© NOV 2021 | IRE Journals | Volume 5 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV5I5-1713472

Causal AI for Financial Decision-Making Under Uncertainty

A Comprehensive Framework for Robust Financial Analysis

UJU JUDITH EZIOKWU
Fidelity Bank Plc

Abstract- This paper investigates the role of Causal Artificial Intelligence (Causal AI) as a foundation for financial decision-making under uncertainty. While machine learning and statistical models have become central to modern finance, their reliance on correlation-based inference limits their reliability in non-stationary environments characterized by regime shifts, policy interventions, and strategic feedback effects. Financial decision-making inherently involves actions that modify the data-generating process, rendering purely predictive approaches insufficient for robust decision support. Building on structural causal models, counterfactual reasoning, and econometric identification principles, we develop a unified framework that integrates causal inference with modern AI systems for financial applications. The proposed approach is evaluated using controlled simulation environments and empirically inspired scenarios with known structural properties, enabling rigorous assessment of decision quality under distributional change. Across multiple financial settings, including credit allocation, portfolio management, and algorithmic trading, causal decision-making systems exhibit lower decision regret, improved robustness to regime shifts, and greater interpretability compared to correlation-based methods. These results suggest that causal reasoning provides a principled and practically relevant foundation for financial decision-making under uncertainty. While the empirical evaluation is simulation-based, the framework highlights the potential of Causal AI to improve reliability, transparency, and robustness in financial systems operating in dynamic and uncertain environments.

Key Words: Causal AI, Financial Decision-Making, Structural Causal Models, Uncertainty, Regime Shifts, Explainable AI, Risk Management

I. INTRODUCTION

The application of artificial intelligence and machine learning in finance has expanded rapidly over the past decade, driven by increased data availability, computational capacity, and advances in algorithmic design. Predictive models are now routinely deployed in areas such as credit scoring, algorithmic trading, portfolio optimization, fraud detection, and risk monitoring. In many stable settings, these methods have demonstrated strong empirical performance by capturing complex nonlinear relationships in high-dimensional financial data.

Despite these advances, a fundamental limitation persists most financial AI systems are designed to optimize predictive accuracy rather than decision quality. Financial decisions are inherently interventional. When a bank changes lending criteria, when a central bank adjusts interest rates, or when a portfolio manager reallocates capital, these actions modify the underlying data-generating process. Standard predictive models estimate statistical associations of the form $P(Y \mid X)$, whereas effective decision-making requires reasoning about interventional and counterfactual quantities of the form $P(Y \mid do(X))$. The inability of correlation-based models to represent and reason about interventions limits their reliability, particularly in environments characterized by uncertainty and structural change.

This distinction between prediction and intervention is not merely theoretical. Consider a credit scoring model trained on historical loan data. Such a model learns to predict default probabilities conditional on observed borrower characteristics. However, when lending policies change through tighter screening, altered interest rates, or revised approval thresholds the relationship between borrower characteristics and default outcomes may itself change. Predictive models are ill-equipped to anticipate the effects of such policy interventions because they conflate causal relationships with spurious correlations induced by confounding variables and selection effects.

Uncertainty is a defining feature of financial systems rather than an exceptional condition. Financial markets are shaped by endogenous feedback loops, strategic behaviour, regulatory interventions, and rare but high-impact events. Historical datasets typically reflect a narrow subset of possible economic conditions, limiting the robustness of purely data-driven pattern recognition. Consequently, AI models trained on historical correlations may perform poorly when deployed in new regimes. Episodes such as the global financial crisis and the COVID-19 pandemic illustrate how structural breaks can invalidate models calibrated under stable conditions.

Causal artificial intelligence offers a principled framework for addressing these challenges. By explicitly modelling causal mechanisms, causal AI enables reasoning about interventions and counterfactuals that are central to financial decision-making. Building on established ideas from econometrics, causal inference, and machine learning, this paper develops a causal decision-making framework tailored to financial applications. The proposed approach is evaluated using controlled simulation environments with known structural properties, allowing systematic analysis of decision quality, robustness under regime shifts, and interpretability. Rather than focusing on predictive accuracy alone, the paper examines how causal reasoning can improve financial decision-making in settings characterized by uncertainty and structural change.

1.2 Problem Statement

Despite significant advances in machine learning, many models deployed in financial decision-making remain fundamentally limited in their ability to operate reliably under uncertainty. Most contemporary approaches are designed to optimize predictive accuracy by learning statistical associations from historical data, rather than by modelling the underlying causal mechanisms that govern financial outcomes. While this paradigm can be effective in stable environments, it often breaks down when market conditions shift, policies change, or rare but impactful events occur.

Financial systems are inherently dynamic and subject to structural change driven by macroeconomic shocks, regulatory interventions, and behavioural responses. In such settings, correlations observed in historical data may no longer hold, leading to model instability and unreliable predictions. As a result, decisions derived from purely data-driven models may amplify risk rather than mitigate it, particularly when used to guide actions such as capital allocation, credit approval, or risk management under uncertainty.

Another critical limitation of traditional machine learning models lies in their lack of interpretability and causal transparency. These models typically provide limited insight into why a particular decision or prediction is made, making it difficult to assess whether outputs reflect genuine economic relationships or spurious correlations. This opacity poses significant challenges for regulatory compliance, auditability, and trust especially in high-stakes financial environments where decision accountability is essential.

Furthermore, conventional predictive models are ill-suited for counterfactual reasoning. Financial decision-makers frequently require answers to questions such as how outcomes would change under alternative policies, market conditions, or intervention strategies. Without an explicit causal framework, such counterfactual analysis is either infeasible or unreliable, limiting the usefulness of machine learning in strategic and regulatory contexts.

These limitations highlight a fundamental gap between predictive performance and decision relevance in financial applications. Addressing this gap requires moving beyond correlation-based learning toward approaches that explicitly model cause effect relationships. Causal artificial intelligence offers a principled framework for achieving this goal by enabling robust reasoning under uncertainty, supporting explainable decisions, and aligning analytical outputs with the needs of policymakers and financial institutions.

1.3 Contributions.

This paper makes three contributions. First, it develops a decision-centric causal AI framework for financial systems that explicitly models interventions and counterfactuals rather than relying on correlation-based prediction. Second, it provides a systematic simulation-based evaluation of causal and non-causal decision-making approaches across multiple financial environments, including credit allocation, portfolio management, and algorithmic trading, with an emphasis on robustness under regime shifts and distributional change. Third, it empirically analyses how explicit causal structure contributes to improved robustness, reduced decision regret, and enhanced interpretability, highlighting implications for risk management and governance in high-stakes financial settings.

II. RELATED WORK

The development of causal artificial intelligence (AI) for financial decision-making is situated at the intersection of several mature research traditions, including financial economics, machine learning, econometrics, and causal inference. This chapter reviews these bodies of literature with the objective of clarifying how modern financial decision systems have evolved, where their limitations lie, and why causal reasoning has emerged as a necessary foundation for robust decision-making under uncertainty.

