8-2024

Empirical Studies on Flexible and Renewable Investment Interaction in the Modernizing Electricity Grid

Seyed Amin Seyed Haeri
Clemson University, sseyedh@clemson.edu

Empirical Studies on Flexible and Renewable Investment Interaction in the Modernizing Electricity Grid

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Business Administration

by
Seyed Amin Seyed Haeri
August 2024

Accepted by:
Dr. Ahmet Colak, Committee Chair
Dr. Şafak Yücel
Dr. M. Serkan Akturk
Dr. Lawrence D. Fredendall

Abstract

In the transition towards sustainable energy sources, the U.S. electricity generation portfolio has seen a significant shift from conventional fossil-fueled generation to renewables such as solar and wind. This evolution presents complex challenges, notably the intermittency of renewable sources and their spatial-dependent investment dynamics. In the first chapter, we delve into the intricate relationship between renewable and conventional generation technologies, with a particular focus on the role of operational flexibility. Utilizing a unique panel data set covering the U.S. electricity industry from 2002 to 2019, we develop a multi-stage empirical strategy to illuminate how operational flexibility influences the interplay between investments in different generation technologies. Our findings reveal that investments in flexible generation technologies are not only stimulating future investments in wind and solar generation capacity but are also responsive to current renewable penetration, particularly wind energy. We uncover a complementary relationship between prior investments in flexible sources and subsequent investment in intermittent generation technologies, highlighting the crucial role of operational flexibility in accommodating renewable intermittency. This complementary dynamic underscores the importance of flexible generation capacity in ensuring grid reliability and efficiency in the face of growing renewable investments. Furthermore, our analysis sheds light on the influence of renewable portfolio standard policy instrument, market competition, and natural gas infrastructure development on the interplay between investments in these generation technologies. By elucidating the economic interaction between renewable and flexible conventional generation technologies, our study contributes to the strategic development of policy and investment frameworks that support the United States' transition towards a more sustainable electricity grid. In the second chapter, we focus on understanding the economic interplay between intermittent renewable technologies (wind and solar) and flexible conventional generation technologies, considering the role of varying infrastructure development levels and market competition within

the U.S. electricity sector from 2002 to 2019. Our methodology employs a comprehensive panel data set, leveraging principal component analysis, k-means clustering, and Tobit models to dissect how prior investments influence subsequent generation capacity expansions under different infrastructural scenarios—categorized as empty, own, and cross. We find that investments in flexible generation significantly boost subsequent renewable investments in scenarios with no pre-existing capacities (empty) or similar pre-existing technologies (own), while discouraging them in scenarios where existing capacities are based on dissimilar technologies (cross). Furthermore, initial renewable investments tend to promote further investments in flexible technologies, particularly in empty and cross scenarios, highlighting the need for enhanced grid stability due to the intermittency of renewable sources. The study's insights offer valuable managerial implications for energy sector stakeholders, including policymakers, investors, and decision-makers. By elucidating the complex dynamics of investment interactions within various infrastructural contexts, this research provides a robust basis for strategic investment decisions aimed at optimizing the integration of renewable resources and enhancing grid reliability. Consequently, stakeholders are better equipped to formulate and implement policies that not only accommodate but actively promote a sustainable and economically efficient transition to renewable energy. In the third chapter we introduce a continuous measure of flexibility, based on startup times of power plants, to empirically evaluate its impact on renewable energy investments. These flexibility scores highlight the uneven distribution of grid capabilities to manage supply-demand mismatches, which is crucial for integrating intermittent renewable sources like wind and solar. Through a robust multi-stage empirical strategy, we demonstrate that regions with higher flexibility scores are significantly more attractive for renewable energy investments, particularly in wind energy. The group fixed-effect (GFE) estimation results indicate that a 1% increase in regional flexibility score correlates with a substantial rise in wind and solar future investments. Our findings highlight the critical importance of enhancing grid flexibility to support the transition to a sustainable energy future. By providing policymakers and investors with a nuanced understanding of the interplay between flexibility and renewable investments, this chapter offers valuable insights for strategic investment decisions and policy formulations aimed at optimizing the integration of renewable resources and ensuring grid reliability. This study not only contributes to the literature on operational flexibility and renewable energy but also underscores the pivotal role of flexible energy infrastructure in the evolving dynamics of the modern electricity grid.

Dedication

To my parents, my sisters, my fiance, and all my friends who tolerated me during my journey.

Acknowledgments

I have been fortunate enough during my academic journey to meet so many incredible people to whom I owe much of who I am today, professionally. First and foremost, I deeply appreciate the time and effort Dr. Ahmet Colak, my dissertation committee chair, dedicated to teaching me the fundamental skills required for any large empirical project. His unwavering support and patience have been invaluable. His unique perspective on empirical research has significantly refined and honed my own research approach. Without his guidance, I would not have been exposed to, nor learned, many sophisticated methodological and programming concepts for which I am forever grateful.

I extend my heartfelt appreciation to Dr. Şafak Yücel, committee member, for his invaluable insights that profoundly impacted my understanding of my research subject. His vast knowledge played a pivotal role in shaping and advancing my research agenda. Moreover, Dr. Yücel's support extended beyond our research, and I am deeply thankful to have had not just an outstanding academic advisor, but an exceptional role model.

My initial independent research during my PhD journey began with the guidance of Dr. M. Serkan Akturk, committee member, from whom I learned a great deal about conducting high-quality empirical research. I am also grateful for Dr. Akturk's help and support during my time at Clemson University, which extended far beyond shaping my academic character.

My heartfelt thanks also go to Dr. Lawrence D. Fredendall, another esteemed committee member, for his incredible patience and kindness. I have worked with Dr. Fredendall in various capacities, where I had the opportunity to learn so much from him. As an international student starting my PhD journey, I faced many new and challenging situations, and I had the privilege of having Dr. Fredendall as my mentor, making things much easier to handle. For this, I am indefinitely grateful for his help and support.

I also want to thank my friends Jaeyoung, Maneesh, Yao, Jingyuan, Hongli, Daniel, and Jonathon for their generous help and support, which significantly impacted my PhD experience. Lastly, I want to express my deepest gratitude to my family for their unconditional love and support throughout this journey. I am especially grateful to my incredibly kind fiancée, who stood by me through the toughest times and always believed in me.

Table of Contents

List of Tables

List of Figures

Chapter 1

Do Energy Generation Flexibility Facilitate More Intermittent Investments? Evidence from the U.S. Electricity Industry

1.1 Introduction

In the past two decades U.S. electricity generation capacity portfolio has undergone major changes. Particularly, more wind and solar generation capacity have been entering the electricity grid at a staggering rate going from nearly zero in 2002, up to covering more than 10% of the total U.S. electricity generation capacity in 2019 (Figure 1.1). This transition is partly driven by the policymakers increasing attention to reduce the greenhouse gas emission of the electricity industry (Borenstein and Kellogg, 2023), as well as renewable electricity generators becoming more economically competitive in the past decades. However, the intermittent nature of solar and wind electricity generation along with their spatial-dependence pose challenges to grid stability and reliability. The generation of solar and wind electricity is contingent upon weather conditions, leading to fluctuations in power supply that do not always match the dynamic electricity demand. Improving electricity grid's flexibility[1] levels through the addition of flexible electricity generation capacity is one of the potential solutions to alleviate adverse consequences of the supply-demand mismatch (Lew and Brinkman 2013, Bird et al. 2013, and Milligan et al. 2015). International Energy Agency (IEA) projects that under a scenario where no changes are made to the existing policy setting of the U.S. government as well as other countries around the world, the flexibility requirements[2] of the U.S. electricity system will rise by 40% from 2020 to 2030 due to investments in solar and wind generators (IEA, 2021). Natural gas, coal, petroleum, and hydro are the main fuel sources of flexible generators in the U.s. Specifically, natural gas-fueled generators are an important source of flexibility with average of slightly more than 45% share of total generation capacity portfolio of the U.S. electricity grid between 2017–2019 (Table 1.2). IEA anticipates that natural gas-fueled power plants will remain central as a source of flexibility in power grids under all potential future scenarios except for a net-zero emission scenario (IEA, 2021). This highlights the increasingly important role of flexibility in achieving long term clean energy-related goals.

The interplay between investments in conventional power plants and solar and wind capacity investments is complex and has been highly debated. More precisely, Penn (2020) in a *New York Times* article underscores how large utility firms are committed to investments in natural gas power plants that allows them to have flexibility in their generation while investing in massive renewable projects. On the other hand, some argue that the fluctuations in natural gas prices may lead to less favorable environment for renewable investments. Specifically, in the presence of lower natural gas prices, such power plants may substitute renewable investments (Kotchen, 2012). Despite the importance of understanding the relationship between these generation technologies, we identify two major gaps in the literature to motivate our research questions. First, there are limited studies in the field of operations management (OM) that explore the role of flexibility in the U.S. electricity market (Kök et al. 2020 and Al-Gwaiz et al. 2017). Nevertheless, these studies do not distinguish between the main fuel type of the conventional flexible generators and only look at two broad groups of flexible versus inflexible without a clearly defined measure of flexibility. This is important as there are various types of conventional power plants with respect to their generation technology with different flexibility levels. Second, investments in solar and wind generators highly

[1]Flexibility is defined as, "the ability of a power system to reliably and cost-effectively manage the variability and uncertainty of demand and supply across all relevant timescales, from ensuring instantaneous stability of the power system to supporting long-term security of supply."(IEA, 2021)

[2]Flexibility requirements are, "the hour-to-hour ramping requirements after removing wind and solar production from electricity demand, divided by the average for the year." (IEA, 2021)

depend on the geographical characteristics of a location and whether a location can provide sufficient solar radiation or wind speed to attract investors (Figure 1.2). Additionally, investments are influenced by the existing generation capacity portfolio in markets. This highlights the importance of understanding the interplay between investments in different generation technologies (i.e., whether certain technologies are substitute or compliment to each other and whether such relationships hold, taking into account spatial heterogeneity).To the best of our knowledge, there is no OM study that empirically investigate the relationship between investments in intermittent (i.e., solar and wind) and flexible generators (natural gas-fueled and other sources of fuels) taking into account the spatial heterogeneity of investment locations.

Understanding the nature of economic interplay between conventional flexible and intermittent generators from an investment perspective is specially important with increasing penetration of intermittent generators changing the electricity market landscape. The coexistence of these two types of generation technologies has been the topic of debate in the past decade and an enhanced understanding of their economic dynamics within investment portfolios ultimately informs policy design and investment decisions. With this motivation, we pose the following research questions: Are investments in flexible and renewable power plants complements or substitutes to each other? We adopt an empirical approach in answering our research question and focus on flexible natural gas-fueled generation capacity that it's share in the U.S. generation capacity portfolio has increased by 8% between 2002–2019 with all other conventional generation shares remaining almost unchanged (Table 1.2).

To answer our research question, we collect a unique panel data set of the U.S. electricity market that includes, power plants geographical locations (i.e., longitude and latitude), generation capacity, generators' technological characteristics (e.g., main fuel type and start-up time), power plants' actual electricity generation, policy factors, solar radiation, wind speed, and other economic control variables for years 2002–2019 (Table 1.1). This data set is obtained from a combination of several publicly available data sets including, Energy Information Administration (EIA), Lawrence Berkeley National Laboratory (LBNL), National Solar Radiation Database (NSRDB), and U.S. Census Bureau. Our data set contains the generator-level annual capacity observations with unique power plant and generator identifiers. We leverage this to identify investments by comparing any power plant's capacity in all two consecutive years that we observe in the data. With this approach, we also identify new power plant entries that were not present in the data in a prior year in a two consecutive year pairwise comparison. Moreover, our data has a rich spatial aspect that report each power plants' zip code and exact geographical location (i.e., longitude and latitude). With this we characterize generation capacity portfolio, investment, and actual generation for all zip codes in the U.S.

One technological characteristic we observe in our data set is the time it takes for a conventional power plant to go from cold shut down to becoming fully operational (ranging from less than 10 minutes to more than 12 hours). We use this as a proxy variable to distinguish the operational flexibility levels of power plants, classifying those with start-up time below 12 hours to have high flexibility and those with start-up time above 12 hours to have low flexibility levels. With this we distinguish the generation capacity, actual generation, and investments into two groups of flexible and inflexible (further separating out the natural gas-fueled flexible generators investments).

We address our research question by hypothesizing that prior investments in intermittent generation capacity (i.e., solar and wind generators) stimulate future investments in natural gas-fueled power plants. In addition, we hypothesize that prior investments in flexible generation capacity also stimulate future investments in intermittent generation capacity. To test these hypotheses, we develop a multi-stage estimation framework, and quantify the impact of any investments in one generation technology on the other technologies of interest. We use linear regression to estimate the effect of investments in one generation technology on another technology while controlling for political and economic factors. However, we face several empirical obstacles such as, investment variables' rareness, un-observable continuous heterogeneity, policy and investment endogeneity. We circumvent these obstacles by following the following estimation strategy: *(i)* Running a first stage estimation model on our policy variable using instrumental variables, *(ii)* Using the combination of principal component analysis (PCA) and k-means clustering to identify the latent cluster structure of observations, *(iii)* Using the grouped fixed effect (GFE) approach to estimate our models and test our hypotheses using cluster fixed effects, *(iv)* Running investment correction models with instrumental variable approach for current investments, and *(v)* Running alternative specifications with current and expected investments. Lastly, we quantify and interpret our findings using a ceteris paribus counterfactual analysis where for example we evaluate what the average future investment in solar generation capacity would have been if there was no prior investment in highly flexible generation capacity.

Our results suggest that for small percentage increases in prior investment in flexible generation capacity relative to the largest investment observed in our data set for flexible generation capacity we expect an approximately 0.145% and 0.061% increase in future investments in wind and solar respectively. We also found that small percentage increase in prior wind investments lead to 0.091% increase in future investments in natural gas-fueled flexible generation capacity. However, we have found no empirical support for the relationship between solar prior investments and future natural gas-fueled investments. These empirical findings underscore the complementary relationship between flexible and intermittent generation technologies. In addition, through our extended analyses our findings were corroborated for the complementary interplay between flexible and wind generation capacity investments. We also found that the complementary relationship of flexible generation capacity prior investments on future intermittent investments are amplified when the investment location has higher variability of renewable electricity generation[3]. We also found empirical evidence that factors such as policy (i.e., renewable portfolio standard (RPS)), ownership sector (i.e., utility firms versus independent power producers), and natural gas infrastructure development (i.e., natural gas pipeline capacity) heterogeneously influence our hypothesized relationships.

Our findings provide insights for both policy makers and investors by shedding light on the dynamics between investments in different electricity generation technologies. This becomes increasingly important where on one hand there is a strong political will towards decarbonizing the electricity industry with renewable investments, while such generation technology is becoming more economically competitive. Developing a spatial understanding of cross-technology investment dynamics enables the policy makers to design better technology-specific subsidy programs that can facilitate the transition towards a cleaner electricity grid. Investors also benefit from our findings by being able to make more informed decisions regarding both the investment location and the generation technology through having a better understanding of the future competition and how they can diversify their generation portfolio to maximize their profits.

The rest of the paper is organized as the following. In Section 1.2 we provide a comprehensive review of the relevant literature. In Section 1.3 we establish our research hypotheses through elaborating on its theoretical background. Section 1.4 summarizes the data collection and variable development processes. In Sections 1.5 we explain our empirical model and estimation strategy. Further, we discuss our results in Section 1.6 and we provide our concluding remarks in Section 1.7.

[3]The variability of intermittent electricity generation can be characterized by solar radiation and wind speed variability.

1.2 Literature Review

Our research relates to two research streams of generation capacity investments and operational flexibility that are explored within two broader streams of *Operations Management (OM)* and *Energy Economics*. In the following we provide a summary of the literature and elaborate how our work pertains to the existing literature and how it contributes to it.

1.2.1 Conventional-Renewable Generation Capacity Link

With more intermittent renewable (i.e., solar and wind) generation capacity entering the electricity grids, interest in the conventional-renewable interaction within the investment portfolio has gained significant momentum. This interest led to an extensive body of research exploring the dynamics of these seemingly-competitive generation technologies. While the majority of the existing body of research in energy economics suggests a substitutive relationship between these generation technologies (Ambec and Crampes, 2012), Lee et al. (2012) argue that since renewable and natural gas-fueled conventional generators have different risk profiles they can have a complementary relation in an investment portfolio. Similarly, Nyangon and Byrne (2023) find that the growth of natural gas-fired generation capacity positively impacts distributed solar photovoltaic investments for the Pennsylvania–New Jersey–Maryland (PJM) Interconnection in the U.S. Some studies reconcile the two opposing relations by emphasizing the importance of moderating factors. Baranes et al. (2017) propose a rather nuanced relationship where for lower natural gas prices conventional and renewable technologies are substitutes and after a certain price threshold is met, the two become complements.

The studies in the energy economics literature have mostly adopted non-empirical methodologies characterizing the conventional-renewable interplay with few exceptions. Bushnell (2010) find under larger wind penetration scenarios, investments are shifted toward thermal generation technologies such as combined-cycle gas turbines (CCGTs) and combustion turbines (CTs) representing a complementary interaction in four sub-regions of the Western Interconnection in the U.S. (Devlin et al. (2017) and Verdolini et al. (2018) suggest a similar complementary renewable-conventional relation using evidence from the United Kingdom and Ireland along with OECD countries). In contrast, Marques et al. (2010) investigate the driving factors of renewable investments for European countries between 1990–2006 and argue that natural gas price has a positive impact on renewable investments suggesting a susbtitutive relation. Our study contributes to this evolving literature by shedding light on the conventional-renewable link over the whole U.S. electricity market. Contrary to the aforementioned studies that use aggregated data, we adopt a spatial approach and use highly granular data to estimate the two-way conventional-renewable relationship.

OM literature explores the driving forces of investments in renewable and conventional electricity generation technologies using various analytical methods (e.g., stochastic equilibrium models, dynamic games, stochastic programming, etc.) taking into account economic and political factors (Parker et al., 2019). More precisely the influence of factors such as, supply intermittency, net-metering, supply-demand data granularity, market liberalization, carbon pricing, electricity pricing (flat vs. peak), production tax credit, investment tax credit, carbon tax, feed-in-tariff, and tax-rebate on investments in both conventional and renewable generation technologies are explored extensively (Hu et al., 2015; Aflaki and Netessine, 2017; Kök et al., 2018; Babich et al., 2020; Alizamir et al., 2021). A less investigated topic in the OM literature is the link between investments in conventional and renewable generation technologies. Kaps et al. (2023) explore strategic investment in renewable generation and storage to meet off-grid energy demands using fossil fuel as backup. They argue that solar and storage capacities are generally strategic complements, but can become substitutes with high generation investment. Peng et al. (2021) examine the investment and operational dynamics of renewable, flexible, and storage energy capacity in the US power system. They identify that storage operation's impact on operating costs, and its constraints in charging or discharging, dictate whether it complements or substitutes other resources. They further find, storage substituting renewables in peak demand and complementing them by storing excess output, with renewables also aiding storage by easing peak demand management. Kök et al. (2020) explore this topic by characterizing the decision-making process of a utility firm investing in renewable and conventional generation capacity focusing on the operational flexibility levels of conventional sources (i.e., flexible and inflexible). They found that flexible and renewable sources have a complementary relationship while inflexible sources serve as substitutes to both technologies. In this paper we build upon the idea that Kök et al. (2020) present for a single utility firm and we empirically evaluate their proposed insights on how the conventional-renewable link in an investment portfolio may be explained by the operational flexibility of the conventional generation technologies on a macro level. To the best of our knowledge our paper is the first empirical OM study of U.S. macro-level electricity grid investments in renewable and conventional generation technologies.

One highly debated policy in the energy economics is the renewable portfolio standard

(RPS). RPS programs are environmental policy instruments used by states to increase the adoption of renewable power generation technologies by utility firms in the U.S. electricity market. These programs are designed and enforced on the state level and no two state programs are identical (Barbose, 2021). These programs, impose a requirement on retail electricity suppliers to procure a minimum percentage of their load from eligible renewable sources (mostly from solar and wind generators) while a penalty of some form is applied to non-complying firms (Barbose, 2021). According to the latest RPS status update report by the Lawrence Berkely National Lab, in year 2021, 30 states and the District of Columbia have adopted RPS policies that account for 58% of the U.S. total retail electricity sales (Barbose, 2021). States are constantly revising their RPS policies. For example, from January 2019 across all states 103 new bills have been introduced out of which 13 are enacted to strengthen the RPS programs of the states. Meanwhile, 30 new bills are also introduced that weaken the states' programs out of which 1 is enacted (Barbose, 2021). Numerous studies in the energy economics literature evaluated different direct and indirect implications of these RPS programs for U.S. market outcomes including, electricity prices, renewable generation, greenhouse gas (GHG) emissions, and renewable generation capacity investment (Lyon, 2016; Bowen and Lacombe, 2017; Carley et al., 2018; Zhu et al., 2020; Mullen and Dong, 2022). The main objective of any RPS program is to increase the share of renewable sources in the electricity generation portfolio to reduce GHG emissions. Therefore, as an indirect effect it is natural to expect – *with increasingly stringent RPS programs being enforced* – not just an increase in the generations from renewable sources but also an increase in renewable investments. Nevertheless, findings in the literature illustrate a rather nuanced relationship between RPS enforcement and renewable deployment. Zhou and Solomon (2020) find that the effectiveness of the RPS programs is a function of renewable resource endowment of a given location where locations with high endowment experience higher above-compliance levels of investments and those with low renewable resource endowment are negatively influenced by the RPS program. While Joshi (2021) argue that there is a positive influence of the RPS program on renewable generation capacity development for the U.S., Feldman and Levinson (2023) contradicts their results by finding small or insignificant influence of these programs on renewable investments. Deschenes et al. (2023) emphasize an effect heterogeneity depending on the renewable generation technology with RPS positively and negatively influencing wind and solar investments respectively. To the best of our knowledge, there is no OM study that takes RPS in to account when evaluating investment decisions made in the U.S. electricity market. We contribute to this literature by proposing a way to correct for the endogeneity of this important policy instrument using a novel two-stage spatial approach when controlling for it's effect on renewable investments.

1.2.2 Operational Flexibility

The notion of operational flexibility has been receiving an increasing attention from both energy economics and OM literature. With more intermittent renewable generation capacity entering electricity grids, flexibility of electricity grids is becoming an indispensable capability to offset supply variability aiding the efficient and reliable operation of electricity value chains (Papaefthymiou and Dragoon, 2016). Kondziella and Bruckner (2016) and Lund et al. (2015) provide a list of options to enhance electricity systems' flexibility that includes but is not limited to, demand response programs, energy storage, flexible generation, transmission network expansion, virtual power plants, and advance forecasting systems. Neetzow (2021) argues that investments in flexible generation capacity (specifically natural gas-fueled generation capacity) is the most cost-efficient approach to improve electricity systems' flexibility.

The energy economics literature underscores the importance of flexible fossil-fueled generation capacity in electricity systems with high levels of renewable penetration and how the existence of such generation technology in the generation capacity portfolio ensures a reliable transition towards a renewable-rich energy landscape (Vithayasrichareon et al., 2017; Koltsaklis et al., 2017; Olsen et al., 2020; Guerra et al., 2022). Nevertheless, there is a dearth of studies that explore the complex conventional-renewable link with a focus on the operational flexibility characteristic of the conventional generation capacity.

In the OM literature, Al-Gwaiz et al. (2017) examine the impact of generation flexibility and renewable energy on power market competition. They explore how these factors influence strategic behavior among market participants. The research offers insights into the intricate interplay between generation capabilities and renewable sources, shedding light on their implications for market dynamics, strategic decisions, and overall competitiveness within the power sector. Angelus (2021) investigates distributed renewable power generation's effects on capacity investment and electricity prices. Their study investigates how decentralized renewable sources impact capacity planning and pricing dynamics, providing insights into their implications for the power sector's investment and economic landscape. Peng et al. (2021) explore the relationships between renewable, flexible, and storage capacity. Investigating their synergies and conflicts, the study offers insights into the intricate balance required for effectively integrating renewable sources, flexible generation, and energy storage within the evolving energy landscape. Kök et al. (2020) shed light on the complicated nature of conventional-renewable link and how the operational flexibility influences a utility firm's generation capacity investment portfolio in an analytical study. They underscore that depending on differences in the operational flexibility levels of conventional generators (i.e., flexible vs. inflexible) the conventional-renewable investment relation will be different for solar and wind generators. In this study, we empirically evaluate the relationships proposed by Kök et al. (2020) and further contribute to the literature by empirically quantifying the conventional-renewable link and how this relation is moderated by economic and political factors. Moreover, our results help policy-makers to design more informed technology-specific policy instruments that accelerate the transition towards a green electricity grid.

1.3 Hypotheses Development

One of the main challenges of any electricity grid is the real-time matching of supply and demand. The main reason is the stochastic nature of demand (Anvari et al., 2022). This challenge becomes more complicated with higher penetration levels of intermittent renewable sources such as wind and solar generation capacity that adds the additional supply-side uncertainty to this already challenging task. Given that peak supply of solar and wind generators are not matched with peak electricity demand – with a lack of an economic grid-level storage solutions – this problem becomes more complicated. Traditionally electricity grids rely on fast-acting fossil fueled generators (mostly natural gas-fueled generators) with short ramp-up times to balance supply and demand in electricity grids with renewable generation capacity. With the reliance of electricity grids on the co-existence of flexible and intermittent renewable generation capacity we hypothesize that there is a two-way relationship between these two distinct generation technologies. We rely on the literature and separately hypothesize about the two-way relationships for the influence of flexible conventional generators on solar and wind (Verdolini et al., 2018). The main reasoning behind this approach is that solar and wind generators have fundamentally different generation intermittency patterns (Wu et al., 2022; Ren et al., 2018), therefore, we expect them to be differently both influence and be influenced by the flexible conventional generators. In the following sub-sections, we further elaborate how we expect each side of the aforementioned two-way relations to work.

1.3.1 Hypothesis 1 & 2: Flexibility Drives Intermittency

We hypothesize that locations with higher prior investments in flexible conventional generation capacity attract more future investments in wind and solar generation capacity controlling for political and economic factors. We explain this hypothesis by the concept of renewable electricity curtailment. Given the non-dispatchable nature of solar and wind generators, there are times that there is excess supply and that the grid manager or the utility firm must curtail the generated electricity by renewable generators. This renewable curtailment leads to opportunity costs for renewable generators as well as delayed return on investments that can potentially lead to disincentivizing renewable investments. (Golden and Paulos, 2015). Solar and wind curtailment has been increasing specifically at locations with more ambitious green electricity objectives such as California. According to a report by the Energy Information Administration (EIA), in year 2022, solar and wind curtailment faced a 63% increase from the same time in 2021 mainly due to over-supply or grid congestion (EIA, 2023). Given that investments in solar and wind are substantially spatial-dependent, it is natural to observe certain locations receive more investments given their favorable geographical attributes (i.e., higher solar radiation and wind speed). However, this effect is diminished with curtailment and may be avoided with existing flexible generation capacity. Then at times of excess supply, instead of renewable curtailment, the grid manager or the utility firm can ramp down the conventional flexible generators while keeping the electricity generated from renewable generators. We hypothesize, assuming all else equal, for locations that have higher prior investments in flexible conventional capacity to attract more future solar and wind investments. Similar to the hypotheses we established in Section 1.3.1 we distinguish solar and wind influences as they have significantly different flexibility requirements.

Hypothesis 1. *Prior flexible generation capacity investment around a focal zip code positively increases future wind investment rates around that focal zip code.*

Hypothesis 2. *Prior flexible generation capacity investment around a focal zip code positively increases future solar investment rates around that focal zip code.*

1.3.2 Hypothesis 3 & 4: Intermittency Drives Flexibility

We hypothesize that a reverse of the relationship that we detailed in Section 1.3.1 also exists. The main driving mechanism of this relationship is the increasing flexibility requirements of electricity grids that is reflected as price signals (i.e., dispatchability[4] price premium). With an increasing share of intermittent renewable generation technology within electricity grids generation capacity portfolio, conventional electricity generators are used to cover the residual demand. This implies that at times when there is no sunshine (or wind), electricity demand is met with more flexible conventional generators that have lower capital cost and higher marginal cost. Such generators have shorter ramp-up times compared to base-load serving generators. It naturally follows that flexible generators are favored in the electricity markets with being able to sell a higher-priced electricity at the times of supply scarcity. This dispatchability price premium is well-documented in the literature (Al-Gwaiz et al., 2017; Rai and Nunn, 2020; Bushnell and Novan, 2021; Glenk and Reichelstein, 2022). Therefore we expect locations with higher prior investments in intermittent renewable generation capacity to incentivize future investments in flexible generation capacity. More specifically, we focus on future investments in natural gas-fueled flexible generation capacity as historically this generation technology is the dominant conventional flexible generation technology. Moreover, Koltsaklis et al. (2017) emphasizes that solar and wind generation technologies create different flexibility requirements leading to the fact that they heterogeneously impact future investments in flexible generation capacity. We formulate the following two hypotheses to evaluate how solar and wind prior capacity investment affect investments in flexible generation capacity:

Hypothesis 3. *Prior wind generation capacity investment around a focal zip code positively increases future natural gas-fueled flexible investment rates around that focal zip code.*

Hypothesis 4. *Prior solar generation capacity investment around a focal zip code positively increases future natural gas-fueled flexible investment rates around that focal zip code.*

1.4 Data

In this section we first explain our main data sources and how we obtained our raw datasets. Second, we elaborate the steps we took to process raw data and prepare our final sample. Third, we rely on the literature (Bourcet, 2020 and Brehm, 2019) defining and constructing our main variables from raw data.

[4] Dispatchable electricity generators are power sources that can be turned on or off, or adjusted to increase or decrease power output according to the demand. An example of such generators is flexible natural gas-fueled generator. Non-dispatchable electricity generators, on the other hand, are sources of power that cannot be controlled in the same way. Their output depends on external factors, such as weather conditions, and cannot be turned on or off at will. Solar and wind energy are prime examples of non-dispatchable sources.

1.4.1 Data Sources

We compile a unique panel data set from various sources as detailed in Table 1.1. We use the data from Energy Information Administration (EIA) forms 860 and 923 (and their retired preceding forms) for years 2002–2019 to create our main data set[5]. Form 860 contains yearly observations for existing utility-scale electricity generators[6] including, unique generator ID number, owner plant's ID number, nameplate capacity, operation status, fuel type, and the time it takes a generator to go from cold shutdown to fully operational (i.e., generators' startup time). Form 860 also details different investment stages of developing projects with their fuel type, generation capacity, and the state in which they will be located. The development stages (i.e., investment stages) are, proposed, pending legislative approvals, under on-time construction, under delayed construction, postponed, and canceled. Form 923, encompasses plants' exact geographical locations, ID numbers, ownership sector[7], and actual electricity generation amount. The geographical aspect of the two aforementioned data sets had missing observations in both zip code of a power plant as well as the longitude and latitude. To fill in missing geographical details, we use geocodio website API services to find the missing longitude-latitude of a location using the available address and zip code (i.e., geocoding) and also find the missing zip codes using the available longitude-latitude (i.e., reverse-geocoding).

