closures in Germany to electricity generation in its neighbors is sufficiently indirect.

While DiD inherently controls for time-invariant unobserved factors, it might be biased in the presence of unobserved factors that change over time and are correlated with the treatment.

12

Unfortunately, as this dataset overlaps with the 2022 Russian war in Ukraine, this is a potential concern. To the extent that the war has a differential impact on the treated and control groups, these estimates may be biased accordingly.

3.3.1.3 Methodological Fit

There are a number of well-documented advantages and disadvantages to the DiD method. Foremost among its advantages is simplicity: the straightforwardness of DiD allows for easy interpretation and communication of results. The approach is well-understood and used broadly across disciplines. Additionally, by comparing within and between groups over time, it inherently accounts for any unobserved factors that remain constant over time. However, the DiD method is not without its weaknesses. The reliance on the parallel trends assumption is a potential pitfall; if pre-treatment trends are not similar, my DiD estimates can be biased. Post-treatment events that disproportionately affect the treated and control groups are another source of potential bias. The approach additionally assumes a consistent treatment effect across time, which might not be the case in certain scenarios. In the next section, I describe an alternative approach to causal inference that addresses some of these concerns.

3.3.2 Synthetic Control Method

An alternative approach that can relax the parallel trends assumption is the synthetic control method. This method constructs a synthetic control unit as a weighted average of the control units, where the weights are chosen to make the synthetic control unit as similar as possible to the treatment unit in terms of the pre-treatment trend and other observed characteristics. Given the unique treatment event in Germany—the closure of nuclear power plants in 2021— the traditional SCM approach provides a compelling framework to estimate the causal effect by comparing the post-treatment outcomes of Germany to those of the synthetic control.

3.3.2.1 Model Specification

The synthetic control is constructed using a combination of control countries, where the weights assigned to each control country are determined by their pre-treatment characteristics and outcomes. Specifically, let $W = (w_2, w_3, ..., w_m)$ be a vector of weights assigned to each control country, where w_i represents the weight of the ith control country in constructing the synthetic control. The goal is to find a set of weights such that the pre-treatment outcomes and characteristics of the synthetic control closely match those of Germany. Before constructing the synthetic control, the outcome variable and other relevant variables are scaled to lie between 0 and 1. This normalization ensures consistent weight determination across different scales of variables.

Mathematically, the synthetic control for Germany at time t can be represented as:

$$y_t^{syn} = \sum_{i=2}^{m} w_i y_{i,t}$$

where y_t^{syn} is the outcome of the synthetic control at time t, and $y_{i,t}$ is the outcome of the ith control country at time t.

To determine the weights W, I employ a weighted least squares optimization approach. The objective is to minimize the root mean squared error (RMSE) between the observed outcomes of Germany and those of the synthetic control. The weights are constrained such that they sum up to 1 and each weight lies between 0 and 1. This constrained optimization problem is applied using the following function defined in R:

```
synthetic_control <- function(X, y) {
  # ...
  res <- nloptr::slsqp(
    x0 = start,
    fn = loss,
    gr = loss_gradient,
    heq = constraints,
    heqjac = constraints_jacobian,
    lower = lower,
    upper = upper
  )
  coef_ <- res$par
  mse <- loss(coef_)
  # ...
}
```

The distance metric used in the optimization is a weighted distance, allowing different pre-treatment characteristics to be given varying importance when determining the weights. This ensures that more crucial characteristics have a higher influence on the synthetic control construction.

For the SCM to provide unbiased estimates, it's crucial that the pre-treatment outcomes and characteristics used in constructing the synthetic control are not affected by the treatment. It's also essential that the chosen control countries did not experience any significant shocks or interventions during the study period that might confound the estimates.

3.3.2.2 Model Validation

I provide a graphical validation by plotting the residuals, which are the differences between observed outcomes and synthetic control outcomes, over time. This is done both for individual

time points and cumulatively. A good match is identified by a residual plot which shows random fluctuations around zero prior to the treatment and a stable post-treatment trend. Additionally, quantitative measures such as the mean squared error (MSE), mean absolute error (MAE), and RMSE between Germany and its synthetic control during the pre-treatment period are used to assess the fit, with a lower MSE indicating a closer match.

To further assess the significance of the estimated treatment effects, placebo tests are conducted. By iteratively treating each control country as if it were treated, constructing a synthetic control for it, and then estimating the post-treatment effect, the distribution of these placebo effects provides a reference to judge the magnitude and significance of the actual treatment effect observed for Germany.

3.3.2.3 Estimation

With the synthetic control validated, the next step is to estimate the causal effect of the nuclear plant closures on electricity generation in Germany. The causal effect at each time point is calculated as the difference between the observed outcome for Germany and the outcome of the synthetic control.

$$\delta_t = y_{1,t} - y_t^{syn}$$

To understand the overall impact of the nuclear plant closures on electricity generation in Germany for the entire year of 2022, the cumulative treatment effect is computed.

$$\delta = \sum_{t=T_0+1}^{T} (y_{1,t} - y_t^{syn})$$

This provides a comprehensive estimate of the total impact of the nuclear plant closures on electricity generation in Germany.

3.3.2.4 Methodological Fit

An important potential model violation for SCM is the assumed absence of spillover effects. If the treatment in Germany affected the outcomes in control countries, it could violate the assumptions of the SCM. For instance, if the closure of nuclear plants in Germany led to significant changes in electricity prices or policies in neighboring countries, the synthetic control might not provide an accurate counterfactual.

Another assumption is the factor model, which posits that the observed outcomes are linear combinations of observed and unobserved factors. If this assumption doesn't hold, the synthetic control might not be an accurate representation. Furthermore, the method assumes that, apart from the treatment, no other events or shocks differentially affect the treated unit

and the synthetic control in the post-treatment period. Any such events can introduce bias into the estimated treatment effects.

The SCM's ability to incorporate multiple control units to create a synthetic counterpart can lead to improved pre-treatment fit, enhancing the credibility of the counterfactual. Additionally, the method's graphical results are intuitive and easy to interpret.

3.3.3 Bayesian Structural Time Series

As described above, the SCM implementation has a number of constraints, including that the weights applied to control units sum to 1 (Doudchenko and Imbens, 2016). This can pose problems when the unit of interest is an outlier relative to the control units for the variable of interest. This is the case in my case as Germany generates much more electricity from fossil fuel sources than any of the control countries. One way to improve on this method is to use a spike-and-slab prior distribution, as in Brodersen et al. (2015).

I implement a synthesis of the SCM and Brodersen's implementation of a Bayesian structural time series model for flexible yet robust causal inference.

3.3.3.1 Model Specification

The BSTS model captures the trend, seasonality, and influence of covariates on the time series using a state-space framework. The observation equation links the observed electricity generation y_t to the trend component μ_t, seasonal component γ_t, covariates x_t, and error term ϵ_t. The state equation governs how the latent state variables like trend and seasonality evolve over time.

The BSTS model, as implemented in the `CausalImpact` package, is designed to infer the causal impact of an intervention by predicting the counterfactual market response in a synthetic control. It does so by leveraging a state-space model that incorporates various structural components, such as trends, seasonality, and regression effects.

In the context of this study, the BSTS model captures the underlying trend in electricity generation in Germany, periodic fluctuations due to seasonality, and the influence of external covariates. The model can be represented as:

$$y_t = \mu_t + \gamma_t + \beta x_t + \epsilon_t$$

Where:

- y_t is the observed electricity generation in Germany at time t.
- μ_t represents the trend component.
- γ_t represents the seasonal component.