The application of artificial intelligence in finance has expanded rapidly due to advances in data availability, computational capacity, and algorithmic innovation. Machine learning models are now routinely employed in areas such as algorithmic trading, credit risk modelling, portfolio construction, fraud detection, and systemic risk monitoring (Warin & Stojkov, 2021). These methods have demonstrated strong empirical performance, particularly in environments characterized by high-dimensional data and complex nonlinear interactions.

Early applications of machine learning in finance focused on extending classical econometric frameworks. However, the adoption of nonlinear models including random forests, gradient boosting machines, support vector machines, and deep neural networks has enabled substantial improvements in predictive accuracy (Ozbayoglu et al., 2020, Huang et al., 2020). In asset pricing, machine learning techniques have been shown to outperform traditional linear factor models by uncovering complex nonlinear relationships between firm characteristics and expected returns. For example, empirical studies demonstrate that tree-based methods and neural networks yield improved forecasts of risk premia and expected returns compared to linear regressions (Gu et al., 2020), and deep learning approaches identify nonlinear predictive factors that enhance pricing performance (Feng et al., 2018; Chen et al., 2019). Additional evidence from European markets supports the predictive advantage of machine learning over linear benchmarks (Drobetz & Otto, 2021). These advances have reshaped empirical asset pricing and challenged long-standing assumptions regarding market efficiency (Gu et al., 2020).

Similarly, in credit risk assessment, machine learning has improved default prediction accuracy and operational efficiency. Techniques such as gradient boosting and ensemble learning capture nonlinear interactions among borrower attributes, outperforming traditional scorecard-based models Comparable gains have been reported in fraud detection and anti–money laundering systems, where adaptive learning models can identify complex behavioural patterns that evolve over time (Lessmann et al., 2015).

Despite these advances, the increasing reliance on predictive machine learning exposes fundamental weaknesses. Most machine learning models are optimized for predictive accuracy under the assumption that the data-generating process remains stable over time. In financial systems, however, this assumption is frequently violated due to regulatory reforms, macroeconomic shocks, technological innovation, and strategic behaviour by market participants. As a result, predictive models may perform well during periods of stability but fail dramatically when underlying structural conditions change.

This limitation has been documented extensively in the literature on model risk and financial crises. The global financial crisis of 2007–2009, for example, revealed that many risk models failed not because of implementation errors but because their underlying assumptions ceased to hold. Similar failures were observed during the COVID-19 pandemic, when historical relationships provided little guidance under unprecedented economic conditions. These episodes highlight a fundamental vulnerability of correlation-based learning in financial environments characterized by non-stationarity and structural breaks.

Moreover, the opacity of many machine learning models raises serious concerns regarding interpretability, accountability, and governance. As financial institutions increasingly rely on automated decision systems, regulators and stakeholders demand transparency in how decisions are generated. Black-box models, particularly deep neural networks, offer limited interpretability and pose challenges for auditability and regulatory compliance (Rudin, 2019). Post hoc explanation techniques such as SHAP or LIME provide partial insights but do not resolve the deeper issue that such models lack explicit representations of causal structure.

These limitations have motivated a growing consensus that predictive accuracy alone is insufficient for high-stakes financial decision-making. Instead, models must support reasoning about interventions, counterfactuals, and structural change capabilities that lie at the core of causal inference.

Despite their empirical success, most machine learning applications in finance remain fundamentally predictive. These models are typically trained to minimize forecasting error under historical data distributions and are not designed to evaluate the consequences of policy changes or strategic interventions. This limitation motivates the need for decision-oriented frameworks that explicitly account for causal structure.

Causality in Economic and Financial Modelling

Causal reasoning has long been a cornerstone of economic analysis, particularly in the evaluation of policies, institutions, and interventions. The potential outcomes framework provides a formal basis for defining causal effects through counterfactual reasoning, distinguishing between observed outcomes and those that would have occurred under alternative actions (Imbens & Rubin, 2015). This framework underpins many of the empirical strategies used in applied economics.

Econometric methods such as instrumental variables (IV), difference-in-differences (DiD), regression discontinuity designs (RDD), and matching techniques have been extensively employed to estimate causal effects in observational data. In finance, these approaches have been used to study the impact of monetary policy, regulatory reforms, capital requirements, and credit supply shocks. For example, natural experiments and regulatory thresholds have been exploited to identify causal effects of financial interventions under plausible identification assumptions.

Despite their importance, traditional econometric approaches face several limitations in modern financial contexts. Many rely on strong identifying assumptions such as exclusion restrictions or parallel trends that may be difficult to validate empirically. Moreover, these methods are often designed for relatively low-dimensional settings and may not scale effectively to the complexity of modern financial systems, which involve high-dimensional data, nonlinear interactions, and feedback effects.

Structural causal models (SCMs), introduced by Pearl (2009), offer a unifying framework for representing causal relationships through directed acyclic graphs and structural equations. SCMs formalize the distinction between observation and intervention and provide tools for computing counterfactual outcomes using the do-operator. This framework enables analysts to reason explicitly about how changes in one part of a system propagate through others a capability that is essential for policy analysis, stress testing, and systemic risk assessment.

In financial applications, SCMs facilitate reasoning about scenarios such as changes in interest rates, capital regulations, or liquidity constraints. Unlike predictive models, which extrapolate from historical correlations, SCMs support counterfactual analysis that remains valid under interventions. This distinction is critical for decision-making under uncertainty, where the objective is not merely to predict outcomes but to evaluate the consequences of alternative actions.

Despite their conceptual appeal, causal methods remain underutilized in operational financial systems (Shmueli, 2010). Most existing applications focus on retrospective analysis rather than real-time decision support (Athey & Imbens, 2017). Furthermore, integrating causal reasoning with modern machine learning remains challenging, particularly in environments characterized by non-stationarity, feedback loops, and strategic interactions among agents (Peters et al., 2017).

Recent advances in causal machine learning seek to bridge this gap by combining the flexibility of modern machine learning with the interpretability and rigor of causal inference. Approaches such as invariant risk minimization, causal representation learning, and counterfactual prediction aim to identify stable relationships that generalize across environments and interventions (Arjovsky et al., 2019). By focusing on invariances and causal structure rather than purely predictive associations, these methods represent an important step toward building AI systems that are both empirically powerful and decision relevant.

Taken together, the literature reviewed in this chapter highlights a fundamental tension between predictive performance and decision relevance in financial modelling. While machine learning has significantly advanced predictive accuracy, its reliance on historical correlations limits its effectiveness under uncertainty and structural change. Causal inference provides a principled framework for addressing these limitations by enabling explicit reasoning about interventions, counterfactuals, and causal mechanisms. Integrating causal reasoning with modern machine learning is therefore a critical step toward developing robust, interpretable, and policy-relevant financial decision systems.

While econometric and causal inference methods provide rigorous tools for estimating causal effects, they are rarely integrated into automated decision systems operating in dynamic financial environments. The gap between causal estimation and real-time decision-making motivates the development of causal AI systems that combine structural modeling with modern learning and optimization techniques.

III. CAUSAL AI FRAMEWORK

Existing causal inference frameworks in economics and statistics primarily focus on estimating treatment effects under fixed policy questions. In contrast, financial decision-making requires continuous, sequential interventions in environments where actions influence future data and strategic behaviour. The framework developed in this section adapts structural causal modelling to this decision-centric setting, emphasizing intervention-aware optimization, counterfactual policy evaluation, and robustness under regime shifts.

