REPRINTED FROM
Risk.net
Risk.net October 2025

Supervised similarity for firm linkages

Supervised similarity for firm linkages

Ryan Samson, Adrian Banner, Luca Candelori, Sebastien Cottrell, Tiziana Di Matteo, Paul Duchnowski, Vahagn Kirakosyan, Jose Marques, Kharen Musaelian, Stefano Pasquali, Ryan Stever and Dario Villani introduce a novel proxy for firm linkages: characteristic vector linkages (CVLs). They use this concept to estimate firm linkages, first through Euclidean similarity and then by applying quantum cognition machine learning (QCML) to similarity learning. They demonstrate that both methods can be used to construct profitable momentum spillover trading strategies, but QCML similarity outperforms the simpler Euclidean similarity

The use of fundamental information as a proxy for firm linkages has been explored in the earlier literature. If investors have limited attention, then news affecting the price of a firm may only slowly be incorporated into the prices of related firms, leading to return predictability across firms. Indeed, for many such firm linkages, it has been shown that the lagged returns of a firm are predictive of the future returns for the firms most similar to it. This effect is sometimes referred to as the momentum spillover effect, or a lead-lag strategy. Momentum spillover has been documented for similarities formed from a variety of fundamental information including industry (Moskowitz & Grinblatt 1999), supply chain (Cohen & Frazzini 2008), analyst coverage (Ali & Hirshleifer 2020) and geography data (Pirinsky & Wang 2006).

Unrelated literature explores the application of machine learning techniques to learning similarity relations between securities, often with the goal of clustering securities for risk management, signal generation or portfolio construction (see, for example, the literature review in Salah & Hayette (2025) for examples of classification and clustering techniques).

Recent work has begun to explore the use of supervised learning techniques to extract similarity relationships. Jeyapaulraj *et al* (2022) proposed a supervised similarity framework based on extracting distance metrics via random forest, and demonstrated an application to clustering corporate bonds. Rosaler *et al* (2025) extended work on the corporate bond clustering problem, demonstrating the application of quantum cognition machine learning (QCML) (Musaelian *et al* 2024), a novel paradigm for both supervised and unsupervised learning tasks rooted in the mathematical formalism of quantum theory, to distance metric learning.

First, we introduce a novel proxy for firm linkages, which we name characteristic vector linkages (CVLs). A distance based on these linkages can be derived from the Euclidean distance of a vector of firm characteristics, which can then be used to form a profitable momentum spillover trading strategy. Next, we demonstrate how an approach similar to that discussed in Rosaler *et al* (2025) can be used to learn the distance relationships across equity securities, which can in turn be used to enhance the momentum spillover trading strategy derived from the simpler Euclidean distance.

Characteristic vector linkages and Euclidean distance

The academic and practitioner literature has identified a large number of stock characteristics that are widely used as cross-sectional predictors of risk or returns (see, for example, Jensen *et al* (2023), which evaluates the robustness of various features proposed in the financial literature, and also serves as an excellent survey of such features). The values of these characteristics, or factors, give us small but important pieces of information about the underlying firms. We propose that, across a broad set of characteristics, the more similar the factor scores of any pair of firms, the more economically similar the underlying pair of firms is, and the more likely they are to display momentum spillover effects. In other words, the similarity of sets of factor scores can be used to measure the strength of firm linkages. We refer to such purported factor- or characteristic-derived linkages as CVLs.

We denote a firm's characteristic observed at time t by $x^c_{t,j}$, where $c = 1, 2, \dots, C$ is the index of the characteristic and $j = 1, 2, \dots, J$ is the index of the firm, $\boldsymbol{x}^c_t \in \mathbb{R}^J$ is the vector over all firms for a single characteristic, $\boldsymbol{x}_{t,j} \in \mathbb{R}^C$ is the vector over all characteristics for a single firm, and $\boldsymbol{x}_t \in \mathbb{R}^{J \times C}$ is the matrix over firms and characteristics (we use similar index-omission notation throughout this article). The distance of factor scores can be defined in a variety of ways, but may simply and naturally be expressed as the Euclidean distance between two firm characteristic vectors. The Euclidean distance between firms i and j at time t is:

$$D_{\text{Euclid}}(\boldsymbol{x}_{t,i}, \boldsymbol{x}_{t,j}) = \sqrt{(\boldsymbol{x}_{t,i} - \boldsymbol{x}_{t,j})^{\mathrm{T}}(\boldsymbol{x}_{t,i} - \boldsymbol{x}_{t,j})} \tag{1}$$

We can then transform this to a similarity measure on $[0, 1]$, as will be explained later.

