Can Texas's Renewables Keep Up? Evidence on AI Data Center Demand and Renewable Supply Response in ERCOT

Master of Quantitative Economics
University of California, Los Angeles

Ketki Prasad Sawant

Faculty Advisor: Professor Randall R. Rojas

June 2026

Contents

1 Abstract **2**

2 Introduction **3**

3 Background and Literature Review **5**

3.1 ERCOT Market Structure and Zonal Pricing 5

3.2 AI Data Centers and Electricity Demand 5

3.3 Renewable Energy and Price Formation 6

4 Data **6**

4.1 Data Sources and Construction 6

4.2 Variable Definitions 7

4.3 Summary Statistics 7

4.4 Proxy Validation for Data Center Demand 7

5 Methodology **10**

5.1 Model 1: OLS Regression Framework 10

5.2 Model 2: Quantile Regression 11

5.3 Model 3: Vector Autoregression and Granger Causality 12

5.4 Model 4: Random Forest 12

5.5 Placebo Tests and Magnitude Validation 13

6 Results **13**

6.1 Measuring AI Electricity Demand 13

6.2 Model 1: Baseline, Interaction, and Structural Break Results 13

6.3 Model 2: Quantile Regression Results 14

6.4 Model 3: VAR and Granger Causality Results 16

6.5 Model 4: Random Forest Feature Importance 16

6.6 Magnitude Validation and Placebo Tests 17

7 Forward Scenario Analysis **19**

8 Conclusion **20**

9 References **21**

1. Abstract

This paper provides market-based evidence on the electricity demand consequences of artificial intelligence data center growth and tests whether West Texas renewable energy generation can effectively offset that demand. Using hourly ERCOT data from January 2023 through December 2025 (25,202 observations), we estimate that AI data center expansion added approximately 3.74 terawatt-hours of annual electricity consumption to the North Texas grid by 2025, implying 534 megawatts of installed capacity — consistent with independently reported industry figures from CBRE (2025). The identification strategy exploits a distinctive baseload signature in ERCOT's North zone, where Dallas–Fort Worth data center growth produced off-peak load increases exceeding on-peak growth by 2.71 percentage points in 2024 — a pattern unique to 24/7 computing loads and absent in all other Texas zones.

Four complementary econometric models reveal a renewable saturation effect: while wind and solar generation significantly suppress the West–North basis spread, their effectiveness diminishes sharply as North zone load increases, with solar's impact approaching zero at high demand levels. The basis spread shifted structurally upward by $5.43 per megawatt-hour following 2024, a regime change consistent with rising demand and not explained by supply-side factors. Quantile regression identifies two distinct market regimes: data center demand drives congestion under normal conditions, while cold weather shocks dominate extreme events. Vector autoregression and Granger causality tests confirm that renewable intermittency predicts daily basis movements while demand operates as structural background pressure. A random forest model independently ranks solar generation as the most important predictor of the basis spread (36.6%), with North zone load ranking fifth (4.6%), confirming demand's role as a structural rather than volatility-driving factor.

Forward scenario analysis, adjusted for geographic and temporal deliverability constraints, suggests that under base growth assumptions, deliverable renewable supply and incremental data center demand converge to near-zero margin by 2030, with a 5.40 terawatt-hour deficit emerging under high growth. These results indicate that the challenge facing the Texas grid is not renewable energy volume but reliable, geographically matched, around-the-clock supply — a dimension along which intermittent wind and solar face inherent structural limitations.

Keywords: electricity markets, ERCOT, artificial intelligence, data centers, renewable energy, basis spread, price congestion, saturation effects.

2. Introduction

Artificial intelligence is reshaping the global economy, and nowhere is this transformation more visible than in electricity markets. The computational infrastructure powering large language models, image recognition systems, and cloud-based AI services requires vast and uninterrupted supplies of electricity. Data centers — the physical facilities housing this infrastructure — consumed approximately 183 terawatt-hours (TWh) of electricity in the United States in 2024, accounting for more than four percent of total national consumption (International Energy Agency [IEA], 2025). By 2030, that figure is projected to reach 426 TWh, driven primarily by continued AI adoption (IEA, 2025). Goldman Sachs (2024) estimates that global data center power demand will grow by 160 percent over the same period. These projections, however, are largely forward-looking and based on announced capacity and industry assumptions. Less is known about how much AI-driven demand has already materialized in observed electricity market data — and what its consequences for grid dynamics have already been.

This paper addresses that gap using Texas as a real-world laboratory. The Electric Reliability Council of Texas (ERCOT) operates one of the most data-rich and competitively structured electricity markets in the United States, with transparent zonal pricing that allows researchers to observe supply and demand conditions at hourly resolution. Texas is simultaneously home to the country's largest installed base of wind and solar generation and one of its fastest-growing data center markets. The Dallas–Fort Worth (DFW) metropolitan area — served by ERCOT's North pricing zone — is the third-largest data center market in the United States, with a total installed inventory of 1,067 MW and 699 MW currently under construction as of the second half of 2025 (CBRE, 2025). ERCOT received 225 new large load interconnection requests in 2025, representing a 270 percent increase in megawatt demand since January of that year, with data centers comprising approximately 73 percent of the 226 gigawatts (GW) of large loads currently seeking interconnection (Electric Reliability Council of Texas, 2025). West Texas, served by ERCOT's West pricing zone, hosts the majority of the state's wind capacity and a rapidly expanding solar fleet. The price differential between these two zones — the West–North basis spread — provides a continuous, market-based signal of the tension between intermittent renewable supply in the west and growing electricity demand in the north.

This paper makes two related contributions. First, we provide a market-based measurement of AI data center electricity demand using observed load growth in ERCOT's North zone. Employing a counterfactual approach that benchmarks North zone load growth against comparable Texas zones without significant data center concentration, we estimate that data center demand added approximately 3.74 TWh of annual electricity consumption to the North Texas grid by 2025 — representing 65 percent growth in a single year and implying approximately 534 MW of installed data center capacity. This estimate compares favorably to CBRE's (2025) reported net absorption of 470.8 MW for the Dallas–Fort Worth market in 2025, supporting the credibility of our market-based approach.

Second, we use this setting to test whether West Texas renewable generation can effectively offset growing North Texas data center demand. Using hourly ERCOT data from January 2023 through December 2025 — comprising 25,202 observations — we estimate four complementary models with the West–North basis spread as the dependent variable. Our central finding is that, while renewable generation significantly reduces the basis spread, its effectiveness diminishes as North zone demand grows. In particular, we document a renewable saturation effect: solar generation's basis-suppressing impact essentially disappears at high levels of North zone load. We further find that the basis spread shifted structurally upward by $5.43 per megawatt-hour (MWh) following 2024, even after controlling for renewable generation, weather, and demand conditions. Quantile regression reveals that the market operates in two distinct regimes: renewable generation and data center demand drive variation under normal conditions, while cold weather shocks dominate the most extreme basis events. Vector autoregression (VAR) and Granger causality tests confirm that renewable generation and cold weather dynamically predict daily basis movements, while North zone load operates as a structural background pressure rather than a day-to-day volatility driver. A random forest model independently confirms these findings, ranking solar generation as the most important predictor of the basis spread, followed by wind generation, with North zone load ranking fifth among thirteen candidate variables.

