Attrition-Aware Causal Inference with GPS Tracking Data: A Study of Germany's 9-Euro-Ticket

Anupriya[1], Claudio Mena Cavieres[1], Liang Ma[1], Heike Link[2], Daniel J. Graham[1], and Daniel Hörcher*[1,3]

[1]Transport Strategy Centre, Imperial College London, London SW7 2AZ
[2]Deutsches Institut für Wirtschaftsforschung, Mohrenstr. 58, Berlin 10117
[3]Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Budapest 1093

*Corresponding author. Email address: d.horcher@imperial.ac.uk

Short summary

Germany's 9-Euro-Ticket, implemented from June to August 2022, attracted exceptional international attention because of both the magnitude of the fare discount and the scale of its implementation. It also sparked intense debate over whether the policy mainly generated additional public transport demand or shifted travel away from private cars. This paper estimates the causal impact of the policy on travel behaviour using a longitudinal smartphone-based GPS panel covering May to August 2022. GPS data offer rich, high-frequency, user-level information on trips, mode use, and distance travelled, making them especially valuable for policy evaluation. At the same time, they pose important challenges for causal inference: some users stop contributing data, reporting may be intermittent, and certain outcomes may be observed selectively in ways that are related to travel behaviour itself. If left unaddressed, these features of the data-generating and measurement process can bias estimated policy effects and obscure their dynamics. To address this problem, we develop an estimation framework that treats panel dropout and selective outcome observation as explicit features of the measurement system. The approach constructs *stabilised inverse-probability weights* on a balanced user-time panel and combines them with a *sharp regression discontinuity in time* design. Applied to the 9-Euro-Ticket, the results show that conclusions depend strongly on how travel demand is measured. Public transport mode share rises substantially, but the frequency of car use does not fall in any trip-purpose or distance-band category. Although public transport distance increases while car kilometres decline, a temporal decomposition shows that these changes do not occur at the same time, providing little evidence of direct mode shift. More broadly, the paper contributes a practical weighting and diagnostics pipeline that can be embedded in other causal designs using smartphone travel data, helping support more credible policy evaluation.
Keywords: Pricing; travel passes; travel behaviour; GPS tracking; causal inference.

1 Introduction

Decarbonising transport is an urgent policy challenge. Road transport remains a major and persistent source of emissions because vehicle fleets turn over slowly, total vehicle-kilometres continue to grow, and efficiency gains have not been sufficient to offset rising travel demand (Jaramillo et al., 2022; IEA, 2025). This has renewed interest in policies that can be deployed quickly to shift everyday trips from private cars towards lower-carbon modes such as public transport, cycling, and walking.

Within this policy toolbox, public transport fares are particularly salient. They are visible to travellers, administratively tractable, and can be adjusted at short notice, allowing governments to send a system-wide price signal. Yet their effects on sustainable travel are empirically ambiguous. Lower fares may shift some trips from road to public transport, but they may also induce additional travel or redistribute demand across times and places, with effects that depend on service quality, frequency, and trip length (Vickerman, 2024; Hörcher & Tirachini, 2021; Paulley et al., 2006; Balcombe et al., 2004). This makes large, temporary fare cuts especially important to evaluate.

We study Germany's *9-Euro-Ticket*, implemented from June to August 2022, which offered unlimited nationwide local and regional public transport for €9 per month. Introduced as part of

a broader cost-of-living relief package during a period of elevated energy prices, the ticket sharply reduced the marginal price of public transport nationwide while service supply remained largely fixed in the short run (Ember, 2024; Adolfsen et al., 2022). It therefore created a rare system-wide fare shock. This setting is policy-relevant because it speaks directly to whether deep but time-limited fare reductions can move travel away from private cars, how quickly such responses emerge, and whether they persist over the intervention window.

Despite intense public and academic attention, rigorous evidence on these behavioural responses remains limited. Existing studies document substantial ridership gains during the three-month period, but provide mixed evidence on car-use displacement and emissions, often relying on descriptive before–after comparisons (see, for instance, Loder et al., 2024; Gohl & Schrauth, 2024; Andor et al., 2023). Such comparisons risk conflating policy effects with concurrent time trends and other shocks unrelated to the intervention (Graham, 2025). At the same time, the policy generated few conventional evaluation data. The rollout was announced at short notice, many trips did not appear in administrative transaction records because the ticket operated through a subscription format, and rapid-response data collection was limited. In this context, smartphone-based revealed-preference panels provide a uniquely valuable source of behavioural evidence (Loder et al., 2024; Link et al., 2023).

