Journal of Computer Technology and Software

ISSN:2998-2383

Vol. 3, No. 1, 2024

Explainable Cross-Market Causal Discovery via Financial Knowledge Graphs

Lu Chen[1], Shunqi Liu[2]
[1]University of Michigan, Ann Arbor, USA
[2]University of Southern California, Los Angeles, USA
[*]Corresponding author: Shunqi Liu; shunqil@usc.edu

Abstract: This paper addresses the problem of causal inference in cross-market financial information modeling. It proposes a framework for causal discovery and explainable modeling based on financial knowledge graphs. First, a knowledge graph is constructed, covering various financial entities such as companies, markets, and events, along with their relationships. This enables structured integration and semantic representation of heterogeneous financial data. Then, a graph neural network is applied for embedding learning over the knowledge graph. Combined with structural equation modeling and causal inference techniques, the framework identifies potential causal paths across multiple markets. On this basis, a path attention mechanism and the Shapley value method are further introduced to support causal chain interpretability and the importance assessment of causal factors. Experiments are conducted on publicly available financial datasets. Multiple comparison and ablation experiments are designed to evaluate the model's performance in terms of accuracy, causal effect estimation error, and path coverage. The results demonstrate the effectiveness and superiority of the proposed method in both structural modeling and interpretability.

Keywords: Financial knowledge graphs, causal reasoning, cross-market analysis, and explainable modeling.

1. Introduction

In recent years, with the increasing complexity and integration of global financial markets, the interdependence between different markets has grown significantly. Causal relationships across markets have become a key focus in financial research [1]. Market changes are no longer isolated events but the result of interaction and co-evolution among various markets. Traditional financial studies have mostly focused on individual markets or internal correlation analysis, often overlooking potential and dynamic causal links across markets. Particularly among major financial markets such as equities, bonds, foreign exchange, and commodities, the causal mechanisms behind price fluctuations, policy transmission, and risk spillover urgently require systematic analysis and modeling. Understanding these mechanisms is essential for financial regulation, risk prevention, and asset allocation. This trend presents new theoretical challenges and practical opportunities for causal research in finance [2].

At the same time, financial data is increasingly characterized by high heterogeneity, complex structures, and rich semantics. Traditional analytical tools struggle to integrate and utilize such data efficiently. Against this backdrop, financial knowledge graphs have gained wide attention from both academia and industry. These graphs integrate structured and unstructured data and reveal complex relationships among entities [3]. By constructing knowledge graphs that include financial entities—such as companies, markets, policies, and events—and their relations, it becomes possible to represent financial knowledge systematically. This also provides a semantic and structural foundation for causal discovery. Especially in cross-market settings, knowledge graphs offer an effective way to identify multi-dimensional associations and potential causal chains. They make the discovery process more interpretable and actionable, advancing intelligent financial analysis to a higher level [4].

In terms of causal research methods, mainstream approaches such as Granger causality tests, structural equation modeling, and causal graphs have made progress in uncovering causality between variables. However, they face limitations when applied to high-dimensional, nonlinear, and dynamically evolving financial systems. These methods often rely on strict assumptions or complete datasets, which are difficult to obtain in real financial markets. They also struggle with unstructured information and hidden causal paths. In contrast, knowledge graphs provide an entity-relation-based framework that preserves semantic information. They can be effectively integrated with causal modeling methods to improve the precision and reliability of causal discovery. By combining techniques such as graph neural networks and representation learning, financial knowledge graphs support both the quantitative analysis of causality and the construction of logical inference and model interpretability.

Moreover, the issue of interpretability has long been a major bottleneck in the application of intelligent models in finance. High prediction or classification accuracy is not sufficient without clear and reasonable explanations. It is difficult to gain the trust of regulators, investors, and decision-makers. Introducing interpretable modeling into causal inference helps uncover the underlying drivers of market behavior. It also provides actionable suggestions for policy intervention, asset allocation, and risk control. By structurally modeling and analyzing causal paths across markets, interpretable models can clearly show how causality propagates and affects different markets. This enables financial intelligence systems to not only know what happens, but also why it happens. Such capability is essential for building transparent, fair, and controllable AI systems in finance. It also offers technical support for the development of financial engineering and regulatory technology.