- x_t is a vector of external covariates at time t, and β is the corresponding vector of coefficients.
- ϵ_t is the error term, typically assumed to follow a normal distribution.

This formulation allows the BSTS model to capture the complex dynamics of time series data, making it particularly suited for estimating the causal impact of interventions in observational studies.

In Bayesian analysis, priors play a crucial role in incorporating external information or beliefs about the parameters. For the BSTS model, the choice of priors can be particularly important to ensure that the model is both flexible and interpretable. In the context of this study, I employ a spike-and-slab prior distribution, as suggested by Brodersen et al. (2015). This type of prior is a mixture of two distributions: a point mass at zero (the "spike") and a continuous distribution (the "slab"). The spike-and-slab prior effectively introduces sparsity into the model, ensuring that only relevant predictors are included in the final model. This is especially useful when dealing with multiple potential control units, as it helps in selecting the most influential controls for constructing the synthetic control. This prevents overfitting by allowing only relevant covariates to influence the synthetic control. The spike concentrates probability mass at zero, while the slab is a diffuse normal distribution.

3.3.3.2 Estimation

The BSTS model, as implemented in the `CausalImpact` package, allows for a comprehensive specification that captures the underlying dynamics of the time series data. The model is formulated in a state-space representation, which decomposes the time series into various structural components.

Given the observed data up to time t, the BSTS model predicts the counterfactual outcome for time $t+1$. This prediction is based on the underlying trend, seasonality, and the influence of external covariates. The difference between the observed and counterfactual outcomes provides an estimate of the causal effect of the intervention.

To estimate the causal effect using BSTS, I employ a Markov chain Monte Carlo (MCMC) algorithm. This Bayesian approach provides a full posterior distribution for each parameter, allowing for a comprehensive understanding of the uncertainty associated with the estimates. The MCMC samples from the posterior distribution of the model parameters, given the observed data, and constructs a synthetic control that represents the counterfactual scenario without the intervention.

3.3.3.3 Validation

Validating the fit of the BSTS model is crucial to ensure that the causal estimates are reliable. One approach to validation is to inspect the in-sample fit of the model. By comparing the model's predictions to the observed data during the pre-intervention period, I can assess how well the model captures the underlying dynamics of the time series.

Another essential aspect of model validation is the assessment of the model's residuals. Ideally, the residuals should be white noise, indicating that the model has adequately captured all the information in the data. Tools like autocorrelation and partial autocorrelation plots can provide insights into the nature of the residuals and indicate whether there's remaining structure that the model hasn't captured.

3.3.3.4 Methodological Fit

One of the primary concerns with BSTS is overfitting. Given its flexibility, the BSTS model can fit the training data very closely, potentially capturing noise rather than the underlying signal. This can lead to overly optimistic in-sample predictions but poor out-of-sample performance. Additionally, the assumption that the residuals are white noise is crucial for the validity of the causal estimates. If there's remaining structure in the residuals, it indicates that the model hasn't captured all the relevant information, potentially leading to biased estimates.

In contrast with the DiD and SCM approaches, BSTS provides a flexible framework that can capture the underlying dynamics of the time series data, including trends, seasonality, and the influence of external covariates. This means that the counterfactual predictions are based on a more comprehensive representation of the information available for analysis. The BSTS approach better accommodates trend, seasonality, and outliers through its state-space framework. Automatic variable selection avoids having to specify covariates a priori. The Bayesian approach also provides more comprehensive uncertainty quantification via posterior samples.

3.3.4 Gaussian Process Regression

3.3.4.1 Model Specification

I model the electricity generation $y_{i,t}$ for country i at time t using a Gaussian process regression model:

$$y_{i,t} = f(X_{i,t}) + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim \mathcal{N}(0, \sigma^2)$$

where $f(\cdot)$ represents an unknown function and $\epsilon_{i,t}$ is Gaussian noise. I place a Gaussian process (GP) prior on the function $f(\cdot)$:

$$f(\cdot) \sim \mathcal{GP}\left(0, k(X, X')\right)$$

The covariance function $k(X, X')$ is defined as:

$$k(X, X') = k_{exp}(X, X') + k_{white}(X, X')$$

where $k_{exp}(X, X')$ is the exponential kernel:

$$k_{exp}(X, X') = \sigma_f^2 \exp\left(-\frac{1}{2\Lambda^{-2}}(X - X')^T(X - X')\right)$$

and $k_{white}(X, X')$ is the white noise kernel. The exponential kernel captures smooth trends while the white noise kernel accounts for noise in the data.

Using the `GPy` python package, I optimize the kernel hyperparameters using only the pre-treatment data to learn the underlying patterns without being influenced by the post-treatment outcomes. A key innovation is using automatic relevance determination (ARD) for input selection as documented in Neal (1996). Here, $\Lambda = \text{diag}(\ell_2, ..., \ell_m)$ is a diagonal matrix containing the characteristic lengthscale parameter ℓ_i for each input i:

$$\ell_i = \frac{1}{\text{FI}_i}$$

Where FI_i is the feature importance of input i. Shorter lengthscales imply greater influence on the GP. I use this to remove unnecessary inputs, avoiding overfitting.

3.3.4.2   Estimation

Once we've identified the relevant features using ARD, the next step involves training the Gaussian Process Regression model. This training process is crucial for ensuring that my model captures the underlying patterns in the data and can make accurate counterfactual predictions.

I find that the White kernel defined above is necessary in this context in order to avoid overfitting. This is likely due to the fact that the data is relatively sparse, with only a relatively small number of observations for each country. The White kernel allows the model to represent the noise in the data, which is especially important given the small sample size.

I begin by fitting the GP model to the pre-treatment data. After the initial fit, I employ time series cross-validation to fine-tune the model's hyperparameters—most importantly, bounds imposed on the lengthscale and noise parameters. This step is essential for two reasons: by holding out portions of my data and testing the model's predictions against these holdouts, I can assess how well my model is likely to perform on unseen data. Second, I'm able to test a large number of hyperparameter combinations to find the best model performance without overfitting to the training data. This is especially important given the relatively small sample size in this study. After selecting hyperparameter values, I do a preliminary evaluatation of model performance by fitting a model using the first 80% of the pre-treatment data and testing the model's predictions on the remaining 20%. This provides a sense of how well the

model generalizes to unseen data without exposing it to the post-treatment outcomes, which I don't necessarily expect the model to capture.

Once I'm ready to fit the final model, I train on the pre-treatment period. With the model trained and the relevant controls identified, I proceed to the crux of my analysis: estimating the causal effect of the nuclear plant closures. I achieve this by making counterfactual predictions for the post-treatment period.

Using the optimized hyperparameters, θ^*, I draw samples from the GP posterior predictive distribution to make predictions on the post-treatment data, making predictions about the counterfactual electricity generation for Germany, $y_{1,t>t_0}(0)$:

$$p(\tilde{y}_{1,t>t_0}|X,y,\theta^*) = \int p(\tilde{y}_{1,t>t_0}|f,X,\theta^*)p(f|X,y,\theta^*)df$$

The estimated treatment effect is then derived by comparing these counterfactual predictions, $\tilde{y}_{1,t>t_0}(0)$, to the observed post-treatment outcomes, $y_{1,t>t_0}(1)$. This comparison provides a quantification of the causal impact of the nuclear plant closures on electricity generation. The Bayesian approach provides uncertainty estimates on these predictions.