However, using such panels for causal inference raises a distinct problem: the set of observed users and the set of observed outcomes may both vary over time in ways related to travel behaviour (Loder et al., 2024; Molloy et al., 2023). We distinguish two selection processes. *Attrition* determines who remains in the panel and continues to provide any usable signal. *Outcome availability* determines which outcomes are observed at a given time, conditional on remaining in the panel. This second process reflects both interruptions to data generation, such as the app or phone being switched off, and the difficulty of distinguishing non-observation from true behavioural inactivity. If these processes change when the policy begins or as responses evolve, analyses based only on observed outcomes may mistake changes in data availability for changes in behaviour.

To address this problem, we develop an attrition- and availability-aware estimation framework that makes these processes explicit. We construct stabilised inverse-probability weights (Wooldridge, 2002) using a panel structure that follows each user over time, including periods in which no outcome is observed. We first estimate *stay-in* weights to account for selective retention, and then *availability* weights to account for selective observation of outcomes among those retained. These components are combined into a single weight that can be embedded in causal estimators. We also report diagnostics on overlap, weight concentration, effective sample size, and covariate balance before and after weighting.

Our empirical analysis uses a longitudinal smartphone panel covering one pre-policy month and the full policy period, with 890,480 trips from 2,436 users across more than 400 districts (Link et al., 2023). The common nationwide start date allows us to implement a sharp regression discontinuity in time (RDiT) at June 1. We integrate the stabilised weights into the RDiT design to recover treatment dynamics net of changing retention and observability. The outcomes include weekly mode-specific trip counts and distances travelled, allowing us to trace both changes in travel intensity and possible changes in modal composition over the three-month intervention.

The paper contributes in two ways. Substantively, it provides causal evidence on the travel-behaviour effects of one of the most prominent fare experiments in recent European transport policy, informing debates on affordability, demand management, and decarbonisation. Methodologically, it offers a portable weighting-and-diagnostics workflow for credible causal inference with app-based mobility panels that can be combined with RDiT and other causal designs.

2 Methods

This section briefly describes the data, the inverse-probability weighting strategy used to address selective observation, the regression discontinuity design, and the outcome models used to estimate policy effects.

Data and outcomes

We use the GPS–survey panel documented in Link et al. (2023), which combines continuous smartphone-based geolocation tracking with linked survey information. Participants were recruited in Germany and agreed to ongoing smartphone tracking; trip processing was carried out by *Intervista*. The raw data contain 890,480 trips made by 2,436 users between 1 May and 31 August 2022,

covering one month before the 9-Euro-Ticket and the full policy period. Each trip record includes timestamps, origin and destination coordinates, NUTS-3 districts, an inferred main mode, trip purpose, distance, duration, and cost. Linked survey waves provide socio-demographic characteristics such as age, gender, employment status, income, and household structure. We also attach weekly district-level weather conditions, fuel prices, school holidays, and an urban–rural typology. Trips are assigned to origin districts and aggregated to ISO weeks. To avoid partial weeks, we exclude 1 May 2022 (week 17) and 29–31 August 2022 (the start of week 35), leaving an analysis window covering ISO weeks 18–34. We distinguish three distance bands: short (≤ 10 km), medium (10–30 km), and long (> 30 km), and we analyse work and leisure trips separately. Modes are harmonised into three groups: private car, bicycle, and public transport, where the latter combines local public transport and train.

We examine two families of outcomes. First, we measure mode-specific weekly trip totals at the district–week–distance-band–purpose level. These outcomes capture changes in the volume of mobility activity. Second, we measure total weekly kilometres travelled by mode at the district–week–purpose level. These outcomes provide a distance-based proxy for potential emissions impacts by showing whether car kilometres fall as lower-carbon modes increase.

Inverse-probability weighting

A central challenge in app-based mobility panels is that the observed sample may change over time in ways related to travel behaviour. We distinguish two selection processes. The first is *attrition*: some users stop contributing data altogether. The second is *outcome availability*: even for users who remain in the panel, outcomes for a given week or user-week subcell may be unobserved because no trip is recorded. If these processes are related to observed characteristics and evolve around the policy change, naïve estimates may confound behavioural responses with changes in who is observed.

