Processing Power: The Effect of Data Centers On Wholesale Electricity Markets*

Owen Kay[1], Robert Reaser[2], and Reid Taylor[1]

[1]Federal Reserve Bank of Dallas
[2]UC Davis

19th February, 2026

[Draft: PLEASE DO NOT CIRCULATE]

Abstract

Artificial-intelligence-driven data centers are reversing two decades of flat U.S. electricity demand. We quantify how this new, geographically concentrated, high-utilization load reshapes wholesale electricity markets using an hourly, unit-level economic dispatch model covering the continental United States. Comparing observed 2021–2025 market conditions to counterfactual scenarios without data center demand, we find that realized load growth has already increased wholesale prices by 3 to 5% on average nationwide, with substantially larger effects in regions hosting major data center corridors. Impacts depend critically on utilization assumptions. Extending the model through 2030, we show that if proposed construction proceeds under high-utilization scenarios, wholesale prices could rise dramatically (100%), while more moderate build-out yields smaller (20%) but still meaningful effects. Finally, we evaluate how optimally reallocating compute across regions would alter prices and emissions, highlighting the importance of siting and utilization decisions for electricity costs and decarbonization outcomes.

*The views expressed here should not be interpreted as reflecting the opinions of the Federal Reserve Board or of any other person associated with the Federal Reserve System.

1 Introduction

After nearly two decades of flat electricity consumption in the United States, load growth has re-emerged as a defining feature of power markets. A central driver of this reversal is the rapid expansion of data centers, fueled by recent advances in artificial intelligence (AI). Unlike traditional sources of demand growth, such as households, commercial buildings, or gradual electrification, data centers arrive in large, discrete increments, cluster geographically, and operate at high utilization rates. Their expansion therefore represents a qualitatively different form of load growth: concentrated rather than diffuse, persistent rather than transitory, and fast relative to the pace of generation and transmission investment.

These features matter because electricity markets are characterized by steep short-run supply curves. Generator outages, transmission constraints, fuel availability, and limited interconnection capacity imply that incremental demand is often met on sharply rising portions of the merit order. When supply is elastic, new load primarily increases production with modest price effects. When supply is tight, even moderate demand increases can produce large changes in generation costs, prices, and emissions. While these mechanisms are well documented in the context of weather shocks and outages, far less is known about how they operate in response to persistent, geographically concentrated demand growth such as that generated by data centers.

This paper quantifies how realized data center expansion has affected wholesale electricity market outcomes in the United States and uses those estimates to evaluate forward-looking scenarios of continued data center build out. We focus on three equilibrium objects: generation costs, wholesale prices, and carbon emissions. Our approach is structural. We construct hourly supply curves for conventional generators using unit-level data and estimates of capacity, outages, fuel costs, operating costs, and emissions. We then solve a least-cost dispatch model that clears regional electricity markets hour by hour. Combining this framework with detailed data on the location, timing, and capacity of data center deployments allows us to compare observed market outcomes to counterfactual scenarios in which data center load is removed.

The first part of the paper is retrospective and isolates the marginal contribution of data centers to market outcomes over the recent period of rapid expansion from 2021-2025. For each hour and region, we compute equilibrium dispatch and prices under observed conditions and under alternative "no–data center" counterfactuals. This comparison yields estimates of the additional generation costs, price impacts, and emissions attributable to data center demand, holding fixed contemporaneous supply conditions and inter-regional power flows. Because data center utilization is not directly observed, we evaluate a range of capacity and utilization scenarios, allowing impacts to vary with both the scale and temporal profile of computing load.

The second part of the paper is forward-looking and extends this framework to forecast outcomes through 2030. Using pre–AI-boom demand projections together with announced generation additions and retirements, we construct hourly demand and supply paths under alternative potential data center growth scenarios. These simulations characterize how future prices and emissions depend on the interaction between data center deployment, utilization patterns, and the pace of supply expansion. Importantly, the forecasting exercise is disciplined by the same dispatch framework used in the retrospective analysis, linking projected impacts directly to short-run market responses rather than engineering assumptions alone.

Our results highlight three main findings. First, realized data center growth has already increased generation costs and wholesale prices in regions facing binding supply constraints, with effects that are nonlinear in system load. Price impacts are largest when data center demand coincides with periods of scarcity or congestion, reflecting the convexity of short-run supply. Second, these price effects are accompanied by shifts in dispatch toward higher-cost and, in many cases, higher-emissions generation, implying that data center expansion has environmental consequences even absent changes in average emissions rates. Third, counterfactual scenarios indicate that future impacts are highly uncertain and effects could be substantially larger if projected data center capacity comes online faster than new generation or transmission.

We contribute to a growing literature on the economic and environmental consequences of electricity-intensive computing infrastructure. We are among the first, to our knowledge, to esti-

mate the electricity market outcomes from data centers. Bogmans et al. (2025) provide quantitative projections of the effect of AI-driven data center growth on electricity prices and carbon emissions at economy-wide scale using a computational DSGE model. Similarly, Wade et al. (2025) forecast the effect of rising demand from data centers and crypto mining in U.S. power markets, but lack the geographic and temporal data on data centers to account for the uncertainty in the build out. Recent work has begun to document localized wholesale price effects of data center expansion in major corridors (Mamkhezri et al. 2025; Feher et al. 2025), while related research on cryptocurrency mining, an extreme form of high-utilization computing load, finds increases in retail electricity prices with only partial fiscal offsets (Benetton et al. 2023). Other work by Gargano and Giacoletti (2025) studies the political economy of attracting data centers showing that state incentives meaningfully shift data center siting and investment, without commensurate increases in local tech employment.

More broadly, our analysis underscores that the consequences of demand growth depend not only on how much electricity is consumed, but on where and when that consumption occurs and on the availability of marginal supply. As data centers continue to expand, understanding these interactions will be essential for infrastructure planning, market design, and climate policy. By quantifying both realized impacts and plausible future trajectories, this paper provides an empirical foundation for those discussions.

2 Data

2.1 Electricity Supply Data

Our primary data source for electricity supply is the Environmental Protection Agency's (EPA) Continuous Emission Monitoring System (CEMS), which reports hourly electricity production and fuel use at nearly every conventional generator in the U.S.[1] CEMS reports generator

[1]CEMS comprises the universe of coal, natural gas, oil, and other combustible fuel generators with capacity over 25 MW, but not nuclear or renewable generators. Following Ham et al. (2025), we drop commercial, industrial, and cogeneration units whose main function is not selling into electricity markets.

gross electricity production, which we convert into net production using EIA-923 data to calculate plant-level ratios. For each generation unit, we observe characteristics including the technology and fuel type, precise geographic location, emissions of regulated air pollutants, and coverage by environmental regulations. We calculate maximum generator capacity as the 99th percentile of observed generation. Generators are grouped into regions based on EPA's eGRID subregions.

We closely follow Hausman (2025) and Ham et al. (2025) to construct hourly supply curves. Generator-specific marginal costs are calculated as the sum of fuel costs, variable operations and maintenance costs and pollution permit prices. Specifically, for generator i in hour t, the marginal cost is given by

$$MC_{it} = HeatRate_i * FuelCost_{it} + VOM_i + Permit_{it}. \quad (1)$$

The heat rate for generator i, $HeatRate_i$, is calculated as the ratio of total fuel consumption to annual net-generation.[2] For coal and coke-fired generators, we use data from the EIA on state specific monthly fuel costs paid by power plants. For natural gas and oil-fired generators, we use daily fuel prices (Henry Hub and West Texas Intermediate) and apply a state-by-month specific markup. We observe environmental trading program permit prices from the EPA and the Regional Greenhouse Gas Initiative (RGGI). Technology specific O&M costs are calibrated from the EIA based on Hausman (2025). Furthermore, power plants are often offline for either routine maintenance or unexpected outages. We follow Hausman (2025); Ham et al. (2025) and uniformly derate generator capacity by the planned outage rate and stochastically apply unplanned outages to individual hours. All dollar values are inflation adjusted using the total CPI index from the Federal Reserve Bank of St. Louis's FRED.

2.2 Data Centers

We obtain data on the characteristics of existing and planned data centers including the precise geographic location, peak power demand (MW), operational dates, capital cost, square footage, and site acreage from the Cleanview data center tracker. The data set covers all categories of data

[2] We censor anomalous heat rates below 4 mmBtu per MWh and above 50 mmBtu per MWh.

centers.

