

A Network and Machine Learning Approach to Factor, Asset, and Blended Allocation

GUEORGUI KONSTANTINOV, ANDREAS CHORUS,
AND JONAS REBMANN

**GUEORGUI
KONSTANTINOV**

is a senior fund manager (multi-asset) at LBBW Asset Management in Stuttgart, Germany.
gueorgui.konstantinov@lbbw-am.de

ANDREAS CHORUS

is a senior fund manager at LBBW Asset Management in Stuttgart, Germany.
andreas.chorus@lbbw-am.de

JONAS REBMANN

is a portfolio manager (multi-asset) at LBBW Asset Management in Stuttgart, Germany.
jonas.rebmann@lbbw-am.de

*All articles are now categorized by topics and subtopics. **View at PM-Research.com.**

KEY FINDINGS

- The authors compare network-based asset and factor allocations and blended strategies and argue that network analysis can be used to derive allocation, monitor interactions, and provide additional layers of risk control mechanisms.
- The authors find that factors and assets are strongly interconnected. Investors should pay close attention to currencies and the Fama–French–Carhart factors (RMW, HML, CMA) because they have large centrality scores.
- Using machine learning and predictive models, investors can find asset and factor allocation solutions. The authors argue that factor exposure is desired within asset allocation.

ABSTRACT: *The main idea of this article is to approach and compare factor and asset allocation portfolios using both traditional and alternative allocation techniques: inverse variance optimization, minimum-variance optimization, and centrality-based techniques from network science. Analysis of the interconnectedness between assets and factors shows that their relationship is strong. The authors compare the allocation techniques, considering centrality and hierarchical-based networks. They demonstrate the advantages of graph theory to explain the advantages to portfolio management and the dynamic nature of assets and factors with their “importance score.” They find that asset allocation can be efficiently derived using directed networks, dynamically driven by both US Treasuries and currency returns with significant centrality scores. Alternatively, the inverse variance weight estimation and correlation-based networks generate factor allocation with favorable risk–return parameters. Furthermore, factor allocation is driven mostly by the importance scores of the Fama–French–Carhart factors: SMB,*

HML, CMA, RMW, and MOM. The authors confirm previous results and argue that both factors and assets are interconnected with different value and momentum factors. Therefore, a blended strategy comprising factors and assets can be defensible for investors. As argued in previous research, factors are much more overcrowded than assets. Therefore, the centrality scores help to identify the crowded exposure and build diversified allocation. The authors run LASSO regressions and show how the network-based allocation can be implemented using machine learning.

TOPICS: *Factor-based models, portfolio theory, portfolio construction**

Atremendous amount of research has been done on factors and assets as information quantities (variables) that pertain to an economic network. An economic network provides a view

of relations in the economy, as argued by Fabozzi and López de Prado (2017). In this article, we define the assets and factors as nodes in a network and the links between them as edges. As a result, we use a network represented by interacting assets and factors.

The amount of research dealing with networks and econometric models has increased recently. However, to the authors' knowledge, a return-based network analysis of both factor and asset graphs has not yet been investigated and applied to multi-asset, multifactor, and blended portfolio allocation. In this article, we investigate the relation between assets and factors using centrality-based techniques. Furthermore, we use both directed and correlation-based networks in traditional and alternative portfolio optimization methods and thus derive allocation. We aim to close the research gap by providing a novel approach and an extensive analysis of building factor and asset allocations for investment managers. We evaluate the strategies and suggest practical ways to benefit from our approach within both traditional methods and those that apply the newest techniques using machine learning and artificial intelligence.

The main idea of our study is not to create and test investment strategies but to compare factor and asset allocation portfolios using both traditional and alternative allocation techniques in the spirit of Simonian (2019). To test asset allocation methods, we apply the newest trends in asset management: network analysis and centrality-based allocation algorithms. Specifically, we compare strategies based on minimum-variance optimization (MVP), inverse volatility weighted (IVW), equal weighting (EW), centrality-based correlation network (CCN), and centrality-based directed network (CDN). We evaluate them based on specific performance and risk metrics. In the spirit of López de Prado (2019), we apply hierarchical clustering to derive the best strategy among the asset allocation methods tested.

There are a few studies related to ours. López de Prado (2016) successfully integrated graph theory and machine learning in the asset management industry. Kaya (2015), Raffinot (2018), and Baitinger and Papenbrock (2017) investigated the interconnectedness of assets using correlation networks and derived asset allocation. Bender, Sun, and Thomas (2019) compared asset and factor allocations in a traditional way. Bender, Blackburn, and Sun (2019) related different weighting schemes for factor portfolios and evaluated their significance. However, Arnott et al. (2019) argued that neither the factors, nor the funds

that have factor exposure to them, are independent. Most recently, Ma, Jacobsen, and Lee (2019) and Simonian and Wu (2019) investigated multi-asset portfolios from a machine learning perspective.

Why is network analysis important? López de Prado and Fabozzi (2017) already highlighted the need for alternative methods in financial analysis. Network analysis provides insightful information regarding factor-based connectedness, relationships, and how risk is transferred between network components. Both academic and practitioner research documents crowdedness in factor investing. Consequently, risks are transferred within the web. Therefore, single entities, as funds carrying risk factor or asset exposure, bear and may transfer risks to the whole market.

Balasubramanian et al. (2019) and Konstantinov and Rebmann (2019) argued that pairwise funds, driven by common factor exposure, might show causal influence on each other. Empirical research shows that the market dynamics are driven by common factors. The most widely used factors are value (high book-to-market—low book-to-market [HML]), size (small minus big, [SMB]), the market minus the risk-free rate (MKT-RF), and momentum (MOM), as documented by Fama and French (1992, 1993) and Carhart (1997). We implement the well-known factors with both strong academic and practitioner foundation in our analysis; these were obtained from AQR Capital Management, LLC, and Kenneth French's website. Furthermore, we use a set of broad assets, similar to the ones widely used in the most recent literature. The relation between factors and assets is one of the issues we investigate.

We conduct our analysis on a few levels. First, we intend to provide some insights into the interaction between assets (equivalently, factors). Second, we aim to investigate the extent to which risk can be transferred from one asset or factor to another asset or factor. Put differently, we measure the importance scores and identify the most linked and interconnected assets and factors and confirm previous findings: Value and momentum are indeed everywhere. Third, we test different factor and asset allocation techniques using both traditional and alternative, network-based methods—machine learning and regularization using shrinkage methods (least absolute shrinkage and selection operator [LASSO]). Specifically, we show that directed networks are applied more efficiently to asset allocation, as showed by Konstantinov and Rusev (2020). For factor allocation, correlation networks earn better results.

We argue that the financial markets can be represented as both directed and undirected networks. We run the process in a sample from January 2003–December 2017 to gain some insight into interconnectedness over time, derive asset allocation strategies, and apply regularization and cross-validation. We focus on practical applications relevant for the investment community. Arnott, Harvey, and Markowitz (2019) emphasized the necessary steps to help improve research culture by using backtests and machine learning. Following them, we aim to produce good research rather than to find a winning strategy. The latter may be topic for future research.

Summarizing, we find that the interconnectedness between factors is much higher than that between assets. In this respect, we argue, in line with previous research, that specific factors are crowded. A decrease in asset returns causes a drop in the interconnectedness and, equivalently, an increase as returns recover. However, factors, being interconnected, offer much broader diversification return in an asset portfolio. Moreover, we find that in a network context, assets and factors are related to momentum and value factors.

What are the possible implications and steps for portfolio management? We show that network-based techniques, derived from centrality scores, provide efficient factor portfolio–optimization techniques compared to the established models. Alternatively, asset allocation is derived better using directed networks or the inverse-variance technique. We apply machine learning algorithms to investigate the exposure of these strategies. Our results are in line with previous findings in regard to asset and factor allocation. Hence, investors should use network-based approaches for allocation, risk management, and monitoring purposes and use machine learning techniques.

DATA

We consider two broad datasets reflecting different economic environments in our analysis, each comprising 16 factors and 16 assets from January 2001–December 2017. The data frequency is monthly. The first set is a pure factor set, as is widely considered in both academic and practitioner research, and consists of the following factors:

- The equity premium as the US equity market excess return, denoted as MKT-RF;

- The standard Fama and French (1992, 1993, 2015) factors for value (high minus low), size (small minus big), profitability (robust minus weak), and investment (conservative minus aggressive), denoted as HML, SMB, RMW, and CMA, respectively.
- The one-month lagged equity premium as the total market value of equity, denoted by $MV(t-1)$; this is an illiquidity factor, existing in equity, credit, and duration.
- The return of time-series value and momentum strategies on commodities, currencies, bonds, and equities based on Asness, Moskowitz, and Pedersen (2013) and Moskowitz, Ooi, and Pedersen (2012) and provided by AQR Capital Management, LLC; these are denoted FI.VALUE, FI.MOM, FX.MOM, FX.VALUE, Comdty.VALUE, and Comdty.MOM.
- The excess returns of the long–short equity betting against beta factor (BAB), which comprise portfolios that are long low-beta securities and short high-beta securities, based on Frazzini and Pedersen (2014) and obtained from AQR Capital Management, LLC.

