3. Empirical Context, Data, and Models

3.1 Data

This study uses a comprehensive panel dataset of monthly automobile sales in China from January 2018 to March 2024. The data were obtained from a publicly available industry database on Kaggle, titled *"China Automobile Monthly Sales Data (2018–2024)"*, which consolidates sales records reported by major automakers.

3.1.1 Vehicle Sales Records

The dataset provides monthly sales by brand and model category, covering over 38,000 individual entries. Each observation includes the year, month, vehicle type, brand country, and units sold.

3.1.2 Policy Context and Time Variable

To identify the policy intervention, we defined a binary indicator, post_policy, which takes the value 1 for all mY_{it} months after April 2020, when China renewed and extended its EV subsidy program.

3.1.3 Data Cleaning and Transformation

The variables "units_sold" and "date" were standardized to numeric and datetime formats. Vehicle types were classified into "EV" and "Gasoline" based on the dummy variable is_ev column. Missing or inconsistent entries were handled using numeric coercion and type conversion in Python (see Appendix A for code)

3.2 Analytical Approach: Difference-in-Differences (DiD)

In order to identify the causal impact of the April 2020 Electric Vehicle Subsidy Reform on vehicle demand, this study applies a *Difference-in-Differences (DiD) approach.* The method compares the change in sales of electric vehicles (treatment group) with gasoline vehicles (control group) before and after the policy expansion.

3.2.1 Difference-in-Differences (DiD) Term Definition

To identify the causal impact of the 2020 Electric Vehicle Subsidy Reform, a difference-in-differences (DiD) framework was applied. The analysis treats **electric vehicles (EVs)** as the treatment group and **gasoline vehicles** as the control group, comparing their sales before (January 2018–March 2020) and after April 2020 the policy expansion.

Formally, the model is specified as:

$$Y_{it} = \alpha + \beta_1 EV_i + \beta_2 Post_t + \beta_3 (EV_i \times Post_t) + \epsilon_{it}$$

where:

- Y_{it} = monthly units sold for vehicle type *i* at time *t*,
- EV_i = 1 if electric vehicle, 0 if gasoline,
- $Post_t$ = 1 for months after April 2020, 0 otherwise,
- $EV_i \times Post_t$ = interaction term capturing the treatment effect,
- β_3 = difference-in-differences estimator, representing the policy's casual impact on EV sales.

This model assumes **parallel pre-policy trends**, meaning that in the absence of the reform, EV and gasoline sales would have followed similar trajectories

4. Empirical Analysis and Results

4.1 Policy Expansion

4.1.1 Pre-Policy Demand Dynamics: Parallel Trends Before 2020

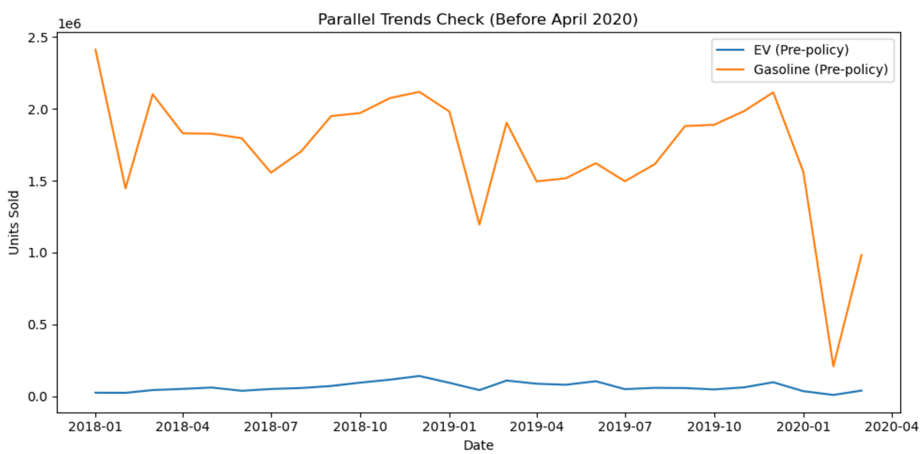

Figure 1. Monthly sales of EV and Gasoline cars in China, 2018–2020/04
(Source: Pre-policy Parallel Trends from "Analysis_china_automobile.csv")

Between 2018 and late 2019, both gasoline and EV sales followed relatively stable paths. Gasoline vehicle sales fluctuated between 1.5 and 2.0 million units per month, while EV sales remained consistently low but steady, indicating broadly parallel trends before government intervention.

