Causal Machine Learning Framework for Market Dynamics Analysis: Methodological Advances Beyond Pattern Recognition

Omaymah Almashaleh[1], Bhavana Bhat[2] and Omid Fatahi Valilai[1*] (ORCID: 0000-0001-7087-6946)
[1]School of Business, Social & Decision Sciences, Constructor University Bremen, Bremen, Germany
[2]Mercedes-Benz AG, Am Wallgraben 114, 70565 Stuttgart, Germany
Emails: {*OAlmashale,OFatahiValilai*}@Constructor.University; bhavana.bhat@mercedes-benz.com

***Abstract*—Understanding causal relationships in market behaviour is essential for strategic decision-making; yet, most artificial intelligence methods remain focused on pattern recognition rather than causal discovery. While reasoning, AI models can generate hypotheses about market dynamics; however, they often lack mechanisms to validate whether observed patterns reflect genuine causal effects or confounding influences. This paper introduces a causal machine learning framework that integrates Double Machine Learning (DML) with causal forests to estimate Conditional Average Treatment Effects (CATE), capturing heterogeneous responses across market conditions. To enhance interpretability and robustness, the framework incorporates validation through Interrupted Time Series (ITS) analysis and Synthetic Control Methods (SCM). Demonstrated through an applied case study using a multi-year dataset of vehicle registrations from 2020 to 2024, the framework identifies and validates causal effects in dynamic automotive market environments. While the current application focusses on the automotive sector, the methodology is generalisable to other domains. The results support the integration of causal estimates into decision support systems, offering a methodological contribution to evidence based market intelligence in industrial applications.**

***Index Terms*—Causal machine learning, Market dynamics analysis, Double machine learning, Causal forests, Interpretable AI, Decision support systems.**

I. Introduction

Advanced artificial intelligence systems, including large language models and deep learning architectures, have transformed market analysis through sophisticated pattern recognition capabilities [1]. These systems can process extensive web data to identify market trends, analyse consumer sentiment from social media, and synthesise competitive intelligence from multiple sources. However, a fundamental limitation constrains their business applications: while reasoning AI models can articulate hypotheses about market dynamics through logical deduction, they do not inherently provide mechanisms to validate whether observed correlations reflect true causal mechanisms or confounding associations. Advanced reasoning models may generate plausible narratives about market drivers, but without rigorous statistical identification from observational data, such conclusions remain unverified hypotheses susceptible to generating confident yet empirically unsupported recommendations. This limitation becomes critical in market intelligence, where decision-makers require validated answers to counterfactual questions about market behaviour under different scenarios [2], [3].This paper addresses this gap through methodological integration rather than algorithmic innovation.

Recent advances in causal machine learning offer potential solutions by integrating graph based reasoning with statistical learning to transcend correlation analysis. Causal forests estimate heterogeneous treatment effects across market segments [4], while DML obtains valid causal estimates in high-dimensional settings [5]. However, machine learning methods—particularly ensemble approaches such as random forests and gradient boosting can suffer from interpretability challenges, raising concerns for business applications that require explainable insights [6], [7].

This paper develops a multi method framework integrating established techniques via DML causal forests estimator with ITS [8] and SCM [9]. The framework estimates CATE to capture heterogeneous responses across market conditions. Through a real world case study based on vehicle registration data from 2020–2024, the analysis demonstrates how causal reasoning could enhance AI driven market analysis systems by enabling validated counterfactual inference while maintaining interpretability through transparent validation. While the empirical focus is on the automotive industry, the framework is designed to be adaptable to other sectors, supporting broader applications in market intelligence.

II. Literature Review

Market analysis requires distinguishing causal mechanisms from spurious correlations in observational data [10], [11]. Modern graphical models establish the foundation for defining causal effects. Applied causal inference, powered by machine learning (ML), offers approaches for analysing observational data [12], with recent comprehensive reviews synthesising advances in causal machine learning for static and dynamic settings [13]. Causality is increasingly recognised in Explainable Artificial Intelligence (XAI) for achieving reliable and interpretable insights [14].