Quantum cognition machine learning distance

A review of QCML. QCML (Musaelian *et al* 2024) has been proposed as a new form of machine learning based on quantum cognition. QCML models learn a representation of the data in quantum states. Recall that in quantum mechanics a state is a vector of unit norm in a Hilbert space and is represented in bra-ket notation by a ket $|\psi\rangle$. The inner product of two states $|\psi_1\rangle$, $|\psi_2\rangle$ is represented by a bra-ket $\langle\psi_1|\psi_2\rangle$. The expectation value of a Hermitian operator O (ie, a quantum observable) on a state $|\psi\rangle$ is denoted by $\langle\psi|\, O\, |\psi\rangle$, representing the expected outcome of the measurement corresponding to O on the state $|\psi\rangle$.

For QCML, we use $|\psi\rangle \in \mathbb{C}^N$ (the dimension of the Hilbert space N is a hyperparameter of QCML), operators $O \in \mathbb{C}^{N \times N}$ and the standard Hermitian inner product, so kets correspond to column vectors, and bras to their conjugate transpose. Combinations of bras, kets and operators are then interpreted using standard matrix-multiplication rules. Hereafter, we omit the ket from ψ unless it is part of such a matrix-multiplication expression.

QCML starts by defining an error Hamiltonian over the data and observables. We have flexibility in how we define the Hamiltonian, as long as the result is non-negative and Hermitian. In this article we define the error Hamiltonian as:

$$H(\boldsymbol{x}_{t,j}, \{A_c\}) = \sum_c (A_c - x^c_{t,j} I)^2 \tag{2}$$

where $\boldsymbol{x}_{t,j}$ is as defined in the section on characteristic vector linkages and Euclidean distance, $\{A_c\} \in \mathbb{C}^{N \times N}$ is the set of C Hermitian observable operators that must be learned, and I denotes the $N \times N$ identity matrix.

In an unsupervised setting, training a QCML model involves iterative updates to the estimated observables $\{\hat{A}_c\}$ so that the ground states $\psi_{t,j}$ of the error Hamiltonian 'cohere' to the data and the overall energy $(E_0(\boldsymbol{x}_{t,j}, \{\hat{A}_c\}) = \langle\psi_{t,j}|\, H\, |\psi_{t,j}\rangle)$ is minimised.[1] Samson *et al* (2024) explains basic training and forecasting with this approach, while Candelori *et al* (2025) gives a detailed discussion of the energy and its decomposition into various components.

In the supervised setting, which is the primary focus of this article, the training process is slightly different. The target variable $y_{t,j} \in \mathbb{R}$ is assigned an N-dimensional quantum 'forecast' observable B, and given a data vector $\boldsymbol{x}_{t,j}$ the corresponding forecast is given by:

$$\hat{y}_{t,j} = \langle\psi_{t,j}|\, B\, |\psi_{t,j}\rangle \tag{3}$$

During the training process, the estimated quantum observables $\{\hat{A}_c\}$ and $\hat{B}$ are updated at each iteration so as to minimise the mean squared error $\sum_{t,j}(\hat{y}_{t,j} - y_{t,j})^2$. This can easily be extended to the case of multiple target variables, and we can also add to the loss any of the components of $E_0(\boldsymbol{x}_{t,j}, \{\hat{A}_c\})$ that we also wish to minimise. To obtain the results in this article we used:

$$\begin{aligned}\text{Loss}_{t,j} &= f(y_{t,j}, \boldsymbol{x}_{t,j}, \{\hat{A}_c\}, \hat{B}, w) \\ &= (\hat{y}_{t,j} - y_{t,j})^2 + w \sum_c (\hat{x}^c_{t,j} - x^c_{t,j})^2 \end{aligned}\tag{4}$$

where w is a hyperparameter influencing how strongly the ground states cohere to the input data. This form of the loss is not essential – for example, our empirical results in subsequent sections are similar, albeit slightly weaker, with $w = 0$ in (4) or with the unsupervised approach.

For results shown in this article, each operator was parameterised as a sum $O_1 + iO_2$, where O_1 (respectively, O_2) is a real symmetric (real antisymmetric) matrix. The dimension of the Hilbert space N is a hyperparameter of the algorithm. Larger values of N typically reduce the loss, but could lead to overfitting and worse out-of-sample performance, while lower dimensions tend to have higher bias and lower variance (Candelori *et al* 2025). The results in this article were obtained using $N = 12$.

■ **QCML distance.** As explained in the previous section, QCML represents each observation as a quantum state. This allows us to have a natural notion of proximity between observations, since the proximity between quantum states can be defined as the quantum fidelity:

$$f(\psi_1, \psi_2) = |\langle\psi_1|\psi_2\rangle|^2 \tag{5}$$

[1] *The ground state $\psi_{t,j}$ is the eigenstate associated with the lowest eigenvalue of the error Hamiltonian, and is also called the quasi-coherent state. Thus, the error Hamiltonian allows us to map the vector $\boldsymbol{x}_{t,j}$ to the ground state $\psi_{t,j}$, giving a representation of data in quantum states. Note that the QCML approach to mapping data into quantum states differs from that of quantum machine learning, where the mapping is given by a quantum circuit and the ability of that circuit to represent complex functions is important for the model's ability to learn. By contrast, in QCML, this mapping is learned from the data, and if the form of the Hamiltonian is not restricted, there is no restriction on the expressivity of the system.*

This can be interpreted as the probability of identifying state ψ_1 with state ψ_2 when performing a quantum measurement designed to test whether a given quantum state is equal to ψ_2 (or vice versa).

