

Quantifying the Association Between Data Center Presence and Retail Electricity Prices

Regression & Hypothesis Testing Report (Automated Output)

Artifact folder: C:\Users\amand\OneDrive\Documents\School\Assignments\WGU\Data-Engineering\Capstone\datacenter-electricity-prices-us\notebooks\outputs\regression\run_20260110_141915

This PDF is generated automatically from the notebook run artifacts (tables, figures, and model output). It is intended to be a polished, reproducible input to the final written capstone report.

Executive Summary

Regression & Hypothesis Testing (PanelOLS)

Model:

Two-way fixed effects panel regression with state fixed effects and month fixed effects.

Outcome: residential price (cents/kWh).

Exposure: data center load intensity (MW per million kWh of residential sales).

Controls: $\log(1+\text{generation})$, fuel mix shares (renew omitted baseline), fuel cost proxy.

Estimation:

- PanelOLS with EntityEffects + TimeEffects
- SEs: two-way clustered by state and month
- N=1671 across 51 states and 34 months

Result:

$\beta = 0.867527$ (SE=0.540843)

two-sided $p = 0.108908$

one-sided $p = 0.054454 \rightarrow$ fail to reject H_0 at $\alpha = 0.05$

Effect size (P25 \rightarrow P75 exposure):

$\Delta x = 0.119246$ implies $\Delta \text{price} = 0.1034$ cents/kWh

(95% CI $\approx [-0.0231, 0.2300]$)

Artifacts written to:

C:\Users\amand\OneDrive\Documents\School\Assignments\WGU\Data-Engineering\Capstone\datacenter-electricity-prices-us\notebooks\outputs\regression\run_20260110_141915

Table 1. Model-Ready Descriptive Summary

variable	group	n	pct_missing	mean	std	min	p25	median	p75	max
res_price_cent	Overall	1,734.00	0.0000	17.245	6.4421	9.2700	13.152	14.910	18.398	45.390
dc_mw_per_res	Overall	1,734.00	0.0000	0.1679	0.4593	0.0000	0.0000	0.0090	0.1145	4.1014
ln_generation	Overall	1,734.00	0.0000	13.392	3.7263	0.0000	13.085	14.895	15.562	17.079
share_coal	Overall	1,671.00	3.6332	0.2206	0.2584	0.0000	0.0000	0.1280	0.3589	0.9815
share_gas	Overall	1,671.00	3.6332	0.3571	0.2686	0.0000	0.1315	0.3279	0.5271	0.9991
share_nuclear	Overall	1,671.00	3.6332	0.1107	0.1676	0.0000	0.0000	0.0000	0.2264	0.6386
fuel_cost_proxy	Overall	1,734.00	0.0000	-0.0108	0.5125	-1.1645	-0.1967	-0.0032	0.1111	11.390
res_price_cent	No DC in state	714.000	0.0000	18.857	7.9642	9.7500	13.050	15.500	23.918	45.390
dc_mw_per_res	No DC in state	714.000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ln_generation	No DC in state	714.000	0.0000	11.494	4.7943	0.0000	9.1330	13.551	14.840	17.079
share_coal	No DC in state	651.000	8.8235	0.2373	0.3010	0.0000	0.0000	0.0707	0.4296	0.9815
share_gas	No DC in state	651.000	8.8235	0.2661	0.2734	0.0000	0.0052	0.1567	0.4725	0.9658
share_nuclear	No DC in state	651.000	8.8235	0.0519	0.1110	0.0000	0.0000	0.0000	0.0000	0.4670
fuel_cost_proxy	No DC in state	714.000	0.0000	-0.0363	0.4754	-1.1645	-0.2147	0.0000	0.0064	4.4206
res_price_cent	Has DC in state	1,020.00	0.0000	16.116	4.8065	9.2700	13.190	14.740	17.145	34.550
dc_mw_per_res	Has DC in state	1,020.00	0.0000	0.2854	0.5703	0.0008	0.0230	0.0957	0.3059	4.1014
ln_generation	Has DC in state	1,020.00	0.0000	14.722	1.7991	7.7463	14.707	15.379	15.749	16.349
share_coal	Has DC in state	1,020.00	0.0000	0.2099	0.2266	0.0000	0.0000	0.1471	0.3304	0.8613
share_gas	Has DC in state	1,020.00	0.0000	0.4152	0.2489	0.0000	0.2317	0.3764	0.5733	0.9991
share_nuclear	Has DC in state	1,020.00	0.0000	0.1483	0.1859	0.0000	0.0000	0.0000	0.3034	0.6386
fuel_cost_proxy	Has DC in state	1,020.00	0.0000	0.0071	0.5365	-0.9154	-0.1938	-0.0201	0.1401	11.390

