

ZIP-Code–Level Drivers of EV Adoption in Washington: Socioeconomics, Urbanization, and Charger Proximity (2020–2025)

Yiwen Xiong^{1*†}, Junqi Ling^{2†}, Yifei Wang^{3†}

¹*School of Management and Economics, University of Electronic Science and Technology of China, Chengdu, China*

²*Woodsworth College, University of Toronto, Toronto, Canada*

³*Raffles Institution, Singapore, Singapore*

**Corresponding Author. Email: cdevan04@std.uestc.edu.cn*

†These authors contributed equally to this work and should be considered co-first authors.

Abstract. Electric vehicles (EVs) offer lower noise and zero tailpipe emissions compared to traditional internal combustion engine (ICE) vehicles. This study quantifies EV adoption across Washington State using ZIP-code–level data from 2020–2025. Multivariable linear regressions relate the monthly change in EV share to socioeconomic factors, urbanization, family structure, and the roll-out of public charging infrastructure. Results indicate that adoption growth is positively associated with higher per-capita income, higher educational attainment, greater urbanization, and larger average family size. Proximity to infrastructure matters: the number of newly built local stations (0–20 miles) is positively associated with adoption growth, whereas a higher number of newly built distant stations (≥ 30 miles) correlates negatively, consistent with rurality and limited access. By providing a comprehensive ZIP-code–level analysis, the study addresses aggregation bias common in county-level work and reveals localized patterns that broader geographies can obscure. The findings offer policy-relevant evidence for charger siting and for targeting incentives.

Keywords: EV adoption, ZIP-code analysis, charger proximity, per-capita income, multivariable linear regression

1. Introduction

The urgency of mitigating greenhouse-gas emissions has focused attention on electrified transport. Transportation accounted for about 28% of U.S. greenhouse-gas emissions in 2022, making it the largest contributor [1]. Because EVs have zero tailpipe emissions, broader adoption can materially lower CO₂; global scenarios project savings on the order of hundreds of millions of tons if current policy pledges are realized [2]. Many governments therefore deploy financial incentives to accelerate adoption; for example, Germany’s federal program led to a 250% increase in subsidized EV purchases between 2020 and 2021 and delivered CO₂ savings of roughly 3 million tons [3]. Similar incentive effects have been documented in China [4].

Prior work highlights income and charging access as central drivers. Studies report higher adoption with higher income [5] and emphasize that insufficient public charging can impede market growth [6]. Within King County, Washington, underinvestment in lower-income areas aligns with lower adoption [7]. Consistent with this literature, the present analysis finds positive correlations between income, charging-station availability, and EV adoption trends.

A common limitation in the literature is spatial aggregation. Analyses at the state or county level can mask local heterogeneity; coarser geographies introduce aggregation error relative to more granular units [8]. ZIP-code-level analysis can therefore reveal patterns that broader unit's obscure.

This study assembles a 2020–2025 ZIP-code-level dataset for Washington State and estimates multivariable linear regressions linking EV adoption to urbanization, per-capita personal income, educational attainment, charger access, and family size. The analysis documents (i) positive associations with income, education, urbanization, the number of newly built local (0–20 miles) charging stations, and family size; and (ii) a negative association with the number of newly built distant (≥ 30 miles) charging stations, consistent with rurality effects. These results underscore the localized nature of EV adoption and inform targeted infrastructure and incentive design.

The remainder of the report is organized as follows. Section 2 describes data sources, cleaning, feature engineering, and merging. Section 3 details model selection, variable screening, and assumption checks. Section 4 presents conclusions. Section 5 discusses limitations and future directions. Section 6 lists references.

2. Data

2.1. Data acquisition and visualization

All data were harmonized to the ZIP-code level for Washington State over 2020–2025. Vehicle registration transactions were obtained from data.wa.gov; ZIP-code geometry and coordinates from SimpleMaps; income, population, age, and household characteristics from NHGIS; urban/rural status from data.census.gov; and public charging-station information from the U.S. DOE Alternative Fuels Data Center (AFDC). In total, twelve datasets were assembled, several exceeding 30 million records. After standardizing identifiers and geographic keys, exploratory data analysis (EDA) guided variable selection and model design.