Fidelity can be transformed to distance in a number of ways. Here, we use the Bures distance:

$$\begin{aligned} D_{\text{QCML}}(\boldsymbol{x}_{t,i}, \boldsymbol{x}_{t,j}) &= \sqrt{2 - 2\sqrt{f(\psi_{t,i}, \psi_{t,j})}} \\ &= \sqrt{2 - 2|\langle\psi_{t,i}|\psi_{t,j}\rangle|} \end{aligned}\tag{6}$$

where the vectors $\boldsymbol{x}_{t,i}$ and $\boldsymbol{x}_{t,j}$ are mapped to ground states $\psi_{t,i}$ and $\psi_{t,j}$ as explained in the previous section. We chose the Bures distance for its similarity to the Euclidean distance, noting that all ground states ψ have unit norm. The empirical results were not particularly sensitive to this choice of distance. For example, we obtained nearly identical results using the geodesic distance, defined as $\arccos|\langle\psi_{t,i}|\psi_{t,j}\rangle|$. Fidelity can also be used directly as a similarity measure, rather than converting states to a distance and then converting distance to similarity. However, directly using these probabilities results in similarities that are too diffuse. Making these transformations allows us to form similarities with distributions comparable with those obtained using the Euclidean distance measure, allowing a fairer comparison of the two.

Converting distance to similarity

Given a notion of distance, we form a similarity matrix following a common transformation:

$$S(\boldsymbol{x}_{t,i}, \boldsymbol{x}_{t,j}) = \begin{cases} \mathrm{e}^{-\gamma D(\boldsymbol{x}_{t,i}, \boldsymbol{x}_{t,j})^2} & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Using this definition and our distance definitions in (1) and (6), we can compute $J \times J$ similarity matrixes by date for all i, j firm pairs.

Note that, in contrast to the Euclidean similarity between two vectors, the QCML similarity is a learned measure, since the representation of the data in quantum states ψ_t has been optimised using the training targets. Thus, in the context of QCML, the similarity measure is estimated on the data via the following steps:

(1) For C input characteristics, learn a set of operators $\{A_c\}$ on a set of training data, as described in the QCML review section.
(2) Compute $D_{\text{QCML}}(\boldsymbol{x}_{t,i}, \boldsymbol{x}_{t,j})$ as defined in (6). This requires first generating the error Hamiltonians $H(\boldsymbol{x}_{t,j})$ from (2) and finding the corresponding ground states $\psi_{t,j}$ as described in the QCML review section.
(3) Compute the $J \times J$ similarity matrix $S_{\text{QCML}}(\boldsymbol{x}_t)$ with elements defined in (7).

Our similarity matrix in (7) also requires a scaling parameter γ, which influences the concentration of the similarity values. For the results that follow, we use $\gamma_{\text{Euclid}} = 1$ when computing Euclidean similarity, after verifying that our results are not sensitive to this parameter and that a value of 1 gives close to optimal results. When computing QCML similarity, we choose $\gamma_{\text{QCML}} = 16$, which results in $\gamma_{\text{Euclid}} D_{\text{Euclid}}(\boldsymbol{x}_t)^2$ and $\gamma_{\text{QCML}} D_{\text{QCML}}(\boldsymbol{x}_t)^2$ having approximately matching median values over the training data.

CVL momentum spillover strategies

■ **Overview.** We now illustrate how to apply our concept of CVLs to form a profitable momentum spillover trading strategy, using an example set of characteristics. We first demonstrate the efficacy of the simple Euclidean similarity measure. We then show how QCML similarity can be used with the same input characteristics to learn similarity relationships across firms, which can in turn be used to enhance the momentum spillover trading strategy.

■ **Data and features.** We collected daily data on a dynamic set of US firms screened for the largest size, liquidity and maturity at each point in time from October 2017 to June 2024. The average number of names available per trading day in this universe is 1,500, but the set of names changes over time to reflect the most mature and liquid names available. This universe of securities is used for both model training and subsequent evaluation.

We create a broad set of features to be used as inputs for our characteristic vectors. We do not take a view on what characteristics are best suited for this task, but here we focus on accounting and valuation ratios, as such characteristics are widely used, easy to compute and fairly stable.

Accounting values are either averaged over the most recent four quarters or are taken as changes over the last four quarters, as applicable. All input characteristics are demeaned by GICS Industry Code to avoid characteristics linking firms based on industry biases. Characteristics are also cross-sectionally z-scored and winsorised at the 1st and 99th percentiles. We use market data sourced from Bloomberg and accounting data sourced from S&P Capital IQ. Our inputs (see Jensen *et al* (2023) or related literature for common definitions; exact definitions are not important to these results) include: Accounting liquidity, Accruals, Book-to-price, Cash-to-assets, Earnings-to-price, EBITDA-to-TEV, Implied equity duration, Leverage, Net leverage, Net operating assets, Operating cashflows, Operating efficiency, Profit margin, Real estate ratio, Revenue-to-price and Tax ratio.

