Subsidy Phase-out and Consumer Demand Dynamics: Evidence from the Battery Electric Vehicle Market in China

Yunyi Hu, Haitao Yin, and Li Zhao

Readme file on data and code

This zipped file includes: (1) data files, including the original data[1] and datasets created by the code in the estimation files; (2) the code (most in Stata and only one in Python) to combine the datasets, estimate parameters, and generate the tables and figures in the paper and the online appendix; and (3) the results generated from the STATA code. The Stata code runs on Stata 17 version.

1. Data files (in the folder "/Datasets")

Summary data

2016.dta and 2019.dta contain the sales of electric vehicles (EV) at the province-level in 2016 and 2019, respectively, which are used to generate Figure A.3.
avg_subsidy.dta is the average subsidy provided by the central government for different driving ranges during 2016 and 2019, which is used to generate Figure 1.
BEV&PHEV_dataset.dta is the aggregated sales data for all vehicle sales in China during 2016 and 2019, which is used to generate Figure A.2.
Charging station.dta contains the number of charging stations at the province level.
china_label and china_map are data sets used to generate the geographic distribution of electric vehicle sales (Figure A.3).
firm_headquarter.dta contains the location of BEV manufacturers' headquarters.
population.dta: socio-demographic information for the city markets.
overview.xlsx is the aggregate sales of BEVs between 2009 and 2019 from https://www.statista.com[2] (used to generate Figure A.4).
Sales_car is the total sales in each city-quarter market.
SELECTED MARKETS.xlsx contains the BEVs stock in selected countries and regions (2014-2019) from https://www.statista.com (used to generate Figure A.1).
subsidy.dta is the subsidy provided by the central government for each month from 2013 to 2019 (used to generate Figure 1).

Intermediate data

balanced_city.dta and balanced_firm.dta are the balanced panel at the city and firm level, respectively, which are used to create the cumulative sales.
cum_q_city.dta is the cumulative sales at each period for the cities hosting BEV producers.
cum_q.dta is the cumulative sales for the BEV producers.
counterfactual_database is the data set created for the counterfactual simulations.

[1] The sales information on city and models are removed and replaced with the code system for the purpose of confidentiality.
[2] These aggregate sales are slightly different from what we used in the paper.

demand estimation data_iv is the data set for structural estimation, which contains the quarterly sales data for each BEV model in each city from the China Banking and Insurance Regulatory Commission, model characteristics, and constructed instrument variables for the variables with endogeneity.
dynamic_elasticity.dta contains the current price elasticity, quantity elasticity, and future price elasticity calculated by using the parameter estimates from the dynamic structural model.
static_elasticity.dta contains the current price elasticity and quantity elasticity calculated by using the parameter estimates from the static structural model.
iv_excess_city and iv_excess_firm contain the IVs for cumulative sales at the city level and at the firm level, respectively, which are used for learning by doing effect estimation.
iv_excess: instrument variable for the BEV stock variable (used in robustness check).
mc_clean.dta and mc.dta contain the computed marginal cost for each model-quarter observation.
mc_matrix folder contains the intermediate data used for calculating the marginal cost.
other_vehicle.dta contains the market share of non-BEVs and the observations are quarterly city-level sales.
sales_monthly and Sales_quarterly are sales of each BEV model in city-quarter markets.
"/Subsidy" folder contains different subsidy schemes used for counterfactual simulation.

2. Estimation files (in the folder "/Do files")

Four main do-files run several other do-files (which also appear in the folder "/Do files"):

Restats_createdata_main runs several do-files to combine the various datasets and create the data sets for estimation (saved as "/Datasets/demand estimation data_iv.dta").

Restats_Figs&Tabs_main.do runs several do-files to create all figures in the main text and the appendix except the counterfactual results in Figure 6, as well as the descriptive statistics (Table 1) and the calculation of the elasticities under the static and dynamic models (Table 5).

Restats_estimation_main.do runs several do-files to estimate the model. This file creates the results of reduced-form analysis (Table 2 and Figure A.6), structural model estimation (Table 3) along with robustness checks (Table A.4), as well as LBD regression (Table 4). It also creates the marginal cost data for further counterfactual analysis.

Restats_ counterfactual_main.do runs several do-files to do the counterfactual simulation. These files create the dataset for simulation (saved as "/Datasets") and simulate the sales under different scenarios (saved as "/Output"). Figure 6 and Table 6 are generated based on those simulations.

Each of these four do-files provides a brief description of the subsequently run do-files.

3. Results files (in the folder "/Output")

This folder contains the estimation results of reduced-form analysis (RF_evidence.tex), structural model analysis (demand_results.tex) along with a robustness check (robustness.tex), and learning

by doing effect (LBD.tex). Counterfactual sales under different scenarios are also included in this folder.