Causal Machine Learning: An Empirical Approach to Supply Chain Management

Authors

*Alfredo Roa-Henriquez, Assistant Professor Supply Chain Management, Challey Institute Fellow, College of Business, North Dakota State University (NDSU), email: alfredo.roa@ndsu.edu.

Juan Fung, Research Economist, Applied Economics Office (AEO), Engineering Lab (EL), National Institute of Standards and Technology (NIST), email: juan.fung@nist.gov.

Ruhaimatu Abudu, Doctoral Student Transportation and Supply Chain Management, College of Business, North Dakota State University (NDSU), email: ruhaimatu.abudu@ndsu.edu.

Jennifer Helgeson, Research Economist, Applied Economics Office (AEO), Engineering Lab (EL), National Institute of Standards and Technology (NIST), email: jennifer.helgeson@nist.gov.

Douglas Thomas, Research Economist, Applied Economics Office (AEO), Engineering Lab (EL), National Institute of Standards and Technology (NIST), email: douglas.thomas@nist.gov.

Citation: Roa-Henriquez, A., Fung, J., Abudu, R., Helgeson, J., Thomas, D. (2026). Causal Machine Learning: An Empirical Approach to Supply Chain Management. In: Khojasteh, Y., Jahan Abadi, S.M. (eds) *Emerging Technologies in Supply Chains*. International Series in Operations Research & Management Science, vol 373 (pp. 327-358). Springer, Cham. https://doi.org/10.1007/978-3-032-01218-0_13

---

* Corresponding author

Abstract: Over the past two decades, artificial intelligence (AI) has revolutionized industries, with machine learning (ML) at its core. While ML has enhanced supply chain management (SCM) in efficiency and resilience, it often relies on correlations, risking decisions based on spurious relationships. Causal machine learning (CML) offers a solution by focusing on cause-and-effect relationships, promising accuracy and better decision-making. Leading companies like Amazon and Walmart are beginning to harness CML for predictive modeling and AI-driven pricing strategies. Despite its potential, CML in SCM remains nascent, with limited empirical research. This chapter delves into the integration of causal inference in SCM, emphasizing methods that avoid pre-selecting covariates. It reviews foundational causal estimation techniques, surveys CML algorithms, and showcases their application through a simulation. By connecting theoretical advancements with practical implementations, the chapter underscores CML's transformative potential for SCM research and practice, and outlines future directions for its development.

1. Introduction

Over the past two decades, artificial intelligence (AI) has experienced substantial developments and entailed transformative changes for businesses globally (Dutta et al., 2024). Its core technology involves algorithms broadly defined as machine learning (ML) methods. These computationally intensive techniques learn from experience (i.e., past data) to make predictions and improve accuracy, performance, and decision-making (Mohri et al., 2018). The evolution of AI has progressed from ML to more elaborate forms like deep learning (DL), natural language processing (NLP), and generative AIs (GenAI) (Goodfellow et al. 2016; Simonite 2023; Burtell and Toner 2024). Although organizations initially employed AI systems for basic tasks, recent theoretical and practical implications for firms and their supply chains have received significant attention, particularly after large investments in AI and ML adoption took place in the late 2010s (Dutta et al., 2024). Today, essential advancements propelled by large language models (LLMs) and GenAI technology have created capabilities that aim to improve the management of supply chains (Dutta et al., 2024; Liddell, 2023).

Some scholars suggest that firms employing AI have opportunities to enhance their supply chain resilience by increasing visibility and improving sourcing, distribution, and last-mile delivery (Modgil et al., 2021). Others indicate that firms implementing AI can gain efficiencies in supply chain processes by taking advantage of digitalization and unphysicalization (Loske & Klumpp, 2022). Another perspective suggests that supply chain management (SCM), in theory and practice, still needs to prepare for this capability and technological transformation because it entails risks and challenges associated with AI-induced bias (Hendriksen, 2023).

The large-scale implementation of ML methods in SCM began in the late 1990s, initially focusing on logistics, inventory management, and demand forecasting (e.g., Carbonneau et al., 2008). As computational power increased and access to big data grew in the 2010s, the adoption of ML in SCM expanded significantly (Ni et al., 2020; Tirkolaee et al., 2021). Most current applications rely on learning algorithms designed to maximize an objective or minimize a loss function central to the models' training and fine-tuning, ensuring they perform as accurately as possible (Mohri et al., 2018). In ML, these "black box" models are created directly from the data provided to the learning algorithms. Often even the designers of these models cannot explain how variables are selected and used to make predictions due to the complexity and opacity involved in the process (Rudin &

Radin, 2019; Wolf & March, 2024). Sometimes, though, using ML methods to address specific problems in different fields of social sciences (e.g., Economics, SCM, and Marketing) requires careful tuning and adaptation to exploit the structure of those problems. This structure refers, among others, to the causal nature of many variables, endogeneity, panel structure, optimization criteria, and selection of confounders (Athey & Imbens, 2019).

In SCM, understanding the structure or underlying causality mechanism that informs decision-making is essential for businesses to improve their competitive position (Helgeson & Roa-Henriquez, 2022). By identifying which actions actually cause improvements, firms can allocate resources more effectively and invest in initiatives that will certainly improve performance rather than relying on intuition or misleading correlations. For example, if a firm learns that adopting an advanced digital analytics tool can causally reduce lead time variability within a supply chain, whereas other factors such as demand volatility, production capacity, and supplier reliability only correlate with lead time variability, the firm can confidently invest in the tool and gain a competitive edge (see Figure 1).

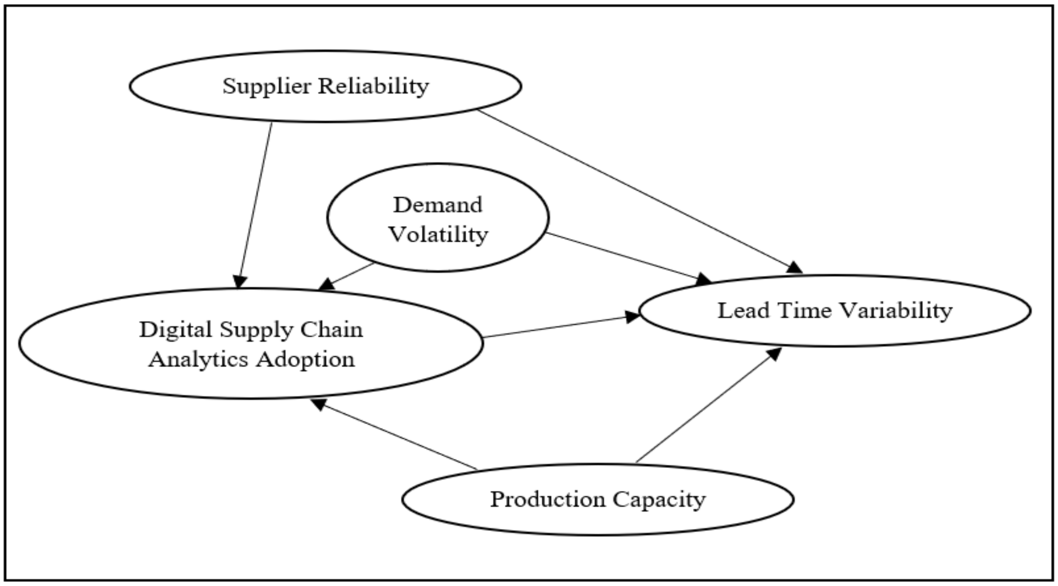


Figure 1. Directed Acyclic Graph (DAG) depicting a hypothetical causal relationship between the causal factor (Digital Supply Chain Analytics Adoption), and the outcome (Lead Time Variability)), including potential confounders. Source: Authors.

Under this scenario, it is possible to combine the advantages provided by ML models to answer prediction questions with the framework provided by causal inference to respond “what if” type of questions (Facure, 2023). Causal machine learning (CML) – or causal AI, more generally – is a recent scientific innovation that goes beyond traditional predictive models by distinguishing between genuine drivers of improvement and spurious relationships of factors that only appear related. In this sense, it addresses the limitations of conventional data or ML models by integrating them with the principles governing causal inference (Athey & Imbens, 2019; Wolf & March, 2024). This new approach has received particular attention from firms in different sectors with capabilities to store, process, and analyze big data, as summarized in Table 1 below.

Table 1. Applications of CML in SCM

Application	Industry	Details	CML Approach
Predicting Supply Chain Disruptions (Bodendorf et al., 2023).	Automotive	Combines deep learning with causal inference to assess supply chain reliability based on supplier data.	Integration of deep learning and causal inference to identify causal disruptions.
Optimal Rework Policies (Schwarz et al., 2024).	Manufacturing	Utilizes data from semiconductor manufacturing to estimate rework policies.	Data-driven causal inference model for policy optimization.
Visual Component Impact (Engin et al., 2023).	Entertainment (Netflix)	Analyzes the causal impact of visual components on viewer engagement.	Causal machine learning to identify key drivers of engagement.
Economic Decision-Making (ALICE, n.d.).	Technology (Microsoft)	Focuses on demand estimation, price optimization, and advertising effectiveness.	Uses causal inference to guide economic strategies.
Logistics and Inventory Efficiency (Götz & Budhathoki, 2022).	E-commerce (Amazon)	Causal analysis of factors affecting delivery times and inventory levels.	Causal inference to enhance logistics operations.
Causal Vector Analysis (Cheikhrouhou, 2023; Chen & Kasiviswanathan, 2020).	E-commerce (Amazon)	Investigates permutations of contextual factors for hypothesis testing.	Contextual vector analysis for causal hypothesis testing.
AI-driven Pricing Models (Pearson, 2021)	Retail (Walmart, Kroger)	Explores the causal effect of AI-powered pricing on sales.	Causal analysis to identify pricing strategies' impact on sales.