To address this problem, we construct *stabilised inverse-probability weights* (Wooldridge, 2002; Robins et al., 2000; Seaman & White, 2013). The weighting procedure starts from complete user-by-week panels that retain all users in all weeks, including weeks with no observed trips. For attrition, we model the probability that a user remains in the panel over time as a function of observed user characteristics, lagged travel behaviour, district-level conditions, and week effects. This yields a cumulative *stay-in* weight that up-weights users who are more likely to disappear from the panel.

We then model outcome availability on complete user–week skeletons defined at the relevant level of analysis. For trip-count outcomes, the availability model is estimated on user–week–distance-band–purpose cells. For kilometre outcomes, it is estimated on user–week–purpose cells. In each case, the probability that an outcome is observed is modelled as a function of observed user characteristics, lagged activity, district-level conditions, and week effects. This yields a second weight that corrects for intermittent observation among those who remain in the panel.

The final analysis weight is the product of the attrition and availability components. Following standard practice, the weights are stabilised using parsimonious numerator models, trimmed to reduce the influence of extreme observations, and normalised within relevant strata. We assess their behaviour using standard diagnostics, including overlap, weight concentration, effective sample size, and balance before and after weighting. These diagnostics are intended to show whether the weighted sample remains informative and whether the weighting succeeds in reducing systematic differences associated with retention and observability.

Causal design

The 9-Euro-Ticket was implemented nationwide on 1 June 2022, leaving no untreated geographic control group. We therefore use a *sharp regression discontinuity in time* (RDiT) design, which exploits the common policy start date as a discontinuity in treatment assignment. Weeks 18–21 form the pre-policy period, and weeks 22–34 form the treated period.

The identifying idea is that, absent the policy, outcomes would have evolved smoothly through the cutoff date. Under this continuity assumption, any discrete change in the conditional expectation of the outcome at the start of week 22 can be attributed to the policy. This design is well suited to a nationwide intervention because it does not require cross-sectional untreated controls, but instead identifies the treatment effect from the break in the outcome trajectory at the policy start date.

Outcome models

We first aggregate weighted user-level observations to district-week cells. For trip outcomes, weighted user-level trip counts are summed to district–week–distance-band–purpose cells. For kilometre outcomes, weighted user-level distances are summed to district–week–purpose cells. Outcome models are then estimated on these aggregated panels.
Our baseline specification is a weighted fixed-effects RDiT model estimated on the log scale:

$$\log(Y_{rtq} + \varepsilon) = \alpha_q \text{Post}_t + \eta_q c_t + \xi_q(\text{Post}_t \times c_t) + \beta' X_{rt} + \psi_r + u_{rtq},$$

where Y_{rtq} denotes the relevant aggregated outcome, Post_t indicates weeks at or after the policy cut-off, c_t is the running time centred at week 22, X_{rt} contains time-varying district controls, and ψ_r are district fixed effects. The coefficient α_q captures the discontinuous policy effect at the cut-off, while the interaction term allows the post-policy slope to differ from the pre-policy trend. Standard errors are clustered by district.
To study treatment dynamics, we also replace the single post-policy indicator with three post-policy period indicators corresponding to early, middle, and late phases of the intervention. This allows the discontinuity to vary over the three months of the ticket. For interpretation, we convert log-scale coefficients into percentage effects and level changes using the weighted pre-policy mean of the corresponding outcome.
In sum, this approach combines a nationwide time-discontinuity design with explicit correction for selective retention and selective observability in smartphone panel data, allowing us to estimate policy effects net of changes in who is observed and when.

3 Results and Discusssion

This section summarises the estimated effects of the 9-Euro-Ticket on weekly trips and kilometres travelled. We focus on the aggregate effects first, as these provide the clearest account of the policy's overall behavioural impact, and then briefly discuss the temporal decomposition. All estimates are based on the weighted sharp RDiT design described above, so the reported discontinuities are interpreted as causal effects at the policy cut-off under the continuity assumption.