Overall, GPR offers a flexible framework by placing a probability distribution over functions without strong assumptions. The ARD technique enables data-driven input selection to prevent overfitting. And the Bayesian nature quantifies uncertainty in a principled way. This makes GPR well-suited for estimating treatment effects in this analysis.

4 Results

My analysis begins with a simple exploration of the raw data.

Table 1: Average monthly electricity generation from fossil fuels (TWh) by country

Country	2015	2016	2017	2018	2019	2020	2021	2022
Austria	0.71	0.73	0.85	0.75	0.87	0.73	0.73	0.78
Belgium	1.79	1.63	1.69	1.73	1.80	1.82	1.50	1.59
Bulgaria	1.71	1.54	1.66	1.58	1.51	1.23	1.58	1.88
Czech Republic	2.92	3.10	3.04	3.00	2.89	2.55	2.75	2.75
Denmark	0.93	1.09	0.68	0.76	0.57	0.51	0.73	0.63
Estonia	0.54	0.64	0.68	0.63	0.33	0.20	0.31	0.33
Finland	1.03	1.08	0.99	1.10	0.94	0.71	0.81	0.67
France	2.50	3.30	3.90	2.80	3.00	2.76	2.82	3.44
Germany	16.81	17.30	16.44	18.01	15.36	13.23	15.11	15.20
Greece	1.93	2.00	2.37	2.15	1.98	1.76	1.90	1.74
Hungary	0.73	0.78	0.84	0.79	0.83	0.85	0.83	0.77
Latvia	0.14	0.16	0.10	0.18	0.19	0.12	0.13	0.08
Lithuania	0.15	0.08	0.05	0.03	0.04	0.11	0.09	0.06
Luxembourg	0.03	0.03	0.02	0.02	0.03	0.02	0.02	0.02
Netherlands	3.59	5.01	4.93	4.67	5.29	5.40	4.68	4.33
Norway	0.21	0.21	0.21	0.22	0.21	0.16	0.07	0.11
Poland	9.47	9.42	9.47	9.57	8.98	8.47	9.70	9.24
Portugal	1.66	1.64	2.22	1.82	1.50	1.33	1.07	1.16
Romania	1.96	1.85	1.98	1.88	1.64	1.41	1.51	1.50
Slovakia	0.24	0.25	0.27	0.28	0.34	0.34	0.41	0.25
Slovenia	0.31	0.35	0.33	0.32	0.31	0.31	0.28	0.24
Spain	7.09	6.00	7.42	6.49	6.57	5.03	4.98	6.30

Germany, central to my research question, displayed only a marginal increase in fossil fuel electricity generation post the closure of its last nuclear plants in December 2021, moving from 15.11 TWh in 2021 to 15.20 TWh in 2022. Given the size of the pandemic-related slowdown and the sizable amount of nuclear energy that was phased out, this minute increment suggests early on that Germany either turned to alternative energy sources, potentially renewables, or made significant gains in energy efficiency. Strikingly, the 2022 figure remains well below every pre-pandemic year for which I have ENTSO-E data. In contrast, France experienced a conspicuous surge in fossil fuel electricity generation, ascending from 2.82 TWh in 2021 to 3.44 TWh in 2022, indicating a shift that may have been influenced by external factors or domestic policy changes. Another nation that stands out is Bulgaria, which underwent a marked jump in its fossil fuel generation from 1.58 TWh in 2021 to 1.88 TWh in 2022, illustrating the potential for changes in energy policies or an upswing in industrial demands. Looking at the table as a whole, it's also noteworthy that Germany is a signficant outlier for fossil fuel electricity generation, with the next highest country, Poland, generating only 60% of Germany's 2022 figure.

Proceeding in order of complexity, I proceed to estimate the causal effect of the nuclear plant closures using the DiD, SCM, BSTS, and GPR models.

Difference-in-Differences

To evaluate the parallel trends assumption, I normalize the outcome variable by country and plot the difference between each potential control unit and my treatment unit. The resulting plot shows the difference between each country's electricity generation and Germany's electricity generation in the pre-treatment period. The parallel trends assumption holds if the differences are relatively stable over time.
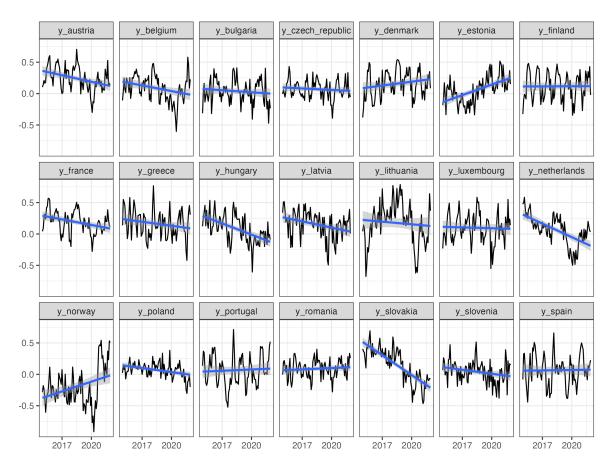

Figure 2: Normalized pre-treatment differences between each potential control unit and Germany

For each country, I fit a linear model for the pre-treatment differences using the time variable as the predictor. I use the resulting coefficients as a measure of a potential control unit's compatibility with the DiD model assumptions. I remove countries with a non-zero slope at the 5% significance level, leaving Bulgaria, Czech Republic, Denmark, Finland, Greece, Lithuania, Luxembourg, Portugal, Romania, and Spain in the final donor pool. The parallel trends assumption is thus determined to hold for these countries, and they are therefore suitable for use as potential controls.

Next, I fit the DiD model and show the resulting model coefficients in Table 2.

Table 2: DiD model coefficients

term	estimate	std.error	statistic	p.value
(Intercept)	2402867.02	361894.05	6.64	0.000
Treatment Period x Treatment Group	-567945.41	296867.90	-1.91	0.056
Treatment Period	-112543.19	89534.48	-1.26	0.209
Treatment Group	19179878.71	544182.08	35.25	0.000
Germany x Temperature	-91970.35	10219.00	-9.00	0.000
Spain x Temperature	50752.92	9924.94	5.11	0.000
Lithuania	-2333068.99	479990.54	-4.86	0.000
Luxembourg	-2349940.05	568087.81	-4.14	0.000
Czech Republic	2018301.34	524899.38	3.85	0.000
Spain	1729496.47	574023.38	3.01	0.003
Portugal x Temperature	42196.52	14323.88	2.95	0.003
Portugal	-2523126.02	880601.84	-2.87	0.004
Temperature	-15571.33	6788.85	-2.29	0.022
Lithuania x Temperature	16038.99	9459.30	1.70	0.090
Greece x Temperature	16905.74	10275.56	1.65	0.100
Czech Republic x Temperature	-14755.40	9968.97	-1.48	0.139
Luxembourg x Temperature	15266.66	10842.30	1.41	0.159
Finland x Temperature	-13375.59	9701.81	-1.38	0.168
Denmark	-507433.46	574284.38	-0.88	0.377
Greece	-497686.06	625646.10	-0.80	0.427
Denmark x Temperature	-7748.78	11179.51	-0.69	0.488
Finland	-213299.11	480201.95	-0.44	0.657
Romania x Temperature	1829.73	9120.90	0.20	0.841
Romania	88616.49	503821.81	0.18	0.860

I show for the DiD estimator of the treatment effect an estimate of -0.57 (p-value = 0.056) with a 95% confidence interval of (-1.15, 0.01). Extrapolated over a year (noting that the true effect is likely not actually constant over time), this would imply that on average, the nuclear plant closures led to a 6.82 TWh decrease in annual fossil fuel electricity generation in Germany. This seems at first glance a modest-sized effect—equal to 3.74% of the country's total electricity generation from fossil fuels in 2022 and 1.57% of the country's total electricity generation from all sources in the same year.