The applications of CML across various industries highlight its transformative potential to derive actionable insights from complex datasets. Each application leverages causal inference to address specific industry challenges, ranging from supply chain disruptions to customer engagement and pricing strategies:

1. **Automotive Industry**: The integration of deep learning with causal inference in predicting supply chain disruptions (Bodendorf et al., 2023) underscores the need for robust models that can simultaneously handle large datasets and identify causative factors. This approach enhances reliability in predicting disruptions, offering a significant advantage in a sector where supply chain continuity is critical.
2. **Manufacturing**: The use of causal models for optimizing rework policies in semiconductor manufacturing (Schwarz et al., 2024) illustrates the potential of CML to improve operational efficiency. By focusing on causal relationships rather than correlations, manufacturers can make informed decisions that directly impact production efficiency and cost-effectiveness.
3. **Entertainment**: Netflix's application of CML to evaluate the impact of visual components (Engin et al., 2023) demonstrates the importance of causality in user engagement strategies. Understanding what visual cues drive engagement enables more personalized content delivery, enhancing user satisfaction and retention.
4. **Technology**: Microsoft's ALICE project highlights the role of causal inference in economic decision-making. By understanding causal factors in demand estimation and pricing, technology firms can develop more effective strategies that align with market dynamics.
5. **E-commerce**: Amazon's exploration of causal factors in logistics and inventory efficiency (Götz & Budhathoki, 2022) showcases CML's ability to optimize supply chain operations. By identifying causative elements, firms can streamline processes, reduce delivery times, and increase customer satisfaction.
6. **Retail**: The use of AI-driven pricing models by Walmart and Kroger (Pearson, 2021) to explore causal effects on sales exemplifies how CML can refine pricing strategies. By focusing on causation, retailers can better understand the impact of pricing decisions on consumer behavior and sales performance.

As Table 1 presents, CML is a recent approach employed in the field of SCM to obtain empirical evidence. Although traditional empirical methods applied to observational studies have been more broadly implemented, the application of these methods is still challenging if endogenous shocks are present, simultaneously affecting inputs and firms' outcomes (e.g., Dormady et al., 2022). In this regard, because supply chains are deeply ingrained into the broader economy, the empirical approach needed for causal identification of effects on firms' outcomes becomes much more complex. Additionally, as the application of these advanced methods mature, they may help respond important policy questions associated with incentive misalignments along supply and value chains when the bearers of disaster impacts in one location are different from those managing much of the infrastructure of supply chains (Thomas and Helgeson 2021, 2022) or when disruptions such as power outages affect delivery downstream (Thomas and Fung, 2022).

In this chapter, we discuss methods not often employed in SCM applications but that supply chain practitioners can use for assessing causal effects without the common practice of pre-selecting covariates. These CML approaches aim to identify treatment effects while providing valid inferences under a data-driven selection of covariates (Huber, 2023). However, before presenting this approach with a simulation example, we provide a review of the causal inference methods under the potential outcome framework.

The chapter is organized as follows. Section 2 presents the classical approach to causal effect estimation employed in Randomized Control Trials (RCT) and from observational data, the potential outcome framework – also known as the Rubin causal model. Section 3 presents a survey of different causal machine learning algorithms that are suitable to be employed in the context of SCM applications. Lastly, section 4 presents a simulation example taken from the literature (Yılmaz et al., 2024) that was formerly implemented through matching methods (i.e., techniques that require the pre-selection of covariates by creating comparable groups for treated and untreated subjects). We employ the same data and implement it by using a causal machine learning approach. In section 5, we summarize future research and provide a conclusion.

2. Causal Inference in SCM

SCM integrates *operations processe*s needed to transform resources into products and *logistics processes* needed to supply inputs and services to transformation centers (e.g., manufacturing plants) and distribute products and services to customers (Jacobs & Chase, 2024). In terms of research, the field has experienced a surge in empirical analysis over the last two decades with most of empirical papers in operations management employing observational data to study causal effects derived from, for instance, the implementation of programs or the adoption of policies (Yılmaz et al., 2024), the analysis of the relationship between visibility and performance (Swift et al., 2019), and the implication of supply chain disruptions on prices of shipping containers (Mechai & Wicaksono, 2024). Given the challenges posed by observational data for the estimation of causal effects, we begin the discussion of how causal inference within the context of randomized experiments can be implemented in the study of supply chains. Many of the foundational concepts used for causal inference with observational data were developed for randomized experiments in the early 20th century (Imbens and Rubin 2015) and remain relevant in this context.

Many people are familiar with randomized control trials; for instance, medical trials for new fever-reducing medication are discussed in the context of randomized control trial efficacy. The classic setup is that volunteers sign up for a medical trial and are randomly assigned to either the treatment group (i.e., receive the medication) or the control group (i.e., receive the placebo drug). Data is collected on all the volunteers to draw inferences on the effect of the medication for the reduction of fevers. Assuming random assignment, any observed fever reductions *on average* can be attributed to the treatment. In particular, the causal effect of the medication can be estimated as the average difference in fevers for those in the treatment group and those in the control group. Moreover, this causal effect is unbiased (Imbens 2024). Random assignment is key because it provides confidence that the observed response to the medication (i.e., fever reduction) is purely due to the treatment and not due other factors (i.e., confounders).

In the context of SCM, similar methodologies are employed. As mentioned in the introduction, retailers often utilize AI to experiment with price adjustments (Pearson, 2021). These price adjustments can be part of strategies to optimize resource allocation, such as inventory management or order fulfillment. For instance, A/B testing, or split testing, serves as a practical example of randomized control trials in this domain (EasyBA.Co 2023).

2.1 Potential Outcomes Framework

This concept is generally intuitive to those without any domain-specific expertise, but there is a mathematical framework that provides the foundation for identifying such causal effects. The potential outcomes framework is based on the assumption that causation is tied to an action (i.e., exposure to treatment) applied to a unit (e.g., individual, household, firm) (Imbens and Rubin 2015). The key observation is that while the individual was exposed to a particular condition (say the medication), they could just as easily have been exposed to another condition (the placebo). In that sense, each individual has a set of *potential* outcomes for each of the possible experimental conditions.

As with the preceding example of a medical trial for a new fever-reducing medication, assume we have a binary exposure (i.e., the treatment variable), $W = \{0,1\}$. Let i be denote an individual volunteer in the medical trial. For each of the potential exposures in $w \in W$, we assume there is a corresponding potential outcome, $Y_i(0)$ and $Y_i(1)$, where $Y_i(w)$ would denote the individual's fever when their exposure is set to $W = w$.

By comparing the two potential outcomes, for instance the measured difference in fever with and without the treatment, $Y_i(1) - Y_i(0)$, we obtain an individual's causal effect from the medication. In other words, the causal effect compares an individuals' actual outcome with their counterfactual outcome: what would plausibly have happened if the individual's exposure had been changed.

However, in a randomized experiment, it is impossible to observe all potential outcomes for each individual, as we can only observe one; this is known as the "fundamental problem of causal inference.[1] Instead, we assign many individuals to one of the exposures to observe outcomes across both conditions. We can then estimate causal effects by comparing observed outcomes across multiple units.

Suppose we have N individuals, indexed $i = 1, \dots, N$ and let W_i denote the exposure indicator. That is, $W_i = 1$ if individual i receives the treatment (medication) and $W_i = 0$ if individual i

---

[1] Note that some experiments can be designed to observe the same units over time (Imbens and Rubin 2015).

receives the control (placebo). Using this notation, the observed potential outcome for each individual is expressed as:

$$Y_i^{\text{obs}} = Y_i(W_i) = \begin{cases} Y_i(0), & \text{if } W_i = 0 \\ Y_i(1), & \text{if } W_i = 1 \end{cases} \tag{1}$$

Without loss of generality, we write $Y_i = Y_i^{\text{obs}}$ to denote the observed outcome. Moreover, we can express the unobserved ("missing") potential outcome as:

$$Y_i^{\text{mis}} = Y_i(W_i) = \begin{cases} Y_i(1), & \text{if } W_i = 0 \\ Y_i(0), & \text{if } W_i = 1 \end{cases} \tag{2}$$

Under a randomized assignment mechanism, the potential outcomes are independent of exposure:

$$W_i \perp\!\!\!\perp \big(Y_i(0), Y_i(1)\big) \tag{3}$$

We can then estimate the causal effect of interest, such as the sample average difference in outcomes:

$$\tau \equiv \frac{1}{N}\sum_{i=1}^{N}\big(Y_i(1) - Y_i(0)\big) \tag{4}$$

This causal effect, τ, is called the **Average Treatment Effect** (ATE) and is one of many potential *estimands* of interest, which we will discuss further in Section 2.5.

The potential outcomes framework allows us to model the counterfactual (the "what if") as the unobserved potential outcome. Importantly, this relies on a critical assumption: that individual i's observed outcome is not influenced by the exposure condition of any other individual. This assumption is known as the **Stable Unit Treatment Value Assumption** (SUTVA), defined by Imbens and Rubin (2015) as follows:

- **No interference**: The outcome of unit i is unaffected by treatment exposure of another unit, i.e., $Y_i\big(W_i, W_{j\neq i}\big) = Y_i(W_i)$;
- **Counterfactual consistency**: Each unit's observed outcome corresponds to the potential outcome, given exposure: that is, if $W_i = w$ then $Y_i(\mathrm{w}) = Y_i$.

While SUTVA is a foundational assumption in causal inference, it is recognized as a strong assumption that does not always hold in practice. There is a body of literature addressing the estimation of causal effects in the presence of spillovers, where the treatment of one unit may affect the outcomes of another (Imbens 2024). However, causal inference in the presence of such spillovers is beyond the scope of this chapter.

Identifying causal effects typically relies on SUTVA and other critical assumptions. Among these are *exclusion restrictions*, which assert that any effect on the outcome is solely through the treatment, with no other pathways. While some exclusion restrictions are testable, others remain untestable and depend on the context and domain knowledge.

Exchangeability implies that conditional on observed covariates, the distribution of potential outcomes is the same for treated and untreated units. **Positivity** requires that there is a non-zero probability of receiving each treatment level for all values of the covariates. Closely related is the assumption of **Conditional Independence** (or selection on observables), which posits that, given a set of observed covariates, the treatment assignment is independent of the potential outcomes. These assumptions are crucial when observing the same or different units over time and underpin many causal inference methodologies. In the subsequent sections, we will delve deeper into the roles of Positivity and Conditional Independence in causal inference.

2.2 Observational Studies

The concept of the counterfactual, or unobserved potential outcome, is fundamental to the potential outcomes framework. It formally asks the question, “What if we had done B instead of A?” In the social sciences, conducting idealized randomized experiments is often challenging, if not impossible. Practical or ethical constraints frequently make it infeasible to randomly assign households or businesses to different treatments solely to observe their responses. For instance, when a small business receives a grant for workforce development, we cannot observe the business outcomes had the grant not been awarded. Similarly, it would be unethical to provide disaster assistance “at random” to small businesses affected by infrastructure damage simply for experimental purposes.