Figure 1 illustrates the strong right-skew in population density, with urban ZIPs anchoring the upper tail. Figure 2 depicts a clear positive association between population density and the share of EVs among registered vehicles, especially apparent once urban ZIPs are highlighted. Figure 3 documents a persistent five-year uptrend in EV adoption, with urban areas leading as early adopters. Figure 4 maps the annual rollout of charging stations, showing concentration in urban areas with gradual penetration into rural regions.

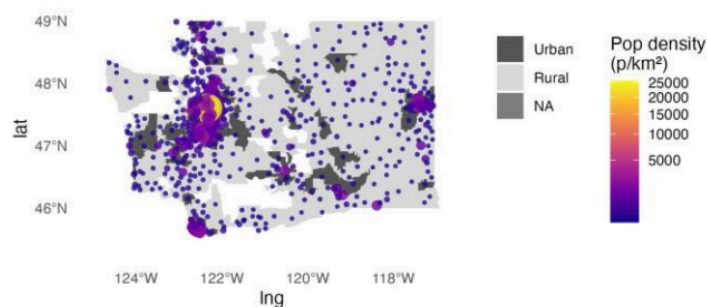


Figure 1. Population density by ZIP centroid with urban/rural background

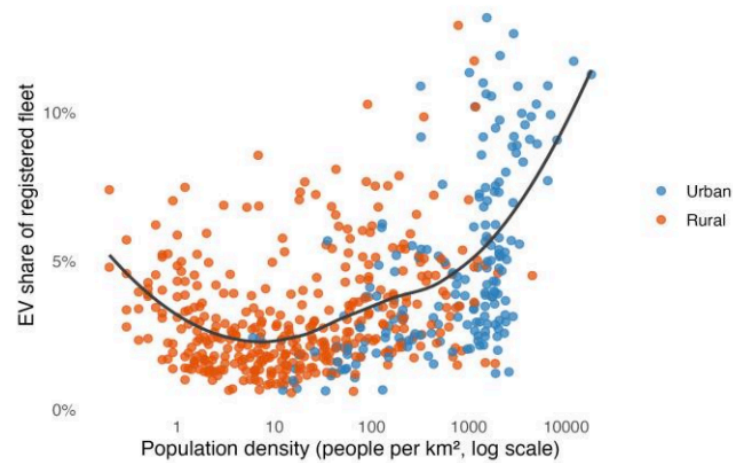


Figure 2. Population density vs. EV share among registered vehicles

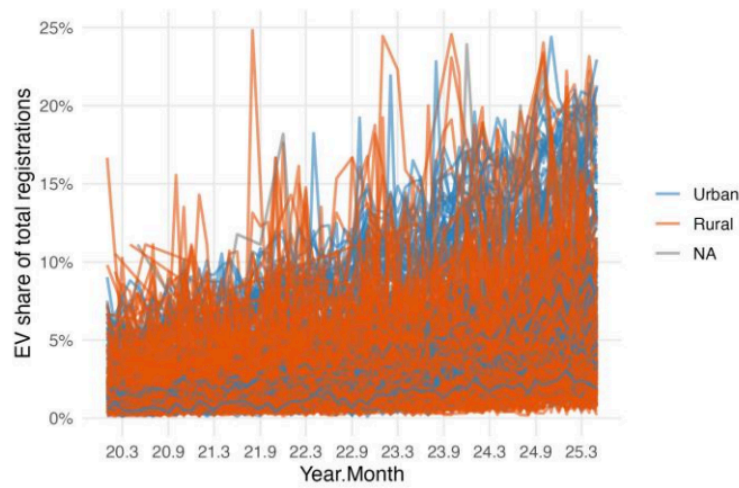


Figure 3. Monthly share of EV registrations by ZIP

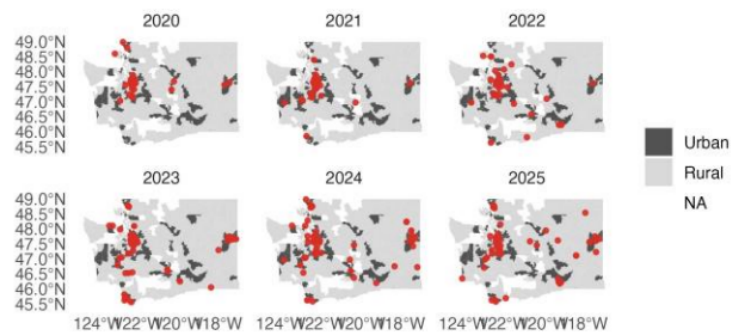


Figure 4. New EV charging stations, urban/rural background (2020–2025)

2.2. Data pre-processing

Data pre-processing ensured internal consistency, reproducibility, and tractable model size. Procedures addressed missing data, exclusion rules for outliers and unreliable observations, feature engineering,

and final merging.

