Causal Machine Learning Approaches for Interpretable Demand Forecasting in Supply Chain Analytics

Rditya Bhopra
Department of Computer Science, University of New Hampshire, Durham, NH, USA.
rditya.work@unh.edu

Ruten Diaz
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
contactruben@ku.edu

Ruben D. Jarvinen
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
rubenwork@missouri.edu

Claudio Terry
School of Computing, Clemson University, Clemson, SC, USA.
claudio.terry@clemson.edu

Abstract

Demand forecasting in modern supply chain analytics faces profound challenges from nonlinear dependencies, latent confounders, and structural interventions such as promotions, disruptions, and policy shifts. Conventional machine learning models, while achieving high predictive accuracy, often function as opaque black boxes that fail to disentangle correlation from causation, thereby limiting their interpretability and robustness under distributional shifts. This paper advances a systems-level framework that integrates causal inference principles with machine learning architectures to produce interpretable demand forecasts that support strategic decision-making. We systematically review the landscape of causal machine learning approaches, including heterogeneous treatment effect estimators, structural causal models, and doubly robust methods, and examine their applicability to supply chain contexts. The discussion emphasizes critical trade-offs between predictive performance, causal identifiability, model complexity, and computational scalability. Architectural considerations such as feature engineering under confounding, model selection for causal effect estimation, and the role of domain knowledge in constructing directed acyclic graphs are explored in depth. Deployment challenges, including data infrastructure requirements, online updating, governance of algorithmic decisions, and sustainability of inference under evolving supply networks, are analyzed from a socio-technical perspective. The paper further draws cross-domain comparisons with healthcare and economics to highlight transferable design principles. Through illustrative cases, we demonstrate how causal machine learning enhances interpretability by providing counterfactual explanations and scenario analyses that are actionable for planners. The conclusion underscores the necessity of embedding causal

reasoning into the forecasting pipeline to achieve resilience, fairness, and long-term operational alignment in supply chain systems.

Keywords

Causal Inference, Interpretability, Demand Forecasting, Supply Chain Analytics, Machine Learning, Robustness.

1. Introduction

Accurate demand forecasting underpins virtually every operational decision in supply chain management, from inventory replenishment and capacity planning to transportation routing and pricing strategies. The increasing availability of granular transactional data, point-of-sale records, and external signals has driven the widespread adoption of machine learning models for forecasting tasks. Methods such as gradient boosting, recurrent neural networks, and attention-based architectures have demonstrated superior predictive performance compared to classical time series approaches [1], [2]. However, these models are fundamentally associational: they learn patterns from historical correlations and are prone to failure when those correlations change due to interventions, policy changes, or external shocks. The lack of interpretability further hampers trust among supply chain managers, who require actionable insights and causal explanations to justify resource allocation and contractual commitments.

The demand for interpretable forecasting has spurred interest in causal machine learning, a rapidly evolving field that combines potential outcomes frameworks, structural causal models, and nonparametric estimation techniques to uncover cause-effect relationships from observational and experimental data [3], [4]. Unlike purely predictive methods, causal approaches aim to answer counterfactual questions: what would demand have been if a promotional campaign had not been run, or if a supplier had not experienced a disruption? Such questions are central to supply chain analytics, where the ability to separate endogenous market dynamics from exogenous interventions is critical for planning. By embedding causal structure into learning algorithms, researchers and practitioners can produce forecasts that are robust to distributional shift and can be explained in terms of underlying mechanisms [5]. This paper adopts a systems perspective to examine how causal machine learning can be operationalized for interpretable demand forecasting, focusing on the structural trade-offs, architectural choices, and governance implications that arise in large-scale supply chain environments.

2. Background and Problem Context

The domain of supply chain demand forecasting is characterized by high-dimensional, temporally dependent, and interventional data. Historical demand observations are the result of complex interactions among consumer preferences, pricing decisions, promotional activities, competitor actions, weather conditions, and macroeconomic trends. Many of these factors act as confounders that simultaneously influence both the treatment of interest (e.g., a price change) and the outcome (demand). Traditional machine learning models, including random forests and deep neural networks, excel at capturing nonlinear interactions but offer limited tools for disentangling these confounding pathways. As a result, estimated effects may be biased and forecasts may extrapolate poorly to new regimes [6].