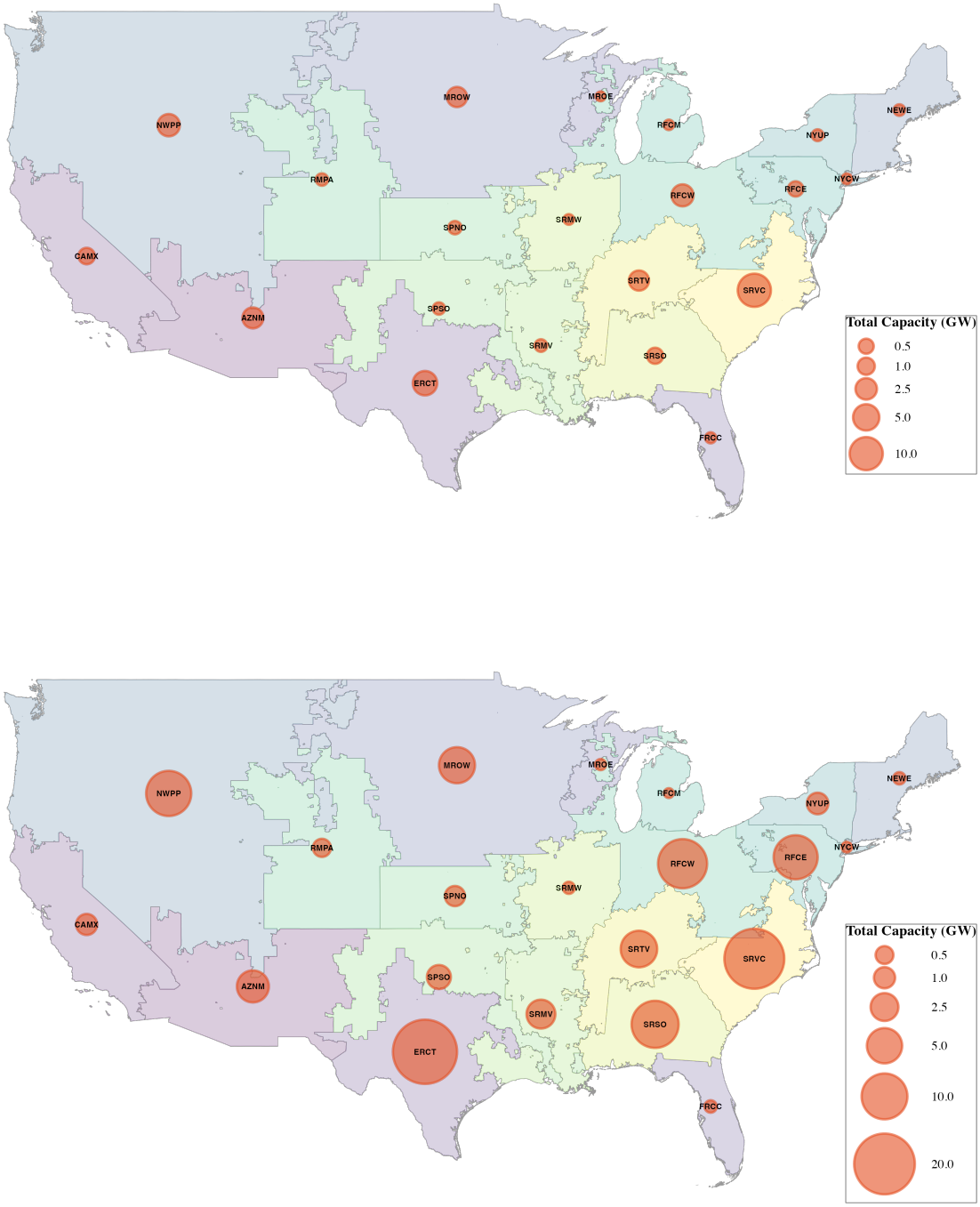
For our model of electricity markets, we require a measure of hourly data center power demand. There are, however, three sources of uncertainty in data center's power demand. First, not all data centers release public information on the size of their infrastructure. Approximately 67% of data centers in our data have a reported power capacity. For those with missing power data, we impute capacity using the year of first operation and site characteristics. Secondly, while the maximum power demand may be known, actual utilization is unknown. Although many data centers publicly state their intent for 100% utilization, industry reports show real world utilization is much lower and varies by the type of data center. This motivates our analysis to test the sensitivity of our results to various utilization assumptions. Lastly, some data centers have installed on-site"behind-the-meter' backup power generation and can produce their own electricity. Information on the size and utilization of "behind-the-meter" sources is not publicly reported. In our various utilization scenarios, we are agnostic as to whether the reduced utilization is due to lack of computing needs or a switch to on-site generation as they have the same impact on the electricity grid.

We capture this uncertainty by constructing four data center demand profiles which allow us to analyze the effect on prices across a variety of potential scenarios. We vary both the power capacity (the maximum power draw at a point in time) and the utilization rate (the percent of maximum used in a given hour). In the Full Capacity scenario, we assume 100% of installed power capacity is used 100% of the time. For the reasons noted above, this scenario is highly unlikely, but provides a useful upper bound. We leverage data from Bloomberg/DC Byte, which provides data on average observed utilization rates and market share by data center type. We scale data center capacity by these rates to construct two scenarios. First, we allocate power equal to the market share for training data centers to every hour as these have a 100% utilization rate. Next, we allocate the remaining portion according to the weighted average utilization rate across to other data center types, having data centers run the number of hours per day to hit the utilization rate every day. In our Mid-Range scenario, we randomly assign individual data center load across hours, either producing their full capacity or 0. In our Mid-Range-Peak scenario, data centers only

operate during business hours, operating at full capacity the required number of hours to meet their utilization rate, centered around 1PM. Lastly, our low scenario reduces the capacity, assuming 50% less capacity than used in the Mid-Range scenario.

While data centers are siting across the country, Figure 1 shows the geographic distribution of operational Data Center capacity in 2025, and operational and proposed data center capacity in 2030. To date, Virginia has led the way with the most operation data centers in the country. However, by 2030, Texas (and thus ERCOT) is slated to be the leader in data center power demand.

Figure 1: Installed and Planned Data Center Capacity: 2025 and 2030



Data centers consider a combination of factors when deciding where to locate. Electricity access is a primary consideration. Grid interconnection queues can often require projects to wait

years for access to connect, driving data centers to areas that will allow them to connect. While low electricity rates are a cost for data center, data centers do not appear to locate in grids with the lowest prices. Using data from the Census Bureau American Community Survey and the EIA, we report demographic variables for the counties where data center siting occurs in Table 1. Data centers locate along major fiber optic cable corridors and often near major urban areas to reduce latency. This can be seen when looking at the differenes between counties with and without data centers. As noted before, counties with data centers have marginally higher electricity rates than areas without.

Table 1: County Characteristics: Data Center vs. Non-Data-Center Counties

	DC counties	Non-DC counties	Difference
Population	508,698	68,056	440,642***
Median household income (USD)	81,069	64,439	16,630***
Poverty rate (%)	12.06	14.48	-2.42***
Unemployment rate (%)	4.70	4.69	0.01
Bachelor's degree or more (%)	33.88	23.01	10.86***
Non-Hispanic White share (%)	62.41	75.26	-12.85***
Median age	38.38	42.01	-3.63***
Industrial electricity price (cents/kWh)	8.07	7.70	0.37*
N counties	271	2751	

Notes: DC counties are counties with at least one non-cancelled data center in the Cleanview sample. Demographics are from ACS 2019–2023 5-year estimates. Industrial electricity price is from EIA-861 and is constructed at the county level using utility-state industrial revenue and sales mapped to counties via service territories, then averaged over 2019–2023. Difference reports mean(DC) - mean(Non-DC). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3 Structural Dispatch Model

To simulate the wholesale market clearing process, whether for our ex post analysis or forecast period, we use our data on marginal costs, de-rated capacities, and residual net-load (demand) for conventional generators to solve a linear program which minimizes the total cost of meeting net

load in each subregion subject to available generation capacity. Formally, we write:

$$\min_{q_{it}} \quad \left\{ \sum_{i \in I_r} mc_{it} q_{it} \right\} \tag{2}$$

$$\text{s.t.} \quad \sum_{i \in I_r} q_{it} \geq d_{rt} \quad \forall r \tag{3}$$

$$0 \leq q_{it} \leq K_{it} \quad \forall i \in I_r. \tag{4}$$

For each region r, let $i \in I_r$ denote a generator in the set of total units in r. Then, for each hour t and region r, we solve for the output q_{it}, for all $i \in I_r$, that minimizes total generation costs, where mc_{it} denotes the marginal cost of unit i in hour t. Crucially, total output for conventional generators in each region, $\sum_{i \in I_r} q_{it}$, must exactly equal residual demand for conventional generators, d_{rt} and generator output is subject to a capacity constraint, where K_{it} represents generator i's derated maximum capacity in hour t. Generator capacity varies at the hourly level in our model because although we uniformly derate all capacity by month to account for expected outages, we apply a stochastic outage component at the hourly level to accont for unexpected outages. Thus, there is a non-zero probability in every hour that $K_{it} = 0$.

Solving the model produces two crucial outputs. We obtain the optimal dispatch quantities for each generator, q^*_{it}, and a regional price, λ^*_{rt}. The regional price is equal to the shadow price of the power balance constraint, given by Equation 3 above. This represents the marginal cost of serving an additional MWh of net load. We interpret λ^*_{rt} as a competitive benchmark for regional wholesale electricity prices since we rely on generator marginal costs –rather than strategic bidding– and omit non-convex physical constraints such as start-up costs and ramp-rates. In Appendix **??** we provide evidence that our simulated prices reasonably approximate their observable analog in the ERCOT region.[3] Finally, it is important to note that much of the western and southeastern US do not have wholesale electricity markets and therefore no wholesale electricity prices. In these regions, which are served by vertically integrated utilities, λ^*_{rt} is still an object of interest and can still be

[3] Hausman (2025) presents similar evidence for the reasonableness of this assumption for the MISO and SPP regions.

interpreted as the marginal cost to the utility of serving an additional MWh of net load. Despite this distinction, we will refer to λ_{rt} as the regional price.

4 Ex Post Analysis

We first characterize how electricity demand from data centers has impacted wholesale electricity markets and generator dispatch since 2021. To do so, we solve the dispatch model described in Section 3 for each hour t, in each subregion r, under alternative assumptions of electricity demand driven by data centers.

Each of our alternatives is compared to the reference case where we set net load for conventional generation equal to observed thermal generation in each region. For the data center utilization scenarios, we subtract implied data center demand from observed net generation in each hour, to produce a no data center residual demand scenario for each counterfactual. We then re-solve the dispatch model for q_{it} and λ_{rt} without data center demand.

By setting our baseline residual demand for thermal generation equal to observed thermal generation in each region, we allow thermal generation to be re-optimized but hold other source of supply and demand fixed. Embedded in this modeling decision is an assumption that other sources of supply such as nuclear, wind, solar, and hydro are always inframarginal. This is consistent with very low marginal costs for these sources of supply.

In addition, we are implicitly fixing the net interchange between regions to their historic levels. Observed generation in each region is equal to regional electricity consumption plus net exports. For the counterfactual scenario that includes data center demand, this simply ensures that net exports between the various regions match observed net exports. In our no data center counterfactuals, we subtract the hourly assumed data center load from observed generation. Thus, we also fix transmission flows between regions to their observed levels, even in the counterfactual without data center demand. This implies that the marginal data center electron is always generated within the same region. In reality, if a neighboring region had cheaper generation available and

transmission lines between regions were not at capacity, we would expect incremental data center load to be partially met by the neighboring region.

These considerations motivate our use of eGRID subregions to define our geographic markets. EPA's methodology for assigning plants to eGRID regions is intended to limit imports and exports within an aggregated area in order to produce regional emissions rates that reflect the electricity consumption within that geographic area. To the extent that the eGRID subregion boundaries successfully limit bias introduced by imports and exports, our main results can be interpreted as the additional costs, emissions, and price impacts of data center load. Because we fix historic trade flows, our model does not capture inter-regional re-dispatch that would occur in a no data center counterfactual. In that respect, our cost estimates will understate the true impacts of data center load. The direct of the bias on emissions and local price impacts is ambiguous. Changes to emissions depend on the regional generation mix. Local price impacts would be biased up in cases where a region is a net-importer and biased down for net-exporters.

