

Supplementary Materials

Understanding the influencing factors of electric vehicle penetration in China: a city-level analysis

Yunfeng Deng^{1,2}, Han Hao^{1,2}, Xin Sun³, Bu Zhao⁴, Hao Dou^{1,2}, Lang Mai^{1,2}, Hanjun Deng⁵

¹State Key Laboratory of Automotive Safety and Energy, Tsinghua University, Beijing 100084, China.

²Tsinghua-Rio Tinto Joint Research Center for Resources Energy and Sustainable Development, Tsinghua University, Beijing 100084, China.

³Integrated Research on Energy, Environment and Society (IREES), Energy and Sustainability Research Institute Groningen (ESRIG), University of Groningen, Groningen 9747AG, the Netherlands.

⁴Department of Environmental and Sustainable Engineering, University at Albany, State University of New York, Albany, NY 12222, USA.

⁵College of Arts and Science, New York University, New York, NY 10012, USA.

Correspondence to: Prof. Han Hao, State Key Laboratory of Automotive Safety and Energy, Tsinghua University, Beijing 100084, China. E-mail: hao@tsinghua.edu.cn

Supplementary Text

Data description

Vehicle dataset: The vehicle dataset includes two key variables: (i) Vehicle sales and (ii) Vehicle stock. To ensure data accuracy and comprehensive coverage, we integrate multiple data sources: (1) The Ministry of Industry and Information Technology of the People's Republic of China (MIIT); (2) China's Compulsory Traffic Accident Liability Insurance (CTALI) database, which provides annual passenger vehicle registration data at the prefecture-level city scale. This dataset contains approximately 25 million entries per year, which we aggregate to construct the final city-level vehicle dataset covering 2019–2024. Throughout, EVs comprise battery electric and plug-in hybrid vehicles.

Climate factors: To characterize climatic conditions, we employed two variables: (i) Average temperature, calculated as the arithmetic mean of monthly average temperatures across the 12 calendar months for each city. All data were obtained from China Gridded Surface Air Temperature Dataset (V2.0); and (ii) Temperature variance, calculated as the variance of these monthly average temperatures, reflecting seasonal fluctuations.

Air Quality factors: Air quality was assessed using the following indicators: (i) Air Quality Index (AQI); and (ii) Concentrations of PM_{2.5}, PM₁₀, NO₂, CO, SO₂ and O₃. All data were obtained from the China Air Quality Online Monitoring and Analysis Platform (<https://www.aqistudy.cn/historydata/>), which provides city-level historical monitoring data.

Energy price factors: Two energy-price variables were considered: (i) Gasoline prices, calculated as the average of RON 92 and RON 95 gasoline prices over the 12 months of each year. All data were obtained from the eastmoney.com (https://data.eastmoney.com/cjsj/oil_default.html) (ii) Electricity price, defined as the government-regulated first-tier residential electricity tariff for standard single-meter households, measured in CNY/kWh. Both gasoline and electricity prices were available only at the provincial level. All cities within the same province were therefore assigned the same annual price. All data come from the provincial Development and Reform Commissions

Infrastructure factors: EV infrastructure was measured using two indicators: (i) Total public charging infrastructure, referring to the number of public charging piles in each city, sourced from the China Electric Vehicle Charging Infrastructure Promotion Alliance (EVCIPA); (ii) Public chargers per EV, defined as the ratio of public chargers to the total number of EVs registered in the city.

Socioeconomic factors: Socioeconomic variables include: (i) GDP per capita, obtained from the China City Statistical Yearbook published by the National Bureau of Statistics. (ii) Population density, defined as the number of permanent residents divided by the city's administrative area. Population and administrative-area data were obtained from the National Bureau of Statistics of China; (iii) Gini coefficient,

calculated based on nighttime lights in the townships using the following equation:

