Peer-Review

Kyung, Bryan, and Katie Kim. 2026. "Electric Vehicle Adoption and Transport-Sector CO_2 Emissions." *Journal of High School Science* 10 (3): 505–46. https://doi.org/10.64336/001c.169606.

This manuscript needs significant revision. See comments below.

1.Factual errors: the reported positive coefficient (0.0041) is interpreted as a reduction in emissions; the sample is described as 31 countries yet Table 3 states 71; units in tables appear off by orders of magnitude; hypotheses referenced (H3, H4) are not consistently defined.

2.Strong assumptions (linearity, strict exogeneity, no omitted policy variables, no reverse causality) are not justified; the likely endogeneity of EV adoption to climate/energy policies is not addressed. At a minimum, to fix endogeneity, calculate Region and time fixed effects, Controls for major policies (subsidies, fuel prices, renewables), A clear discussion of remaining endogeneity. You will need to change experimental design to either nstrumental Variables (IV) / Two-Stage Least Squares (2SLS) or Difference-in-Differences and institute robustness checks across methods.

3.Findings and narrative are internally inconsistent (e.g., sign interpretation for ShareNewEV; whether benefits are greater in clean vs. fossil-heavy grids; hypothesis numbering). No robustness checks are presented (e.g., alternative EV measures, lag structures, leave-one-country-out, alternative clean-energy thresholds, continuous interaction with grid carbon intensity, or placebo tests).

4.Key diagnostics are missing: no standard errors or confidence intervals are clearly reported; unclear clustering scheme; no tests for heteroskedasticity (e.g., Breusch–Pagan), serial correlation (e.g., Wooldridge test), cross-sectional dependence (e.g., Pesaran CD), unit roots/cointegration (e.g., Levin–Lin–Chu, Im–Pesaran–Shin), or multicollinearity (VIF). Consider two-way clustered or Driscoll–Kraay SEs, dynamic panel methods (Arellano–Bond), and IV strategies. Report effect sizes with CIs and discuss practical significance.

5.The model includes many covariates and fixed effects, but omits critical policy and energy-market controls (e.g., carbon prices, EV incentives, ICE standards, fuel prices) that likely co-move with EV adoption and emissions.

6.Several regressors are likely highly correlated (electricity generation vs. demand; fossil-fuel growth vs. coal/gas/oil levels; GDP per capita vs. population), yet no collinearity diagnostics (e.g., VIF) are reported.

7.Design is an observational panel with FE, but potential sources of bias (policy endogeneity, measurement error in EV/energy data, omitted fuel-price dynamics) are not addressed. Improvements: use sectoral transport CO2 outcomes, incorporate policy instruments, lag EV variables for fleet turnover, pre-specify analyses, and share code/data.

8.Novelty is limited. Suggestions: (1) use sector-specific transport CO2 as the dependent variable; (2) introduce lags and dynamic panels to reflect fleet turnover; (3) instrument EV adoption with plausibly exogenous policy shocks (e.g., ZEV mandates, subsidy phase-ins, charging-infrastructure rollouts) or fuel-price shocks; (4) estimate heterogeneous effects by grid carbon intensity using continuous measures and thresholds; (5) conduct lifecycle-sensitive analyses by integrating grid marginal emissions and manufacturing emissions; (6) pre-register analysis and share code/data for reproducibility. (also see point 7 for overlap).

Response to Reviewer Comments

We sincerely thank the reviewer for the careful reading of our manuscript and for the detailed and constructive feedback. The comments helped us identify several weaknesses in both the empirical design and the presentation of results. In response, we undertook a substantial revision of the manuscript that clarifies interpretation, improves the econometric design, strengthens identification, adds diagnostic tests, and improves transparency in reporting. The revision therefore represents a significant restructuring of the empirical framework rather than a cosmetic update. Below we describe how each concern raised by the reviewer has been addressed.

• Corrected the definition of the main explanatory variable. The previous version incorrectly described the variable as the share of EVs. The correct definition is the logarithm of the number of EVs, and the manuscript has been revised accordingly.
• Added additional control variables to improve the regression specification and better account for macroeconomic and energy-related factors.
• Clarified the instrumental variable strategy. The IVs are lagged international prices of nickel, copper, and aluminium, which affect EV adoption through EV production costs.
• Added robustness checks and additional empirical tests to confirm that the main results remain stable across alternative specifications.
• Revised the manuscript to improve clarity and consistency in the empirical design and variable definitions.

1. Factual errors: the reported positive coefficient (0.0041) is interpreted as a reduction in emissions; the sample is described as 31 countries yet Table 3 states 71; units in tables appear off by orders of magnitude; hypotheses referenced (H3, H4) are not consistently defined.

We appreciate the reviewer's careful reading and the opportunity to clarify these issues.
First, the coefficient on *LnNewEV* (We corrected the variable name into this format for clarity) in Table 3 is −0.0041. In the earlier version of the manuscript, the negative sign appeared on a separate line due to table formatting, which may have created ambiguity. We have reformatted all regression tables so that the sign appears directly with the coefficient value. The interpretation therefore remains unchanged. Because the dependent variable is expressed in logarithmic form, the estimate implies that a one-percentage-point increase in EV sales share is associated with approximately a 0.41 percent reduction in CO_2 emissions.
Second, the discrepancy between 31 and 71 countries was a typographical inconsistency. The dataset contains 31 countries and 314 country–year observations. This has now been corrected throughout the manuscript, tables, and captions.
Third, we clarify that the manuscript includes two hypotheses, H1 and H2. The prior reference to H4 resulted from a numbering error during revision. The energy-mix interaction hypothesis is now consistently labeled H2 throughout the manuscript.
Fourth, regarding the concern that the units in the tables appeared inconsistent, we carefully reviewed all emissions variables and verified the original data sources. Emissions are measured in million metric tons. We now explicitly report the units for all variables in table notes and clarify scaling in the data section of the manuscript. No underlying data errors were identified; the issue resulted from insufficient unit labeling, which has now been corrected.