2.2.1. Handling missing values

One of the primary concerns in any dataset is the presence of missing values, and complete datasets were required in the regression analysis. To address this issue, missing values were assessed before and after merging. Prior to merging, ZIPs with extensive missing fields—rare and concentrated in very sparsely populated areas with multiple unreliable statistics—were removed to avoid undue influence from noisy observations. After merging, mean imputation was applied to education proportions when necessary, using the statewide ZIP mean; imputed ZIPs constituted less than 10% of the analytic sample. This approach preserves coverage while acknowledging a potential attenuation of estimated effects, discussed in the discussion part.

2.2.2. Outliers and exclusions

EDA (boxplots and scatterplots) revealed a small number of manifest anomalies. To reduce spurious leverage, observations (ZIPs) were excluded when cumulative registrations over 2020–2025 were below 1,000, when the EV share exceeded 24.9% (implausible for the period), when average family size exceeded 10 (indicative of data error), or when the outcome construction (Section 2.2.3) yielded a non-significant trend (slope p -value > 0.05). Thresholds were set *ex ante* based on distributional evidence; sensitivity to alternative cutoffs is reported elsewhere.

2.2.3. Feature engineering

Categorical variables were encoded numerically and, where appropriate, regrouped to mitigate multicollinearity. The outcome variable is the monthly trend in EV share at the ZIP level. For each ZIP i , a monthly EV-share series (EV registrations divided by total registrations) was constructed and the following OLS was estimated (Eq. Equation 1):

$$EVShare_{it} = \alpha_i + \beta_i t + \varepsilon_{it}. \quad (1)$$

The slope β_i (percentage points per month) is retained when statistically significant at $p \leq 0.05$.

Education composition was expressed as ZIP-level percentages grouped into mid-low (\leq high school), mid-high (high school to some college), and high (\geq college), with the very-low category suppressed to limit collinearity. Age composition was expressed as ZIP-level percentages for the bands 1–20, 21–30, ..., 71–80, with ages ≥ 81 suppressed for the same reason. Urbanization was encoded as a binary indicator (Urban.Rural = 1 for urban, 0 for rural). Charging-infrastructure access was measured as the number of new public charging stations relative to the ZIP centroid in fixed distance bands of 0–10, 10–20, 20–30, and ≥ 30 miles (num_0_10, num_10_20, num_20_30, num_30_plus). Distances were computed as great-circle distances from ZIP centroids; with fixed radii, counts are directly interpretable, and density-based normalizations are examined in robustness checks.

2.2.4. Final analysis dataset

The final merged dataset comprises the following variables: the ZIP identifier; population density (people per km²); per-capita personal income (USD); education shares (edu_midlow, edu_midhigh, edu_high); age-band shares from 1–20 through 71–80; an urbanization indicator (Urban.Rural, with 1 denoting urban); counts of new charging stations within 0–10, 10–20, 20–30, and ≥ 30 miles

(num_0_10, num_10_20, num_20_30, num_30_plus); average family size (people); and the outcome slope (percentage-point change in EV share per month). Table 1 provides concise definitions and units.

Table 1. Variable definitions and units used in the analysis dataset

Variable	Description	Unit	Range
zip	US ZIP code (identifier)	—	98000–99400
population_density	Population density	people per km ²	0.2–18200
PCPIE	Per-capita personal income	USD	15090–129100
edu_midlow	Education: middle or lower	% of pop.	0.002425–0.2156
edu_midhigh	Education: middle–high	% of pop.	0.1102–0.8525
edu_high	Education: high	% of pop.	0.03178–0.8823
age_1_20	Age 1–20	% of pop.	0.01864–0.4503
age_21_30	Age 21–30	% of pop.	0.005293–0.3963
age_31_40	Age 31–40	% of pop.	0.004869–0.3119
age_41_50	Age 41–50	% of pop.	0.02876–0.3495
age_51_60	Age 51–60	% of pop.	0.0199–0.5155
age_61_70	Age 61–70	% of pop.	0.03433–0.399
age_71_80	Age 71–80	% of pop.	0.009093–0.2983
Urban.Rural	Urban / rural classification	—	0–1
num_0_10	New EV charging stations within 10 miles	stations	0–1195
num_10_20	New EV charging stations between 10–20 miles	stations	0–1318
num_20_30	New EV charging stations between 20–30 miles	stations	0–1185
num_30_plus	New EV charging stations ≥30 miles	stations	290–2731
family_size	Average family size	people	1.459–4.417
slope	Slope of EV time-series regression	pp/month	7.842e-05–0.002535