The second dataset is composed of assets and is closely related to asset datasets used in previous research. We suggest that investors are interested in broad diversification, including traditional and alternative assets. The set includes emerging markets, international equities, corporate bonds, private equity, commodity returns, currency, and hedge fund returns:

- The S&P 500 returns, denoted as S&P 500
- MSCI Emerging Markets Index, denoted as EM
- The ICE BofAML US Corporates Index, denoted as US.Corp.
- ICE BofAML US High Yield Corporate Bond Index, denoted as US.HY
- JPMorgan US Treasuries Index, denoted as UST
- ICE BofAML US Inflation-linked Index, denoted as INF.L
- Russell 2000 Index for the US Small Cap Equities, denoted as Russell2000
- HFRX Hedge Fund Global Index, denoted as HFRX
- The Gold Index, denoted as Golds
- The Thomson Reuters Private Equity Index, denoted as PE

- GS Commodity Index, denoted as Comdty
- The US Dollar Spot Index as average international exchange rate, denoted as DXY
- The DJ Euro Stoxx 50 Index, for European equities, denoted as EStoxx50
- Nikkei 225 Index, for Japanese Equities, denoted as Nikkei225
- ICE BofAML Euro Corporate Bond Index, denoted as EU.Corp
- ICE BofAML US Treasury Bills Index, denoted as UST.Bills

HOW TO MEASURE ASSET AND FACTOR CONNECTEDNESS?

Correlation Networks versus Directed Networks

There are two widely accepted methods to consider assets and factors as a web of nodes with links between them. Das (2016) argued that there are two different methods: directed and undirected, or correlation, networks. Whereas Baitinger and Papenbrock (2017) and Kaya (2015) used correlation networks to capture mutual relationships and derive allocations, Konstantinov and Rusev (2020) argued that directed networks successfully capture asymmetric relationships as a web of interacting nodes. Why should we consider both of them? The answer is straightforward. Whereas modern portfolio theory considers correlations between assets, as argued in the seminal article by Markowitz (1952), factor investing stresses that the factors considered in a portfolio should be unrelated. However, Simonian et al. (2019) and Arnott et al. (2019) showed that the factors are related and interconnected. Nonetheless, a relationship might be asymmetric and thus directed. Put differently, a factor might influence another factor, but not necessarily should the latter be related to the former. Specifically, the HML factor might be related to MOM, but this relationship might be one directional. Asness, Moskowitz, and Pedersen (2013) found that there is value and momentum everywhere.

Consider the following example: A strong HML exposure might translate to a strong momentum exposure in the overall equity index, thus with underlying stocks and instruments being related by their factors. Alternatively, a strong fixed-income value factor might be related to fixed-income momentum and to the US

Treasury index. Previous research highlighted the relations among the Fama–French–Carhart (FFC) factors in a machine learning context. However, we argue that financial analysis should consider both directed and undirected networks. We argue that financial markets comprise both.

How can we evaluate networks and thus connectedness? The most basic properties of a network $G = (n, m)$ are the number of *nodes*, or *vertices* n (factors/assets), and *links*, or *edges* (connections) m , where $n \in \mathbb{R}^n$ and $m \in \mathbb{R}^{n \times n}$ ($m_{i,j} \neq m_{j,i}$). As opposed to an undirected correlation network, the properties of the nodes in the network are different. Specifically, a factor i might be linked to a factor j , but this does not mean that the former is directly related (by a correlation coefficient) to the latter (consider, for example, the social network Twitter). Put another way, pairwise connectedness is not symmetrical: It is directed, as showed by Das (2016) and Konstantinov and Rebmann (2019). Directed relationships exist widely in the market. For example, hedge fund replication products replicate return strategies and apply factor or algorithmic approaches. However, their return profiles diverge from the original strategies.

A focal point is to define an appropriate tool for measuring the dynamic nature of directed interconnectivity. In a machine learning context, balancing the trade-off between the flexibility and interpretability of models capturing relations (inference) and making predictions is imperative. Therefore, we use a linear regression as a statistical inference model to capture mutual connectedness. More precisely, we regress each factor (asset) return against each of the other factors (assets) in the sample period, based on the inference model suggested by Konstantinov and Rebmann (2019) for return-based directed networks. Because we are interested in the mutual relation in an unweighted network, the significance of the parameter is relevant and not the magnitude of the coefficients. We use the F-statistic to test for any association between the predictors and response (see James et al. 2013).

$$R_{i,t} = \alpha + \beta_{ij,t} r_{j,t} + \varepsilon_t \quad (1)$$

Here, $R_{i,t}$ is the return of factor (asset) i regressed against the returns of the factor (asset) j at time t ($r_{j,t}$). Thus, Equation 1 serves as an inference model assigning the observations to a class or, more precisely, to a binary

category of relation between the variables.¹ We model this relationship in Equation 2.

A network is represented by nodes (factors/assets) joined in pairs by directed links. The size of the network is given by the number of nodes available, and the links are identified by pairs (i, j) , using Equation 1. We consider the significant coefficients in Equation 1 to construct the adjacency matrix, mapping the links between the assets and factors continuously through the sample period with a time-step of one monthly observation. An $n \times n$ adjacency matrix denoted by A is estimated using following rule:

$$a_{ij,t} = \begin{cases} 1 & \text{if Factor/Asset } i \text{ and } j \text{ are connected} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where a_{ij} are the components of the adjacency matrix time t . If factor/asset i does not have a significant influence on factor/asset j , the value in the matrix is set to zero.

Correlation networks represent a different view and are widely applied in recent research because of their usefulness and easy implementation. In an undirected network, $G^* = (n^*, m^*)$ are the number of nodes (or vertices) n and links (or edges; connections) m where $n^* \in \mathbb{R}^n$ and $m^* \in \mathbb{R}^{n \times n}$ ($m_{i,j}^* = m_{j,i}^*$). This network is symmetrical and using a distance metric d_{ij} , as suggested by Mantegna (1999), allows for adjustment of the correlation coefficients ρ_{ij} at time t .

$$d_{ij,t} = \sqrt{2(1 - \rho_{ij})}, \text{ with } i \neq j \quad (3)$$

Network Properties

A classical financial network has specific properties, as shown by Newman (2010). We use relevant network measures that are widely accepted in the literature and suitable for directed networks. In a social network context, Facebook is example of an undirected web. The average degree $k_{i,t}$ of a node i and node j k_j at time t is a basic structural property representation of the connections of a node in a directed network ($k_{i,t} = k_{j,t} = m_{ij,t}/n$). The relation between the pairs of factors or assets in a network is

given by the value of *reciprocity*. The reciprocity measures whether a factor (asset) i is linked to a factor (asset) j and whether a factor (asset) j is also linked to a factor (asset) i . The values range between 0 and 1 (with 1 for fully reciprocal network). Given the parameter of the regression model, number of links m , and thus the entries in the adjacency matrixes, the reciprocity coefficient ϱ at time t is given by

$$\varrho_t = \frac{\sum_{ij} a_{ij} a_{ji}}{\sum_{ij} m_{ij}} \quad (4)$$

A pair (i, j) is called reciprocal if ties exist between both factors (assets) in both directions and measures the mutual connectedness on a network level, as stressed by Das (2016).

Research on financial networks shows extensive usage of different network measures to capture and explain interconnectedness. However, Borgatti (2005) showed that the different types of networks require specific metrics, depending on the flow process between nodes. Specifically, money flows require metrics such as alpha centrality, as opposed to the classical geodesic metrics, paths, or trail flows relevant for other types of networks.

It is imperative to identify the importance score of the single node in the network. Kinlaw, Kritzman, and Turkington (2012) highlighted the centrality measure in determining the vulnerability of portfolio components to systemic risks and shocks and stressed the role of “important” nodes (assets, or factors) in a correlation-based network. The *eigenvector centrality* is suitable metric for such webs.

$$\Phi x = Ax \quad (5)$$

However, the *alpha centrality* score is preferred to eigenvector centrality because of its usefulness for mapping asymmetric relationships on a single node level in directed networks, as suggested by Bonacich and Lloyd (2001). Specifically, the alpha centrality assigns a score to every node (asset, or factor) that is proportional to the sum of the scores of its neighbor nodes.

$$x = \text{alpha}Ax + e \quad (6)$$

¹ For example, applying generative or discriminative classification methods using naïve Bayes classifiers or a logit model would be topic for further research (see James et al. 2013).

In Equations 5 and 6, α reflects the relative importance of endogenous versus exogenous variables in the determination of centrality, and Φ is a diagonal matrix of eigenvalues; e is a vector of exogenous information and in our case is a vector of ones, and the matrix A is the adjacency matrix.

FROM NETWORKS TO PORTFOLIO MANAGEMENT

Portfolio Allocation

To derive the long-only allocation in a portfolio, we use a number of traditional and alternative allocation algorithms recently applied by, for example, Raffinot (2018). Our sample runs from January 2003 to December 2017, with walk-forward estimations (period $t - 24$ through t) beginning in January 2003. This is the same sample period we use for training and tests within machine learning applications in the following sections.

As a benchmark, we run the $w = 1/N$, or EW approach. We run a set of alternative allocation techniques. A traditional allocation approach is the MVP technique developed by Markowitz (1952). Rather than incorporating a correlation matrix in the MVP optimization, we follow Sandhu, Georgiou, and Tannenbaum (2016) and consider the distance metrics from Equation 3. This approach has been widely applied in previous research.