Interpretability, in the context of supply chain analytics, encompasses multiple dimensions: transparency of the model's internal logic, the ability to attribute predictions to specific input features, and the provision of counterfactual explanations that allow planners to simulate alternative scenarios. Linear models and decision trees provide a degree of transparency but often sacrifice predictive accuracy. Post-hoc explanation techniques such as SHAP and LIME offer approximations of feature importance but cannot guarantee causal validity, as they rely on the same associational assumptions as the underlying model [7]. Causal machine learning directly addresses this gap by explicitly modeling the data-generating process and estimating conditional average treatment effects or structural parameters that have a clear causal interpretation.

From a practical standpoint, supply chains operate as distributed socio-technical systems with multiple stakeholders, heterogeneous data sources, and time-varying structures. Forecasts are used not only for operational decisions but also for strategic planning, contract negotiations, and regulatory compliance. The stakes for interpretability are therefore high: an opaque forecast that fails under a new tariff policy or a sudden raw material shortage can lead to costly stockouts or excess inventory. Causal approaches offer the promise of modularity, where the effects of different drivers can be isolated and validated through domain expertise, thus fostering trust and enabling adaptive decision-making.

3. Causal Inference and Machine Learning: A Taxonomy of Approaches

A broad taxonomy of causal machine learning methods relevant to demand forecasting can be organized along two axes: the type of causal estimand and the flexibility of the estimation procedure. On one axis, researchers distinguish between average treatment effects, conditional average treatment effects, and structural parameters derived from causal graphs. On the other axis, methods range from parametric (e.g., linear instrumental variables) to fully nonparametric (e.g., causal forests, neural networks with propensity score weighting). For supply chain applications, the most relevant estimands are often heterogeneous treatment effects, because the impact of a price change or a promotion may vary across product categories, regions, or customer segments.

One prominent class of methods is the meta-learner framework, which uses base learners such as random forests or gradient boosting to model outcomes and/or treatment assignment [8]. The T-learner, S-learner, and X-learner each adapt differently to the challenge of learning treatment effect heterogeneity while controlling for confounding. These approaches are appealing for their modularity: any predictive algorithm can serve as the base learner, enabling practitioners to leverage existing forecasting infrastructure. However, they require careful specification of the propensity score model and can be sensitive to model misspecification. Doubly robust estimators, which combine outcome regression with propensity score weighting, provide protection against misspecification of either component, a property that is valuable in high-stakes supply chain settings where both data quality and domain knowledge may be imperfect [9].

Structural causal models offer a complementary perspective by explicitly encoding the causal relationships among variables as directed acyclic graphs. In supply chain analytics, such graphs can be constructed from domain knowledge about the timing of promotions, the ordering of procurement decisions, and the flow of information across echelons. Once the graph is specified, machine learning algorithms can be used to estimate the conditional distributions associated with each causal mechanism. This approach yields interpretable forecasts because each variable's prediction can be traced back to its direct causes, and interventions can be simulated by altering the graph's structure [10]. The main limitation is the difficulty of learning the graph from observational data in the presence of cycles, unobserved confounders, and measurement error.

Recent advances in deep learning have extended causal effect estimation to high-dimensional settings. Neural network architectures with separate treatment and control branches, as well as variational autoencoders for latent confounding, have been proposed [11]. These methods can capture complex interactions but introduce additional hyperparameters and require large training samples. In supply chain contexts with moderate data sizes and strong temporal dependencies, simpler ensemble-based methods often outperform deep architectures in terms of stability and interpretability. The choice of method must therefore be guided by the specific trade-offs between model capacity, data availability, and the need for causal identification.

4. Structural Trade-offs and Architectural Considerations

The integration of causal reasoning into a demand forecasting system involves architectural decisions that affect both the upstream data pipeline and downstream decision support. A fundamental trade-off exists between the fidelity of causal modeling and computational tractability. Full specification of a structural causal model with latent variables may be theoretically desirable but becomes infeasible when dealing with thousands of stock-keeping units (SKUs) across multiple distribution centers. Practical architectures often adopt a two-stage approach: first, a causal graph is constructed using domain knowledge and automated causal discovery algorithms, then a reduced set of confounders and instruments is selected for downstream estimation. The selection of adjustment variables requires balancing bias reduction against variance inflation, a problem that is exacerbated in high-dimensional settings [12].