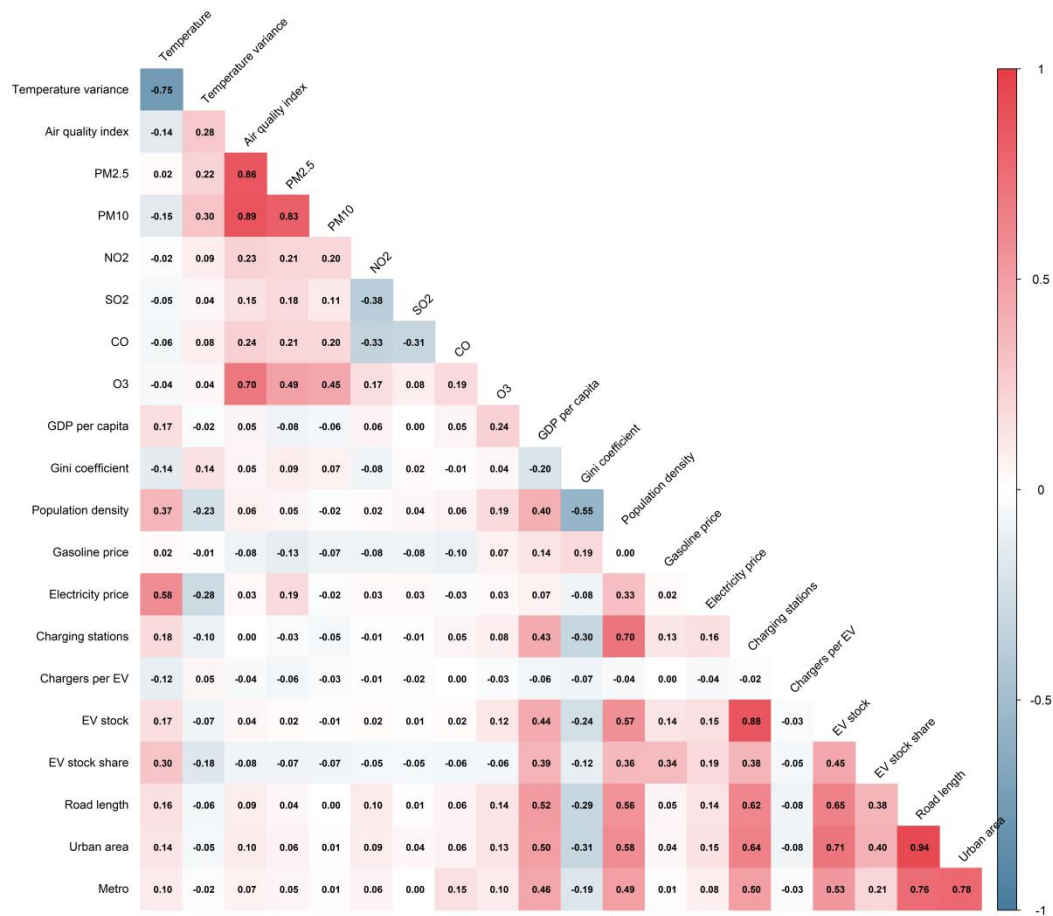
$$G_{jj} = \Delta_{jj} / (2\bar{Y}_j) = \left(\frac{1}{2\bar{Y}_j} \sum_{i=1}^{n_j} \sum_{r=1}^{n_j} |y_{ji} - y_{jr}| \right) n_j^2$$

where Δ_{jj} defines the Gini mean difference within the j -th city, i.e., the mean of the absolute value of the difference in nighttime light intensity for the binary combination of $n \times n$ of town units belonging to city j . y_{ji} and y_{jr} represent nighttime light intensity in towns i and r of city j , respectively. n_j represents the number of towns in city j . $\bar{Y}_j$ represents the average intensity of nighttime lights in city j . Annual nighttime-light data were obtained from the VIIRS nighttime-light composites produced by the Earth Observation Group (<https://eogdata.mines.edu/products/vnl/>). Township-level administrative units included towns, townships, ethnic townships, and subdistricts located within each prefecture-level city. Township units with zero nighttime-light intensity were retained because zero radiance represents little or no detected nighttime activity.

