# Causal Network Representations in Factor Investing*

Clint Howard[1], Harald Lohre[1,2] and Sebastiaan Mudde[3]

[1] *Robeco Quantitative Investments*
[2] *Lancaster University Management School*
[3] *Erasmus University Rotterdam*

December 14, 2024

## Abstract

This paper explores the application of causal discovery algorithms to factor investing, addressing recent criticisms of correlation-based models. We create novel causal network representations of the S&P 500 universe and apply them to three investment scenarios. Our findings suggest that causal approaches can complement traditional methods in areas such as stock peer group identification, factor construction, and market timing. While causal networks offer new insights and sometimes outperform correlation-based methods in terms of risk-adjusted returns, they do not consistently surpass traditional approaches. The causal method though shows promise in identifying unique market relationships and potential hedging opportunities. However, its practical implementation presents challenges due to computational complexity and interpretation difficulties. Our study demonstrates the potential value of causal discovery in factor investing, while also identifying areas for further research and refinement.

*JEL classification: C32, C38, G11, G12*

*Keywords: causal discovery, factor investing, asset pricing, financial networks, market timing*

*Corresponding author: Harald Lohre, *E-mail: h.lohre@robeco.nl*, *Address:* Weena 850, 3014DA, Rotterdam, South Holland, The Netherlands.
We are grateful to David Blitz, Amit Goyal, Matthias Hanauer, Iman Honarvar, Tobias Hoogteijling, Pim van Vliet, Dean Walsh, and the Robeco Research Seminar for helpful discussions and comments. Note that this paper expresses the authors' views which do not necessarily coincide with those of Robeco.

## 1. Introduction

Financial markets are complex systems where understanding relationships between assets is crucial for constructing effective investment strategies. Mantegna (1999) envisioned financial markets as highly complex networks and demonstrated that representations, other than the covariance matrix, can be of economic importance. Traditionally, these relationships have been modeled using correlation-based approaches to measure the similarity of pairs of assets.

In the context of financial markets, and particularly factor investing, the limitations of correlation-based approaches have become increasingly apparent. Factor investing, which aims to capture systematic sources of return by targeting specific stock characteristics or "factors", has traditionally relied heavily on correlation-based methods to identify and exploit these factors. However, the correlation coefficient does not condition on anything and is susceptible to spurious association (Simon, 1954). The latter caveat can lead to significant issues in factor investing, including overfitting and false discoveries around the underlying drivers of financial markets (López de Prado, 2023a).

The use of correlation-based methods in factor investing faces several key limitations. These methods often fail to distinguish between genuine causal relationships and mere statistical associations (López de Prado, 2023b), struggle to determine the direction of influence between variables, and cannot account for complex, non-linear relationships in financial markets. Moreover, their static nature fails to capture the dynamic, evolution of financial markets, potentially leading to outdated or ineffective investment strategies as market conditions change. These shortcomings highlight a significant gap in the current approach to factor investing: the need for more sophisticated methodologies that can better capture the true underlying relationships driving financial markets, account for their dynamic nature, and provide more robust insights for investment decision-making.

Causal discovery algorithms offer a potential solution to these challenges that arise when using correlation-based methods. Causal methods aim to uncover causal relationships from observational data, and to thus provide a more nuanced understanding of financial markets. By conditioning on all available data, allowing for lead-lag relations, and determining the directionality of comovement, causal approaches may offer richer insights than traditional correlation-based methods. This is particularly pertinent for factor investing applications, potentially leading to a richer understanding of factor dynamics. We employ DYNOTEARS, a representative causal discovery algorithm developed by Pamfil et al. (2020), which can identify both immediate and time-lagged relationships between stock returns, offering insights into their comovement patterns.

We apply the DYNOTEARS algorithm to create causal network representations of the S&P 500 investment universe and leverage these representations in three investment-based experiments. Using constituents of the S&P 500 investment universe over the period of December 1989 to December 2022, we estimate monthly a causal-based network representation based on the returns of each stock in the estimation universe. We leverage these representations in three empirical experiments related to factor investing.

Our study thus makes several contributions to the literature on causal discovery techniques in factor investing.[1] First, we introduce a novel method for peer group identification using causal networks, enhancing factor neutralization in long–short equity strategies.[2] Second, we develop a "low centrality" factor based on a stock's position within the causal network, providing new insights into how network structure relates to returns. Third, we propose a market timing indicator derived from causal network density, demonstrating the potential of these techniques for return prediction.

This work extends the growing literature on causal discovery in finance by demonstrating its practical utility in factor investing. While early research focused on market index returns due to computational constraints (e.g., Dornau, 1999; Stavroglou et al., 2017), recent studies have explored causal relationships between individual assets (e.g., Gao and en Ren, 2013; Diebold and Yılmaz, 2014; Fiedor, 2014; Upadhyay et al., 2020). Durcheva and Tsankov (2021) and Shirokikh et al. (2022) explore the application of causal techniques to uncover networks between U.S. firms, observing significant changes in causal networks around major global events. We build upon this stream of literature by constructing causal network

[1] There is an extensive literature on the broader application of causal discovery techniqes in finance, e.g., Billio et al. (2012); Gnecco et al. (2020); Cornell et al. (2024); Dong et al. (2024); Sadeghi et al. (2024); Sokolov et al. (2023).

[2] A peer group neutral long–short strategy is an investment approach that aims to neutralize unwanted exposure to specific peer groups (e.g., industries or sectors) while maintaining exposure to desired factors. In this strategy, long and short positions are taken within each peer group, typically going long on stocks expected to outperform and short on those expected to underperform within their respective groups.

representations of S&P 500 constituents and showcasing their investment applications.

Our approach is most closely related to Pamfil et al. (2020), who proposed the DYNOTEARS algorithm. While they focused on demonstrating the algorithm's superiority, we extend their work by applying DYNOTEARS-derived causal networks to real investment scenarios. In doing so, we bridge the gap between theoretical causal discovery methods and their practical implementation in factor investing.

In the context of factor investing, we complement emerging literature that applies causal techniques to identify influential items for return forecasting (Yang et al., 2014) and latent return-driving factors (Liu et al., 2023). D'Acunto et al. (2021) apply causal discovery to existing risk factors (such as the market, value, momentum, and size) and show certain factors tend to dominate the overall causal process. Our use of causal discovery-based network representations for clustering peer groups contributes to research on financial network clustering algorithms (López de Prado, 2016; Konstantinov et al., 2020; Tristan and Ong, 2021; French, 2023). This novel approach to peer group identification may better capture true stock relationships, potentially leading to more effective factor neutralization strategies.

The remainder of this paper is structured as follows: Section 2 provides the related work for our experiments and a conceptual framework for causal discovery algorithms in financial markets. Section 3 presents the detailed methodology for estimating our causal networks, and Section 4 details the investment experiments we conduct. Section 5 presents the results of our three experiments. Section 6 concludes with a summary of our findings, limitations, and suggestions for future research.

## 2. Related work and conceptual framework

This section presents the related work and conceptual framework behind factor investing, peer group selection, Bayesian networks, and causal networks. We introduce the concept of factor investing and detail related work behind the three empirical experiments we conduct. We present a comprehensive review of peer group selection strategies. Then we provide the conceptual frameworks behind Bayesian networks, before introducing the conceptual framework of causal networks.

## 2.1. Factor investing and its corresponding strategies

### *2.1.1. What is factor investing?*

Factor investing is a systematic approach to portfolio construction that aims to capture specific drivers of asset returns. These drivers, known as factors, are characteristics or attributes of securities that have been shown to explain cross-sectional differences in returns over time.[3] By targeting specific factors, investors aim to achieve better risk-adjusted returns, increased diversification, and more precise control over portfolio characteristics. Factor investing offers a rules-based approach that can be implemented at scale while still aiming to outperform market-cap weighted indices (Blitz and Vidojevic, 2019).

Key factors that have garnered significant attention in the literature include momentum (Jegadeesh and Titman, 1993), value (Fama and French, 1993), quality/profitability (Asness et al., 2018; Novy-Marx, 2013), short-term reversal (Jegadeesh, 1990; Lehmann, 1990; Blitz et al., 2023), and low-volatility (Black, 1972; Blitz and van Vliet, 2007) or low-beta (Frazzini and Pedersen, 2014). These factors form the basis of widely adopted models such as the Fama and French (1993) three-factor model and its subsequent extensions (Carhart, 1997; Fama and French, 2015). As the number of identified factors has proliferated, a secondary literature has emerged aiming to determine their validity. Cochrane (2011) introduced the term "factor zoo" to describe this growth, with subsequent research suggesting that the hundreds of proposed factors can likely be reduced to a more manageable number (Bartram et al., 2021; Swade et al., 2023; Jensen et al., 2023).

### *2.1.2. Highlighted strategies and related work*

In this section, we introduce three empirical experiments in factor investing that we will apply the causal discovery algorithms to. These experiments span key areas of factor investing where causal insights could provide enhancements to existing methodologies. First, we examine peer group neutralization in long–short equity strategies, a fundamental technique in factor construction that aims to isolate specific return drivers. Second, we investigate

[3] There are also macroeconomic drivers of returns, such as interest rates, and oil prices, which can be used to explain asset returns.

the use of network centrality measures derived from causal graphs to construct a novel factor, potentially offering a new perspective on risk and return dynamics. Lastly, we explore how causal network density can serve as an indicator for market timing. Each of these applications builds upon established factor investing concepts while introducing novel causal results, aiming to demonstrate how causal discovery algorithms can complement and enhance traditional approaches in factor investing.

#### *2.1.2.1 Experiment 1: Peer group neutralization*

Our first experiment of financial networks is to explore alternative peer groups for the neutralization of cross-sectional long–short portfolio strategies. In the analytical framework of Ehsani et al. (2023), the predictive power of firm characteristics can be split into two components: an across-industry component and a within-industry component. They show that long–short portfolios generally benefit from sector neutralization (i.e., removing the across-industry component). The reason that the across-industry component often does not carry a premium compared to the within-industry component is well-illustrated by Vyas and van Baren (2021). They document that some equity factors suffer from unpriced industry exposure due to industry tilts. That is, the industry composition of the tail distribution, for a given firm characteristic (e.g., Book-to-Price ratio), can be highly homogeneous. An example of this homogeneity can be found in Blitz and Hanauer (2020) who show that the value factor displays major industry bias in favor of the Utilities sector, and thus incurs unnecessary risk. In addition, they find that the value factor is systematically short in sectors that have a large amount of intangible assets, e.g., the Information Technology sector. In fact, Asness et al. (2000), Bender et al. (2019), and Cohen and Polk (1996) all show that most, if not all, value characteristics suffer from unrewarded industry exposure. In this experiment, we investigate the potential of alternative peer groups in the context of peer group neutralization, with a specific focus on causality-based peer groups.

#### *2.1.2.2 Experiment 2: Centrality as a cross-sectional long–short equity factor*

Our second experiment explores the role of financial network centrality in the cross-section of equity returns. First, we define what we exactly mean by (node) centrality. Freeman et al.

(2002) posit that the center of a star network, as seen in Figure 1, is the purest example of a central node. They understand central nodes as the nodes that have more ties, can reach other nodes more quickly, and more often lie on the shortest path between other nodes. Classical centrality measures such as degree, closeness, and betweenness centrality respectively embody these properties. For instance, degree centrality simply counts the number of the in- and out-going edges as a measure of centrality.

<Insert Figure 1 about here>

The general sentiment in the literature is that stocks with high centrality are undesirable, with many authors arguing that incorporating centrality can yield more effective portfolio diversification in the framework of Markowitz (1952). For instance, Peralta and Zareei (2016) formally show that, both in minimum-variance and mean-variance portfolios, large eigenvector centrality scores correspond to low optimal weights. Notably, this result relies on the financial network being formulated as a correlation network. On the empirical side, Pozzi et al. (2013) find that peripheral companies offer lower risk and higher returns than central companies. They contextualize this finding by positing that central companies carry more risk during market crashes. Their network formulation is a filtered correlation network like in Mantegna (1999).[4]

Výrost et al. (2019) argue that highly central stocks offer less diversification than peripheral stocks. Additionally, they argue that central stocks are riskier because market shocks will not only affect the highly central asset itself, but also its neighborhood. Konstantinov et al. (2020) are among the first to comment on the relation between centrality and factors, however they do not use stock-level data but multi-asset multi-factor data. They do so by including the factor returns into their network analysis and find that the RMW, HML, and CMA factors of Fama and French (2015) are highly central nodes in the financial network.

#### *2.1.2.3 Experiment 3: Market timing*

Our final experiment investigates the market timing ability of the stock network density. We hypothesize that network density reflects the systematic risk of a financial network and

[4]The authors explore two types of filtered correlation networks are explored: Minimum Spanning Trees (MSTs) and Planar Maximally Filtered Graphs (PMFGs).

may carry return predictability. This builds on previous work exploring the relationship between network topology and market behavior.

Kaya (2015) found that network topology becomes denser prior to market crashes, aligning with Lenzu and Tedeschi's (2012) observation that denser topologies are problematic from a systematic risk perspective. To test this hypothesis, we develop a causal network-based proxy for financial network density, focusing on relative changes rather than absolute levels of centrality to account for structural breaks in network topology.

Our approach differs from previous studies in two key aspects. Unlike Eng-Uthaiwat (2018) or Lohre et al. (2014), who used correlation-based methods and graph diameter as a measure of network topology, we employ causal discovery algorithms and average eigenvector centrality. This choice addresses limitations of using graph diameter in financial networks, which typically exhibit small-world properties (Watts and Strogatz, 1998; Gao et al., 2013; Haldane, 2013; Sun and Chan-Lau, 2017). Eigenvector centrality is particularly suited for causal networks (Dablander and Hinne, 2019) as it accounts for the varying importance of network connections.

## 2.2. Peer selection methods

To motivate our exploration of causal networks, we explore the notion of a stock peer group. A peer group is supposed to consist of companies that are similar or homogeneous in some sense. Due to the myriad of ways to measure company similarity, there are numerous peer group identification schemes. These identification schemes can be broadly divided into two classes: discretionary classification schemes and systematic classification schemes.

### *2.2.1. Discretionary classification schemes*

Discretionary classification schemes are typically formed from a qualitative perspective on a company's business activities, linking it to other companies performing similar activities. These schemes often rely on expert judgment and manual classification.

Notable examples include the Standard Industrial Classification (SIC), North American Industry Classification System (NAICS), and the Global Industry Classification Standard (GICS). GICS, developed in 1999 by S&P and MSCI, has become particularly popular among financial practitioners. Its widespread adoption can be attributed to its desirable properties, as demonstrated by Bhojraj et al. (2003), who showed that GICS outperforms other classification systems in explaining cross-sectional variations in out-of-sample returns and firm-level characteristics.

