

参赛学生姓名：

蔡子乐、包韵晨、刘奕辰

中学：

上海外国语大学附属外国语学校

省份：

上海市

国家/地区：

中国

指导老师姓名：

朱皓斌

指导老师单位：

上海外国语大学附属外国语学校

论文题目：

电力先于算力：电力约束与数据中心相关产业发展的阶段性关系

Power Before Compute: Stage-Dependent Relationships Between Electricity Constraints and Data-Center-Related Industry Development

Cai Zile, Bao Yunchen, Liu Yichen

Abstract

As generative artificial intelligence services expand, timely electricity connections and a stable, continuous supply have become important conditions for data center location. This paper examines the relationship between electricity constraints and industry agglomeration across 51 U.S. state-level jurisdictions through a multi-stage analysis. Based on data from 2013-2025, we combine the Attribute Hierarchical Model (AHM) with Criteria Importance Through Intercriteria Correlation (CRITIC) to construct a Power Constraint Index (PCI) with four dimensions and 11 indicators. A fixed-effects Poisson Pseudo-Maximum Likelihood (PPML) model compares agglomeration relationships in the early, middle, and recent periods. A Slacks-Based Measure model with variable returns to scale (VRS-SBM) then assesses industry outputs relative to the regional economy and talent base. The PPML estimates describe conditional associations and are not interpreted as causal effects of electricity constraints.

Stronger electricity constraints correspond to a lower conditional mean number of establishments in data-center-related industries. The negative association is stronger in the recent period than in the early period. This association is weaker in regions with better broadband and cloud-provider support, while the moderating relationship of the tax environment remains unclear. Relative gaps from reference levels are larger for employment and payroll than for establishment counts, indicating a need to assess establishment agglomeration separately from employment and income outcomes. These findings suggest that regional planning should consider electricity delivery times, continuity of supply, and digital support together in location assessments, while using firm size, job structure, and payroll outcomes to evaluate the local economic benefits of industry development.

Keywords: data centers; electricity constraints; industry agglomeration; multi-stage analysis; regional efficiency

Contents

1. Introduction.....	1
2. Literature review.....	3
2.1 Research on data center location.....	3
2.2 Electricity constraints and data center operations.....	4
2.3 Industry development efficiency.....	4
2.4 Combined framework and research hypotheses.....	5
3. Methods.....	5
3.1 Constructing the PCI.....	5
3.2 Location model.....	8
3.3 Model for DEA analysis of regional efficiency.....	11
4. Results.....	12
4.1 Sample and data.....	12
4.2 PCI measurement.....	14
4.3 Industry location.....	17
4.4 Regional industry efficiency analysis.....	21
4.5 Combined findings.....	26
5. Discussion.....	27
5.1 Electricity constraint mechanisms and regional decisions.....	27
5.2 Implications for practical planning.....	28
5.3 Limitations and further research.....	28
6. Conclusion.....	28
References.....	30

1. Introduction

Generative AI has become part of everyday question answering, writing, and programming. Models usually process inputs and generate answers by dividing text into units called tokens, and some services also charge on this basis. In this paper, the term "token economy" refers to demand for computing services arising from increased model use. These services depend on data center equipment, cooling, and electricity (IEA, 2025), and their expansion raises the question of whether electricity can be connected on schedule.

As demand for computing expands, does the relationship between electricity conditions and industry agglomeration also change? Bick et al. (2026) document the rapid adoption of generative AI, while Yoo et al. (2024) discuss new questions in digital innovation research. For data centers, this growth in demand involves both model scale and ongoing service needs. Because energy use depends on models, hardware, and operating practices, we examine industry location through regional electricity conditions.

Existing estimates show how much demand electricity supply planning must accommodate. IEA (2025) estimates that data centers worldwide used about 415 TWh in 2024 and could use nearly 945 TWh by 2030. Shehabi et al. (2024) estimate that U.S. data centers used 176 TWh in 2023, about 4.4% of national electricity consumption, with a possible substantial increase by 2028. These estimates cover AI and other computing tasks, indicating the need for regional electricity supply to accommodate rapidly growing data center demand.

To compare industry agglomeration across states, we use the annual average number of establishments in NAICS 518210, "Data Processing, Hosting, and Related Services," from the U.S. Bureau of Labor Statistics' Quarterly Census of Employment and Wages (QCEW).

Construction timelines often differ between computing facilities and electricity infrastructure. Servers can be installed relatively quickly, while substations, transmission lines, and grid connection upgrades often take longer (U.S. Department of Energy, 2024). Even a region with many customers and good fiber connectivity may face project delays because of electricity connection queues. IEA (2025) notes that concentrated growth can put pressure on local electricity systems. Esram and Elliott (2024) further discuss how operators, utilities, and local governments can coordinate electricity supply. We therefore consider electricity prices, supply reliability, and whether power can arrive on time.

This assessment also depends on cooling and resource conditions. Cooling methods influence both water and electricity use (Shehabi et al., 2024). Recent research examines liquid-cooling challenges and the role of ambient temperature in cooling operations (Azarifar et al., 2024; Qu et al., 2024). Land, permits, and local planning also affect site selection (Cáceres et al., 2024). These conditions can coincide: more electricity generation does not necessarily mean easier local connections, and lower prices do not necessarily mean lower outage or drought risks.

The PCI is therefore organized into four dimensions. Cost and uncertainty cover industrial electricity prices and their variation; capacity and reliability cover supply tightness, outages, dependence on electricity imports, and transmission and distribution losses. NERC (2024) discusses longer-term electricity resource adequacy, while Donnellan and Lawrence (2024) examine outage risk. Separate measures of cost and reliability help distinguish price advantages from the ability to maintain supply.

The clean electricity dimension includes carbon intensity and the renewable energy share. Storage operations and renewable procurement can have different implications for emissions and energy efficiency (Dupont et al., 2026; Awaysheh et al., 2026). The cooling and resources dimension uses cooling degree days and drought severity to describe climate and water pressure, drawing on evidence about cooling operations and data center resource use (Qu et al., 2024; Shehabi et al., 2024). The four dimensions together provide a basis for comparing regional electricity constraints.

Broadband, cloud-provider presence, market size, talent, and the tax environment enter the location model as controls or moderators. The PCI focuses on electricity and resource constraints. The model examines how network services and operating conditions alter their relationship with industry agglomeration.

Regions with similar establishment counts can still differ substantially in employment and payroll. This paper uses DEA analysis of regional efficiency to assess how closely industry outputs match the regional economy and talent base. It interprets regional differences in agglomeration through firm size and the structure of employment and income.

Existing research approaches this question from several directions. Studies of capacity reservation and deployment examine demand uncertainty in cloud services (Chen et al., 2024; Liu et al., 2025). Location studies consider renewable energy and local planning conditions (Ayyildiz et al., 2025; Cáceres et al., 2024), while policy research examines investment responses to corporate taxes and regional subsidies (Link et al., 2024; Siegloch et al., 2025). Differences in geographical scales and outcome variables across these studies make it difficult to assess directly whether the same electricity constraints have different associations across periods.

Recent firm-level research links AI investment with growth and product innovation (Babina et al., 2024) and examines how IT outsourcing relates to the performance of young businesses (Jin & McElheran, 2025). These findings motivate a separate regional analysis of the infrastructure supporting digital activity. We construct the PCI on a common state-year scale. A multi-stage analysis covering 2014–2019, 2020–2023, and 2024–2025 then compares changes in the relationship between electricity constraints and industry agglomeration directly. A single full-period association may mask differences that arise as computing demand grows.

U.S. state-level statistics provide annual measures of electricity, labor, and regional economic conditions. Our design covers 51 jurisdictions and links the outcome variable to the previous year's PCI and controls. The design supports both cross-region comparisons and examination of changes within the same region over time.

The study relates electricity conditions to industry location and regional outputs. The PCI puts electricity and resource indicators on a common scale for these comparisons. PPML compares establishment distributions conditional on population and operating conditions, with interaction terms in the multi-stage analysis testing whether the negative association strengthens over time. Interactions with broadband, cloud-provider presence, and taxes explain differences in regional supporting conditions. DEA compares industry outputs with the regional economy and talent base to assess how closely they match.

Figure 1-1 shows how measuring electricity constraints, analyzing industry agglomeration, and assessing regional efficiency are linked in the research design. AHM and CRITIC jointly determine PCI weights; fixed-effects PPML and period interactions compare agglomeration relationships; and VRS-SBM assesses how closely employment and payroll outputs match regional foundations (Ma et al., 2018; Diakoulaki et al., 1995; Correia et al., 2020; Mergoni et al., 2025).

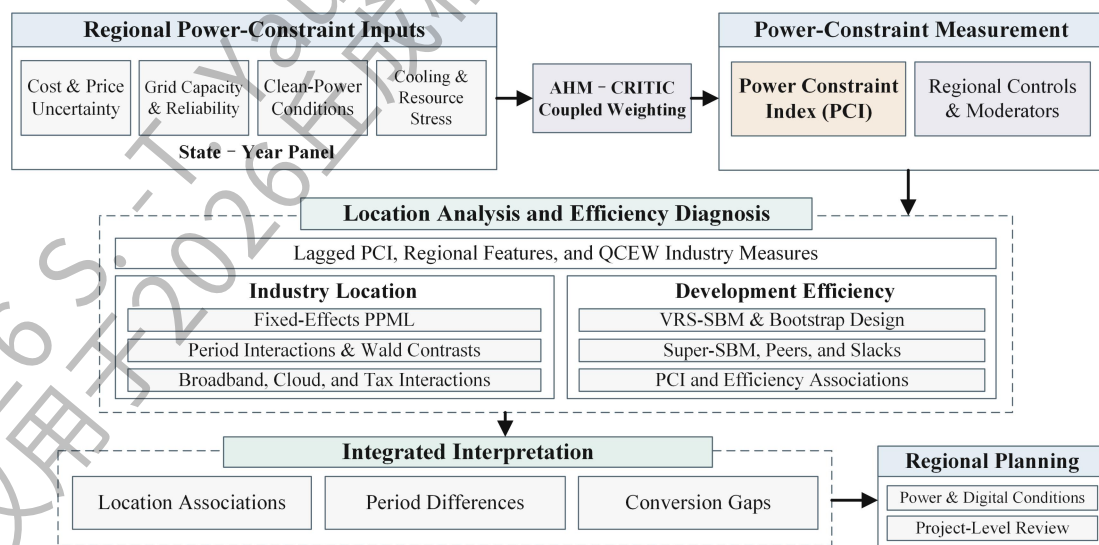


Figure 1-1 Research framework

The sequence in Figure 1-1 moves from measuring constraints, through their association with establishment stocks, to comparing employment and payroll outputs. The PCI summarizes

electricity and resource conditions through four dimensions and 11 indicators. The location model examines associations and their changes across periods, while DEA identifies output differences beyond establishment counts. Keeping broadband, cloud-provider presence, and the tax environment as separate variables in the subsequent analysis allows regional digital conditions and electricity weaknesses to be interpreted individually. PCI components identify supply problems; establishment, employment, and payroll records help identify gaps in industry development.

We pose the following three research questions.

RQ1: What is the relationship between electricity constraints and the agglomeration of data-center-related industries, and how does it change across the early, middle, and recent periods?

RQ2: How closely do industry establishment counts, employment, and payroll match the regional economy and talent base, and how is this alignment related to electricity constraints?

RQ3: How do broadband, cloud-provider presence, and the tax environment moderate the relationship between electricity constraints and industry agglomeration?

AHM and CRITIC jointly determine PCI weights: AHM reflects the subjective importance of each constraint, while CRITIC assigns weights based on data variation and indicator correlations. An equally weighted PCI is used to assess whether the weighting method changes the estimated association.

In the multi-stage analysis, association coefficients are compared on a common PCI scale. Wald tests assess period differences directly, and stability is checked under alternative weights, lag lengths, and sample coverage. These steps help distinguish persistent regional differences from changes arising as demand expands.

Regional comparisons can guide focused investigations of electricity connection times and employment structure, with project assessments drawing on local utility and site records.

Section 2 reviews research on data center location, electricity constraints, and industry efficiency and develops the hypotheses. Section 3 introduces PCI construction, the multi-stage analysis of industry location, and the model for DEA analysis of regional efficiency. Section 4 reports indicator weights, agglomeration relationships, period differences, and regional efficiency results. Section 5 discusses the links between electricity supply, digital support, and regional industry decisions. Section 6 summarizes the main findings.

2. Literature review

2.1 Research on data center location

Data center locations require access to digital markets as well as the physical conditions for operation. Research on critical IT infrastructure examines investment spillovers and coordination across districts (Guo et al., 2024), while work on shared digital suppliers shows how supplier relationships can transmit digitalization knowledge between firms (Hou et al., 2026). These studies help explain why connectivity and service ecosystems matter, although they do not directly estimate data center location effects. Established regions offer customers, specialized suppliers, and skilled workers. Individual projects also require sufficiently large sites, electricity, cooling, and predictable permitting. Location analysis therefore considers industry agglomeration alongside these operating conditions.

Network connectivity and the cloud service ecosystem can strengthen a region's advantages. Compatibility and multicloud procurement shape operating choices (Hosseini et al., 2026; Jain et al., 2025), while shared security responsibilities make providers and users interdependent (Clement & Arce, 2025). Capacity reservation and deployment decisions respond to demand uncertainty (Chen et al., 2024; Liu et al., 2025). Platform pricing and service interruptions also affect the capacity offered to users (Zhai et al., 2026), and task-to-server assignment constraints influence resource pooling in queueing systems (Cardinaels et al., 2024). Computing capacity alone therefore does not describe the services available in a region. Network access and cloud-provider presence enter the analysis as separate regional controls.

Climate, land, access to clean energy, and fiscal policy also influence location choices. Cooling

technology faces thermal-management and implementation constraints (Azarifar et al., 2024), while water use depends on the cooling system and operating conditions (Shehabi et al., 2024). Renewable energy has also been included as an explicit criterion in location models (Ayyildiz et al., 2025). The policy literature examines investment responses to corporate taxes and the regional effects of firm subsidies (Link et al., 2024; Siegloch et al., 2025). These criteria inform regional comparisons, but regions that attract investment may still face grid or resource bottlenecks. The PCI therefore records constraints. Network connectivity, cloud-provider presence, market conditions, and the tax environment remain separate explanatory variables.

