FERI INFORMATION

Special Issue on Frontiers of Causal Inference

– Explanatory Notes (Part II) on SBI Research Review Vol.8 –

Yutaka Soejima, Principal Research Fellow of SBI Financial and Economic Research Institute

This article is a continuation of Part I.

1. Izumi Paper

The first paper authored by Izumi explores the application of causal inference in the financial domain. Statistical causal inference, including econometric techniques, has long been employed to evaluate the effects of monetary policy, analyze financial market fluctuations, and investigate corporate behavior and governance. The paper positions statistical causal inference—including econometric methods—as the first major trend in financial economics, and then introduces two emerging trends that have received relatively little attention in the field.

One is a new wave known as "Causal AI," which has emerged from the integration of causal inference with machine learning. Improvements in AI performance driven by deep learning, along with advances in machine learning techniques, have enabled the discovery of latent causal structures within high-dimensional and complex datasets. These developments have made it possible to estimate causal effects and generate counterfactual scenarios. Seasoned professionals have long relied on heuristics and intuition—forms of pattern recognition and causal perception refined through years of practical experience—to anticipate market fluctuations, assess creditworthiness, and forecast political and economic developments. Efforts to embed such tacit expertise into AI and machine learning systems have been ongoing for some time. As the digitization of virtually all aspects of society continues to advance, these technologies now enable comprehensive exploration of vast datasets, unlocking powerful capabilities for causal discovery. They have already achieved levels of performance that surpass human capabilities. While the domains differ, one clear example of AI surpassing human capabilities is its dominance in games such as shogi (Japanese chess) and go (another traditional board game). Comparable levels of superhuman accuracy and performance have also been achieved in fields including radiographic image analysis, drug discovery, logistics optimization, and anomaly detection.

The other emerging trend centers on causal inference based on NLP. This field focuses on extracting and analyzing causal relationships from textual data such as financial reports,

FERI INFORMATION (25.09.10) - Posted September 1, 2025 in Japanese

earnings releases, news articles, and social media posts. It specifically deals with causal structures expressed in natural language that are directly interpretable by humans. Unlike statistical causal inference or Causal AI, this method offers distinct advantages and significance. As the paper succinctly observes: *"In domains like economic phenomena, where human behavior plays a central role, statistical causality alone is insufficient to explain events. This is primarily because human cognition and behavioral rules—how people perceive and respond to events—form the core of causal structures."* The paper introduces techniques for extracting causality from textual data through concrete examples. For instance, earnings reports from publicly listed companies often reveal how external conditions and corporate strategic decisions causally influence financial outcomes. A system designed to explore such causal chains has been implemented and made publicly available, with illustrative use cases provided.

Rather than merely introducing techniques, the paper organizes causal inference into two conceptual frameworks—*"statistical causality"* and *"empirical causality"*—and highlights their complementarity. Statistical causality is a formal approach grounded in counterfactual models and structural equations, valued for its reproducibility and quantifiability. Empirical causality, by contrast, reflects an understanding of causality that incorporates human cognition and contextual awareness, supporting intuitive judgment in rare or unfamiliar situations. The paper argues that an integrated perspective combining both frameworks is increasingly vital in today's complex and non-stationary financial environment.

Another notable feature of the paper is its breadth of technical domains. It bridges diverse fields including AI, machine learning, NLP, statistics, econometrics, multi-agent simulation, alternative data, and domain-specific financial knowledge. This intellectual cross-pollination has the potential to push forward the frontier of causal inference. The editorial direction of this special volume reflects and embodies this multidisciplinary vision.

2. Ichise et al. Paper

This paper belongs to the third emerging wave identified in the introductory article of this volume, which outlines three major waves of causal inference in the financial domain—statistical methods, causal AI, and NLP—and presents a novel technique developed by the authors. To illustrate, securities reports typically contain a section of Management Discussion and Analysis (MD&A), which provides management's perspective on the company's financial condition and performance. Suppose corporate executives are classified into two groups: those who prioritize shareholders and those who emphasize stakeholders such as business partners and employees. Would their interpretations of causality differ? For instance, did improved corporate governance lead to enhanced profitability, or did the emergence of surplus profitability provide the opportunity for the company to pursue governance reforms? As these

questions suggest, conflicting causal directions may coexist. The former may reflect a shareholder-oriented management, while the latter may indicate a stakeholder-focused management. Can such causal reasoning, embedded in the minds of executives, truly be extracted from the textual content of MD&A sections? The central claim of this paper is that it can, and the authors introduce a tool they have developed to accomplish this.