Causal forests, introduced by Athey and Imbens [15], estimate Conditional Average Treatment Effects (CATE) by maximising treatment effect heterogeneity. This approach identifies market segments with varying behavioural responses and has been applied to analyse heterogeneous treatment effects (HTE) in randomised controlled trial data.

Double Machine Learning (DML) [16] employs sample splitting and cross-fitting to obtain orthogonal moment conditions, reducing bias from model misspecification. Almashaleh and Fatahi Valilai [5] demonstrate that integrating DML with the DoWhy framework yields robust causal estimates in high-dimensional social media data, revealing measurable causal effects of content formats on user engagement. Recent comprehensive comparisons confirm the viability of DML for population-level treatment effects [17], with robust software implementations facilitating practical applications [18].

Despite these advances, recent applications in consumer behaviour analysis reveal methodological limitations [19]. Studies in digital marketing [20] often rely on single-method approaches without comprehensive robustness validation [21]. Similarly, in dynamic business contexts, adaptive meta-learning approaches are crucial for continuously updating models as market conditions evolve [22], especially when consumer preferences shift rapidly [23]. However, most studies rely on single-method approaches without comprehensive robustness validation, and reliance on predictive models alone can lead to decisions based on spurious correlations.

Quasi-experimental designs offer complementary approaches for market analysis when observable events exhibit sharp temporal discontinuities. Interrupted Time Series (ITS) analysis evaluates market dynamics by comparing trajectories before and after events [8]. The Synthetic Control Method (SCM) [9] constructs weighted combinations of control units to approximate counterfactual trajectories. Recent methodological advances include augmented synthetic control approaches that combine outcome modelling with weighting [24] and multi-outcome extensions that reduce bias through shared factor structures [25]. These methods provide transparent validation for understanding market adoption patterns, where comprehensive conditions and consumer heterogeneity influence outcomes [26].

A persistent limitation in existing research is the tendency to examine individual factors in isolation, with limited use of complementary validation methods and insufficient attention to heterogeneous effects across dynamic market conditions. Building on prior work integrating causal inference with market intelligence [2], [11], the proposed framework addresses these gaps through methodological integration—combining machine learning flexibility with quasi-experimental transparency for robust causal identification in market intelligence applications.

III. Methodology

This section outlines the implementation of the proposed causal machine learning framework through an empirical case study based on publicly available vehicle registration data. The study design considers three distinct market regimes observed between 2020 and 2024: (i) the period of an enhanced purchase incentive programme, (ii) the introduction of fuel cost mechanisms, and (iii) the subsequent termination of the incentive programme. The primary outcome variable is market share, defined as the proportion of registrations for a specific vehicle category relative to total monthly registrations. The estimation strategy is grounded in a causal inference framework that explicitly controls for confounding temporal trends, demand variability, and competitive market dynamics.

A. Data Sources and Description

This study employs publicly available datasets spanning January 2020 to December 2024. The data integration architecture implements systematic protocols to ensure methodological reproducibility and validity.

1) Vehicle Registration Data: Monthly vehicle registration statistics were obtained from the Kraftfahrt-Bundesamt (KBA) [27]. The dataset comprises comprehensive registration counts disaggregated by powertrain type. This dataset provides complete market coverage through mandatory reporting requirements. For the analytical period, annual aggregate totals were temporally disaggregated to a monthly frequency using cubic spline interpolation methods, a standard econometric technique that preserves aggregate consistency while maintaining smooth temporal transitions and minimising information loss.

2) Economic Indicators: Supplementary data on per capita carbon emissions were obtained from the comprehensive environmental database [28]. These indicators serve as control variables within the causal framework to account for broader awareness trends that may confound the relationship between incentive interventions and market outcomes.