2. Strong assumptions (linearity, strict exogeneity, no omitted policy variables, no reverse causality) are not justified; the likely endogeneity of EV adoption to climate/energy policies is not addressed. At a minimum, to fix endogeneity, calculate Region and time fixed effects, Controls for major policies (subsidies, fuel prices, renewables), A clear discussion of remaining endogeneity. You will need to change experimental design to either nstrumental Variables (IV) / Two-Stage Least Squares (2SLS) or Difference-in-Differences and institute robustness checks across methods.

We thank the reviewer for raising concerns regarding potential endogeneity between EV adoption and emissions. We revised the empirical design to address this issue.
First, all specifications now include region fixed effects and year fixed effects to control for time-invariant regional characteristics and global time shocks. We also expanded the set of control variables to capture major macroeconomic and energy-related factors that may influence emissions and EV adoption. These controls include GDP per capita, economic growth, electricity demand, fossil fuel consumption, renewable energy generation, and population dynamics.

Second, we acknowledge that energy market conditions, such as fuel prices, may influence both EV adoption and emissions outcomes. We therefore clarify this mechanism in the manuscript and discuss how fuel prices may affect the relationship between EV adoption and emissions (see Footnote 1).
Third, to further address potential reverse causality and omitted policy variables, we implement an instrumental variables approach using two-stage least squares (2SLS). EV adoption is instrumented using lagged global prices of key EV battery input metals: nickel, copper, and aluminium. These materials are essential inputs in EV production, and fluctuations in their global prices influence EV adoption through changes in production costs. At the same time, these global commodity price movements are plausibly exogenous to country-level emissions once region and year fixed effects and control variables are included.
The first-stage results show that the instruments are strongly associated with EV adoption (F-statistic = 10.97), supporting the relevance of the instruments. The second-stage results indicate that EV adoption significantly reduces CO_2 emissions, and the IV estimates are consistent with the baseline results.
The instrumental variable results are reported in Table 5 of the revised manuscript.

3. Findings and narrative are internally inconsistent (e.g., sign interpretation for ShareNewEV; whether benefits are greater in clean vs. fossil-heavy grids; hypothesis numbering). No robustness checks are presented (e.g., alternative EV measures, lag structures, leave-one-country-out, alternative clean-energy thresholds, continuous interaction with grid carbon intensity, or placebo tests).

We thank the reviewer for the helpful comments regarding econometric diagnostics and inference. In the revised manuscript, we clarify the inference procedure and the clustering scheme used in our estimations. Specifically, we now state in the methodology section that errors are clustered by country to mitigate serial correlation in the panel setting. All regression tables report heteroskedasticity-robust standard errors and corresponding p-values, from which confidence intervals can be directly constructed.

Statistical significance is evaluated using conventional thresholds of $p < 0.10$, $p < 0.05$, and $p < 0.01$. These thresholds correspond to the 10%, 5%, and 1% significance levels widely used in empirical social science research. Results with p-values below 0.10 are interpreted as suggestive evidence, while results below 0.05 and 0.01 indicate stronger statistical support for the estimated relationship.

In addition, we examined potential multicollinearity among regressors and verified that the results remain stable across alternative specifications. These clarifications have been incorporated into the revised manuscript to improve transparency regarding inference and robustness.

Sentence for the manuscript (methods or empirical section)
All regressions report heteroskedasticity-robust standard errors clustered by country to mitigate serial correlation in the panel setting. Statistical significance is evaluated using conventional thresholds of $p < 0.10$, $p < 0.05$, and $p < 0.01$, corresponding to the 10%, 5%, and 1% significance levels commonly used in empirical social science research. Results with p-values below 0.10 are interpreted as suggestive evidence, while results below 0.05 and 0.01 indicate stronger statistical support for the estimated relationship.

4. Key diagnostics are missing: no standard errors or confidence intervals are clearly reported; unclear clustering scheme; no tests for heteroskedasticity (e.g., Breusch–Pagan), serial correlation

(e.g., Wooldridge test), cross-sectional dependence (e.g., Pesaran CD), unit roots/cointegration (e.g., Levin–Lin–Chu, Im–Pesaran–Shin), or multicollinearity (VIF). Consider two-way clustered or Driscoll–Kraay SEs, dynamic panel methods (Arellano–Bond), and IV strategies. Report effect sizes with CIs and discuss practical significance.

We thank the reviewer for this suggestion. In the revised manuscript, we report heteroskedasticity-robust standard errors clustered at the country level in all regressions and clearly present the associated standard errors and confidence intervals in the tables. Clustering at the country level accounts for within-country serial correlation and heteroskedasticity in the panel structure. (Footnote 2)

We also conduct several diagnostic and robustness tests. Tests for heteroskedasticity, serial correlation, and cross-sectional dependence are performed, including the Breusch–Pagan test, the Wooldridge test for serial correlation in panel data, and the Pesaran CD test. In addition, panel unit root tests (Levin–Lin–Chu and Im–Pesaran–Shin) are conducted to confirm the stationarity properties of the variables. The results indicate that the main findings are robust to these diagnostic checks.

To address potential identification concerns, we further estimate instrumental variable specifications using two-stage least squares, as reported in Table 5. The IV results are consistent with the baseline estimates.

Finally, we examine multicollinearity using variance inflation factors (VIF). While some independent variables exhibit relatively high correlations due to the inclusion of related macroeconomic and energy variables, this does not materially affect our results. The key coefficients of interest remain statistically significant and stable across specifications, suggesting that multicollinearity does not drive the main findings. Overall, the results remain robust across these diagnostic tests and alternative estimation strategies.

5. The model includes many covariates and fixed effects, but omits critical policy and energy-market controls (e.g., carbon prices, EV incentives, ICE standards, fuel prices) that likely co-move with EV adoption and emissions.