Another example discussed in the paper concerns managerial overconfidence. Prior research suggests that overconfident executives tend to overestimate the accuracy of their decisions and underestimate uncertainty in probabilistic events. These executives are presumed to exhibit a tendency to credit favorable outcomes to their own abilities while attributing unfavorable results to external circumstances—an inclination known as *self-serving attribution bias*. Cognitive biases of this kind, including overconfidence, can significantly shape corporate strategic decisions, capital structure, investment behavior, and forecasts of performance and stock prices. An analysis of MD&A text enables researchers to identify which companies' executives display a stronger propensity toward such behavior. The author's study finds a positive correlation between self-serving attribution bias and overconfidence or optimistic forecasting. Moreover, the bias tends to diminish among CEOs who are older, have longer tenure, or possess academic training in science or engineering. Empirical evidence further shows that firms with high levels of this bias are more likely to reduce dividends, increase share buybacks, maintain high financial leverage, and engage in value-destroying M&A activity.

Professionals engaged in corporate evaluation within the financial industry are likely to find this methodology highly compelling. It enables the formalization and automation of executive assessment—an area that has traditionally depended on human intuition, required intensive manual effort, and resisted scientific systematization. This advancement also opens the door to empirical comparisons between human judgment and algorithmic processing. The paper explains how to extract causal relationships from text, structure them into causal knowledge graphs, and identify differences in causal perspectives. It further introduces FinCaKG, a framework developed by the authors that employs machine learning to automatically generate causal knowledge graphs from the perspective of financial experts. From an econometric perspective, ETE-FinCa provides a compelling approach: it employs computational text analysis to systematically identify instrumental variables for causal inference from an extensive corpus of candidates.

A broader contribution of this methodology lies in its assertion that causal inference depends more critically on the observer's perspective than on the data or models employed. Consider the question: *"How did individuals who received financial product recommendations or financial education alter their decision-making based on that information?"* In such a research setting, the cognitive and interpretive processes of individuals themselves become the object of causal

inference. This requires constructing models that incorporate how humans perceive and understand the world—that is, models informed by cognitive causal reasoning. Such an approach aligns with insights from behavioral economics and cognitive science, enabling the redesign of causal inference from a more human-centered perspective.

3. Sannai et al. Paper

The central theme of this paper is the integration of Large Language Models (LLMs) with statistical causal discovery. Traditional causal inference methods that rely solely on statistical data may, when applied mechanically, produce causal relationships that contradict intuitive understanding. For example, a spurious correlation suggesting that increased ice cream sales lead to more drowning incidents may emerge if relevant causal channels—such as temperature and recreational water activities—are omitted from the analysis. Indeed, there is no guarantee that all relevant variables are appropriately included in the statistical dataset.

The reason such outcomes are rare in practice is that analysts, whether consciously or unconsciously, formulate hypotheses on the basis of their prior knowledge. However, this reliance on the analyst's background inevitably introduces the possibility of bias. The key challenge lies in identifying the origin of a causal narrative. While the ice cream example can be easily dismissed through common sense and would never be considered a viable hypothesis, more complex causal hypotheses—such as the notion that prolonged economic stagnation was caused by insufficient fiscal stimulus—require consideration of numerous factors and cannot be validated with a single graph. Yet, many individuals hold strong biases that lead them to infer causality in such cases. Understanding why these biases arise is itself an important subject of causal inference within the broader scope of sociology.

Returning to the main topic, the authors of foundational research behind this paper focus on the fact that LLMs have been trained on vast amounts of information across diverse domains, encompassing expert knowledge and common sense from various business fields. The authors develop a method for integrating statistical techniques with LLMs in causal inference, which unfolds in the following steps:

1. Initial Causal Discovery: Conduct causal exploration based on statistical data and construct a causal knowledge graph, supported by the application of various statistical analysis techniques.

2. Mechanism Exploration via LLM Prompting: For each detected causal relationship—such as between events A and B—elicit knowledge already embedded in LLM. Specifically, prompt the LLM with a question like: *"Please explain the mechanism by which A influences B, based on domain-specific expertise."* At the same time, provide the results of relevant statistical

analyses, and employ Chain-of-Thought prompting to leverage the LLM's reasoning capabilities. In other words, use prompts that encourage the LLM to explicitly articulate its reasoning step-by-step.

3. Validation Using LLM Reasoning: Present the inferred causal relationships derived from the preceding analysis to the LLM and prompt it to assess their validity through a binary Yes/No response. This approach effectively leverages the LLM's knowledge base and reasoning capabilities in a multi-layered fashion. Given the inherent variability in outputs of probabilistic language models, the same question should be posed multiple times, and the average probability of affirmative responses should be employed.

4. Feedback Loop: Incorporate the information obtained from the LLM into statistical methods to re-examine causal relationships among all events (variables). Through this iterative analytical process, causal discovery and data processing based on statistical techniques are integrated with the LLM's knowledge and reasoning capabilities in a mutually reinforcing and cross-referenced manner.

