The Energy Cost of AI and Data Centers*

Olivier Darmouni
HEC Paris

Clemens Lehner
Columbia Business School

Yuqi Zhang
Columbia Business School

October 14, 2026

Abstract

We study the spillover effects of rising AI and data center electricity demand on the rest of the U.S. economy over the next decade. Combining project-level data on planned data center and power plants with regional power supply curves built from individual power plants, we calibrate a multi-region spatial equilibrium model of electricity prices, output in non-data-center sectors, and emissions. Despite a large increase in demand, our baseline predicts only modest electricity price increases on average by 2030, with output losses in non-data-center sectors below 0.1% of GDP. Emissions rise significantly as new electricity demand is met largely by existing and new gas plants. The energy sector emerges as a large winner. However, the effects are highly uneven across regions. Price increases and output losses are concentrated in a small number of exposed areas, such as Texas, Virginia and the Carolinas. The magnitudes of the effects would be several times larger if all announced data center projects were built by 2035. The results show the importance of the spatial and technological composition of new power investment in driving the energy costs of AI.

Keywords: Data centers, artificial intelligence, energy transition, regional spillovers
JEL codes: Q41, Q43, Q56, R11, L94

*Comments welcome. Contact: Olivier Darmouni darmouni@hec.fr; Clemens Lehner clehner28@gsb.columbia.edu; Yuqi Zhang yzhang25@gsb.columbia.edu.

1 Introduction

The rapid expansion of artificial intelligence has triggered an unprecedented wave of data center investment across the United States (Van Nieuwerburgh, 2026).[1] The new generation of AI models demands so much computing power that newly announced data center projects are far larger and more energy-intensive than existing facilities built for cloud and search. For instance, OpenAI's "Stargate" project alone is projected to rival the electricity demand of a city like Los Angeles by 2030. At the national level, the IEA projects that U.S. data center electricity consumption will increase by about 240 TWh, or 130%, between 2024 and 2030. Data centers would then consume more electricity than all energy-intensive manufacturing combined and acccount for almost half of the growth in U.S. electricity demand (IEA, 2025).

These striking magnitudes have raised concerns that the electricity costs of AI will spill over into the rest of the economy, through higher electricity prices, lower output in non-data-center sectors, and higher carbon emissions (Feher et al., 2025; Heiss et al., 2026; Benetton et al., 2025).[2] Citizens, politicians, and regulators have responded with a flurry of initiatives, and policymakers at both the federal and state levels are drafting new rules for large data center loads. The central question is thus no longer whether the AI build-out will materially affect U.S. power markets and the rest of the economy, but how.

Quantifying these potential costs raises two challenges. First, most of the projected build-out has not happened yet. Announced projects will come online over the coming years, so today's electricity prices contain limited information about the equilibrium response. Second, the size of the spillover hinges on the shape of *local* power supply. The slope of the supply curve determines how much a demand shift raises prices, while emissions depend on the sources dispatched to meet new demand. This is complicated by the fact that power supply is inherently local. Electricity is costly to transport and store, so the marginal cost

[1] "Big Tech's AI Push Is Costing a Lot More Than the Moon Landing" (*Wall Street Journal*, 2026).

[2] See e.g., "AI Data Centers Are Sending Power Bills Soaring" (*Bloomberg*, 2025); "Who Pays? AI Boom Sparks Fight Over Soaring Power Costs" (*Wall Street Journal, 2025*); "AI demand drives record electricity supply costs in largest U.S. market" (*Financial Times, 2025*); "The True Cost of AI Data Center Development" (*Business Insider*, 2025).

of meeting additional demand at a given site depends on the dispatch order of generators in that balancing area, the spare capacity of those generators, and the cost of building new ones nearby (Hausman, 2025). The same gigawatt of incremental data center load can therefore have very different price and emissions consequences depending on where it lands, making it essential to account for local supply curves near large data center projects. The build-out is also highly geographically concentrated, plausibly leading to uneven effects on prices, output, and investment across regions.

To address these challenges, we combine project-level micro-data on the planned U.S. data center and energy build-out with a multi-region spatial equilibrium model of electricity markets. Our baseline horizon is 2030, and we also consider the full announced build-out through 2035. We ask two central questions. What is the incidence of the AI-driven electricity demand surge across regions, sectors, and the energy supply chain? And how does it depend on the scale and technology composition of new power built to meet it?

Our data on data centers come from Cleanview, which tracks 3,451 U.S. data center projects (399 GW of capacity) through environmental and air-quality permits, interconnection queue filings, zoning applications, press releases, SEC filings, and local news. We combine these projections with data on existing and planned energy sources. We complement 2023 existing generator-level data from the EIA with new planned generation to construct local power supply curves. Power plant projects data includes both on-the-grid and off-the-grid ("behind the meter") investments. To build supply curves, we compute the marginal cost of each generator operational in the country by 2030 and sort within each balancing area (Hausman, 2025).

We start with key facts that emerge from this micro-data. The pace of the data center build-out is staggering and, if anything, has accelerated in recent months. Data centers that have come online since 2024 or are scheduled to do so by 2030 add 157 GW of new capacity, and including all announced projects raises the figure to 359 GW by 2035. The build-out is highly geographically concentrated. Only 202 U.S. counties host new capacity by 2030 and

the top 20 counties account for 40% of it. Some of that concentration is driven by city-sized mega-projects, used for training frontier AI models. Investment in power generation is also rising much faster than in previous years, driven by a revival of gas. After years of remarkable renewable energy growth, gas-fired power plants have become the largest source of planned new generation, in part through "behind-the-meter" gas projects to avoid delays in connecting a data center to the power grid. Even so, planned new generation falls short of projected data center demand in the largest data center markets.

Building on these facts, we develop a multi-region spatial model to quantify the effects of the data center build-out on electricity prices, output, and emissions. The economy consists of three sectors. First, a data center sector with exogenous, perfectly inelastic electricity demand from new data centers. Second, a non-data center sector that uses local electricity, labor, and capital to produce a final good. Third, an energy sector that consists of electricity producers and fossil fuel suppliers. Electricity markets are regional, with heterogeneous generators producing via "economic dispatch" (i.e. lower marginal cost first) within each region. For tractability, we assume that regional wholesale markets are competitive and that final good producers pay a constant retail markup on top of the wholesale price. We do not explicitly model existing cost-allocation rules for retail rates as the current regulatory landscape is being upended by the data center boom and is likely to change considerably going forward.[3]

The supply response, meaning investment in new power generation to meet data center demand, is central to our analysis. We parametrize the supply response following three dimensions: how much is built (new supply as a share $\alpha \in [0, 1]$ of new demand), what is built (the share $\beta \in [0, 1]$ of renewables), and how it is built (the share $\gamma \in [0, 1]$ off the grid). This parameterization allows us to evaluate the costs and tradeoffs associated with different supply responses, from relying entirely on existing capacity to building out new generation

[3]In November 2025, Virginia's State Corporation Commission approved a new rate class (GS-5) requiring data centers and other 25 MW-plus customers to enter 14-year take-or-pay contracts covering at least 85% of contracted transmission and distribution demand and 60% of generation demand. Ohio adopted similar minimum-billing rules earlier in the year and pending data center interconnection requests there fell from 30 GW to 13 GW within months.

of varying technology composition. We discipline these parameters with project-level data on planned power plants and on generation built directly for data centers. By 2030, these data imply that new supply meets roughly $\alpha = 0.6$ of new data center demand, about half of it clean ($\beta = 0.5$) and half of it behind the meter ($\gamma = 0.5$), which we take as our baseline. No new supply ($\alpha = 0$) and new supply equal to new demand, all on the grid ($\alpha = 1$), bound the spillover to the rest of the economy. The remaining model parameters are disciplined by generator-level data on fuel costs, efficiency and capacity, as well as county-level data on electricity consumption and local economic conditions.

Our first finding is that the aggregate energy cost of new data center demand by 2030 appears smaller than commonly assumed. In our baseline, data centers pay about $78 billion per year for grid electricity, but the indirect cost to the rest of the economy is limited. Higher electricity prices reduce income in non-data-center sectors by about $10 billion per year, and output by 0.04% of U.S. GDP. Even if no new power generation was built, the loss would be $31 billion, or 0.12% of GDP. Two features of today's electricity landscape explain this modest aggregate effect. First, existing fossil fuel generators operate below capacity in many regions, so supply curves are relatively flat around current demand levels and can absorb additional demand without large average price increases. New investment in power supply flattens them further. Second, electricity expenditures represent a relatively small share of aggregate production costs, limiting the pass-through to output in non–data center sectors.

In contrast, the energy sector emerges as a clear beneficiary of the data center boom, throughout the value chain. Electricity generator profits increase by $63 billion and upstream fuel producers gain about $22 billion. Equipment manufacturers also stand to benefit, as building new power supply requires large capital expenditures. Our baseline supply response implies about $296 billion in capital expenditure by 2030, and meeting all new data center demand with new capacity would cost $373–773 billion, depending on the technology mix

($856–1,544 billion for the full build-out through 2035).[4]

The environmental cost is however likely to be much higher. Emissions rise as new demand is met largely by gas, whether by expanding existing plants or building new ones. We estimate an aggregate emission cost of approximately $88 billion per year, given a social cost of carbon of $200 per ton. Building renewables instead of gas plants would naturally cut these emissions substantially. However, the technology mix matters little for electricity prices and output, since both technologies have relatively similar effects given existing spare capacity. In contrast, the mix matters greatly for the distribution of energy sector profits. Building renewables rather than gas redistributes profits from fuel providers to electricity generators, since low-cost renewables are dispatched first but do not require purchased fuels. These findings speak to an active debate about who captures the economic rents generated by AI-driven electricity demand and whether this demand is primarily met through investments in renewables or fossil fuels (Heiss et al., 2026).

These aggregate results conceal substantial regional heterogeneity. In regions with concentrated data center deployment and limited spare capacity, electricity prices can increase sharply even in the presence of new supply. These price increases generate localized output losses and distributional effects across sectors. In our baseline, 50% of the output loss is concentrated in three areas: Most of Texas, Virginia/Carolinas, Lower Mississippi Valley. The largest wholesale price increases are in Most of Michigan (+23%), Most of Texas (+17%), Texas Panhandle-Oklahoma (+11%), against a median of 5% across balancing areas. Even regions with moderate data center presence can experience significant price pressures if their local supply curves are steep, as in Michigan.

The full announced build-out through 2035 is much larger, at 359 GW, and about half of it consists of projects with no announced online date. Under a supply response calibrated the same way, output losses reach 0.14% of GDP and wholesale prices rise by up to 68% in

[4]These figures use current National Renewable Energy Laboratory Annual Technology Baseline capital costs for wind, solar and storage, which have risen substantially since 2022, and recent utility filings for combined-cycle gas plants ($2,200 per kW).

the most exposed areas (Northwest US (+68%), Texas Panhandle-Oklahoma (+61%), Most of Texas (+54%)), where large undated projects concentrate. Our baseline estimates are thus likely to underestimate both the economic and environmental costs if the data center build-out keeps accelerating.

We also emphasize that the AI energy boom is occurring in an unfavorable macroeconomic environment. The 2022 inflation crisis and supply chain bottlenecks have increased the cost of building power plants by over 50%.[5] Higher gas prices due to supply disruptions in the Middle East would generate large spillovers, since much of the new demand is likely to be met by building new and ramping up existing gas plants.[6] Finally, a higher degree of electrification in the rest of the economy (EVs, heat pumps, cooling, etc.) also increases the energy cost of AI, since data centers then compete more fiercely with other users.

Our findings also mask the amplification of two structural risks that are more difficult to capture within our model. The first is reduced grid reliability. Existing spare capacity in power supply is important for avoiding price spikes and blackouts during exceptional events such as extreme weather (Kacperczyk and Licher, 2025). Much larger loads from data centers reduce the ability of local power grids to handle such episodes.[7] The second is increased dependence on fossil fuels, and in particular gas, whose price is highly volatile and influenced by geopolitical events. Any medium-run transition away from fossil fuels is made more difficult by the large increase in electricity demand from AI.

Our findings carry implications for policy and the energy transition. We relate to policy debates over who will bear the energy costs of AI and data centers. The concentration of economic costs in a few regions underscores the need to develop new mechanisms that limit energy spillovers to the rest of the economy. Moreover, the effects of the AI boom on the path of the energy transition are an important open question. Although using more fossil

[5]For example Wood Mackenzie estimates that combined-cycle gas plant costs have risen 71% since 2022, with delivery times for gas turbines now exceeding four years (Wood Mackenzie, 2026).

[6]The IEA projects that natural gas will be the largest source of new U.S. data-center electricity supply through 2030, adding over 130 TWh of annually compared to about 110 TWh from renewables (IEA, 2025).

[7]See "America's Biggest Power Grid Operator Has an AI Problem: Too Many Data Centers" (*Wall Street Journal*, 2026).

fuels seems unavoidable in the short run, the massive increase in electricity demand from data centers might stimulate technological innovation in the long run. We are currently seeing an acceleration in the development of new low-carbon energy sources, such as geothermal and nuclear, in parallel with the data center boom.[8]

The remainder of the paper is organized as follows. Section 2 discusses related literature. Section 3 describes the data and key facts. Section 5 presents the model and calibration. Section 6 reports counterfactual results. Section 7 discusses policy implications.

2 Related Literature

AI build-out A growing body of work documents the physical, financial, and local dimensions of the AI build-out. Van Nieuwerburgh (2026) documents both its scale and the financial architecture sustaining it: a layered capital structure of REITs, private credit and securitized vehicles. We relate most closely to recent works empirically analyzing the effect of the AI build-out on electricity prices and energy. Feher et al. (2025) empirically analyze the effect of data centers on local electricity prices and real outcomes. They find little effect of the build-out so far, except in the very highest exposed areas, consistent with the existence of spare capacity in the grid. Using wholesale market data, Kay et al. (2026) find that existing data centers have already raised wholesale prices by a few percentage points on average, with larger effects in the main data center corridors. Evidence on retail bills is mixed. Meeks et al. (2026) find data center entry increases retail rates, with effects concentrated in investor-owned utilities and deregulated states, and Scalera et al. (2026) find that data centers raise household electricity prices, with households bearing a disproportionate share of the increase. Watten et al. (2026) instead find that retail rates fell modestly between 2015 and 2024 as fixed grid costs were spread over more load. On the environmental side, Muller (2026) estimates that air pollution and greenhouse gas damages from data center electricity use reached about $25

[8]See e.g., "Three Mile Island Nuclear Plant Will Restart to Power Microsoft AI" (*Bloomberg*, 2024); "Google invests in Fervo's $462M round to unlock even more geothermal energy" (*TechCrunch*, 2025).

billion in 2025. Benetton et al. (2025) study the electricity demand from Bitcoin mining in upstate NY and find that it increased bills by around 7% and reduced consumer surplus. Heiss et al. (2026) exploit the public release of ChatGPT as a plausibly exogenous shock to show that data center regions attract substantially greater investment in renewables and battery storage rather than gas-fired generation, with hyperscaler net-zero commitments and long-term PPAs as the enabling mechanisms. Mytton et al. (2023) document the grid capacity constraints that this surge in demand creates, highlighting the tension between rapid data center expansion and the pace of transmission and generation investment. We contribute to this line of work by using a model-based counterfactual approach building on Zhang (2025) to project the effects ten years out, alleviating the concern that the acceleration of data center growth has only recently started. Our focus on the spatial distribution of data center electricity demand also relates to Fang and Greenstein (2020), who studies the economic geography of data centers, documenting how access to cheap power, infrastructure and local regulatory conditions jointly shape their geographic concentration. Gargano and Giacoletti (2025) show that state tax subsidies are a meaningful determinant of data center location. Alvarez et al. (2026) find that data centers bring lasting gains in local employment and house prices, while their effect on local electricity prices emerges only with a lag. Relative to these studies, which measure the effects of the data centers built so far, we quantify the much larger build-out that is still to come.

Energy costs and industrial competitiveness. A substantial literature studies how energy prices affect industrial output, firm competitiveness, and employment. Hausman and Kellogg (2015) examine the welfare and distributional consequences of energy price shocks for households and industries. Arkolakis and Walsh (2023) show that clean power expansion can support economic growth and industrial competitiveness. Arkolakis and Walsh (2024) analyze how clean power expansion shapes local electricity prices and spatial equilibrium outcomes, including regional wages and output. Together, these papers provide the empirical and con-

ceptual backdrop against which we evaluate the costs and spillovers of data center demand across U.S. regions and sectors.

Electricity markets and price formation. Our paper builds on an extensive line of work studying how market design and transmission constraints shape electricity prices. Cicala (2015) and Cicala (2022) show that deregulation and market-based dispatch improved productive efficiency in U.S. electricity generation, while MacKay and Mercadal (2022) emphasize that these efficiency gains can be offset by higher wholesale markups and ambiguous retail price effects. Davis and Hausman (2016) provide direct evidence on the price and welfare effects of a localized supply shock (the San Onofre nuclear closure) by showing how a localized supply shock propagates through a transmission constrained power market. Hausman (2025) and Ham et al. (2026) more broadly quantify the allocative costs of limited interregional transmission and the resulting spatial price wedges. We build on this literature by treating the AI-driven surge in electricity demand as a series of localized demand shocks and using a spatial equilibrium model to trace how these propagate through segmented power markets.

3 The data center and energy infrastructure build-out

The AI boom has set off a historic investment cycle in data centers. U.S. Census data show that monthly spending on data center construction reached $5.4 billion in August 2026 (in 2021 dollars), more than six times its level when ChatGPT was released in November 2022 (Figure 1). Data centers have gone from a small niche to nearly two thirds of all private office construction and spending in the first eight months of 2026 was more than a third higher than a year earlier.