We then compute counterfactual dispatch under "no-data-center" counterfactual demand scenarios. To do so, we subtract data center electricity demand from the measure of total conventional generator electricity demand used in the base case. We only observe capacities for existing data centers, not utilization rates, so we calculate total data center electricity demand under a number of alternative assumptions of utilization rates described in section 2. For each"no-data-center" demand scenario, we solve the dispatch model for generator production levels, q_{it}^{-DC}, and wholesale prices, λ_{rt}^{-DC}, at each hour had there been no data center demand.

4.1 Additional Generation Costs

We first calculate the difference in total generation costs under the baseline and no data center counterfactuals, $\sum_i mc_{it} q_{it} - \sum_i mc_{it} q_{it}^{-DC}$. We compare our no data center counterfactual to our simulated baseline rather than observed generation to ensure we only capture differences in dispatch that result from removing data center load. [4]

[4]Generator specific production in our base case could differ from observed production for several reasons including our simulated outages not matching the timing of observed outages, the fact that we do not model ramping

Figure 2: Percentage Increase in Monthly Generation Costs
Attributed to Data Center Load, 2021–2025Q3

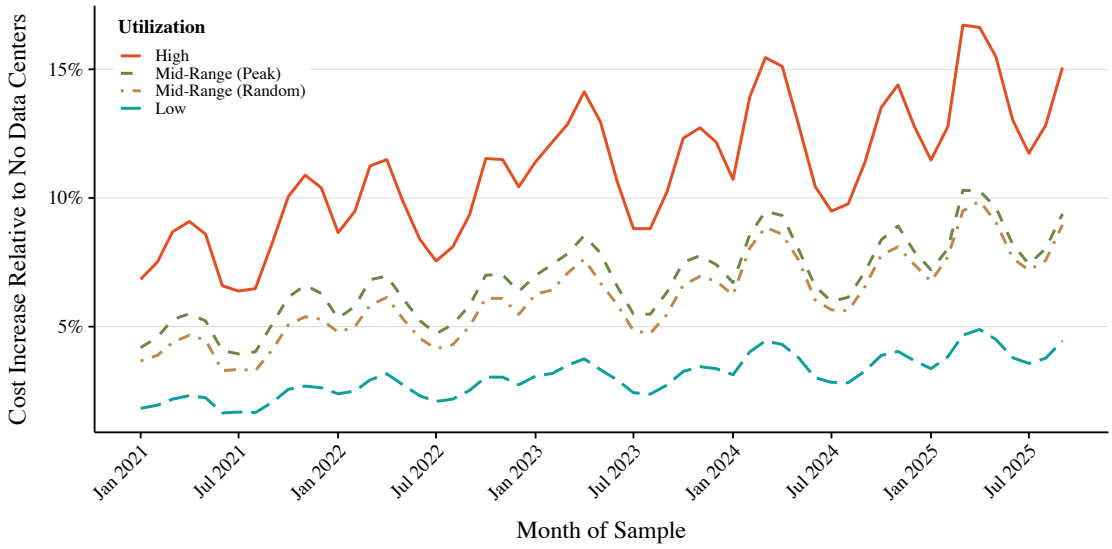


Figure 2 displays the percentage increase in monthly generation costs attributed to additional data center load, relative to a no data center counterfactuals. In our high-end case where 100% of operational data centers are operating 100% of the time, we estimate that by 2025, data centers had increased generation costs by 10 to 15% nationwide. We can bound our potential outcomes with the low-end case which shows that data centers added at most 5% additional cost of generation.

Focusing on the mid-range scenarios, which are calibrated to observed utilization rates, we estimate additional costs of 5% at the beginning of our sample, increasing to a peak of 10% in spring of 2025. Our scenario which concentrates DC demand in the peak business hours of the day results in an additional 1% of generation costs compared to the scenario where usage is spread throughout the day.

These results also reveal an important pattern. The cost increase in summer and winter months tends to be several percentage points *lower* than in the effects in the shoulder seasons. The spring and fall seasons align with the maintenance schedules of many plants as traditionally these are periods of lower demand due to milder temperatures and weather events. In these periods, the thermal supply curves tend to be more convex. As data center load expanded over our sample,

constraints, and imperfect modeling of the underlying transmission network.

markets were forced to clear at steeper portions of the supply curve relative to where they would have cleared in the absence of this constant load.

Our model allows us to decompose the welfare impacts of data centers. We calculate the total competitive wholesale payments made to generators, their total production costs, and the competitive, inframarginal rents earned by generators. Total payments are $\sum_{rt} \lambda_{rt} Q_{rt}$, where λ_{rt} is the competitive wholesale price of subregion r in hour of sample t, and $Q_{rt} = \sum_i q_{it}$ is total thermal net generation in hour t. Production costs are total variable costs (fuel, variable O&M, emissions permits) summed across all dispatched generators and hours. Competitive rents equal the difference between wholesale payments and production costs.

Table 2: Total Payments to Generators, Production Costs, and Competitive Rents With and without Data Centers in Millions of 2024 Dollars, 2021–2025

	Year					
	2021	2022	2023	2024	2025	Cumulative
Payments	104,662	146,553	81,491	73,834	64,605	471,145
Payments, No DC	98,828	136,508	75,253	67,740	58,426	436,754
Δ *Payments*	*5,834*	*10,045*	*6,239*	*6,094*	*6,179*	*34,391*
Production Costs	78,624	104,296	65,697	58,448	51,546	358,611
Production Costs, No DC	74,903	98,561	61,456	54,392	47,511	336,824
Δ *Production Costs*	*3,721*	*5,735*	*4,240*	*4,056*	*4,035*	*21,787*
Competitive Rents	26,037	42,257	15,795	15,386	13,059	112,534
Competitive Rents, No DC	23,924	37,946	13,796	13,348	10,914	99,930
Δ *Rents*	*2,113*	*4,310*	*1,998*	*2,038*	*2,144*	*12,604*

This table reflects the competitive wholesale payments, production costs, and inframarginal, competitive rents associated with the thermal generation in our sample under the observed and no data center counterfactuals. The main object of interest is the change, Δ, in each of these objects, driven by data centers, which is calculated by the difference in the observed and no data center cases. All values above use the Mid-Range Peak scenario and are in millions of 2024 dollars.

Table 2 presents these three objects computed from our observed dispatch and our no data center scenarios. In addition, we include the change in payments, production costs, and rents attributable to data centers.[5] Notably, we find a cumulative welfare transfer from wholesale consumers to suppliers of $12.6 billion from 2021 to 2025 that was driven by data center demand.

[5]Past work, such as Borenstein et al. (2002), similarly decompose these welfare objects. Because we only compute a competitive equilibrium in our setting, we make no claims about the market power effects of data centers.

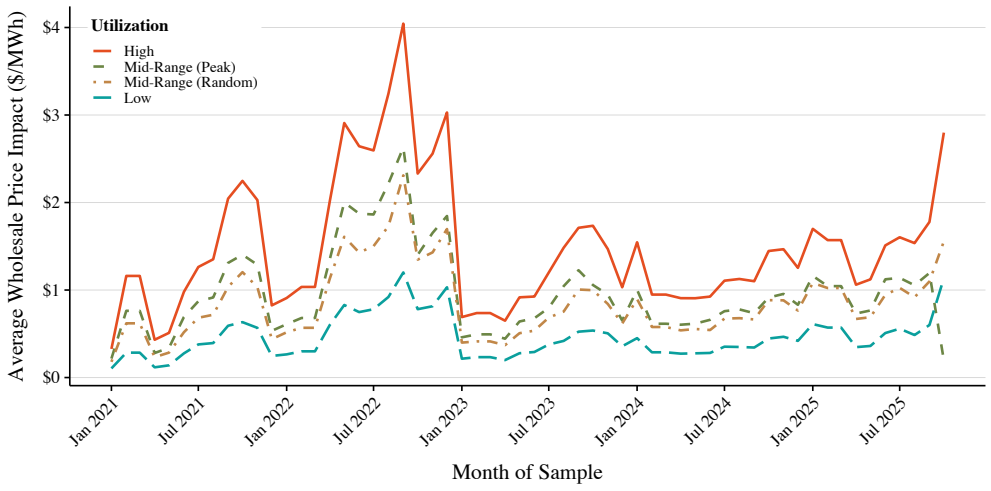
Total production costs totalled over $21 billion.

4.2 Competitive Wholesale Price Effects

While cost of generation is a relevant measure for the areas of the country covered by vertically integrated utilities, for the majority of the country covered by competitive wholesale electricity markets, the estimated market price is relevant both for generator profits and the price paid by utilities to procure electricity for retail customers.

Figure 3 contains our estimates of the nationwide load-weighted average monthly price effects of data centers under the same 4 data center demand scenarios. Similar to our characterization of additional costs, we use the difference between the baseline competitive wholesale price using observed net load and our counterfactual scenarios. Our estimates reveal price impacts from 2023 onward ranging from approximately $0.50 to just over $3 with our Mid-Range estimates falling in-between. By late 2025Q3, the marginal generator dispatched to meet electricity demand was approximately $1.50 to $3.00 more expensive, on average, than what would have been dispatched in a the counterfactual world without data center load.