However, discretionary schemes like GICS have limitations. They are often relatively static in nature (Costa and De Angelis, 2011), with classification changes typically taking more than a year to materialize. This can lead to investors acting on outdated information. Additionally, these schemes often focus primarily on end-products to determine company similarity (Phillips and Ormsby, 2016), potentially neglecting other important factors such as supply chain relationships. Despite these limitations, discretionary classification schemes remain widely used due to their intuitive nature and established track record in portfolio management.

### *2.2.2. Systematic classification schemes*

Systematic classification schemes employ quantitative methods to measure similarity between companies and group them accordingly. These schemes typically involve three components: a data input (such as stock returns, financial ratios, or company-related text), an algorithmic process to measure similarity, and an output of stock peer groups.

These schemes can be broadly categorized into three types:

1. **Fundamentally-based methods** use financial statement information to group companies. For example, Hoberg and Phillips (2016) use product descriptions from 10-K filings to create dynamic industry classifications, while Lee et al. (2016) incorporate various fundamental-based information sources (such as company filings and analyst coverage).

2. **Statistically-based methods** typically use return correlations or other statistical measures. Many of these methods build on the seminal work of Mantegna (1999), clustering stocks based on return correlation. However, Chan et al. (2007) found that these correlation-based methods are often outperformed by GICS in capturing out-of-sample return covariation. There are also efforts using non-correlation-based

statistical measures such as Vermorken et al. (2010) or Jung and Chang (2016).

3. **Graph-based methods** leverage graph theory to identify patterns and group stocks. For instance, Yang and Cogill (2013) apply graph similarity techniques to study structural differences in financial statements, while Zhang et al. (2022) use graph representation learning to measure similarity in industry supply-chain networks. Noels et al. (2023) apply a novel distance measure to studying the similarity of financial statements.

The main advantage of systematic classification schemes is their ability to quickly respond to market changes and incorporate a wide range of information beyond just the end-product of a company. However, they can be computationally intensive and may produce results that are less intuitive or interpretable compared to discretionary schemes.

In the following sections, we will explore how causal discovery algorithms can be used as a novel systematic classification scheme.

### 2.3. Bayesian networks

While systematic classification schemes offer advantages in terms of adaptability and data incorporation, they often rely on correlation-based or other statistical measures that may not fully capture the underlying relationships between stocks. To address these limitations and potentially provide a more nuanced understanding of stock interactions, we turn to the framework of Bayesian Networks (BNs). BNs offer a powerful approach to modeling complex relationships and dependencies, conceivably providing a more accurate representation of stock peer groups and market dynamics. The theoretical framework underlying causal algorithms, which we will use to construct our peer groups, is quite extensive and rooted in the concept of BNs. As such, we will first provide a concise explanation of the most crucial elements of BNs, discuss the limitations behind constraint-based BNs compared to score-based BNs, before delving into their application in causal discovery algorithms.

Pearl (1985) details that a given *Directed Acyclic Graph* (DAG), over a set of $p$ random variables $X$, constitutes a Bayesian Network if:

$$P(X) = \prod_{i=1}^{p} P(X_i | Pa(X_i)), \quad (1)$$

where $Pa(Z)$ denotes the parents of node $Z$. This equation expresses that each random variable $X_i$ only depends on its direct parents $Pa(X_i)$ and is independent of other variables. This condition is also referred to as the *local Markov condition*. An additional assumption that can be made is the *faithfulness assumption*. This assumption states that for a DAG $D$, a distribution $P$ is called faithful w.r.t. $D$ IFF all conditional independences are encoded by $D$ (Spirtes et al., 2000). An equivalent formulation of this assumption can be constructed via the concept of d-separation. In a DAG $G$, sets of nodes $X$ and $Y$ are d-separated by the set $Z$, if $Z$ is blocking all paths between $X$ and $Y$.[5]

In practice, a stronger variant of the faithfulness assumption is applied. The rigorous definition (Zhang and Spirtes, 2002) of strong faithfulness, in the Gaussian case, is as follows: for a DAG $D = (V, E)$, a Gaussian distribution $P$ is called $\lambda$-strongly faithful w.r.t. $D$ if:

$$\min\{|\operatorname{corr}(X_i, X_j | X_S)|,\ j \text{ not d-separated from } i \mid S,\ \forall i,\ j,\ S\} > \lambda,$$

where $\lambda \in (0, 1)$, $i$, $j \in V$ and $S \subset V \setminus \{i, j\}$. Essentially, this assumption dictates that (sets of) variables that are *d-connected* in the underlying DAG should have this association reflected in the data above a certain threshold. Uhler et al. (2013) argue that this requirement is a rather restrictive condition for most DAGs. When this assumption does not hold, some causal algorithms produce inconsistent estimates.[6] Thus, algorithms that rely on this assumption are less credible.

It is precisely this assumption that constitutes a major difference between the two main classes of causal algorithms: constraint-based and score-based algorithms. *Constraint-based algorithms* employ a conditional independence test to form a causal graph, whilst *score-based algorithms* iterate over all possible DAGs and select the best-scoring DAG. An example of a score function is the Bayesian Information Criterion (BIC). Most of the constraint-based algorithms, such as the PC algorithm, rely on the strong faithfulness assumption for their

[5] The equivalent formulation reads: a DAG $G$ satisfies the faithfulness assumption if for every $X$, $Y$, $Z$, if $X$ and $Y$ are conditionally independent given $Z$ then $Z$ d-separates $X$ and $Y$.

[6] In this context, consistency means that the estimated graph $\hat{G}_n$ converges to the true graph $G$ as $n \to \infty$.

consistency. In contrast, most score-based algorithms, e.g., Van de Geer and Bühlmann (2013), do not rely on the (strong) faithfulness assumption.

Regular Bayesian Networks do not take the temporal dimension into account. They assume that the observations stem from one point in time and thus have no sequential ordering. The class of BNs that allow for a sequential ordering of the data are referred to as Dynamic Bayesian Networks (DBNs). In DBNs, not only the variables themselves are included in the network, but also their lags up to some order $k$. The edges connecting variables from different periods are called inter-slice edges, whilst edges connecting variables in the same period are called intra-slice edges.

### 2.4. Causal networks

Causal networks extend the concept of BNs by explicitly focusing on identifying cause-and-effect relationships, which is particularly relevant in the context of financial markets where understanding the drivers of asset returns is crucial. However, it is important to note that the existing literature on causal algorithms can be ambiguous in its terminology, as it encompasses both causal discovery and causal inference. For our purposes in modeling stock relationships, we will focus on causal discovery algorithms, which aim to estimate a causal network from observational data, rather than causal inference, which focuses on estimating the causal effect of interventions via do-calculus (Pearl, 2010). This approach allows us to model a causal network of stocks, potentially offering deeper insights into market dynamics than traditional correlation-based methods or standard BNs. Before introducing the DYNOTEARS formulation, we briefly discuss the limitations of constraint-based Bayesian networks, which necessitate the usage of score-based methods in the financial markets setting.

A well-known example of a constraint-based algorithm is the PC (Peter-Clark) algorithm (Spirtes et al., 2000). The PC algorithm begins with the so-called skeleton estimation phase, wherein it prunes a fully connected undirected graph based on a conditional independence test. The edge between node $A$ and $B$ is removed, if $A$ and $B$ are found to be independent, conditioning on a set of nodes $C$. In the orientation phase, the PC algorithm assigns directions to the edges based on a set of rules.

Unfortunately, the feasibility of the PC algorithm for modeling financial markets is severely limited. As with most constraint-based methods, the run time of the PC algorithm is exponential to the number of nodes, which makes it unsuitable for high-dimensional financial data. Moreover, the PC algorithm assumes causal sufficiency, meaning that all common drivers are assumed to be included in the observed data. This assumption is challenging in the context of financial markets due to its factor structure; yet, in the absence of this assumption, one cannot guarantee the consistency of the PC algorithm. These limitations illustrate the immense difficulty the literature has had in effectively adopting (constraint-based) causal algorithms in the context of financial markets.

Recent advances in score-based methods have enabled practitioners to use them for high-dimensional financial data. The run time of score-based methods used to be a challenge given the acyclicity constraint on the optimization problem. This constraint renders the optimization problem a Combinatorial Optimization Problem (COP), and it is generally infeasible to exhaustively search the solution space of a COP (Korte et al., 2011). The seminal work of Zheng et al. (2018) reformulated the acyclicity constraint to be smooth, continuous, and exact. The authors named this reformulation NOTEARS which stands for *Non-combinatorial Optimization via Trace Exponential and Augmented Lagrangian for Structure earning*. The crucial benefit of NOTEARS is that standard solvers such as L-BFGS-B (Byrd et al., 1995; Zhu et al., 1997) can be applied, making score-based methods more tractable.

## 3. Causal network estimation

In this section we present the methodology we use to construct the causal networks for our experiments, the method to convert these causal networks into peer groups, and the empirical details for our three experiments.

### 3.1. DYNOTEARS model

Several papers have extended the NOTEARS methodology of Zheng et al. (2018), with notable examples including DYNOTEARS (Pamfil et al., 2020), NTS-NOTEARS (Sun et al., 2023), and GraphNOTEARS (Fan et al., 2023). The DYNOTEARS method uses NOTEARS for estimating a Dynamic Bayesian Network, and this network can be formulated as a structural vector autoregressive (SVAR) model[7]:

$$\mathbf{X} = \mathbf{X}\mathbf{W} + \mathbf{Y}_1\mathbf{A}_1 + ... + \mathbf{A}_p\mathbf{Y}_p + \mathbf{Z}, \tag{2}$$

where $\mathbf{X}$is an $n \times d$ matrix, $p$ is the autoregressive order, the $\mathbf{Y}_i$ are time-lagged versions of $\mathbf{X}$, and $\mathbf{Z}$ are the errors. For the resulting DBN to be acyclic, only the intra-slice weights $\mathbf{W}$ need to be acyclic. A computationally efficient way of enforcing the acyclicity constraint on $\mathbf{W}$ was introduced by Zheng et al. (2018), where they show that the matrix $\mathbf{W}$ is a DAG if and only if:

$$h(\mathbf{W}) = trace(e^{\mathbf{W}\circ\mathbf{W}} - d) = 0, \tag{3}$$

where $e^{\mathbf{W}}$ is the matrix exponential of $\mathbf{W}$ and $d$ is the number of columns of $\mathbf{W}$. The $\mathbf{A}_i$ are inherently acyclic as they point forward in time and thus cannot create cycles.

### 3.2. Model implementation

For estimating the causal financial networks, we opt for the DYNOTEARS method. The benefit of this method is that the underlying NOTEARS technique makes it feasible for our high-dimensional dataset, unlike many of the other methods available in the literature. Furthermore, Pamfil et al. (2020) show that DYNOTEARS outperforms other causal discovery algorithms in retrieving the causal structure of simulated datasets and the DREAM4 (Marbach et al., 2009) datasets.[8]

Having detailed the model specification, we now discuss our model implementation choices. First, we set the number of lags $p$ to zero, and thus estimate the causal effect of stock $i$ on stock $j$ as the intra-slice weight $\hat{w}_{ij}$.[9] This choice follows from the observation that at almost all estimation points the inter-slice weights for the off-diagonal elements of $\mathbf{A}_i$ are almost always zero. This is consistent with Pamfil et al. (2020) who make a similar observation for

[7]We opt for an SVAR model using a rolling window estimation as is common in the financial literature to balance capturing time-variation with computational feasibility. An alternative approach would be to use a SVAR model with time-varying parameters, such as the time-varying-parameter SVAR (TVP-SVAR).

[8]Indeed, in unreported results we find the DYNOTEARS algorithm to outperform simple correlation-based methods in recovering simulated causal structures. These results are available upon request.

[9]Notably, the intra-slice weights constitute a DAG $G$. This graph representation will be essential for our investment applications.

the S&P 100, and suggest that the contemporaneous returns are the most relevant. Second, we apply a log transformation to the stock returns and normalize the resulting log returns to have a mean of zero, and a variance of one. Without this normalization, the regularization scheme would cause the method to prefer low-variance stocks over high-variance stocks. Third, we use a sliding window approach to allow for a time-varying network. We set the window size to four years and the window increment to one month. This window size is roughly in line with Pamfil et al. (2020) in their descriptive analysis of the S&P 100. Fourth, due to the computational burden of this method, it is infeasible to tune the regularization parameter $\lambda_{\mathbf{W}}$. Therefore, we set $\lambda_{\mathbf{W}}$ to 0.1 which is in line with Pamfil et al. (2020), who calibrate this parameter via a grid search approach. These implementation choices yield the final optimization problem:

$$\min_{W} \frac{1}{2n} \|\mathbf{X} - \mathbf{X}\mathbf{W}\|_F + \lambda_{\mathbf{W}} \|\mathbf{W}\|_1 \quad \text{s.t. } \mathbf{W} \text{ is acyclic,} \tag{4}$$

where $X$ are the normalized log returns, $\frac{1}{2n} \|\mathbf{X} - \mathbf{X}\mathbf{W}\|_F$ is the least-squares loss, and $\lambda_{\mathbf{W}}$ is the $L_1$ regularization parameter. DYNOTEARS clusters are obtained by applying node2vec (Grover and Leskovec, 2016) to the DYNOTEARS estimated causal network.[10]

## 4. Experimental setup

In this section we detail the methodology used in each of the three empirical experiments. Specifically, we define the objective of the experiment, the method used to obtain the results, and the different performance evaluation approaches.

### 4.1. Experiment 1: Peer group neutralization

The aim of this experiment is to evaluate the efficacy of different peer group selection methods in the context of peer group neutralization for long–short equity strategies.

[10] Further detail on clustering is given in Appendix A.1.

#### *4.1.1. Factor selection*

We select twelve well-known and representative firm characteristics: twelve-minus-one month momentum, one month reversal, beta (using sixty months of returns), book-to-price, cash-to-assets, earnings-to-price, EBITDA-to-EV, one-year forward earnings-to-price, gross profitability-to-assets, residual twelve-minus-one month momentum (Blitz et al., 2011), and return on equity.[11] These characteristics were chosen based on their prevalence in the literature and their representation of various aspects of firm performance and valuation.

#### *4.1.2. Peer group selection*

We compare three peer group selection methods in our study. The first is the Global Industry Classification Standard (GICS), which serves as our baseline industry classification. The second method is DYNOTEARS, implemented as described in Section 3.1. For DYNOTEARS parameters, we use a window size of 4 years, a window increment of 1 month, and regularization parameters $\lambda_W = \lambda_A = 0.1$. The third method is Statistical Clustering (SC), which uses hierarchical clustering on PCA-transformed returns, following Avellaneda and Serur (2020). For SC, we use the number of principal components $K$ to match the GICS sector count (typically 10 or 11), employ Ward's method for clustering, and use the same window size and increment as DYNOTEARS. We chose this simple correlation-based statistical clustering approach as a representative technique, given the prevalence of using asset correlations in portfolio construction.[12]

The choice of hierarchical clustering on PCA-transformed returns as our Statistical Clustering (SC) method represents a widely used correlation-based approach in financial literature. This method captures the essence of correlation-based techniques by leveraging the covariance structure of returns, which is fundamentally what correlation measures aim to capture. By applying PCA before clustering, we reduce noise and focus on the most significant patterns in return correlations, making it a robust representation of correlation-based methods. This approach contrasts with DYNOTEARS in several key aspects. While SC

[11] For one month reversal and beta, we flip the sign of the characteristic (i.e., low values are associated with high expected returns). Financial ratios are all estimated using the most recently available data as at the end of each month.

