CAUSALRANK: DISCOVERING CAUSAL FACTORS FOR CROSS-SECTIONAL STOCK RETURNS

Anonymous authors
Paper under double-blind review

ABSTRACT

Empirical asset pricing in the finance literature requires not only identifying numerous predictive factors but also uncovering the reliable and explainable pricing mechanisms behind return predictability. To answer this question, we propose CausalRank, a foundation model designed for parent node ranking. For a given target variable, CausalRank aims to simultaneously identify its direct causal parents from a massive pool of candidate variables and rank them based on the strength of their causal effects. Specifically, we train CausalRank on synthetic and semi-synthetic data with known causal structures and intervention effects. This training enables the model to learn transferable causal inference capabilities across various causal structures and generating functions. CausalRank distinguishes direct parent variables from correlated noncausal ones by assessing each pairwise candidate-target relationship in the context of the other candidates. This enables it to predict both whether a direct causal relationship exists and its effect. Using the U.S. stock market data, we show that the investment strategy using the characteristics identified by CausalRank can achieve a higher risk premium with lower turnover. Meanwhile, it ranks these characteristics by causal effect strength, which is crucial for evaluating a factor's role in empirical asset pricing and understanding the complexity of financial markets.

1 INTRODUCTION

Machine learning has been widely used in empirical asset pricing as a tool to analyze linear and nonlinear relationships between high-dimensional stock-level characteristics and expected returns (Freyberger et al., 2020; Gu et al., 2020; 2021; Leippold et al., 2022; Chen et al., 2024). Existing approaches typically train the model from predictive associations of candidate characteristics with expected returns. However, strong predictive performance does not imply that a certain characteristic has a causal relationship with the expected returns. Such a characteristic with strong predictive performance may merely serve as a proxy for an underlying causal driver. Moreover, when the distribution of the proxy characteristics shifts, predictive signals built on such associations may also change (Green et al., 2017), thereby undermining the stability of the model across time.

Causal learning offers a different perspective to solve this problem (Rojas-Carulla et al., 2018; Brouillard et al., 2020). Whereas predictive associations are inherently tied to a particular observational distribution, causal discovery seeks to uncover structural relationships implied by the underlying data-generating mechanism (Heinze-Deml et al., 2018). If the generating function of expected returns remains stable while the marginal distribution of candidate characteristics changes, causal relationships may provide more stable and pricing-relevant information than purely associational patterns (Huang et al., 2020; Schölkopf et al., 2021). Moreover, causal drivers can differ substantially in effect magnitude (Tramontano et al., 2024; Huang et al., 2025). Causal discovery identifies which characteristics are direct causes of expected returns, while causal effect estimation quantifies their strength (Schubert et al., 2025). Combining the two provides a view of both the causal mechanisms underlying expected returns and the relative importance of the identified characteristics, beyond predictive association alone.

Recent advances in causal foundation models provide a useful tool toward this objective (Lorch et al., 2022). Methods such as DAG-FM (Chen et al., 2026), Arrow (Thompson et al., 2026), and TabCausal (Li et al., 2026) are pretrained on large collections of synthetic datasets generated under diverse

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

causal graphs and structural mechanisms (Ke et al., 2023). They provided a theoretical foundation for learning causal inference from structurally generated pretraining tasks. Through such training, these models acquire causal inference capabilities that can transfer to other separate datasets, enabling them to infer causal structure directly from observations in the testing sample (Lachapelle et al., 2019; Ma et al., 2026). This paradigm is particularly appealing when ground-truth causal structures are unavailable and informative real-world data are limited. In empirical asset pricing, these challenges arise naturally, and both the number of stocks on the market and the historical length of the market are inherently limited. Furthermore, existing causal foundation models primarily focus on recovering causal structures or identifying causal variables (Kim et al., 2025; Toth et al., 2022), while the joint identification of direct parents and their ranking by the strength of causal effect remains largely unexplored. Unlike recovering an entire causal graph is the primary task for the existing research, the target of interest is the future excess returns of stocks in the cross-section for empirical asset pricing, so we formulate this objective as target-specific causal ranking.

Motivated by this formulation, we propose CausalRank, a foundation model for causal discovery and ranking, which is trained to reveal the pricing mechanisms in stock markets. Its central algorithmic design evaluates each candidate-target relationship against a candidate specific conditional context formed by the remaining variables, allowing the model to separate direct parent evidence from target related information shared by correlated candidates. The resulting parent-aware evidence is used to jointly estimate direct causal existence and intervention defined effect strength. In the financial context, as illustrated in Figure 1, the candidates are stock-level characteristics, and the target is future stock excess returns. CausalRank performs direct parent identification and effect based ranking in a single forward pass.

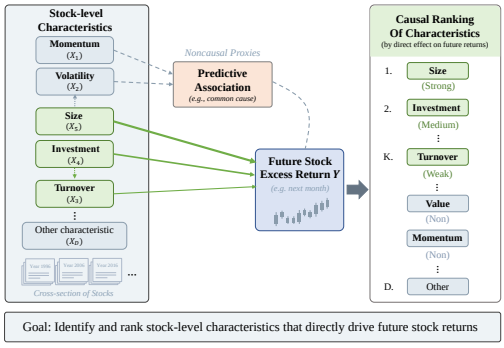


Figure 1: Target-specific causal ranking.

Our main contributions are summarized as follows.

- We formulate target-specific causal ranking as a new causal learning task that unifies direct parent identification with ranking based on the strength of underlying causal effect.
- We develop CausalRank, a foundation model that is trained for empirical asset pricing. By comparing each candidate-target relationship with a leave-one-out conditional context, it constructs parent-aware evidence to distinguish direct causal drivers from correlated noncausal variables and rank the identified parents by intervention defined effect strength.
- We construct fully synthetic and semi-synthetic financial panels to demonstrate CausalRank's ability to identify and rank direct causal parents and distinguish them from highly predictive noncausal proxies. We further show the merits of CausalRank in identifying the characteristics that drive the cross-section of stock returns in the U.S. equity market.

2 Related Work

Traditional causal discovery methods typically apply conditional independence tests, structure search, or continuous structure optimization separately to each dataset. Although these methods can recover causal structures under appropriate assumptions, structure search, distribution estimation, and graph pruning often need to be repeated for new datasets, leading to computational costs in high-dimensional settings (Bello et al., 2022; Rolland et al., 2022). Causal foundation models provide a different paradigm by pretraining on large-scale simulated causal tasks and transferring the learned causal inference capability to new data (Lorch et al., 2022; Wu et al., 2025; Ma et al., 2026; Thompson et al., 2026). DAG-FM further decomposes full DAG recovery into leaf node prediction and target-parent prediction, and recursively reconstructs the causal graph Chen et al. (2026). These studies show that simulated causal tasks can provide transferable inductive biases for causal inference.

Local causal discovery and causal feature selection instead focus on the local structure or causal variables associated with a specified target (Bodik & Chavez-Demoulin, 2026; Aliferis et al., 2010).

Other methods identify causal predictors through invariance across environments (Quinzan et al., 2023). TCD-DL trains neural models on simulated data to identify potential causal variables for a target without full graph recovery, and scales this paradigm to high-dimensional systems (Kim et al., 2025). These methods usually return a set of causal variables, direct parents, or causal confidence scores. Less attention has been given to ranking multiple direct parents by their structural effect strength. Rank-Learner (Arno et al., 2026) shows that directly optimizing the relative ordering of causal effects can differ from accurately estimating effects first and ranking them afterward. Its focus, however, is treatment effect ranking across individuals, while we rank the direct causal effects of multiple candidate variables on the same target.

Machine learning has also been widely used for cross-sectional return prediction and asset pricing (Gu et al., 2020; Chen et al., 2024; Jensen et al., 2026). High-dimensional asset pricing studies address large sets of stock-level characteristics through latent factor extraction and factor selection. IPCA uses stock-level characteristics to model time varying factor loadings (Kelly et al., 2019), while (Feng et al., 2020) study the incremental pricing information of new factors after controlling for a high-dimensional set of existing factors. These methods mainly rely on predictive accuracy, statistical significance, or pricing explanatory power, which do not necessarily identify direct causal drivers of future returns. Recent financial studies have begun to use causal relationships for feature selection and show that causal features can improve stock prediction and investment performance (Zhang et al., 2014; Zhou et al., 2026).

3 PRELIMINARIES

3.1 FUNCTIONAL CAUSAL MODEL

We use a Functional Causal Model (FCM) to characterize the structural relationships between candidate variables and the target variable. For the variable set $\mathcal{V} = \{X_1, \ldots, X_D, Y\}$, an FCM represents each variable as a function of its direct parents and exogenous noise, written as $V_k = f_k(Pa(V_k), U_k)$, where $Pa(V_k)$ and U_k denote the direct parents and exogenous noise of V_k.

We focus on the local causal structure of the target variable Y, whose structural equation is

$$Y = f_Y\left(X_{Pa(Y)}, U_Y\right), \tag{1}$$

where $Pa(Y) \subseteq \{X_1, \ldots, X_D\}$ denotes the direct parents of Y, and $X_{Pa(Y)}$ denotes the corresponding structural input. For each candidate variable X_j, we define the direct parent label as $z_j = \mathbf{1}[X_j \in Pa(Y)]$. Thus, $z_j = 1$ indicates that X_j directly enters the structural equation of Y, while $z_j = 0$ indicates that it is not a direct parent.

The FCM also characterizes causal effects through structural interventions. The intervention $do(X_j = x)$ replaces the generating mechanism of X_j with the constant x while keeping the other structural equations unchanged. We focus on the direct structural effect of X_j on Y. When comparing intervention levels, we therefore hold the other direct parents of Y fixed and use this intervention to define direct causal strength in Section 3.3.

3.2 PROBLEM SETUP

Consider a financial panel containing N stocks. Within a historical window of length T, each stock has D candidate financial features. We denote the observed feature tensor as $\mathbf{X} \in \mathbb{R}^{N\times T\times D}$, where $\mathbf{X}_{i,t,j}$ is the observation of candidate variable X_j for stock i at time t. Here, N, T, and D denote the number of cross-sectional assets, the historical window length, and the number of candidate features. For each stock and time point, the target observation corresponding to the future t' period excess return is defined as $\mathbf{Y}_{i,t} = R^e_{i,t+t'}$. The realized target returns within the historical window are denoted by $\mathbf{Y} \in \mathbb{R}^{N\times T}$. When $t' = 1$, $\mathbf{Y}_{i,t}$ corresponds to the next period excess return. During inference, the model receives the financial panel $\mathcal{P} = (\mathbf{X}, \mathbf{Y})$. Here, X_j and Y denote random variables in the FCM, while $\mathbf{X}$ and $\mathbf{Y}$ denote their observations in the financial panel. Let $\mathcal{X} = \{X_1, \ldots, X_D\}$ denote the candidate variable set, with $Pa(Y) \subseteq \mathcal{X}$.

Our goal is to identify $Pa(Y)$ from $\mathcal{X}$ and compare the causal effect strengths of different parents on Y. Within each input window, we assume that the local causal structure and generating mechanism of the target remain unchanged, while the observational distributions and correlation structures of

the candidate variables may vary over time. We define this problem as target-specific causal factor ranking. Given a financial panel $\mathcal{P} = (\mathbf{X}, \mathbf{Y})$, the model learns the mapping

$$F_\theta : (\mathbf{X}, \mathbf{Y}) \to s, \tag{2}$$

where $s = (s_1, \ldots, s_D) \in \mathbb{R}^D$ denotes the causal ranking scores and induces the ordering

$$X_{\pi_1} \succ X_{\pi_2} \succ \cdots \succ X_{\pi_D}, \qquad s_{\pi_1} \geq s_{\pi_2} \geq \cdots \geq s_{\pi_D}. \tag{3}$$

CausalRank therefore performs direct causal identification and ranking with respect to a specified target variable without recovering the full causal graph among the candidate variables.