These buildings will need power. After two decades of almost no growth in U.S. electricity consumption, the EIA now projects demand to rise by about 11% between 2023 and 2030 (Figure 2, panel a). Data centers are the main driver of this acceleration. Projections of their electricity use in 2030 range from about 180 TWh in the EIA's own outlook, up about 70%

from 2025, through about 425 TWh in the IEA's projection, up about 130% from 2024, to nearly 600 TWh in the medium scenario of the Electric Power Research Institute (EPRI), more than four times its 2023 level (panel b).

These increases are large relative to new supply. The EIA projects U.S. net generation to grow by only about 43 TWh a year between 2025 and 2030. Data center demand growing at the IEA's pace, about 40 TWh a year, would absorb nearly all of this new generation, and growth at EPRI's medium pace, about 65 TWh a year, would exceed it by half. If data centers grow the pace many forecasts expect, the power system will need far more new supply than official projections assume. Meeting this demand therefore requires an investment boom in power generation of its own.

The rest of this section uses project-level data to document both the pace of the data center build-out and the new power generation planned.

3.1 The pace of the data center build-out

Our data on data center projects come from Cleanview, which tracks announced and operating U.S. data centers from public records. Its records report each project's location, capacity, developer, development status and expected online date, drawing on environmental and air-quality permits, interconnection queue filings, zoning applications, press releases, investor presentations, SEC filings and local news. The database is updated continuously. We compare the pipeline of planned projects at two points in 2026, in February and in September.

Announcements have accelerated and spread across the country. Figure 3 shows that capacity planned or under construction and due online by 2030 more than doubled, from 60 GW in February to 129 GW in September.[9] Over the same seven months, the number of counties with a project due by 2030 also roughly doubled, from 109 to 215. Figure 4 compares the location of the pipeline in February and September. Established hubs such as Northern Virginia, central Utah and Phoenix kept growing, while new projects appeared throughout

[9]Part of this increase reflects broader coverage of smaller projects in the later data.

the Ohio Valley, the Midwest, Texas and the Southeast.

However, of the 28.1 GW of projects scheduled in February to come online during 2026, only 4.8 GW (17%) was operating or under construction by mid-September. Another 11.3 GW was still planned but had been pushed back to a later date. Construction nonetheless keeps pace with mainstream forecasts. Figure 5 shows the data center capacity coming online each year. It reached 13.1 GW in 2025 and 9.4 GW in the first eight and a half months of 2026, against roughly 9 GW a year in EPRI's medium scenario. The build-out runs late relative to announcements, but on track relative to forecasts, and it shows no sign of slowing down. For our model, this means that the capacity scheduled for 2030 is best read as the build-out delivered on schedule, an upper envelope for new data center demand rather than a central forecast.

3.2 Concentration and mega-projects

Not all data centers are alike. Projects range from small colocation facilities serving cloud and enterprise customers to campuses of several gigawatts built to train and serve frontier AI models. For scale, one gigawatt is roughly the output of a large nuclear reactor. At a 70% utilization rate, a 1 GW campus uses as much electricity as about half a million U.S. households. The largest campus in the pipeline, at around 12 GW, would consume as much as 6.8 million households. Figure 6 shows the size distribution of announced projects. Most of the Giga-scale projects are planned for 2035 or later.

What makes the build-out consequential for power markets is its geography. Only 202 counties are projected to host new data center capacity by 2030, and the top 20 hold 40% of it (Figure 7). Major clusters include Northern Virginia, central and southern Ohio, metropolitan Phoenix, and northeastern Louisiana. The map also shows the regional electricity-market boundaries used in our model. Because electricity is priced within regional markets, concentration across counties translates into large demand increases in a relatively small number of power markets.

Where data centers go is not random. Table 1 compares the 86 counties accounting for approximately 80% of new capacity by 2030 with the remaining 3,035 counties. High-capacity counties have, on average, 5.5 times as much employment and 4.7 times as much population, and annual pay is approximately $8.1 thousand higher. They also have large existing fossil infrastructure, with roughly five times the natural-gas capacity. In contrast, differences in local wind and solar capacity are not statistically significant. These comparisons suggest that developers are choosing places where firm power and transmission already exist, not places where new clean power would be cheapest, in line with evidence that power access, infrastructure, and local taxes and regulation jointly shape data center location (Fang and Greenstein, 2020; Gargano and Giacoletti, 2025). For our analysis, the central implication is that the national electricity requirement must be evaluated against the supply available in the particular markets where projects locate.

3.3 Power investment: the revival of gas

To analyze how power supply is responding to data center growth, we collect micro-data on planned power plant projects, both those that will connect to the grid and those built behind the meter for specific data centers.

We first analyze projects connecting to the grid using data from form EIA-860M. Figure 8 shows how planned generation has changed since the advent of AI. After years of stagnation, expected generation from planned plants nearly doubled, from 393 TWh per year in December 2022 to 775 TWh in August 2026 (panel a).[10] Gas drives most of this growth. Panel (b) shows that gas dominates new proposals since 2025, after ten years of renewables domination. Planned gas generation tripled and accounts for about two thirds of the increase. Nuclear investment is also back, after years of absence, but is quantitatively smaller, and we abstract from it in the rest of the paper.

[10]Note that planned projects are reported in terms of capacity (MW), which is not comparable across technologies due to differences in capacity factors. The figure makes them comparable by using fixed, average national capacity factors. This however does not fully capture variations in utilization due to local resources availability or demand response.

Figure 9 shows the spatial distribution of new power generation against planned data center load. Many areas are below the 45 degree line, suggesting that this new supply might fall short of the new data center demand. This is particularly true of the largest data center markets. Planned plants would cover 20% of new data center load in the Ohio Valley and 10% in Virginia and the Carolinas, against 69% in Texas.

One reason is that connecting to the grid takes time. In fact, interconnection delays now exceed five years in many parts of the U.S. Given the race to power AI, many data centers plan to co-locate with their own power supply by building a power plant nearby. This new power is said to be "behind the meter" (BTM) in the sense that it is not connected to the rest of the grid.[11] We build a project-level dataset of planned BTM plants by combining Cleanview's reports, Global Energy Monitor's gas plant tracker, the Environmental Integrity Project's inventory of gas plants for data centers, and state air-permit records, and we verify each project individually against permits and news coverage. A plant is included if it is built primarily to serve a specific data center campus, while utility-owned plants serving data center load, standby generators, and merchant plants without a named data center customer are excluded. For each project, we record its capacity in the current phase, its technology, and its status, from announced to permitted, under construction or operating. Panel (a) of Figure 10 shows their take-off since 2025. with 94 GW announced by October 2026 including 33 GW permitted, under construction or operating. Panel (b) shows the geographical distribution: Texas stands out as representing almost half of all BTM planned capacity.

Almost all of this capacity is gas-fired. For illustration, Table 2 lists the largest projects in our data. For example, the GW Ranch project, currently under construction to power data centers for Amazon, includes 750 MW of solar capacity and 1.8 GW of battery storage. This is however dwarfed by the scale of its gas power of 7.7 GW. There are currently no large data center projects powered fully by renewable in the current pipeline. The physical footprint and battery capital costs make a pure renewable BTM architecture economically impractical today.

[11] See e.g., "AI Data Centers, Desperate for Electricity, Are Building Their Own Power Plants" (*Wall Street Journal*, 2025); "AI boom drives clash over grid power vs. going it alone" (*Axios, 2026*).

Access to local gas explains why Texas and other gas-rich regions dominate BTM planned projects. Note that many of these projects cannot rely on state-of-the-art gas turbines, due to supply bottlenecks, and thus often turn to less efficient sources of gas generation. Only about a third of planned BTM gas capacity is identified as combined cycle, and the rest consists mostly of simple-cycle or aeroderivative turbines, some of them repurposed aircraft engines.

4 Data

Our analysis combines project-level data on planned data center developments with detailed information on electricity generation, fuel costs, electricity consumption, and local economic conditions. Our model is calibrated to 2023 as the baseline period, capturing the state of U.S. electricity markets before the projected wave of data center demand.

Data center demand: Data center demand in the model comes from the September 2026 version of the Cleanview data, which lists 3,451 projects covering almost 400 GW of installed or announced capacity. New data center capacity by a given horizon comprises projects that came online from 2024 onward or are scheduled to come online by that horizon. Projects under construction count as planned and projects with no announced online date (181 GW) are assigned to 2035. This gives 157 GW of new capacity by 2030, our baseline horizon, and 359 GW by 2035. We convert capacity into electricity demand with a 70% utilization factor following Shehabi et al. (2024), which yields 966 TWh a year by 2030, and assign each project to a balancing area through its county. Since announced online dates tend to be optimistic (Section 3), we read this demand as the announced build-out delivered on schedule. It is an upper envelope rather than a central forecast.

Electricity generation and fuel costs: We characterize the existing U.S. electricity generation fleet using generator-level data from the U.S. Energy Information Administration (EIA). EIA Form 860 provides information on each generator's nameplate capacity, prime mover technology, fuel type, and geographic location. EIA Form 923 reports plant-level data

on monthly fuel consumption, fuel costs, and net electricity generation. Together these data allow us to construct generator-level marginal costs and regional supply curves. Fuel prices are national averages by fuel type, weighted by quantity, so plants differ in marginal cost through their heat rates and local wages rather than through regional fuel prices.

Electricity consumption and market geography: The geographic unit for electricity markets in our model are balancing areas based on the EPA eGRID subregions (Figure 11). These boundaries, based on the EPA's Emissions and Generation Resource Integrated Database (eGRID), proxy for transmission-related market segmentation and capture the fact that electricity markets in the United States are not fully integrated. Each county is assigned to the balancing area that covers most of its area. Electricity flows freely within each balancing area in the model but not across them. Electricity consumption by balancing area is constructed from EIA Form 861, which reports retail electricity sales and revenues by utility and state. We allocate each utility's sales in a state to the counties in its service territory in proportion to county employment, and sum counties to balancing areas.

Local economic data: County-level labor supply is measured using total employment from the 2023 Quarterly Census of Employment and Wages (QCEW), published by the Bureau of Labor Statistics. County-level wages are from the Bureau of Economic Analysis (BEA) Regional Economic Accounts for 2023.

Capital costs of new generation: To estimate the investment required to build new power generation, we use technology-specific baseline capital cost projections from the 2024 National Renewable Energy Laboratory (NREL) Annual Technology Baseline (ATB). The ATB provides a consistent set of cost and performance parameters for electricity generation and storage technologies in the United States, with projections through 2050. The cost estimates used in our counterfactual exercises are averages of the ATB's annual capital cost projections from 2024 to the horizon year (2030 or 2035). For combined-cycle gas plants, the ATB lags the recent surge in turbine prices, so we instead use $2,200 per kW, the median of 2028–31 utility combined-cycle filings (GridLab, Energy Futures Group and Halcyon, 2025;

BloombergNEF reports a 2025 U.S. average of $2,157 per kW). We also construct an alternative set of cost estimates based on pre-2022 ATB vintages to quantify the impact of recent supply chain disruptions and input cost inflation.

5 Model

5.1 Framework

We develop a multi-region model in which data center electricity demand is introduced into regional electricity markets. The economy consists of three sectors: a data center sector, a final good sector, and an energy sector. The model characterizes a static equilibrium, and there are no explicit dynamics. There is a set of regions indexed by i, where data centers are built, the final good is produced, and workers reside. There are a set of balancing areas indexed by b, within which electricity flows freely and without loss. Each region i belongs to a balancing area $b(i)$.

5.1.1 Data Center Sector

Data centers belong to specific regions and can purchase electricity from their balancing area's electricity market. Their electricity demand is exogenous and perfectly inelastic. For each balancing area b, there is E_b^{dc} units of exogenously-given data center electricity demand.

Data centers do not produce an output in the model. They generate an exogenous change in electricity demand that varies across space.

5.1.2 Final Good (Non-Data Center) Sector

In each region i, a representative firm produces a homogeneous final good using a Cobb-Douglas production technology

$$Y_i = A_i E_i^a L_i^b K_i^{1-a-b}$$

where E_i is electricity purchased from the balancing area at the local retail price $\tilde{p}_{b(i)}$, L_i is labor that is hired at the local wage rate w_i, K_i is capital rented at a fixed rate $\bar{r}$, and A_i captures region-specific productivity.[12]

We assume there are no productivity gains from AI that accrue to this sector. Markets for labor are local; there is no labor mobility across regions.

5.1.3 Energy Sector

Electricity Generation. Electricity is produced from two technology classes: fuel-based energy sources (e.g. fossil fuels), and renewable sources (e.g. wind and solar). Electricity generated from each technology are perfect substitutes. Each fuel-based electricity generator ι operates under a Leontief technology using fuels and labor as inputs, with a capacity constraint. The electricity produced by each fuel-based generator follows:

$$E_\iota = \min\{\nu_\iota F_\iota, \xi L_\iota, \bar{C}_\iota\} \tag{1}$$

where F_ι is the quantity of fuel purchased at the exogenously specified unit price c_ι, and L_ι is the amount of labor hired at the regional wage $w_{i(\iota)}$. The fuel cost for a generator may vary based on the type of fuel that the generator uses (e.g. coal, natural gas). The generator-specific coefficient $\nu_\iota > 0$ governs the conversion of fuel into electricity, and may incorporate technological factors of a fuel-based generator such as its heat rate. The coefficient $\xi > 0$ reflects electricity produced per unit of labor. The term $\bar{C}_\iota$ denotes the generator's capacity, which is a physical limit on the total output of the generator. With full use of inputs, each generator has a constant marginal cost up to the capacity constraint of $MC_\iota = \nu_\iota^{-1} c_\iota + \xi^{-1} w_{i(\iota)}$, encompassing both fuel and operating expenses.

Each renewables generator ι operates with Leontief technology using labor as an input, with a capacity constraint. Unlike fuel-based generators, renewables do not use purchased fuels,

[12] In the quantification, the rental rate of capital is fixed and normalized to one, with any constant absorbed into the productivity parameter without affecting equilibrium allocations.

but rather use a free locally-available resource, which is indexed by $\theta_{\iota,t(\iota)}$. The electricity produced by each renewables generator follows:

$$E_\iota = \min(\vartheta_{\iota,t(\iota)} L_\iota, \theta_{\iota,t(\iota)} \bar{C}_\iota) \tag{2}$$

The factor $\theta_{\iota,t(\iota)} > 0$ reflects the physical productivity of the generator, which incorporates the wind speed or solar irradiance based on the location of the generator. Marginal costs are constant for each generator at $MC_\iota = \vartheta^{-1}_{\iota,t(\iota)} w_{i(\iota)}$.

New Power Generation. There is investment in new power generation that is parameterized along three dimensions. First, $\alpha \in [0,1]$ is the fraction of data center electricity demand in each balancing area that is met by new local generation in the same balancing area. When $\alpha = 0$, there is no new supply and data center demand must be served by existing generators. When $\alpha = 1$, data center electricity demand is exactly equal to the amount of new generation. Second, $\beta \in [0,1]$ is the share of new generation that consists of clean power versus fossil fuels. When $\beta = 0$, all new generation is fossil fuels that creates carbon emissions; when $\beta = 1$ all new generation is clean. Third, $\gamma \in [0,1]$ is the share of new generation that is built off the grid, just to serve local data centers. This captures the effect of growing trend of "co-location" or "behind the meter" power generation. This share of new supply does not affect regional supply curves, and can potentially have a different renewable share than the average β.

Market. Each balancing area is modeled as a competitive wholesale market, where generators are dispatched according to economic dispatch. That is, they are ranked in order of marginal cost and each generator produces up to its potential capacity until total demand in the balancing area is met.

Let total electricity demand in balancing area b be $E_b = E_b^{\text{dc}} + \sum_{i \in \mathcal{R}_b} E_i$. The first term is demand from data centers and the second term is demand from final good producers. Generators can be ordered in increasing marginal cost, creating an upward-sloping supply curve. Figure 12 illustrates the result of this procedure in the case of ERCT region covering

most of Texas. For comparison, the figure also includes average hourly consumption in 2023 as well as projected demand increase from data centers. The price p_b^E that clears the market in each balancing area is the marginal cost of the last generator dispatched, which may run below its capacity. Markets clear on annual quantities, and there is no trade between balancing areas. The retail price is then equal to $\tilde{p}_b^E = p_b^E + \eta_b$, where η_b is a constant transmission and distribution markup.

5.1.4 Equilibrium

An equilibrium consists of final good input demands $\{E_i, L_i, K_i\}$, generator outputs and input demands $\{E_\iota, F_\iota, L_\iota\}$, wages $\{w_i\}$, and electricity prices $\{p_b^E, \tilde{p}_b^E\}$ such that each representative firm maximizes profits, generators are dispatched according to economic dispatch with p_b^E equal to the marginal cost of the last dispatched generator; retail prices satisfy $\tilde{p}_b^E = p_b^E + \eta_b$; labor markets clear in each region i; electricity markets clear in each balancing area b.

5.2 Quantification

Regions correspond to counties in the contiguous United States, and balancing areas follow EPA eGRID boundaries, which proxy for transmission constraints (Hausman, 2025). Labor supply $\bar{L}_i$ is measured using total employment from the 2023 QCEW. The output elasticities of labor and electricity are set to $b = 0.6$ and $a = 0.02$, calibrated to match the economy-wide cost shares of labor and electricity, respectively. Data center electricity demand E_b^{dc} is constructed from Cleanview project-level data, aggregated to the balancing area, with a 70% utilization factor applied to convert capacity into electricity demand (Shehabi et al., 2024).