[12] This is a common approach in the literature, see Mantegna (1999) and Bonanno et al. (2003, 2004).

relies on symmetric measures of co-movement (correlations), DYNOTEARS aims to uncover directed, potentially asymmetric relationships between stocks. Furthermore, SC is based on contemporaneous relationships, whereas DYNOTEARS can capture both contemporaneous and lagged relationships, potentially offering a richer representation of stock interactions. While we acknowledge that there are numerous correlation-based methods available, our chosen SC approach serves as a representative benchmark against which to compare the causal-based DYNOTEARS method. This comparison allows us to highlight the possible benefits of causal discovery in peer group selection, while recognizing that a more comprehensive evaluation of multiple correlation-based methods could provide additional insights in future research.

#### *4.1.3. Portfolio construction and performance evaluation*

For each characteristic and classification scheme, we form long–short portfolios within each peer group or industry. At time $t$, we generate a ranking of stocks in each cluster based on the characteristic. We then go long in the top 20% and short in the bottom 20%, employing a quintile long–short strategy.[13] We rebalance these portfolios monthly and construct both equal-weighted (EW) and value-weighted (VW) versions. We consider both one-month and twelve-month holding periods, with the latter following the overlapping portfolio approach of Jegadeesh and Titman (1993). As a benchmark, we include a trading strategy that does not use peer group neutralization. This benchmark strategy generates a ranking based on a given characteristic over all stocks and produces a quintile long–short strategy using this ranking.

### 4.2. Experiment 2: Centrality as a cross-sectional long–short equity factor

The aim of this experiment is to investigate the role of stock-level centrality in the cross-section of equity returns by constructing and evaluating a low centrality factor.

[13] Additionally for beta 60M we follow Blitz, van Vliet, and Baltussen (2020) and make the portfolio beta neutral by ensuring the beta of the long leg and short legs have full-sample market betas of one.

#### 4.2.1. Centrality measure

For our primary measure of stock-level centrality, we employ eigenvector centrality. For each month in our sample, we compute the eigenvector centrality for each stock based on the DYNOTEARS causal network. This choice is motivated by several factors. Eigenvector centrality accounts for both direct and indirect connections in the network, aligning well with the intuition that connections to more important nodes contribute more to a node's own importance. Moreover, Dablander and Hinne (2019) demonstrate that for causal networks, eigenvector centrality is a superior choice compared to other node centrality measures.

#### 4.2.2. Factor portfolio construction and performance evaluation

We construct a monthly low centrality factor using the following approach. First, we calculate eigenvector centrality scores for each stock in the S&P 500 using the causal network estimated by DYNOTEARS. We then rank stocks based on the inverse of their centrality scores, assigning higher ranks to more peripheral (less central) stocks. Using these rankings, we form value-weighted quintile portfolios. The low centrality factor is constructed as a long position in the top quintile (most peripheral stocks) and a short position in the bottom quintile (most central stocks). We rebalance this factor portfolio at the end of each month.

To evaluate the performance of the low centrality factor, we use standard performance measures on the long–short portfolio such as annualized return and standard deviation of returns. Additionally, we perform spanning time-series regressions against several common factor models to test the uniqueness of the low centrality factor. Specifically, we report alpha, which is the intercept from regressions of the form:

$$r_{p,t} = \alpha_t + \sum_{i=1}^{K} \beta_i f_{i,t} + \epsilon_t, \tag{5}$$

where $r_{p,t}$ is the long–short centraliy factor return at time $t$ and $f_{i,t}$ is the return of control factor $i$ at time $t$.

For control factors we include the market return in excess of the risk-free rate (MKT-Rf), and the Fama-French factors: Small-Minus-Big (SMB), High-Minus-Low (HML), Up-Minus-Down (UMD), Robust-Minus-Weak (RMW), Conservative-Minus-Aggressive (CMA), Short-Term Reversal (STR), and Long-Term Reversal (LTR) from the Kenneth French data library.[14] We also include the Betting-Against-Beta (BAB) factor from the AQR data library[15], and the Hou et al. (2021) $q^5$ factors of Investment (AI), Size (ME), Expected growth (EG), and Return on equity (ROE) from the Global-Q database.[16]

### 4.3. Experiment 3: Market timing

The aim of this experiment is to investigate the market timing ability of a causal network density indicator for predicting U.S. stock market returns.

#### *4.3.1. Network density indicator*

We construct our causal network density indicator ($d_t$) as:

$$d_t = \frac{1}{|U_t|} \sum_{i \in U_t} c_{i,t} - \frac{1}{|U_{t-1}|} \sum_{j \in U_{t-1}} c_{j,t-1}, \tag{6}$$

where $c_{i,t}$ is the eigenvector centrality measure of stock $i$ at time $t$ and $U_t$ is the investment universe at time $t$. This construction is like the systematic risk index of Kaya (2015) who computes the density as the average network centrality but uses an eccentricity measure instead of an eigenvector measure. We use the change in average centrality rather than the absolute level, as structural breaks in network topology might cause the absolute level to be uninformative for predictive purposes.

#### *4.3.2. Performance evaluation*

To assess the timing efficacy of the network density indicator, we follow established market timing literature methodologies (Welch and Goyal, 2008; Neely et al., 2014; Hammerschmid and Lohre, 2018). We measure in-sample return predictability using bivariate predictive regressions of the form:

$$r_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1}, \tag{7}$$

[14] Available here: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.
[15] Available here: https://www.aqr.com/Insights/Datasets/Betting-Against-Beta-Equity-Factors-Monthly.
[16] Available here: https://global-q.org/index.html.

where $r_t$ is the excess market (S&P 500) return at time $t$, and $x_t$ is the value of the indicator at time $t$.

To gauge the efficacy of our timing indicator for future performance, we follow Campbell and Thompson (2008) to estimate an out-of-sample R-squared, denoted $R^2_{OS}$. We assess economic significance by calculating the certainty equivalent return (CER) gain, as per Neely et al. (2014). The CER gain represents the annual management fee an investor would be willing to pay to access the network density forecast instead of the historical average forecast.[17]

For robustness, we conduct a multivariate regression that includes common macroeconomic variables and technical indicators as controls. We incorporate eleven macroeconomic indicators from the Welch and Goyal (2008) dataset and ten technical indicators following Hammerschmid and Lohre (2018). These technical indicators include lagged values of the excess market return (accounting for autocorrelation in market returns) and moving-average signals.[18] To manage this large set of control predictors and multicollinearity, we employ the PCA approach of Neely et al. (2014), reducing the twenty-one control predictors to three orthogonal factors.

We then use these orthogonal factors in a multivariate regression:

$$r_{t+1} = \alpha + \beta_C x_{C,t} + \sum_{k=1}^{K} \beta_k x_{k,t} + \epsilon_{t+1}, \tag{8}$$

where $x_{C,t}$ is our causal network density indicator at time $t$, and $x_{k,t}$ is a principal component factor $k$ at time $t$.

This comprehensive approach allows us to evaluate the predictive power of our causal network density indicator while controlling for a wide range of established predictors, providing a robust assessment of its efficacy in market timing.

## 5. Experiments

In this section, we present and discuss the results of the causal network investment applications. First, we compare the performance of several classification schemes for the

[17] We provide additional details in Appendix B.2.

[18] The full list of controls can be found in Appendix B.3.

purpose of peer group neutralization when constructing long–short investment strategies. Subsequently, we evaluate the economic value of stock centrality by creating a low centrality factor derived from a causal network representation of stock returns. Lastly, we assess the return predictability of a causal network density-based market timing indicator.

### 5.1. Data

In our empirical analysis, we use constituents of the S&P 500 index as a representative universe to build causal networks over. We source daily and monthly stock and index returns from Refinitiv Datastream for the period of December 1989 to December 2022. We source fundamental and company information data from S&P Compustat. Our results run from January 1993 to December 2022, because the first three years of return data are used to calibrate models. In our estimation process at the end of each month, we include all stocks in the S&P 500 as at the end of the month and have a full history of returns over the estimation window. This can lead to a different number of nodes for each estimation window. Table 1 presents summary statistics of the investment universe that we construct causal networks over. We have an average of 357 stocks in each network, which is due to our requirement of a full history of stock returns at each estimation point. We can also see that our sample of stock returns exhibits common properties, such as large tails (time-series average kurtosis of 7.55) and a wide dispersion of monthly return outcomes.

<Insert Table 1 about here>

### 5.2. Experiment 1: Peer group neutralization

#### *5.2.1. Results*

In Table 2 we present the annualized results for value-weighted trading strategies based on different classification schemes. We also report the causal alpha, which is the annualized $\alpha$ derived from a regression of the form:

$$R_t^{\text{Causal}} = \alpha + \beta_{MKT} R_{MKT} + \beta R_t^{\text{c}} + \varepsilon_t,$$

where $R_t^{\text{Causal}}$ is the long–short return of the causal neutralization and $R_t^{\text{c}}$ is the long–short return of GICS/correlation/no-neutralization based strategy. Table 2 presents results for one-month holding periods.[19] First, the strategy without neutralization (labeled None) often performs worse from a volatility perspective than its neutralized counterparts, suggesting that these characteristics suffer from unpriced industry exposure. For instance, for the 12-1M Momentum characteristic, the benchmark volatility is 20.92%, whilst the neutralized strategy volatility is between 14.64% and 16.42%. The underperformance of the no-neutralization strategy is often caused by its high volatility, causing its Sharpe ratio to decrease in turn. This increased volatility likely stems from its lack of diversification. In summary, for these specific characteristics, we find substantial economic value to peer group neutralization. Figure 2 demonstrates the risk-adjusted return improvements of the causality-based approach over the GICS and SC approaches, highlighting how the only scenario in which the causal approach underperforms, from a return perspective, is for the cash-to-assets strategy. However, from a volatility reduction perspective, we generally find that GICS and SC outperform the causality-based approach.

<Insert Table 2 about here>

Second, turning our attention to the SC variant, we find that its performance is generally worse than that of the causal and GICS strategies but, at times, better than the no-neutralization strategy. For instance, for the Book-to-price characteristic, both the causal variant and the GICS variant outperform the SC variant. Moreover, for five of the characteristics, the causal variant attains a significant alpha over the SC variant with annualized alphas up to 3.74%.

<Insert Figure 2 about here>

Third, when comparing the causal variant to GICS, we find that the causal variant generally outperforms GICS in terms of Sharpe ratio. For instance, for both the FY1 Earnings-to-price characteristic and the book-to-price characteristic, the causal variant outperforms GICS. Moreover, in both cases, the causal variant holds significant return predictability over

[19] In unreported results, we also test a twelve-month holding period and find consistent results.

the GICS variant. The causal variant has a considerable risk-adjusted return edge over the GICS variant for (most of) these characteristics. We find that these results hold for both one-month holding periods, as well as for longer holding periods. This highlights the general applicability of these causal clustering techniques for peer group neutralization.

In Table 3 we present the results of running Ledoit and Wolf's (2008) Sharpe ratio test to compare the significance of the Sharpe ratio of the causal strategies and the other strategies. In general, we find that the Sharpe ratio difference between the causal strategy and other strategies is greater than zero but not statistically significant. This reflects the fact that the Sharpe ratios of the strategies in general are less than 1.0, and thus establishing statistically significant differences using thirty years of monthly returns is challenging. Nevertheless, our results show that the causal-neutralized strategies demonstrate some statistically significant outperformance.

<Insert Table 3 about here>

Finally, our previous results are estimated using aggregated returns over a roughly thirty-year period. In unreported results, we split our sample up into decades and into high, normal, and low volatility regimes. Within each of these regimes we re-compute the Sharpe ratio and causal alpha results. The results observed within each regime are generally consistent with the full sample results. We do not find any obvious patterns across the regimes, but our results show how the outperformance of the causal strategies is not consistent across the full sample.

#### *5.2.2. Discussion*

Our findings on the efficacy of causal-based peer group neutralization contribute to the ongoing debate in the factor investing literature about the role of sector neutralization. Ehsani et al. (2023) analytically and empirically demonstrated that long–short portfolios generally benefit from sector neutralization, arguing that the across-industry component often does not carry a premium compared to the within-industry component. Our results largely support this conclusion, as we find that applying peer group neutralization generally improves risk-adjusted returns compared to non-neutralized strategies.

However, our study extends beyond traditional sector-based neutralization by introducing a causal discovery approach. While Vyas and van Baren (2021) highlighted the issue of unpriced industry exposure in equity factors due to industry tilts, our causal-based approach offers a potential solution by identifying peer groups based on complex interdependencies rather than predefined sectors. This aligns with the work of Blitz and Hanauer (2020), who showed that traditional value factors display major industry biases. Our causal approach can potentially capture more nuanced relationships between stocks, and may thus help mitigate such biases.

Our study primarily compares one causal approach (DYNOTEARS) with one representative correlation-based method (Statistical Clustering, SC) and the industry-standard GICS. While SC is a widely used method in the literature, it represents only one of many possible correlation-based approaches, and thus broad-based conclusions around causal versus correlation-based peer groups cannot be drawn. Nonetheless, it is worth considering why the SC approach performed poorly compared to DYNOTEARS and GICS. One explanation can be found in MacMahon and Garlaschelli (2015) who argue that applying community detection algorithms to correlation matrices leads to inherently poor/biased results.[20] A second explanation for this phenomenon is the "apples and oranges" narrative, that the diversification power of correlation-based ACS is harmed by its tendency to cluster companies that should not be clustered. We find support for this hypothesis, as the stock-selection power of the correlation method is markedly lower than the stock-selection power of the other variants.

Interestingly, our findings suggest that while causal-based peer groups often lead to higher Sharpe ratios, this stems primarily from a superior ability to select stocks. Traditional GICS-based neutralization still offers superior volatility reduction in many cases. This result suggests that different neutralization approaches may offer distinct benefits, and that a combination of methods might be optimal in practice. While causality-based peer groups have a stronger theoretical foundation by attempting to capture direct influence relationships rather than just statistical associations, the practical benefits appear to be context-dependent.

[20]Specifically, the authors show that modularity-oriented community detection algorithms, when applied to correlation matrices, cannot retrieve the community structure even under optimal (simulation) conditions.