The paper presents empirical validation using datasets commonly used in causal discovery research—such as those on automobile fuel efficiency and climate-topography variables—as well as a proprietary health screening dataset that was not part of the LLM's training data. The authors assess the extent to which the proposed method enhances the accuracy of causal discovery and confirm that the LLM's embedded knowledge and reasoning capabilities contribute substantively to this improvement.

Notably, while the LLM does not have access to the health screening dataset itself, it has already been trained on expert-curated medical knowledge. Therefore, its medical expertise and reasoning are considered to contribute to the analysis. As the author of these explanatory notes, in my own empirical research I have extracted graduate-level econometric knowledge, mathematical formulations, and programming implementations—including code syntax as well as the identification and application of relevant libraries—from LLMs. When cross-checked against academic textbooks, such outputs have exhibited little to no hallucination. This approach, which integrates the rapid evolution of LLMs into causal inference, suggests that further performance enhancements can be expected as these models continue to advance.

4. Saito Paper

In statistical causal inference, several methods are commonly employed: randomized controlled trials (RCTs); difference-in-differences (DID), which compares changes before and after an intervention under the assumption that other conditions remain constant; matching methods, which pair treated and untreated units with similar characteristics; regression

discontinuity designs, which exploit naturally occurring environmental thresholds; and instrumental variable techniques. However, depending on the nature of the problem, these methods are not always applicable. For example, in personalized services such as e-commerce or music and video recommendation systems, the number of possible interventions is so vast that methods premised on a single intervention are fundamentally inapplicable. Other limitations inherent in these methods include the breakdown of assumptions regarding the comparability of external environments and influencing factors, the inability to generalize beyond pre- and post-intervention contexts, and the absence of appropriate instrumental variables. Moreover, RCTs and A/B testing may degrade user experience, and in domains such as medicine, ethical constraints often make experimental interventions infeasible.

In such cases, decision-makers must select among multiple intervention strategies based solely on data collected under specific environmental and policy conditions in the past. The methodology that enables this is known as *Off-Policy Evaluation* (OPE), and it is already being implemented in companies that use recommendation systems.

This paper provides an overview of the fundamental concepts of OPE, its major estimators, and recent research developments. In addition to conceptual explanations, it offers clear and accessible descriptions of the underlying computations, presented through mathematical formulas. The term *past environment* refers to what are known in machine learning as features—such as user viewing and purchasing histories, or historical stock price movements. Based on the observation of these features, an intervention—specifically, a policy decision such as executing a recommendation algorithm or selecting a stock portfolio—is undertaken. This intervention is referred to as an action. Given the features and actions, a reward (e.g., purchase, viewing time, investment return) is realized, typically estimated using functions from machine learning and related methods. The goal of a decision-making policy is to maximize expected rewards.

The objective of OPE is to develop estimators that can accurately predict the expected reward of policies not yet implemented, using only log data collected under past conditions in which rewards were observed following specific actions based on given features. Although inherently challenging, this approach offers notable advantages: it relies exclusively on historical data, incurs no experimental cost, and allows for rapid execution. As a result, OPE is already being applied in practice, and ongoing research efforts directed toward improving these estimators.

The paper introduces three representative estimators and provides quantitative explanations of their advantages and disadvantages using real-world analytical examples. The accuracy of each estimator depends on factors such as the precision of the reward estimation function, the level of noise in the observed data, the volume of available observations (with larger datasets

generally improve precision), and the number of policies to be evaluated (more policies tend to reduce precision). In practice, these factors must be carefully weighted when selecting the most appropriate estimator for a given problem.

Each estimator is evaluated along two dimensions: accuracy in terms of expected value and variance stability, i.e., low fluctuation of error. This is analogous to the mean–variance approach used in asset portfolio selection. As more estimators are developed, practitioners face the challenge of selecting the most appropriate one based on the nature of the problem. One approach to evaluation entails clarifying practitioner's preferences concerning the trade-off between mean and variance prior to selecting an estimator. For example, in medical treatment selection, where high risk is unacceptable, minimizing variance takes precedence. This line of reasoning closely parallels the approach used in selecting portfolio management strategies. Another contribution of this paper is the introduction of a new metric for policy selection. Drawing on the concept of the Sharpe ratio, the author proposes a risk-adjusted performance evaluation index and explains its practical applicability.

5. Masujima and Namba Paper

The final paper in this volume presents a case study of causal inference using RCTs and instrumental variable techniques, based on survey data collected by the institute, that is, *Next Generation Finance Survey*. These methods represent classical applications of statistical causal inference, and the paper demonstrates that these foundational techniques—widely adopted in the field—are also actively implemented by the institute. The full details of the study are available in the institute's working paper series.