The 2020 Electric Vehicle Subsidy Reform

The sharp downturn from January to March 2020 reflects the immediate impact of the COVID-19 outbreak, which disrupted automobile production and demand nationwide. Because this occurred prior to the April 2020 subsidy expansion, it represents a temporary macroeconomic shock rather than a structural market change.

As the economy began to recover, the subsidy policy acted as a catalyst for demand, accelerating EV adoption more quickly than the rebound in gasoline vehicles and laying the groundwork for their subsequent divergence.

4.1.2 Post-policy Demand Shifts

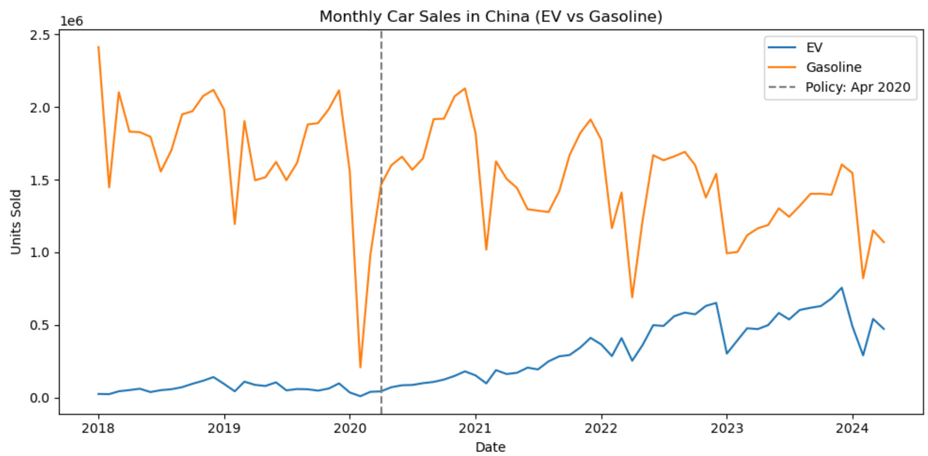

Figure 2. Monthly sales of EV and Gasoline cars in China, 2018–2024
(Source: Full-trends from "Analysis_china_automobile.csv")

Following the April 2020 subsidy expansion, the relationship between EV and gasoline demand shifted sharply. EV sales rose rapidly and remained high, increasing from around 60,000 units per month before the reform to over 600,000 units per month by 2022–2023. In contrast, gasoline vehicle sales declined steadily, falling from roughly 1.7 million to 1.4 million units per month.

This widening gap marks a clear structural change in consumer behavior. Rather than reflecting a general post-COVID recovery, the data suggest a reallocation of demand as households increasingly shifted from gasoline vehicles toward EVs.

4.2 Causal Estimation (DiD): Measuring the Policy Effect

Group	Pre-Policy Mean	Post-Policy Mean	Change (Post-Pre)
Treatment (EV)	64422.0	361139.0	296717.0
Control (Gasoline)	1712053.0	1453311.0	-258742.0

Estimated DiD effect : 555,459 additional EV units/month

Table 1. Policy Effects from DiD Table
(Source: DiD Summary table from "Analysis_china_automobile.csv")

After the 2020 subsidy expansion, EV demand rose sharply while gasoline sales declined. EV sales increased from about **64 000 to 361 000 units per month**, whereas gasoline dropped from **1.7 million to 1.45 million**. The estimated **DiD effect of +555 000 units per month** shows that, even after accounting for market slowdowns, the policy drove a significant and lasting boost in EV adoption.

Appendix

Appendix A: Data Preparation

```
# Import libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# Load dataset
car_df =
pd.read_csv('/Users/smileloukkade/Downloads/Analysis_china_automobile.csv')

# Convert date and numeric columns
car_df['units_sold'] = pd.to_numeric(car_df['units_sold'], errors= 'coerce') #if
it found N/A turns it into NaN
car_df['date'] = pd.to_datetime(car_df['year_date'], errors= 'coerce')

# Create vehicle_type based on is_ev_dummy (1: EV and 0: Gasoline)
car_df['vehicle_type'] = car_df['is_ev_dummy'].apply(lambda x: 'EV' if x == 1
else 'Gasoline')
```

Code A1. Python script for cleaning and standardizing the dataset for analysis.
Description: Dataset cleaned and standardized for analysis. All rows were filtered to include only EV and gasoline vehicles, ensuring comparability for Difference-in-Differences estimation.