For instance, we observe that causal neutralization tends to perform better for characteristics that have clear economic transmission mechanisms between firms (like momentum and value metrics) but shows less advantage for firm-specific attributes (like quality metrics) where inter-firm relationships may be less relevant. However, it is difficult to identify a universal pattern across our results which could pinpoint some latent characteristic which indicates whether the causal model will outperform.

## 5.3. Experiment 2: Centrality as a cross-sectional long–short equity factor

### *5.3.1. Results*

Figure 3 displays the cumulative return of the low centrality factor over time. The performance of the low centrality factor seems highly driven by the business cycle, as it is best before NBER recessions and worst after these recessions. This pattern is consistent with a growth strategy, as growth stocks display strong performance during bull markets, whilst this performance tends to evaporate during bear markets. Before characterizing the low centrality factor in terms of existing factors, we first establish that the low centrality factor adds value beyond these factors by investigating the alphas of the spanning regressions in Table 4 relative to several common factor models.

<Insert Figure 3 about here>

The low centrality factor attains an annualized alpha of 2.21% in the CAPM (Panel (a) column (i)), 2.78% in the FF4 model (Panel (a) column (iii)), 5.56% in the FF6 model (Panel (a) column (v)), and 5.02% in the $q^5$ model (Panel (b) column (iii)), whilst the annualized Sharpe Ratio is negative. At first glance, one might expect a higher Sharpe Ratio given the high alphas, but the factor loadings in Table 4 help rationalize this contradiction. The low centrality factor loads negatively on the MKT and the HML/IA factors, both of which carry positive risk premia. Therefore, by negatively loading on these factors, the expected return and the Sharpe Ratio of the low centrality factor are lowered. Notably, the $q^5$ alpha is higher than the FF4 alpha, in part because the low centrality factor loads more negatively on the investment factor than on the value factor.

<Insert Table 4 about here>

Given that its alphas are significant and sizable, the low centrality factor adds orthogonal information to factor models such as Fama and French (1993) and Hou et al. (2021). We visualize the cumulative alpha (or alpha-add) over time in Figure 3. We define the cumulative alpha of the low centrality factor as the cumulative sum of $\hat{\alpha} + \varepsilon_{:t}$, where $\varepsilon_{:t}$ are the residuals up to and including time $t$. Beyond the difference in trend, we observe that the alpha-add in the FF6 model is similar to the alpha-add in the $q^5$ model. This is to be expected as the factor loadings on MKT-Rf, SMB/ME, and HML/IA are similar. For both models, the alpha-add over time is not consistent, as the alpha-add of the low centrality factor seems roughly related to its performance (as visualized in Figure 3).

Next, we characterize the low centrality factor in terms of existing factors. Indeed, based on the left partition of Table 4, the conjecture that the low centrality factor resembles a growth strategy rings true. Namely, the low centrality factor negatively loads on the HML value factor with a significant coefficient of -0.34 in the FF3 model. This anti-value tilt is consistent with the growth definition of Fama and French (2007). In their valuation-based perspective, growth stocks are defined as stocks with an inflated valuation, e.g., as reflected in a low book-to-price ratio. In this definition, a growth stock is simply the opposite of a value stock.

However, the definition of a growth stock is somewhat contested in the broader literature. Rather than a valuation-based perspective, others have taken a profitability-based or an investment-based perspective on growth. For instance, Novy-Marx (2013) argues that strategies based on profitability, as measured by the gross profits-to-assets ratio, are growth strategies. In this perspective, profitability/growth strategies can substantially improve the performance of value strategies by acting as a hedge. Hou et al. (2021) take an investment-based perspective on growth and argue that trading on the (estimated) growth rate of investment-to-assets should be considered a growth strategy. Therefore, based on Panel (b) of Table 4, the low centrality factor should not be deemed a growth strategy as the coefficient of the expected growth factor is insignificant.

The spanning regressions offer an insightful characterization of central stocks. Given that the low centrality factor operates on a reverse sort of centrality (i.e., we buy peripheral stocks and short central stocks), we can effectively reverse the factor loadings to get a better understanding of central stocks. The foremost conclusion is that central stocks are large, value-heavy companies. This finding is supported by Buraschi and Porchia (2012) who find that the book-to-price ratio is positively related to their centrality measure, and thus that central stocks are generally value-heavy. Additionally, Konstantinov et al. (2020) conclude that the value factor has a central position in the financial network, albeit at the asset allocation level. Thus, our finding is an important step in clarifying the financial interpretation of centrality measures.

### *5.3.2. Discussion*

Our findings on the low centrality factor contribute to and extend the growing body of literature on network-based approaches in asset pricing. The significant alpha generated by our factor across various model specifications aligns with previous studies that have found network centrality measures to be informative for asset returns. For instance, Pozzi et al. (2013) found that peripheral companies in correlation-based networks offer lower risk and higher returns than central companies, which is consistent with the performance of our low centrality factor. However, our use of causal networks, as opposed to correlation networks, provides a novel perspective on this relationship.

Our results also resonate with the work of Výrost et al. (2019), who argued that highly central stocks offer less diversification than peripheral stocks. While they focused on diversification benefits, our study extends their insight by demonstrating that this characteristic translates into a priced factor in the cross-section of returns. Methodologically, our use of the DYNOTEARS algorithm for causal discovery sets our work apart from most existing studies in financial networks, which typically rely on correlation-based or Granger causality approaches (e.g., Billio et al., 2012). This causal approach potentially captures more nuanced and directional relationships between assets, offering a new perspective on how information flows and risk propagate through financial markets.

For example, we find that the alpha-add is much larger during stable regimes than that in (the aftermath of) unstable regimes. This suggests that in stable regimes, the existing factors fail to incorporate centrality more so than in unstable regimes. In turn, the low centrality factor can effectively act as a cheap hedge as it does not underperform, but it controls for risk in other factors which are known to perform well. However, the performance of the low centrality factor in the 2020 to 2023 period highlights that this is not a free hedge, as it significantly underperforms.

Finally, Figure 4 presents a graphical depiction of the causal network at selected months. We selected months with particularly extreme densities, as a means to demonstrate how the network density varies. Generally, a dense network means that more of the underlying stocks are connected to each other, and the coefficient associated with these edges are larger. Generally, networks tend to become more dense around crisis periods, and early detections of these shifts in density can assist with market timing. We explore the latter avenue in the next sub-section.

<Insert Figure 4 about here>

### 5.4. Experiment 3: Market timing

#### *5.4.1. Results*

We first review the bivariate predictive regression on network density, which yields an $R^2$ of 0.75% (as seen in column (i) of Table 5). This suggests that the network density carries some return predictability, as the $R^2$ exceeds the often-mentioned $R^2$ threshold of 0.5% (Zhou, 2010; Neely et al., 2014). Moreover, the coefficient $\hat{\beta}$ possesses the expected sign and is significant, albeit at a 10% level. Namely, the coefficient is negative, indicating that an increase in network density is associated with lesser expected returns. [21]

<Insert Table 5 about here>

Validating the predictability of the network density indicator out-of-sample, we find an $R^2_{OS}$ of 0.55%, which is significant at a 10% level. The level of out-of-sample return predictability is like the return predictability of a number of technical indicators seen in Neely et al. (2014) and Hammerschmid and Lohre (2018). Particularly, Neely et al. (2014) find that moving-average and volatility-based technical indicators carry significant (at the 10% level) predictive value out-of-sample. For these indicators, they find $R^2_{OS}$ ranging from 0.44%

[21] In unreported results, we find similar results when performing a simple multi-variate regression incorporating lagged market return values alongside the centrality density measure.

to 0.88%, some of which are even significant at the 5% level. Likewise, Hammerschmid and Lohre (2018) also find that moving-average and volatility-based technical indicators carry significant predictive value out-of-sample. They find $R^2_{OS}$ ranging from 0.59% to 1.03%, none of which are significant at a 5% level.

We evaluate the economic utility of the network density proxy by computing the CER gain ($\Delta$) which amounts to 1.83% for the bivariate forecast. Similar to the $R^2_{OS}$, this value is roughly in line with the value of the technical indicators in Neely et al. (2014) and Hammerschmid and Lohre (2018). Particularly, for technical indicators that have a significant $R^2_{OS}$, Hammerschmid and Lohre (2018) find annualized CER gains ranging from 1.9% to 2.6% whilst Neely et al. (2014) find annualized CER gains ranging from 1.5% to 2.9%. Based on the value of $\Delta$, we conclude that the network density proxy has economically substantial predictive power for the (monthly) equity risk premium.

Given that network density carries a similar amount of predictive value, it is interesting to evaluate how different network density is from these technical indicators. To this end, column (ii) of Table 5 presents the same results but including a set of three principal component control factors. We find qualitatively similar results as the bivariate regression in column (i), whereby we have a negative coefficient on $\hat{\beta}$ that is significant at the 10% level. Hence, the predictability of the causal network-based density indicator is not explained by the information contained in common control variables related to macroeconomic variables and technical indicators.

*5.4.2. Discussion*

Our findings on the predictive power of the causal network density indicator contribute to the extensive literature on market timing and return predictability. The observed $R^2$ of 0.75% in our bivariate predictive regression exceeds the often-cited threshold of 0.5% for economic significance (Zhou, 2010; Neely et al., 2014), suggesting that network density carries meaningful predictive information for market returns. This aligns with the findings of Rapach et al. (2022), who argued for the importance of considering interconnectedness in financial markets for return prediction.

A potential explanation for this relatively high predictability is that the network density is a good indicator for recessions. Intuitively, one would expect more variance during recessions than during expansionary periods. Therefore, if a predictor functions well in recessions, it will likely have a higher $R^2$ than predictors that function equally well in expansionary periods. This reasoning is empirically supported by Henkel et al. (2011) who show that the return predictability for many predictors is concentrated in recessionary periods.

Comparing our results to those of traditional predictors, we find that the performance of our network density indicator is competitive with, and in some cases superior to, established macroeconomic and technical indicators. For instance, Neely et al. (2014) reported $R^2_{OS}$ values ranging from 0.44% to 0.88% for various technical indicators, while our indicator achieves an $R^2_{OS}$ of 0.55%. Similarly, our annualized CER gain of 1.83% is comparable to the range of 1.5% to 2.9% reported by Neely et al. (2014) for significant technical indicators.

Our findings align with and extend the work of Kaya (2015), who used aggregate network eccentricity as a proxy for network density. While Kaya focused on network topology becoming denser prior to market crashes, our approach provides a more continuous measure of market conditions. Furthermore, our use of causal networks, as opposed to correlation-based networks, potentially captures more nuanced and directional relationships between assets.

## 6. Conclusion

In this paper, we explore the potential of a representative causal discovery algorithm (DYNOTEARS) in factor investing, focusing on its application to financial network modeling. We applied DYNOTEARS-derived causal networks to three common investment applications, aiming to complement and extend the prevalent correlation-based framework of Mantegna (1999).

Our findings suggest that causal discovery algorithms can offer valuable insights in factor investing. DYNOTEARS provides alternative peer groups that, in many cases, lead to improved risk-adjusted returns when constructing peer group neutral long–short strategies compared to traditional methods. However, it is important to note that established methods like industry classifications remain relevant, particularly for volatility reduction. We also demonstrate that a causal network-based low centrality factor can serve as a useful hedging tool when combined with common factor models such as Fama-French and $q^5$. Additionally, our causal network-based market timing indicator shows promise in predicting S&P 500 excess returns, performing comparably to existing indicators in the literature. These results indicate that causal discovery algorithms can be a valuable addition to the factor investing toolkit, offering novel perspectives and insights to traditional factor investing paradigms. However, it is crucial to view these techniques as complementary to, rather than superior to, traditional methods. Each approach has its strengths and may be more suitable in different contexts.

While our study demonstrates the potential of causal discovery algorithms in factor investing, several limitations should be acknowledged. Causal discovery algorithms, including DYNOTEARS, can be computationally intensive, especially when applied to large datasets. Our study focused on a specific causal discovery algorithm (DYNOTEARS) and a limited set of investment applications, which does not fully represent the breadth of potential applications in factor investing. Finally, we do not explicitly horse race causal discovery algorithms against alternative benchmarks, instead we opt to present a broad range of applications and show results in line with previous literature.

Future work in this area could address these limitations and further explore the potential of causal discovery in factor investing. More comprehensive comparisons between different causal discovery algorithms and traditional methods across a wider range of investment applications would provide a clearer picture of their relative strengths and weaknesses. Finally, developing tools and methodologies to improve the interpretability and computational complexity of causal discovery outputs could increase their adoption among practitioners.

## References


Asness, C. S., A. Frazzini, and L. H. Pedersen (2018). Quality minus junk. *Review of Accounting Studies 24*(1), 34–112. doi:10.1007/s11142-018-9470-2.

Asness, C. S., R. B. Porter, and R. L. Stevens (2000). Predicting stock returns using industry-relative firm characteristics. *SSRN working paper no. 213872*. doi:10.2139/ssrn.213872.

Avellaneda, M. and J. A. Serur (2020). Hierarchical PCA and modeling asset correlations. *arXiv working paper no. 2010.04140*. doi:10.48550/arXiv.2010.04140.

Bartram, S., H. Lohre, P. Pope, and A. Ranganathan (2021, 04). Navigating the factor zoo around the world: an institutional investor perspective. *Journal of Business Economics 91*, 1–49. doi:10.1007/s11573-021-01035-y.

Bender, J., R. Mohamed, and X. Sun (2019). Country and sector bets: Should they be neutralized in global factor portfolios? *Journal of Beta Investment Strategies 10*(1), 60–74. doi:10.3905/jii.2019.1.069.

Bhojraj, S., C. M. Lee, and D. K. Oler (2003). What's my line? A comparison of industry classification schemes for capital market research. *Journal of Accounting Research 41*(5), 745–774. doi:10.1046/j.1475-679X.2003.00122.x.

Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics 104*(3), 535–559. doi:10.1016/j.jfineco.2011.12.010.

Black, F. (1972). Capital market equilibrium with restricted borrowing. *Journal of Business 45*(3), 444–455. doi:10.1086/295472.

Blitz, D. and M. X. Hanauer (2020). Resurrecting the value premium. *Journal of Portfolio Management 47*(2), 63–81. doi:10.3905/jpm.2020.1.188.

Blitz, D., M. X. Hanauer, I. Honarvar, R. Huisman, and P. van Vliet (2023). Beyond Fama-French factors: Alpha from short-term signals. *Financial Analysts Journal 79*(4), 96–117. doi:10.1080/0015198x.2023.2173492.

Blitz, D., J. Huij, and M. Martens (2011). Residual momentum. *Journal of Empirical Finance 18*(3), 506–521. doi:10.1016/j.jempfin.2011.01.003.

Blitz, D., P. van Vliet, and G. Baltussen (2020). The volatility effect revisited. *Journal of Portfolio Management 46*(2), 45–63. doi:10.3905/jpm.2019.1.114.