We augment our main data set with the in-flow and out-flow capacity of natural gas pipeline at each county in the U.S. using EIA-Natural Gas data set. We use the data from the Lawrence Berkeley National Laboratory (LBNL) to collect state-level renewable energy certificate (REC) requirements[8] of each state, mandated by their respective RPS programs' obligations.

To distinguish geographical locations' potential for solar and wind investment, we collect the zip code level observations of solar radiation and wind speed every 30-minutes for 2002–2019 that compiles in to over 10 billion observations. We use the National Solar Radiation Database (NSRDB) API service to obtain two measures of Global Horizontal Irradiance[9] (GHI), and wind speed at the geographical center of each zip code.

[5]According to EIA Electric Power Data - *March 2018*, *"the data are most consistent, and of the highest quality for the period beginning with the 2002 data."*

[6]All power plants with nameplate capacity above 1 megawatts that is defined as, "the maximum rated output of a generator, prime mover, or other electric power production equipment under specific conditions designated by the manufacturer."

[7]Power plants can be owned by one of the following sectors, independent power producers (IPP), utility firms, commercial, or industrial

[8]This data set report the net new REC requirements in each year for each state so they can stay in compliance with their RPS programs' requirements.

[9]GHI represents the total amount of shortwave radiation received from above by a surface which is horizontal (parallel) to the ground. GHI is the most important parameter for calculation of PV electricity yield.

We collect state-level observations of residential electricity prices using EIA form 826, along with states' energy intensity[10] from EIA State Energy Data System (SEDS). We obtain country-level observations of natural gas price from EIA form 826.

1.4.2 Data Pre-processing

To prepare our sample we make several adjustments to our raw data set. First, we standardized the spatial data by aligning zip codes and geographical coordinates, enabling us to merge all collected datasets. Second, we set the unit of analysis to be a 100 miles radius around the center of each zip code with 3-years time window aggregation. Third, we leverage categorical variables to create generation technology-specific variables. Forth, we make several adjustments to our raw data measurement scales. In the following sub-sections we discuss this process in greater details.

1.4.2.1 Unit of Analysis

Generation technology: In our data set, we observe the generation capacity associated with their fuel type including: solar, wind, natural gas, coal, petroleum, hydro, nuclear, and bio-fuels (details on the U.S. generation capacity portfolio is provided in Table 1.2). We leverage this categorical variable to create technology-specific generation capacity variables that show technological distribution of any focal region. This allows us to observe technology-specific investment behaviors as we explain in Section 1.4.2.3.

Operational flexibility: In electricity systems, flexibility corresponds to the system's capability to match supply and demand at each point in time. We adopted an OM approach in defining a a generator's flexibility that is aligned with its definition in the context of the electricity industry. The literature of OM defines a specific dimension of flexibility in manufacturing as volume flexibility to be a system's *"ability to be operated profitably at different overall output levels"* (Sethi and Sethi, 1990). In manufacturing, volume flexibility has two main aspects of, response speed and production variation range. Nevertheless, in the electricity systems, given that the commodity is homogeneous (electricity), the former can be used as an appropriate measure of flexibility for a generator. Therefore, we use the *time it takes for a generator to go from cold shutdown to fully operational* as a proxy measure to differentiate generators' levels of flexibility. In our data set, this attribute is a categorical

[10]Energy intensity is defined as the total energy consumption divided by real gross domestic product (GDP) for each state.

variable with 4 levels including: less than 10 minutes, between 10–60 minutes, between 1–12 hours, and over 12 hours. Using this measure, we divide conventional generators into two groups of flexible with start-up time below 12 hours and inflexible generators with startup time over 12 hours.

Geographical aggregation: Our raw data set has several units of observation as granular as generator-level to aggregated national-level observations. We first aggregate our generator- and plant-level observations to be on the zip code level. Later, we adopt 100 miles radius around each focal zip code to be our unit of analysis that we hereafter refer to as *"region"* for four reasons. First, in contrast to zip code level observation of investments that lead to a highly skewed distribution with many zero values, setting observation level as regional allows us to have a far less skewed distribution. Second, since investment in renewable technologies is a function of the geographical location, it is reasonable to have fixed locations to observe investment behaviors across time. Third, locations differ with respect to the existing infrastructure development which substantially influences investments in solar and wind (from the viewpoint of available transmission lines), as well as investments in natural gas-fueled generators (from the perspective of access to natural gas pipeline). With selecting regions as our unit of analysis, we control whether the appropriate infrastructure exists in a given location or not through the existing capacity of similar generation technology. Fourth, we leverage this unit of analysis to separate the cross effect that investment in different generation technologies may have on each other from the possible synergistic effects that investment in one generation technology may have across time. In Table 1.5 we represent each variable group and their initial granularity levels as well as how we have made them to be on higher aggregation levels.

Time frame aggregation: In our data set we have several frequency of observations ranging from 30-minute intervals (for solar radiation and wind speed of each zip code) to yearly observations (of main generators' variables). We aggregate all of our variables to be averages over 3-years time windows (e.g., 2002–2004 is a 3-year time window). Additionally, we construct leading investment variables that represent average investments in generation capacity in the next 3-year time window taking into account the time it takes for new project developments.

1.4.2.2 Data Transformation

Data filtering: Form 860 contains detailed annual generator-level information about operating, proposed, and canceled (or retired) generators. This form reports more than 40 variables for operating generators (i.e., the focus of our study) containing details on the environmental equipment associated with the generators. Given the scope of this study we do not use these variables and we only focus on the ones listed in Table 1.1. Similarly, form 923 reports monthly fuel consumption and actual generation of power plants where for the purpose of this study we only use the actual monthly generation data.

Data identifiers: In this study we treat zip codes as the staple to bind different individual data sets together. Our main data set (i.e., data obtained from forms 860 and 923) involves rich spatial information for generators and power plants such as zip code, postal address, and geographical coordinates (i.e., longitude and latitude). There are missing observations for some plants though. Namely, there are plants with only zip code available in the raw data set along with some other lacking zip code data. We overcome this issue by geo-coding (i.e., transforming zip code and states into geographical coordinates) and reverse-geocoding (i.e., transforming geographical coordinates into zip code and states) using geocodio website's API service. As detailed in Table 1.3, we find the missing coordinates of 390 generators and use the whole geographical coordinate data set to find zip code and state data of all observations to have a consistent geographical data set as the raw data set has some inconsistencies in reporting zip codes and states. Lastly, since the natural gas pipeline capacity data is observed on the county level, we gather data of the geographical coordinates for the center of each county in the U.S. Then we use each zip code's geographical coordinate to locate all counties within 100 miles radius of the center of a focal zip code. Using this approach, we assign a total natural gas pipeline capacity to any region based on the average availability of pipeline capacity in its 100 miles vicinity. For any other variable with lower granularity resolution, we aggregate into higher levels such as state or national level.

Data categorization: We combine two categorical variables of generation technology and startup time to categorize the variables such as, generation capacity, capacity utilization, and investments. We focus on 5 major generation technology-flexibility categories including: solar, wind, flexible (encompassing all fuel types), natural gas-fueled flexible (NG-flexible), and inflexible (encompassing all fuel types). We distinguish flexible and NG-flexible for the purpose of answering our research question. When we are looking at prior flexible investments, we consider all types of fuels for flexible generators, however, when we are evaluating future flexible investments we only take into account the NG-flexible technology.

Data normalization: We normalize all of our investment variables to be per capita by dividing them by the population of their focal region. We use the summation of all zip codes' populations within a 100 miles radius of a focal zip code to be the total population of a given region that is measured in million people. Thus all investment-related variables are measured per million individual. We also normalize the state-level RPS obligation variable by dividing state-level RPS obligations with the number of unique zip codes in a given state. After this normalization the measurement unit of the RPS obligation variable becomes Megawatt-hours per zip code. This takes into account the size heterogeneity of states, therefore, RPS obligations are more comparable with this normalization. Lastly, we do a similar normalization for the state-level investment stages.

1.4.2.3 Variable Construction

Dependent variables: We use 3-year leading investments in generation capacity of solar, wind, and NG-flexible technologies at any focal region with a unique zip code at it's center as our dependent variable. These variables capture next 3-year mean investments in any of the three generation technologies of interest.

Independent variables: We define two sets of independent variables including *prior* and *current* investments. We calculate prior investments of wind, solar, and flexible technologies while for the current investment in addition to wind and solar we focus on a subset of investments in flexible sources that is the natural gas-fueled generation technology. Prior to the time frame aggregation, we calculate the existing generation capacity of the three generation technologies of interest at the beginning of each year by subtracting that year's investments (Table 1.4). Then we aggregated the existing generation capacity at the beginning of each year for 3-years time windows. Aligned with our research objective and our empirical setting, we denote these independent variables of interest as *prior investments* while we denote the additional generation capacity *current investments.*

Control variables: We use a suite of control variables measured on different granularity levels. Except for variable calibration process that we explain in the following sub-section, we use the following variables as we collected them: solar radiation, wind speed, natural gas price, RPS obligation, energy intensity, region's population, and residential electricity prices. In Table 1.4 we summarize the construction steps of the other variables while providing more details in the following: *(i) Capacity utilization:* Initially we created this variable for each plant, then computed the weighted average of utilization of plants inside their respective zip codes then we averaged them on the 100 miles radius of the focal zip code to calculate the regional utilization. We measure plant utilization by assuming a plant can generate 24/7 in a given year and multiply hours in a year (i.e., 8,760) by it's generation capacity. Then dividing the actual generation by this value gives us a proxy variable for utilization as actual utilization takes in to account the maintenance and down times but we do not have such data. *(ii) Work-in-progress (WIP) index:* As we explained in Sections 1.4.1 we observe 6 stages of project development for three generation technologies of, solar, wind, and natural gas. We omit the *canceled projects* and combine the other 5 stage capacity into index variables. We do this by assigning weights to each stage with higher weights assigned to stages closer to completion. We assign the following weights to each stage: planned (9%), postponed (9%), pending approval (18%), under delayed construction (27%), and under on-time construction (36%). We provide the 3-year averages of these variables between 2002–2019 in Table 1.A1. Values in this table represent average capacity (measured in gigawatts) in each stage of development per state given that we aggregated these variables to be on the state-level. *(iii) Utility ownership:* We leveraged the ownership that classifies generators' owners in to one of the utility, independent power producer (IPP), commercial, or industrial groups to create the utility ownership variable. Since the ownership ratio of commercial and industrial sectors are rather small and constant across time, we only focus on utility and IPP ownership. We define and calculate utility ownership as the ratio of all capacity at zip codes within a region that are owned by utility firms. We found that IPP and utility ownership show almost perfect correlation with one another thus we only incorporate utility ownership in our sample. *(iv) Regional total capacity:* To account for the size of each region, we calculate the total available generation capacity regardless of technology simply by summing up all available capacity at zip codes within a 100 miles radius of a focal zip code. *(v) Capacity surplus:* We define this proxy variable to capture how much a given region is of an electricity exporter or importer to the other regions. We first compute the per capita generation capacity that is required to generate the actual electricity (i.e., total per capita supply divided by 8,760) and then multiplying this value by the peak-to-mean ratio of 2. Lastly, we divide each region's per capita generation capacity by the aforementioned value. This measure allows us to identify how much generation capacity exists in a given region compared to the generation capacity required to meet peak demand of the U.S. *(vi) Intermittency level:* We define intermittency level to use it as a proxy measure for potential renewable intermittency caused by under-supply during a day at a given location. We use the latest EIA hourly electricity demand data for February 2024 to identify daily average U.S. demand pattern as the percentage of total demand in each hour of the day[11] (we denote this as demand ratio). For the renewable supply side, we

[11]The average demand daily pattern of the U.S. has not significantly changed over time. For this reason, we used

use solar radiation and wind speed as proxy variables that represent supply variability of renewable generators (we denote these as supply ratios). Similarly, for wind and solar separately we calculate the ratio of solar radiation and wind speed in each hour of a day. Next, we subtract demand ratio from supply ratio and only keep the observations where demand was larger than supply. *(vii) Regional total capacity:* We construct a total generation capacity variable that characterizes the total available generation capacity in a given region. The variation of this measure helps us to control for the regional size heterogeneity. We construct this variable by summing up all the generation capacity within a focal region irrespective of their generation technology.

Variable calibration: Given that in our data set we have variables of vastly different measurement units and distributions, to avoid such characteristics influencing our estimation, we make several adjustments to the scales of our variables as we have summarized in Table 1.5 to prepare our sample for analysis. After time grame we explained previously,First, we use a Maximum Absolute Scaling approach[12] to re-scale all variables to be between $0 - 1$ then we transform them into percentages by multiplying them with 100. Second, we calculated the natural logarithm of all variables. Third, we take an additional winsorization step in preparing our sample with variables that have highly skewed distributions, to avoid outliers that influence our analyses. In Table 1.5 we present each individual variable winsorization level ranging from 90% to 95%.

1.4.3 Descriptive Statistics

In Table 1.2 we present the dynamic distribution of U.S. generation capacity technological portfolio between 2002–2019. Solar and wind combined generation capacity went from covering less than 1% of the U.S. total capacity portfolio up to covering more than 10% of it. This shift of investment toward renewable sources coincided with a continuous 8% increase in share of flexible natural gas-fueled and 11% decrease in share of coal-fired generators. We observe that investments in natural gas-fueled generators have been dominated by flexible generators with inflexible share staying rather constant across the time frame of our study. With investments being focused only on intermittent and flexible natural gas technologies, Table 1.2 underscores the importance of studying the interplay between these generation technologies as they are the main forces reshaping the U.S.

the latest pattern and switching this pattern will not create meaningful variation in our data, therefore, our findings are robust to changes in daily demand pattern.

[12]Since all of our variables are strictly positive, we divide every observation of any given variable by the maximum value of that variable.

generation capacity portfolio.

We organize our variables into 5 technology-dependent groups of, wind, solar, flexible, NG-flexible, inflexible, and one non technology-dependent group as summarized in Table 1.6. The reported summary statistics are for regional values aggregated over 3-year time windows and are percentages scaled adopting maximum absolute scaling approach. This final sample that we will use in our analysis contains 168,114 region-time period observations. One important observation from this table is that solar and wind prior and current generation capacity investments are observed only in the top quartile of our sample while flexible natural gas-fueled generation capacity investments are present across our sample. This observation implies how solar and wind investments are both a function of technological advancements and geographical location. This observation is further corroborated by Figure 1.A1 where we illustrate states' generation capacity investment heterogeneity and that widespread solar investments specifically started only after year 2010.

1.5 Models and Estimation Strategy

In this section, we provide an overview of the estimation models we developed to test our hypotheses. Next, we identify several empirical concerns and address them in following sub-sections.

1.5.1 Flexible Investment Impact on Intermittent Investment Model

We construct the following multi-level OLS model represented in Equation (1.1) to evaluate the hypothesis we developed in Section 1.3.1:

$$\begin{aligned} ln[\mathbf{Investment}_{r,e,t'}^{Intermittent}] = \alpha_0 + \alpha_1 \cdot ln[\text{Prior Investment}_{r,t}^{Flexible}] + \\ \boldsymbol{\alpha_2} \cdot ln[\mathbf{Controls}_{r,e,t}] + \epsilon_{r,e,t} \end{aligned} \tag{1.1}$$

In Equation (1.1), e denote the types of intermittent generation technologies where $e \in \{\text{Solar}, \text{Wind}\}$. Moreover, r denotes the region, and t represents the current 3-year time period while t' identifies future 3-year period($t+1$, $t+2$, $t+3$). We add a suite of control variables shared for both solar and wind investments as dependent variables detailed in Table 1.9. In addition, depending on the dependent variable (wind or solar future investments), we also add solar radiation, wind speed, prior investments in solar or wind technologies, solar or wind capacity utilization, and solar or wind work-in-progress index variables. Lastly, we add several interaction terms between our control variables that are measured on different granularity levels to allow for a more nuanced and flexible relationship between these variables. We provide a list of these interaction terms used in each estimation model in Table 1.10.

We test hypotheses 1 and 2 developed in Section 1.3.1 by separately estimating α_1 from Equation (1.1) for different wind and solar dependent variables. With this estimation, we evaluate how prior investments in flexible technology at a focal region with a unique zip code at it's center in the current period can influence the future investments in wind and solar technologies.

1.5.2 Intermittent Investment Impact on Flexible Investment Model

To test our second set of hypotheses, we develop a multi-level OLS model (Equation (1.2)) that is similar to the one we developed before:

$$
\begin{aligned}
ln[\text{Investment}_{r,t'}^{NG-Flexible}] = \beta_0 + \boldsymbol{\beta_1} \cdot ln[\textbf{Prior Investment}_{r,e,t}^{Intermittent}] + \\
\boldsymbol{\beta_2} \cdot ln[\textbf{Controls}_{r,e,t}] + \nu_{r,e,t}
\end{aligned}
\tag{1.2}
$$

We estimate $\boldsymbol{\beta}_1$ from Equation (1.2) to test hypotheses 3 and 4 from Section 1.3.2. With this estimation we identify how prior investments in wind and solar technologies may influence future investments in NG-flexible technology. One distinction of this model with the one we showed previously is the composition of the control variables. We provide a list of control variables as well as interaction terms in Tables 1.9 and 1.10.

1.5.3 Estimation Strategy

We recognize several concerns in estimating $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$, therefore, we develop an estimation strategy to overcome these challenges and concerns. Particularly we address three empirical concerns: *i)* RPS endogeneity, *ii)* Unobserved continuous heterogeneity of regions, and *iii)* Investment decisions endogeneity. RPS requirements are designed to directly stimulate investments in renewable generation technologies while indirectly dis-incentivizing investments in fossil fuels. The design of such programs depends on each states' political orientation as well as their existing renewable potential. Existing renewable potential of a state can be thought of as their existing capacity portfolio mix, solar radiation, and wind speed. With our definition of regions from Section 1.4.2.2 we naturally expect to have un-observed heterogeneity due to omitted variables. This unobserved heterogeneity not controlled for leads to biased estimates of $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$. For investments, the endogeneity concern arises due to un-observable confounding factors potentially influence investments in both intermittent and flexible technologies. In following sub-sections we explain our estimation strategy in greater details.

1.5.3.1 RPS Correction

The RPS annual obligations set by state legislators can correlate with investments in all types of generation technologies. To circumvent bias in our estimations of Equations (1.5) and (1.6), we first regress the RPS obligations on the drivers of the solar and wind investments as well as their work-in-progress index variables along with a suite of control variables. Equation (1.3) represents the first step of our first stage:

$$ln[\text{RPS}_{r,t}^{Adopted}] = \rho_0 + \boldsymbol{\rho_1} \cdot ln[\textbf{Drivers}_{r,e,t}^{Intermittent}] + \boldsymbol{\rho_2} \cdot ln[\textbf{Controls}_{r,e,t}] + \tau_{r,t} \tag{1.3}$$

In Equation (1.3), $\text{RPS}_{r,t}^{Adopted}$ denotes the values of RPS obligations for each region r measured for each 3-year time period t. Vector of $\textbf{Drivers}_{r,e,t}^{Intermittent}$ represents driving factors including: prior solar and wind generation capacity investments, actual electricity generation by solar and wind sources, solar radiation, and wind speed. $\textbf{Controls}_{r,e,t}$ is a vector of control variables measured on different aggregation levels and are listed in Table 1.9.

We run this model only for regions that have adopted the RPS program (i.e., $\text{RPS}_{r,t}^{Adopted}$). Then we use the estimated coefficients to extrapolate RPS requirements for all regions including the ones that have not adopted the RPS program. We use the RPS residual values $\widehat{\tau}_{r,t}$ replacing the original variable in Equation (1.5) from the controls vector to circumvent RPS endogeneity. The smaller values of RPS residual indicate a region with less restrictive RPS program while high RPS residuals are regions with most restrictive RPS programs. We update Equation (1.5) as the following:

$$ln[\mathbf{Investment}^{Intermittent}_{r,e,t'}] = \alpha_0 + \alpha_1 \cdot ln[\text{Prior Investment}^{Flexible}_{r,t}] + \boldsymbol{\alpha_2} \cdot ln[\mathbf{Controls}_{r,e,t}] + \underbrace{\alpha_3 \cdot ln[\widehat{\tau}_{r,t}]}_{\text{RPS Residual}} + \tilde{\epsilon}_{r,e,t} \tag{1.4}$$

Note that we estimate Equation (1.3) over our whole U.S. sample.

1.5.3.2 Two-Step Grouped Fixed Effect (GFE)

The influence of unobserved time-varying heterogeneity on estimating generation capacity investment effects is well justified with these decisions taking place in a dynamic context of competing economic and political forces with only a subset of them observable to the econometrician. Bonhomme and Manresa (2015) emphasize the practical challenges of flexibly modeling unobserved time-varying heterogeneity while keeping a parsimonious model specification. Since we expect regions in our sample to exhibit substantial heterogeneity, we use the *"two-step grouped fixed effect"* estimation approach proposed by Bonhomme et al. (2022) to alleviate the estimation bias due to the unobserved heterogeneity. Using this approach in a first step we discretize the continuous unobserved heterogeneity using a combination of the *"principal component analysis"* (PCA) and *"k-means"* clustering approach to identify the latent clustering structure of unobserved heterogeneity. Second, we use the identified clusters as fixed-effects and run our estimation models. We update Equations (1.1) and (1.2) in the following to reflect our two-step estimation approach:

$$ln[\mathbf{Investment}^{Intermittent}_{r,e,t'}] = \alpha_0 + \alpha_1 \cdot ln[\text{Prior Investment}^{Flexible}_{r,t}] + \boldsymbol{\alpha_2} \cdot ln[\mathbf{Controls}_{r,e,t}] + \kappa_r + \epsilon^*_{r,e,t} \tag{1.5}$$

$$ln[\text{Investment}^{NG-Flexible}_{r,t'}] = \beta_0 + \boldsymbol{\beta_1} \cdot ln[\mathbf{Prior\ Investment}^{Intermittent}_{r,e,t}] + \boldsymbol{\beta_2} \cdot ln[\mathbf{Controls}_{r,e,t}] + \kappa_r + \nu^*_{r,e,t} \tag{1.6}$$

In Equations (1.5) and (1.6), $\kappa_{r,e}$ is the clustering fixed effect that we add to both models for estimation. We use PCA and k-means clustering approach as it is proved that using PCA for dimension reduction prior to clustering data using k-means provides more accurate clustering (Ding and He, 2004). In the first step of this approach we run the PCA on our sample using 11 regional features including: RPS obligation residual, wind speed, solar radiation, wind intermittency, solar intermittency, natural gas pipeline capacity, utility ownership, total generation capacity, population, energy intensity, and residential price. Then we use the first principal components that explain more than 70% of cumulative variation as our sample's dimensions to run k-means clustering algorithm and identify clusters.

As an extension we estimate $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ separately over k-means clusters to illustrate the distribution of effects across clustering factors as well as showcasing possible non-linear effects.

1.5.3.3 Investment Decisions Correction

Firstly, our panel data structure and leading investment dependent variables allow us to alleviate investment decisions endogeneity concerns arising in our original model. We extend our analysis by creating a new set of independent investment variables for regions that present *expected* investments using instrumental variables. We adopt this approach for three reasons: First, investors decision to invest at a certain location can be influenced by both the prior investments in generation capacity and their expectations of the future investments of their competitors. Second, realized investments have different predictability and investors may adjust their investment decision responses based on how much the realized investment in a given region was *expected*. This expectation signals how quickly developing projects are becoming operational that ultimately impacts investors' decision-making process. Third, investors choose to invest at a location that maximizes their short and long term profits. In our data we have a censored investment that are the ones which maximized investors' profits. This censorship can potentially lead to correlations between the *investment* variables and the *error* terms in Equations (1.1) and (1.2) (i.e., $\epsilon_{r,e,t}$ and $\nu_{r,e,t}$).

We use our clustered sample to run OLS regression models on the investment dependent variables using instrumental variables. We use the following OLS models detailed in Equations (1.7) and (1.8) to find predicted investment values for wind, solar and flexible natural gas fueled generators:

$$ln[\mathbf{Investment}^{Intermittent}_{r,e,t}] = \theta_0 + \boldsymbol{\theta_1} \cdot ln[\mathbf{Instruments}^{Intermittent}_{r,e,t}] + \theta_2 \cdot ln[\widehat{\tau}_{r,e,t}] + \boldsymbol{\theta_3} \cdot ln[\mathbf{Controls}_{r,e,t}] + \zeta_{r,e,t} \tag{1.7}$$

$$ln[\text{Investment}^{NG-Flexible}_{r,t}] = \lambda_0 + \boldsymbol{\lambda_1} \cdot ln[\mathbf{Instruments}^{NG-Flexible}_{r,t}] + \boldsymbol{\lambda_2} \cdot ln[\mathbf{Controls}_{r,t}] + \eta_{r,t} \tag{1.8}$$

$\mathbf{Investment}^{Intermittent}_{r,e,t}$ represent the amount of investments a region r receives for generation technology type $e \in \{\text{Solar}, \text{Wind}\}$ at time t. Similarly, $Investment^{NG-Flexible}_{r,t}$ denotes the amount of investments at region r, in flexible natural gas-fueled generation technology at time t. In Equation (1.7) we include the following instruments: solar radiation, wind speed, work-in-progress indices while we use natural gas price and pipeline capacity instruments in Equation (1.8). We add several control variables listed in Table 1.9. Since these controls are measured at various levels and our dependent variables are measured on the region level, to have better predictions of investments as part of our controls, we interact variables with state variation with those with regional variation (these variables are listed in Table 1.10). These equations are estimated separately across identified clusters and predictions are made inside each cluster. We then update Equations (1.6) and (1.4) in the following to test our hypotheses with extended analysis of predicted investment variables:

$$ln[\mathbf{Investment}^{Intermittent}_{r,e,t'}] = \alpha_0 + \alpha_1 \cdot ln[\text{Predicted Investment}^{NG-Flexible}_{r,t}] + \boldsymbol{\alpha_2} \cdot ln[\mathbf{Controls}_{r,e,t}] + \underbrace{\alpha_3 \cdot ln[\widehat{\tau}_{r,t}]}_{\text{RPS Residual}} + \underbrace{\kappa_r}_{\text{Cluster FE}} + \epsilon^{\backprime}_{r,e,t} \tag{1.9}$$

$$ln[\text{Investment}^{NG-Flexible}_{r,t'}] = \beta_0 + \boldsymbol{\beta_1} \cdot ln[\textbf{Predicted Investment}^{Intermittent}_{r,e,t}] + \\ \boldsymbol{\beta_2} \cdot ln[\textbf{Controls}_{r,e,t}] + \underbrace{\kappa_r}_{\text{Cluster FE}} + \nu^{`}_{r,e,t} \tag{1.10}$$

In our extended analysis, we additionally estimate Equations (1.9) and (1.10) for each cluster and evaluate the distribution of effects across clusters. Further, we also run the raw current investments separately instead of the predicted investments to show that our results are robust to alternative specifications. We have summarized all estimation models with their corresponding dependent and independent variables in Table 1.8.

1.6 Results

In this section we provide a summary of our findings following our empirical strategy starting from RPS correction, next we detail our clustering results followed by an explanation of our main findings for hypotheses 1–4. Later, we detail our extended analysis. Lastly, we provide the managerial implications of our findings along with the limitations of our study.

1.6.1 RPS Residuals

We summarize the RPS endogeneity correction model results in Table 1.B1. We use the parameters estimated on a sub-sample of regions that has not adopted RPS program to predict RPS obligations for the whole U.S. sample. Using the predicted values of RPS, we calculate the RPS residuals. We plot the residuals for each state in Figure 1.3. Darker shaded areas represent cases where the actual RPS obligation is higher than the predicted values signaling a more stringent RPS program compared to the country's average expected RPS obligations. Meanwhile, lighter shaded regions are the cases where the actual RPS obligation is below the predicted RPS obligation that characterizes regions with less stringent RPS program.

1.6.2 k-means Clustering

We use the principal component analysis as a dimensionality reduction method using 11 variables that we listed in Section 1.5.3.2. We found the first 5 principal components (PCs) explain 78% of the total variation in our sample data set. Therefore , we used these 5 PCs as an input to our k-means clustering algorithm. To identify the number of clusters that assures capturing homogeneity within clusters, we run the k-means algorithm to generate cluster sizes of 1–20 and for each we calculated the *"within-cluster sum of squares"* (WCSSs). Then we re-scaled these variables with using the maximum absolute value approach using the value of WCSS for only having one cluster and present this in Figure 1.4. For example we illustrate that having 2 clusters reduces 25% of WCSS compared to having 1 clusters instead. We use 16 clusters based on the elbow method where a point is selected from which there is no significant reduction in the values of WCSS. Table 1.7 represents the coefficient of variation[13] of the 11 clustering features along with the cluster sizes that is the number of region-year observations inside each cluster. In this table we have sorted our clusters in an ascending order based on the average amount of clusters' RPS obligation. Our smallest cluster has more than four thousand observations with our largest cluster covering more than twenty three thousand observations. Additionally, in Figure 1.5 we illustrate the distribution of clustering features within each cluster with clusters sorted with smallest mean RPS residual value to the larges mean RPS value. We observe significant heterogeneity across the clusters along with small variation within clusters justifying our clustering approach in estimating main effects.

1.6.3 Main Results (GFE Estimates)

We illustrate a summary of the GFE estimates for our research hypotheses in Table 1.11. We find strong empirical support for hypotheses 1–3. The estimated coefficients are the elasticity of future investments on prior investments. Our estimation suggests if regions prior investments in flexible generation capacity increases by one percent, then they will get 0.145% and 0.061% more future investment ratio compared to the largest investment we observe in our data set in wind and solar generation capacity respectively. Further, we find if regions' prior investment in wind generation capacity increases by one percent they will on average get 0.091% more future

[13] We calculate the coefficient of variation for each cluster by dividing the standard deviation of a given feature by the average of that feature within that cluster then multiply it by 100. This statistic shows us how our data is dispersed around the mean for each clustering feature.

investments ratio compared to maximum investment observed in flexible generation capacity. We find the coefficient for solar prior investments is negative as opposed to our hypothesis. In summary we found empirical evidence supporting that flexibility drives intermittent investments regardless of the generation technology while the other way around is only true for wind generation technology.