Appendix B: Data Aggregation

```
monthly = car_df.groupby(['date', 'vehicle_type'], as_index = False)
['units_sold'].sum()
pivot = monthly.pivot(index = 'date', columns = 'vehicle_type', values =
'units_sold').fillna(0)
```

Code B1. Python script for monthly aggregation and pivot table creation.
Description: The data were aggregated at the monthly level to align EV and gasoline sales on a consistent time scale, facilitating comparison of demand trends between 2018 and 2024.

Appendix C: Policy Variables

```
# Create policy dummy (April 2020)
policy_date = pd.to_datetime("2020-04-01")
car_df['post_policy'] = (car_df['date'] >= policy_date).astype(int)
car_df['is_ev_dummy'] = (car_df['vehicle_type'] == "EV").astype(int)

# Check result
car_df.head(5)
```

Code C1. Python script defining the April 2020 policy dummy for DiD analysis.
Description: A binary indicator for the April 2020 policy expansion distinguishes pre-policy (0) and post-policy (1) periods, providing a clear treatment-control structure for DiD estimation.

Appendix D: Key Findings

1. Parallel Pre-Policy Trends

The following code reproduces **Figure 1** from the main text, showing pre-policy trajectories of EV and gasoline sales before April 2020.

```
# Plot: Pre-policy parallel trends
pre = pivot[pivot.index < policy_date]
plt.figure(figsize=(10,5))
plt.plot(pre.index, pre['EV'], label='EV (Pre-policy)')
plt.plot(pre.index, pre['Gasoline'], label='Gasoline (Pre-policy)')
plt.title('Parallel Trends Check (Before April 2020)')
plt.xlabel('Date')
plt.ylabel('Units Sold')
plt.legend()
```

```
plt.tight_layout()
plt.show()
```

Code D1. Python script for generating pre-policy parallel trends.
Figure 1. Parallel pre-policy movement followed by divergence post-policy; EV sales rise sharply while gasoline plateaus.

2. Pre- vs Post-Policy Demand Comparison

The following code reproduces **Figure 2** from the main text, showing how electric-vehicle (EV) and gasoline sales evolved before and after the April 2020 subsidy expansion

```
# Plot: Full trend
plt.figure(figsize=(10,5))
plt.plot(pivot.index, pivot['EV'], label='EV')
plt.plot(pivot.index, pivot['Gasoline'], label='Gasoline')
plt.axvline(policy_date, linestyle='--', color='gray', label='Policy: Apr 2020')
plt.title('Monthly Car Sales in China (EV vs Gasoline)')
plt.xlabel('Date')
plt.ylabel('Units Sold')
plt.legend()
plt.tight_layout()
plt.show()
```

Code D2. Python script for generating the full-trend comparison of EV and gasoline sales.
Figure 2. Monthly Car Sales in China (2018–2024). This figure illustrates monthly EV and gasoline car sales from 2018 to 2024, with the dashed line marking the April 2020 subsidy expansion. After the policy change, EV sales increased sharply while gasoline sales declined, showing a clear demand reallocation toward EVs.

3. Market Share Trends

3.1 EV Share

```
pivot['EV_share'] = pivot['EV'] / (pivot['EV'] + pivot['Gasoline'])
pivot['EV_share'].plot()
```

Code D3. Python code to compute and plot EV market share relative to total vehicle sales

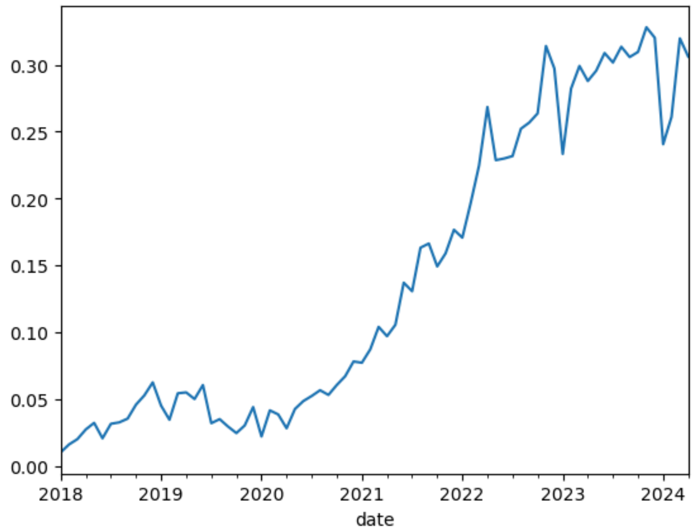

Figure A3. EV market share in China rose from less than 5% in 2018 to about 30% by 2023, reflecting rapid adoption after policy support.25–30% by 2023. This steady increase suggests that government incentives and improved technology drove a lasting shift toward EV adoption.