We also create the following set of controls for our portfolio construction process, using market data sourced from Bloomberg as well as accounting and analyst ratings data sourced from S&P Capital IQ. All controls except GICS dummies are winsorised at the 1st and 99th percentiles.

■ **Analyst-connected stock momentum:** the momentum spillover signal formed on the basis of shared stock analyst coverage, following the approach in Ali & Hirshleifer (2020). We create five versions of this control, each with input returns matched to the horizon of the input returns to the CVL signal being tested (see the 'Momentum spillover signal construction' section for details of our CVL signal construction). All versions are demeaned by GICS Industry Code and cross-sectionally z-scored.

■ **Analyst coverage:** the number of analysts covering a stock, cross-sectionally z-scored (Parsons *et al* 2020).[2]

■ **Beta:** the rolling 252-day time series estimated Beta to S&P 500 Index.

■ **Momentum:** returns from 21 days ago to 252 days ago, cross-sectionally z-scored.

■ **Short-term reversal:** returns from 1 day ago to 100 days ago, exponentially weighted with a 10-day half-life, with returns within 1 day of earnings announcements replaced by beta-adjusted market returns, demeaned by GICS Industry Code, cross-sectionally z-scored.

■ **Size:** the log of Market Cap, cross-sectionally z-scored.

■ **Subindustry momentum:** the log of the Market Cap average weighted returns for the firm's GICS subindustry from 1 day ago to 200 days ago, exponentially weighted with a 20-day half-life, cross-sectionally z-scored.

[2] *As detailed in Parsons* et al *(2020), analyst coverage serves as an important proxy for investor attention, and some previously documented spillover signals are substantially weaker when controlling for analyst coverage.*

■ **Revenue-to-price:** Revenues, scaled by Market Cap.

■ **GICS dummies:** the dummy variables formed based on the GICS industry group.

■ **Similarity construction.** As explained in the section on converting distance to similarity, our Euclidean similarity measure is not a learned measure and requires no training. Given the inputs defined in the 'Data and features' section, we can directly compute $S_{\text{Euclid}}(\boldsymbol{x}_t)$ on each day for which the measure is needed during the test period (January 2014 to June 2024). In contrast, our QCML similarity measure requires first training a QCML model, as explained in the section on converting distance to similarity.

QCML is trained using daily data from October 2007 to August 2013. For model training, we are required to choose a target variable. Note that the trained QCML model can be used to directly forecast the chosen target using (3). However, in this article we are not interested in the efficacy of the direct forecast, and thus we do not include herein any analysis of its performance. Instead, we will show that the ground states $\psi_{t,j}$ can be used to create a superior distance measure using the technique outlined in the section on QCML distance.

It stands to reason that stocks that have similar future returns driven by observed characteristics are also likely to display stronger future comovement due to the similarity derived from those characteristics. Therefore, we choose as our target variable the 63-day forward returns, cross-sectionally z-scored. Other choices of target could be made. For example, future revenue/profitability growth or earnings surprises may more closely link stocks for fundamental reasons.

The inputs are defined in the 'Data and features' section, and are the same as those used for the Euclidean similarity measure. To create more robust similarity measures, we train 50 unique QCML models with varying seeds. Each seed randomly selects 10% of the available training names to use. We form five distinct subgroups of training dates (equally spaced), and seeds rotate through the subgroup selected for training. Signals produced from each seed are averaged with equal weight across all seeds. While QCML models can be updated online or through rolling/expanding retraining, for simplicity we keep the parameters static after the initial training. Periodic retraining or online updating of QCML parameters could lead to stronger results than those we present herein.

Our QCML models used a Hilbert space dimensionality of 12. The training loop and loss function described in the algorithm in the QCML review section were implemented in PyTorch, and models were trained using the Adam optimiser. All the results and figures for this article were obtained on a laptop equipped with a 32-core 13th Generation Intel Core i9-13950HX processor with 64 GB of memory, supplemented by an NVIDIA RTX 5000 Ada Generation Laptop graphics processing unit (GPU).

Training a single QCML model on the laptop GPU takes approximately 15 seconds for the data and hyperparameters we selected. Once we had successfully trained our QCML models on the training data, the QCML parameters remained static over the course of the test period (January 2014 to June 2024). As explained in the section on converting distance to similarity, we can then compute $S_{\text{QCML}}(\boldsymbol{x}_t)$ for each day the measure is needed.

Computing $S_{\text{QCML}}(\boldsymbol{x}_t)$ for the full test period, also leveraging the laptop GPU, takes approximately 50 seconds per QCML model. This means that, using only the laptop, all 50 seeds could be trained and the final signals constructed in approximately one hour. A cloud computing approach could run all seeds in parallel, providing a final signal in less than two minutes. In this article, we use 50 seeds for our final results, but 20 seeds gives sufficient convergence for hyperparameter exploration, further reducing the computational burden.