2.2 Electricity constraints and data center operations

Electricity constraints involve cost, supply resilience, and sustainability. Research on electricity demand explains why power planning needs to account for data center growth (IEA, 2025). Facilities with continuous large loads face both current electricity prices and within-year price changes. Studies of distributed energy aggregation and battery storage show that market rules and combinations of energy services also affect electricity costs (Gao et al., 2024; Bae et al., 2026). Evidence from small firms in Uganda further illustrates how backup resources can limit disruption losses (Kundu et al., 2024); its relevance here is the continuity-of-service mechanism. For regional comparisons, electricity availability, expected costs, and supply risks therefore need to be considered together.

A project needs a feasible connection plan even when electricity is inexpensive. Longer-term electricity resource adequacy affects supply (NERC, 2024), as do local connection requirements and infrastructure coordination (U.S. Department of Energy, 2024). Renewable generation, flexible capacity, and storage may complement or substitute for one another under different operating conditions (Peng et al., 2024). Asset-level investment and operating decisions also matter for the clean energy transition (Secomandi, 2025). Evidence from U.S. utilities shows that digital energy management can affect network losses and outage duration, with organizational capabilities shaping its benefits (Meeks et al., 2026). ESRAM and Elliott (2024) explain why utilities and regional policymakers need to coordinate supply arrangements for data center loads. The PCI therefore treats capacity and reliability as a dimension separate from price.

The remaining dimensions describe the generation mix and cooling resources. Storage can have unintended carbon effects because its operation depends on the generation mix and market incentives (Dupont et al., 2026). Renewable procurement and energy efficiency are also distinct: evidence from industrial facilities finds that their relationship depends on how renewable electricity is procured (Awaysheh et al., 2026). In data centers, cooling technology, ambient temperature, and thermal loads affect operating choices (Azarifar et al., 2024; Qu et al., 2024). These findings support retaining separate indicators for carbon intensity, renewable energy, and cooling conditions, so that similar PCI totals can be traced to different supply and resource pressures.

2.3 Industry development efficiency

Establishment counts describe industry scale; employment and wage payments per establishment require additional information. Firm-level research links AI investment with growth and product innovation (Babina et al., 2024), while IT outsourcing affects the performance dynamics of young businesses (Jin & McElheran, 2025). Cloud sourcing also relates to expansion through mergers and acquisitions (Matta et al., 2026). Studies of collaborative work technologies examine changes in managerial intensity within firms (Gulati et al., 2026). Together, these findings suggest that digital activity can change firm size and workforce composition without a proportional change in establishment counts. Regional efficiency analysis therefore compares establishment counts, employment, and real payroll with regional economic size, digital talent, and training capacity.

Incentive policies raise a similar question. Link et al. (2024) study how investment responds to corporate tax increases. Siegloch et al. (2025) examine the spillover, efficiency, and equity effects of regional firm subsidies. A policy's value depends on both the investment it attracts and who bears the costs of electricity grids and water systems. A state may have relatively few establishments while still having relatively high industry employment compared with its economic base. A frontier model with multiple inputs and outputs compares this input-output relationship.

The Slacks-Based Measure model with variable returns to scale (VRS-SBM) allows input excesses and output shortfalls to adjust in different proportions without assuming a common price-based production function; [Mergoni et al. \(2025\)](#) review the relevant DEA modeling choices. Its scores measure the performance of the selected industry outputs relative to regional foundations. Because some resource slack can provide useful buffers ([Bogetoft & Kerstens, 2024](#)), a measured gap is interpreted as a relative input-output imbalance, not an instruction to cut regional resources mechanically.

2.4 Combined framework and research hypotheses

This paper uses PPML to analyze the association between the one-year-lagged PCI and QCEW establishment stocks. Period interactions compare the early, middle, and recent periods, and the analysis estimates the moderating relationships of broadband, cloud-provider presence, and taxes. DEA analysis of regional efficiency assesses how closely industry outputs match the economic and talent base. These relationships motivate the following research hypotheses.

H1: *Stronger electricity constraints are associated with weaker agglomeration of the data-processing industry.*

H2: *Under the specified regional input-output system, stronger electricity constraints are associated with higher period-average DEA inefficiency.*

H3: The broadband subscription rate, cloud-provider count, and general corporate tax conditions weaken the adverse associations of electricity constraints with industry agglomeration and regional efficiency.

Hypothesis 1 is tested using the baseline PPML coefficient and its confidence interval; Wald tests then assess whether this association differs across periods. For Hypothesis 2, inefficiency is regressed on the PCI. Hypothesis 3 is examined by adding moderator interactions separately to the location and efficiency models. Period comparisons use coefficient differences and their uncertainty, rather than relying solely on the significance of each period's estimate.

3. Methods

3.1 Constructing the PCI

The Power Constraint Index (PCI) summarizes the electricity and resource conditions that may constrain data-center-related investment across regions and years. The index contains four dimensions and 11 indicators: industrial electricity prices and price volatility; the System Average Interruption Duration Index (SAIDI), System Average Interruption Frequency Index (SAIFI), capacity tightness, dependence on net electricity imports, and transmission and distribution (T&D) losses; carbon intensity and the renewable energy share; and cooling degree days and drought severity. Broadband, cloud-provider presence, the tax environment, market size, and talent are excluded from the index.

The modules share regional coverage, but their time windows are defined separately. The overall 2010–2025 span contains 816 (51×16) possible region-year combinations; the 2013–2025 window for the 11 indicators contains 663 (51×13); and the 2014–2025 location and efficiency window contains 612 (51×12). QCEW provides annual state-level establishment statistics for detailed industries. Sources are merged by common regions and years. The baseline and period-interaction PPML estimations use 612 region-year observations over 2014–2025, with current outcomes and lagged covariates available for the same observations. The efficiency analysis further aggregates the sample into 51 period-average decision-making units (DMUs).

3.1.1 Data preprocessing

Aggregation choices can conceal differences among a composite indicator's components and affect comparisons between units ([Mariani et al., 2024](#)). Indicator selection should therefore follow a clear measurement boundary. The PCI accordingly covers electricity and resource constraints; broadband, cloud-provider presence, tax conditions, market size, and talent enter later modules as moderators or controls. Section 2 supplies the constraint mechanisms, while EIA, BEA, SEDS, the U.S. Drought Monitor, and other sources supply the indicator definitions. The definitions, directions, and sources of 11 indicators are organized across four dimensions in Table 3-1: cost and uncertainty,

capacity and reliability, clean electricity, and cooling and resources.

Table 3-1 PCI indicator system

Dimension	Indicator	Relationship between the raw value and constraints	Definition and source
Cost and uncertainty	Industrial electricity price	Higher values indicate stronger constraints	Annual state industrial retail electricity prices in constant 2024 dollars (Forms EIA-861 and EIA-861M; BEA).
Cost and uncertainty	Electricity price volatility	Higher values indicate stronger constraints	Within-year standard deviation of log monthly industrial electricity prices, requiring at least ten monthly observations (Form EIA-861M).
Capacity and reliability	System Average Interruption Duration Index (SAIDI)	Higher values indicate stronger constraints	Annual interruption duration excluding major events (EIA-861).
Capacity and reliability	System Average Interruption Frequency Index (SAIFI)	Higher values indicate stronger constraints	Annual interruption frequency excluding major events (EIA-861).
Capacity and reliability	Capacity tightness	Higher values indicate stronger constraints	Electricity sales divided by theoretical annual generation capacity (EIA).
Capacity and reliability	Net import dependence	Higher values indicate stronger constraints	The positive part of electricity sales minus net generation, divided by electricity sales (EIA).
Capacity and reliability	Transmission and distribution loss rate (T&D)	Higher values indicate stronger constraints	Estimated transmission and distribution losses divided by total electricity supply (EIA-861).
Clean electricity	Carbon intensity	Higher values indicate stronger constraints	Power-sector CO ₂ emissions divided by net generation (EIA).
Clean electricity	Renewable generation share	Higher values indicate weaker constraints	Renewable electricity generation divided by total net generation (EIA).
Cooling and resources	Cooling degree days	Higher values indicate stronger constraints	Annual population-weighted cooling degree days from EIA's State Energy Data System (SEDS).
Cooling and resources	Drought severity	Higher values indicate stronger constraints	Annual mean of the weekly Drought Severity and Coverage Index (DSCI), divided by 500 (U.S. Drought Monitor).

In Table 3-1, 5 indicators measure capacity and reliability, compared with 2 for each of the other three dimensions. Together, the dimensions describe supply assurance, operating costs, energy use structure, and cooling resources. Among the 11 indicators, a higher renewable generation share indicates weaker constraints, while higher raw values for the other 10 indicate stronger constraints.

Their directions must therefore be aligned before index construction. Electricity price volatility requires observations for at least 10 months, and outage indicators exclude major events; these requirements affect whether annual data can enter the PCI. To avoid directly adding values with different directions or definitions, data preparation should retain the original units, year coverage, and processing rules for each indicator.

Let $X=(x_{ij})$ be the region-year $\times$ indicator matrix. The raw series are first aligned to common region identifiers, observation years, industry definitions, and units, then checked for duplicates, signs, ranges, continuity, and missing values. Release dates are recorded separately from observation years. Continuous variables are winsorized at the 1st and 99th percentiles. Period comparisons use the same winsorization bounds, normalization limits, and PCI weights within the common analysis window. To construct interactions, each continuous variable is standardized separately before multiplication.

Constraint-increasing indicators include industrial electricity prices, price volatility, SAIDI, SAIFI, capacity tightness, dependence on external electricity, network losses, carbon intensity, cooling degree days, and drought severity. Higher raw values indicate stronger constraints, and these indicators are normalized using Equation (1). The analysis panel is winsorized before this calculation. A zero denominator is flagged for review.

$$z_{ij} = \frac{x_{ij} - \min_i x_{ij}}{\max_i x_{ij} - \min_i x_{ij}} \quad (1)$$

The renewable share of electricity generation is a constraint-reducing indicator. Equation (2) reverses its direction to match the other components. Higher normalized values therefore always indicate less favorable electricity and resource conditions.

$$z_{ij} = \frac{\max_i x_{ij} - x_{ij}}{\max_i x_{ij} - \min_i x_{ij}} \quad (2)$$

Price volatility is calculated as the within-year standard deviation of log industrial electricity prices, using only years with at least 10 monthly observations. SAIDI and SAIFI use series explicitly reported as excluding major events. Coverage measures are also reported so that data gaps can be checked. Capacity, trade, losses, emissions, generation, climate, and drought follow the annual definitions used by their respective sources. Missing PCI observations are not imputed in constructing the descriptive index.

3.1.2 Attribute Hierarchical Model (AHM)

The Attribute Hierarchical Model (AHM) converts pairwise judgments of indicator importance into subjective weights (Ma et al., 2018). For the planned expert assessment, 3 experts from universities and data center location research were invited to complete questionnaires independently. Each expert had at least 10 years of relevant research experience. A consistency ratio of 0.037 is reported.

The hierarchy consists of the PCI objective, the four dimensions, and the 11 indicators in Table 3-1. Pairwise preferences at each level are written as a positive reciprocal matrix $A=(a_{jk})$, whose elements satisfy $a_{jj}=1$ and $a_{jk}=1/a_{kj}$.

Off-diagonal ratios are converted into relative attribute measures, while diagonal elements are set to zero:

$$q_{jk} = \frac{a_{jk}}{1 + a_{jk}}, \quad j \neq k; \quad q_{jj} = 0 \quad (3)$$

Because each pair satisfies $q_{jk}+q_{kj}=1$, summing each row and normalizing by the total gives the AHM weights:

$$w_j^A = \frac{\sum_{k=1}^J q_{jk}}{\sum_{l=1}^J \sum_{k=1}^J q_{lk}}, \quad \sum_{j=1}^J w_j^A = 1 \quad (4)$$

3.1.3 Objective weighting through indicator correlations (CRITIC)

CRITIC stands for Criteria Importance Through Intercriteria Correlation. It determines objective weights from correlations among indicators by considering an indicator's variation and its informational overlap with other indicators (Diakoulaki et al., 1995). An indicator usually receives a higher weight when greater variation is accompanied by less overlap with the others.

For indicator j , contrast intensity is measured by its standard deviation in the panel:

$$\sigma_j = \sqrt{\frac{1}{m} \sum_{i=1}^m (z_{ij} - \bar{z}_j)^2} \quad (5)$$

Let r_{jk} be the Pearson correlation coefficient between indicators j and k . CRITIC combines dispersion and conflict to calculate the information content of an indicator:

$$C_j = \sigma_j \sum_{k=1}^J (1 - r_{jk}) \quad (6)$$

Greater variation across regions and over time increases σ_j ; a larger sum of $(1 - r_{jk})$ indicates less overlap with the information in other components. Both increase the indicator's weight. Normalizing C_j gives the CRITIC weight vector in Equation (7).

$$w_j^C = \frac{C_j}{\sum_{k=1}^J C_k}, \quad \sum_{j=1}^J w_j^C = 1 \quad (7)$$

All CRITIC weights are nonnegative and sum to one. Sensitivity analysis examines how the choice of panel affects these weights.

3.1.4 Coupled weights and the PCI

The AHM and CRITIC weights are coupled through Equation (8), so that the PCI reflects both subjective importance judgments and objective indicator information.