1.6.4 Extended Analyses Results

As we explained in Section 1.5.3.2 we extend our analysis to elicit managerial implications from our findings as well as testing the robustness of our findings in the following sub-sections. First we show the results of individually estimating coefficients across different clusters for each hypothesis test. Second, we quantify the impacts of prior investments on future investments for the individually estimated models and demonstrate how they are influenced by high(or low) levels of clustering features. Third, we depict how our findings might change if we look at alternative specifications with current and expected investment independent variables.

1.6.4.1 Between-Cluster Estimates

In Table 1.12 we demonstrate the correlations of future and prior investments in wind, solar, flexible and NG-flexible generation capacity based on our 4 research hypotheses as we developed in Section 1.3. We depict the correlations for the whole sample along with the distribution of correlations across clusters. We observe an overall positive correlation for all 4 hypotheses in the whole sample, while the distribution of correlations across clusters signal the existence of a more nuanced relation between our DVs and IVs. Even though we observe some negative correlations for all hypotheses across all clusters, but the correlations in the top quartile of our clusters is always positive.

Justified by the clusters' heterogeneity illustrated in Figure 1.5 and the correlation distribution across k-means (Table 1.12) we estimate prior investment coefficients across k-means clusters for each hypothesis to explore potential relationship nuances. We summarize the coefficients of the log-log models we estimated for each individual cluster evaluating hypotheses 1–4 in Table 1.13. Similar to our fixed effect estimations, we observe that the majority of the coefficients are positive for all clusters across hypotheses 1–3 validating our hypotheses with very few exceptions. In contrast, for hypothesis 4 we observe that the results are more mixed with 7 out of 13 significant coefficients to be negative. This finding emphasizes the importance of regional heterogeneity and that there is still empirical evidence for our hypothesis 4 but it relies on clusters' attributes that we will explore in more details in the following sub-sections. Table 1.13 also reports how well our models are fitted for each cluster by reporting r-squares for each estimation model of clusters. The values of r-squares are between 7.5% and 73.7% of explained variation.

1.6.4.2 Impact Quantification

To better interpret our findings, we conduct a series of *"ceteris paribus"* counterfactual analyses (Bray et al., 2019) quantifying the impact of prior investments on future investments. We conduct these counterfactual analyses by holding all else constant and comparing two predictions of same models assuming prior investments to be as they were against no prior investments. After calculating the impacts across clusters for a sub-set of our clustering features including RPS residuals, NG pipeline capacity, and utility ownership we find the mean value of each clusters for a given clustering feature. Next, we sort all clusters' average values in an ascending order and find the median value. All clusters with averages below the median of a given clustering feature is deemed to encompass regions with *low* levels of that clustering feature while those above the median are the ones with *high* levels of that clustering feature. Lastly, we take the average of all impacts for clusters that belong to the low and high levels and report them in Table 1.14. Values in this table are percentages of impacts. For example, for locations with low RPS residual levels, regions with prior investments in flexible technologies have 3.46% more future investments in wind generation technology compared to the case without any prior investment in flexible sources. Therefore, in this table positive values identify a complementary relationship meaning that more prior investment in one generation technology leads to more future investments in the other.

From Table 1.14 we observe a complementary relationship between prior investments and future investments at locations with less stringent RPS programs (shown by positive impacts across) compared to regions with highly stringent RPS programs. Moreover, regions with high levels of natural gas infrastructure development (high levels of natural gas pipeline capacity access) showcase a complementary relationship between prior investments and future investments while regions having low levels of this specific infrastructure development represent a substitute relationship. Lastly, going from regions with low levels of utility firms ownership of the total generation capacity to those with higher levels of utility firms ownership, we observe a stronger complementary relationship between prior investments in flexible generation capacity and future intermittent investments. More specifically, having average impact of 0.65% and 0.19% for wind and solar future investments respectively in regions with low utility ownership to 6.07% and 5.37% in regions with higher utility ownership.

As we argued in Section 1.3, solar and wind generation technologies have dissimilar intermittency patterns that potentially is one of the driving forces of the heterogenous relationship of these technologies and flexible generation technologies. We evaluate this by examining the distribution of the impacts for each hypothesis under high and low levels of regional renewable potential. The intermittency pattern dissimilarity effects are amplified under these two extreme levels as measured by average solar radiation and wind speed for each cluster. Figure 1.6 demonstrate the distribution of impacts classified for two groups of high and low renewable potential. High renewable potential group encompasses regions (within clusters) with above median average wind speed and solar radiation. Conversely low renewable potential group involves regions with below median average wind speed and solar radiation. We observe from Figure 1.6 a sharp distinction between high and low groups' median impact. Wind and solar future investments benefit more from flexible prior investments in regions with higher renewable potential. We observe a similar effect for wind prior investments on future flexible natural gas-fueled generation capacity investments.

1.6.4.3 Alternative Specification Estimates

In addition to our main models, we have used two substitutes for our main independent variables (i.e., prior investments) with control variables and interaction terms listed in Tables 1.9 and 1.10 respectively. These substitutes are *"current investments"* and *"expected investments"* as we discussed in Section 1.5.3. Current investments are the additions to the prior investments while expected investments are the predictions made for investments in each generation technology from OLS estimation models using instrumental variables explained in Equations (1.7) and (1.8). We summarize these estimation results in Table 1.B2. We provide the GFE estimation results of these alternative specifications in Table 1.C1 which illustrate consistent results for the influence of wind prior investments on future NG-flexible investments. These results may be due to the timing of realized investments with respect to the future investments. This implies the possibility of future investments to be a function of older investments realized as opposed to newer investments.

1.7 Conclusion

We contribute to the existing body of knowledge that examines the interplay of investments in different electricity generation technologies. To the best of our knowledge this is the first empirical study that specifically examines the role of operational flexibility in these investment decision from a macro level. Our study offers a fresh perspective unraveling investment dynamics in the U.S. electricity industry. We conduct an empirical analyses to illustrate that locations with prior investments in flexible electricity generation technologies are visited with more future investments in wind and solar electricity generation technologies. We also found that prior investments in wind electricity generation technology leads to more future investments in natural gas-fueled flexible electricity generation technology. These findings underscore the inevitable role of the natural gas-fueled flexible generators in greening the U.S. electricity grid in the foreseeable future along with the possibility that more renewable investments can be translated into more future emissions with wind investments creating strong positive investment signals for natural gas-fueled flexible electricity generators.

Our research provides significant managerial implications for both policy makers and investors. For policy makers, the main take away of this study is that a potential pathway to accelerate the transition towards a green grid is through incentivizing investments in certain types of conventional power plants that are flexible enough to offset supply-side fluctuations. Essentially, our empirical approach that leverages spatially granular data, emphasizes the importance of technology-specific policies while taking geographical heterogeneity into account suggesting a more disaggregated technology- and geographical-dependent policy design and enforcement. Additionally, our findings showcase the necessity of investments in appropriate natural gas-related infrastructure to make sure that a smooth transition toward a greener electricity grid takes place. Lastly, our findings suggest that electricity market structures also play an important role in moderating the investment dynamics between flexible and intermittent generation technologies. Current U.S. market structures range from monopolistic markets to highly competitive ones and we found that the complementary relation is much stronger in regions with high levels of utility firm ownership that characterizes more monopolistic markets. This finding emphasizes the importance of competition levels in markets and how it influences investment dynamics in the electricity industry.

Our study provides crucial insights for investors in the U.S. electricity industry, particularly highlighting the interdependent nature of investments in renewable energy sources and natural gas-fueled flexible generation capacity. Our empirical analysis reveals that investments in wind energy not only contribute to the green energy transition but also stimulate further investments in flexible, natural gas-powered plants, which are essential for maintaining grid reliability and stability amidst the variability of renewable energy sources. This complementary relationship underscores the strategic value of considering the broader investment landscape rather than focusing on isolated asset classes. For investors, this implies a nuanced investment strategy that balances the push towards decarbonization with the operational imperatives of the current electricity grid. By recognizing these dynamics, investors can make informed decisions that align with both environmental goals and the practical needs of energy markets, thereby contributing to a more sustainable, reliable, and economically viable energy future.

Nevertheless, our study has several limitations. First, we adopted a regional approach to be the granularity level of our observations. To this end, we have aggregated firm level observations. With this aggregating approach, firm heterogeneity is overlooked and as a result our estimations might be less accurate compared with a study on the firm level. Second, one of our most important instrumental variables (WIP indices) are measured on the state level. Having geographical location of WIP indices can significantly improve our estimations of investment effects. Third, our study lacks granular supply, demand, and price variables that are important driving factors of investment decisions.

Briefly, our study highlights the importance of operational flexibility role in transitioning toward a greener electricity grid with a focus on natural gas-fueled flexible generation technology along with wind and solar technologies. We identified and quantified the effect of prior investments and future investments in these generation technologies. Our findings suggest that there is a great opportunity for future research to improve our understanding of investments dynamics as renewable penetrations into electricity generation portfolios increases. Specifically, one direction is to focus on the flexibility measure and take into account the heterogeneity of flexible electricity generators as not all flexible generators are the same with respect to how quickly they can respond to supply and demand fluctuations. Another direction that resolves one of our study's limitations is to obtain granular electricity price and re-evaluate similar relationships to our research. This approach will produce a more comprehensive understanding of investment dynamics in the electricity industry.

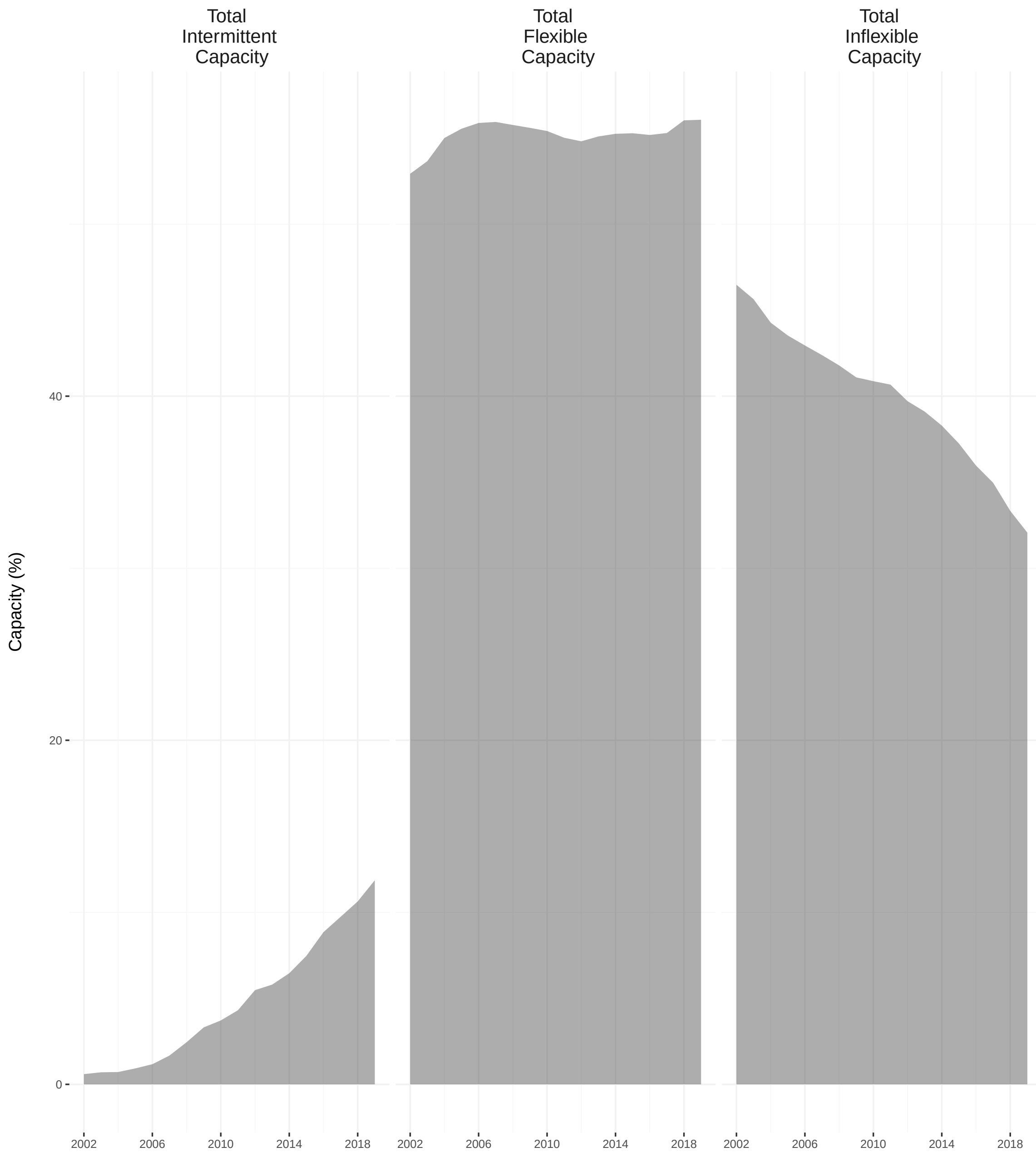
Exhibits

Table 1.1: Summary of data collection sources

	Raw Variables	Source Data Granularity Levels	Source Data Observation Frequencies	Source Data Years	Data Sources	Suorce Data Total Observations
Technology-	Main fuel type	Generator	Yearly	2002–2019	EIA-860	338.32K
dependent	Generator start-up time	Generator	Yearly	2002–2019	EIA-860	338.32K
variables	Nameplate capacity	Generator	Yearly	2002–2019	EIA-860	338.32K
	Nameplate capacity (work-in-progress)	Generator	Yearly	2002–2019	EIA-860	35.44K
	Developmen stage* (work-in-progress)	Generator	Yearly	2002–2019	EIA-860	35.44K
	Actual generation	Plant	Monthly	2002–2019	EIA-923	118.32K
	Solar radiation	Zip code	Semi-hourly	1999–2019	NSRDB	10.42B
	Wind speed	Zip code	Semi-hourly	1999–2019	NSRDB	10.42B
	NG pipeline capacity	County	Yearly	2002–2019	EIA-NG	6.64K
	NG price	National	Yearly	2002–2019	EIA-NG	18
Non technology-	Unique ID number	Generator	Yearly	2002–2019	EIA-860	338.32K
dependent		Plant	Yearly	2002–2019	EIA-860	118.32K
variables	Geographical coordinates	Plant	Yearly	2002–2019	Geocodio	118.32K
	Ownership sector	Plant	Yearly	2002–2019	EIA-860	118.32K
	Population	Zip code	Yearly	2020	Census Bureau	28.30K
	Geographical coordinates	County	Yearly	2002–2019	Opendatasoft	58.09K
	RPS obligation	State	Yearly	2002–2019	LBNL	558
	Residential price	State	Yearly	1990–2019	EIA-861	1.53K
	Energy intensity	State	Yearly	2002–2019	EIA-SEDS	918

Note: Energy Information Administration (EIA) provide publicly available data on both generator and plant levels. EIA State Energy Data System (EIA-SEDS) provides state-level time series data on three main dimensions of energy consumption, prices, expenditure, and production. EIA-Natural gas reports involve natural gas annual prices based on delivery at Henry Hub at Louisiana and are calculated based on the official daily closing prices from the New York Mercantile Exchange (NYMEX). Lawrence Berkeley National Laboratory (LBNL) collects and reports net Renewable Energy Certificates (RECs) required by each state as a part of their Renewable Portfolio Program (RPS). This requirement is net of the existing generation from eligible sources in the year prior to the RPS being enforced. Geocodio apis are used for the purpose of geocoding and reverse-geocoding to improve the accuracy of the plants' geographical location data. Opendatasoft website provides the coordinates for the center of each US county that enables us to match our county-level datasets with zip-level datasets. National Solar Radiation Database (NSRDB) captures half an hour observations of solar radiation and wind speed that are accessible via API requests. We've collected solar radiation and wind speed for each zip code in our dataset using the center coordinates of the zip codes.
*Development stages (i.e., investment stages) are, proposed, pending legislative approvals, under on-time construction, under delayed construction, postponed, and canceled.

Figure 1.1: Trends in flexible, intermittent, and inflexible electricity generation capacity of the U.S.



Note: In this plot we show how the power grid generator portfolio mix has been changing between 2002 to 2019. Each area plot shows the percentage of U.S. installed capacity that belongs to a certain type of generators given their flexibility in electricity generation.

Table 1.2: Fuel type distribution of U.S. generation capacity portfolio (%)

		2002–2004	2005–2007	2008–2010	2011–2013	2014–2016	2017–2019
Wind	Intermittent	0.64	1.24	3.16	4.97	6.41	8.20
Solar	Intermittent	0.05	0.04	0.07	0.34	1.35	2.77
Natural Gas	Flexible	32.36	35.91	36.52	37.03	38.34	40.37
	Inflexible	5.28	5.53	5.26	5.29	5.57	5.32
Coal	Flexible	6.67	6.14	5.83	5.52	4.88	3.87
	Inflexible	27.04	25.23	24.41	23.84	21.88	18.70
Petroleum	Flexible	3.85	3.58	3.50	3.13	2.82	2.45
	Inflexible	1.71	1.84	1.75	1.46	0.92	0.72
Hydro	Flexible	10.62	9.77	9.33	8.88	8.77	8.69
	Inflexible	0.00	0.00	0.00	0.00	0.00	0.00
Nuclear	Inflexible	11.77	10.71	10.18	9.54	9.08	8.90

Note: In this table we show the percentage of each energy type grouped by its generation flexibility level changing across 3-year time windows within 2002–2019. Note that in this table, the capacity with unknown flexibility is omitted as well as other smaller fractions of the total capacity coming from other renewable/non-renewable sources such as bio-fuel, bio-gas, etc. We use generators' start up time (i.e., the time it takes a generator to go from cold shutdown to being fully operational.) as a proxy to measure and categorize the flexibility levels of generation capacity. To this end, for high flexibility group this time is less than 10 minutes, moderate flexibility is 10-60 minutes, low flexibility is 1-12 hours, and for inflexible group it is more than 12 hours.

Figure 1.2: Spatial dynamics of generation capacity investment in the U.S. electricity grid (2002-2019)

Note: In the plot above we capture the geographical nature of investments in intermittent generators that are clustered in specific regions. Further, we emphasize how investments in solar and wind are location-dependent as well as how more flexible generators are mostly developed to counter the generation uncertainty of solar generators.

Table 1.3: Summary of plant geographical location data cleaning process

	Plant Observations		
	Existing	Missing	Geo-coding (Reverse geo-coding)
State and zip code	118,317	-	118,317
Longitude and latitude	117,927	390	390

Note: This table shows our process of filling in for missing geographical location data for each power plant in the 2002–2019 time frame. For all plant-level observations in our data set we have state and zip code values. However, we observe some inconsistencies in the existing longitude and latitude with the state and zip code data, therefore, we use geo-coding and reverse geo-coding to convert locations identified by state and zip code in to longitude and latitude and vice versa. We used these approaches along with the API service from Geocodio.io to create a consistent data set of all observed power plants along with their exact geographical locations.

Table 1.4: Summary of variable construction process

	Raw Variables	Formulas	Granularity Levels
Capacity investment	Nameplate capacity	$(\text{Capacity}_{g,e,t}) - (\text{Capacity}_{g,e,t-1})$	Generator
Capacity utilization	Actual generation and capacity	$(\text{Supplye}_{p,e,t})/(\text{Capacity}_{p,e,t} \times 8,760)$	Plant
Work-in-progress index	Work-in-progress stages capacity	$\sum(w_{s,e} \times \text{WIP}_{s,e,t})$	State
Utility ownership	Ownership sector and capacity	$(\sum \text{Utility-owned Capacity}_{z,t})/(\text{Total Capacity}_{z,t})$	Zip code
Capacity surplus	Actual generation and peak-to-mean ratio	$(\text{Mean Req. Cap.}_{z,t})/(\text{Peak Req. Cap.}_{z,t})$	Zip code
Intermittency level	Solar radiation, wind speed, generation	$\sum\sum(\text{Under Supply}_{z,e,h,d})$	Zip code
Regional total capacity	Nameplate capacity	$\sum(\text{Total Capacity}_{z,t})$	Regional

Note: This table summarizes our variable construction approach using the most granular levels of measurements observed in our data set. In all the formulas mentioned in the table, g denotes generator, e is the generation technology based on the main fuel (solar, wind, conventional), t represents the yearly time period, p shows power plants, while z and s denote zip code and state. Lastly, h is the hour in a day and d is the day in a year. We provide a detailed overview of the construction process of each variables mentioned in the table above in Section 1.4.2.3.

Table 1.5: Summary of variables calibration process

		Raw Data Measurement Unit	Granularity	Aggregated Granularity	Transformed Data Measurement Unit	Level of Winsorization
Technology-dependent variables	Capacity investment	Megawatt	Generator	Regional	Megawatt/Million people	90%–95%
	Capacity utilization	Ratio (unitless)	Plant	Regional	Ratio (unitless)	90%–95%
	Solar radiation	Watt/Square meter	Zip code	Regional	Watt/Square meter	-
	Wind speed	Meter/Second	Zip code	Regional	Meter/Second	-
	Intermittency level	Ratio (unitless)	Zip code	Regional	Ratio (unitless)	-
	NG pipeline capacity	Million cubic feet/Day	County	Regional	Million cubic feet/Day	95%
	Work-in-progress index	Megawatt	State	State	Megawatt/Zip code	95%
	NG price	Dollars/Million Btu	National	National	Dollars/Million Btu	-
Non technology-dependent variables	Capacity surplus	Ratio (unitless)	Zip code	Regional	Ratio (unitless)	95%
	Regional total capacity	Megawatt	Generator	Regional	Megawatt/Million people	95%
	Utility ownership	Ratio (unitless)	Zip code	Regional	Ratio (unitless)	-
	Population	Thousand people	Zip code	Regional	Million people	95%
	RPS obligation	Megawatt-hours	State	State	Megawatt-hours/Zip code	-
	Residential price	Cents/Kilowatt-hours	State	State	Cents/Kilowatt-hours	-
	Energy intensity	Kilowatt-hours/Dollar	State	State	Kilowatt-hours/Dollar	-

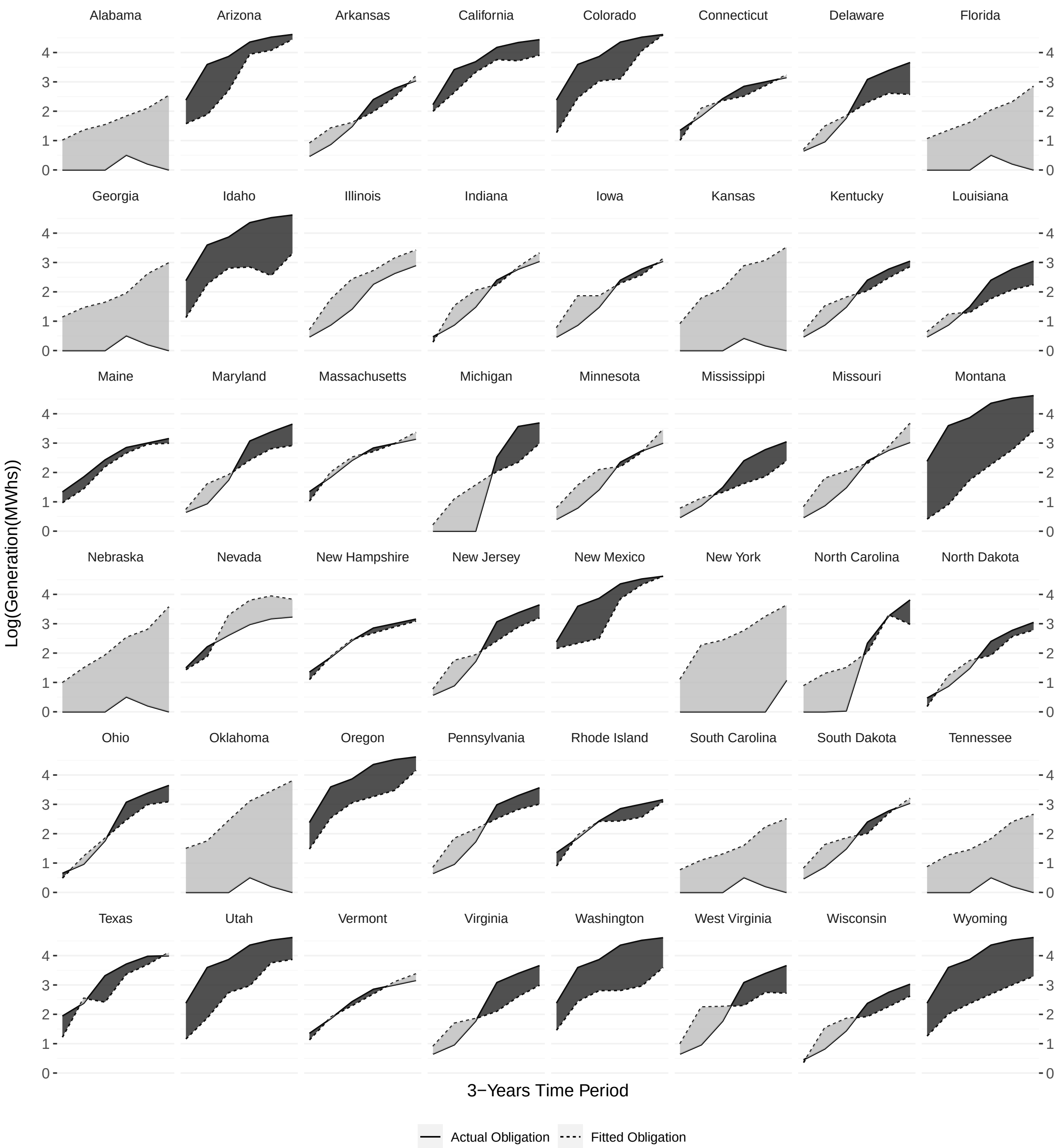
Note: This table provides a summary of our data calibration steps dividing our variables by the ones that vary with the generation technology (i.e., solar, wind, flexible, ng-flexible, and inflexible) and those independent of generation technology. Our data calibration involves 5 stages of granularity aggregation, *per capita* normalization, winsorization, maximum absolute re-scaling, and log-transformation. In this table we are reporting the first 3 stages as they differ across variables. However, all variables reported in this table, will lastly go through the maximum absolute re-scaling and log-transformation. Note all variables are aggregated over 3-year time windows prior to going through the calibration process.

Table 1.6: Sample descriptive statistics

		Mean	*Q1*	*Q2*	*Q3*	*Std. Dev*
Wind technology variables	Prior investment	11.4	0.0	0.0	8.7	25.1
	Current investment	6.6	0.0	0.0	2.2	18.6
	Capacity utilization	10.5	0.0	0.0	11.6	20.9
	WIP index	19.6	1.2	8.1	26.1	26.3
	Wind speed	48.1	25.4	48.0	74.3	26.1
	Wind intermittency	32.2	25.0	29.3	36.0	12.3
Solar technology variables	Prior investment	8.1	0.0	0.0	1.8	22.4
	Current investment	5.2	0.0	0.0	1.2	16.7
	Capacity utilization	7.5	0.0	0.0	3.2	19.8
	WIP index	9.1	0.0	0.2	3.9	23.3
	Solar radiation	73.0	66.2	70.7	78.9	8.7
	Solar intermittency	96.0	94.9	96.1	97.2	1.6
Flexible technology variables	Prior investment	38.6	20.1	32.8	51.8	25.6
	Capacity utilization	38.7	17.9	33.1	52.8	26.9
NG-flexible technology variables	Current investment	16.7	0.0	3.0	23.1	26.1
	NG price	61.8	42.5	56.3	76.7	21.7
	WIP index	33.6	13.5	28.1	45.3	26.0
	NG pipeline capacity	33.7	7.4	25.4	53.0	29.9
Inflexible technology variables	Prior investment	30.2	7.5	23.5	44.2	27.5
	Current investment	7.8	0.0	0.0	2.5	20.2
	Capacity utilization	40.9	19.4	38.4	59.2	27.7
Non technology-dependent variables	RPS obligation	17.4	0.6	8.2	24.4	24.1
	Energy intensity	39.2	29.5	36.8	48.3	15.4
	Utility ownership	18.0	8.7	16.8	24.9	11.5
	Regional total capacity	39.9	21.2	35.0	53.3	25.0
	Population	38.6	13.6	28.1	58.4	31.2
	Residential price	53.7	42.9	52.5	59.9	14.3
	Capacity surplus	44.0	27.7	39.4	55.5	23.2

Note: In this table we present a comprehensive list of all variables we include in our analysis with their descriptive statistics. These variables are divided into generation technology-dependent and non technology dependent groups of variables. We report the values of these variables after winsorization, absolute maximum re-scaling, and multiplying them by 100 so all the values shown are in percentages.