3.3 Direct Causal Effect and Ranking Target

The parent label z_j indicates whether a candidate variable directly enters the target generating mechanism. Since different direct parents may have different causal effect strengths, we further define a continuous measure of direct causal strength using the known structural equation. For each candidate variable X_j, we select low and high intervention levels from its empirical distribution in the current causal environment, denoted by $q_j^L = Q_{0.2}(X_j)$ and $q_j^H = Q_{0.8}(X_j)$. Here, $Q_\alpha(X_j)$ denotes the empirical α-quantile of X_j in the current task. For $X_j \in Pa(Y)$, while holding the other direct parents and the exogenous noise U_Y fixed, we define the structural intervention difference as

$$\Delta_j = f_Y\left(X_{Pa(Y)\setminus\{X_j\}}, q_j^H, U_Y\right) - f_Y\left(X_{Pa(Y)\setminus\{X_j\}}, q_j^L, U_Y\right). \tag{4}$$

The two structural evaluations hold the remaining direct parents and U_Y fixed and differ only in the intervention on X_j, Δ_j isolates the direct structural effect of X_j on Y. Its sign indicates the effect direction. We further define the direct causal strength as $\tau_j = \mathbb{E}\left[|\Delta_j|\right]$, where the expectation is taken over the other parents and exogenous noise. For $X_j \notin Pa(Y)$, we define $\tau_j = 0$. Thus, τ_j is an intervention defined measure of direct causal strength and serves as the supervision signal for ranking.

Given $\tau = (\tau_1, \ldots, \tau_D)$, the ground-truth causal ranking is induced by τ. For any two candidate variables X_j and X_k,

$$X_j \succ X_k \quad \Longleftrightarrow \quad \tau_j > \tau_k. \tag{5}$$

Equivalently, let $\pi^* = \mathrm{argsort}(-\tau)$. Then, $\tau_{\pi_1^*} \geq \tau_{\pi_2^*} \geq \cdots \geq \tau_{\pi_D^*}$.

No fixed internal ordering is required for candidate variables with the same causal strength. CausalRank learns continuous scores s whose induced ordering recovers the ground-truth ranking defined by τ, while the parent labels z provide explicit supervision for the existence of direct causal relationships.

4 CausalRank: A Ranking Causal Foundation Model

We propose CausalRank to identify the direct parents of a target variable from an observational panel and rank them according to their causal effects. CausalRank evaluates each candidate-target relationship in the context of the remaining candidates to distinguish direct parents from correlated noncausal variables, and then jointly estimates direct parent existence and causal effect strength. In financial applications, the model takes an observed feature panel $\mathbf{X} \in \mathbb{R}^{N\times T\times D}$ and historical target returns $\mathbf{Y} \in \mathbb{R}^{N\times T}$ as input, and outputs causal ranking scores $s = (s_1, \ldots, s_D)$ for all candidate financial features. The overall architecture is illustrated in Figure 2. Implementation details are provided in Appendix A.

4.1 Model Architecture

4.1.1 Panel Representation Encoder

The Panel Representation Encoder first jointly encodes cross-sectional candidate variables $\mathbf{X}$ and target returns $\mathbf{Y}$ at each time pointand aggregates their representations over the historical window. We further construct target-factor statistics from several nonlinear correlations and fuse them with the panel representations to retain explicit candidate-target information. The encoder outputs candidate representations $\bar{H} = (\bar{h}_1, \ldots, \bar{h}_D)$ and the target representation h_Y, summarized as

$$\Psi_{\mathrm{enc}}(\mathbf{X}, \mathbf{Y}) \to (H, h_Y) \tag{6}$$

Full construction details are provided in the Appendix A.1.

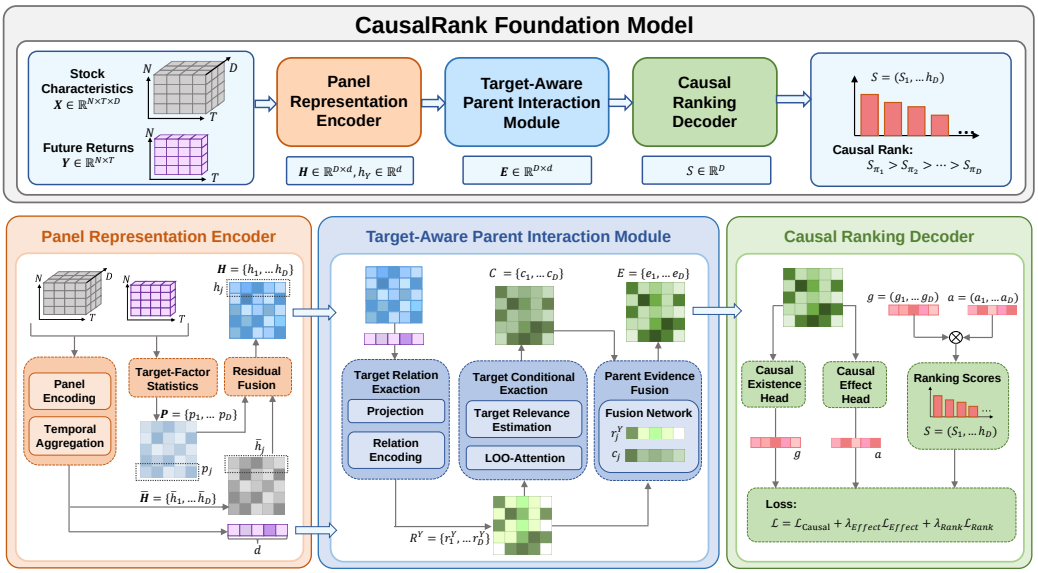


Figure 2: Overview of the CausalRank Fondation Model.

4.1.2 Target-Aware Parent Interaction

The Target-Aware Parent Interaction Module is designed to distinguish direct parents from correlated nonparent variables. It first models how each candidate relates to the target and then evaluates this relationship in the context of the remaining candidates.

Target Relation Extraction We first constructs a relation representation for each candidate X_j and the target Y. It combines the candidate and target representations with their difference, element-wise interaction, and gated target information v_j^Y

$$r_j^Y = \phi_Y\left(h_j, h_Y, h_j - h_Y, h_j \odot h_Y, v_j^Y\right). \tag{7}$$

The resulting r_j^Y captures the relationship between X_j and Y. However, a correlated nonparent can exhibit a similar target relationship when it shares information with a true parent.

To provide additional conditional information, the model extracts conditional incremental evidence directly from the raw observational panel. This includes the marginal correlation with the target, the ridge regression coefficient when all candidate variables are included jointly, and the partial correlation with the target after controlling for the remaining candidates. These statistics and valid observation information are mapped to d_j^{cond} and fused with the target relation representation as $\widetilde{r}_j^Y = \text{Norm}(r_j^Y + d_j^{\text{cond}})$.

Target Conditional Competition CausalRank next evaluates each candidate against the target related information carried by the remaining candidates. Let $\widetilde{R}^Y = (\widetilde{r}_1^Y, \ldots, \widetilde{r}_D^Y)$. A learnable relevance function estimates target relevance as $\alpha_k = \sigma(f_{\text{rel}}(\widetilde{r}_k^Y))$, and defines the weighted representation $\hat{r}_k^Y = \alpha_k \widetilde{r}_k^Y$. For candidate X_j, $\hat{R}_{-j}^Y = \{\hat{r}_k^Y \mid k \neq j\}$ denotes the representations of other candidates. The model uses the current candidate as the query and reads context from the remaining candidates

$$c_j = \text{Attention}\left(\widetilde{r}_j^Y, \hat{R}_{-j}^Y, \hat{R}_{-j}^Y\right). \tag{8}$$

Because X_j is excluded from $\hat{R}_{-j}^Y$, the context c_j contains only target related information provided by the other candidates. Comparing $\widetilde{r}_j^Y$ with c_j indicates whether the relationship between X_j and Y remains distinctive in the presence of competing variables.

Parent Evidence Fusion CausalRank finally combines the candidate-target relation with the competition context and explicit conditional evidence to construct parent-aware evidence

$$e_j = \phi_P\left(\widetilde{r}_j^Y, c_j, \widetilde{r}_j^Y - c_j, \widetilde{r}_j^Y \odot c_j, d_j^{\text{cond}}\right). \tag{9}$$

The difference $\widetilde{r}_j^Y - c_j$ captures evidence that remains distinctive to the current candidate relative to other target related variables. The interaction $\widetilde{r}_j^Y \odot c_j$ captures how the candidate-target relationship aligns with the conditional context. Together with d_j^{cond}, these terms form the parent-aware representation e_j used for direct parent identification and effect-strength estimation. The final output is $E = (e_1, \ldots, e_D)$.

4.1.3 Causal Ranking Decoder

The parent-aware representation e_j is passed to the Causal Ranking Decoder. For each candidate variable, the model predicts a direct causal existence score $g_j = \sigma(f_g(e_j))$, where $g_j \in [0, 1]$ measures the support for the direct relationship $X_j \to Y$. The model then predicts a non-negative causal strength component $a_j = \text{Softplus}(f_a(e_j))$ and defines the final ranking score as $s_j = g_j a_j$. Candidates with weak direct causal support are suppressed by g_j, while a_j differentiates relative effect strengths among candidates with stronger parent support.

4.2 Causal Task Generation

To obtain training tasks with causal supervision derived from known structural mechanisms, we construct both fully synthetic and financial semi-synthetic causal tasks. For fully synthetic tasks, we generate an observed panel $\mathbf{X} \in \mathbb{R}^{N \times T \times D}$ from a predefined distribution over FCMs, and sample the parent set $Pa(Y)$ together with the target structural equation f_Y. All observations within the same task share the same target causal mechanism, and the target is generated as $\mathbf{Y}_{i,t} = f_Y(\mathbf{X}_{i,t,Pa(Y)}, U_{i,t}^Y)$. Across tasks, we vary contemporaneous DAGs, structural parameters, node nonlinearities, temporal persistence, target mechanisms, and correlation patterns.

To reduce the distribution gap between fully synthetic data and real financial data, we further construct financial semi-synthetic tasks. We directly use real financial feature panels and control only the target generating mechanism. By sampling $Pa(Y)$ from real candidate variables and generating $\mathbf{Y}$ synthetically, these tasks preserve real financial covariate distributions while providing exact causal ground truth for the target mechanism.

Since $Pa(Y)$ and f_Y are known for every task, we obtain exact structural supervision, with $z_j = \mathbf{1}[X_j \in Pa(Y)]$ identifying direct parent set and τ_j computed from the intervention contrast in 4. Training follows two stages: pretraining on large-scale fully synthetic tasks and fine-tuning on financial semi-synthetic tasks. The candidate systems naturally contain correlated nonparents, while controlled proxies are explicitly constructed for correlation-trap evaluation. Test-task generation and sampling are detailed in Appendix B.2.

4.3 Causal Ranking Training

For each training task, CausalRank receives supervision $(\mathbf{X}, \mathbf{Y}, z, \tau)$. Given the parent-aware representation e_j, the Causal Ranking Decoder predicts the direct causal existence probability g_j and the non-negative causal strength a_j, and defines the final ranking score as $s_j = g_j a_j$.

Training combines three complementary objectives. $\mathcal{L}_{\text{causal}}$ uses the parent labels z to identify whether a candidate directly enters the target generating mechanism. $\mathcal{L}_{\text{effect}}$ supervises the predicted strength of true direct parents using the ground-truth intervention strength τ. $\mathcal{L}_{\text{rank}}$ directly constrains the relative ordering of candidates with different causal strengths. The overall objective is

$$\mathcal{L} = \mathcal{L}_{\text{causal}} + \lambda_{\text{effect}}\mathcal{L}_{\text{effect}} + \lambda_{\text{rank}}\mathcal{L}_{\text{rank}}, \tag{10}$$

where λ_{effect} and λ_{rank} control the contributions of effect-strength and ranking supervision. Together, these objectives train CausalRank to infer direct parent existence and effect strength from observational panels. After training, the model requires only a new observational panel $(\mathbf{X}, \mathbf{Y})$ to rank candidate variables, without access to parent labels, intervention strengths, or structural equations.

Table 1: Standard causal identification, ranking, and inference results, rounded to two decimals.