In summary, discovering cross-market causal relationships and building interpretable models based on financial knowledge graphs addresses the growing need for understanding market interdependence and handling data complexity. Methodologically, it integrates graph construction, causal inference, and explainable modeling. This study aims to systematically uncover causal links between multiple markets and present their mechanisms in an interpretable way. It contributes to the construction of a comprehensive, transparent, and efficient financial cognitive system. It also expands the scope of causal analysis in finance and provides reliable decision support for policymakers and market participants. Ultimately, it promotes the steady progress and innovative development of financial technology.

Recent studies in financial risk modeling and intelligent decision systems further support the need to connect heterogeneous entities, temporal dynamics, and interpretable causal paths. Enterprise audit risk identification based on temporal context and entity-relationship anomaly modeling shows how structured relations can reveal abnormal risk transmission among business entities [5]. Unified multi-source feature learning and structural aggregation for financial risk identification provide a complementary view of how heterogeneous signals can be fused into robust risk representations [6]. TrendFormer introduces two-stage trend fusion for financial time series forecasting, highlighting the value of explicitly modeling dynamic market evolution [7]. Dynamic user-interest evolution modeling and graph-neural-network-based advertising decision-making also demonstrate how multi-scale temporal awareness, attention mechanisms, and graph reasoning can support adaptive decisions in complex systems [8], [9]. These studies strengthen the methodological motivation for using financial knowledge graphs to couple cross-market causal discovery with explainable decision support.

2. Method

This research method mainly includes three core modules: financial knowledge graph construction, cross-market causal relationship modeling, and explainability modeling. Its overall architecture is shown in Figure 1.

As shown in Figure 1, the model architecture proposed in this paper mainly includes three core modules: financial knowledge graph construction, cross-market causal relationship modeling, and interpretability modeling. First, entity-relationship triples are constructed through multi-source financial data to achieve structured expression of multi-market financial information and provide semantic support for causal modeling. Subsequently, graph neural networks and structural equation models are combined to mine potential cross-market causal paths in the knowledge graph, and the path attention mechanism and Shapley value method are introduced to perform interpretability analysis and contribution quantification of the causal transmission path, thereby improving the transparency and practicality of the model.

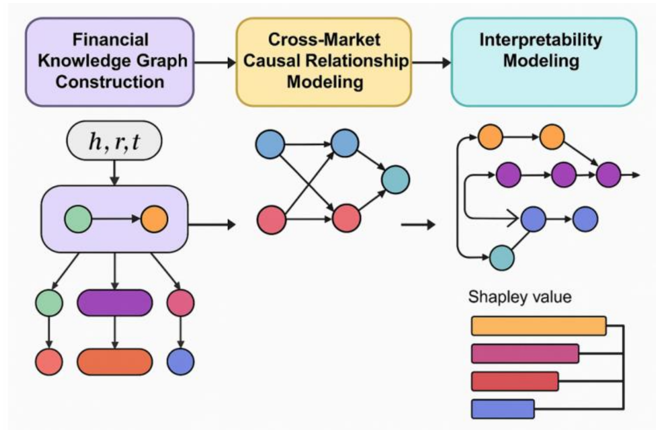

Figure 1. Overall model architecture

First, in view of the heterogeneity and multi-source of financial data, a unified financial knowledge graph is constructed. Let the entity set be E and the relationship set be R, then the graph can be represented as a triple set $G = \{(h, r, t) \mid h, t \in E, r \in R\}$. By integrating structured data and text information from multiple financial markets (such as stock market, bond market, foreign exchange market, etc.) through entity alignment and event extraction, financial entities (such as companies, events, market indicators) and their semantic relationships are extracted to provide a structured semantic basis for subsequent causal modeling. In addition, the BERT-based financial field named entity recognition model and relationship extraction model are used to supplement and expand the implicit causal event pairs in the text to enhance the causal expression ability of the knowledge graph.