However, the 95% confidence interval is also quite wide and includes zero. This suggests that the DiD estimator is not very precise and that the true treatment effect could be much larger or smaller than the point estimate. Surprisingly though, since the upper bound of the intervals is only just above zero, the results lend weak evidence to there having been not been an increase in fossil fuel electrcity generation following the plant closures. The model doesn't appear to be a poor fit—the adjusted R-squared value is 0.959, meaning the temperature control and country fixed effects capture a large amount of the variability in the outcome variable. However, on inspection of a plot of the fitted values, it's apparent that the model fit varies considerably by country. The model's residuals show some structure, indicating that the model may not be capturing all the relevant information in the data. This is to be expected, given the model's simplicity and the complexity of the underlying dynamics.

I additionally show the DiD predicted and observed values in Figure 3 with the start of the treatment effect (the date of the 2021 nuclear plant closures) denoted with a dotted line.
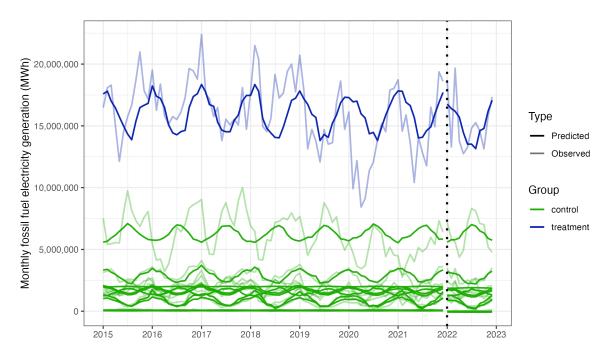

Figure 3: Predicted values from DiD model

Finally, I perform a sensitivity analysis to assess the robustness of the DiD model by removing one country at a time from the donor pool and refitting the model. I then compare the resulting treatment effect estimates to the original estimate. I find these hold-one-out estimates to be quite stable, all falling within around 10% of the original estimate. This means that the DiD model is relatively insensitive to the particular composition of the donor pool, which is a positive sign for the model's robustness.

Synthetic Control Method

The synthetic control method provides an alternative approach to estimating the treatment effect. I construct a synthetic Germany as a weighted average of other European countries that matches the pre-treatment electricity generation trends.

The optimization routine produces weights for each control candidate country and the subset of temperature measurements that correspond to Germany. The results are shown in Table 3.

I find satisfyingly intuitive results from the optimization routine. The countries with the highest weights are geographically proximate to Germany—a characteristic I would expect from countries which share similar generation trends, as climate is one of the most important factors in determining electricity demand. The optimization routine places the highest weight

Table 3: Weights for each control candidate country obtained from SCM optimization routine

Country	Weight
Poland	0.386
Slovenia	0.197
Estonia	0.156
Czech Republic	0.141
Norway	0.068
Belgium	0.052
Slovakia	0.000
Spain	0.000
Finland	0.000
Netherlands	0.000
Latvia	0.000
Hungary	0.000
Austria	0.000
Bulgaria	0.000
Denmark	0.000
France	0.000
Greece	0.000
Lithuania	0.000
Luxembourg	0.000
Portugal	0.000
Romania	0.000

on Poland at 0.39, with the next highest weight going to Slovenia at 0.2. The weights for most of the remaining countries are pushed nearly to zero.

As shown in Figure 4, the model shows a reasonable fit to the German generation trend over the pre-treatment period.

In the post-treatment period, I observe a somewhat different result than found using DiD. As shown in Figure 5 and Figure 6, the trajectories of observed and synthetic Germany diverge, with observed Germany showing a modest increase relative to the synthetic counterfactual.

The estimated causal effect accumulates to 13.04 TWh in total over the 12 post-treatment months. This implies the nuclear phase-out increased fossil fuel generation in Germany by TWh on average per month in 2022. The estimated effect appears constant over time. This seems likely to be a model artifact, as a large majority of residuals from the treatment period are positive.

Placebo tests were performed by randomly assigning the treatment to each of the other countries in the donor pool and refitting the model. These tests using other countries show the post-treatment RMSE for Germany is not unusually high, yielding a p-value of 0.82. Despite the sign change from the previous approach, this lends additional evidence to the conclusion that the observed effect for Germany is not statistically meaningful and plausibily spurious. Overall, the causal analysis so far has not indicated the nuclear phase-out led to a meaningful increase in fossil fuel electricity generation in Germany during the first year.
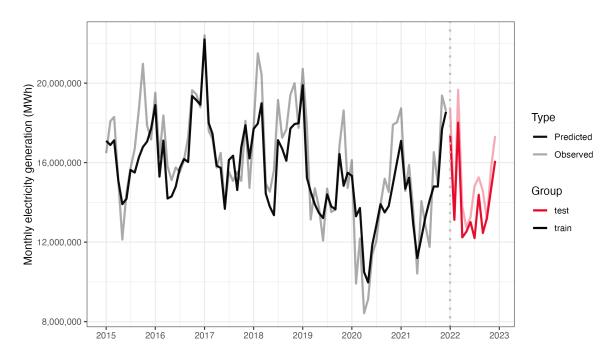
Figure 4: Predicted values from SCM model

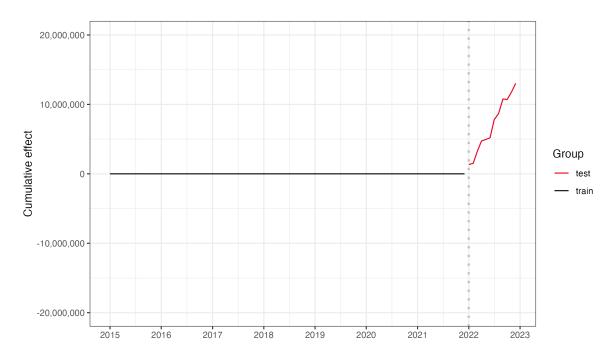Figure 5: Treatment effect estimated using SCM model

Figure 6: Cumulative treatment effect estimated using SCM model

Bayesian Structural Time Series

As stated above, I implement the Bayesian structural time series model using the `CausalImpact` R package. I show the model's predicted values compared with the observed values in Figure 7. Like the SCM implementation, the model predictions appear to follow the general trend in the data during the pre-treatment period—now complemented with the notable benefit of Bayesian uncertainty intervals.

In the post-intervention period, Germany's electricity generation from fossil fuels averaged approximately 15.2 TWh. With `CausalImpact`, I estimate that if Germany's nuclear plants had not been closed, generation would have averaged 15 TWh. The difference between these values, 0.2 TWh, represents the estimated causal effect of the nuclear plant closures. While this suggests that the closures led to an increase in fossil fuel electricity generation, the 95% prediction interval ranges from -0.8 TWh to 1.1 TWh, indicating a high level of uncertainty around this point estimate.