The coupled weights w are nonnegative and sum to one. With equal importance assigned to AHM and CRITIC, we minimize the differences between the final weights and these two sets of weights:

$$\begin{aligned} & \min_{w \in \Delta} \left\{ \frac{1}{2} D_{\text{KL}}(w \| w^A) + \frac{1}{2} D_{\text{KL}}(w \| w^C) \right\} \\ & = \min_{w \in \Delta} \sum_{j=1}^J w_j \ln \frac{w_j}{\sqrt{w_j^A w_j^C}}, \quad \Delta = \left\{ w : w_j \geq 0, \sum_{j=1}^J w_j = 1 \right\}. \end{aligned} \quad (8)$$

The first-order conditions give the following normalized geometric-mean form:

$$w_j^{AC} = \frac{\sqrt{w_j^A w_j^C}}{\sum_{k=1}^J \sqrt{w_k^A w_k^C}} \quad (9)$$

Each coupled weight is the normalized geometric mean of the two weighting results. To calculate the PCI, component directions are first aligned and their values normalized; the components are then combined as a weighted sum. The robustness analysis includes an equally weighted PCI to assess sensitivity to the weighting method.

$$\text{PCI}_i = \sum_{j=1}^J w_j^{AC} z_{ij} \quad (10)$$

3.2 Location model

The industry outcome consists of nonnegative, strongly right-skewed counts. Poisson Pseudo-Maximum Likelihood (PPML) specifies a multiplicative conditional mean and accommodates zero outcomes. We use PPML with region and year fixed effects, following the implementation described by [Correia et al. \(2020\)](#). These fixed effects control for time-invariant regional heterogeneity and common shocks, respectively. The conditional mean specification, rather than equality between the mean and variance, underlies the use of PPML. We interpret the estimated coefficients as conditional associations.

3.2.1 Baseline specification

Let y_{it} be the annual average number of establishments in NAICS 518210 from QCEW. QCEW is used because its annual average establishment counts match the study's focus on establishment stocks rather than new establishment entry. The main explanatory variable is the standardized *PCI* lagged by one year. The previous year's electricity conditions are linked to the current year's industry distribution because adjustments in location and operations usually take time.

$$E(y_{it} | \mathcal{F}_{i,t-1}, \text{Population}_{it}) = \exp[\ln(\text{Population}_{it}) + \beta z(\text{PCI}_{i,t-1}) + \gamma' z(X_{i,t-1}) + \alpha_i + \tau_t]. \quad (11)$$

The model includes region and year fixed effects, with current population used as an exposure term to account for differences in regional size. Controls lagged by one year include real GDP per capita, the digital talent share, rental Regional Price Parity (RPP), the broadband subscription rate, and the general corporate tax proxy. Cloud-provider counts enter only the corresponding moderator model. Conditional on these factors, the resulting β describes the association between the *PCI* and industry agglomeration.

Because the *PCI* is standardized, β is the semi-elasticity associated with a one-standard-deviation increase in the one-year-lagged *PCI*. The corresponding percentage association is reported using Equation (12).

$$\Delta\%E(y) = 100 [\exp(\beta \Delta \text{PCI}) - 1] \quad (12)$$

In the PPML specification ([Correia et al., 2020](#)), $\exp(\beta)$ is the incidence rate ratio (IRR) for the expected establishment count used to measure industry agglomeration, conditional on the included covariates, fixed effects, and exposure term. Since population enters as an exposure, this ratio compares expected establishment counts after accounting for population size.

3.2.2 Other count models

Count outcomes may be overdispersed, especially in established industry agglomerations. The analysis therefore uses a negative binomial model with a specified dispersion parameter to check sensitivity to the distributional assumption. Fixed-effects PPML remains the comparison baseline.

$$\text{Var}(y_{it} | \cdot) = \mu_{it} + \alpha \mu_{it}^2, \quad \alpha = 1.0 \quad (13)$$

PPML is the main estimator because its identifying restriction concerns the conditional mean, without requiring equality of the conditional mean and variance. A $\log(1+y)$ two-way fixed-effects regression checks sensitivity to a different functional form.

The robustness analysis includes a two-year *PCI* lag, region-specific trends, exclusion of major industry centers, an equally weighted *PCI*, a negative binomial comparison, a $\log(1+y)$ two-way fixed-effects specification, and a placebo test using future *PCI*. These comparisons could later be extended to facility-level outcomes.

3.2.3 Moderating mechanisms

A separate model is estimated for each moderator: the annual household broadband subscription rate, the number of major cloud providers active each year, or a general corporate tax proxy scored from 0–4. Continuous variables are standardized before interaction terms are constructed. The renewable energy share remains in the *PCI*, while the tax proxy describes the general corporate tax environment.

$$E(y_{it} | \cdot) = \exp[\ln(\text{Population}_{it}) + \beta_1 \text{PCI}_{i,t-1} + \beta_2 M_{i,t-1} + \beta_3 (\text{PCI}_{i,t-1} \times M_{i,t-1}) + \gamma' X_{i,t-1} + \alpha_i + \tau_t] \quad (14)$$

In Equation (14), β_1 represents the *PCI* association when the moderator is at its mean; β_3 describes how that association changes with a one-standard-deviation increase in the moderator. The three interactions are estimated separately, and their *p*-values are adjusted with the Holm method to reduce the chance of false conclusions from multiple comparisons.

3.2.4 Multi-stage analysis of period differences

The multi-stage analysis uses the same PPML specification and divides outcome years into early (2014–2019), middle (2020–2023), and recent (2024–2025) periods. The periods respectively provide a pre-pandemic reference, include the pandemic and subsequent adjustment, and correspond to expanding demand for computing services. This division compares changes over time in the relationship between electricity constraints and industry agglomeration. With the one-year lag, the recent window pairs outcomes in 2024–2025 with electricity conditions in 2023–2024.

The complete 11-indicator *PCI* begins in 2013; with a one-year lag, the first outcome year is 2014. All three periods use the same *PCI* construction and standardization. Recalculating weights or standard deviations for each period would alter what a one-unit increase represents. In the following interaction specification, $z_{i,t-1}$ is the *PCI* standardized on a common scale; D_{1t} equals one in 2020–2023, and D_{2t} equals one in 2024–2025. Year fixed effects absorb the period indicators, but their interactions with the *PCI* can still vary across regions.

$$\log \mu_{it} = \log \text{Population}_{it} + \alpha_i + \lambda_t + \beta z_{i,t-1} + \delta_1 (z_{i,t-1} D_{1t}) + \delta_2 (z_{i,t-1} D_{2t}) + \gamma' X_{i,t-1} \quad (15)$$

Here, μ_{it} is the model's expected QCEW establishment count. The control vector, exposure term, fixed effects, and region-clustered covariance follow the baseline specification. The period slopes are $b_0 = \beta$, $b_1 = \beta + \delta_1$, and $b_2 = \beta + \delta_2$. Thus, δ_1 compares 2020–2023 with the reference period, δ_2 compares 2024–2025 with the reference period, and $\delta_2 - \delta_1$ compares the latter two periods. Estimating the period terms jointly preserves the covariance needed for each comparison.

Let $b = (b_0, b_1, b_2)'$ and let V be its covariance matrix. For a row vector c defining a pairwise contrast, the coefficient difference is $d = c'b$ and its standard error is $\sqrt{c'Vc}$. The Wald statistic in Equation (16) tests whether this difference is zero. The joint test uses two independent contrast rows in C to test whether all three slopes are equal.

$$W = \frac{(\hat{cb})^2}{\hat{c'Vc}} (df = 1) \quad (16)$$

$$W_{\text{joint}} = (\hat{Cb})' (\hat{CVC})^{-1} (\hat{Cb}) (df = 2)$$

The three pairwise-comparison *p*-values form one family for Holm adjustment; the plotted 95% confidence intervals are unadjusted pointwise intervals. Period differences are assessed through direct contrasts of the interaction terms. A difference in significance labels does not itself imply a significant coefficient difference.

The robustness checks add region-specific trends and exclude major industry centers to assess sensitivity to regional trends and influential regions.

3.2.5 Identification limits and robustness checks

Robustness checks use region and year effects, lagged explanatory variables, region-specific trends, exclusion of industry centers, alternative functional forms, clustered standard errors, and a lead-variable placebo test to examine how coefficients depend on timing assumptions, trends, sample selection, functional form, and the placebo specification. These checks do not create exogenous variation in *PCI*. The PPML coefficients, including the period-specific slopes, are therefore interpreted as conditional associations.

3.3 Model for DEA analysis of regional efficiency

Data Envelopment Analysis (DEA) compares multiple regional foundations with industry outputs. [Mergoni et al. \(2025\)](#) review the model choices and methodological developments in this field. Here, variable returns to scale allow comparisons among regions of different sizes; the Slacks-Based Measure (SBM) describes gaps in individual inputs and outputs, and Super-SBM ranks regions on the frontier. This paper uses desirable outputs. Extensions for environmental efficiency can also accommodate undesirable outputs, but carbon emissions and water use are not included as outputs in the present model.

3.3.1 Decision-making units and variables

The main frontier contains 51 decision-making units, one for each state-level jurisdiction in the sample. Each DMU uses period averages within the 2014–2025 efficiency window. The available period includes only years jointly covered by the specified inputs and QCEW outputs. Annual frontiers are also included to assess the temporal robustness of the comparison based on period averages.

The inputs comprise real regional gross domestic product (GDP), employment in computer and mathematical occupations under Standard Occupational Classification (SOC) code 15-0000, and annual program completions in "Computer and Information Sciences" under Classification of Instructional Programs (CIP) code 11. Desirable outputs are the annual average number of establishments, employment, and real payroll in the data-processing industry. These inputs represent regional foundations. After the scores have been calculated, PCI is used for association analysis.

3.3.2 SBM and Super-SBM

SBM allows inefficiency to take the form of input excesses or output shortfalls in different proportions. Let x_o and y_o denote the input and desirable-output vectors for region o , while X and Y collect observations for all reference regions.

Under variable returns to scale, the reference technology is defined by $X\lambda \leq x_o$, $Y\lambda \geq y_o$, $\lambda \geq 0$, and $\mathbf{1}'\lambda = 1$. Input and output slacks represent mathematical distances between the evaluated region and the fitted frontier. When inputs are regional GDP and broad talent measures, input slacks describe relative gaps between industry scale and regional foundations.

$$P = \{(x, y) : x \geq X\lambda, y \leq Y\lambda, \mathbf{1}'\lambda = 1, \lambda \geq 0\} \quad (17)$$

The non-oriented SBM objective combines normalized input excesses and output shortfalls into an efficiency measure. Since adjustment proportions can differ, the model can separately describe gaps in talent, training, establishment counts, employment, and payroll. A region with a large nondigital economy may receive a lower score because its data-processing industry is relatively small, even when its overall economy performs well.

$$\rho_o = \min \frac{1 - \frac{1}{m} \sum_{i=1}^m \frac{s_i^-}{x_{io}}}{1 + \frac{1}{s} \sum_{r=1}^s \frac{s_r^+}{y_{ro}}}, \quad (18)$$

$$x_o = X\lambda + s^-, \quad y_o = Y\lambda - s^+, \\ \mathbf{1}'\lambda = 1, \quad \lambda \geq 0, \quad s^- \geq 0, \quad s^+ \geq 0.$$

Conventional efficiency scores lie in (0,1]. A score of one places the DMU on the frontier; a score below one indicates at least one input-side or output-side gap relative to the frontier. The score is then used for bias correction and analysis of its association with electricity constraints.

To rank frontier regions, the evaluated DMU is removed from its own reference set and the model is re-estimated according to the Super-SBM approach. Super-efficiency scores can exceed one, while nonfrontier DMUs retain their conventional scores below one.

$$\rho_o^{SE} = \min \frac{\frac{1}{m} \sum_{i=1}^m \frac{\bar{x}_i}{x_{io}}}{\frac{1}{s} \sum_{r=1}^s \frac{\bar{y}_r}{y_{ro}}}, \quad (19)$$

$$X_{-o}\lambda \leq \bar{x}, \quad Y_{-o}\lambda \geq \bar{y}, \quad \bar{x} \geq x_o,$$

$$0 < \bar{y} \leq y_o, \quad \mathbf{1}'\lambda = 1, \quad \lambda \geq 0.$$

Model outputs include conventional SBM scores, Super-SBM scores and ranks, reference regions, input-side reference gaps, output gaps, and annual frontier results, showing regional rankings and gaps relative to the corresponding benchmarks.

3.3.3 Efficiency score correction and the relationship with electricity constraints

Efficiency scores compare evaluated regions with reference regions. With a small reference sample, leading regions within the sample may receive inflated scores. Recent methodological reviews discuss bootstrap inference and second-stage estimation in DEA (Mergoni et al., 2025). Our design separates these tasks. A resampling correction procedure repeatedly generates reference samples and re-estimates efficiency scores; changes in the scores are used to estimate finite-sample bias. To analyze the relationship between efficiency and electricity constraints, a truncated-regression resampling procedure is then designed to measure coefficient uncertainty. The first procedure corrects efficiency scores; the second assesses the precision of the association coefficients.

$$\hat{\rho}_o^{BC} = 2\hat{\rho}_o - \frac{1}{B} \sum_{b=1}^B \hat{\rho}_o^{*(b)} \quad (20)$$

In the association regression, inefficiency is defined as the reciprocal of the corrected efficiency score. For example, when the same industry output corresponds to a larger regional input base, relative efficiency is lower and inefficiency is higher. This transformation allows a direct examination of whether stronger electricity constraints are associated with larger regional output gaps.

$$I_i = \frac{1}{\rho_i^{BC}}$$

$$I_i = \alpha + \delta z(PCI_i) + \eta z(M_i) + \kappa z(PCI_i)z(M_i) + \gamma z(Rent_i) + \varepsilon_i \quad (21)$$

$$\varepsilon_i \sim N(0, \sigma^2), \text{ with lower truncation } I_i \geq 1$$

Separate regressions of inefficiency on electricity constraints include $PCI \times$ broadband, $PCI \times$ cloud-provider count, or $PCI \times$ general corporate tax proxy. Continuous variables are standardized separately before multiplication to construct interactions. The main PCI coefficient represents the association at the moderator's mean.

3.3.4 Efficiency diagnostics

Sensitivity checks compare period-average with annual frontiers and bootstrap-corrected with original scores, and examine reference frequencies and slack ratios to assess how the frontier definition affects efficiency levels, rankings, and reference gaps.