3) Market Condition Timeline Variables: Timeline variables were constructed, documenting three observable market condition periods based on publicly available government announcements:

1) **Enhanced Incentive Period**: Purchase support programme active from July 2020 to December 2022 [29].
2) **Fuel Cost Mechanism**: Carbon-based fuel pricing was introduced in January 2021 [30].
3) **Incentive Termination**: Programme cessation effective January 2024 [31].

B. Data Processing

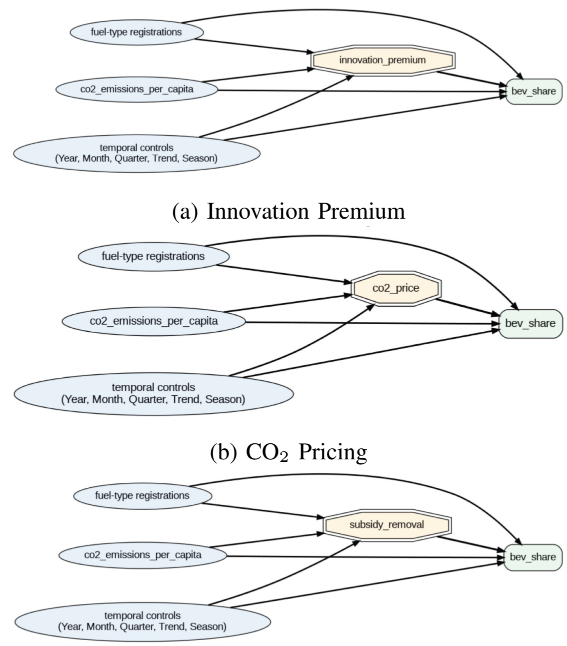
Data processing implemented a medallion architecture comprising bronze, silver, and gold layers to ensure data quality, traceability, and analytical validity [32], [33]. Annual registration data were interpolated to monthly frequency using cubic spline interpolation [34]. Environmental indicators were

TABLE I: Variable Categorization

Variable Name	Description
bev_registrations	Monthly count of new BEV registrations
bev_share	Market share of BEVs relative to total registrations
innovation_premium	Binary indicator for incentive period (Jul 2020–Dec 2022)
co2_price	Binary indicator for CO_2 pricing policy (Jan 2021 onward)
subsidy_active	Binary indicator for financial subsidy availability
total_registrations	Total monthly vehicle registrations
fuel_type_registrations	Registrations by powertrain category
co2_emissions_per_capita	Environmental indicator (tonnes per capita)
temporal_controls	Time-based control variables (trend, seasonality)

TABLE II: Descriptive Statistics (2020-2024, $N = 60$))

Variable	Mean	SD	Min	Max
BEV registrations	28,500	7,200	16,000	44,000
Market share (%)	14.0	5.3	5.1	22.4
Total registrations	210,000	40,000	150,000	290,000
CO_2 per capita (tonnes)	7.9	0.4	7.3	8.5
Innovation premium	0.50	0.50	0	1
CO_2 price	0.80	0.40	0	1



(a) Innovation Premium

(b) CO_2 Pricing

(c) Subsidy Removal

Fig. 1: Causal diagrams illustrating relationships between market conditions, control variables, and market share.

forward-filled for 2024. For variable selection in the 28-variable dataset, theory-driven approaches were prioritised over automated feature selection methods [35], [36], as causal inference requires variables justified by causal mechanisms rather than predictive power alone. The final dataset created a balanced panel with 60 monthly observations. Following established conventions in causal inference methodology [37], variables were categorised according to their theoretical roles within the causal structure: outcome variables, treatment variables, and confounding covariates, as presented in Table I. Descriptive statistics presented in Table II summarise the distributional properties and temporal variation of key variables across the complete observation period (N=60 months).

C. Causal Identification Strategy

The causal relationships were structured using Directed Acyclic Graphs (DAGs), clarifying assumptions about temporal sequencing, confounding factors, and outcome dependencies. Figure 1 illustrates the causal structures underlying the three market conditions considered in this study: innovation premium, CO_2 pricing, and subsidy removal. Identification relies on the assumption that, conditional on these observed covariates, no unblocked back-door paths remain between the market condition variables and the outcome. Accordingly, the included control variables are assumed to sufficiently account for observed confounding influences that jointly affect policy exposure and market share.

