Groß-Klußmann, Axel

Article — Published Version

Learning deep news sentiment representations for macro-finance

Digital Finance

Suggested Citation: Groß-Klußmann, Axel (2024) : Learning deep news sentiment representations for macro-finance, Digital Finance, ISSN 2524-6186, Springer International Publishing, Cham, Vol. 6, Iss. 3, pp. 341-377, https://doi.org/10.1007/s42521-024-00107-2

Digital Finance (2024) 6:341–377
https://doi.org/10.1007/s42521-024-00107-2

ORIGINAL ARTICLE

Learning deep news sentiment representations for macro-finance

Axel Groß-Klußmann[1]

Received: 19 September 2023 / Accepted: 17 February 2024 / Published online: 30 March 2024

Abstract
This paper introduces custom neural network techniques to the problem of latent economic factor extraction for voluminous news analytics data. In the context of macro-financial news, we derive low-dimensional representations of time series that arise in textual sentiment analyses spanning various topics. We explore three applications for compressed news sentiment data: nowcasting GDP growth, explaining asset class returns in a panel data analysis, and time series momentum investment. Our empirical study shows that nonlinear data representations based on supervised autoencoder architectures compare favorably to alternatives across all applications. In specific, we demonstrate that augmenting autoencoders with supervision tasks based on common asset class returns and market characteristics disciplines the dimension reduction and naturally supports the transparency of resulting representations. Taken together, our findings position supervised autoencoders as attractive competitor models alongside PCA and PLS approaches.

Keywords Neural networks · Sentiment analysis · Semi-structured data · Dimension reduction · Latent factor extraction · High-frequency macro-data

JEL Classification C45 · C55 · C58 · C38

1 Introduction

An ever-increasing number of data vendors in finance seek to capture salient features of news texts like the underlying author sentiment toward economic themes. The use of these investment signals based on news from macroeconomic, political, and financial contexts has become commonplace on financial markets. Alternative data products typically organize the sentiment signals in layered topic taxonomies and hierarchies corresponding to overarching narratives. By nature, many of the topics

✉ Axel Groß-Klußmann
axel.gross-klussmann@quoniam.com

[1] Research Department, Quoniam Asset Management GmbH, Frankfurt am Main, Germany

are short-termed and scarcely correlated among each other. This poses substantial challenges for subsequent data analyses as the dimensionality and heterogeneity of the topic space across vendors and taxonomy layers becomes huge. Extant dimension reduction methods often yield poor results against this backdrop. While data compression based on the conventional PCA lacks interpretability, methods enforcing sparsity tend to discard large chunks of the data. However, in contrast to 'hard' macroeconomic data primarily observed on monthly frequencies, country-level news sentiment data, for instance, can be aggregated to small, even intra-daily intervals. This leads to considerable increases in the data volume available for training statistical models. As a result, news sentiment lends itself naturally to answering in how far more expressive statistical approaches scale up to the complexity of news data.

This paper addresses the challenge of extracting and analyzing meaningful low-dimensional news sentiment representations for topics from macro-finance contexts. Our approach is inspired by macroeconomic factor models suggesting that a few common factors can explain the dynamics of multiple macroeconomic variables. In a similar way, the number of individual macro-financial topics in sentiment analyses is considerably larger than the number of major risks driving the markets. To capture complex data features in a compressed format, we aim to capitalize on the ability of neural networks to non-linearly encode news data in a lower dimensional space.

The pre-training of lower dimensional data representations with neural networks has long been known to improve the performance of subsequent supervised tasks (see Erhan et al. (2010)). More recently, Kelly et al. (2021) show that increasingly complex modeling approaches always benefit out-of-sample financial return forecast accuracy when optimal regularization is applied. Yet, little is known in the finance domain about the value of neural network-based sentiment representations employed in subsequent macro-financial applications like, e.g., nowcasting.

We contribute to a burgeoning literature based on harnessing financial news data to supplement hard macroeconomic data. Works like Calomiris and Mamaysky (2019), Ellingsen et al.(2022), Ter Ellen et al. (2021), and Thorsrud (2020) utilize monthly time series of news topics to model and forecast economic activity. However, these studies abstain from a distinct dimension reduction step for topics and instead employ statistical latent factor models highly customized to the problem at hand. Further, the data aggregation frequencies are monthly, thereby disregarding the precision and larger training data offered by higher frequencies. In contrast, our approach is predicated on the need for observed multi-purpose low-dimensional representations of daily news topic sentiment which can benefit multiple downstream tasks. To the best of our knowledge, we are the first to generate and analyze representations of macro-news sentiment obtained via custom artificial neural networks.