Aggregate treatment effects

The aggregate results reveal a clear asymmetry between public transport uptake and car-use displacement. As shown in Figure 1, the 9-Euro-Ticket generates a substantial increase in public transport use, but this increase is not matched by a corresponding fall in the frequency of car travel. This is the central empirical finding of the paper.
For weekly trip totals, the policy leads to a strong and robust rise in public transport trips. This pattern is consistent with the interpretation that the fare reduction successfully stimulated use of local and regional public transport during the three-month intervention. By contrast, the estimated effects on car-trip frequency are small and statistically indistinguishable from zero in the aggregate specifications. Bicycle trips also show no comparably large or systematic response. Taken together, the trip-based results suggest that the dominant behavioural margin affected by the ticket is additional public transport use rather than a broad substitution away from private car travel.
This distinction is important for policy interpretation. Higher public transport ridership does not automatically imply decarbonisation if the increase mainly reflects induced travel or reallocation from other non-car modes. On the trip margin, the evidence points to limited direct displacement of car use: people used public transport more often, but not at the expense of making fewer car trips.
A comparison with the non-IPW estimates (figures omitted for brevity) indicates why correcting for selective retention and observability matters. In the unweighted models, the qualitative pattern is similar in broad terms, but the magnitudes and precision of some effects differ, suggesting that changes in panel composition and outcome availability are not innocuous for inference. The weighted specification is therefore preferred because it better isolates behavioural change from changes in who remains observed in the app data over time. At the same time, the fact that the main substantive message is preserved across weighted and unweighted specifications strengthens confidence that the increase in public transport use is not an artefact of the weighting procedure.

The kilometre-based outcomes provide a more nuanced picture. Figure 1 also shows that public transport kilometres increase while car kilometres decline. This pattern is more consistent with substitution than the trip-count results, suggesting that some behavioural adjustment may occur on the intensive margin, that is, in how far people travel by each mode rather than how frequently they use them. From an environmental perspective, this matters because total distance travelled is more directly linked to emissions than trip frequency alone.
At the same time, the kilometre results do not by themselves establish a one-for-one mode shift from car to public transport. Read jointly, the two outcome families indicate that the policy changed travel behaviour, but not in a simple way. Public transport use clearly rose, yet the absence of a corresponding fall in car-trip frequency implies that much of this increase reflects induced demand or more complex behavioural responses rather than straightforward substitution. Accordingly, headline ridership gains alone would overstate the extent to which the policy reduced reliance on the car.

Temporal dynamics

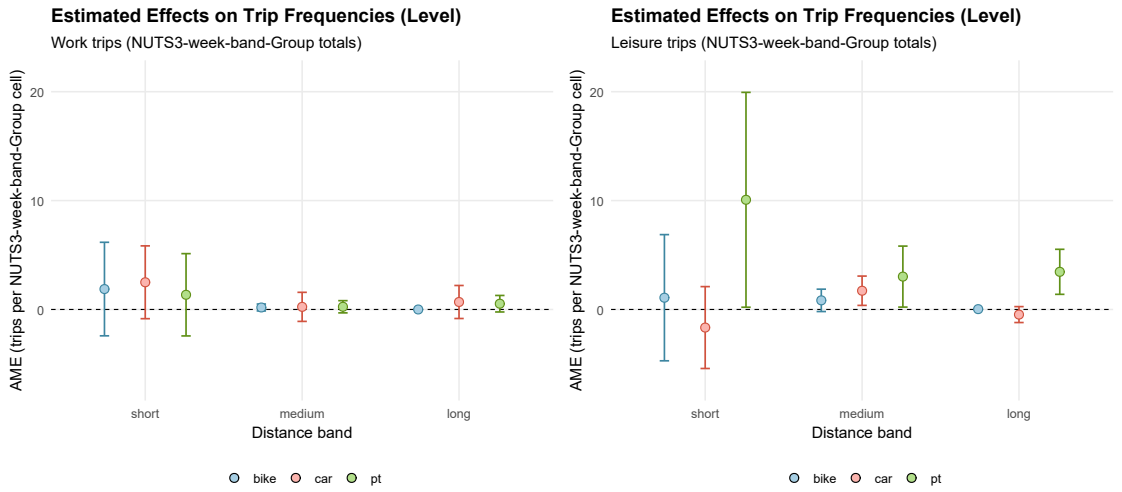
Figure 2 summarises the dynamic specifications by dividing the intervention into early, middle, and late phases. The increase in public transport trips appears quickly after the introduction of the ticket and remains visible throughout the policy period, although its magnitude varies across phases. This pattern is consistent with an immediate response to the large fare reduction, followed by some adjustment over time.
By contrast, car-trip frequency remains close to zero in all post-policy periods. There is no stage of the intervention in which car-trip counts show a clear and systematic decline. This reinforces the core message of the aggregate analysis: the 9-Euro-Ticket substantially increased public transport use, but did not reduce how often people used cars.
The kilometre-based dynamics are somewhat more suggestive of substitution, but the timing is not tightly aligned across modes. Public transport kilometres rise during the intervention, and car kilometres fall in some post-policy periods, yet these changes do not occur in a strongly synchronised way. This temporal mismatch weakens the interpretation of an immediate and direct shift from car to public transport. If the main behavioural effect had been a simple contemporaneous transfer of trips from one mode to the other, the increase in public transport and the decrease in car travel would be expected to move together more closely over time.
The dynamic results are therefore best interpreted as evidence of a more complex adjustment process. The policy clearly increased public transport demand, but the temporal structure of the responses suggests heterogeneous behavioural channels, including induced travel and only limited displacement of private car use.