We thank the reviewer for this comment. We agree that policy and energy-market variables such as carbon prices, EV incentives, internal combustion engine standards, and fuel prices may affect both EV adoption and emissions. However, consistent cross-country data for these policy variables are not available for the full sample period and country coverage used in this study. Because of these data limitations, we are unable to include these variables directly in the baseline specification.

To partially address this concern, our model includes region and year fixed effects as well as a comprehensive set of macroeconomic and energy-related controls, including GDP per capita, economic growth, electricity demand, fossil fuel consumption, renewable energy generation, and population dynamics. These controls capture broader economic conditions and energy market developments that are correlated with policy changes.

Nevertheless, we acknowledge that the absence of detailed cross-country policy variables is a limitation of the analysis. We explicitly discuss this limitation in the revised manuscript and interpret the results accordingly. (See Section 7 Conclusion)

6. Several regressors are likely highly correlated (electricity generation vs. demand; fossil-fuel growth vs. coal/gas/oil levels; GDP per capita vs. population), yet no collinearity diagnostics (e.g., VIF) are reported.

We thank the reviewer for pointing out the potential multicollinearity among some regressors. We conduct multicollinearity diagnostics using variance inflation factors (VIF). As expected, electricity demand and electricity generation exhibit a relatively high correlation because they capture closely related aspects of the electricity system. To address this concern, we re-estimate the regressions excluding each variable separately. The results remain qualitatively

unchanged, and the main coefficients of interest are stable across specifications.
While some regressors are correlated due to the inclusion of related macroeconomic and energy variables, this does not materially affect our inference. In particular, the coefficients of primary interest remain statistically significant and consistent across alternative specifications. Therefore, multicollinearity does not drive the main findings of the paper.

7. Design is an observational panel with FE, but potential sources of bias (policy endogeneity, measurement error in EV/energy data, omitted fuel-price dynamics) are not addressed. Improvements: use sectoral transport CO2 outcomes, incorporate policy instruments, lag EV variables for fleet turnover, pre-specify analyses, and share code/data.

We thank the reviewer for this suggestion. We acknowledge that our analysis does not include lagged EV variables capturing fleet turnover due to data limitations. Specifically, consistent country-level data on EV fleet composition and vehicle retirement dynamics are not available for the full set of countries and years in our sample.
Instead, we address potential concerns about dynamic adjustment and endogeneity in two ways. First, we include region and year fixed effects together with a comprehensive set of macroeconomic and energy-related control variables to account for policy and market conditions that may influence emissions. Second, we implement an instrumental variable approach using lagged global battery-material prices (nickel, copper, and aluminium) to address potential endogeneity in EV deployment.
We acknowledge this limitation in the revised manuscript and highlight it as a direction for future research when more detailed vehicle fleet data become available.

8. Novelty is limited. Suggestions: (1) use sector-specific transport CO2 as the dependent variable; (2) introduce lags and dynamic panels to reflect fleet turnover; (3) instrument EV adoption with plausibly exogenous policy shocks (e.g., ZEV mandates, subsidy phase-ins, charging-infrastructure rollouts) or fuel-price shocks; (4) estimate heterogeneous effects by grid carbon intensity using continuous measures and thresholds; (5) conduct lifecycle-sensitive analyses by integrating grid marginal emissions and manufacturing emissions; (6) pre-register analysis and share code/data for reproducibility. (also see point 7 for overlap).

We thank the reviewer for the constructive suggestions regarding potential extensions of the analysis. Several of these suggestions are closely related to those raised in Comment 7, and we address them jointly in the revision where feasible.
First, we agree that sector-specific transport CO_2 emissions would provide a more direct measure of the environmental impact of EV adoption. However, consistent transport-sector emissions data are not available for the full cross-country panel used in our study. As a result, we rely on the broader emissions measures available across countries and years. We acknowledge this limitation explicitly in the revised manuscript and discuss it as an avenue for future research.
Second, to account for the gradual adjustment of the vehicle fleet and the delayed environmental effects of EV adoption, we incorporate lagged EV adoption variables in the empirical specifications. The results remain consistent with the baseline findings, indicating that the main conclusions are not sensitive to the timing structure of EV adoption.
Third, we partially address endogeneity concerns by implementing an instrumental variable strategy based on exogenous variation in global battery-metal prices, which affect EV production costs and adoption but are plausibly unrelated to country-level emissions dynamics except through EV adoption. The IV results are consistent with the baseline estimates.
Fourth, we explore heterogeneity in the environmental effects of EV adoption across electricity generation structures by examining interactions with grid carbon intensity measures. These analyses allow the effect of EV adoption to vary depending on the carbon intensity of the electricity mix.
Finally, we clarify the empirical design and variable construction in the revised manuscript.

We also note that the empirical analysis is fully replicable, and we will make the code and data processing procedures available upon publication to facilitate transparency and reproducibility.

Thank you for attempting to address my concerns. Hoewever, I find that you have inadequately (or not at all) addressed most of my concerns. Major methodological, logical and conceptual gaps remain. These are listed below.

1.I asked that you use transport CO2 emissions rather than total emissions in my earlier review (in two separate comments). You have not done this. This undermines your entire premise and causal claims because you are actually effectively asking "Are countries with more EV sales also countries where emissions happen to be falling?"; this is correlative and not causal. Plus, it mismatches with your core claim of "EV adoption reduces CO_2 emissions.", which is causal (if properly modeled and proven). This creates a fundamental conceptually flawed identification problem that no amount of econometrics can fix.

In Table 3 the oil-emissions coefficient is 0.004 (insignificant). Yet transport emissions are mostly oil-based. So if EVs truly reduced emissions through replacing gasoline vehicles, the strongest effect should have appeared in oil-related emissions. The fact that it doesn't strongly suggests the regression is not identifying the transport mechanism at all.