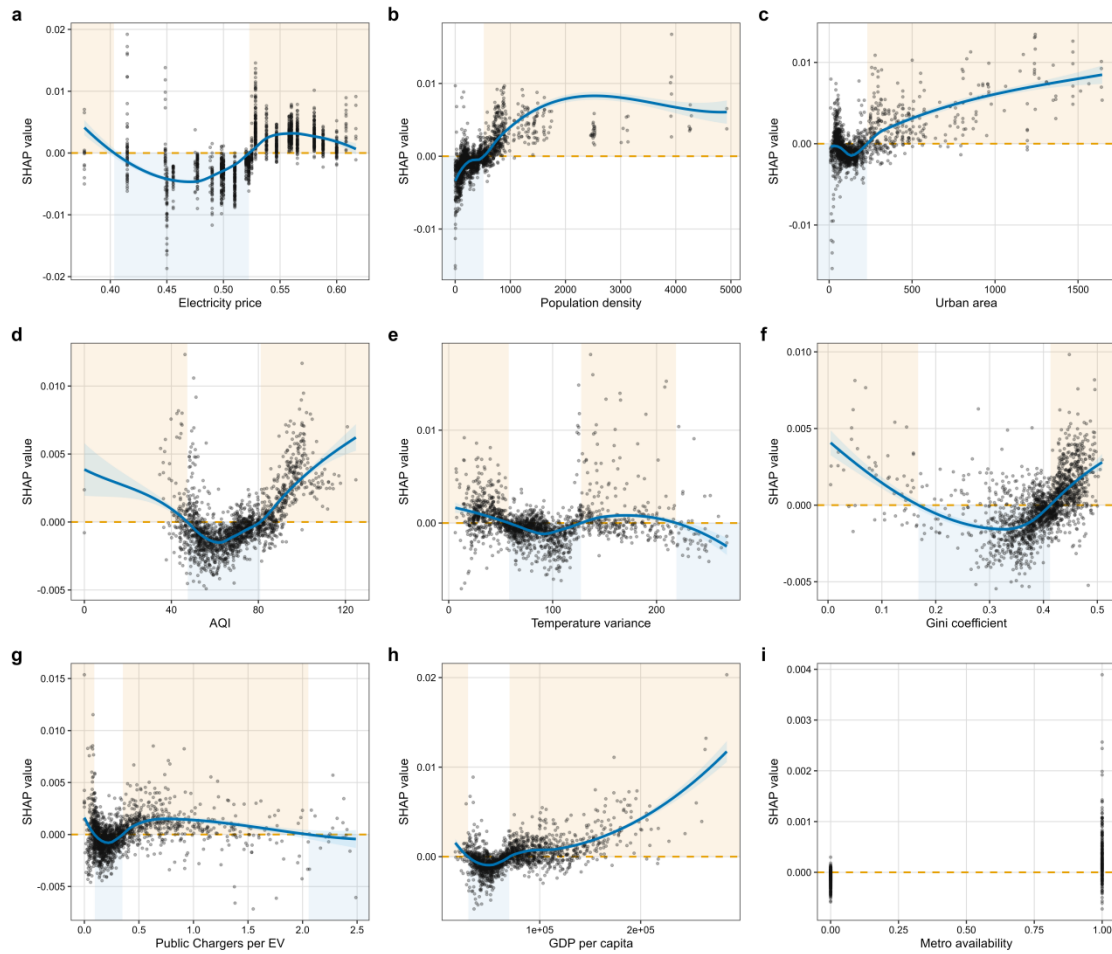
Urban form factors: (i) Urban form was represented by three city-level variables obtained from the China City Statistical Yearbook. Urban area was defined as the built-up area of each city, measured in km². (ii) Road length was defined as the total length of urban roads reported for each city, measured in km. (iii) Metro availability was a binary indicator equal to 1 if the city had an operational urban rail transit system in the corresponding year and 0 otherwise.

Market maturity factors: We included two variables: (i) EV stock, the number of EVs in the local vehicle fleet, derived from the vehicle dataset described above; (ii) EV stock share, defined as the ratio of EV stock to the total vehicle stock in each city, indicating the prevalence of EVs within the urban fleet. To avoid data leakage in model training, the EV stock share for each specific year was calculated using data available only up to the previous year, thereby ensuring that no future information was incorporated into the modeling process.

Supplementary Figures



Supplementary Figure 1. Pearson correlation matrix of the 21 candidate predictors considered during variable screening. Only the lower triangle of the matrix is shown; the value and color of each cell denote the same correlation coefficient, ranging from +1 (perfect positive linear correlation) through 0 (no linear correlation) to -1 (perfect negative linear correlation).



Supplementary Figure 2. SHAP dependence plots for the remaining predictors excluding four leading predictors. The panels show electricity price, population density, urban area, air quality index, temperature variance, Gini coefficient, chargers per EV, GDP per capita, and metro availability. Black points represent individual city–year observations, blue curves indicate LOESS-smoothed relationships with 95% confidence intervals, and orange dashed lines denote a SHAP value of zero. Positive and negative SHAP values indicate contributions toward higher and lower predicted EV penetration rates, respectively.