Blitz, D. and M. Vidojevic (2019). The characteristics of factor investing. *Journal of Portfolio Management 45*(3), 69–86. doi:10.3905/jpm.2019.45.3.069.

Blitz, D. C. and P. van Vliet (2007). The volatility effect. *Journal of Portfolio Management 34*(1), 102–113. doi:10.3905/jpm.2007.698039.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment 2008*(10), P10008. doi:10.1088/1742-5468/2008/10/P10008.

Bonanno, G., G. Caldarelli, F. Lillo, and R. N. Mantegna (2003). Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E 68*(4), 046130. doi:10.1103/PhysRevE.68.046130.

Bonanno, G., G. Caldarelli, F. Lillo, S. Micciche, N. Vandewalle, and R. N. Mantegna (2004). Networks of equities in financial markets. *European Physical Journal B 38*, 363–371. doi:10.1140/epjb/e2004-00129-6.

Buraschi, A. and P. Porchia (2012). Dynamic networks and asset pricing. *SSRN working paper no. 2024483*. doi:10.2139/ssrn.2024483.

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing 16*(5), 1190–1208. doi:10.1137/0916069.

Campbell, J. Y. and S. B. Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies 21*(4), 1509–1531. doi:10.1093/rfs/hhm055.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance 52*(1), 57–82. doi:10.1111/j.1540-6261.1997.tb03808.x.

Chan, L. K., J. Lakonishok, and B. Swaminathan (2007). Industry classifications and return comovement. *Financial Analysts Journal 63*(6), 56–70. doi:10.2469/faj.v63.n6.4927.

Clark, T. E. and K. D. West (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics 138*(1), 291–311. doi:10.1016/j.jeconom.2006.05.023.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance 66*(4), 1047–1108. doi:10.1111/j.1540-6261.2011.01671.x.

Cohen, R. B. and C. Polk (1996). An investigation of the impact of industry factors in asset-pricing tests. *Kellogg Graduate School of Management working paper no. 96-02*.

Cornell, C., L. Mitchell, and M. Roughan (2024). Enhancing causal discovery in financial networks with piecewise quantile regression. *arXiv Working Paper No. 2408.12210*. doi:10.48550/arXiv.2408.12210.

Costa, M. and L. De Angelis (2011). Sector classification in stock markets: A latent class approach. In *Classification and Multivariate Analysis for Complex Data Structures*, pp. 229–236. Springer. doi:10.1007/978-3-642-13312-1_23.

Dablander, F. and M. Hinne (2019). Node centrality measures are a poor substitute for causal inference. *Scientific Reports 9*(1), 6846. doi:10.1038/s41598-019-43033-9.

D'Acunto, G., P. Bajardi, F. Bonchi, and G. De Francisci Morales (2021). The evolving causal structure of equity risk factors. In *Proceedings of the Second ACM International Conference on AI in Finance*, pp. 1–8. doi:10.1145/3490354.349437.

De Meo, P., E. Ferrara, G. Fiumara, and A. Provetti (2011). Generalized Louvain method for community detection in large networks. In *2011 11th International Conference on Intelligent Systems Design and Applications*, pp. 88–93. IEEE. doi:10.1109/ISDA.2011.6121636.

Diebold, F. X. and K. Yılmaz (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics 182*(1), 119–134. doi:10.1016/j.jeconom.2014.04.012.

Dong, X., H. Dai, Y. Fan, S. Jin, S. Rajendran, and K. Zhang (2024). On the three demons in causality in finance: Time resolution, nonstationarity, and latent factors. *arXiv Working Paper No. 2401.05414*. doi:10.48550/arXiv.2401.05414.

Dornau, R. (1999). Shock around the clock—on the causal relations between international stock markets, the strength of causality and the intensity of shock transmission: an econometric analysis. *Intelligent Systems in Accounting, Finance and Management 8*(4), 253–270. doi:10.1002/(SICI)1099-1174(1).

Durcheva, M. and P. Tsankov (2021). Granger causality networks of S&P 500 stocks. *AIP Conference Proceedings 2333*(1), 110014. doi:10.1063/5.0041747.

Ehsani, S., C. R. Harvey, and F. Li (2023). Is sector neutrality in factor investing a mistake? *Financial Analysts Journal 79*(3), 1–23. doi:10.1080/0015198X.2023.2196931.

Eng-Uthaiwat, H. (2018). Stock market return predictability: Does network topology matter? *Review of Quantitative Finance and Accounting 51*, 433–460. doi:10.1007/s11156-017-0676-3.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56. doi:10.1016/0304-405X(93)90023-5.

Fama, E. F. and K. R. French (2007). The anatomy of value and growth stock returns. *Financial Analysts Journal 63*(6), 44–54. doi:10.2469/faj.v63.n6.4926.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116* (1), 1–22. doi:10.1016/j.jfineco.2014.10.010.

Fan, S., S. Zhang, X. Wang, and C. Shi (2023). Directed acyclic graph structure learning from dynamic graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Volume 37, pp. 7512–7521. doi:10.1609/aaai.v37i6.2591.

Ferreira, M. A. and P. Santa-Clara (2011). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics 100* (3), 514–537. doi:10.1016/j.jfineco.2011.02.003.

Fiedor, P. (2014). Causal non-linear financial networks. *arXiv Working Paper No. 1407.5020*. doi:10.48550/arXiv.1407.5020.

Frazzini, A. and L. H. Pedersen (2014). Betting against beta. *Journal of Financial Economics 111* (1), 1–25. doi:10.1016/j.jfineco.2013.10.005.

Freeman, L. C. et al. (2002). Centrality in social networks: Conceptual clarification. *Social Networks 1* (3), 238–263. doi:10.1016/0378-8733(78)90021-7.

French, R. (2023). Peer group identification in factor portfolios: A data-driven approach. *Journal of Portfolio Management 50*, 149–173. doi:10.3905/jpm.2023.1.566.

Gao, B. and R. en Ren (2013). The topology of a causal network for the Chinese financial system. *Physica A: Statistical Mechanics and its Applications 392* (13), 2965–2976. doi:10.1016/j.physa.2013.02.015.

Gao, Y.-C., Z.-W. Wei, and B.-H. Wang (2013). Dynamic evolution of financial network and its relation to economic crises. *International Journal of Modern Physics C 24* (02), 1350005. doi:10.1142/S0129183113500058.

Gnecco, N., N. Meinshausen, J. Peters, and S. Engelke (2020). Causal discovery in heavy-tailed models. *arXiv Working Paper No. 1908.05097*. doi:10.48550/arXiv.1908.05097.

Grover, A. and J. Leskovec (2016). node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 855–864. doi:10.1145/2939672.2939754.

Haldane, A. G. (2013). *Rethinking the financial network.* Springer. doi:10.1007/978-3-658-02248-8_17.

Hammerschmid, R. and H. Lohre (2018). Regime shifts and stock return predictability. *International Review of Economics & Finance 56*, 138–160. doi:10.1016/j.iref.2017.10.021.

Henkel, S. J., J. S. Martin, and F. Nardari (2011). Time-varying short-horizon predictability. *Journal of Financial Economics 99* (3), 560–580. doi:10.1016/j.jfineco.2010.09.008.

Hoberg, G. and G. Phillips (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy 124*(5), 1423–1465. doi:10.1086/688176.

Hosseini, S. S., N. Wormald, and T. Tian (2021). A weight-based information filtration algorithm for stock-correlation networks. *Physica A: Statistical Mechanics and its Applications 563*, 125489. doi:10.1016/j.physa.2020.125489.

Hou, K., H. Mo, C. Xue, and L. Zhang (2021). An augmented q-factor model with expected growth. *Review of Finance 25*(1), 1–41. doi:10.1093/rof/rfaa004.

Hou, K., C. Xue, and L. Zhang (2014). Digesting anomalies: An investment approach. *Review of Financial Studies 28*(3), 650–705. doi:10.1093/rfs/hhu068.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance 45*(3), 881–898. doi:10.1111/j.1540-6261.1990.tb05110.x.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48*(1), 65–91. doi:10.1111/j.1540-6261.1993.tb04702.x.

Jensen, T. I., B. Kelly, and L. H. Pedersen (2023). Is there a replication crisis in finance? *Journal of Finance 78*(5), 2465–2518. doi:10.1111/jofi.13249.

Jung, S. S. and W. Chang (2016). Clustering stocks using partial correlation coefficients. *Physica A: Statistical Mechanics and its Applications 462*, 410–420. doi:10.1016/j.physa.2016.06.094.

Kaya, H. (2015). Eccentricity in asset management. *Network Theory in Finance 1*(1), 1–32. doi:10.21314/JNTF.2015.003.

Konstantinov, G., A. Chorus, and J. Rebmann (2020). A network and machine learning approach to factor, asset, and blended allocation. *Journal of Portfolio Management 46*(6), 54–71. doi:10.3905/jpm.2020.1.147.

Korte, B. H., J. Vygen, B. Korte, and J. Vygen (2011). *Combinatorial optimization*, Volume 1. Springer. doi:10.1007/978-3-662-56039-6.

Laloux, L., P. Cizeau, J.-P. Bouchaud, and M. Potters (1999). Noise dressing of financial correlation matrices. *Physical Review Letters 83*(7), 1467. doi:10.1103/PhysRevLett.83.1467.

Ledoit, O. and M. Wolf (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance 15*(5), 850–859.

Lee, C., P. Ma, and C. C. Wang (2016). The search for peer firms: When do crowds provide wisdom? *Harvard Business School Accounting & Management Unit Working Paper* (15-032), 14–46. doi:10.2139/ssrn.2517496.

Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *Quarterly Journal of Economics 105*(1), 1–28. doi:10.2307/2937816.

Lenzu, S. and G. Tedeschi (2012). Systemic risk on different interbank network topologies. *Physica A: Statistical Mechanics and its Applications 391*(18), 4331–4341. doi:10.1016/j.physa.2012.03.035.

Liu, J., H. Li, M. Hai, and Y. Zhang (2023). A study of factors influencing financial stock prices based on causal inference. *Procedia Computer Science 221*, 861–869. doi:10.1016/j.procs.2023.08.062.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory 28*(2), 129–137. doi:10.1109/TIT.1982.1056489.

Lohre, H., J. Papenbrock, and M. Poonia (2014). The use of correlation networks in parametric portfolio policies. *SSRN working paper no. 2505732*. doi:10.2139/ssrn.2505732.

López de Prado, M. (2016). Building diversified portfolios that outperform out of sample. *Journal of Portfolio Management 42*(4), 59–69. doi:10.3905/jpm.2016.42.4.059.

López de Prado, M. (2023a). *Causal Factor Investing: Can Factor Investing Become Scientific?* Cambridge University Press. doi:10.1017/9781009397315.

López de Prado, M. (2023b). Where are the factors in factor investing? *Journal of Portfolio Management 49*(5), 6–20. doi:10.3905/jpm.2023.1.477.

MacMahon, M. and D. Garlaschelli (2015). Community detection for correlation matrices. *Physical Review. X 5*(2), 021006. doi:10.1103/PhysRevX.5.021006.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *European Physical Journal B-Condensed Matter and Complex Systems 11*, 193–197. doi:10.1007/s100510050929.

Marbach, D., T. Schaffter, C. Mattiussi, and D. Floreano (2009). Generating realistic in silico gene networks for performance assessment of reverse engineering methods. *Journal of Computational Biology 16*(2), 229–239. doi:10.1089/cmb.2008.09TT.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance 7*(1), 77–91. doi:10.2307/2975974.

Neely, C. J., D. E. Rapach, J. Tu, and G. Zhou (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science 60*(7), 1772–1791. doi:10.1287/mnsc.2013.1838.

Noels, S., S. De Ridder, S. Viaene, and T. De Bie (2023). An efficient graph-based peer selection method for financial statements. *Intelligent Systems in Accounting, Finance and Management 30*(3), 120–136. doi:10.1002/isaf.1539.

Novy-Marx, R. (2013). The other side of value: Good growth and the gross profitability premium. *Journal of Financial Economics 108*(1), 1–28. doi:10.1016/j.jfineco.2013.01.003.

Pamfil, R., N. Sriwattanaworachai, S. Desai, P. Pilgerstorfer, K. Georgatzis, P. Beaumont, and B. Aragam (2020). DYNOTEARS: Structure learning from time-series data. In *International Conference on Artificial Intelligence and Statistics*, pp. 1595–1605. PMLR.

Pearl, J. (1985). Bayesian netwcrks: A model self-activated memory for evidential reasoning. In *Proceedings of the 7th Conference of the Cognitive Science Society*, pp. 15–17.

Pearl, J. (2010). Causal inference. *Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008*, 39–58.

Peralta, G. and A. Zareei (2016). A network approach to portfolio selection. *Journal of Empirical Finance 38*(A), 157–180. 10.1016/j.jempfin.2016.06.003.

Phillips, R. L. and R. Ormsby (2016). Industry classification schemes: An analysis and review. *Journal of Business & Finance Librarianship 21*(1), 1–25. doi:10.1080/08963568.2015.1110229.

Pozzi, F., T. Di Matteo, and T. Aste (2013). Spread of risk across financial markets: Better to invest in the peripheries. *Scientific Reports 3*(1), 1665. doi:10.1038/srep01665.

Ros, F. and S. Guillaume (2019). A hierarchical clustering algorithm and an improvement of the single linkage criterion to deal with noise. *Expert Systems with Applications 128*, 96–108. doi:10.1016/j.eswa.2019.03.031.

Sadeghi, A., A. Gopal, and M. Fesanghary (2024). Causal discovery in financial markets: A framework for nonstationary time-series data. *arXiv Working Paper no. 2312.17375*. doi:10.48550/arXiv.2312.17375.

Shirokikh, O., G. Pastukhov, A. Semenov, S. Butenko, A. Veremyev, E. L. Pasiliao, and V. Boginski (2022). Networks of causal relationships in the us stock market. *Dependence Modeling 10*(1), 177–190. doi:10.1515/demo-2022-0110.

Simon, H. A. (1954). Spurious correlation: A causal interpretation. *Journal of the American Statistical Association 49*(267), 467–479. doi:10.2307/2281124.

Sokolov, A., F. Sabelli, B. A. Faraz, W. Li, and L. A. Seco (2023). Towards automating causal discovery in financial markets and beyond. *SSRN working paper no. 4679414*. doi:10.2139/ssrn.4679414.

Spirtes, P., C. N. Glymour, R. Scheines, and D. Heckerman (2000). *Causation, prediction, and search.* MIT press. doi:10.1007/978-1-4612-2748-9.

Stavroglou, S., A. Pantelous, K. Soramaki, and K. Zuev (2017). Causality networks of financial assets. *Journal of Network Theory in Finance 3*(2), 17–67. doi:10.21314/jntf.2017.029.

Sun, A. J. and J. A. Chan-Lau (2017). Financial networks and interconnectedness in an advanced emerging market economy. *Quantitative Finance 17*(12), 1833–1858. doi:10.1080/14697688.2017.1357976.