The inverse-variance (σ_i^{-2}) procedure assigns weights w_i as shown by the following equation:

$$w_i = \frac{\sigma_i^{-2}}{\sum_{i=1} \sigma_i^{-2}} \quad (7)$$

We implement alpha centrality metrics for the directed asset and factor allocations—CDN allocation using the alpha centrality scores from Equation 6. Konstantinov and Rebmann (2019) were among the first to run a similar approach. Alternatively, we apply eigenvector centrality for the correlation network-based asset and factor allocation—the correlation-based correlation network (CCN) using the eigenvector scores from Equation 5. The intuition behind this is straightforward and has already been successfully implemented by Aldridge (2019). Specifically, the approach weights the assets (factors) that are most weakly connected as given by the centrality score. Previous research identified the

power of non-optimization-based tools. We highlight some metrics next to compare the strategies.

Comparing Portfolio Metrics

To compare the strategies defined previously, we apply the notions suggested in recent research. We apply the Sharpe ratio (SR), the adjusted SR (ASR), the certainty equivalent return (CEQ), the average turnover per rebalancing (TO), the Herschman–Herfindhal Index (HHI), and the maximum drawdown (MDD) of each strategy.

The ASR suggested by Pezier and White (2008) considers the skewness s and the kurtosis k , and the original SR (Sharpe 1994) (with standard deviation σ , mean μ , and the risk free rate r_f) is given by the following expression:

$$\begin{aligned} \text{ASR} &= \text{SR} \left[1 + \frac{s}{6} \text{SR} - \frac{(k-3)}{24} \text{SR}^2 \right], \text{ with} \\ \text{SR} &= \frac{(\mu - r_f)}{\sigma} \end{aligned} \quad (8)$$

The CEQ, as defined by DeMiguel, Garlappi, and Uppal (2009), considers the risk-free return for an investor with quadratic utility and risk aversion parameter λ compared to the risky portfolio and is given by the following equation:

$$\text{CEQ} = (\mu - r_f) - \frac{\lambda}{2} \sigma^2 \quad (9)$$

The TO is given as the absolute difference in the portfolio weights after rebalancing over the total number of out-of-sample estimations F . High turnover is often associated with high transaction costs, which might be undesired by investors.

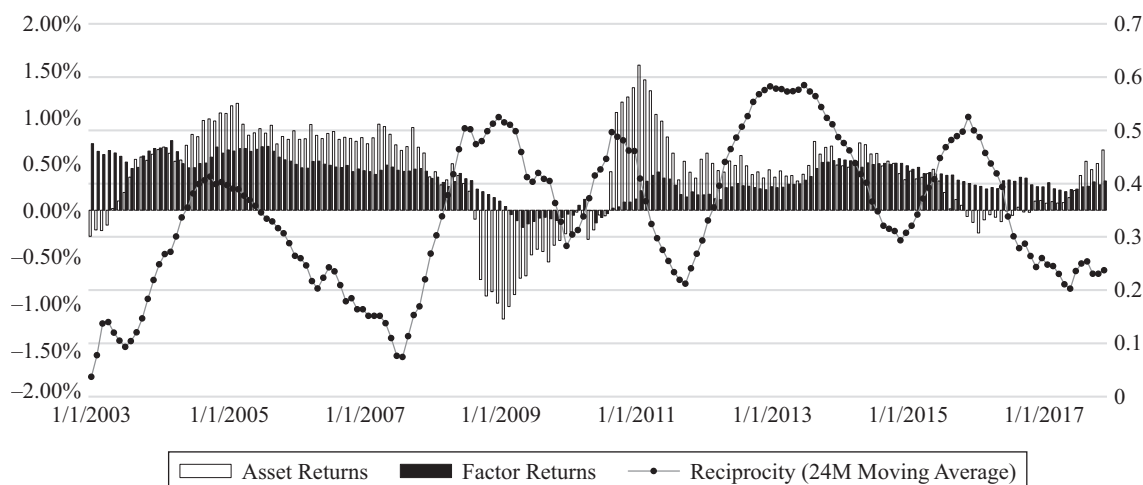
$$\text{TO} = \frac{1}{F} \sum_{t=2} |w_t - w_{t-1}| \quad (10)$$

The HHI is widely used in investment management as an indicator for diversification in a portfolio. It is given by following notation:

$$\text{HHI} = \frac{1}{F} \sum_{t=1} w_t^2 \quad (11)$$

EXHIBIT 1

Asset, Factor Returns, and Total Reciprocity (January 2003–December 2017)



Notes: The total reciprocity (right-hand axis) is calculated using Equation 4 with period $t - 24$ through t , with a rolling time-step of one-month using the total assets and factors in the analysis (32). Asset and factor returns are on the left-hand axis. The first 24 observations are omitted owing to the regression estimation, according to Equation 1. The sample contains 180 monthly observations from January 2003–December 2017. The R^2 is 0.14, with an average total reciprocity (network level connectedness) of 0.35 and standard deviation of 0.19. Significance is at the 1% level.

Sources: Prepared by the authors from Bloomberg Finance LP, AQR Capital Management, LLC, and Kenneth French's website data.

The MDD is an indicator for the maximum loss of capital an investor can suffer from the peak to the bottom in the net asset value of a portfolio.

An algorithm involving our analysis would include the following steps, considering issues stressed by López de Prado (2018) and Arnott, Harvey, and Markowitz (2019):

1. Run inference models to the data: assets, factors, big data, and economic variables
2. Capture connectedness: adjust t -statistics for false positives, build adjacency matrixes
3. Compute appropriate centrality metrics for the data, identify clusters
4. Run portfolio optimizations with one-period forward-walk step
5. Divide the data into k -fold training and test folds, adjusting for data leakage (*purging*)
6. Run regularizations and compute the mean squared error (MSE) for the strategies.

ARE FACTORS AND ASSETS (UN-)RELATED?

We investigate an important question: Are factors and assets unrelated? López de Prado and Fabozzi (2017)

argued that economics is a science of relations, which is serious motivation for us. We extend previous network-based analysis to a broad spectrum. Applying graph theory and directed network analysis, we investigate the relations between factors and assets. We find that our factors and assets are, in fact, related and thus interconnected. Asness, Moskowitz, and Pedersen (2013) argued in their seminal article that there is value and momentum everywhere. The authors' findings alone suggest that there is an underlying factor relation, which we investigate next.

To compare asset and factor allocation, an assessment of the relations between assets and factors is essential from both portfolio management and risk points of view. Whereas Simonian et al. (2019) showed that factors are interconnected, a broad investigation across assets and factors would highlight important allocation issues. We investigate the network structure on both the node and the edge level. In Exhibit 1, we show the connectedness as measured by the reciprocity of the asset and factor set and compare the mutual connectedness to the rolling average returns of the assets and factors. Konstantinov and Rebmann (2019) argued that reciprocity—as a measure of interconnectedness on the network level—mimics the carry, value, and trend-crowded trades of currency exposure. Konstantinov and Rusev (2020) showed that

EXHIBIT 2

Predictive Relationship between Returns and Total Reciprocity

Asset Relationship			Factor Relationship		
	Slope	Intercept		Slope	Intercept
Coef.	−0.0072	0.0064	Coef.	−0.0054	0.0053
t-Stat.	−2.612	6.201	t-Stat.	−4.945	13.042
R ²		0.04	R ²		0.12

Notes: Predictive regression of factor and asset returns against the reciprocity computed using Equation 4. Significance at the 1% level.

reciprocity is a proxy for crowdedness on a factor level. The average degree (node level connectedness) is seven, and its shape is similar to the reciprocity (unreported). Put differently, a node is connected to roughly seven other nodes. A visual inspection of Exhibits 1 and 2 confirms the negative relationship between reciprocity and the factor and asset returns.

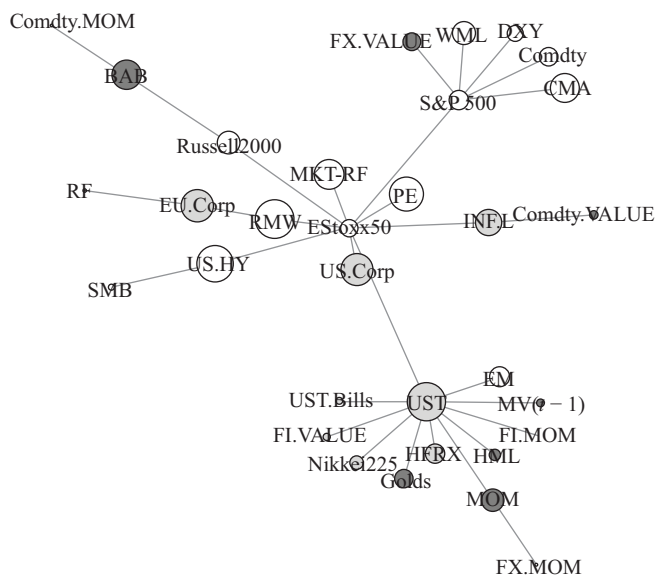
For example, a drop in lagged total reciprocity is associated with a decline in both asset returns by roughly 72 bps (*t*-statistic of −2.612). Alternatively, a decrease in the reciprocity momentum is associated with a factor return decline of 54 bps (*t*-statistic of −4.945). Thus, interconnectedness is concerning in times of return deterioration. The average total reciprocity value is 0.35, and the median is 0.36. We argue that both factors and assets are, in fact, a crowded place. In contrast, asset returns are much more volatile; however, the mutual relation dries out as return momentum decreases. Unreported tests show that the return meltdown during the financial crisis in 2008 resulted in even higher factor interconnectedness. Additional unreported tests with lagged data confirm the negative relationship. A central question is which factors are the most relevant for investors.