3.2 EV and Gasoline Market Share

```
# EV vs Gasoline Market share
plt.figure(figsize=(10,5))
plt.plot(pivot.index, pivot['EV_share'], label='EV Share', color='blue')
plt.plot(pivot.index, pivot['Gasoline_share'], label='Gasoline Share',
color='orange')

# Add policy reference line
plt.axvline(pd.to_datetime('2020-04-01'), color='gray', linestyle='--',
label='Policy: Apr 2020')

plt.title("EV vs Gasoline Market Share in China (2018-2024)")
plt.xlabel("Date")
plt.ylabel("Market Share")
plt.legend()
plt.tight_layout()
plt.show()
```

Code D4. Python script comparing EV and gasoline market shares over time.

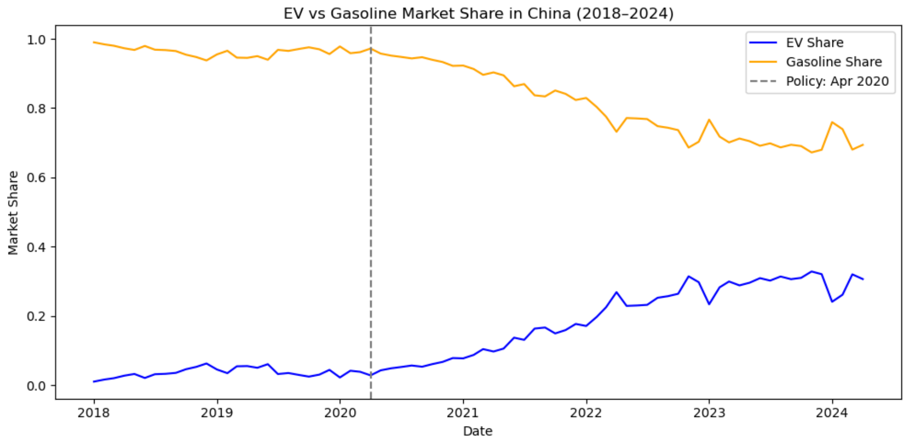

Figure 3. EV and Gasoline Market share chart
(Source: EV vs Gasoline Market share from "Analysis_china_automobile.csv")

Figure 3: **EV and Gasoline Market share chart.** After April 2020, EV share increases steadily while gasoline share declines, reflecting a clear shift in consumer demand toward electric vehicles.

Appendix E: DiD Setup and Results

1. DiD Estimation

```
mask_post = pivot.index >= policy_date
ev_pre = pivot.loc[~mask_post, 'EV'].mean()
ev_post = pivot.loc[mask_post, 'EV'].mean()
gas_pre = pivot.loc[~mask_post, 'Gasoline'].mean()
gas_post = pivot.loc[mask_post, 'Gasoline'].mean()

did_effect = (ev_post - ev_pre) - (gas_post - gas_pre)
print(f"Estimated DiD effect: {did_effect:,.0f} additional EV units/month")
```

```
Estimated DiD effect: 555,459 additional EV units/month
```

Code E1. Python script computing the Difference-in-Differences (DiD) estimate for the 2020 EV subsidy reform.

2. DiD Summary Table

```
# DiD Summary Table
# Show the before-after averages clearly
summary = pd.DataFrame({
```

```
    'Group' : ['EV (Treatment)', 'Gasoline (Control)'],
    'Pre-Policy Mean' : [ev_pre, gas_pre],
    'Post-Policy Mean' : [ev_post, gas_post],
    'Change (Post-Pre)' : [ev_post - ev_pre, gas_post - gas_pre]
})

print(summary.round(0))
print("\nEstimated DiD effect: {:,.0f} additional EV
units/month".format(did_effect))
```

```
                Group  Pre-Policy Mean  Post-Policy Mean  Change (Post-Pre)
0      EV (Treatment)          64422.0          361139.0           296717.0
1  Gasoline (Control)        1712053.0         1453311.0          -258742.0

Estimated DiD effect: 555,459 additional EV units/month
```

Code E2: DiD Summary Table

Table 1. Policy Effects from DiD Table

Description: The estimated difference-in-differences (DiD) effect of ≈ 555 000 additional EV units per month isolates the causal impact of the 2020 policy beyond general market fluctuations. This finding confirms that the reform significantly accelerated EV adoption relative to gasoline demand.