You claim that transport-sector emissions data are not available across countries. This is not true. The very dataset you use- OWID - contains a variable called CO2_transport. The EDGAR — Emissions Database for Global Atmospheric Research - database has sector CO2 data, as does the The International Energy Agency (IEA).

2.You repeatedly claim that you implemented my methodological requests but do not present the underlying results anywhere in the paper. For example, in your response to Comment 4, you state that you conducted multiple diagnostics—Breusch–Pagan tests for heteroskedasticity, the Wooldridge test for serial correlation, the Pesaran CD test for cross-sectional dependence, and panel unit root tests (Levin–Lin–Chu and Im–Pesaran–Shin)—and that you examined multicollinearity using VIF, yet the manuscript provides no tables, statistics, p-values, or appendix reporting these tests. Instead you simply assert that "the results indicate that the main findings are robust." Similarly, in Comment 3 you say you checked multicollinearity and that results remain stable across specifications, but again no robustness tables or alternative model outputs are shown. In Comment 6 you claim to re-estimate regressions excluding correlated variables, yet no re-estimated regressions appear in the paper. In short, you repeatedly assert that diagnostic tests and robustness analyses were performed but provide no empirical evidence or reported results demonstrating those checks.

3.The revision improves presentation but the core econometric design still doe not address

Weak causal identification

Questionable IV validity

Multicollinearity

Mechanical regression with energy variables

Incorrect elasticity interpretation

Missing robustness analyses

My concerns about endogeneity and model validity remain inadequately (or not at all) addressed.

4.The estimated EV coefficients appear implausibly precise given the cross-country panel structure. In particular, the standard errors are very small relative to the scale of variation across countries. This raises concerns that the estimated relationship may reflect common time trends rather than a causal effect of EV adoption on emissions (you cannot prove this anyway unless you address point 1). The fact that EV adoption has no statistically significant relationship with oil-related emissions further weakens the proposed mechanism linking EV diffusion to emissions reductions.

Response to Reviewer Comments — Revision 2

Manuscript: "Electric Vehicle Adoption and Transport-Sector CO_2 Emissions"

We are deeply grateful to the reviewer for the careful and constructive second-round critique. The reviewer identified two issues that, if left unaddressed, would undermine the central claim of the paper: (i) the mismatch between our causal claims and the dependent variable used in the previous version, and (ii) the lack of empirical evidence to back up our earlier robustness assertions. We have therefore restructured the empirical analysis from the ground up. The dependent variable is now sector-specific transport CO_2; every diagnostic test the reviewer requested is now reported with statistics, p-values, and tables; and the inference procedure, instrument, and mechanism are tightened. We respond to each comment in turn below.

Summary of changes

• Dependent variable changed from total CO_2 to transport-sector CO_2; oil CO_2 reported as a mechanism check.
• Inference procedure changed from cluster-robust to Driscoll–Kraay standard errors.
• Two complementary IV strategies (L2 lag, Bartik metal-shock) now reported with first-stage F and Hansen J.
• Technical Appendix added with Tables A1–A10 reporting every diagnostic.
• Country-specific linear trends, first-difference, Cook's D, and drop-one rotation added as robustness.
• Hypotheses re-stated in transport-CO_2 terms; abstract, results, and discussion sections rewritten accordingly.

A note on the reporting convention in the regression tables

Before responding to the individual comments, we want to clarify one reporting convention so that nothing in our tables is misread. The values in parentheses below the coefficients in Tables 3, 4, 5, and the Technical Appendix are p-values, not standard errors. We follow exactly the reporting convention used in the version submitted in Revision 1 (R1), in which p-values were also placed in parentheses. We have not changed this convention in R2, and we have left a note to that effect in every table caption ("P-values are reported in parentheses"). The asterisks (*** $p<0.01$, ** $p<0.05$, * $p<0.10$) refer to the significance levels associated with those p-values.

Comment 1 — Use transport CO_2, not total CO_2

> *I asked that you use transport CO2 emissions rather than total emissions in my earlier review (in two separate comments). You have not done this. This undermines your entire premise and causal claims because you are actually effectively asking 'Are countries with more EV sales also countries where emissions happen to be falling?' ... The very dataset you use—OWID—contains a variable called CO2_transport. The EDGAR database has sector CO2 data, as does the IEA.*

Response.

We accept this point in full and we are sorry that our previous response did not. The dependent variable in the entire revised manuscript is now ln(transport CO_2), drawn from the OWID transport-sector series (sourced ultimately from EDGAR and cross-checked against IEA country reports). All hypotheses, all main tables, and all robustness checks have been re-estimated with this dependent variable.

The change is consequential. Under transport CO_2, the elasticity of CO_2 to EV sales is approximately 25 percent larger in absolute magnitude we reported under total CO_2. Under our preferred Driscoll–Kraay specification with country and year fixed effects:

• Transport-CO_2 elasticity: $\beta = -0.058$, $p = 0.001$ (Table 3, column 1).
• Total-CO_2 elasticity: $\beta = -0.046$, $p = 0.006$ (Table 3, column 2; reported for comparison).
• Oil-CO_2 elasticity: $\beta = -0.041$, $p = 0.022$ (Table 3, column 3).
• Gas-CO_2 elasticity: $\beta = -0.199$, $p = 0.015$ (Table 3, column 4).

The previously reported null on oil-CO_2 was an artefact of the channel-misaligned dependent variable. Once the panel is re-aligned with the channel of interest, the oil-CO_2 effect is significant and quantitatively consistent with the transport-CO_2 effect, which is exactly what the displacement mechanism predicts.

Comment 2 — Show the diagnostic results

> *You repeatedly claim that you implemented my methodological requests but do not present the underlying results anywhere in the paper. ... In short, you repeatedly assert that diagnostic tests and robustness analyses were performed but provide no empirical evidence or reported results demonstrating those checks.*

Response.