Plotting the network of assets and factors, we show these links for the directed network from January 2003–December 2017 in Exhibit 3. We use the minimum spanning trees (MST) technique developed by Prim (1957) and implemented in investment management by López de Prado (2016).

Our results confirm the findings of Bergeron, Kritzman, and Sivitsky (2018) that factor exposure is desired within asset allocation. The network structure easily reveals the interconnectedness between assets and factors. Specifically, an observant reader can easily

EXHIBIT 3

Clustered Asset and Factor MST Network (January 2003–December 2017)



Notes: The MST network comprises 32 nodes (both assets and factors) from January 2003–December 2017; the weight of a node size depends on the links. The current node is associated with degree centrality; the larger the node, the higher the number of links attached to it. We use the cluster spinglass algorithm, suitable for directed graphs with highest modularity score. The first cluster (white) is composed of the following factors and assets: EStoxx50, S&P 500, Russell2000, EM, DXY, US.HY, Comdty, PE, MKT-RF, SMB, RMW, CMA, and WML. The second cluster (light gray) is composed of HFRX, UST, US.Corp, INF.L, EU.Corp, UST.Bills, Nikkei225, RF, FI.VALUE, and FI.MOM. The third cluster (dark gray) is composed of Golds, HML, BAB, MV ($t - 1$), FX.VALUE, FX.MOM, Comdty.VALUE, Comdty, MOM, and MOM. We used the “igraph” package in software R.

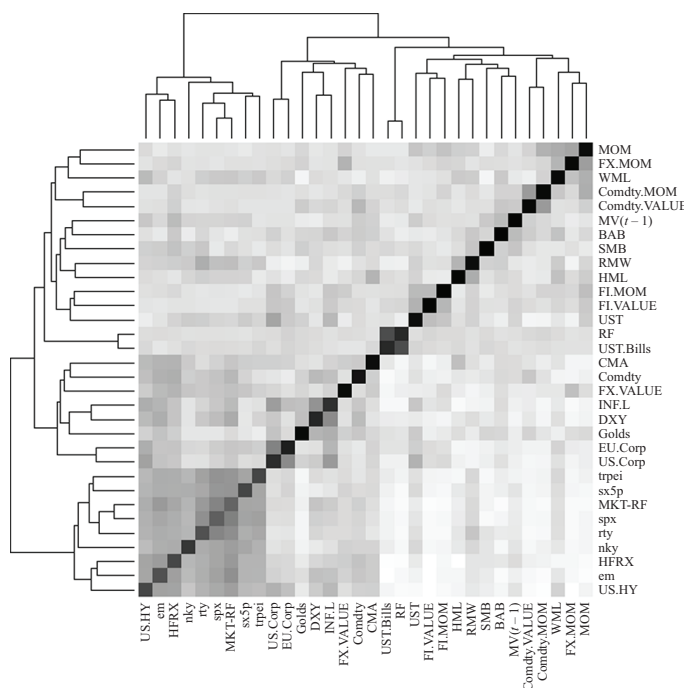
Source: Prepared by the authors from Bloomberg Finance LP, AQR Capital Management, LLC, and Kenneth French's website data.

identify that every cluster and tree in the network comprises a value and momentum factor, linked to other factors and to assets. Every node in the network is connected to seven other nodes on average (average degree). For example, the FX-Value factor, which is an explanatory variable of the DXY, is connected to the S&P 500 index. Sliding down, note the connection to the RMW and the Euro Stoxx 50 Index, which is closely linked to commodity value factor (Comdty.VALUE) and inflation-linked securities. In addition, the US Treasuries (UST) are closely linked to hedge funds (HFRX); the liquidity premium gained as exposure in the MV($t - 1$); both the fixed-income value and

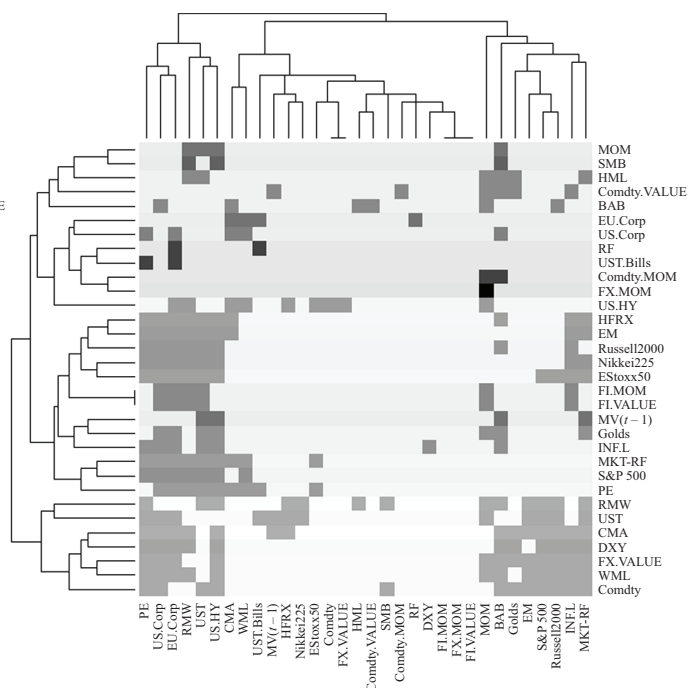
EXHIBIT 4

Hierarchical Clustering of Assets and Factors: January 2003–December 2017

Panel A: Correlation-Based Network



Panel B: Directed Network



Sources: Prepared by the authors from Bloomberg Finance L.P., AQR Capital Management, LLC, and Kenneth French's website data.

fixed-income momentum (FI.VALUE and FI.MOM); and the FX momentum factor (FX.MOM). We confirm findings from previous research and argue that a blended, combined factor and asset allocation strategy might reflect efficiently the relations between factors and assets.²

What are the possible implications for investment professionals? Predictive models for link prediction between assets and factors would help to solve risk and allocation issues. In addition, implementing classical regression analysis helps to predict both asset and factor returns. Thus, in the spirit of López de Prado (2018), investment managers can solve the *Meta Labeling* paradigm thinking data-analytically, as emphasized by Provost and Fawcett (2013), thus separating allocation and risk management decisions, for example.

² Adjusting the p-values for the family-wise error rate (t -statistic of 3.066), the total links of the assets and factor over the sample period drop from 236 to 94 (see the Appendix). The most connected—as measured by the centrality scores—factors and assets are BAB, US.Corp, MOM, and EU.Corp. The degree centrality assigns maximum scores to US.HY, RMW, and EU.Corp.

Now, we take an alternative route and show the relations in a hierarchical structure, as noted by López de Prado (2016). In Exhibit 4, Panel A, we show the correlation structure of the web of factors and assets. The lighter the color, the stronger the relation between the components. Visual inspection of Exhibit 4, Panel B, reveals that the dependencies for both networks are very strong.³ Following Raffinot (2018), a classical hierarchical allocation would be to split the weights in half for every step in the *dendrogram* tree. However, investigating hierarchical clustering strategies is beyond the scope of our article because it is well documented in previous research. Might it not be better for investors to avoid interconnected factors? We think not. Investors should consider both assets and factors in their search for diversified solutions. We provide supportive arguments in the next sections.

³ Provost and Fawcett (2013) and James et al. (2013) stressed that the choice of statistical learning algorithms depend on data, can be highly subjective, and result in different estimations.

EMPIRICAL RESULTS FOR PORTFOLIO ALLOCATION

Pure Factor Allocation

We compute the alpha centrality scores for the factors to gain information regarding the dynamic importance during the sample period. In line with previous research, we find that the most central or “importance” score is associated with the factors FX.VALUE, Comdty.VALUE, MKT-RF, and the RMW. Previous research documented a significant role for both equity and bond funds of the HML factor. However, specific value scores dominate over MOM factors during the sample period. This suggest that portfolio allocation remains crucial for managing risk. Using boxplots, we show the distribution of the scores in Exhibit 5.

The results for the factor portfolios show that the inverse-variance allocation earns the best results in terms of ASR, MaxDD, TO, and HHI. Both CCN and IVN outperform the benchmark strategy (EW). However, the results for the CCN with high ASR and low MaxDD suggest that it is an appropriate defensible and defensive factor-based strategy. We show the performance metrics in Exhibit 6.

Pure Asset Allocation

Computing the alpha centrality scores, we find that the currency returns of basket currencies versus US dollar (DXY), commodities (Comdty), and UST.Bills—as a proxy for the risk-free investments within the asset allocation framework—have the highest importance scores (see Exhibit 7). Put differently, currencies and bonds determine the network relations between assets. These results are not surprising. Therefore, managers should pay close attention to currencies (with large variance and few outliers in the scores), given their strong interconnectedness, because both fixed income and equities heavily rely on foreign exchange exposure, as showed by Konstantinov (2017), de Boer (2016), and most recently by Pojarliev (2019).