Forward scenario analysis, adjusted for the geographic and temporal deliverability constraints of renewable energy, suggests that under a high-growth scenario consistent with observed 2024–2025 demand trends, deliverable renewable supply will fall short of incremental data center demand by approximately 5.40 TWh annually by 2030. Under a base scenario, the margin narrows to near zero. These results suggest that the challenge facing the Texas grid is not simply one of total renewable energy volume, but of reliable, geographically matched, around-the-clock supply — precisely the dimension along which intermittent wind and solar generation face inherent limitations.

The remainder of the paper proceeds as follows. Section 3 reviews the relevant literature on ERCOT market structure, AI electricity demand, and renewable energy price dynamics. Section 4 describes the data and validates our proxy for data center demand. Section 5 presents the empirical methodology. Section 6 reports the main empirical results. Section 7 presents the forward scenario analysis. Section 8 concludes and discusses policy implications.

3. Background and Literature Review

3.1 ERCOT Market Structure and Zonal Pricing

The Electric Reliability Council of Texas operates as an independent system operator managing approximately 90 percent of Texas's electric load in near-complete isolation from neighboring grids (Potomac Economics, 2025). Unlike most major United States electricity markets, ERCOT operates as an energy-only market without a formal capacity market, relying on high real-time prices during shortage conditions to provide investment signals for new generation resources.

The market employs nodal pricing, with settlement point prices aggregated into four primary load zones — North, South, West, and Houston — each with distinct generation and demand characteristics. Triolo and Wolak (2022) document that ERCOT's transition to a nodal market design in December 2010 improved operational efficiency, reducing average energy prices and ancillary service costs. The West–North basis spread — the difference between West and North zone settlement point prices — is therefore a market-based signal of transmission congestion and supply-demand imbalances between the two zones. A persistent West premium, observed in approximately 76 percent of hours in our sample, indicates that West Texas generation faces barriers to reaching North Texas load due to physical transmission constraints, renewable intermittency, or demand conditions. Understanding the determinants of this spread is the central empirical objective of this paper.

3.2 AI Data Centers and Electricity Demand

The rapid proliferation of artificial intelligence has fundamentally altered the trajectory of electricity demand in the United States after two decades of near-flat growth. Lawrence Berkeley National Laboratory (2024) estimates that United States data centers consumed approximately 176 TWh of electricity in 2023 and projects this figure to reach between 325 and 580 TWh by 2028. The Electric Power Research Institute (2024) similarly projects that data centers could account for between 4.6 and 9.1 percent of all United States electricity consumption by 2030. At the global level, the IEA (2025) projects that data center electricity consumption will nearly double from 415 TWh in 2024 to approximately 945 TWh by 2030.

A key distinguishing characteristic of AI-focused data centers is their near-constant load profile. Unlike residential and commercial consumers, data centers processing AI workloads operate continuously at high and stable rates around the clock (Lawrence Berkeley National Laboratory, 2024), requiring reliable dispatchable generation at all hours rather than simply during peak periods. The geographic concentration of data center development amplifies these grid impacts: in 2023, approximately 80 percent of national data center load was concentrated in 15 states, with clusters forming around major population centers, low electricity costs, and favorable regulatory environments (E3, 2024). Despite the scale of data center buildout in markets like Dallas–Fort Worth — which reported 1,067 MW of installed capacity and 699 MW under construction as of late 2025, ranking third na-

tionally (CBRE, 2025) — relatively little academic literature has directly measured the market impact of this demand growth on electricity prices using observed data. This paper contributes to filling that gap.

3.3 Renewable Energy and Price Formation

The relationship between renewable energy and electricity prices is well-established through the merit order effect: because wind and solar generation have near-zero marginal costs, their dispatch displaces higher-cost thermal generation and reduces the market-clearing price. Antweiler and Muesgens (2021) demonstrate that this effect is robust across markets including Germany, Spain, and Australia, though long-run equilibrium effects depend on conventional capacity adjustment. In the ERCOT context, high renewable generation drives West zone prices sharply lower, often to zero or negative, directly widening the West–North basis spread. Antweiler and Muesgens (2021) further note that variable renewables increase price volatility even as they reduce average prices — the diurnal solar pattern creates systematic intraday variation in the basis that motivates our use of hour-of-day fixed effects.

What remains less studied is the interaction between renewable intermittency and demand-side shocks of the scale introduced by AI data center growth. Most existing merit order literature examines supply-side changes in isolation, without accounting for the simultaneous growth of large, inelastic baseload consumers. This paper addresses that gap.

4. Data

4.1 Data Sources and Construction

Our empirical analysis draws on five primary data sources covering the period January 1, 2023 through December 31, 2025, yielding 25,202 hourly observations and 1,096 daily observations.

Hourly settlement point prices (SPPs) and actual integrated load (AIL) for ERCOT's four primary load zones — North, South, West, and Houston — were obtained from the ERCOT public data portal. Settlement point prices represent the nodal-aggregated marginal cost of electricity at the load zone level and are used for settlement of ERCOT's real-time energy market. Load measurements are real-time metered values, not forecasts. Hourly wind and solar generation totals represent system-wide instantaneous generation in megawatts (MW). Daily Henry Hub natural gas spot prices were obtained from the U.S. Energy Information Administration (2025) and merged to hourly observations by forward-filling. Natural gas prices are relevant because gas-fired generation sets the marginal price during a majority of non-renewable hours in ERCOT.

Hourly temperature data for Dallas–Fort Worth (latitude 32.78, longitude −96.80) and Midland, Texas (latitude 31.99, longitude −102.08) were obtained from Open-Meteo (2025), which provides ERA5-based reanalysis data at high geographic resolution aligned to the America/Chicago time-

zone. We average the two series and compute Cooling Degree Days (CDD) as $\max(0,\, T - 65)$ and Heating Degree Days (HDD) as $\max(0,\, 65 - T)$, where T is average hourly temperature in degrees Fahrenheit (base of 65°F $\approx$ 18°C). The Texas Coincident Index (series: TXPHCI), a monthly index of real economic activity, was obtained from the Federal Reserve Bank of St. Louis (2025) FRED database and first-differenced prior to use in regression analysis to achieve stationarity (ADF statistic: -158.94, $p < 0.001$).

We exclude several candidate variables on grounds of multicollinearity. ERCOT total system load is excluded due to near-perfect correlation with North zone load ($r = 0.94$). The NCENT load zone is excluded due to high collinearity with West zone load ($r = 0.93$). The FWEST zone is excluded due to high collinearity with the Texas Coincident Index ($r = 0.93$). Raw temperature is excluded because it is collinear with CDD and HDD by construction. Variance inflation factor (VIF) diagnostics confirm that no included variable exceeds a VIF of 10 in the baseline specification.