4. Results

4.1 Sample and data

4.1.1 Study area

Research on electricity demand shows that data center growth is concentrated in some regions (IEA, 2025). This paper compares U.S. state-level jurisdictions in terms of electricity prices, grid conditions, generation mix, climate, water stress, and regional economic foundations.

The study covers the 50 U.S. states and the District of Columbia, giving 51 state-level geographic units; R001–R051 are internal codes assigned consistently in this paper. The overall data organization period is 2010–2025, and the location and efficiency analysis window is 2014–2025. The main data sources are organized by analytical use in Table 4-1, following the variable

definitions and data catalogs of public statistical agencies; QCEW provides establishment, employment, and wage statistics for detailed industries.

Table 4-1 Data sources and variable uses

Data provider or system	Variables covered	Website
U.S. Energy Information Administration (EIA)	Electricity prices and volatility; SAIDI/SAIFI; capacity, electricity inflows and outflows, losses, generation mix, and carbon intensity	eia.gov
U.S. Bureau of Economic Analysis (BEA)	State GDP, industry output, investment-related inputs, and regional economic controls	bea.gov
U.S. Bureau of Labor Statistics (BLS)	Establishments, employment, and wages for NAICS 518210 in QCEW; occupational employment and wages in Occupational Employment and Wage Statistics (OEWS)	bls.gov
U.S. Census Bureau	Population and demographic controls; digital-market proxies	census.gov
Integrated Postsecondary Education Data System (IPEDS)	Scale of higher education and completion counts in relevant fields	nces.ed.gov/ipeds
U.S. Drought Monitor	Annual state drought severity, measuring changes in water-resource pressure over time	droughtmonitor.unl.edu

The research design in Table 4-1 draws on 6 categories of sources: EIA provides electricity conditions, BEA supplies regional economic variables, and BLS provides industry and occupational statistics, while the U.S. Census Bureau, IPEDS, and the U.S. Drought Monitor add information on population, talent training, and water resource pressure, respectively. Each source supports variable design for the PCI, location model, and DEA according to its analytical role. BLS establishment, employment, and wage measures for NAICS 518210 maintain a consistent industry definition, although records from different agencies still require matching by region, observation year, and unit.

4.1.2 Data processing

Preprocessing aligns region identifiers, observation years, industry definitions, and units and checks duplicates, ranges, directions, continuity, and missing values. QCEW values suppressed for confidentiality, missing values, and zeros are recorded separately. For the analysis panel of 51 regions over 12 years from 2014–2025, variable distributions are described by means, standard deviations, medians, and ranges in Table 4-2.

Table 4-2 Panel descriptive statistics

Variable	<i>N</i>	Mean	<i>Standard deviation</i>	Median	Minimum	Maximum
Data-processing industry establishments	612	642.30	518.70	431.00	27.00	4218.00
Broadband subscription rate	612	86.20	7.30	87.10	60.80	97.10
Number of cloud providers	612	0.75	0.98	0.00	0.00	3.00
General corporate tax environment proxy	612	2.03	1.12	2.01	0.00	4.00
Real GDP per capita	612	67.40	20.80	63.10	37.20	176.50
Digital workforce share	612	3.11	1.10	2.93	1.02	7.18
Rental regional price parity (RPP)	612	98.60	19.50	94.20	58.70	168.20

Variable	N	Mean	Standard deviation	Median	Minimum	Maximum
Population	612	6.55	7.48	4.63	0.59	39.82

The right-skewed establishment counts in Table 4-2 have a mean of 642.30, a median of 431.00, and a maximum of 4218.00; a few regions with high counts raise the mean. Broadband subscription rates range from 60.80% to 97.10%, while the cloud-provider count has a median of 0 and a maximum of 3, indicating substantial differences in networks and service locations. Table 4-2 summarizes the 612 region-year observations used in the baseline estimation. Broadband and the digital talent share are expressed as percentages; GDP per capita is measured in thousands of constant 2024 U.S. dollars, population in millions, and rental RPP as an index. The subsequent models retain the establishment count scale, using population exposure to distinguish population size from industry agglomeration.

The location model uses the PCI, moderators, and continuous controls lagged by one year, with current population as the exposure. Broadband is measured by the household subscription rate, cloud-provider presence by the number of major providers active each year (0–3), and the general corporate tax environment on a 0–4 scale. Each module's annual coverage requirements, lag structure, and outcome definitions determine the analysis windows and period boundaries in Figure 4-1.

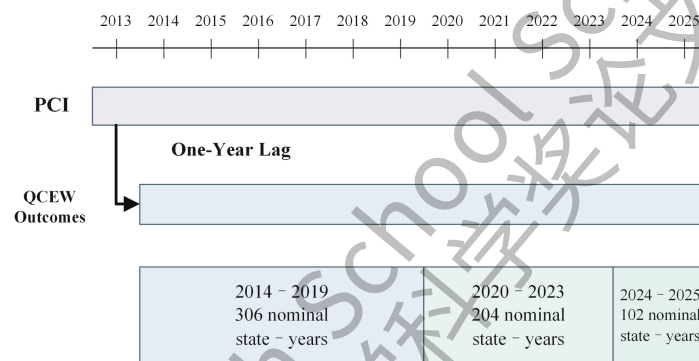


Figure 4-1 Annual coverage, period boundaries, and the QCEW measure

The windows in Figure 4-1 distinguish overall data organization over 2010–2025, complete PCI construction over 2013–2025, and location analysis over 2014–2025. The last window is divided into 2014–2019, 2020–2023, and 2024–2025, with actual estimation samples of 306, 204, and 102 region-year observations, respectively. The 2013 PCI provides the first lagged explanatory observation for the 2014 outcome. QCEW establishment stock remains the outcome, and period comparisons retain a common PCI scale. The model sample includes only regions and years with both current outcomes and previous-year covariates.

4.2 PCI measurement

4.2.1 PCI weighting

The AHM and CRITIC weights obtained from the comparison matrix and indicator information are coupled using Equation (9). Figure 4-2 summarizes the shares of individual indicators and dimensions: the outer ring represents the weights of the 11 indicators, and the inner ring the combined weights of the four dimensions.

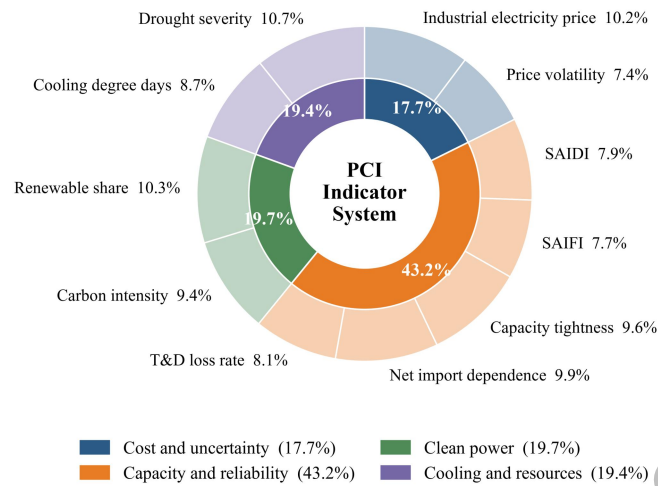


Figure 4-2 Coupled PCI weights

Capacity and reliability receive the largest coupled weight in Figure 4-2, at 43.2%, compared with 19.7% for clean electricity, 19.4% for cooling and resources, and 17.7% for cost and uncertainty. Individual weights range from approximately 7.4% to 10.7%. These weights give outages, capacity tightness, and electricity delivery a substantial role in the PCI while retaining information on prices, clean electricity, and cooling resources. Capacity and reliability comprise 5 indicators, so the dimension share also reflects the indicator count; comparisons of individual importance should use indicator weights. Regional comparisons should examine both the overall score and its components to determine whether a high PCI mainly reflects connection constraints, supply continuity, or operating costs.

The weighting results from the same comparison matrix and indicator information give AHM, CRITIC, and coupled weights for each indicator. Figure 4-3 compares them to examine how subjective importance judgments and differences in the data affect the final weights.

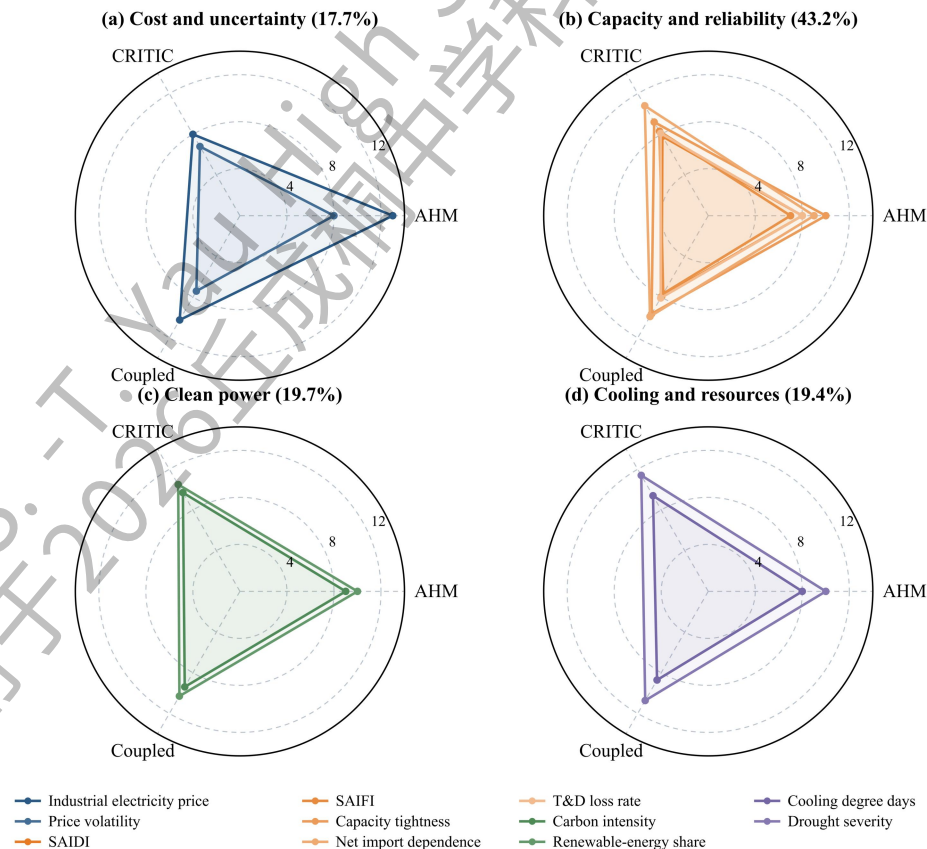


Figure 4-3 Comparison of AHM, CRITIC, and coupled indicator weights

For industrial electricity prices, Figure 4-3 reports an AHM weight of 13.0%, a CRITIC weight of 8.0%, and a coupled weight of 10.25%. The coupled weights for drought severity and renewable energy constraints are 10.73% and 10.29%, respectively. These relatively high weights are close, indicating that cooling provision, the energy mix, and operating costs all contribute to the PCI. The difference between the two initial weights for industrial electricity prices also indicates that importance judgments do not always align with indicator dispersion and correlation. In high-PCI regions, outage records and drought indicators can be used to identify constraints on supply reliability and cooling water, respectively.

4.2.2 PCI distribution

The PCI results for the complete indicator window of 2013–2025 describe regional differences in electricity and resource constraints mapped in Figure 4-4. Higher PCI values indicate stronger constraints in the selected measures.



Figure 4-4 PCI distribution, 2013–2025

The PCI values in Figure 4-4 differ substantially across the 51 regions in the 2013–2025 display window. Similar PCI values can reflect different sources of pressure, such as high operating costs or weaker supply reliability. The spatial distribution identifies regional differences, and the 11 indicators distinguish their specific sources. The subsequent model includes region and year fixed effects to account for relatively stable regional differences and common annual changes when examining the association between PCI changes and establishment stocks.

Annual summaries of the same PCI series track changes in constraints over 2013–2025 in Figure 4-5, complementing the regional differences in the spatial distribution.

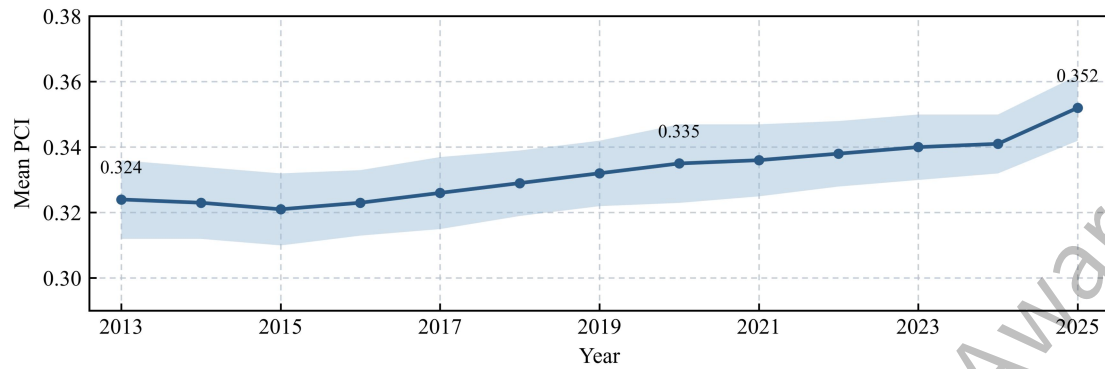


Figure 4-5 PCI trend, 2013–2025

The gradual rise in the mean PCI in Figure 4-5, from 0.324 in 2013 to 0.335 in 2020 and 0.352 in 2025, indicates an increase in overall constraints. Together with the regional differences in Figure 4-4, this trend suggests that location conditions vary both across regions and within a region over time. For expanding computing businesses, capacity and cooling conditions that initially met project needs may become tighter as demand grows. Connection timing and subsequent capacity expansion therefore require continued monitoring. Planning assessments should be updated accordingly: current electricity prices and reliability should be checked alongside progress in capacity construction relative to planned commissioning dates. Annual means indicate the overall pattern, while individual indicators and connection records identify the sources of pressure in a particular region.