Sun, X., O. Schulte, G. Liu, and P. Poupart (2023). NTS-NOTEARS: Learning nonparametric DBNs with prior knowledge. In *International Conference on Artificial Intelligence and Statistics*, pp. 1942–1964. PMLR.

Swade, A., M. X. Hanauer, H. Lohre, and D. Blitz (2023). Factor zoo (.zip). *Journal of Portfolio Management 50*(3), 11–31. doi:10.3905/jpm.2023.1.561.

Tristan, L. and C. S. Ong (2021). Portfolio diversification using shape-based clustering. *Journal of Financial Data Science 3*(1), 111–126. doi:10.3905/jfds.2020.1.054.

Uhler, C., G. Raskutti, P. Bühlmann, and B. Yu (2013). Geometry of the faithfulness assumption in causal inference. *Annals of Statistics 41*(2), 436–463. doi:10.1214/12-AOS1080.

Upadhyay, S., A. Banerjee, and P. K. Panigrahi (2020). Causal evolution of global crisis in financial networks. *Physica A: Statistical Mechanics and its Applications 554*, 124690. doi:10.1016/j.physa.2020.124690.

Van de Geer, S. and P. Bühlmann (2013). $\ell_0$-penalized maximum likelihood for sparse directed acyclic graphs. *Annals of Statistics 41*(2), 536–567. doi:10.1214/13-AOS1085.

Vermorken, M., A. Szafarz, and H. Pirotte (2010). Sector classification through non-Gaussian similarity. *Applied Financial Economics 20*(11), 861–878. doi:10.1080/09603101003636238.

Vyas, K. and M. van Baren (2021). Should equity factors be betting on industries? *Journal of Portfolio Management 48*(1), 73–92. doi:10.3905/jpm.2021.1.297.

Výrost, T., Štefan Lyócsa, and E. Baumöhl (2019). Network-based asset allocation strategies. *North American Journal of Economics and Finance 47*, 516–536. doi:10.1016/j.najef.2018.06.008.

Wang, G.-J., C. Xie, and S. Chen (2017). Multiscale correlation networks analysis of the US stock market: A wavelet analysis. *Journal of Economic Interaction and Coordination 12*(3), 561–594. doi:10.1007/s11403-016-0176-x.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association 58*(301), 236–244. doi:10.1080/01621459.1963.10500845.

Watts, D. J. and S. H. Strogatz (1998). Collective dynamics of 'small-world' networks. *Nature 393*(6684), 440–442. doi:10.1038/30918.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508. doi:10.1093/rfs/hhm014.

Yang, J., Y. Tong, X. Liu, and S. Tan (2014). Causal inference from financial factors: Continuous variable based local structure learning algorithm. In *2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics*, pp. 278–285. doi:10.1109/CIFEr.2014.6924084.

Yang, S. and R. Cogill (2013). Balance sheet outlier detection using a graph similarity algorithm. In *2013 IEEE Conference on Computational Intelligence for Financial Engineering & Economics*, Volume 59, pp. 135—-142. IEEE. doi:10.1109/cifer.2013.6611709.

Zhang, B., C. Yang, H. Zhang, Z. Wang, J. Sun, L. Wang, Y. Zhao, and Y. Wang (2022). Graph representation learning for similarity stocks analysis. *Journal of Signal Processing Systems 94*, 1283–1292. doi:10.1007/s11265-022-01755-6.

Zhang, J. and P. Spirtes (2002). Strong faithfulness and uniform consistency in causal inference. In *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence*, pp. 632–639. doi:10.5555/2100584.2100661.

Zheng, X., B. Aragam, P. K. Ravikumar, and E. P. Xing (2018). DAGs with NO TEARS: Continuous optimization for structure learning. *Advances in neural information processing systems 31*, 9492–9503.

Zhou, G. (2010). How much stock return predictability can we expect from an asset pricing model? *Economics Letters 108*(2), 184–186. doi:10.1016/j.econlet.2010.05.008.

Zhu, C., R. H. Byrd, P. Lu, and J. Nocedal (1997). Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS) 23*(4), 550–560. doi:10.1145/279232.279236.

**Table 1: Summary statistics of investment universe**

This table reports summary statistics for several properties of the investment universe we study. At the end of each month in our sample we construct the cross-sectional average market capitalization ($ Billions), number of stocks, average returns (%), dispersion of returns (%), skewness of returns, and kurtosis of returns. We then calculate the summary statistics using this time-series of cross-sectional statistics. $q_y$ is the value representing the $y$-th percentile. Our sample runs from January 1993 to December 2022.

| | Market Cap. | Count | Avg. Returns | Dispersion | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| Mean | 29.96 | 357 | 1.03 | 8.11 | 0.48 | 7.55 |
| Stdev. | 19.59 | 21 | 4.55 | 2.71 | 1.60 | 18.61 |
| Min. | 5.33 | 298 | -19.6 | 4.06 | -2.77 | -0.33 |
| $q_{0.01}$ | 5.41 | 302 | -10.84 | 4.77 | -2.26 | 0.20 |
| $q_{0.25}$ | 18.88 | 345 | -1.52 | 6.43 | -0.29 | 1.87 |
| Median | 24.26 | 360 | 1.37 | 7.36 | 0.25 | 3.33 |
| $q_{0.75}$ | 38.14 | 374 | 3.60 | 8.84 | 0.86 | 6.18 |
| $q_{0.99}$ | 86.62 | 386 | 11.58 | 17.78 | 7.49 | 98.76 |
| Max. | 92.14 | 388 | 17.99 | 23.98 | 14.08 | 243.26 |

**Table 2: Performance of peer group neutralized trading strategies**

This table reports the performance statistics associated with various long–short investment strategies, each using different classification scheme (Causal, GICS, or Correlation) as well as the benchmark strategy with no-neutralization. For each of the characteristics we apply a long–short value-weighted quintile strategy within each cluster of the given peer group classification. Our sample runs from January 1993 to December 2022. We report the annualized return, annualized volatility, Sharpe ratio, the annualized CAPM alpha and beta, the maximum drawdown (DD), and the Causal Alpha (i.e., the alpha when regressing the causal strategy returns on the non-causal strategy returns.)*,** ,*** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

| Characteristic | Neutral. | Return | Volatility | SR | CAPM $\alpha$ | Beta | DD | Causal $\alpha$ |
|---|---|---|---|---|---|---|---|---|
| 1M Reversal | Causal | 7.97% | 11.97% | 0.665 | 6.87% | 0.13 | -19.8% | |
| | GICS | 6.63% | 11.01% | 0.602 | 5.19% | 0.17 | -37.0% | 2.75%*** |
| | Correlation | 6.48% | 10.54% | 0.614 | 4.95% | 0.18 | -17.3% | 3.07%** |
| | None | 5.98% | 15.46% | 0.387 | 4.17% | 0.22 | -33.7% | 4.15%** |
| 12-1M Momentum | Causal | 3.55% | 16.42% | 0.216 | 4.89% | -0.16 | -64.8% | |
| | GICS | 2.41% | 14.82% | 0.162 | 3.34% | -0.11 | -58.6% | 1.74%* |
| | Correlation | 2.65% | 14.64% | 0.181 | 3.37% | -0.09 | -58.9% | 1.77% |
| | None | 3.21% | 20.92% | 0.153 | 5.95% | -0.33 | -62.4% | 0.90% |
| Beta 60M | Causal | 4.56% | 13.06% | 0.349 | 4.57% | 0.00 | -40.9% | |
| | GICS | 2.93% | 10.38% | 0.282 | 2.93% | 0.00 | -41.3% | 2.08% |
| | Correlation | 3.08% | 11.69% | 0.263 | 3.09% | 0.00 | -37.6% | 1.84% |
| | None | 5.97% | 18.33% | 0.325 | 5.99% | 0.00 | -57.7% | 1.13% |
| Book / Price | Causal | 1.35% | 13.32% | 0.101 | 1.00% | 0.04 | -57.6% | |
| | GICS | 1.17% | 13.26% | 0.088 | 0.82% | 0.04 | -53.1% | 0.27% |
| | Correlation | -0.78% | 12.62% | -0.062 | -1.25% | 0.06 | -62.8% | 2.13%* |
| | None | 0.07% | 15.01% | 0.004 | -0.44% | 0.06 | -73.0% | 1.35% |
| Cash / Assets | Causal | 5.69% | 9.79% | 0.581 | 5.10% | 0.07 | -27.5% | |
| | GICS | 3.53% | 9.17% | 0.385 | 3.59% | -0.01 | -23.0% | 0.64% |
| | Correlation | 4.72% | 10.11% | 0.467 | 4.99% | -0.03 | -27.7% | -0.50% |
| | None | 4.98% | 13.54% | 0.367 | 1.87% | 0.37 | -64.9% | 2.01%* |
| Earnings / Price | Causal | 1.95% | 12.21% | 0.159 | 2.58% | -0.08 | -49.3% | |
| | GICS | 1.90% | 11.43% | 0.166 | 2.65% | -0.09 | -49.9% | 0.15% |
| | Correlation | 0.60% | 11.00% | 0.055 | 1.20% | -0.07 | -65.0% | 1.41% |
| | None | 0.60% | 14.91% | 0.040 | 2.26% | -0.20 | -63.2% | 0.93% |
| EBITDA / EV | Causal | 5.01% | 13.04% | 0.384 | 5.36% | -0.04 | -43.5% | |
| | GICS | 3.70% | 13.11% | 0.282 | 4.35% | -0.08 | -53.4% | 1.73%* |
| | Correlation | 2.26% | 12.39% | 0.183 | 3.18% | -0.11 | -57.7% | 2.60%** |
| | None | 4.28% | 14.54% | 0.295 | 5.02% | -0.09 | -59.6% | 1.56% |
| FY1 Earnings / Price | Causal | 3.07% | 12.35% | 0.249 | 3.98% | -0.11 | -49.0% | |
| | GICS | 0.62% | 11.83% | 0.052 | 1.14% | -0.06 | -60.7% | 2.92%*** |
| | Correlation | 0.92% | 12.03% | 0.076 | 1.56% | -0.08 | -64.1% | 2.64%** |
| | None | 2.48% | 14.44% | 0.172 | 4.47% | -0.24 | -68.6% | 0.84% |
| Gross Profitability / Assets | Causal | 2.93% | 10.72% | 0.273 | 3.87% | -0.11 | -34.4% | |
| | GICS | 2.29% | 10.46% | 0.219 | 3.39% | -0.13 | -47.9% | 1.14% |
| | Correlation | 2.89% | 10.48% | 0.276 | 3.96% | -0.13 | -36.3% | 0.75% |
| | None | 3.10% | 11.73% | 0.264 | 3.30% | -0.02 | -34.5% | 1.59% |
| Res. 12-1M Momentum | Causal | 2.53% | 9.72% | 0.260 | 3.20% | -0.08 | -37.3% | |
| | GICS | 2.10% | 9.00% | 0.234 | 2.46% | -0.04 | -20.8% | 1.14% |
| | Correlation | 2.17% | 8.89% | 0.244 | 2.67% | -0.06 | -28.1% | 1.00% |
| | None | 1.92% | 10.52% | 0.182 | 2.51% | -0.07 | -35.8% | 1.32% |
| Return on equity | Causal | 3.07% | 9.90% | 0.309 | 3.84% | -0.09 | -41.7% | |
| | GICS | 1.78% | 10.19% | 0.175 | 2.75% | -0.12 | -49.9% | 1.68%* |
| | Correlation | 1.04% | 10.11% | 0.103 | 1.79% | -0.09 | -54.4% | 2.46%** |
| | None | 2.36% | 11.95% | 0.197 | 3.91% | -0.19 | -37.4% | 1.17% |

**Table 3: Sharpe ratio difference tests**

This table reports the annualized Sharpe ratio difference for the causal-neutralized long–short quintile strategies when compared against No-neutralization, GICS-neutralized, and Correlation-neutralized strategies using the Lediot and Wolf (2008) Sharpe ratio test. We report results for value-weighted strategies. We also report results using a holding period of one-month or twelve-months following the overlapping portfolio approach of Jegadeesh and Titman (1993). Our sample runs from January 1993 to December 2022. *,** ,*** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

| | | Value-Weighted | | Equal-Weighted | |
|---|---|---|---|---|---|
| Strategy | Comparison | 1M | 12M | 1M | 12M |
| 1M Reversal | None | 0.278*** | 0.189** | 0.057 | 0.092* |
| | Correlation | 0.051 | 0.260** | 0.002 | 0.084 |
| | GICS | 0.063 | 0.222** | -0.018 | 0.021 |
| 12-1M Momentum | None | 0.063 | 0.072 | 0.054 | 0.044 |
| | Correlation | 0.036 | -0.057 | 0.043 | -0.024 |
| | GICS | 0.054 | -0.045 | 0.094* | 0.001 |
| Beta 60M | None | 0.024 | -0.002 | 0.061 | 0.020 |
| | Correlation | 0.086 | 0.098 | 0.042 | 0.054 |
| | GICS | 0.067 | 0.122 | 0.058 | -0.035 |
| Book / Price | None | 0.097 | 0.055 | 0.073 | 0.041 |
| | Correlation | 0.163* | 0.136* | 0.129* | 0.097** |
| | GICS | 0.013 | -0.007 | 0.012 | -0.044 |
| Cash / Assets | None | 0.213 | 0.156 | 0.089 | 0.197*** |
| | Correlation | 0.113 | 0.117 | -0.055 | 0.060 |
| | GICS | 0.196 | 0.162 | 0.146 | 0.168 |
| Earnings / Price | None | 0.119 | 0.050 | 0.141* | 0.040 |
| | Correlation | 0.104 | 0.143 | 0.110 | 0.051 |
| | GICS | -0.007 | 0.023 | -0.084 | -0.103 |
| EBITDA / EV | None | 0.090 | -0.035 | 0.048 | 0.012 |
| | Correlation | 0.201* | 0.152* | 0.119 | 0.108** |
| | GICS | 0.102 | -0.050 | -0.025 | -0.078 |
| FY1 Earnings / Price | None | 0.077 | -0.019 | 0.111 | -0.021 |
| | Correlation | 0.172 | 0.111 | 0.073 | 0.053 |
| | GICS | 0.196** | 0.100* | 0.059 | 0.055 |
| Gross Profitability / Assets | None | 0.009 | -0.057 | 0.139 | 0.031 |
| | Correlation | -0.003 | -0.025 | 0.218** | 0.190** |
| | GICS | 0.054 | 0.032 | 0.193* | 0.155 |
| Res. 12-1M Momentum | None | 0.078 | 0.115 | 0.045 | 0.132** |
| | Correlation | 0.016 | 0.034 | 0.002 | 0.054 |
| | GICS | 0.026 | 0.067 | 0.034 | 0.005 |
| Return on equity | None | 0.112 | 0.004 | 0.091 | 0.003 |
| | Correlation | 0.206* | -0.030 | 0.130* | 0.003 |
| | GICS | 0.134 | 0.100 | 0.129* | 0.049 |

**Table 4: Spanning regression results for the low centrality factor**
This table presents the results of time-series spanning regressions of the value-weighted top-minus-bottom quintile portfolio return of the low centrality factor over various factor models. We report the results of regressing the return of the long–short centrality factor on the returns of the Fama-French factor models (panel a) and the Hou et al. (2014) $q^5$ model (panel b). The long–short portfolios are rebalanced monthly over the period January 2003 to December 2022. Newey-West adjusted $t$-statistics are reported in parentheses. $^{*}$,$^{**}$,$^{***}$ indicate statistical significance at the 10%, 5%, and 1% level, respectively.