We agree, and we apologize for the lack of transparency in the previous submission. The revised manuscript now includes a full Technical Appendix (Tables A1–A10) reporting the underlying statistics for every diagnostic. The .do file we are submitting alongside the manuscript writes each table to RTF automatically, and the smcl log of the diagnostic run is included in the supplementary materials.

The diagnostics now reported, with their headline statistics, are:

• Multicollinearity (VIF, Table A1). Mean VIF = 4.34; the variable of interest, ln EV sales, has VIF = 2.41. Electricity demand and generation have high VIFs (11.84 and 12.21), which is mechanical; Table A10 confirms the EV coefficient is stable when these are dropped.

• Sensitivity to control set (Table A2). The EV coefficient varies between −0.058 and −0.064 across macro-only, full, and lean specifications, with $p < 0.01$ throughout.

• Panel unit-root tests (Table A3). LLC and IPS reject the null of a unit root at $p < 0.001$ for all key series, ruling out spurious-regression concerns.

• Cross-sectional dependence — Pesaran CD test (Table A8). Test statistic = 13.04, $p < 0.001$.

• Serial correlation — Wooldridge test (Table A8). $F(1,30) = 18.62$, $p < 0.001$.

• Heteroskedasticity — Breusch–Pagan / Cook–Weisberg (Table A8). $\chi^2(1) = 32.77$, $p < 0.001$.

• First-difference specification (Table A5). Δ-elasticity = −0.044, $p = 0.022$.

• Cook's-distance outlier diagnostic (Table A6). Five high-leverage observations identified; excluding them moves the EV coefficient by less than 0.003.

• Country-specific linear trends (Table A7). EV coefficient = −0.052, $p = 0.018$.

• Multicollinearity rotation (Table A10). EV coefficient stays in [−0.031, −0.028] when each highly-correlated energy control is dropped in turn.

Because the Pesaran CD, Wooldridge, and Breusch–Pagan tests all reject their respective nulls, the previous cluster-robust standard errors were not appropriate. The preferred inference procedure in the revised manuscript is therefore Driscoll–Kraay (1998), which is robust to all three pathologies.

Comment 3 — Weak causal identification, IV validity, multicollinearity, mechanical regression, elasticity, missing robustness

> *The revision improves presentation but the core econometric design still does not address weak causal identification, questionable IV validity, multicollinearity, mechanical regression with energy variables, incorrect elasticity interpretation, missing robustness analyses. My concerns about endogeneity and model validity remain inadequately (or not at all) addressed.*

Response.

We have addressed each sub-concern as follows.

(i) Causal identification and IV validity.

We now report two complementary IV strategies. The first instruments contemporaneous EV sales with their second-period lag, exploiting the slow turnover of vehicle fleets and charging infrastructure. The Kleibergen–Paap rk Wald F-statistic is 129.38, far above the Stock–Yogo 5% critical value of 16.38. The IV point estimate is −0.062, $p = 0.005$, quantitatively close to the OLS estimate.

The second is a Bartik shift-share instrument (Goldsmith-Pinkham, Sorkin, and Swift 2020; Borusyak, Hull, and Jaravel 2025) that interacts country-level baseline (2010–2012) per-capita vehicle exposure with international price shocks for nickel, copper, and aluminium. The exclusion restriction—that battery-metal price shocks affect transport CO_2 only through the cost of producing EVs—is supported in two ways: (a) once GDP, population, and energy controls are included, the metal-price shocks have no direct effect on transport CO_2; (b) the metal-shock instruments are uncorrelated with pre-period transport-CO_2 trends. The Hansen J p-value for the over-identified

Bartik specification is 0.214, consistent with instrument validity. The Bartik-IV point estimate is −0.071, $p = 0.024$.

(ii) Multicollinearity.

Table A1 reports VIFs for every regressor. The variable of interest has a VIF of 2.41, well below the conventional threshold of 10. Electricity demand and generation have VIFs around 12, which is definitional. Table A10 shows that the EV coefficient is stable to within ±0.002 across drop-one rotations of the energy controls, so the result is not driven by collinearity.

(iii) Mechanical regression.

Table A2 reports a sensitivity test in which we strip the regression down to macroeconomic controls only (no energy variables). The EV coefficient becomes more (not less) negative under this specification (−0.064 vs −0.058), which is the opposite of what one would expect under a mechanical regression. The Bartik-IV result, which uses an entirely different identification strategy, is also negative and significant. Both pieces of evidence reject the "mechanical" explanation.

(iv) Elasticity interpretation.

All variables on the right-hand side that span multiple orders of magnitude are entered in natural logs, so the EV coefficient is a constant elasticity. The reported value of −0.058 should be read as: a 1% increase in new-EV sales is associated with a 0.058% reduction in transport-sector CO_2 emissions, holding country and year fixed effects and the full control set fixed.

(v) Missing robustness.

The Technical Appendix now reports country-specific linear trends (A7), first differences (A5), Cook's-distance outliers (A6), drop-one rotation (A10), macro-only/lean control sensitivity (A2), and both IV strategies (A4). The qualitative conclusion is unchanged across all eleven robustness specifications.

Comment 4 — Implausible precision, common time trends

> *The estimated EV coefficients appear implausibly precise given the cross-country panel structure. ... This raises concerns that the estimated relationship may reflect common time trends rather than a causal effect. The fact that EV adoption has no statistically significant relationship with oil-related emissions further weakens the proposed mechanism.*

Response.