Horizon. Our baseline horizon is 2030, for two main reasons. First, the planned supply side is largely observed rather than extrapolated. Developers file new power plants with the EIA only three to four years before operation, so 90% of the expected generation in the August 2026 EIA-860M pipeline is due online by 2030. Second, data center demand by 2030 consists

entirely of planned projects with a stated online date. We report the full announced build-out through 2035 as an alternative scenario in Section 6.5.

Supply response. We calibrate (α, β, γ) to the new power supply that is already operating or planned, on the model's energy basis (Table 5). The model's fleet is that of 2023, so the grid side starts with the 144 GW of generators that came online between January 2024 and August 2026. They would generate 250 TWh a year, 90% of it from solar, wind and nuclear. To these we add every generator planned in the August 2026 EIA-860M and due by the horizon. We convert capacity to annual energy with technology-specific capacity factors, namely 64% for combined-cycle gas, 10% for simple-cycle turbines and engines, 24% for solar, and 37% for wind. Three adjustments follow. First, planned online years beyond the EIA's three-to-four-year window are thin. To approximate account for future planned projects, we fill 2029 and 2030 up to the 2027-28 level of scheduled generation. Second retirements since the 2023 fleet are subtracted, both realized and planned. Third, demand growth outside data centers is subtracted, since the model holds that demand at its 2023 level. It equals the growth in U.S. retail sales projected by the EIA minus the growth in data center use, taken from EPRI to 2025 and from the EIA's own data center projection thereafter. This gives 219–249 TWh by 2030.

The behind-the-meter side is our list of on-site generation built for specific data center campuses, about 90% of it gas. Each side is reported at face value and adjusted for completion. The adjustment weights planned grid units by the share of the December 2022 planned pipeline that was operating by August 2026 (49% for gas, 51% for solar and 33% for wind), units under construction by 95%, and behind-the-meter projects by status.

The implied share of new data center demand met by new supply is $\alpha \in [0.57, 0.90]$ by 2030. We take $\alpha = 0.6$, near the completion-adjusted end. The face-value end assumes that every planned plant is built on schedule, while turbine order backlogs extend to about 2030 and only about half of the 2022 planned cohort was operating four years later. The clean share of

new grid supply is $\beta \in [0.63, 0.69]$, high because the solar and wind built in 2024–26 dominate realized additions. We nevertheless take $\beta = 0.5$, below this range, for two reasons. The plants entering the pipeline since 2025 are only 0.39 clean, which suggests that the marginal supply built for data centers is gas-heavier than recent additions. A lower clean share is also the conservative choice, since it raises emissions and exposure to gas prices. Behind-the-meter generation is $\gamma \in [0.44, 0.48]$ of new supply, and we take $\gamma = 0.5$, at the top of this range, with a clean share of $\beta_{\text{BTM}} = 0.1$. No new supply ($\alpha = 0$) and new supply equal to new demand, all on the grid with $\beta = 0.5$ ($\alpha = 1$), bracket the spillover to the rest of the economy.

Fuel prices are national quantity-weighted averages by fuel type from EIA Form 923 (2023), with nuclear fuel costs taken from EPA estimates. Efficiency parameters ν_ι are set to the ratio of annual electricity output to fuel usage from Form 923. Capacity $\bar{C}_\iota$ is each unit's summer capacity, available throughout the year. Wind, solar and hydro units produce their 2023 output, which pins down θ. New fuel generators are assumed to be natural gas combined cycle plants, with efficiency $\bar{\nu} = 0.13$ MWh/MMBtu and fuel cost $\bar{c}$ equal to the 2023 average natural gas price, both from EIA. The two new plants in each balancing area are placed in one of its counties. Their operating cost includes the same labor cost per MWh as existing plants, but we do not count their employment, so new plants do not compete for local workers. The labor coefficient ξ converts operating workers per MW of fossil capacity in 2023 into output per worker at full utilization. For wind and solar, labor coefficients are operating workers per MW of capacity from Zhang (2025).

We use the model's first-order conditions to back out the retail price in each balancing area from its 2023 wage bill and electricity generation, and county productivity A_i from county wages, employment and electricity consumption (EIA Form 861). The markup η_b is the gap between this retail price and the marginal cost of the plant that clears the market at 2023 generation. It absorbs transmission and distribution costs as well as net imports, which the model does not trace across balancing areas.

We make several modeling choices to keep the analysis tractable. First, the equilibrium

is static: we do not model the timing of new plant construction or the path of the data-center build-out over time. Second, we take data-center electricity demand as exogenous and perfectly inelastic, even though hyperscalers may respond to local prices and reliability through siting choices, demand response, and behind-the-meter generation. Third, there is no labor mobility and trade across counties, so any spatial heterogeneity is not attenuated by migration or exports. Fourth, electricity markets clear on annual quantities. We do not model hourly dispatch, storage, ramping constraints, or the time-of-day correlation between renewables output and load. Together with the full-year availability of fossil plants, this means that our supply curves contain more spare capacity than the grid has in its tightest hours, so our price effects understate scarcity. Fifth, balancing area boundaries and inter-area transmission are taken as given. We do not model interconnection-queue congestion or transmission investment. Within a balancing area we assume competitive wholesale dispatch with no market power or strategic withholding, and a constant retail markup that absorbs transmission, distribution, and the regulatory allocation of fixed costs. Allowing markups to respond to demand shocks, as in MacKay and Mercadal (2022), would amplify the incidence on non-data-center ratepayers. Finally, the non-data-center sector is a single Cobb-Douglas producer per county with a common electricity elasticity, even though energy-cost shares vary sharply across industries. Extending the model in these directions is an important avenue for future research.

6 Counterfactuals

6.1 Aggregate

We first discuss how aggregate quantities change with the deployment of new data centers by 2030. Table 4 reports our baseline, $(\alpha, \beta, \gamma) = (0.6, 0.5, 0.5)$, together with two reference cases, no new supply and new supply equal to new demand. In our baseline, data centers pay about \$78B per year for grid electricity. The final good sector loses about \$10B per year in

labor and capital income, and output falls by 0.04% of GDP. Without new supply, the loss would be $31B (0.12% of GDP); with new supply equal to new demand, $5B. The energy sector gains. Electricity sector income, the profits of generators plus the margin utilities earn on each MWh they deliver, rises by $63B, and fuel suppliers' revenue by $22B. The emissions cost is $88B per year, $21B of which come from behind-the-meter gas plants.

To illustrate the mechanisms, we then study how these aggregate effects change when we build new power generation (indexed by parameter α in the notation of the previous section). We discuss two opposite cases. First, we assume that if any power generation is built, it only involves gas plants and no new clean power (i.e. $\beta = 0$ in the notation of the previous section). Second, we assume that any new power built is clean power (i.e. $\beta = 1$). We assume that all new power is connected to the grid and defer the discussion of co-location and "behind the meter" generation to Section 6.3. The dashed line in each panel marks our baseline α.

No new clean power ($\beta = 0$): Figure 13a shows the effects of new data center deployment on five aggregate quantities. The y-axis shows the change in these quantities relative to a baseline of no additional data center electricity demand. The x-axis is the fraction α of new demand that is met by new (local) power generation (in this case, fully by building new gas plants).

We first see that there is a direct cost to data centers: without new supply, they pay an electricity bill of about $115B per year (orange line). The bill does not change much even if new power is built. This would add to any other cost of running the data center sector, although this is outside the model. Second, and perhaps more importantly, we estimate an indirect cost to the final good sector through a loss of output (blue line). In the case of no new power generation ($\alpha = 0$), the output loss amounts to about $31B per year. Interestingly, this amount is not very large relative to total output. It is about 0.12% of US GDP in 2023. This output loss can be reduced to near zero if new power generation is built to offset new

demand.[13]

What explains such effects? There are two main mechanisms. First, in the aggregate, despite the large increase in demand, electricity prices do not increase much. The reason is that there is significant spare capacity in power supply. In virtually all regional electricity markets, the marginal generator is fossil-fuel based (usually gas). Fossil-fuel generators in recent years run significantly below capacity. The most efficient "base load" plants typically run at around 60% of their maximum capacity. The least efficient "peaker plants" run closer to 40%.[14] That means that the existing supply curve is relatively flat around its intersection with the 2023 demand curve. Building new gas plants (increasing α) still matters. Moving from no new supply to $\alpha = 0.6$ with all-gas supply cuts the loss from \$31B to \$13B and the median price increase from 12% to 6%. There can however be large regional differences, which we explore in detail below in Section 6.2. Second, the final good sector output elasticity with respect to electricity prices is relatively small in our 2023 calibration. That is because with Cobb-Douglas technology, this elasticity is related to the expenditures share of electricity in GDP. In the data, this share is small, at around 2% in the aggregate. A small increase in electricity prices combined with a small elasticity explain a small change in output in the final good sector.

In contrast, we also see that the energy sector is a big winner. From its perspective, AI is a boom. Generators and utilities sell more electricity at a moderately higher price. Electricity sector income, generator profits plus the retail margin on the additional MWh delivered, rises by about \$87B (red line). The revenue of upstream fuel suppliers (i.e. gas extraction) also rises significantly, by about \$28B (green line). (The sum is roughly equal to the data center energy bill, since the final good sector is relatively inelastic). These profits are relatively unchanged when new gas plants are built. This is because, in any case, the new electricity demand is met with gas plants in this scenario without new clean power ($\beta = 0$). Omitted

[13] With all-gas new supply, a loss of \$8B remains at $\alpha = 1$. In some balancing areas, most notably Michigan, new combined-cycle plants cost more than the price at which the market clears, so they stay idle and existing plants still serve the new demand. With all-clean new supply, the loss vanishes at $\alpha = 1$.

[14] See EIA.

here are suppliers of energy equipment (i.e. gas turbine manufacturers, construction firms) that also stand to benefit from new power, but that is the focus of Section 6.4.

Finally, there is also a high social cost of emissions. Figure 13 uses a social cost of carbon of \$200/ton of CO2, which amounts to about \$117B (purple line). In terms of magnitude, this is several times the drop in output in the final good sector and the additional profits of the gas extraction sector. When no new power generation is built ($\alpha = 0$), all the new power demand is met with fossil fuel generators, since renewables are inframarginal and cannot ramp up capacity. Emissions decline when new gas plants are built, because modern combined-cycle plants replace older and less efficient coal and gas units. At $\alpha = 1$ the emissions cost falls from \$117B to \$92B per year, still far above what clean supply would achieve.

All new power is clean ($\beta = 1$): Figure 13b shows these aggregate effects in the polar opposite case: any new power built is a renewable plant (according to the 2023 local wind/solar mix). We can understand the aggregate effects of the *composition* of new power by comparing it to the previous figure. The first result is that building renewables instead of gas barely changes the data center bill (orange line), which rises by \$111B with all-gas and \$111B with all-clean supply at $\alpha = 0.6$. Spare fossil capacity keeps the marginal plant, and hence the price, close to the $\beta = 0$ case in most balancing areas. The spillover to the rest of the economy does fall (blue line), from \$13B to \$8B at $\alpha = 0.6$. The difference comes from the few markets where new gas plants would cost more than the incumbents they compete with. In Michigan, the wholesale price rises by 27% with all-gas supply and by 16% with all-clean supply.

The first main effect of renewables is to redistribute profits within the energy sector. Fuel producers lose since less of the new demand is met with gas, since low-cost renewable generation is dispatched first. Electricity generators as a whole win because they have a similar increase in revenue but lower costs (since renewables do not require buying inputs). There is a direct conflict between the fossil fuel and renewable sector for capturing the boom of AI. Of course, building renewables might have a higher construction cost, that is the subject of

Section 6.4. The second main effect of renewables is to limit the increase in emissions. This is natural since renewables are always dispatched and crowd out fossil fuel generation. At $\alpha = 0.6$, all-clean supply avoids $53B per year in emissions costs relative to all-gas supply.

6.2 Spatial distribution

There is, however, substantial spatial variation across the United States. Figure 14 shows the distribution of output loss in the final-good sector across balancing areas, in our 2030 baseline (panel a) and without new power supply (panel b). The x-axis measures the projected electricity demand from data centers in each balancing area. Two patterns stand out. First, the effects are highly heterogeneous. Data center demand is concentrated in a small number of states, and output losses are concentrated there as well. Most of Texas, Virginia/Carolinas, Lower Mississippi Valley together account for 50% of the aggregate output loss in our baseline. Without new supply, Most of Texas, Ohio Valley, Virginia/Carolinas account for 48%.

Second, there is some non-monotonicity. Higher data center demand does not necessarily translate into more output loss. What matters is how much spare capacity is locally available. When data center demand is relatively small (like in most states), output loss is typically near zero. However, Virginia and the Carolinas host the largest data center demand relative to their electricity use, yet in our baseline their output loss relative to the size of their economy is smaller than in Texas or the Lower Mississippi Valley, which face less demand. This highlights the importance of thinking jointly about demand and existing supply.

We can explain these spatial patterns by understanding the effects on regional electricity markets. Figure 15 shows that electricity prices follow the same patterns in terms of heterogeneity and non-monotonicity. In our baseline, the median balancing area experiences a 5% increase in market ("wholesale") electricity prices. This is because regional power supply has enough spare capacity to absorb additional demand.[15]

[15]The Mid-Atlantic is the only balancing area where the baseline price falls. This reflects the discreteness of its supply curve rather than an economic effect, since new supply there lands on a flat segment that our price grid cannot clear exactly.

The largest increases are in Most of Michigan (+23%), Most of Texas (+17%), Texas Panhandle-Oklahoma (+11%). Strikingly, Michigan faces the largest increase in spite of modest data center demand. Its 2023 price is set by inexpensive coal plants, so new demand moves it up a steep part of the supply curve, and new combined-cycle gas plants, whose marginal cost exceeds the resulting price, are not dispatched. This is why Michigan's price still rises by 23% when new supply matches new demand with half of it gas, and falls back only when new supply is clean.[16] Regions with steeper supply curves are more affected by new demand. The same merit-order logic ties our main results together. Flat supply curves over spare fossil capacity keep the average effect small. Steep local curves create hotspots. And new gas plants help only where they are cheaper than the plants they would replace, which is why clean supply matters most in places like Michigan.

6.3 Co-location and "behind the meter" power generation

Given the "speed-to-power" paradigm, more and more data centers plan to co-locate with their own power supply. "Behind the meter" (BTM) power is not connected to the rest of the grid and thus does not add to regional power supply. At the same time, it removes data center demand from the local grid. How would that change the energy cost of AI and data centers? To study this, we vary two additional parameters. First, a fraction γ of new power built is "behind the meter". Second, the share of renewables for this "behind the meter" power is given by β_{BTM}. This can potentially differ from the "front of the meter" β given the data centers' preference for firm 24/7 power.

We first consider the case in which new BTM power is mostly gas ($\beta_{BTM} = 0.1$), as in our data and our baseline. Figure 16a fixes our baseline $\alpha = 0.6$ and $\beta = 0.5$ and varies the BTM share γ. The dashed line marks our baseline $\gamma = 0.5$. More BTM generation shrinks electricity sector income, since data centers buy less power from the grid. Fuel suppliers' revenue and the emissions cost rise, because BTM plants are almost all gas. The emissions cost rises from

[16] The relative change for "retail" prices will be lower, since they include large transmission and distribution costs (that we keep unchanged in our counterfactual).

$80B with all new supply on the grid to $95B with all of it behind the meter. The spillover to the rest of the economy changes only modestly, from $11.0B to $8.2B. This is largely by construction. The grid must absorb the same excess demand, $(1-\alpha)E_b^{\text{dc}}$, either way. The loss falls slightly because behind-the-meter plants serve data center load directly, even in markets such as Michigan where a new grid plant would be too expensive to be dispatched. The only way to limit the increase in fossil fuel use is to have new BTM generation to use a high share of renewables, as illustrated in Figure 16b.

6.4 Capital costs of new power

The previous section showed the effects of building more power supply to meet the new demand. In this section, we estimate the aggregate cost of building this new power supply for different scenarios. Figure 17a shows capital costs depending on how much is built (parameter α) and what is built (parameter β). Capital costs for wind, solar and storage are taken from the 2024 NREL Annual Technology Baseline (ATB), averaged over its 2024–2030 projections, and combined-cycle gas plants cost $2,200 per kW (Section 4). New capacity is sized to deliver the same annual energy as the model's new supply.

The first finding is that building new supply to meet all of the new demand ($\alpha = 1$) is expensive. Renewables are more expensive to build per unit of output, so that meeting all new demand with renewables (and storage) would cost $773B ($\beta = 1$). Gas is cheaper to build per unit of output, so the initial capital cost would drop to $373B. Our baseline supply response costs about $296B (star). This lies below the $\beta = 0.5$ line because its behind-the-meter half is mostly gas. To size the magnitude of these numbers, we compare it to some other benchmark. First, these energy investment costs appear to be an order of magnitude smaller than the cost of data center investment themselves. Second, most of it would have to be spent before 2030, at a time when turbine supply is constrained. This helps explain the recent rise in private "behind the meter" investment by technology firms, which buy speed rather than lower cost.

Is this additional capital investment "worth it"? A back-of-the-envelope comparison helps.