*Panel (a): Fama-French models*

| | **(i)** | **(ii)** | **(iii)** | **(iv)** | **(v)** | **(vi)** |
|---|---|---|---|---|---|---|
| Alpha (ann.) | 2.21% | 3.26% | 2.78% | 6.11% | 5.56% | 5.29% |
| | (1.09) | (1.81) | (1.60) | (3.43) | (3.14) | (2.94) |
| MKT-Rf | -0.04 | -0.09 | -0.07 | -0.18 | -0.15 | -0.14 |
| | (-0.69) | (-2.12) | (-1.64) | (-4.47) | (-3.69) | (-3.56) |
| SMB | | 0.16 | 0.15 | 0.07 | 0.06 | 0.05 |
| | | (1.7) | (1.77) | (0.85) | (0.79) | (0.60) |
| HML | | -0.34 | -0.33 | -0.12 | -0.09 | -0.11 |
| | | (-5.12) | (-4.42) | (-1.90) | (-1.33) | (-1.72) |
| UMD | | | 0.05 | | 0.08 | 0.05 |
| | | | (1..00) | | (1.74) | (0.94) |
| RMW | | | | -0.28 | -0.29 | -0.33 |
| | | | | (-2.93) | (-3.23) | (-3.91) |
| CMA | | | | -0.39 | -0.42 | -0.46 |
| | | | | (-3.91) | (-4.42) | (-4.29) |
| STR | | | | | | -0.05 |
| | | | | | | (-0.99) |
| LTR | | | | | | 0.03 |
| | | | | | | (0.35) |
| BAB | | | | | | 0.09 |
| | | | | | | (1.46) |
| $R^2$ | 0.0% | 18.0% | 18.3% | 24.5% | 25.5% | 25.9% |

*Panel (b): $q^5$model*

| | **(i)** | **(ii)** | **(iii)** |
|---|---|---|---|
| Alpha (ann.) | 2.21% | 4.85% | 5.02% |
| | (1.09) | (2.73) | (2.84) |
| MKT-Rf | -0.04 | -0.15 | -0.17 |
| | (-0.69) | (-3.69) | (-4.21) |
| ME | | 0.15 | 0.13 |
| | | (1.35) | (1.29) |
| IA | | -0.66 | -0.63 |
| | | (-6.23) | (-6.36) |
| EG | | 0.02 | 0.14 |
| | | (0.23) | (1.26) |
| ROE | | | -0.19 |
| | | | (-2.26) |
| $R^2$ | 0.0% | 22.8% | 24.3% |

**Table 5: Predictive regression on network density**

This table presents the estimated coefficients for the predictive regressions in Eq. (7) and Eq. (8). $R^2_{OS}$ is estimated following Campbell and Thompson (2008). The $\Delta$ statistic is the annualized CER gain for an investor who opts to use the predictive regression forecast instead of the historical average forecast. Our sample runs from January, 1993 to December, 2022. Heteroskedasticity robust $t$-values are reported in parentheses. $^{*}$,$^{**}$,$^{***}$indicate statistical significance at the 10%, 5%, and 1% level, respectively.

| | **Bivariate** | **Multivariate** |
|---|---|---|
| $\alpha$ | 0.012 | 0.013 |
| | (5.57)$^{***}$ | (5.81)$^{***}$ |
| $\beta_{Centrality}$ | -4.06 | -4.10 |
| | (-1.62) | (-1.69)$^{*}$ |
| Controls | N | Y |
| $R^2$ | 0.75% | 5.98% |
| $R^2_{OS}$ | 0.55%$^{*}$ | 3.40% $^{**}$ |
| $\Delta$ | 1.83% | 8.16% |

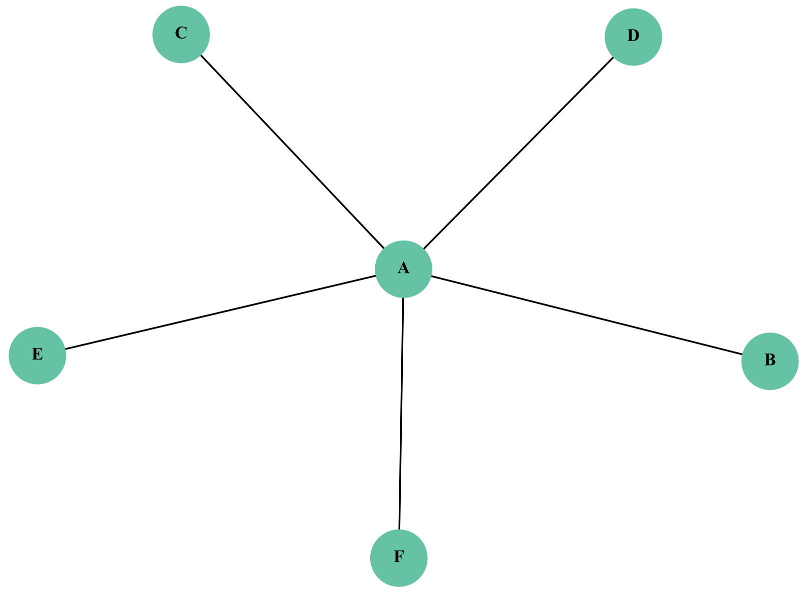


**Figure 1: Star network illustration**
Example of a star network with a central node (A) and five connected nodes (B–F).

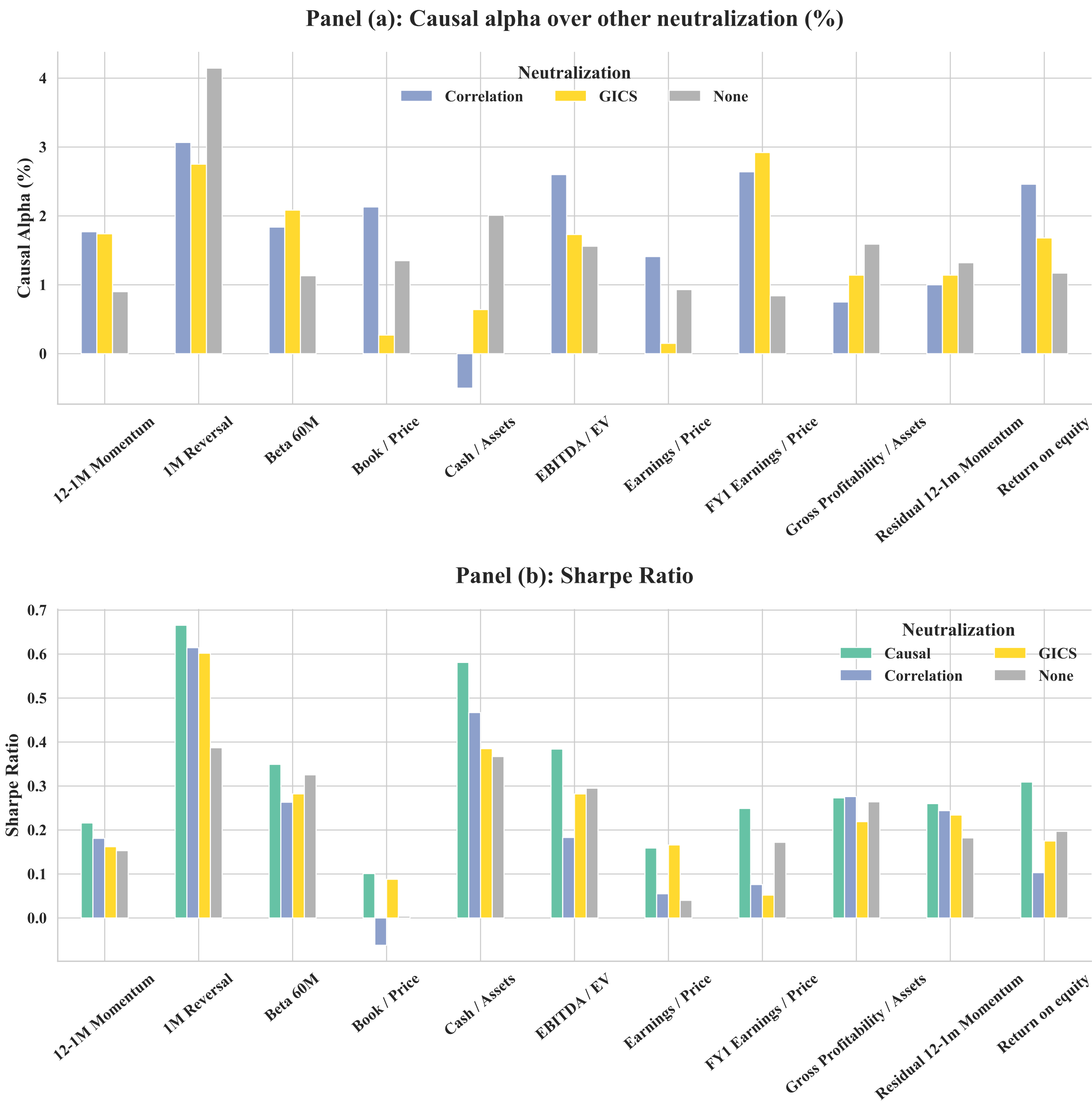


**Figure 2: Peer group neutralization alpha and Sharpe ratio comparison**
Panel (a) plots the alpha of the causality-based peer group neutralization method over the statistical correlation, GICS, and no-neutralization approaches. Panel (b) plots the Sharpe Ratio of each approach. Results are calculated on the S&P 500 universe over the sample of January 1993 to December 2022. Note that for SR, we plot the absolute Sharpe ratio for ease of comparison.

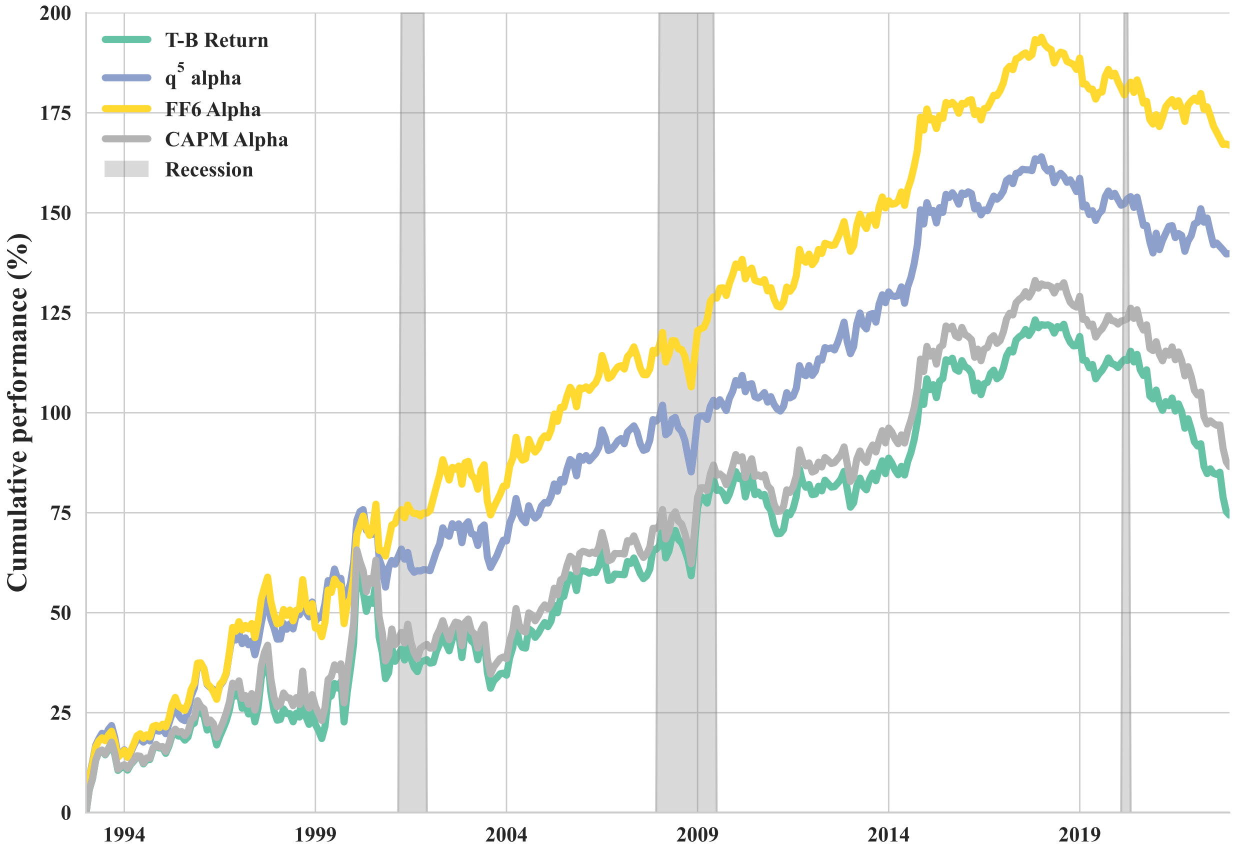


**Figure 3: Cumulative return and alpha of the low centrality factor**

This figure shows the cumulative top-minus-bottom (T-B) centrality factor return and associated cumulative alphas, in percentage points. The cumulative alpha at time $t$ is calculated as the cumulative sum of $\hat{\alpha} + \varepsilon_{:t}$, where $\varepsilon_{:t}$ are the residuals up to and including time $t$. The residuals are derived from a regression of the T-B centrality factor returns on the market return (CAPM), Fama-French six-factor model (FF6), and the Hou et al. (2014) $q^5$ model. NBER recessionary periods are highlighted with gray shading. The sample runs from January 1993 to December 2022.