Supplementary Tables

Supplementary Table 1. Descriptive statistics of the outcome and 13 predictors across 296 Chinese prefecture-level cities, 2019-2024

Variable	Mean	SD	P0	P25	P50	P75	P100
Temperature (°C)	14.9	5.3	-3.9	11.1	16.1	18.3	25.9

Temperature variance (°C ²)	87.8	50.6	6.3	52.5	78.0	111.1	267.6
Air quality index	69.6	16.9	34.7	55.2	68.8	81.7	124.8
GDP per capita (CNY)	73312.	38338.	16302.	45755.	62154.	89791.	28533
Gini coefficient (× 10 ⁻²)	4	2	4	4	1	0	8.0
Population density (persons/km ²)	39.1	7.2	5	36.6	40.3	43.6	50.9
Gasoline price (CNY/L)	493.0	730.1	1.4	155.2	317.8	585.7	8908.4
Electricity price (CNY/kWh)	7.50	0.90	5.83	7.01	7.93	8.18	9.69
Charging stations	0.526	0.046	0.377	0.500	0.528	0.560	0.617
Chargers per EV	5170.2	17209.	0.0	315.8	1008.0	3172.0	31772
EV stock share (%)		0					3.4
Urban area (km ²)	0.600	2.093	0.000	0.141	0.241	0.484	67.000
Metro	2.36	3.05	0.00	0.33	1.13	3.37	32.17
EV penetration rate (%)	178.8	237.8	8.9	62.5	96.1	174.7	1645.4
	0.13	0.34	0.00	0.00	0.00	0.00	1.00
	13.2	11.6	0.0	2.5	11.6	19.6	77.7

Notes: N = 1,774 city-year observations. SD denotes the sample standard deviation; P0 and P100 denote the minimum and maximum. EV penetration rate and EV stock share are expressed as percentages; the Gini coefficient is multiplied by 100. Metro is a binary indicator.

Supplementary Table 2. Variance inflation factors for the 21 candidate predictors before variable screening, based on 1,772 complete city-year observations from 2019-2024

Variable	VIF
AQI	19.0742
Urban area	12.0250
Road length	9.3210

PM ₁₀	7.6098
EV stock	7.4161
PM _{2.5}	7.3609
Infrastructure stock	6.9981
Average temperature	6.3157
O ₃	4.5581
Temperature variance	4.3129
Population density	4.1397
NO ₂	3.5575
CO	3.5242
SO ₂	3.2366
EV stock share	3.0659
Metro availability	2.7847
Gini coefficient	2.2987
Gasoline price	2.1825
GDP per capita	1.9089
Electricity price	1.8127
Public Chargers per EV	1.1431

Note: VIF denotes variance inflation factor. The elevated values for several air-pollution, urban-form, and market-maturity variables motivated the correlation-based screening described in Section 2.4.

Supplementary Table 3. Variance inflation factors for the 13 predictors retained in the final random forest model, based on 1,774 city-year observations from 296 Chinese prefecture-level cities, 2019-2024

Variable	VIF
Average temperature	5.174
Temperature variance	3.564
Population density	3.546
Urban area	3.514
Infrastructure stock	3.006
Metro availability	2.675
EV stock share	2.052

Gini coefficient	1.937
Electricity price	1.733
GDP per capita	1.589
Gasoline price	1.329
Air quality index	1.206
Public chargers per EV	1.127

Note: VIFs were calculated across the 13 retained substantive predictors. Calendar-year indicators were not included in this diagnostic.

Supplementary Table 4. Model performance under the city-grouped validation design and a random observation-level split for the final random forest model with 13 predictors and calendar-year indicators, 296 Chinese prefecture-level cities, 2019-2024

Validation design	Training set	Test set	OOB R2 (training)	Test R2	Test RMS E	Test MAE
City-grouped split (main)	80% of cities (237 cities, all years)	20% of cities held out entirely (59 cities)	0.925	0.902	0.048	0.027
Random observation-level split (robustness)	80% of observations	20% of observations	0.940	0.941	0.035	0.024

Notes: Both designs use an 80/20 partition and identical hyperparameter-tuning procedures. Under the city-grouped design, all years of a held-out city are excluded from training, whereas the random split can place different years of the same city in both sets. OOB, out-of-bag; RMSE, root-mean-square error; MAE, mean absolute error.