Annualized at a 7% cost of capital over 30 years, the baseline investment costs about $24B per year. In return, it reduces the annual loss in the final good sector from $31B to $10B and the emissions cost from $117B to $88B. The avoided spillover alone roughly pays for the investment, and the avoided emissions make the case clear. Building all-clean rather than all-gas supply to meet all new demand costs an extra $32B per year in annualized capital, against $92B per year in avoided emissions, before counting the fuel it saves.

Finally, we isolate the effects of the recent surge of building new power generation.[17] Supply chain bottlenecks and general inflation mean that upfront capital costs have skyrocketed for all technologies. Averaged over 2024–2030, NREL's current projections are 40% higher for wind, 60% higher for solar with storage, and 69% higher for combined-cycle gas than its pre-2022 projections. Figure 17b compares the two sets of ATB projections for all technologies. Meeting all new demand would cost $290–773B at current ATB costs, against $172–501B at pre-2022 costs. Market prices for gas turbines have risen even faster than the ATB, which is why our main estimates use $2,200 per kW for gas.

6.5 Additional results

Fuel-price risk: higher natural gas prices: Natural gas is the marginal generator in most U.S. balancing areas, so the spillovers documented above are sensitive to the path of future gas prices. The EIA's 2026 Annual Energy Outlook projects natural gas prices for U.S. electricity generators to rise from $3.81/MMBtu in 2025 to an average of about $4.6/MMBtu over 2026–2035, peaking near $5.40/MMBtu in 2032 (2025 dollars).[18] Our calibration uses the 2023 average cost of gas delivered to power plants, $3.80/MMBtu. To quantify how the data-center spillovers depend on this fuel-price path, we re-solve the model with the fuel cost of natural gas generators increased by 20% (to $4.56/MMBtu), leaving all other parameters at their 2023 calibration.

[17] See "Gas-Turbine Prices Surge, Crimping Efforts to Power Data Centers" (*Bloomberg*, 2026); "Gas turbine prices soar 195% as market faces supply-demand crisis" (*Wood Mackenzie*, 2026).

[18] See the EIA's Annual Energy Outlook 2026, Counterfactual Baseline case, Table 3 (energy prices by sector).

The price effect follows the role of gas in each market. On its own, without new data centers, the shock raises wholesale prices most where gas plants set the price for much of the year, in New York (+20%), Most of Florida (+20%), New England (+18%). Where coal, nuclear or hydro sit at the margin, as in Michigan, prices barely move, and the median balancing area sees an increase of only 1%. Gas-fired generation falls by 9% as cheaper plants move up the merit order, and generators gain $3.8B in inframarginal rents. The final good sector loses $10.5B per year, or 0.04% of output. A 20% rise in the price of gas thus costs the rest of the economy about as much as the entire 2030 data center build-out in our baseline. Figure 18 layers this shock on our 2030 baseline. Each column measures the data center spillover relative to the same world without new data centers, so it isolates how the shock changes the cost of the data center build-out. Higher gas prices raise the loss in the final good sector caused by data centers from $10B to $22B per year, since much of the new demand is met by gas plants that set the marginal price. Building clean power hedges this exposure, since renewables do not burn gas. This re-frames the renewable-vs-gas margin as a *hedging* choice in addition to an emissions choice.

Higher electricity share in production: In our baseline specification the output elasticity of electricity in the production function is calibrated to $a = 0.02$ to match the empirical 2023 economy-wide electricity cost share.[19] Long-run electrification scenarios project this share to rise as electrified transport, cooling, etc. grow with some studies projecting it to roughly double by mid-century.[20] We re-solve the model with $a = 0.04$ as a high-electrification benchmark, recalibrating the productivity term A so that 2023 baseline output remains consistent with the new a.

The last two columns of Figure 18 layer this benchmark on our 2030 baseline. Each is measured against an economy with the same higher electricity share but no new data centers, which separates the data center spillover from the effect of the recalibration itself. The data

[19]The output elasticity of electricity a in the Cobb-Douglas $Y = A\,E^a L^b K^{1-a-b}$ governs how sharply output responds to changes in electricity costs. Under cost-minimizing input choice, a equals the electricity share of production costs, which we observe directly.

[20]See e.g. the NREL Electrification Futures Study (Mai et al., 2018) and the IEA's *Net Zero by 2050*.

center spillover to the final good sector is then \$14B per year, against \$10B at 2023 cost shares, and \$29B when combined with higher gas prices. Put differently, the aggregate spillovers we report are conditional on 2023 cost shares, and a more electrified economy is more exposed to the same data center demand surge.

The full announced build-out by 2035: Our 2030 baseline excludes projects scheduled after 2030 and the 181 GW of projects with no announced online date. Including all announced projects raises new data center capacity to 359 GW by 2035, which we treat as an upper bound on the build-out. We calibrate the supply response as for 2030 and take $(\alpha, \beta, \gamma) = (0.5, 0.5, 0.4)$. The 2035 evidence implies $\alpha \in [0.33, 0.68]$, a clean share of new grid supply of $\beta \in [0.55, 0.60]$, and a behind-the-meter share of $\gamma = 0.39$. It is thinner than for 2030, since 45% of grid supply by 2035 must be filled in for years the EIA does not yet see, at the composition of recent, gas-heavier entries. We keep β at its 2030 value, below the evidence for the same reasons as in 2030, and set γ at the evidence.

Table 6 reports the results. The final good sector loses \$37B per year and output falls by 0.14% of GDP, against 0.04% in the 2030 baseline. Wholesale prices rise by 11% in the median balancing area and by up to 68% in the most exposed areas (Northwest US (+68%), Texas Panhandle-Oklahoma (+61%), Most of Texas (+54%)). The most exposed areas are those where large projects without an announced online date concentrate, many of them multi-gigawatt campuses in the Northwest and Texas. There, the existing fleet would run close to full capacity, and prices rise until other users cut back. Meeting all new demand with new supply would cost \$856–1,544B. Internet Appendix Figures IA.5–IA.7 report the counterparts of our main figures for 2035.

7 Implications

To wrap up, we sketch some implications and avenues for future research.

Our analysis suggests that the spillover cost on electricity prices and output is likely to be

moderate *on average*. However, these average effects are likely to mask the amplification of two underlying structural risks. First, the AI and data center boom can endanger grid stability. Spare capacity in power supply plays an important stabilizing role, absorbing demand spikes from exceptional events like extreme weather without triggering price spikes or blackouts. Because data centers require much larger average loads, they reduce the ability of local power grids to handle such extreme episodes. This is especially concerning as the electrification of other sectors will also increase electricity demand in the economy. Quantifying this risk would require a much higher-frequency model of electricity markets, as well as incorporating demand flexibility on the part of data centers. Second, the boom raises exposure to fossil-fuel price volatility, particularly natural gas. The market price of gas is highly volatile and influenced by geopolitical events, and the IEA projects that natural gas will be the single largest source of new U.S. data-center electricity supply through 2030. The large increase in electricity demand from AI therefore raises both the share of hours in which gas plants set the marginal price and the volume of fossil dispatch in those hours, magnifying pass-through from gas price shocks into the rest of the economy. This implies that the economy will be more exposed to fluctuations in fossil fuel prices, especially as electrification grows. Growth in non-fossil power generation can help hedge this risk.

We also relate to the active policy debate over who bears the energy cost of the AI build-out. Historical rules for cost allocation are being re-examined, along with new mechanisms to limit energy spillovers to non-data-center sectors (Feher et al., 2025). One way is to ring-fence the data center economy via "behind-the-meter" privately owned generators or new types of power purchase agreements (PPAs). A non-exclusive alternative is to update the rules of utility regulation. That would involve redistributing costs to the data sector away from other end-users. A recent example is the GS-5 Rate Class policy in Virginia, which made data centers foot the bill for infrastructure. A broader, even more visible initiative is the Ratepayer Protection Pledge, initiated by the White House and signed by the largest technology companies in early

March 2026.[21] In our current baseline, we assume that transmission markups are fixed, but extending that is an important avenue for future research. The underlying economics are subtle and the recent evidence on retail rates is mixed (Scalera et al., 2026; Watten et al., 2026; Meeks et al., 2026).

Finally, the AI boom is likely to alter the path of the energy transition. Exactly how remains a key question for future research. In the short run, there are concerns that data centers are turning to inefficient gas generators (like repurposed jet engines), such that using more fossil fuels seems unavoidable. At the same time, the massive increase in electricity demand coming from data centers might stimulate technological innovation in the long run. We are witnessing an acceleration of investments in new forms of "firm" clean power such as geothermal and nuclear. Now that demand is strong, directed technological change might make these new power sources viable at scale earlier than previously thought. Another margin is the potential adoption of AI by firms and grids to improve energy efficiency by better balancing demand and supply in real time.

[21] See "Ratepayer Protection Pledge" (*The White House*, 2026).

References

F. E. Alvarez, D. Argente, J. Chow, and D. Van Patten. Data centers and local economies in the age of AI: A shift–share approach. NBER Working Paper 35194, National Bureau of Economic Research, 2026.

C. Arkolakis and C. Walsh. Clean growth. NBER Working Paper 31615, National Bureau of Economic Research, Aug. 2023. Revise and resubmit at the Journal of Political Economy.

C. Arkolakis and C. Walsh. The economic impacts of clean energy. *Brookings Papers on Economic Activity*, 55(2):161–238, 2024.

M. Benetton, G. Compiani, and A. Morse. When technology processors come to town: High electricity use spillovers to the local economy. *NBER Working Paper*, 31312, 2025.

S. Cicala. When does regulation distort costs? Lessons from fuel procurement in US electricity generation. *American Economic Review*, 105(1):411–444, 2015. doi: 10.1257/aer.20131377.

S. Cicala. Imperfect markets versus imperfect regulation in US electricity generation. *American Economic Review*, 112(2):409–441, 2022. doi: 10.1257/aer.20172034.

L. Davis and C. Hausman. Market impacts of a nuclear power plant closure. *American Economic Journal: Applied Economics*, 8(2):92–122, 2016.

T. P. Fang and S. Greenstein. Where the cloud rests: The economic geography of data centers. Working Paper 21-042, Harvard Business School, Sept. 2020. Revised August 2025.

A. Feher, E. Garcia-Appendini, and R. Mihet. Is ai trained on public money? evidence from us data centers. *Evidence from US Data Centers (September 05, 2025). Swiss Finance Institute Research Paper*, (25-73), 2025.

A. Gargano and M. Giacoletti. Subsidizing the cloud: U.S. state incentives to data centers. Working Paper SSRN 5881105, 2025.

D. Ham, O. Kay, and C. Hausman. Transmission lowers US generation costs, but generator incentives are not aligned. *Proceedings of the National Academy of Sciences*, 123(9):e2524463123, 2026. doi: 10.1073/pnas.2524463123.

C. Hausman. Power flows: Transmission lines, allocative efficiency, and corporate profits. *American Economic Review*, 115(8):2574–2615, 2025. doi: 10.1257/aer.20240276.

C. Hausman and R. Kellogg. Welfare and distributional implications of shale gas. *Brookings Papers on Economic Activity*, 46(1):71–139, 2015.

A. Heiss, Z. Sautner, and T. Schmid. Powering AI: How do data centers affect renewable energy investment? Technical report, March 2026. Swiss Finance Institute Research Paper No. 26-29.

IEA. Energy and AI. Technical report, International Energy Agency (IEA), Paris, 2025.

M. T. Kacperczyk and D. Licher. Infrastructure capacity, risk, and firm value: Evidence from us electricity tightness. *HKU Jockey Club Enterprise Sustainability Global Research Institute Paper*, (2025/136), 2025.

O. Kay, R. Reaser, and R. Taylor. Processing power: The effect of data centers on wholesale electricity markets. Working Paper 2606, Federal Reserve Bank of Dallas, 2026.

A. MacKay and I. Mercadal. Deregulation, market power, and prices: Evidence from the electricity sector. Working Paper 2022-008, MIT Center for Energy and Environmental Policy Research, April 2022.

R. Meeks, J. Pless, Z. Qi, and Z. Wang. Who pays for growth? evidence from data centers and the grid. Working Paper 6967558, SSRN, 2026.

N. Z. Muller. Measuring the impact of data centers in the United States economy: Monetary damage from air pollution and greenhouse gas emissions. NBER Working Paper 35100, National Bureau of Economic Research, 2026.

D. Mytton, M. Ashtine, S. Wheeler, and D. Wallom. Stretched grid? Managing data center energy demand and grid capacity. *Oxford Open Energy*, 2:oiad014, 2023. doi: 10.1093/ooenergy/oiad014.

C. Scalera, V. Bosetti, and F. Pecci. Large-load electricity customers and spillovers on household prices: Evidence from U.S. data centers. CEPR Discussion Paper 21854, Centre for Economic Policy Research, 2026.

A. Shehabi, S. J. Smith, A. Hubbard, A. Newkirk, N. Lei, M. A. B. Siddik, B. Holecek, J. G. Koomey, E. R. Masanet, and D. A. Sartor. 2024 United States Data Center Energy Usage Report. Technical Report LBNL-2001637, Lawrence Berkeley National Laboratory, Berkeley, California, 2024.

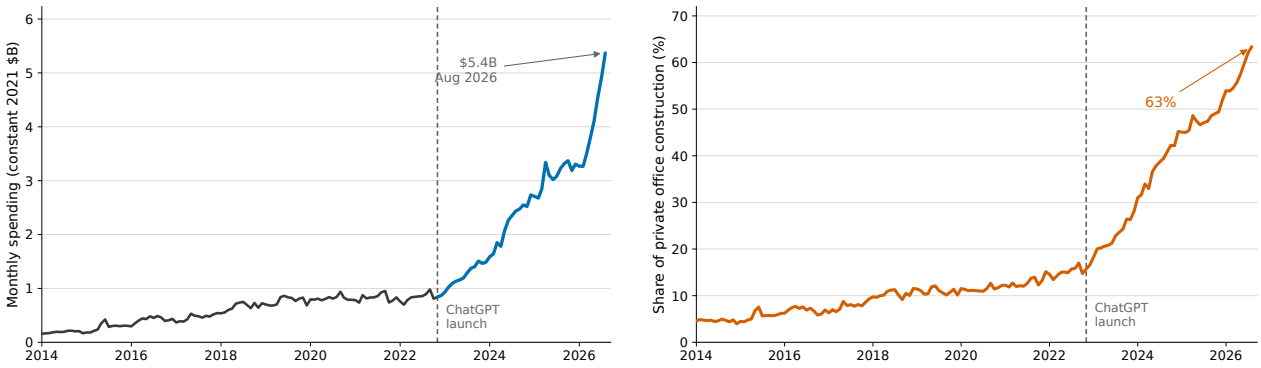
S. Van Nieuwerburgh. Financing the AI buildout. Working paper, Columbia Business School, March 2026. Draft: March 20, 2026.

A. Watten, J. Bistline, and G. Blanford. Have data centers raised your electric bill? causal evidence

from the United States. arXiv:2606.19777, 2026.

Y. Zhang. Clean energy subsidies and capital misallocation. Technical report, 2025.

Figure 1: Data center construction spending in the United States

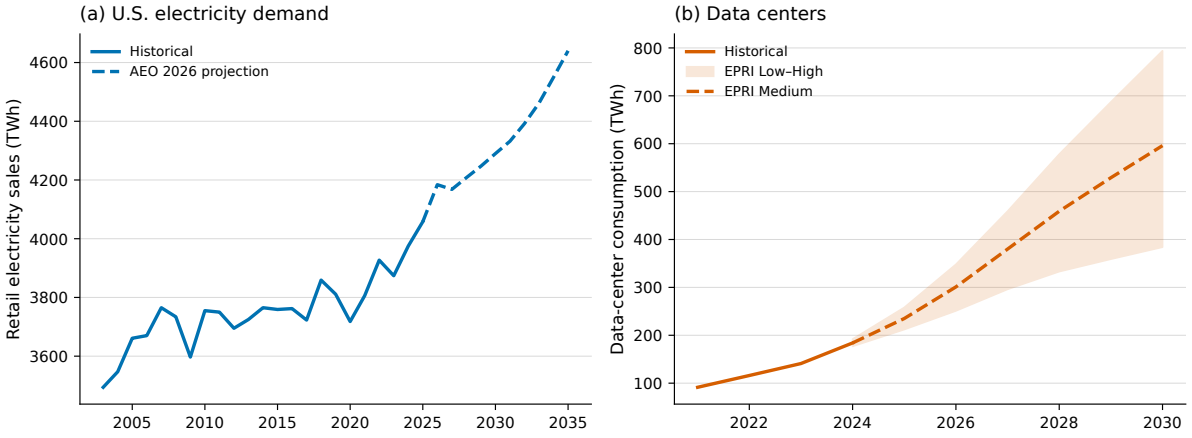


(a) Monthly spending (constant 2021 $B)

(b) Share of private office construction

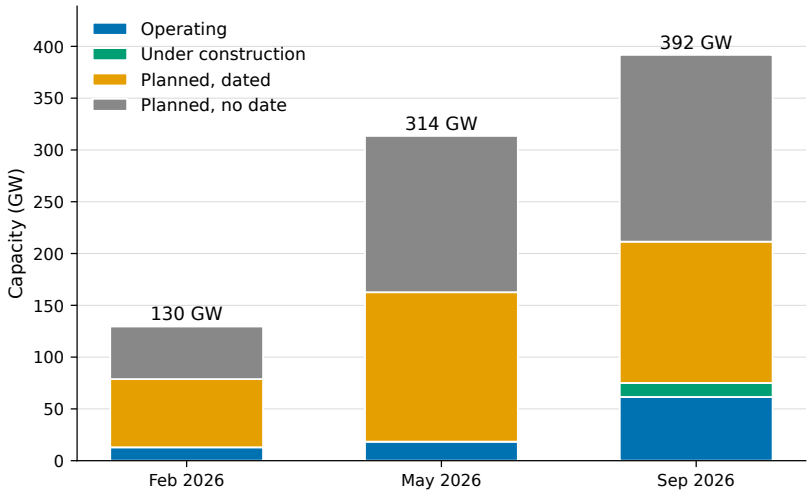
Notes: Monthly value of data center construction put in place, U.S. Census Bureau (C30), 2014 to August 2026. Panel (a) is in constant 2021 dollars using the Our World in Data deflator. Panel (b) divides nominal data center spending by nominal private office construction, the Census category in which data centers are reported. The dashed line marks the release of ChatGPT (November 2022).