3. Methods

3.1. Model building

The empirical strategy proceeds from a saturated linear specification to progressively more parsimonious models, with formal tests at each step. Let the dependent variable be the monthly change in EV share at the ZIP level (Section 2.2.3). The initial (“full”) model, M1, includes socioeconomic composition, urbanization, and charging-access measures:

$$\begin{aligned}
 \text{slope} &= -4.295 \times 10^{-4} \\
 &+ 7.130 \times 10^{-9} \text{population_density} + 8.130 \times 10^{-9} \text{PCPIE} \\
 &+ 7.990 \times 10^{-5} \text{edu_midlow} + 5.448 \times 10^{-5} \text{edu_midhigh} \\
 &+ 1.313 \times 10^{-3} \text{edu_high} - 7.058 \times 10^{-4} \text{age_1_20} \\
 &- 5.184 \times 10^{-4} \text{age_21_30} + 7.083 \times 10^{-4} \text{age_31_40} \\
 &+ 8.928 \times 10^{-4} \text{age_41_50} + 1.854 \times 10^{-4} \text{age_51_60} \\
 &- 4.113 \times 10^{-4} \text{age_61_70} + 8.954 \times 10^{-5} \text{age_71_80} \\
 &+ 6.657 \times 10^{-5} \text{Urban.Rural} + 2.628 \times 10^{-7} \text{num_0_10} \\
 &+ 3.531 \times 10^{-7} \text{num_10_20} + 5.221 \times 10^{-8} \text{num_20_30} \\
 &- 3.780 \times 10^{-8} \text{num_30_plus} + 1.315 \times 10^{-4} \text{family_size} .
 \end{aligned} \tag{2}$$

Coefficient-wise inference uses two-sided t tests of $H_0 : \beta_j = 0$ against $H_1 : \beta_j \neq 0$. Guided by these tests at $\alpha = 0.05$, we retain predictors with $p < 0.05$ to form M2 via backward elimination.

The resulting specification M2 is

$$\begin{aligned}
 \text{slope} = & -2.955 \times 10^{-4} + 9.329 \times 10^{-9} \text{PCPIE} + 1.230 \times 10^{-3} \text{edu_high} \\
 & + 7.901 \times 10^{-5} \text{Urban.Rural} + 3.045 \times 10^{-7} \text{num_0_10} + 3.991 \times 10^{-7} \text{num_10_20} \\
 & - 4.794 \times 10^{-8} \text{num_30_plus} + 7.184 \times 10^{-5} \text{family_size} .
 \end{aligned} \tag{3}$$

All retained coefficients in M2 are statistically significant at the 5% level. Variance inflation factors (VIFs) computed on M2 do not exceed 5, so the VIF-screened model M3 coincides with M2.

To evaluate parsimony against goodness of fit, a deliberately reduced model M4 is constructed by dropping two predictors from M3 (those with the smallest absolute t -statistics). A partial- F test is used for the nested comparison of M3 (full) versus M4 (reduced), complemented by information criteria. The test strongly favors M3(full) specification (P-value = 9.824×10^{-7}); in addition, M3 attains not only higher R^2 (0.8397 vs. 0.8320) and adjusted R^2 (0.8373 vs. 0.8302), but also lower AIC (-6676.96 vs. -6652.82) and BIC (-6639.55 vs. -6623.72), indicating a superior fit without undue complexity.