Method	Full AUPRC	Full NDCG@20	Full Parent ρ	Semi AUPRC	Semi NDCG@20	Semi Parent ρ
Pearson	0.45 ± 0.04	0.46 ± 0.04	0.26 ± 0.04	0.22 ± 0.02	0.16 ± 0.04	-0.05 ± 0.14
Correlation	0.49 ± 0.04	0.50 ± 0.05	0.36 ± 0.06	0.59 ± 0.07	0.60 ± 0.05	0.44 ± 0.02
MI	0.44 ± 0.02	0.45 ± 0.03	0.33 ± 0.05	0.39 ± 0.03	0.40 ± 0.03	0.29 ± 0.13
Lasso	0.97 ± 0.00	0.94 ± 0.01	0.66 ± 0.04	0.25 ± 0.03	0.20 ± 0.05	0.05 ± 0.09
LightGBM	0.93 ± 0.01	0.89 ± 0.00	0.61 ± 0.02	0.73 ± 0.04	0.73 ± 0.04	0.41 ± 0.08
DirectLiNGAM	0.76 ± 0.05	0.75 ± 0.02	0.41 ± 0.03	0.26 ± 0.01	0.22 ± 0.03	-0.01 ± 0.08
NOTEARS	0.69 ± 0.01	0.70 ± 0.01	0.34 ± 0.07	0.24 ± 0.02	0.18 ± 0.04	0.11 ± 0.06
TCD-DL	0.26 ± 0.00	0.22 ± 0.02	0.03 ± 0.05	0.23 ± 0.03	0.18 ± 0.04	-0.01 ± 0.16
CausalRank	$\mathbf{0.98 \pm 0.00}$	$\mathbf{0.96 \pm 0.00}$	$\mathbf{0.81 \pm 0.02}$	$\mathbf{0.77 \pm 0.01}$	$\mathbf{0.78 \pm 0.02}$	$\mathbf{0.48 \pm 0.05}$
Existence-Only	0.93 ± 0.01	0.90 ± 0.00	0.63 ± 0.04	0.67 ± 0.02	0.60 ± 0.02	0.40 ± 0.08

5 EXPERIMENTS

We evaluate CausalRank in three settings: causal identification and ranking on financially calibrated synthetic and semi-synthetic data, correlation-trap tests with highly correlated proxies, and out-of-sample evaluation on U.S. stock data. We use $T = 60$ and $D = 94$. Details are provided in Section 3, Section 4 and Appendix B.

5.1 STANDARD CAUSAL IDENTIFICATION AND CAUSAL RANKING

Experimental Protocol. We evaluate direct parent identification and causal strength ranking on independent fully synthetic data and Financial Semi-Synthetic test sets, with $N = 64$ and $K_{\mathrm{pa}} = 20$ in the main setting. Fully synthetic evaluates generalization to unseen synthetic causal environments, while financial semi-synthetic retains the joint distribution of real financial features. Both test sets provide exact direct parent labels and intervention defined causal strengths from their known target mechanisms. We compare CausalRank with baselines (Shimizu et al., 2011; Zheng et al., 2018; Kim et al., 2025; Ke et al., 2017), together with CausalRank-Existence.

Main Results. CausalRank achieves the best parent identification and causal ranking performance on both test sets. On fully synthetic data, its AUPRC, NDCG@20, and Parent Spearman reach 0.9835, 0.9647, and 0.8093, respectively. In particular, NDCG@20 and Parent Spearman improve by approximately 2.9% and 22.8% over the strongest baselines. This result shows that the model can not only identify direct parents, but also recover their relative effect strengths. On Semi-Synthetic data, CausalRank achieves an AUPRC of 0.7666, an NDCG@20 of 0.7754, and a Parent Spearman of 0.4847, outperforming all compared methods. The semi-synthetic results show that the same causal ranking framework remains effective when its task distribution is fitted to real financial covariates, providing a practical way to adapt CausalRank to domain-specific observational distributions. Although TCD-DL performs strongly in large-scale targeted cause discovery, its training objective neither distinguishes direct parents from indirect causes nor supervises causal effect strength. Consequently, despite calibrating its continuous model input configuration, it still not performs well on our task.

Effect Supervision and Efficiency. The difference between CausalRank and CausalRank-Existence further demonstrates the role of continuous effect supervision. On synthetic data, the full model improves NDCG@20 by approximately 6.9% and Parent Spearman by approximately 27.8% relative to the variant trained only with parent labels. This suggests that continuous intervention strength provides additional information for ranking different direct parents. Meanwhile, CausalRank requires only a single forward pass to generate scores for all candidate variables. As shown in Table 2, average inference time on synthetic data is only 0.042 seconds, which is lower than other methods.

Table 2: Mean latency.

Method	Full(s)	Semi(s)
Correlation	0.14	0.06
Lasso	0.10	0.08
LightGBM	0.94	0.48
NOTEARS	300.89	361.77
TCD-DL	0.10	0.11
CausalRank	**0.04**	**0.03**

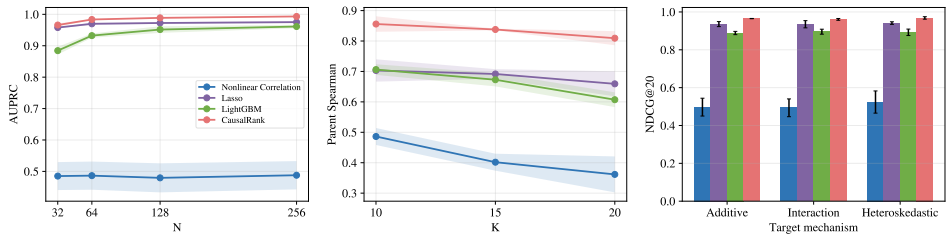


Figure 3: Generalization performance under different task scales and causal mechanisms.

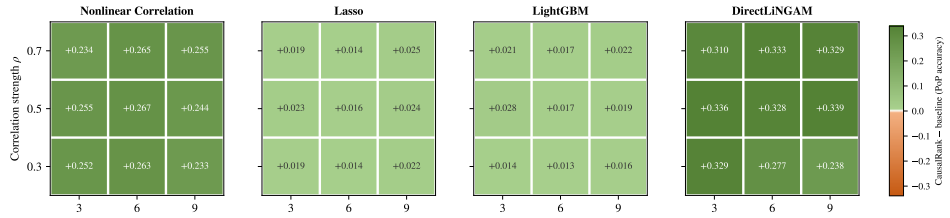


Figure 4: Results for single-parent proxies.Cells report CausalRank's PoP advantage over baselines.

Generalization across Causal Environments. We further vary the cross-sectional size N, the number of direct parents K_{pa}, and the target generating mechanism. The results are shown in Figure 3. CausalRank maintains the highest AUPRC across all sample sizes $N \in \{32, 64, 128, 256\}$ and the highest Parent Spearman at every evaluated parent count, remaining above 0.80 at $K_{\text{pa}} = 20$. Under the additive, interaction, and heteroscedastic mechanisms, its NDCG@20 values are 0.9650, 0.9608, and 0.9684, respectively. CausalRank remains stable across causal tasks. The exact controlled variables and task sampling procedures are provided in Appendix B.2.

5.2 Correlation Trap Experiment

Experimental Protocol. We construct correlation trap tasks to test whether CausalRank can distinguish true direct parents from highly correlated nonparent proxies. We fix $N = 128$ and $K_{\text{pa}} = 20$, and vary the number of proxies $P \in \{3, 6, 9\}$ and proxy strength $\rho \in \{0.30, 0.50, 0.70\}$. Each proxy is correlated with a true parent but does not enter the target structural equation, so $z_p = 0$ and $\tau_p = 0$.

Main Results. As the correlation between proxy variables and true parents increases, the ability of predictive methods to distinguish them gradually decreases. For example, when ρ increases from 0.30 to 0.70, the PoP of Lasso decreases from 98.15% to 96.99%, while that of LightGBM decreases from 98.61% to 96.76%. In comparison, CausalRank achieves PoP values of 100.00%, 100.00%, and 98.84% at the three correlation strengths. Under the strongest correlation setting with $\rho = 0.70$, its PoP is 1.85 percentage points higher than that of the best non-CausalRank baseline. When $\rho = 0.70$ and $P = 3$, CausalRank achieves a Proxy FPR@20 of 2.08%, which is 40.1% lower than that of the best non-CausalRank baseline on this metric. Its average Rank Gap reaches 39.20. Proxy FPR provides complementary evidence. Together with the positive Rank Gap, it shows that CausalRank places parents above their proxies while admitting fewer proxies into the Top-20 set. These results show that CausalRank learns to distinguish direct causal parents from statistically informative nonparent variables across different proxy structures and correlation strengths Extended results with different proxy counts and multi-parent aggregate proxies are provided in Appendix C.2.

5.3 Real Financial Experiment

Experimental Protocol. We evaluate the real-data rankings on the U.S. stocks using the annual recursive design of Gu et al. (2020), with 1957–1974 for initial training, 1975–1986 for validation, and 1987–2016 for out-of-sample testing. Feature rankings are updated annually using only historical data available at time t, and the Top-20 features are used for the subsequent 12 months. All ranking methods share the same annually updated Elastic Net (Zou & Hastie, 2005) downstream predictor.

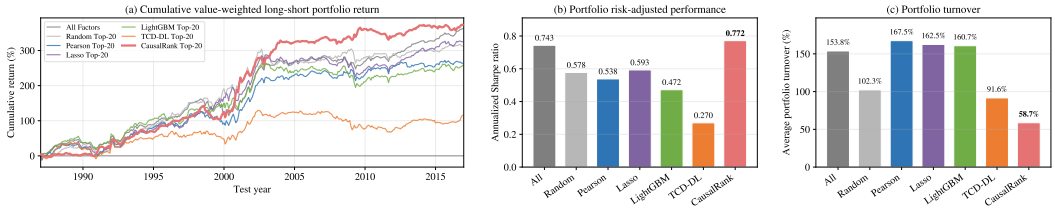


Figure 5: Out-of-sample economic performance on the U.S. stock market from 1987 to 2016. The left panel reports cumulative returns of value-weighted decile 10–1 long-short portfolios. The middle panel reports the Sharpe ratio, and the right panel reports drift-adjusted average monthly turnover.

Table 3: Performance on financial data (percentage for R^2_{OOS}, Return and Turnover).

Method	R^2_{OOS}	Sharpe	Return	Turnover
All Factors	**0.26**	0.74	12.08	153.80
Random	0.23	0.58	10.42	102.26
Pearson	0.24	0.54	8.80	167.48
Lasso	0.23	0.59	10.83	162.49
LightGBM	0.23	0.47	8.58	160.75
TCD-DL	0.22	0.27	3.84	91.63
CausalRank	0.19	**0.77**	**12.43**	**58.72**

Predictive and Economic Performance. CausalRank Top-20 achieves a positive out-of-sample R^2 of 0.19%, showing that 20 highly ranked features retain predictive information. Its long-short portfolio achieves a Sharpe ratio of 0.77, an annualized return of 12.43%, and an average monthly turnover of 58.72%, delivering the strongest risk-adjusted performance among the Top-20 methods. Compared with Lasso, CausalRank improves the Sharpe ratio and annualized return by approximately 30.1% and 14.7%, respectively. It reduces average monthly turnover by 35.9% relative to TCD-DL. Compared with using all 94 features, CausalRank achieves a higher Sharpe ratio and annualized return using only 20 features, while reducing turnover by 61.8%. These results show that causal ranking extracts a compact characteristic set with higher risk-adjusted returns and substantially lower trading intensity.

Performance over Time. Figure 5 shows that CausalRank also performs more consistently over the 1987–2016 out-of-sample period. Its cumulative long-short return reaches 372.91%, slightly above the 362.28% of All Factors and higher than all other Top-20 methods. The main difference comes from weaker downside performance rather than substantially higher upside returns. CausalRank earns an average return of 18.56% in positive-return years, close to 18.19% for All Factors, but its average loss in negative-return years is only -4.43%, compared with -12.40% for All Factors. In all six years when the All-Factors portfolio records a negative return, CausalRank performs better and still delivers an average return of 0.60%. This pattern is also reflected at longer horizons. 36-month rolling cumulative return of CausalRank is positive in 92.0% of the available windows, versus 84.3% for All Factors, while its maximum drawdown is lower at 31.99% compared with 37.93%. Together with a Sharpe ratio of 0.77 and average monthly turnover of 58.72%, these results suggest that the compact feature set selected by CausalRank retains more stable information for cross-sectional expected returns while requiring less portfolio rebalancing.