Figure 2: U.S. electricity demand and data center consumption



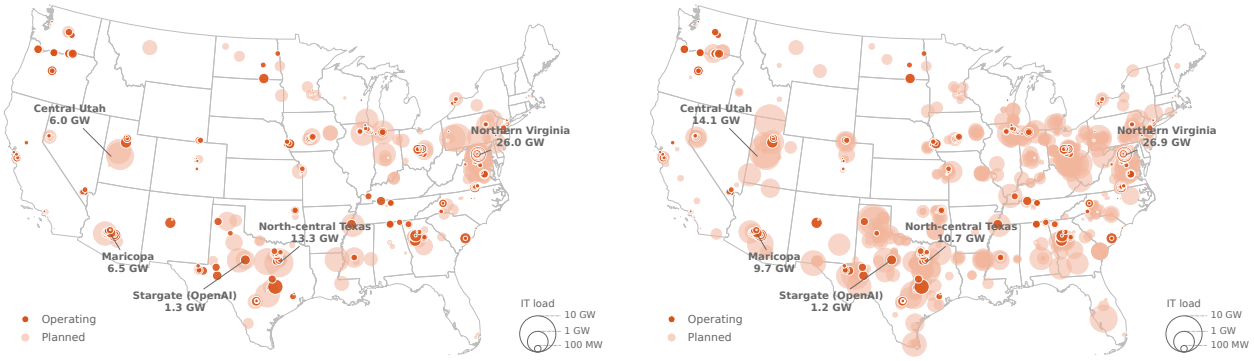
Notes: Panel (a) shows U.S. retail electricity sales, historical through 2025 (EIA Monthly Energy Review) and the EIA's Annual Energy Outlook 2026 Counterfactual Baseline case thereafter. Panel (b) shows data center electricity consumption, historical and in the EPRI *Powering Intelligence* (2026) scenarios to 2030. The shaded band spans the Low and High scenarios.

Figure 3: Data center capacity in three Cleanview pulls, 2026



Notes: Total capacity of U.S. data center projects by status in the February and May 2026 exports and the September 2026 update, excluding canceled projects. Planned projects are split by whether they have an announced online date. The update covers more existing and smaller data centers than the earlier exports, so part of the increase reflects coverage.

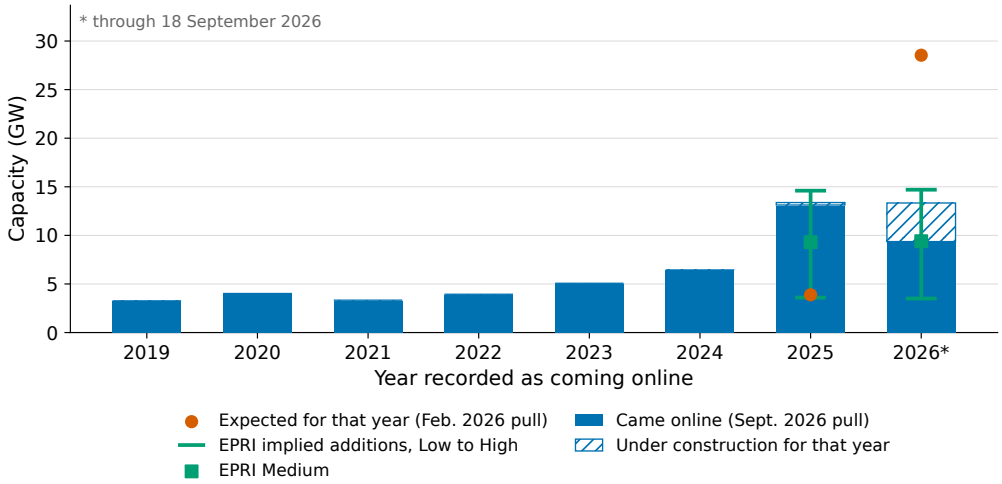
Figure 4: Planned data center pipeline, February versus September 2026



(a) February 2026 (b) September 2026

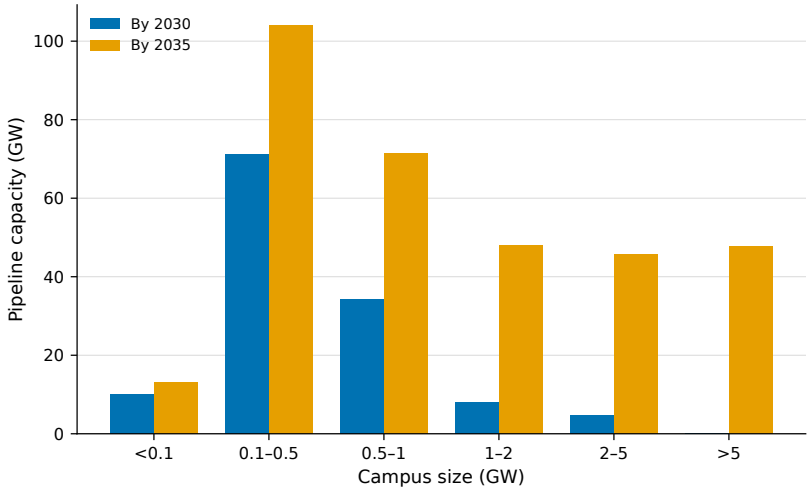
Notes: Both maps show the same operating data centers (online by the end of 2025, from the Cleanview update, dark circles) and the pipeline in each pull (light circles), meaning projects planned, under construction, or that came online from 2026 onward. Projects with no announced online date are included, and projects dated after 2035 are excluded. Circle area is proportional to IT load.

Figure 5: Data center capacity coming online, and the 2026 announcements



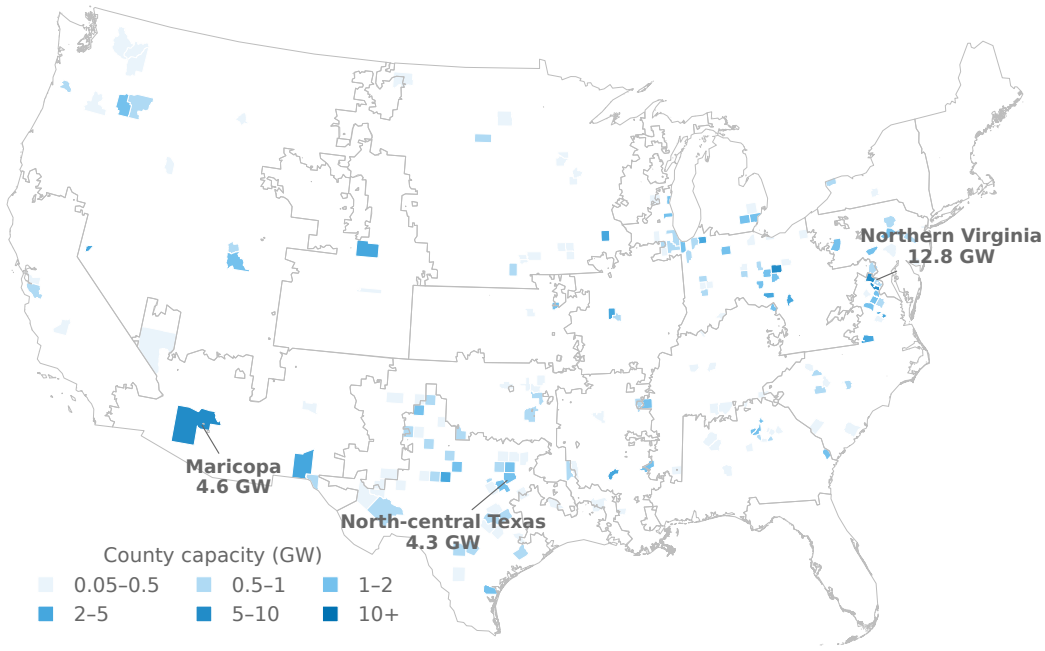
Notes: Bars show capacity recorded as coming online in each year in the Cleanview update, with capacity under construction for 2026 hatched. Dots show the capacity the February 2026 export expected to come online in that year. Green markers show implied annual additions in EPRI's *Powering Intelligence* (2026) scenarios (Medium, with the Low–High range). 2026 is through 18 September.

Figure 6: Planned data center capacity by campus size



Notes: Capacity of planned and under-construction data center campuses due online by 2030 and by 2035, by campus size. Projects with no announced online date are assigned to 2035. A campus groups the records Cleanview lists separately for each phase or building of the same developer at the same location. Cleanview update.

Figure 7: New data center capacity by 2030, by county and balancing area



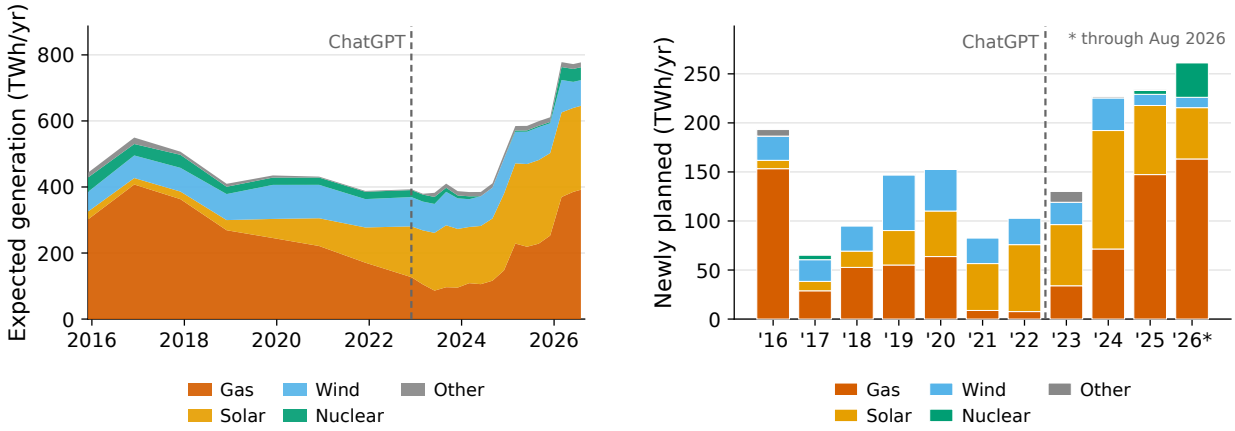
Notes: County shading shows new data center capacity by 2030, from projects that came online from 2024 onward or are scheduled to come online by 2030 (Cleanview update). This is the demand that enters our 2030 baseline. Lines mark the 20 balancing areas (EPA eGRID subregions) that define electricity markets in the model. Internet Appendix Figure IA.4 shows the full announced build-out by 2035.

Table 1: County covariates: low vs. high projected data center demand (2030).

Variable	**Low-DC mean** ($N = 3,035$)	**Low-DC SD**	**High-DC mean** ($N = 86$)	**High-DC SD**	**Diff**
Projected DC capacity (MW)	10	59	1,470	1,232	1,460***
GDP per capita ($k)	107.1	1,613.5	87.6	81.2	-19.5
Employment (thousands)	43.7	153.9	238.4	496.4	194.7***
Avg annual pay ($k)	51.6	11.4	59.7	12.9	8.1***
Unemployment rate (%)	3.57	1.21	3.57	0.85	-0.00
Wind capacity (MW)	43.9	168.5	99.7	308.7	55.8
Solar capacity (MW)	22.0	131.5	71.5	238.8	49.5*
Natural gas capacity (MW)	163.7	561.4	850.4	1,917.3	686.8***
Residential electricity price ($/kWh)	0.133	0.025	0.130	0.019	-0.003
Number of utilities	3.77	1.87	4.20	2.43	0.44
Avg wind CF	0.401	0.081	0.407	0.075	0.006
Avg solar CF	0.262	0.027	0.267	0.032	0.005
Dist. to load (km)	132.9	102.4	102.3	89.4	-30.6***
Population (thousands)	94.8	300.6	441.4	892.7	346.6***
Population density (per km^2)	101.6	701.0	211.0	332.8	109.3***
Land area (km^2)	2,466	3,381	2,588	3,799	122

Notes: High-DC counties are the top ranked counties whose cumulative 2030 projected data center capacity reaches $\geq 80\%$ of the U.S. total (80.3% of 157,468 MW, $N = 86$); Low-DC counties are the remaining $N = 3,035$. Difference is High - Low. Two-sample t-test; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. *Sources*: County-level 2030 DC capacity from CleanView (September 2026). GDP and GDP per capita: BEA (2023). Employment and avg. pay: BLS QCEW (2023). Unemployment rate: BLS LAU (2022). Wind, solar, and natural-gas capacity: EIA Form 860 (2022). Residential price and number of utilities: EIA Form 861 (2022). Wind/solar capacity factors and distance to load: NREL ReEDS supply curves (2030 reference). Population: Census Bureau estimates (2021, latest available).

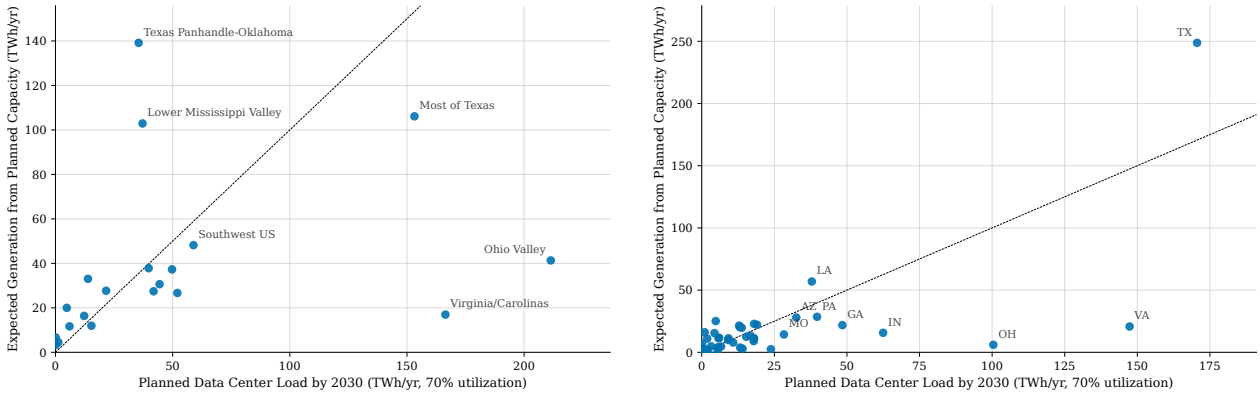
Figure 8: Planned power generation in the United States



(a) Expected generation from planned plants (b) Generation entering the pipeline each year

Notes: Generators listed as planned in EIA-860M, by energy source, measured by expected annual generation. Capacity is converted with national capacity factors by source, namely 64% for gas (new combined cycle), 24% for solar, 37% for wind and 90% for nuclear. Storage adds no net generation and is excluded. Panel (a) shows the pipeline at each release (December releases through 2022, quarterly since 2023, latest August 2026). Panel (b) shows units entering the pipeline in each year, dated by a unit's first appearance, with 2026 through August. Coal and hydro are included in Other. The dashed line marks the release of ChatGPT.

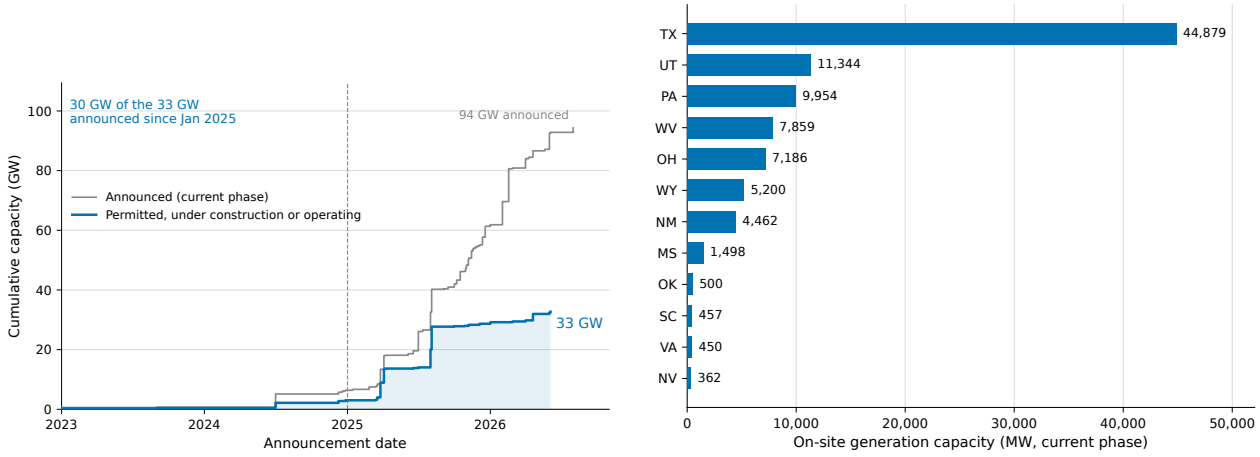
Figure 9: Planned data centers and planned power generation across balancing areas and states



(a) Balancing areas (b) States

Notes: Each point is a balancing area (panel a) or a state (panel b). The horizontal axis is data center electricity demand by 2030 (Cleanview update, 70% utilization). The vertical axis is expected annual generation from all generators planned in the August 2026 EIA-860M, using capacity factors of new plants by region and source (zero for storage). The dashed line is equality.