Figure 3: Average Monthly Competitive Wholesale Price Impact, 2021–2025Q3



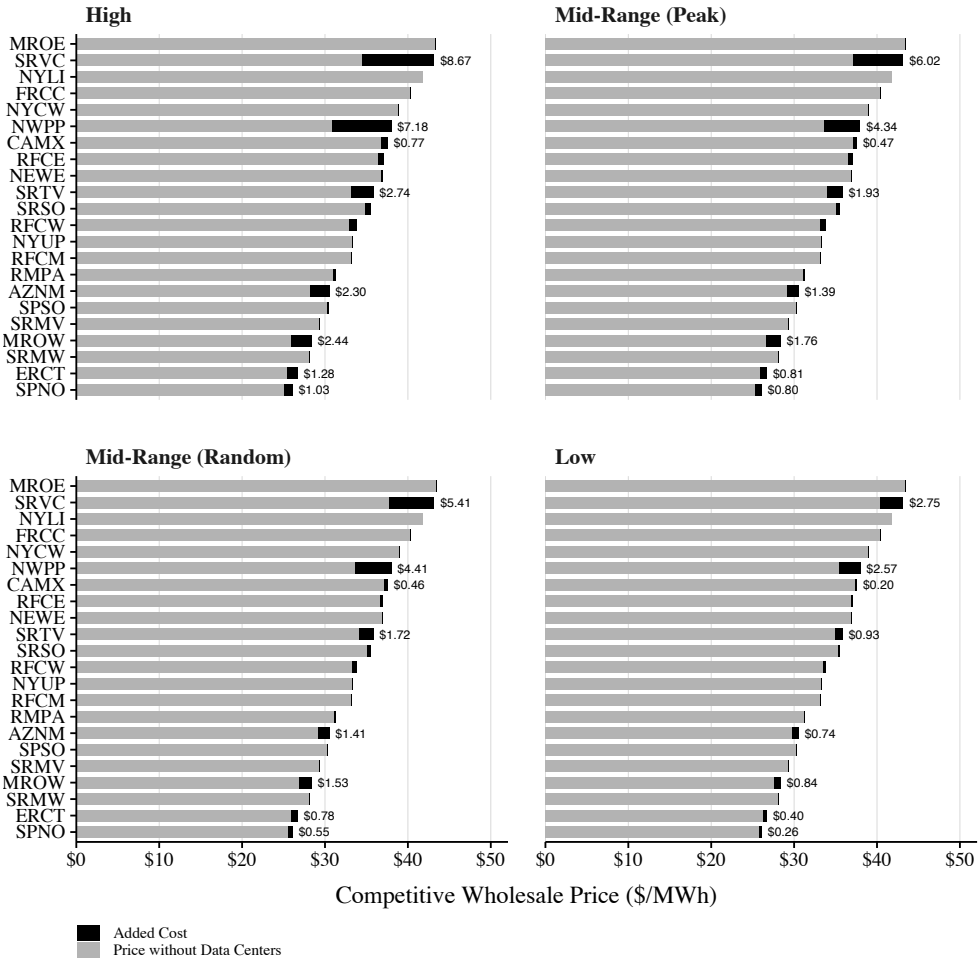
The highest estimated price effects occurred during the summer of 2022. Large spikes to natural gas prices caused by the Russian invasion of Ukraine increased wholesale electricity prices

as natural gas generators are often the marginal generator, and set the market price. Our estimates show that during this period, data center demand added an additional $1 to $4 to the nationwide wholesale price.

The distribution of data center load across the U.S. and the heterogeneity in regional power plant fuel composition means that there is substantial variation in results by region. In Figure 3 we show the average price impacts by EGRID subregion from 2023 to 2025Q3. The largest effects are seen in SRVC which is home to "Data Center Alley" in northern Virginia, the densest collection of data centers in the country. Northwest Power Pool (NWPP), which cover the majority of the pacific northwest also has significant effects from data centers. Our mid-range scenarios indicate data centers were responsible for a $5.41 to $6.02 increase in competitive wholesale prices in SRVC, a greater than 10% increase, and a $4.34 to $4.41 increase in NWPP.

ERCOT (ERCT), which is the grid covering the vast majority of Texas has very little price effect despite containing the second high amount of data centers behind SRVC. Data centers have added less than $1 to wholesale prices, likely driven by nation-leading supply build out. We also find that data centers had little impact in the subregion covering most of California, CAMX. While Californians face the highest average retail electricity prices in the nation during this period, even our upper-bound estimates yield an impact of less than $0.60.

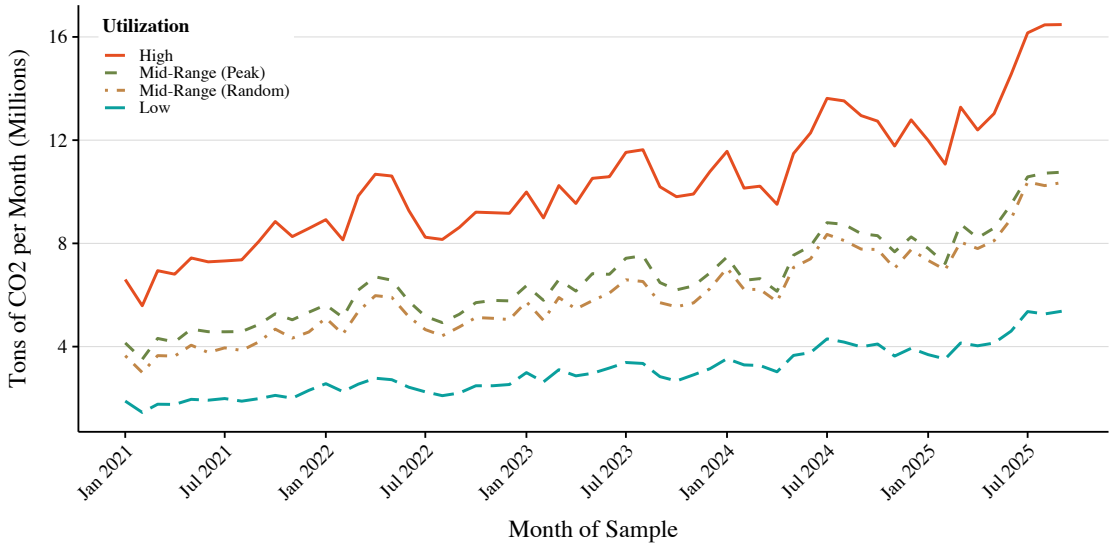
Figure 4: Average Regional Competitive Wholesale Price Impact, 2024-2025Q3



4.3 Emissions

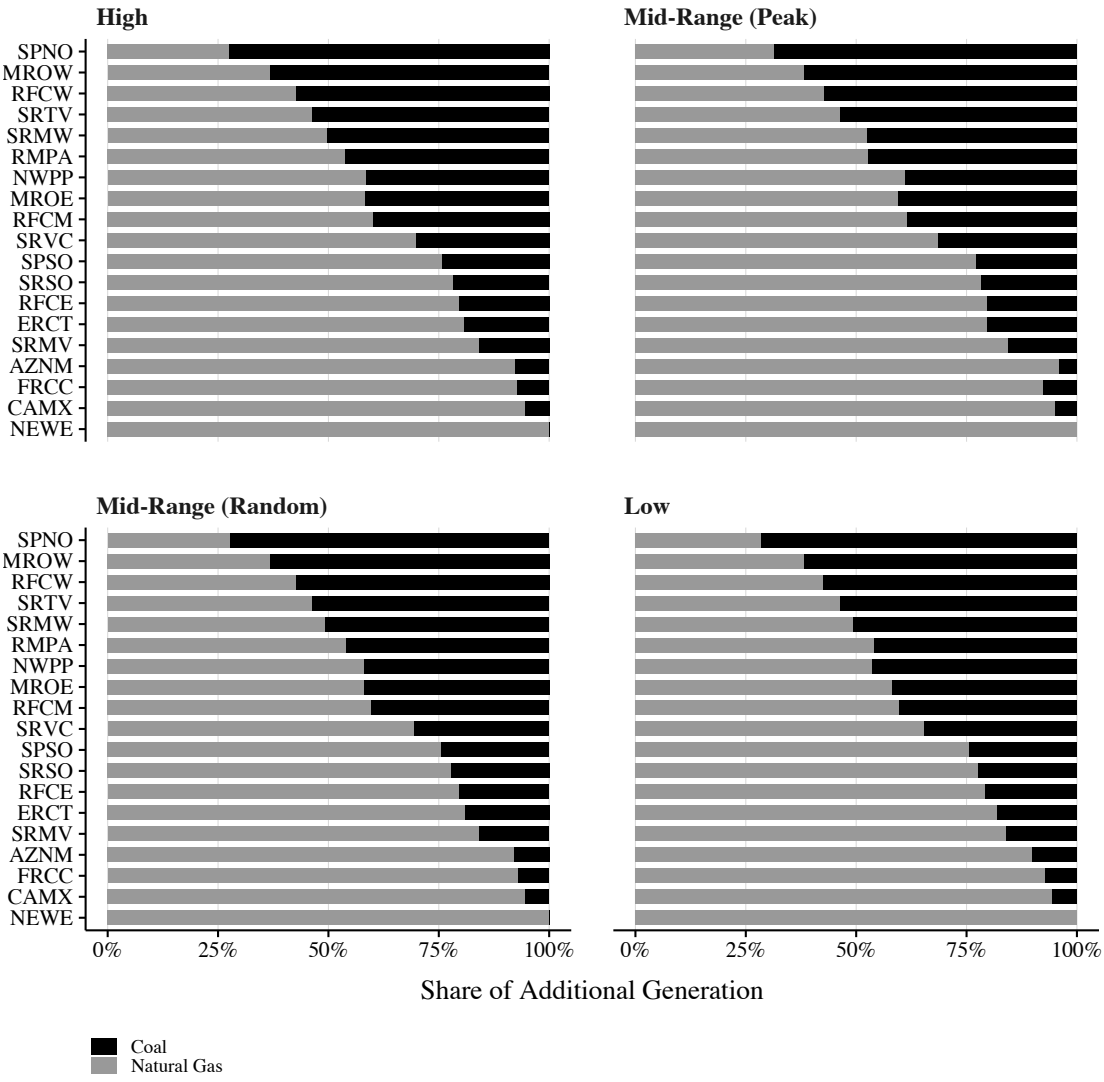
From our model results, we observe the additional output required to meet data center load. Using plant specific emissions rates, we can estimate the incremental emissions, measured as tons of CO_2, directly related to data center load growth. Our results, shown in Figure 5, suggest that incremental emissions have been steadily increasing over our sample period as data center load continues to increase, suggesting that data center demand growth has generally outpaced the rate at which new renewable generating capacity has connected to the grid.

Figure 5: Incremental Monthly CO_2 Emissions, 2021–2025Q3



By 2025, data centers were generating between 5 and 16 million additional tons of CO_2 per month. For reference, a moderate social cost of carbon of $252, results in damages ranging from $12 to $48 billion annually depending on the data center demand assumption. This must necessarily be weighed against the positive externalities and productivity gains created by data centers, an exercise which is a promising avenue for future research.

Figure 6: Share of Incremental Load Met by Coal vs. Natural Gas
By Region, 2024–2025Q3



5 Forecast

This section quantifies how projected data center expansion changes wholesale market outcomes through 2030. Two results emerge. First, effects are highly state-dependent: under the high-end assumption, data center load drastically raises generation costs and wholesale prices relative to the no-new-data-center benchmark. Second, the effects are geographically uneven, with larger price effects in regions that are both most exposed to new data center construction and face

supply constraints.

