Causal Network Representations in Factor Investing*

Clint Howard[1,2], Harald Lohre[1,3] and Sebastiaan Mudde[4]

[1] *Robeco Quantitative Investments*
[2] *University of Technology Sydney*
[3] *Lancaster University Management School*
[4] *Erasmus University Rotterdam*

January 4, 2024

Abstract

The most common methods for studying the comovement of assets hinge on correlation-based measures. Given the mounting criticism of correlation-based models in the factor investing literature, we use causal discovery algorithms to revisit three salient investment applications through the lens of novel causal network representations. First, we benchmark causality-based peer groups in the context of industry neutralization of long-short investment strategies, finding that they often outperform correlation-based or industry peer groups from a Sharpe ratio perspective. However, industry peer groups remain important from a volatility reduction perspective. Second, we build a long-short equity factor guided by stocks' centrality in the causal network. We characterize peripheral stocks as potential hedges against value, whilst central stocks tend to be larger, value companies. Finally, we explore market timing of the S&P 500 using a causal network density indicator as a gauge for systematic risk and market resilience. Indeed, a shrinking network density predicts negative market returns and exhibits return predictability with an out-of-sample R^2 of 0.55%. Although causal networks thus provide novel insights in factor investing, their implementation calls for cautious pioneering as they bring about computational complexity and render interpretability more challenging.

JEL classification: C32, C38, G11, G12

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

*Corresponding author: Clint Howard, *E-mail:* c.howard@robeco.nl, *Address:* Weena 850, 3014DA, Rotterdam, South Holland, The Netherlands.
We are grateful to David Blitz, Amit Goyal, Matthias Hanauer, Iman Honarvar, 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

The seminal work of Mantegna (1999) envisioned financial markets as highly complex networks and demonstrated that representations, other than the covariance matrix, can be of economic importance. Network measures and graph invariants that had long been available to network scientists have since been applied extensively to financial networks to gain further insight into their functioning.[1] Despite the revolutionary nature of Mantegna's (1999) work, the corresponding network formulation may not fully be seen as an accurate representation of the true financial network. The foremost concern is the usage of the correlation coefficient, which is considered to be a poor measure of comovement (Li, 1990). In the context of financial markets, and particularly factor investing, the correlation coefficient does not condition on anything and is susceptible to spurious association (Simon, 1954). López de Prado (2023) argues that these associational claims have led to significant overfitting and false discoveries in the factor investing and asset pricing literatures. In this paper, we apply causal discovery algorithms to create causal network representations of the S&P 500 investment universe and leverage these representations for three investment applications. Through these applications, we demonstrate the potential value-add and current limitations of causal discovery algorithms in factor investing.

Causal discovery algorithms aim to learn causal relationships based on observational data. A critical component of these algorithms is their support for a broad range of comovement patterns. These methods look beyond simple statistical association, and instead condition on all available data, allow for lead-lag relations, and determine the directionality of comovement (Jiang et al., 2018). Although causal discovery algorithms have been popularized in fields outside finance, such as biology and neuroscience (Spirtes and Zhang, 2016), they remain underutilized in the study and practice of finance. We assess the economic value of causal discovery algorithms and their associated network representation via three investment applications. First, we choose a representative causal discovery algorithm, the DYNOTEARS approach of (Pamfil et al., 2020). DYNOTEARS is a score-based causal algorithm which can

[1] See for example Bonanno et al. (2003, 2004); Pozzi et al. (2007); Matteo et al. (2010); Buraschi and Porchia (2012); Ahern (2013); Gao et al. (2013); Haldane (2013); Pozzi et al. (2013); Papenbrock and Schwendner (2015); Bianchi et al. (2015); Kaya (2015); Peralta and Zareei (2016).

estimate both contemporaneous and time-lagged relationships between given observations (e.g., stock returns), and hence may provide richer insights into the comovement patterns of stocks.

We apply these resulting networks to three investment applications using data on the S&P 500 constituents from 1990 to 2023. First, we utilize the estimated causal financial network to create causality-based peer group clusters for the purpose of industry-neutralization in the context of long-short equity strategies. Industry-neutralized strategies avoid the industry bias present in some firm characteristics by diversifying over these industries. To evaluate the industry-neutralization ability of the causality-based clusters, we construct and assess the performance of industry-neutralized long-short portfolios for a variety of stock characteristics. We compare the industry-neutralization ability of the DYNOTEARS classification scheme against a standard classification scheme, the Global Industry Classification Standard (GICS), and a statistical correlation-based clustering approach. We find that at both one-month and twelve-month holding periods, the causality-based peer groups generally outperform their comparison methods from a Sharpe ratio perspective. However, from a volatility reduction perspective, the incumbent GICS is difficult to beat.

In the second investment application, we investigate the role of centrality in financial networks. To this end, we construct long-short portfolios based on the centrality scores of stocks, where we buy peripheral stocks and sell central stocks. To characterize centrality in terms of common equity factors, we use spanning regressions based on Fama and French (1993) and Hou et al. (2021). Standalone, the long-short centrality factor yields mixed results and is highly dependent on the business cycle. However, judging from spanning regression results, this centrality factor emerges as a potential hedge against value factors, delivering significant alpha (4% to 6% annualized) on top of extant factor models. However, the relationship between centrality and value also appears to be time-varying, thus giving pause to its reliability as a constant value hedge. Lastly, we propose a novel market timing indicator, based on the density of the network topology. We evaluate its in-sample and out-of-sample return predictability, as well as its economic relevance. We find an in-sample R^2 of 0.75% and an out-of-sample R^2 of 0.55%, and this predictability of the centrality timing indicator is not explained by well-known macroeconomic or technical indicators (Welch and

Goyal, 2008; Campbell and Thompson, 2008; Neely et al., 2014; Hammerschmid and Lohre, 2018).

We contribute to several streams of literature. The foundation of our research is most closely related to the literature on financial networks, particularly the application of novel techniques to build alternate financial network representations. Pamfil et al. (2020), who propose the DYNOTEARS causal discovery algorithm, apply the technique to returns of S&P 100 companies. They find that whilst the resulting representation is like a sector-based approach, there are important differences, such as the connection between Amazon (a Consumer Discretionary company) and many technology stocks (such as Facebook, Google, and Microsoft). We extend upon this work by using a richer S&P 500 dataset and extending the application of the causal network representations to real investment applications. Proposing a causal discovery-based representation of a financial network for clustering of alternate peer groups, we further contribute to the literature which subsequently applies clustering algorithms to financial network representations (López de Prado, 2016; Konstantinov et al., 2020; Tristan and Ong, 2021; French, 2023).