Another architectural consideration is the handling of temporal dependencies and feedback loops. Supply chains are inherently dynamic: a promotion today influences inventory levels tomorrow, which in turn affect future promotions. Standard causal inference methods assume that treatments and outcomes are i.i.d. or that sequential ignorability holds, assumptions that are violated in time series with carryover effects. Recent work on causal inference in time series using recurrent neural networks and state-space models offers a pathway forward, but these methods introduce additional complexity in model selection and hyperparameter tuning [13]. For practical deployment, a common solution is to aggregate data over appropriate time windows and to include lagged versions of both treatments and confounders as features, effectively approximating a nonparametric autoregressive structure.

The architecture of the forecasting pipeline also must account for data governance and versioning. Causal estimates are only as valid as the assumptions underlying the identification strategy. When new data streams, such as social media sentiment or real-time point-of-sale data, become available, the causal graph may need to be revised. Continuous learning pipelines that retrain causal models on rolling windows can introduce concept drift in the estimated effects unless change detection mechanisms are in place. The use of causal discovery algorithms that are robust to distribution shift, such as those based on invariance principles, is an active area of research [14]. In practice, a hybrid approach that combines periodic manual review of the causal structure with automated updating of effect estimates offers a pragmatic balance between agility and reliability.

5. Deployment, Governance, and Sustainability

Deploying causal machine learning models for demand forecasting at organizational scale raises significant governance questions. Forecasts derived from causal methods are often used to justify budget allocations, rebate contracts, and inventory commitments, which creates incentives for stakeholders to influence model design. Ensuring that causal assumptions are auditable and that estimates are reproducible requires transparent documentation of the data provenance, causal graph, and estimation protocol. This is akin to the concept of model cards but extended to include causal identification assumptions and sensitivity analyses [15]. Furthermore, fairness considerations emerge when causal models are used to allocate resources across regions or customer segments. If a causal model estimates that a certain demographic group responds less to a promotion, that estimate could be biased by unobserved confounders related to income or access. Causal machine learning can actually help uncover such disparities when carefully applied, but it also requires vigilance against discriminatory treatment effect estimates that are driven by structural inequalities [16].

Sustainability of the forecasting infrastructure over time depends on the robustness of causal estimates to changes in the supply chain network. Mergers, supplier exits, and new product introductions alter the underlying causal graph, potentially invalidating previously learned effects. Deploying causal models that are modular in the sense that the effect of a single intervention (e.g., a price change) is estimated using a separate submodel allows for selective retraining when only that intervention changes. This modularity aligns with the concept of microservices in software architecture, where each causal effect is a standalone service that can be updated independently. However, maintaining such an architecture requires disciplined version control and coordination among data engineering, modeling, and business teams.

The operational carbon footprint of large-scale causal machine learning pipelines also warrants attention. Training complex models, especially those that involve multiple base learners or deep architectures, consumes considerable energy. In supply chain settings where forecasts need to be updated daily or hourly, the computational cost can be substantial. Simpler methods such as causal forests and doubly robust estimators are often more computationally efficient than deep causal networks, and they produce results that are easier to inspect. From a sustainability perspective, prioritizing interpretable and lightweight causal models not only reduces energy consumption but also simplifies the governance burden.

6. Case Illustrations and Cross-Domain Comparisons

To ground the discussion, we consider two illustrative scenarios from supply chain practice. The first involves a large retailer seeking to evaluate the effect of targeted promotions on store-level demand. Historical data show that promotions are more frequently applied to high-traffic stores, creating a confounding bias. A causal forest is deployed to estimate store-specific treatment effects while controlling for store traffic, day-of-week patterns, and concurrent markdowns. The resulting estimates reveal substantial heterogeneity: some low-traffic stores respond strongly but were historically underserved by promotions. This insight enables the retailer to reallocate promotional spend, increasing overall revenue while reducing waste. The interpretability of the causal forest, through variable importance measures and partial dependence plots, allows category managers to understand why specific stores are predicted to respond differently, building trust in the recommendations.