Consequently, many causal inference studies rely on observational data. However, the real world presents complexities, and in most cases, potential outcomes derived from observational data are not independent of exposure (Imbens 2024). While observational data is not generated through an experimental process, we can sometimes make certain assumptions about the data-generating process so that the observed data appears “as if” random.

The literature on causal inference with observational data is rich and diverse, spanning various fields and providing insights into numerous complex relationships. Classical examples include Angrist and Krueger's (1991) study on the returns to schooling, which utilized the quarter of birth as proxy (or instrumental variable, to be defined below) to address endogeneity in educational attainment. In public health, observational studies often explore the effects of lifestyle factors on health outcomes, such as diet and exercise, on cardiovascular health, assuming that the selection of observables can adequately control for confounding variables.

In the context of SCM, Bodendorf et al. (2023) investigate the impacts of plausibly exogenous factors, such as changes in political stability and variations in upstream and downstream precipitation and wind speeds, on delivery reliability. This study highlights the use of observational data to infer causal relationships by focusing on external factors that are likely independent of the supply chain's internal dynamics. Another example in SCM might involve assessing the impact of supplier diversity on procurement costs, assuming selection on observables to control for firm size and industry characteristics that influence both supplier diversity and costs. These examples illustrate the breadth of applications and the potential of observational data to yield valuable causal insights across different domains.

A major challenge in observational studies is confounding, which involves disentangling the causal effect of a treatment from other factors that influence the outcome. Addressing these challenges requires key identification assumptions to condition observed data, allowing us to interpret comparisons between treated and control units as causal effects. We discuss these assumptions below, noting that identifying a causal effect depends on a combination of both testable and untestable assumptions. In some instances, interpreting observed outcomes as causal effects may be impossible due to unobserved confounding.

2.3 Selection on Observables

In controlled experiments, treatment is assigned randomly to the population under a random assignment mechanism. In this setting, if potential outcomes are independent of the exposure variable and the two groups (treated and control) are similar regarding outcomes, the causal interpretation is justified (Imbens and Rubin 2015).

With observational data, units such as individuals, households, or firms have agency over their exposure: they can select the treatment. For example, treatments in Bodendorf et al. (2023) include shipping modes and changes in trading-partner economic conditions, which may not be plausibly exogenous. Although we may not have a controlled experiment, we can compare outcomes when the treated and non-treated groups differ only by observable characteristics.

A key assumption that justifies a causal interpretation in such cases is the **Conditional Independence Assumption**, also known as **Selection on Observables**:

$$W_i \perp\!\!\!\perp \big(Y_i(0), Y_i(1)\big)|X_i \tag{5}$$

where X_i denotes a vector of pretreatment variables observed by the analyst. This assumption posits that treatment assignment is independent of potential outcomes, conditional on observables. It is also referred to as Unconfounded Assignment, as opposed to randomized or individualized assignment (Imbens and Rubin 2015).[2]

This assumption is typically paired with the assumption of **positivity**, sometimes called overlap:

$$e(x) = pr(W_i = w|X_i = x) \in (c, 1-c), \quad \text{for some } c > 0 \tag{6}$$

Here, $e(\cdot)$ is known as the **propensity score** and models the exposure, conditional on observables. For many problems, the propensity score is central to causal inference, as we show in the discussion of estimators. Together, the conditional independence assumption and positivity are known as the strong ignorability (or ignorable treatment) assumption.

In practice, a common mistake when selecting variables for estimating the Average Treatment ATE under Conditional Independence is to use variables that are causally influenced by the treatment or the outcome (Imbens, 2024). For example, in SCM, consider a scenario where a

---

[2] Individualized assignment refers to a treatment assignment mechanism where the decision to assign a treatment is based on individual characteristics and not randomly determined. This contrasts with randomized assignment, where treatment is assigned randomly, and unconfounded assignment, treatment is assigned randomly conditional on observed variables (Imbens and Rubin 2015).

company implements a new inventory management system (the treatment) to improve delivery times (the outcome). If analysts mistakenly include delivery times as a variable to condition on when estimating the ATE of the new system, they risk biasing the results because delivery times are directly affected by the treatment itself. Instead, they should focus on variables that are not influenced by the treatment or outcome, such as historical demand patterns.

2.4 Causal Inference without Unconfounded Assignment

In observational studies, the conditional independence assumption that treatment assignment is independent of potential outcomes often does not hold, leading to endogeneity and bias. Endogeneity is a pervasive issue, particularly in economics and other social sciences, arising from omitted variables, measurement error, and simultaneity. These issues can lead to biased estimators, complicating causal inference. For example, omitted variables correlated with both the treatment and the outcome can confound the estimated treatment effect.

To address selection on unobservables, researchers employ several methods to infer causality in natural or quasi-experimental settings. These methods include Instrumental Variables (IV), Regression Discontinuity (RD), Difference-in-Differences (DiD), and Synthetic Control Designs. Each method has specific strengths and weaknesses, depending on the context and available data. We refer the reader to Imbens and Rubin (2015) for more details.

1. **IV:** IV methods are used to address endogeneity by identifying variables (instruments) that are correlated with the treatment but not directly with the outcome. The key conditions for a valid IV are relevance (the instrument is correlated with the treatment) and exogeneity (the instrument affects the outcome only through the treatment).
2. **Regression Discontinuity (RD):** RD designs exploit situations where treatment assignment is based on a cutoff score, such as a test score or income level. By comparing individuals just above and below the cutoff, researchers can estimate the causal effect of the treatment in a local neighborhood around the cutoff.
3. **Difference-in-Differences (DiD):** DiD methods compare the changes in outcomes over time between a treatment group and a control group. This method assumes that, in the absence of treatment, the average change in the outcome would have been the same for both groups.

4. **Synthetic Control Designs:** Synthetic control methods construct a weighted combination of control units to serve as a counterfactual for the treated unit. This approach is particularly useful in comparative case studies where a single treated unit is compared to a synthetic control group.

2.4.1 Instrumental Variables

For illustrative purposes, we present the IV approach to selection on unobservables. While other methods for unconfounded assignment (including RD, DiD, and synthetic control) are also pivotal in causal inference, our emphasis is motivated in part by its natural extension from the estimators case study we present in this chapter. Understanding the IV approach also complements the CML techniques we explore, as there is research on ML for IV (Chen and Hsiang (2019); Wang et al. (2022)). For detailed discussions on other methods, we refer the reader to Imbens (2024).

IV methods address endogeneity by identifying instruments that are correlated with the treatment but not directly with the outcome. The key conditions for a valid IV are:

- **Relevance**: The instrument must be correlated with the treatment.
- **Exogeneity**: The instrument affects the outcome only through the treatment.

Despite their potential, IV methods pose theoretical and empirical challenges. One challenge in using IVs is finding instruments that satisfy both conditions. Theoretically, researchers must prove instrument validity, while empirically, they must find suitable instruments. Conducting robustness checks and sensitivity analyses is crucial to validate findings and ensure reliable causal estimates. Weak instruments, which show a weak correlation with the treatment, can introduce bias. Diagnostic tests, such as the F-statistic, help identify weak instruments. Researchers can use multiple instruments or the Limited Information Maximum Likelihood (LIML) estimator to address this issue. Angrist and Pischke (2009) cover these and other challenges in depth.

For researchers considering IVs or other methods to address selection on unobservables, careful instrument selection, diagnostic testing, and cautious interpretation are paramount. Robustness checks and sensitivity analyses are essential to ensure the reliability of causal estimates. Moreover, recent advances have integrated machine learning techniques with IV estimation to improve robustness and accuracy. Additionally, combining IV with other methods, like RD, enhances causal effect identification.

2.5 Estimands and Estimators

In the potential outcomes framework, we define a range of potential estimands of interest. An *estimand* is the target quantity we aim to estimate, such as the ATE introduced earlier. An estimand may be estimated using various *estimators*, which are methods for quantifying the estimand.

Estimands are defined with respect to specific populations of interest.

- ATE measures the average effect of a treatment across the entire population:

  $$\tau_{ATE} = E[Y(1) - Y(0)].$$

- The Average Treatment Effect on the Treated (ATT) measures the average effect of a treatment on those who have received it:

  $$\tau_{ATT} = E[Y(1) - Y(0)|W = 1].$$

- The Average Treatment effect on the Untreated (ATU) can be defined analogously, focusing on those that did not receive the treatment:

  $$\tau_{ATU} = E[Y(1) - Y(0)|W = 0].$$

- More generally, the Conditional Average Treatment Effect (CATE) targets heterogeneous treatment effects and is defined as the average effect within a subpopulation characterized by specific pretreatment observables $X = x$

  $$\tau_{CATE}(x) = E[Y(1) - Y(0)|X = x].$$

Estimators are methods used to estimate these estimands and can be parametric, semi-parametric, or non-parametric. In CML, the focus is often on developing robust estimators that minimize the need for stringent parametric assumptions, allowing for more flexible modeling of complex relationships.

- Parametric Estimators: Rely on specific functional forms and distributional assumptions.

- Semi-parametric Estimators: Combine parametric and non-parametric elements, offering a balance between flexibility and interpretability.
- Non-parametric Estimators: Make minimal assumptions about the functional form, allowing for greater flexibility in capturing complex patterns.

CML is particularly concerned with developing and applying estimators that can accurately capture causal effects while reducing reliance on parametric assumptions.

In practice, estimands are obtained from a sample of the population of interest, calculated analogously to the sample ATE presented in Section 2.1. Understanding and selecting appropriate estimators is crucial for accurate causal inference, especially in complex data environments.

2.6 Mediation Analysis

In this section, we focus on mediation analysis. We refer the reader to Hernan and Robins (2024) for a detailed discussion on causal mediation. In the context of supply chain management, by identifying mediators (i.e., variables that transmit the effect of a predictor, such as a supply chain disruption, to an outcome, such as supply chain performance), this analysis helps uncover the indirect pathways through which resilience is achieved. This is the case of supply chain resilience, for instance. The resilience of a supply chain might be influenced not only directly by the robustness of its suppliers but also indirectly through factors such as inventory management practices or the flexibility of logistics operations. Understanding these mediating effects can provide deeper insights into the mechanisms by which supply chains absorb and adapt to disruptions. This understanding allows supply chain managers to identify factors and design more effective interventions that enhance resilience by strengthening direct and indirect pathways.