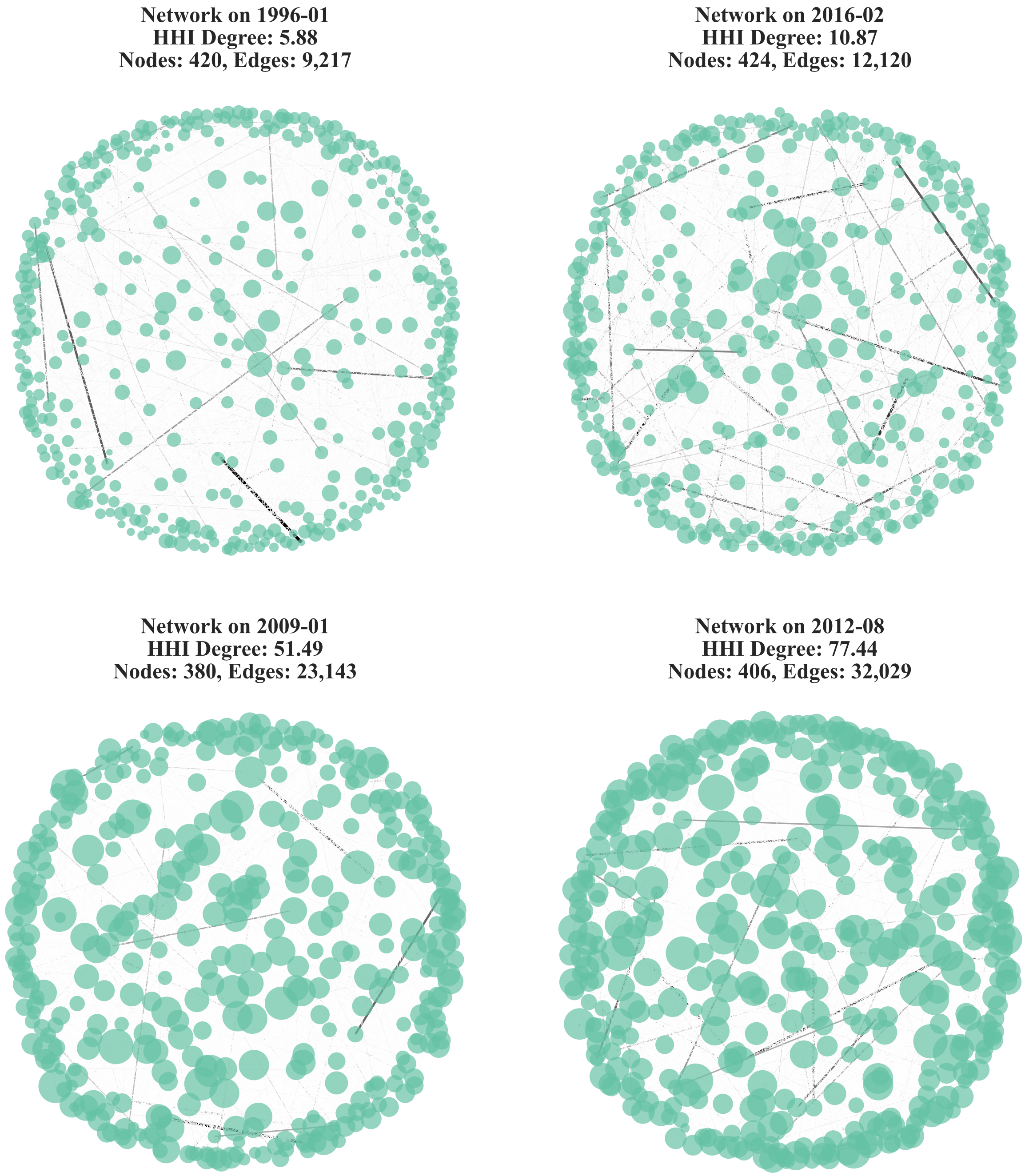


**Figure 4: Selected causal network examples**

This figure shows the causal network as at four different months. For each network we plot a spring-layout representation, where each node represents a stock, and each edge represents a connection between nodes. The size of each node corresponds to the degree centrality of each node and the edges correspond to the coefficient estimate from the causal network algorithm. For each plot we present the Herfindahl-Hirschman Index (HHI), the number of nodes, and the number of edges.

# Appendix

## A. Additional technical details for causal network estimation

### A.1. Specification of DYNOTEARS clusters

For our first investment applications, we wish to organize causal network graphs into clusters of similar stocks. When presented with such a graph, we recognize two approaches to clustering. The first approach is to directly apply some graph clustering algorithm. However, conventional graph clustering algorithms, such as Strongly Connected Components (SCC) and the Louvain method, have undesirable properties. For instance, Shirokikh et al. (2022) show that the SCC algorithm usually returns one super-cluster when applied to financial networks. This finding is in line with common asset pricing models such as the CAPM, but is undesirable when it comes to identifying sub-clusters in financial networks.

A prominent example of a graph clustering algorithm that can precisely do this, is Louvain clustering (Blondel et al., 2008).[22] The Louvain method is well-known for its performance and scalability (De Meo et al., 2011). It is a greedy technique that produces clusters by iteratively optimizing the modularity of these clusters, i.e., optimizing the density of links in-cluster relative to the density of links out-of-cluster. The Louvain method does not allow for choosing the number of clusters ex-ante, which hinders a fair comparison between clustering methodologies.

The second approach to produce clusters from a graph is what we refer to as graph representation clustering. Instead of clustering the graph using its natural form, we cluster the graph using continuous feature representations of the nodes. That is, we learn a mapping of nodes to an embedding space that preserves the network neighborhoods of the nodes. These embeddings can then be used by common clustering algorithms, such as K-means, to produce clusters. There are three major benefits to graph representation clustering. First, one can use a large variety of common clustering methods. Second, many of these clustering methods allow choosing the number of clusters ex-ante. Third, these embeddings are well-suited for

[22] Wang et al. (2017) and Hosseini et al. (2021) investigate the (in)ability of the Louvain method to detect sub-clusters in financial networks.

visualization.

To create the DYNOTEARS clusters, we make use of Node2Vec in combination with K-means. First, we apply Node2Vec to the estimated DAG $G$ to obtain embeddings for each node. To this end, we need to configure the hyperparameters of Node2Vec (Grover and Leskovec, 2016): embedding dimensionality, walk length, number of random walks, return parameter ($p$), and in-out parameter ($q$). In essence, the parameters $p$ and $q$ jointly control the random walk behavior. While $p$ dictates the probability that a just-visited node is visited once more, $q$ regulates the incentive to explore. Namely, if $q$ is sufficiently large, the random walk is biased towards nodes close to the source node.[23] Second, these embeddings are used by K-means to generate clusters. Given a predefined number of clusters $K$, K-means optimizes the following loss function:

$$L = \sum_{k=1}^{K} \sum_{i \in C_k} ||x_i - \mu_k||_2^2, \tag{A. 1}$$

where $C_k$ contains all stocks in cluster $k$, $x_i$ is the embedding belonging to stock $i$, $\mu_k$ is the mean of cluster $k$, and $||\cdot||_2$ is the Euclidean norm. The exact optimization methodology can be found in Lloyd (1982). We restrict the number of clusters $K$ to be either the past ($K = 10$) or current number ($K = 11$) of GICS sectors, to allow for a fair comparison.

### A.2. Specification of correlation benchmark

We compare the performance of peer group neutralized trading strategies based on GICS, DYNOTEARS, and statistical clustering (SC). Therefore, we require a representative clustering method for the SC strategy. As is common in this literature, we apply hierarchical clustering to the correlation matrix of the log returns to obtain statistical clusters.[24] In particular, we measure stock return correlations via the Pearson correlation coefficient.

Hierarchical clustering takes a matrix of distances as the sole input and produces a hierarchical tree (dendrogram) as output. For instance, Mantegna (1999) computes the pairwise distances as $d_{ij} = 1 - \rho_{ij}^2$, where $\rho$ is the Pearson correlation measure. Hierarchical clustering can either be agglomerative or divisive in nature. In agglomerative clustering,

[23] We refer the reader to Grover and Leskovec (2016) for a technical review of the Node2Vec algorithm.
[24] See, for example, Mantegna (1999); Bonanno et al. (2003), and Bonanno et al. (2004).

each observation is assigned its own cluster, and pairs of clusters are merged based on some linkage function. We opt for Ward's linkage function (Ward, 1963) to determine which clusters are merged at each iteration. This linkage function minimizes the total within-cluster variance at each iteration. The benefit of Ward's linkage function over other linkage functions is that it produces relatively compact clusters (Ros and Guillaume, 2019). From the estimated hierarchical tree, we can extract a number of clusters $K$ by choosing the appropriate height/distance.

Instead of performing hierarchical clustering on the raw correlation matrix, we first utilize Principal Component Analysis (PCA) to transform the stock returns into their factor exposures. In this transformation, we only use the first $K$ principal components, such that the transformed matrix has dimensions $N$ (number of assets) by $K$. Then, we can compute the pairwise distance between stock $i$ and stock $j$ as $d_{ij} = \|C_i - C_j\|_2$, where $C_i$ is the i$^{\text{th}}$ row of the PCA-transformed matrix. Equivalently, we compute the distance between stocks as the Euclidean distance between their factor exposures. Similarly to the DYNOTEARS method, we restrict $K$ to be either the past or current number of GICS sectors. Likewise, to allow for a fair comparison, we use the same sliding window methodology as DYNOTEARS.

The main benefit of this application of PCA is that it reduces the statistical uncertainty associated with the sample correlation matrix. In the context of financial markets, it is well-established that there is substantial noise present in the correlation matrix. For instance, Laloux et al. (1999) analyze the correlation matrix of the S&P 500 and find that the lowest eigenvalue-eigenvector pairs are dominated by noise. Therefore, the clear benefit of only keeping the first $K$ components is that some noise is removed. In turn, the stability and quality of the hierarchical clusters should improve as they are based on less noisy distances.

## B. Additional methodology for empirical applications

### B.1. Peer group neutralization

To investigate the efficacy of our trading strategies, we employ some well-known performance metrics. For each portfolio and its associated returns $R^P$ (with length $T$), we report the mean return, the volatility of the returns, and the annualized Sharpe Ratio:

$$SR(R^P) = \sqrt{D}\,\frac{\mu(R^P)}{\sigma(R^P)}, \tag{B. 1}$$

where $D$ is the number of trading periods in a year. We ignore the risk-free rate in the Sharpe Ratio computation, as long–short trading strategies are self-financing. To test whether the causal strategy has significant return predictability over the other classification schemes, we perform spanning regressions of the form:

$$R_t^{\text{Causal}} = \alpha + \beta R_t^{\text{c}} + \varepsilon_t, \tag{B. 2}$$

where $R_t^{\text{Causal}}$ are the portfolio returns for the DYNOTEARS classification scheme and $R_t^{\text{c}}$ are the portfolio returns based on some other classification scheme.

### B.2. Market timing

In-sample return predictability does not guarantee out-of-sample return predictability. As such, we validate the out-of-sample predictability of the network density proxy using the $R^2_{OS}$ of Campbell and Thompson (2008) which is defined as:

$$R^2_{OS} = 1 - \frac{MSFE_1}{MSFE_0}, \tag{B. 3}$$

where $MSFE_1$ is the Mean Squared Forecast Error (MSFE) based on the scrutinized timing indicator[25], whilst $MSFE_0$ is the MSFE based on the historical average forecast. To evaluate whether the $R^2_{OS}$ is significant, we utilize the $MSFE$-adjusted statistic of Clark and West (2007), which is defined as the average of:

$$f_{t+1} = (r_{t+1} - \bar{r}_{t+1})^2 - [(r_{t+1} - \hat{r}_{t+1})^2 - (\bar{r}_{t+1} - \hat{r}_{t+1})^2], \tag{B. 4}$$

where $\bar{r}_{t+1}$ is the historical average forecast, and $\hat{r}_{t+1}$ is the forecast based on our proxy. With this test statistic, we test the one-tailed hypothesis $H_0 : R^2_{OS} \leq 0$ against $H_1 : R^2_{OS} > 0$.

Return predictability alone does not speak to the economic relevance of the proxy.

[25] We estimate the required forecast errors using an expanding window methodology. To ensure a sufficiently large in-sample estimation period, we opt for a burn-in period of one-hundred months. For an example, see Hammerschmid and Lohre (2018).

Following Campbell and Thompson (2008); Ferreira and Santa-Clara (2011); Neely et al. (2014), and Hammerschmid and Lohre (2018), we assess its economic utility by computing the certainty equivalent return (CER) gain. We compute the CER for a mean-variance investor who allocates across equities and risk-free bills based on some forecast of the excess market return. Each month $t$, this investor allocates the following weight to equities (Neely et al., 2014):

$$w_t = \frac{1}{\gamma} \frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}, \tag{B. 5}$$

where $\gamma$ is the risk aversion coefficient, $\hat{r}_{t+1}$ is the forecast of the equity risk premium, and $\hat{\sigma}^2_{t+1}$ is a forecast of its variance. Moreover, a weight of $1 - w_t$ is allocated to risk-free bills. Similar to Neely et al. (2014) or Hammerschmid and Lohre (2018), the variance is computed using a moving-average window of five years, the risk aversion coefficient is set to five, and the portfolio weights $w_t$ are restricted to be between 0 and 1.5. Neely et al. (2014) argue that this restriction is reflective of realistic portfolio constraints, as it excludes short selling and taking on more than 50% leverage. Subsequently, the CER can be computed as:

$$\text{CER} = \hat{\mu}_p - \frac{1}{2}\gamma\hat{\sigma}^2_p, \tag{B. 6}$$

where $\hat{\mu}_p$ and $\hat{\sigma}^2_p$ are the mean and variance of the portfolio based on the weights in Eq. (B. 5). The CER gain can now be computed as the difference between the CER of the forecasts using the network density and the CER of the forecasts using the historical average forecast. We multiply the difference between these CERs by 1,200 such that it is an annualized percentage. This difference can be interpreted as the annual management fee the investor would be willing to pay to have access to the network density forecast instead of the historical average forecast.

### B.3. Market timing indicators

We construct ten technical indicators following Neely et al. (2014) and Hammerschmid and Lohre (2018). Specifically, we create the momentum and moving average signals. To construct the momentum signals, we use the below rule:

$$MOM_m = \begin{cases} 1 & \text{if} \quad P_t \geq P_{t-m} \\ 0 & \text{if} \quad P_t < P_{t-m} \end{cases}, \tag{B. 7}$$

where $P_t$ is the index price at time $t$, and we evaluate this signal for the values of $m = 1, 3, 6, 9, 12$ months.

To construct the moving average signal, we take the simple moving average of the S&P 500 index value as:

$$MA_{j,t}^P = \frac{1}{j} \sum_{i=0}^{j-1} P_{t-i} \quad \text{for} \quad j = s, l, \tag{B. 8}$$

where $s$ and $l$ are the lengths of the short and long moving averages. We use values of $s = 1, 2, 3$ months and $l = 9, 12$ months. We then compare these moving averages to produce six indicators as follows:

$$MA_{s,l} = \begin{cases} 1 & \text{if} \quad MA_{s,t}^P \geq MA_{l,t}^P \\ 0 & \text{if} \quad MA_{s,t}^P < MA_{l,t}^P \end{cases}. \tag{B. 9}$$