Figure 1.3: Renewable portfolio standard (RPS) obligation residuals of an OLS model



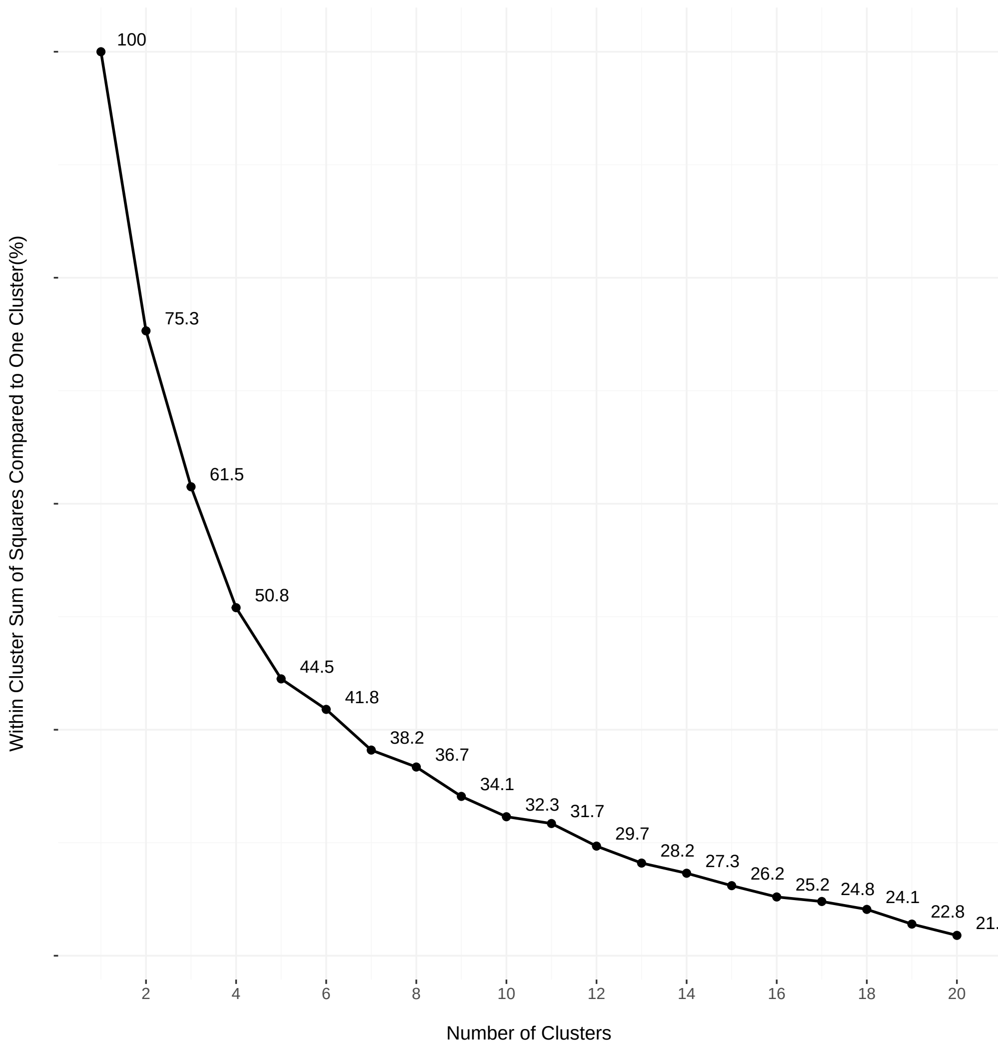
Note: This figure represents the residuals of the fitted values for regions renewable energy certificate (REC) requirements that corresponds to each state's unique RPS program. The darker shaded areas show the under-expected obligations while the lighter shaded areas are the over-expected obligations. These demonstrate when states have less and more stringent RPS programs respectively. We estimated the model first only using locations with active RPS requirements and then used the estimated parameters of that model to find the expected requirements of locations without an active RPS program.

Table 1.7: Regional features coefficient of variation across k-means clusters

	Cluster Size	RPS Residual	Solar Radiation	Solar Intermittency	Wind Speed	Wind Intermittency	NG Pipe Capacity	Utility Ownership	Generation Capacity	Total Population	Energy Intensity	Residential Price
Cluster 1	4,880	−22.7	3.8	0.6	46.2	20.6	40.6	47.4	44.8	67.8	11.6	8.1
Cluster 2	20,225	−536.2	6.1	0.8	22.0	10.6	50.7	32.1	48.8	51.7	14.8	17.5
Cluster 3	4,923	1927.8	4.3	0.9	24.7	8.9	44.5	35.3	40.9	48.6	20.5	13.8
Cluster 4	9,245	233.9	3.6	0.9	36.1	17.1	38.2	47.0	44.7	32.6	21.5	16.2
Cluster 5	5,112	210.0	5.7	0.8	62.6	25.0	43.6	36.1	49.0	60.8	30.4	17.7
Cluster 6	13,538	194.2	4.7	0.9	33.9	22.2	83.9	53.0	43.7	43.5	25.7	10.6
Cluster 7	6,264	135.3	3.0	0.7	33.2	14.9	43.6	35.7	28.4	14.8	19.5	13.9
Cluster 8	17,346	121.0	2.9	0.4	57.9	19.2	18.6	84.4	33.5	48.6	18.8	12.9
Cluster 9	9,416	121.1	4.5	1.0	77.3	24.4	90.5	54.5	52.9	45.9	29.8	19.3
Cluster 10	7,853	189.2	4.8	0.9	19.0	16.0	81.6	41.7	60.9	49.1	29.1	18.9
Cluster 11	4,749	108.2	2.3	0.3	5.6	6.1	38.3	18.4	39.9	30.5	20.1	16.4
Cluster 12	7,521	115.2	2.7	0.5	51.8	16.8	80.2	55.1	42.4	77.3	18.7	11.4
Cluster 13	23,403	49.1	3.8	0.6	17.4	25.6	89.6	51.0	60.0	56.0	14.5	13.6
Cluster 14	15,392	126.3	11.5	1.7	24.6	31.3	157.3	71.4	88.2	125.1	28.6	20.7
Cluster 15	9,271	88.1	7.6	1.3	28.8	27.3	83.2	48.8	71.4	84.3	31.3	16.7
Cluster 16	8,976	43.4	9.1	1.0	52.9	23.6	75.6	36.9	87.5	77.1	31.1	19.9

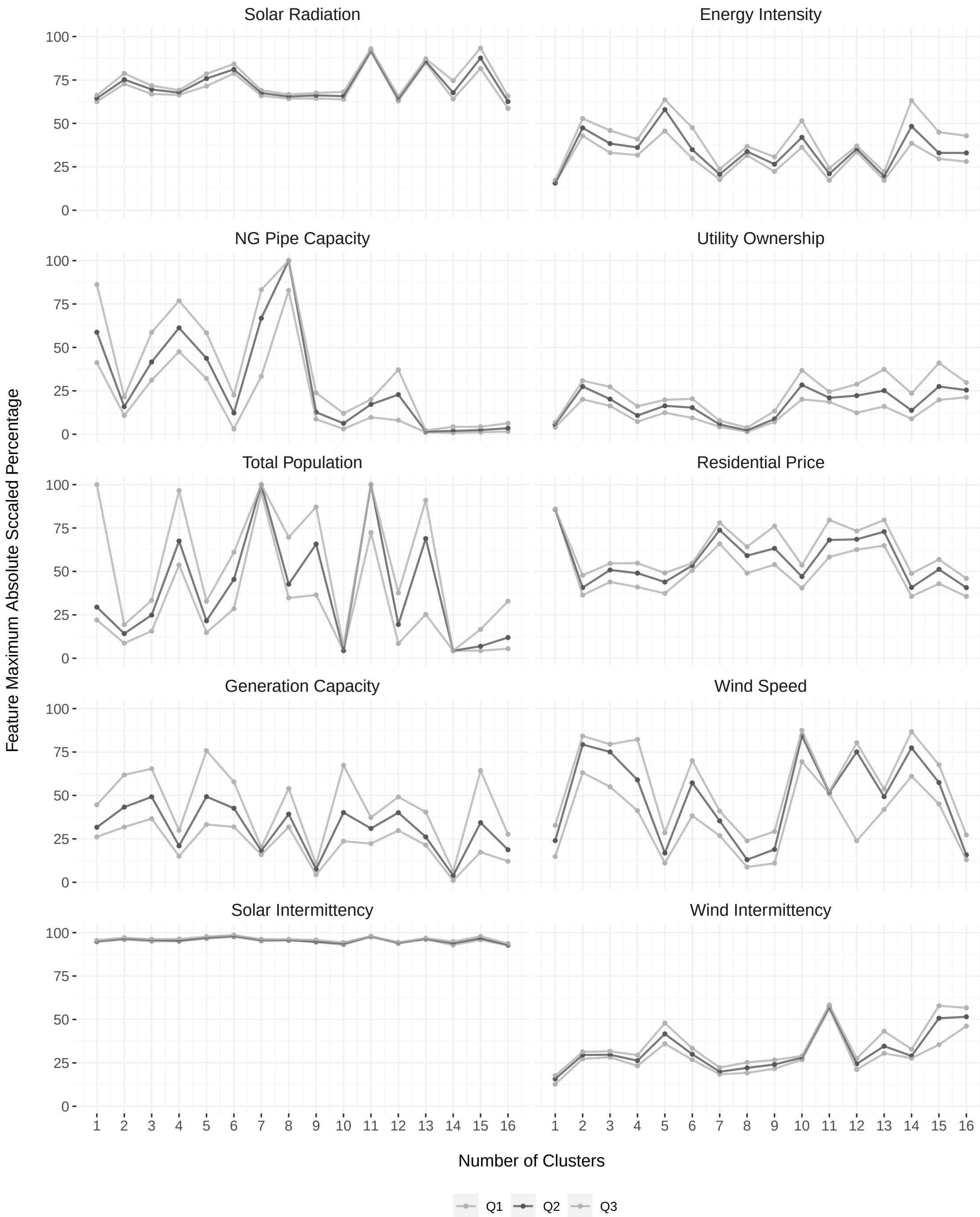
Note: This table represents the coefficient of variation for all nine features across 16 K-means clusters of regions. For the purpose of clustering regions with zip codes' as their centers, we use 9 features including solar radiation (measured by Global Horizontal Index) mean, wind speed mean, total natural gas pipeline capacity available within a focal region, ratio of generation capacity owned by utility companies in a focal region (i.e., utility ownership), renewable portfolio standard obligation in a focal region, energy intensity of states, electricity residential prices on the state level, regional populations, and regional total capacity available. Cluster size indicates the number of region-period (i.e., 3-years time windows) observations we have in each K-means clusters.

Figure 1.4: K-means within-cluster sum of square (WCSS) elbow plot



Note: In this plot we illustrate the percentage of within-cluster sum of squares (WCSS) compared to the one-cluster only case (with one-cluster case having the maximum WCSS). For example, the point on having 2 clusters shows that compared to having only 1 cluster, we can reduce the within-cluster sum of squares by 25%. We decided to have 16 clusters as the cumulative reduction of within-cluster sum of squares for more than 16 clusters is marginal compared to having 16 clusters.

Figure 1.5: Regional feature distributions across k-means clusters



Note: In this plot we show the distribution of clustering features across 16 k-means clusters. In this plot, values are re-scaled using the maximum absolute re-scaling method and are measured in percentages. The unique shape of each plot amplifies the uniqueness of each cluster with respect to the clustering features. Moreover, the low difference between the first and third quartile represent that observations within the same cluster are sufficiently homogeneous.

Table 1.8: Summary of hypotheses testing models

			Dependent Variables Future Investments		
Hypotheses	Main Independent Variables		Wind	Solar	NG-Flexible
H1 and H2	Flexible Investment	Pior	✓	✓	-
	NG-Flexible Investments	Current	✓	✓	-
		Expected	✓	✓	-
H3 and H4	Wind Investments	Prior	-	-	✓
		Current	-	-	✓
		Expected	-	-	✓
	Solar Investments	Prior	-	-	✓
		Current	-	-	✓
		Expected	-	-	✓

Note: In this table we show a summary of all the models we run using our estimation framework test our research hypotheses. In the first column we indicate how each combination of independent and dependent variables belong to a hypotheses group (i.e., H1–H4). The second column denotes the main independent variable of interest that are different investments including, prior, current, and expected investments. In the last 3 columns we show our dependent variables that are future investments in each generation technology and check marks identify the models we estimated.

Table 1.9: Detailed overview of control variables used in estimation models

		RPS Obligation	Prior Investments			Current Investments			Expected Investments			Current Investments Correction		
		Correction	Wind	Solar	Flexible	Wind	Solar	NG-Flexible	Wind	Solar	NG-Flexible	Wind	Solar	NG-Flexible
Wind technlogy	Capacity utilization	Yes	Yes	–	–	Yes	–	–	Yes	–	–	Yes	–	–
vaiables	WIP index	Yes	Yes	–	–	Yes	–	–	Yes	–	–	Yes	–	–
	Wind speed	Yes	Yes	–	–	Yes	–	–	Yes	–	–	Yes	–	–
	Wind intermittency	Yes	Yes	–	–	Yes	–	–	Yes	–	–	Yes	–	–
Solar technology	Capacity utilization	Yes	–	Yes	–	–	Yes	–	–	Yes	–	–	Yes	–
variables	WIP index	Yes	–	Yes	–	–	Yes	–	–	Yes	–	–	Yes	–
	Solar radiation	Yes	–	Yes	–	–	Yes	–	–	Yes	–	–	Yes	–
	Solar intermittency	Yes	–	Yes	–	–	Yes	–	–	Yes	–	–	Yes	–
Flexible tech. variables	Capacity utilization	Yes	–	–	–	–	–	–	–	–	–	–	–	Yes
NG-Flexible technology	NG price	Yes	–	–	Yes	–	–	Yes	–	–	Yes	–	–	Yes
variables	WIP index	Yes	–	–	Yes	–	–	Yes	–	–	Yes	–	–	Yes
	NG pipeline capacity	Yes	–	–	Yes	–	–	Yes	–	–	Yes	–	–	Yes
Inflexible technology	Prior investment	–	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	–	–	–
variables	Capacity utilization	–	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	–	–	–
Non-technology	RPS residuals	–	Yes	Yes	–	Yes	Yes	–	Yes	Yes	–	Yes	Yes	–
dependent	Energy intensity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
variables	Utility ownership	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Regional total capacity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Population	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Residential price	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Capacity surplus	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table provides a list of all control variables we used in each estimation model. On each column we specify the type of estimation model and the corresponding generation technology.

Table 1.10: Detailed overview of interaction terms used in estimation models

Interaction Terms	RPS Obligation Correction	Prior Investments Wind	Prior Investments Solar	Prior Investments Flexible	Current Investments Wind	Current Investments Solar	Current Investments NG-Flexible	Expected Investments Wind	Expected Investments Solar	Expected Investments NG-Flexible	Current Investments Correction Wind	Current Investments Correction Solar	Current Investments Correction NG-Flexible
Wind prior investment × Wind capacity utilization	Yes	–	–	–	–	–	–	–	–	–	Yes	–	–
Solar prior investment × Solar capacity utilization	Yes	–	–	–	–	–	–	–	–	–	–	Yes	–
Flexible prior investment × Flexible capacity utilization	Yes	–	–	–	–	–	–	–	–	–	–	–	Yes
Inflexible prior investment × Inflexible capacity utilization	–	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	–	–	–
Wind prior investment × Energy intensity	Yes	–	–	–	–	–	–	–	–	–	Yes	–	–
Solar prior investment × Energy intensity	Yes	–	–	–	–	–	–	–	–	–	–	Yes	–
Flexible prior investment × Energy intensity	–	–	–	–	–	–	–	–	–	–	–	Yes	Yes
Wind prior investment × Residential price	Yes	–	–	–	–	–	–	–	–	–	Yes	–	–
Solar prior investment × Residential price	Yes	–	–	–	–	–	–	–	–	–	–	Yes	–
Flexible prior investment × Residential price	–	–	–	–	–	–	–	–	–	–	–	Yes	Yes
Wind prior investment × Utility ownership	–	–	–	–	–	–	–	–	–	–	Yes	–	–
Solar prior investment × Utility ownership	–	–	–	–	–	–	–	–	–	–	–	Yes	–
Flexible prior investment × Utility ownership	–	–	–	–	–	–	–	–	–	–	–	–	Yes
Flexible prior investment × NG price	–	–	–	–	–	–	–	–	–	–	–	Yes	–
NG price × NG pipeline capacity	–	–	–	Yes	–	–	Yes	–	–	Yes	–	–	Yes
NG pipeline capacity × Utility ownership	–	–	–	–	–	–	–	–	–	–	–	–	Yes
NG pipeline capacity × Utility ownership	–	–	–	–	–	–	–	–	–	–	–	–	Yes
Wind speed × Utility ownership	–	–	–	–	–	–	–	–	–	–	Yes	–	–
Solar radiation × Utility ownership	–	–	–	–	–	–	–	–	–	–	–	Yes	–

Note: This table provides a list of interaction terms we used in each of the estimation models. The first column shows the name of the models following our estimation strategy steps. Starting from the RPS endogeneity correction first stage. Then the second stage estimations of prior and current investments. We then show the interaction terms of the current investment correction models followed by the expected investment estimation models. In last 5 columns we show the technology-specific dependent variable of interest for each estimation model.

Table 1.11: Group fixed-effect (GFE) estimation results of prior investment effects

	Coefficient (Log-Log)	R-square (%)
H1: Wind ~ Flexible	0.145^*	54.83
H2: Solar ~ Flexible	0.061^*	52.20
H3: NG-Flexible ~ Wind	0.091^*	25.19
H4: NG-Flexible ~ Solar	-0.075^*	24.89

Note: This table summarizes the prior investment in generation capacity coefficients and r-squares. We used linear regression models with leading investment dependent variables and k-means cluster fixed effects. We log-transformed all variables. We observe empirical support for hypotheses 1–3 while we do not observe such support for hypothesis 4. Significance level: $^*p < 0.01$.

Table 1.12: Detailed overview of future and prior investment correlations

	Raw Sample Correlations (%)	Clustered Sample Correlations (%)						
		Min	*Q1*	*Q2*	*Q3*	*Max*	*Mean*	*Std. Dev.*
H1: Wind ~ Flexible	7.72	-45.36	7.52	15.08	19.69	33.79	10.92	17.44
H2: Solar ~ Flexible	8.01	-34.18	-4.33	7.04	16.03	59.98	8.98	24.20
H3: NG-Flexible ~ Wind	4.55	-56.31	-13.51	-0.68	10.36	32.52	-3.62	24.67
H4: NG-Flexible ~ Solar	22.87	-49.51	-5.05	3.13	16.58	65.29	5.53	24.88

Note: This table represents the correlations (%) between future investments and prior investments as DVs and IVs respectively for each of our 4 research hypotheses. In the first column we show these correlations over the whole sample without clustering. We observe an overall positive correlation between the dependent and independent variables in our whole sample. In other columns, we illustrate the distribution of correlations across 16 k-means clusters.

Table 1.13: Between-cluster estimation results of prior investment effects

	H1: Wind ~ Flexible		***H2: Solar ~ Flexible***		***H3: NG-Flexible ~ Wind***		***H4: NG-Flexible ~ Solar***	
	Coefficient	R-square	Coefficient	R-square	Coefficient	R-square	Coefficient	R-square
Cluster 1	0.867^*	44.5	-0.353^*	75.5	-0.114^*	71.9	0.031	69.9
Cluster 2	0.250^*	69.2	0.038^*	44.3	0.131^*	18.0	0.713^*	17.3
Cluster 3	0.130^*	62.4	0.350^*	34.6	0.173^*	31.8	0.371^*	31.2
Cluster 4	0.049^*	23.7	0.099^*	29.0	-0.126^*	16.6	0.219^*	16.7
Cluster 5	-0.034^*	71.6	0.243^*	56.5	0.149^*	29.5	-0.087^*	29.1
Cluster 6	0.122^*	40.9	0.045^*	41.3	-0.204^*	36.1	0.056^*	28.8
Cluster 7	0.060^*	49.6	0.066^*	42.7	0.008	14.7	0.104	14.7
Cluster 8	0.037^*	31.8	0.685^*	62.3	0.024	56.2	-0.142^*	53.9
Cluster 9	0.287^*	59.2	0.259^*	40.5	0.297^*	73.7	0.197^*	73.1
Cluster 10	0.132^*	10.2	0.027^*	43.7	0.011	7.5	-0.107^*	7.8
Cluster 11	0.017	29.8	0.057^*	61.4	-0.127^*	52.9	-0.057^*	52.7
Cluster 12	0.371^*	48.2	-0.108^*	58.5	-0.133^*	28.8	0.042	28.1
Cluster 13	0.618^*	52.2	0.577^*	67.5	-0.374^*	28.8	-0.111^*	26.4
Cluster 14	0.321^*	46.3	0.251^*	47.3	0.192^*	28.4	-0.073^*	26.6
Cluster 15	0.033	54.4	0.446^*	72.2	0.061^*	65.2	-0.199^*	62.1
Cluster 16	0.493^*	34.2	0.006	47.2	-0.007	16.0	0.349^*	16.4

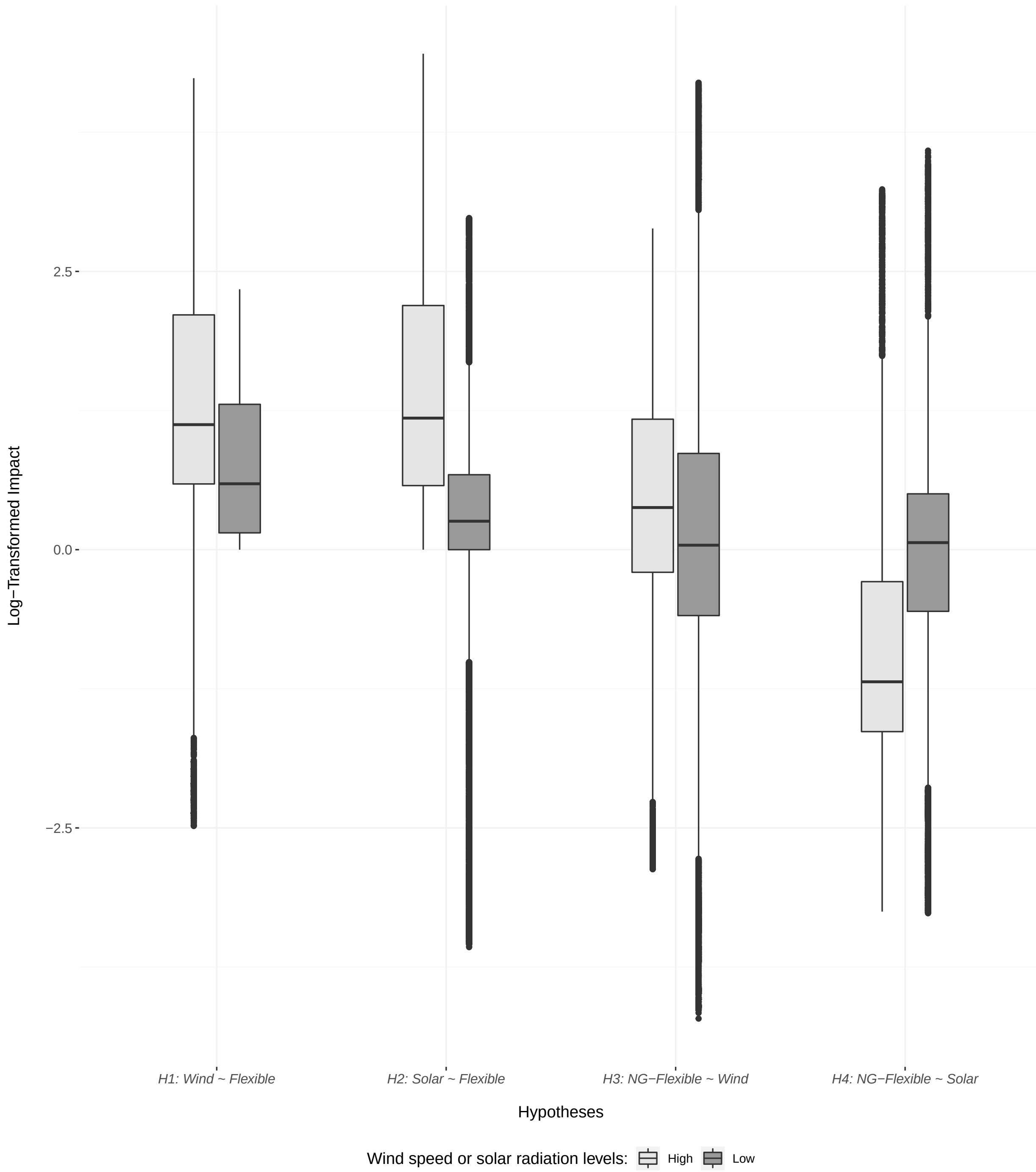
Note: This table represents the individual estimates of prior investment coefficients and their corresponding r-squares as independent variable for future investments as dependent variables. Clusters are sorted in an ascending order based on the RPS obligation residuals (i.e., cluster 1 contains the least stringent regions while cluster 16 involves the most stringent regions). Significance level: $^*p < 0.01$.

Table 1.14: Summary of prior investment impact quantification over clustering feature levels

	Hypothesis 1: Wind ∼ Flexible		***Hypothesis 2: Solar ∼ Flexible***		***Hypothesis 3: NG-Flexible ∼ Wind***		***Hypothesis 4: NG-Flexible ∼ Solar***	
	Low	High	Low	High	Low	High	Low	High
RPS residual	3.46	3.25	0.31	5.26	1.94	−3.35	0.87	−1.62
NG pipeline capacity	4.34	2.37	6.64	−1.08	−2.40	1.00	−1.46	0.70
Utility ownership	0.65	6.07	0.19	5.37	1.30	−2.71	−0.09	−0.66

Note: In this table we show the average impact of the independent variables (i.e., prior investments) across high and low levels of clustering features. We identify clusters with high and low levels of features by finding the median of average values of clustering features and setting all clusters with average value below the median as those with low levels of the clustering feature and conversely those with values beyond the median as the ones with high levels of clustering feature. For example, from the table we observe that clusters with high average levels of RPS residual (i.e., more stringent RPS program) show a larger positive impact for future investments in solar generation technology. This represents that in the presence of more stringent RPS programs future solar investments benefit more from prior investments in flexible generation technologies.

Figure 1.6: Renewable potential effect on prior investment impacts



Note: This plot demonstrate the distribution of estimated impacts for hypothesis 1–4 across high and low renewable potential levels. We log-transformed the impacts (i.e., percentages of maximum absolute scaled values) preserving their signs. We do this to better visualize the impact trends visible in the center of each distribution avoiding skewness caused by extreme values. We characterize renewable potential with average wind speed and average solar radiation for wind and solar generation technologies respectively. We identify the high and low levels by dividing our 16 clusters based on the median of the renewable potential and assigning those below the median as low renewable potential and those above the median as clusters with high renewable potential.

Chapter 2

The Foundation of Transition: Analyzing Infrastructure's Impact on Renewable Deployment and Flexible Conventional Resources

2.1 Introduction

Within the past two decades numerous factors have been reshaping the electricity industry globally including governmental policy instruments, advancing generation technologies, emerging technologies, and electricity market reforms. Governments across the world have been pushing toward higher adoption of renewable generation technologies using different political instruments in order to tackle the adverse environmental implications of the electricity industry. In 2019 electricity generation was the largest contributor to greenhouse gas (GHG) emissions with accounting for 34% of total GHG emissions (Dhakal et al., 2022). Particularly according to an Environmental Protection Agency (EPA) report, in the U.S. during 2022 following the transportation sector that was responsible for 28% of total GHG emission, electricity generation sector was responsible for 25% of total emissions (EPA, 2024). To this end, there has been an unprecedented penetration of renewable sources in the electricity generation capacity portfolio. In 2019, solar and wind generation capacity accounted for more than 10% of the U.S. total generation capacity portfolio.

Although renewable generators such as wind and solar have strong influence on reducing GHG emissions of the electricity grids, they come with an intricate challenge for the grid managers and investors that is due to their intermittent generation pattern which is a function of the climate. This implies that as we move toward a greener electricity grid, there will be an increasing amount of added uncertainty in supply in addition to the traditional demand uncertainties that must be handled properly by the grid managers to ensure the supply-demand balance. For this reason, we observe that as a solution the past investment trends in wind and solar generation capacity has been accompanied by another investment trend in flexible electricity generators and more specifically natural gas fueled flexible (NG-Flexible) generators as we established in the first chapter (Figure 1.1). Additionally, on the demand side with the advent of Artificial Intelligence (AI) systems and their massive adoption new demand patterns are emerging. According to the International Energy Agency's 2024 report, electricity demand of AI and cryptocurrency sectors can double by 2026 where they can go from 460 terawatt-hours (TWh) in 2022 to exceeding 1,000 TWh in 2026 that is almost equal to Japan's total electricity demand (IEA, 2024). On the other hand, in the U.S. due to the massive electricity demand of data centers that are powering AI machines, electricity demand forecasts has become extremely unreliable and challenging. According to a *Washington Post* article, Georgia's most recent demand forecast for the next decade has proved to be 17 times lower than what it was just recently predicted with other states such as Arizona and Virginia struggling with a similar challenge (Post, 2024).

These figures suggest a global hard-to-predict increase in electricity demand that is coupled with the necessity of sustainable development that is translated into investments in wind and solar generation technologies. However, there are strong arguments that the demand increase might outpace the infrastructure developments and generation capacity additions (Dive, 2023). With this motivation, we pose the following research question, *"How does strategic investment choices influence the economic interplay between flexible, wind, and solar investment?"* We adopt a spatial empirical approach to answer this research question by evaluating distinct strategic investment scenarios that characterize two aspects of market structure including, competition and infrastructure development. We distinguish strategic investments in a location where there is no existing generation capacity of any technology (i.e., empty) from those where an established market exists with different electricity generation technologies. We categorize the latter in to locations that had similar existing generation technology with future investments received in that location (i.e., own) and the ones without similar existing generation technology (i.e., cross). This categorization of future investment scenarios allows us to identify how the existence of infrastructure and different competition levels in a given location moderates the economic interplay between future and prior investments in wind, solar, and flexible generation technologies. Our hypotheses are structured around the influence of prior investments in flexible generation technologies on subsequent investments in intermittent renewable technologies (wind and solar), and vice versa, across three types of strategic investment scenarios: empty, cross, and own. For wind and solar future investments we hypothesize that prior investments in flexible technologies enhance future investments in both wind and solar in scenarios where there are no pre-existing generation capacities (empty) or where existing capacities are of a similar type (own). Conversely, in cross scenarios—where existing capacities differ from the new investments—we expect a negative impact, suggesting that prior flexible investments could deter new renewable investments. For flexible future investments we postulate that previous investments in renewable technologies (wind and solar) positively affect subsequent investments in flexible technologies in empty and cross scenarios, reflecting the need for grid stability due to the intermittency of renewables. However, in own scenarios, where flexible capacities already exist, we anticipate a negative impact, indicating a potential saturation or reduced necessity for additional flexible capacity. These hypotheses are driven by the operational need to balance the grid in the face of renewable intermittency and the economic impacts of existing infrastructure on new investments. The underlying mechanisms involve the mitigation of curtailment risks and the enhancement of dispatchability, which are crucial for integrating high shares of intermittent renewables into the energy system.

Particularly, we answer our research question by collecting and combining a unique high-resolution panel data set for the U.S. electricity industry from 2002–2019 that encompasses numerous characteristics of investments in different generation technologies as well as spatial attributes of the locations where the investments take place. We leverage this panel data set with a multi-staged empirical approach to examine our hypotheses combining principal component analysis (PCA), k-means clustering, and Tobit models to handle the censored nature of the investment data. Our empirical approach allows us to capture the nuances of strategic investment decisions influenced by the existing generation capacity and their impacts on future energy infrastructure development.