Figure 10: Behind-the-meter power plants for data centers



(a) Cumulative capacity by announcement date (b) Capacity by state

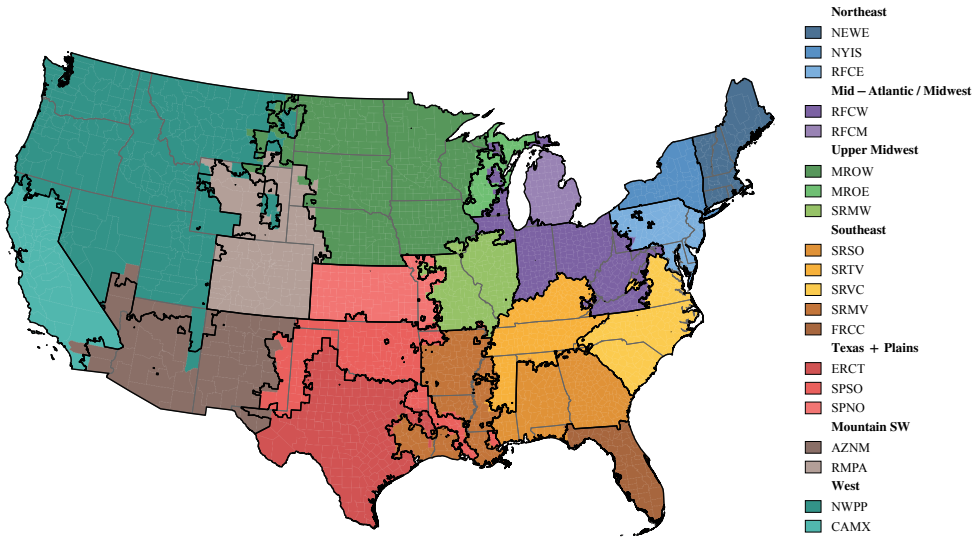
Notes: On-site or dedicated generation built primarily to serve a specific data center campus, from our audited list (October 2026). Canceled projects and sub-projects of larger campuses are excluded. Capacity is the current phase, not the full build-out. In panel (a), the grey line is all announced capacity and the blue line is the part covered by an issued permit, under construction or operating. Panel (b) shows the twelve states with the most capacity.

Table 2: Largest announced behind-the-meter power plants serving U.S. data centres

#	**Project**	**Developer / tenant**	**Location**	**Capacity (MW)**	**Full build (MW)**	**Status**
1	Stratos (Wonder Valley)	Stratos Project / MIDA	Box Elder, UT	9,000	–	Announced
2	GW Ranch	Pacifico Energy / Amazon	Pecos, TX	7,654	–	Construction
3	Nexus Hubbard	Nexus / Anthropic	Hill, TX	7,230	–	Mixed
4	HyperGrid (Project Matador)	Fermi America	Carson, TX	6,000	11,280	Construction
5	Homer City Energy Campus	Knighthead / Amazon	Indiana, PA	4,396	–	Construction
6	Shippingport	Frontier Group / Aligned	Beaver, PA	3,600	–	Mixed
7	Kilby	Chevron / Microsoft	Reeves, TX	2,868	5,000	Announced
8	Cheyenne Power Hub	Tallgrass / Google	Laramie, WY	2,700	–	Construction
9	Vantage Frontier	VoltaGrid / Oracle–OpenAI	Shackelford, TX	2,582	–	Construction
10	Project Jupiter	Acoma / OpenAI–Oracle	Doña Ana, NM	2,462	–	Mixed
11	Adams Fork, site I	TransGas Development	Mingo, WV	2,457	–	Permitted
12	Adams Fork, site II	TransGas Development	Mingo, WV	2,457	–	Permitted
13	Dove Creek	Beacon Data Centers	Tom Green, TX	2,409	–	Announced
14	Monarch Compute Campus	Nscale / Fidelis	Mason, WV	2,160	8,160	Construction
15	Circe West Texas	Circe Energy	Ward, TX	2,000	–	Announced
Top 15 subtotal				59,976		63%
All 76 projects				94,639		
of which operating				2,458		3%
of which under construction				29,028		31%

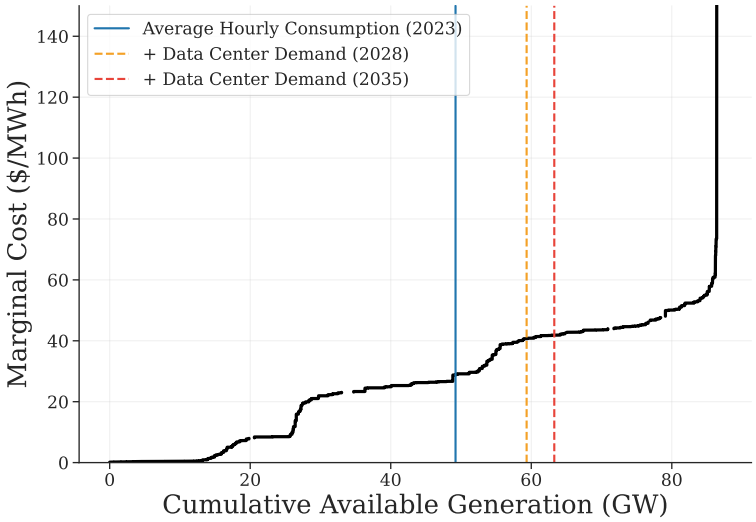
Notes: Behind-the-meter (BTM) generation is on-site capacity built to serve a specific data-centre campus, off-grid or with limited grid export. *Capacity* is the nameplate of the phase actually permitted or formally announced; *Full build* reports the larger campus ambition where one is marketed, and is shown as – where the two coincide. Utility-owned plants contracted to a named data centre (Entergy Louisiana for Meta Hyperion, Georgia Power for the Effingham County Stargate campus) are grid-connected and therefore excluded. Capacity is generation nameplate, not data-centre IT load. Own compilation as of 30 September 2026, combining the Cleanview behind-the-meter report, the Global Energy Monitor Global Oil and Gas Plant Tracker (August 2026 release), the Environmental Integrity Project inventory of proposed BTM gas plants (30 April 2026) and state air-permit records.

Figure 11: Spatial units of analysis: balancing areas and counties



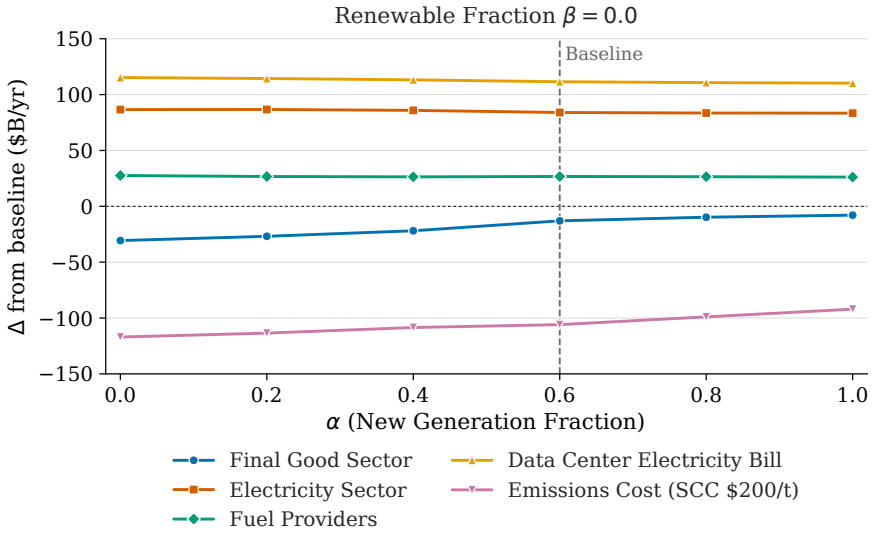
Notes: County-level map of the contiguous United States. The 20 balancing areas are shaded in different colours. Bold black lines denote balancing-area boundaries; thin gray lines denote state boundaries. *Sources*: EPA eGRID 2023 (subregion definitions); Census 2023 (county and state boundaries).

Figure 12: Local Supply Curve: ERCT

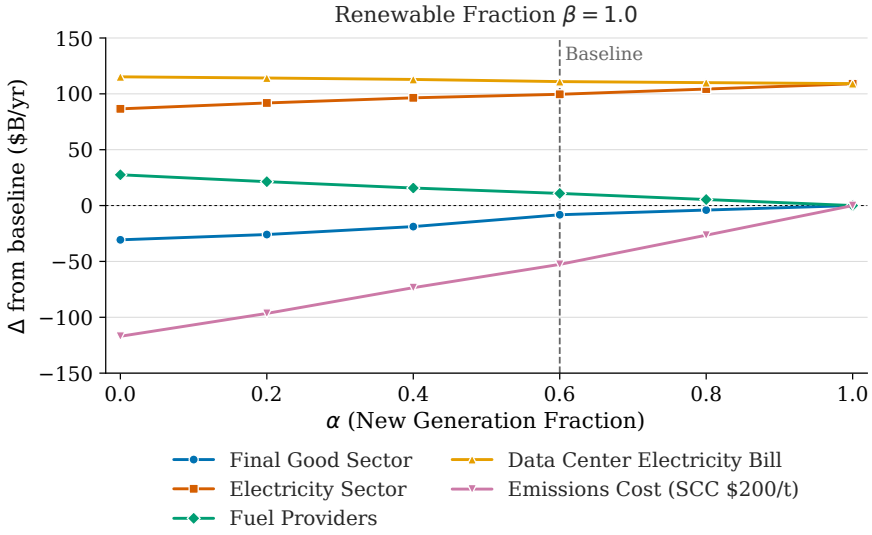


Notes: This figure shows the local supply curve for the EPA EGrid region *ERCT*, which covers most of Texas. The (step-wise) supply curve uses estimates of generator-level marginal costs, which comprise labor and fuel costs. Each step's width reflects the generator's available capacity, which is measured using summer nameplate capacity. Vertical lines indicate average hourly electricity consumption in 2023 (EIA), with projected data center demand added for 2028 and 2035 (Cleanview).

Figure 13: Economic impact of data-center expansion under the projected 2030 demand



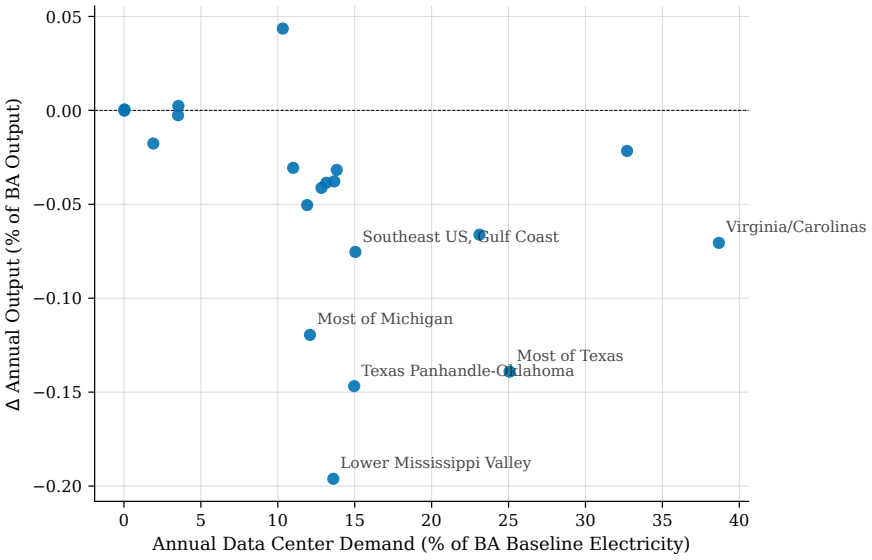
(a) All-gas new build ($\beta = 0$)



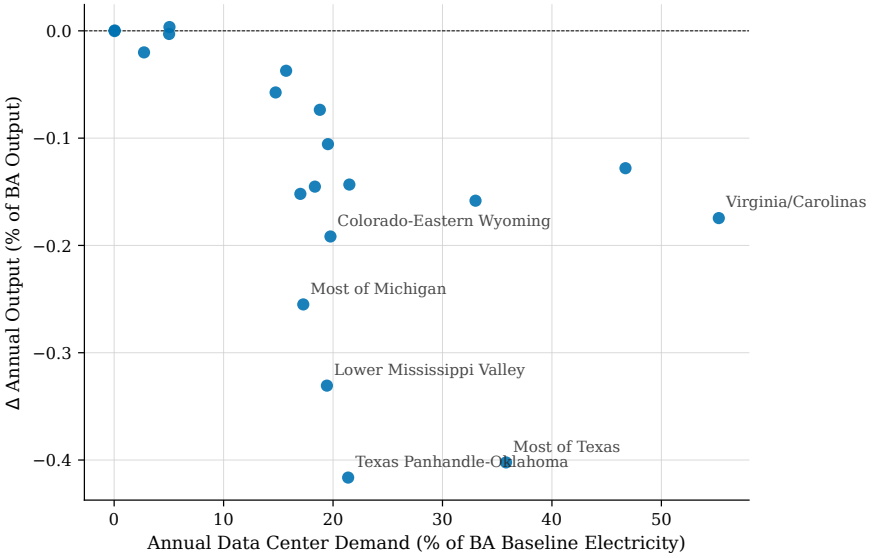
(b) All-renewable new build ($\beta = 1$)

Notes: Change in five aggregate quantities relative to the 2023 baseline as a function of the new-generation fraction α under the 2030 projected data-center demand, with all new generation on the grid ($\gamma = 0$). The dashed line marks our baseline $\alpha = 0.6$. Panel (a) holds the renewable share of new build at $\beta = 0$ (all-gas); panel (b) at $\beta = 1$ (all-renewable). *Final Good Sector* is the sum of labor income and other capital income. *Electricity Sector* is generator profits plus the retail and transmission margin η_b on electricity delivered. *Fuel Providers* is fuel suppliers' revenue, equal to generators' fuel spending. *Data Center Electricity Bill* is data center grid purchases at the retail price. *Emissions Cost* is the monetized CO_2 externality computed by applying fuel-specific emissions factors to each generator's fuel use and valuing emissions at a social cost of carbon of $200 per metric ton.

Figure 14: Output change versus data-center demand across balancing areas



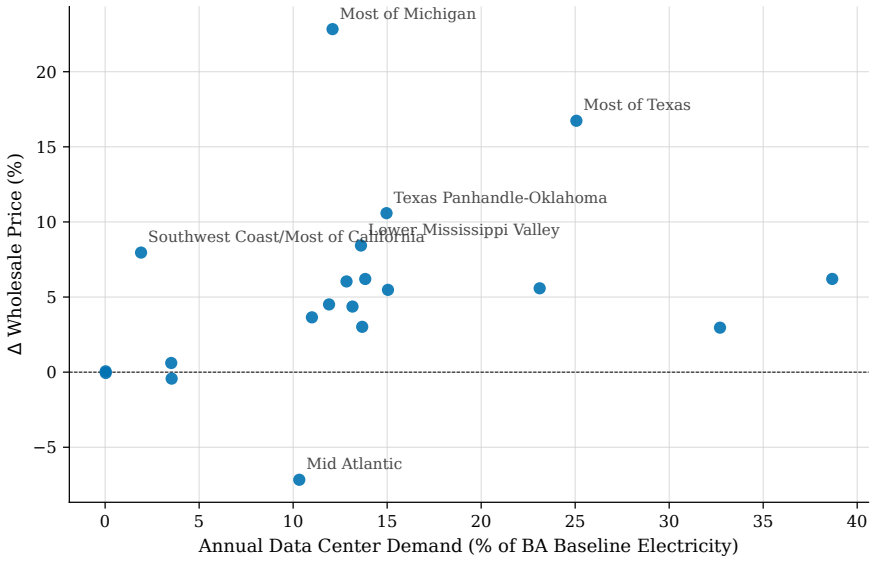
(a) Baseline, 2030



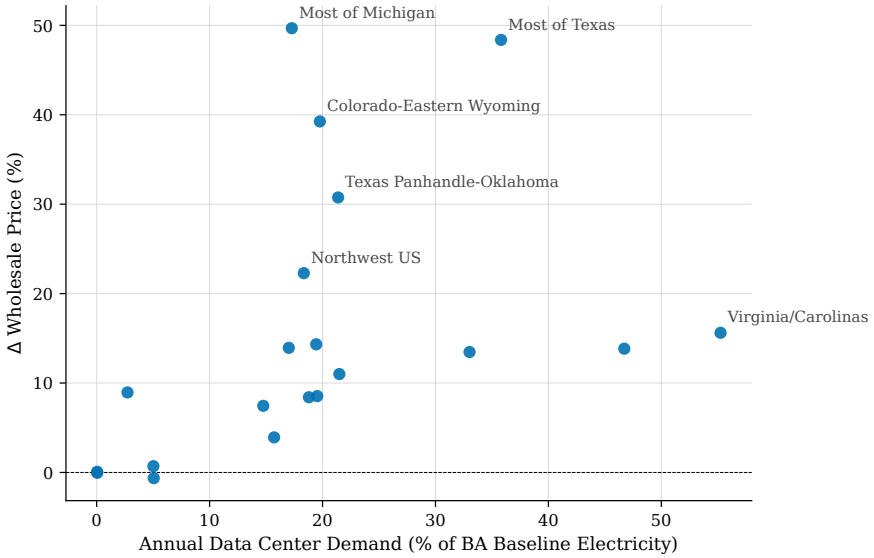
(b) No new power supply ($\alpha = 0$), 2030

Notes: Each point is a balancing area, plotting the change in annual non-data-center output (% of BA output) against new on-grid data center electricity demand (% of BA baseline electricity). Panel (a) is our 2030 baseline, $(\alpha, \beta, \gamma) = (0.6, 0.5, 0.5)$, in which behind-the-meter generation serves part of data center demand off the grid. Panel (b) assumes no new power generation ($\alpha = 0$). Labelled balancing areas are those whose output loss exceeds a given threshold.