Motivated by exploratory evidence that local charger supply may operate differently in urban settings, an interaction between urbanization and nearby charging growth is introduced. The interaction-augmented model M5 adds $\text{Urban.Rural} \times \text{num_0_10}$ to M3 and is compared to M3 via a partial- F test. The improvement in P-value (0.000742) is statistically significant, while AIC, BIC, R^2 , and adjusted R^2 remain similar in magnitude, supporting inclusion of the interaction. So the final specification M5 is

$$\begin{aligned}
 \text{slope} = & -2.501 \times 10^{-4} + 9.026 \times 10^{-9} \text{PCPIE} + 1.189 \times 10^{-3} \text{edu_high} \\
 & + 1.017 \times 10^{-4} \text{Urban.Rural} + 1.190 \times 10^{-6} \text{num_0_10} + 3.688 \times 10^{-7} \text{num_10_20} \\
 & - 4.892 \times 10^{-8} \text{num_30_plus} + 5.966 \times 10^{-5} \text{family_size} \\
 & - 8.786 \times 10^{-7} (\text{Urban.Rural} \times \text{num_0_10}) .
 \end{aligned} \tag{4}$$

All of the model parameters are concluded in Table 2, with M2 being the same as M3 and M5 being our final model. Note that: * indicates $P < 0.1$, ** indicates $P < 0.05$, *** indicates $P < 0.01$.

Table 2. Model coefficients with p-values

Variable	Model_1	Model_3	Model_4	Model_5
(Intercept)	-4.295e-04 (p = 0.423278)	-2.955e-04 *** (p = 0.000481)	-4.869e-05 (p = 0.230178)	-2.501e-04 ** (p = 0.003319)
pop_density	7.130e-09 (p = 0.516168)	—	—	—
PCPIE	8.130e-09 *** (p = 5.21e-13)	9.329e-09 *** (p < 2e-16)	8.937e-09 *** (p = 4.46e-16)	9.026e-09 *** (p < 2e-16)
edu_midlow	7.990e-05 (p = 0.869053)	—	—	—
edu_midhigh	5.448e-05 (p = 0.812045)	—	—	—
edu_high	1.313e-03 *** (p = 2.56e-07)	1.230e-03 *** (p < 2e-16)	1.191e-03 *** (p < 2e-16)	1.189e-03 *** (p < 2e-16)
age_1_20	-7.058e-04 (p = 0.189066)	—	—	—
age_21_30	-5.184e-04 (p = 0.322986)	—	—	—
age_31_40	7.083e-04 (p = 0.201556)	—	—	—
age_41_50	8.928e-04 (p = 0.114511)	—	—	—
age_51_60	1.854e-04 (p = 0.722269)	—	—	—
age_61_70	-4.113e-04 (p = 0.469325)	—	—	—
age_71_80	8.954e-05 (p = 0.887181)	—	—	—
Urban.Rural	6.657e-05 ** (p = 0.004749)	7.901e-05 *** (p = 0.000353)	—	1.017e-04 *** (p = 1.29e-05)
num_0_10	2.628e-07 *** (p = 0.000274)	3.045e-07 *** (p = 5.97e-06)	3.740e-07 *** (p = 1.99e-08)	1.190e-06 *** (p = 0.000112)
num_10_20	3.531e-07 *** (p = 6.21e-12)	3.991e-07 *** (p = 2.68e-15)	4.425e-07 *** (p < 2e-16)	3.688e-07 *** (p = 4.24e-13)
num_20_30	5.221e-08 (p = 0.216048)	—	—	—
num_30_plus	-3.780e-08 ** (p = 0.009471)	-4.794e-08 *** (p = 0.000989)	-5.733e-08 *** (p = 0.000102)	-4.892e-08 *** (p = 0.000706)
family_size	1.315e-04 ** (p = 0.002568)	7.184e-05 ** (p = 0.003806)	—	5.966e-05 * (p = 0.016694)
Urban.Rural:num_0_10	—	—	—	-8.786e-07 ** (p = 0.003134)
R_squared	0.85	0.8397	0.832	0.8427
adj_Rsquared	0.8441	0.8373	0.8302	0.84
AIC	—	-6676.959	-6652.816	-6677.992

BIC	—	-6639.546	-6623.717	-6636.423
-----	---	-----------	-----------	-----------

3.2. Regression diagnostics

Model adequacy is evaluated against standard linear-model conditions: linearity of the conditional mean, uncorrelated errors, homoskedastic disturbances, and approximately normal residuals. Plots of observed responses against fitted values show close alignment with the 45-degree line, consistent with an adequate linear signal; deviations are small and unsystematic (Figure 5).