D. Model Specification

A causal forest estimator within a DML framework is employed. This approach allows flexible estimation while controlling for high-dimensional confounders. Market share at time t is denoted by Y_t. Similarly, the binary market condition is represented by $T_t \in \{0, 1\}$, while the vector of observed control variables is captured by X_t.

The primary estimand is the CATE, defined as the expected difference between potential outcomes conditional on observed covariates. This measure captures how market responses vary across conditions. In particular, treatment effects are allowed to differ across contextual features rather than being assumed to be constant.

Estimation is conducted using the Causal Forest DML estimator implemented in the `econml` library. Specifically, two nuisance components are learnt. First, the expected outcome conditional on covariates is estimated as

$$m(X) = \mathbb{E}[Y \mid X]. \tag{1}$$

Second, treatment assignment is modelled through the propensity score.

$$e(X) = \mathbb{P}(T = 1 \mid X). \tag{2}$$

Subsequently, orthogonalised pseudo-outcomes are constructed to separate treatment effects from nuisance estimation errors. As a result, the estimated causal effects remain robust even in the presence of complex and high-dimensional confounding.

E. Condition-Specific Estimation

CATE estimates are computed separately for each market condition. For the discrete incentive termination event, validation uses ITS regression in Equation 3 and Synthetic Control gap analysis. These quasi-experimental methods are applied exclusively to the termination event because it represents a sharply timed structural change, satisfying the identifying assumptions required for both ITS and Synthetic Control designs. In contrast, the remaining market conditions evolve gradually over time, for which time-varying CATE estimation provides a more appropriate identification strategy.

$$Y_t = \beta_0 + \beta_1 \cdot \mathrm{time}_t + \beta_2 \cdot \mathrm{post}_t + \beta_3 \cdot (\mathrm{time}_t \times \mathrm{post}_t) + \varepsilon_t, \quad (3)$$

IV. Results and Discussion

The observed market share trajectory over time reveals distinct phases, as shown in Fig. 2. Prior to mid-2020, the levels remained relatively stable. A pronounced increase was observed in early 2021, temporally associated with the implementation of enhanced purchase incentives and revised fuel cost mechanisms. This phase was characterised by persistently elevated levels throughout 2022 and 2023. In early 2024, immediately following programme termination, an abrupt inflexion was detected, after which market share stabilised at substantially lower values. Causal effects were estimated using CATE from a

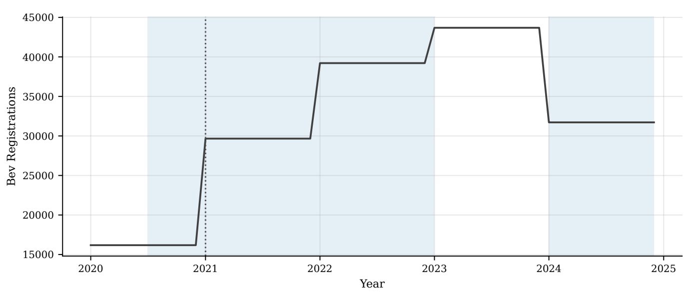


Fig. 2: Observed adoption pattern and market condition.

Causal Forest model within the DML framework, as illustrated in Fig. 3 for the three variables: Innovation Premium, CO2 Pricing, and Subsidy Removal.The enhanced incentive period was associated with gradual positive effects during 2020–2022, suggesting that financial support mechanisms influenced market development patterns. A stronger and more sustained positive effect was attributed to the fuel cost mechanism, which became more pronounced from 2023 onwards, indicating that expected long-run operating costs influenced decisions. Following the termination of the incentive in January 2024, a negative effect was observed. The decline emerged immediately and persisted throughout the subsequent observation period.