4.2 Variable Definitions

The primary dependent variable is the West–North basis spread:

$$\text{Basis}_{W,N,t} = P_{W,t} - P_{N,t} \tag{1}$$

where $P_{W,t}$ and $P_{N,t}$ denote the settlement point prices for the West and North zones at hour t, respectively.

4.3 Summary Statistics

Table 1 presents descriptive statistics for the key variables, and Figure 1 illustrates the distributional and time-series properties of the basis spread. The basis spread averages \$8.83/MWh over the full sample with substantial dispersion (standard deviation of \$16.56/MWh), reflecting the intermittent nature of renewable supply. The figure reveals a clear diurnal pattern driven by solar generation, persistent positive skew, and an upward shift in the spread level beginning in 2024. ERCOT solar generation more than doubled over the sample period, rising from 3,797 MW in 2023 to 8,044 MW in 2025.

4.4 Proxy Validation for Data Center Demand

A central identification assumption in this paper is that load growth in ERCOT's North zone is attributable primarily to data center expansion, while load growth in other zones reflects conventional demand drivers such as population growth, economic activity, and weather responsiveness. We validate this assumption through two complementary approaches: a baseload signature test and a counterfactual demand measurement.

Table 1: Summary Statistics (Hourly Data, January 2023 – December 2025)

Variable	N	Mean	Std. Dev.	Min	Max
West–North Basis ($/MWh)	25,202	8.83	16.56	−68.70	169.34
North Load (MW)	25,202	1,459.76	353.14	739.30	2,732.09
ERCOT Wind (MW)	25,202	12,578.66	6,673.74	37.93	28,264.50
ERCOT Solar (MW)	25,202	5,838.13	7,499.66	0.00	29,503.09
Gas Price ($/MMBtu)	25,202	2.77	1.08	1.21	13.20
Cooling Degree Days	25,202	8.85	10.21	0.00	42.33
Heating Degree Days	25,202	5.75	9.51	0.00	55.68
Texas Coincident Index	25,202	174.16	4.52	166.17	181.18

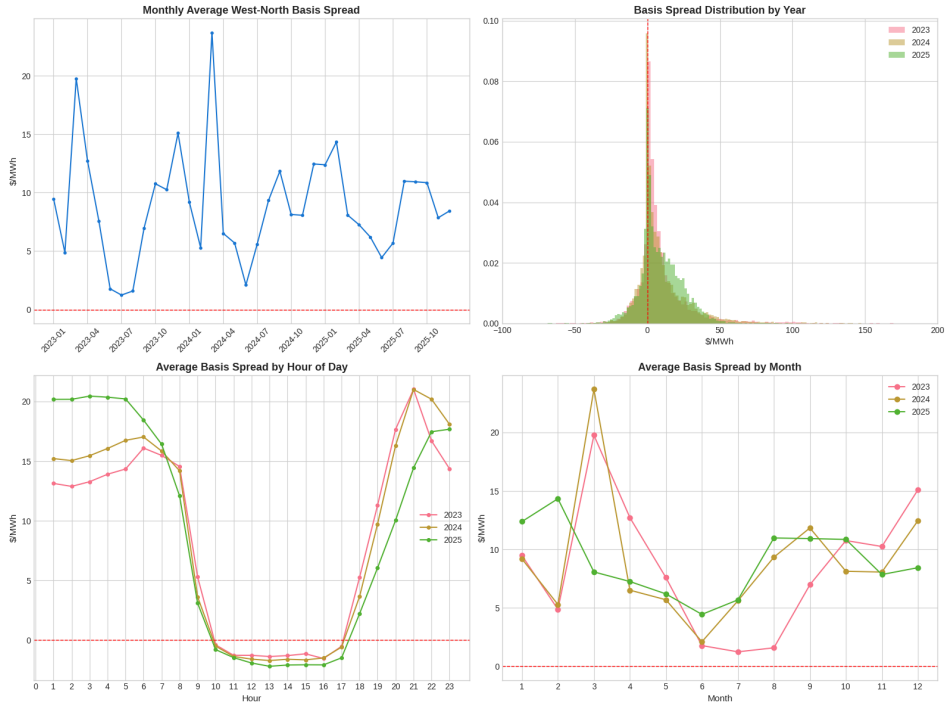


Figure 1: West–North Basis Spread: Monthly Average, Annual Distribution, Intraday Pattern, and Seasonal Pattern (January 2023 – December 2025). Midday solar generation systematically collapses the spread; the upper-right panel shows a rightward distributional shift in 2024–2025, consistent with structural demand growth (Electric Reliability Council of Texas, 2025).

4.4.1 Baseload Signature Test

Data centers processing AI workloads draw power at high, stable rates around the clock, producing a near-flat load profile with minimal diurnal variation. Conventional electricity consumers, by contrast, exhibit demand that peaks during business hours and on days of extreme temperature, with off-peak hours generating substantially lower load. If the North zone is experiencing data center-driven growth, off-peak load growth should exceed on-peak load growth — a pattern we refer to as the baseload signature.

We define peak hours as 4 PM to 9 PM on weekdays and classify all remaining hours as off-peak. We then compute year-over-year growth in average load for each period by zone and calculate the off-peak minus on-peak growth gap. A positive gap indicates that off-peak hours are growing faster than peak hours, consistent with continuous, non-weather-responsive data center load.

Figure 2 presents the results. North zone is the only zone in which off-peak growth consistently exceeds on-peak growth, with a gap of 2.71 percentage points in 2024 and 1.26 percentage points in 2025. All other zones — South, Houston, and West — exhibit negative gaps in 2024, meaning on-peak growth exceeded off-peak growth, consistent with weather-driven, peak-concentrated demand. This pattern supports our interpretation of North zone excess growth as reflecting 24/7 data center operations rather than conventional load growth.

4.4.2 Counterfactual Demand Measurement

To quantify the magnitude of data center demand, we construct a counterfactual baseline for North zone load using the average growth rate of zones without significant data center concentration. We use the average of the West and North-Central (NCENT) zones as our primary baseline. Excess North zone growth above this baseline is attributed to data center demand:

$$\widehat{\text{DC Demand}}_t = \overline{\text{NORTH}}_t - \overline{\text{NORTH}}_{2023} \times \prod_{s=2024}^{t} (1 + g_s^{\text{baseline}}) \tag{2}$$

where $\overline{\text{NORTH}}_t$ is the annual average North zone load in year t, and g_s^{baseline} is the average growth rate of the West and NCENT zones in year s.