The second scenario involves a manufacturer forecasting demand for a critical component after a raw material supply disruption. Standard time series models underestimate the long-term impact because they treat the disruption as a one-time shock. A structural causal model that encodes the propagation of inventory shortages through the production process is used to generate counterfactual forecasts: what would demand have been without the disruption? The model provides a transparent decomposition of the forecast into effect channels (e.g., production slowdown, price increase, supplier switch) that can be discussed with procurement teams. This case highlights the value of causal modeling for scenario planning and risk management.

Drawing cross-domain comparisons, the integration of causal machine learning in healthcare for personalized treatment effects shares many structural similarities with supply chain demand forecasting. In both domains, the goal is to estimate how a unit-level intervention (a treatment or a promotion) affects an outcome in the presence of confounding, and to identify subgroups that benefit most [17]. Econometric methods for policy evaluation, such as difference-in-differences and regression discontinuity, have also been adapted to supply chain contexts, though their applicability depends on the availability of clean natural experiments. The supply chain domain further contributes unique challenges: the presence of competition effects (promotions of one product affecting others) and spatial spillovers that require network causal models [18]. These challenges call for continued methodological innovation that borrows from network science and multi-agent systems.

7. Conclusion

Causal machine learning offers a paradigm shift for demand forecasting in supply chain analytics, moving beyond associational correlations toward a deeper understanding of cause-effect relationships that underpin operational decisions. This paper has argued that interpretability, robustness, and governance are not merely desirable properties but essential requirements for forecasting systems that operate in dynamic, intervention-rich environments. By taxonomic analysis of causal methods, discussion of architectural trade-offs, and consideration of deployment realities, we have illuminated a path for integrating causal inference into the forecasting pipeline in a manner that respects both statistical rigor and practical constraints. The cross-domain comparisons with healthcare and economics underscore the transferability of causal principles while highlighting domain-specific complexities such as temporal carryover and network spillovers.

Future work should focus on automating causal discovery in the presence of long time series and high-dimensional confounders, developing scalable estimation procedures that adapt to changing supply networks, and embedding causal reasoning into interactive decision support systems. The ultimate goal is to create forecasting infrastructures that are not only accurate but also explainable, fair, and sustainable, thereby enabling supply chain professionals to make better-informed decisions that align with broader organizational and societal objectives.

References

1. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. PLOS ONE, 13(3), e0194889.

2. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

3. Pearl, J. (2009). Causality: Models, reasoning, and inference (2nd ed.). Cambridge University Press.

4. Imbens, G. W., & Rubin, D. B. (2015). Causal inference for statistics, social, and biomedical sciences: An introduction. Cambridge University Press.

5. Athey, S., & Imbens, G. W. (2016). Recursive partitioning for heterogeneous causal effects. Proceedings of the National Academy of Sciences, 113(27), 7353–7360.

6. Shmueli, G. (2010). To explain or to predict? Statistical Science, 25(3), 289–310.

7. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.

8. Künzel, S. R., Sekhon, J. S., Bickel, P. J., & Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. Proceedings of the National Academy of Sciences, 116(10), 4156–4165.

9. Bang, H., & Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. Biometrics, 61(4), 962–973.

10. Peters, J., Janzing, D., & Schölkopf, B. (2017). Elements of causal inference: Foundations and learning algorithms. MIT Press.

11. Johansson, F. D., Shalit, U., & Sontag, D. (2016). Learning representations for counterfactual inference. Proceedings of the 33rd International Conference on Machine Learning, 3020–3029.

12. Kyono, T., Zhang, Y., & van der Schaar, M. (2020). CASTLE: Regularization via auxiliary causal graph. Proceedings of the 37th International Conference on Machine Learning, 5524–5534.

13. Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: A survey. Philosophical Transactions of the Royal Society A, 379(2194), 20200209.

14. Arjovsky, M., Bottou, L., Gulrajani, I., & Lopez-Paz, D. (2019). Invariant risk minimization. arXiv preprint arXiv:1907.02893.

15. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019). Model cards for model reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency, 220–229.

16. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, 214–226.

17. Athey, S. (2017). Beyond prediction: Using big data for policy problems. Science, 355(6324), 483–485.

18. Toulis, P., & Kao, E. (2013). Estimation of causal peer influence effects. Proceedings of the 30th International Conference on Machine Learning, 1487–1495.

19. Hernán, M. A., & Robins, J. M. (2020). Causal inference: What if. Chapman & Hall/CRC.

20. Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523), 1228–1242.