Unlike the ex-post exercise, the forecast requires constructing hourly net load from total forecast demand and non-thermal supply. For each subregion-hour, forecast net load equals projected total demand minus projected output from wind, solar, hydro, storage, and other non-fossil baseload resources. We then solve the same least-cost thermal dispatch problem used in Section 3. The forecast incorporates expected additions, retirements, derates/uprates, outages, and fuel-cost paths from EIA-based inputs.[6] All effects are reported relative to a no-data center counterfactual.

We evaluate the same four demand scenarios (`high`, `mid_peak`, `mid_rand`, `low`) that differ in effective utilization and intraday load shape. However, to reflect the additional uncertainty with data centers not completing construction or being 'behind-the-meter', we scale down the capacity used for the mid scenarios by 40% of announced. This in turn lowers the low-end case. The `mid_peak` scenario is our preferred central case with `high` and `low` bounding the effects.

5.1 Costs and Wholesale Prices

Figure 7 reports monthly percentage changes in generation cost, and Figure 8 reports monthly wholesale price effects in levels. The very large estimated effects in the high-end case reflect the explosive announced growth in data centers. If they all arrive as planned, and run continuously, generation costs are set to increase by 80% and prices over $40 higher than a no-data center counterfactual. While this scenario is highly improbable, it reflects the tail upside risk data centers pose to electricity prices. The mid scenarios hours imply more modest, yet still significant, 20-30% increases in generation costs by 2030 with substantial month-to-month variation. Interesting, there is little differentiation between the different utilization scenarios. Even in the low-end case, generation costs are estimated to grow by 10-15%, with prices rising by around 10% on average.

[6] Full details on demand and supply forecasts are detail in Appendix A

Figure 8: Forecasted Average Monthly Competitive Wholesale Price Impacts, 2026–2030

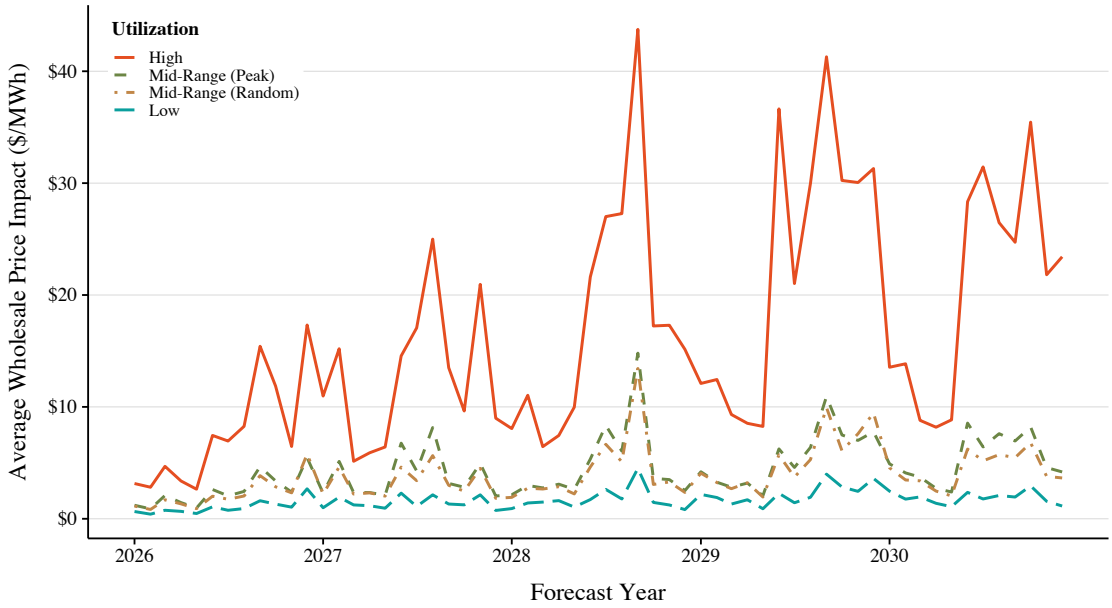
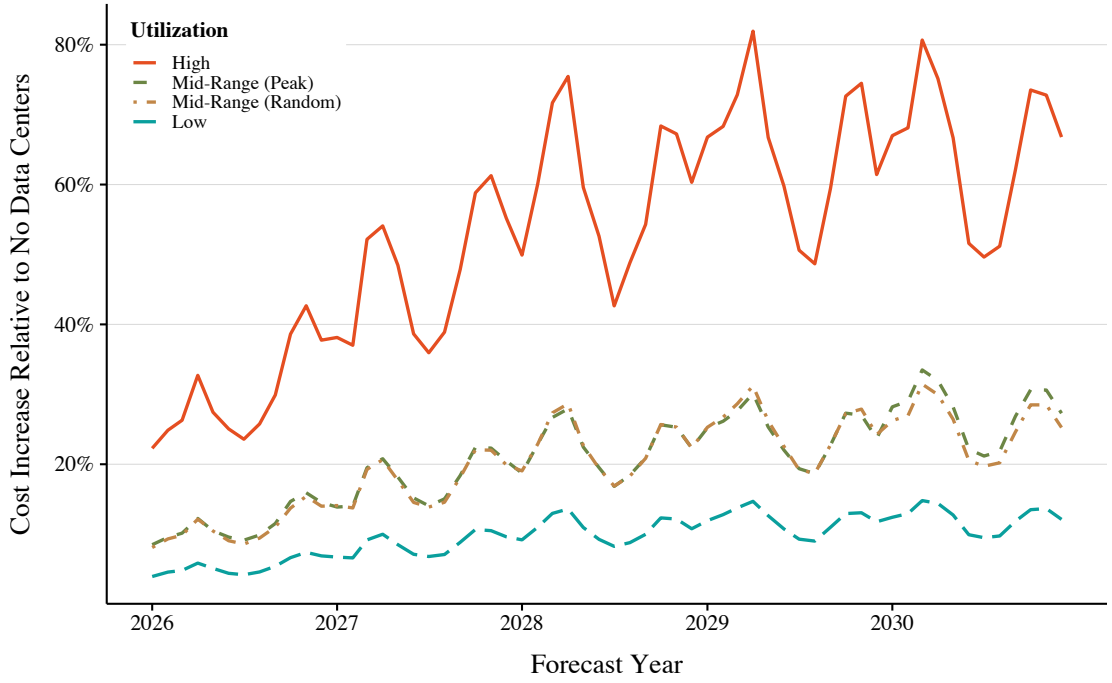


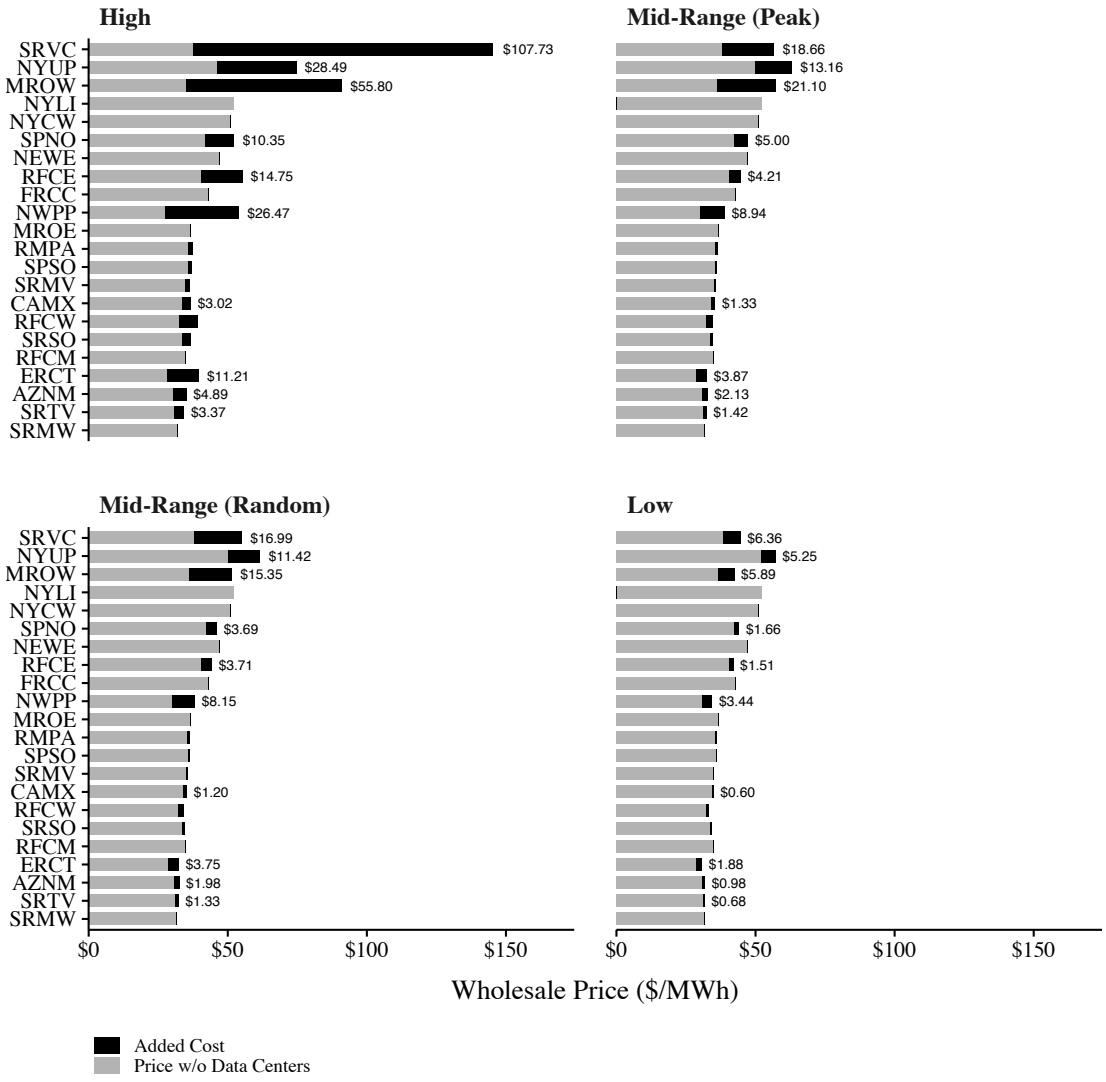
Figure 7: Forecasted Percentage Increase in Generation Costs, 2026–2030



5.2 Regional Incidence

Figure 9 the significant geographic concentration of effects in a subset of regions. This heterogeneity is consistent with differences in marginal thermal supply conditions and the geographic concentration of data center demand. Unlike the previous ex post analysis, there are significant effects in regions beyond SRVC. Slow interconnection queues in Northern Virginia have led data centers to explore new locations around the country, and will lead to wider spread impacts. The large effects still seen in SRVC reflect the tight supply constraint and the extreme convexity of marginal costs at the tail of the supply curve. Of note, MROW which constitutes much of the upper great plains has large forecast effects despite modest impacts from data centers to date. Lastly, ERCOT has modest forecast data center impacts despite leading the country in proposed data center expansion, again driven by high levels of new supply set to enter over the forecast period.