ITS analysis provided validation, as shown in Fig. 4. The fitted counterfactual trend continues upward, whereas the observed share declines immediately after termination. The ITS and Synthetic Control validation applies specifically to incentive termination, as this represents a sharply timed discrete event suitable for structural break analysis. For gradual changes, time-varying CATE estimation provides the appropriate identification strategy. These results demonstrate three analytical insights from the empirical analysis. First, financial incentives showed an association with early market development through reduced upfront costs. Second, changes in long-term operating costs, such as fuel pricing, played a key role in sustaining market growth by influencing consumer decision-making over time. Third, the sudden removal of support showed an immediate, measurable decline in shares.

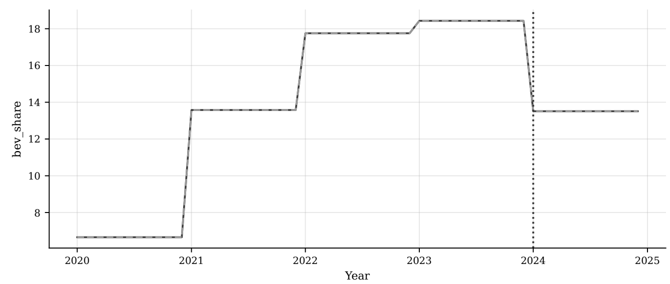


Fig. 4: ITS validation for termination event.

The Synthetic Control comparison provided additional validation. Fig. 5 presents the Synthetic Control gap analysis for the subsidy termination event. The gap measures the difference between the observed BEV market share and its estimated counterfactual trajectory. Before 2024, the gap remains close to zero; therefore, the synthetic control achieves a strong pre-intervention fit. After the termination, however, the gap shifts sharply and remains negative. Consequently, the observed market share falls below the counterfactual path, indicating a negative causal effect following subsidy removal.

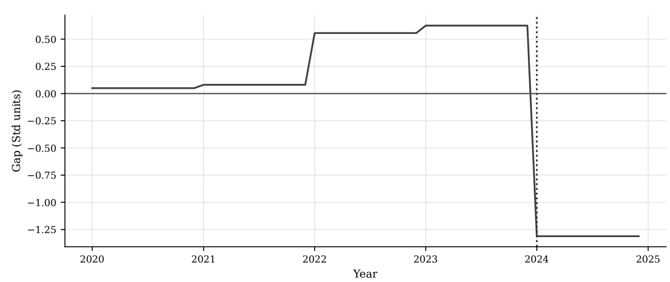


Fig. 5: Synthetic Control gap analysis for termination.

A. Analytical Implications and Robustness

Robustness was examined through a series of sensitivity analyses. First, alternative outcome specifications were considered, including market share and absolute registration levels, and the estimated effect patterns remained consistent