Our research elucidates the complex interactions between renewable and flexible generation technologies across varied infrastructural landscapes in the U.S. energy sector. We discovered that prior investments in flexible generation capacities generally encourage subsequent investments in renewable technologies like wind and solar, particularly in areas without pre-existing capacities (empty) or where existing capacities align with the new investments (own). However, in settings where pre-existing capacities involve different technologies (cross), these investments tend to discourage new renewable initiatives. This study significantly advances the existing literature by mapping the economic interdependencies of energy investments relative to infrastructure, offering a nuanced understanding of how strategic investment decisions can impact the economic and operational dynamics of the energy sector. The findings hold crucial managerial implications, suggesting that energy policymakers and investors should consider existing infrastructural contexts when planning new investments. This perspective supports the development of more tailored strategies to optimize the integration of renewable resources, ultimately fostering a more stable and efficient energy grid while also enhancing the economic feasibility of transitioning towards renewable energy. This approach not only informs investment strategies but also aids in designing policies that accommodate the complexities of modern energy markets.

The rest of this paper is structured as the following. Section 2.2 sets the stage by providing a comprehensive literature review which contextualizes our research within the fields of energy economics, operations management, and strategic management. In Section 2.3, we develop our hypotheses based on the theoretical underpinnings discussed earlier. Section 2.4 details our data collection methodology and variable development, essential for understanding the empirical analysis that follows. The empirical approach and model specifications are elaborated in Section 2.5, where we describe the analytical methods used to test our hypotheses. Our findings are then presented and discussed in Section 2.6, highlighting the key results from the analysis. The paper concludes in Section 2.7 with a summary of the findings, their implications for both theory and practice, and suggestions for future research.

2.2 Literature Review

Our research relates and contributes to three research streams of, *"energy economics"*, *"operations management"*, and *"strategic management"*. In the following, we first provide an overview of the relevant research streams and then we build upon them to explain our research hypotheses.

Energy Economic: Infrastructure investment role in reducing curtailment of renewable energy is a critical theme within energy economics. Curtailment, or the under-utilization of available renewable energy due to grid limitations, poses significant economic and operational challenges. Khorramfar et al. (2024) and Henni et al. (2021) highlight how infrastructure developments can mitigate these issues by enhancing the grid's ability to absorb and redistribute intermittent renewable energies like wind and solar. Investments in grid infrastructure, such as upgraded transmission lines and enhanced grid connectivity, play a pivotal role in minimizing renewable energy curtailment. According to Jayadev et al. (2020), enhancing grid infrastructure not only supports the direct integration of higher shares of renewables but also stabilizes the energy supply by reducing the variability impacts on the grid. These enhancements are crucial for regions experiencing rapid growth in renewable deployments, where existing infrastructure may not suffice. Our study extends these discussions by examining the interplay between prior and future investments in wind, solar, and flexible generation technologies under different infrastructure development levels. With this study we provide an enhanced spatial understanding of the role of infrastructure development in moderating the economic interplay between different generation technologies investments on a relatively high resolution level.

Operations Management: Conventional-Renewable generation capacity economic interaction is critical given the increasing incorporation of intermittent renewable energy into electricity grids. We provided a comprehensive review of the existing studies in the field of operations management that examine the relationship between these two generation technologies as well as the role of operational flexibility in Sections 1.2.1 and 1.2.2. In summary, earlier research predominantly suggests a competitive, substitutive relationship between these two generation technologies, recent insights propose a nuanced view. These studies highlight that the relationship between conventional and renewable generation can vary significantly based on economic factors such as natural gas prices. Under certain conditions, these two generation types transition from being substitutes to complements, which is essential for evolving energy policies and investment strategies aimed at enhancing grid stability and accommodating renewable expansion. Additionally, operational flexibility is increasingly recognized as a vital aspect of modern electricity systems, especially with the growth of intermittent renewable generation sources like wind and solar. Flexibility in electricity grids helps mitigate the challenges posed by the variability of supply from renewable sources. Various strategies used to enhance system flexibility, include demand response programs, energy storage, and flexible generation capacities.

This literature emphasizes the crucial role of flexible conventional generation technologies in ensuring a reliable energy transition towards more renewable sources. Moreover, operational flexibility is linked with the strategic behaviors in the electricity market, affecting investment decisions and competition dynamics.

Strategic Management: Within the literature of strategic management our research pertains to studies evaluating entry decision of firms and the influence of market structure and policy instruments on such strategic decisions. For example, Kapoor and Furr (2015) explore the strategic decisions behind technology choices by new entrants in the solar photovoltaic industry. It finds that firms entering a new market assess both the performance potential of different technologies and the availability of necessary complementary assets. Established companies entering the market (diversifying entrants) often prioritize the availability of these assets over superior technological performance, leveraging their existing capabilities to reduce market entry risks and costs. In contrast, startups tend to focus on technological superiority, sacrificing immediate asset support to distinguish themselves from established competitors. The study highlights how the interplay between company type and market entry strategy influences technological evolution and industry dynamics. Gohdes et al. (2022) emphasize the critical role of Power Purchase Agreements (PPAs) and counter-party credit quality in reducing entry costs for renewable energy projects in Australia's national electricity market. Llobet and Padilla (2018) investigate the interaction between conventional power generation and the entry of renewable energy sources within liberalized electricity markets. It focuses on the consequences of integrating renewable energy, such as solar and wind, which, although beneficial for environmental goals, tends to lower wholesale electricity prices and reduce the profitability of conventional thermal plants. These changes compel market regulators to adjust capacity remuneration mechanisms to maintain financial incentives for maintaining and investing in conventional power facilities, ensuring reliable energy supply despite the volatility of renewable production. Thy discuss capacity auctions as a tool to promote investment in conventional power, balancing the grid against the unpredictable nature of renewable energy output. In another study Shittu and Weigelt (2022) examine how established companies adapt to the adoption of wind energy, differentiating between direct ownership and contracting strategies. The research identifies several factors influencing these entry decisions: the physical distance of innovations, existing policy frameworks, and the characteristics of the market space. Firms tend to own wind power resources if they are physically distant from their core operations or if supportive policies are stable and longstanding. Conversely, firms with a significant stake in traditional technologies are less likely to own wind farms, preferring contractual arrangements. This behavior highlights how entrenched interests and policy stability shape the strategic entry paths of incumbents into renewable energy markets. Despite the richness of the existing studies focused on evaluating different market entry decisions, to the best of our knowledge, this study is the first to evaluate the interplay of investments in two distinct intermittent and flexible generation technologies. We add to this literature by showing how different investment strategies with respect to the market structure (i.e., existing generation technologies in a given location) can affect the relationship between prior and future investments in wind, solar, and flexible generation technologies.

2.3 Hypotheses Development

Building upon our arguments in Section 1.3, we keep our original hypotheses structure with slight modifications that address our research question.

2.3.1 Investment Strategies Moderating Flexibility Impact on Intermittency

In Section 1.3, we established electricity curtailment to be the main driving mechanism behind the positive impact of prior investments in flexible generation technologies on the future investments in intermittent generation technologies (i.e., wind and solar). The role of perceived potential curtailment in impeding future renewable investments is well established in the literature. Egli (2020) argues as renewable energy penetration deepens, the likelihood of curtailment—where energy production must be halted due to grid limitations—increases. As curtailment risks become more prominent, they can deter investment unless mitigated through adopting grid flexibility enhancement measures. Newbery (2023) analysis complements this by discussing the economic implications of renewable energy curtailment and arguing that the economic costs associated with curtailment can make renewable energy sources less competitive unless integrated with robust grid management strategies or pricing adjustments that compensate for the potential lost revenue during periods of curtailment. We characterize three investment strategies based on their spatial characteristics including: empty, cross, and own. These respectively show investments in locations where there is no existing generation technology, locations with existing generation capacity but dissimilar to the new investment generation technology, and locations with existing generation capacity that includes a similar generation technology to the new investments. Based on this theoretical underpinning and our definition of investment strategies, we hypothesize that future investments in locations where it is the first time that they are receiving wind and solar investments (i.e., empty), experience a larger positive impact (or smaller negative impact) from prior flexible investments compared to locations receiving wind and solar investments with existing generation capacity that lacks wind or solar existing generation capacity (i.e., cross). This is due to two distinct opposing forces. Empty locations benefit stems from being less crowded therefore they experience less competition. Conversely, as the size of the existing generation capacity portfolio increases we expect a given cross location to become closer to it's carrying capacity and therefore the curtailment risk offsetting positive force is trumped in favor of the increased competition. This hypothesis is well supported by four string of theories including: the multimarket contact theory (Gimeno and Woo, 1999), agglomeration spillover theories (Myles Shaver and Flyer, 2000), density dependence theory (Hannan and Freeman, 1977), and institutional theory of organizational behavior (Haveman, 1993). In summary, these theories propose that the existence of rivals make a given location attractive for future investments only up to a certain point from which that location becomes increasingly less attractive for future investments.

The theoretical foundation explained above, helps with developing hypotheses for empty and cross markets. However, in own markets we hypothesize that another force opposes the negative impact of increased existing rivals and that is the positive influence of existing generation capacity of wind and solar. This signals that sufficient infrastructure to integrate future investments in wind and solar are already developed and are successfully accommodating the existing renewable generation capacity. Within this context, we expect that the positive curtailment risk offsetting effect of prior investments in flexible sources is more pronounced at own locations. In the following, we re-formulate our previous hypotheses 1 and 2 separately hypothesizing about wind and solar future investments as their different intermittency patterns call for their effects to be heterogeneous:

Hypothesis 1 (Empty & Own). *Prior flexible generation capacity investment around a focal zip code increases future wind investment rates at empty and own zip codes around that focal zip code.*

Hypothesis 1 (Cross). *Prior flexible generation capacity investment around a focal zip code decreases future wind investment rates at cross zip codes around that focal zip code.*

Hypothesis 2 (Empty & Own). *Prior flexible generation capacity investment around a focal zip code increases future solar investment rates at empty and own zip codes around that focal zip code.*

Hypothesis 2 (Cross). *Prior flexible generation capacity investment around a focal zip code decreases future solar investment rates at cross zip codes around that focal zip code.*

2.3.2 Investment Strategies Moderating Intermittency Impact on Flexibility

As we established in Section 1.3, price dispatchability premium is the main driving mechanism behind the hypothesized positive influence of prior investments in wind and solar generation technologies on future investments in NG-flexible generation technology. We argue this positive influence diminishes as we move from future investments at empty locations to cross, and own locations. Our reasoning builds upon the four theory strings we mentioned in Section 2.3.1. We hypothesize that due to the existing generation capacity inc cross locations that compete for the same dispatchability price premium (but with lower flexibility levels), future investments in NG-flexible generation technology will be left with less of the total available profit due to the intermittency of prior investments in wind and solar generation technologies. Therefore, we hypothesize that the interplay between future investments in NG-flexible generation technology and prior investments in wind and solar generation technologies will still be positive but lower than empty markets. Lastly, we argue that own markets will be the most hostile of all locations with existing generation technology of NG-flexible that will leave future investments in NG-flexible with even lower total price premium left. This leads to our hypothesis that the relationship between prior intermittent investments and future NG-flexible investments to be negative indicating that the complementary relationship of these technologies can become substitutive under certain conditions. In the following we provide modified versions of hypotheses 3 and 4 that reflect our new set of hypotheses:

Hypothesis 3 (empty & cross). *Prior wind generation capacity investment around a focal zip code increases future natural gas-fueled flexible investment rates at empty and cross zip codes around that focal zip code.*

Hypothesis 3 (own). *Prior wind generation capacity investment around a focal zip code* *decreases future natural gas-fueled flexible investment rates at own zip codes around that focal zip code.*

Hypothesis 4 (empty & cross). *Prior solar generation capacity investment around a focal zip code increases future natural gas-fueled flexible investment rates at empty and cross zip codes around that focal zip code.*

Hypothesis 4 (own). *Prior solar generation capacity investment around a focal zip code decrease future natural gas-fueled flexible investment rates at own zip codes around that focal zip code.*

2.4 Data and Variables

2.4.1 Variable Construction

To answer our research question through testing hypotheses we developed in Section 2.3, we collect and combined a unique panel data set of the U.S. electricity sector as we detailed in chapter 1 in Section 1.4. This data set provide a comprehensive image of the U.S. electricity industry for years 2002–2019 from both economic and political perspectives involving different data resolutions going from highly granular observations (i.e., at generator level) to more aggregated observations (i.e., on the national level). We provided a list of all variables compiled in Table 1.1 and we elaborated the steps we took to clean and further prepare a sample for our estimation steps in Section 1.4. Aligned with our empirical setting, we decompose investments in three generation technologies of wind, solar, and NG-flexible into three groups of empty, cross, and own as we described in Section 2.3 based on the existing generation capacity portfolio at each location prior to investments being realized. Similar to our empirical setting in chapter 1, we are setting our main observation level to be region that we define to be an aggregation level of a 100 miles radius around a focal zip code entailing all zip code level observations with centers falling within this 100 miles radius area. With this, all observations with granularity level below zip codes are aggregated to be on the zip code level first and then are averaged to be on the region level. All variables with granularity level of higher than zip code are used as they were. We further aggregate on the time dimension with creating 3-year time windows to account for the lead time between the establishment of an investment decision and it's realization. In the following we explain the construction process of a the new set of investment variables in addition to the ones we summarized in Section 1.4.2.3.

Empty: Investment variables are initially aggregated into zip code level observations based on the difference between any two consecutive years generation capacity portfolio (detailed formula presented in Table 1.4). Equations 2.1 and 2.2 represent a statistical definition of our investment variables that are the amounts invested in any generation technology conditional on having no generation capacity of any generation technologies already installed. $\textbf{Empty Investment}_{z,e,t}^{Intermittent}$ and $\text{Empty Investment}_{z,t}^{NG-Flexible}$ are the investment amounts at zip code z in generation technologies $e \in \{wind, solar\}$ and NG-flexible respectively at time t (here t denotes yearly observations). $\text{Prior Investment}_{z,t}$ shows the amount of existing total generation capacity of any generation technology at zip code z at time t.

$$\textbf{Empty Investment}_{z,e,t}^{Intermittent} = [\textbf{Investment}_{z,e,t}^{Intermittent} \mid (\text{Prior Investment}_{z,t} = 0)] \tag{2.1}$$

$$\text{Empty Investment}_{z,t}^{NG-Flexible} = [\text{Investment}_{z,t}^{NG-Flexible} \mid (\text{Prior Investment}_{z,t} = 0)] \tag{2.2}$$

Cross: We distinguish this set of investment variables from investments we defined as empty with switching the condition in to having locations with existing generation capacity. However, the prior investments can be in any generation technology with the exception of the one for which we observe investments. Equations 2.3 and 2.4 showcase the statistical expression of cross investment conditions. For example, when constructing wind cross investments, we only capture investment amounts that meet two conditions of having non-zero prior investments and no prior investments in wind generation technology. Similarly, calculate this for solar and NG-flexible generation technologies.

$$\begin{aligned}\textbf{Cross Investment}_{z,e,t}^{Intermittent} =& [\textbf{Investment}_{z,e,t}^{Intermittent} \mid (\text{Prior Investment}_{z,t} \neq 0, \\ & \textbf{Prior Investment}_{z,e,t}^{Intermittent} = 0)]\end{aligned} \tag{2.3}$$

$$\begin{aligned}\text{Cross Investment}_{z,t}^{NG-Flexible} =& [\text{Investment}_{z,t}^{NG-Flexible} \mid (\text{Prior Investment}_{z,t} \neq 0, \\ & \text{Prior Investment}_{z,t}^{NG-Flexible} = 0)]\end{aligned} \tag{2.4}$$

Own: Lastly, we characterize own investments as locations with prior investments, removing the restriction on similarity of generation technology of prior investments at that location. Equations 2.5 and 2.6 represent this characterization. To illustrate, when constructing wind own investments, we isolate investments that took place at locations with prior investments in any generation technology but making sure that there exist some prior investments in wind generation technology.

$$\begin{aligned}\textbf{Own Investment}^{Intermittent}_{z,e,t} =& [\textbf{Investment}^{Intermittent}_{z,e,t} \mid (\text{Prior Investment}_{z,t} \neq 0, \\ & \textbf{Prior Investment}^{Intermittent}_{z,e,t} \neq 0)]\end{aligned} \tag{2.5}$$

$$\begin{aligned}\text{Own Investment}^{NG-Flexible}_{z,t} =& [\text{Investment}^{NG-Flexible}_{z,t} \mid (\text{Prior Investment}_{z,t} \neq 0, \\ & \text{Prior Investment}^{NG-Flexible}_{z,t} \neq 0)]\end{aligned} \tag{2.6}$$

2.4.2 Descriptive Statistics

In Figure 2.1 we depict strategic investment heterogeneity across states combining all generation technologies of wind, solar, and NG-flexible for years 2002–2019. With this figure we underscore the spatial heterogeneity of strategic investments in the U.S. that justifies the importance of geographical location in strategic investments in the electricity sector. Each state represents a unique strategic investment profile with extreme points such as the state of Georgia with roughly 90% of new investments in wind, solar, and NG-flexible taking place at own locations while in the state of New Hampshire approximately 90% of investments where in empty locations.

Figure 2.2 illustrates a snapshot of investment trends based on generation technology and our investment decomposition approach. In this figure we observe how investments in each year are distributed between three main categories of empty, own, and cross. Considering the case of 2019, 64% of all investments in wind generation technology took place at empty locations, with only 10%, and 26% for cross and own locations respectively. We observe that this has been a rather consistent distributional trend for investments in wind sources while solar trends have been more volatile. In 2005 almost all of solar investments were at own locations while in 2007 the majority of investments were at empty locations and in 2019 we observe a roughly 50%-50% distribution of solar investments at own and empty locations. Lastly, investments in NG-flexible generation technology demonstrates an opposite relatively consistent pattern to wind with the majority of investments taking place at own locations. This plot presents the strategic investment heterogeneity across different generation technologies and how for a generation technology like solar, it has been evolving through time.

2.5 Empirical Approach

We develop a multi-staged empirical approach to test our research hypotheses, avoiding similar empirical concerns we elaborated in chapter 1. Our empirical setting differs with the one we detailed in Section 1.5.3 with respect to our dependent variables of interest. We decompose the previously developed technology-specific dependent variables based on spatial existence of generation technologies to evaluate how different investment strategies' influences are reflected on the interplay between investments in intermittent and flexible generation technologies. Similar to chapter 1 where we also controlled for the influence of political instruments, in this study we also take this into account by controlling for the effect of the most prominent and highly debated renewable portfolio standard (RPS) (Extensive details on how this policy influences investments are provided in Section 1.5.3). We follow the steps below to evaluate our hypotheses:

1. **RPS Endogeneity Correction:** To control for the influence of the most important state-level political instrument, we run a first stage OLS regression model on the renewable portfolio standard (RPS) obligations to calculate RPS residuals.

2. **PCA and K-means Clustering:** Using a set of 11 variables (Table 1.7), we create 16 clusters for our region-year observations.

3. **Grouped Fixed Effect Estimation:** Running grouped fixed effect (GFE) OLS regression models using control variables and interaction terms we summarized in Tables 1.9 and 1.10 with k-means cluster fixed effects.

4. **Between-Cluster Estimation:** We run the same regression models we estimated in Step 3 for each individual k-means cluster.

5. **Counterfactual Analysis (i.e., impact quantification):** We conduct a set of *ceteris paribus* counterfactual analyses to better interpret our findings as well as providing further managerial implications.

Nevertheless, our empirical setting involve dependent variables that have a large number of zero values due to decomposition of investment variables. Using OLS regression models to evaluate our hypotheses can lead to biased estimation of coefficients. To circumvent this issue, we use Tobit model to test our hypotheses as the Tobit model is particularly appropriate for dealing with dependent variables that are censored (in our study at zero). Tobit model is well-suited for scenarios where investment values cannot fall below zero, reflecting conditions such as non-feasibility, lack of resources, or regulatory constraints that prevent any investment. In contrast, OLS would treat these zero values as typical observations, potentially leading to biased estimates since it fails to account for the special treatment required by censored data. Tobit model involves maximizing a likelihood function that considers both the presence of non-zero values and the accumulation of observations at the censoring point. This allows for more accurate modeling of the relationship between independent variables and the observed censored investments. The coefficients derived from a Tobit model reflect the effect on the latent propensity for investment, which is particularly useful for understanding how prior investment variables influence the decision for future investments, given that the future investment amount is positive. To this end, in addition to the OLS estimation in step 3 we estimate the same model specification using the Tobit approach. We modify Equations 1.5 and 1.6 in the following where Equations 2.7 and 2.8 represent the hypotheses sets 1 and 2 evaluation while Equations 2.9 and 2.10 are for hypotheses sets 3 and 4 evaluation with Tobit approach:

$$\mathbf{y}^*_{l,r,e,t'} = \begin{cases} \ln[\mathbf{Investment}^{Intermittent}_{l,r,e,t'}] & \text{if } \mathbf{y}^*_{l,r,e,t'} > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2.7}$$

$$\begin{aligned} \mathbf{y}^*_{l,r,e,t'} = \alpha_0 + \alpha_1 \cdot \ln[\text{Prior Investment}^{NG-Flexible}_{r,t}] + \\ \boldsymbol{\alpha_2} \cdot \ln[\mathbf{Controls}_{r,e,t}] + \underbrace{\alpha_3 \cdot \ln[\hat{\tau}_{r,t}]}_{\text{RPS Residuals}} + \underbrace{\kappa_r}_{\text{Cluster FE}} + \epsilon_{l,r,e,t} \end{aligned} \tag{2.8}$$

$$z^*_{l,r,t'} = \begin{cases} \ln[\text{Investment}^{Flexible}_{l,r,e,t'}] & \text{if } z^*_{l,r,t'} > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2.9}$$

$$\begin{aligned} z^*_{l,r,t'} = \beta_0 + \boldsymbol{\beta_1} \cdot \ln[\mathbf{Prior\ Investment}^{Intermittent}_{r,e,t}] + \\ \boldsymbol{\beta_2} \cdot \ln[\mathbf{Controls}_{r,e,t}] + \underbrace{\kappa_r}_{\text{Cluster FE}} + \zeta_{l,r,t} \end{aligned} \tag{2.10}$$

In equations above, $\mathbf{y}^*_{l,r,e,t'}$ and $z^*_{l,r,t'}$ are investments in intermittent (with $e \in \{Wind, Solar\}$) and NG-flexible respectively representing investment strategy $l \in \{Empty, Cross, Own\}$, at region r

in a 3-year future time period t'. Equations 2.7 and 2.9 show that investment variables are censored at 0. In Equations 2.8 and 2.10 we show that all independent variables are observed at the 3-year time period t. This setting is equivalent to lagging all independent variables where we allow for the lead time it takes for project developments as we do not observe the exact time that a given decision is made while only observing when a decision is realized (i.e., an investment becoming operational).

2.6 Results

In this section we provide our findings first based on the estimation results of the grouped fixed effect for all four hypothesis across different investment strategies (i.e., empty, cross, own). Later we elaborate our ceteris paribus counterfactual analysis with which we quantify impacts based on the results of the between-cluster estimated models.

2.6.1 GFE Tobit Estimates

Prior to estimating α and β parameters, in Table 2.1 we provide an overview of investments in each generation technology across different investment strategies. Moreover, in Table 2.2 we provide an overview of 16 clusters that we have created as we detailed in Section 1.6.2. In this table we provide a list of clusters sorted based on RPS residual values from least restrictive to most restrictive. This table justifies the use of the Tobit approach as we provide the percentages of non-zero observations of investments in wind, solar, and NG-flexible across different investment strategies. We show that the percentages of non-zero observations are quite variable ranging between 0.1% to 98.9% across clusters.

We estimated the grouped fixed effect models listed in Table 2.3 using both OLS and Tobit approaches and we provide a summary of the estimated coefficients in Tables 2.4 and Tables 2.5 respectively. From the results we observe as expected that the estimated coefficients from the OLS models have a downward bias, however, the the direction of effects are rather consistent across the two models.

2.6.2 Hypotheses 1 & 2 Results

We hypothesized for future investments in wind and solar sources to have a positive relationship with prior investments in flexible generation technologies under empty and own investment strategies. However, our results provide empirical support for wind while solar future investment analyses do not support our hypotheses for empty and own. The estimated elasticity of future investments for wind with respect to prior investment in flexible sources are 0.06 and 0.063 across empty and own investments strategies respectively. Based on the log-log setting of our estimation approach along with Tobit modeling assumptions, these numbers are interpreted by if prior investments in flexible generation technologies doubles in a focal region, conditional on future investments being non-zero, we expect a 6% and 6.2% increases in future wind investments across empty and own strategies. We further hypothesized that prior investments in flexible generation technologies will have a negative influence on future wind and solar investments under cross investment strategy. From Table 2.5 we found empirical support for these hypotheses with estimated effects of 0.07 and 0.197 for wind and solar respectively. This illustrates that if prior investments in flexible generation capacity in a given region doubles, future investments in wind will decrease by 7% while solar investments will decrease by 19.7% conditional on non-zero investments for cross investment strategy.

2.6.3 Hypotheses 3 & 4 Results

In examining the influence of renewable energy prior investments on NG-flexible future investments, we proposed two main hypotheses across different investment strategies. Hypothesis 3 suggested that prior wind generation capacity investments in a focal region are expected to decrease future NG-flexible investments under own strategy within the same region while increasing such investments under empty and cross strategic approaches. We estimated these coefficients to be 0.282 for the former with 0.020 and 0.039 for the latter showing empirical support only for the empty and cross strategies. Similarly, in hypothesis 4 we posited that prior solar generation capacity investments would negatively impact future NG-flexible investments, under own investment strategy and positively impact future NG-flexible investments under empty and cross strategies. Our results show empirical support for own and cross strategies with estimated coefficients of -0.019 and 0.251 respectively.

2.6.4 Counterfactual Analysis

To provide a better interpretation of our results, we first conduct a between-cluster estimation of effects. To this end, we run the Tobit models on each individual k-means clusters using the same model specification for the fixed effect models. We provide a list of coefficients in Table 2.6. The estimated coefficients across k-means clusters under similar investment strategies reveal the possibility of a non-linear relationship as the sign of the coefficients change. Later, to better represent this and quantify impacts we conduct a set of ceteris paribus counterfactual analyses where we calculate two predictions of our dependent variable. One prediction using the actual values of all independent variables, and one holding all independent variables constant and only setting the prior investment values to zero (similar to Bray et al. (2019) and Akkas et al. (2019)). Next, we calculate the difference between these two predicted values to uncover the impact of the prior investments on future investments for different generation technologies under different investment strategies. The positive values of differences signal the positive influence of prior investments on future investments while a negative difference value signals the opposite.

We first explore how these effects are moderated by the renewable potential of regions that we measure by average wind speed and solar radiation. For this purpose, we identify k-means clusters with high and low levels of renewable potential by calculating the average wind speed and solar radiation of each cluster, then, we find the median value of cluster averages. Next, we assign clusters with renewable potential below the median to be the ones with low renewable potential with the ones with solar radiation or wind speed above the median assigned to have high renewable potential. With this we plotted the quantified impacts based on high and low groups for each hypotheses under each investment strategy in Figure 2.3. We show that irrespective of the investment strategy, impact of prior investments in flexible generation technologies on future investments in solar sources is more pronounced in locations with high solar radiation. This is consistent with our findings in chapter 1 and further underscores the importance of flexibility demand of renewable sources, specifically solar generation technologies.

Lastly, we aggregate the quantified impacts across investment strategies for each hypotheses. We do this by calculating the average impact of all clusters across empty, own, and cross investment strategies as demonstrated in Figure 2.4. With this plot we provide empirical support for all of our hypotheses. We observe similar patterns when evaluating the influence of prior investments in flexible generation technologies (i.e., the top panels of the plot) on future investments in intermittent generation technologies. It is evident that future wind and solar investments in locations where there is no existing generation capacity (i.e., empty) or existing generation capacity with wind and solar (i.e., own) are better complemented with prior investments in flexible generation technologies. However, the bottom panel of Figure 2.4 depicts a diminishing relationship between prior investments in intermittent generation technologies and future investments in NG-flexible sources.

2.7 Conclusion

In this study, we rigorously explored the economic interplay between strategic investments in renewable energy sources—specifically wind and solar—and flexible conventional generation capacities, within different levels of infrastructure development across the U.S. This inquiry was grounded in a spatial empirical analysis that integrated a detailed, high-resolution panel data set spanning from 2002 to 2019, encompassing various attributes of generation investments and their geographic specifics. We set out to dissect how strategic investment choices impact the economic interactions between flexible, wind, and solar investments. Through a multi-staged empirical framework, we distinguished investment strategies into three categories based on the existing infrastructure: empty, cross, and own. These categories served as a basis to examine the moderating effects of prior investments on subsequent ones within these distinct infrastructural contexts. Our findings reveal several critical insights into the dynamics of energy infrastructure investment. We discovered that prior investments in flexible generation capacities substantially enhance the integration of intermittent renewables by providing the necessary support to maintain grid stability and reduce curtailment risks. Specifically, investments in flexible technologies tend to boost subsequent investments in renewable technologies in scenarios where previous infrastructure either did not exist or was complementary. Furthermore, we found that the interactions between different types of investments are highly dependent on the pre-existing infrastructural and technological landscape. For instance, in areas with existing flexible capacities (own scenarios), the introduction of new renewable capacities could either complement or compete with these established investments, suggesting a complex economic landscape that varies significantly across different geographic and strategic contexts.