4.3 Industry location

We use the location model to examine the association between lagged PCI and the stock of establishments recorded in QCEW, comparing results for the full period with those for individual periods.

4.3.1 Baseline results

The baseline and moderation-model coefficients are reported as PPML point estimates, standard errors, confidence intervals, and IRRs in Table 4-3. The baseline model uses 612 region-year observations and includes region and year fixed effects, standard errors clustered by region, lagged controls, and current population as the exposure. The PCI coefficient describes the association for a one-standard-deviation increase in constraints lagged by one year.

Table 4-3 Associations with PCI

Group	Dependent variable	Term	β	SE	95% confidence interval	IRR	<i>p</i>
	Industry agglomeration	<i>PCI lagged by one year</i>	-0.071	0.033	[-0.136, -0.006]	0.931	0.031
	Industry agglomeration	<i>PCI × broadband</i>	0.049	0.016	[0.018, 0.080]	1.050	0.002
	Industry agglomeration	<i>PCI × number of cloud providers</i>	0.044	0.015	[0.015, 0.073]	1.045	0.003
	Industry agglomeration	<i>PCI × corporate tax proxy</i>	0.019	0.018	[-0.016, 0.054]	1.019	0.291

Note: The baseline estimation uses 612 observations. The Holm-adjusted *p*-values for the broadband, cloud-provider presence, and tax interactions are 0.006, 0.006, and 0.291, respectively. The three interactions are adjusted as one family of tests.

The lagged PCI coefficient in Table 4-3 is -0.071, with a 95% confidence interval of [-0.136, -0.006], *p*=0.031, and IRR=0.931, corresponding to a 6.9% lower conditional mean establishment count. The broadband and cloud-provider presence interaction coefficients are 0.049 and 0.044, respectively, with Holm-adjusted *p*-values of 0.006 for both. The tax interaction is 0.019, with an

adjusted p -value of 0.291. These estimates indicate that the negative association with electricity constraints coexists with buffering associations for digital conditions; the tax interaction interval includes zero.

The baseline PCI coefficient and its 95% confidence interval in Table 4-3 define the estimated direction and range of the association between lagged electricity constraints and establishment stocks plotted in Figure 4-6.

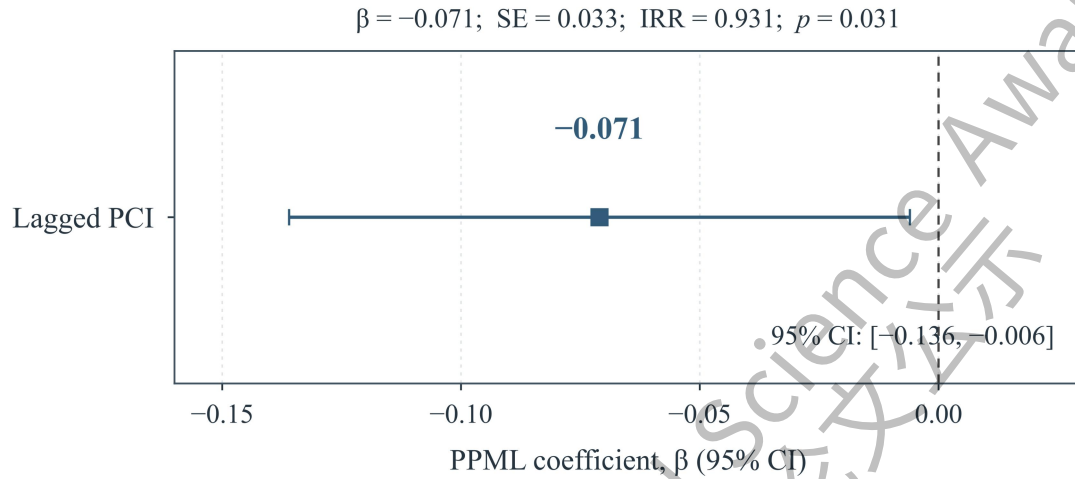


Figure 4-6 Baseline association with PCI

The baseline estimate in Figure 4-6 is negative: the coefficient of -0.071 and its 95% confidence interval of $[-0.136, -0.006]$ lie below zero. An IRR of 0.931 corresponds to a 6.9% lower conditional mean establishment count after accounting for population and other variables. Regions with stronger electricity constraints may face longer connection waits, greater investment in backup supply, and difficulties scheduling expansion. These issues can be examined when assessing their appeal to firms. The full-period coefficient summarizes the average association, while period-specific coefficients and their differences compare the recent and early periods.

4.3.2 Moderating associations

The three moderation-model coefficients in Table 4-3 describe the interactions of PCI with broadband, cloud-provider presence, and the general corporate tax proxy. Figure 4-7 compares the directions and intervals of these moderating associations.

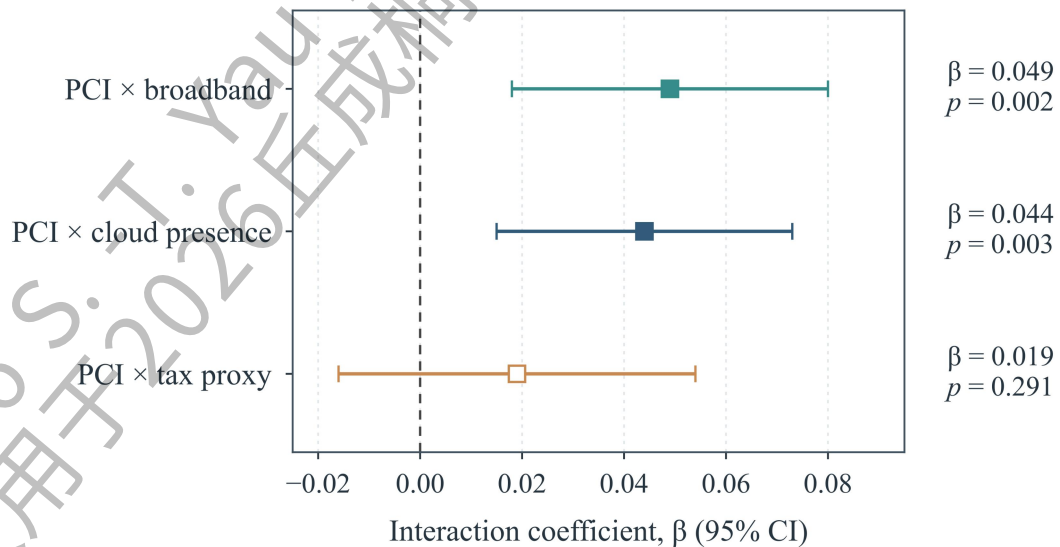


Figure 4-7 Moderating associations

The digital-condition interactions in Figure 4-7 are positive: the PCI interaction coefficients for broadband and cloud-provider presence are 0.049 and 0.044, with raw p -values of 0.002 and 0.003 and Holm-adjusted p -values of 0.006 for both. The tax interaction is 0.019, with $p=0.291$ and

an interval that includes zero. The weaker negative association between electricity constraints and establishment stocks in regions with better digital services is consistent with the expectation for digital conditions in Hypothesis 3. Broadband helps firms reach customers and organize business across regions; existing cloud services may provide technical support and opportunities for collaboration. These conditions may increase firms' willingness to accept connection waits or additional investment in electricity supply. Regional investment promotion can therefore consider electricity connections and network services together. The actual benefits of tax policies depend on when incentives are delivered, eligibility, and grid upgrade costs. Tax scores summarize the general environment, while project assessments can examine these specific terms.

4.3.3 Robustness checks

The alternative-specification coefficients are compared across different lags, regional trends, sample coverage, an equally weighted *PCI*, a negative binomial model, a $\log(1+y)$ fixed-effects model, and a future-*PCI* placebo specification in Figure 4-8. The count models and the $\log(1+y)$ model are displayed separately to retain their respective coefficient scales.

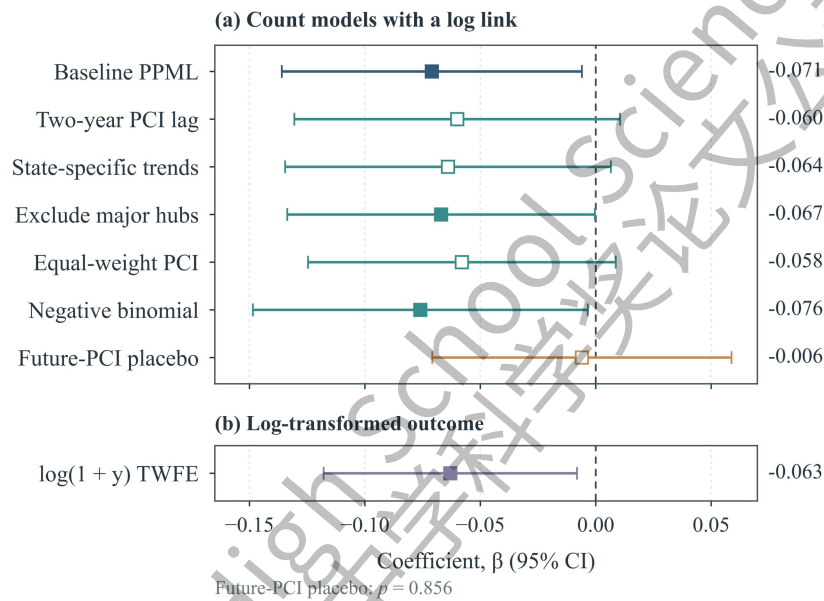


Figure 4-8 Robustness checks

In Figure 4-8, all 7 non-placebo coefficients are negative, with point estimates between -0.058 and -0.076 . The intervals for the baseline, the specification excluding industry hubs, the negative binomial model, and the $\log(1+y)$ model exclude zero, although some alternative specifications have wider intervals. The future-*PCI* coefficient is -0.006 , with $p=0.856$ and an interval that includes zero, indicating no clear lead association. The direction of the associations is consistent, while interval widths reflect differences in estimation precision. Coefficients from the count models and the $\log(1+y)$ model are interpreted on their own scales. Both sample coverage and model form affect the estimates.

In the component comparisons, capacity and reliability have the largest negative point estimate in absolute value ($\beta=-0.061$, $p=0.002$), followed by cost and uncertainty ($\beta=-0.042$, $p=0.027$). Clean electricity and cooling resources also have negative coefficients but lower estimation precision. This ordering suggests that supply continuity and operating costs may correspond to different location constraints. The following analysis examines changes in the overall *PCI* association across periods.

4.3.4 Multi-stage analysis of differences in associations

The multi-stage analysis examines whether the *PCI* association differs across 2014–2019, 2020–2023, and 2024–2025. Calculations using the coefficients and covariances yield the slopes, 95% confidence intervals, p -values, and conditional mean differences for the three periods, summarized in Table 4-4.

Table 4-4 Stage-specific *PCI* slopes

Period	N^*	β	Standard error	95% confidence interval	p	Conditional mean difference (%)
2014–2019	306	−0.021	0.029	[−0.078, 0.036]	0.4690	−2.1%
2020–2023	204	−0.078	0.032	[−0.141, −0.015]	0.0148	−7.5%
2024–2025	102	−0.186	0.046	[−0.276, −0.096]	<0.0001	−17.0%

Note: N is the actual estimation sample size: 306, 204, and 102 region-year observations across 51 regions in the three periods, respectively. Conditional mean differences are calculated as $100[\exp(\beta)-1]$, using the same PCI standardization scale in every period. Intervals and p -values are calculated from the coefficients and covariance matrix.

The period estimates in Table 4-4 become progressively more negative: the PCI slopes for 2014–2019, 2020–2023, and 2024–2025 are −0.021, −0.078, and −0.186, corresponding to conditional mean differences of −2.1%, −7.5%, and −17.0%. The early-period interval includes zero, whereas the middle- and recent-period intervals lie below zero. All three periods use the same PCI scale and contain 306, 204, and 102 observations, respectively; the recent window is the shortest. Direct tests of period differences use slope differences and the corresponding covariances.

The calculated slopes and confidence intervals for the three periods in Table 4-4 allow changes across periods to be compared on the same standardized PCI scale in Figure 4-9.

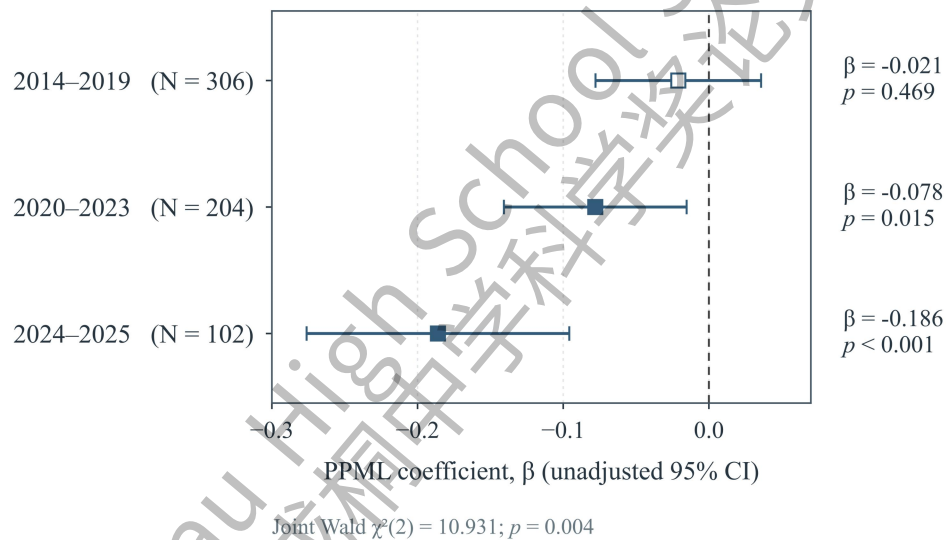


Figure 4-9 PCI associations by period

The estimates plotted in Figure 4-9 put the early-period coefficient close to zero at −0.021, followed by −0.078 for the middle period and −0.186 for the recent period. The early-period 95% interval is [−0.078, 0.036], compared with [−0.276, −0.096] for the recent period. These changes suggest that electricity conditions may be more closely associated with industry location as demand expands. Existing firms can use their current connection capacity, while new or expanding activities require additional electricity. Changes in constraints may therefore appear in establishment stocks after an adjustment period. Project connection, expansion, and commissioning records can be used to test this mechanism; differences between period slopes are compared directly below.