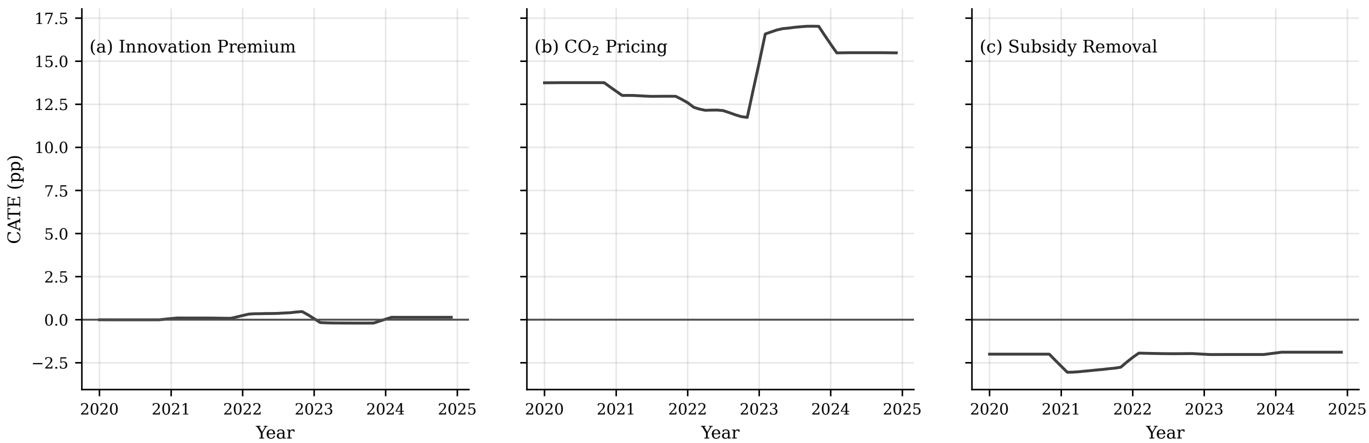


Fig. 3: Estimated conditional average treatment effects (CATE) across market conditions.

across specifications. In addition, visual inspection of the pre-intervention period indicated stable trends, supporting the validity of the identifying assumptions underlying the Interrupted Time Series analysis. For the Synthetic Control analysis, robustness with respect to donor pool composition was evaluated, and comparable post-intervention gaps were observed across alternative donor combinations.

Beyond these robustness cheques, the estimated Conditional Average Treatment Effects display noticeable variation in magnitude over the observation period. This dispersion suggests that market responses to the examined conditions are heterogeneous rather than uniform. As a result, analytical relevance is driven by the distribution of effects across policy phases rather than by a single average estimate. This finding underscores the importance of context-sensitive interpretation when analysing dynamic market behaviour.

Despite the applied controls, potential limitations remain. Unobserved factors, such as concurrent marketing activities or supply-side constraints, may coincide with policy changes and influence market outcomes. If such factors are correlated with both the market condition variables and the outcome, residual bias cannot be fully excluded. Accordingly, the reported estimates should be interpreted as causal effects conditional on the observed data and the stated identification assumptions.

From an analytical perspective, these findings demonstrate how causal machine learning frameworks can enhance market intelligence for strategic decision-making. The results show that market dynamics respond differently across conditions, with time-varying effects revealing shifts in market sensitivity over time. Moreover, the integration of causal forests with quasi-experimental validation improves interpretability by combining flexible estimation with transparent robustness evidence.

From a deployment perspective, the proposed framework can be incorporated into decision support systems through periodic data updates and scheduled model re-estimation. Market data may be refreshed at regular intervals to maintain temporal relevance, while recalibration can be performed when structural changes in market conditions are observed. In this context, model monitoring should prioritise the stability of estimated causal effects rather than predictive accuracy alone. Overall, the framework is designed to support human decision-making by providing validated causal insights, thereby emphasising transparency and governance over automated execution.

V. Conclusion

This research presents a methodological integration of established causal machine learning techniques addressing a fundamental limitation in artificial intelligence for market analysis: the inability to statistically validate causal claims from observational data. While reasoning capable AI models can articulate hypotheses about market dynamics, they cannot rigorously distinguish true causal mechanisms from confounding associations. The proposed framework integrates DML causal forests with transparent quasi experimental validation, potentially transforming AI from pattern recognition into validated market intelligence systems. Using a real world, multi year dataset as a test case, the framework demonstrates methodological advances. Analysis reveals heterogeneous market responses across different conditions, with validated effect estimates that enable an understanding of patterns. This multi method convergence provides evidence that causal machine learning, when properly validated, could potentially distinguish genuine market dynamics from spurious correlations. A key contribution of this work is the development of an integrated causal machine learning architecture that combines the flexibility of DML causal forests estimator with transparent quasi experimental validation. Although the empirical evaluation focuses on the automotive domain, the framework is designed to be domain-agnostic and applicable to other industries where observational data and policy interventions are present.These extensions would further establish the framework's utility as a generalizable tool for evidence-based market intelligence and strategic decision-making.

References

[1] T. Ameer and O. Fatahi Valilai, "Cloud-native causal AI for supply chain KPI monitoring: A GCP framework to diagnose out-of-stock events," *Machine Learning with Applications*, vol. 22, p. 100765, Dec. 2025.