6 Conclusion

In this paper, we propose CausalRank to formulate direct parent identification and causal effect estimation as a unified target-specific causal ranking task. Experimental results show that CausalRank outperforms the compared methods in standard causal identification, causal strength ranking, and correlation-trap tests. In the application to empirical asset pricing, the stock characteristics selected by CausalRank achieve economically significant risk-adjusted returns with substantially lower portfolio turnover. Although CausalRank produces continuous scores for all candidate variables, several limitations still remain. In particular, the current framework requires a prespecified number of parents, K_{pa}, to select the candidate variables as direct causal features. Because the score is not calibrated across tasks, CausalRank lacks a selection criterion that is auto-determined by the data for separating parents from nonparents. From the perspective of empirical asset pricing, our future work will explore portfolio optimization utilizing the causal ranking scores of the causal characteristics.

AI USE STATEMENT

We used AI tools for language polishing, grammar checking, and a limited amount of translation during the preparation of this manuscript. All AI-assisted edits and translations were reviewed by the authors, who take full responsibility for the final content of the paper.

REFERENCES

Constantin F Aliferis, Alexander Statnikov, Ioannis Tsamardinos, Subramani Mani, and Xenofon D Koutsoukos. Local causal and markov blanket induction for causal discovery and feature selection for classification part i: algorithms and empirical evaluation. *Journal of Machine Learning Research*, 11(1), 2010.

Henri Arno, Dennis Frauen, Emil Javurek, Thomas Demeester, and Stefan Feuerriegel. Rank-learner: Orthogonal ranking of treatment effects. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=xM56OFSUaN.

Kevin Bello, Bryon Aragam, and Pradeep Ravikumar. Dagma: Learning dags via m-matrices and a log-determinant acyclicity characterization. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 8226–8239. Curran Associates, Inc., 2022. doi: 10.52202/068431-0598. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/36e2967f87c3362e37cf988781a887ad-Paper-Conference.pdf.

Juraj Bodik and Valérie Chavez-Demoulin. Structural restrictions in local causal discovery: identifying direct causes of a target variable. *Biometrika*, 113(1):asaf042, 2026.

Philippe Brouillard, Sébastien Lachapelle, Alexandre Lacoste, Simon Lacoste-Julien, and Alexandre Drouin. Differentiable causal discovery from interventional data. *Advances in Neural Information Processing Systems*, 33:21865–21877, 2020.

Luyang Chen, Markus Pelger, and Jason Zhu. Deep learning in asset pricing. *Management Science*, 70(2):714–750, 2024.

Yikang Chen, Zhengkang Guan, Haoyuan Qian, Xingxuan Zhang, Peng Cui, Yi Yang, Fei Wu, and Kun Kuang. Dag-fm: A foundation model for causal discovery under heterogeneous causal mechanisms. *arXiv preprint arXiv:2607.11510*, 2026.

Guanhao Feng, Stefano Giglio, and Dacheng Xiu. Taming the factor zoo: A test of new factors. *The Journal of Finance*, 75(3):1327–1370, 2020.

Joachim Freyberger, Andreas Neuhierl, and Michael Weber. Dissecting characteristics nonparametrically. *The Review of Financial Studies*, 33(5):2326–2377, 2020.

Jeremiah Green, John RM Hand, and X Frank Zhang. The characteristics that provide independent information about average us monthly stock returns. *The Review of Financial Studies*, 30(12): 4389–4436, 2017.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273, 2020.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. Autoencoder asset pricing models. *Journal of Econometrics*, 222(1):429–450, 2021.

Christina Heinze-Deml, Jonas Peters, and Nicolai Meinshausen. Invariant causal prediction for nonlinear models. *Journal of Causal Inference*, 6(2):20170016, 2018.

Biwei Huang, Kun Zhang, Jiji Zhang, Joseph Ramsey, Ruben Sanchez-Romero, Clark Glymour, and Bernhard Schölkopf. Causal discovery from heterogeneous/nonstationary data. *Journal of Machine Learning Research*, 21(89):1–53, 2020.

Shimeng Huang, Niklas Pfister, and Jack Bowden. Sparse causal effect estimation using two-sample summary statistics in the presence of unmeasured confounding. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025. URL https://openreview.net/forum?id=ipopfqHYxN.

Theis Ingerslev Jensen, Bryan Kelly, Semyon Malamud, and Lasse Heje Pedersen. Machine learning and the implementable efficient frontier. *The Review of Financial Studies*, pp. hhag022, 2026.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 2017.

Nan Rosemary Ke, Silvia Chiappa, Jane X Wang, Jorg Bornschein, Anirudh Goyal, Melanie Rey, Theophane Weber, Matthew Botvinick, Michael Curtis Mozer, and Danilo Jimenez Rezende. Learning to induce causal structure. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=hp_RwhKDJ5.

Bryan T Kelly, Seth Pruitt, and Yinan Su. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524, 2019.

Jang-Hyun Kim, Claudia Skok Gibbs, Sangdoo Yun, Hyun Oh Song, and Kyunghyun Cho. Large-scale targeted cause discovery via learning from simulated data. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=NVgy29IQw8.

Sébastien Lachapelle, Philippe Brouillard, Tristan Deleu, and Simon Lacoste-Julien. Gradient-based neural DAG learning. In *NeurIPS 2019 Workshop on Solving Inverse Problems with Deep Networks*, 2019. URL https://openreview.net/forum?id=ryl6nX398r.

Markus Leippold, Qian Wang, and Wenyu Zhou. Machine learning in the chinese stock market. *Journal of Financial Economics*, 145(2):64–82, 2022.

Zi-Rong Li, Si-Yang Liu, Tian-Zuo Wang, and Han-Jia Ye. Tabcausal: Pretraining across causal environments for tabular causal discovery. *arXiv preprint arXiv:2605.31156*, 2026.

Lars Lorch, Scott Sussex, Jonas Rothfuss, Andreas Krause, and Bernhard Schölkopf. Amortized inference for causal structure learning. *Advances in Neural Information Processing Systems*, 35: 13104–13118, 2022.

Yuchen Ma, Dennis Frauen, Emil Javurek, and Stefan Feuerriegel. Foundation models for causal inference via prior-data fitted networks. In *International Conference on Learning Representations*, volume 2026, pp. 79065–79098, 2026.

Francesco Quinzan, Ashkan Soleymani, Patrick Jaillet, Cristian R Rojas, and Stefan Bauer. Drcfs: Doubly robust causal feature selection. In *International Conference on Machine Learning*, pp. 28468–28491. PMLR, 2023.

Mateo Rojas-Carulla, Bernhard Schölkopf, Richard Turner, and Jonas Peters. Invariant models for causal transfer learning. *Journal of Machine Learning Research*, 19(36):1–34, 2018.

Paul Rolland, Volkan Cevher, Matthäus Kleindessner, Chris Russell, Dominik Janzing, Bernhard Schölkopf, and Francesco Locatello. Score matching enables causal discovery of nonlinear additive noise models. In *International Conference on Machine Learning*, pp. 18741–18753. PMLR, 2022.

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021.

Mátyás Schubert, Tom Claassen, and Sara Magliacane. SNAP: Sequential non-ancestor pruning for targeted causal effect estimation with an unknown graph. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025. URL https://openreview.net/forum?id=0gEjlLdjK9.

Shohei Shimizu, Takanori Inazumi, Yasuhiro Sogawa, Aapo Hyvarinen, Yoshinobu Kawahara, Takashi Washio, Patrik O Hoyer, Kenneth Bollen, and Patrik Hoyer. Directlingam: A direct method for learning a linear non-gaussian structural equation model. *Journal of Machine Learning Research*, 12(Apr):1225–1248, 2011.

Ryan Thompson, He Zhao, Daniel M Steinberg, and Edwin V Bonilla. Arrow: A foundation model for causal discovery. *arXiv preprint arXiv:2605.07204*, 2026.

Christian Toth, Lars Lorch, Christian Knoll, Andreas Krause, Franz Pernkopf, Robert Peharz, and Julius Von Kügelgen. Active bayesian causal inference. *Advances in Neural Information Processing Systems*, 35:16261–16275, 2022.

Daniele Tramontano, Yaroslav Kivva, Saber Salehkaleybar, Mathias Drton, and Negar Kiyavash. Causal effect identification in liNGAM models with latent confounders. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=C1iNBLIClt.

Menghua Wu, Yujia Bao, Regina Barzilay, and Tommi Jaakkola. Sample, estimate, aggregate: A recipe for causal discovery foundation models. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=h434zx5SX0.

Xiangzhou Zhang, Yong Hu, Kang Xie, Shouyang Wang, EWT Ngai, and Mei Liu. A causal feature selection algorithm for stock prediction modeling. *Neurocomputing*, 142:48–59, 2014.

Xun Zheng, Bryon Aragam, Pradeep K Ravikumar, and Eric P Xing. Dags with no tears: Continuous optimization for structure learning. *Advances in Neural Information Processing Systems*, 31, 2018.

Jiamei Zhou, Hongxu Wu, and Shaoze Li. Information-theoretic causal feature selection via markov blanket discovery for stock return direction prediction: Evidence from china. *Entropy*, 28(8):847, 2026.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2):301–320, 2005.

A CausalRank Method Details

This appendix provides additional details of the CausalRank architecture, causal task construction, and training objectives.

A.1 Panel Representation Encoder

A.1.1 Cross-Sectional and Temporal Representations

Given the observed panel $\mathbf{X} \in \mathbb{R}^{N\times T\times D}$ and historical targets $\mathbf{Y} \in \mathbb{R}^{N\times T}$, the encoder first jointly processes the D candidate variables and the target variable at each time point. At time t, the cross-sectional encoder takes the joint table formed by all valid asset observations as input and outputs cross-sectional representations for the candidate variables and the target variable.

The model then independently aggregates the representation sequence of each candidate variable over the historical window. The temporal encoder introduces temporal positional information and models dependencies across different historical periods. A learnable temporal aggregation operation is then used to obtain the base representations of the candidate variables

$$\bar{H} = (\bar{h}_1, \ldots, \bar{h}_D) \tag{11}$$

and the target representation h_Y. This process allows the model to jointly use contemporaneous cross-sectional relationships and temporal dynamics without binding different assets or candidate variables to fixed ordering positions.

A.1.2 Target-Factor Statistics

Relying only on deep panel representations may weaken low-order dependency information with clear statistical meaning between candidate variables and the target. We therefore construct explicit target-factor statistics for each candidate variable.

For a candidate variable X_j, we consider the following four fixed transformations

$$\phi_1(x) = x, \tag{12}$$
$$\phi_2(x) = \tanh(2x), \tag{13}$$
$$\phi_3(x) = \mathrm{sign}(x)x^2, \tag{14}$$
$$\phi_4(x) = \sin(\pi x). \tag{15}$$

For each transformation, we use only observations for which both X_j and Y are valid and compute

$$\rho_{j,m} = \mathrm{Corr}\left(\phi_m(\mathbf{X}_j), \mathbf{Y}\right), \qquad m = 1, \ldots, 4. \tag{16}$$

In addition to the signed correlations, we retain their absolute values to represent both the direction and strength of the relationship. We also use r_j^{obs} to describe the valid observation coverage shared by the candidate variable and the target. We then construct

$$u_j = \left[\rho_{j,1}, \ldots, \rho_{j,4}, |\rho_{j,1}|, \ldots, |\rho_{j,4}|, r_j^{\mathrm{obs}}\right]. \tag{17}$$

These statistics are passed through a shared mapping

$$p_j = \phi_{\mathrm{stat}}(u_j), \tag{18}$$

and are fused with the panel representation through a residual connection

$$h_j = \mathrm{Norm}\left(\bar{h}_j + p_j\right). \tag{19}$$

The Panel Representation Encoder finally outputs $H = (h_1, \ldots, h_D)$ and h_Y. Importantly, this explicit statistical branch is computed only from $(\mathbf{X}, \mathbf{Y})$, which are available to the model at inference time. It does not use the direct parent labels z, the ground-truth effects τ, or any data-generating parameters.