Figure 15: Wholesale electricity-price change versus data-center demand across balancing areas



(a) Baseline, 2030



(b) No new power supply ($\alpha = 0$), 2030

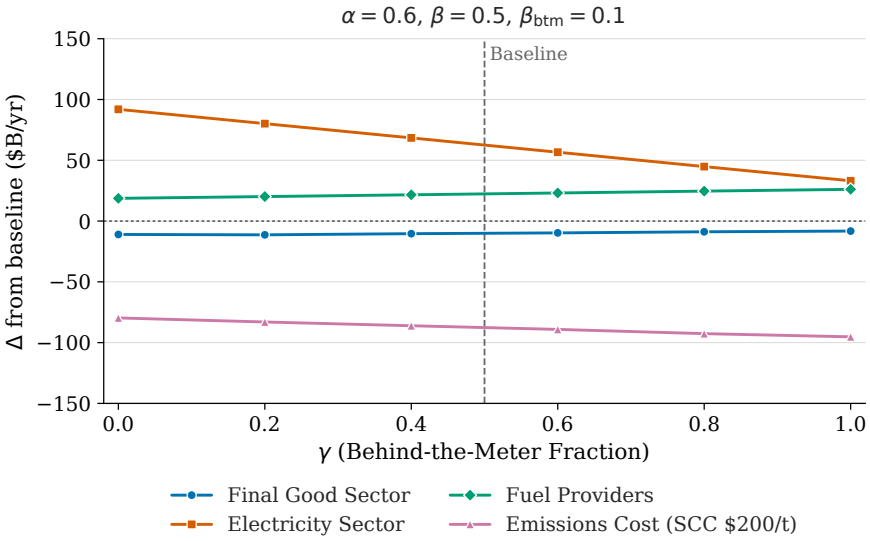
Notes: Each point is a balancing area, plotting the percent change in wholesale electricity price against new on-grid data center electricity demand (% of BA baseline electricity). Panel (a) is our 2030 baseline, and panel (b) assumes no new power generation ($\alpha = 0$). Labelled balancing areas are those whose price change exceeds a given threshold.

Table 3: Shifting data-center activity to New York and California, 2030

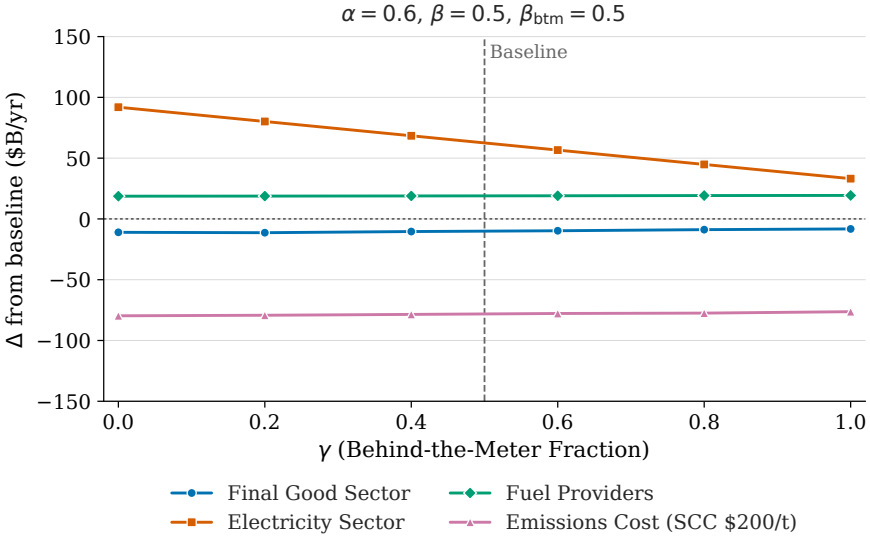
	DC bill vs. actual locations (%)	Wholesale price change		Final-good loss ($B/yr)
		New York (%)	California (%)	
No new supply ($\alpha = 0$)	44	21	53	31
Baseline	47	9	38	11

Notes: 20% of 2030 data center demand in every other balancing area is moved, in equal parts, to NYISO and CAISO, so total demand is unchanged. The first column compares the data center electricity bill with the same scenario at actual locations. The bill is at retail prices, so it reflects the high retail markups in New York and California. Price changes and the final-good loss are relative to the 2023 baseline. With actual locations, the final-good loss is $31B without new supply and $10B in the baseline.

Figure 16: Behind-the-meter generation, 2030 ($\alpha = 0.6$, $\beta = 0.5$)



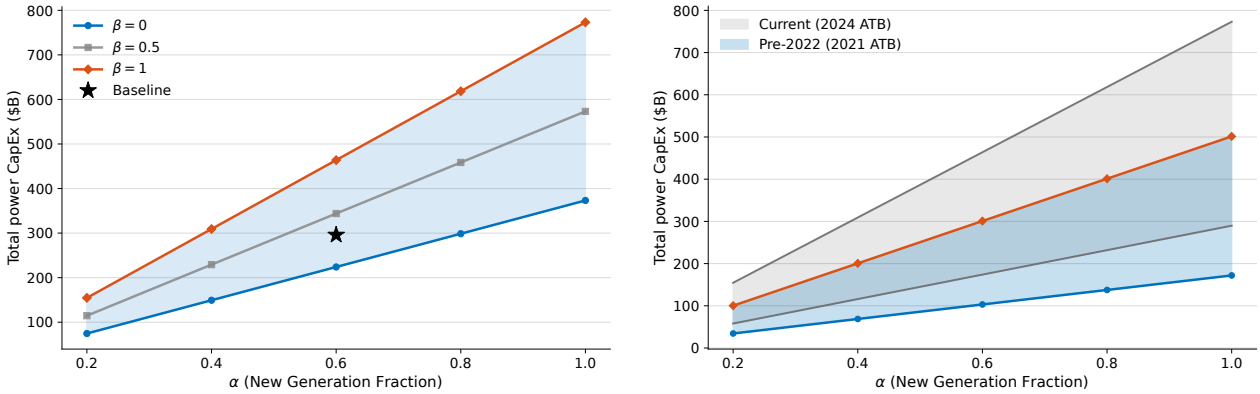
(a) Mostly gas BTM ($\beta_{\text{BTM}} = 0.1$)



(b) Mixed BTM ($\beta_{\text{BTM}} = 0.5$)

Notes: Change in aggregate quantities relative to the 2023 baseline, varying the BTM fraction γ at our baseline $\alpha = 0.6$ and $\beta = 0.5$ under the 2030 projected data-center demand. The dashed line marks our baseline $\gamma = 0.5$. The BTM fraction γ is the share of new generation that is off-grid. β_{BTM} is the share of BTM generation that is renewables. Panel (a) takes $\beta_{\text{BTM}} = 0.1$, as in our baseline, and panel (b) takes $\beta_{\text{BTM}} = 0.5$. *Final Good Sector* is the sum of labor income and other capital income. *Electricity Sector* is on-grid generator profits plus the retail and transmission margin η_b on electricity delivered. *Fuel Providers* is fuel suppliers' revenue from on-grid and BTM generators. *Emissions Cost* is the monetized CO_2 externality computed by applying fuel-specific emissions factors to each generator's fuel use and valuing emissions at a social cost of carbon of $200 per metric ton.

Figure 17: Capital Costs of New Power Generation

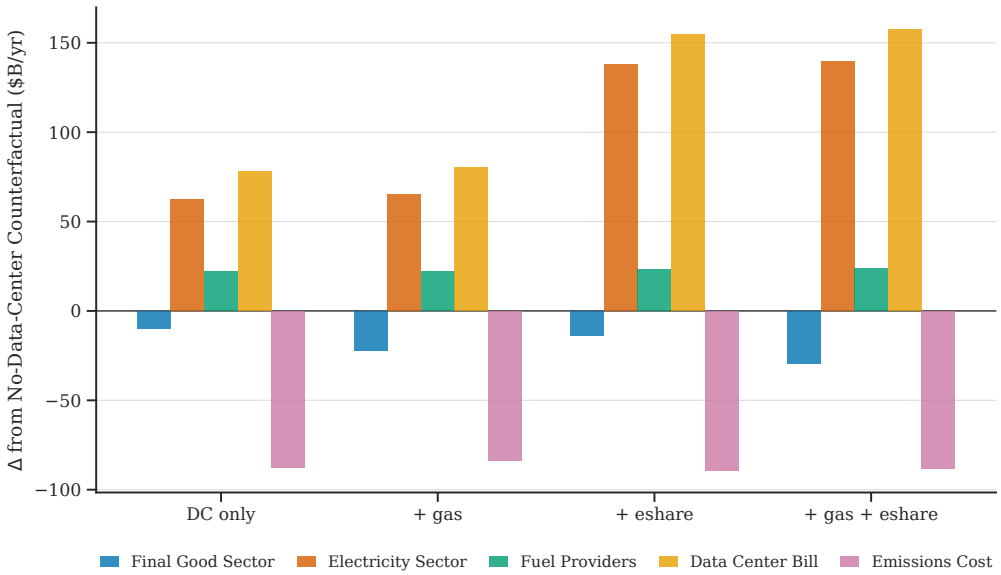


(a) Current capital cost estimates

(b) Pre-2022 capital cost estimates

Notes: This figure shows the range of total capital expenditure (CapEx) required to build new electricity generation capacity to meet data center demand, varying the parameters α (the share of demand met by new generation) and β (the share of new generation capacity built from renewables, with the remainder supplied by natural gas). The horizontal axis reports α, while the vertical axis reports total generation investment in billions of U.S. dollars. Investment is what is needed to meet 2030 data center demand. Capacity is sized to deliver the model's new energy at balancing-area capacity factors. Panel (a) uses current capital cost estimates, namely the 2024 NREL ATB for wind, solar and storage and $2,200 per kW for combined-cycle gas. The star marks our baseline, in which the behind-the-meter half of new supply is mostly gas ($\beta_{\text{BTM}} = 0.1$). Panel (b) compares the 2024 and pre-2022 (2021) ATB projections for all technologies.

Figure 18: Structural risks layered on the 2030 baseline



Notes: Change in five aggregate quantities caused by 2030 data center demand under our baseline supply response, $(\alpha, \beta, \gamma) = (0.6, 0.5, 0.5)$ with $\beta_{\text{BTM}} = 0.1$. Each column compares the economy with and without new data centers under one structural risk. The columns are no additional risk (2023 calibration), natural gas at \$4.56/MMBtu, an electricity output elasticity of $a = 0.04$ instead of 0.02, and both. Measuring each column against its own no-data-center counterfactual isolates the data center spillover from the direct effect of the risk itself. With the higher electricity share, the recalibration implies a higher retail electricity price, which mechanically raises the data center bill and electricity sector income in the last two columns. *Final Good Sector* is the sum of labor income and other capital income. *Electricity Sector* is generator profits plus the retail and transmission margin η_b on electricity delivered. *Fuel Providers* is fuel suppliers' revenue, equal to generators' fuel spending. *Data Center Bill* is data center grid purchases at the retail price. *Emissions Cost* values CO_2 at \$200 per metric ton.

Table 4: Aggregate effects of 2030 data center demand

	No new supply ($\alpha = 0$)	Baseline	New supply = demand ($\alpha = 1$)
Change from 2023, $B per year			
Data center electricity bill	115	78	110
Final good sector loss	31	10	5
Electricity sector gain	87	63	96
Fuel providers gain	28	22	13
Emissions cost	117	88	48
Output loss (% of GDP)	0.12	0.04	0.02
Wholesale price, median balancing area (%)	12	5	0
Wholesale price, largest increase (%)	50	23	23

Notes: Changes relative to the 2023 calibration without new data center demand. The baseline is $(\alpha, \beta, \gamma) = (0.6, 0.5, 0.5)$. The right column builds new supply equal to new demand, all on the grid, with $\beta = 0.5$. In the baseline, a share γ of new supply is built behind the meter with renewable share $\beta_{\text{BTM}} = 0.1$. The final good sector loss is the fall in labor and capital income outside data centers. The electricity sector gain is generator profits plus the retail and transmission margin η_b on electricity delivered. The fuel providers gain is the increase in fuel suppliers' revenue, including fuel for behind-the-meter plants. The data center bill covers grid purchases at the retail price, and the emissions cost includes $21B from behind-the-meter gas plants. Output loss is the change in county output summed across the country, as a share of total 2023 output in the model. The largest wholesale price increases in the baseline are in Most of Michigan (+23%), Most of Texas (+17%), Texas Panhandle-Oklahoma (+11%).

Table 5: Evidence on the supply response to data center demand

	2030		2035	
TWh/yr	Face value	Adjusted	Face value	Adjusted
Data-center demand	966	966	2,200	2,200
Planned grid generation (EIA-860M)	697	505	772	564
Not-yet-filed years	133	98	989	726
Retirements since 2023	220	138	311	183
Non-DC load growth	125	337	214	578
Behind-the-meter generation	381	264	580	284
α (new supply / DC demand)	0.90	0.41	0.83	0.37
β (clean share of grid supply)	0.45	0.49	0.42	0.44
γ (behind-the-meter share)	0.44	0.67	0.32	0.35
Chosen (α, β, γ)	0.6 / 0.5 / 0.5		0.5 / 0.5 / 0.4	

Notes: Annual energy (TWh). Grid generation online since 2024 covers units that started operating from January 2024 to August 2026, and planned grid generation the units in the August 2026 EIA-860M due by the horizon. Capacity factors are 64% (combined-cycle gas), 10% (simple-cycle gas), 24% (solar), 37% (wind), 90% (nuclear) and 0 (storage). Planned years after 2028 are topped up to the 2027–28 average at the clean share of entries since 2025. Retirements are those realized since 2024 plus those planned by the horizon, half of the latter deferred in the adjusted case. Non-DC load growth is U.S. retail sales growth (EIA) minus data center growth (EPRI to 2025, AEO servers thereafter, $\times 1.4$ for cooling at face value). Behind-the-meter generation is our audited list at 70% utilization, without announced-only projects at face value by 2030. Adjusted columns weight planned grid units by the completion rate of the December 2022 cohort, units under construction by 0.95, and BTM projects by status (operating 1, building 0.95, partly operating 0.5 or 0.6, permitted 0.4 or 0.5, for 2030 or 2035, announced 0). α is (grid supply − retirements − non-DC growth + BTM) over DC demand, γ the BTM share of that supply, and β the clean share of gross grid supply.

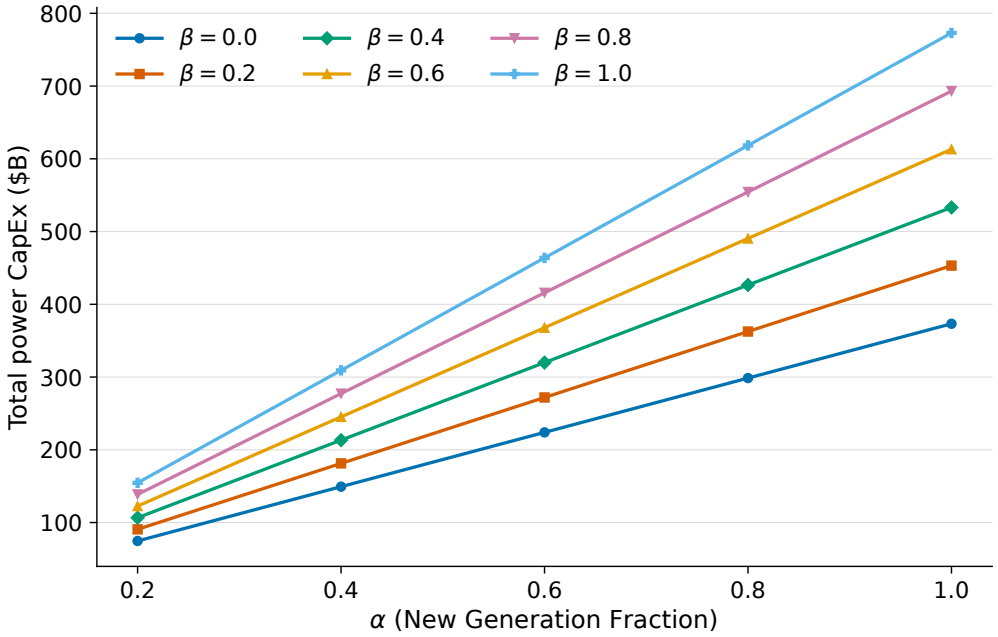
Table 6: The full announced build-out by 2035

	No new supply ($\alpha = 0$)	Alternative	New supply = demand ($\alpha = 1$)
Change from 2023, $B per year			
Data center electricity bill	300	215	254
Final good sector loss	86	37	8
Electricity sector gain	229	173	223
Fuel providers gain	69	53	30
Emissions cost	247	192	101
Output loss (% of GDP)	0.33	0.14	0.03
Wholesale price, median balancing area (%)	17	11	0
Wholesale price, largest increase (%)	271	68	26

Notes: As Table 4, with $(\alpha, \beta, \gamma) = (0.5, 0.5, 0.4)$, for all announced data center projects by 2035 (359 GW), including 181 GW of projects with no announced online date. The largest wholesale price increases under the alternative are in Northwest US (+68%), Texas Panhandle-Oklahoma (+61%), Most of Texas (+54%).