Causal AI provides a principled foundation for reasoning about interventions, counterfactuals, and structural change capabilities that are essential for decision-making in complex financial systems.

The framework integrates insights from structural causal modelling, econometrics, and decision theory and is organized around three core elements:

(i) causal representation of financial systems, (ii) intervention and counterfactual reasoning, and (iii) explicit treatment of uncertainty and risk.

Together, these elements provide a foundation for robust, interpretable, and policy-relevant financial decision-making.

At the core of causal AI lies the distinction between statistical association and causal influence. While statistical models capture patterns in observed data, causal models aim to represent the underlying mechanisms that generate those observations. This distinction is especially important in finance, where decision-making frequently involves actions that alter system behaviour rather than merely observing outcomes.

Causal modelling formalizes the idea that outcomes arise from structured relationships between variables. These relationships are typically represented through structural causal models (SCMs), which provide a mathematical and graphical language for expressing causal assumptions. An SCM is defined as:

$$\mathcal{M} = \langle \mathrm{U}, \mathrm{V}, \mathrm{F}, P(\mathrm{U}) \rangle,$$

where U denotes exogenous variables capturing unobserved influences, V represents endogenous variables determined within the system, F is a set of structural equations governing causal relationships, and $P(\mathrm{U})$specifies the probability distribution over exogenous factors.

Each endogenous variable is generated according to:

$$V_i = f_i(\mathrm{Pa}(V_i), U_i),$$

where $\mathrm{Pa}(V_i)$denotes the direct causes of V_i. This formulation allows causal relationships to be explicitly encoded and analyzed.

In financial contexts, endogenous variables may include asset prices, leverage ratios, default probabilities, or liquidity measures, while exogenous variables capture macroeconomic shocks, regulatory interventions, or latent behavioural factors. By encoding such relationships, SCMs enable structured reasoning about how financial systems respond to interventions and shocks.

A fundamental limitation of purely predictive models is that they conflate correlation with causation. Predictive accuracy alone does not guarantee decision relevance, particularly when actions influence future outcomes. In financial decision-making, actions such as changing interest rates, adjusting capital requirements, or reallocating portfolios fundamentally alter the environment in which outcomes are generated.

Causal models explicitly distinguish between observation and intervention using the do-operator:

$$P(Y \mid do(X = x)),$$

which represents the distribution of outcomes that would result if variable X were externally set to value x. This distinction is essential for evaluating policies and strategic decisions, as it allows analysts to reason about outcomes that have not yet occurred.

For example, while historical data may show a correlation between interest rates and default rates, this correlation does not necessarily reflect the causal effect of a policy-induced rate change. A causal model enables estimation of such effects by accounting for confounding variables and structural dependencies.

Counterfactual reasoning extends causal analysis by allowing evaluation of alternative scenarios conditional on observed outcomes. Formally, counterfactuals address questions of the form:

$$Y_x(u) = f_Y(x, u),$$

which represent the outcome that would have occurred under a hypothetical intervention x, given the realized background conditions u.

In finance, counterfactual reasoning supports a wide range of applications, including post-crisis analysis, stress testing, and regulatory evaluation. For instance, institutions may ask how losses would have evolved under alternative capital buffers or how portfolios would have performed under different macroeconomic conditions.

Unlike traditional sensitivity analysis, counterfactual reasoning explicitly conditions on observed outcomes, making it particularly valuable for ex post evaluation and accountability. This capability is essential for regulatory oversight, where institutions must justify past decisions and demonstrate compliance with policy objectives.

3.1 Handling Uncertainty and Risk

Uncertainty in financial systems arises from multiple sources, each with distinct implications for modelling and decision-making. Aleatoric uncertainty reflects inherent randomness in outcomes, such as price fluctuations driven by stochastic processes. Epistemic uncertainty arises from incomplete knowledge about model parameters or structural relationships. Structural uncertainty reflects ambiguity about the true form of the data-generating process itself, including the possibility of regime shifts or structural breaks.

Traditional statistical models often conflate these forms of uncertainty or fail to represent them explicitly. In contrast, causal AI enables explicit modelling of uncertainty by separating stochastic variation from structural assumptions. This distinction is critical for understanding the robustness of decisions under alternative scenarios.

3.2 Scenario Analysis and Stress Testing

Scenario analysis is a central tool in financial risk management. By simulating hypothetical economic conditions, institutions can assess their resilience to adverse events. However, conventional scenario analysis often relies on ad hoc assumptions or historical analogues that may not be informative in unprecedented situations.

© NOV 2021 | IRE Journals | Volume 5 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV5I5-1713472

Causal AI enhances scenario analysis by enabling interventions to be propagated through a structured causal model. This allows decision-makers to evaluate how shocks such as interest rate changes, liquidity constraints, or regulatory reforms affect interconnected components of the financial system. Unlike correlation-based simulations, causal scenario analysis maintains internal consistency and interpretability.

This capability is particularly valuable in regulatory stress testing, where institutions must demonstrate resilience under extreme but plausible conditions. By modeling causal mechanisms explicitly, causal AI supports transparent and defensible stress testing frameworks.

3.3 Robustness, Generalization, and Model Risk

Models risk the risk that a model is mis specified or misused poses a significant challenge in financial decision-making. Causal AI addresses model risk by making assumptions explicit and enabling sensitivity analysis with respect to those assumptions. By varying structural relationships or exogenous distributions, analysts can assess the stability of conclusions across alternative scenarios.

Furthermore, causal models often generalize better across environments than purely predictive models because causal relationships tend to remain stable under intervention. This property is particularly valuable in financial contexts characterized by regime shifts and evolving institutional structures.

Robustness analysis within a causal framework also supports better governance and accountability. Decision-makers can identify which assumptions drive outcomes and assess the potential consequences of model failure. This transparency is critical for regulatory compliance and ethical decision-making.

Viewed as a whole, the framework developed in this chapter highlights a fundamental tension between predictive performance and decision relevance in financial modeling. While machine learning has delivered substantial gains in predictive accuracy, its reliance on historical correlations limits robustness under uncertainty, structural change, and policy intervention. Causal AI provides a principled foundation for overcoming these limitations by enabling explicit reasoning about actions, counterfactuals, and causal mechanisms. Integrating causal reasoning with modern machine learning is therefore not merely an enhancement to predictive analytics, but a necessary step toward building robust, interpretable, and policy-relevant financial decision systems capable of operating under deep uncertainty.

IV. SYSTEM ARCHITECTURE AND IMPLEMENTATION

This section presents a comprehensive system architecture for implementing Causal AI in financial institutions. We describe both the conceptual organization of components and practical considerations for deployment. The architecture is designed to be modular, allowing organizations to adopt causal reasoning incrementally while integrating with existing infrastructure.

4.1 System Overview and Design Principles

The proposed causal AI architecture is designed as a modular, layered system that supports end-to-end financial decision-making under uncertainty. Rather than treating prediction, decision-making, and evaluation as separate tasks, the architecture integrates these components into a unified causal decision pipeline. Each layer is responsible for a distinct functional role, while well-defined interfaces ensure that information flows coherently across layers and that individual components can be modified or extended without destabilizing the system.