Under our primary baseline (West + NCENT average growth of 1.9 percent annually), excess North zone load attributable to data centers is estimated at 259 MW in 2024 and 427 MW in 2025, corresponding to annual electricity consumption of approximately 2.27 TWh and 3.74 TWh respectively, and implying installed data center capacity of approximately 324 MW and 534 MW at an assumed 80 percent capacity factor. The growth in implied capacity from 2024 to 2025 — approximately 210 MW, or 65 percent year-over-year — is consistent with the rapid pace of data center development reported for the Dallas–Fort Worth market during this period. Figure 2 and Figure 3 together illustrate the baseload signature and counterfactual measurement. To assess sensitivity to baseline choice, Table 2 reports results across three scenarios. The estimates are robust to baseline choice, ranging from 3.26 to 3.77 TWh in 2025. Our primary estimate of 534 MW compares favorably to CBRE's (2025) reported net absorption of 471 MW for the Dallas–Fort Worth market — a difference of approximately 13 percent — validating our identification strategy.

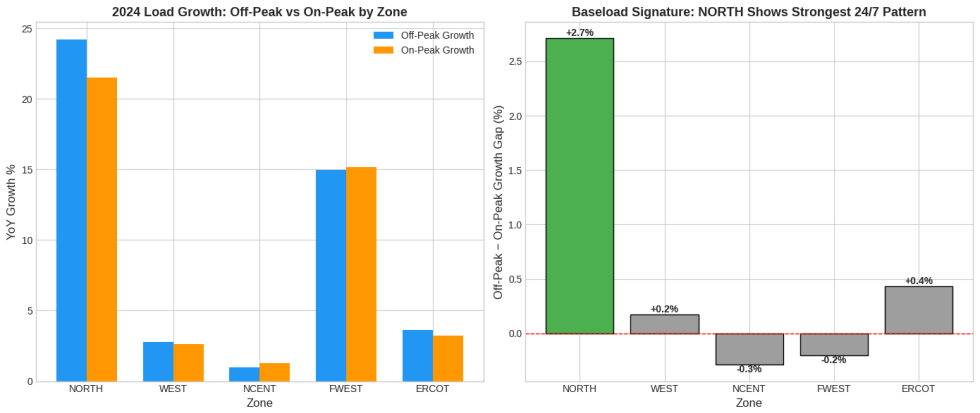


Figure 2: 2024 Load Growth by Zone: Off-Peak vs. On-Peak (left) and Baseload Signature Gap (right). North zone is the only zone with a positive off-peak minus on-peak growth gap, consistent with 24/7 data center load (Electric Reliability Council of Texas, 2025).

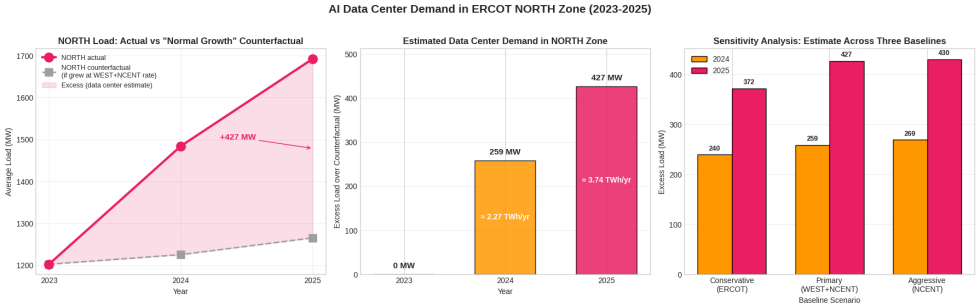


Figure 3: AI Data Center Demand in ERCOT North Zone (2023–2025). Actual vs. counterfactual North zone load (left), estimated excess demand by year (center), and sensitivity analysis across three baseline scenarios (right) (Electric Reliability Council of Texas, 2025).

Table 2: Data Center Demand Estimates by Scenario (2025)

Scenario	Baseline Growth	Excess MW	Annual TWh	Implied Capacity (MW)
Conservative (ERCOT)	3.52%	372	3.26	465
Primary (West + NCENT)	1.90%	427	3.74	534
Aggressive (NCENT)	1.06%	430	3.77	538

5. Methodology

5.1 Model 1: OLS Regression Framework

The baseline empirical specification is:

$$\begin{aligned}\text{Basis}_t = \beta_0 + \beta_1 \text{Wind}_t + \beta_2 \text{Solar}_t + \beta_3 \text{NORTH}_t + \beta_4 \text{WEST}_t \\ + \beta_5 \text{Gas}_t + \beta_6 \text{CDD}_t + \beta_7 \text{HDD}_t + \beta_8 \Delta\text{TxIndex}_t + \boldsymbol{\delta}_t + \varepsilon_t\end{aligned} \tag{3}$$

where $\boldsymbol{\delta}_t$ denotes a vector of time fixed effects for hour of day, month, and day of week. Standard errors are Newey–West HAC with a 24-lag bandwidth to account for heteroskedasticity and autocorrelation in hourly electricity data. Breusch–Pagan tests confirm heteroskedasticity and a Durbin–Watson statistic of 0.25 confirms strong autocorrelation, both justifying the use of HAC standard errors.

We extend Model 1 in two directions. First, to test the saturation hypothesis — that renewable effectiveness diminishes as North zone demand grows — we augment the baseline with four interaction terms. All variables entering interactions are mean-centered prior to multiplication to preserve interpretability of main effect coefficients. The interaction terms are:

$$+\gamma_1(\widetilde{\text{Wind}} \times \widetilde{\text{NORTH}}) + \gamma_2(\widetilde{\text{Solar}} \times \widetilde{\text{NORTH}}) + \gamma_3(\widetilde{\text{Wind}} \times \widetilde{\text{HDD}}) + \gamma_4(\widetilde{\text{Gas}} \times \widetilde{\text{Wind}}) \quad (4)$$

where tildes denote mean-centered variables. The first two interactions test whether wind and solar generation become less effective at suppressing the basis spread as North zone load increases. The third tests whether wind generation matters more during cold weather events. The fourth tests whether gas prices have a larger effect when wind generation is low, consistent with gas-fired plants setting the marginal price during renewable scarcity.

Second, to test whether the market fundamentally changed following the acceleration of data center development in 2024, we add a post-2024 structural break indicator and its interactions:

$$+\delta_0 \text{Post}_{2024} + \delta_1(\widetilde{\text{Wind}} \times \text{Post}_{2024}) + \delta_2(\widetilde{\text{Solar}} \times \text{Post}_{2024}) + \delta_3(\widetilde{\text{NORTH}} \times \text{Post}_{2024}) \quad (5)$$

where Post_{2024} equals one for observations in 2024–2025. The coefficient δ_0 captures any level shift in the basis spread after 2024 not explained by observable covariates, while δ_1, δ_2, and δ_3 test whether the effects of renewables and demand changed after 2024.