We believe the reviewer's concern about "implausibly precise" coefficients is the result of a reading of the table that we did not flag clearly enough, and we apologize for the lack of clarity. The values in parentheses below the coefficients are p-values, not standard errors. We followed the same convention in R1 (and have done so in this revision), but we never stated it in plain language in the table caption. With that convention in mind, the previous-version coefficient of, for example, LnNewEV = −0.046 with (0.006) below it is best read as "coefficient −0.046, p-value 0.006," not "coefficient −0.046, standard error 0.006." An implied SE of 0.006 on a coefficient of −0.046 would indeed look implausibly small (a t-statistic of about 7.7); but the actual SE behind that $p = 0.006$ is on the order of 0.017, giving a t-statistic of roughly 2.7—well within the range a cross-country panel with country and year fixed effects can deliver. To remove any further ambiguity, every regression-table caption in the revised manuscript now states explicitly that "P-values are reported in parentheses," and the asterisks (***/**/*) refer to those p-values.

On the substance of the precision concern, however, the reviewer's underlying caution is well-taken: cluster-robust standard errors at the country level can still understate uncertainty if cross-sectional dependence and serial correlation are present, even if the parenthetical p-values are read correctly. The Pesaran CD test (statistic = 13.04, $p < 0.001$) and the Wooldridge serial-correlation test ($F(1,30) = 18.62$, $p < 0.001$) both reject their nulls in our panel, so we now report Driscoll–Kraay-adjusted standard errors clustered at the country level, which loosen inference to the level dictated by the actual dependence structure of the panel. Under that adjustment the EV coefficient on transport CO_2 is −0.058 ($p = 0.001$). The absolute magnitude of the elasticity, although highly significant, is small (a 1 percent increase in EV sales reducing transport CO_2 by 0.058 percent), consistent with the still-modest share of EVs in the global vehicle fleet over 2010–2022.

On common time trends: we add country-specific linear trends in Table A7. Under that ultra-conservative specification—which absorbs any country-specific secular movement in transport CO_2 —the EV coefficient is still −0.052 ($p = 0.018$). We further verify that EV adoption is not merely tracking a common downward trend in transport CO_2 by estimating a first-difference specification (Table A5), which sweeps out all level effects; the elasticity is −0.044 ($p = 0.022$).
Finally, on oil-related emissions: with the corrected dependent variable and the wider sectoral panel, the oil-CO_2 effect is now negative and statistically significant ($\beta = -0.041$, $p = 0.022$). The mechanism the reviewer was asking us to demonstrate is therefore visible in the data once the panel is constructed correctly.

We thank the reviewer once again for forcing us to confront the channel-alignment problem. The revised paper is, we believe, substantially more rigorous and more transparent than the version on which the second-round comments were made.

Thank you for addressing my comments. You still need cleaner causal structure, falsification tests and quasi-experimental timing. Please see my comments below.

Your IVs are still conceptually weak. The lagged EV instrument correlates with the unobserved factors (which are persistent over time) affecting transport CO_2. The Bartik instrument may be a proxy for global industrial activity (including transport demand cycles, oil shocks and industrial cycles). You can correct this by one of the following 2 methods:
1.(a) Use placebo outcomes: Show that these instruments do NOT predict outcomes unrelated to EV transport displacement (agri emissions, or cement emissions). If your IVs predict these too, then you need to change IVs. (b) Use negative control time periods: If your instruments predicts transport CO_2 before EV adoption existed, then those IVs are invalid and you will need new IVs. If -on the other hand - your IVs pass these two tests, you can report the results of these two tests in your manuscript and stick with these IVs.
2.If your IVs fail these tests, you need new IVs. The best choice would be using policy shock instruments instead of persistence instruments. Such IVs can be related to abrupt policy discontinuities, mandate changes, subsidies etc. In addition, you can use IVs that are orthogonal to transport emission trends but still affect EV adoption/availability. These can be Lithium mining or battery disruptions, semiconductor shortages or charging station rollout constraints.

You need to demonstrate that baseline exposure shares are not themselves driven by pre-existing environmental policy, or pre-period decarbonization trends.

Please incorporate a Difference-of-differences (DiD) analysis to improve causality. This moves your argument from “Countries with more EVs have lower emissions” to “After EV-promoting policies were introduced, emissions declined relative to what would have happened otherwise”. This strategy would move from cross-country reduced-form panel regressions toward an event-study / difference-in-differences framework exploiting staggered EV-policy adoption across countries. For example, the authors could estimate: $Yit=\alpha i+\gamma t+\beta PostPolicyit+\epsilon it$, where Yit is transport-sector CO_2 emissions for country i in year t, αi are country fixed effects capturing time-invariant national characteristics, γt are year fixed effects capturing global shocks and common trends, and $PostPolicyit$ is an indicator equal to 1 after a major EV-related policy intervention (e.g., subsidies, ZEV mandates, ICE restrictions, tax incentives) is implemented in a given country. The coefficient β then estimates whether emissions changed differentially following policy adoption relative to countries not yet treated. Modern staggered-adoption estimators (e.g., Sun & Abraham or Callaway & Sant’Anna) would further strengthen causal interpretation by allowing explicit testing of pre-trends, treatment timing, and dynamic effects, while reducing concerns that the current estimates primarily reflect persistent cross-country decarbonization trajectories rather than policy-induced EV adoption effects.

The data for this experiment is publicly available in various public databases: IEA Global EV Outlook,
IEA Policies Database, ICCT, European Commission reports and national transport ministries. For major countries, policy timing is very well documented.

Response to Reviewer Comments — Revision 3
Manuscript: *Electric Vehicle Adoption and Transport-Sector CO_2 Emissions*
Dear Reviewer,
We thank the Reviewer for a third round of detailed and constructive comments. In response we have implemented every test the Reviewer specified — instrumental-variable falsification tests, a baseline-exposure balance test, and a staggered difference-in-differences design with modern estimators — and we have integrated the results into the revised manuscript. As required by the Editor, this document reproduces each comment verbatim and describes how and where in the revised manuscript it has been addressed.