Momentum spillover signal construction. Once we have constructed a similarity matrix $S(\boldsymbol{x}_t)$, we can create a simple momentum spillover signal. The signal uses $S(\boldsymbol{x}_t)$ to construct firm linkage weights, and the lagged returns of linked firms as additional inputs. Let $\boldsymbol{r}_{t-l:t-1,i}$ represent the returns for firm i from $t-l$ to $t-1$ for some prior number of days l, and let $w_{t,j,i}$ represent the signal input weight from firm i for the signal to be computed for firm j on date t. We proceed as follows:

(1) Define $w_{t,j,i} = S(\boldsymbol{x}_t)_{j,i} / \sum_i S(\boldsymbol{x}_t)_{j,i}$.

(2) The spillover signal using input return days l on trading date t for firm j is $f_{l,t,j} = \sum_i (w_{t,j,i} \boldsymbol{r}_{t-l:t-1,i})$.

(3) We construct $f_{l,t,j}$ using both Euclidean and QCML similarity, for l equal to 21 days, 63 days, 126 days and 252 days. We also construct composite (over l) signals that average each of these four input returns, and we use the resultant $f_{\bar{l},t,j}$ as input to signal production.

(4) Each $f_{l,t,j}$ is individually demeaned by GICS Industry Code and cross-sectionally z-scored prior to signal evaluation.

Signal evaluation. To test performance, we use the covariance estimated daily from daily returns to produce Markowitz optimal investment portfolios,[3] with portfolio weight $w = V^{-1}Rf$ given covariance V, spillover signal forecast f and projection operator R,[4] to give risk optimal portfolio weights with no exposure to our controls.[5]

Investment portfolios are smoothed over 21 days to proxy for the fact that realistic investment portfolios must trade into new forecasts gradually. We compute daily returns to the smoothed portfolios, and analyse portfolio performance from January 2014 to June 2024. We do not remove the transaction costs, as these forecasts are not intended to represent a stand-alone investment strategy.

Results

Euclidean similarity CVL signals. We first examine the properties and performance of the Euclidean similarity CVL signals. Table A shows the average daily half-lives of the signals.[6] Relative to the horizon of the input returns, half-lives are shorter than would be expected from a simple average of returns, particularly for longer-dated input returns. This shows that linkages based on our accounting-data-focused inputs are fairly stable at a monthly horizon but are an important source of variation over longer horizons.

Table B shows the Sharpe ratios of portfolios formed from each of the signals. We plot the cumulative returns to the signals in figure 1. Recall, as explained in the subsection above, that our market-neutral investment portfolios are formed to have zero exposure to analyst-connected stock momentum, analyst coverage, beta, momentum, short-term reversal, size, subindustry momentum, revenue-to-price ratio and GICS industry groups, and thus these features make no linear contribution to returns.

The Euclidean similarity signals all show positive performance on average, although their performance significantly deteriorates in the final subperiod. We also note that the signals formed with longer input horizon returns perform substantially worse than the signals for short input horizon returns.

A. Average daily half-lives of the Euclidean similarity CVL signals with varying input return horizons

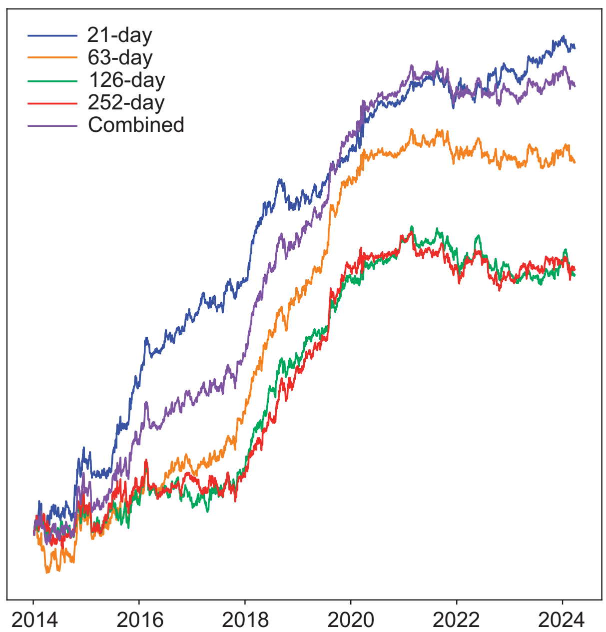
Period	21-day	63-day	126-day	252-day	Combined
Jan 2014–Jun 2024	9.1	19.5	27.0	36.3	14.8
Jan 2014–Jun 2017	9.2	19.7	26.9	35.6	14.9
Jul 2017–Dec 2020	9.0	19.3	28.2	40.3	15.0
Jan 2021–Jun 2024	9.0	19.4	25.9	33.5	14.5