Figure 9: Forecasted Average Regional Competitive Wholesale Price Impact, 2029–2030

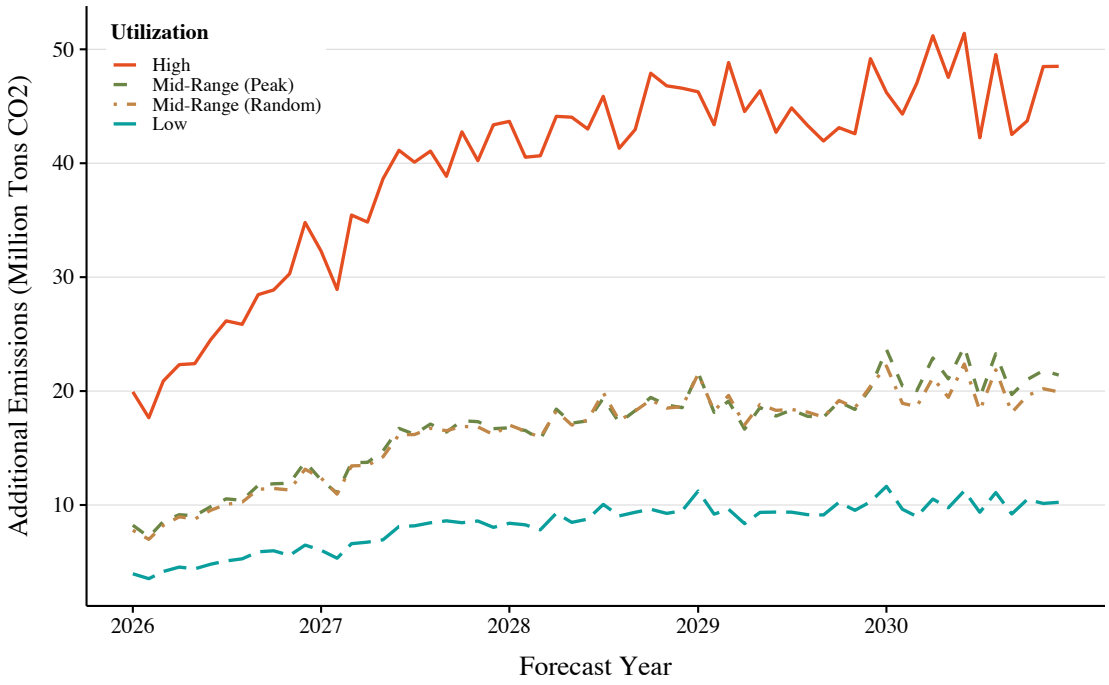


5.3 Emissions

Figure 10 reports incremental monthly CO_2 emissions from the required additional generation. In the high-demand scenario, data center growth greatly increases emissions, resulting in 45 to 50 million tons of CO_2 per month. In the mid scenario, emissions increases are still high, reaching 20 million tons of CO_2 per month by 2030. It is important to note we are measuring grid generated emissions. Part of our assumption in lowering the proposed capacity is due to data centers creating their own on-site generation 'behind-the-meter'. These sources of energy are often

more heavily polluting than on-grid assets, and would be additional to what is shown in Figure 10. The mid cases result in $60 billion dollars in additional damages.

Figure 10: Forecast Incremental Monthly CO_2 Emissions, 2026–2030

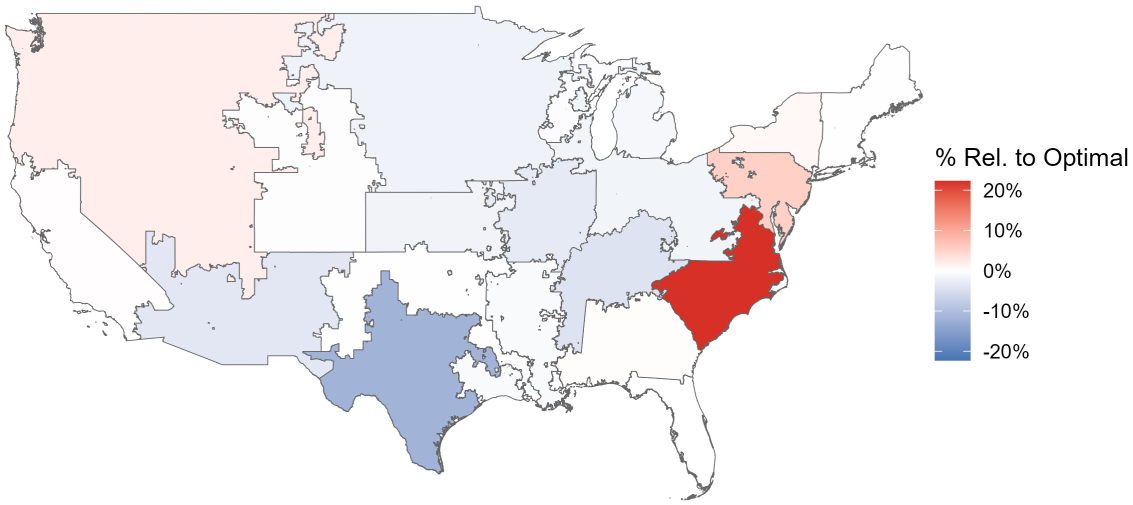


5.4 A Spatial Efficiency Benchmark

We finish our forecast analysis by benchmarking observed siting against a planner that allocates compute geographically to minimize short-run electricity operating costs. This counterfactual abstracts from latency, interconnection, construction, labor, and transmission constraints that may determine where data centers would prefer to locate. We construct this counterfactual by solving a national dispatch model (without regional constraints) using the excess thermal supply *not* called upon to operate in our no data center forecast scenario and the total amount of data center demand for each hour. The solution will choose the cheapest plants, nationwide, that are available, but idle, until the amount of additional generation meets the required demand from data centers. The resulting hourly unit-level dispatch quantities can can then be compared to the distribution of actual installed data centers across regions.

Figure 11 highlights large wedges between installed and cost-minimizing compute shares in the using our mid-peak demand scenario. For example, SRVC accounts for about 25.7% of installed compute but only about 3.4% under the least-cost allocation, while ERCOT rises from about 19% installed to about 31% under the benchmark. These wedges imply that non-energy siting constraints are first-order for realized geography and, by extension, for the distribution of wholesale price pressure.

Figure 11: Distribution of Compute Compared to Optimal



Taken together, the forecasts imply that data center expansion can create meaningful regional wholesale-price pressure in high-demand states of the world, while average national effects in central scenarios are more moderate, yet meaningful. Because impacts are concentrated in specific regions and high-price periods, policy design is likely to be more effective when targeted to local system conditions rather than uniform nationwide rules.

6 Conclusion

This paper provides one of the first empirical assessments of wholesale market impacts of expanding data center load in the United States. Our least-cost dispatch model covering the continental U.S. allows us to isolate the specific contribution of data center load on generator costs,

competitive wholesale electricity prices, and carbon dioxide emissions. We document a steady and significant rise in all three of these market outcomes.

Our results highlight the non-linearity of data center impacts. Their high utilization rates raise the floor of system demand, pushing the market equilibrium up increasingly steep portions of the supply curve. This effect is often most acute not during summer hours when demand often reaches its peak, but during shoulder seasons when more baseload generation is off-line for maintenance. As AI load continues to grow, the available slack in the system may diminish year-round, potentially forcing older fossil-fuel generation to remain online longer in a world without this load growth.

Relatedly, our forecast results illustrate the friction that exists between this load growth and the slow pace of infrastructure deployment. Because our model holds inter-regional transmission flows fixed, this dynamic is exacerbated in our forecasted outcomes. Therefore, the high prices observed in our forecasted scenarios should not be interpreted as a deterministic prediction, but rather as a potential signal of the shadow value of transmission.

References

Benetton, Matteo, Giovanni Compiani, and Adair Morse, "When Technology Processors Come to Town: High Electricity Use Spillovers to the Local Economy," Working Paper 31312, National Bureau of Economic Research June 2023. NBER Working Paper No. 31312.

Bogmans, Christian, Patricia Gomez-Gonzalez, Ganchimeg Ganpurev, Giovanni Melina, Andrea Pescatori, and Sneha Thube, "Power Hungry: How AI Will Drive Energy Demand," IMF Working Paper 2025/081, International Monetary Fund April 2025.

Borenstein, Severin, James B Bushnell, and Frank A Wolak, "Measuring market inefficiencies in California's restructured wholesale electricity market," *American Economic Review*, 2002, *92* (5), 1376–1405.

Bushnell, James, "A mixed complementarity model of hydrothermal electricity competition in the western United States," *Operations research*, 2003, *51* (1), 80–93.

Feher, Adam, Emilia Garcia-Appendini, and Roxana Mihet, "Is AI Trained on Public Money? Evidence from U.S. Data Centers," November 2025. Working paper.

Gargano, Antonio and Marco Giacoletti, "Subsidizing the Cloud: U.S. State Incentives to Data Centers," Working Paper 25-588, SSRN eLibrary 2025.

Ham, Dasom, Owen Kay, and Catherine Hausman, "Power Flows, Part 2: Transmission Lowers US Generation Costs, But Generator Incentives Are Not Aligned," Technical Report, Resources for the Future 2025.

Hausman, Catherine, "Power flows: Transmission lines, allocative efficiency, and corporate profits," *American Economic Review*, 2025, *115* (8), 2574–2615.