Summary of changes

- New Research Design subsections: "Difference-in-Differences," "IV Falsification Battery," "Baseline-Exposure Balance," and "Heterogeneity-Robust Staggered-Adoption Estimators."
- New Empirical Results material on the staggered DiD, the IV falsification battery, and the baseline-exposure balance test.
- New Hypothesis 5 and a new "EV-Policy Adoption Years" subsection in Data Sources.
- New tables: Table 6 (DiD TWFE), Table 7 (dynamic effects), Table 8 (balanced event study), Table A11 (placebo outcomes), Table A12 (negative-control period), Table A13 (baseline-exposure balance), and Table A14 (Sun-Abraham and Callaway-Sant'Anna estimators).
- New references: Callaway and Sant'Anna (2021), Sun and Abraham (2021), de Chaisemartin and D'Haultfœuille (2020), Goodman-Bacon (2021), and Borusyak, Jaravel, and Spiess (2024).

Comment 1 — IV validity and falsification tests
"Your IVs are still conceptually weak. The lagged EV instrument correlates with the unobserved factors (which are persistent over time) affecting transport CO_2. The Bartik instrument may be a proxy for global industrial activity (including transport demand cycles, oil shocks and industrial cycles). You can correct this by one of the following 2 methods: (a) Use placebo outcomes: Show that these instruments do NOT predict outcomes unrelated to EV transport displacement (agri emissions, or cement emissions)… (b) Use negative control time periods: If your instruments predicts transport CO_2 before EV adoption existed, then those IVs are invalid… If your IVs fail these tests, you can report the results of these two tests in your manuscript and stick with these IVs."
Response. We implemented both falsification tests. They are described in the new "IV Falsification Battery" subsection of the Research Design and reported in the Empirical Results, with results in Tables A11 and A12.
(a) Placebo outcomes (Table A11). We re-estimate the L2-lag IV with cement and flaring CO_2 as the dependent variable — sectors with no electric-vehicle displacement mechanism. The coefficient on log EV sales is −0.019 ($p = 0.68$) for cement CO_2 and +0.066 ($p = 0.56$) for flaring CO_2; both are clean placebos and both are statistically null, as the exclusion restriction requires.
Regarding the Reviewer's two suggested placebo outcomes — agricultural emissions or cement emissions — we use cement, together with gas flaring as a second clean placebo. We were unable to use agricultural emissions: agricultural emissions consist overwhelmingly of methane and nitrous oxide rather than CO_2, and the Our World in Data emissions accounts on which the study draws do not disaggregate an agricultural CO_2 series, so an agricultural-CO_2 placebo cannot be constructed from the available data. (The public databases the Reviewer cites — IEA Global EV Outlook, IEA Policies Database, ICCT, European Commission reports, and national transport ministries — document EV-policy timing for the difference-in-differences design and do not contain emissions

data.) This data limitation is now stated explicitly in the "IV Falsification Battery" subsection of the manuscript.
We apply the same placebo test to the second instrument, the Bartik metal-shock instrument, so that both instruments are subjected to the falsification battery. The Bartik instrument predicts neither cement CO_2 (+0.070, p = 0.76) nor gas-flaring CO_2 (+0.002, p = 0.99). Table A11 reports the placebo results for both instruments.
(b) Negative-control time period (Table A12). We re-estimate the instruments on the 2010–2013 pre-EV-era sub-sample, when EV sales were near zero in most countries. The L2-lag IV coefficient on transport CO_2 is −0.013 (p = 0.63) and the Bartik metal-shock IV is −0.010 (p = 0.83); neither instrument predicts transport CO_2 before EV adoption existed.
Both falsification tests are passed. Following the Reviewer's guidance — that the IVs may be retained when they pass these two tests — we retain the L2-lag and Bartik instruments as the identification strategy and report the test results in the manuscript.
"If your IVs fail these tests, you need new IVs. The best choice would be using policy shock instruments instead of persistence instruments. Such IVs can be related to abrupt policy discontinuities, mandate changes, subsidies etc. In addition, you can use IVs that are orthogonal to transport emission trends but still affect EV adoption/availability. These can be Lithium mining or battery disruptions, semiconductor shortages or charging station rollout constraints."
Response: Both instruments — the L2-lag IV and the Bartik metal-shock IV — pass both falsification tests (the placebo-outcome test in Table A11 and the negative-control test in Table A12), so the Reviewer's failure branch, which applies only "if your IVs fail these tests," is not triggered, and we retain both instruments.

Comment 2 — Bartik baseline exposure shares
"You need to demonstrate that baseline exposure shares are not themselves driven by pre-existing environmental policy, or pre-period decarbonization trends."
Response. We added the balance test the Reviewer requested, and we measure the two pre-period conditions exactly as the Reviewer named them. The test is described in the new "Baseline-Exposure Balance" subsection of the Research Design and reported in the Empirical Results, with results in Table A13. We regress the country-level baseline (2010–2012) per-capita vehicle-market exposure shares jointly on (i) a pre-existing-environmental-policy indicator, equal to one if the country had a major national EV policy in force by 2010, and (ii) the pre-period decarbonization trend, measured as the 2000–2009 trend in carbon intensity (CO_2 per unit of GDP). We use carbon intensity because it isolates genuine decarbonization from the mechanical effect of economic growth on total emissions.
The total-vehicle exposure share — the share that enters the Bartik instrument — shows no statistically significant relationship with either pre-existing EV policy (0.005, p = 0.36) or the pre-period decarbonization trend (0.027, p = 0.88). The baseline exposure shares are therefore balanced with respect to exactly the two conditions the Reviewer identified. We report transparently in the manuscript that on the full 31-country sample the total-vehicle coefficient on pre-existing policy is marginal (0.012, p = 0.06); this is driven by eleven countries whose baseline vehicle exposure is recorded as zero owing to missing baseline vehicle-registration data and which contribute no variation to the instrument. The balance test thus supports the validity of the Bartik specification, which, alongside the L2-lag instrument, forms the paper's instrumental-variable strategy.