Source: table1_model_ready.csv

Panel Regression Results (Coefficients)

param	coef	std_err	t_stat	p_value	ci_95_low	ci_95_high
Intercept	12.453	3.2755	3.8019	0.0001	6.0285	18.878
dc_mw_per_res_mkwh_w	0.8675	0.5408	1.6040	0.1089	-0.1933	1.9284
ln_generation_mwh_total	0.3512	0.2587	1.3576	0.1748	-0.1562	0.8585
share_coal	-0.4226	1.4467	-0.2921	0.7702	-3.2603	2.4151
share_gas	-1.0273	0.9335	-1.1004	0.2713	-2.8584	0.8038
share_nuclear	-0.9960	1.1134	-0.8945	0.3712	-3.1799	1.1879
fuel_cost_proxy	0.1496	0.1586	0.9434	0.3456	-0.1615	0.4607

Source: *panelols_coefficients_clustered.csv*

Model Summary (PanelOLS)

PanelOLS Estimation Summary			
=====			
Dep. Variable:	res_price_cents_kwh	R-squared:	0.0209
Estimator:	PanelOLS	R-squared (Between):	-0.1883
No. Observations:	1671	R-squared (Within):	0.0087
Date:	Sat, Jan 10 2026	R-squared (Overall):	-0.1356
Time:	14:19:21	Log-likelihood	-2451.9
Cov. Estimator:	Clustered		
		F-statistic:	5.6315
Entities:	51	P-value	0.0000
Avg Obs:	32.765	Distribution:	F(6,1581)
Min Obs:	6.0000		
Max Obs:	34.000	F-statistic (robust):	1.1335
		P-value	0.3402
Time periods:	34	Distribution:	F(6,1581)
Avg Obs:	49.147		
Min Obs:	47.000		
Max Obs:	51.000		

Parameter Estimates						
=====						
	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI

Intercept	12.453	3.2755	3.8019	0.0001	6.0285	18.878
dc_mw_per_res_mkwh_w	0.8675	0.5408	1.6040	0.1089	-0.1933	1.9284
ln_generation_mwh_total	0.3512	0.2587	1.3576	0.1748	-0.1562	0.8585
share_coal	-0.4226	1.4467	-0.2921	0.7702	-3.2603	2.4151
share_gas	-1.0273	0.9335	-1.1004	0.2713	-2.8584	0.8038
share_nuclear	-0.9960	1.1134	-0.8945	0.3712	-3.1799	1.1879
fuel_cost_proxy	0.1496	0.1586	0.9434	0.3456	-0.1615	0.4607
=====						

F-test for Poolability: 462.78
P-value: 0.0000
Distribution: F(83,1581)