A.2 TARGET-AWARE PARENT INTERACTION

The goal of Target-Aware Parent Interaction is to further determine which target related candidate variables exhibit conditional evidence that is more consistent with direct parent relationships. The module consists of target relation modeling, explicit conditional incremental evidence, candidate competition, and parent evidence fusion.

A.2.1 TARGET RELATION EXTRACTION

For the candidate representation h_j and target representation h_Y, the model first projects them into a shared relation space

$$q_j = W_q h_j, \tag{20}$$

$$k_Y = W_k h_Y. \tag{21}$$

It then constructs the element-wise relation

$$b_j = q_j \odot k_Y \tag{22}$$

and applies a target-aware gate

$$\gamma_j = \sigma(W_g b_j), \tag{23}$$

$$v_j^Y = \gamma_j \odot W_v h_Y. \tag{24}$$

The relation representation between the candidate variable and the target is obtained as

$$r_j^Y = \phi_Y \left(h_j, h_Y, h_j - h_Y, h_j \odot h_Y, v_j^Y \right). \tag{25}$$

This representation combines information from the candidate variable itself, the target, their difference, and their interaction.

A.2.2 CONDITIONAL INCREMENTAL EVIDENCE

Marginal target relationships alone are insufficient to distinguish true parents from candidate variables that share highly overlapping information with those parents. We therefore extract additional explicit conditional incremental evidence from the raw observational panel.

Within each task, the model constructs standardized candidate variables and target observations using valid data. We first compute the marginal relationship between each candidate variable and the target. We then include all valid candidate variables in a regularized linear projection and denote the coefficient associated with X_j by β_j. The general form is

$$\hat{\beta} = \left(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}} + \lambda I \right)^{-1} \widetilde{\mathbf{X}}^\top \widetilde{\mathbf{Y}}, \tag{26}$$

where $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{Y}}$ denote standardized data after valid-observation processing within the current task.

We further obtain conditional correlation information from the regularized joint covariance structure of the candidate variables and the target. Let Ω denote the corresponding precision matrix. The conditional correlation statistic between candidate variable X_j and target Y is written as

$$\rho_{jY}^{\text{partial}} = -\frac{\Omega_{jY}}{\sqrt{\Omega_{jj}\Omega_{YY}}}. \tag{27}$$

The conditional statistical representation of each candidate variable combines its marginal correlation, joint regression coefficient, conditional correlation, corresponding magnitude information, and valid observation coverage. A shared mapping is then used to obtain

$$d_j^{\text{cond}} = \phi_{\text{cond}} \left(u_j^{\text{cond}} \right). \tag{28}$$

This conditional evidence is then fused with the target relation representation

$$\widetilde{r}_j^Y = \text{Norm} \left(r_j^Y + d_j^{\text{cond}} \right). \tag{29}$$

As with the target-factor statistics, this branch uses only the currently observable $\mathbf{X}$, $\mathbf{Y}$, and their valid-observation information. It does not access z or τ.

A.2.3 TARGET-CONDITIONED COMPETITION

True parents and highly correlated nonparents may both have strong target relationships. The model therefore further introduces target-conditioned competition among candidate variables.

We first estimate the importance of each candidate representation as a target related competitor

$$\alpha_k = \sigma\left(f_{\mathrm{rel}}(\widetilde{r}_k^Y)\right), \tag{30}$$

$$\hat{r}_k^Y = \alpha_k \widetilde{r}_k^Y. \tag{31}$$

For the current candidate variable X_j, the model is allowed to read context only from the other candidate variables

$$c_j = \mathrm{Attention}\left(\widetilde{r}_j^Y, \hat{R}_{-j}^Y, \hat{R}_{-j}^Y\right), \tag{32}$$

where

$$\hat{R}_{-j}^Y = \left\{\hat{r}_k^Y \mid k \neq j\right\}. \tag{33}$$

The attention operation follows a leave-one-out constraint, so the current candidate variable cannot read its own representation through the competition branch. Therefore, c_j describes the conditional context provided by other target related candidate variables that compete with the current candidate variable.

A.2.4 PARENT EVIDENCE FUSION

Finally, the model combines the target relation of the current candidate variable, the conditional context formed by other candidate variables, and the explicit conditional incremental evidence

$$e_j = \phi_P\left(\widetilde{r}_j^Y, c_j, \widetilde{r}_j^Y - c_j, \widetilde{r}_j^Y \odot c_j, d_j^{\mathrm{cond}}\right). \tag{34}$$

Here, $\widetilde{r}_j^Y - c_j$ represents the distinctive information retained by the candidate variable relative to the competing context, while $\widetilde{r}_j^Y \odot c_j$ describes the interaction between the current target relationship and the overall conditional context. The final parent-aware representations are

$$E = (e_1, \ldots, e_D). \tag{35}$$

A.3 CAUSAL RANKING DECODER

The Causal Ranking Decoder maps each parent-aware representation e_j into two complementary outputs.

First, the existence branch produces

$$g_j = \sigma(f_g(e_j)), \tag{36}$$

where $g_j \in [0, 1]$ represents the degree of support provided by the current observational evidence for the direct relationship $X_j \to Y$.

Second, the effect branch produces a non-negative strength

$$a_j = \mathrm{Softplus}\left(f_a(e_j)\right). \tag{37}$$

The final causal ranking score is defined as

$$s_j = g_j a_j. \tag{38}$$

This decomposition assigns different roles to direct causal existence and effect strength. When a candidate variable has weak evidence of being a direct parent, its final score is suppressed by g_j. For variables with strong direct parent support, a_j further distinguishes their relative effect strengths. The model finally ranks all candidate variables in descending order according to s_j.

A.4 CAUSAL RANKING TRAINING OBJECTIVE

For each training task, let $m_j \in \{0, 1\}$ indicate whether candidate variable X_j has valid supervision information. The training objective contains three components corresponding to direct causal existence, direct causal strength, and relative causal ranking.

A.4.1 DIRECT CAUSAL EXISTENCE LOSS

The model uses the parent label $z_j = \mathbf{1}[X_j \in Pa(Y)]$ to supervise direct causal existence. Because direct parents and nonparents may be class-imbalanced, we use a positively weighted binary cross-entropy loss

$$\mathcal{L}_{\text{causal}} = -\frac{1}{\sum_{j=1}^{D} m_j} \sum_{j=1}^{D} m_j \left[\omega_+ z_j \log g_j + (1 - z_j) \log(1 - g_j)\right]. \tag{39}$$

Here, ω_+ controls the relative weight of positive-class supervision. This loss directly learns whether a candidate variable enters the target structural equation. Therefore, even when a nonparent variable has a strong statistical relationship with the target or a true parent, an excessively high existence score is still penalized by the supervision.

A.4.2 DIRECT CAUSAL STRENGTH LOSS

For true direct parents, we further supervise the effect branch using the direct causal strength τ_j computed from the known structural equation

$$\mathcal{L}_{\text{effect}} = \frac{\sum_{j=1}^{D} m_j z_j (a_j - \tau_j)^2}{\sum_{j=1}^{D} m_j z_j}. \tag{40}$$

This term is applied only to true direct parents. The effect branch of a nonparent variable is not separately required to regress to zero. Instead, its final ranking score is constrained by the existence gate and the ranking supervision. In this way, effect learning mainly focuses on differences in effect strength among true direct parents.

A.4.3 PAIRWISE CAUSAL RANKING LOSS

Effect-strength regression provides supervision on the absolute scale, but the final objective of CausalRank also requires recovery of the relative causal ordering among candidate variables. We therefore construct the ordered pair set

$$\mathcal{Q} = \{(j, k) \mid m_j = m_k = 1,\ \tau_j > \tau_k\}, \tag{41}$$

and use

$$w_{jk} = \tau_j - \tau_k \tag{42}$$

to represent the ground-truth effect difference between each pair.

Because the final score satisfies $s_j = g_j a_j > 0$, we define a strictly order-preserving score

$$\widetilde{s}_j = \log g_j + \log a_j = \log s_j. \tag{43}$$

The logarithmic transformation does not change the final ranking and allows the existence and effect-strength branches to participate in ranking optimization on a unified additive scale.

The weighted pairwise ranking objective is defined as

$$\mathcal{L}_{\text{rank}} = \frac{\sum_{(j,k)\in\mathcal{Q}} w_{jk} \log\left[1 + \exp\left(-(\widetilde{s}_j - \widetilde{s}_k)\right)\right]}{\sum_{(j,k)\in\mathcal{Q}} w_{jk}}. \tag{44}$$

Therefore, when $\tau_j > \tau_k$, the training objective encourages $s_j > s_k$. Since all nonparent variables are defined to have $\tau_j = 0$, true parents and nonparent variables also naturally form ranking constraints. Highly correlated nonparent variables therefore do not require a separate correlation-specific penalty.

A.4.4 Overall Objective

The final training objective is

$$\mathcal{L} = \mathcal{L}_{\text{causal}} + \lambda_{\text{effect}}\mathcal{L}_{\text{effect}} + \lambda_{\text{rank}}\mathcal{L}_{\text{rank}}, \tag{45}$$

where λ_{effect} and λ_{rank} control the contributions of effect supervision and ranking supervision, respectively.

The three losses correspond to three complementary objectives. $\mathcal{L}_{\text{causal}}$ learns the existence of direct parents, $\mathcal{L}_{\text{effect}}$ learns the direct effect strengths among true parents, and $\mathcal{L}_{\text{rank}}$ directly optimizes the relative causal ordering of candidate variables.

During training, the known structural equations can be used to obtain z and τ, but these quantities serve only as supervision signals. At inference time, the model does not access parent labels, ground-truth intervention effects, or structural generation parameters. For a new observational panel, the model only takes $(\mathbf{X}, \mathbf{Y})$ as input and outputs

$$s = (s_1, \ldots, s_D), \tag{46}$$

which is then used to perform direct causal ranking for the specified target variable.

B Additional Experimental Details

B.1 Common Evaluation Protocol, Baseline Scores, and Metric Definitions

Common Evaluation Protocol. This section describes the evaluation rules shared by the Full-FCM, Financial Semi-Synthetic, and correlation trap experiments. Each test task corresponds to an independent observational panel. Within the same task, the direct parent set of the target variable, the structural function, and its parameters remain fixed. Different stocks and months form repeated observations within the task. All methods receive exactly the same $(\mathbf{X}, \mathbf{Y})$. The direct parent labels z, intervention strengths τ, and target structural parameters are used only for evaluation and are not provided as inputs during testing.

Baseline Score Conversion. For consistent evaluation, the native outputs of different methods are converted into continuous scores such that a higher score indicates that a candidate variable is more likely to be a direct parent of the target. Pearson uses the absolute marginal correlation coefficient. Nonlinear Correlation uses the maximum absolute correlation after applying the predefined nonlinear transformations. Mutual Information uses mutual information for continuous variables. Lasso uses the absolute standardized regression coefficient. LightGBM uses feature gain importance. DirectLiNGAM and NOTEARS use the absolute directed edge weight from each candidate variable to the target node. TCD-DL uses its target-specific causal score. CausalRank uses the final ranking score s_j. All ranking metrics are computed directly from the continuous scores without prior binarization.

Evaluation Metrics. Parent identification is mainly evaluated using AUPRC, while AUROC and F1 are also computed. Suppose a task contains K_{pa} true parents. For all Top-k metrics, we set $k = K_{\mathrm{pa}}$. NDCG@k uses the ground-truth direct causal strength τ_j as the relevance gain

$$\mathrm{NDCG}\,@k = \frac{\sum_{r=1}^{k} \tau_{\pi_r} / \log_2(r+1)}{\sum_{r=1}^{k} \tau_{\pi_r^*} / \log_2(r+1)}. \tag{47}$$

Here, π denotes the predicted ranking and π^* denotes the ideal ranking induced by τ. Parent Spearman and Parent Kendall compare the predicted scores with τ only among true parent variables. They therefore measure the ranking of effect strengths among different direct parents rather than the separation between parents and zero-effect nonparents.