Causal mediation analysis also facilitates targeted improvements in supply chain resilience by distinguishing between direct and indirect effects of risk management strategies. For example, a strategy aimed at diversifying suppliers might directly benefit from reducing dependency on any single supplier. However, its full impact might be mediated through improvements in lead time variability and overall flexibility of the supply chain. By quantifying these mediating effects, organizations can prioritize investments in areas that yield the most significant overall resilience gains, making the supply chain more resistant to disruptions and more adaptive to changing market conditions.

3. Causal Machine Learning in Supply Chain Management

If firms are solely interested in “good” predictions (relative to the decision maker’s loss function), machine learning might be sufficient, as it focuses on finding patterns in data to make accurate forecasts without explicitly understanding why those patterns exist. However, CML becomes crucial when firms need to understand the underlying structure of cause-and-effect relationships that drive these patterns. Causal insights allow firms to make informed, strategic decisions that go beyond correlation (Athey and Imbens 2019; Brand et al. 2023). This can be critical when economic conditions change and spurious correlations are broken. For instance, spurious correlations may no longer apply as a result of an economic downturn, a natural hazard, or other supply chain disruption. When spurious correlations no longer apply and there are limited data observations to correct specification, a model can make suboptimal predictions. Additionally, it can provide insight into what factors can be leveraged to change outcomes and the magnitude of their effects.

Machine learning is typically described as a discipline focused on prediction rather than inference (Breiman 2001; Athey and Imbens 2019). In causal inference, the goal is to understand how an exposure (a feature or input into the model) affects the outcome of interest, providing confidence in the magnitude of the estimand, such as through standard errors or confidence intervals. In contrast, the goal of machine learning is to predict the outcome of interest “really well,” which typically means optimizing metrics such as accuracy or precision based on a loss function. The causal relationship is interesting only to the extent that it minimizes prediction error, with less focus on the magnitude or confidence of estimated parameters like treatment effects.

3.1 Why Causal Machine Learning?

In the last decade, causal inference has increasingly incorporated machine learning methods while maintaining the primary goal of inference. As highlighted in the introduction, notable applications including Amazon's delivery-time analysis and Microsoft’s model focusing on demand estimation and price optimization demonstrate the value of CML.

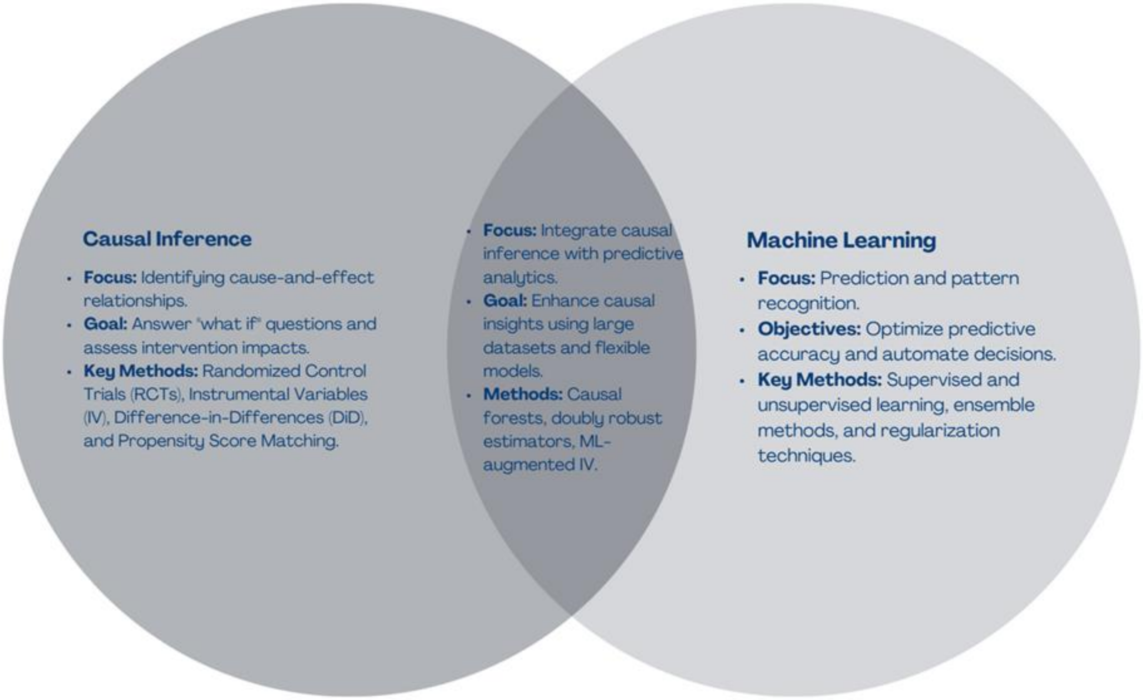
While causal inference fundamentally relies on identification assumptions rather than specific estimators, machine learning can offer estimators with desirable properties, such as efficient first-stage estimators, that can be seamlessly integrated into causal inference frameworks. Despite these advances, inference remains a challenge, with ongoing differences between the econometrics-statistics focus on inference and the machine learning focus on prediction.

Machine learning facilitates “model free” causal inference. Typically, estimating an estimand requires unnecessary parametric assumptions on the effect of X on the outcome. For instance, in a model such as $Y_i = \tau W_i + \beta X_i$, under reasonable conditions, we can relax the parametric assumption and use machine learning methods to learn the functional form of the relationship between X and the outcome: $Y_i = \tau W_i + f(X_i)$. Under certain conditions, we can achieve *consistent* estimators and *asymptotic normality*, providing confidence intervals for inference.

Parametric assumptions often lead to modeling choices that constrain the functional form of f. In a linear model, it can be challenging to include many interaction effects or to estimate heterogeneous treatment effects. Relaxing parametric assumptions allows for more flexible modeling of the outcome and estimation of a broader range of estimands. However, semi- or non-parametric model, or more complex machine learning models, may not always improve inference. As we relax parametric assumptions, more observations (N) are needed to estimate f. More flexibility also allows us to include more features, increasing the dimension K of X. Both factors can introduce the curse of dimensionality, resulting in biased or inefficient (i.e., not consistent) estimators as N or K (or both) increase.

In the next subsection, we elaborate on a few key semi- and non-parametric estimators that are suitable in settings with selection on observables. We then discuss a few key machine learning techniques, including regularization and ensemble methods, which can be used to implement a broad range of estimators.

Figure 2 illustrates the intersection of machine learning and causal inference and some of the topics we cover (as well as topics beyond the current chapter scope).



Causal Inference

- **Focus:** Identifying cause-and-effect relationships.
- **Goal:** Answer "what if" questions and assess intervention impacts.
- **Key Methods:** Randomized Control Trials (RCTs), Instrumental Variables (IV), Difference-in-Differences (DiD), and Propensity Score Matching.

- **Focus:** Integrate causal inference with predictive analytics.
- **Goal:** Enhance causal insights using large datasets and flexible models.
- **Methods:** Causal forests, doubly robust estimators, ML-augmented IV.

Machine Learning

- **Focus:** Prediction and pattern recognition.
- **Objectives:** Optimize predictive accuracy and automate decisions.
- **Key Methods:** Supervised and unsupervised learning, ensemble methods, and regularization techniques.

Figure 2.Intersection of Causal Inference and Machine Learning. Causal Inference identifies causal links, while Machine Learning excels in prediction. Causal Machine Learning leverages both to provide deeper causal insights Source: Authors.

To set the stage, Table 2 below provides an overview of the machine learning models we discuss, categorizing them into traditional and advanced models. For each model, we provide an explanation, use cases, strengths, and weaknesses. This table is intended to serve as a reference point, aiding readers in comparing and contrasting the various models discussed.

Table 2. Overview of machine learning methods, including use cases, strengths, and weaknesses.

Model	Explanation	Use Cases	Strengths	Weaknesses
Linear Regression	A basic regression model that assumes a linear relationship between variables.	Predicting sales, demand forecasting	Simplicity, easy interpretability	Assumes linearity, sensitive to outliers
Decision Trees	A model that splits data into subsets based on feature value decisions.	Classification, risk assessment	Interpretability, handles both numerical and categorical data	Prone to overfitting, can be unstable
Random Forests	An ensemble of decision trees to improve prediction accuracy.	Demand forecasting, anomaly detection	Robust to overfitting, handles high dimensionality	Less interpretable, computationally intensive
Gradient Boosting	An ensemble model that iteratively improves weak models.	Price optimization, customer behavior	High accuracy, handles complex relationships	Prone to overfitting, requires careful tuning
Neural Networks	Deep learning models capable of capturing complex patterns in data.	Image recognition, demand prediction	High flexibility, can model non-linear relationships	Requires large datasets, opaque model structure
Ridge Regression	Linear regression with L2 regularization to prevent overfitting by penalizing large coefficients.	Handling multicollinearity in supply chain data; improving model robustness.	Reduces model variance; stabilizes estimates in presence of multicollinearity.	Does not perform variable selection; all variables are retained in the model.
Lasso Regression	Linear regression with L1 regularization to enforce sparsity in features by shrinking some coefficients to zero.	Feature selection, predictive modeling	Reduces overfitting, performs automatic feature selection	May exclude relevant features, sensitive to choice of regularization
Elastic Net	Combines L1 and L2 regularization for feature selection and model stability.	Predictive modeling	Balances feature selection and model stability, handles correlated features	Requires careful tuning, more complex than Lasso or Ridge

3.2 Key Estimators in Causal Inference

Inverse-Propensity Weighting (IPW) adjusts for confounding by estimating treatment probabilities based on observable covariates, assuming selection on observables and SUTVA (Rosenbaum and Rubin 1984). Let $\hat{e}(x) = P(W = 1|x)$ denote the estimator of the exposure model (or propensity score), or the probability of treatment given observable covariates. The IPW estimator of the ATE is

$$\hat{\tau}_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{W_i Y_i}{\hat{e}(X_i)} - \frac{(1 - W_i)Y_i}{1 - \hat{e}(X_i)}\right) \tag{7}$$

Under correct specification of the propensity score model and the assumption of selection on observables, the IPW estimator is consistent for the ATE (Imbens 2004). The efficiency of the IPW estimator is contingent upon accurately estimating the propensity scores. Misspecification can lead to inefficient estimates. Finally, while IPW can handle high-dimensional covariates, it is sensitive to extreme propensity score values, which can inflate the variance.