4 Conclusions

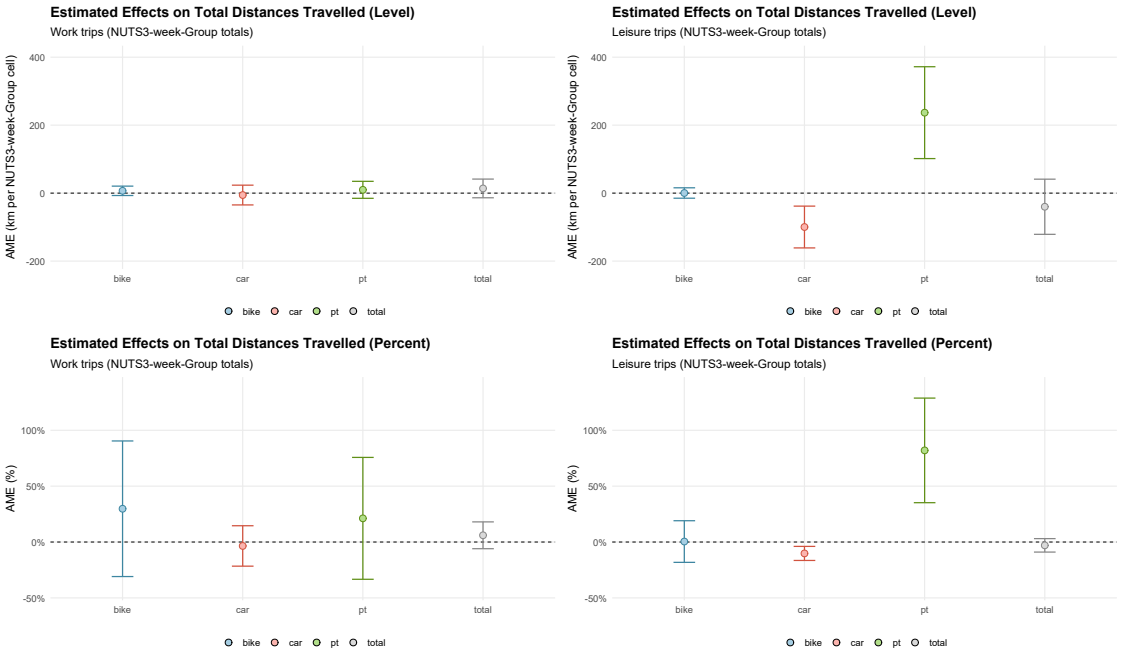
Overall, the results support three conclusions. First, the 9-Euro-Ticket caused a substantial increase in public transport use. Second, this increase did not translate into a corresponding reduction in the frequency of car travel. Third, kilometre-based measures provide some evidence of lower car use, but the timing of these effects does not align closely with the increase in public transport demand. The policy therefore appears to have been effective in boosting public transport uptake, but less effective in producing an immediate and unambiguous shift away from the car.
These findings matter both substantively and methodologically. Substantively, they suggest that large temporary fare cuts can stimulate public transport demand, but may not on their own deliver strong decarbonisation gains through car-use displacement. Methodologically, they demonstrate the value of evaluating such policies using multiple behavioural outcomes and with explicit correction for selective retention and observability in app-based mobility panels.