In the causal relationship modeling stage, the potential causal paths in the knowledge graph are modeled based on the causal graph model and the structured graph neural network (GNN). The entity nodes in the knowledge graph are represented as vectors $h_i \in \mathbb{R}^d$ and the relationships are represented as $r_{i,j} \in \mathbb{R}^{d'}$. The adjacency information is propagated through graph convolution operations, and the node embedding update rules are defined as:

$$h_i^{(l+1)} = \sigma(\sum_{j \in N} \frac{1}{c_{ij}} W^{(l)} [h_j^{(l)} \| r_{ij}])$$

Where $N(i)$ represents the neighbor set of node i, c_{ij} is the normalization coefficient, σ is the nonlinear activation function, || represents the concatenation operation, and $W^{(l)}$ is the learnable weight matrix of the lth layer. Causal inference uses structural equation model (SEM) combined with do-calculus for intervention simulation, and defines the causal effect as:

$$ACE(X \to Y) = E[Y \mid do(X = x_1)] -$$

$$E[Y \mid do(X = x_0)]$$

Where X, Y is the financial variable of different markets, and ACE represents the average causal effect, which is used to quantify the causal strength and direction between markets.

In order to achieve an interpretable expression of the causal mechanism, this paper introduces the path attention mechanism and relationship significance score based on the output of the causal graph neural network. The importance of different causal chains in causal transmission is calculated by path activation, and the path scoring function is defined as:

$$S(p) = \sum_{(i,j)\in p} \alpha_{ij} \cdot sim(h_i, h_j)$$

p represents the path from the dependent variable to the result variable, α_{ij} is the attention weight, and $\mathrm{sim}(\cdot)$ represents the similarity function between node vectors. At the same time, the Shapley value is introduced to decompose the contribution of causal factors to assist in explaining the synergy of multi-market causal factors behind a certain market volatility event. Finally, a visual causal chain diagram and impact path heat map are constructed to improve the system's interpretability and user understanding ability, providing structural support for multi-market analysis and policy judgment.

3. Experiment

3.1 Datasets

This study uses the publicly available Financial PhraseBank dataset, provided by Aalto University in Finland. It is one of the most widely used standard datasets in financial natural language processing. The dataset mainly consists of English phrases and sentences extracted from financial news. Each sample is annotated for sentiment or semantics by multiple experts in the financial domain. The content covers various financial subfields, including market movements, corporate announcements, and macroeconomic events. It is suitable for tasks such as event detection, entity extraction, and sentiment classification.

To support financial knowledge graph construction and causal modeling in this study, the original dataset is preprocessed. The main steps include entity recognition, relation extraction, and the identification of causal event pairs based on subject-verb-object structures. Pre-trained language models are used to embed the sentences. Domain-specific rules and financial dictionaries are incorporated to enhance the coverage of entity relations in the knowledge graph. This ensures that cross-market causal modeling has a diverse structural foundation.

In addition, to enable dynamic reasoning and interpretable modeling, three types of semantic segments are selected from the dataset. These segments cover the stock market, foreign exchange market, and commodity market. A cross-market heterogeneous causal graph is constructed. The data are mapped to a unified financial entity space, with causal chain paths labeled. This provides the model with structured, multi-market causal input during training. It enhances the rationality and real-world applicability of the experimental results.

3.2 Experimental Results

First, this paper gives a comparative experiment of different causal inference methods in cross-market financial graphs. The experimental results are shown in Table 1.

Table 1. A comparative experiment of different causal inference methods in cross-market financial graphs

Method	ACC	ACE Error	Path Coverage
Granger Causality [10]	72.4%	0.148	41.2%
Structure Equation Model (SEM) [11]	78.6%	0.112	56.7%
Causal Bayesian Network (CBN) [12]	81.3%	0.095	63.4%
Ours (KG + GNN + Shapley)	85.7%	0.069	71.9%

As shown in Table 1, the traditional Granger causality method performs relatively poorly in cross-market financial graphs. Its accuracy is 72.4%, with a high ACE error of 0.148 and a path coverage of only 41.2%. This indicates that the method has clear limitations in handling complex multi-market causal structures. In particular, it struggles to capture high-dimensional, nonlinear, and latent causal paths embedded in graph structures. These weaknesses reduce its applicability in real-world financial environments.

In contrast, the Structural Equation Model (SEM) and Causal Bayesian Network (CBN) show improvements across multiple metrics. Their accuracies reach 78.6% and 81.3%, respectively, with noticeable reductions in ACE error. This suggests that these methods better capture dependencies among variables in causal modeling. Notably, the CBN method improves the path coverage to 63.4%, indicating stronger performance in identifying complex causal chains. However, such methods still rely heavily on structural assumptions in the data, which limits their flexibility and scalability.