The Augmented Inverse-Propensity Weighting (AIPW) estimator, introduced in Robins, Rotnitzky, and Zhao (1994), additionally incorporates outcome estimation for each of the potential outcomes $\hat{Y}_1(x)$ and $\hat{Y}_0(x)$. Like IPW, AIPW assumes selection on observables and SUTVA. Combining both the propensity score model and the outcome model to estimate the ATE is intended to mitigate potential biases from misspecification. The AIPW estimator of the ATE is

$$\hat{\tau}_{AIPW} = \frac{1}{n}\sum_{i=1}^{n}\left(W_i\frac{Y_i - \hat{Y}_1(X_i)}{\hat{e}(X_i)} - (1 - W_i)\frac{Y_i - \hat{Y}_0(X_i)}{1 - \hat{e}(X_i)} + \left(\hat{Y}_1(X_i)_{-}\hat{Y}_0(X_i)\right)\right) \tag{8}$$

The AIPW estimator often achieves higher efficiency compared to IPW alone by leveraging information from both models. Moreover, by incorporating outcome models, the AIPW estimator can reduce bias when the propensity score model is misspecified, offering more reliable causal estimates. Finally, the AIPW estimator is **doubly robust**.[3]

---

[3] We refer the reader to the excellent course notes by Stefan Wager (2022).

Doubly robust estimators, including AIPW and Targeted Maximum Likelihood Estimation (TMLE),[4] combine the exposure model $\left(W = g(Z)\right)$ with the outcome model $\left(Y = f(X)\right)$ to estimate causal effects. A notable feature of these estimators is that the models can take any functional form and the sets X and Z do not need to coincide. This flexibility allows for more tailored modeling approaches.

One advantage of doubly robust estimators is the property of **weak double robustness**, which refers to their ability to provide consistent estimates if either the exposure model, g, or the outcome model, f, is correctly specified, though not necessarily both (Funk et al. 2011).[5] This property provides a safeguard against misspecification, making these estimators particularly useful in complex observational settings. More importantly, when both the exposure and outcome models are correctly specified, the estimator achieves **strong double robustness**, allowing the estimator to exhibit optimal statistical properties, such as consistency and asymptotic normality, ensuring minimal variance.

Doubly robust estimators are desirable in settings where there is uncertainty about which model (exposure or outcome) is correctly specified. They are particularly valuable in observational studies with high-dimensional data or when there is potential for complex interactions and non-linear relationships among covariates, allowing for the use of semi- and non-parametric machine learning methods for the exposure or the outcome. Moreover, inference properties hold even when the estimator uses cross-validation for model selection. The ability to achieve consistent estimation even with partial model specification knowledge makes these estimators a powerful tool in causal inference.

However, despite their advantages, doubly robust estimators also have several potential disadvantages. Implementing these estimators can be computationally intensive, particularly in large datasets or when using complex models, such as those involving machine learning

---

[4] Targeted Maximum Likelihood Estimation (TMLE) is an approach that iteratively refines initial estimates of the outcome model using the propensity score, ensuring that the final estimator is doubly robust. TMLE optimizes both the bias and variance trade-offs, leading to efficient and consistent estimates under correct model specifications (van der Laan and Rose, 2011).

[5] More precisely, if one of the models is *consistent*.

techniques. While they can be consistent if either model is correct, the choice of models and their specifications can be challenging. Poorly chosen models can lead to inefficiencies and increased variance. In practice, doubly robust estimators may exhibit poor performance in small samples, where the theoretical properties are less likely to hold. Although designed to be robust, the performance of doubly robust estimators can degrade if both models are severely misspecified, leading to biased estimates.

In settings with selection on unobservables, doubly robust estimators are typically not suitable. These estimators rely on the assumption of selection on observables, meaning all relevant confounding variables must be observed and accounted for in the models. When there is selection on unobservables, the underlying assumption that either the treatment or outcome model sufficiently adjusts for confounding is violated. In such cases, alternative methods that account for unobserved confounders, such as instrumental variable approaches or structural modeling techniques, may be more appropriate.

3.3 Advanced Machine Learning Techniques

Before diving into the specific techniques, it is important to understand why techniques such as regularization and ensemble methods hold significance in causal inference, particularly within SCM applications. Regularization methods, such as ridge and lasso, are powerful tools that help prevent overfitting by penalizing overly complex models, thus improving the generalizability of causal estimations. These techniques are especially useful in high-dimensional settings where the number of potential covariates is large, providing a disciplined approach to model selection and variable shrinkage. Similarly, ensemble methods like random forests offer robustness against overfitting by combining multiple models to produce more stable and reliable estimates. Their ability to capture complex interactions and non-linear relationships makes them valuable in exploring causal effects, particularly when traditional parametric models may fall short.

3.3.1 Regularization Techniques

Regularization is a critical technique in machine learning that addresses the challenge of overfitting by adding a penalty term to the loss function. This penalty discourages complex models that fit the training data too closely, enhancing the model's generalization to new data (Hastie et al., 2009).

Regularization techniques help in feature selection and can handle multicollinearity among predictors (Tibshirani, 1996).

For simplicity, we discuss regularization in the context of ordinary least squares (OLS) regression, where $f(X) = x_i^T\beta + \varepsilon$ and the objective is to minimize the sum of squared residuals, $(Y_i - X_i^T\beta)^2$. Note that these techniques have been extended to the full family of generalized linear models (GLMs) that include logistic (binomial), Poisson, and gamma regression and the objective is to minimize the (negative) log-likelihood function.[6]

Ridge regression (also known as L2 regularization) adds a penalty equal to the square of the magnitude of coefficients:

$$L_2 = \sum_{i=1}^{n}(Y_i - X_i^T\beta)^2 + \lambda\sum_{j=1}^{K}\beta^2 \tag{9}$$

Lasso regression (L1 regularization) adds a penalty equal to the absolute value of the magnitude of coefficients:

$$L_1 = \sum_{i=1}^{n}(Y_i - X_i^T\beta)^2 + \lambda\sum_{j=1}^{K}|\beta| \tag{10}$$

Elastic net is a regularization technique that combines the penalties of both ridge and lasso, providing a more flexible approach that can handle correlated variables and performs well in high-dimensional settings (Zou and Hastie, 2005).

$$L_\alpha = \sum_{i=1}^{n}(Y_i - X_i^T\beta)^2 + \lambda\left(\alpha\sum_{j=1}^{K}|\beta| + \frac{1-\alpha}{2}\sum_{j=1}^{K}\beta^2\right) \tag{11}$$

The regularization term in the elastic net is thus a mixture of L1 and L2, with $\lambda \geq 0$ serving as a tuning parameter and $0 \leq \alpha \leq 1$ as the hyperparameter that determines the mixture between L1 and L2.

---

[6] For further details on GLMs and elastic net regularization for GLMs, see Tay et al. (2023).

For causal inference, regularization can be particularly useful in selecting covariates and ensuring that the estimation models are not overly complex, which could lead to biased estimators (Belloni et al., 2014).

3.3.2 Ensemble Methods

Ensemble methods are a class of machine learning techniques that combine multiple models to improve predictive performance and robustness. By aggregating the outputs of various models, ensemble methods can mitigate the variance and bias typically associated with individual models, leading to more reliable estimations. Common ensemble techniques include bagging, boosting, and stacking. Bagging, or bootstrap aggregating, enhances model stability by training multiple versions of a model on different subsets of the data and averaging their predictions (Breiman, 1996). Boosting, on the other hand, sequentially applies models to the data, focusing on correcting the errors made by previous models, thus improving overall accuracy (Friedman, 2001). Stacking involves training a meta-model to combine the predictions of several base models, potentially capturing complex patterns that single models might miss (Wolpert, 1992).

In the realm of causal inference, ensemble methods like causal forests—a variant of random forests—can be particularly valuable. They estimate heterogeneous treatment effects by leveraging the ensemble's ability to handle complex interactions and non-linear relationships (Athey et al., 2019). This is especially useful in SCM, where diverse and intricate causal factors often influence outcomes. Ensemble methods thus provide a robust framework for exploring causal relationships, allowing researchers and practitioners to derive actionable insights from complex data landscapes.

3.3.3 Random Forests

Random forests are an ensemble learning method that constructs multiple decision trees during training and outputs the mode of their classes (for classification) or mean prediction (for regression) (Breiman, 2001). This method is known for its robustness and ability to handle large datasets with higher dimensionality, making it suitable for causal inference applications where complex interactions and non-linearities are present.

In the context of CML, random forests can be adapted for causal inference through approaches such as causal forests, which are specifically designed to estimate heterogeneous treatment effects

(Athey et al., 2019). These methods allow for flexible modeling of the treatment effect heterogeneity without the need for pre-specifying interaction terms.

3.4 Summary of strengths and weaknesses

Machine learning techniques are increasingly applied in causal inference to address complex supply chain challenges. Each technique has its unique strengths and limitations, making it crucial for practitioners to understand when and how to use them effectively. Table 3 provides a comprehensive overview of these aspects, alongside recommendations for practitioners.