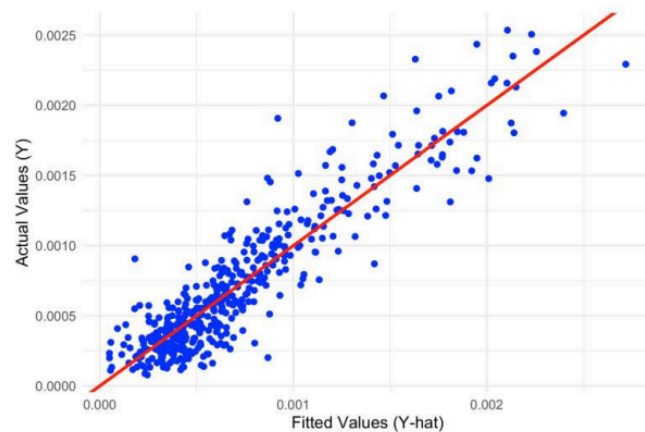


Figure 5. Actual vs. fitted values (linearity of the conditional mean)

Pairwise associations among retained predictors appear roughly linear. Given the visual correlation between per-capita income (PCPIE) and the high-education share, we compute variance inflation factors (VIFs) to assess multicollinearity. All VIFs are below 5, indicating limited collinearity and no need to exclude additional predictors (Figure 6).

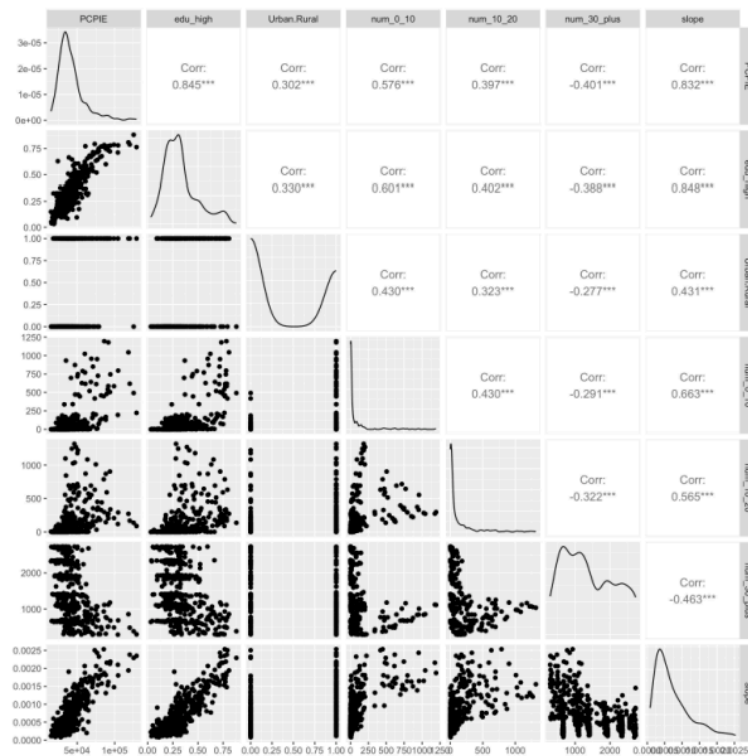


Figure 6. Pairwise relationships among predictors (screening for multicollinearity)

Residual-versus-fitted diagnostics show residuals centered near zero without pronounced curvature. A mild fan-shape at higher fitted values suggests slight heteroskedasticity; to guard against size distortions in hypothesis tests, we report heteroskedasticity-consistent (HC3) standard errors alongside conventional ones (Figure 7).

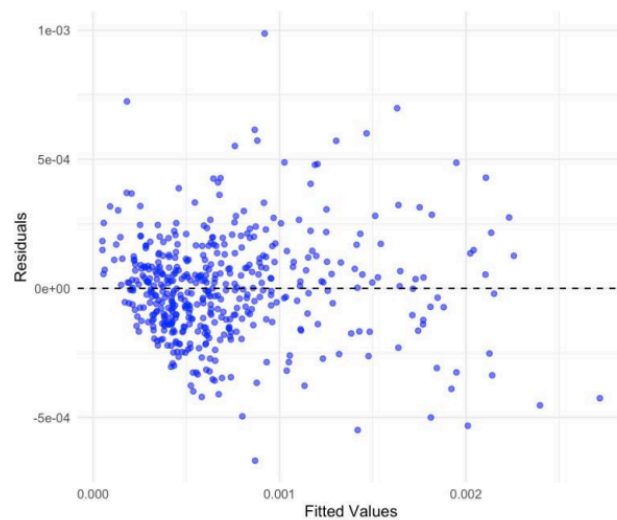


Figure 7. Residuals vs. fitted (checking linearity and constant variance)

The normal Q–Q plot indicates residuals track the reference line closely with modest tail deviations, supporting approximate normality and the reliability of large-sample inference (Figure 8).