B. Sharpe ratios of returns to the Euclidean similarity CVL signals with varying input return horizons

Period	21-day	63-day	126-day	252-day	Combined
Jan 2014–Jun 2024	1.35	1.03	0.71	0.73	1.24
Jan 2014–Jun 2017	1.60	0.53	0.28	0.29	1.02
Jul 2017–Dec 2020	1.63	2.50	2.00	2.00	2.41
Jan 2021–Jun 2024	0.61	−0.13	−0.38	−0.34	0.00

Strategy portfolios are market neutral and have zero exposure to analyst-connected stock momentum, analyst coverage, beta, momentum, short-term reversal, size, subindustry momentum, revenue-to-price ratio and GICS industry groups, and thus these features make no linear contribution to returns

1 Cumulative returns to the Euclidean similarity CVL signals with varying input return horizons



Strategy portfolios are market-neutral and have zero exposure to analyst-connected stock momentum, analyst coverage, beta, momentum, short-term reversal, size, subindustry momentum, revenue-to-price ratio and GICS industry groups, and thus these features make no linear contribution to returns. Returns have been scaled by the full sample realised volatility of each time series

[3] *Covariance is estimated using a technique proprietary to Duality Group, the details of which are not relevant to this article.*

[4] *If we wish to solve for portfolio weights that maximise expected returns, with a penalty for expected portfolio variance, while maintaining zero exposure to a set of controls, then for weights w, spillover signal forecast f, asset variance V, risk aversion μ and controls M, we need to solve for a w that minimises $-w^{\mathrm{T}} f + 0.5\mu w^{\mathrm{T}} V w$ such that $w^{\mathrm{T}} M = 0$. The solution is $w = V^{-1}Rf$, where $R = I - M(M^{\mathrm{T}} V^{-1} M)^{-1} M^{\mathrm{T}} V^{-1}$. Thus, R is a projection operator away from M, consistent with the desired investment process, while Rf is the residual from an inverse variance weighted regression of f on M.*

[5] *Given a matrix of control features M, we have $f^{\mathrm{T}} M = 0$ and $w^{\mathrm{T}} M = 0$.*

[6] *We define signal decay d as the average cross-sectional correlation of f_t with f_{t-1}. We approximate the half-life as $\ln 0.5 / \ln d$.*

C. Average daily half-lives of the QCML similarity CVL signals with varying input return horizons

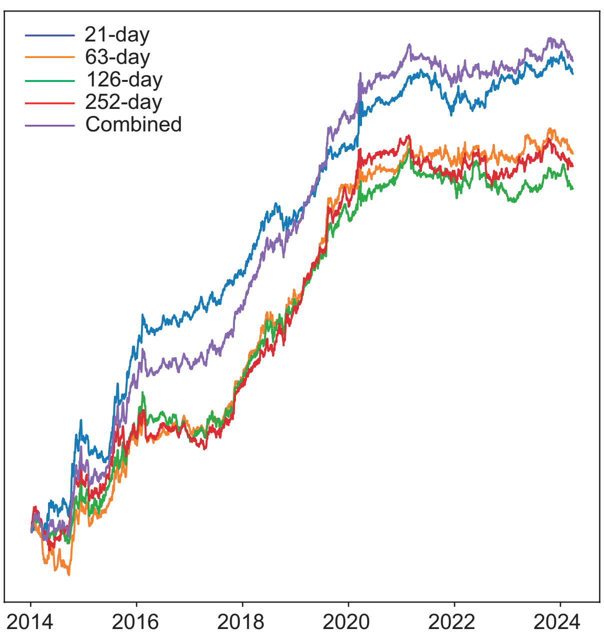
Period	21-day	63-day	126-day	252-day	Combined
Jan 2014–Jun 2024	11.5	32.6	53.5	90.9	23.5
Jan 2014–Jun 2017	12.5	34.9	54.6	90.2	25.0
Jul 2017–Dec 2020	11.5	33.3	61.5	120.5	25.4
Jan 2021–Jun 2024	10.6	29.7	46.0	72.2	20.5

D. Sharpe ratios of returns to the QCML similarity CVL signals with varying input return horizons

Period	21-day	63-day	126-day	252-day	Combined
Jan 2014–Jun 2024	1.38	1.14	1.04	1.10	1.42
Jan 2014–Jun 2017	1.84	0.89	0.88	0.78	1.47
Jul 2017–Dec 2020	1.68	2.14	2.14	2.48	2.31
Jan 2021–Jun 2024	0.28	0.14	−0.25	−0.28	0.08

Strategy portfolios are market neutral and have zero exposure to analyst-connected stock momentum, analyst coverage, beta, momentum, short-term reversal, size, subindustry momentum, revenue-to-price ratio and GICS industry groups, and thus these features make no linear contribution to returns

2 Cumulative returns to the QCML similarity CVL signals with varying input return horizons



Strategy portfolios are market neutral and have zero exposure to analyst-connected stock momentum, analyst coverage, beta, momentum, short-term reversal, size, subindustry momentum, revenue-to-price ratio and GICS industry groups, and thus these features make no linear contribution to returns. Returns have been scaled by the full sample realised volatility of each time series