The differences calculated from the slopes and covariances for the three periods provide pairwise comparisons of the associations; Holm adjustment is applied to the three p -values as one family in Table 4-5. The 95% confidence intervals remain unadjusted pointwise intervals.

Table 4-5 Pairwise differences between periods

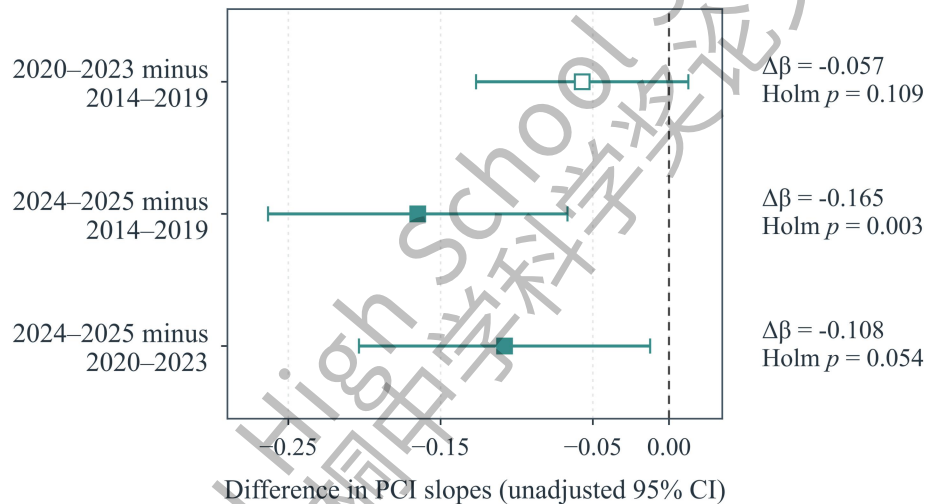
Contrast	$\Delta\beta$	Standard error	95% confidence interval	Raw p -value	Holm-adjusted p -value	Below 0.05?
P1 – P0	−0.057	0.036	[−0.127, 0.013]	0.1090	0.1090	No

Contrast	$\Delta\beta$	Standard error	95% confidence interval	Raw p -value	Holm-adjusted p -value	Below 0.05?
P2 – P0	-0.165	0.050	[-0.263, -0.067]	0.0010	0.0030	Yes
P2 – P1	-0.108	0.049	[-0.204, -0.012]	0.0268	0.0537	No

Note: P0=2014–2019, P1=2020–2023, and P2=2024–2025. The three pairwise tests are adjusted as one family using Holm's method; the confidence intervals remain unadjusted pointwise intervals. The joint test of equality of the three coefficients gives $\chi^2(2)=10.931$, $p=0.0042$.

The clearest of the three comparisons in Table 4-5 is the recent-to-early coefficient difference of -0.165 , with a pointwise 95% confidence interval of $[-0.263, -0.067]$ and a Holm-adjusted p -value of 0.0030 . The recent-to-middle difference is -0.108 , with a raw p -value of 0.0268 and an adjusted value of 0.0537 . The middle-to-early difference is -0.057 , with an adjusted p -value of 0.1090 . The joint test of equality of the three coefficients gives $\chi^2(2)=10.931$, $p=0.0042$. These multiple-comparison results indicate that stage differences are most evident between the recent and early periods.

The calculated slope differences and their pointwise 95% confidence intervals in Table 4-5 are displayed as the three comparisons relative to zero in Figure 4-10, and period differences are assessed alongside the Holm-adjusted p -values.



Pointwise 95% CIs; Holm p across three contrasts

Figure 4-10 Differences between period slopes

The recent-to-early difference plotted in Figure 4-10, -0.165 , lies below zero together with its interval, consistent with the Holm-adjusted p -value of 0.0030 . Although the pointwise interval for the recent-to-middle comparison excludes zero, its adjusted p -value of 0.0537 does not meet the 0.05 threshold. The middle-to-early adjusted p -value of 0.1090 likewise provides insufficient evidence of a clear difference. The plotted intervals should therefore be read alongside the multiple-comparison results in Table 4-5. Further investigation can prioritize the recent-to-early change. Organizing project records in the sequence of connection application, electricity availability, and commissioning can help examine how electricity conditions enter firms' expansion plans.

Supply capacity and reliability may explain this period pattern. Even regions with good fiber connections may face inadequate substation capacity or connection queues. When firms need to begin operating quickly, the timing of electricity availability is particularly relevant to expansion plans. Project application, connection, and commissioning records could help distinguish the roles of electricity conditions, industry composition, and long-term regional trends.

Changes in the establishment stock also depend on how firms organize their activities. Adding equipment to an existing campus may leave the establishment count unchanged, while the entry of small service firms raises it. Employment, payroll, and facility records can help distinguish changes in firm numbers from changes in operating scale.

4.4 Regional industry efficiency analysis

The DEA analysis of regional efficiency compares establishment counts, employment, and real payroll with the regional economy, digital workforce, and training base. It identifies regions where industry outputs are relatively well matched to these foundations and determines which outputs account for the largest gaps.

4.4.1 Efficiency distribution

The VRS-SBM uses means for 2014–2025 to define 51 decision-making units (DMUs). The reference frontier and mean efficiency score depend on this averaging window. Inputs are real GDP, employment in computer and mathematical occupations, and the number of CIP 11 program completions; outputs are data-processing industry establishments, employment, and real payroll. Super-SBM ranks regions on the frontier, while annual comparisons examine changes that period means may conceal. Corrected efficiency scores for 51 regions describe the spatial distribution of period-average efficiency in Figure 4-11.

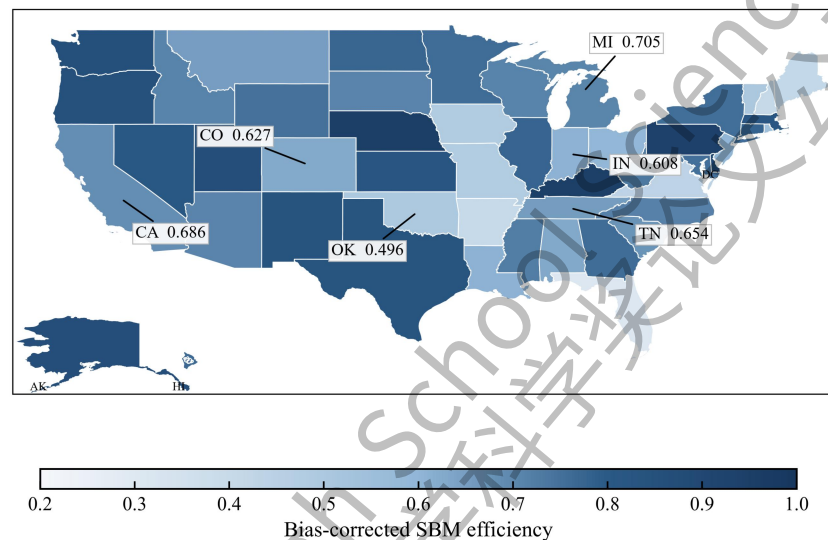


Figure 4-11 Distribution of bias-corrected SBM efficiency

Across the regions shown in Figure 4-11, corrected efficiency scores average about 0.692 and range from 0.311 to 0.955, indicating clear regional differences in relative output performance. The scores compare establishment counts, employment, and real payroll with the regional economy, digital workforce, and training base. Similar scores may arise from different industry structures: one region may have many small establishments, while another has fewer establishments but provides more jobs and payroll. Regional gaps therefore require examination of both scores and output composition. Where establishment counts approach the reference level but employment and payroll remain relatively low, further investigation should consider business activities, workforce size, and the mix of local job levels. Where a strong talent base coincides with relatively low industry outputs, the correspondence between business needs and training provision warrants attention. Regions with higher relative scores also allow comparisons of industry scale in relation to the regional resource base. These comparisons translate the mapped differences into specific research questions and guide the subsequent matching of firm data and project records. Figure 4-12 further separates input and output slacks to identify whether the gaps mainly concern establishments, employment, or payroll.

4.4.2 Input and output gaps and efficiency score correction

Super-SBM rankings and input and output slacks describe the relative ordering of frontier regions and their reference gaps in Figure 4-12. Examining these components helps explain the industry performance underlying the overall scores.

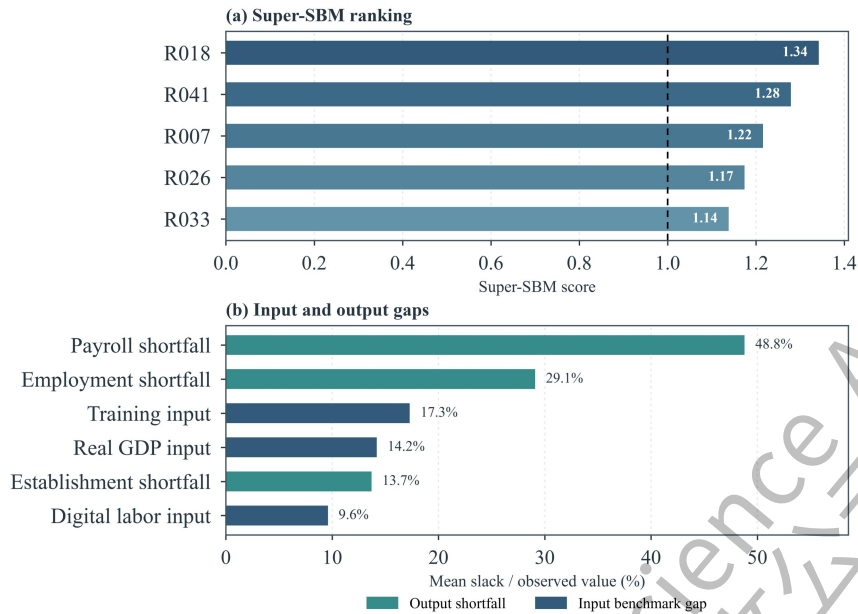


Figure 4-12 Super-SBM and slacks

The leading regions in Figure 4-12 are R018 (1.342), R041 (1.279), R007 (1.216), R026 (1.174), and R033 (1.138) in the Super-SBM ranking. Average input-side reference gaps are 14.2% for real GDP, 9.6% for the digital workforce, and 17.3% for training. Average output gaps are 13.7% for establishments, 29.1% for employment, and 48.8% for real payroll. Payroll has the largest gap, followed by employment. Employment and income gaps relative to the reference level are therefore larger than the establishment-count gap. Regional investigation should examine firm size, business activities, job skills, and pay structures. Inputs such as GDP describe broad regional foundations, and their reference gaps measure how industry outputs align with those foundations, not input reduction percentages.

The 51 pairs of original and corrected scores are compared in a paired scatterplot to show the direction of score adjustments in Figure 4-13.

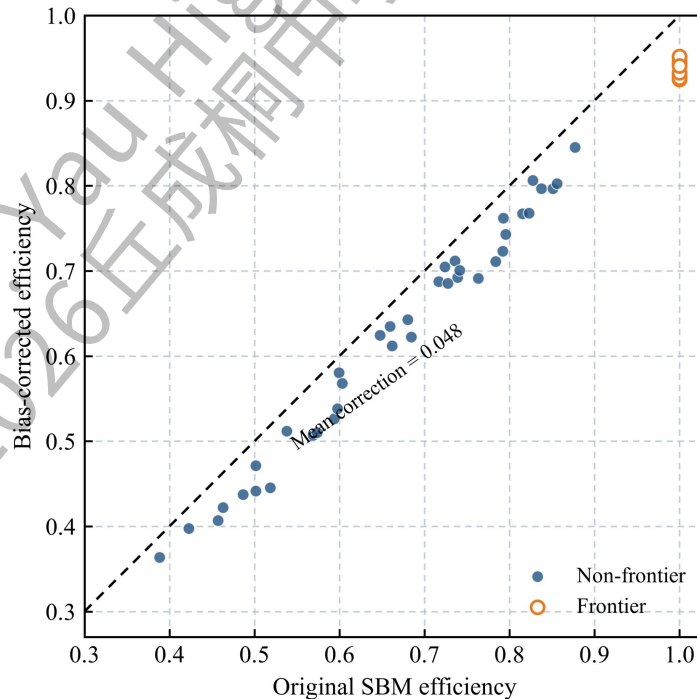


Figure 4-13 Paired scores before and after correction

The correction shown in Figure 4-13 lowers all 51 scores relative to their original values. Every point lies below the 45-degree line, and the mean decrease is about 0.048. The 45-degree line

denotes equal scores before and after correction; the distance from it indicates the adjustment for each pair. Pairing the scores preserves the correspondence between regions and reveals variation in adjustments beyond the mean change. Bias correction addresses the influence of a finite reference sample on relative scores. When reference regions change, regions previously close to the frontier face a different comparison, and their scores and rankings may also change. Regional evaluation should therefore examine both the corrected scores and whether gaps in individual outputs persist. A relatively stable payroll gap calls for closer examination of firm size, job skills, and pay structures. If changes mainly concern reference regions or rankings, employment, payroll, and establishment records should also be checked to determine whether industry performance has changed materially. Matching these records to the original and corrected scores helps distinguish structural output gaps from changes in the reference sample, making regional comparisons easier to interpret.

4.4.3 Association between efficiency and electricity constraints

The association analysis uses the reciprocal of the corrected score as the inefficiency measure and includes *PCI*, the moderator, their interaction, and rental RPP. Higher inefficiency indicates lower industry outputs relative to the selected regional foundations.