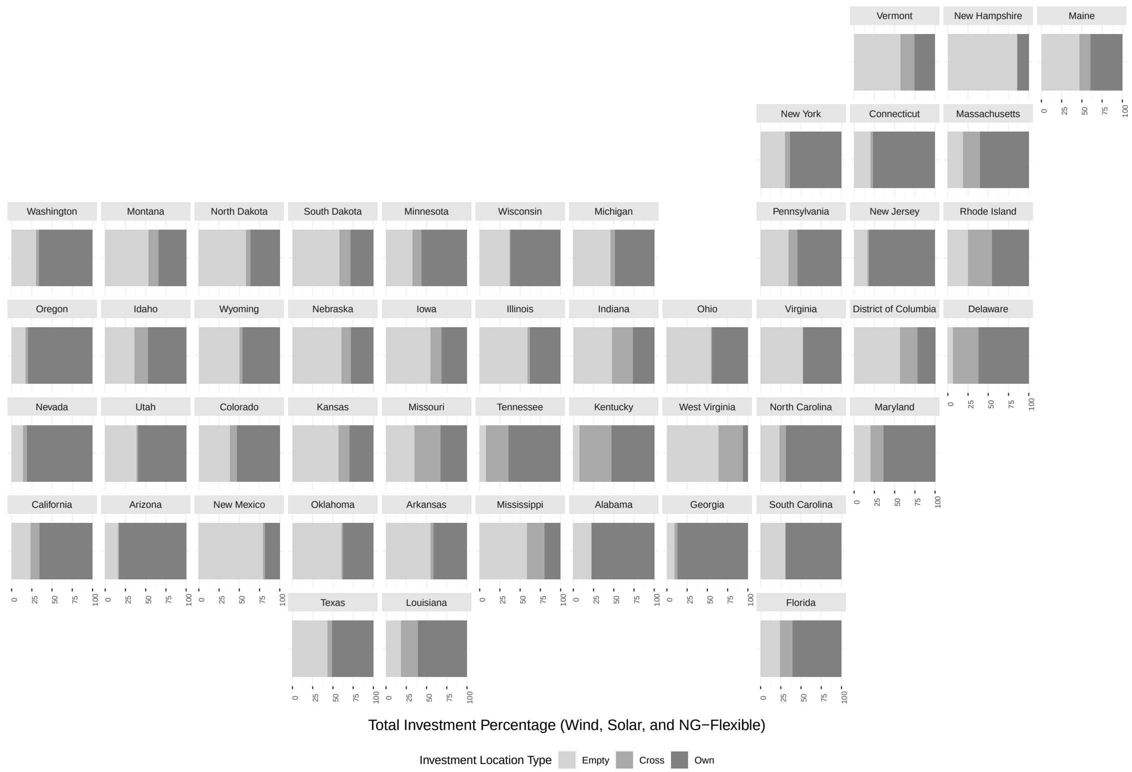
From a managerial perspective, our study offers substantial implications for both policymakers and investors. The nuanced understanding of investment interdependencies provides valuable insights for strategic planning and policy formulation. By acknowledging these complex interactions, policymakers can design more effective strategies to encourage sustainable energy investments, while investors can make more informed decisions that account for the long-term economic impacts of their investment choices on energy grid stability and growth.

Despite the robustness of our analysis, we recognize certain limitations that suggest avenues for future research. Our focus on the U.S. market may limit the applicability of our findings to other regions with different regulatory and market conditions. Additionally, the analysis is somewhat constrained by the temporal scope of the data, which does not extend beyond 2019. As the energy sector is rapidly evolving, incorporating more recent data could provide a more current perspective on these dynamics. Another limitation of our study is that we look at the market characteristic-driven investment decisions while not taking into account the firm characteristic driven entry decisions. Therefore, future studies could benefit from a closer examination of firm-specific strategies and decision-making processes, which could offer deeper insights into how companies individually respond to policy and market changes. Understanding these micro-level dynamics would add a significant layer to our comprehension of market behaviors and investment interactions in the energy sector.

In conclusion, our research significantly advances the academic and practical understanding of strategic investment interactions in the energy sector. By mapping out the economic relationships between various types of energy investments and their infrastructural contexts, we provide a foundation for more strategic investment and policy decisions aimed at fostering a resilient and sustainable energy future. This study not only enriches the existing body of knowledge but also underscores the critical need for continued exploration into the complex mechanisms that drive energy market developments.

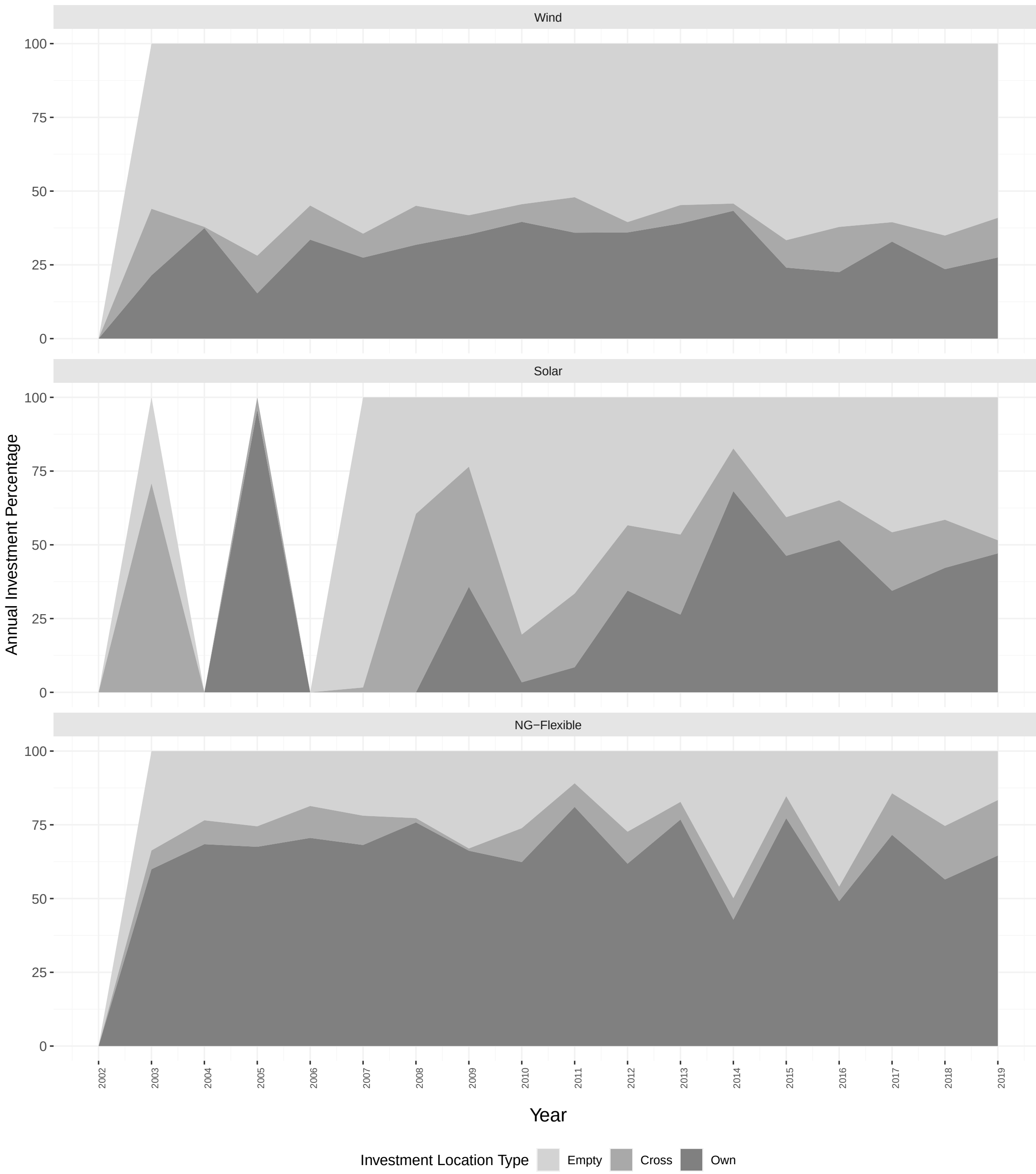
Exhibits

Figure 2.1: State-level total investment ratios in different region types aggregated within 2002–2019



Note: This plot shows a snapshot of the distribution of all wind, solar, and natural gas-fueled flexible investments combined for each state that took place between 2002–2019.

Figure 2.2: Technology-specific investment trends in different region types



Note: This plot illustrates trends in technology-specific investments at different regions (i.e., empty, own, and cross) in the U.S. For each electricity generation technology, we show how investments are distributed between these three region types.

Table 2.1: Overview of technology-specific investment types across k-means clusters

	Wind			Solar			NG-Flexible		
	Empty	Own	Cross	Empty	Own	Cross	Empty	Own	Cross
Cluster 1	2.30	2.21	2.60	4.79	0.91	2.55	11.16	23.33	1.53
Cluster 2	16.48	8.00	8.84	0.21	0.00	0.37	20.63	21.39	7.38
Cluster 3	11.97	3.56	3.41	1.31	0.17	2.55	7.80	17.21	1.68
Cluster 4	2.52	0.74	0.37	1.57	0.39	1.40	11.08	19.10	3.06
Cluster 5	0.36	0.12	0.40	2.36	0.31	0.83	18.44	25.98	8.45
Cluster 6	2.92	2.44	1.02	8.43	3.36	8.98	13.34	27.85	2.97
Cluster 7	0.19	0.10	0.23	8.35	1.81	4.93	7.90	30.75	4.34
Cluster 8	2.51	1.80	2.31	2.03	0.30	1.63	15.89	15.32	5.48
Cluster 9	1.54	0.45	0.93	5.79	1.63	4.96	17.48	26.13	9.39
Cluster 10	22.56	18.68	12.11	1.31	0.00	1.93	6.89	17.34	5.00
Cluster 11	0.60	6.08	0.03	13.46	20.46	16.54	11.61	29.97	4.97
Cluster 12	8.27	4.50	7.40	2.91	0.42	8.75	6.39	11.20	0.39
Cluster 13	0.69	4.74	0.25	10.26	8.01	17.37	9.10	21.87	6.56
Cluster 14	9.14	3.34	1.61	1.53	0.41	1.43	5.93	6.50	1.46
Cluster 15	10.23	9.64	4.30	12.07	4.32	8.79	14.18	25.64	2.25
Cluster 16	3.45	4.80	1.48	0.37	0.00	1.09	7.92	12.45	0.30

Note: This table provides an overview of investments in our three generation technologies of interest including, wind, solar, and natural gas-fueled flexible across k-means clusters. These investments are further classified into three groups of, empty, own, and cross based on the existing generation capacity portfolio of the locations that are the recipient of investments. Specifically, investment into empty locations characterize zip codes that do not have any existing electricity generation capacity. Own locations identify zip codes with existing generation technology that includes ones with similar generation technologies with the new investments. For example, when we show investments in wind generation technology at "own" locations, we are showing investments in locations that have existing generation capacity in any generation technology as well as wind generation technology. Similarly, cross locations are zip codes with existing generation capacity but it uses a generation technology that is different with the new investments. Values in the table are after the maximum absolute normalization and are measured in percentages of the maximum value of investments in a certain technology and location type. Clusters are sorted based on the RPS residuals and are from least to most stringent.

Table 2.2: Overview of non-zero technology-specific investment types across k-means clusters

	Total Number of Observations	Wind Investment Non-Zero Ratio (%)			Solar Investment Non-Zero Ratio (%)			NG-Flexible Investment Non-Zero Ratio (%)		
		Empty	Own	Cross	Empty	Own	Cross	Empty	Own	Cross
Cluster 1	5,660	48.1	23.3	27.8	52.0	33.5	47.4	89.9	98.9	35.2
Cluster 2	14,228	50.7	26.7	22.5	4.1	0.1	5.1	63.3	88.4	19.1
Cluster 3	13,310	47.1	30.9	27.5	16.4	3.7	18.9	73.3	96.4	24.1
Cluster 4	12,211	47.4	28.3	10.5	39.6	7.4	28.4	95.0	98.8	35.8
Cluster 5	20,284	6.8	3.5	0.8	17.9	5.8	8.7	72.6	88.0	20.5
Cluster 6	16,588	15.4	8.8	4.0	43.3	22.2	34.3	72.9	96.2	23.5
Cluster 7	7,821	49.2	20.2	44.2	80.0	48.2	71.9	94.9	99.9	67.9
Cluster 8	7,388	68.2	42.5	26.4	45.6	14.1	34.7	96.3	90.9	49.2
Cluster 9	3,977	46.7	22.6	30.6	52.2	31.5	45.8	85.7	99.1	52.1
Cluster 10	7,382	53.4	44.1	30.1	9.4	0.1	6.1	35.1	68.7	21.6
Cluster 11	3,965	37.5	82.8	14.1	78.2	55.3	58.3	87.8	97.4	61.0
Cluster 12	5,366	68.3	47.2	29.5	36.3	15.4	36.3	69.3	98.5	15.3
Cluster 13	4,094	42.5	72.7	35.0	64.5	54.3	62.1	89.1	97.0	41.5
Cluster 14	4,508	31.0	14.0	8.8	5.5	3.1	3.0	34.9	37.0	10.3
Cluster 15	8,069	26.9	18.7	13.5	30.2	16.2	19.7	44.0	73.4	10.3
Cluster 16	5,244	17.3	12.3	4.8	11.1	0.0	11.1	61.7	81.9	6.0

Note: This table showcases the ratio of non-zero observations of investments in wind, solar, and natural gas-fueled generation technologies that take place in empty, own, and cross zip codes within regions. Values in the table characterize the percentages of non-zero investment observations out of total number of region-year observations in each k-means cluster for each generation technology. For example, on the first row we show out of 5,660 total region-year observations in cluster 1, 48.1% of all observed investments in wind generation capacity are only non-zero values that took place in empty zip codes. We sort clusters in this table based on the RPS residual values, going from cluster of regions with least to most stringent RPS programs.

Table 2.3: Model overview for hypotheses testing

Hypotheses	Main Independent Variables	Location Type	Main DVs Future Investments Wind	Solar	NG-Flexible
H1 and H2	Flexible prior investments	Empty	✓	✓	-
		Own	✓	✓	-
		Cross	✓	✓	-
H3 and H4	Wind prior investments	Empty	-	-	✓
		Own	-	-	✓
		Cross	-	-	✓
	Solar prior investments	Empty	-	-	✓
		Own	-	-	✓
		Cross	-	-	✓

Note: This table presents a summary of dependent and independent variables of interest that we use in evaluating our research hypotheses. Each check mark identifies the group fixed-effect estimation models using both OLS and Tobit approaches. We further conduct a within-cluster estimation of all models specified with check marks.

Table 2.4: Group fixed-effect (GFE) OLS estimation results of prior investment effects

	Empty Coefficient	R-square	Own Coefficient	R-square	Cross Coefficient	R-square
H1: Wind ~ Flexible	0.015^{*}	0.49	-0.001	0.37	-0.029^{*}	0.23
H2: Solar ~ Flexible	-0.104^{*}	0.49	-0.045^{*}	0.48	-0.056^{*}	0.42
H3: NG-Flexible ~ Wind	$0.005^{\dagger}$	0.13	0.034^{*}	0.18	0.056^{*}	0.13
H4: NG-Flexible ~ Solar	-0.067^{*}	0.14	-0.019^{*}	0.18	0.039^{*}	0.13

Note: This table provides an overview of the estimated coefficients to test hypotheses 1–4. The estimated coefficients are grouped fixed-effect estimates using an OLS approach. We report both coefficients of interest and the value of r-square for each estimated model. Significance level: $^{*}p < 0.01$; $^{\dagger}p < 0.05$.

Table 2.5: Group fixed-effect (GFE) Tobit estimation results of prior investment effects

	Empty		Own		Cross	
	Coefficient	Log-Likelihood	Coefficient	Log-Likelihood	Coefficient	Log-Likelihood
H1: Wind ∼ Flexible	0.060^{*}	-148012.42	0.062^{*}	-117141.01	-0.070^{*}	-103705.98
H2: Solar ∼ Flexible	-0.336^{*}	-153847.40	-0.302^{*}	-92236.20	-0.197^{*}	-146227.89
H3: NG-Flexible ∼ Wind	0.020^{*}	-222930.40	0.039^{*}	-236829.32	0.282^{*}	-129207.85
H4: NG-Flexible ∼ Solar	-0.104^{*}	-222802.35	-0.019^{*}	-236883.38	0.251^{*}	-129549.14

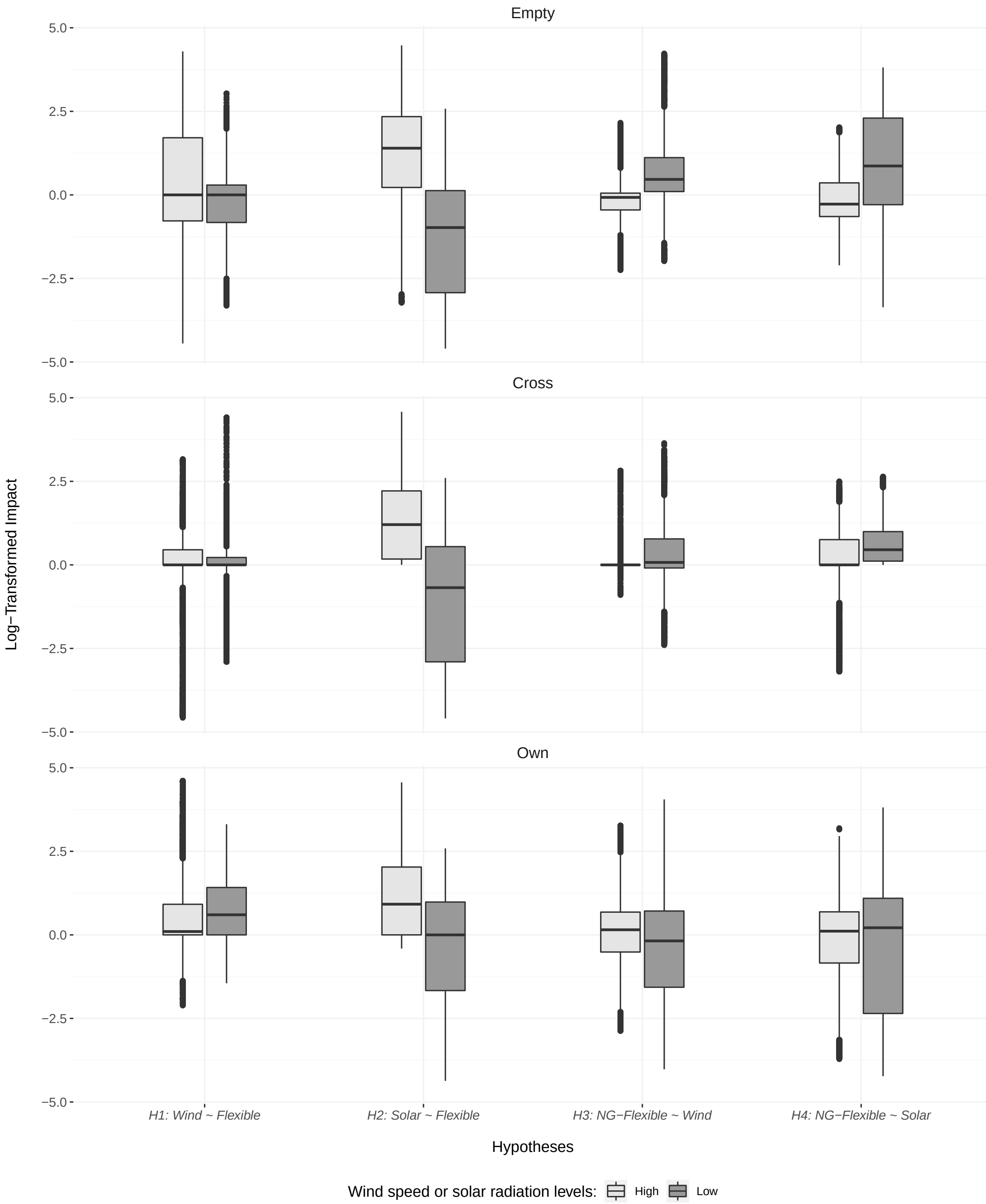
Note: In this table we provide an overview of estimated group fixed-effect coefficients for each hypothesis (1–4) using the Tobit estimation approach. We also provide a summary of log-likelihood values for each estimation model. Significance level: $^{*}p < 0.01$.

Table 2.6: Between-cluster Tobit estimation results of prior investment effects

	H1: Wind ∼ Flexible			***H2: Solar ∼ Flexible***			***H3: NG-Flexible ∼ Wind***			***H4: NG-Flexible ∼ Solar***		
	Empty	Own	Cross	Empty	Own	Cross	Empty	Own	Cross	Empty	Own	Cross
Cluster 1	0.43^{*}	1.12^{*}	0.81^{*}	-0.59^{*}	-0.34^{*}	-0.75^{*}	$0.05^{\dagger}$	-0.15^{*}	0.68^{*}	0.49^{*}	-0.24^{*}	0.36^{*}
Cluster 2	0.27^{*}	-0.02	0.08	-0.25^{*}	-0.04	$0.11^{\dagger}$	-0.07^{*}	$0.02^{\dagger}$	0.22^{*}	1.00^{*}	0.80^{*}	1.06^{*}
Cluster 3	0.12^{*}	0.38^{*}	0.12^{*}	0.65^{*}	1.53^{*}	0.94^{*}	-0.08^{*}	0.24^{*}	-0.17^{*}	0.16^{*}	-0.14^{*}	0.51^{*}
Cluster 4	0.35^{*}	3.38^{*}	4.72^{*}	0.21^{*}	0.77^{*}	0.67^{*}	0.41^{*}	-0.27^{*}	1.40^{*}	-0.49^{*}	-0.03	-1.11^{*}
Cluster 5	-0.34^{*}	0.91^{*}	0.52	0.26^{*}	0.24^{*}	0.23^{*}	0.03	-0.03^{*}	0.58^{*}	-0.21^{*}	0.06^{*}	0.62^{*}
Cluster 6	-0.01	0.07^{*}	0.05	0.03	0.24^{*}	-0.18^{*}	-0.40^{*}	-0.21^{*}	-0.01	-0.41^{*}	0.22^{*}	0.36^{*}
Cluster 7	-0.39^{*}	0.02	-0.65^{*}	-0.60^{*}	-0.36^{*}	-0.07	$0.05^{\dagger}$	0.02	0.67^{*}	-1.17^{*}	0.71^{*}	0.32
Cluster 8	0.12^{*}	0.12^{*}	0.19^{*}	-0.15^{*}	0.32^{*}	$0.04^{\dagger}$	1.00^{*}	0.41^{*}	0.81^{*}	0.42^{*}	-0.30^{*}	0.18^{*}
Cluster 9	-0.30^{*}	0.24^{*}	-0.41^{*}	0.23^{*}	$-0.15^{\dagger}$	0.44^{*}	0.43^{*}	-0.13^{*}	0.82^{*}	0.35^{*}	0.24^{*}	0.15^{*}
Cluster 10	0.03	-0.83^{*}	0.76^{*}	-0.20	0.36^{*}	1.00^{*}	0.15^{*}	0.23^{*}	0.39^{*}	-0.65^{*}	-0.23^{*}	-0.60
Cluster 11	0.01	-0.31^{*}	-0.18^{*}	-0.19^{*}	-0.09^{*}	-0.25^{*}	0.58^{*}	0.19^{*}	0.59^{*}	-0.56^{*}	0.40^{*}	0.10^{*}
Cluster 12	-0.18^{*}	1.06^{*}	-0.42^{*}	-0.79^{*}	-1.15^{*}	-0.76^{*}	0.09^{*}	-0.06^{*}	0.02	0.85^{*}	0.18^{*}	0.36^{*}
Cluster 13	$0.10^{\dagger}$	0.92^{*}	0.40^{*}	0.71^{*}	0.90^{*}	0.43^{*}	-0.08^{*}	-0.37^{*}	-0.11^{*}	0.06^{*}	-0.15^{*}	0.12^{*}
Cluster 14	0.37^{*}	0.34^{*}	1.76^{*}	0.38^{*}	0.14^{*}	0.86^{*}	$0.07^{\dagger}$	0.12^{*}	0.99^{*}	-0.38^{*}	-0.10^{*}	-0.04
Cluster 15	-0.18^{*}	0.48^{*}	0.16^{*}	0.72^{*}	0.91^{*}	0.89^{*}	$0.03^{\dagger}$	0.09^{*}	-0.03	-0.21^{*}	0.04^{*}	-0.75^{*}
Cluster 16	1.04^{*}	1.75^{*}	0.29	-0.04	–	-0.34^{*}	0.12^{*}	0.06^{*}	1.55^{*}	-0.79^{*}	0.06	$-1.23^{\dagger}$

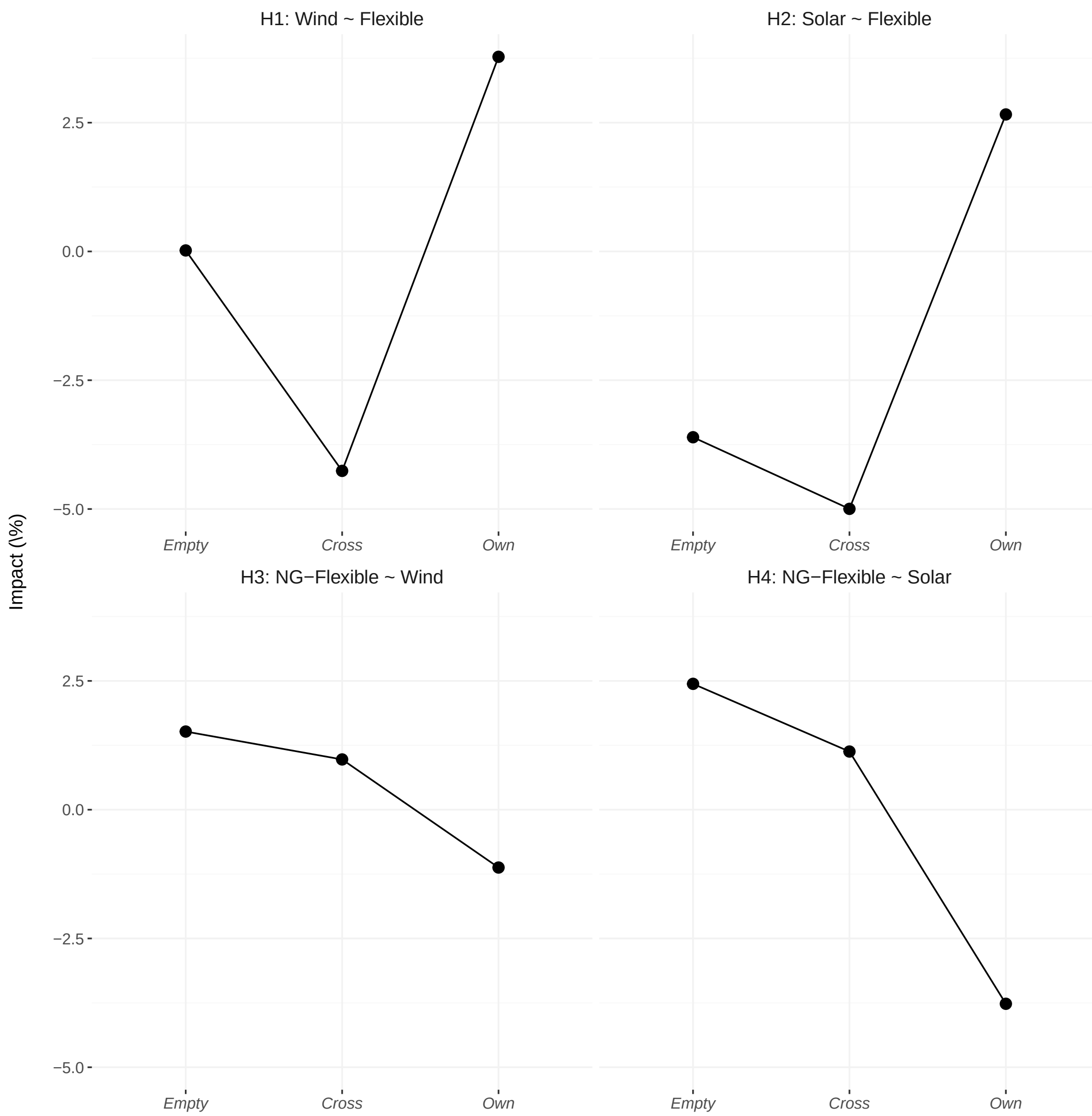
Note: This table summarizes a list of all estimated coefficients using Tobit approach for each cluster and each investment type. Each of the numbers presented above, identifies a single estimated model that compiles into a total of 191 individually estimated models. We do not have an estimated coefficient for cluster 16 when evaluating our second hypothesis for investments that took place in own zip codes. The reason is that in that cluster we do not observe any non-zero investments in solar technology that took place in own locations. Significance level: $^{*}p < 0.01$; $^{\dagger}p < 0.05$.

Figure 2.3: Prior investments impact distribution over different renewable potential levels



Note: This plot illustrates the distribution of prior investments impacts on future investments that take place in, empty, own, and cross zip codes moderated by the renewable potential of each regions. We have classified clusters into 2 groups of high and low renewable potential based on their average levels of solar radiation or wind speed. Depending on the type of renewable technology of interest in each of our hypotheses, we sort clusters based on their average wind speed or solar radiation and find the median value. Subsequently, all clusters with average solar wind speed (or solar radiation) below the median value are considered as locations with low levels of renewable potential and those with averages above the median value are considered as locations with high renewable potential.

Figure 2.4: Prior investments average impact over different investment strategies



Note: This plot represents the percentage impact of prior investments on future investments across different investment strategies (i.e., empty, cross, and own). The impacts are calculated based on a set of ceteris paribus counterfactual analysis done using within cluster estimated models. Impact values showcase the difference between the predicted values of future investments based on the actual observations of the prior investment and the predicted values of future investments assuming prior investment had not existed (i.e., setting it equal to zero).

Chapter 3

Flexibility Unpacked: Startup Times and Their Impact on the Economics of Energy Mix Strategies

3.1 Introduction

The landscape of the U.S. electricity sector is undergoing a significant transformation, propelled by the rapid integration of renewable energy sources. This shift is driven by the dual imperatives of reducing carbon emissions and enhancing energy sustainability. However, the intermittent nature of renewable energy sources such as wind and solar introduces substantial challenges to grid stability and reliability. Addressing these challenges necessitates a reevaluation of the grid's operational flexibility, particularly in the context of conventional power generation.

Recent forecasts, including a comprehensive analysis by Goldman Sachs, predict a sharp increase in electricity demand, primarily fueled by the exponential growth of AI and data centers (Davenport et al., 2024). The report estimates that by 2030, data centers will account for approximately 8% of total U.S. power demand, up from about 3% in 2022. To accommodate this surge, an estimated \$50 billion investment in new power generation capacity is required, with a significant portion allocated to renewable sources. This development underscores a critical issue: the increasing reliance on renewable energy amplifies the grid's vulnerability to fluctuations in power supply, thus elevating the importance of flexibility in conventional power generation. These unprecedented shifts in the energy landscape emphasizes the importance of understanding the economic interaction between the two major electricity generation technologies including intermittent (i.e., wind and solar) and flexible. In the first and second chapters we have established the nuanced relationship between these generation technologies while we delve deeper into this relationship by arguing that not all flexible generators are built the same dependent on their technological capabilities.