Expressed in relative terms, the increase amounts to approximately 1%, with a 95% prediction interval ranging from a decrease of -5% to an increase of 8%. This wide range underscores the inherent uncertainties associated with the causal inference process in this context. While the results seem to indicate a positive effect of the nuclear plant closures on fossil fuel electricity generation, the effect is not statistically significant over the entire post-intervention period, as the p-value stands at 0.01. Such a finding suggests that such a result could plausibly be the result of random variability. This is visible in Figure 8 and Figure 9, which show that
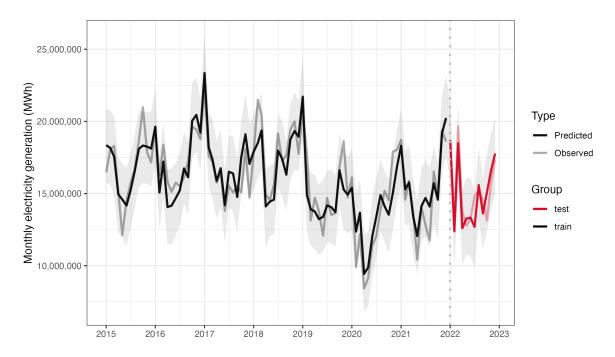
27

Figure 7: Predicted values from BSTS model

the uncertainty intervals for the pointwise and cumulative treatment effects include zero.

Gaussian Process Regression

In implementing the GPR approach, I adopt a combination of exponential and white noise kernels. The choice was made given the assumption that the data might exhibit smooth trends but might also contain noise requires explicit accounting for. A series of constraints were applied to the kernel hyperparameters, ensuring the stability and interpretability of the model's results.

After loading the data and preprocessing, I scaled the dataset between 0 and 1 to ease the optimization process. The data was divided into pre-treatment (for training) and post-treatment (for testing) periods, respecting the date of the nuclear plant closures in 2021 as the pivotal moment.

To construct the initial GPR model, I utilized all available features, which included observations for the variable of interest for each country along with temperature measurements for Germany. The exponential kernel's length scales and the variances of both exponential and white noise kernels were bounded to prevent both overfitting and overflow errors, and to ensure the results remained within an interpretable range. After fitting the model on the entire pre-treatment dataset, a preliminary assessment revealed the model's performance on both training and test data, which was quantified using RMSE.
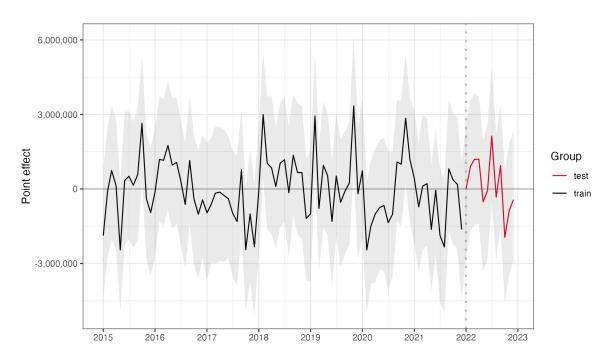
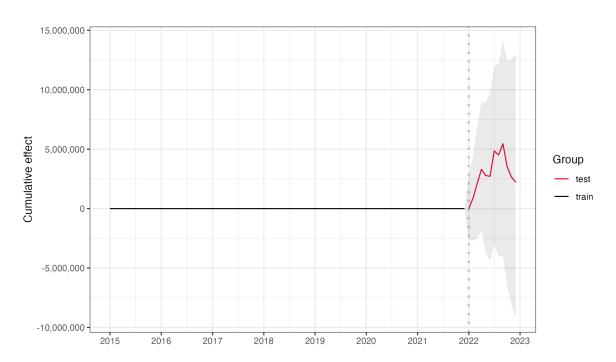Figure 8: Treatment effect estimated using BSTS model

Figure 9: Cumulative treatment effect estimated using BSTS model

Subsequently, I prioritized the exploration of feature importance. GPR provides a natural, if unitless, way to assess the relevance of different predictors via the length scales of the kernel. Features with shorter length scales have more impact on predictions, whereas features with longer length scales have less impact. The cumulative importance of features was calculated, and I observed that only a subset of the initial features was required to produce nearly 100% of the model's predictive power.

Based on this observation, I refined the GPR model by incorporating only those features that cumulatively contributed significantly to the model's explanatory power. A second round of constraints was identified and applied to the kernel hyperparameters to tailor the model to this reduced set of features. This streamlined model was then re-optimized and assessed on both the training and testing datasets.

Visual representations of the GPR predictions for Germany's fossil fuel electricity generation were created. In Figure 10 the model's predictions are shown alongside observed values and 95% credible intervals produced using samples from the posterior predictive distribution. It was more clear than with any other model that GPR replicated the seasonal patterns and the overarching trend in the data.
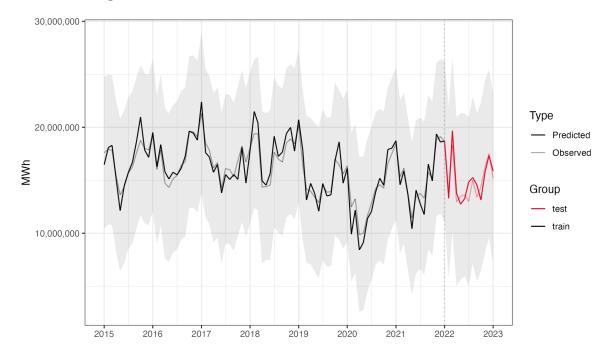

Figure 10: Predicted values from GPR model

However, the results tell much the same story as before. The model estimated the average causal effect to be 0.19 TWh. This corresponds with an estimated cumulative effect of 2.44 TWh over the 12-month post-treatment period. As before, zero is well inside the credible intervals for both values.

Sensitivity analyses were additionally performed. Various features were sequentially omitted
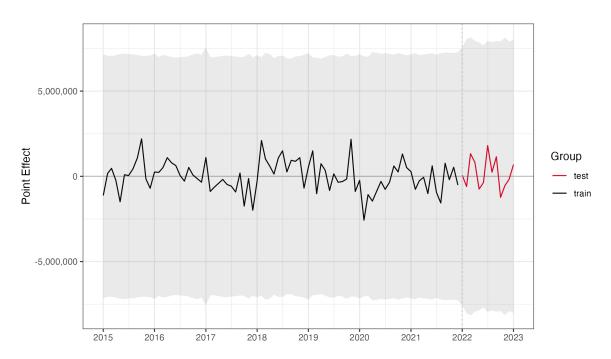
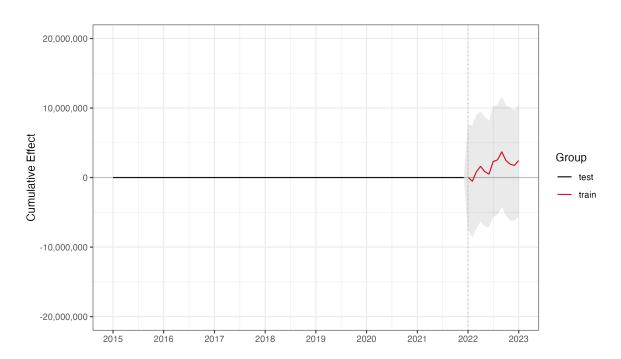Figure 11: Treatment effect estimated using GPR model

Figure 12: Cumulative treatment effect estimated using GPR model

to gauge the robustness of the model's predictions. These exercises indicated that the GPR model's estimates were consistent, thereby reinforcing my confidence in its capability to offer meaningful insights into the effects of the nuclear power plant closures on fossil fuel electricity generation in Germany.