By applying causal discovery algorithms to stock returns, we illustrate the novelty of these causal algorithm approaches over extant correlation-based approaches. In our industry-neutralization investment application, we first use causal discovery algorithms to produce alternate peer groups of stocks. We are among the first to compare the industry-neutralization ability of this novel industry classification scheme against the industry-neutralization ability of standard classification schemes such as GICS and correlation-based approaches. We further demonstrate the insights that a causality-based representation of financial networks offers by extracting centrality-based information from this representation. As a result, we are able to characterize the factor-based characteristics of central and peripheral stocks in the S&P 500, as well as propose a market timing indicator which is orthogonal to existing indicators.

The remainder of this paper is structured as follows. Section 2 lays the foundation of causal discovery algorithms in the context of financial markets. Section 3 presents each investment application, with a brief introduction, methodology, and empirical results. Section 4 then concludes.

2 Causal network estimation

The existing literature on causal algorithms can be ambiguous in its terminology, as it encompasses both causal discovery and causal inference. The former focuses on estimating a causal network from observational data, whilst the latter focuses on estimating the causal effect of an intervention via do-calculus (Pearl, 2010). As our goal is to model a causal network of stocks rather than the causal effect of interventions, we restrict our presentation of causal algorithms to causal discovery algorithms.

2.1 Financial peer groups

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: Industry Classification Schemes (ICS) and Algorithmic Classification Schemes (ACS). Notable examples of ICS include the Standard Industrial Classification (SIC), the North American Industry Classification System (NAICS), or the Fama-French industries (Fama and French, 1997). Arguably the most popular classification scheme for financial practitioners is the Global Industry Classification Standard (GICS). Developed in 1999 by S&P and MSCI, GICS was targeted towards the investment community and has since then been adopted by many financial institutions (Phillips and Ormsby, 2016). The widespread adoption of GICS can be explained by its desirable properties. Bhojraj et al. (2003) show that GICS outperforms SIC, NAICS and the Fama-French industries in explaining cross-sectional variations in out-of-sample returns and firm-level characteristics. Furthermore, the classification is reviewed annually and feedback from stakeholders is taken into account during this review, ensuring that GICS accurately reflects the ever-changing market structure (Phillips and Ormsby, 2016).

Yet, there are also limitations to the GICS scheme. First, GICS is fairly static in nature (Costa and De Angelis, 2011), as classification changes often take more than a year to materialize. Hence, investors relying on GICS are potentially acting on outdated information.

Figure 1 presents an example thereof this lag. The top figure shows the S&P 500 constituents and their associated GICS membership (by color) as of November 4th, 2015. The spatial clustering of the figure is a result of applying a causality-based ACS (the details not being important here). Generally, we observe that the causal clustering agrees with the GICS classifications, except for the highlighted Financial sector (in green), where we observe two distinct clusters. If we then observe the bottom figure, which is the same information as of October 3, 2016, we see that one of the separate Financial clusters was reclassified as Real Estate. Thus, the causal based ACS identified, at least a year in advance of GICS, that the Real Estate sector was distinct from the Financial sector, at least from a returns perspective.

<Insert Figure 1 about here>

A second limitation is that GICS is fixated on the end-product to determine company similarity (Phillips and Ormsby, 2016), and thus neglects similarity stemming from the supply chain. In contrast, ACS aim to quickly respond to market changes and do not (solely) take the end-product of a company into account for computing company similarity. Many of these ACS are inspired by the seminal work of Mantegna (1999) and as such cluster on the basis of return correlation.[2] However, the efficacy of these correlation-based ACS has been questioned by Chan et al. (2007), who find that statistical clustering methods are outperformed by GICS in their ability to capture out-of-sample return covariation. Vermorken et al. (2010) argue that the non-Gaussian nature of stock returns might cause correlation-based ACS to be less accurate.[3] Unfortunately, their argument does not give any insight into how this issue distorts the estimation of clusters. In the absence of such a mechanism, it becomes difficult to hypothesize about the improvements other ACS could offer.

We posit three mechanisms that detail how the undesirable properties of the correlation measure distort the estimation of clusters. First, the symmetry of the correlation measure causes suppliers and their clients to be clustered together. As an example, food producers in the Consumer Staples sector have spillover effects on restaurants, fast-food chains, and hotels in the Consumer Discretionary sector. The substantial correlation resulting from these

[2]For some examples, see Onnela et al. (2003) and Musmeci et al. (2015).

[3]Markowitz and Usmen (1996a, 1996b) provide some evidence that daily log-returns in the S&P 500 follow a Student's t-distribution rather than a normal distribution.

spillover effects will likely cause these companies to be clustered together in correlation-based ACS. In asymmetric approaches, food producers are clearly distinct from their clients because there is only one-way traffic. Second, the unconditioned nature of the correlation measure causes clients of the same supplier to be clustered together. For instance, producers of microchip (machines) have clients in different sectors such as Health Care, Information Technology, and Communication Services. The pairwise correlations between members of these sectors will increase when shocks occur to these producers because they all suffer from spillover effects. This increase in correlation depends to an extent on the frequency and magnitude of these shocks. Therefore, members of these sectors could be clustered together due to the spurious correlation (Simon, 1954) originating from shared suppliers. In causal discovery algorithms, this issue is non-existent because spurious association does not affect the similarity measure. Based on these two mechanisms, it is clear that correlation-based ACS will sometimes group stocks together that should not be grouped together. We will refer to these two mechanisms and the associated conclusion as the "apples and oranges" narrative.

2.2 Causal network framework

The theoretical framework underlying causal algorithms is quite extensive. It contains elements and assumptions that are often not utilized by modern methods. As such, we limit the exposition to a concise explanation of the most-used elements. The most fundamental notion within this framework is that of *Bayesian Networks* (BNs). Pearl (1985) details that a given *Directed Acyclic Graph* (DAG), over a set of p random variables X, constitutes a Bayesian Network if:

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

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.[4]

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 λ-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.[5] 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 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

[4]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.

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

variables from different periods are called inter-slice edges, whilst edges connecting variables in the same period are called intra-slice edges.

2.3 Causal discovery algorithms

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.

Several papers have extended the 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:

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

where 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, necessitating all diagonal entries of $\mathbf{W}$ to be zero. The $\mathbf{A}_i$ are inherently acyclic as they point forward in time and thus cannot create cycles.

2.4 Estimation method

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.