Flexible generators are one of the main sources of flexibility in electricity grids. These are capable of changing their output levels in relatively short time spans to respond to supply and demand fluctuations. However, there are differences in how fast a generator can react to such changes through ramping up (down). In the U.S. we do not observe the ramping capabilities of prior investments or new investments in flexible generations technologies. Nevertheless, we observe how much time it takes a conventional flexible power plant to go from cold shut down to fully operational. In this study we leverage this observation and create a continuous flexibility level variable that serves as a proxy to distinguish flexibility capabilities of electricity generators. In the U.S. power grid, there are flexible generators with start-up times ranging from less than 10 minutes to more than 12 hours. According to the Energy Information Administration, in 2020 66% of total U.S. generation capacity had a start-up time of more than 1 hour with 25% having less than 1 hour start-up time (Comstock, 2024). This represents a flexibility heterogeneity that may differently influence future investments in solar and wind generators given their intermittency pattern dissimilarities. To this end, in this study we pose the following research question. *How flexibility heterogeneity of conventional electricity generators moderate the relationship between investments in intermittent and flexible generation technologies?*

We answer this research question by the data set we collected and combined in chapter 1, utilizing a panel data of U.S. electricity industry spanning from 2002 to 2019 representing operational and investment dynamics of different electricity generation technologies. We adopt an empirical strategy similar to the ones we detailed in chapters 1 and 2, while developing a new continuous flexibility measure to account for the flexibility heterogeneity of the conventional electricity generators.

The rest of this chapter is organized as the following. In Section 3.2 we provide a summary of the existing literature and how this study contributes to the existing literature while building upon it and develop our research hypotheses. Next, we detail our data collection and variable construction in Section 3.3 and we present an overview of our empirical strategy in Section 3.4. Lastly we illustrate our findings in Section 3.5 followed by concluding remarks in Section 3.6.

3.2 Literature Review and Hypotheses

The concept of operational flexibility in the power sector has been explored extensively through diverse lenses in operations management and energy economics. Each field contributes unique insights into how flexibility can be leveraged to improve system efficiency, particularly as energy markets evolve with increased integration of renewable energy sources. This literature review synthesizes findings from these domains, setting the stage for a deeper examination of how flexibility, specifically defined by startup times of power generation units, affects economic and operational outcomes in energy systems.

Operations management research has traditionally focused on flexibility as a critical element of manufacturing and service operations, emphasizing its role in enhancing responsiveness to market dynamics and improving operational efficiency. Upton (1994) broadly defined operational flexibility as the ability to change or adapt operations in a timely manner, at a reasonable cost, and without significant performance degradation. This foundational perspective was further detailed by Sethi and Sethi (1990), who categorized flexibility into several types, including machine, labor, and routing flexibility, each pertinent to manufacturing settings. In the context of energy operations, the notion of flexibility extends to how power generation systems can adjust to changes in demand and supply, particularly with the intermittent nature of renewable energies like wind and solar. Studies such as De Toni and Tonchia (1998) and Vokurka and O'Leary-Kelly (2000) have explored the implications of flexibility in industrial settings, noting that operational agility can significantly reduce costs and improve service levels. These insights are particularly relevant to the electricity sector, where demand variability is high, and the integration of renewables introduces additional uncertainty. Jack and Raturi (2003) applied these principles to the energy sector, investigating how volume flexibility—defined as the ability to modulate output without significant cost impacts—can enhance the economic performance of power plants. Their study highlighted that volume flexible firms are better equipped to handle market variability, leading to improved profitability and competitiveness. This is especially pertinent in energy markets increasingly dominated by renewable energy sources, which introduce significant variability in power generation.

Our research builds on this foundation by quantitatively exploring how the startup flexibility of power plants contributes to economic efficiencies in energy markets. By introducing a continuous flexibility measure based on startup times, this study extends the operational flexibility discussion in operations management to the specific challenges faced by the power sector, particularly under high renewable penetration scenarios.

In energy economics, the focus shifts to examining how flexibility impacts the integration of renewable energy and the overall stability and efficiency of power systems. The increasing share of renewables in energy mixes worldwide necessitates a reevaluation of traditional power system operations. Wang et al. (2016) and Zhang et al. (2016) have both studied the operational and economic challenges posed by the intermittent and unpredictable nature of renewable energy sources. Their research underscores the need for power systems to possess sufficient flexibility to manage fluctuations in power generation and maintain reliability and stability. Bistline (2018) further explored the economic value of operational flexibility in electricity markets, particularly through the lens of reducing minimum load levels on coal- and gas-fired power plants. His analysis revealed that more flexible generation assets could significantly reduce system costs, decrease renewable curtailment, and enhance system reliability. These findings are crucial in contexts where policy goals, such as reductions in greenhouse gas emissions, are prioritized alongside maintaining economic competitiveness and system reliability.

In this study we extend these insights by focusing on a specific aspect of flexibility—startup times—and its quantifiable impact on the economic interplay between investments in flexible and intermittent sources. By developing a nuanced measure of flexibility that reflects the real-world operational capabilities of power plants, our research contributes to a more detailed understanding of how flexibility can be optimized to support the economic integration of renewables into the energy mix.

We build upon the aforementioned studies as well as the detailed study of the literature that we've provided in Sections 1.2 and 2.2 to formulate our research hypotheses. We keep the hypotheses structure we developed in Section 2.3 for clarity sake and further hypothesize that the influence of prior investments in flexible generation technologies is positively moderated by flexibility levels.

That is to say, the higher the flexibility level of prior investments in flexible generation technologies, the larger the positive impact they have on future investments in intermittent renewable generation technologies (i.e., wind and solar). Given that wind and solar generation technologies represent different intermittency patterns we argue that subsequently they have different flexibility requirement profiles. Thus, we formulate the same hypothesis for wind and solar separately arguing that flexibility level has a more pronounced impact on future investments in solar generation technology compared to future investments in wind technology.

Hypothesis 1. *Higher flexibility levels of prior flexible generation capacity investment around a focal zip code positively increases future wind investment rates around that focal zip code.*

Hypothesis 2. *Higher flexibility levels of prior flexible generation capacity investment around a focal zip code positively increases future solar investment rates around that focal zip code.*

Similarly, we hypothesize a reverse relationship in a fashion we did in chapters 1 and 2. This means, we argue that the economic interplay between investments in flexible and intermittent generation technologies works in both ways. However conversely, we hypothesize that the positive influence of prior investments in intermittent generation technologies on future investments in flexible generation technologies – *driven by increases in flexibility requirements that leads to higher dispatchability premiums* – is undermined by the increased competition and a saturation effect. Therefore, we expect higher levels of flexibility to negatively moderate the effect of prior investments in intermittent generation technologies on future investments in flexible natural gas-fueled electricity generators. We formulate these hypotheses as the following separately hypothesizing for wind and solar generation technologies.

Hypothesis 3. *Higher flexibility levels of prior wind generation capacity investment around a focal zip code decreases future flexible natural gas-fueled investment rates around that focal zip code.*

Hypothesis 4. *Higher flexibility levels of prior solar generation capacity investment around a focal zip code decreases future flexible natural gas-fueled investment rates around that focal zip code.*

3.3 Data and Variable Construction

Since we are building upon the hypotheses examination structure we developed in chapter 1, we use the same data set to test the hypotheses we developed in Section 3.2. For the sake of brevity, we eschew from detailing all of the data collection and pre-processing steps that are already provided in Sections 1.4.1 and 1.4.2, instead we explain how we develop our continuous flexibility level measure. As we detailed in Section 1.4.2.1, we use the *"time it takes for a conventional electricity generator to go from cold shutdown to fully operational"* reported in Energy Information Administration form 860, as a proxy for the operational flexibility of electricity generators. That said, this variable in its raw form is a categorical variable with 4 levels including: startup times less than 10 minutes, between 10–60 minutes, 1–12 hours, and beyond 12 hours. Similar to our empirical setting in chapter 1 and 2, we consider electricity generators with startup times above 12 hours as inflexible and inline with the scope of our study we focus on the three groups of startup times below 12 hours.

We set the level of analysis to be on the region-year level with regions defined as a 100-miles radius around any focal zip code in the U.S. that encompasses all zip codes with their center coordinates falling within this range. This implies that after collecting generator and plant level data, we first aggregate into zip code level observations and then in to regional level observations. With this approach, for each zip code in the U.S. we first calculate the amount of generation capacity with different flexibility levels. Namely, we create 3 groups of capacity variables including: low flexibility (startup time between 1 to 12 hours), moderate flexibility (startup time between 10 to 60 minutes), and high flexibility (startup time below 10 minutes). Next, we propose a non-linear transformation approach to assign numerical values to the three flexibility groups. We adopt a non-linear approach since our flexibility group observations have significantly different widths (a 12 hour versus 10 minutes widths for the two extreme groups).

We use the logarithmic scale to map startup times more effectively by exaggerating the differences between smaller values (shorter startup times) and compressing differences between larger values. The following steps summarize our approach:

1. **Calculate Logarithmic Scores:** Use the logarithm of the startup time to compute scores. We use Equation 3.1 to calculate the flexibility score:

$$\text{Flexibility Score} = A - B \times \log(\frac{\text{Startup Time}}{Min(\text{Startup Time})}) \tag{3.1}$$

Where A and B are constants that are set in a way to scale the flexibility scores between 0 and 100 appropriately[1].

2. **Determine Constants A and B:** A is set in a way so that the maximum flexibility score for the smallest startup time is 100, while B is calculated in a way to ensure that the minimum flexibility score is slightly above 0 (in this study we set it to be the arbitrary value of 10). We derive B by setting the minimum and maximum levels of flexibility score to be 10 and 100 respectively through solving Equation 3.2 in the following:

$$B = \frac{A - 10}{\log(\frac{Max(\text{Startup Time})}{Min(\text{Startup Time})})} \tag{3.2}$$

This formula for B scales the logarithmic differences in startup times into a linear score range defined by our minimum and maximum scores, ensuring a meaningful and interpretable distribution of scores based on flexibility. With these calculation, we assign the values of 100, 62.54, and 10 to high, moderate, and low flexibility groups.

After computing numerical values for each flexibility groups in each zip code, we characterize how much generation capacity resides in each zip code. Subsequently, with the combination of these values we can calculate average flexibility levels for regions using a weighted average method presented in Equation 3.3.

$$\text{Flexibility Score}_{r,t} = \sum \text{Capacity Weight}_{r,f,t} \times \text{Flexibility Score}_{f} \tag{3.3}$$

For each region r we calculate the flexibility score at time period t by finding the capacity weight of different flexibility groups $f \in \{\text{High, Moderate, Low}\}$ in each region. Leveraging this approach, we compute a continuous measure of flexibility for each region r at each time period t that is the weighted average of all flexible sources in a given region with a zip code as its center that captures how much flexibility is available from different sources based on the time it takes for these generators to go from cold shutdown to fully operational.

In Figure 3.1 we illustrate the flexibility score of each state averaged over regions and time periods (i.e., between 2002 to 2019). This figure effectively illustrates the geographical distribution of flexibility in power generation. This visualization underscores the variability in electricity generation responsiveness, with some states achieving higher flexibility scores, indicative of a greater

[1] Nevertheless, we avoid the value 0 and set the minimum to be 10 since a value of 0 is not an accurate representation to of the existing flexibility level.

prevalence of how electricity generators are capable of faster start-up times. In addition to the observed heterogeneity across states with respect to the flexibility scores, in Figure 3.2 we showcase the within-state variability of flexibility scores for regions in each state. In this plot we sorted states in an ascending order where states with highers levels of average flexibility are at the bottom of the plot and those with lowest average flexibility scores are at the top of the plot. This figure indicate that we can identify two distinct patterns of flexibility score distributions. There are states with a high range of flexibility levels such as New York, Connecticut, North Carolina, Delaware, South Carolina, and California opposed to states with very small ranges of flexibility levels such as, West Virginia, Oklahoma, Texas, Tennessee, New Hampshire, Maine, and Vermont. These descriptive statistics highlight the importance of data granularity when examining the role of operational flexibility in moderating the economic interplay between flexible and intermittent electricity generation technologies.

3.4 Empirical Strategy

We build upon the multi-staged empirical approach that we proposed in chapter 1, detailed in Section 1.5.In the following we explain how we modify this approach to answer our research question. We augment our empirical approach proposed in chapter 1 with a novel continuous flexibility measure that is introduced into our estimation framework via an interaction term. In the following we summarize our empirical strategy: *(ii)* We correct the endogeneity of the renewable portfolio standard (RPS) obligations with a first stage estimation and calculate RPS residuals that encapsulate states' RPS programs stringency levels. *(ii)* We combine the principal component analysis (PCA) and k-means clustering to create clusters of homogeneous regions using 11 features (Table 1.7 provides a list of clustering features). *(iii)* We run grouped fixed effect (GFE) OLS estimation of the influence of prior investments on future investments for flexible and intermittent generation technologies while including an interaction term for flexibility score. *(iv)* Additionally, we run each of the estimation models for each of the clusters to examine the distribution of effects across clusters.

In the following, we provide modified versions of models we estimated in chapter 1 to examine hypotheses we developed in Section 3.2:

$$
\begin{aligned}
ln[\mathbf{Investment}^{Intermittent}_{r,e,t'}] = \alpha_0^{'} + \alpha_1^{'} \cdot ln[\text{Prior Investment}^{Flexible}_{r,t}] + \\
\alpha_2^{'} \cdot ln[\text{Flexibility Score}_{r,t}] + \\
\alpha_3^{'} \cdot ln[\text{Prior Investment}^{Flexible}_{r,t}] \times ln[\text{Flexibility Score}_{r,t}] + \\
\boldsymbol{\alpha_4^{'}} \cdot ln[\mathbf{Controls}_{r,e,t}] + \underbrace{\alpha_4^{'} \cdot ln[\hat{\tau}_{r,t}]}_{\text{RPS Residuals}} + \underbrace{\kappa_r}_{\text{Cluster FE}} + \zeta_{r,e,t}
\end{aligned}
\tag{3.4}
$$

$$
\begin{aligned}
ln[\text{Investment}^{NG-Flexible}_{r,t'}] = \beta_0^{'} + \boldsymbol{\beta_1^{'}} \cdot ln[\textbf{Prior Investment}^{Intermittent}_{r,e,t}] + \\
\boldsymbol{\beta_2^{'}} \cdot ln[\text{Flexibility Score}_{r,t}] + \\
\boldsymbol{\beta_3^{'}} \cdot ln[\textbf{Prior Investment}^{Intermittent}_{r,e,t}] \times ln[\text{Flexibility Score}_{r,t}] + \\
\boldsymbol{\beta_4^{'}} \cdot ln[\mathbf{Controls}_{r,e,t}] + \underbrace{\kappa_r}_{\text{Cluster FE}} + \eta_{r,e,t}
\end{aligned}
\tag{3.5}
$$

Equations 3.4 and 3.5 model the relationship between future investments in intermittent (flexible) generation technologies and prior investments in flexible (intermittent) generation technologies. By estimating $\alpha_3^{'}$ and $\boldsymbol{\beta_3^{'}}$, we examine our research hypotheses. In these models we use a 3-year aggregated time frame with leading dependent variables examined at time period $t^{'}$ and independent variables at period t. We use model specifications we detailed in chapter 3 as represented in Tables 1.9 and 1.10.

With this empirical strategy similar to chapter 1 we overcome several empirical challenges. Firstly, we tackle the issue of endogeneity in renewable portfolio standard (RPS) obligations by employing a two-stage instrumental variable approach. In the first stage, we regress RPS obligations on key drivers of intermittent generation investments, such as prior investments in solar and wind capacity, actual electricity generation by these sources, and relevant control variables including solar radiation and wind speed. This allows us to extract the residuals of the RPS obligations, which we then incorporate into our main regression models to correct for potential endogeneity. Secondly, we manage the unobserved continuous heterogeneity of regions, through implementing a two-step grouped fixed effect (GFE) estimation method that allows us to initially identify latent clusters based on regional characteristics such as natural gas pipeline capacity, renewable energy intermittency, and regional renewable potentials. By incorporating these clusters as fixed effects in our regression models, we effectively control for unobserved heterogeneity that could bias our results. These methodological steps collectively enable us to produce robust and unbiased estimates of the moderating impact of flexibility levels on the economic interplay between investments in intermittent and flexible electricity generation technologies.

3.5 Results

Our analysis provides comprehensive insights into the role of startup times and flexibility in shaping the economics of energy mix strategies. The regional flexibility score descriptive statistics, presented in Table 3.1, reveal substantial variation across different regions in the U.S. This broad range indicates significant disparities in generation flexibility across the U.S., which likely influences the economic interplay between prior and future investments in flexible and intermittent electricity generation technologies. The mean flexibility score of clusters range from 26.3 to 78.8 highlight this diversity, underscoring flexibility heterogeneity of regions in the U.S.

The group fixed-effect (GFE) estimation results in Table 3.2 show the impact of flexibility levels on subsequent renewable energy investments, using a fixed-effects model to control for unobserved heterogeneity. Our key findings indicate that the flexibility score of regions positively moderate the relationship between prior investments in flexible generation technologies and future investments in intermittent generation technologies providing empirical support for hypotheses 1 and 2. These results underscore a strong positive relationship between electricity generation flexibility levels and renewable energy investments, with wind energy benefitting more significantly from increased flexibility (illustrated by larger estimated coefficient). This disparity may stem from the differing operational characteristics and integration challenges associated with wind and solar technologies. Moreover, we estimated a negative coefficient for the moderating role of the flexibility levels on the relationship between prior investments in solar generation technology and future investments in flexible natural gas-fueled generation technology. This finding supports our hypothesis 4 and suggesting the existence of a saturation effect with respect to flexibility availability in a given region. We couldn't find empirical support for our third hypothesis.

Further insights are provided by the between-cluster estimation results in Table 3.3, which analyze the impact of flexibility levels across different clusters of regions. This approach allows us to capture the nuanced effects of flexibility levels in varying regional contexts. The cluster-specific results emphasize that the majority of estimated coefficients are positive for hypotheses 1 and 2 while conversely the majority of coefficients are negative for hypotheses 3 and 4 providing further empirical support for our research hypotheses. These coefficients suggest that the moderating relationship might be non-linear.

Figure 3.3 visually support these statistical findings by illustrating the geographical distribution and evolution of flexibility levels and their correlation with renewable energy investments. More specifically, we observe from this figure that states with a steady increases in the flexibility score have steadily invested in intermittent generation technologies namely, California, Colorado, Connecticut, North Carolina, Iowa, and Texas. Additionally, we observe a distinct behavior when higher flexibility scores are followed by larger investments in intermittent generation technologies such as in D.C., Arkansas, South Carolina, and Delaware.

In summary, our results robustly demonstrate that regions with higher generation flexibility levels are better positioned to attract renewable energy investments, particularly in wind energy. These findings provide critical insights for policymakers and investors, highlighting the need for targeted investments in generation flexibility to facilitate the transition to a sustainable energy future. By fostering greater flexibility, we can better accommodate the growing share of renewable energy, ensuring a reliable and efficient electricity grid for the future.

3.6 Concluding Remarks

Our findings in this chapter provide valuable insights into the impact of startup times on the economics of energy mix strategies, particularly emphasizing the importance of generation flexibility capabilities in attracting renewable energy investments. We used a unique panel data set to examine the moderating role of conventional electricity generators' technological capability to respond to electricity supply (and demand) fluctuations. We proposed a novel continuous flexibility measure that quantifies this technological capability of conventional electricity generators leveraging the time it takes a power generator to go from cold shutdown to fully operational. We utilized this data set along with a multi-staged empirical strategy to shed light on how this particular technological capability influences the future energy landscape of the U.S. We found empirical evidence that higher levels of flexibility positively moderates the relationship between prior investments in flexible generation technologies and future investments in wind and solar. This is while higher flexibility levels negatively moderate the interplay between prior investments in intermittent generation technologies and future investments in natural gas-fueled generation technology.

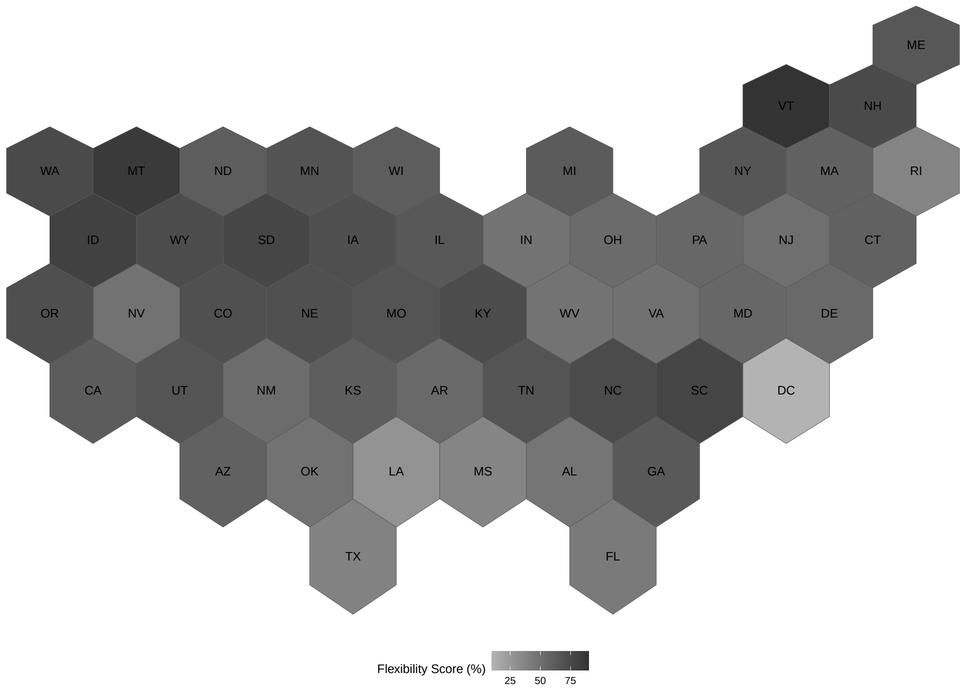
Nevertheless, several limitations of this study warrant discussion and offer directions for future research. Firstly, while our analysis identifies a positive correlation between flexibility and renewable energy investments, the data is constrained to the U.S. electricity market from 2002 to 2019. This temporal and geographical limitation suggests that the results may not be entirely generalizable to other regions or periods with different regulatory environments, technological advancements, or market conditions. Future studies could extend this analysis to include a broader range of countries and more recent data to validate the robustness of our findings. Secondly, while we account for major control variables, there are potential unobserved factors such as local policy changes, technological innovations, and specific market behaviors that could influence both flexibility and renewable investments. These unobserved variables might introduce biases into our estimations. Future research could incorporate advanced econometric techniques or experimental designs to better isolate the causal effects of grid flexibility on renewable energy investments.

Moreover, the interaction between different types of renewable energy sources and their specific flexibility requirements remains an underexplored area. While our study distinguishes between wind and solar investments, it does not delve into the synergistic or antagonistic effects of integrating multiple renewable sources simultaneously. Investigating these interactions could yield valuable insights into optimizing energy mix strategies and enhancing overall grid resilience.

In summary, while this study highlights the pivotal role of grid flexibility in promoting renewable energy investments, it also underscores the need for further research to address the identified limitations. By expanding the geographical scope, employing more granular data, considering unobserved factors, examining long-term impacts, and exploring interactions between different renewable sources, future studies can build upon our findings and contribute to a more comprehensive understanding of the dynamics shaping the modern electricity grid.

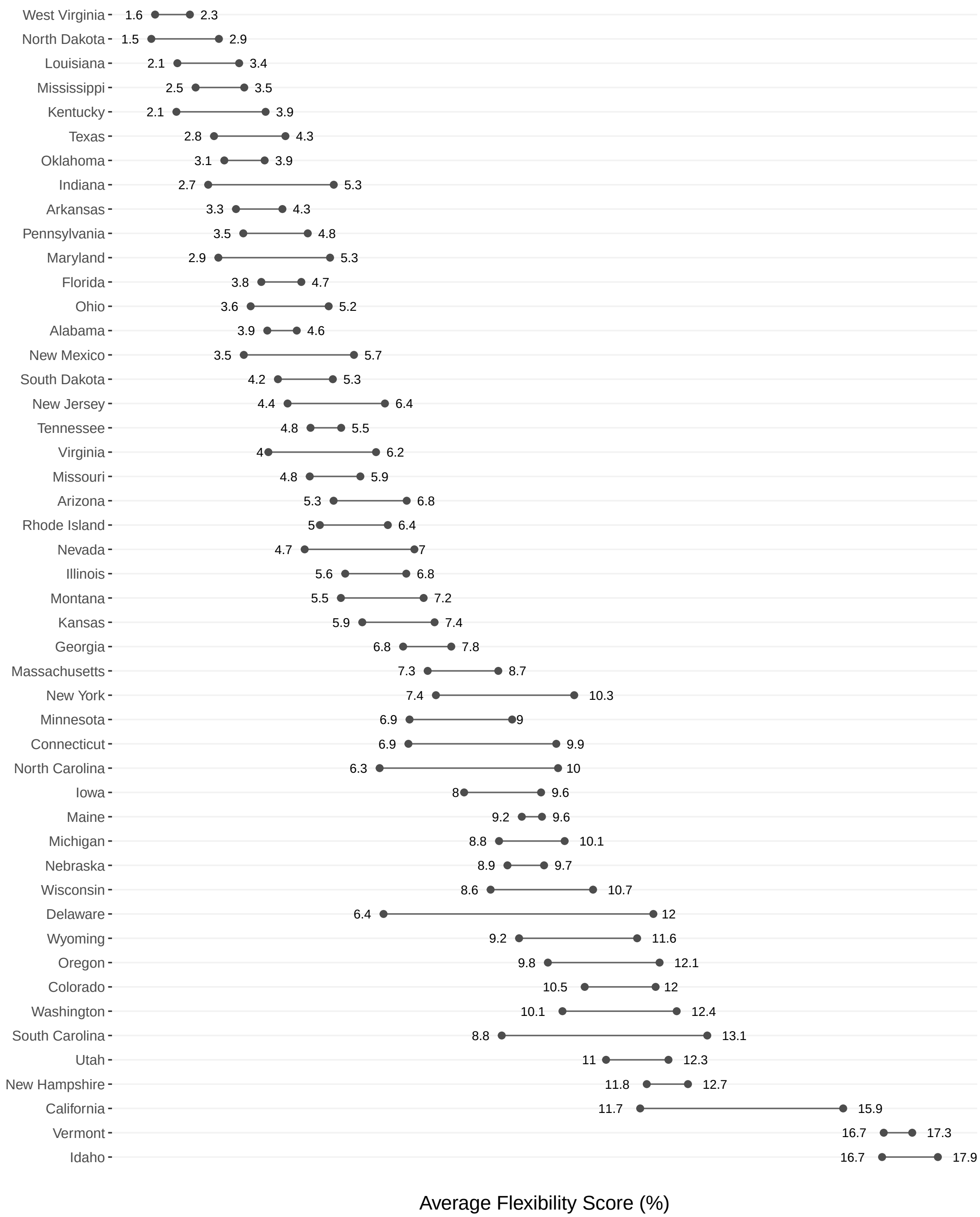
Exhibits

Figure 3.1: State-by-state average flexibility levels between 2002–2019



Note: In this plot we illustrate a snapshot of the U.S. flexibility profile by plotting the average flexibility scores of regions within the each state that ranges from 0 to 100. States with darker shades, represent a higher level of flexibility with states showing lighter shades are the ones with lower flexibility levels.

Figure 3.2: Average flexibility score distribution across states between 2002–2019



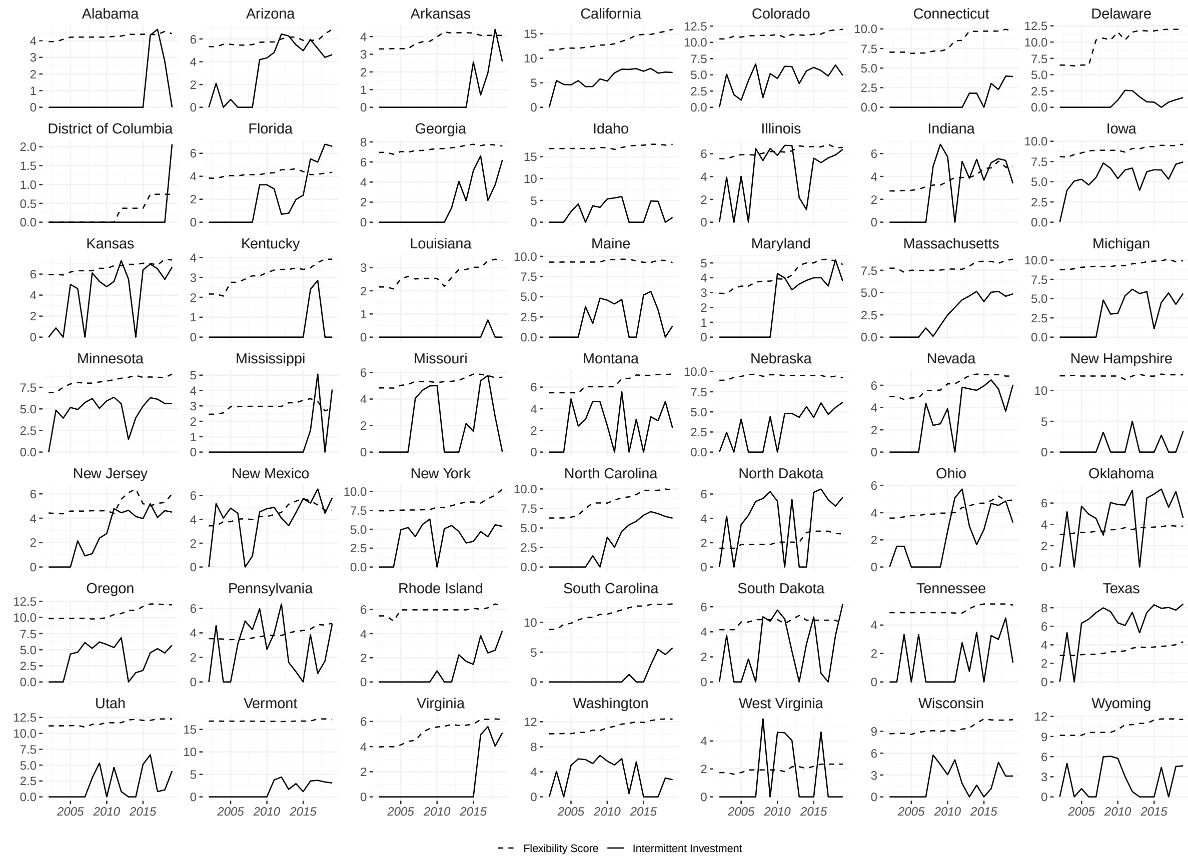
Note: In this plot we depict the distribution of flexibility scores for each state between 2002–2019. To this end, we calculate the average values of flexibility scores for each state in each year based on the flexibility scores of the individual zip codes. Next we identify the min and max values of flexibility scores across years 2002–2019 for each state. We plot them in an ascending order of total average flexibility scores representing states with lowest values of flexibility score on the top and as we move down, the average state-level flexibility score increases. For example, West Virginia has the lowest average flexibility score ranging between 1.6% to 2.3% while Idaho has the highest average flexibility score ranging between 16.7% to 17.9%.