Table 3. Summary of ML Models and Methods for Causal Inference

Model	Strengths	Weaknesses	Recommendations
IPW	Straightforward adjustment for confounding; suitable for high-dimensional data	Sensitive to extreme propensity scores; requires correct model specification	Use when selection on observables is plausible and data is well-controlled; avoid with extreme propensity values
AIPW	Combines IPW with outcome modeling for bias reduction; doubly robust	Computationally intensive; sensitive to model misspecification	Employ in settings with complex interactions; leverage machine learning for outcome modeling
Doubly Robust Estimators	Protection against model misspecification; consistent if one model is correct	High computational cost; challenging model specification	Ideal for high-dimensional data and potential interactions; ensure robust model validation
Regularization	Prevents overfitting; aids in feature selection	May oversimplify models if not carefully tuned	Suitable for high-dimensional settings; use cross-validation to select hyperparameters
Ensemble Methods	Robustness against overfitting; captures complex interactions	Lack of interpretability; computationally demanding	Best for complex causal structures; use stacking to mitigate reliance on specific models
Random Forests	Handles non-linear relationships; estimates heterogeneous treatment effects	Interpretability challenges; requires large datasets	Effective for exploring heterogeneous effects; ensure sufficient data to avoid overfitting

Integrating machine learning techniques such as regularization and ensemble methods with causal inference tools like IPW, AIPW, and doubly robust estimators offers substantial benefits but also presents challenges. The guidance for application of these methods in SCM can be summarized as follows:

- **IPW and AIPW**: These methods are foundational in adjusting for confounding variables, especially in high-dimensional settings. IPW is relatively straightforward but can be sensitive to extreme propensity scores, which destabilizes the variance of estimates. AIPW enhances this approach by incorporating outcome models, providing robustness against misspecification of either the exposure or outcome model. In SCM applications, where complex interactions are common, AIPW is particularly beneficial, although computationally intensive.

- **Doubly Robust Estimators**: These estimators are valuable in settings with uncertainty about model specification. By achieving consistency if either the treatment or outcome model is correctly specified, they offer a safeguard against potential misspecification errors. In practice, this requires careful consideration of model choices and validation processes to ensure reliability, especially in high-dimensional data environments.
- **Regularization Techniques**: Regularization methods like ridge, lasso, and elastic net are essential for handling multicollinearity and preventing overfitting, particularly in high-dimensional SCM data. They are useful for feature selection and can be tuned using cross-validation to optimize performance. However, practitioners must be cautious not to oversimplify models, especially when capturing complex causal relationships.
- **Ensemble Methods**: Techniques such as bagging, boosting, and stacking provide robustness by aggregating model outputs. They are particularly effective in scenarios with complex causal structures, offering enhanced predictive performance and stability. However, their lack of interpretability and computational demand should be carefully managed. Stacking, which combines multiple models, is recommended for reducing reliance on any single model specification.
- **Random Forests and Causal Forests**: These methods are adept at modeling non-linear relationships and estimating heterogeneous treatment effects. In SCM, where diverse and intricate causal factors influence outcomes, these approaches are powerful tools. Nonetheless, they require large datasets to avoid overfitting and maintain interpretability. Practitioners should ensure data sufficiency and consider interpretability trade-offs when employing these models.

By understanding these strengths and limitations, practitioners can make informed decisions about the appropriate use of these methods in their causal inference efforts within SCM, ensuring both the reliability and relevance of their findings.

4. Application with synthetic data

With the purpose of highlighting the usefulness for the field of supply chain and operations management and guiding scholars on how to approximate randomized experiments by using synthetic control and matching methods, Yılmaz et al. (2024) presents implementation details in a simulation study comparing matching algorithms. The hypothetical scenario describes a brick-and-mortar store that is expanding to online sales, offering exclusive SKUs unavailable in-store. The observations are customer transactions. The outcome of interest is customer spending. The exposure is a promotion (special deal) that the retailer is offering to incentivize online sales. Thus, we are interested in the causal impact of the deal on spending. The data-generating process (DGP) follows a selection-on-observables model (satisfying SUTVA and the other key identification assumptions), as per Heckman and Robb Jr. (1985).

Table 4 presents a summary of the variables, including details on the outcome and exposure models. The outcome model is:

$$OnlineSpending = \tau Discount + f(X) + u \tag{12}$$

The DGP specifies a linear relationship between $OnlineSpending$ and X, where X includes all the covariates in Table 4 (with continuous covariates in log form).

The exposure model is

$$Discount^{*} = g(X') + v \tag{13}$$

The DGP specifies a linear relationship between X' and the latent variable, $Discount^{*}$, where X' includes Age, log($Distance$), and log($PastVisits$). The observed binary exposure, $W = Discount$, is determined by a cutoff value, κ, such that $Discount = 1$ if $Discount^{*} > \kappa$.

We make a minor modification to the exposure model by adding a squared Age term. This simple modification allows us to explore the impacts of misspecification of our empirical models. In a sense, we have a semi-parametric exposure model: $Discount^{*} = \theta Z + g(Age) + v$, where $Z = (X' \setminus Age)$ and g is the linear combination of Age and Age^{2}. Given the new latent exposure, we use an imbalanced treatment rule such that κ is a random number between the 8th and 9th decile of the latent exposure. More details on the data generating process, including model equations and code, we refer to Yılmaz et al. (2024).

Table 4. Descriptive statistics for synthetic data, based on Yılmaz et al. (2024).

		All (N = 10,000)				**Control (N=8637)**	**Treated (N=1363)**
Variable	**Definition**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**Mean**
OnlineSpending	Spending at the first online transaction	50.59	18.68	18.46	239.01	48.59	63.25
Male	1 if customer is male, 0 otherwise	0.45	0.50	0	1	0.45	0.44
Age	Customer age	38.33	8.55	18	80	34.67	49.57
Income	Customer household income	59.92	35.03	10	250	55.70	73.90
PastStoreVisits	Number of store visits last year	5.96	3.15	1	23	5.94	6.32
PastStoreSpent	Amount spent in the store last year	208.22	314.51	60	4000	194.62	315.57
Midterm	Medium-term customer (24-47 months)	0.20	0.40	0	1	0.20	0.21
Longterm	Long-term customer (48+ months)	0.40	0.49	0	1	0.40	0.40
Distance	Distance to the store	20.00	2.99	9.01	30.95	20.05	19.64

To provide a clearer understanding of the causal relationships inherent in our synthetic scenario, we present a causal graph in Figure 3. This graph visually represents the assumptions about the relationships between the treatment (Discount), the outcome (Online Spending), and relevant covariates. The causal graph, constructed using a Directed Acyclic Graph (DAG), helps to illustrate how we believe the promotional offer influences online spending, accounting for potential confounding variables such as Age and Distance.

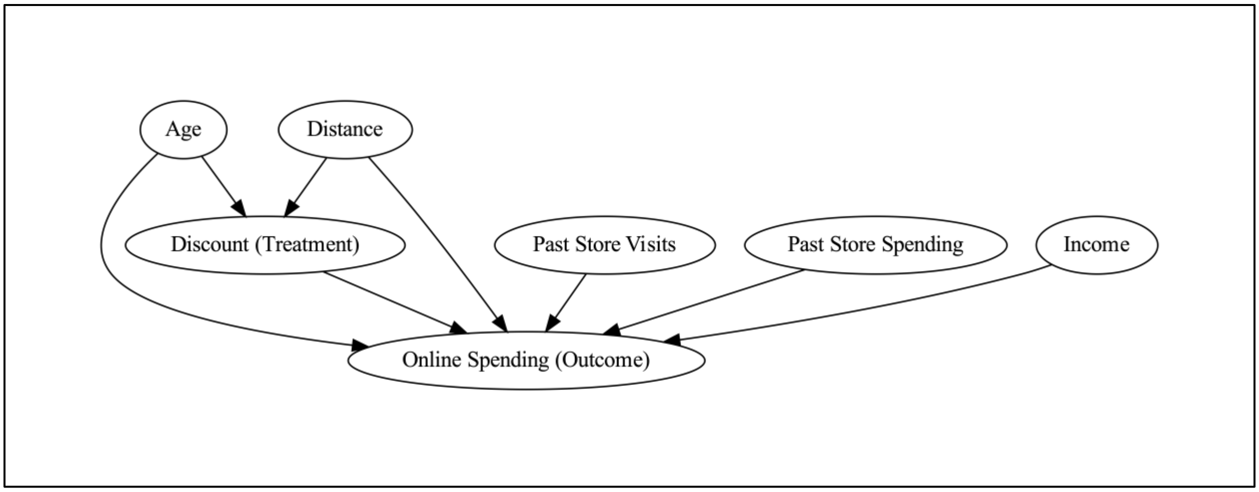


Figure 3. Directed Acyclic Graph (DAG) depicting causal relationships between the covariates, the treatment (Discount), and the outcome (Online Spending), including potential confounders. Source: Authors.

The causal graph in Figure 3 depicts the hypothesized relationships among the treatment, outcome, and covariates. The directed edges (arrows) represent causal assumptions, where the direction indicates the presumed causal influence. For example, the treatment Discount is expected to directly affect the Online Spending outcome. Additionally, covariates such as Age and Distance are assumed to have causal effects on both the treatment and the outcome, suggesting their potential role as confounders. Understanding these relationships is crucial for identifying causal effects and ensuring that our analyses adjust appropriately for confounding.

4.1 Application setup

Our estimand of interest is the ATE. In the true DGP, we know the ATE is $\tau = 10$. We begin our case study by illustrating naïve estimation of the ATE: a GLM for the outcome model that is used to predict both potential outcomes. The ATE is then the difference between the predicted potential outcomes. We then explore the impact of misspecification (of the outcome model) on the results from naïve estimation. To incorporate an exposure model, we then explore IPW estimates of the ATE. This time, we can explore the impact of misspecification of the exposure model as well as the outcome model. Finally, we explore doubly robust estimation with AIPW, where again we examine the impacts of misspecification. Moreover, we illustrate a few of the key machine learning techniques (i.e., regularization, random forests, and ensembles with stacking).

Our analysis uses the R programming language (version 4.4.0, build aarch64-apple-darwin20 for Mac OS Sequioa), including the following packages:

- `boot` and `rsample`: These packages were used to construct bootstrap standard errors and confidence intervals. The `boot` package allows for resampling with replacement, while `rsample` facilitates the creation of resampling objects, supporting various types of bootstrap strategies.
- `WeightIt`: This package was employed for implementing IPW in our analysis. It provides a flexible interface for creating weights that adjust for confounding, helping in estimating causal effects from observational data.
- `AIPW` and `SuperLearner`: For the AIPW analysis, we utilized the `AIPW` package. Coupled with `SuperLearner`, a framework for ensemble learning, we enhanced the robustness of our causal estimates by stacking multiple models and optimizing both bias and variance.

The authors did not generate the synthetic data used in this analysis. We adhered to the exact Stata code provided in the cited paper by Yılmaz et al. (2024) to replicate the data generation process, except for modifying the exposure model with the addition of the Age^2 term.