The survey covers not only traditional financial assets such as equities but also emerging assets like crypto assets and stablecoins. Identical surveys have been conducted in the United States, China, and Germany, thereby enabling international comparisons. In terms of scale, the survey on crypto assets ranks among the largest undertaken by Japanese companies or organizations. Aggregate results from three rounds of the survey have also been published.

This paper focuses on analyzing the characteristics of individuals who hold crypto assets. The findings indicate a general tendency for men, younger individuals, high-income earners, those with lower educational attainment, and those with lower risk aversion to be more likely to invest in crypto assets. Furthermore, individuals with higher inflation expectations and greater uncertainty regarding inflation and growth prospects tend to demonstrate stronger investment propensities.

A key contribution of this paper is its analysis of the impact of financial literacy on crypto investment, accounting for reverse causality and endogeneity arising from omitted variable

bias. Prior research—also published as a working paper by other institute staff—had identified a nonlinear relationship between financial literacy and crypto investment. Specifically, the proportion of individuals with investment experience in crypto assets was higher among those with moderate financial literacy scores and lower among those with either high or low scores. Using instrumental variable techniques, the paper demonstrates that this observed relationship may be spurious. Instead, the analysis indicates that higher financial literacy is positively associated with both the likelihood of having investment experience in crypto assets and the share of crypto holdings within an individual's financial portfolio.

This empirical analysis uncovers causal relationships that cannot be captured through simple correlations or standard regression techniques, thereby demonstrating the effectiveness of instrumental variable methods. A notable feature of this study is its approach to a common limitation of the instrumental variable method, namely, the difficulty of identifying valid instruments, by embedding research planning into the survey design from the outset.

Another notable feature of the study is its use of RCTs, also embedded in the survey design. Specifically, the paper investigates how the provision of information influences crypto purchasing behavior. Respondents were randomly assigned to two groups: a control group that received no information, and a treatment group that was provided with data on Bitcoin's historical returns—namely, that its price had increased more than sixfold over the past five years and more than one hundredfold over the past ten years. Respondents were then asked to indicate their preferred portfolio composition one year later. The results suggest that the provision of information increased both the willingness to purchase crypto assets and the proportion of crypto holdings. This analysis implies that providing information about financial assets can encourage investment behavior.

6. Study Group Report

The last section of this volume presents the second set of policy recommendations—titled "*A Blueprint for Next-Generation Financial Infrastructure*"—from the Study Group on the Creation of Next-Generation Financial Infrastructure, organized by the institute. The study group's core concerns are closely aligned with the theme of this volume. The opening statement of the report highlights the following issue:

"With the advancement of a digital society driven by innovations in information technology—such as the use of financial APIs, blockchain, and big data—new forms of payment and remittance, digital financial assets like crypto assets, and decentralized financial services (DeFi) are emerging. These developments are transforming both the providers and delivery methods of financial services. Today, data has become a key source of revenue, and enhancing information

production capabilities has become a critical management challenge for financial institutions."

The recommendations emphasize the following key points:

- Visualization of user needs through the utilization of data spanning both financial and non-financial sectors
- Development of robust data infrastructure and integration into IT systems to facilitate effective data use.
- Application of user behavioral data—such as accounting records, purchase histories, and production activity information—from non-financial sectors in financial services
- Monetization of customer data, originally generated in the course of business services, by repurposing it for alternative uses; this enhances personalization services including demand forecasting, dynamic pricing, and recommendation systems, thereby contributing to greater profitability.
- Utilization of data science (e.g., AI, machine learning, causal inference) and modern IT system development methods to enable a rapid cycle of continuous improvement based on business implementation and feedback from operational results
- Acceleration of timely and efficient service enhancement and discovery of new offerings through this feedback-driven cycle.
- Legislation of user data rights, known as Consumer Data Rights (CDR), to establish legal frameworks ensuring user control over personal data.

In essence, these recommendations convey a clear message: financial services are increasingly integrated with non-financial services through data sharing, evolving into both a data-circulation business model and a broader data-driven system.

Readers who have followed these explanatory notes thus far will likely have recognized the connection between the recommendations and causal inference. Causal inference is a key component in the next generation financial infrastructure. The following elements are considered critical:

1. Robust IT infrastructure for data observation, storage, and utilization, supported by skilled professionals.
2. Promotion of standardization and interoperability of data and IT systems, together with secure data distribution.
3. Collaboration between data scientists and professionals with domain-specific business

expertise to advance analytics.

4. Strategic vision and leadership from executives and policymakers to drive implementation across the ecosystem.

We hope that this volume's special feature, *"The Frontiers of Causal Inference,"* will make a meaningful contribution to advancing these initiatives.