Table 3.1: Regional flexibility score descriptive statistics

	Min	*Q1*	*Q2*	*Q3*	*Max*	*Std. D.*	*Mean*
Cluster 1	10.9	20.4	23.5	28.9	95.0	10.2	26.3
Cluster 2	5.2	18.2	24.8	35.8	100.0	17.6	30.3
Cluster 3	0.0	20.3	30.6	40.2	100.0	20.2	34.9
Cluster 4	11.7	26.9	33.5	46.5	80.5	12.7	36.6
Cluster 5	0.0	16.4	30.4	55.5	100.0	27.1	36.8
Cluster 6	3.9	28.4	40.6	49.6	100.0	15.9	40.2
Cluster 7	13.8	30.5	40.3	55.3	86.0	15.1	43.2
Cluster 8	17.0	29.7	39.9	48.2	100.0	19.6	43.9
Cluster 9	16.9	36.7	41.2	54.3	87.9	11.1	45.3
Cluster 10	0.0	28.3	50.3	95.3	100.0	32.1	56.3
Cluster 11	0.0	38.0	57.0	75.9	100.0	25.6	56.3
Cluster 12	0.0	44.4	59.3	77.9	100.0	22.0	61.2
Cluster 13	0.0	52.1	61.0	73.0	100.0	14.1	62.9
Cluster 14	7.6	55.7	67.9	79.1	100.0	18.5	69.2
Cluster 15	7.1	53.5	68.5	89.2	100.0	21.4	70.0
Cluster 16	18.7	58.2	83.9	100.0	100.0	21.4	78.8

Note: This table summarizes descriptive statistics of regional flexibility scores for each of the clusters we identified using the combination of principal component analysis (PCA) and k-means clustering. Clusters are sorted in an ascending order based on the average flexibility score of clusters. Cluster 1 has the lowest average flexibility score while cluster 16 has the largest flexibility score.

Figure 3.3: State-by-state average flexibility level and intermittent investment trends



Note: This table illustrates the trends of changes in states' average flexibility score and their total investments in intermittent generation technologies (i.e., wind and solar) combined capacity. The investment values are logged for the sake of comparability.

Table 3.2: Group fixed-effect (GFE) estimation results of flexibility level interaction term

	Interaction (Log-Log)	R-square (%)
H1: Wind ∼ Flexible	0.016^{*}	54.93
H2: Solar ∼ Flexible	0.090^{*}	52.81
H3: NG-Flexible ∼ Wind	0.004	25.32
H4: NG-Flexible ∼ Solar	-0.139^{*}	25.14

Note: This table provides an overview of the estimated interaction terms to test hypotheses 1–4. The estimated coefficients are grouped fixed-effect estimates using an OLS approach. We report both coefficients of interest and the value of r-square for each estimated model. Significance level: $^{*}p < 0.01$; $^{\dagger}p < 0.05$.

Table 3.3: Between-cluster estimation results of interaction terms

	H1: Wind ∼ Flexible		***H2: Solar ∼ Flexible***		***H3: NG-Flexible ∼ Wind***		***H4: NG-Flexible ∼ Solar***	
	Interaction	R-square (%)	Interaction	R-square (%)	Interaction	R-square (%)	Interaction	R-square (%)
Cluster 1	0.17^{*}	60.0	0.26^{*}	43.1	-0.05	73.7	-1.87^{*}	74.4
Cluster 2	-0.08^{*}	23.6	0.21^{*}	29.5	-0.05	16.7	0.78^{*}	17.5
Cluster 3	-0.18^{*}	71.9	0.26^{*}	56.9	-0.00	29.1	0.07^{*}	28.5
Cluster 4	0.49^{*}	42.2	-0.18^{*}	41.7	-0.37^{*}	36.6	-0.43^{*}	29.1
Cluster 5	0.18^{*}	11.2	-0.04^{*}	44.1	-0.01	8.8	0.04	9.0
Cluster 6	-0.09^{*}	69.7	0.12^{*}	44.8	-0.07^{*}	17.8	-1.78^{*}	17.7
Cluster 7	0.50^{*}	61.6	-0.40^{*}	36.6	0.32^{*}	28.3	-0.91^{*}	28.9
Cluster 8	1.65^{*}	49.0	-0.32^{*}	75.7	-0.16^{*}	72.5	0.40^{*}	70.8
Cluster 9	0.17^{*}	32.7	0.70^{*}	63.5	-0.50^{*}	56.3	-0.15^{*}	53.9
Cluster 10	-0.18^{*}	45.5	0.13^{*}	48.4	-0.14^{*}	30.1	-0.05^{*}	26.6
Cluster 11	-0.20^{*}	51.6	-0.05^{*}	43.8	0.06^{*}	16.6	1.75^{*}	16.7
Cluster 12	0.15^{*}	31.5	0.12^{*}	63.8	0.07^{*}	54.9	0.09^{*}	54.6
Cluster 13	0.18^{*}	52.7	0.63^{*}	67.4	0.21^{*}	29.1	$-0.09^{\dagger}$	26.4
Cluster 14	-0.23^{*}	48.3	0.34^{*}	61.5	-0.04	30.0	0.10	29.5
Cluster 15	-0.25^{*}	33.6	$0.07^{\dagger}$	48.0	0.11^{*}	22.0	-0.89^{*}	22.8
Cluster 16	0.26^{*}	54.6	0.70^{*}	72.3	-0.13^{*}	65.5	$0.06^{\dagger}$	62.2

Note: This table represents the individual estimates of prior investment coefficients and their corresponding r-squares as independent variable for future investments as dependent variables. Clusters are sorted in an ascending order based on the average flexibility level of each cluster with clusters having the highest average flexibility score being on the bottom of the table (i.e., cluster 16) and those with least flexibility score on top of the table (i.e., cluster 1). Significance level: $^{*}p < 0.01$.

Appendices

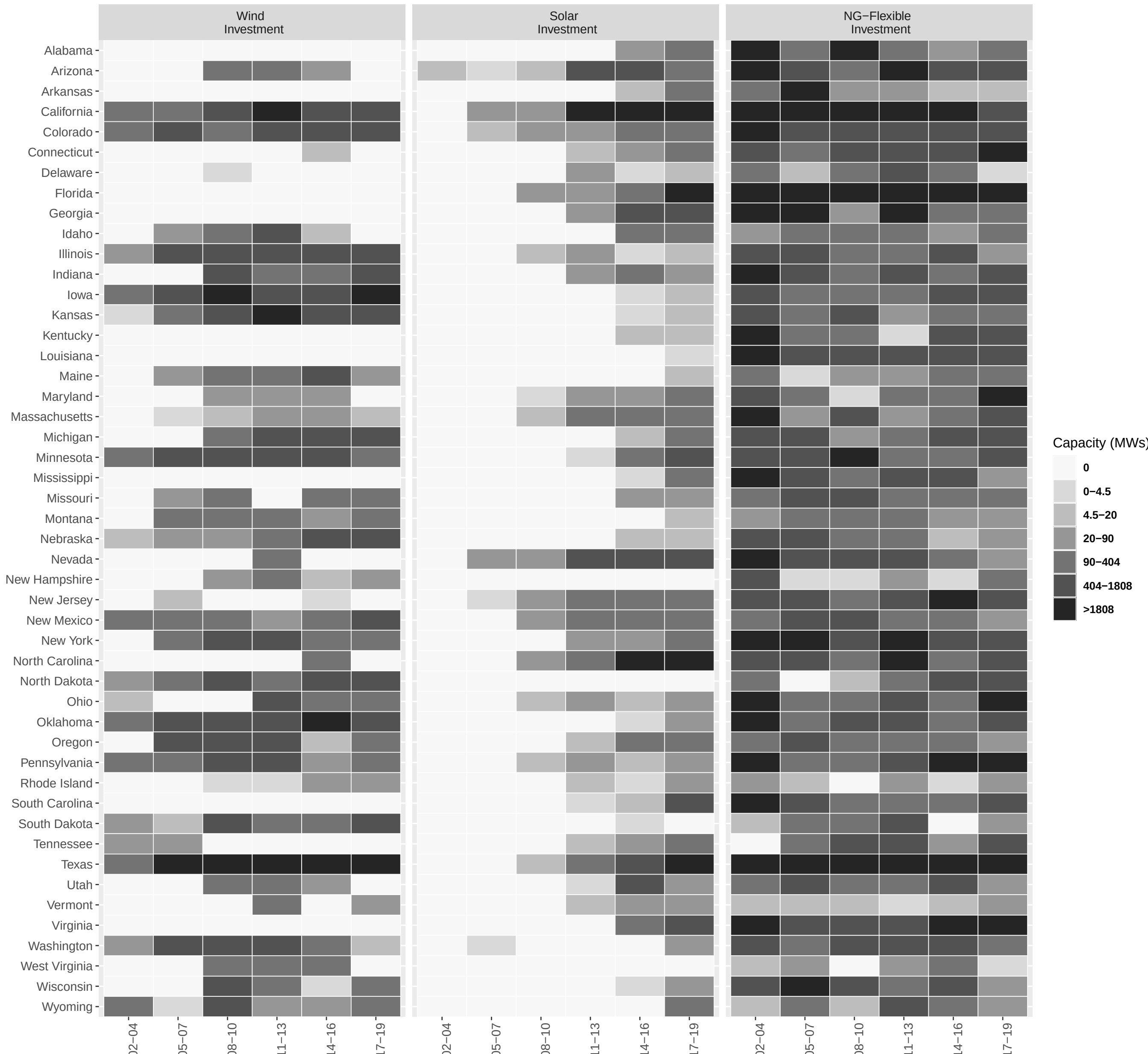
Appendix A Descriptive Statistics

Table 1.A1: Summary statistics of the U.S. electricity grid new generation capacity investment stages

		2002–2004	2005–2007	2008–2010	2011–2013	2014–2016	2017–2019
State-level	Solar	0.00	17.12	64.97	46.82	49.03	166.42
Proposed	Wind	13.52	61.51	77.38	72.67	126.97	189.44
Capacity	Natural gas	1392.15	545.88	277.87	207.52	318.93	243.71
State-level	Solar	0.00	1.96	33.32	69.06	96.70	121.59
Approval Pending	Wind	0.00	18.30	34.12	103.81	148.50	162.23
Capacity	Natural gas	114.27	56.08	157.24	213.69	379.36	171.24
State-level	Solar	0.01	0.76	13.95	66.80	91.84	171.02
Capacity Under	Wind	9.37	79.41	96.69	99.51	123.83	191.57
Ontime Construction	Natural gas	387.17	227.41	242.23	316.52	525.57	395.13
State-level	Solar	0.00	0.01	13.00	33.59	42.13	63.78
Capacity Under	Wind	2.16	20.51	41.02	35.50	113.07	125.27
Delayed Construction	Natural gas	471.03	229.72	216.40	98.65	180.80	251.81
State-level	Solar	0.00	0.00	13.85	42.54	32.42	19.13
Postponed	Wind	0.02	5.08	20.74	46.91	65.64	62.92
WIP Capacity	Natural gas	2793.20	3735.72	2969.45	996.73	1012.31	827.68
State-level	Solar	0.00	0.00	0.00	0.00	2.15	17.16
Canceled	Wind	0.00	0.00	0.00	0.00	5.78	61.67
WIP Capacity	Natural gas	0.00	0.00	0.00	0.00	27.31	350.64
State-level	Solar	0.00	2.19	21.84	54.13	69.87	118.56
WIP Index	Wind	5.22	43.85	61.47	75.61	120.38	156.26
	Natural gas	670.52	544.78	470.90	290.33	430.42	340.89

Note: In this table we capture 3-year mean of the variables that represent investment stages across all states. All values are measure in Megawatts. Note that these variables does not have the flexibility level identifier for natural gas fueled power plants and these variables involve all kinds of flexibility levels. Proposed capacity shows total capacity planned for a state that must go through the approval process upon which the construction may begin. Approval pending capacity represents the total amount of capacity that is awaiting legislative approvals. After construction begins for a proposed power plant, depending on it being constructed on schedule or with some deviations from the original proposed time frame, we may have generation capacity under ontime construction or with delays. Postponed capacity shows the total amount of generation capacity that is indefinitely postponed with the possibility of getting back into a state's planning in the future, while canceled capacity will not get back into a state's planning. Work-in-progress (WIP) index variables are the weighted average of the WIP variables in each generation technology group that is used to reduce the dimensionality.

Figure 1.A1: State-level investment distribution in wind, solar, and natural gas-fueled technologies



Note: In this table we represent the investments in three different flexibility levels of the natural gas-fueled generators. Investments are measured in the amount of added capacity (MWs) to the power grid including newly developed power plants and additions to the capacity of the existing power plants. Note that the District of Columbia is omitted.

Appendix B Instrumental Variable Regression Results

Here we provide detailed tables for the first stage estimations as outlined in Section 1.5. The following tables are for the RPS requirement first stage linear regression estimation and the current investment correction models of wind, solar, and natural gas-fueled technologies.

Table 1.B1: RPS obligation correction model results

		RPS Obligations
Main IVs	Solar prior investment	0.79*
		(0.09)
	Wind prior investment	0.18*
		(0.06)
Controls	Energy intensity	Y
	Regional-state interactions	Y
	Residential price	Y
	Solar radiation	Y
	Time trend	Y
	Utility ownership	Y
	Wind speed	Y
	Wind WIP index	Y
Statistics	Observations(K)	137.94
	R-square	0.51

Note: In this table we use RPS obligations set by state legislators on an annual basis as the dependent variable for both states. We provide a list of control variables and interaction terms in Table 1.9 and 1.10 respectively. We estimate this model using observations for regions that adopted the RPS program and then use the estimated model to find the predicted RPS obligation values for all regions and subsequently we calculate the residuals that we use in later estimation models as we explain in Section 1.5.3.1. Significance levels: $^{*}p < 0.01$, $^{\dagger}p < 0.05$.

Table 1.B2: Current investment correction model results across k-means clusters

	Wind Investments		Solar Investments		NG-Flexible Investments	
	Intercept	R-square(%)	Intercept	R-square(%)	Intercept	R-square(%)
Cluster 1	−2.65*	0.42	−0.25	0.61	2.00*	0.11
Cluster 2	−2.14*	0.77	−26.11*	0.79	5.01*	0.31
Cluster 3	3.02*	0.35	−26.88*	0.84	16.90*	0.65
Cluster 4	−7.44*	0.59	106.81*	0.68	−1.28	0.30
Cluster 5	20.40*	0.61	0.47	0.88	19.16*	0.52
Cluster 6	−8.19*	0.77	−44.47*	0.58	3.63*	0.35
Cluster 7	5.12*	0.68	1.54	0.85	11.73*	0.10
Cluster 8	11.32*	0.80	−85.58*	0.74	15.14*	0.22
Cluster 9	0.38*	0.42	−36.64*	0.89	1.02	0.56
Cluster 10	−15.50*	0.54	−68.72*	0.78	28.35*	0.23
Cluster 11	20.70*	0.61	−1128.84*	0.83	79.60*	0.47
Cluster 12	7.30*	0.68	−95.66*	0.95	65.27*	0.64
Cluster 13	0.18	0.27	−3.06	0.68	14.56*	0.28
Cluster 14	−10.44*	0.79	−25.23*	0.77	−4.15†	0.19
Cluster 15	4.27*	0.67	−42.70*	0.71	−60.96*	0.54
Cluster 16	−5.93†	0.57	−23.27*	0.78	1.30	0.18

Note: In this table we represent the results of the first stage regression model on wind, solar, and natural gas-fueled investments using instrumental variables. We represent a list of control variables and interaction terms in Table 1.9 and 1.10 respectively. We use this model estimated parameters to develop a new set of investment variables (i.e., expected investment) for each generation technology. Significance levels: $^{*}p < 0.01$, $^{\dagger}p < 0.05$.

Appendix C Extended Analyses Results

Table 1.C1: Grouped fixed effect (GFE) estimation results of current and expected investment effects

	Current Investment		Expected Investment	
	Coefficient	R-square	Coefficient	R-square
H1: Wind $\sim$ Flexible	-0.029^{*}	55.02	-0.134^{*}	55.58
H2: Solar $\sim$ Flexible	0.003	52.20	-0.032^{*}	52.25
H3: NG-Flexible $\sim$ Wind	0.108^{*}	25.36	0.069^{*}	25.11
H4: NG-Flexible $\sim$ Solar	-0.090^{*}	25.02	-0.134^{*}	25.18

Note: In this table we show the coefficients and r-squares of each linear regression model that is run with k-means clusters fixed-effects for both current and expected investments as independent variables. Significance level: $^{*}p < 0.01$.

Bibliography

Aflaki S, Netessine S (2017) Strategic investment in renewable energy sources: The effect of supply intermittency. *Manufacturing & Service Operations Management* 19(3):489–507.

Akkas A, Gaur V, Simchi-Levi D (2019) Drivers of product expiration in consumer packaged goods retailing. *Management Science* 65(5):2179–2195.

Al-Gwaiz M, Chao X, Wu OQ (2017) Understanding how generation flexibility and renewable energy affect power market competition. *Manufacturing & Service Operations Management* 19(1):114–131.

Alizamir S, Iravani F, Yücel Ş (2021) Investment in wind energy: The role of subsidies. *Georgetown McDonough School of Business Research Paper* (3868573).

Ambec S, Crampes C (2012) Electricity provision with intermittent sources of energy. *Resource and Energy Economics* 34(3):319–336.

Angelus A (2021) Distributed renewable power generation and implications for capacity investment and electricity prices. *Production and Operations Management* 30(12):4614–4634.

Anvari M, Proedrou E, Schäfer B, Beck C, Kantz H, Timme M (2022) Data-driven load profiles and the dynamics of residential electricity consumption. *Nature communications* 13(1):4593.

Babich V, Lobel R, Yücel Ş (2020) Promoting solar panel investments: Feed-in-tariff vs. tax-rebate policies. *Manufacturing & Service Operations Management* 22(6):1148–1164.

Baranes E, Jacqmin J, Poudou JC (2017) Non-renewable and intermittent renewable energy sources: Friends and foes? *Energy Policy* 111:58–67.

Barbose GL (2021) Us renewables portfolio standards 2021 status update: Early release .

Bird L, Milligan M, Lew D (2013) Integrating variable renewable energy: Challenges and solutions. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States).

Bistline JE (2018) Turn down for what? the economic value of operational flexibility in electricity markets. *IEEE Transactions on Power Systems* 34(1):527–534.

Bonhomme S, Lamadon T, Manresa E (2022) Discretizing unobserved heterogeneity. *Econometrica* 90(2):625–643.

Bonhomme S, Manresa E (2015) Grouped patterns of heterogeneity in panel data. *Econometrica* 83(3):1147–1184.

Borenstein S, Kellogg R (2023) Carbon pricing, clean electricity standards, and clean electricity subsidies on the path to zero emissions. *Environmental and Energy Policy and the Economy* 4(1):125–176.

Bourcet C (2020) Empirical determinants of renewable energy deployment: A systematic literature review. *Energy Economics* 85:104563.

Bowen E, Lacombe DJ (2017) Spatial dependence in state renewable policy: Effects of renewable portfolio standards on renewable generation within nerc regions. *The Energy Journal* 38(3):177–194.

Bray RL, Serpa JC, Colak A (2019) Supply chain proximity and product quality. *Management science* 65(9):4079–4099.

Brehm P (2019) Natural gas prices, electric generation investment, and greenhouse gas emissions. *Resource and Energy Economics* 58:101106.

Bushnell J (2010) Building blocks: Investment in renewable and non-renewable technologies. *Harnessing renewable energy in electric power systems: theory, practice, policy* 159.

Bushnell J, Novan K (2021) Setting with the sun: The impacts of renewable energy on conventional generation. *Journal of the Association of Environmental and Resource Economists* 8(4):759–796.

Carley S, Davies LL, Spence DB, Zirogiannis N (2018) Empirical evaluation of the stringency and design of renewable portfolio standards. *Nature Energy* 3(9):754–763.

Comstock O (2024) About 25% of u.s. power plants can start up within an hour. *U.S. Energy Information Administration (EIA)* Principal contributor: Owen Comstock.

Davenport C, Singer B, Mehta N, Lee B, Mackay J, Corbett B, Ritchie J, Jaiya J, Venugopal V, Cash N, Halferty O, Miller J, Delaney M, Modak A, Hari T, Revich J (2024) Generational growth ai, data centers and the coming us power demand surge. Equity research report, Goldman Sachs.

De Toni A, Tonchia S (1998) Manufacturing flexibility: a literature review. *International journal of production research* 36(6):1587–1617.

Deschenes O, Malloy C, McDonald G (2023) Causal effects of renewable portfolio standards on renewable investments and generation: The role of heterogeneity and dynamics. *Resource and Energy Economics* 75:101393.

Devlin J, Li K, Higgins P, Foley A (2017) Gas generation and wind power: A review of unlikely allies in the united kingdom and ireland. *Renewable and Sustainable Energy Reviews* 70:757–768.

Dhakal S, Minx J, Toth F, Abdel-Aziz A, Figueroa Meza M, Hubacek K, Jonckheere I, Kim Y, Nemet G, Pachauri S, et al. (2022) Emissions trends and drivers (chapter 2) .

Ding C, He X (2004) K-means clustering via principal component analysis. *Proceedings of the twenty-first international conference on Machine learning*, 29.

Dive U (2023) Electricity load growing twice as fast as expected: Grid strategies report. *Utility Dive* URL `https://www.utilitydive.com/news/electricity-load-growing-twice-as-fast-as-expected-Grid-Strategies-report/702366/`.

Egli F (2020) Renewable energy investment risk: An investigation of changes over time and the underlying drivers. *Energy Policy* 140:111428.

EIA (2023) Solar and wind power curtailments are rising in california. `https://www.eia.gov/todayinenergy/detail.php?id=60822`, (Accessed on 02/08/2023).

EPA EPA (2024) Inventory of u.s. greenhouse gas emissions and sinks: 1990-2022. URL `https://www.epa.gov/ghgemissions/`

inventory-us-greenhouse-gas-emissions-andsinks-1990-2022.

Feldman R, Levinson A (2023) Renewable portfolio standards. *The Energy Journal* 44(5):1–20.

Gimeno J, Woo CY (1999) Multimarket contact, economies of scope, and firm performance. *Academy of Management Journal* 42(3):239–259.

Glenk G, Reichelstein S (2022) The economic dynamics of competing power generation sources. *Renewable and Sustainable Energy Reviews* 168:112758.

Gohdes N, Simshauser P, Wilson C (2022) Renewable entry costs, project finance and the role of revenue quality in australia's national electricity market. *Energy Economics* 114:106312.

Golden R, Paulos B (2015) Curtailment of renewable energy in california and beyond. *The Electricity Journal* 28(6):36–50.

Guerra K, Haro P, Gutiérrez R, Gómez-Barea A (2022) Facing the high share of variable renewable energy in the power system: Flexibility and stability requirements. *Applied Energy* 310:118561.

Hannan MT, Freeman J (1977) The population ecology of organizations. *American journal of sociology* 82(5):929–964.

Haveman HA (1993) Follow the leader: Mimetic isomorphism and entry into new markets. *Administrative science quarterly* 593–627.

Henni S, Staudt P, Kandiah B, Weinhardt C (2021) Infrastructural coupling of the electricity and gas distribution grid to reduce renewable energy curtailment. *Applied Energy* 288:116597.

Hu S, Souza GC, Ferguson ME, Wang W (2015) Capacity investment in renewable energy technology with supply intermittency: Data granularity matters! *Manufacturing & Service Operations Management* 17(4):480–494.

IEA (2021) *World Energy Outlook 2021.*

IEA IEA (2024) Electricity report 2024 analysis and forecast to 2026. URL https://iea.blob.core.windows.net/assets/ddd078a8-422b-44a9-a668-52355f24133b/Electricity2024-Analysisandforecastto2026.

Jack EP, Raturi AS (2003) Measuring and comparing volume flexibility in the capital goods industry. *Production and operations management* 12(4):480–501.

Jayadev G, Leibowicz BD, Kutanoglu E (2020) Us electricity infrastructure of the future: Generation and transmission pathways through 2050. *Applied energy* 260:114267.

Joshi J (2021) Do renewable portfolio standards increase renewable energy capacity? evidence from the united states. *Journal of Environmental Management* 287:112261.

Kapoor R, Furr NR (2015) Complementarities and competition: Unpacking the drivers of entrants' technology choices in the solar photovoltaic industry. *Strategic Management Journal* 36(3):416–436.

Kaps C, Marinesi S, Netessine S (2023) When should the off-grid sun shine at night? optimum renewable generation and energy storage investments. *Management Science* 69(12):7633–7650.

Khorramfar R, Mallapragada D, Amin S (2024) Electric-gas infrastructure planning for deep decarbonization of energy systems. *Applied Energy* 354:122176.

Kök AG, Shang K, Yücel Ş (2018) Impact of electricity pricing policies on renewable energy investments and carbon emissions. *Management Science* 64(1):131–148.

Kök AG, Shang K, Yücel Ş (2020) Investments in renewable and conventional energy: The role of operational flexibility. *Manufacturing & Service Operations Management* 22(5):925–941.

Koltsaklis NE, Dagoumas AS, Panapakidis IP (2017) Impact of the penetration of renewables on flexibility needs. *Energy Policy* 109:360–369.

Kondziella H, Bruckner T (2016) Flexibility requirements of renewable energy based electricity systems–a review of research results and methodologies. *Renewable and Sustainable Energy Reviews* 53:10–22.

Kotchen M (2012) Cheap gas is a trap. *The New York Times* .

Lee A, Zinaman O, Logan J, Bazilian M, Arent D, Newmark RL (2012) Interactions, complementarities and tensions at the nexus of natural gas and renewable energy. *The Electricity Journal* 25(10):38–48.

Lew D, Brinkman G (2013) The western wind and solar integration study phase 2 (executive summary). Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States).

Llobet G, Padilla J (2018) Conventional power plants in liberalized electricity markets with renewable entry. *The Energy Journal* 39(3).

Lund PD, Lindgren J, Mikkola J, Salpakari J (2015) Review of energy system flexibility measures to enable high levels of variable renewable electricity. *Renewable and sustainable energy reviews* 45:785–807.

Lyon TP (2016) Drivers and impacts of renewable portfolio standards. *Annual Review of Resource Economics* 8:141–155.

Marques AC, Fuinhas JA, Manso JP (2010) Motivations driving renewable energy in european countries: A panel data approach. *Energy policy* 38(11):6877–6885.

Milligan M, Frew B, Zhou E, Arent DJ (2015) Advancing system flexibility for high penetration renewable integration (chinese translation). Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States).

Mullen JD, Dong L (2022) Effects of state and federal policy on renewable electricity generation capacity in the united states. *Energy Economics* 105:105764.

Myles Shaver J, Flyer F (2000) Agglomeration economies, firm heterogeneity, and foreign direct investment in the united states. *Strategic management journal* 21(12):1175–1193.

Neetzow P (2021) The effects of power system flexibility on the efficient transition to renewable generation. *Applied Energy* 283:116278.

Newbery DM (2023) High renewable electricity penetration: Marginal curtailment and market failure under "subsidy-free" entry. *Energy Economics* 126:107011.

Nyangon J, Byrne J (2023) Estimating the impacts of natural gas power generation growth on solar electricity development: Pjm's evolving resource mix and ramping capability. *Wiley Interdisciplinary Reviews: Energy and Environment* 12(1):e454.

Olsen KP, Zong Y, You S, Bindner H, Koivisto M, Gea-Bermúdez J (2020) Multi-timescale data-driven method identifying flexibility requirements for scenarios with high penetration of renewables. *Applied Energy* 264:114702.

Papaefthymiou G, Dragoon K (2016) Towards 100% renewable energy systems: Uncapping power system flexibility. *Energy Policy* 92:69–82.

Parker GG, Tan B, Kazan O (2019) Electric power industry: Operational and public policy challenges and opportunities. *Production and Operations Management* 28(11):2738–2777.

Peng X, Wu OQ, Souza G (2021) Renewable, flexible, and storage capacities: Friends or foes? *Available at SSRN 3983678* .

Penn I (2020) The next energy battle: renewables vs. natural gas. *The New York Times* .

Post TW (2024) Ai data centers are power hungry. here's why that's a problem. *The Washington Post* URL `https://www.washingtonpost.com/business/2024/03/07/ai-data-centers-power/`.

Rai A, Nunn O (2020) Is there a value for "dispatchability" in the nem? yes. *The Electricity Journal* 33(3):106712.

Ren G, Wan J, Liu J, Yu D, Söder L (2018) Analysis of wind power intermittency based on historical wind power data. *Energy* 150:482–492.

Sethi AK, Sethi SP (1990) Flexibility in manufacturing: A survey. *International Journal of Flexible Manufacturing Systems* 2:289–328.

Shittu E, Weigelt C (2022) When the wind blows: Incumbents' sourcing strategies for wind power. *IEEE Transactions on Engineering Management* .

Upton DM (1994) The management of manufacturing flexibility. *California management review* 36(2):72–89.

Verdolini E, Vona F, Popp D (2018) Bridging the gap: Do fast-reacting fossil technologies facilitate renewable energy diffusion? *Energy Policy* 116:242–256.

Vithayasrichareon P, Riesz J, MacGill I (2017) Operational flexibility of future generation portfolios with high renewables. *Applied energy* 206:32–41.

Vokurka RJ, O'Leary-Kelly SW (2000) A review of empirical research on manufacturing flexibility. *Journal of operations management* 18(4):485–501.

Wang Q, Wu H, Florita AR, Martinez-Anido CB, Hodge BM (2016) The value of improved wind power forecasting: Grid flexibility quantification, ramp capability analysis, and impacts of electricity market operation timescales. *Applied Energy* 184:696–713.

Wu C, Zhang XP, Sterling M (2022) Solar power generation intermittency and aggregation. *Scientific Reports* 12(1):1363.

Zhang X, Che L, Shahidehpour M, Alabdulwahab A, Abusorrah A (2016) Electricity-natural gas operation planning with hourly demand response for deployment of flexible ramp. *IEEE Transactions on Sustainable Energy* 7(3):996–1004.

Zhou S, Solomon BD (2020) Do renewable portfolio standards in the united states stunt renewable electricity development beyond mandatory targets? *Energy Policy* 140:111377.

Zhu C, Fan R, Lin J (2020) The impact of renewable portfolio standard on retail electricity market: A system dynamics model of tripartite evolutionary game. *Energy Policy* 136:111072.