4.2 Naive ATE

First, we "naively" fit a linear outcome model (in particular, a GLM, with Gaussian link, which is equivalent to classical linear regression, OLS). We use the linear model fit to predict the outcome:

- Y_1, when we assume all units are treated ($W_i = 1, \forall i$)
- Y_0, when we assume all units are untreated ($W_i = 0, \forall i$)

We then compute the ATE as $Y_1 - Y_0$. To obtain valid standard errors (and confidence intervals), we bootstrap the process (based on $n_{boot} = 4000$ replications).

The results are summarized in Table 5. The correctly specified model yields an ATE of 10.55 with a standard error of 0.103, and a 95% confidence interval of (10.34, 10.74). However, omitting a key variable, Age, results in a misspecified model, producing a biased ATE of 8.67 and increased standard errors, highlighting the sensitivity to model specification.

Table 5. Comparison between the correctly specified and missepecified models for ATE.

Model Specification	**ATE**	**Standard error**	**95% Confidence interval**		**Bootstrap replicates**	**Calculation scale**
			Lower interval	Upper interval		
Naive ATE	10.55	0.103	10.34	10.74	4000	Original scale
Naive ATE, misspecified	8.67	0.154	8.36	8.97	4000	Original scale

4.3 Inverse Probability Weighted estimates

Using IPW, we estimate separate exposure and outcome models, allowing for flexible modeling. Results are shown in Table 6. With both models correctly specified, the estimated ATE is 10.08, but misspecification of either model leads to biased results. When the outcome model is misspecified, the ATE is underestimated (9.18) with a lower standard error. Misspecifying the exposure model also results in bias, though less severe, with a reduced standard error.

Table 6. IPW-estimated ATE, standard error, and 95% confidence interval with correct and misspecified outcome and exposure models.

Model Specification	**Estimated ATE**	**Standard error**	**95% Confidence interval**	
			Lower interval	Upper interval
IPW with correct outcome and exposure model	10.08	0.501	9.10	11.07
IPW with **misspecified** outcome model	9.18	0.358	8.48	9.88
IPW with misspecified **exposure** model	9.99	0.235	9.53	10.45

4.4 Doubly robust estimates with AIPW

The doubly robust AIPW estimator is applied to further address model misspecification. We explore stacking—a form of ensemble learning combining multiple models (e.g., linear models, and random forests)—to mitigate restrictive assumptions. Stacking provides flexibility by reducing reliance on precise model specification.

The results are summarized in Table 7. When both models are correctly specified and stacked with differences in means and GLM, the estimated ATE is 10.09. Stacking with random forests and regularization demonstrate stable ATE estimates, though slightly larger standard errors.

With misspecification, the estimated ATE is biased downward, with a misspecified outcome model causing larger bias than a misspecified exposure model. However, stacking shows resilience against misspecification, maintaining fairly stable estimates despite increased standard errors.

Table 7. AIPW-estimated ATE, standard error, and 95% confidence interval, with stacking and correct and misspecified outcome and exposure models.

Model Specification	**Estimated ATE**	**Standard error**	**95% Confidence interval**	
			Lower interval	Upper interval
Correct (stacked with stacked differences in means and GLM)	10.09	0.069	9.953	10.223
Correct (stacked with stacked difference in means, GLM, and RF)	10.13	0.114	9.903	10.349
Correct (stacked with stacked difference in means, GLM, RF, and GLMnet)	10.00	0.114	9.78	10.226
Misspecified **outcome** (stacked with differences in mean and GLM)	8.77	0.069	8.637	8.909
Misspecified **outcom**e (stacked with differences in mean, GLM, and RF)	8.89	0.111	8.67	9.106
Misspecified **outcome** (stacked with differences in mean, GLM, RF, and GLMnet)	8.87	0.108	8.658	9.080
Misspecified **exposure** (stacked with differences in mean and GLM)	9.95	0.149	9.653	10.238
Misspecified **exposure** (stacked with differences in mean, GLM, and RF)	9.94	0.149	9.649	10.234
Misspecified **exposure** (stacked with differences in mean and GLM, RF, and GLMnet)	9.90	0.149	9.61	10.193

Note: GLM = Generalized Linear Model; GLMnet = Generalized Linear Model with elastic net regularization; RF = random forests.

4.5 Key Takeaways

Both naive and IPW methods demonstrate significant sensitivity to model specification, underscoring the importance of including relevant covariates. Machine learning methods, especially AIPW with regularization and ensemble techniques, exhibit enhanced robustness to model misspecification. Stacking, in particular, reduces dependency on precise parametric forms, providing more reliable estimates.

While the case study uses synthetic data, the findings illustrate the potential value of CML methods, like regularization and ensemble techniques, in SCM. These methods offer robust tools for estimating causal effects, even amidst complex interactions and high-dimensional data. By applying these approaches, practitioners can achieve more accurate causal inferences, aiding strategic decision-making in dynamic environments like supply chain management.

5. Conclusions and Future Directions

CML is a recent scientific innovation that has emerged as a transformative approach for supply chain management, enabling more accurate predictions and deeper understanding into the underlying mechanism and drivers of supply chain performance. Traditional machine learning methods often excel at pattern recognition but struggle to establish causation, a crucial aspect for supply chains where decisions often depend on understanding the causal impact of various strategies. By integrating causal inference techniques into ML, CML provides a more robust framework for predicting how interventions (e.g., changes in inventory policies, supplier adjustments, or transportation optimizations) will affect outcomes, including cost, lead time, and service levels. SCM has experienced and benefited from the application of recent advancements of causal inference techniques, such as propensity score matching, synthetic controls, and instrumental variable approaches, which have been adapted to SCM settings. These methods allow managers to evaluate the impact of interventions across the supply chain, even in the absence of randomized experiments. For example, CML has been applied to assess the effects of quality assurance programs (e.g., Giamattei et al., 2024), analyze demand forecasting methods' impacts on inventory levels (e.g., Wibowo, 2024), determine how transportation mode and order size influence delivery performance (Bo and Xiao, 2024), and analyze the causal mechanism driving supplier delivery delays (Wyrembek et al., 2025).

However, the adoption of CML in SCM is still in its early stages. The dynamic, interdependent nature of supply chains poses challenges regarding data availability, the complexity of confounding factors, and the contextual uniqueness of supply chain interventions. To enhance the robustness and applicability of CML, future research should prioritize the development of models that can handle the high-dimensional and often non-linear nature of SCM data. Advanced methods could improve the precision of causal estimates and better account for the complexities in supply chains. A crucial aspect is that CML holds significant potential to revolutionize the field of SCM by shifting the focus from descriptive and predictive analytics to truly prescriptive analytics based on causal understanding. This approach may help link disparate supply chains to help identify weak points of intersection across spatial and temporal dimensions. Continued research into more adaptable, scalable, and interpretable causal models will likely drive broader adoption in industry, ultimately leading to more resilient, efficient, and responsive supply chains.

6. References

Angrist, J.D., and Krueger, A.B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics*, 106(4), 979–1014. https://doi.org/10.2307/2937954

Angrist, J. D., and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

ALICE. (n.d.). Microsoft Research. Retrieved August 12, 2024, from https://www.microsoft.com/en-us/research/group/alice/

Athey, S., and Imbens, G. W. (2019). Machine Learning Methods That Economists Should Know About. *Annual Review of Economics*, *11*, 685–725. https://doi.org/10.1146/annurev-economics-080217-053433

Athey, S., Tibshirani, J. and Wager, S. (2019). Generalized Random Forests. *The Annals of Statistics*, 47(2), pp. 1148-1178. doi:10.1214/18-AOS1709.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives*, *28*(2), 29–50. https://doi.org/10.1257/jep.28.2.29

Bo, S., & Xiao, M. (2024). Root cause attribution of delivery risks via causal discovery with reinforcement learning. *Algorithms, 17*(11), 498.

Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3), 199–231. https://doi.org/10.1214/ss/1009213726

Breiman, L. (1996). Bagging Predictors. *Machine Learning*, 24(2), pp. 123-140. doi:10.1023/A:1018054314350.

Bodendorf, F., Sauter, M., and Franke, J. (2023). A mixed methods approach to analyze and predict supply disruptions by combining causal inference and deep learning. *International Journal of Production Economics*, *256*, 108708. https://doi.org/10.1016/j.ijpe.2022.108708

Brand, J. E., Zhou, X., and Xie, Y. (2023). Recent Developments in Causal Inference and Machine Learning. *Annual Review of Sociology*, *49*, 81–110. https://doi.org/10.1146/annurev-soc-030420-015345

Burtell, M., and Toner, H. (2024). *The Surprising Power of Next Word Prediction: Large Language Models Explained, Part 1*. Georgetown Center for Security and Emerging Technology. https://cset.georgetown.edu/article/the-surprising-power-of-next-word-prediction-large-languagemodels-explained-part-1/

Carbonneau, R., Laframboise, K., and Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, *184*(3), 1140–1154. https://doi.org/10.1016/j.ejor.2006.12.004

Cheikhrouhou, H. (2023, September 15). *How big tech companies use ML and Causal Inference to make data-driven decisions*. Medium. https://medium.com/@haytham99cheikhrouhou/how-amazon-facebook-netflix-and-spotify-use-causal-inference-to-make-data-driven-decisions-6290cd6d6dec

Chen, J.E., and Hsiang, C.W. (2019). Causal random forests model using instrumental variable quantile regression. *Econometrics*, 7(4), 49. https://doi.org/10.3390/econometrics7040049

Chen, S., and Kasiviswanathan, S. (2020, September 9). *Machine learning tools increase power of hypothesis testing*. Amazon Science. https://www.amazon.science/blog/machine-learning-tools-increase-power-of-hypothesis-testing

Dormady, N. C., Rose, A., Roa-Henriquez, A., and Morin, C. B. (2022). The cost-effectiveness of economic resilience. *International Journal of Production Economics*, *244*, 108371. https://doi.org/10.1016/j.ijpe.2021.108371

Dutta, S., Abielmona, A., Rana, P., Wasan, S., Chowdhary, K., and Anurag, R. (2024). *Implications of AI and generative AI for the future of supply chains*. EY Center for Executive Leadership. https://assets.ey.com/content/dam/ey-sites/ey-com/en_us/topics/coo/ey-implications-of-ai-and-generative-ai-for-the-future-of-supply-chains.pdf

EasyBA.Co. (2023). *Elevate Your Supply Chain Management Strategy with A/B Testing Results Insights.* EasyBA.Co. https://easyba.co/blog/business-analysis/elevate-your-supply-chain-management-strategy-with-a-b-testing-results-insights/

Engin, B., Lan, Y., Tang, G., Segalin, C., Griggs, K., and Iyengar, V. (2023, January 14). *Causal Machine Learning for Creative Insights*. Medium. https://netflixtechblog.com/causal-machine-learning-for-creative-insights-4b0ce22a8a96

Facure, M. (2023). *Causal Inference in Python.* O'Reilly Media, Inc.