Before detailing our implementation of this method, we first provide a rigorous model specification. Two important components of this model specification are the identification and the regularization scheme. Typically, the identifiability of SVAR models is not guaranteed. However, Pamfil et al. (2020) argue that the parameters of this model are identified in two special cases. Both of these cases rely to some extent on the acyclicity of the contemporaneous effects $\mathbf{W}$. First, when the errors $\mathbf{Z}$ are standard normal variables, the identifiability follows from Peters and Bühlmann (2014).[6] They prove that linear Gaussian structural equation models, which correspond to a DAG, are fully identified if the errors have equal variance. Second, when the errors $\mathbf{Z}$ are independent non-Gaussian variables, the identifiability follows

[6]For the origin of the parameters $\mathbf{W}$ and $\mathbf{Z}$, we refer the reader to Eq. (2).

from Hyvärinen et al. (2010). They show that non-Gaussian models generally cannot produce a unique ordering of the components. However, by assuming the acyclicity of $\mathbf{W}$, they show that a unique ordering can be found. For the identifiability of the DYNOTEARS model, we assume that one of these cases applies. As to the regularization scheme, Pamfil et al. (2020) apply L_1 (LASSO) regularization to both the intra-slice and the inter-slice weights, ensuring that the estimated DBN is sufficiently sparse.

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 $\hat{w}_{ij}$.[7] This choice follows from the weak form of the Efficient Market Hypothesis (Fama, 1970), which posits that the market is efficient to all information contained in historical prices. 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:

$$\min_{W} \frac{1}{2n} \|\mathbf{X} - \mathbf{X}\mathbf{W}\|_F + \lambda_{\mathbf{W}} \|\mathbf{W}\|_1 \quad \text{s.t. } \mathbf{W} \text{ is acyclic,} \tag{3}$$

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.

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

3 Causal network applications in portfolio management

In this section, we introduce and discuss the results of the causal network investment applications. First, we compare the performance of several classification schemes for the purpose of industry-neutralization when constructing long-short investment strategies. Subsequently, we evaluate the economic value of stock centrality by creating a 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.

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 any models. We follow our previously described methodology to estimate the DYNOTEARS causal network over the S&P 500 constituents, and we then subsequently use the Node2Vec algorithm to translate this causal network representation into clusters of stocks. Further detail is given in the Online Appendix.

3.1 Industry neutralization

Our first investment application of financial networks is to provide peer groups for industry-neutralization. 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 analytically and empirically that long-short portfolios generally benefit from sector-neutralization (i.e., removing the across-industry component).[8] 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 find 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

[8]Note that we use the terms industry-neutralization and sector-neutralization interchangeably.

(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 application, we investigate the potential of ACS in the context of industry-neutralization, with a specific focus on causality-based ACS.

3.1.1 Methodology

To study the efficacy of different peer groups in the context of industry neutralization of long-short portfolios, we take twelve well-known 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.[9] For each characteristic and each different classification scheme, we form long-short portfolios within each peer group/industry. At a given time t and for a given classification scheme, we generate a ranking of stocks, in each cluster, based on a characteristic, and go long in the top 20% and go short in the bottom 20%.[10] That is, within each cluster, we apply a quintile long-short strategy. Furthermore, we employ monthly rebalancing to these long-short portfolios. We report results for both equal-weighted (EW) and value-weighted (VW) portfolios with a one-month holding period and twelve-month holding period (following the overlapping portfolio approach of Jegadeesh and Titman (1993)). As a benchmark, we include a trading strategy that does not use industry-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.

We compare the performance of industry-neutralized trading strategies based on GICS,

[9]For 1M reversal and beta 60M, 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.

[10]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.

DYNOTEARS, and statistical clustering (SC). DYNOTEARS clusters are provided by the method described in Section 3.1, and thus we require a representative clustering method for the SC strategy. We apply hierarchical clustering to the correlation matrix of the log returns to obtain $K = 10$ or 11 statistical clusters.[11] In particular, we measure the stock's correlation using the Pearson correlation coefficient. Instead of performing hierarchical clustering on the raw correlation matrix, we utilize Principal Component Analysis (PCA) to transform the stock returns into their factor exposures (Avellaneda and Serur, 2020). In this transformation, we use the first $K = 10$ or 11 principal components, such that the transformed matrix has dimensions N (number of assets) by K. Then, we compute the pairwise distance between stock i and stock j as $d_{ij} = \|C_i - C_j\|_2$, where C_i is the i^{th} row of the PCA-transformed matrix. Equivalently, we compute the distance between stocks as the Euclidean distance between their factor exposures. Similar to the DYNOTEARS method, we restrict K to be either the past or current number of GICS sectors. We use the same sliding window methodology as DYNOTEARS.

3.1.2 Results

In Table 1 we present the Sharpe ratio rank (one being highest, four being lowest) for the DYNOTEARS long-short strategies, relative to the GICS and SC neutralized strategies and the benchmark strategy with no industry-neutralization. We present the Sharpe ratio rank for both EW and VW portfolios, as well as for differing holding periods. From a risk-adjusted return perspective, we generally find that the DYNOTEARS strategies are outperforming the other strategies across all construction methods and characteristics. In forty-eight possible comparisons, DYNOTEARS is ranked first thirty times, second fifteen times, and third three times. Importantly, we do not find any strong biases for specific types of characteristics. We find similar performance for value characteristics (such as book-to-price) as well as quality characteristics (such as Cash-to-assets and ROE).

<Insert Table 1 about here>

In Table 2 and Table 3, we present the annualized results for value-weighted trading

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

strategies based on different classification schemes. We also report the causal alpha, which is an annualized α 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^{Causal} is the long-short return of the causal neutralization and R_t^{c} is the long-short return of GICS/correlation/no-neutralization based strategy. Table 2 presents results for one-month holding periods and Table 3 for twelve-month holding periods. 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 industry-neutralization. Figure 3 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%. We identify two mechanisms that explain the poor performance of the SC variant. The first 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.[12] The 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.

<Insert Figure 3 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 the GICS variant. This outperformance mostly seems to derive from increased stock-selection power, considering that the mean returns of the causal variant are consistently higher. This aligns well with the illustration in Figure 1, namely that due to a lack of timeliness, and an excessive focus on the end-product, the diversification power of GICS is potentially lower than that of causal algorithms. Thus, the causal variant has a significant 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 industry-neutralization. However, given the volatility reduction power of GICS, this suggests there is still an important place for GICS when applying industry-neutralization.

<Insert Table 3 about here>

3.2 Centrality as a cross-sectional long-short equity factor

Our second investment application 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 2, is the purest example of a central node. They understand central nodes as the nodes that have more ties,

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

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 2 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).[13]

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.

3.2.1 Methodology

To investigate the role of centrality in financial networks, we construct a monthly centrality factor using our causal network. For this centrality factor, we monthly sort S&P 500 stocks into value-weighted quintile long-short portfolios based on the inverse of their eigenvector

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

centrality score.[14] Thus, peripheral stocks will belong to the top portfolio, whilst central stocks will belong to the bottom portfolio.