Result Aggregation. All metrics are first computed at the individual task level and then averaged within each random seed. When multiple target mechanisms are involved, results are first equally averaged across mechanisms and then averaged over three independent random seeds. The $\pm$ values in the main text and the error bars in the figures represent the sample standard deviation across the three random seeds.

B.2 Standard Causal Test Sets and Generalization Settings

Full-FCM Test Tasks. The candidate variables in Full-FCM are generated by dynamic sparse structural equations. Their cross-sectional correlation, temporal persistence, and structural sparsity are calibrated to the statistical properties of financial feature panels. However, the specific graph structures, parameters, noise variables, and parent identities are resampled for the test tasks. The target variable is generated only from the direct parents selected for the current task. Other variables may be statistically dependent on the parents, but they do not enter the target structural equation.

The target generating mechanisms include additive, interaction, and heteroscedastic mechanisms. The additive mechanism randomly uses linear, saturating, quadratic, periodic, and threshold-type univariate functions. The interaction mechanism further introduces multiplicative interactions between parents. The heteroscedastic mechanism allows parent variables to modulate the noise scale of the target. All three mechanisms use the same definitions of direct parents and direct intervention strength, and the target signal-to-noise ratio is fixed at 5. The generator rejects and resamples any task in which the quantile-based direct-strength contrast degenerates to zero for a sampled parent.

Financial Semi-Synthetic Test Tasks. Financial Semi-Synthetic data directly preserve the joint distribution, tail behavior, cross-sectional correlation, and temporal dependence of real financial features, while the structural equation of the target variable is artificially controlled. Each task samples stock trajectories from the unseen period, selects direct parents from the candidate variables, and generates a synthetic target. Therefore, both z and τ remain available as exact ground truth.

The Semi part of Table 1 uses a pure nonlinear target mechanism. Each direct parent enters the target equation through a cubic nonlinear basis function after removing its first-order linear projection, which reduces the usefulness of simple linear correlation for parent identification. This experiment therefore evaluates nonlinear direct-cause identification under real financial covariate distributions with exact structural supervision.

Main Benchmark Settings. For both Full-FCM and Financial Semi-Synthetic in Table 1, we fix $N = 64$, $T = 60$, $D = 94$, and $K_{\mathrm{pa}} = 20$. Full-FCM equally averages over additive, interaction, and heteroscedastic mechanisms. Each random seed contains 16 independent tasks for each mechanism. The Semi setting uses the pure nonlinear mechanism, with four independent tasks for each random seed.

Generalization Settings. Figure 3 uses only Full-FCM. When varying the cross-sectional size, we set $N \in \{32, 64, 128, 256\}$ and fix $K_{\mathrm{pa}} = 20$. When varying local causal density, we set $K_{\mathrm{pa}} \in \{10, 15, 20\}$ and fix $N = 64$. The target mechanism experiment reports additive, interaction, and heteroscedastic mechanisms separately. All other parameters are kept fixed within each comparison group so that performance differences arise only from the factor being examined.

B.3 Proxy Construction and Evaluation for the Correlation Trap

Proxy Construction. In the correlation trap experiment, proxy variables occupy nonparent positions in the original candidate set. Therefore, they do not change the candidate dimension and do not enter the target structural equation. The identities of parents, proxy variables, and other nonparent variables are randomly reassigned for each task.

For a single-parent proxy, each proxy X_p is paired with one true parent $X_{c(p)}$ and is constructed as

$$X_p = \rho X_{c(p)} + \sqrt{1-\rho^2}\eta_p, \tag{48}$$

where the independent residual η_p is standardized and orthogonalized before mixing. For a multi-parent aggregate proxy, each proxy corresponds to a set G_p containing three true parents. A weighted joint signal is first formed from the parents in the group, after which an independent residual is added in the same manner. Regardless of how strongly a proxy is statistically related to the target or the true parents, it satisfies $z_p = 0$ and $\tau_p = 0$ as long as it does not enter the target structural equation.

Correlation Strength. The parameter ρ represents the signal mixing strength before rank transformation. For a single-parent proxy, it directly controls the relationship between the proxy and its corresponding parent. For an aggregate proxy, it controls the relationship between the proxy and the joint parent signal. Therefore, the proxy is not required to have the same pairwise correlation with every source parent.

Correlation-Trap Metrics. For single-parent proxies, Parent-over-Proxy Accuracy is defined as

$$\mathrm{PoP}_{\mathrm{single}} = \frac{1}{P}\sum_{p=1}^{P} \mathbf{1}[s_{c(p)} > s_p]. \tag{49}$$

For multi-parent aggregate proxies, we use all source parent and proxy pairs

$$\mathrm{PoP}_{\mathrm{aggregate}} = \frac{1}{\sum_p |G_p|}\sum_{p=1}^{P}\sum_{c\in G_p} \mathbf{1}[s_c > s_p]. \tag{50}$$

Rank Gap uses the same parent-proxy pairs and is defined as the proxy rank minus the corresponding parent rank. A positive value therefore indicates that the true parent is ranked ahead of the proxy. Proxy FPR@k measures the proportion of proxy variables that enter the Top-k ranking.

Experimental Settings. The experiment fixes $N = 128$, $T = 60$, $D = 94$, and $K_{\mathrm{pa}} = 20$, and varies $P \in \{3, 6, 9\}$ and $\rho \in \{0.3, 0.5, 0.7\}$. Single-parent proxies appear in the main text, with multi-parent aggregate proxies reported in Appendix C.2.

B.4 Recursive Evaluation Protocol for Real Financial Data

Rolling Sample Split. The real financial experiment uses an annual recursive out-of-sample design. The initial training period is 1957–1974, the initial validation period is 1975–1986, and the out-of-sample test period is 1987–2016. As the test year advances, the training period expands annually, while the validation period always contains the 12 years immediately preceding the test year. The training sample is used to estimate the Elastic Net parameters, and the validation sample is used to select the regularization parameters. Each annual estimation produces out-of-sample predictions for the subsequent 12 months.

Annual CausalRank Aggregation. For each test year y, CausalRank uses only the corresponding expanding training period to generate the annual ranking. We divide the history into 60-month panel blocks and select the 256 largest stocks by market capitalization at each block end. Block-level log ranking scores are aggregated by the number of assets

$$\widetilde{s}_y = \frac{\sum_{b \in \mathcal{B}_y} n_b \widetilde{s}_{y,b}}{\sum_{b \in \mathcal{B}_y} n_b}. \tag{51}$$

Here, $\mathcal{B}_y$ denotes the panel blocks for test year y, and n_b is the number of assets in block b. The Top-20 features selected by $\widetilde{s}_y$ remain fixed throughout that test year.

Downstream Prediction and Preprocessing. All methods use the same annually updated Elastic Net as the downstream predictor, together with the same training, validation, and testing splits and stock universe. Stock-level characteristics are transformed into cross-sectional ranks each month and mapped to $[-1, 1]$, with missing values filled with zero. The compared methods include All Factors, Random Top-20, Pearson, Lasso, LightGBM, TCD-DL, and CausalRank.

Portfolio Evaluation. Predictive performance is evaluated using out-of-sample R^2 relative to a zero-forecast benchmark. For economic evaluation, we form decile portfolios based on predicted returns and construct a value-weighted 10–1 long-short portfolio. We report the annualized Sharpe ratio, annualized return, and drift-adjusted average monthly turnover of this long-short portfolio over the full out-of-sample period.

Table 4: Full-precision standard causal identification, ranking, and inference results.

Method	Full AUPRC	Full NDCG@20	Full Parent ρ	Semi AUPRC	Semi NDCG@20	Semi Parent ρ
Pearson	0.4512 ± 0.0391	0.4558 ± 0.0447	0.2601 ± 0.0394	0.2249 ± 0.0201	0.1595 ± 0.0424	-0.0533 ± 0.1422
Nonlinear Correlation	0.4865 ± 0.0448	0.5048 ± 0.0499	0.3616 ± 0.0591	0.5929 ± 0.0735	0.5977 ± 0.0495	0.4420 ± 0.0152
Mutual Information	0.4384 ± 0.0207	0.4539 ± 0.0268	0.3306 ± 0.0493	0.3936 ± 0.0324	0.3992 ± 0.0313	0.2865 ± 0.1326
Lasso	0.9698 ± 0.0033	0.9374 ± 0.0121	0.6593 ± 0.0411	0.2480 ± 0.0309	0.1996 ± 0.0497	0.0545 ± 0.0858
LightGBM	0.9322 ± 0.0069	0.8923 ± 0.0020	0.6071 ± 0.0242	0.7331 ± 0.0430	0.7329 ± 0.0429	0.4143 ± 0.0763
DirectLiNGAM	0.7596 ± 0.0501	0.7483 ± 0.0234	0.4125 ± 0.0302	0.2617 ± 0.0109	0.2192 ± 0.0349	-0.0114 ± 0.0756
NOTEARS	0.6863 ± 0.0111	0.6986 ± 0.0105	0.3413 ± 0.0676	0.2423 ± 0.0154	0.1786 ± 0.0352	0.1055 ± 0.0606
TCD-DL	0.2625 ± 0.0048	0.2157 ± 0.0156	0.0293 ± 0.0460	0.2295 ± 0.0255	0.1849 ± 0.0444	-0.0102 ± 0.1615
CausalRank-Existence	0.9310 ± 0.0053	0.9021 ± 0.0027	0.6334 ± 0.0415	0.6714 ± 0.0209	0.6023 ± 0.0182	0.4002 ± 0.0826
CausalRank	$\mathbf{0.9835 \pm 0.0009}$	$\mathbf{0.9647 \pm 0.0004}$	$\mathbf{0.8093 \pm 0.0231}$	$\mathbf{0.7666 \pm 0.0069}$	$\mathbf{0.7754 \pm 0.0162}$	$\mathbf{0.4847 \pm 0.0481}$

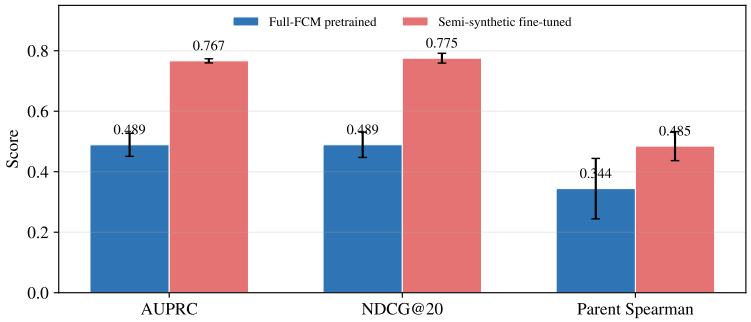


Figure 6: Causal identification and ranking before and after Financial Semi-Synthetic fine-tuning. Bars and error bars report the mean and sample standard deviation across three random seeds.

C Supplementary Experimental Results

This section follows the three experiments in the main text and provides additional results for standard causal identification and ranking, correlation traps, and real financial data. Synthetic results use the same aggregation protocol as the main text and report means and sample standard deviations across three random seeds.

C.1 Standard Causal Identification and Causal Ranking

Full-Precision Main Results. Table 4 reports the unrounded values underlying Table 1. All comparisons and numerical analyses in the main text use these full-precision results.

Effect of Financial Semi-Synthetic Fine-Tuning. We compare the Full-FCM pretrained model with the Financial Semi-Synthetic fine-tuned model on the same pure-nonlinear semi-synthetic test set, fixing $N = 64$, $T = 60$, $D = 94$, and $K_{\mathrm{pa}} = 20$. As shown in Figure 6, fine-tuning raises AUPRC from 0.4890 to 0.7666, NDCG@20 from 0.4891 to 0.7754, and Parent Spearman from 0.3441 to 0.4847, corresponding to relative improvements of 56.8%, 58.5%, and 40.9%.

The improvement is accompanied by lower cross-seed variation: the standard deviations of AUPRC, NDCG@20, and Parent Spearman decrease by 81.9%, 61.2%, and 51.9%, respectively.

Complementary Identification and Ranking Metrics. Table 5 supplements the main metrics with AUROC, F1@20, and Parent Kendall. CausalRank achieves the highest mean in all six comparisons. On Full-FCM, Parent Kendall increases from 0.5059 for the strongest baseline to 0.6638. On Financial Semi-Synthetic data, CausalRank reaches an AUROC of 0.9029, an F1@20 of 0.7125, and a Parent Kendall of 0.3421, exceeding the corresponding strongest-baseline values of 0.8909, 0.6958, and 0.3079.