Mamkhezri, Jamal, Xiaochen Sun, and Yuting Yang, "The Hidden Cost of the Cloud: Data Centers and Electricity Market Inefficiency," Working Paper 25-660, USAEE / IAEE Working Paper Series 2025.

Wade, Cameron, Mike Blackhurst, Joe DeCarolis, Anderson de Queiroz, Jeremiah Johnson, and Paulina Jaramillo, "Electricity Grid Impacts of Rising Demand from Data Centers and Cryptocurrency Mining Operations," Report, Scott Institute for Energy Innovation, Carnegie Mellon University June 2025.

A Appendix: Forecast Model

In order to solve for market prices in future pieces, we must construct hourly demand and supply forecasts out through 2030. We detail their construction below.

A.1 Demand Forecast

To calculate a demand forecast through 2030, we combine data on current hourly production from EPA eGrid and EIA-930, and annual electricity consumption from the EIA Annual Electricity Outlook (AEO).

Our goal is to forecast electricity demand from all sources except data centers. We start with the 2021 observed annual net generation (TWh) by eGrid region as reported in the 2021 EPA eGrid model. As this was prior to the massive performance improvements in AI and subsequent buildout of data centers, we use the 2021 number as representing a demand absent modern large-scale data centers. We use annual forecasts of electricity demand by electricity market module region from the 2021 EIA AEO to calculate year-over-year growth rates from 2021 through 2030. We use the 2021 vintage of the AEO to forecast electricity demand for the same reason as the most recent (2025) AEO forecast would include demand growth from the build-out of data centers. The 2021 vintage of the AEO would have been completed before the boom in AI specific data centers, but still include expectations of growth in demand from electrification and the energy transition. Year over year growth rates are calculated for the EMM regions and then mapped to EPA Egrid regions.[7] We then apply these growth rates to the 2021 observed net generation figures above to estimate annual demand for years 2026-2030 *net* of data center demand.

To apportion this annual demand number across hours within a year we use data from the EIA-930 form which reports hourly load by balancing authority for January 2021 through 2025. For each balancing authority, we generate a load factor for each hour of the year which is the

[7]For the egrid regions that span multiple EMM regions, we use the sum of demand over the underlying regions for the growth rate.

percent of annual demand consumed in that specific hour. Using these observed load shapes, we then construct a panel of forecast load shapes for our forecast years by performing a block bootstrap at the month level to allow for temporal correlation in demand. This results in a 5-year panel (2026-2030) where the load shape for each calendar month is a random selection from the corresponding donor months. The selected month from the bootstrap draw is consistent across balancing authorities to account for common shocks. We assign the load shape to our EGrid geography by using the shape of the most represented balancing authority within the EGrid region.[8] Lastly, each hourly load factor is then multiplied by the annual electricity demand number to yield the hourly load.

For each of our demand scenarios we use the same data source as our ex-post analysis, using reported power and operational dates for all operational and propsed data centers from CleanView. For the demand scenarios, we use 100% of reported data centers operating at 100% capacity as our highest case. For the two mid-range cases, we allocate compute across hours as before using estimated market shares and utilization rates from Bloomberg/DC Byte. Unlike before, there is an additional source of uncertainty in data center demand. there is significant uncertainty in which data centers will be actually constructed as data center owners have submitted duplicate bids and plans and many operators have constructed (or plan to) on-site primary and back-up generation. To represent this uncertainty, we first scale down the data center power number by 40% before constructing our mid-range scenarios. The low-end case is now a 50% further reduction from this. Hourly data center demand is added to the hourly demand constructed above when calculating the appropriate counterfactuals.

A.2 Supply Forecast

Constructing forecasted thermal supply is straight forward. We use the EIA's Form 860m to gather the most up-to-date information on planned retirements, uprates, and derates to apply to our existing thermal panel at the monthly-level. We use the last available heat rate, capacity,

[8]The majority of eGrid regions are largely or entirely comprised of one balancing authority.

and permit prices for each existing generating unit.[9] Form 860m also contains information on new planned thermal generation by month. We apply average heat rates from existing plants of the same generator type and of newer construction vintage (since 2017) to capture technological innovation. We apply the permit prices from units of the same subregion to planned units. Plants are derated and have stochastic outages as in our baseline approach.

In order to construct unit-level marginal costs, we require data on fuel costs. We retrieve the Henry Hub and WTI prices corresponding to the months drawn in our bootstrap, and for each day, we compute the deviation in observed price from the donor month's average annual natural gas or oil price. We multiply these coefficients by the annual natural gas and petroleum prices for electric power projected in the 2025 AEO. We apply the state-specific markups to these daily natural gas and oil prices to capture the time-invariant spacial differences in prices faced by our panel of thermal units. Coal and coke plants receive the same treatment, utilizing the fuel cost data from the EIA Form 923 that corresponds to each donor month and projected coal prices from the 2025 AEO.[10]

For renewables, we use data from the EIA-930 to estimate month by hour capacity factors for installed solar and wind. We then apply these to the total amount of installed wind and solar from existing units and planned units in EIA-860m.

We perform a similar exercise for hydroelectric units. We create average hourly generation profiles for each month by using hourly data from the EIA-930. For each subregion-hour within a donor month, we compute the average share of that month's total hydro generation that is utilized. We then apply those shares to the total monthly output of each hydroelectric facility, as reported in the EIA-930, for that donor month.[11] We use average hourly capacity factors by subregion for

[9] We ignore unit degradation over time. Given our forecast is over only a five-year period, we don't expect unit heat rates to change much over this time. To the extent that they do, our results will tend to under-estimate costs and emissions across thermals.

[10] The AEO does not include projected coke prices for delivered power. Because coal and coke are substitutes, the price of coke is highly correlated with coke. We observe that over our sample, the cost for coke is, on average, $0.40 per MMBTU higher than coal. Thus, we use the price of coal with a $0.40 markup for the small number of coke units in our panel.

[11] If the shares work out such that a facility ever produces above its nameplate capacity, we instead force that facility to have uniform output for all hours within a month.

planned hydro units. We take this approach in order to preserve some of the complex seasonality and environmental constraints inherent to hydro operations without explicitly modeling reservoir shadow prices as in Bushnell (2003).

All nuclear, cogeneration plants, and other, less common generation technologies, such as geothermal plants, receive the same treatment. While output from nuclear plants does show some variation across months due to factors such as maintenance outages/derates, their hourly production tends not to fluctuate, as it can be quite costly for these units to ramp up and down. Cogeneration plants are typically used in industrial processes, and thus typically do not respond to wholesale market signals. Other generation types, such as geothermal, tend to operate similarly to nuclear plants, or are so uncommon (e.g. wood-burning generators), that the benefit to modeling their supply decision directly would provide little-to-no benefit to our analysis. We use the EIA-923 to compute average monthly capacity factors for all of these units, and assume they constantly generate at that fraction of their nameplate capacity.[12]

Along with solar, grid-scale storage capacity is expanding at a rapid pace. But, storage is unique in its role as an electricity arbitrageur. And, batteries are constrained by their state-of-charge, round-trip efficiency, and rated power capacity (their upper limit on charging and discharging), rather than by fuel costs or daily weather patterns. Thus, we cannot treat it similarly to the technologies described above.[13] We treat the battery fleet in each subregion as a single price-taking agent who seeks to smooth the daily net load profile by charging in low net load periods and discharging during periods of high net load.

We construct the aggregate battery fleet in each subregion by combining data on the existing fleet with planned additions from EIA Form 860m. Planned additions report rated power capacity, but not the planned duration, or energy capacity, of the batter –the length of time, at which, they can continuously discharge at their maximum power capacity before having to recharge. To handle

[12]For cogeneration units, we first compute a monthly capacity factor, and multiply this by the average fraction of their reported net generation that is sold to the grid –a number that is reported annually in the EIA-923 data.

[13]Hydroelectric resources also act as arbitrageurs, shifting energy delivery to periods of highest value. However, unlike battery storage which typically operates on a short-term, typically daily, cycle. Hydro scheduling is governed by additional intertemporal constraints spanning seasons (e.g., reservoir management and environmental flows). Modelling these constraints are beyond the scope of this paper. See Bushnell (2003) for further discussion.

this, we probabilistically assign a duration of two or four hours.

We assume the objective of this subregional battery operator is to flatten the daily net load curve rather than to arbitrage prices across time. However, because we have no intra-regional transmission constraints, our wholesale prices will be convex with respect to net load. Thus, this objective function mirrors that of a profit maximizing agent. In addition, we assume this battery operator has perfect foresight over the net load within a 24-hour period.[14] The battery operator's "peak-shaving problem" is described as follows:

For each day, d, the operator observed hourly net load, d_t and chooses x_t and y_t, charge and discharge, to minimize the peak net load and maximize the net load trough. Formally, for each hour, t:

$$x_t = \min(K, \max(0, \underline{d} - d_t)), \tag{5}$$

$$y_t = \min(K, \max(0, d_t - \overline{d})), \tag{6}$$

where we use binary search to numerically solve our peak and trough thresholds, $\underline{d}$ and $\overline{d}$, to subject to a daily energy balance constraint:

$$\begin{aligned} &\frac{1}{\eta}\sum_{t=0}^{24} y_t = \eta \sum_{t=0}^{24} x_t, \quad \text{where} \\ &\frac{1}{\eta}\sum_{t=0}^{24} y_t \leq E \\ &\eta \sum_{t=0}^{24} x_t \leq E \end{aligned} \tag{7}$$