5.2 Model 2: Quantile Regression

To examine whether the determinants of basis spread variation differ across the distribution — in particular, whether extreme scarcity events are driven by different factors than normal market conditions — we estimate quantile regressions at five quantiles: $\tau \in \{0.25, 0.50, 0.75, 0.90, 0.95\}$. Following Koenker and Bassett (1978), the quantile regression estimator minimizes the asymmetrically weighted sum of absolute deviations:

$$\hat{\beta}(\tau) = \arg\min_{\beta} \sum_{t=1}^{T} \rho_\tau \left(\text{Basis}_t - \mathbf{x}_t' \beta \right) \quad (6)$$

where $\rho_\tau(u) = u(\tau - \mathbf{1}_{u<0})$ is the check function. We use the same specification as the Model 1 baseline, including all time fixed effects, and report bootstrapped standard errors with 500 replications. By comparing coefficients across quantiles, we identify whether renewable generation and demand effects are uniform across market conditions or exhibit heterogeneous behavior — particularly whether extreme tail events (the 90th and 95th percentiles representing the most severe congestion events) are driven by fundamentally different mechanisms than normal operations. A key hypothesis is that cold weather shocks should dominate the extreme upper tail, reflecting the well-documented pattern of Texas grid stress during winter weather events, while data center demand — being largely invariant to weather — should matter more in the median and lower quantiles.

5.3 Model 3: Vector Autoregression and Granger Causality

To examine the dynamic relationships between the basis spread, renewable generation, demand, and weather — and specifically to test whether North zone load has predictive power for daily basis movements — we estimate a VAR model on daily aggregated data (1,096 observations). Lag length is selected using the Akaike Information Criterion (AIC), which identifies seven lags as optimal, capturing one week of daily dynamics. We then conduct pairwise Granger causality tests for each variable with respect to the basis spread. Under the null hypothesis that variable X does not Granger-cause the basis spread, the coefficients on all seven lags of X in the basis equation are jointly zero, tested using an F-test (Sims, 1980).

5.4 Model 4: Random Forest

As a nonparametric complement to our regression analysis, we estimate a random forest model (Breiman, 2001) to assess the relative predictive importance of each variable in explaining basis spread variation. The random forest algorithm constructs an ensemble of decision trees, each trained on a bootstrap sample of the data with a random subset of features considered at each split. By averaging predictions across trees, random forests reduce variance and produce stable importance rankings that are robust to overfitting.

The model is trained on 80 percent of the hourly sample (20,161 observations) and evaluated on the remaining 20 percent (5,041 observations) using 100 decision trees. Variable importance is measured by mean decrease in impurity (Gini importance), which captures the total reduction in node impurity attributable to each feature across all trees, normalized to sum to one. Unlike OLS coefficients, random forest importance rankings are robust to nonlinear relationships, interactions among variables, and non-Gaussian error distributions — all of which are likely present in high-frequency electricity price data. This model is used for predictive importance ranking only and does not support causal interpretation; its role is to provide an independent, model-free confirmation of the variable importance hierarchy established by our regression models.

5.5 Placebo Tests and Magnitude Validation

To validate that our NORTH coefficient reflects data center demand specifically rather than generic demand growth, we conduct two placebo tests designed to falsify our identification strategy.

Placebo A: Alternative Dependent Variable. We replace the West–North basis spread with the West–Houston basis spread as the dependent variable, re-estimating the baseline Model 1 specification. Houston is not a primary data center hub — it is predominantly a petrochemical and industrial demand center — so the North zone coefficient should reflect generic system-wide demand effects rather than data center-specific effects. If our baseline NORTH coefficient captures DFW-specific data center demand, the coefficient in the Houston specification should differ in magnitude or direction.

Placebo B: Alternative Demand Variable. We replace North zone load with NCENT load (Austin–San Antonio corridor) as our demand proxy, re-estimating the baseline specification. NCENT has population and economic activity broadly comparable to the Dallas–Fort Worth area but substantially lower data center concentration, lacking the infrastructure, connectivity, and incentive profile that has made DFW the third-largest data center market nationally (CBRE, 2025). If our baseline NORTH effect reflects data center demand specifically, the NCENT coefficient should be significantly weaker and potentially negative, as NCENT load growth is driven by traditional residential and commercial demand that responds differently to price differentials.

In addition to the placebo tests, we validate the magnitude of our counterfactual demand estimate by comparing it against independently reported industry figures from CBRE (2025), providing an external cross-check on the credibility of our market-based measurement approach.

6. Results

6.1 Measuring AI Electricity Demand

North zone load grew 23.45 percent in 2024 and 14.01 percent in 2025 — more than five times the average growth rate of other ERCOT zones, which ranged from 1 to 4 percent. This divergence is consistent with the rapid expansion of data center capacity in the Dallas–Fort Worth metropolitan area documented in Section 4.4. Under our primary counterfactual approach, data center demand added 427 MW of average load and 3.74 TWh of annual electricity consumption to the North Texas grid by 2025, implying approximately 534 MW of installed data center capacity at an 80 percent capacity factor.

6.2 Model 1: Baseline, Interaction, and Structural Break Results

Table 3 presents results from all three Model 1 specifications. The baseline OLS ($R^2 = 0.4765$) shows that each additional 1,000 MW of wind reduces the basis by \$1.13/MWh ($t = -26.44$, $p < 0.001$), consistent with the merit order effect documented by Antweiler and Muesgens (2021). So-

lar generation similarly suppresses the basis by \$0.45/MWh per 1,000 MW ($t = -11.31$, $p < 0.001$). North zone load increases the basis by \$3.26/MWh per 1,000 MW ($t = 3.23$, $p < 0.001$), capturing the market-clearing effect of data center demand on zonal price differentials. The peak hours indicator is negative and highly significant (−\$9.12/MWh, $t = -16.22$), reflecting the systematic collapse of the basis during midday solar generation peaks. Gas price, CDD, HDD, and the Texas Coincident Index are not statistically significant in the baseline, suggesting their effects operate primarily through the renewable and demand channels already captured. These results are robust to day-level clustered standard errors (1,096 clusters) and to arcsinh transformation of the dependent variable ($R^2 = 0.6082$).

The interaction specification ($R^2 = 0.4948$, improvement of 0.0183) confirms the saturation hypothesis. The Wind × North interaction ($t = 4.41$, $p < 0.001$) and Solar × North interaction ($t = 7.11$, $p < 0.001$) are both positive and significant. Figure 4 illustrates these marginal effects. At mean North load (approximately 1,460 MW), each additional 1,000 MW of wind reduces the basis by \$1.34/MWh; at high North load (approximately 2,400 MW), the same wind addition reduces the basis by only \$0.66/MWh — a 51 percent reduction in effectiveness. The solar saturation effect is even more pronounced: at high North zone loads, the solar effect approaches zero, indicating that solar generation becomes essentially ineffective at suppressing congestion precisely when data center demand is elevated. The Wind × HDD ($t = -3.80$) and Gas × Wind ($t = 5.52$) interactions are also significant at the 1 percent level, consistent with gas-fired plants setting the marginal price during periods of renewable scarcity and cold weather.