Supplementary Table 5. One-way ANOVA results testing regional differences in city-level mean values of EV penetration and 12 continuous predictors across the Northeast, East, Central, and West regions, based on 296 Chinese prefecture-level

cities averaged over 2019-2024

Variable	Source	Df	Sum Sq	Mean Sq	F value	P value	Sig.
EV penetration rate	Region	3	0.4456	0.1485	42.458	8.44e-2	***
	Residuals	292	1.0215	0.0035		3	
Average Temperature	Region	3	3225.23	1075.08	64.555	4.78e-3	***
	Residuals	292	4862.91	16.6538		2	
Temperature variance	Region	3	338871	112957	81.946	1.73e-3	***
	Residuals	292	402502	1378.43		8	
AQI	Region	3	10080.7	3360.24	14.305	1.01e-8	***
	Residuals	292	68591.5	234.902			
GDP per capita	Region	3	6.59e+10	2.20e+10	19.197	2.17e-1	***
	Residuals	292	3.34e+11	1.15e+9		1	
Population density	Region	3	2.67e+7	8.89e+6	20.069	7.45e-1	***
	Residuals	292	1.29e+8	442913		2	
Gasoline price	Region	3	0.3554	0.1185	3.8749	0.0097	**
	Residuals	292	8.9278	0.0306			
Electricity price	Region	3	0.2725	0.0908	77.242	8.26e-3	***
	Residuals	292	0.3434	0.0012		7	
Infrastructure	Region	3	5.15e+9	1.72e+9	8.62	1.68e-5	***

stock							
	Residuals	292	5.81e+10	1.99e+8			
Public Chargers per EV	Region	3	74.4143	24.8048	11.099	6.40e-7	***
	Residuals	292	652.603	2.2349			
EV stock share	Region	3	0.0301	0.01	30.887	2.27e-17	***
	Residuals	292	0.0949	3.25e-4			
Urban area	Region	3	1.93e+6	643611	12.906	6.10e-8	***
	Residuals	292	1.46e+7	49868			
Gini coefficient	Region	3	0.2031	0.0677	20.449	4.69e-12	***
	Residuals	292	0.9667	0.0033			

Notes: Significance codes are *** $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$. Each city contributes one observation, defined as its 2019-2024 mean, to avoid pseudo-replication from repeated annual observations.

Supplementary Table 6. Tukey's HSD pairwise comparisons of regional means for the outcome and predictors tested in Table S5, with 95% confidence intervals and family-wise adjusted p-values

Variable	Comparison	Difference	Lower CI	Upper CI	Adjusted P	Sig.
EV penetration rate	East-Northeast	0.1214	0.0905	0.1522	7.09e-13	***
	Central-Northeast	0.0862	0.055	0.1175	5.05e-11	***
	West-Northeast	0.0508	0.0202	0.0814	1.43e-4	***
	Central-East	-0.0351	-0.0587	-0.0115	8.53e-4	***

Average Temperature	West-East	-0.0706	-0.0933	-0.0479	8.43e-13	***
	West-Central	-0.0355	-0.0587	-0.0122	5.83e-4	***
	East-Northeast	10.869	8.7401	12.999	6.73e-13	***
	Central-Northeast	9.2545	7.0958	11.413	6.93e-13	***
	West-Northeast	6.502	4.3918	8.6123	9.40e-13	***
Temperature variance	Central-East	-1.6149	-3.2438	0.014	0.053	
	West-East	-4.3674	-5.9314	-2.8033	2.87e-11	***
	West-Central	-2.7525	-4.3564	-1.1485	7.71e-5	***
	East-Northeast	-111.99	-131.36	-92.619	6.73e-13	***
	Central-Northeast	-97.909	-117.55	-78.269	6.73e-13	***
AQI	West-Northeast	-103.9	-123.09	-84.697	6.73e-13	***
	Central-East	14.082	-0.7374	28.902	0.0693	
	West-East	8.0949	-6.1349	22.325	0.4571	
	West-Central	-5.9873	-20.58	8.6053	0.714	
	East-Northeast	7.359	-0.6378	15.356	0.0836	
GDP per capita	Central-Northeast	16.212	8.1044	24.319	2.63e-6	***
	West-Northeast	2.7785	-5.1468	10.704	0.8017	
	Central-East	8.8529	2.7352	14.971	0.0013	**
	West-East	-4.5804	-10.455	1.2937	0.1849	
	West-Central	-13.433	-19.457	-7.4093	1.26e-7	***
Population	East-Northeast	44328	26672	61983	2.24e-9	***
	Central-Northeast	18704	803.79	36603	0.0367	*
	West-Northeast	14713	-2784.5	32211	0.1333	
	Central-East	-25624	-39131	-12117	9.36e-6	***
	West-East	-29614	-42584	-16645	6.01e-8	***
Population	West-Central	-3990.3	-17290	9309.5	0.8656	
	East-Northeast	738.75	391.51	1086	5.03e-7	***