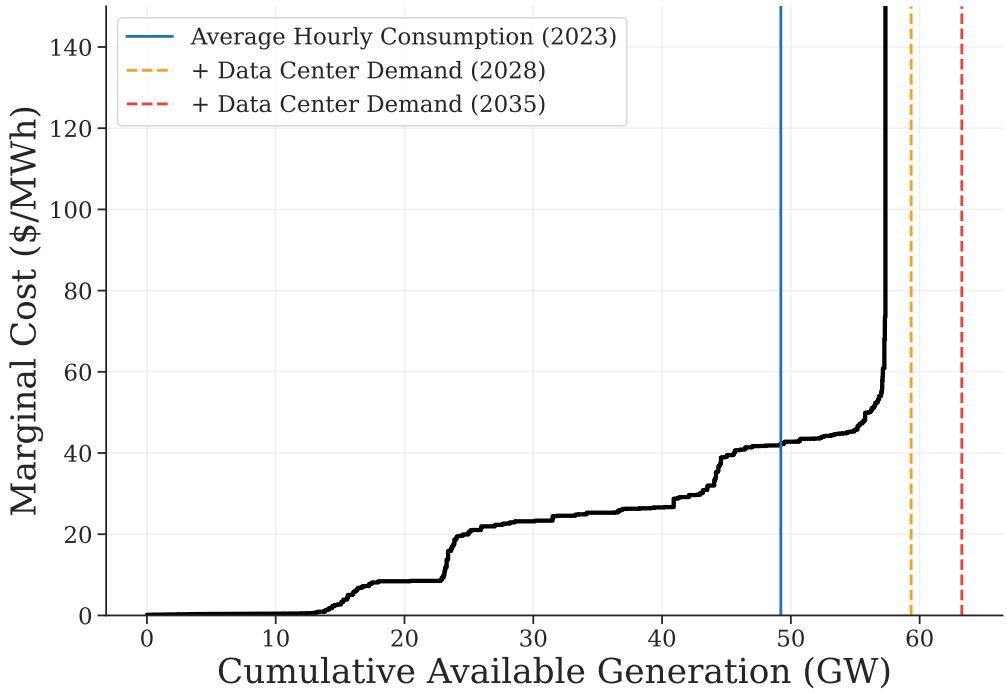
Internet Appendix for Energy Cost of AI and Data Centers

Figure IA.1: Capital Costs of New Power Generation by Renewable Share



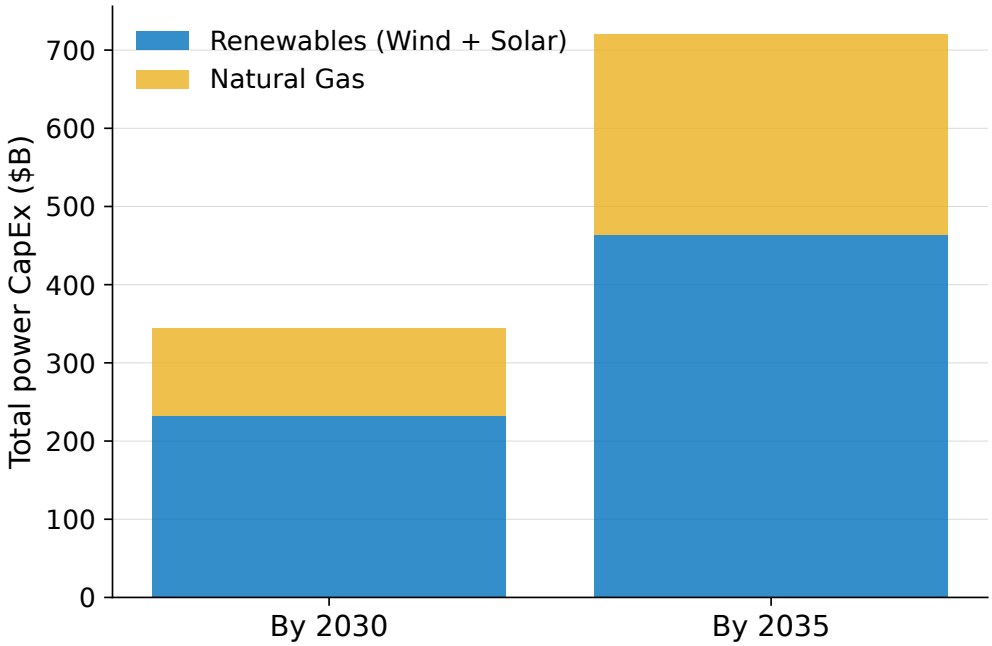
Notes: This figure shows total capital expenditure (CapEx) required to build new electricity generation capacity to meet 2030 data center demand, varying the parameters α (the share of demand met by new generation) and β (the share of new generation capacity built from renewables, with the remainder supplied by natural gas). The horizontal axis reports α, while the vertical axis reports total generation investment in billions of U.S. dollars. Each line corresponds to a different value of β, illustrating how total capital costs increase with both the share of demand met by new generation and the renewable share.

Figure IA.2: Local Supply Curve: ERCT, Historical Generation-Based Capacity



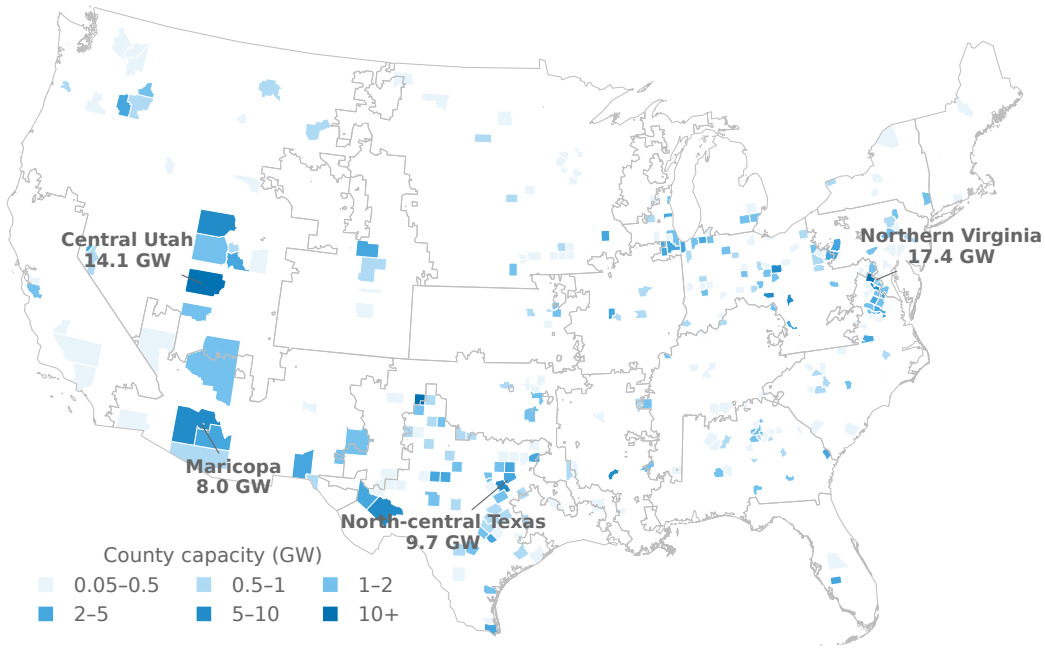
Notes: This figure shows the local supply curve for the EPA EGrid region *ERCT*, which covers most of Texas. The (step-wise) supply curve uses estimates of generator-level marginal costs, which comprise labor and fuel costs. Each step's width reflects the generator's available capacity, measured as the maximum annual generation observed since 2008. Vertical lines indicate average hourly electricity consumption in 2023 (EIA), with projected data center demand added for 2028 and 2035 (Cleanview).

Figure IA.3: Capital Costs of New Power Generation by Technology



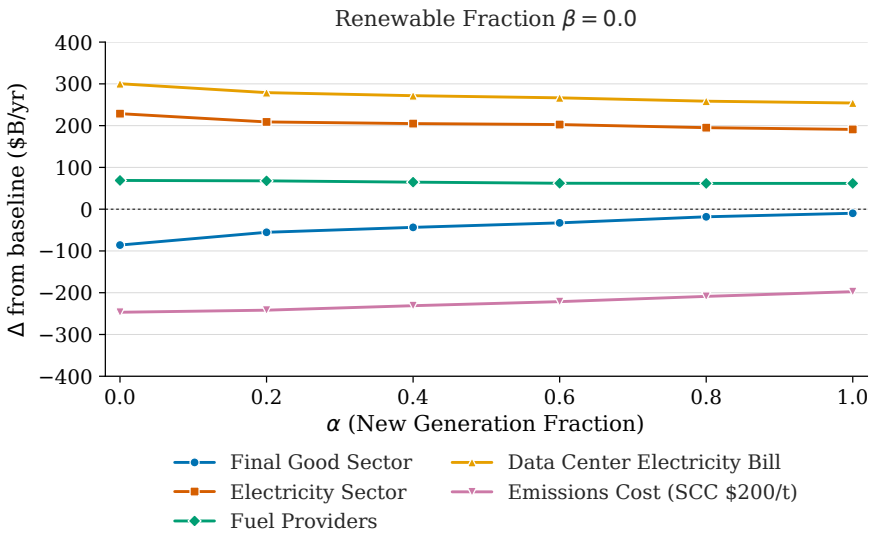
Notes: This figure shows the breakdown of capital expenditure (CapEx) on new electricity generation capacity needed to meet projected data center demand by 2030 and 2035, split between renewables (wind and solar) and natural gas. Each bar averages over the grid of α and β values in Figure IA.1. The vertical axis reports total generation investment in billions of U.S. dollars.

Figure IA.4: New data center capacity by 2035, by county and balancing area

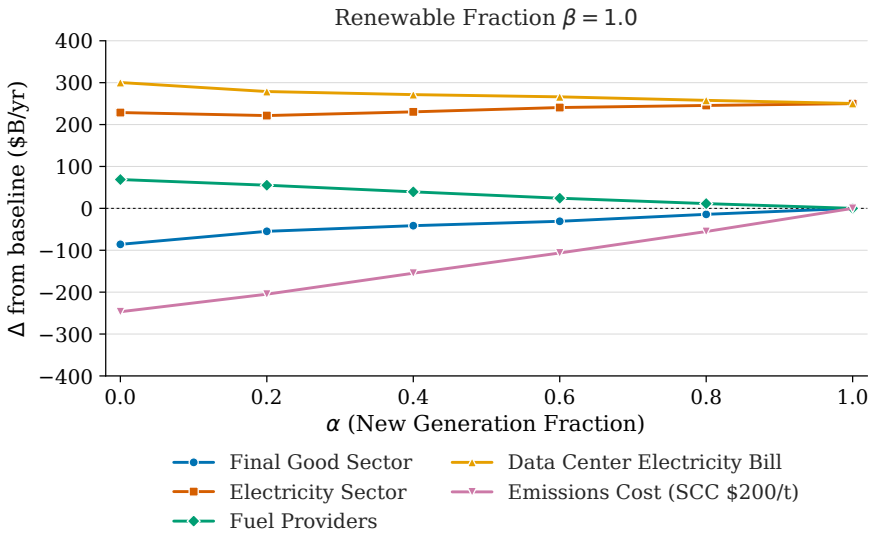


Notes: As Figure 7, for all announced projects by 2035, including projects with no announced online date (Section 6.5).

Figure IA.5: The full announced build-out by 2035: economic impact by new-generation fraction



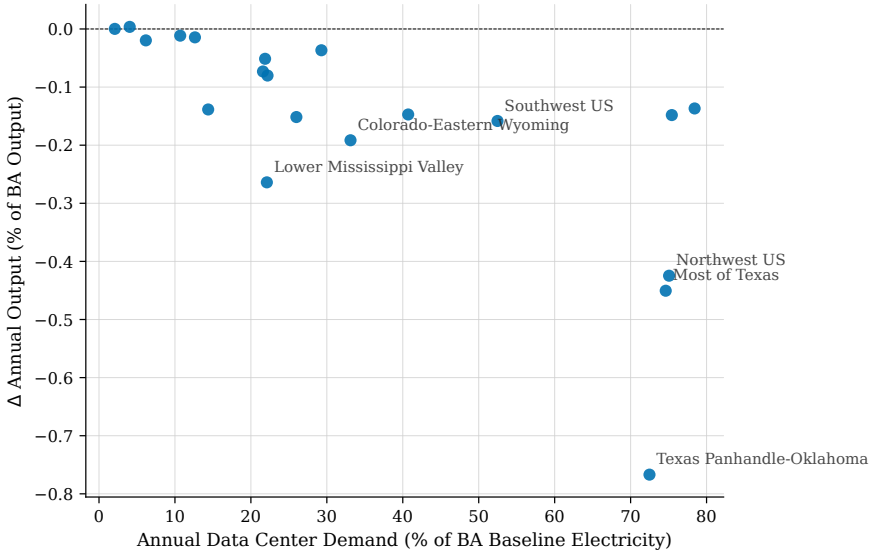
(a) All-gas new build ($\beta = 0$)



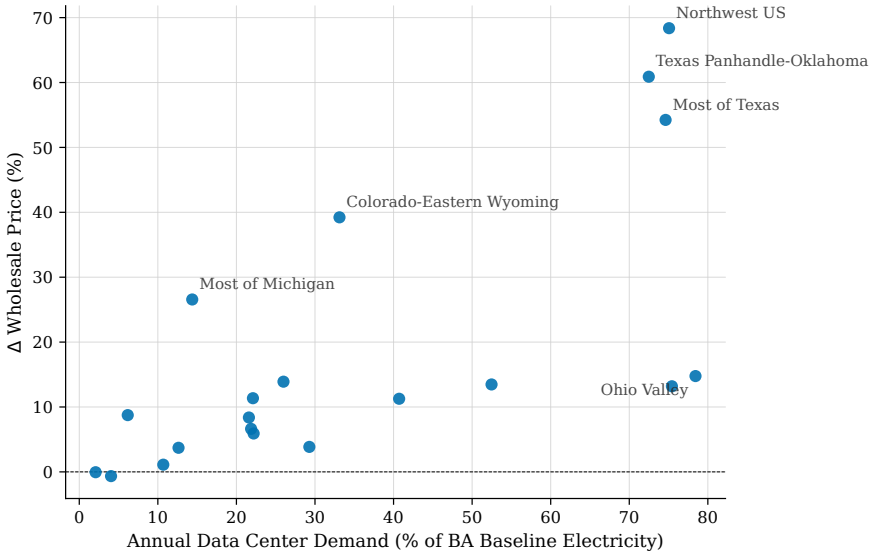
(b) All-renewable new build ($\beta = 1$)

Notes: As Figure 13, for all announced data center projects by 2035 (Section 6.5).

Figure IA.6: The full announced build-out by 2035: output and wholesale prices across balancing areas



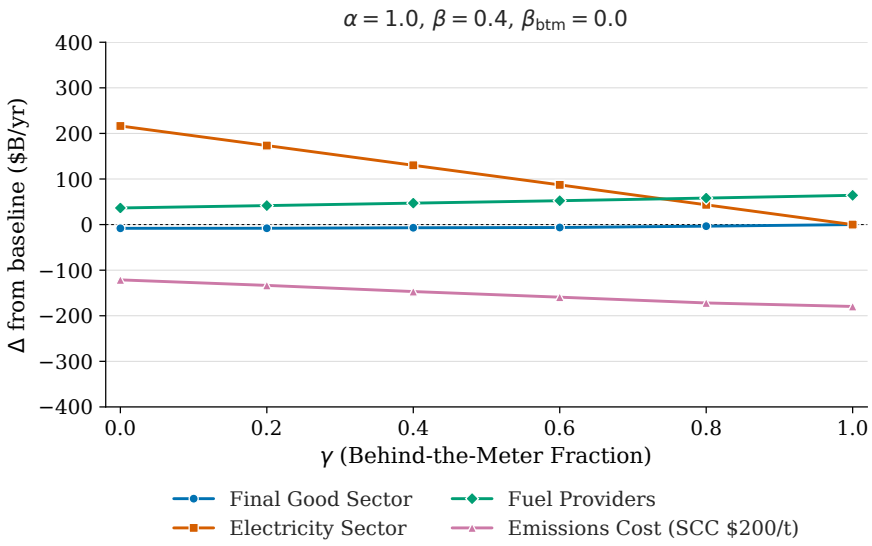
(a) Output change



(b) Wholesale price change

Notes: As Figures 14 and 15, panel (a), under the 2035 alternative $(\alpha, \beta, \gamma) = (0.5, 0.5, 0.4)$.

Figure IA.7: The full announced build-out by 2035: behind-the-meter generation ($\alpha = 1$, $\beta = 0.4$)



Notes: As Figure 16, panel (a), for all announced data center projects by 2035, with all new supply built ($\alpha = 1$), a renewable share of $\beta = 0.4$ for grid supply, and all-gas behind-the-meter generation.