■ **QCML similarity CVL signals.** We next examine the properties and performance of the QCML similarity CVL signals. In table C we show the average daily half-lives of the signals. The signal half-lives here follow a similar pattern to those obtained with the simpler Euclidean similarity in table A, but with the important distinction that the ones here are universally longer, and to a greater extent the longer the horizon of the input returns. In fact, the average half-life of the 252-day input return QCML similarity signal (90.9 days) is 2.5 times the half-life of the 252-day input return Euclidean similarity signal (36.3 days). This implies that, rather than adding instability from a highly parameterised learning approach, QCML is likely to have focused on the more resilient relationships that can be inferred from the input data.

Table D shows the Sharpe ratios of portfolios formed from each of the signals. We plot the cumulative returns to the signals in figure 2. Recall, as explained in the section on signal evaluation, that our market-neutral investment portfolios (like the Euclidean similarity investment portfolios) are formed to have zero exposure to analyst-connected stock momentum, analyst coverage, beta, momentum, short-term reversal, size, subindustry momentum, revenue-to-price ratio and GICS industry groups, and thus these features make no linear contribution to returns.

The performance is similar to that obtained with the simpler Euclidean similarity in table B (the averages of the daily cross-sectional correlations of the underlying signals range from 0.74 to 0.78). However, the signals derived from QCML similarity all outperform those from Euclidean similarity, and the extent of their overperformance increases as the horizon of the input returns increases. While the 21-day input return QCML only slightly outperforms the benchmark, the 252-day input return QCML signal shows a Sharpe ratio of 1.10 compared with 0.73 for the Euclidean signal (a greater than 50% improvement in the Sharpe ratio), despite the fact the QCML signal has a half-life (90.9 days) 2.5 times longer than the Euclidean signal (36.3 days). The QCML combined input return signal gives a Sharpe ratio of 1.42, an improvement over the Euclidean combined input return signal, which has a Sharpe ratio of 1.24, while the QCML signal has a half-life (23.5 days) nearly 1.6 times longer than the Euclidean signal (14.8 days).

In addition, the returns of most of the QCML similarity signals show somewhat better subperiod consistency than the Euclidean benchmark. These findings again imply that QCML is likely to have learned more resilient cross-firm relationships than those computed using the simpler Euclidean similarity. Note that to obtain the results in this article we trained QCML a single time on data prior to the test period. Periodic retraining or online updating of QCML parameters could lead to further improved results.

Conclusion

We first introduced a novel form of firm linkages, which we refer to as CVLs. We then demonstrated how both simple Euclidean similarity CVLs and QCML similarity CVLs can be leveraged to capture momentum spillover effects in equity markets.

We found that CVLs constructed from vectors of firm characteristics provide a meaningful basis for identifying economically linked firms, evidenced by the positive Sharpe ratios achieved across the various input return horizons and similarity types, displayed in tables B and D.

The Euclidean similarity approach, despite its simplicity, generates risk-controlled portfolio returns with Sharpe ratios ranging from 0.73 to 1.37 depending on the input horizon, with shorter-horizon signals generally outperforming longer-horizon variants.

Our supervised QCML similarity approach demonstrates meaningful improvements over the simple Euclidean method, particularly for longer-horizon signals. While the performance differences are modest for short-term signals, with 252-day input returns the QCML similarity measure achieves a Sharpe ratio of 1.12 compared with 0.76 for the corresponding Euclidean measure, while having a half-life (90.9 days) 2.5 times longer than that of the Euclidean signal (36.3 days). The combined QCML signal delivers a Sharpe ratio of 1.43, an improvement over the Euclidean combined signal's 1.26, while it has a half-life (23.5 days) nearly 1.6 times longer than that of the Euclidean signal (14.8 days).

This improved performance, as well as the considerably longer average half-lives observed for QCML signals, suggest that the supervised learning approach successfully identifies more robust and persistent firm relationships. In addition, to obtain the results in this article we trained QCML a single time, on data prior to the test period. Periodic retraining or online updating of QCML parameters could likely lead to stronger outperformance, especially during the most recent sub-period. There is no such possible avenue to improve the performance of the Euclidean similarity signals.

As discussed in detail in Candelori *et al* (2025), QCML learns non-commutative quantum models for the data manifold. These quantum models have the ability to abstract the fundamental features of the geometry of the data manifold in a manner that is robust to noise. QCML therefore offers an elegant framework for representing complex, nonlinear patterns in data and subsequently making more accurate measurements of distance between data points. We take advantage of QCML's abilities to ultimately learn relationships between firms that may not be captured by traditional distance metrics. By mapping firm characteristics to quantum states and leveraging quantum fidelity to derive similarity measures, QCML appears to uncover more accurate connections, enhancing the predictive power of momentum spillover strategies.