Included effects: Entity, Time

Diagnostics Figures

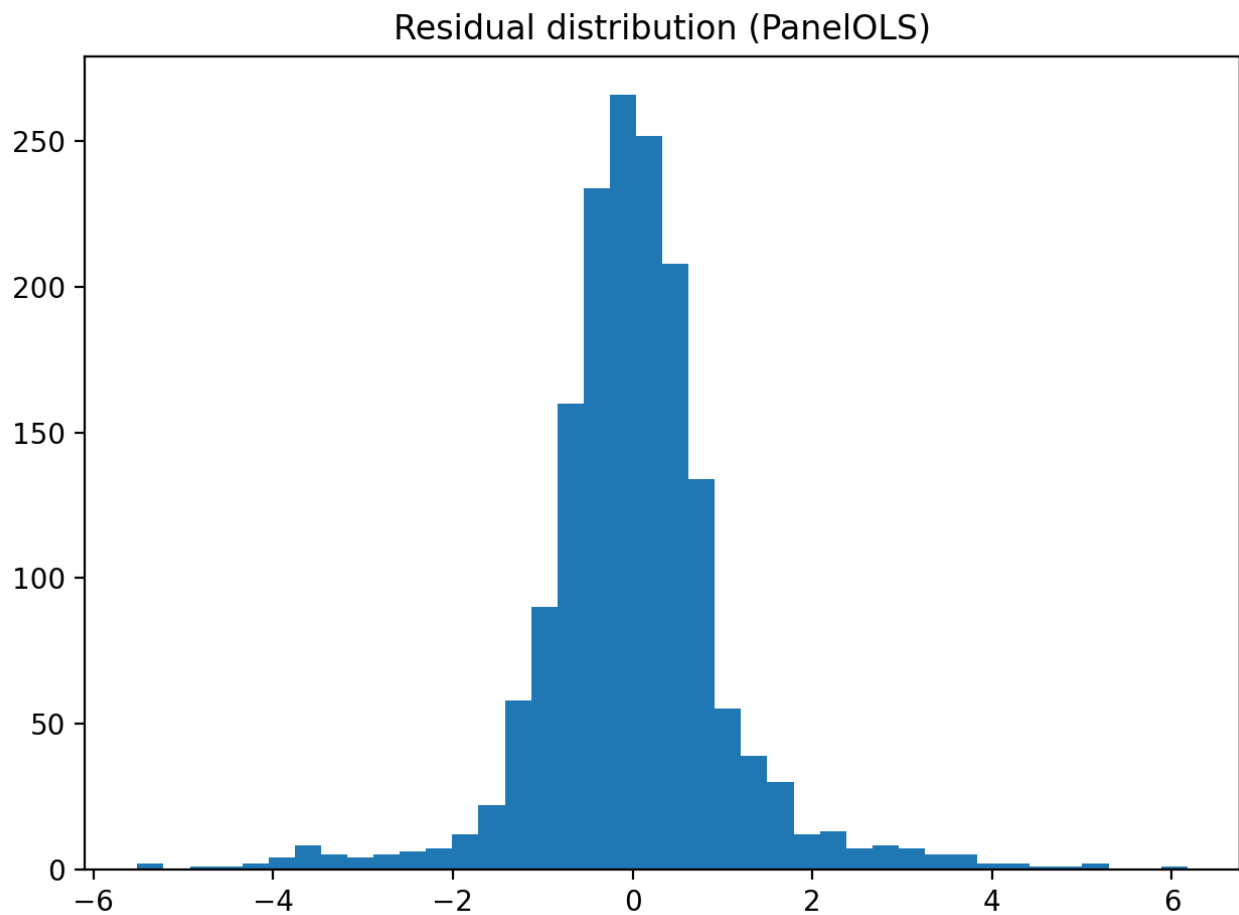


Figure 1. Histogram of regression residuals.