[2] N. Mohammadian and O. Fatahi Valilai, "Decoding Consumer Insights: A Python-Powered Analysis of Social Media for FMCG Production Strategy," in *Intelligent Systems and Applications* (K. Arai, ed.), (Cham), pp. 314–324, Springer Nature Switzerland, 2024.

[3] O. Almashaleh and O. F. Valilai, "Enhancing digital marketing with causal machine learning: Insights from social media engagement in textile circularity," in *2025 International Conference on Software, Knowledge, Information Management & Applications (SKIMA)*, pp. 1–6, IEEE, 2025.

[4] S. Wager and S. Athey, "Estimation and inference of heterogeneous treatment effects using random forests," *Journal of the American Statistical Association*, vol. 113, no. 523, pp. 1228–1242, 2018.

[5] O. Almashaleh and O. Fatahi Valilai, "Causal Drivers of Sustainable Social Media Engagement in the Textile Industry: A Double Machine Learning Approach," *IEEE Transactions on Engineering Management*, vol. 73, pp. 495–509, 2026.

[6] S. Athey and G. W. Imbens, "Machine learning methods that economists should know about," *Annual Review of Economics*, vol. 11, pp. 685–725, 2019.

[7] O. Mashalh, A. Gharaibeh, and A. Khalifeh, "Bandwidth usage forecasting and network anomaly detection based on neural network approach," in *Proceedings of the 3rd International Conference on Advances in Artificial Intelligence*, pp. 42–46, 2019.

[8] J. L. Bernal, S. Cummins, and A. Gasparrini, "Interrupted time series regression for the evaluation of public health interventions: a tutorial," *International Journal of Epidemiology*, vol. 46, no. 1, pp. 348–355, 2017.

[9] A. Abadie, A. Diamond, and J. Hainmueller, "Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program," *Journal of the American Statistical Association*, vol. 105, no. 490, pp. 493–505, 2010.

[10] N. Mohammadian, Y. Uygun, and O. F. Valilai, "Optimizing Product Lifecycle Management with Omnichannel Strategies, Social Media Insights, and Extended Producer Responsibility: A Case Study on Efficient Plastic Waste Recycling," in *Flexible Automation and Intelligent Manufacturing: Manufacturing Innovation and Preparedness for the Changing World Order* (Y.-C. Wang, S. H. Chan, and Z.-H. Wang, eds.), (Cham), pp. 349–367, Springer Nature Switzerland, 2024.

[11] O. Almashaleh, H. Wicaksono, and O. Fatahi Valilai, "A framework for social media analytics in textile business circularity for effective digital marketing," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 11, p. 100544, June 2025.

[12] V. Chernozhukov, C. Hansen, N. Kallus, M. Spindler, and V. Syrgkanis, "Applied causal inference powered by ml and ai," *arXiv preprint arXiv:2403.02467*, 2024.

[13] J.-e. Chen and A. Jing, "Recent advances in causal machine learning and dynamic policy learning," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 17, no. 4, p. e70050, 2025.

[14] G. Carloni, A. Berti, and S. Colantonio, "The Role of Causality in Explainable Artificial Intelligence," *WIREs Data Mining and Knowledge Discovery*, vol. 15, no. 2, p. e70015, 2025. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.70015.

[15] S. Athey and G. Imbens, "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences*, vol. 113, no. 27, pp. 7353–7360, 2016.

[16] V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins, "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, vol. 21, no. 1, pp. C1–C68, 2018.

[17] M. Lechner and J. Mareckova, "Comprehensive causal machine learning," *arXiv preprint arXiv:2405.10198*, 2024.

[18] P. Bach, V. Chernozhukov, M. S. Kurz, and M. Spindler, "Doubleml-an object-oriented implementation of double machine learning in python," *Journal of Machine Learning Research*, vol. 23, no. 53, pp. 1–6, 2022.