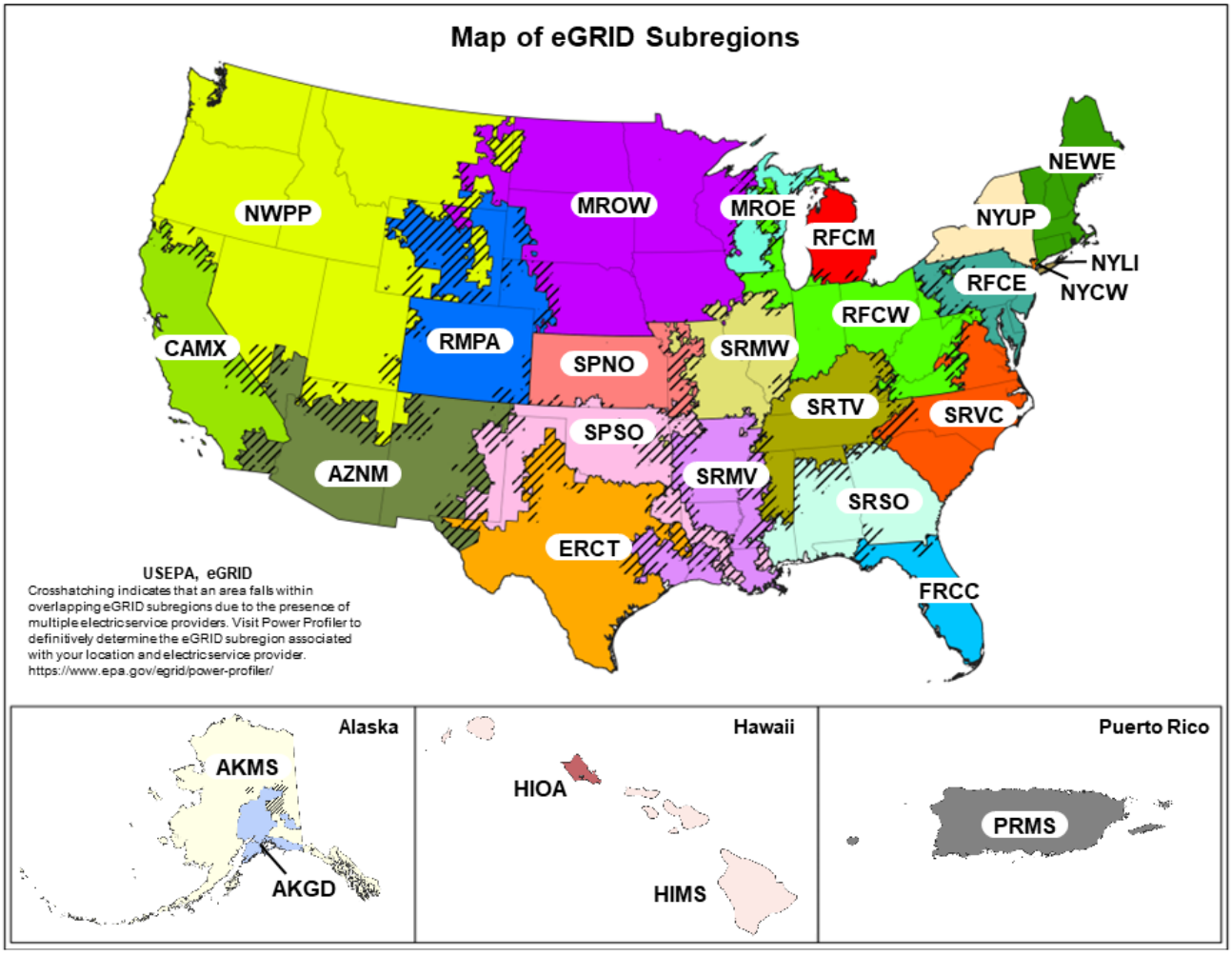
Equation 7 ensures that the battery operator charge and discharge is exactly equal and does not exceed it total energy capacity (in MWh), E. The parameter η represents the round-trip efficiency of batteries, which we set at 0.85.[15]

[14] At the time of solving for battery charge and discharge, net load is equal to our forecasts for regional demand minus solar, wind, hydro, nuclear, cogeneration, and other thermal plants.

[15] Batteries lose some energy when charging and discharging. This value is typically 0.8 to 0.9 for grid-scale storage.

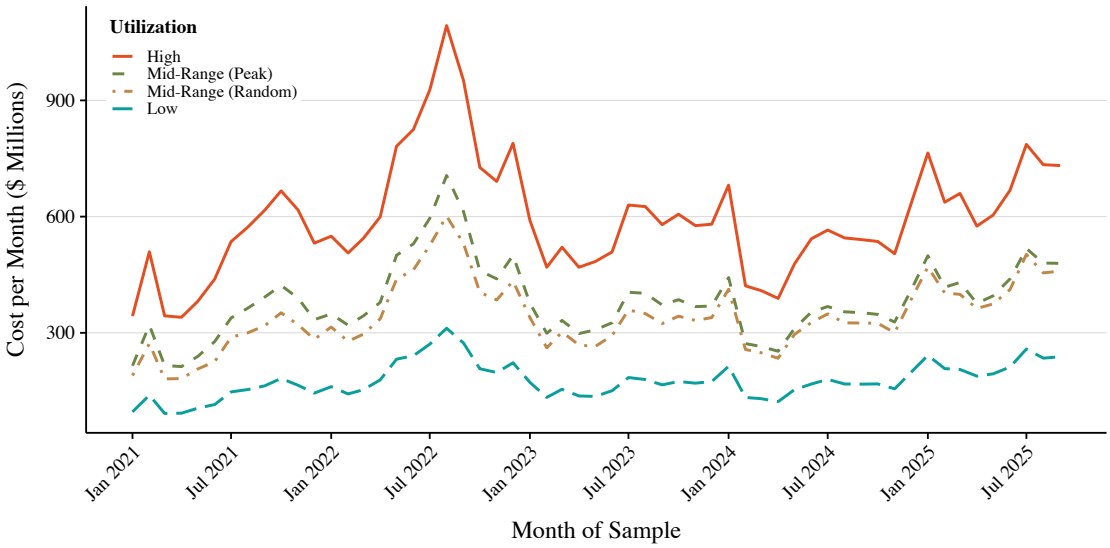
B Appendix: Figures

Appendix Figure A1

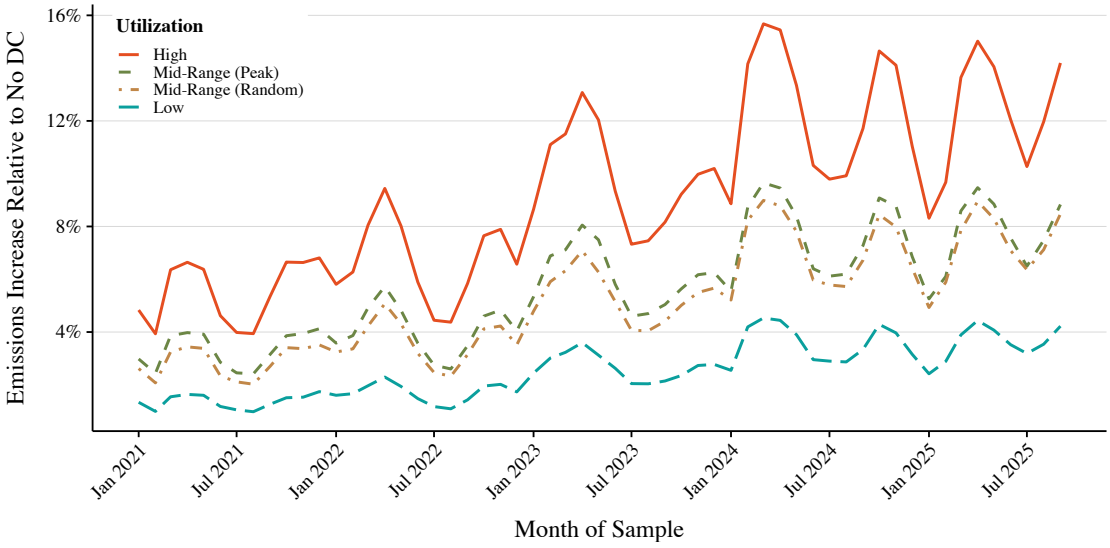


Notes:

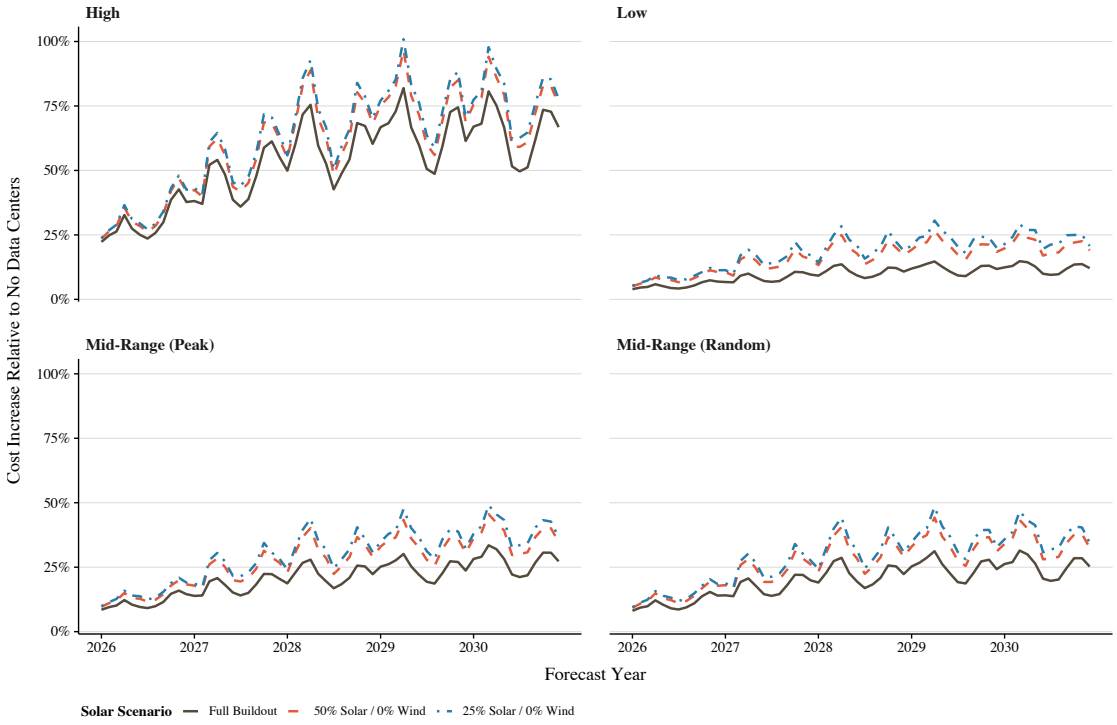
Appendix Figure A2: Monthly Incremental Costs, 2021–2025Q3



Appendix Figure A3: Monthly Precentage Increase in CO_2 Emissions, 2021–2025Q3



Appendix Figure A4: Forecasted Percentage Increase in Generation Costs Under Alternative Renewables Buildout, 2026–2030



Appendix Figure A5: Forecasted Incremental Carbon Emissions under Alternative Renewables Buildout, 2026–2030