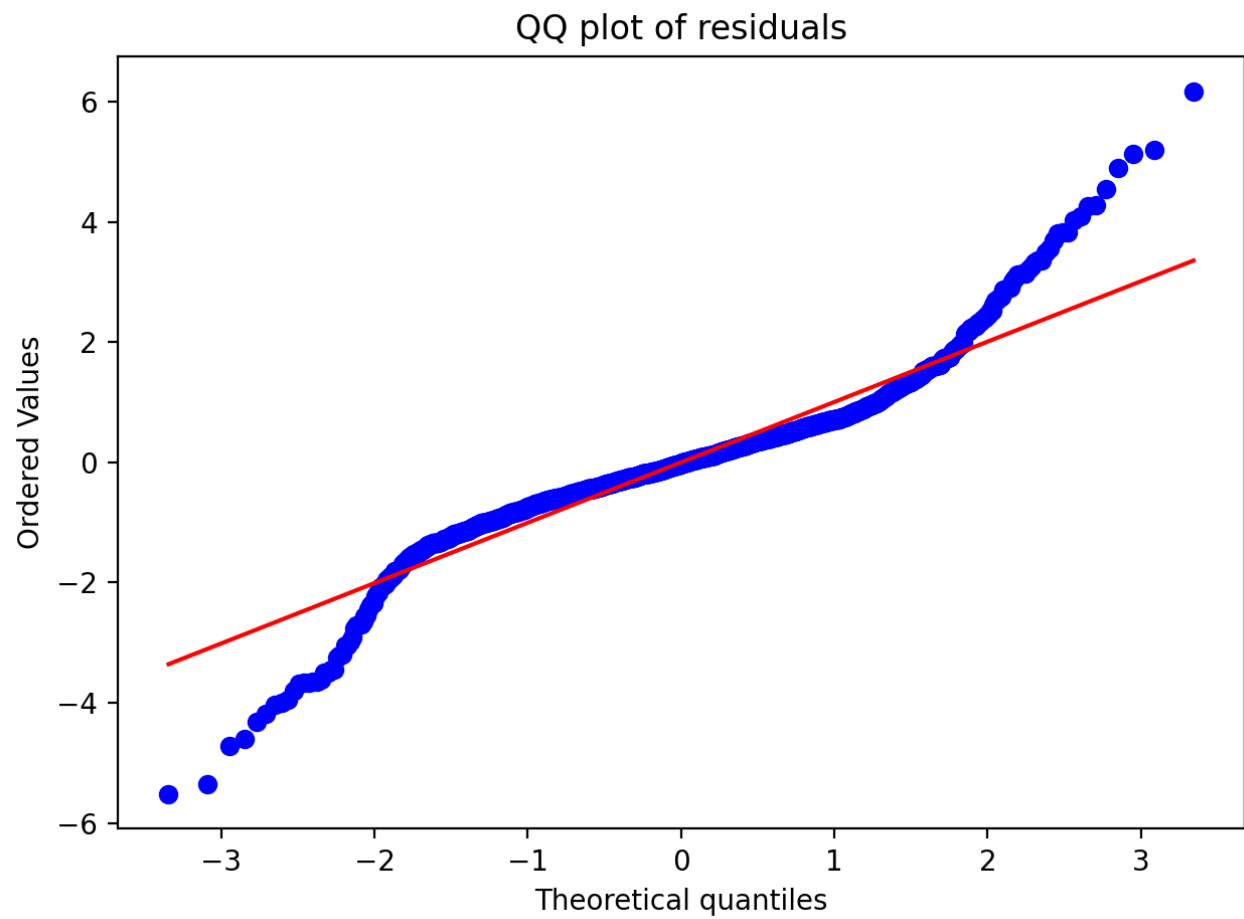


Figure 2. Q-Q plot of residuals (normality check).