To extract economically meaningful sentiment representations, we take up the idea of leveraging external data as supervising variables to guide the data compression. The inclusion of outside data in dimension reduction is an established strategy to improve efficiency and interpretability of estimates of latent factors driving data dynamics. Prominent examples in this vein are given by the supervised PCA approaches proposed by Bair et al. (2006) and Giglio

et al. (2021) as well as the projected PCA of Fan et al. (2016). In an asset pricing context, Bybee et al. (2022) and Bybee et al. (2021) bring to bear the instrumented PCA (IPCA, see Kelly et al. (2019)) to connect the cross-section of returns to news topics.

Gu et al. (2020), Gu et al. (2021) and Spilak and Härdle (2023) show that autoencoders and, more generally, feedforward neural networks can be successfully employed in the analysis of financial market data. Along these lines, our study introduces supervised autoencoder networks put forward by Le et al. (2018) to the dimension reduction problem for news sentiment across multiple topics. We propose to add supervised learning losses based on both asset class returns and hard macroeconomic data to the reconstruction error loss of an autoencoder. Casting the dimension reduction problem in such a multi-task framework allows us to construct interpretable representations where individual dimensions correspond to major market risks. Further, blending different losses serves as another regularization of the complex model and can be expected to improve generalization.

Ultimately, we aim to answer two main questions.

1. Can we establish the empirical equivalence of models employing non-linear, neural network-based macro-news sentiment representations and those employing conventional linear data compression approaches?
2. In how far are results obtained for neural news representations robust and stable across different applications?

To this end, we explore the properties of daily news sentiment representations in three distinct applications for a panel of seven large geographical regions. First, we aim to use sentiment representations to explain daily returns in three major asset classes—equities, fixed-income, and currencies. Second, we use the representations to nowcast GDP growth in the seven regions considered. Third, given the trading objective inherent in directional sentiment signals, we employ the sentiment representations in a time series momentum exercise along the lines of Moskowitz et al. (2012). In all applications, we consider representations from several competing approaches. When linearity is imposed on the data projection, a natural benchmark to the supervised autoencoder is the partial least-squares (PLS, see Wold (1966)) approach. Letting go of the reconstruction error in the supervised autoencoder altogether leads to supervised feedforward networks as competing models. In terms of exclusively unsupervised approaches, we turn to the PCA as benchmark. Further, upon observing sample autocorrelation in time series of sentiment measures, we employ recurrent neural network encoders which can exploit this property.

Our study shows that non-linear low-dimensional sentiment representations encoded by neural networks outperform linear competitors across our three applications. We conclude that such data representations can indeed serve as multi-purpose variables to be employed in various subsequent supervised tasks: Among the neural representations, the supervised autoencoder yields most

favorable results across tasks. In the context of shallow supervised autoencoders, we demonstrate that a pre-training based on the reconstruction loss before training based on all losses sacrifices little of the variance explained while improving both the explanatory power and the interpretability of the representations. Taken together, our findings position supervised autoencoders as attractive latent factor extraction methods for news analytics data alongside the PLS and PCA.

The remainder of the paper is organized as follows. Section 2 describes the underlying data sets for sentiment and hard economic data. In Sect. 3, we provide details on the representation learning and extraction via neural networks and linear projectors. Section 4 characterizes representations and outlines results for the three applications. Section 5 concludes.

2 Data

2.1 Macroeconomic data

To use macroeconomic data as supervising variables in the representation learning, we collect economic data from the public vendors of the OECD (https://stats.oecd.org/) and the ECB (https://sdw.ecb.europa.eu). Our data cover the regions Canada (CA), China (CN), Europe (EU), Japan (JP), the United Kingdom (UK), and the United States (US).

We exclusively retrieve real-time vintage data to avoid potential look-ahead biases from data revisions common in macroeconomic data. The monthly data vintages contain data that were available up to the specific month. Combining the latest vintage data per month we construct real-time economic data for the following variables across the regions. We include the consumer price index, the harmonized unemployment rates, the industrial production index, the production in construction index, the index of retail trade volume, the international trades in goods (exports, imports), as well as the broad money monetary aggregate. Except for the case of China, these variables are available from 01/2001 onwards.

Our data processing consists of de-trending and de-seasonalization steps. In the de-trending step, we divide by the 6 months trend, while the de-seasonalization is based on dividing by the trailing mean per month. In addition, we standardize all variables. Next, to make sure all variables have the same directionality, we flip the sign of the unemployment variable and the broad monetary aggregate variable. All variables now share the same underlying economic rationale in that a positive value corresponds to a favorable macroeconomic environment. Finally, we construct compact macroeconomic factors per region by computing the first principal component (PC) of the variables. To avoid look-ahead biases, we chain together the PCs retrieved from an expanding window application of the PCA. To avoid arbitrarily flipping signs of the PCs, we fix the signs as discussed in Sect. 3.1.1. All of our macroeconomic variables have daily within-months publication time stamps. This allows us to derive daily time series of macro-factors from the daily time series with last observations carried forward to fill gaps.