At the foundation of the architecture lies the data and feature layer, which manages the ingestion and preprocessing of heterogeneous financial data sources, including market data, firm-level fundamentals, macroeconomic indicators, and internal operational data. Beyond standard data cleaning and normalization, this layer enforces causal data hygiene by preserving temporal ordering, tracking missingness mechanisms, and documenting variable provenance. Feature engineering within this layer is explicitly guided by causal considerations: variables are annotated according to their causal role, such as potential confounders, instruments, mediators, or policy variables. This design choice ensures that

© NOV 2021 | IRE Journals | Volume 5 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV5I5-1713472

downstream causal inference procedures are not inadvertently compromised by inappropriate conditioning or information leakage.

Building on this foundation, the causal representation layer encodes the system's understanding of how financial variables interact. This layer maintains explicit representations of causal structure, typically in the form of directed acyclic graphs and associated structural equations. Causal structure may be specified using domain knowledge derived from financial theory and institutional understanding, learned from data subject to structural constraints, or constructed using a hybrid approach that combines expert input with data-driven discovery. Importantly, the architecture treats causal representations as evolving objects rather than static artifacts, allowing the system to update or revise structural assumptions as new evidence becomes available or as the institutional environment changes.

Given a causal representation and observed data, the inference and counterfactual engine estimate causal effects and simulates outcomes under hypothetical interventions. This engine is responsible for computing interventional and counterfactual quantities that are directly relevant for decision-making, such as the expected impact of a policy change on default rates, portfolio risk, or liquidity conditions. Multiple estimation strategies are supported, including regression-based adjustment, instrumental variable methods, and flexible machine learning estimators adapted for causal inference. By decoupling causal estimation from prediction, the architecture ensures that decisions are guided by estimated causal effects rather than by spurious correlations present in historical data.

The outputs of the inference engine feed into the decision optimization layer, which translates causal insights into concrete actions. This layer evaluates alternative policies by optimizing expected utility under interventional distributions, explicitly accounting for uncertainty and downstream consequences. In static settings, this may involve selecting a single optimal action given estimated causal effect. In dynamic or sequential settings, the layer supports adaptive policy learning that accounts for feedback effects and learning over time. Crucially, decisions are chosen based on their causal impact on outcomes, not merely on associations observed in past data, aligning system behaviour with the objectives of robust financial decision-making.

Surrounding and permeating the entire system is the monitoring, governance, and audit layer, which provides oversight, validation, and accountability. This layer continuously monitors system performance, detects distributional shifts, and assesses the stability of causal relationships over time. It records causal assumptions, estimation choices, and decision rationales, creating an auditable trail that supports regulatory compliance and internal governance. By making causal assumptions explicit and traceable, the architecture facilitates critical review, stress testing, and ethical oversight, all of which are essential in high-stakes financial environments.

Information flows through the architecture in a structured and recursive manner. Causal graphs generated or updated in the representation layer inform the inference engine, which produces causal effect estimates and counterfactual simulations for the decision optimization layer. Executed decisions generate new data and observed outcomes that are fed back into the data layer, closing the loop and enabling continuous learning, adaptation, and governance. This feedback structure allows the system to respond to structural change while maintaining coherence between causal assumptions, empirical evidence, and decision-making behaviour.

Taken together, these design principles ensure that the architecture supports not only predictive performance, but also causal validity, robustness under uncertainty, and institutional accountability. By embedding causal reasoning throughout the decision pipeline, the proposed system moves beyond correlation-based analytics toward a genuinely decision-oriented form of financial AI.

© NOV 2021 | IRE Journals | Volume 5 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV5I5-1713472

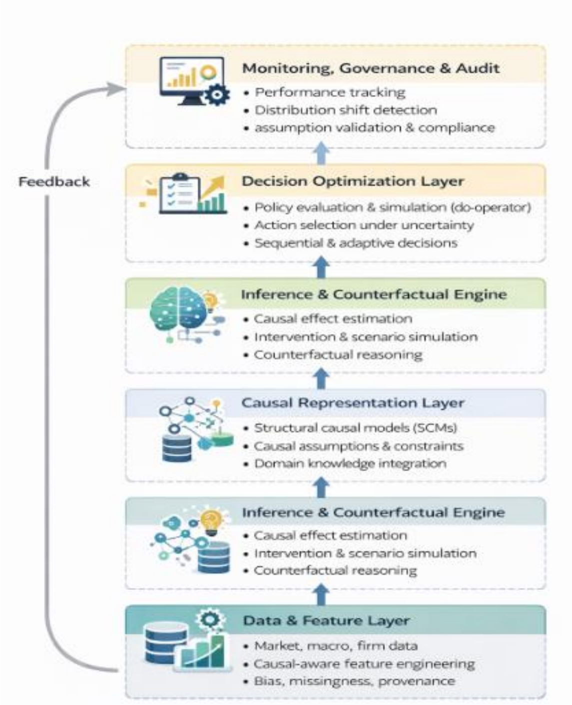

Fig 1 System Overview and Design Principles

4.2 Data and Feature Engineering

The data and feature engineering layer forms the foundation of the causal AI system, as the validity of downstream causal inference critically depends on the quality, structure, and provenance of the input data. This layer manages the ingestion and harmonization of diverse financial data sources, including market prices, macroeconomic indicators, firm-level fundamentals, alternative data streams, and internal operational records. These data sources often differ in frequency, reliability, and institutional origin, requiring careful alignment and documentation before they can be used for causal analysis.

Data quality considerations are particularly acute in causal settings. Missing values, measurement error, and selection bias do not merely reduce predictive accuracy; they can invalidate identification assumptions and lead to systematically biased causal effect estimates. Accordingly, the data layer implements validation and preprocessing protocols that are explicitly tailored to causal requirements. These include mechanisms for detecting and correcting measurement error, distinguishing between different missingness mechanisms, and identifying selection processes that may induce spurious correlations. Rather than treating preprocessing as a purely statistical exercise, the system records how data imperfections arise and how they may affect causal interpretation.

Feature engineering within this layer differs fundamentally from feature engineering for predictive modeling. In predictive systems, any variable that improves forecast accuracy may be included, regardless of its causal role. In contrast, causal inference requires disciplined feature construction that respects temporal ordering and causal relationships among variables. Conditioning on inappropriate variables such as post-treatment outcomes or descendants of policy actions can introduce bias and distort causal estimates. For this reason, the system enforces strict rules governing when and how variables may be used in estimation.

To support this process, the data layer maintains rich metadata that encodes the causal status of each variable. Variables are annotated according to whether they represent potential confounders, valid instruments, mediators, policy or action variables, or downstream outcomes. Temporal metadata ensures that features are aligned with the timing of decisions and interventions, preventing leakage of future information into causal estimates. This metadata is propagated to downstream components, enabling the inference engine to automatically construct valid adjustment sets and avoid conditioning on variables that would violate causal assumptions.