3.2.2 Results

Figure 4 displays the cumulative return of the centrality factor over time. The performance of the 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 evaporates during bear markets. Before characterizing the centrality factor in terms of existing factors, we first establish that the centrality factor adds value beyond these factors by investigating the alphas of the spanning regressions in Table 4 relative to several common factor models. Specifically, we take the market return in excess of the risk free rate (MKT-Rf), and the Fama-French 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. We take the Betting-Against-Beta (BAB) from the AQR data library. We take 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.

<Insert Figure 4 about here>

The 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 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 centrality factor are lowered. Notably, the q^5 alpha is higher than the FF4 alpha, in part because the centrality factor loads more negatively on the investment

[14] We find quantitatively similar results for equal-weighted portfolios.

factor than on the value factor.

<Insert Table 4 about here>

Given that its alphas are significant and sizable, the 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 4. We define the cumulative alpha of the 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 centrality factor seems roughly related to its performance (as visualized in Figure 4). 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 centrality factor can effectively act as a cheap hedge since it does not underperform, but it controls for risk in other factors which are known to perform well. However, the performance of the centrality factor in the 2020 to 2023 period highlights that this is not a free hedge, as it significantly underperforms.

Next, we characterize the centrality factor in terms of existing factors. Indeed, based on the left partition of Table 4, the conjecture that the centrality factor resembles a growth strategy rings true. Namely, the 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 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 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.

3.3 Market timing

For our last investment application, we investigate the market timing ability of the causal network density. The rationale is that network density is reflective of the systematic risk of a financial network and thus might carry some return predictability. Since the seminal work of Cowles (1933), many scholars have investigated return predictability of the U.S. stock market. Kaya (2015) investigates the role of aggregate network eccentricity.[15] They interpret the average eccentricity (aggregated over the nodes) of the network as a proxy for the density of the topology. In their interpretation, a lower average eccentricity corresponds to a denser topology and vice versa. They observe that the topology becomes denser prior to market crashes and that returns, after a period of increasing density, are more left-skewed and leptokurtic than returns after a period of decreasing density. This finding is in line with

[15] The eccentricity of a node v is the maximum distance to any other node in the network. Therefore, eccentricity is the inverse of a centrality measure, such that lower eccentricities correspond to more central nodes.

Lenzu and Tedeschi (2012) who explore how a denser topology is rather problematic from a systematic risk perspective.

There is good reason to believe that the density of a financial network can be understood as a measure of its resiliency against crises. To test this hypothesis, we perform an equity market-timing case, using a causal network-based proxy for financial network density. Naively, this proxy could be defined as the average centrality in the network. However, we expect this proxy to have little market timing ability, as structural breaks in the network topology likely cause the absolute level of centrality to be uninformative for predictive purposes. Therefore, we use the relative level (change) of centrality, rather than the absolute level. Concretely, this proxy is defined as:

$$p_t = \frac{1}{|U_t|} \sum_{i \in U_t} c_{i,t} - \frac{1}{|U_{t-1}|} \sum_{j \in U_{t-1}} c_{j,t-1}, \tag{4}$$

where $c_{i,t}$ is the eigenvector centrality measure. This construction is like the *systematic risk index* of Kaya (2015) who also computes the density as the average network centrality but uses the eccentricity measure instead of the eigenvector measure.

The work most closely related to our market timing efforts is Eng-Uthaiwat (2018). They investigate how much return predictability the network topology offers in the S&P 100. However, two critical aspects of their methodology diverge from our methodology. First, their network formulation follows the correlation-based PMFG methodology of Tumminello et al. (2005). While our network formulation is based on a causal discovery algorithm. Second, they choose the graph diameter as a measure of network topology, whilst we choose the average eigenvector centrality. The graph diameter is defined as the maximum eccentricity, or equivalently, the length of the shortest path between the most distant nodes. The choice for the graph diameter as a measure of network topology poses a subtle issue in the context of financial networks. This issue stems from the small-world property of financial networks.[16] Small-world networks, as introduced by Watts and Strogatz (1998), are characterized by a low degree of separation and significant local clustering. This means that the diameter of such networks will consistently be low and therefore be of little explanatory value. For

[16] The small-world property of financial networks has ostensibly been confirmed by Haldane (2013), Gao et al. (2013), and Sun and Chan-Lau (2017).

instance, the diameter of the financial network in Sun and Chan-Lau (2017) is only two. Therefore, we opt for the average eigenvector centrality. Ruhnau (2000) explains that the main benefit of the eigenvector centrality approach is that the importance of an edge is no longer static. Instead, out-going edges from influential nodes are deemed more important than out-going edges from less influential nodes. Moreover, Dablander and Hinne (2019) show that for causal networks, eigenvector centrality is a better choice compared to other node centrality measures.

3.3.1 Methodology

In testing the timing efficacy of the network density indicator, we follow existing market timing literature (Welch and Goyal, 2008; Neely et al., 2014; Hammerschmid and Lohre, 2018). Specifically, we measure in-sample return predictability using bivariate predictive regressions of the following form:

$$r_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1}, \tag{5}$$

where r_t is the excess market (S&P 500) return at time t, and x_t is the value of some indicator at time t. The predictors used to explain stock returns in the literature can roughly be divided into two classes: macroeconomic variables and technical indicators. To give an example of the former, the dividend-price ratio is a macroeconomic variable used by Fama and French (1988) and Cochrane (2008). An example of the latter can be found in Neely et al. (2014) who use fourteen technical indicators based on moving averages, momentum, and volume. 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}. Like Neely et al. (2014), we calculate the certainty equivalent return (CER) gain for the timing indicator to get a sense of the indicator's economic relevance. Specifically, the CER 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. We provide additional details in Online Appendix B.

In addition to using our network centrality measure in the bivariate regression of Eq. (5), we also conduct a multivariate regression where we include common macroeconomic variables and technical indicators as controls. Specifically, we include eleven macroeconomic indicators using the Welch and Goyal (2008) dataset, and ten technical indicators following Hammerschmid and Lohre (2018).[17] To reduce this set of twenty-one control predictors, we use the PCA approach of Neely et al. (2014) to arrive at three principal components. Specifically, we run PCA over all twenty-one control predictors to generate three orthogonal factors. With these factors, we can use the following multivariate regression:

$$r_{(t+1)} = \alpha + \beta_C x_{C,t} + \sum_{k=1}^{K} \beta_k x_{(k,t)} + \epsilon_{(t+1)}, \tag{6}$$

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

3.3.2 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. 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.