2.2 Asset-specific time series

We aim to connect news sentiment to daily asset return series. For this purpose, we retrieve stock index and stock index futures returns for the ASX index (AU), the TSE60 index (CA), the Hang Seng index (representing China), the Eurostoxx 50 (EU), the TOPIX (JP), the FTSE 100 (UK), and the S &P 500 (US) from the Bloomberg data vendor. Further, we collect 10-year zero-bond yields and 10-year bond futures returns for the regions of interest with the exception of China. To cover the currency space, we retrieve FX forward returns for the free-floating currencies against the US dollar as quote currency. In specific, we use the crosses AUD/USD, CAD/USD, EUR/USD, JPY/USD, and GBP/USD. Alongside individual crosses, we further consider the Dollar index (DXY) measuring the generic strength of the US Dollar against a currency basket. All fixed-income and currency data stem from Refinitiv's datastream service. To be able to encode news sentiment as it concerns commodities, we resort to the Refinitiv/CoreCommodity CRB Index which tracks a weighted basket of commodity futures.

2.3 (Macro-financial) news sentiment data

Sentiment analysis utilizes Natural Language Processing techniques to infer the positive or negative tonality from textual information. We combine news for a global macroeconomic context from two prominent sentiment datasets based on financial (Ravenpack) and general news texts (GDELT). The two datasets are chosen to complement each other in that the Ravenpack dataset is targeted at a finance customer, while the GDELT data mainly focus on politics.

First, we use data from the commercial Ravenpack global macro-package (RPA 1.0, https://www.ravenpack.com/). Each row of these data contains sentiment scores, time stamp, entities mentioned, and further analytics for a news article. The news articles are assigned to the country/region they predominantly pertain to. Two types of main polarity scores, each with values in $[-1, 1]$, are given. The Event Sentiment Score (ESS) is attached to news that can be traced back to the so-called news events, defined as major events with a quick and broad real-time coverage. The Composite Sentiment Score (CSS) captures sources of news that often cannot be connected to specific events. An example of the latter concept is given by finance-related tweets on the Twitter platform. The themes underlying the news are organized in a topic hierarchy. We consider the two topmost aggregates of themes, called 'groups' and 'types'. Appendix A gives an overview of the individual themes in these two aggregates. Further, two relevance scores (relevance, event relevance) in $[0, 1]$ are given to measure the news article's relevance to an event as well as its relevance to the entities tagged. Our study is based on news with relevance and event relevance exceeding 0.7. The threshold is chosen to avoid both discarding too many news articles and including too many irrelevant ones. After filtering data for the countries/regions (AU, CA, CN, EU, JP, UK, US), we group the data by (day, region, topic-group)—and (day, region, and topic-type)—tuples and compute group-based averages of the

Table 1 Summary statistics for hard and sentiment macro-data in case of the US. Averages given except for max and min and PC1 − 3 columns. PC1 − 3 gives the explained variation of the first 3 principal components

	mean	std.dev.	max	min	ρ	AC1	AC2	AC3	PC1-3
Sent.	– 0.01	0.09	1.32	– 1.09	0.06	0.19	0.11	0.09	0.13
Macro	– 0.02	0.48	2.92	– 15.49	0.35	0.86	0.73	0.59	0.81

two sentiment scores. This procedure yields daily country-level time series of two sentiments (CSS, ESS) for news topics in two hierarchies (group, type). To form the EU-region, we group together both the European countries of the Eurozone and the Eurozone country aggregate already given in the data package.

Second, we retrieve data from the open access Global Database of Events, Language, and Tone (GDELT, see https://www.gdeltproject.org/). Similar to the Ravenpack product, the GDELT 'global knowledge graph' 2.0 data gives NLP-derived sentiment scores, time stamps, affected countries/regions, and themes mentioned for news articles. The data are crawled every 15 min from global news vendors. In our analysis, the GDELT sentiment score with values in $[-1, 1]$ is based on the Loughran and McDonald (2011) dictionary (LMCD henceforth) which represents an established source for sentiment analysis in a financial context. In contrast to Ravenpack's global macro-package, the GDELT data are not per se designed for a financial use case. To make the data more appropriate for our purposes, we require the news articles in the our GDELT dataset to touch economically and politically relevant themes. Our hand-selected themes for this purpose are described in appendix A. Ultimately, we group the data rows by (day, region, and theme)-tuples and compute group-based averages of the LMCD sentiment scores.