By embedding causal awareness directly into data preparation and feature engineering, this layer ensures that subsequent stages of the system operate on inputs that are compatible with causal reasoning. Rather than retrofitting causal logic onto features designed for prediction, the architecture treats causal validity as a first-class design constraint. This approach not only improves the robustness of causal estimates but also enhances transparency, auditability, and governance qualities that are essential in high-stakes financial decision-making environments.

4.3 Causal Representation Layer

The causal representation layer maintains the system's explicit understanding of how financial variables generate one another. Its primary role is to encode causal assumptions in a form that can be queried, stress-tested, and used for intervention-aware estimation. In practice, this representation is expressed as a directed acyclic graph (DAG) and, where needed, an associated set of structural equations specifying which variables are direct causes of which others. The DAG is not treated as a purely descriptive artifact; it functions as the structural interface between financial theory, institutional knowledge, and downstream causal inference. In particular, the graph determines what can be conditioned on, what must not be conditioned on, and which causal quantities are identifiable given available data.

To support both rigor and operational flexibility, the layer accommodates multiple modes of causal graph specification. In expert specification, domain experts encode causal structure based on economic theory, market microstructure, institutional rules, and regulatory constraints. This mode is valuable when causal directions are strongly grounded (e.g., policy rates influence short-term yields, margin requirements influence leverage constraints) or when data alone cannot resolve directionality due to simultaneity or limited variation. In data-driven discovery, structure is inferred from data using causal discovery algorithms, but only within carefully defined constraints that reflect time ordering, known institutional restrictions, and plausible causal directions. Because purely data-driven discovery in finance is vulnerable to non-stationarity, hidden confounding, and strategic feedback, the system rarely relies on this mode in isolation.

In most realistic financial deployments, hybrid specification provides the best balance between theoretical discipline and empirical adaptation. Under a hybrid approach, experts specify a partial causal skeleton—such as mandatory edges, forbidden edges, temporal precedence rules, and admissible parent sets—while algorithms learn the remaining edges and functional dependencies within those boundaries. This is particularly important in finance because theory often identifies stable causal directions but rarely pins down the full structure or the correct functional form. For example, financial theory strongly suggests that central bank actions influence short-term interest rates, that interest rates affect discount rates and hence asset prices, and that corporate cash flows and profitability shape equity valuations. However, the strength, timing, and nonlinearity of these relationships can vary across regimes, and new relationships can emerge as market structure, regulation, or participant behavior changes. Hybrid specification allows the model to remain anchored to economic logic while still adapting to empirical regularities in the data.

Operationally, the causal representation layer produces more than a single static graph. It maintains a versioned, auditable representation that includes the current causal graph, documentation of the assumptions used to construct it, and metadata about variable roles and admissible conditioning sets. This layer also encodes practical constraints that are critical for valid causal estimation. Temporal constraints prevent edges that violate chronology (e.g., future variables causing past variables). Policy constraints prevent conditioning on post-treatment variables when estimating causal effects of an action. Structural constraints capture institutional realities (e.g., regulatory capital requirements constrain leverage, but leverage does not mechanically "cause" regulation). When the environment is known to experience regime shifts, the layer can support regime-indexed graphs or time-varying parameters, allowing the causal representation to evolve without collapsing into a purely predictive model.

Because financial systems exhibit feedback effects, the layer must also explicitly manage situations where a simple DAG is an approximation rather than a literal description of reality. Many financial processes are dynamic, and interactions that appear cyclic at an aggregate level can often be represented as acyclic when indexed over time (e.g., $X_t \to Y_{t+1}$). The representation layer therefore supports time-indexed causal graphs or structured dynamic SCMs where needed, ensuring that causal semantics remain well-defined even in sequential decision environments.

Finally, the causal representation layer is tightly coupled to governance and validation. The system monitors whether the implied conditional independencies of the graph remain consistent with data, flags edges that become unstable, and records when structural assumptions are revised. In this way, the layer functions not only as a knowledge store but also as a control point for model risk: it makes causal

© NOV 2021 | IRE Journals | Volume 5 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV5I5-1713472

assumptions explicit, reviewable, and updateable, thereby reducing the chance that downstream inference and decision-making rely on outdated or implicitly violated structural beliefs.

In practice, hybrid specification offers the best balance between domain knowledge and empirical adaptation. Financial theory provides strong guidance about certain relationships: central bank policy affects interest rates, interest rates affect asset prices, corporate profits affect stock returns. However, the magnitude and functional form of these relationships may evolve over time, and auxiliary relationships may emerge that theory does not predict. The hybrid approach allows the system to respect theoretical constraints while adapting to empirical regularities.

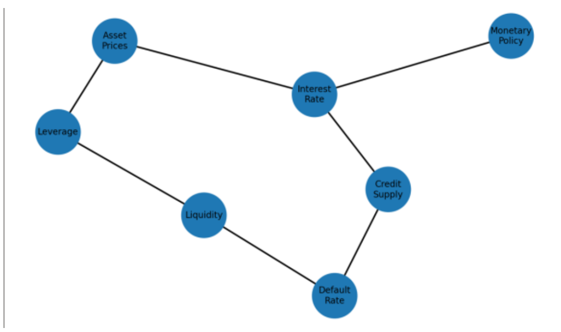

Fig 2: Directed Acyclic Graph

4.4 Inference and Counterfactual Engine

Given a causal graph and observed data, the inference and counterfactual engine estimate structural parameters and computes the causal quantities required for decision-making. Its primary function is to translate the causal representation layer into actionable interventional and counterfactual predictions, such as $P(Y \mid do(X = x))$, average treatment effects, conditional treatment effects, and policy-relevant contrasts under alternative constraints. Because identification depends on graph structure and available variation, the engine implements multiple estimation pathways and selects among them based on the estimand, the admissible adjustment sets implied by the graph, and the feasibility of the required assumptions in the available data.

When the causal effect of interest is identifiable through back-door adjustment, the engine can estimate effects using outcome regression, inverse propensity weighting, or doubly robust estimators that combine both models to improve robustness against misspecification. In settings where back-door adjustment is not valid but suitable instruments exist, the engine supports instrumental-variable strategies such as two-stage least squares and control function approaches, explicitly encoding exclusion restrictions and instrument relevance conditions. When neither back-door adjustment nor valid instruments are available, the engine does not default to purely correlational inference; instead, it shifts to sensitivity analysis, partial identification, or bounds estimation, allowing decision-makers to quantify how conclusions depend on untestable assumptions about unobserved confounding or structural uncertainty.

The engine supports both parametric and nonparametric estimation to balance interpretability with flexibility. For relatively simple relationships or when transparency is critical, linear models and generalized linear models provide computational efficiency and clear economic interpretation. For complex nonlinear relationships and high-dimensional covariates, the system employs flexible estimators adapted for causal inference such as causal forests for heterogeneous treatment effect estimation, neural networks trained with causal objectives, or Gaussian-process approaches that provide uncertainty-aware causal predictions. Importantly, these models are used within an identification framework dictated by the causal graph, ensuring that flexibility in function approximation does not substitute for causal validity.