Summary statistics for 51 period-average decision-making units describe efficiency levels and output gaps; the association coefficients give the main parameters of three regression models in Table 4-6. Broadband, cloud-provider presence, and taxes enter separate models, each controlling for costs through rent RPP.

Table 4-6 DEA scores and associations with electricity constraints

Grp.	Indicator/model	Term	Estimate	SE	95% confidence interval	<i>p</i>
A	Mean original SBM score	—	0.728	—	—	—
A	Mean bias-corrected score	—	0.692	—	—	—
A	Number of frontier regions	—	13/51	—	—	—
A	Mean correction	—	0.036	—	—	—
A	Highest Super-SBM score	—	1.342	—	—	—
A	Relative payroll gap	—	48.8%	—	—	—
B	Broadband model	<i>PCI</i>	0.486	0.146	[0.200, 0.772]	<0.001
B	Broadband model	<i>PCI</i> × broadband	-0.205	0.085	[-0.372, -0.038]	0.016
C	Cloud-provider model	<i>PCI</i>	0.431	0.149	[0.139, 0.723]	0.004
C	Cloud-provider model	<i>PCI</i> × number of cloud providers	-0.174	0.073	[-0.317, -0.031]	0.017
D	Corporate tax model	<i>PCI</i>	0.394	0.158	[0.084, 0.704]	0.013
D	Corporate tax model	<i>PCI</i> × corporate tax proxy	-0.112	0.078	[-0.265, 0.041]	0.151

The original and corrected mean efficiency scores in Table 4-6 are 0.728 and 0.692, respectively, a decrease of 0.036. There are 13/51 original frontier regions, the highest Super-SBM score is 1.342, and the relative payroll gap is 48.8%. The descriptive statistics also report original and corrected medians of 0.701 and 0.668, an original interquartile range of 0.582 – 0.876, and a corrected range of 0.311 – 0.955. The *PCI* coefficients in the three models are 0.486, 0.431, and 0.394, respectively, each indicating lower relative output performance under stronger electricity constraints. The broadband and cloud-provider interaction coefficients are -0.205 and -0.174, while

the interval for the tax interaction, -0.112 , includes zero. Examining efficiency levels alongside output gaps helps distinguish a shortfall in industry scale from weaker correspondence between regional foundations and the structure of employment and payroll.

The three inefficiency regressions in Table 4-6 report coefficients and intervals for the PCI main effects and its interactions with broadband, cloud-provider presence, and taxes. Their directions are compared in Figure 4-14. Inefficiency is defined as the reciprocal of the corrected score; higher values indicate lower relative output performance.

The inefficiency regressions plotted in Figure 4-14 yield PCI coefficients for the broadband, cloud-provider, and tax models of 0.486 ($p<0.001$), 0.431 ($p=0.004$), and 0.394 ($p=0.013$), respectively. The broadband and cloud-provider interaction coefficients are -0.205 ($p=0.016$) and -0.174 ($p=0.017$). The tax interaction is -0.112 ($p=0.151$), with an interval that includes zero. These association directions are consistent with H2 and the digital-condition component of H3: stronger electricity constraints accompany greater inefficiency, and better digital conditions weaken that association. Regional investigation can jointly examine electricity supply, customer connectivity, and local skill requirements.

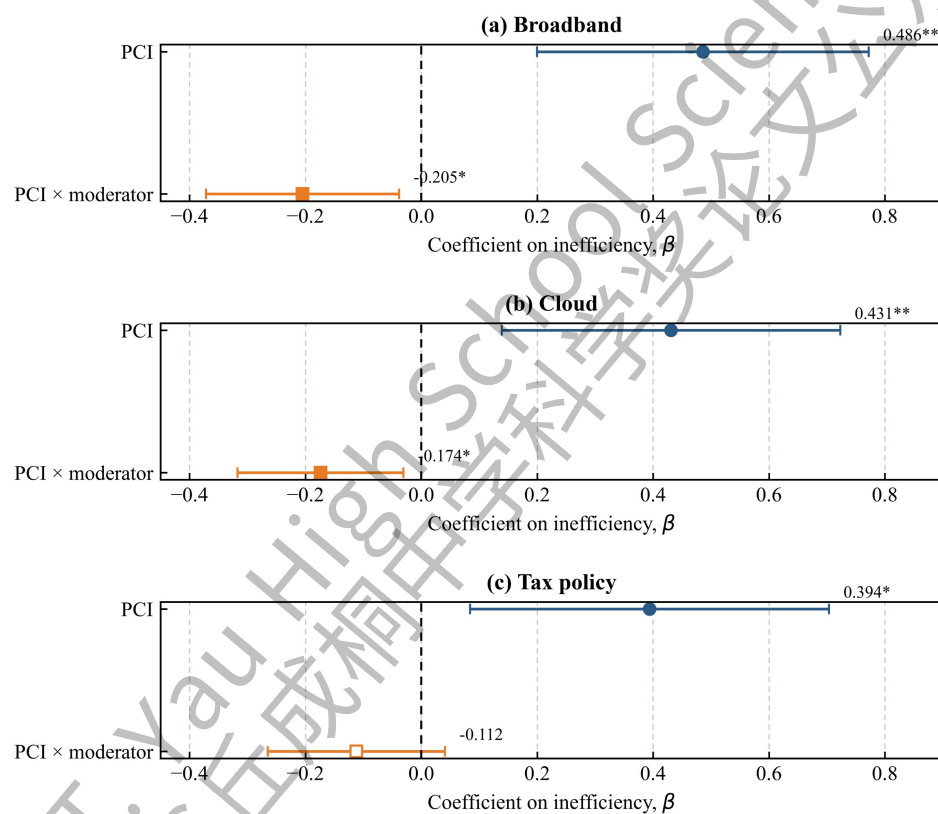


Figure 4-14 Association between efficiency and PCI

4.4.4 Robustness of annual frontiers

Annual efficiency comparisons track yearly mean and median scores, the number of frontier regions, and the correlation between annual and period-average rankings in Figure 4-15. Together these measures help identify annual changes that period averages may conceal.

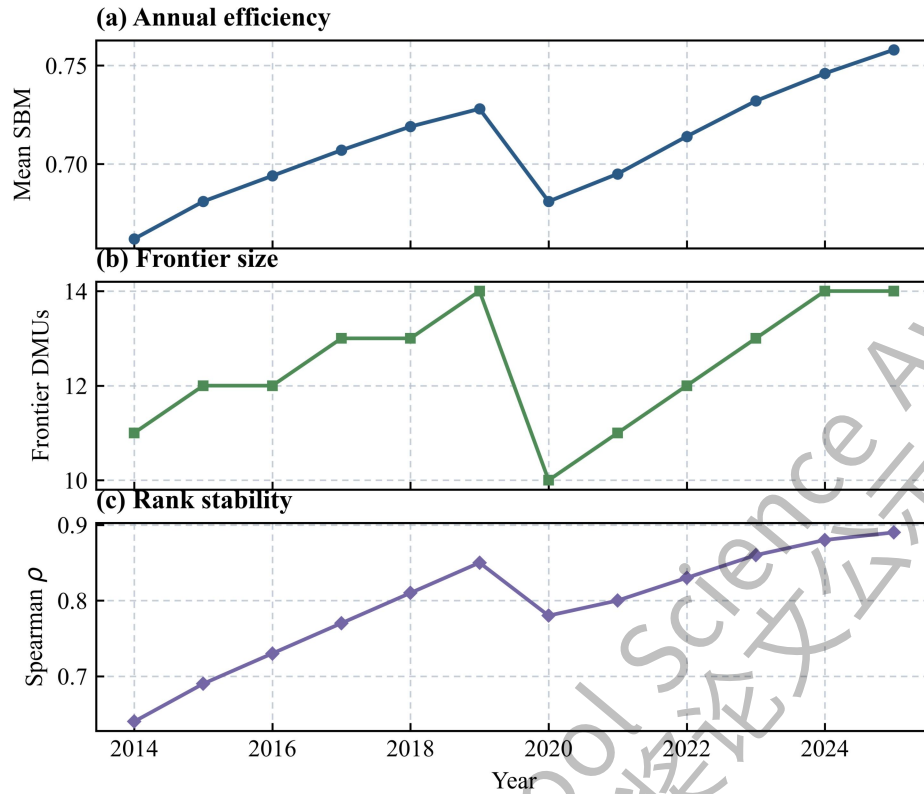


Figure 4-15 Annual frontier stability

Mean annual efficiency in Figure 4-15 rises from 0.662 in 2014 to 0.728 in 2019, falls to 0.681 in 2020, and then recovers to 0.758 in 2025. The number of frontier decision-making units varies between 10 and 14. The Spearman correlation between annual and period-average rankings increases from 0.64 to 0.89. Later annual rankings are thus closer to the period-average ordering, although positions still vary by year.

Mean annual efficiency, frontier size, and ranking correlation in Figure 4-15 describe different aspects of performance. The mean describes the overall relative level, frontier size counts regions that reach the reference boundary in a given year, and the correlation measures the correspondence between annual and longer-term rankings. Annual scores depend on both local output changes and the performance of reference regions. A region may add employment and payroll yet receive a lower relative score if its reference regions grow faster. Conversely, an improvement in its relative score should also be interpreted alongside actual output records. Period averages summarize more persistent regional differences, while annual records help establish when business expansion, workforce adjustment, and payroll changes occur. Regional planning can therefore track establishments, employment, and payroll over successive years. For regions with large ranking changes, investigation should identify the outputs and types of firms that account for the movement. For regions that remain close to the reference level, it should examine the fit between electricity arrangements, network services, and skill requirements. These records extend evaluation beyond a single year's ranking to industry structure and its adjustment over time.

4.5 Combined findings

The preceding PCI calculations, model association coefficients, multi-stage difference tests, and DEA analysis of regional efficiency inform the main findings and practical implications summarized in Table 4-7.

Table 4-7 Analytical findings

Analysis	Data and method	Result/status	Interpretation
PCI measurement	Eleven indicators; AHM–CRITIC	Capacity and reliability weight: 43.2%	Higher PCI indicates stronger constraints; the complete indicator window begins in 2013.
Industry location	QCEW and PPML design	<i>PCI coefficient: $\beta = -0.071$</i>	Conditional association with the establishment stock.
Moderation model	Separate interaction models	Positive broadband and cloud-service interactions; imprecise tax estimate	Moderating associations conditional on other factors.
Regional industry efficiency analysis	SBM for 51 DMUs	Mean score of 0.728 before correction and 0.692 after correction	How industry outputs compare with regional foundations and the associated relative gaps.
Period comparison	PPML stage interactions and Wald tests of differences	Recent-to-early β difference: -0.165 ; Holm $p=0.0030$	Direct comparison of coefficients across periods.

The capacity and reliability weight in Table 4-7 is 43.2%, making it the largest component of the PCI. The baseline coefficient of -0.071 summarizes the association in the location model, while the recent-to-early difference of -0.165 and Holm $p=0.0030$ provide a direct comparison across stages. The DEA values of 0.728 and 0.692 summarize efficiency for interpreting relative output differences beyond establishment counts. Together, these analyses suggest a sequence for subsequent regional investigation: use PCI components to identify prices, capacity, reliability, and resource pressures requiring verification; examine establishments, employment, and payroll to characterize industry performance; and assess the explanations using project records on connection dates, cost allocation, and employment commitments.

5. Discussion

5.1 Electricity constraint mechanisms and regional decisions

The multi-stage analysis finds a stronger negative association between electricity constraints and industry agglomeration in the recent period than in the early period. A possible explanation for this pattern lies in the different timelines of infrastructure construction and demand growth. Server deployment and business expansion often need to respond quickly to markets, whereas substation expansion, line upgrades, and connection approval require longer implementation schedules. When demand is relatively stable, firms can adjust equipment and loads within existing capacity. When demand rises sharply, new projects and expansions by existing firms seek capacity at the same time. Previously minor connection difficulties may then affect the order in which projects begin operating. A region's attractiveness therefore depends both on the supply capacity it can deliver and on whether electricity arrangements fit firms' operating schedules.

Electricity reliability also affects ongoing operating costs. Firms providing continuous hosting, data processing, and cloud services face outage risks that may increase their requirements for backup power, redundant systems, equipment maintenance, and emergency arrangements. A region with low electricity prices may still impose substantial operating costs if interruptions are frequent or electricity becomes available late. The relatively high weight assigned to capacity and reliability in the PCI is consistent with this interpretation. Comparing expected connection dates, available capacity, outage records, and electricity prices gives site decisions a basis in verifiable supply conditions.

The local economic benefits of establishment agglomeration also depend on firm organization and business structure. The same number of establishments may belong to a few large operators or many small service providers, with different activities, staffing needs, and pay levels. The larger employment and payroll gaps suggest that regions that attract projects should continue to track average firm size, the share of technical jobs, and local service procurement. Training provision should reflect the different skills required for operations, networking, cooling, and electrical roles. Simply increasing completions in related programs may not produce a corresponding increase in the region's ability to support the industry.

Broadband and cloud-provider presence may improve firms' operating conditions under electricity constraints through customer access, business connections, and specialized services. Firms must operate equipment continuously while delivering computing services reliably to customers. Better network and cloud-service provision may give regional firms more opportunities to use existing electricity capacity for business. Electricity planning and digital infrastructure therefore warrant joint assessment: connection arrangements determine when a project can operate, while network and service provision influence the activities it can undertake and the customers it can serve.

Identifying the main constraints helps regions choose appropriate measures. Where outages and capacity indicators drive a high PCI, regions should first examine capacity expansion and reliability upgrades. Where drought and cooling pressures dominate, regions should compare cooling technologies, water arrangements, and seasonal risks. Where network conditions are favorable but connection times are long, electricity construction schedules can guide project phasing. Where electricity conditions are favorable but industry employment is low, further investigation should consider customer demand, firm size, and skill matching.

The effects of tax incentives also depend on how closely the benefits match the project's actual costs and schedule. Regions with similar general tax scores may differ in payment timing, eligibility thresholds, employment conditions, and the allocation of grid-upgrade costs. A comparison of the complete project arrangement can therefore establish when usable electricity becomes available, when the firm incurs construction and operating costs, when incentives are received, and what employment or payroll commitments are required. These milestones, together with the responsible parties, help assess whether an incentive addresses the firm's specific constraints.