Compared to the pure factor portfolios, the CDN algorithm, incorporating directed networks, shows the best results in terms of the performance metrics. For asset allocation, the IVW strategy earns superior results in terms of the highest diversification, SR, and low MaxDD, which makes it suitable for investors with

EXHIBIT 5

Factor Alpha Centrality Scores: January 2003–December 2017

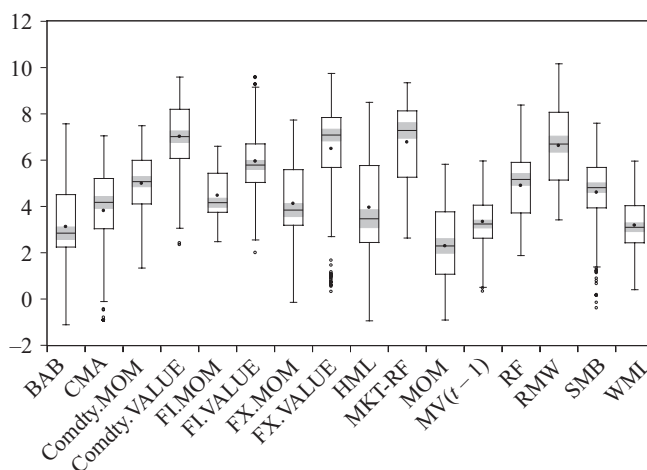


EXHIBIT 6

Performance Metrics for the Pure Factor Allocation

	MVP	CCN	IVW	EW	CDN
SR	0.1865	0.4774	0.5504	0.4538	0.3289
ASR	0.1794	0.4789	0.5598	0.4443	0.3320
CD	0.0032	0.0031	0.0010	0.0033	0.0042
TO	0.2457	0.0122	0.0250	—	0.6100
HHI	0.2326	0.0762	0.6467	0.0625	0.1310
MaxDD	-0.1411	-0.0511	-0.0154	-0.0575	-0.0692

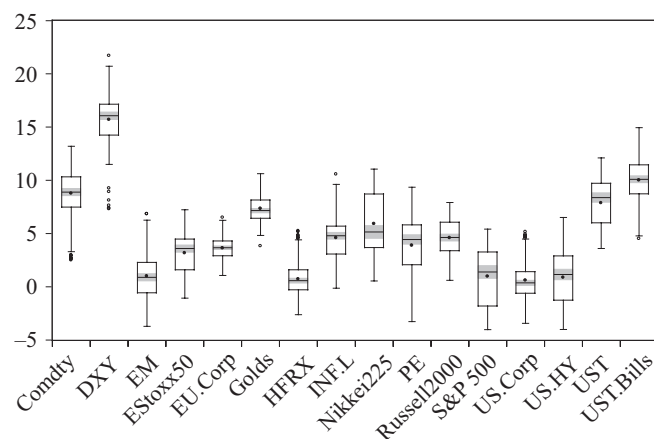
greater risk aversion. However, the CDN has highest ASR and CD. A drawback is the high TO. We argue that the CDN is suitable for aggressive positioning in asset allocation, with high tolerance for risk. We show the performance metrics of the asset strategies in Exhibit 8.

Blended Allocation

Bergeron, Kritzman, and Sivistky (2018) proposed a framework to add factor exposure to asset allocation. Thus, investors can earn the highest benefits from both approaches. Most recently, Bass, Gladstone, and Ang (2017) showed an integrated approach to selecting and deriving factor and asset allocation. More precisely, the authors showed how to derive factors from assets. Madhavan, Sobczyk, and Ang (2018) documented the factor exposure of major equity indexes. We follow

EXHIBIT 7

Asset Alpha Centrality Scores: January 2003–December 2017



previous research and test a mixed strategy of both asset allocation and factor exposure. To investigate the impact of factor exposure on a portfolio, we add the Fama–French–Carhart factors, MOM, SMB, HML, RMW, and CMA, to the 16 assets and investigate the same five allocation techniques.⁴ A visual inspection of the boxplots in Exhibit 9 reveals important information on the factor ability to gain importance in a portfolio. Comparing the centrality metrics, we document large importance scores of factors, suggesting crowdedness as noted by Kinlaw, Kritzman, and Turkington (2012). As argued in previous literature, factors are greatly overcrowded compared to assets, and their centrality scores help identify the exposure to crowdedness risk. The factor scores are higher than the asset scores. Specifically, because DXY remains the most important asset and might be hedged, the scores of the RWM, SMB, and CMA are elevated. We argue that factors achieve diversification. In fact, they absorb portfolio risks. Given the documented factor crowdedness, blended strategies remain challenging and require additional monitoring and investigation. This could be promising topic for further research.

We show the performance metrics for the blended strategy in Exhibit 10. The CCN, IVW, and CDN show

⁴The number of possible combinations of different blended strategies is large and beyond the scope of our article. Our aim is to highlight the applications of network and machine learning techniques for factor and asset allocation and not to develop a “true” trading strategy.

EXHIBIT 8

Performance Metrics for the Pure Asset Allocation

	MVP	CCN	IVW	EW	CDN
SR	0.1468	0.1798	0.2798	0.2035	0.2667
ASR	0.1441	0.1650	0.1957	0.1886	0.2375
CD	0.0052	0.0027	0.0015	0.0042	0.0051
TO	0.1474	0.1067	0.0236	–	0.4907
HHI	0.2473	0.1158	0.6501	0.0625	0.0949
MaxDD	–0.4305	–0.1721	–0.0699	–0.2375	–0.1743

higher SR than the benchmark EW strategy. Which strategy is preferred now?

Following Harvey and Liu (2014), we translate the SRs into a t -statistic. Adjusting the p -values for the tests according to a hierarchical clustering (see the Appendix), the threshold t -statistic is 3.066. Put differently, we would need to observe a t -statistic of at least 3.066 to declare a true discovery with 99% confidence. An observant reader can easily calculate the metrics for the strategies shown. For example, the observed t -statistic for the CDN blended strategy is $3.61(0.2693/\sqrt{180})$, well above 3.066. Put otherwise, the minimum threshold SR for the strategies is 0.2285.⁵ It is worth mentioning that applying the probabilistic SR would be topic for further research and valuable extension of our analysis.

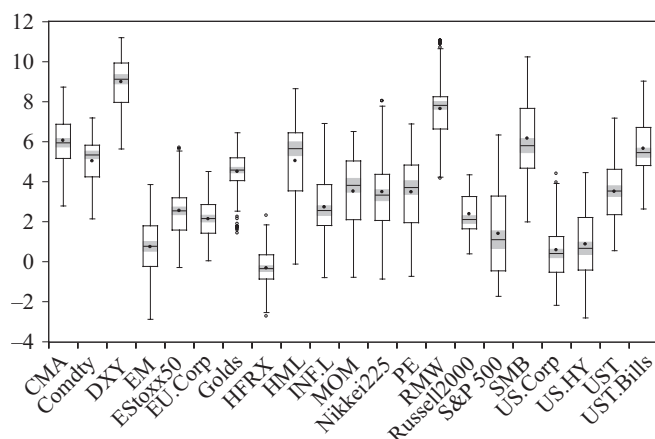
MACHINE LEARNING IN PORTFOLIO MANAGEMENT

In the spirit of Bergeron, Kritzman, and Sivitsky (2018), we add the FFC (HML, SMB, RMW, CMA, and MOM) factors to the primary set of 16 assets and run all models. We find that the factors become an integral part of the portfolio. What are the possible implications for portfolio management? The applications of machine learning to factor and asset portfolio allocation may have an important impact on the

⁵Recent research points to the usefulness of a different correction methods (see Harvey and Liu 2014). We adjust the p -values according to the Bonferroni correction (see the Appendix). For example, we randomly exchange assets with factors in five strategies (not reported but available upon request). Considering all combinations of 16 assets and the five FFC factors would result in $21!/16!5! = 20,349$ possible strategies. Assuming fully independent 20,349 possible strategies, the t -statistic after the Bonferroni correction is 5.219 ($0.01/20349 = 0.0000005$) and the threshold SR is $0.38 = 5.219/\sqrt{180}$. However, with this threshold only the IVW would survive.

EXHIBIT 9

Asset and Factor Alpha Centrality Scores:
January 2003–December 2017



markets and investors. For example, the CDN is more suitable for asset allocation than for factor allocation. A possible explanation is that the correlation-based networks are derived using the distant metric from Equation 3, which focuses on return profiles rather than return magnitude. This also explains the broad feature selection and diversification.

Portfolio Allocation: Feature Importance

In their seminal analysis, Feng et al. (2019) applied quantitative process using LASSO to determine the significance of factors in an empirical analysis. We follow the procedure for machine learning algorithms, as suggested by Provost and Fawcett (2013), and apply predictive models for both the training and test data. Specifically, we perform cross-validation, splitting the original data into training folds and handout folds. We use the *glmnet* package in R.

First, applying LASSO to the whole data (in-sample) asset allocation, the DXY and EM coefficients are zero and the intercept is 14 bps for the CDN allocation. However, none of the other allocations delivers a significant positive intercept (except EW). Nevertheless, the CCN and the IVW show distinct allocations to all assets (see Exhibit 11, Panel A). For factor allocation, LASSO predicts the highest intercept of 6 bps for the CCN allocation algorithm. However, the CDN is very close with 5 bps and shrinkage of the RMW and RF factors, which have strong influence, as measured by the

EXHIBIT 10

Performance of the Blended Allocation

	MVP	CCN	IVW	EW	CDN
SR	0.1399	0.3473	0.4612	0.2498	0.2693
ASR	0.1379	0.3329	0.3828	0.2292	0.2351
CD	0.0050	0.0031	0.0016	0.0040	0.0041
TO	0.2209	0.1055	0.0376	–	0.6865
HHI	0.2563	0.0763	0.5629	0.0476	0.0750
MaxDD	–0.4775	–0.0672	–0.0328	–0.1722	–0.1460

centrality scores and shown in Exhibit 7, thus making the strategy less vulnerable to risks associated with these factors. Summarizing, we have evidence of strong allocations with significant intercepts and shrinkage of the assets and factors with the highest centrality scores. Therefore, we argue that network analysis, combined with predictive models, earns robust asset allocation highlighting the *feature importance*.