The structural break specification ($R^2 = 0.5023$) shows that the basis spread was \$5.43/MWh higher in 2024–2025 than in 2023 after controlling for all observable covariates ($t = 6.96$, $p < 0.001$). This shift is not explained by any observable covariate in the model and represents a structural change in the basis spread level coinciding with the acceleration of data center development in the Dallas–Fort Worth market. The North × Post-2024 interaction is positive and significant ($t = 2.30$, $p < 0.05$), indicating that data center demand had a larger effect on the basis spread after 2024. The wind and solar interactions with the post-2024 indicator are not statistically significant, consistent with a demand-side rather than supply-side source of the structural change.

6.3 Model 2: Quantile Regression Results

Table 4 presents selected coefficients from Model 2, revealing two distinct market regimes separated by the upper tail of the distribution.

Under normal market conditions ($\tau = 0.25$ through 0.75), wind and solar generation consistently suppress the basis spread, and North zone load consistently increases it. The North zone coefficient is positive and significant across the lower three quantiles ($p < 0.001$), confirming that data center demand contributes to congestion under typical operating conditions.

At the extreme upper tail ($\tau = 0.90$ and 0.95), the pattern fundamentally changes. The North zone coefficient turns negative and becomes statistically significant at the 95th percentile (−0.0033,

Table 3: Model 1: Baseline, Interaction, and Structural Break Results (Dependent Variable: West–North Basis Spread, $/MWh)

	Baseline		Interactions		Structural Break	
Variable	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.
Wind (per 1k MW)	−1.127	***	−1.134	***	−1.121	***
Solar (per 1k MW)	−0.445	***	−0.452	***	−0.441	***
North (per 1k MW)	3.257	***	3.301	***	3.189	***
Peak	−9.122	***	−8.934	***	−9.089	***
Wind × North			+0.000	***		
Solar × North			+0.000	***		
Wind × HDD			−0.000	***		
Gas × Wind			+0.000	***		
Post-2024					5.433	***
North × Post					0.004	**
R^2	0.4765		0.4948		0.5023	

$N = 25{,}202$. HAC SE: Newey–West 24 lags. Time FE: hour, month, day-of-week.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

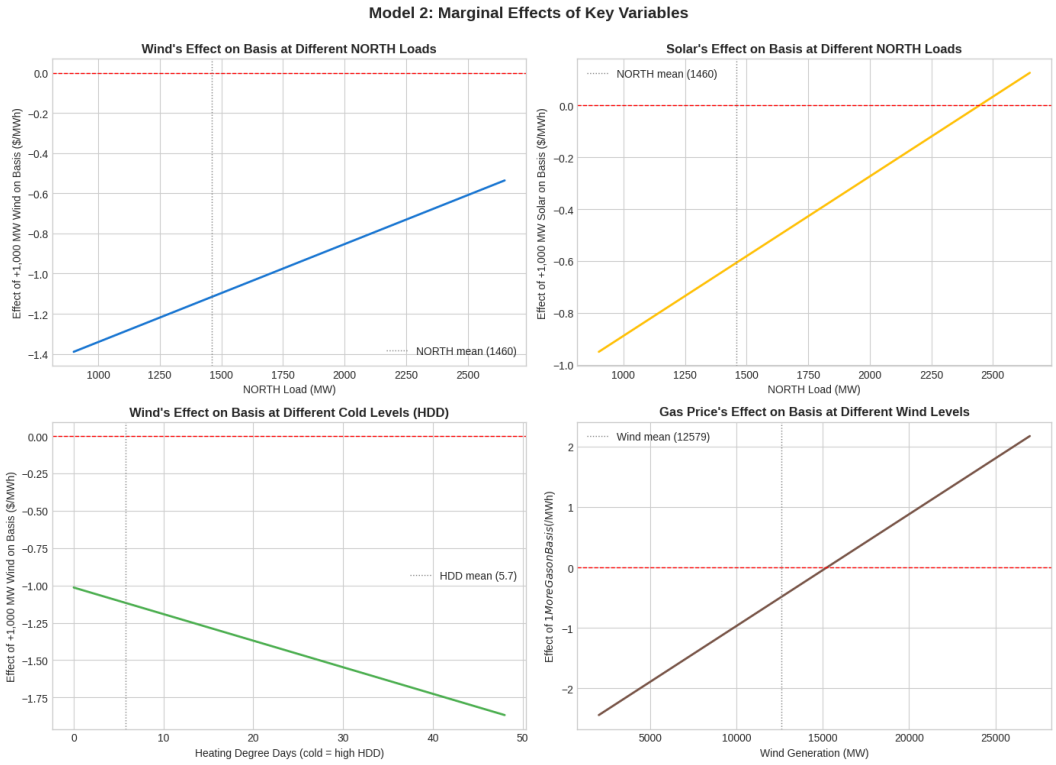


Figure 4: Model 1 (Extended): Marginal Effects of Key Variables at Different Conditioning Values. Top panels show wind and solar effectiveness declining sharply as North zone load increases (saturation effect). Bottom-left shows wind more effective during cold weather. Bottom-right shows gas prices matter more when wind is scarce (Electric Reliability Council of Texas, 2025).

$p < 0.001$), indicating that data center demand no longer drives the most extreme basis events. Instead, the HDD coefficient increases sharply with quantile: at the 95th percentile, HDD has a coefficient of 0.2704 compared to -0.0028 at the 25th percentile — a 96-fold increase in magnitude. This pattern indicates that the most extreme basis spike events are driven by cold weather shocks in combination with low renewable output, not by data center demand growth. The implication is that while data centers contribute to persistent structural pressure on the basis spread, the worst scarcity events are weather-driven and would occur regardless of data center load.

Table 4: Model 2: Quantile Regression Results (Selected Coefficients)

Variable	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	$\tau = 0.95$
Wind	-0.0008^{***}	-0.0008^{***}	-0.0009^{***}	-0.0009^{***}	-0.0010^{***}
Solar	-0.0003^{***}	-0.0004^{***}	-0.0004^{***}	-0.0003^{***}	-0.0002^{***}
North	$+0.0045^{***}$	$+0.0052^{***}$	$+0.0043^{***}$	-0.0003	-0.0033^{***}
HDD	-0.0028	-0.0067	$+0.0320^{**}$	$+0.1861^{***}$	$+0.2704^{***}$

Bootstrapped standard errors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

6.4 Model 3: VAR and Granger Causality Results

Table 5 presents Granger causality test results from Model 3. Wind generation ($F = 2.882$, $p = 0.005$), solar generation ($F = 7.625$, $p < 0.001$), and cold weather (HDD: $F = 3.202$, $p = 0.002$) all Granger-cause the basis spread at conventional significance levels, confirming that past values of these variables contain predictive information about future daily basis movements. By contrast, North zone load does not Granger-cause the basis spread ($F = 0.598$, $p = 0.758$), and neither does natural gas price ($p = 0.772$).