The work in this article could be extended in a number of ways. As noted in the 'Similarity construction' section, our QCML models are only trained a single time, on data available prior to the test period. A more dynamic approach allowing for periodic retraining or online updating of parameters may provide improved estimates of similarity. In addition, as mentioned in the 'Data and features' section, we did not take a view regarding which sort of firm characteristics are most likely to be meaningful for capturing CVLs. Future research could hypothesise economic mechanisms driving the ability of CVLs to capture firm linkages, and thereby refine the categories of input factors used or the target selected. The simplicity of the CVL approach makes testing such data variations trivial.

This work introduces CVLs as a promising new approach for capturing firm relationships and demonstrates the potential for QCML to more accurately measure distances between data points and thus capture similarity relationships. As markets continue to evolve and traditional strategies face increasing competition, the ability to identify and exploit subtle firm linkages through advanced similarity learning may become increasingly valuable for quantitative investment strategies. ■

Ryan Samson, Vahagn Kirakosyan and Luca Candelori are directors of research at Qognitive in Miami Beach, Florida, while Kharen Musaelian and Dario Villani are the chief science officer and the chief executive officer, respectively, there. Candelori is also associate professor in the mathematics department at Wayne State University, and Musaelian and Villani are, in addition to their Qognitive roles, chief investment officer and chief executive officer of the Duality Group, respectively. Ryan Stever, Jose Marques, Paul Duchnowski and Adrian Banner are, respectively, the chief investment officer, the chief executive officer, a portfolio Manager and senior researcher, and a portfolio manager and head of Australia at Intech Investment Management in West Palm Beach, FL. Tiziana Di Matteo ia a professor of econophysics at King's College London, a visiting professor at Centro Ricerche Enrico Fermi, and external faculty at the Complexity Science Hub Vienna. Stefano Pasquali is a senior vice president and the head of financial services solutions at Domyn in New York. Sebastien Cottrell is the global head of securitized products trading at Deutsche Bank, also based in New York. Collectively, they thank Hongli Lan for his contributions to this article. The views expressed here are those of the authors alone and not of Deutsche Bank, Domyn, Duality Group, Intech Investment Management or Qognitive. The contents of this article do not constitute investment advice.
Email: ryan.samson@qognitive.io, vahagn.kirakosyan@qognitive.io, luca.candelori@qognitive.io, kharen@qognitive.io, dario@qognitive.io, rstever@intechinvestments.com, jmarques@intechinvestments.com, pduchnowski@intechinvestments, abanner@intechinvestments.com, tiziana.di_matteo@kcl.ac.uk, stefano.pasquali@domyn.com, sebastien.cottrell@db.com.

REFERENCES

Ali U and D Hirshleifer, 2020
Shared analyst coverage: unifying momentum spillover effects
Journal of Financial Economics 136(3), pages 649–675

Candelori L, AG Abanov, J Berger, CJ Hogan, V Kirakosyan, K Musaelian, R Samson, JET Smith, D Villani, MT Wells and M Xu, 2025
Robust estimation of the intrinsic dimension of data sets with quantum cognition machine learning
Scientific Reports 15(1), article 6933

Cohen L and A Frazzini, 2008
Economic links and predictable returns
Journal of Finance 63(4), pages 1977–2011

Jensen TI, B Kelly and LH Pedersen, 2023
Is there a replication crisis in finance?
Journal of Finance 78(5), pages 2465–2518

Jeyapaulraj J, D Desai, P Chu, D Mehta, S Pasquali and P Sommer, 2022
Supervised similarity learning for corporate bonds using random forest proximities
Preprint, arXiv, https://arxiv.org/abs/2207.04368

Moskowitz TJ and M Grinblatt, 1999
Do industries explain momentum?
Journal of Finance 54(4), pages 1249–1290

Musaelian K, A Abanov, J Berger, L Candelori, V Kirakosyan, R Samson, J Smith and D Villani, 2024
Quantum cognition machine learning: AI needs quantum
Working paper, Qognitive Inc, available at www.qognitive.io/papers/QCML%20-%20Qognitive,%20Inc.pdf

Parsons CA, R Sabbatucci and S Titman, 2020
Geographic lead-lag effects
Review of Financial Studies 33(10), pages 4721–4770

Pirinsky C and Q Wang, 2006
Does corporate headquarters location matter for stock returns?
Journal of Finance 61(4), pages 1991–2015

Rosaler J, L Candelori, V Kirakosyan, D Mehta, K Musaelian, S Pasquali, R Samson and MT Wells, 2025
Supervised similarity for high-yield bonds
Risk.net, www.risk.net/7961606

Salah A and G Hayette, 2025
A meta-analysis of supervised and unsupervised machine learning algorithms and their application to active portfolio management
Expert Systems with Applications 271, article 126611

Samson R, J Berger, L Candelori, V Kirakosyan, K Musaelian and D Villani, 2024
Quantum cognition machine learning: financial forecasting
Risk.net, www.risk.net/7960053