[19] O. Almashaleh and O. F. Valilai, "An influence quantification model utilizing social media analytics for business advertising and digital marketing in sustainable fashion industry," in *Intelligent Systems Conference*, pp. 421–436, Springer, 2025.

[20] C. Farronato, L. Johnson, and R. Smith, "Green behavior through mobile applications: A causal analysis," *Environmental Technology & Innovation*, vol. 28, pp. 103–121, 2024.

[21] N. Egami and E. Hartman, "Elements of external validity: Framework, design, and analysis," *American Political Science Review*, vol. 117, no. 3, pp. 1070–1088, 2023.

[22] B. Bhat, S. Suriyanarayanan, and S. Stathel, "Adaptive meta-learning approach for building damage evaluation expert system in dynamic supply chain scenarios," in *Intelligent Systems and Applications* (K. Arai, ed.), vol. 1553 of *Lecture Notes in Networks and Systems*, Cham: Springer, 2025.

[23] J. Grimmer, S. Messing, and S. J. Westwood, "Estimating heterogeneous treatment effects and the effects of heterogeneous treatments with ensemble methods," *Political Analysis*, vol. 32, no. 1, pp. 1–20, 2024.

[24] T. Krajewski and M. Hudgens, "The augmented synthetic control method in public health and biomedical research," *Statistical methods in medical research*, vol. 33, no. 3, pp. 376–391, 2024.

[25] L. Sun, E. Ben-Michael, and A. Feller, "Using multiple outcomes to improve the synthetic control method," *Review of Economics and Statistics*, pp. 1–29, 2025.

[26] E. Muehlegger and D. S. Rapson, "Subsidizing low-and middle-income adoption of electric vehicles: Quasi-experimental evidence from california," *Journal of Public Economics*, vol. 216, p. 104752, 2023.

[27] Kraftfahrt-Bundesamt (KBA), "Fahrzeugzulassungen - neuzulassungen von kraftfahrzeugen nach umwelt-merkmalen." https://www.kba.de/DE/Statistik/Fahrzeuge/Neuzulassungen/neuzulassungen_node.html, 2024.

[28] H. Ritchie, M. Roser, and P. Rosado, "Co_2 and greenhouse gas emissions." https://ourworldindata.org/co2-and-greenhouse-gas-emissions, 2020. Our World in Data.

[29] Federal Ministry for Economic Affairs and Climate Action, "Innovation premium for electric vehicles," policy document, Federal Ministry for Economic Affairs and Climate Action, Berlin, Germany, 2020.

[30] German Emissions Trading Authority, "National emissions trading system for fuel emissions," policy implementation report, German Emissions Trading Authority (DEHSt), Berlin, Germany, 2021.

[31] Federal Ministry for Economic Affairs and Climate Action, "Termination of electric vehicle purchase subsidies," government announcement, Federal Ministry for Economic Affairs and Climate Action, Berlin, Germany, 2023.

[32] Databricks, "What is a data lakehouse?." https://www.databricks.com/glossary/medallion-architecture, 2023. Databricks Documentation.

[33] O. Almashaleh, H. Wicaksono, and O. F. Valilai, "A case study of social media analytics for effective digital marketing in textile circularity," in *2024 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)*, pp. 1–7, 2024.

[34] F. T. Denton, "Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization," *Journal of the American Statistical Association*, vol. 66, no. 333, pp. 99–102, 1971.

[35] T. B. Bhat, M. V. Trupthi, and M. Satish, "Multi-featured data analysis algorithm for feature subset selection based on clustering," *International Journal of Computer Applications*, 2024. In press.

[36] M. Satish, T. B. Bhat, M. V. Trupthi, K. Vishu, and C. N. Chinnaswamy, "Efficient design of fast feature subset selection algorithm for multidimensional data based on clustering," *International Journal of Advanced Computer Science and Applications*, 2024. In press.

[37] M. A. Hernán and J. M. Robins, *Causal Inference: What If.* Chapman & Hall/CRC, 2020.