<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

[17] The full list of controls can be found in Online Appendix B.

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% 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 (Δ) 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 Δ, 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.

4 Conclusion

In this paper, we set out to advance financial network modeling by introducing causal discovery algorithms to the factor investing literature. To motivate this effort, we highlight the potential improvements that causal discovery algorithms offer over the prevalent correlation-based framework of Mantegna (1999). We structured this effort by selecting a representative causal discovery algorithm (DYNOTEARS) and applying it to three common investment applications.

We found that DYNOTEARS offers orthogonal industry-neutralization strategies compared to GICS and statistical clusters, typically resulting in higher risk-adjusted returns compared to competing methods. However, the incumbent GICS still remains relevant from a volatility reduction perspective. We further demonstrate that a causal network-based stock centrality factor acts as a hedging factor in combination with the common Fama-French and q^5 factor models. Finally, we find that a causal network-based market timing indicator successfully predicts forward S&P 500 excess returns, often on par with existing indicators from the literature.

Ultimately, these findings lead us to conclude that causal discovery algorithms can be of substantial economic value and thus represent a valuable tool for practitioners. Such algorithms are able to provide novel and unique insights, which go above and beyond what is possible with standard correlation-based measures. Despite these opportunities, causal discovery algorithms remain limited by their computational complexity and interpretability. Thus, we advocate for the continued exploration of causal discovery techniques in the factor investing literature.

References

Ahern, K. R. (2013). Network centrality and the cross section of stock returns. *SSRN working paper no. 2197370*.

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

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

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.

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.

Bianchi, D., M. Billio, R. Casarin, and M. Guidolin (2015). Modeling contagion and systemic risk. *Systemic Risk Tomography working paper series no. 10*.

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

Blitz, D., J. Huij, and M. Martens (2011). Residual momentum. *Journal of Empirical Finance 18*(3), 506–521.

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

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.

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.

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.

Buraschi, A. and P. Porchia (2012). Dynamic networks and asset pricing. *SSRN working paper no. 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.

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.

Chan, L. K., J. Lakonishok, and B. Swaminathan (2007). Industry classifications and return comovement. *Financial Analysts Journal 63*(6), 56–70.

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.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies 21*(4), 1533–1575.

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*.

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.

Cowles, A. (1933). Can stock market forecasters forecast? *Econometrica 1*(3), 309–324.

Dablander, F. and M. Hinne (2019). Node centrality measures are a poor substitute for causal inference. *Scientific Reports 9*(1), 6846.

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.

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

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

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance 25*(2), 383–417.

Fama, E. F. and K. R. French (1988). Dividend yields and expected stock returns. *Journal of Financial Economics 22*(1), 3–25.

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.

Fama, E. F. and K. R. French (1997). Industry costs of equity. *Journal of Financial Economics 43*(2), 153–193.

Fama, E. F. and K. R. French (2007). The anatomy of value and growth stock returns. *Financial Analysts Journal 63*(6), 44–54.

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

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.

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.

Freeman, L. C. et al. (2002). Centrality in social networks: Conceptual clarification. *Social Networks 1*(3), 238–263.

French, R. (2023). Peer group identification in factor portfolios: A data-driven approach. *Journal of Portfolio Management Forthcoming*.

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.

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.

Haldane, A. G. (2013). *Rethinking the financial network*. Springer.

Hammerschmid, R. and H. Lohre (2018). Regime shifts and stock return predictability. *International Review of Economics & Finance 56*, 138–160.

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

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.

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.

Hyvärinen, A., K. Zhang, S. Shimizu, and P. O. Hoyer (2010). Estimation of a structural vector autoregression model using non-gaussianity. *Journal of Machine Learning Research 11*(56), 1709–1731.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance 48*(1), 65–91.

Jiang, L., K. Wu, and G. Zhou (2018). Asymmetry in stock comovements: An entropy approach. *Journal of Financial and Quantitative Analysis 53*(4), 1479–1507.

Kaya, H. (2015). Eccentricity in asset management. *Network Theory in Finance 1*(1), 1–32.

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.

Korte, B. H., J. Vygen, B. Korte, and J. Vygen (2011). *Combinatorial optimization*, Volume 1. Springer.

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

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

Li, W. (1990). Mutual information functions versus correlation functions. *Journal of Statistical Physics 60*, 823–837.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory 28*(2), 129–137.

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

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

MacMahon, M. and D. Garlaschelli (2015). Community detection for correlation matrices. *Physical Review. X 5*(2).

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

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.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance 7*(1), 77–91.

Markowitz, H. M. and N. Usmen (1996a). The likelihood of various stock market return distributions, Part 1: Principles of inference. *Journal of Risk and Uncertainty 13*, 207–219.

Markowitz, H. M. and N. Usmen (1996b). The likelihood of various stock market return distributions, Part 2: Empirical results. *Journal of Risk and Uncertainty 13*, 221–247.

Matteo, T. D., F. Pozzi, and T. Aste (2010). The use of dynamical networks to detect the hierarchical organization of financial market sectors. *The European Physical Journal B: Condensed Matter and Complex Systems 73*(1), 3–11.

Musmeci, N., T. Aste, and T. Di Matteo (2015). Relation between financial market structure and the real economy: Comparison between clustering methods. *PloS one 10*(3), e0116201.

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.

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

Onnela, J.-P., A. Chakraborti, K. Kaski, J. Kertesz, and A. Kanto (2003). Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E 68*(5), 056110.

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.

Papenbrock, J. and P. Schwendner (2015). Handling risk-on/risk-off dynamics with correlation regimes and correlation networks. *Financial Markets and Portfolio Management 29*, 125–147.

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.

Peters, J. and P. Bühlmann (2014). Identifiability of gaussian structural equation models with equal error variances. *Biometrika 101*(1), 219–228.

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

Pozzi, F., T. Aste, G. Rotundo, and T. Di Matteo (2007). Dynamical correlations in financial systems. *Proceedings of SPIE 6802*.

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.

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.

Ruhnau, B. (2000). Eigenvector-centrality—a node-centrality? *Social Networks 22*(4), 357–365.

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.

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

Spirtes, P., C. N. Glymour, R. Scheines, and D. Heckerman (2000). *Causation, prediction, and search.* MIT press.

Spirtes, P. and K. Zhang (2016). Causal discovery and inference: Concepts and recent methodological advances. In *Applied Informatics*, Volume 3, pp. 1–28. SpringerOpen.

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.

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.

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

Tumminello, M., T. Aste, T. Di Matteo, and R. Mantegna (2005). A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences of the United States of America 102*(30), 10421.