density						
Gasoline price	Central-Northeast	268.57	-83.479	620.61	0.2013	
	West-Northeast	35.21	-308.93	379.35	0.9935	
	Central-East	-470.19	-735.83	-204.54	4.19e-5	***
	West-East	-703.54	-958.62	-448.47	4.91e-11	***
	West-Central	-233.36	-494.93	28.218	0.0992	
	East-Northeast	0.0402	-0.051	0.1314	0.6659	
Electricity price	Central-Northeast	-0.009	-0.1015	0.0835	0.9944	
	West-Northeast	0.0759	-0.0145	0.1663	0.1345	
	Central-East	-0.0492	-0.119	0.0206	0.2653	
	West-East	0.0357	-0.0313	0.1027	0.5157	
	West-Central	0.0849	0.0162	0.1536	0.0085	**
	East-Northeast	0.0352	0.0173	0.0531	3.96e-6	***
Infrastructure stock	Central-Northeast	0.05	0.0318	0.0681	5.13e-11	***
	West-Northeast	-0.0225	-0.0402	-0.0047	0.0065	**
	Central-East	0.0148	0.0011	0.0285	0.0285	*
	West-East	-0.0577	-0.0708	-0.0445	6.85e-13	***
	West-Central	-0.0725	-0.0859	-0.059	6.73e-13	***
	East-Northeast	10656	3295	18018	0.0013	**
Public Chargers per EV	Central-Northeast	2974.6	-4488.6	10438	0.7321	
	West-Northeast	1336.5	-5959.1	8632	0.9649	
	Central-East	-7681.8	-13313	-2050.3	0.0028	**
	West-East	-9319.9	-14727	-3912.5	7.09e-5	***
	West-Central	-1638.1	-7183.4	3907.2	0.8709	
	East-Northeast	-1.5477	-2.3278	-0.7677	3.20e-6	***

EV stock share	Central-Northeast	-1.6852	-2.476	-0.8944	4.80e-7	***
	West-Northeast	-1.3647	-2.1377	-0.5916	4.42e-5	***
	Central-East	-0.1375	-0.7342	0.4593	0.9335	
	West-East	0.1831	-0.3899	0.756	0.8424	
	West-Central	0.3205	-0.2671	0.9081	0.4944	
	East-Northeast	0.0302	0.0208	0.0396	7.38e-13	***
	Central-Northeast	0.0173	0.0078	0.0268	2.54e-5	***
	West-Northeast	0.0098	4.41e-4	0.0191	0.0361	*
	Central-East	-0.0129	-0.0201	-0.0057	3.09e-5	***
	West-East	-0.0205	-0.0274	-0.0135	2.47e-12	***
Urban area	West-Central	-0.0075	-0.0146	-4.43e-4	0.0324	*
	East-Northeast	161.54	45.023	278.05	0.0022	**
	Central-Northeast	-6.9816	-125.11	111.15	0.9987	
	West-Northeast	-25.316	-140.79	90.158	0.9419	
	Central-East	-168.52	-257.66	-79.384	1.01e-5	***
	West-East	-186.85	-272.44	-101.27	2.39e-7	***
	West-Central	-18.335	-106.11	69.436	0.9492	
	East-Northeast	-0.0597	-0.0898	-0.0297	2.96e-6	***
	Central-Northeast	-0.0037	-0.0341	0.0268	0.9895	
	West-Northeast	-0.0024	-0.0321	0.0274	0.9969	
Gini coefficient	Central-East	0.0561	0.0331	0.079	6.21e-9	***
	West-East	0.0574	0.0353	0.0794	5.64e-10	***
	West-Central	0.0013	-0.0213	0.0239	0.9989	

Notes: Significance codes are *** $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$. P-values are adjusted for the family of six pairwise comparisons within each variable using Tukey's honestly significant difference procedure.