Acknowledgements

Daniel Hörcher acknowledges funding from the National Research, Development and Innovation Office of Hungary (NKFIH), under Excellence24 grant #151498.

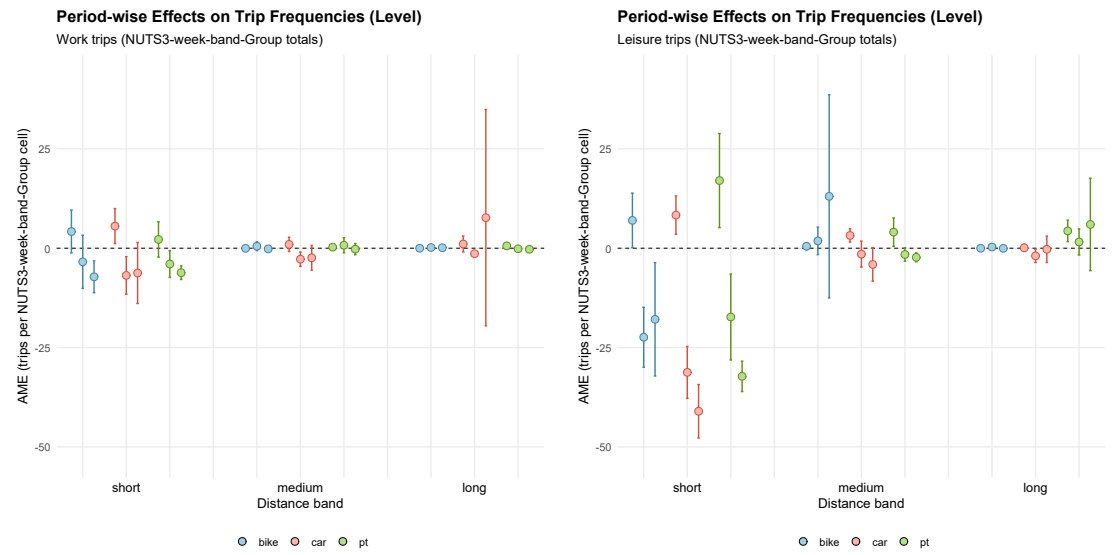


(a) Weekly trips by mode

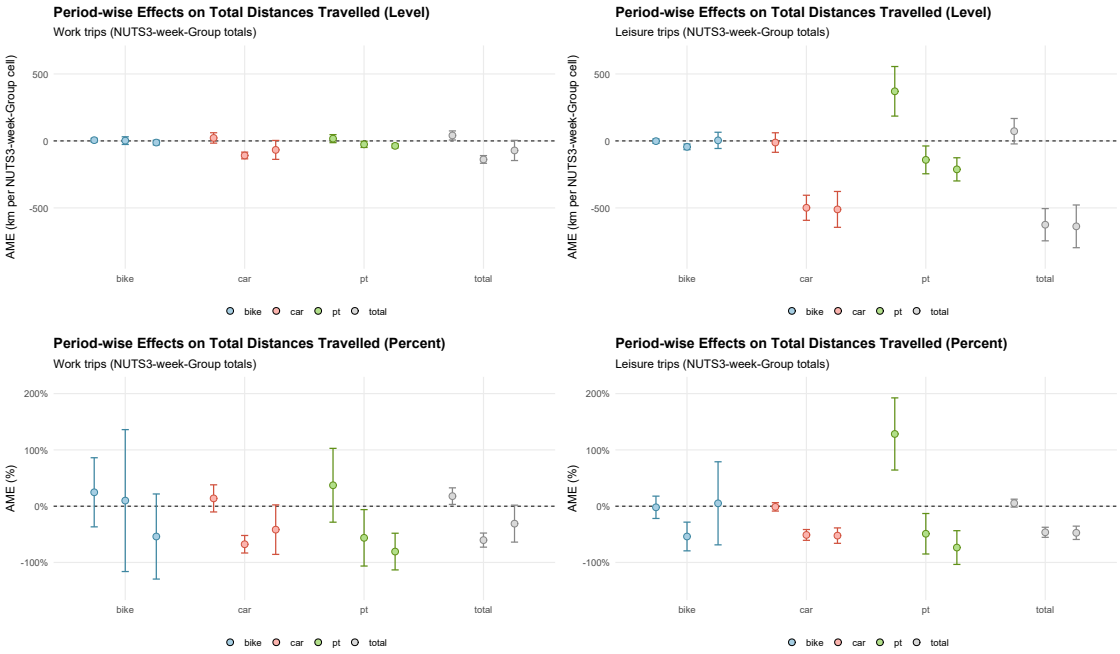


(b) Weekly kilometres by mode

Figure 1: Aggregate estimated effects of the 9-Euro-Ticket on weekly trips and kilometres travelled, by mode. Points show average marginal effects and whiskers indicate 95% confidence intervals.



(a) Weekly trips by mode



(b) Weekly kilometres by mode

Figure 2: Dynamic estimated effects of the 9-Euro-Ticket across three post-policy periods. Panels report average marginal effects by mode, with whiskers indicating 95% confidence intervals.

References

Adolfsen, J. F., Kuik, F., Lis, E. M., & Schuler, T. (2022). *The impact of the war in Ukraine on Euro Area Energy Markets.* Retrieved from `https://tinyurl.com/bdeznpnn`

Andor, M. A., Dehos, F., Gillingham, K., Hansteen, S., & Tomberg, L. (2023). *Public transport pricing: An evaluation of the 9-euro ticket and an alternative policy proposal.* Ruhr Economic Papers.

Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., ... White, P. (2004). The demand for public transport: a practical guide.

Ember. (2024). *Average monthly electricity wholesale price in Germany from January 2019 to June 2024 (in euros per megawatt-hour).* Retrieved from `https://tinyurl.com/yczmeb3e`

Gohl, N., & Schrauth, P. (2024). Jue insight: Ticket to paradise? the effect of a public transport subsidy on air quality. *Journal of Urban Economics*, *142*, 103643.

Graham, D. J. (2025). Causal inference for transport research. *Transportation Research Part A: Policy and Practice*, *192*, 104324.

Hörcher, D., & Tirachini, A. (2021). A review of public transport economics. *Economics of transportation*, *25*, 100196.