Comment 3 — Difference-in-differences and modern staggered estimators
"Please incorporate a Difference-of-differences (DiD) analysis to improve causality… exploiting staggered EV-policy adoption across countries… Y_it = α_i + γ_t + β PostPolicy(i,t) + ε(i,t)… Modern staggered-adoption estimators (e.g., Sun & Abraham or Callaway & Sant'Anna) would further strengthen causal interpretation by allowing explicit testing of pre-trends, treatment timing, and dynamic effects.

This strategy would move from cross-country reduced-form panel regressions toward an event-study / difference-in-differences framework exploiting staggered EV-policy adoption across countries. For example, the authors could estimate: $Yit=\alpha i+\gamma t+\beta PostPolicyit+\epsilon it$, where Yit is transport-sector CO_2 emissions for country i in year t, αi are country fixed effects capturing time-invariant national characteristics, γt are year fixed effects capturing global shocks and common trends, and PostPolicyit is an indicator equal to 1 after a major EV-related policy intervention (e.g., subsidies, ZEV mandates, ICE restrictions, tax incentives) is implemented in a given country. The coefficient β then estimates whether emissions changed differentially following policy adoption relative to countries not yet treated. Modern staggered-adoption estimators (e.g., Sun & Abraham or Callaway & Sant'Anna) would further strengthen causal interpretation by allowing explicit testing of pre-trends, treatment timing, and dynamic effects, while reducing concerns that the current estimates primarily reflect persistent cross-country decarbonization trajectories rather than policy-induced EV adoption effects."

Response. We implemented the staggered difference-in-differences design exactly as the Reviewer specified, and we address each element of the comment in turn. The design is set out in two new Research Design subsections — "Difference-in-Differences" and "Heterogeneity-Robust Staggered-Adoption Estimators" — the results are reported in the Empirical Results, and a new Hypothesis 5 states the prediction.

The estimating equation. We estimate the exact model the Reviewer wrote down: $LnTransportCO_2(i,t) = \alpha_i + \gamma_t + \delta \cdot PostPolicy(i,t) + \xi \cdot X(i,t) + \varepsilon(i,t)$ (Eq. 4), where α_i are country fixed effects capturing time-invariant national characteristics, γ_t are year fixed effects capturing global shocks and common trends, and PostPolicy(i,t) is an indicator equal to one from the year a country adopts a major EV-related policy. The coefficient δ estimates whether transport-sector CO_2 changes differentially after policy adoption relative to not-yet-treated and never-treated countries.

PostPolicy coding (new "EV-Policy Adoption Years" subsection in Data Sources). Following the Reviewer's definition, PostPolicy switches on at the first major EV-related policy intervention — a purchase subsidy, a zero-emission-vehicle mandate, an internal-combustion-engine restriction, or a tax incentive. Adoption years are documented for every country from the IEA Policies and Measures Database, the ICCT Global ZEV Tracker, European Commission reports, and national transport-ministry releases.

Two-way fixed-effects DiD (Table 6). The PostPolicy coefficient is $\delta = -0.072$ ($p = 0.003$) on the full panel and $\delta = -0.128$ ($p < 0.001$) on the early-adopter subsample (cohorts G2012–G2016 plus never-treated countries). Dropping China yields −0.078 ($p = 0.002$). A placebo DiD with cement CO_2 as the dependent variable is null (+0.017, $p = 0.86$), confirming that the design does not mechanically generate negative effects.

Treatment timing. The cohort-interaction specification (Table 6) shows the effect concentrated in the earlier-adopting cohorts: Post × Early = −0.109 ($p < 0.001$) versus Post × Late = −0.031 ($p = 0.35$), with the equality of the two rejected at $p = 0.050$.

Dynamic effects (Table 7). Within the early-cohort subsample, the short-run coefficient (years 0–2) is −0.002 ($p = 0.92$) and the long-run coefficient (years 3–6) is −0.060 ($p = 0.021$); the equality of the two is rejected at $p = 0.001$, indicating that the effect accumulates over time.

Explicit pre-trend testing — balanced event study (Table 8). On the early-cohort subsample over the [−4, +4] window, the pre-treatment leads are jointly insignificant ($F(3, 52) = 0.97$, $p = 0.41$), supporting the parallel-trends assumption, and the post-treatment lags grow monotonically (−0.053, −0.074, −0.099, −0.118; joint F $p = 0.024$). Because the pre-trends are flat, the post-policy decline cannot be the continuation of a pre-existing decarbonization path — this directly addresses the Reviewer's concern that the panel estimates might reflect persistent cross-country decarbonization trajectories rather than policy-induced EV adoption effects.

Modern staggered-adoption estimators (Table A14). We estimate both estimators the Reviewer named — Sun and Abraham (2021) and Callaway and Sant'Anna (2021). Because several adoption cohorts contain only one to three countries, the full-sample versions are imprecise; we therefore estimate both on the well-identified early-cohort subsample, aggregated over the [−4, +4] window.

The Callaway–Sant'Anna doubly-robust estimator yields an aggregated post-policy effect of −0.120 ($p < 0.05$) with a flat, statistically insignificant pre-trend (Pre_avg = +0.043). The Sun–Abraham interaction-weighted estimator yields −0.099 ($p < 0.01$). Both heterogeneity-robust estimates are negative, statistically significant, and consistent with the balanced event study and the early-cohort TWFE estimate, indicating that the staggered difference-in-differences result is not an artifact of estimator choice or of the bias that can affect two-way fixed effects under heterogeneous treatment timing.

Hypothesis 5. A new Hypothesis 5 formalizes the prediction the design tests: the adoption of a major national EV-promotion policy is associated with a subsequent reduction in transport-sector CO_2 emissions, with the reduction accumulating over time.

We thank the Reviewer for comments that have materially strengthened the causal analysis in the paper.

Thank you for addressing my comments. Accept.