3. Causal Estimation (Difference-in-Differences)

3.1 Manual Ordinary Least Square Coefficient (OLS)

```
# OLS Coefficient (manual)
panel = monthly.copy()
panel['is_ev'] = (panel['vehicle_type'] == "EV").astype(int)
panel['post'] = (panel['date'] >= policy_date).astype(int)
panel['interaction'] = panel['is_ev'] * panel['post']

X = np.column_stack([np.ones(len(panel)), panel['is_ev'], panel['post'],
panel['interaction']])
y = panel['units_sold'].values
beta = np.linalg.lstsq(X, y, rcond=None)[0]

# Define the names for each regression coefficient (to make the output
readable)
```

```
coef_names = ['Intercept', 'is_ev', 'post', 'is_ev×post (DiD)']
print("\nOLS Coefficients:")


for name, b in zip(coef_names, beta):
    print(f"{name:20s}: {b:,.2f}")
```

```
OLS Coefficients:
Intercept           : 1,712,053.37
is_ev               : -1,647,631.78
post                : -258,742.31
is_ev×post (DiD)    : 555,459.23
```

Code E3: Manual OLS Coefficient for DiD estimation

Term	Coefficient	Interpretation
Intercept	1,712,053.37	Baseline gasoline sales pre-policy
is_ev	-1,647,631.78	EVs sold less before policy
post	-258,742.31	Gasoline decline post-policy
Is_ev x post (DiD)	555,459.23	Policy-driven EV increase

Table E1: OLS Coefficient Estimates and Interpretation

Description: Estimated coefficients from the manual OLS estimation. The interaction term *(is_ev × post)* captures the Difference-in-Differences effect, showing that EV sales rose by approximately 555,000 units per month after the 2020 subsidy expansion.

```
# Confirm the effect of DiD
print(panel.groupby(['is_ev','post'])['units_sold'].mean())
```

```
is_ev  post
0      0       1.712053e+06
       1       1.453311e+06
1      0       6.442159e+04
       1       3.611385e+05
Name: units_sold, dtype: float64
```

Code E4: Verification of Group Means for DiD Structure

Description: This code confirms the Difference-in-Differences setup by reporting the mean monthly sales of EV and gasoline vehicles before and after the April 2020 policy. The results show that EV sales increased substantially post-policy, while gasoline sales declined, consistent with the DiD estimation logic.

3.2 Statistical Version: OLS Regression Results

```python
# Formal statistical version: OLS Regression Results
import statsmodels.formula.api as smf
model = smf.ols('units_sold ~ is_ev + post + is_ev:post', data =
panel).fit()
print(model.summary())
```

Code E5: Statistical OLS Regression (Formal DiD Estimation)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             units_sold   R-squared:                       0.852
Model:                            OLS   Adj. R-squared:                  0.849
Method:                 Least Squares   F-statistic:                     283.6
Date:                Mon, 13 Oct 2025   Prob (F-statistic):           3.88e-61
Time:                        03:11:07   Log-Likelihood:                -2119.2
No. Observations:                 152   AIC:                             4246.
Df Residuals:                     148   BIC:                             4258.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   1.712e+06   5.36e+04     31.970      0.000    1.61e+06    1.82e+06
is_ev      -1.648e+06   7.57e+04    -21.756      0.000    -1.8e+06    -1.5e+06
post       -2.587e+05   6.67e+04     -3.880      0.000   -3.91e+05   -1.27e+05
is_ev:post  5.555e+05   9.43e+04      5.889      0.000    3.69e+05    7.42e+05
==============================================================================
Omnibus:                       48.956   Durbin-Watson:                   2.181
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              211.958
Skew:                          -1.090   Prob(JB):                     9.41e-47
Kurtosis:                       8.359   Cond. No.                         8.16
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Table E2. Statistical OLS Regression Results for the 2020 EV Subsidy Reform

Description:

This model formally estimates the Difference-in-Differences (DiD) regression using *statsmodels*. The specification regresses monthly vehicle sales on treatment status (*is_ev*), post-policy period (*post*), and their interaction (*is_ev* × *post*).

The coefficient on the interaction term (≈ **555,000**, $p < 0.001$) represents the average increase in EV sales attributable to the 2020 subsidy expansion, controlling for baseline differences and general market changes.

The high R^2 (0.85) indicates that the model explains most of the variation in sales, while the significance of all coefficients supports the robustness of the policy effect.