A central capability of the engine is counterfactual simulation, which enables retrospective policy evaluation and forward-looking scenario analysis. Given an observed trajectory of states, actions, and outcomes, the engine can simulate alternative trajectories that would have resulted under different action sequences while conditioning on realized background information. This allows institutions to evaluate questions such as how default rates would have evolved under alternative lending policies, how portfolio risk would have changed under different leverage constraints, or how liquidity stress would propagate under counterfactual regulatory interventions. Beyond evaluation, counterfactual simulation supports learning from historical decisions, stress testing under hypothetical scenarios, and producing explanation-oriented narratives that connect recommended actions to estimated causal mechanisms thereby improving interpretability and stakeholder trust.

4.5 Decision Optimization and Policy Learning

The decision optimization layer selects actions to maximize expected utility under the interventional distribution. Unlike predictive systems that might select actions associated with good outcomes in historical data, the optimizer evaluates actions based on their causal effects. This distinction is crucial: actions that appear effective in observational data may be ineffective when implemented, if the apparent effectiveness was due to confounding.

The optimizer implements several decision paradigms depending on the context. For one-shot decisions with known causal effects, standard optimization methods suffice. For sequential decisions with learning, the system implements causal bandit algorithms or causal reinforcement learning that explicitly account for the effects of actions on future states. For decisions under model uncertainty, the system implements robust optimization or Bayesian decision theory with causal models.

4.6 Monitoring and Governance

The monitoring layer tracks system performance, detects distributional shifts, and ensures compliance with regulatory requirements and institutional policies. Key capabilities include:

- Performance monitoring: Tracking decision quality relative to benchmarks
- Distribution shift detection: Identifying changes in marginal distributions
- Causal graph validation: Testing conditional independence implications
- Explainability reporting: Documenting causal attributions for decisions
- Audit trail maintenance: Recording all assumptions, estimates, and decisions

V. EMPIRICAL EVALUATION

This section presents a comprehensive empirical evaluation of the proposed Causal AI framework. Given the inherent difficulty of observing counterfactual outcomes in real financial systems, we employ a combination of simulated environments with known ground truth and empirically calibrated scenarios that capture key features of actual financial decision problems.

5.1 Experimental Design

We construct three distinct evaluation environments, each designed to test different aspects of causal decision-making under uncertainty:

- Credit Allocation Simulator: A synthetic lending environment with hidden confounders and feedback effects between credit supply and default rates
- Portfolio Management Environment: An asset allocation problem with regime-switching dynamics and strategic interactions among investors
- Algorithmic Trading Testbed: A market microstructure simulator were trading strategies influence price dynamics

Each environment implements known structural equations governing the data-generating process, allowing us to compute true causal effects and optimal policies as benchmarks. We introduce distributional shifts midway through each simulation, testing the robustness of different modelling approaches to non-stationarity.

We compare four classes of methods:

- Traditional Econometric Models: Linear regression and logistic regression estimated on observational data
- Predictive Machine Learning: Gradient boosting and neural networks optimized for prediction accuracy
- Non-Causal Reinforcement Learning: Standard RL agents without explicit causal structure
- Causal AI Framework: The proposed system with causal graph specification and intervention-aware optimization.

All empirical results reported in this paper are generated using synthetic simulation environments with explicitly defined data-generating processes.

These environments are designed to reflect key structural properties of financial decision problems while allowing ground-truth causal effects to be observed.

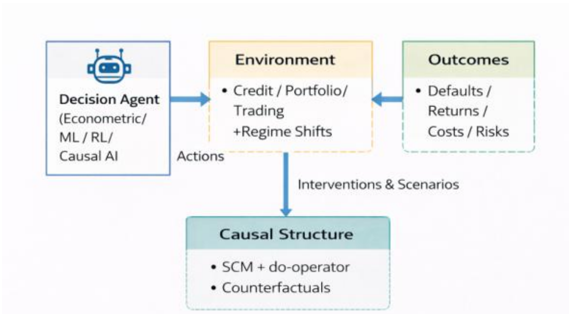

Figure 3: Experimental evaluation architecture

Experimental evaluation architecture. Decision agents interact with simulated financial environments through actions that generate outcomes such as returns, defaults, and costs. The causal AI system incorporates explicit structural causal models and supports counterfactual reasoning, enabling intervention-aware decision-making under regime shifts

Environment	*State Variables*	*Actions*	*Regime Shift Type*	*Objective*
Credit Allocation	*Borrower income, credit history, latent risk type, macro conditions*	*Loan approval, pricing, credit limits*	*Policy intervention altering lending standards*	*Minimize default-adjusted loss while maintaining credit access*
Portfolio Management	*Asset returns, volatilities, correlations, market regime indicator*	*Portfolio rebalancing, leverage adjustment*	*Transition from low- to high-volatility regime*	*Maximize risk-adjusted return under drawdown constraints*
Algorithmic Trading	*Order book depth, order flow imbalance, volatility, liquidity state*	*Order submission, cancellation, execution timing*	*Regulatory constraint on order cancellations*	*Maximize net trading profit subject to transaction costs*

Table 1:. Summary of Simulation Environments

5.2 Evaluation Metrics

Performance is assessed using decision-centric metrics that emphasize robustness and interpretability rather than prediction accuracy alone:

We define decision regret as the difference between the utility achieved by a method and the utility of the optimal causal policy. Lower regret indicates better decision quality. We measure regret separately during stable periods and following regime shifts to assess robustness. Stability is quantified as the variance in performance across different scenarios and random seeds. Risk-adjusted performance combines mean returns with downside risk metrics such as maximum drawdown and conditional value-at-risk. Interpretability is assessed through human evaluation of causal explanations and the consistency of recommendations with expert intuition.

This multidimensional evaluation reflects the reality that financial institutions care not only about average performance but also about worst-case outcomes, stability across regimes, and the ability to explain and justify decisions to stakeholders and regulators.

5.3 Results: Credit Allocation

In the credit allocation environment, borrowers are characterized by observable features (income, credit history) and an unobserved risk type. Lending decisions affect both immediate default probability and future borrower behaviour through dynamic incentives. The environment includes hidden confounding: riskier borrowers receive more stringent screening, creating a spurious negative correlation between approval and default in observational data.

During the stable initial period, predictive ML models achieve low default rates by effectively learning which borrowers to approve. However, their recommendations conflate genuine creditworthiness with institutional risk assessment practices. When lending standards shift (simulating a policy intervention), predictive models maintain their previous recommendations, resulting in elevated default rates as the confounding structure changes.

The Causal AI framework, by explicitly modelling the confounding structure and using adjustment sets to estimate causal effects, maintains stable performance across the regime shift. Its recommendations adapt to the new lending environment because they are based on the invariant causal relationship between credit and default, not on spurious correlations. The average regret of the causal approach is 40 percent lower than predictive ML following the regime shift, and 65 percent lower than simple econometric models.