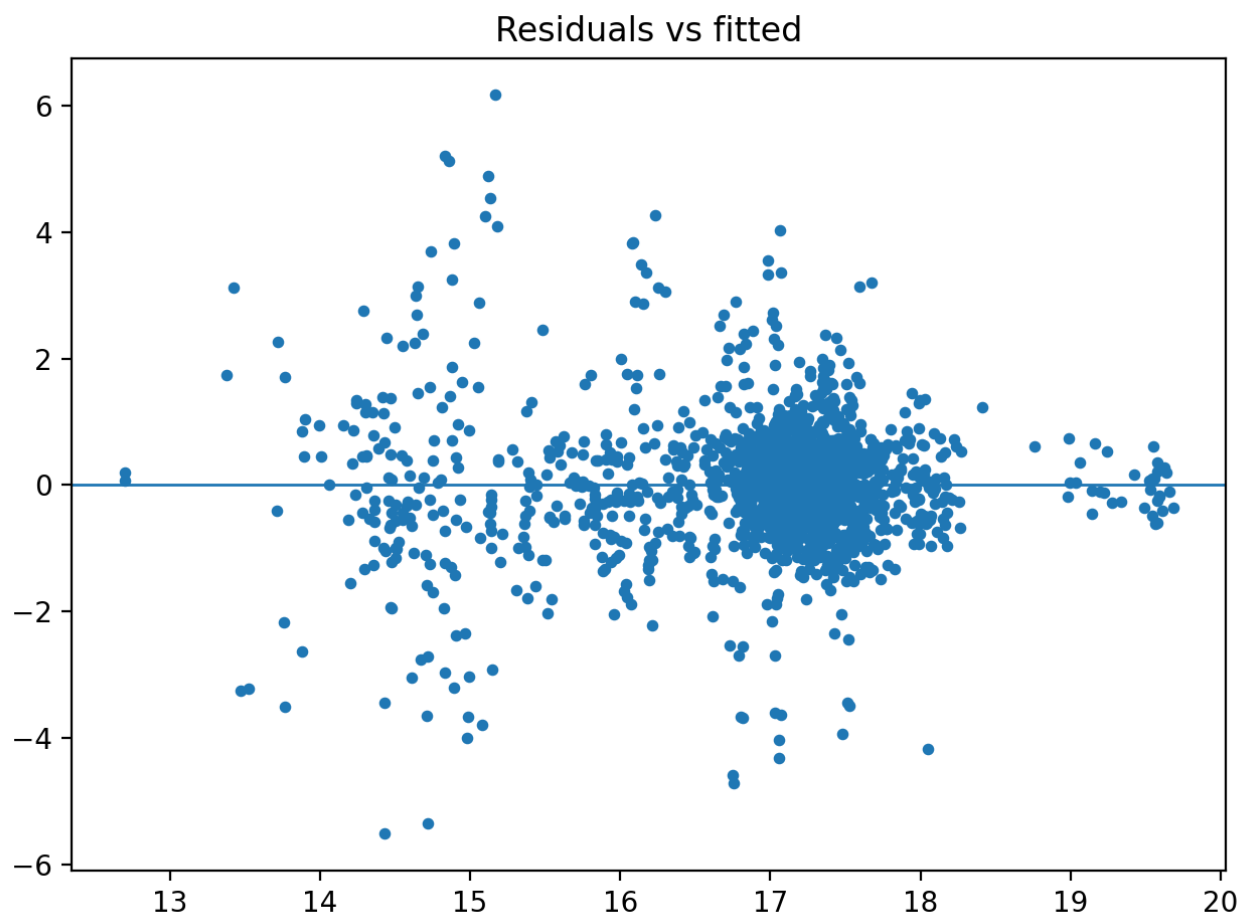


Figure 3. Residuals vs fitted values.

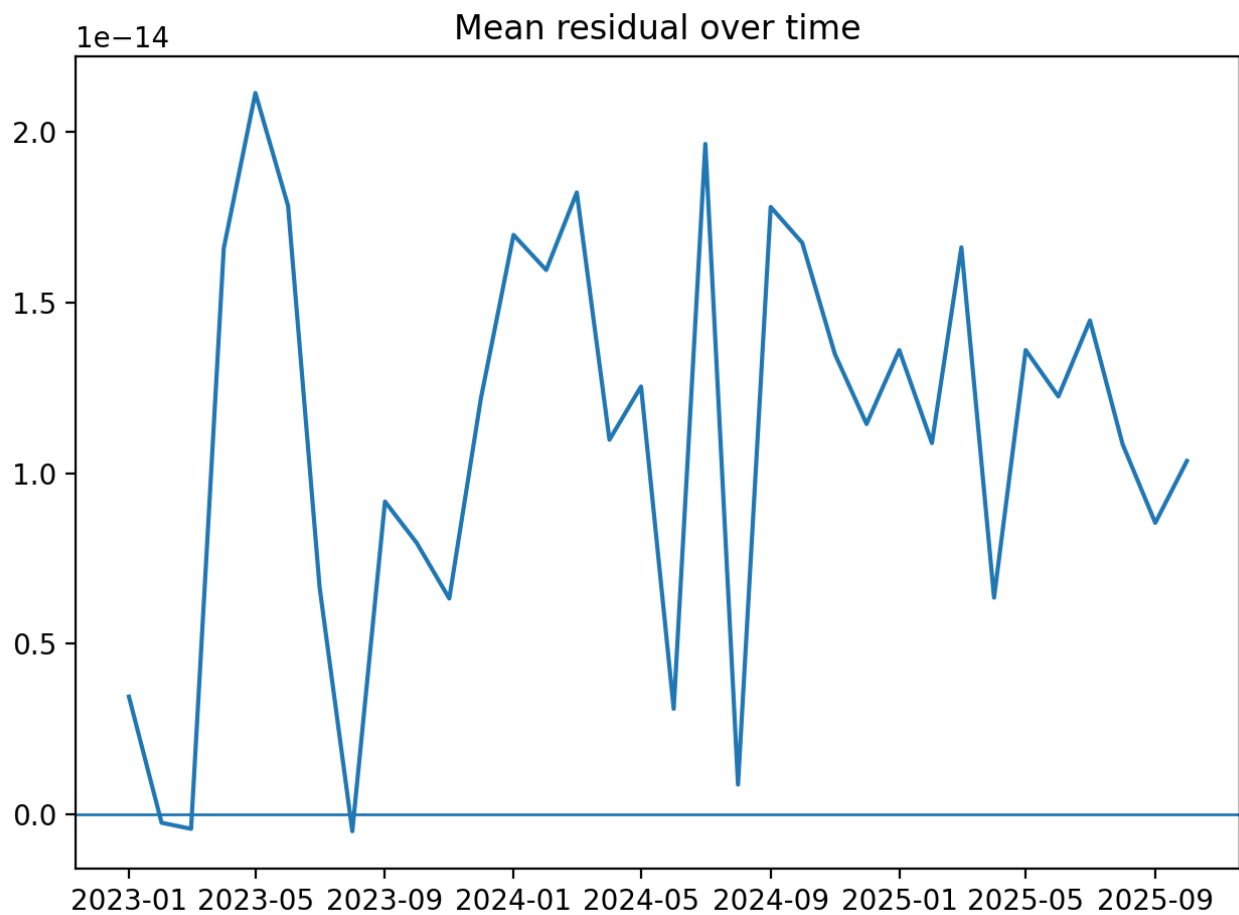


Figure 4. Mean residual over time.

Figure: Missing image: acf_monthly_mean_residual.png