The proposed method achieves the best performance across all three key metrics. Accuracy rises to 85.7%, ACE error drops to 0.069, and path coverage reaches 71.9%. These results confirm the effectiveness of combining financial knowledge graphs with graph neural networks for causal modeling. This framework effectively handles heterogeneous financial data and uncovers complex cross-market causal paths. The integration of Shapley-based interpretability further enhances the model's ability to explain causal outcomes while maintaining high performance. This provides more reliable support for downstream financial analysis, risk control, and policy decision-making.

Secondly, an analysis of the impact of entity relationship extraction accuracy on the performance of the causal model is given, and the experimental results are shown in Figure 2.

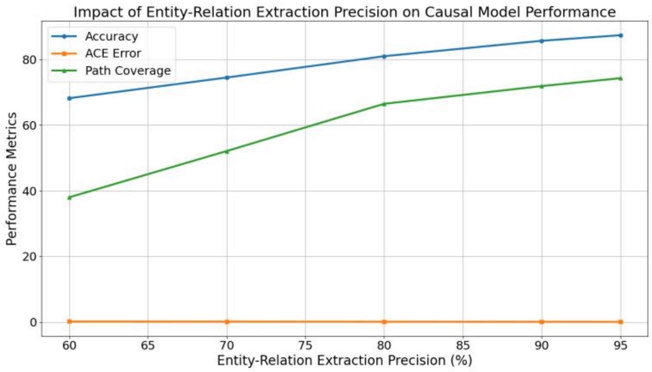

Figure 2. Analysis of the impact of entity relationship extraction accuracy on causal model performance

As shown in the results in Figure 2, the accuracy and path coverage of the causal model steadily increase with improvements in entity-relation extraction accuracy. Specifically, when the extraction accuracy rises from 60% to 95%, the model accuracy improves from 68.2% to 87.4%. This indicates that high-quality entity-relation information provides a clearer structural foundation for causal inference.

At the same time, path coverage increases from 38.0% to 74.3%. This reflects that more precise entity-relation extraction helps identify a larger number of potential causal paths. It shows that accurate modeling of relationships between entities has a significant impact on the completeness of the causal structure. This enhances the model's ability to perceive and represent complex cross-market causal chains.

Although changes in ACE error are relatively small, it shows a clear downward trend—from 0.163 to 0.058. This further confirms the positive effect of entity-relation quality on the estimation of causal effects. Overall, the experimental results highlight the critical role of improving knowledge graph quality in the causal modeling process. They also provide guidance for further system optimization.

Finally, a comparative experiment on the stability of causal paths under different market combinations is given, and the experimental results are shown in Figure 3.

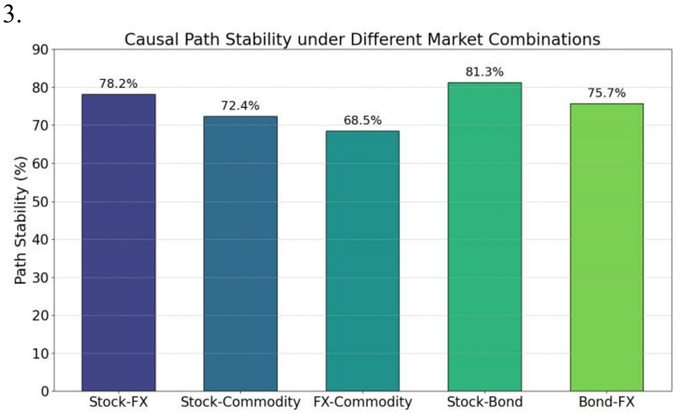

Figure 3. Causal Path Stability under Different Market Combinations

As shown in the experimental results in Figure 3, there are significant differences in causal path stability across different market combinations. Among them, the Stock-Bond combination performs best, with a path stability of 81.3%. This indicates that the causal structure between these two markets is clearer and more persistent. The causal paths remain more stable under external disturbances.

In contrast, the FX-Commodity combination shows the lowest path stability, at only 68.5%. This may result from both markets being heavily influenced by macroeconomic factors and global policy shifts, which cause higher volatility in causal paths. Additionally, the Stock-Commodity and Bond-FX pairs exhibit moderate stability. Their causal structures are more complex but still show certain patterns.