4.0.1 Summary

Table 4 compares the out-of-sample predictive performance of the four causal inference methods on both the pre-treatment and post-treatment data. Predictive accuracy is quantified using the root mean squared error metric.

On the pre-treatment data, Gaussian process regression achieved the lowest RMSE, indicating it most accurately captured the underlying generative process during this period. The SCM and BSTS models attained higher but comparable errors. As expected, the simplicity of the differences-in-differences approach resulted in the poorest out-of-sample fit.

The relative predictive accuracy remains consistent in the post-treatment period. Gaussian process regression again outperformed the other techniques with the lowest RMSE of 904,171. BSTS saw only a marginal increase in error to 1.09 million. SCM and DiD experienced more substantial deteriorations in fit to approximately 1.35 and 2.47 million, potentially indicative of their stronger modeling assumptions.

Overall, the GPR model's superior predictive performance underscores the value of flexible nonparametric modeling for this time series forecasting task. BSTS also provides reasonable fits, lending support for its state space approach to distinguishing trends and seasonality. The SCM and DiD methods appear inadequately flexible to accurately extrapolate patterns beyond the training data.

Table 4: Performance comparison of causal inference methods

Method	Treatment	RMSE
GPR	pre	904171.1
BSTS	pre	1248023.1
SCM	pre	1409268.7
DiD	pre	1856929.6
GPR	post	896523.6
BSTS	post	1088262.8
SCM	post	1349784.9
DiD	post	2468849.8

Turning to the Table 5, it is apparent that the four methods produce inconsistent point estimates of the treatment effect. This likely reflects their different modeling assumptions and biases. However, it is noteworthy that the estimates broadly span a range centered around a minor positive or negative effect.

GPR, and BSTS yield modest positive point estimates for the mean monthly effect of 188,000 and 438,000 MWh. SCM gives a higher value of 1.1 million MWh. DiD appears the outlier at -568,000 MWh. Examining the cumulative effects tells a similar story.

Given the uncertainties involved, placing emphasis on the precise point estimates risks overinterpreting noisy signals. But the overall convergence around marginal effects provides mutual corroboration. This triangulation across varied methods strengthens confidence in the conclusion that Germany's nuclear phase-out did not drastically increase fossil fuel reliance over the examined time horizon.

Table 5: Estimate of causal effect from each model

Model	Estimate	Value
DiD	mean	-567945.4
BSTS	mean	184908.5
GPR	mean	187737.0
SCM	mean	1086281.9
DiD	cumulative	-6815345.0
BSTS	cumulative	2218901.6
GPR	cumulative	2440581.1
SCM	cumulative	13035382.4

4.1 Discussion

The results obtained from the four methods showcase some notable contrasts, while also reinforcing a consistent narrative regarding the effects of Germany's nuclear phase-out. Across the board, the estimates imply at most a marginal impact on fossil fuel electricity generation in the year following the closure of the country's remaining reactors. However, differences emerge in the precision and certainty ascribed to these findings.

4.1.1 Causal Effect

The simplicity of the differences-in-differences approach yields an interpretable point estimate suggesting a minor decrease in fossil fuel generation post-intervention. Yet the wide confidence interval underscores the uncertainty in quantifying the effect. The parallel trends assumption appears reasonably satisfied, but the model does not account for time- and potentially space-varying confounders like the Ukraine war that could bias results. The synthetic control method provides improved fit over the pre-treatment period. But the estimated positive effect seems spuriously constant, and the method lacks inherent uncertainty quantification.

The Bayesian structural time series model adds more flexibility through its latent components. The predictive intervals widen compared to the point estimates from the previous models, accurately reflecting the uncertainties involved. However, its linear regression basis may not capture intricate temporal dynamics as effectively as nonparametric alternatives. This is evidenced by its weaker predictive performance. Gaussian process regression combines flexible nonparametric modeling with principled Bayesian uncertainty characterization. The automatic relevance determination technique for input selection prevents overfitting. The end result is an interpretable causal effect estimate with credibility intervals that hedge against overconfident conclusions.

4.1.2 Feature Selection

A key differentiator between methods is how they perform feature selection, or determine which control units and covariates to include in the model. DiD uses pre-trends and statistical tests to filter control countries. But this may still include extraneous variables. The SCM optimization provides intuitive weighting, but lacks uncertainty and may overfit. GPR's automatic relevance determination technique offers a more robust data-driven approach. By learning the lengthscales as feature importance measures, it incorporates only predictors vital to the model's fit. This prevents overfitting while retaining flexibility. Moreover, the reduced feature set lightens computational demands and enhances interpretability without sacrificing performance. The GPR model with ARD-based feature selection thus demonstrates the power of principled Bayesian approaches for input variable selection in causal inference.

The differences-in-differences methodology filters control units based on statistical tests for pre-treatment parallel trends. However, while removing countries with significantly different pre-period outcomes guards against violations of a key assumption, this approach risks excluding useful information. The binary decision of retaining or dropping units ignores nuances in their potential similarities. DiD does not leverage continuous measures of pre-treatment fit when selecting controls. In contrast, the SCM directly incorporates fit via the weight optimization. But its hard thresholding of choosing only the best fitting units can lead to overfitting and spurious matches. The GPR model with ARD takes a more graded approach. By learning a continuous importance measure through the lengthscales, it soft thresholds inputs based on their explanatory power. Less relevant inputs are not discarded but simply downweighted. This retains useful signals without overfitting to noise.

The probabilistic nature of GPR's feature selection allows for principled uncertainty characterization regarding the relevance of inputs. In a frequentist paradigm like DiD or SCM, covariates are either included or excluded based on thresholding. This binary view cannot capture uncertainties. But GPR's Bayesian derivation provides posterior credibility intervals around lengthscale estimates. Narrower intervals indicate higher confidence in an input's low or high importance. Wider intervals reflect greater ambiguity. This principled measure of uncertainty is lost in traditional feature selection but adds valuable nuance regarding the role of different predictors. Overall, GPR with ARD highlights the advantages of Bayesian modeling for input variable selection - offering a more robust, flexible, and uncertainty-aware approach than alternatives like pre-trend filtering or weight optimization.

Figure 13 shows a somewhat crude comparison of the different methods used for feature selection. The portion labeled DiD displays bars at heights one and zero for countries which have been included or excluded from the donor pool, respectively. The SCM portion shows the weights assigned to each country by the optimization routine. The BSTS portion shows the inclusion probability for each country, calculated by `CausalImpact`. The conditional probability that a coefficient is positive is also available, but it is not shown here for simplicity. The GPR portion shows the inverse lengthscales of the kernel for each input variable. The inverse lengthscales are scaled between zero and one for ease of comparison.

Viewed all together, a somewhat gratifying picture emerges. The methods appear to agree on
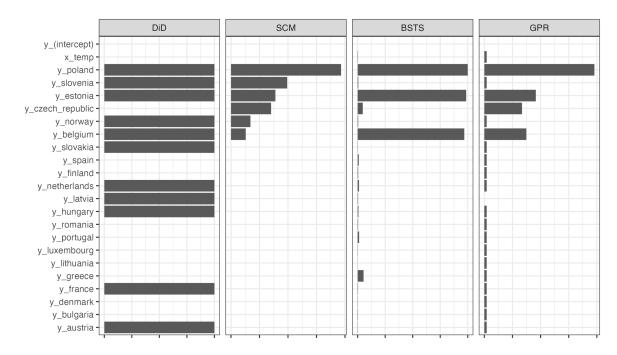
Figure 13: Feature selection comparison

the "importance" of the most proximate countries to Germany. Poland, Estonia, and Belgium are assigned high importance by all four methods. Curiously the temperature covariate is assigned low importance by all models. The SCM and DiD models do not explicitly incorporate temperature measurements, but they do include countries with similar climates to Germany.