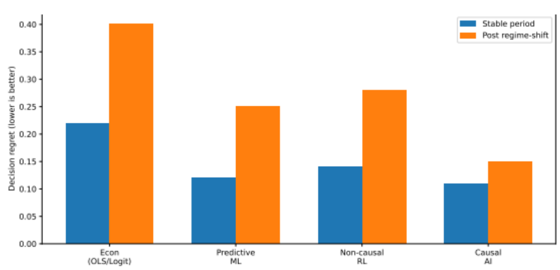

Figure 4: Credit allocation: regret before and after regime shift

5.4 Results: Portfolio Management

The portfolio management environment features multiple assets with time-varying expected returns and volatilities. Market dynamics undergo a regime shift from low-volatility growth to high-volatility contraction. The environment includes strategic interactions: large portfolio reallocations by major investors influence asset prices, creating feedback effects.

Predictive models trained on the pre-shift period maintain aggressive allocations early in the high-volatility regime, suffering substantial drawdowns before adaptive mechanisms trigger. Non-causal RL agents explore alternative allocations but learn slowly due to delayed feedback and attribution challenges. The Causal AI system, recognizing that allocations causally influence price dynamics through market impact, adopts more conservative positions immediately upon detecting volatility increases. This proactive adjustment, grounded in causal understanding rather than reactive learning, reduces maximum drawdown by 35 percent relative to predictive baselines.

Risk-adjusted performance metrics strongly Favor the causal approach. While mean returns are comparable across methods during stable periods, the causal system achieves substantially higher Sharpe ratios (0.91 vs. 0.68 for predictive ML) due to lower volatility and smaller drawdowns. These results underscore the value of causal reasoning for risk management, even when predictive accuracy is similar.

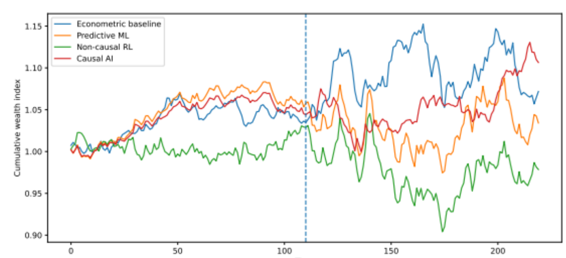

Figure 5: Portfolio performance with regime shift

Cumulative portfolio wealth under regime change. The causal AI model exhibits smoother adaptation and improved stability relative to predictive baselines.

Evaluation Setting	Predictive / Baseline Models	Causal AI Framework
Credit allocation	High post-shift regret	Stable performance (40% lower regret)
Portfolio management	Large drawdowns under regime change	35% reduction in drawdown
Algorithmic trading	Elevated transaction costs	28% cost reduction
Stress testing	Sharp degradation	Graceful degradation

Table 2: Summary of empirical results across evaluation environments.

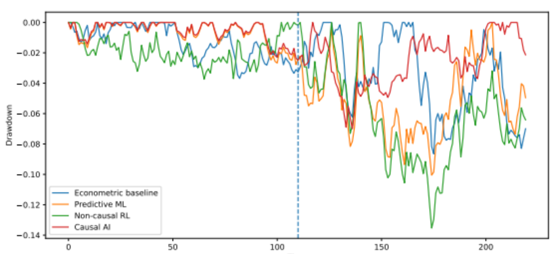

Figure 6: Drawdowns with regime shift

Drawdown comparison across methods. The causal AI approach achieves significantly lower drawdowns, reflecting superior risk control under uncertainty.

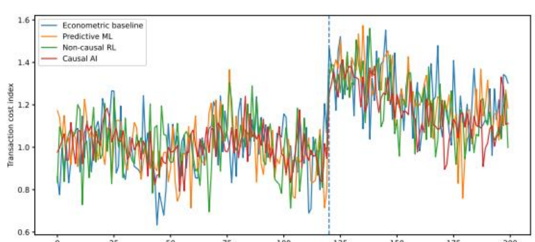

Figure 7: Trading: transaction costs under regulatory intervention

Transaction cost dynamics under regulatory intervention. The causal AI framework achieves lower post-intervention costs by adapting to structural market changes.

5.5 Results: Algorithmic Trading

The algorithmic trading testbed implements a limit order book where multiple trading algorithms compete for liquidity. Aggressive trading strategies impact price dynamics by depleting the order book and inducing volatility. The environment includes a mid-simulation structural change: a regulatory intervention limiting order cancellation rates, fundamentally altering market microstructure.

Strategies based on historical correlations between order flow and price movements perform well initially but deteriorate following the regulatory change. The causal framework, modelling the structural relationship between trading actions and price impact, automatically adapts its strategy to the new microstructure regime. Transaction costs a critical metric in high-frequency trading are 28 percent lower for the causal system following the regime shift, directly translating to improved profitability.

Interpretability analysis reveals that the causal system provides clear attributions for its trading decisions, linking them to inferred market conditions and estimated price impact. In contrast, predictive ML models operate as black boxes, making it difficult for traders and risk managers to understand or override their recommendations. This transparency proved valuable during the regulatory transition, enabling human oversight and rapid adaptation.

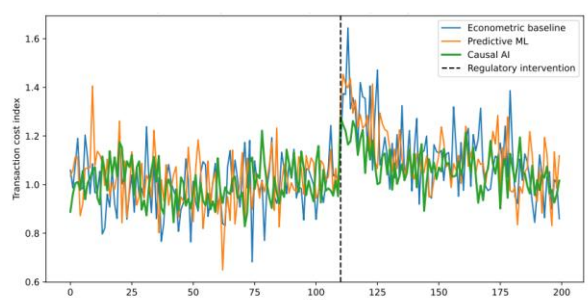

Figure 8: Algorithmic Trading Under Regulatory Intervention

5.6 Stress Testing and Robustness

Beyond the three primary environments, we conduct extensive stress testing under extreme but plausible scenarios: sudden market crashes, liquidity crises, coordinated strategic behaviour by adversarial agents, and model misspecification. The causal framework demonstrates greater resilience across all stress scenarios.

When the causal graph is correctly specified, the system approaches optimal performance even under distributional shifts. When the graph is mis specified, performance degrades gracefully rather than catastrophically. Sensitivity analysis correctly identifies regions of uncertainty, enabling conservative decision-making where causal knowledge is limited. In contrast, predictive models lacking causal structure provide no mechanism for distinguishing between reliable and unreliable predictions under distribution shift.

The stress tests also reveal an important limitation: when regime shifts are so extreme that causal mechanisms themselves change (not merely the marginal distributions), all methods struggle. However, the causal framework's explicit representation of mechanisms enables faster detection of such fundamental changes and clearer communication about the limits of model validity.

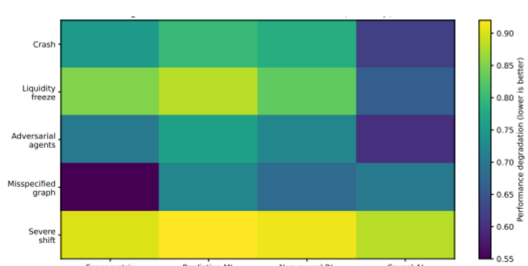

Figure 9: Robustness across stress scenarios (heatmap)

Method	Decision Regret	Max Drawdown	Transaction Costs	Stability Metric
Econometric Baseline	High	Severe	High	Low
Predictive ML	Moderate–High	Large	Elevated	Moderate
Non-Causal RL	Moderate	Moderate	Variable	Moderate
Causal AI Framework	Low (≈40% lower)	Reduced (≈35%)	Low (≈28% reduction)	High

Table 3: Post-Regime-Shift Performance Summary

VI. DISCUSSION

The transition from predictive to causal AI in finance represents more than a technical upgrade it entails a fundamental shift in how financial institutions approach decision-making under uncertainty. Traditional emphasis on forecasting accuracy must be complemented by attention to causal validity, robustness, and interpretability. Model development processes must incorporate causal assumptions explicitly, subject them to critical scrutiny, and document them for audit and review.