The blended strategy earns similar results. Specifically, the shrinkage in the S&P500, PE, UST, and UST.Bills has a positive impact because of the strong centrality scores. However, the blended allocation with CDN has a monthly intercept of 8 bps, compared to the 11 bps in the CCN, outperforming the no-information EW benchmark strategy. Therefore, correlation-based networks show better results (intercepts) and balanced allocations for blended strategies.

Cross-Validation

We run the cross-validation tests for all strategies, dividing the training set into four folds and applying a model for the holdout fold. Specifically, we split the sample of 180 observations into four training sets (each comprising 144 observations) and a test set consisting of 36 monthly observations for every allocation strategy. It is not possible for a single month to appear in both the training and the test set, but it is possible for a single month to appear in more than one training set. We run independent LASSO for the training and test sets. An intuitive reader would immediately recognize that we run multiples of tests for the strategies, adjusted for purging (see López de Prado 2018). For the sake of brevity, we do not report all results and instead focus on the best models, which we select after computing the MSE and the alpha of the models, but all results are available upon request.

EXHIBIT 11

LASSO Regressions for the Asset and Factor Allocation Strategies

Panel A: Asset Allocation						Panel B: Factor Allocation					
Coef.	MVP	CCN	IVW	EW	CDN	Coef.	MVP	CCN	IVW	EW	CDN
(Intercept)	0.0004	0.0003	0.0005	0.0003	0.0014	(Intercept)	0.0013	0.0006	0.0002	0.0002	0.0006
EStoxx50	0.0209	0.0216	0.0096	0.0676	0.0249	MKT-RF	0.0814	0.0771	0.0059	0.0584	0.0519
S&P 500	0.0338	–	0.0216	0.0575	0.0552	SMB	0.1051	0.0588	0.0270	0.0577	0.0010
Russell2000	0.0187	0.0421	0.0046	0.0622	0.0401	HML	–	0.0499	0.0184	0.0596	0.0199
EM	0.0103	–	0.0075	0.0604	–	RMW	0.1103	0.0667	0.0263	0.0491	–
HFRX	–	0.1144	0.0688	0.0727	0.2054	CMA	0.0071	0.0491	0.0161	0.0559	0.2023
DXY	–	0.2062	0.0348	0.0164	0.0105	RF	0.5886	0.1960	0.7792	0.0266	–
Golds	–	0.0776	0.0042	0.0588	0.0917	WML	0.2870	0.0118	0.0124	0.0615	0.0249
UST	–	0.1217	0.0095	0.0423	0.1832	BAB	0.0871	0.0243	0.0108	0.0612	0.0941
US.Corp	–	–	0.1448	0.0872	0.0267	MV($t-1$)	0.0049	0.0523	0.0061	0.0620	0.0655
US.HY	0.0202	0.0315	0.0398	0.0566	0.0561	FX.VALUE	–	0.0914	0.0075	0.0573	0.0726
INFL	–	0.0640	0.0073	0.0311	0.0143	FX.MOM	0.1160	0.0447	0.0151	0.0604	0.0376
EU.Corp	–	0.0456	0.0854	0.0612	0.1606	FI.VALUE	0.0784	0.0714	0.0444	0.0518	0.1619
Comdty	–	0.1158	0.0084	0.0581	0.0773	FI.MOM	–	0.0594	0.0385	0.0569	0.2626
PE	0.0949	0.0616	0.0048	0.0563	0.0286	Comdty.VALUE	–	0.0932	0.0032	0.0606	0.1332
UST.Bills	–	0.1975	0.7918	–	0.1851	Comdty.MOM	0.0087	0.0598	0.0062	0.0613	0.1706
Nikkei225	–	0.0641	0.0129	0.0620	0.0867	MOM	0.1322	0.0477	0.0035	0.0626	0.0343

EXHIBIT 12

LASSO Regression for the Blended Allocations

Coef.	MVP	CCN	IVW	EW	CDN
(Intercept)	–0.0022	0.0011	–0.0003	0.0004	0.0008
EStoxx50	0.1011	0.0417	0.0008	0.0498	0.0247
S&P 500	0.3286	–	0.0095	0.0371	–
Russell2000	0.3010	0.0083	0.0064	0.0525	0.0669
EM	0.0382	–	0.0025	0.0464	0.0219
HFRX	–0.0807	–	0.0309	0.0576	0.2069
DXY	–0.1487	0.0919	0.0070	0.0189	0.0382
Golds	–0.0284	0.0457	0.0034	0.0447	0.0697
UST	0.0507	0.0390	0.0390	0.0480	–
US.Corp	0.0962	–	0.0815	0.0591	0.1273
US.HY	0.0971	–	0.0119	0.0443	0.0042
INFL	–0.1308	0.0273	–0.0145	0.0251	0.0457
EU.Corp	0.2469	0.0297	–0.0202	0.0498	0.0861
Comdty	0.0009	0.0626	0.0060	0.0454	0.0459
PE	0.0655	0.0622	0.0133	0.0419	–
UST.Bills	0.0769	–	0.7615	–	–
Nikkei225	0.0370	0.0282	0.0060	0.0477	0.0444
SMB	–0.1248	0.0694	0.0354	0.0293	0.0079
HML	0.2853	0.0389	0.0484	0.0389	0.0002
RMW	0.3027	–0.0184	0.0345	0.0226	0.0177
CMA	–0.3632	0.0258	–0.0194	0.0260	0.0277
MOM	0.0559	0.0390	0.0082	0.0463	0.0559

Defining a training fold and a test fold for every of the five strategies (MVP, CCN, CDN, EW, and IWR) results in 50 models (25 training and 25 test) for each of the factor, asset, and the blended allocations (total 150). As noted by James et al. (2013), we compare the MSE as the difference between the training and test models and select the best model and thus the sweet spot (lowest MSE between test and training data). We show the LASSO coefficients for the blended allocation strategies in Exhibit 12 and the LASSO coefficients for the selected strategies in Exhibit 13.

Taking the squared difference of the MSE of the training and the test models, we show that the strategies with the highest intercepts, lowest number of factors, and lowest MSE have exposure to the S&P 500 Index and incorporate alternative investments. Inflation-linked exposure, Gold, PE, and the HFRX investments are an integral part of both asset and blended allocation. Factor portfolios using machine learning incorporate equity-related factors. For example, the blended strategy has exposure to the CMA factor and, in fact, momentum is everywhere. As Provost and Fawcett (2013) pointed out, the value of predictive models stems from understanding them rather than focusing on exact predictions.

EXHIBIT 13

Cross-Validation: In-Sample and Out-of-Sample LASSO Regressions

Panel A: Asset Allocation			Panel B: Factor Allocation			Panel C: Blended Allocation		
Coef.	Training	Test	Coef.	Training	Test	Coef.	Training	Test
(Intercept)	0.0014	0.0008	(Intercept)	0.0008	0.0013	(Intercept)	0.0009	0.0018
EStoxx50	—	—	MKT-RF	0.0634	0.0600	EStoxx50	0.0185	—
S&P 500	0.0332	0.1245	SMB	0.0561	0.0078	S&P 500	—	0.1694
Russell2000	0.0236	0.0626	HML	0.0282	0.1077	Russell2000	0.0546	—
EM	0.0071	0.0091	RMW	0.0300	0.0399	EM	0.0286	—
HFRX	0.1924	0.1450	CMA	0.0332	0.0477	HFRX	0.2293	0.1503
DXY	—	—	RF	−0.2206	—	DXY	0.0344	—
Gold	0.0708	0.1328	WML	0.0257	—	Gold	0.0655	0.0822
UST	0.0012	0.1051	BAB	0.0181	0.0070	UST	—	—
US.Corp	0.0820	0.1540	MV($t-1$)	0.0514	0.0619	US.Corp	0.1066	—
US.HY	0.0639	—	FX.VALUE	0.0862	—	US.HY	0.0074	—
INFL	0.0457	—	FX.MOM	0.0579	0.0244	INFL	0.0512	0.0759
EU.Corp	0.2316	—	FL.VALUE	0.1167	—	EU.Corp	0.1170	—
Comdty	0.0731	—	FL.MOM	0.0998	—	Comdty	0.0445	—
PE	0.0245	0.0729	Comdty.VALUE	0.0951	0.0705	PE	—	—
UST.Bills	—	—	Comdty.MOM	0.0648	0.0180	UST.Bills	—	—
Nikkei225	0.1000	0.0474	MOM	0.0411	0.0431	Nikkei225	0.0499	—
						SMB	—	—
						HML	—	—
						RMW	—	—
						CMA	0.0423	—
						MOM	0.0520	0.0257

Notes: Panel A refers to the CDN strategy. In Panel B, we refer to the CCN strategy, and in Panel C the CDN strategy.

CONCLUSION

The main idea of our study is not to create and test investment strategies but to compare factor and asset allocation portfolios, using both traditional and alternative allocation techniques. For this purpose, we use graph theory and centrality-based optimization techniques and show that assets and factors are themselves strongly connected. Specifically, we determined the underlying connectedness, considering both correlation-based and directed networks.