Regional screening can use verifiable records of electricity conditions, digital services, and expected industry outputs. Comparisons should state the source and period of validity of each item of evidence. Revising candidate rankings and construction arrangements as connection capacity,

project progress, and business needs change allows industry planning to respond to actual conditions.

5.2 Implications for practical planning

The analysis suggests that land and incentive decisions should follow a joint assessment of electricity and digital conditions. Alongside prices, this means examining available capacity, reliability, dependence on electricity imports, grid losses, and connection timing, as well as fiber access and workforce availability. Low prices offer little support for a firm's planned start of operations if electricity is still unavailable. Regional comparisons would better reflect project conditions if they recorded expected connection dates and responsibility for upgrade costs.

Assessing a specific incentive package requires a review of the conditions applicable to the project and its contractual terms.

Project records should include statutory eligibility, incentive-award timing, electricity-supply milestones, responsibility for grid-upgrade costs, employment and payroll commitments, clean-electricity procurement, and water-use obligations. Arranging these records by the effective date of each term would support further analysis of policy associations with project progress, costs, and regional outputs.

Subsequent policy assessment should record the specific measures and comparison conditions, define the observation period, and identify who bears the costs. These records can link changes in electricity constraints to changes in outputs such as employment and payroll.

5.3 Limitations and further research

First, the state-level analysis uses QCEW establishment counts to represent the development of data-center-related industries. This stock measure does not directly capture new establishment entry, individual data centers, installed capacity, or computing power. State-level averages may therefore hide local differences in grid access, land, and water conditions. Future research can combine regional data with facility-level records. An entry analysis using BDS establishment births would require a sample window matched to the available birth records.

Second, combining electricity and resource conditions in the Power Constraint Index makes regional comparisons easier, but it may hide differences between specific constraints. For example, high electricity prices and limited grid capacity may affect industry development in different ways. Future studies can compare their effects by examining these components separately.

Third, the recent window in the period analysis is relatively short. The estimates for 2024–2025 consequently use fewer annual observations than those for the earlier period. The results can indicate changes in the relationship over time, but they do not establish a permanent structural change. More years of data are needed to test whether the recent pattern continues. Further analysis can assess sensitivity to nearby period boundaries while holding the regional sample fixed.

Finally, some moderators rely on simple proxies: annual cloud-provider counts measure only regional presence, while the 0–4 tax variable represents the general business-tax environment. They do not capture the timing, scale, or detailed rules of individual projects. Future research can use cloud-region opening dates, interconnection records, project incentives, electricity contracts, water use, permits, and site conditions. These data may help explain more directly why electricity constraints affect regions differently.

6. Conclusion

This paper links electricity constraint measurement to multi-stage industry location analysis and regional efficiency evaluation. The PPML analysis finds a negative conditional association between electricity constraints and data-processing industry agglomeration, with the clearest difference between the recent and early periods. For 2024–2025, a one-standard-deviation increase in PCI corresponds to a 17.0% lower conditional mean establishment count; the early-period estimate is not statistically distinguishable from zero. This association is not interpreted as a causal effect. Capacity and reliability form the largest weighted dimension of the PCI, suggesting that location assessment should consider when electricity becomes available and whether supply can be maintained.

The association between electricity constraints and industry agglomeration is weaker in regions with better-developed broadband and cloud-provider services. The moderating relationship for tax conditions remains unclear. The DEA analysis finds larger relative gaps in employment and payroll than in establishment counts, suggesting a need to evaluate firm size, jobs, and income structures alongside establishment agglomeration.

Regional planning can combine an assessment of electricity availability dates, reliability, and digital support with an examination of firm size and job structure to identify industry development needs.

The study compares how electricity conditions relate to industry location across stages of development and interprets regional differences through employment and payroll performance. Electricity supply and digital services describe the conditions under which projects operate, while industry scale, employment, and income measure development outcomes.

References

- Awaysheh, A., Chen, C. J., & Wu, O. Q. (2026). Does renewable energy renew energy efficiency? *Management Science*. Advance online publication. <https://doi.org/10.1287/mnsc.2023.02165>
- Ayyildiz, E., Yildirim, B., & Aydin, N. (2025). Location selection methodology for data center with renewable energy integration. *Renewable Energy*, 250, 123270. <https://doi.org/10.1016/j.renene.2025.123270>
- Azarifar, M., Arik, M., & Chang, J.-Y. (2024). Liquid cooling of data centers: A necessity facing challenges. *Applied Thermal Engineering*, 247, 123112. <https://doi.org/10.1016/j.applthermaleng.2024.123112>
- Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, 151, 103745. <https://doi.org/10.1016/j.jfineco.2023.103745>
- Bae, J., Kapuscinski, R., & Silberholz, J. (2026). Cost-saving synergy: Energy stacking in battery energy storage systems. *Management Science*, 72(5), 3906–3924. <https://doi.org/10.1287/mnsc.2022.03477>
- Bick, A., Blandin, A., & Deming, D. J. (2026). The rapid adoption of generative AI. *Management Science*. Advance online publication. <https://doi.org/10.1287/mnsc.2025.02523>
- Bogetoft, P., & Kerstens, P. J. (2024). Distinguishing useful and wasteful slack. *Operations Research*, 72(4), 1556–1573. <https://doi.org/10.1287/opre.2022.2415>
- Cáceres, C. R., Sandberg, M., & Sotoca, A. (2024). Planning data center locations in Swedish municipalities. A comparative case study of Luleå and Stockholm. *Cities*, 150, 105063. <https://doi.org/10.1016/j.cities.2024.105063>
- Cardinaels, E., Borst, S., & van Leeuwen, J. S. H. (2024). Heavy-traffic universality of redundancy systems with assignment constraints. *Operations Research*, 72(4), 1539–1555. <https://doi.org/10.1287/opre.2022.2385>
- Chen, S., Lei, J., & Moinszadeh, K. (2024). Cost optimization in cloud computing: Capacity reservation for intermittent random demand surges. *Production and Operations Management*, 33(6), 1265–1284. <https://doi.org/10.1177/10591478241251614>
- Clement, N., & Arce, D. (2025). Dynamics of shared security in the cloud. *Information Systems Research*, 36(2), 916–943. <https://doi.org/10.1287/isre.2023.0256>
- Correia, S., Guimarães, P., & Zylkin, T. (2020). Fast Poisson estimation with high-dimensional fixed effects. *The Stata Journal*, 20(1), 95–115. <https://doi.org/10.1177/1536867X20909691>
- Diakoulaki, D., Mavrotas, G., & Papayannakis, L. (1995). Determining objective weights in multiple criteria problems: The CRITIC method. *Computers & Operations Research*, 22(7), 763–770. [https://doi.org/10.1016/0305-0548\(94\)00059-H](https://doi.org/10.1016/0305-0548(94)00059-H)
- Donnellan, D., & Lawrence, A. (2024). *Annual outage analysis 2024*. Uptime Institute. <https://intelligence.uptimeinstitute.com/resource/annual-outage-analysis-2024>
- Dupont, C., Heijmans-Kuryatnikova, O., Ghiassi-Farrokhfal, Y., & Bunn, D. (2026). The unintended carbon impacts of large-scale electricity storage. *Manufacturing & Service Operations Management*, 28(2), 663–685. <https://doi.org/10.1287/msom.2024.1411>
- Esram, N. W., & Elliott, R. N. (2024). *Turning data centers into grid and regional assets: Considerations and recommendations for the federal government, state policymakers, and utility regulators*. American Council for an Energy-Efficient Economy. <https://www.aceee.org/resource/turning-data-centers-grid-and-regional-assets-considerations/>
- Gao, Z., Alshehri, K., & Birge, J. R. (2024). Aggregating distributed energy resources: Efficiency and market power. *Manufacturing & Service Operations Management*, 26(3),

- 834–852. <https://doi.org/10.1287/msom.2021.0539>
- Gulati, P., Marchetti, A., & Puranam, P. (2026). Collaborative work management technologies and managerial intensity in U.S. corporations: An examination. *Management Science*. Advance online publication. <https://doi.org/10.1287/mnsc.2023.04127>
- Guo, H., Liu, Y., & Nault, B. R. (2024). Join up or stay away? Coalition formation for critical IT infrastructure. *Information Systems Research*, 35(3), 1344–1362. <https://doi.org/10.1287/isre.2021.0463>
- Hosseini, L., Farahani, M. H., & Kumar, S. (2026). Is multicloud the future? Desirability of compatibility in cloud computing market. *Management Science*. Advance online publication. <https://doi.org/10.1287/mnsc.2023.01359>
- Hou, T., Chen, Y., Feng, J., & Li, J. J. (2026). Digital transformation learning: The value of learning from the digitalization of traditional industries through shared digital suppliers. *MIS Quarterly*. Advance online publication. <https://doi.org/10.25300/misq/2026/19113>
- International Energy Agency. (2025). *Energy and AI*. <https://www.iea.org/reports/energy-and-ai>
- Jain, T., Hazra, J., & Gopal, R. (2025). Procuring cloud services: An economic analysis of multi-cloud strategy. *Production and Operations Management*, 34(9), 2793–2813. <https://doi.org/10.1177/10591478251326421>
- Jin, W., & McElheran, K. (2025). Economies before scale: I.T. strategy and performance dynamics of young U.S. businesses. *Management Science*, 71(9), 7449–7473. <https://doi.org/10.1287/mnsc.2019.03116>
- Kundu, A., Anderson, S. J., & Ramdas, K. (2024). Disruptions, redundancy strategies, and performance of small firms: Evidence from Uganda. *Management Science*, 70(12), 8265–8283. <https://doi.org/10.1287/mnsc.2023.4978>
- Link, S., Menkhoff, M., Peichl, A., & Schüle, P. (2024). Downward revision of investment decisions after corporate tax hikes. *American Economic Journal: Economic Policy*, 16(4), 194–222. <https://doi.org/10.1257/pol.20220530>
- Liu, R. P., Mellou, K., Gong, E. X.-Y., Li, B., Coffee, T., Pathuri, J., Simchi-Levi, D., & Menache, I. (2025). Efficient cloud server deployment under demand uncertainty. *Manufacturing & Service Operations Management*, 27(2), 425–440. <https://doi.org/10.1287/msom.2023.0372>
- Ma, L., Chen, H., Yan, H., Li, W., Zhang, J., & Zhang, W. (2018). Post evaluation of distributed energy generation combining the attribute hierarchical model and matter-element extension theory. *Journal of Cleaner Production*, 184, 503–510. <https://doi.org/10.1016/j.jclepro.2018.02.223>
- Mariani, F., Ciommi, M., & Recchioni, M. C. (2024). A new class of composite indicators: The penalized power mean. *European Journal of Operational Research*, 317(3), 1015–1035. <https://doi.org/10.1016/j.ejor.2024.04.032>
- Matta, M., Kwon, H. E., Ravindran, K., & Ray, G. (2026). Seizing growth opportunities: A risky business? Effects of cloud sourcing on mergers and acquisitions. *MIS Quarterly*, 50(1), 269–298. <https://doi.org/10.25300/misq/2025/18085>
- Meeks, R., Pless, J., & Wang, Z. (2026). Can digitalization improve public services? Evidence from innovation in energy management. *Management Science*. Advance online publication. <https://doi.org/10.1287/mnsc.2024.06349>
- Mergoni, A., Emrouznejad, A., & De Witte, K. (2025). Fifty years of Data Envelopment Analysis. *European Journal of Operational Research*, 326(3), 389–412. <https://doi.org/10.1016/j.ejor.2024.12.049>
- North American Electric Reliability Corporation. (2024). *2024 long-term reliability assessment*. https://nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/NERC_Long%20Term%20Reliability%20Assessment_2024.pdf

- Peng, X., Wu, O. Q., & Souza, G. C. (2024). Renewable, flexible, and storage capacities: Friends or foes? *Manufacturing & Service Operations Management*, 26(5), 1730–1749. <https://doi.org/10.1287/msom.2023.0068>
- Qu, S., Duan, K., Guo, Y., Feng, Y., Wang, C., & Xing, Z. (2024). Real-time optimization of the liquid-cooled data center based on cold plates under different ambient temperatures and thermal loads. *Applied Energy*, 363, 123101. <https://doi.org/10.1016/j.apenergy.2024.123101>
- Secomandi, N. (2025). OM Forum—Asset-level perspectives on the clean energy transition. *Manufacturing & Service Operations Management*, 27(1), 8–20. <https://doi.org/10.1287/msom.2024.1421>
- Shehabi, A., Smith, S. J., Hubbard, A., Newkirk, A., Lei, N., Siddik, M. A. B., Holecek, B., Koomey, J., Masanet, E., & Sartor, D. (2024). *2024 United States data center energy usage report*. Lawrence Berkeley National Laboratory. <https://doi.org/10.71468/PIWC7Q>
- Siegloch, S., Wehrhöfer, N., & Etzel, T. (2025). Spillover, efficiency, and equity effects of regional firm subsidies. *American Economic Journal: Economic Policy*, 17(1), 144–180. <https://doi.org/10.1257/pol.20220667>
- U.S. Department of Energy, Secretary of Energy Advisory Board. (2024). *Recommendations on powering artificial intelligence and data center infrastructure*. <https://www.energy.gov/sites/default/files/2024-11/Powering%20AI%20and%20Data%20Center%20Infrastructure%20Recommendations%20July%202024.pdf>
- Yoo, Y., Henfridsson, O., Kallinikos, J., Gregory, R., Burtch, G., Chatterjee, S., & Sarker, S. (2024). The next frontiers of digital innovation research. *Information Systems Research*, 35(4), 1507–1523. <https://doi.org/10.1287/isre.2024.editorial.v35.n4>
- Zhai, Y., Stinchcombe, M. B., & Whinston, A. B. (2026). Strategic throttling in large cloud computing platforms. *Information Systems Research*. Advance online publication.