Inspection of the VAR coefficients in the basis equation reinforces these findings. The first lag of the basis spread is large and significant (coefficient = 0.600, $t = 14.85$), indicating strong day-to-day persistence. Solar generation has significant coefficients at lags one, two, and three, suggesting that solar output shocks propagate through the basis spread over multiple days. North zone load has no significant coefficient at any of the seven lags. These results establish that data center demand operates as a structural background pressure — shifting the level of the basis upward as documented in Models 1 — without creating predictable day-to-day fluctuations. Daily volatility is instead driven by renewable intermittency and weather shocks, precisely the variables that cannot be controlled or scheduled by grid operators.

6.5 Model 4: Random Forest Feature Importance

Table 6 presents Model 4 results. The random forest achieves an out-of-sample R^2 of 0.5358 (MAE = \$5.82/MWh, RMSE = \$8.04/MWh). Solar generation ranks first (36.6%), wind second (23.0%), and the sine term of month (seasonality) third (13.8%); together with its cosine term (3.5%),

Table 5: Model 3: Granger Causality Tests (Null: X Does Not Granger-Cause Basis)

Variable	F-statistic	p-value	Reject H_0?
Wind Generation	2.882	0.005	Yes***
Solar Generation	7.625	<0.001	Yes***
HDD	3.202	0.002	Yes***
North Load	0.598	0.758	No
Gas Price	0.581	0.772	No

VAR lag order = 7 (AIC). Daily data, $N = 1{,}096$.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

seasonality accounts for 17.4% of importance. Gas price ranks fourth at 8.2%. North zone load ranks fifth at 4.6%, independently confirming that renewable supply and seasonality are the primary drivers of basis spread variation, while data center demand operates as a secondary, structural factor. Figure 5 illustrates the full importance rankings. The random forest rankings are robust to nonlinear relationships and interactions that OLS may not fully capture, providing independent confirmation of our central finding.

Table 6: Model 4: Random Forest Feature Importance Rankings

Rank	Variable	Importance	Share (%)
1	ERCOT Solar	0.3658	36.6
2	ERCOT Wind	0.2298	23.0
3	Month (sine)	0.1383	13.8
4	Gas Price	0.0817	8.2
5	North Load	0.0458	4.6
6	Month (cosine)	0.0354	3.5
7	HDD	0.0249	2.5
8	CDD	0.0236	2.4

Out-of-sample $R^2 = 0.5358$. MAE = \$5.82/MWh. 80/20 train–test split.
Top 8 of 13 features. Month enters as sine and cosine terms.

6.6 Magnitude Validation and Placebo Tests

Our primary 534 MW estimate aligns closely with CBRE's (2025) reported net absorption of 471 MW for Dallas–Fort Worth — a difference of 13 percent — and falls well within the range of under-construction pipeline capacity (700 MW) and the total existing DFW market (approximately 1,000 MW), supporting the credibility of our identification approach.

In Placebo A, replacing the dependent variable with the West–Houston basis spread yields a North zone coefficient of 0.00490, larger than the baseline estimate of 0.00326 ($R^2 = 0.5111$). This result is inconclusive: it does not show that the North effect is specific to the West–North corridor, and may instead reflect system-wide pricing effects of North zone demand. Identification therefore

rests primarily on Placebo B. In Placebo B, replacing North zone load with NCENT load yields a coefficient of -0.000787 — 124 percent weaker in magnitude and negative in sign ($R^2 = 0.4776$) — confirming that the positive, significant North effect is specific to the Dallas–Fort Worth data center hub and does not reflect generic Texas demand growth.

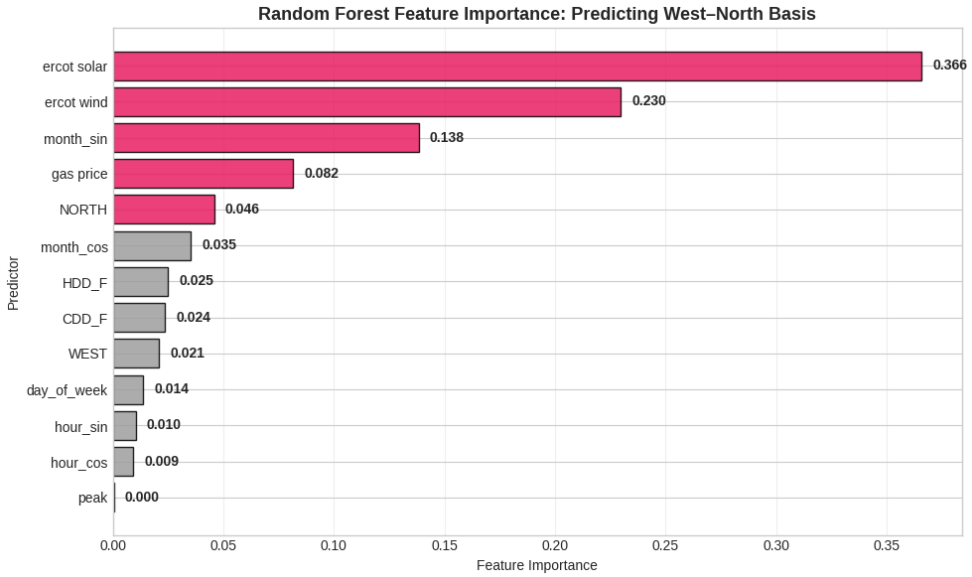


Figure 5: Model 4: Random Forest Feature Importance Rankings for Predicting the West–North Basis Spread. Solar (36.6%) and wind (23.0%) dominate. North zone load ranks fifth at 4.6%, consistent with its role as a structural background variable rather than a short-term volatility driver (Electric Reliability Council of Texas, 2025).

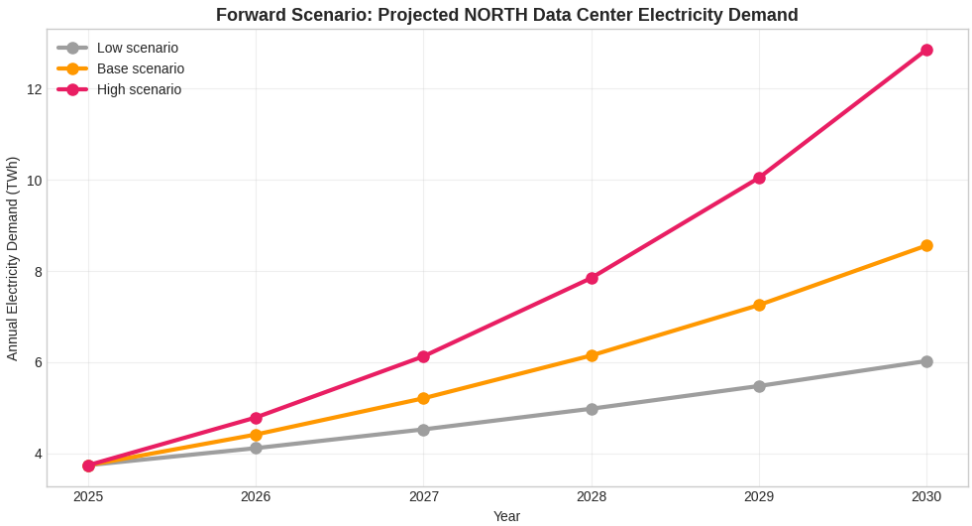


Figure 6: Forward Scenario: Projected North Zone Data Center Electricity Demand (2025–2030) Under Low, Base, and High Growth Assumptions. Data center demand grows at 10, 18, and 28 percent annually in the low, base, and high scenarios (Electric Reliability Council of Texas, 2025; IEA, 2025).