4.1.3 Uncertainty

The marginal positive effects seen in some models could be spurious rather than indicative of an actual post-intervention increase in fossil fuel generation. Several factors support this conservative interpretation. The uncertainties around the point estimates are substantial, as revealed by DiD and GPR—despite models seeming relatively robust to changes in thier specifications, as visible under leave-one-out perturbations of the synthetic control. Known contemporaneous shocks like the Ukraine war and gas supply constraints likely impacted the dynamics in ways not fully captured by the models. In any case, the apparent lack of an unambiguous rise in fossil fuel generation contrasts with the aftermath of Germany's 2011 nuclear phase-out decision. Several factors may explain this divergence. In the interim, investments in renewables expanded Germany's wind and solar capacity, potentially offsetting the nuclear shortfall. Market liberalization enabled greater cross-border power trading to meet demand. Energy efficiency policies may have tempered electricity requirements. Furthermore, the plants shuttered in 2021 accounted for a smaller share of supply versus 2011's

immediate closure of 8 reactors. The lead time could have allowed for gradual adaptation. Lastly, exogenous events like the pandemic and Ukraine war make discerning the isolated effect of the closures difficult. Disentangling these myriad influences necessitates a holistic examination beyond the models employed presently. Ultimately, the hypothesis that nuclear reductions inexorably increase fossil fuels requires re-evaluation given Germany's changing energy landscape. A future approach could incorporate a richer set of covariates to better capture the complex dynamics at play; another approach could be using a method like multi-task Gaussian process regression to jointly model the effects of the nuclear phase-out on different energy sources—particularly renewables and imports—which could provide a more complete picture. More granular data or more powerful modeling tool will allow for more definitive claims around the closure effects amidst these compounded uncertainties. But the current results demonstrate the feasibility of data-driven causal inference on this practically relevant issue.

5 Conclusion

5.1 Summary of Key Findings

This study aimed to estimate the causal effect of Germany's 2021 nuclear power plant closures on the country's electricity generation from fossil fuel sources during 2022. The analyses implemented four complementary causal inference techniques - differences-in-differences, synthetic control modeling, Bayesian structural time series, and Gaussian process regression - on a panel dataset of European countries' electricity generation over time.

The results consistently point to a marginal impact, at most, of the 2021 nuclear shutdowns on fossil fuel reliance in Germany. The differences-in-differences model yielded a point estimate implying a minor decrease in fossil fuel generation following the closures. However, wide confidence intervals indicated substantial uncertainty around quantifying the effect precisely. The synthetic control method provided improved pre-treatment fit but overestimated post-treatment effects as implausibly constant over time. It also lacked inherent uncertainty characterization. Bayesian time series modeling added more flexibility through latent components and more appropriate predictive intervals. Gaussian process regression combined nonparametric flexibility with principled Bayesian uncertainty estimates. Its automatic relevance determination technique for input selection warded off overfitting concerns.

Across techniques, the nuclear phase-out does not appear to have caused a statistically significant increase in fossil fuel-based electricity generation during the first year. While marginal positive effects emerged in some models, uncertainty levels give reason to doubt these represent true effects rather than spurious signals. The results appear relatively robust to perturbations in model specifications. Overall, the hypothesis that nuclear reductions inexorably increase fossil fuel reliance requires re-evaluation given Germany's changing energy landscape.

5.1.1 Substantive and Methodological Implications

Substantively, the findings provide insights relevant to nuclear policy and Germany's energy transition, while warranting prudent qualification. The lack of clear evidence that phasing out nuclear generation increased fossil fuel reliance contrasts with some expectations. However, Germany's investments in renewables and interconnected grid may represent an overly optimistic case. Nuclear closures could still risk increasing emissions in less prepared systems. Policymakers should therefore update beliefs on the interconnectedness of nuclear, renewables, and fossil fuels cautiously. Though the results suggest possible pathways to mitigate trade-offs, phasing out low-carbon nuclear energy in isolation likely remains inadvisable from a climate perspective.

Methodologically, applying multiple complementary causal inference techniques proved illuminating. While all methods have limitations, their combined perspectives lent greater confidence in the consistency of the findings. Gaussian process regression demonstrated particular value through its flexible, nonparametric modeling and automatic relevance determination

for input selection. It signficantly improved on other methods in capturing the variability in the outcome of interest. Its principled uncertainty quantification avoided overconfident conclusions. The analysis elucidated the relative merits of differences-in-differences, synthetic controls, Bayesian time series, and Gaussian processes for inference on observational time series data. It exemplified how leveraging their respective strengths provides a richer understanding than any single method alone.

5.1.2 Limitations and Future Work

While the study provides valuable insights, it has limitations that could be addressed by future work. The data consists of country-level electricity generation at a monthly resolution. More granular data below the country level could support analysis of regional dynamics within countries (at higher frequencies, signals within the data become difficult to detect as the noise increases considerably). The time period available also constrains the pre-treatment window for training models—resource-intensive approaches as in Jarvis et al. (2019) can help bypass limitations on data availability in this case. Additionally, the set of predictor variables is limited to only one, hampering control of potential confounders. This analysis faced the challenge of considering a treatment effect with broad-reaching impacts, disqualifying many traditionally valuable covariates for this context, such as electricity prices, demand levels, and power imports. All of these would certainly have changed in response to the 2021 power plant closures and so could not act as controls within this framework.

The models themselves also have shortcomings, which have been discussed in the relevant sections. The differences-in-differences estimates could suffer from biases if time-varying confounders affected the parallel trends assumption. The synthetic control approach risks overfitting and assumes no spillovers between units. The Bayesian structural time series as implemented in `CausalImpact` relies on modeling assumptions about latent components and assumes that the "relationship between covariates and treated time series, as established during the pre-period, remains stable throughout the post-period." Gaussian process regression faces computational burdens scaling to massive datasets. Applying recent advances like deep Gaussian processes could alleviate this.

Future work could strengthen the analysis along multiple dimensions. Region- or even plant-level data on electricity generation, prices, and external drivers would support more granular modeling. Expanding the panel to include additional nuclear-phase out countries could bolster synthetic controls. Incorporating richer sets of covariates and applying methods robust to observed confounders like regression discontinuity could isolate the effects of interest. Shifting to a multi-task modeling framework—for example, multi-task Gaussian processes—could capture treatment effects across energy sources, providing a more holistic view. Overall, the limitations of data and methods leave room for significant analytical refinements to better disentangle the complex dynamics around nuclear energy transitions. But the feasibility demonstrated points towards productive avenues for further investigation.

5.1.3 Concluding Remarks

This study estimated the causal effect of Germany's 2021 nuclear power plant closures on the country's fossil fuel-based electricity generation in 2022. The results across four complementary causal inference techniques consistently point to a marginal impact, at most. While limitations of the data and models necessitate judicious interpretation, the findings challenge expectations that phasing out nuclear generation inevitably increases fossil fuel reliance in the short term. Germany's changing energy landscape underscores the complexity of managing transitions away from nuclear power.