Uhler, C., G. Raskutti, P. Bühlmann, and B. Yu (2013). Geometry of the faithfulness assumption in causal inference. *Annals of Statistics*, 436–463.

Van de Geer, S. and P. Bühlmann (2013). ℓ_0-penalized maximum likelihood for sparse directed acyclic graphs. *Annals of Statistics 41*, 536–567.

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

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

Vỳrost, T., Š. Lyócsa, and E. Baumöhl (2019). Network-based asset allocation strategies. *North American Journal of Economics and Finance 47*, 516–536.

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.

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

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

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.

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.

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.

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.

Table 1: Performance ranking for causal-neutralized long-short strategies
This table reports the rank (out of four) of the Sharpe ratio for the causal-neutralized long-short quintile strategies when compared against No-neutralization, GICS-neutralized, and Correlation-neutralized strategies. A rank of one corresponds to the highest Sharpe ratio, and a rank of four to the lowest. We report results for equal-weighted and 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.

Weighting Method	**Equal-weighted**		**Value-weighted**	
Holding Period	**1M**	**12M**	**1M**	**12M**
1M Reversal	2	1	1	1
12-1M Momentum	1	2	1	3
Beta 60M	1	2	1	2
Book-to-price	1	2	1	2
Cash-to-Assets	2	1	1	1
Earnings-to-price	2	2	2	1
EBITDA-to-EV	2	2	1	3
FY1 Earnings-to-price	1	2	1	2
Gross Profitability / Assets	1	1	2	3
Residual 12-1M Momentum	1	1	1	1
ROE	1	1	1	2

Table 2: Backtest results for industry-neutralized factors - 1M holding period
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 α	Beta	DD	Causal α
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%
Residual 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: Backtest results for industry-neutralized factors - 12M holding period

This table reports the performance statistics associated with various long-short investment strategies, each using different classification schemes (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. We calculate returns over a 12M holding period using the overlapping portfolio approach of Jegadeesh and Titman (1993). 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 annualized 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 α	Beta	DD	Causal α
1M Reversal	Causal	0.19%	4.33%	0.044	-0.16%	0.04	-15.4%	
	GICS	-0.68%	4.10%	-0.165	-0.96%	0.03	-25.7%	-0.36%
	Correlation	-0.84%	4.12%	-0.204	-1.16%	0.04	-24.2%	-0.52%
	None	-0.72%	5.43%	-0.133	-1.31%	0.07	-24.7%	-0.43%
12-1M Momentum	Causal	2.02%	12.04%	0.168	2.14%	-0.01	-52.0%	
	GICS	2.27%	10.66%	0.213	2.20%	0.01	-41.9%	-0.79%
	Correlation	2.48%	11.01%	0.225	2.59%	-0.01	-41.9%	-0.47%
	None	1.46%	15.16%	0.096	1.98%	-0.06	-64.1%	0.36%
Beta 60M	Causal	3.72%	10.90%	0.342	3.74%	0.00	-34.3%	
	GICS	1.96%	8.89%	0.220	1.96%	0.00	-37.7%	1.60%*
	Correlation	2.40%	9.85%	0.244	2.41%	0.00	-34.3%	1.78%*
	None	6.01%	17.52%	0.343	6.03%	0.00	-57.9%	0.33%
Book / Price	Causal	0.41%	11.49%	0.036	0.63%	-0.03	-52.7%	
	GICS	0.50%	11.65%	0.043	0.73%	-0.03	-52.9%	1.31%*
	Correlation	-1.14%	11.39%	-0.100	-0.98%	-0.02	-61.8%	-0.28%
	None	-0.27%	13.92%	-0.020	0.02%	-0.04	-71.6%	0.38%
Cash / Assets	Causal	4.22%	8.28%	0.510	3.39%	0.10	-26.3%	
	GICS	2.92%	8.39%	0.348	3.03%	-0.01	-29.2%	0.43%
	Correlation	3.63%	9.22%	0.393	3.85%	-0.03	-31.4%	0.91%
	None	4.65%	13.14%	0.354	1.45%	0.38	-63.4%	2.28%
Earnings / Price	Causal	0.99%	9.40%	0.105	1.81%	-0.10	-40.7%	
	GICS	0.75%	9.13%	0.082	1.51%	-0.09	-47.8%	1.41%*
	Correlation	-0.33%	8.67%	-0.038	0.12%	-0.05	-53.5%	0.19%
	None	0.73%	13.40%	0.055	2.82%	-0.25	-55.0%	-0.21%
EBITDA / EV	Causal	1.82%	11.21%	0.162	2.69%	-0.10	-46.6%	
	GICS	2.44%	11.48%	0.213	3.31%	-0.10	-47.4%	1.77%**
	Correlation	0.11%	10.75%	0.010	0.98%	-0.10	-56.0%	-0.28%
	None	2.59%	13.14%	0.197	3.80%	-0.14	-55.6%	-0.33%
FY1 Earnings / Price	Causal	0.44%	10.90%	0.040	1.69%	-0.15	-57.2%	
	GICS	-0.66%	10.99%	-0.060	0.30%	-0.11	-64.5%	1.69%
	Correlation	-0.75%	10.67%	-0.071	-0.01%	-0.09	-68.4%	1.40%**
	None	0.77%	12.88%	0.060	3.03%	-0.27	-65.7%	-0.69%
Gross Profitability / Assets	Causal	1.56%	9.56%	0.163	2.20%	-0.08	-31.6%	
	GICS	1.37%	9.72%	0.141	1.96%	-0.07	-42.8%	-0.05%
	Correlation	1.92%	9.66%	0.198	2.73%	-0.10	-28.7%	0.64%
	None	2.58%	11.22%	0.230	2.52%	0.01	-35.0%	0.51%
Residual 12-1M Momentum	Causal	1.11%	6.43%	0.173	1.62%	-0.06	-22.3%	
	GICS	0.65%	6.11%	0.106	1.01%	-0.04	-22.2%	0.39%
	Correlation	0.91%	6.51%	0.139	1.36%	-0.06	-22.5%	0.64%*
	None	0.42%	7.26%	0.058	0.99%	-0.07	-37.4%	0.81%*
Return on equity	Causal	1.74%	8.47%	0.206	2.21%	-0.06	-36.6%	
	GICS	0.91%	8.56%	0.106	1.34%	-0.05	-42.7%	-0.13%
	Correlation	2.01%	8.53%	0.236	2.18%	-0.02	-27.8%	0.62%
	None	2.23%	11.03%	0.202	3.07%	-0.10	-35.9%	-0.31%

Table 4: Spanning regression results for the centrality factor

This table presents the results of spanning regressions of the value-weighted top-minus-bottom quintile portfolio return of the centrality factor over various factor models. 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^5model

	(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. (5) and Eq. (6). R^2_{OS} is estimated following Campbell and Thompson (2008). The Δ 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**
α	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% **
Δ	1.83%	8.16%

Figure 1: GICS classification as of November 2015 (upper) and October 2016 (lower)
This figure visualizes the causal network embeddings and GICS sector classification of 4th November 2015 (top panel) and 3rd October 2016 (bottom panel). In the top panel, Real Estate was still part of the Financial GICS sector. In the bottom panel, Real Estate was moved into a new sector in GICS. Each point corresponds to a constituent of the S&P 500 and is colored by the GICS sector it was a member of at that point in time. The spatial clustering is based on the causal network representation of returns.