Supplementary Table 7. Composition of the four geographic regions used in the

analysis, including provincial-level administrative units and numbers of prefecture-level cities in the 296-city analytical sample

Region	Provincial-level administrative units included	Cities
Northeast	Heilongjiang, Jilin, Liaoning	34
East	Beijing, Tianjin, Shanghai, Hebei, Shandong, Jiangsu, Zhejiang, Fujian, Guangdong, Hainan	88
Central	Shanxi, Henan, Anhui, Hubei, Hunan, Jiangxi	80
West	Chongqing, Inner Mongolia, Guangxi, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang	94

Note: City counts sum to the 296 prefecture-level cities retained after data cleaning.

Supplementary Table 8. Agreement between SHAP-based and permutation-based variable-importance rankings in the main model, averaged over 300 city-block bootstrap replicates

Variable	Mean SHAP	SHAP rank	%IncMS E	Permutatio n rank	Rank difference
EV Stock Share	0.0559	1	44.8	1	0
Gasoline Price	0.0282	2	27.0	2	0
Charging Infrastructure	0.0111	3	24.0	3	0
Stock Average	0.0059	4	20.5	6	-2
Temperature					
Electricity Price	0.0046	5	20.7	5	0
Population Density	0.0030	6	20.1	7	-1
Gini Coefficient	0.0027	7	17.7	8	-1
Public Chargers per EV	0.0024	8	17.1	10	-2
Temperature Variation	0.0017	9	22.2	4	5
Air Quality Index	0.0017	10	17.3	9	1
GDP per Capita	0.0013	11	13.5	12	-1

Urban Area	0.0013	12	15.3	11	1
Metro Availability	0.0001	13	5.9	13	0

Notes: Mean |SHAP| denotes mean absolute SHAP value; %IncMSE denotes the mean percentage increase in mean squared error upon permutation. Rankings refer to substantive predictors. Calendar-year indicators were included in all model fits as trend controls but are excluded from the ranking. The rankings agree closely, with Spearman $\rho = 0.90$ in the main model and $\rho = 0.84$ in the specification excluding EV stock share

Supplementary Table 9. Results of the dynamic (replacement) specification, in which EV stock share is replaced by the one-year-lagged EV penetration rate
Panel A. Permutation importance (%IncMSE) of all predictors in the dynamic specification

Rank	Variable	%IncMSE
1	Lagged EV penetration rate (t-1)	84.3
2	Charging infrastructure stock	27.8
3	Year (trend control)	27.2
4	Gasoline price	23.7
5	Average temperature	20.9
6	Population density	20.1
7	Air quality index	15.7
8	Electricity price	15.6
9	Public chargers per EV	15.5
10	Temperature variation	15.0
11	Urban area	14.5
12	Gini coefficient	13.2
13	GDP per capita	8.8
14	Metro availability	4.4

The dynamic specification is estimated on the exclusion-only model (EV stock share removed) with the one-year-lagged EV penetration rate added in its place, so that the two collinear variables (correlation ≈ 0.88) never enter the model together. The lagged penetration rate is the most important predictor (%IncMSE = 84.3), approximately three times that of the next-ranked predictor (charging infrastructure stock, %IncMSE =

27.8). Year is retained as a trend control. The panel covers 2020–2024 (the first year is lost to the one-year lag).

Panel B. Permutation importance (%IncMSE) of selected predictors across the three specifications.

Variable	Main model	Excl. EV stock share	Dynamic (replaced)
EV stock share	58.5	-	-
Lagged EV penetration rate (t-1)	-	-	84.3
Gasoline price	34.8	41.7	23.7
Charging infrastructure stock	30.6	42.4	27.8
Average temperature	28.8	37.2	20.9
Public chargers per EV	21.6	33.9	15.5
GDP per capita	16.0	22.5	8.8

Main model, all 13 predictors plus year controls; Excl. EV stock share, EV stock share removed; Dynamic (replaced), EV stock share replaced by the lagged penetration rate. Dashes indicate variables not present in a given specification. The importance of charging infrastructure stock, public chargers per EV, and GDP per capita rises when EV stock share is removed and then reverts to or below its main-model level once the lagged penetration rate is included, indicating that this rise partly reflected these variables proxying for prior market development. Values differ slightly from a single fit because of the stochastic nature of random forests.