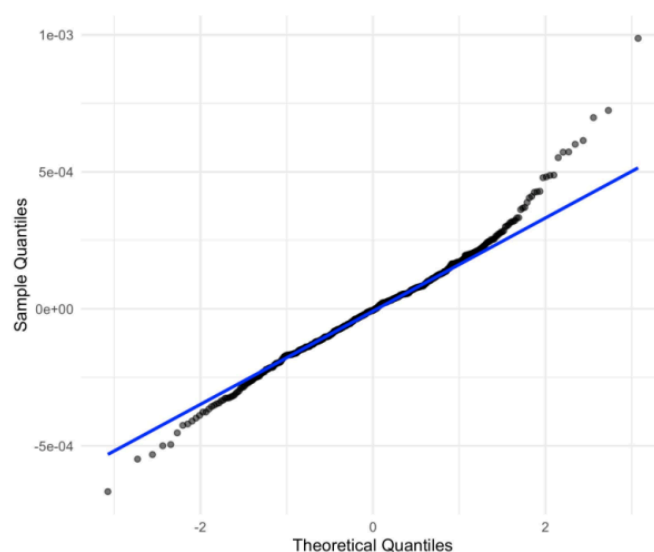


Figure 8. Normal Q–Q plot of residuals (approximate normality)

Outliers, leverage points, and influential observations should be assessed to ensure model reliability. A data point is considered an outlier if its standardized residual lies outside the range of -4 to 4 . A leverage point is one that is far from the mean of the predictor values and can substantially affect the fitted regression line. The threshold for leverage is defined as: $2 \times (\text{number of predictors} + 1) / n$, where n is the sample size.

Influential observations are identified using Cook's Distance. A data point is considered influential if its Cook's Distance exceeds the 50th percentile of the F distribution with degrees of freedom $(p+1, n-p-1)$, where p is the number of predictors and n is the total number of observations.

By results, most standardized residuals fall within -4 and 4 , with only a few potential outliers (e.g., observation 84). Cook's Distance shows most points have low influence, with only a few exceeding the cutoff. Overall, the model is stable with limited influential cases.

4. Conclusion

This study quantifies ZIP-code-level EV adoption dynamics in Washington State over 2020–2025 by relating the monthly change in EV share to socioeconomic composition, urbanization, and the roll-out of public charging infrastructure. Multivariable linear models indicate that higher per-capita income, a larger share of highly educated residents, and urban status are each positively associated with faster growth in EV share. Access to charging infrastructure is salient at near distances: growth in the number of new stations within 0–10 miles and 10–20 miles is positively associated with adoption growth, whereas a greater concentration of distant stations (≥ 30 miles) is negatively associated with adoption, consistent with the distinct constraints of rural geographies. An interaction between urbanization and nearby station growth shows a weaker marginal association in urban ZIPs than in non-urban ZIPs, suggesting diminishing returns where baseline access is already high. These results answer the central questions posed in the Introduction by identifying the principal correlates of adoption and by demonstrating that spatial granularity at the ZIP level reveals localized patterns that broader geographies can obscure. Policy implications follow directly: charger siting decisions that prioritize proximity—particularly in underserved, lower-access areas—are likely to be more effective than strategies that expand distant infrastructure, and incentives targeted by local socioeconomic profiles may accelerate uptake where the marginal response is greatest. Overall, the findings provide an

empirically grounded basis for fine-grained infrastructure planning and for designing geographically targeted incentives, while also motivating further causal analysis of infrastructure–adoption dynamics.

5. Discussion

Several considerations qualify the interpretation and indicate directions for further work. First, the outcome is a ZIP-level trend (the slope of monthly EV share), which improves comparability across places but does not by itself resolve endogeneity between charger deployment and demand. Infrastructure may be installed preferentially where adoption is already rising, potentially biasing naive associations upward. Future analyses should therefore incorporate identification strategies—such as panel fixed effects with time-varying controls, difference-in-differences around exogenous policy shocks, or instrumental variables exploiting program eligibility rules or grid siting constraints—to sharpen causal interpretation.