IEA. (2025). *Global EV Outlook 2025, Licence: CC BY 4.0.* International Energy Agency Paris, France. Retrieved from `https://www.iea.org/reports/global-ev-outlook-2025`

Jaramillo, P., Kahn Ribeiro, S., Newman, P., Dhar, S., Diemuodeke, O. E., Kajino, T., ... others (2022). Climate change 2022 - mitigation of climate change: Working group iii contribution to the sixth assessment report of the intergovernmental panel on climate change. In (chap. Transport (Chapter 10)). Cambridge University Press.

Link, H., Gaus, D., Murray, N., Guajardo Ortega, M. F., Gervois, F., von Waldow, F., & Eigner, S. (2023). Combining GPS tracking and surveys for a mode choice model: Processing data from a quasi-natural experiment in Germany.

Loder, A., Cantner, F., Adenaw, L., Nachtigall, N., Ziegler, D., Gotzler, F., ... others (2024). Observing Germany's nationwide public transport fare policy experiment "9-Euro-Ticket"–Empirical findings from a panel study. *Case Studies on Transport Policy*, *15*, 101148.

Molloy, J., Castro, A., Götschi, T., Schoeman, B., Tchervenkov, C., Tomic, U., ... Axhausen, K. W. (2023). The mobis dataset: a large gps dataset of mobility behaviour in switzerland. *Transportation*, *50*(5), 1983–2007.

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., ... White, P. (2006). The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transport policy*, *13*(4), 295–306.

Robins, J. M., Hernan, M. A., & Brumback, B. (2000). *Marginal structural models and causal inference in epidemiology* (Vol. 11) (No. 5). Lww.

Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical methods in medical research*, *22*(3), 278–295.

Vickerman, R. (2024). The transport problem: The need for consistent policies on pricing and investment. *Transport policy*, *149*, 49–58.

Wooldridge, J. M. (2002). Inverse probability weighted m-estimators for sample selection, attrition, and stratification. *Portuguese economic journal*, *1*(2), 117–139.