Effect Supervision and Candidate Competition. The comparison with CausalRank-Existence isolates the role of continuous effect supervision. On Full-FCM, adding effect and ranking supervision

Table 5: Complementary metrics on the standard causal test sets.

Method	Full AUROC	Full F1@20	Full Parent Kendall	Semi AUROC	Semi F1@20	Semi Parent Kendall
Pearson	0.6867 ± 0.0204	0.3976 ± 0.0396	0.1876 ± 0.0261	0.4699 ± 0.0176	0.1750 ± 0.0451	-0.0395 ± 0.0942
Nonlinear Correlation	0.7179 ± 0.0255	0.4253 ± 0.0412	0.2626 ± 0.0414	0.8558 ± 0.0428	0.5708 ± 0.0315	0.3070 ± 0.0169
Mutual Information	0.6883 ± 0.0191	0.3826 ± 0.0139	0.2374 ± 0.0340	0.6663 ± 0.0336	0.3375 ± 0.0331	0.2158 ± 0.1066
Lasso	0.9763 ± 0.0047	0.9434 ± 0.0049	0.5059 ± 0.0358	0.4995 ± 0.0261	0.2042 ± 0.0382	0.0386 ± 0.0691
LightGBM	0.9770 ± 0.0035	0.8795 ± 0.0106	0.4581 ± 0.0256	0.8909 ± 0.0204	0.6958 ± 0.0505	0.3079 ± 0.0479
DirectLiNGAM	0.8155 ± 0.0302	0.6896 ± 0.0476	0.3080 ± 0.0231	0.5039 ± 0.0190	0.2333 ± 0.0315	-0.0114 ± 0.0323
NOTEARS	0.7780 ± 0.0047	0.6080 ± 0.0069	0.2367 ± 0.0427	0.4908 ± 0.0048	0.2042 ± 0.0505	0.0145 ± 0.0194
TCD-DL	0.5116 ± 0.0047	0.2347 ± 0.0085	0.0209 ± 0.0336	0.4771 ± 0.0496	0.2042 ± 0.0402	-0.0123 ± 0.1173
CausalRank	$\mathbf{0.9942 \pm 0.0013}$	$\mathbf{0.9469 \pm 0.0093}$	$\mathbf{0.6638 \pm 0.0219}$	$\mathbf{0.9029 \pm 0.0069}$	$\mathbf{0.7125 \pm 0.0217}$	$\mathbf{0.3421 \pm 0.0502}$

increases NDCG@20 from 0.9021 to 0.9647 and Parent Spearman from 0.6334 to 0.8093, relative improvements of 6.9% and 27.8%. On Financial Semi-Synthetic data, the same components raise NDCG@20 from 0.6023 to 0.7754 and Parent Spearman from 0.4002 to 0.4847.

Removing the candidate competition module reduces AUPRC, Pairwise Accuracy, and Top-20 Precision to 0.7948, 0.8463, and 0.6956, respectively.

Amortized Inference Efficiency. CausalRank requires 0.042 seconds per Full-FCM task and 0.034 seconds per Financial Semi-Synthetic task. Its Full-FCM inference is approximately $2.3\times$ faster than Lasso and TCD-DL and $22\times$ faster than LightGBM. On Financial Semi-Synthetic data, DirectLiNGAM and NOTEARS require 141.703 and 361.774 seconds per task.

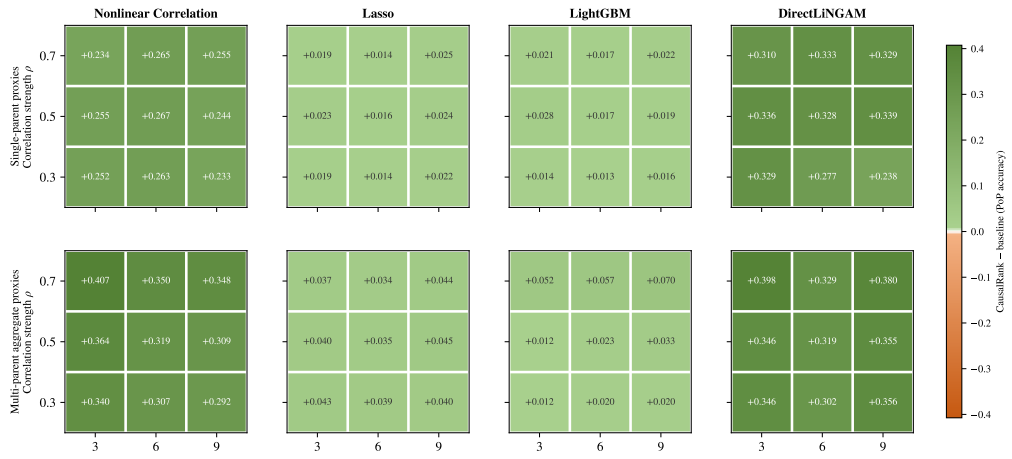


Figure 7: CausalRank's PoP advantage across proxy structures, counts, and strengths. Rows show single-parent and multi-parent aggregate proxies, columns correspond to the baselines.

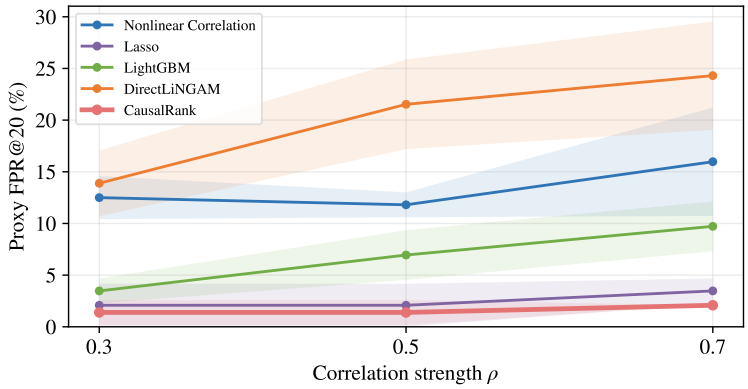


Figure 8: Proxy FPR@20 under additive single-parent correlation traps with $N = 128$, $K_{\text{pa}} = 20$, and $P = 3$. Lower is better, bands show sample standard deviations across three seeds.

C.2 Extended Correlation-Trap Results

Robustness across Proxy Structures. Figure 7 extends the single-parent heatmaps in Figure 4 to multi-parent aggregate proxies. Across two proxy structures, three proxy counts, three correlation strengths, and four baselines, all 72 PoP differences are positive. For single-parent proxies, the advantage over Lasso and LightGBM ranges from 0.013 to 0.028. Under the multi-parent construction, the corresponding ranges are 0.034–0.045 and 0.012–0.070.

Fixed-Budget Proxy Rejection. Figure 8 reports Proxy FPR@20 for additive single-parent traps with $P = 3$. At $\rho = 0.3$, 0.5, and 0.7, CausalRank obtains 1.39%, 1.39%, and 2.08%, respectively, below every baseline at each strength. The corresponding Lasso values are 2.08%, 2.08%, and 3.47%, while LightGBM rises from 3.47% to 9.72%.

Table 6: Long-horizon performance of value-weighted 10–1 portfolios. Pos.-36M is the percentage of 36-month windows with positive cumulative return, Neg.year is the average return over each method's negative-return years.

Method	Cum. return	Pos.-36M	Max DD	Neg.-year
All Factors	362.28%	84.3%	37.93%	−12.40%
Random Top-20	312.71%	88.9%	37.40%	−8.48%
Pearson Top-20	264.11%	83.4%	32.69%	−9.74%
Lasso Top-20	325.01%	75.7%	46.30%	−10.18%
LightGBM Top-20	257.38%	73.8%	49.97%	−9.48%
TCD-DL Top-20	115.08%	64.0%	46.25%	−12.88%
CausalRank Top-20	**372.91%**	**92.0%**	**31.99%**	**−4.43%**

Table 7: Most frequently selected CausalRank characteristics across 30 recursive test years.

Characteristic	Years selected	Frequency
`mvel1`	30/30	100.0%
`rd_sale`	30/30	100.0%
`divi`	30/30	100.0%
`securedind`	29/30	96.7%
`divo`	27/30	90.0%
`rd_mve`	27/30	90.0%
`pctacc`	26/30	86.7%
`turn`	26/30	86.7%

C.3 Additional Evidence on Real Financial Data

Downside and Intertemporal Portfolio Performance. Table 6 extends the financial comparison to long-horizon persistence and downside behavior.

Its cumulative return exceeds All Factors by 10.63%, and its 92.0% positive-window frequency is 3.1% above the next-best Random Top-20 result. Its 31.99% maximum drawdown is 0.70% below the next-best Pearson Top-20 portfolio, while its average negative-year loss is 4.05% smaller than the next-best Random Top-20 result. CausalRank also outperforms All Factors in all six years in which the All-Factors portfolio loses money and earns an average return of 0.60% in those years.

Recurring Characteristics in Recursive Selection. Table 7 reports the characteristics most frequently selected across the 30 annual recursive rankings from 1987 to 2016.

The recurring characteristics cover complementary economic dimensions: market equity (`mvel1`), R&D intensity (`rd_sale` and `rd_mve`), dividend policy (`divi` and `divo`), financing structure (`securedind`), accrual quality (`pctacc`), and trading activity (`turn`).

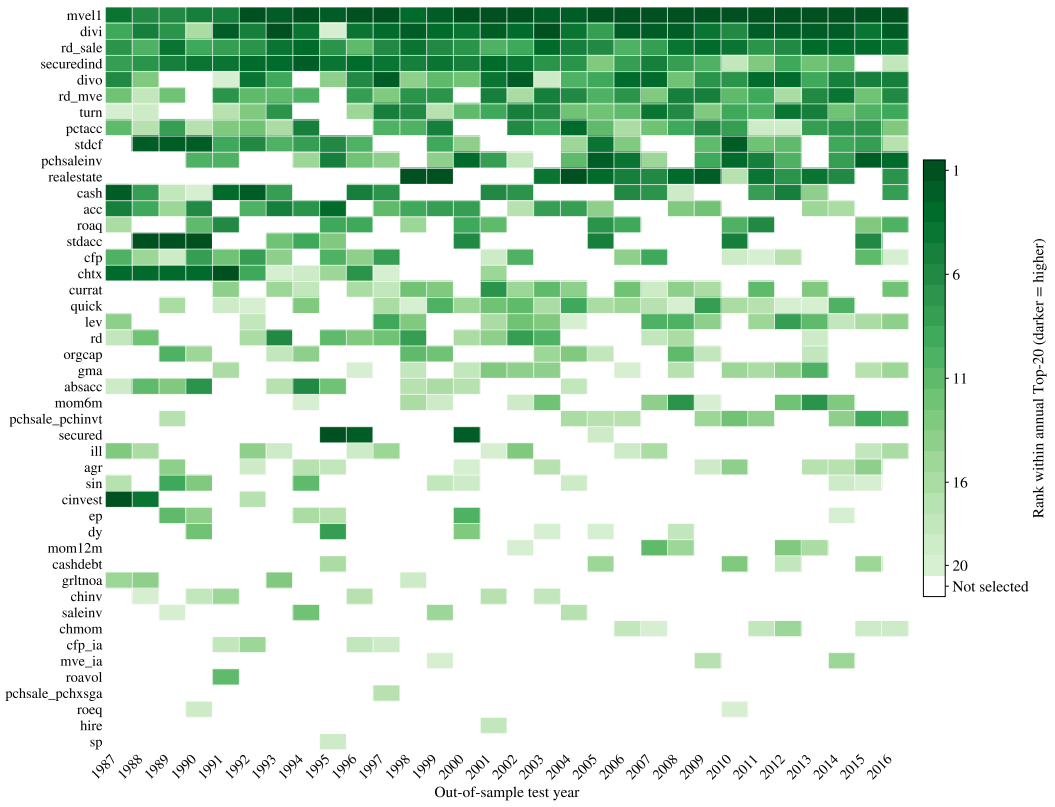


Figure 9: Yearly selected Top-20 Characteristics by CausalRank. Rows show the 46 characteristics selected at least once from 1987 to 2016, ordered by cumulative Top-20 rank strength. Columns correspond to out-of-sample test years. Darker blue cells indicate higher within-year ranks, while white cells indicate that the characteristic was not selected in that year.