Second, the construction of the analytic dataset necessarily involved handling missing data and excluding manifest anomalies. Mean imputation for education shares was applied to fewer than ten percent of ZIPs to preserve coverage. This approach preserves coverage while acknowledging a potential attenuation of estimated effects, because mean imputation can shrink cross-sectional variance and bias coefficients toward zero; reporting heteroskedasticity-consistent standard errors (HC3) and comparing results to complete-case estimates partly mitigates concerns but does not eliminate them. Relatedly, exclusion thresholds—minimum cumulative registrations, extreme EV shares, implausible family sizes, and non-significant outcome trends—were set *ex ante* based on distributional evidence; sensitivity to alternative cutoffs is reported elsewhere and leaves the principal findings intact. Nonetheless, replication with alternative outlier rules and multiple-imputation procedures would further validate robustness.

Third, spatial structure may matter. Although fixed-radius counts of new stations offer transparent interpretation, spatial spillovers and network effects can induce dependence across neighboring ZIPs. Extensions using spatial error or spatial lag specifications, geographically weighted regression, or gravity-based accessibility indices could capture these interactions more explicitly. Likewise, aligning supply with potential demand via per-capita or per-vehicle accessibility, and distinguishing rapid-charging from lower-power stations, may refine effect heterogeneity. The documented weaker marginal association of nearby station growth in urban areas suggests diminishing returns in already dense networks; modeling nonlinearity (e.g., spline terms for charger access) would test this mechanism directly.

Finally, measurement and scope limitations remain. Registration and station inventories may contain dating or classification errors; charger operability and reliability are not observed; and important covariates—such as local electricity prices, parking availability, housing type, or model availability—are not yet integrated. Incorporating these factors, expanding beyond one state, and exploiting policy discontinuities would strengthen external validity and causal claims. Taken together, the present results underscore the importance of proximity-based infrastructure for accelerating adoption while highlighting methodological avenues—causal identification, richer spatial modeling, and improved data—to advance the evidence base.

Acknowledgement

Yiwen Xiong and Junqi Ling contributed equally to this work and should be considered co-first authors.

References

- [1] United States Environmental Protection Agency, “Transportation sector emissions | US EPA.” Mar. 2025, [Online]. Available: <https://www.epa.gov/ghgemissions/transportation-sector-emissions>.
- [2] International Energy Agency, “Global EV outlook 2022,” IEA, Paris, May 2022. [Online]. Available: <https://www.iea.org/reports/global-ev-outlook-2022>.
- [3] M. Schulthoff, P. Anstett, J. Lange, D. Arend, E. Gencer, and M. Kaltschmitt, “Federal transformation costs of e-mobility in germany: Effectiveness and efficiency of EV incentives between 2015 and 2023,” *Transportation Research Interdisciplinary Perspectives*, vol. 31, p. 101435, May 2025, doi: 10.1016/j.trip.2025.101435.
- [4] T. Lu, E. Yao, F. Jin, and Y. Yang, “Analysis of incentive policies for electric vehicle adoptions after the abolishment of purchase subsidy policy,” *Energy*, vol. 239, p. 122136, 2022, doi: 10.1016/j.energy.2021.122136.
- [5] Z. He et al., “Examining the spatial mode in the early market for electric vehicles adoption: Evidence from 41 cities in china,” *Transportation Letters*, vol. 14, no. 6, pp. 640–650, 2021, doi: 10.1080/19427867.2021.1917217.
- [6] D. L. Greene, E. Kontou, B. Borlaug, A. Brooker, and M. Muratori, “Public charging infrastructure for plug-in electric vehicles: What is it worth?” *Transportation Research Part D: Transport and Environment*, vol. 78, p. 102182, 2020, doi: 10.1016/j.trd.2019.11.011.
- [7] S. Ding and L. Wu, “Disparities in electric vehicle charging infrastructure distribution: A socio-spatial clustering study in king county, washington,” *Sustainable Cities and Society*, vol. 121, p. 106193, Mar. 2025, doi: 10.1016/j.scs.2025.106193.
- [8] L. Luo, S. McLafferty, and F. Wang, “Analyzing spatial aggregation error in statistical models of late-stage cancer risk: A monte carlo simulation approach,” *International Journal of Health Geographics*, vol. 9, p. 51, 2010, doi: 10.1186/1476-072X-9-51.