The analysis elucidates how leveraging multiple methods provides richer insights than any one approach alone. It particularly highlights the value of nonparametric Bayesian modeling with automatic relevance determination for input selection. The feasibility of data-driven causal inference on this empirical issue demonstrates promise for strengthening future investigations. With care taken in extrapolating the context-specific implications, the study offers lessons for both academic literature and policy discussions around nuclear power transitions. While substantial questions remain unresolved, the multi-faceted perspectives provided by complementing flexible modeling with principled uncertainty quantification move constructively towards illuminating the interconnected dynamics shaping decarbonization pathways.

Bibliography

Abadie, A. et al. (2010) Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association.* [Online] 105 (490), 493–505. [online]. Available from: https://doi.org/10.1198/jasa.2009.ap08746.

Abadie, A. & Gardeazabal, J. (2003) The economic costs of conflict: A case study of the basque country. *American Economic Review.* [Online] 93 (1), 113–132. [online]. Available from: https://www.aeaweb.org/articles?id=10.1257/000282803321455188.

Alkousaa, R. (2023) Atomic angst over? Germany closes last nuclear plants. *Reuters.* [online]. Available from: https://www.reuters.com/world/europe/atomic-angst-over-germany-closes-last-nuclear-plants-2023-04-14/.

Alquthami, T. et al. (2022) A performance comparison of machine learning algorithms for load forecasting in smart grid. *IEEE Access.* PP1–1. [online]. Available from: https://api.semanticscholar.org/CorpusID:248463707.

Anon (2021) Renewable energy. United Nations [online]. Available from: https://www.un.org/en/climatechange/raising-ambition/renewable-energy (Accessed 1 September 2022).

Binder, M. et al. (2001) *Green industrial restructuring: International case studies and theoretical interpretations.* Springer Berlin.

Brodersen, K. H. et al. (2015) Inferring causal impact using bayesian structural time-series models. *The Annals of Applied Statistics.* 9 (1), 247–274. [online]. Available from: http://www.jstor.org/stable/24522418 (Accessed 27 July 2023).

Carlson, D. (2018) Modeling without conditional independence: Gaussian process regression for time-series cross-sectional analyses. PhD thesis thesis. University of Washington, St. Louis.

Carlson, D. & Brown, D. (2020) *Estimating a counter-factual with uncertainty through gaussian process projection.*

Doudchenko, N. & Imbens, G. W. (2016) *Balancing, regression, difference-in-differences and synthetic control methods: A synthesis.* [online]. Available from: http://www.nber.org/papers/w22791.

Eli Ben-Michael, A. F., David Arbour (2023) Estimating the effects of a California gun control program with multitask Gaussian processes. *The Annals of Applied Statistics.* [Online] 17 (2), 985–1016. [online]. Available from: https://doi.org/10.1214/22-AOAS1654.

Ernst, B. et al. (2009) Wind power and photovoltaic prediction tools for balancing and grid operation. *2009 CIGRE/IEEE PES Joint Symposium Integration of Wide-Scale Renewable Resources Into the Power Delivery System.* 1–9. [online]. Available from: https://api.semanticscholar.org/CorpusID:34026977.

Giudice, G. et al. (2022) *Inference on causal effects of interventions in time using gaussian processes.* [online]. Available from: https://arxiv.org/abs/2210.02850.

Häfner, D. (2019) The future is still unwritten – history too. *Energiepolitik und Klimaschutz. Energy Policy and Climate Protection.* [online]. Available from: https://api.semanticscholar.org/CorpusID:199340799.

Heering, J. & Gustafson, T. (2021) Germany's energiewende at a crossroads. *German Pol-*

itics and Society. [online]. Available from: https://api.semanticscholar.org/CorpusID: 239743713.

International Atomic Energy Agency (2023) *Power reactor information system.* [online]. Available from: https://pris.iaea.org/pris/.

Jarvis, S. et al. (2019) '*The Private And External Costs Of Germany's Nuclear Phase-Out*', in [Online]. December 2019 National Bureau of Economic Research.

Kim, S. et al. (2020) Bayesian synthetic control methods. *Journal of Marketing Research.* 57831–852. [online]. Available from: https://api.semanticscholar.org/CorpusID: 115149966.

Menchetti, F. & Bojinov, I. (2022) Estimating the effectiveness of permanent price reductions for competing products using multivariate Bayesian structural time series models. *The Annals of Applied Statistics.* [Online] 16 (1), 414–435. [online]. Available from: https://doi.org/10.1214/21-AOAS1498.

Mez, L. (2020) 'Rise and fall of the nuclear dream in germany', in Lutz Mez et al. (eds.) *The ecological modernization capacity of japan and germany: Comparing nuclear energy, renewables, automobility and rare earth policy.* [Online]. Wiesbaden: Springer Fachmedien Wiesbaden. pp. 87–101. [online]. Available from: https://doi.org/10.1007/978-3-658-27405-4_7.

Michaelides, E. E. (2019) Fossil fuel substitution with renewables for electricity generation – effects on sustainability goals. *European Journal of Sustainable Development Research.* [online]. Available from: https://api.semanticscholar.org/CorpusID:214350155.

Neal, R. M. (1996) *Bayesian learning for neural networks.* Lecture notes in statistics. Vol. 118. New York: Springer.

Nelkin, D. & Pollak, M. (1980) Political parties and the nuclear energy debate in france and germany. *Comparative politics.* 12127. [online]. Available from: https://api.semanticscholar.org/CorpusID:155582972.

Norman, J. & Sweeney, D. (2010) Germany's 'hot autumn' of nuclear discontent. *The Sydney Morning Herald.* [online]. Available from: https://www.smh.com.au/politics/federal/germanys-hot-autumn-of-nuclear-discontent-20100914-15aag.html.

Rasmussen, C. E. & Williams, C. K. I. (2006) *Gaussian processes for machine learning.* Adaptive computation and machine learning. MIT Press.

Rocholl, N. & Bolton, R. (2016) Berlin's electricity distribution grid: An urban energy transition in a national regulatory context. *Technology Analysis & Strategic Management.* 281182–1194. [online]. Available from: https://api.semanticscholar.org/CorpusID: 4245000.

Rode, D. C. et al. (2023) Modifying the EPA's new power plant rules to eliminate unnecessary reliability risks. *Environmental Science & Technology.* [Online] 57 (30), 10904–10906. [online]. Available from:https://doi.org/10.1021/acs.est.3c04608 .

Rossnagel, A. & Hentschel, A. (2012) The legalities of a nuclear shutdown. *Bulletin of the Atomic Scientists.* [Online] 68 (6), 55–66. [online]. Available from:https://doi.org/10.1177/0096340212464361.

Scherwath, T. et al. (2020) Nuclear decommissioning after the german nuclear phase-out an integrated view on new regulations and nuclear logistics. *Energy Policy.* [Online]

137111125. [online]. Available from: https://www.sciencedirect.com/science/article/pii/S0301421519307128.

Schreurs, M. A. (2013) Orchestrating a low-carbon energy revolution without nuclear: Germany's response to the fukushima nuclear crisis. *Theoretical Inquiries in Law*. [Online] 14 (1), 83–108. [online]. Available from: https://doi.org/10.1515/til-2013-006.

Wehrle, N. P. B. (2022) The cost of renewable electricity and energy storage in germany. *European Journal of Business Science and Technology*. [online]. Available from: https://api.semanticscholar.org/CorpusID:251206397.

Xu, Y. (2016) Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*. 2557–76. [online]. Available from: https://api.semanticscholar.org/CorpusID:2542789.