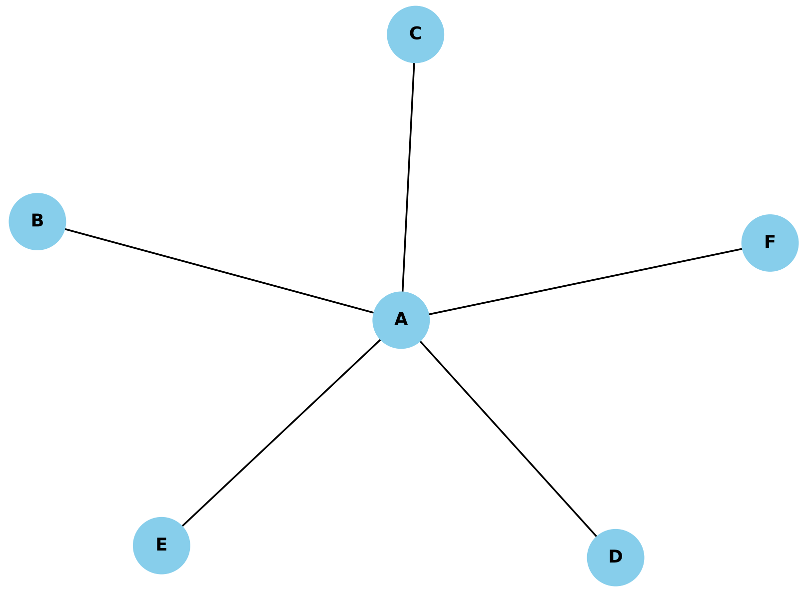

Figure 2: Star network illustration
Example of a star network with a central node (A) and five connected nodes (B–F).

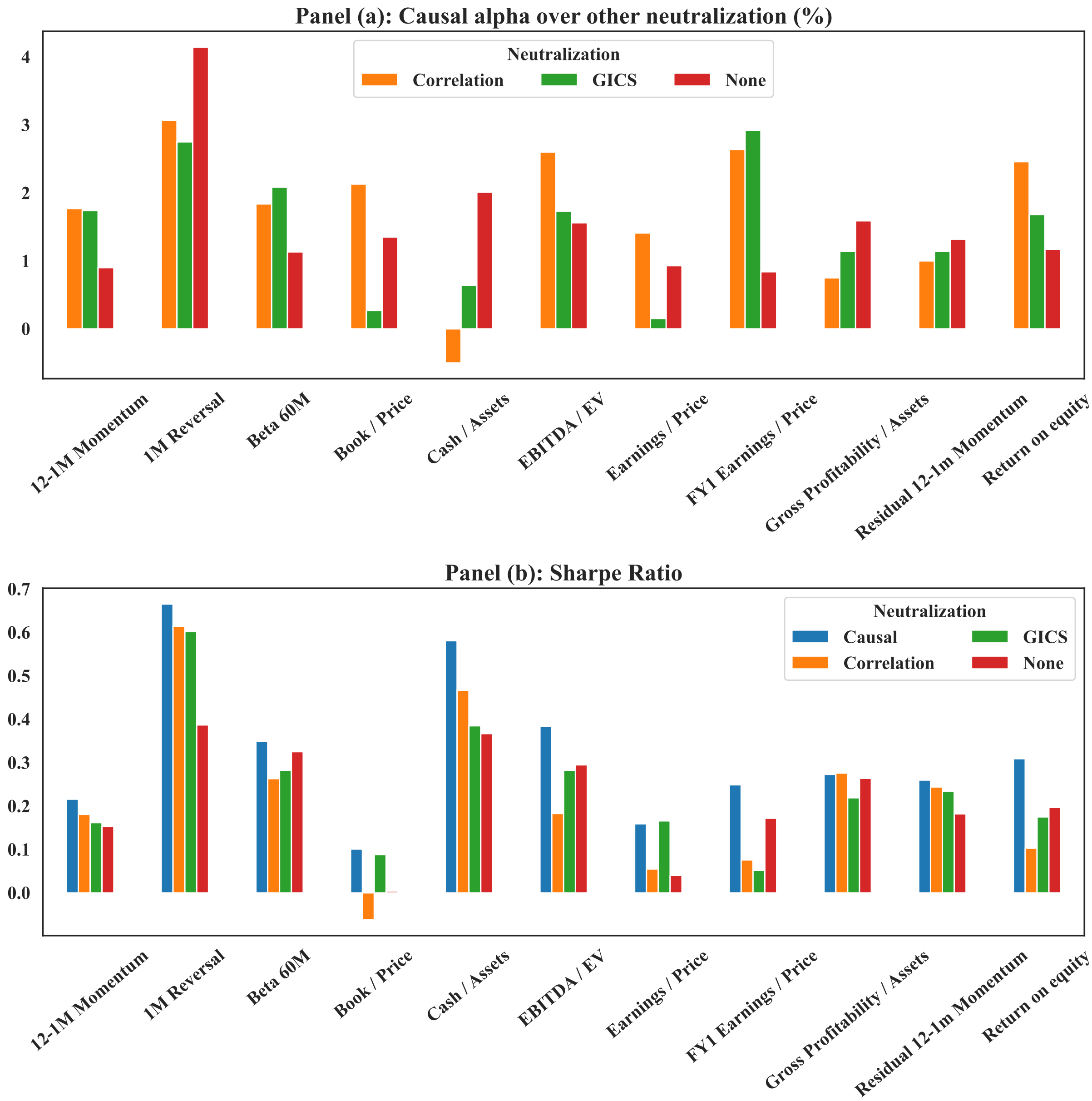

Figure 3: Industry neutralization alpha and Sharpe ratio comparison
Panel (a) plots the alpha of the causality-based industry neutralization method over the statistical correlation, GICS, and no-neutralization approaches. Panel (b) plots the Sharpe Ratio of each approach. Results are calculated on the 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.