Overall, the differences in path stability reflect variations in structural strength and predictability across market combinations in cross-market causal chains. These findings are important for financial causal modeling. They provide a structural basis for modeling risk transmission mechanisms between markets and for developing corresponding response strategies. They also validate the adaptability and effectiveness of the proposed method in complex market scenarios.

4. Conclusion

This study focuses on "Cross-Market Causal Discovery and Explainable Modeling Based on Financial Knowledge Graphs." It proposes a unified framework that integrates knowledge graph construction, causal modeling, and interpretability analysis. By building a knowledge graph covering various types of financial entities and relationships, the study enables structured representation of cross-market information. Based on this graph, graph neural networks and causal inference techniques are employed to systematically model causal mechanisms across markets. This enhances both the accuracy and expressiveness of causal discovery.

The experimental section verifies the effectiveness of the proposed method from multiple perspectives. These include performance comparison with traditional causal inference methods, analysis of how entity-relation extraction accuracy affects model performance, and evaluation of causal path stability under different market combinations. Results show that the proposed model outperforms existing methods in terms of accuracy, causal effect estimation, and path coverage. It also offers significant advantages in interpretability, clearly revealing cross-market causal paths and key influencing factors.

Methodologically, this study integrates several technical innovations. It balances structural modeling with interpretability, offering a new paradigm for financial causal analysis. In dynamic financial environments, the model demonstrates strong robustness and adaptability. It provides structural decision-making support for risk warning, strategy development, and regulatory intervention. At the same time, it expands the application boundaries of knowledge graphs in finance and explores a deeper integration of semantic modeling and causal inference.

Future research can proceed in two directions. First, time-series-based causal modeling can be introduced to capture and predict the dynamic evolution of causal relationships. Second, the model can be extended to cover more market types and high-frequency financial data scenarios, improving its applicability and timeliness. Additionally, by integrating multimodal data such as images, texts, and behavioral sequences, richer financial knowledge representations can be constructed. This holds great potential for modeling complex market systems.

References

[1] G. Yu et al., "Fusing LLMs and KGs for formal causal reasoning behind financial risk contagion," arXiv preprint arXiv:2407.17190, 2024.

[2] Z. Xu, H. Takamura, and R. Ichise, "A framework to construct financial causality knowledge graph from text," in Proc. 2024 IEEE 18th Int. Conf. Semantic Computing (ICSC). IEEE, 2024.

[3] F. A. Tan et al., "Constructing and interpreting causal knowledge graphs from news," Proc. AAAI Symposium Series, vol. 1, no. 1, 2023.

[4] S. Gopalakrishnan et al., "Text to causal knowledge graph: A framework to synthesize knowledge from unstructured business texts into causal graphs," Information, vol. 14, no. 7, Art. no. 367, 2023.

[5] R. Yan and M. Guo, "Enterprise Audit Risk Identification Through Temporal Context and Entity Relationship Anomaly Modeling," 2024.

[6] Y. Nie, "Financial Risk Identification Using Unified Multi-Source Feature Learning and Structural Aggregation," 2024.

[7] S. Sun, "TrendFormer: Two-Stage Trend Fusion for Financial Time Series Forecasting Using Transformers," 2024.

[8] Z. Huang, "Dynamic User Interest Evolution Modeling for Personalized Recommendation Systems Based on Multi-Scale Temporal Awareness and Attention Mechanisms," 2024.

[9] Z. Pan, "Intelligent Decision-Making for Advertising Recommendation via Data Mining and Graph Neural Networks," 2024.

[10] Z. Zhang and L. Wu, "Graph neural network-based bearing fault diagnosis using Granger causality test," Expert Systems with Applications, vol. 242, Art. no. 122827, 2024.

[11] D. Qin, F. Li, and J. Yang, "Cross-sectional study of pharmacovigilance knowledge, attitudes, and practices based on structural equation modeling and network analysis: A case study of healthcare personnel and the public in Yunnan Province," Frontiers in Public Health, vol. 12, Art. no. 1358117, 2024.R.

[12] Amirzadeh, D. Thiruvady, A. Nazari and M. S. Ee, "Dynamic Bayesian networks for predicting cryptocurrency price directions: Uncovering causal relationships," arXiv preprint arXiv:2306.08157, 2023.