7. Forward Scenario Analysis

Building on our empirical estimates, we project the adequacy of West Texas renewable generation to offset incremental data center electricity demand in the North zone through 2030 under three growth scenarios. We parameterize data center demand growth at 10 percent annually (low), 18 percent (base), and 28 percent (high), starting from our 2025 estimate of 3.74 TWh. Each scenario pairs demand growth with a renewable growth rate that moves in the opposite direction — 12, 10, and 7 percent annually — reflecting the risk that interconnection queues and transmission constraints slow renewable build-out precisely when large-load demand is growing fastest. Renewable growth is applied to 2025 ERCOT wind and solar output (184.1 TWh). To account for the geographic and temporal deliverability constraints of intermittent renewables, we apply a local renewable share of 20 percent and a deliverability factor of 25 percent — the fraction of incremental local renewable output that coincides with data center demand (IEA, 2025). The net gap is incremental data center demand minus the deliverable portion of incremental local renewable output.

Figure 6 presents projected data center electricity demand under all three scenarios and Table 7 summarizes the 2030 adequacy gap.

Table 7: Forward Scenario Analysis: 2030 Renewable Adequacy Summary

Scenario	2030 DC Demand (TWh)	Capacity (MW)	Renewable Offset (TWh)	Net Gap (TWh)
Low (10% / 12%)	6.02	860	7.02	−4.73
Base (18% / 10%)	8.56	1,222	5.62	−0.80
High (28% / 7%)	12.85	1,835	3.71	+5.40

Growth rates shown as data center demand / renewables, per year.
Net gap = incremental DC demand − deliverable renewable offset.

Under the low scenario, deliverable renewable growth comfortably exceeds incremental data center demand through 2030, with a surplus of 4.73 TWh. Under the base scenario, the margin narrows to near zero (−0.80 TWh), leaving essentially no buffer for forecast error or demand acceleration. Under the high scenario, a 5.40 TWh annual deficit emerges by 2030, equivalent to approximately 620 MW of unmet continuous power demand.

These projections highlight a critical distinction between total renewable capacity and deliverable renewable supply. Even as Texas continues to lead the nation in renewable energy deployment, the geographic separation between West Texas generation and North Texas demand — combined with the intermittent nature of wind and solar — constrains how much of that capacity can reliably serve 24/7 data center loads. The deliverability constraints embedded in our assumptions are conservative; actual deliverable fractions may be lower if transmission constraints worsen or data center load factors increase above historical levels.

8. Conclusion

This paper uses three years of hourly ERCOT market data and four complementary econometric models to ask whether Texas's renewable energy infrastructure can keep pace with AI data center demand. The answer is nuanced. Using a counterfactual approach, we estimate that data center demand added 3.74 TWh of annual electricity consumption to the North Texas grid by 2025, implying approximately 534 MW of installed capacity — consistent with CBRE's (2025) independently reported net absorption of 471 MW. The AI demand surge is not a future projection; it has already materialized in observed market data.

Our empirical results reveal a consistent pattern across all models. Renewable generation significantly suppresses the basis spread, but its effectiveness diminishes sharply as North zone load grows — the solar saturation effect approaches zero at high demand levels. The basis shifted structurally upward by $5.43/MWh after 2024, a regime change consistent with rising demand and not explained by supply-side factors. Extreme congestion events are weather-driven, not demand-driven, while day-to-day volatility is governed by renewable intermittency. Under the base growth scenario, deliverable renewable supply and incremental data center demand converge to near-zero margin by 2030; under high growth, a 5.40 TWh annual deficit emerges.

The core finding is that the challenge is not total renewable volume but reliable, geographically matched, 24/7 supply — precisely the dimension along which intermittent wind and solar face inherent limits. Addressing this requires more than additional renewable capacity: grid planners need dispatchable resources such as battery storage and natural gas backup, targeted transmission investment between West and North Texas, and incorporation of data center baseload profiles into long-term resource adequacy planning. Data center operators can contribute through demand flexibility programs that shift workloads to periods of high renewable output. Future research could extend this market-based measurement approach to other major data center clusters, or incorporate facility-level interconnection data to distinguish individual operators and workload types.

9. References

Antweiler, W., & Muesgens, F. (2021). On the long-term merit order effect of renewable energies. *Energy Economics, 99*, Article 105275. https://doi.org/10.1016/j.eneco.2021.105275

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

CBRE. (2025). *North America data center trends H2 2025*. https://www.cbre.com/insights/books/north-america-data-center-trends-h2-2025

E3 (Energy and Environmental Economics). (2024). *Load growth is here to stay, but are data centers?* https://www.ethree.com/wp-content/uploads/2024/07/E3-White-Paper-2024-Load-Growth-Is-Here-to-Stay-but-Are-Data-Centers-2.pdf

Electric Power Research Institute. (2024). *Powering intelligence: Analyzing artificial intelligence and data center energy consumption*. https://www.epri.com/research/products/3002028905

Electric Reliability Council of Texas. (2025). *Long-term load forecast 2025–2031*. https://www.ercot.com

Federal Reserve Bank of St. Louis. (2025). *Texas coincident economic activity index* [Data set]. FRED. https://fred.stlouisfed.org/series/TXPHCI

Goldman Sachs. (2024). *AI is poised to drive 160% increase in data center power demand*. https://www.goldmansachs.com/insights/articles/AI-poised-to-drive-160-increase-in-power-demand

International Energy Agency. (2025). *Energy and AI*. https://www.iea.org/reports/energy-and-ai

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica, 46*(1), 33–50. https://doi.org/10.2307/1913643

Lawrence Berkeley National Laboratory. (2024). *2024 United States data center energy usage report*. U.S. Department of Energy. https://eta-publications.lbl.gov/sites/default/files/2024-12/lbnl-2024-united-states-data-center-energy-usage-report_1.pdf

Open-Meteo. (2025). *Historical weather API*. https://open-meteo.com

Potomac Economics. (2025). *2024 state of the market report for the ERCOT electricity markets*. https://www.potomaceconomics.com/markets-monitored/ercot/

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica, 48*(1), 1–48. https://doi.org/10

.2307/1912017

Triolo, R., & Wolak, F. A. (2022). Quantifying the benefits of a nodal market design in the Texas electricity market. *Energy Economics, 112*, Article 106154. https://doi.org/10.1016/j.eneco.2022.106154

U.S. Energy Information Administration. (2025). *We expect rapid electricity demand growth in Texas and the mid-Atlantic*. https://www.eia.gov/todayinenergy/detail.php?id=65844