Friedman, J.H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5), pp. 1189-1232. doi:10.1214/aos/1013203451.

Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., and Davidian, M. (2011). Doubly Robust Estimation of Causal Effects. *American Journal of Epidemiology*, *173*(7), 761–767. https://doi.org/10.1093/aje/kwq439

Giamattei, L., Guerriero, A., Pietrantuono, R., & Russo, S. (2024). Causal reasoning in Software Quality Assurance: A systematic review. *Information and Software Technology,* 107599.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

Götz, P., and Budhathoki, K. (2022, May 31). *AWS contributes novel causal machine learning algorithms to DoWhy Python library*. Amazon Science. https://www.amazon.science/blog/aws-contributes-novel-causal-machine-learning-algorithms-to-dowhy

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd edn. New York: Springer. doi:10.1007/978-0-387-84858-7.

Helgeson, J. F., and Roa-Henriquez, A. (2022). A Theoretical Framework for Supply Chain Resilience Planning. In Y. Khojasteh, H. Xu, and S. Zolfaghari (Eds.), *Supply Chain Risk Mitigation: Strategies, Methods and Applications* (pp. 159–189). Springer International Publishing. https://doi.org/10.1007/978-3-031-09183-4_8

Hendriksen, C. (2023). Artificial intelligence for supply chain management: Disruptive innovation or innovative disruption? *Journal of Supply Chain Management*, *59*(3), 65–76. https://doi.org/10.1111/jscm.12304

Hernan, M. A., & Robins, J. M. (2024). *Causal inference: What if* (First edition). Taylor and Francis.

Huber, M. (2023). *Causal analysis: Impact evaluation and causal machine learning with applications in R*. The MIT Press.

Imbens, G.W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86(1):4–29.

Imbens, G. W. (2024). Causal Inference in the Social Sciences. *Annual Review of Statistics and Its Application*, *11*, 123–152. https://doi.org/10.1146/annurev-statistics-033121-114601

Imbens, G. W., and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press. https://doi.org/10.1017/CBO9781139025751

Jacobs, F. R., & Chase, R. B. (2024). *Operations and supply chain management* (Seventeenth edition). McGraw Hill.

Liddell, P. (2023, December 11). *Supply chain trends 2024: The digital shake-up - KPMG Global*. KPMG. https://kpmg.com/xx/en/home/insights/2023/12/supply-chain-trends-2024.html

Loske, D., and Klumpp, M. (2022). Verifying the effects of digitalisation in retail logistics: An efficiency-centred approach. *International Journal of Logistics Research and Applications*, *25*(2), 203–227. https://doi.org/10.1080/13675567.2020.1815681

Mechai, N., and Wicaksono, H. (2024). Causal Inference in Supply Chain Management: How Does Ever Given Accident at the Suez Canal Affect the Prices of Shipping Containers? *Procedia Computer Science*, *232*, 3173–3182. https://doi.org/10.1016/j.procs.2024.02.133

Modgil, S., Singh, R. K., and Hannibal, C. (2021). Artificial intelligence for supply chain resilience: Learning from Covid-19. *The International Journal of Logistics Management*, *33*(4), 1246–1268. https://doi.org/10.1108/IJLM-02-2021-0094

Mogstad, M., and Torgovitsky, A. (2018). Identification and Extrapolation of Causal Effects with Instrumental Variables. *Annual Review of Economics*, *10*, 577–613. https://doi.org/10.1146/annurev-economics-101617-041813

Mohri, M., Rostamizadeh, A., and Talwalkar, A. (2018). *Foundations of Machine Learning, second edition*. MIT Press. https://books.google.com?id=dWB9DwAAQBAJ

Ni, D., Xiao, Z., and Lim, M. K. (2020). A systematic review of the research trends of machine learning in supply chain management. *International Journal of Machine Learning and Cybernetics*, *11*(7), 1463–1482. https://doi.org/10.1007/s13042-019-01050-0

Pearson, B. (2021, September 7). *Personalizing Price With AI: How Walmart, Kroger Do It*. Forbes. https://www.forbes.com/sites/bryanpearson/2021/09/07/personalizing-price-with-ai-how-walmart-kroger-do-it/

Robins, J.M., Rotnitzky, A., and Zhao, L.P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866.

Rosenbaum, P.R., and Rubin, D.B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79(387):516–524.

Rudin, C., and Radin, J. (2019). Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From an Explainable AI Competition. *Harvard Data Science Review*, *1*(2). https://doi.org/10.1162/99608f92.5a8a3a3d

Schwarz, P., Schacht, O., Klaassen, S., Grünbaum, D., Imhof, S., and Spindler, M. (2024, June 17). *Management Decisions in Manufacturing using Causal Machine Learning – To Rework, or not to Rework?* https://doi.org/10.48550/arXiv.2406.11308

Simonite, T. (2023). *The WIRED Guide to Artificial Intelligence*. WIRED. https://www.wired.com/story/guide-artificial-intelligence/

Swift, C., Guide, V. D. R., & Muthulingam, S. (2019). Does supply chain visibility affect operating performance? Evidence from conflict minerals disclosures. *Journal of Operations Management*, *65*(5), 406–429. https://doi.org/10.1002/joom.1021

Tay, J. K., Narasimhan, B., and Hastie, T. (2023). Elastic Net Regularization Paths for All Generalized Linear Models. *Journal of Statistical Software*, 106(1), 1–31. https://doi.org/10.18637/jss.v106.i01

Thomas, D., and Fung, J.F. (2022). Measuring downstream supply chain losses due to power disturbances. *Energy Economics*. 114, 106314. https://doi.org/10.1016/j.eneco.2022.106314

Thomas, D., and Helgeson, J. (2021). The effect of natural/human-made hazards on business establishments and their supply chains. *International Journal of Disaster Risk Reduction*. 51(1), 102257. https://doi.org/10.1016/j.ijdrr.2021.102257

Thomas, D., and Helgeson, J. (2022). The Effect of Natural Hazard Damage on Manufacturing Value Added and the Impact of Spatiotemporal Data Variations on the Results. *International Journal of Disaster Risk Science*. 13, 691-704. https://link.springer.com/article/10.1007/s13753-022-00438-x

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tirkolaee, E. B., Sadeghi, S., Mooseloo, F. M., Vandchali, H. R., and Aeini, S. (2021). Application of Machine Learning in Supply Chain Management: A Comprehensive Overview of the Main Areas. *Mathematical Problems in Engineering*, *2021*, 1–14. https://doi.org/10.1155/2021/1476043

Wager, S. (2022). *STATS 361: Causal Inference*. Stanford University Course Notes. https://web.stanford.edu/~swager/stats361.pdf

Wang, G., Li, J., and Hopp, W. J. (2022). An instrumental variable forest approach for detecting heterogeneous treatment effects in observational studies. *Management Science*, 68(5), 3399-3418. https://doi.org/10.1287/mnsc.2021.4084

Wibowo, B. S. (2024). When and how to adjust statistical forecasts in supply chains? Insight from causal machine learning. *Journal of Business Analytics, 7*(1), 25-41.

Wolf, T., and March, C. (2024, June 24). *Decision Intelligence and Causal AI are transforming the future of manufacturing*. New SCMR. https://www.scmr.com/article/decision-intelligence-and-causal-ai-are-transforming-the-future-of-manufacturing

Wolpert, D.H. (1992). Stacked Generalization. *Neural Networks*, 5(2), pp. 241-259. doi:10.1016/S0893-6080(05)80023-1.

Wyrembek, M., Baryannis, G., & Brintrup, A. (2025). Causal machine learning for supply chain risk prediction and intervention planning. *International Journal of Production Research,* 1-20.

Yılmaz, Ö., Son, Y., Shang, G., & Arslan, H. A. (2024). Causal inference under selection on observables in operations management research: Matching methods and synthetic controls. *Journal of Operations Management*, *70*(5), 831–859. https://doi.org/10.1002/joom.1318

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), pp.301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Alfredo Roa-Henriquez: Alfredo is a scholar for the Challey Institute for Global Innovation and Growth and an Assistant Professor of logistics in the Department of Transportation, Logistics, and Finance at North Dakota State University. His research interests include the resilience of businesses and supply chains, organizational economics, and the intersection of behavioral economics and supply chain management.

Juan Fung: Juan is a research economist in the Applied Economics Office of the Engineering Laboratory (EL) at the National Institute of Standards and Technology (NIST). Juan is currently working on selected topics in community resilience, including estimating costs of retrofitting government buildings for seismic events and quantifying a community's indirect benefits from disaster resilience.

Ruhaimatu Abudu: Ruhaimatu is a doctoral student in Transportation and Supply Chain Management at North Dakota State University (NDSU). Her interests include machine learning applications in supply chains resilience.

Jennifer Helgeson: Jennifer is a research economist in the Applied Economics Office of the Engineering Laboratory (EL) at the National Institute of Standards and Technology. She currently leads the office's work on the "Economics of Community Resilience Planning" and is a member of the National Construction Safety Team (NCST). Her research interests are focused around survey assessments, economic analyses that consider behavioral aspects and approaches to dealing with environmental issues, resilience to hazards in the built environment, with consideration for cost-effectiveness of community- scale mitigation and adaptation efforts. She is the developer of the Economic Decision Guide Software (EDGe$) Tool Online which provides ex ante economic evaluation of resilience planning options.

Douglas Thomas: Douglas is an economist in the Applied Economics Office at the National Institute of Standards and Technology. His work focuses on manufacturing industry costs, risks, and economic decision making. His work utilizes input-output analysis, econometrics, and survey techniques to measure costs and risks faced by manufacturing establishments.