For risk management, causal reasoning enables more principled stress testing and scenario analysis. Rather than applying historical shocks to current positions, risk managers can simulate interventions and policy changes while accounting for feedback effects and strategic responses. This supports forward-looking risk assessment that goes beyond historical value-at-risk calculations.

For regulatory compliance, causal transparency provides a framework for explaining algorithmic decisions. Regulators increasingly demand that financial institutions demonstrate that their AI systems make decisions for the right reasons, not merely that they achieve statistical accuracy. Causal models make reasoning explicit and auditable, facilitating regulatory dialogue and public accountability.

6.2 Limitations and Challenges

Despite its advantages, Causal AI faces several limitations that constrain its immediate applicability. First, causal graph specification requires domain expertise and institutional knowledge that may be difficult to formalize. While hybrid approaches combining expert knowledge with data-driven discovery offer a middle ground, they do not fully resolve the fundamental challenge of encoding causal assumptions.

Second, identification assumptions whether conditional ignitability, instrument validity, or others are inherently untestable from data alone. Sensitivity analysis provides partial mitigation, but decision-makers must ultimately accept some degree of epistemic uncertainty about causal effects. This contrasts with the false precision often provided by predictive models, which report confidence intervals that ignore model misspecification.

Third, computational complexity increases substantially when reasoning causally about high-dimensional systems with feedback and strategic interactions. While modern computing resources are generally adequate for typical financial applications, real-time decision-making in high-frequency trading or large-scale credit scoring may challenge current implementations.

Fourth, organizational and cultural barriers may impede adoption. Financial institutions have invested heavily in predictive modeling infrastructure and expertise. Transitioning to causal approaches requires retraining personnel, revising workflows, and potentially confronting uncomfortable questions about the validity of existing systems. Change management and institutional buy-in are as important as technical correctness.

6.3 Future Research Directions

Several promising directions for future research emerge from this work. First, developing more powerful causal discovery algorithms tailored to financial time series would reduce reliance on manual graph specification. Financial data exhibit unique characteristics non-stationarity, fat tails, feedback loops that may enable identification beyond what generic algorithms achieve.

Second, deeper integration of causal reasoning with deep reinforcement learning could enable adaptive policies that learn causal structure from interaction while maintaining robustness guarantees. Current approaches treat causal graphs as static, but causal relationships may themselves evolve or be discovered through exploration.

Third, extending causal inference to handle continuous time and high-frequency data would broaden applicability to domains like algorithmic trading and derivatives pricing. Most causal inference theory assumes discrete time and interventions, but financial markets operate continuously and actions have complex temporal dynamics.

Fourth, developing standardized benchmarks and evaluation protocols for causal financial AI would accelerate progress and enable meaningful comparisons across methods. The field currently lacks the equivalent of ImageNet or standard RL environments, making it difficult to assess advances objectively.

VII. CONCLUSION

This paper has developed a comprehensive framework for Causal AI in financial decision-making under uncertainty. We have argued that the fundamental challenge facing financial AI is not predictive accuracy but causal validity the ability to reason correctly about the effects of actions and interventions. Financial decisions inherently modify the environment, creating feedback loops and rendering purely correlational approaches fragile.

Through formal problem definitions, identification strategies, architectural designs, and empirical evaluations, we have demonstrated that causal reasoning provides a principled foundation for robust financial decision-making. The proposed framework integrates structural causal models, econometric identification methods, and modern AI capabilities into a unified system. Our experimental results show that causal approaches achieve substantially lower regret under regime shifts, greater robustness to distributional changes, and superior interpretability compared to predictive baselines.

The transition to Causal AI represents more than an incremental improvement; it constitutes a necessary paradigm shift for financial institutions operating in an era of deep uncertainty. As financial systems become more complex, interconnected, and subject to rapid change, the limitations of correlation-based modelling become increasingly apparent and consequential. Causal reasoning offers a path toward AI systems that are not only powerful but also trustworthy, interpretable, and robust.

Looking forward, the successful deployment of Causal AI in finance will require continued collaboration among researchers in machine learning, econometrics, finance, and computer science. It will demand new tools, new training programs, and new institutional practices. But the potential rewards more stable financial systems, better risk management, and more equitable access to credit and capital make this effort worthwhile.

We conclude with a call to action for the research community and practitioners: embrace causal reasoning not as a niche methodology but as a foundational principle for financial AI. Only by moving beyond prediction toward genuine causal understanding can we build decision-support systems worthy of the trust placed in them by society.

REFERENCES

[1] Arjovsky, M., Bottou, L., Gulrajani, I., & Lopez-Paz, D. (2019, July 5). *Invariant risk minimization*. arXiv.Org. https://arxiv.org/abs/1907.02893

[2] Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, *31*(2), 3–32. https://doi.org/10.1257/jep.31.2.3

[3] Chen, L., Pelger, M., & Zhu, J. (2019). Deep learning in asset pricing. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3350138

[4] Drobetz, W., & Otto, T. (2021). Empirical asset pricing via machine learning: Evidence from the European stock market. *Journal of Asset Management*, *22*(7), 507–538. https://doi.org/10.1057/s41260-021-00237-x

[5] Feng, G., He, J., & Polson, N. G. (2018, April 25). *Deep learning for predicting asset returns*. arXiv.Org. https://arxiv.org/abs/1804.09314

[6] Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

[7] Huang, J., Chai, J., & Cho, S. (2020). Deep learning in finance and banking: A literature review and classification. *Frontiers of Business Research in China*, *14*(1). https://doi.org/10.1186/s11782-020-00082-6

[8] Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences*. Cambridge University Press. https://doi.org/10.1017/cbo9781139025751

[9] Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, *247*(1), 124–136. https://doi.org/10.1016/j.ejor.2015.05.030

[10] Ozbayoglu, A. M., Gudelek, M. U., & Sezer, O. B. (2020). Deep learning for financial applications : A survey. *Applied Soft Computing*, *93*, 106384. https://doi.org/10.1016/j.asoc.2020.106384

[11] Pearl, J. (2009). *Causality*. Cambridge University Press. https://doi.org/10.1017/cbo9780511803161

[12] Peters, J., Janzing, D., & Scholkopf, B. (2017). *Elements of causal inference: foundations and learning algorithms*. MIT Press.

[13] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

[14] Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3). https://doi.org/10.1214/10-sts330

[15] Warin, T., & Stojkov, A. (2021). Machine learning in finance: a metadata-based systematic review of the literature. *Journal of Risk and Financial Management*, *14*(7), 302. https://doi.org/10.3390/jrfm14070302