Factors play a central role in the market. However, assets connect clusters. Put differently, connections between assets are the bridges by which both factors and assets and thus markets can be most affected. Our findings have important implications for institutional investors. First, deriving asset allocation using network-based approaches earns the best results and can be integrated in an investment process and a machine learning algorithm. Second, traditional factor allocation approaches easily help to build diversified factor portfolios, and network

metrics help to monitor the interaction between the most well-known factors—HML, SMB, MOM and the MKT-RF. We show how managers can use the relations between factors and assets, deriving integrated asset allocations based on a hierarchical clustering, centrality scores, traditional techniques, and the LASSO procedure. Therefore, network analysis and LASSO provide efficient solutions to asset and factor allocation in portfolio management. As Provost and Fawcett (2013) stressed, the value of predictive models stems from understanding them rather than focusing on their predictions. We show direct results for a machine learning implementation with training and test sets and are aware of overfitting. Centrality-based allocation earns superior results. Identifying profitable trading strategies (true positives) remains a topic for further research. We argue that investors should use both traditional techniques and alternative allocation methods to derive the best allocation according to their desired utility. Graph theory is an efficient tool for such an exercise.

APPENDIX

REPORTING FAMILY-WISE-ERROR-RATE (FWER)

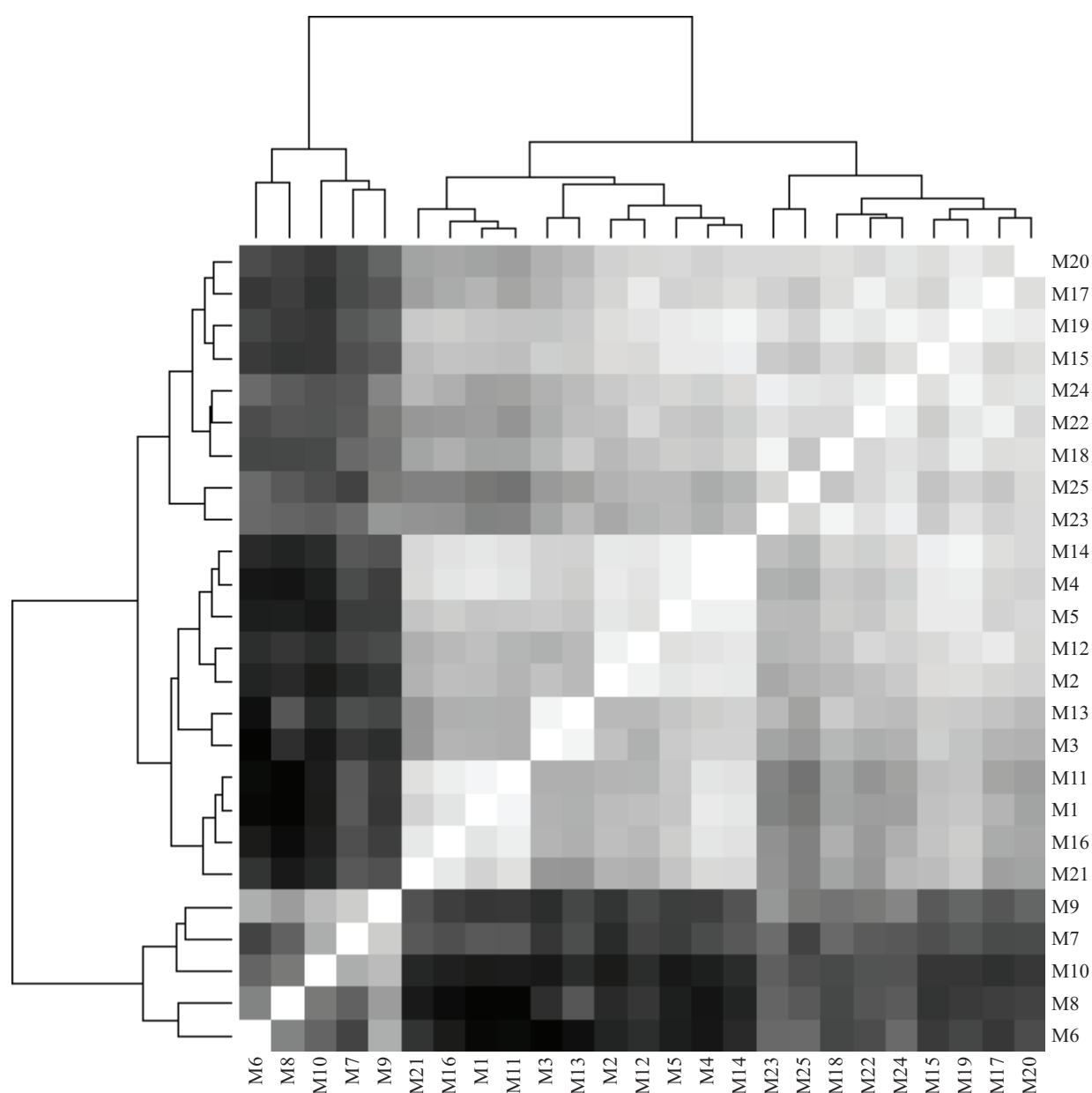
We follow Fabozzi and López de Prado (2018) and López de Prado (2019) and disclose the FWER—the probability of a coming to at least one false conclusion (Type I error) with the strategies tested. Prior to the FWER adjustments,

we use significance at the 1% level. We compute the hierarchical clustering of 25 return time-series of all strategies (Exhibit A1). Dark shading indicates that the correlation between returns of two strategies is high.

We argue that the number of clusters ($c = 4$) reflects the four theoretical foundations of the strategies tested. We adjust the p -values (Bonferroni correction) by dividing the original acceptance threshold by the number of clusters to find the statistical significance of each test. A stringent correction based on the approximating inference model used in

EXHIBIT A1

Family-Wise Error Rate—A Hierarchical Clustering of Tests



our analysis would be $0.01/25 = 0.0004$ and corresponds to 22.22% probability of Type I error ($1 - 0.99^{25}$). The relevant t -statistic is ± 3.608 . This would correspond to a threshold Sharpe ratio of 0.2685. The adjusted p -value ($0.0025 = 0.01/4$), taking into account the four clusters, corresponds to a t -statistic of ± 3.066 . We use the FWER t -statistic to evaluate the allocation strategies. Based on these calculations, the average reciprocity drops from 0.49 to 0.2. All test results are available upon request.

ACKNOWLEDGMENT

We are grateful to the editor, Frank J. Fabozzi. We thank Otto Loistl for helpful comments.

REFERENCES

- Aldridge, I. 2019. "Big Data in Portfolio Allocation: A New Approach to Successful Portfolio Optimization." *The Journal of Financial Data Science* 1 (1): 45–63.
- Arnott, R., C. R. Harvey, V. Kalesnik, and J. Linnainmaa. 2019. "Alice's Adventures in Factorland: Three Blunders That Plague Factor Investing." *The Journal of Portfolio Management* 45 (4): 18–36.
- Arnott, R., C. R. Harvey, and H. Markowitz. 2019. "A Back-testing Protocol in the Era of Machine Learning." *The Journal of Financial Data Science* 1 (1): 64–72.
- Asness, C. S., T. Moskowitz, and L. H. Pedersen. 2013. "Value and Momentum Everywhere." *The Journal of Finance* 68 (3): 929–985.
- Baitinger, E., and J. Papenbrock. 2017. "Interconnectedness Risk and Active Portfolio Management: The Information-Theoretic Approach." *Journal of Network Theory in Finance* 3 (4): 25–47.
- Balasubramanian, K., H. Sundaresh, D. Wu, and K. Kearns. 2019. "Do Stocks Stalk Other Stocks in Their Complex Network? A Complex Network Approach to Stock Market Dynamics" *The Journal of Financial Data Science* 1 (2): 35–55.
- Bass, R., S. Gladstone, and A. Ang. 2017. "Total Portfolio Factor, Not Just Asset, Allocation." *The Journal of Portfolio Management* 43 (5): 38–53.
- Bender, J., T. Blackburn, and X. Sun. 2019. "Clash of the Titans: Factor Portfolios versus Alternative Weighting Schemes." *The Journal of Portfolio Management* 45 (3): 38–49.
- Bender, J., J. L. Sun, and R. Thomas. 2019. "Asset Allocation vs. Factor Allocation—Can We Build Unified Method?" *The Journal of Portfolio Management* 45 (2): 9–22.
- Bergeron, A., M. Kritzman, and G. Sivitsky. 2018. "Asset Allocation and Factor Investing: An Integrated Approach." *The Journal of Portfolio Management* 44 (4): 32–38.
- Bonacich, P., and P. Lloyd. 2001. "Eigenvector-like Measures of Centrality for Asymmetric Relations." *Social Networks* 23: 191–201.
- Borgatti, S. P. 2005. "Centrality and Network Flow" *Social Networks* 27: 55–71.
- Carhart, M. M. 1997. "On Persistence in Mutual Fund Performance." *The Journal of Finance* 52 (1): 57–82.
- Das, S. R. 2016. "Matrix Metrics: Network-Based Systemic Risk Scoring." *The Journal of Alternative Investments* 18 (4): 33–51.
- De Boer, S. 2016. "Smart Currency Hedging for Smart Beta Global Equities." *The Journal of Investing* 25 (4): 64–78.
- DeMiguel, V., L. Garlappi, and R. Uppal. 2009. "Optimal Versus Naïve Diversification: How Inefficient Is the 1/N Portfolio Strategy?" *The Review of Financial Studies* 22 (5): 1915–1953.
- Fabozzi, F. J., and M. López de Prado. 2017. "Who Needs a Newtonian Finance?" *The Journal of Portfolio Management* 44 (1): 1–3.
- Fabozzi, F. J., and M. López de Prado. 2018. "Being Honest in Backtest Reporting: A Template for Disclosing Multiple Tests." *The Journal of Portfolio Management* 45 (1): 141–147.
- Fama, E. F., and K. R. French. 1992. "The Cross-Section of Expected Stock Returns." *The Journal of Finance* 47 (2): 427–465.
- . 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33 (1): 3–56.
- . 2015. "A Five-Factor Asset Pricing Model." *Journal of Financial Economics* 116 (1): 1–22.
- Feng, G., S. Giglio, and D. Xiu. "Taming the Factor Zoo: A Test of New Factors." Working paper, Chicago Booth Research Article No. 17–04, 2019.