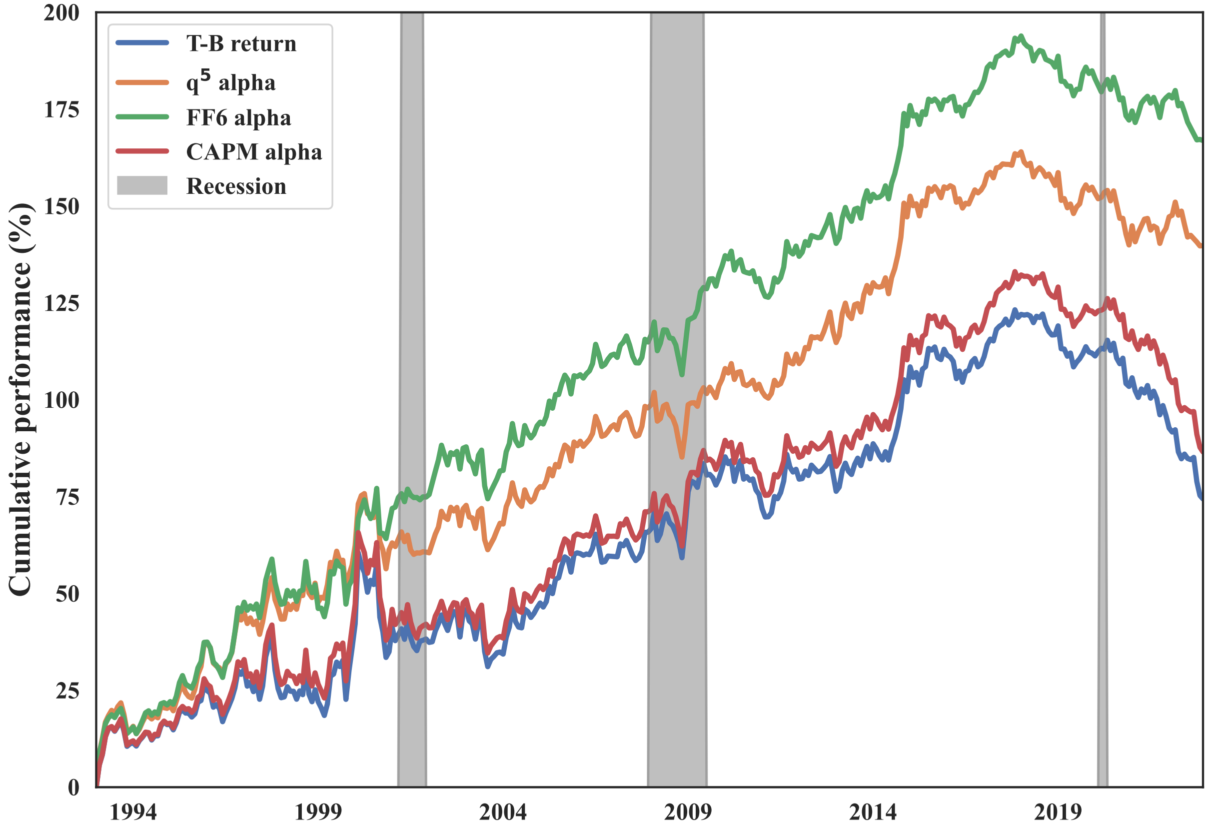

Figure 4: Cumulative alpha of centrality factor
This figure shows the cumulative alpha, in percentage points, of the spanning regressions. The cumulative alpha at time t is calculated as the cumulative sum of $\hat{\alpha} + \varepsilon_{:t}$, where $\varepsilon_{:t}$ are the residuals up to and including time t. NBER recessionary periods are highlighted with gray shading. The sample runs from January 1993 to December 2022.

Causal network representations in factor investing

-Online Appendix-

Clint Howard[1,2], Harald Lohre[1,3] and Sebastiaan Mudde[4]

[1] *Robeco Quantitative Investments*

[2] *University of Technology Sydney*

[3] *Lancaster University Management School*

[4] *Erasmus University Rotterdam*

Appendix A: Additional technical details for causal network estimation

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).[18] 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

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

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 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.[19] 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, μ_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.

[19]We refer the reader to Grover and Leskovec (2016) for a technical review of the Node2Vec algorithm.

Specification of correlation benchmark

We compare the performance of industry-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.[20] 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 ρ is the Pearson correlation measure. Hierarchical clustering can either be agglomerative or divisive in nature. In agglomerative clustering, 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,

[20] See, for example, Mantegna (1999); Bonanno et al. (2003), and Bonanno et al. (2004).

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.

Appendix B: Additional methodology for empirical applications

Industry 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:

$$SR(R^P) = \sqrt{D}\,\frac{\mu(R^P)}{\sigma(R^P)}, \tag{A. 2}$$

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:

$$R_t^{\text{Causal}} = \alpha + \beta R_t^{\text{c}} + \varepsilon_t, \tag{A. 3}$$

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

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:

$$R^2_{OS} = 1 - \frac{MSFE_1}{MSFE_0}, \tag{A. 4}$$

where $MSFE_1$ is the Mean Squared Forecast Error (MSFE) based on the srutinized timing indicator[21], 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:

$$f_{t+1} = (r_{t+1} - \bar{r}_{t+1})^2 - [(r_{t+1} - \hat{r}_{t+1})^2 - (\bar{r}_{t+1} - \hat{r}_{t+1})^2], \tag{A. 5}$$

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. 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):

$$w_t = \frac{1}{\gamma} \frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}, \tag{A. 6}$$

where γ 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:

[21] 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).

$$\text{CER} = \hat{\mu}_p - \frac{1}{2}\gamma\hat{\sigma}_p^2, \tag{A. 7}$$

where $\hat{\mu}_p$ and $\hat{\sigma}_p^2$ are the mean and variance of the portfolio based on the weights in Eq. (A. 6). 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.

Marketing 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:

$$MOM_m = \begin{cases} 1 & \text{if} \quad P_t \geq P_{t-m} \\ 0 & \text{if} \quad P_t < P_{t-m} \end{cases}, \tag{A. 8}$$

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:

$$MA_{j,t}^P = \frac{1}{j}\sum_{i=0}^{j-1} P_{t-i} \quad \text{for} \quad j = s, l, \tag{A. 9}$$

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:

$$MA_{s,l} = \begin{cases} 1 & \text{if} \quad MA_{s,t}^P \geq MA_{l,t}^P \\ 0 & \text{if} \quad MA_{s,t}^P < MA_{l,t}^P \end{cases}. \tag{A. 10}$$