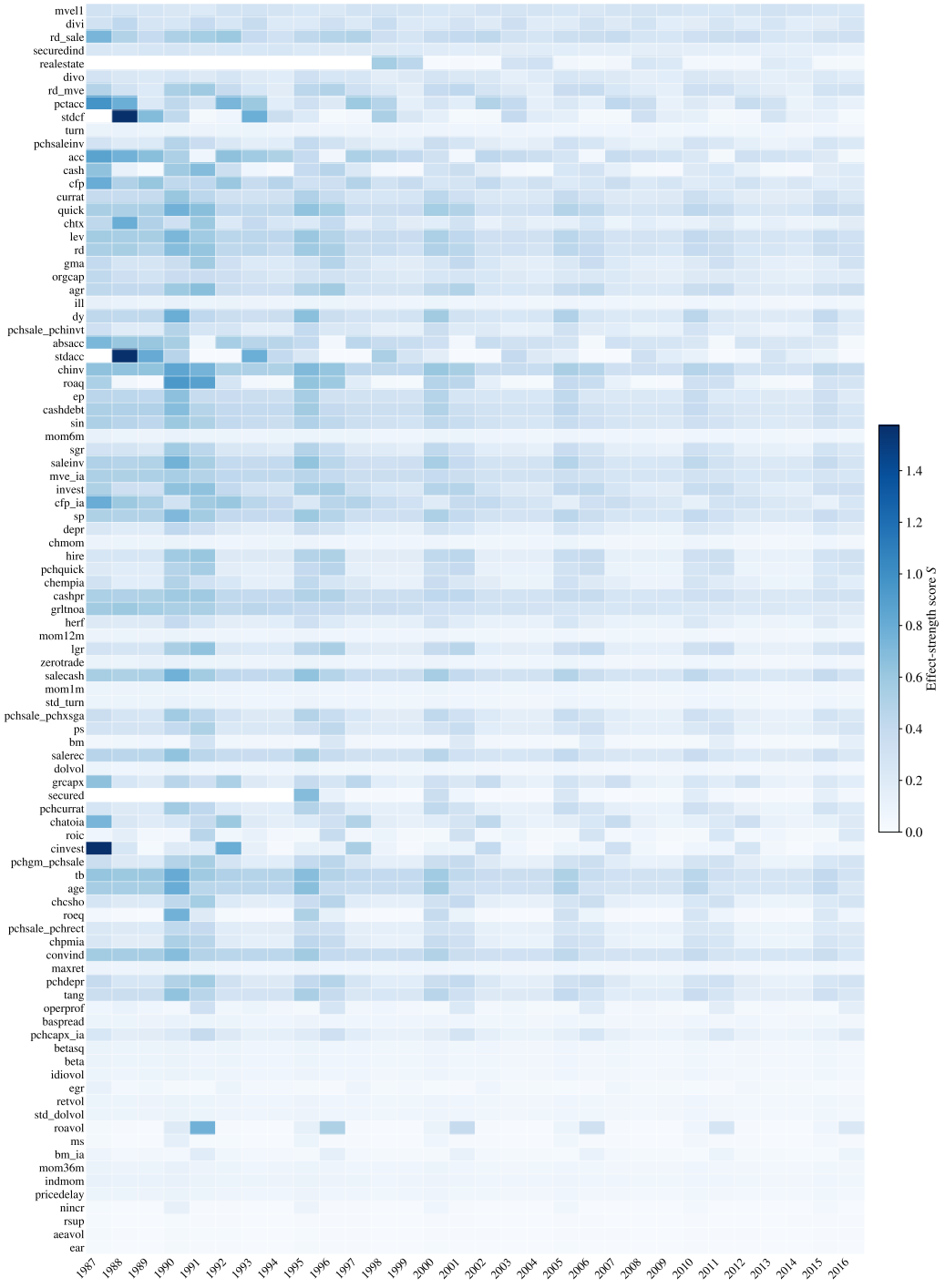


Figure 10: Yearly strength of the causal effect estimated by CausalRank. Rows show all 94 characteristics, ordered by their mean log causal ranking scores, and columns correspond to out-of-sample test years from 1987 to 2016. Darker blue cells indicate larger effect strength a_j

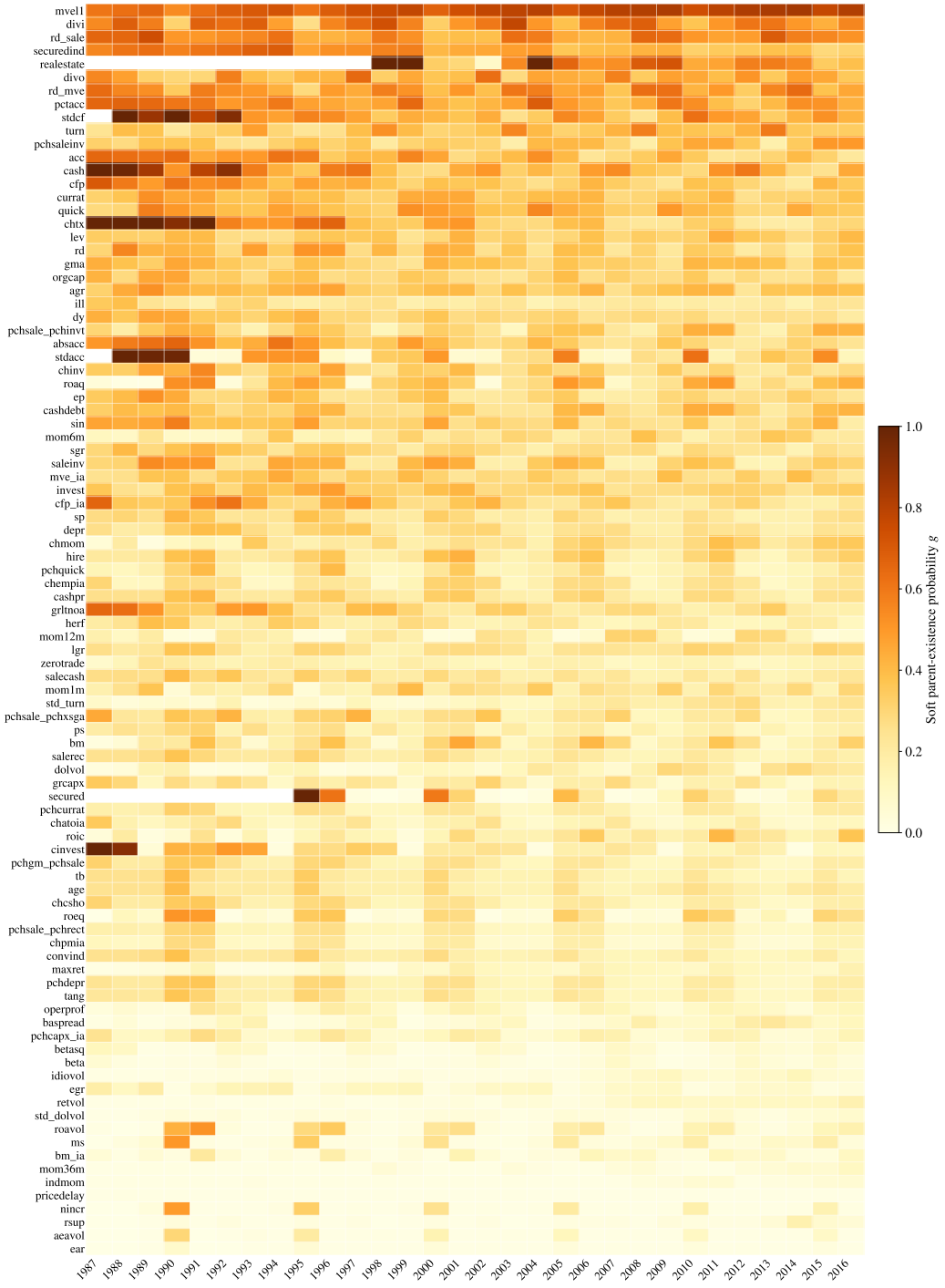


Figure 11: Yearly direct parent existence probabilities estimated by CausalRank. Rows show all 94 characteristics in the same order as Figure 10, and columns correspond to out-of-sample test years from 1987 to 2016. The color gradient ranges from low probabilities in light yellow to high probabilities in dark brown.

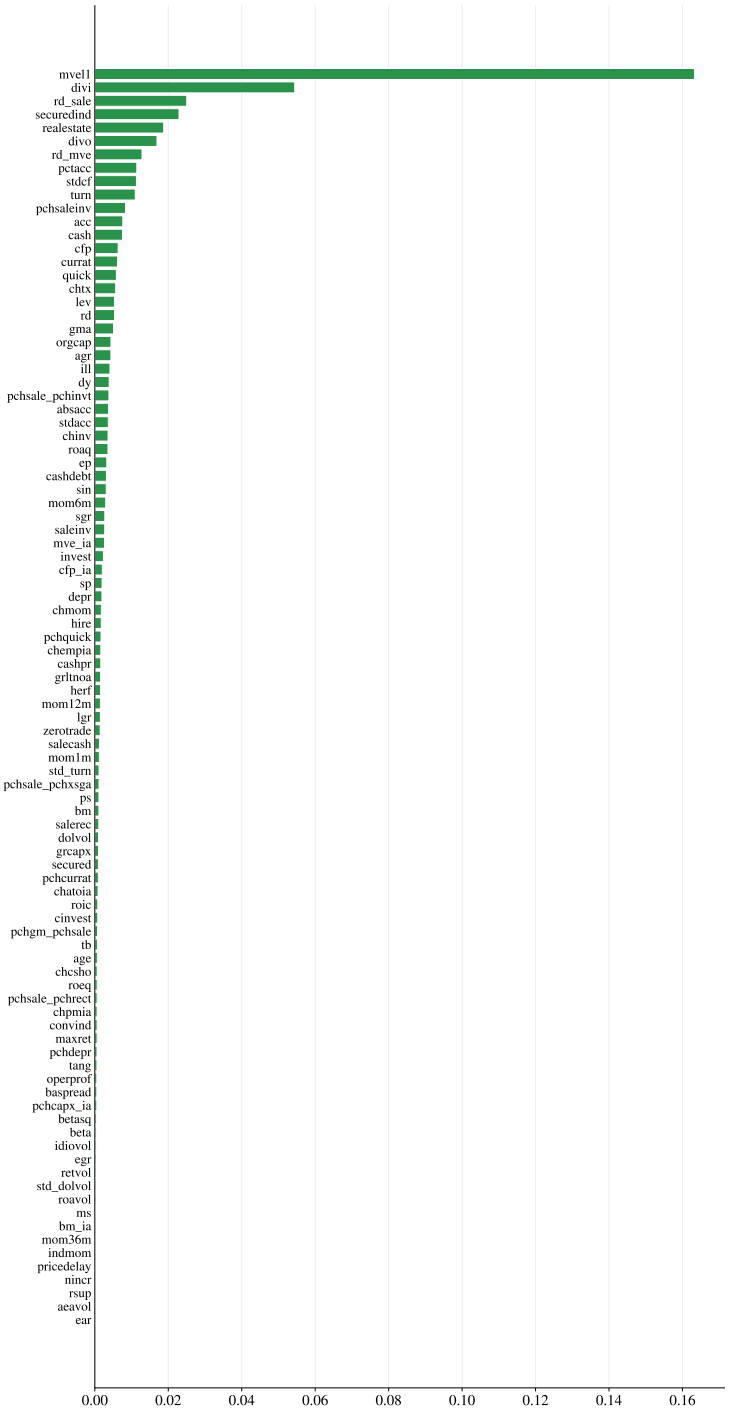


Figure 12: Causal ranking scores produced by CausalRank. Rows show all 94 characteristics ordered by the mean of their yearly scores $s_j = g_j a_j$ over 1987–2016.