- Frazzini, A., and L. Pedersen. 2014. "Betting Against Beta." *Journal of Financial Economics* 111 (1): 1–25.
- Harvey, C. R., and Y. Liu. 2014. "Evaluating Trading Strategies." *The Journal of Portfolio Management* 40 (5): 110–118.
- James, G., D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. New York City, NY: Springer Science + Business Media, 2013.
- Kaya, H. 2015. "Eccentricity in Asset Management." *Journal of Network Theory in Finance* 1 (3): 1–32.
- Kinlaw, W., M. Kritzman, and D. Turkington. 2012. "Toward Determining Systemic Importance." *The Journal of Portfolio Management* 38 (4): 100–111.
- Konstantinov, G. 2017. "Currency Crowdedness Generated by Global Bond Funds." *The Journal of Portfolio Management* 43 (2): 123–135.
- Konstantinov, G., and J. Rebmann. 2019. "From Risk Factors to Networks: A Case Study on Interconnectedness Using Currency Funds." *The Journal of Financial Data Science* 1 (3): 108–123.
- Konstantinov, G., and M. Rusev. 2020. "The Bond–Equity Fund Relation Using the Fama–French–Carhart Factors: A Practical Network Approach." *The Journal of Financial Data Science* 2 (1): 22–44.
- López de Prado, M. 2016. "Building Diversified Portfolios That Outperform Out of Sample." *The Journal of Portfolio Management* 42 (4): 59–69.
- . 2018. "The 10 Reasons Most Machine Learning Funds Fail." *The Journal of Portfolio Management* 44 (5): 120–133.
- . 2019. "A Data Science Solution to the Multiple-Testing Crisis in Financial Research." *The Journal of Financial Data Science* 1 (1): 99–110.
- López de Prado, M., and F. J. Fabozzi. 2017. "Who Needs Newtonian Finance?" *The Journal of Portfolio Management* 44 (1): 1–3.
- Ma, C., B. Jacobsen, and W. Lee. 2019. "Pest Control: Eliminating Nuisance Allocations Through Empirical Asset Class Identification." *The Journal of Financial Data Science* 1 (3): 29–40.
- Madhavan, A., A. Sobczyk, and A. Ang. 2018. "What is in Your Benchmark? A Factor Analysis of Major Market Indexes." *The Journal of Portfolio Management* 44 (4): 46–59.
- Mantegna, R. N. 1999. "Hierarchical Structure in Financial Markets." *The European Physical Journal B: Condensed Matter and Complex Systems* 11 (1): 60–70.
- Markowitz, H. 1952. "Portfolio Selection." *The Journal of Finance* 7 (1): 77–91.
- Moskowitz, T., Y. Ooi, and L. Pedersen. 2012. "Time Series Momentum." *Journal of Financial Economics* 104 (2): 228–250.
- Newman, M. E. J. *Networks: An Introduction*. Oxford, UK: Oxford University Press, 2010.
- Pezier, J., and A. White. 2008. "The Relative Merits of Alternative Investments in Passive Portfolios." *The Journal of Alternative Investments* 10 (4): 37–39.
- Pojarliev, M. "Some Like It Hedged." CFA Institute Research Foundation, 2019.
- Prim, R. C. 1957. "Shortest Connection Networks and Some Generalizations." *Bell System Technical Journal* 37: 1389–1401.
- Provost, F., and T. Fawcett. *Data Science for Business*. O'Reilly Media Inc., 2013.
- Raffinot, T. 2018. "Hierarchical Clustering-Based Asset Allocation." *The Journal of Portfolio Management* 44 (2): 89–99.
- Sandhu, R. S., T. T. Georgiou, and A. R. Tannenbaum. 2016. "Ricci Curvature: An Economic Indicator for Market Fragility and Systemic Risk." *Science Advances* 2 (5): e1501495.
- Sharpe, W. 1994. "The Sharpe Ratio." *The Journal of Portfolio Management* 21 (1): 49–58.
- Simonian, J. 2019. "Proofs and Cross-Validations: Three Lessons for Financial Data Science." *The Journal of Financial Data Science* 1 (4): 1–7.
- Simonian, J., and C. Wu. 2019. "Minsky vs. Machine: New Foundations for Quant–Macro Investing." *The Journal of Financial Data Science* 1 (2): 94–110.
- Simonian, J., C. Wu, D. Itano, and V. Narayanam. 2019. "A Machine Learning Approach to Risk Factors: A Case Study Using the Fama–French–Carhart Model." *The Journal of Financial Data Science* 1 (1): 32–44.

Disclaimer

The email addresses are used by the authors for no other purpose other than to indicate their professional affiliations, as is customary in publications. The views and opinions in this article are those of the authors only and do not necessarily represent the official views and opinions of LBBW Asset Management or any of its affiliates or employees. LBBW Asset Management may or may not apply investment techniques or approaches of analysis similar to those described in this article.

To order reprints of this article, please contact David Rowe at d.rowe@pageantmedia.com or 646-891-2157.

ADDITIONAL READING

Big Data in Portfolio Allocation: A New Approach to Successful Portfolio Optimization

IRENE ALDRIDGE

The Journal of Financial Data Science

<https://jfds.pm-research.com/content/1/1/45>

ABSTRACT: In the classic mean-variance portfolio theory as proposed by Harry Markowitz, the weights of the optimized portfolios are directly proportional to the inverse of the asset correlation matrix. However, most contemporary portfolio optimization research focuses on optimizing the correlation matrix itself, and not its inverse. In this article, the author demonstrates that this is a mistake. Specifically, from the Big Data perspective, she proves that the inverse of the correlation matrix is much more unstable and sensitive to random perturbations than is the correlation matrix itself. As such, optimization of the inverse of the correlation matrix adds more value to optimal portfolio selection than does optimization of the correlation matrix. The author further shows the empirical results of portfolio reallocation under different common portfolio composition scenarios. The technique outperforms traditional portfolio allocation techniques out of sample, delivering nearly 400% improvement over the equally weighted allocation over a 20-year investment period on the S&P 500 portfolio with monthly reallocation. In general, the author demonstrates that the correlation inverse optimization proposed in this article significantly outperforms the other core portfolio allocation strategies, such as equally weighted portfolios, vanilla mean-variance optimization, and techniques based on the spectral decomposition of the correlation matrix. The results presented in this article are novel in the data science space, extend far beyond financial data, and are applicable to any data correlation matrixes and their inverses, whether in advertising, healthcare, or genomics.

Alice's Adventures in Factorland: Three Blunders That Plague Factor Investing

ROB ARNOTT, CAMPBELL R. HARVEY, VITALI KALESNIK, AND JUHANI LINNAINMAA

The Journal of Portfolio Management

<https://jpm.pm-research.com/content/45/4/18>

ABSTRACT: Factor investing has failed to live up to its many promises. Its success is compromised by three problems that are often underappreciated by investors. First, many investors develop exaggerated expectations about factor performance as a result of data mining, crowding, unrealistic trading cost expectations, and other concerns. Second, for investors using naive risk management tools, factor returns can experience downside shocks far larger than would be expected. Finally, investors are often led to believe their factor portfolio is diversified. Diversification can vanish, however, in certain economic conditions when factor returns become much more correlated. Factor investing is a powerful tool, but understanding the risks involved is essential before adopting this investment framework.

A Backtesting Protocol in the Era of Machine Learning

ROB ARNOTT, CAMPBELL R. HARVEY, AND HARRY MARKOWITZ

The Journal of Financial Data Science

<https://jfds.pm-research.com/content/1/1/64>

ABSTRACT: Machine learning offers a set of powerful tools that holds considerable promise for investment management. As with most quantitative applications in finance, the danger of misapplying these techniques can lead to disappointment. One crucial limitation involves data availability. Many of machine learning's early successes originated in the physical and biological sciences, in which truly vast amounts of data are available. Machine learning applications often require far more data than are available in finance, which is of particular concern in longer-horizon investing. Hence, choosing the right applications before applying the tools is important. In addition, capital markets reflect the actions of people, who may be influenced by the actions of others and by the findings of past research. In many ways, the challenges that affect machine learning are merely a continuation of the long-standing issues researchers have always faced in quantitative finance. Although investors need to be cautious—indeed, more cautious than in past applications of quantitative methods—these new tools offer many potential applications in finance. In this article, the authors develop a research protocol that pertains both to the application of machine learning techniques and to quantitative finance in general.