A combination of the two datasets after pre-processing results in daily time series for about 1,005 topics per region. The last data processing step consists of an expanding window demeaning applied to the final data frame. The data availability of the two sentiment datasets differs: While the RPA 1.0 data are available from 01/01/2001, the GDELT GKG 2.0 data stream starts in 03/01/2015. When combining both datasets, we zero-pad all missing information.

Table 1 gives descriptive statistics of both the news sentiment and hard macroeconomic data. The sheer breadth of financial topics in the news data is reflected in the low pairwise correlation as well as the low portion of variance explained by the first three principal components. Together with the moderate autocorrelation in news sentiment series, these empirical facts motivate the search for sophisticated representation learners encoding the data in lower dimensionality.

3 Methodology

In our study, we seek to compress daily sentiments for economic topics in seven regions (AU, CA, CN, EU, JP, UK, and US) to time series of five-dimensional representations each. The following subsections review unsupervised and supervised approaches to this problem. In the supervised setting, we focus on tasks where the sentiment representations are trained to explain (signs of) contemporaneous financial returns. The individual elements of supervised representations are based on the following considerations:

1. Supervision with 'r_t^{Eq}': One element of each representation will encode the region's sentiment data, such that it pertains to the corresponding stock index return. The idea is to represent news that primarily explain the daily stock market return.
2. Supervision with 'r_t^{FI}': Another representation element is supposed to encode regional news sentiment concerning daily changes in yields for ten-year zero bonds.
3. Supervision with 'Δf_t^{Macro}': We further encode the news sentiment to relate to monthly changes of the first principal component (PC) of our (hard economic) macro-variables (see 2.1). To match the frequency of the sentiment data, the monthly changes of the PC are computed daily.
4. Supervision with 'r_t^{FX}': A fourth element of the representation is trained to capture daily news sentiment toward a day's regional currency forward return. In case of the US, the dollar index return is used.
5. Supervision with 'r_t^{Cmdty}': To encode commodity-themed news, we utilize the return of the CRB commodity index for supervision.

While not exhaustive, the chosen variables provide a parsimonious yet comprehensive view of a day's market environment.

3.1 Linear dimension reduction for news sentiment

3.1.1 Principal component analysis

A workhorse of applied statistics, the PCA can be derived based on the singular value decomposition (SVD) of the $(n \times p)$ data matrix $\mathbf{X}$ after demeaning. $\mathbf{X}$ is mapped to a lower dimensionality $k < p$ via $\mathbf{X}_k = \mathbf{X}\mathbf{W}_k$, where $\mathbf{W}_k$ contains the k singular vectors corresponding to the k largest singular values.

A drawback of PCA implementations in empirical work lies in the fact that the signs of the singular value/vector pairs used for principal component construction are arbitrary. The SVD will often interchangeably output $\mathbf{W}_{\cdot i}$ or $-\mathbf{W}_{\cdot i}$ for the ith column of $\mathbf{W}$ in a rolling window application of the PCA. To align the signs of the principal components, we set $\mathbf{W}_{\cdot i} := -\mathbf{W}_{\cdot i}$ whenever the average of $\mathbf{W}_{\cdot i}$ is negative. In doing so, we impose a positive average weighting of the sentiment topics in each PC.

3.1.2 Projections based on partial least squares

Similar to the PCA, the partial least-squares (PLS) algorithm by Wold (1975) linearly maps the data matrix to lower dimensional scores $\mathbf{T}$ via $\mathbf{T} = \mathbf{XW}$, where $\mathbf{W}$ is of dimension $(p \times j)$, $j < p$. However, while the PCA represents a fully unsupervised approach to dimension reduction, PLS takes into account the relationship of predictors and target variables in the construction of the latent components $\mathbf{T}$. In this sense, the PLS mechanism differs from the PCA as the PCA only seeks to explain the variation in $\mathbf{X}$.

To project news sentiment such that each component in the final representation distinctly explains only one of our five financial market variables, we apply the PLS five times. In each run, $\mathbf{y}$ consists of one of the market variables and $\mathbf{X}$ is projected to one latent component. The final score $\mathbf{T}$ simply collects the five one-dimensional projections per data point.

3.2 Encoding news sentiment with neural networks

Some of the good performance of neural networks can be traced down to their ability to learn strong internal representations of data features (see Erhan et al. (2010)). Bengio et al. (2013) as well as Zhuang et al. (2015) note that particularly strong representations are found when label (supervision) information from multiple tasks can be used in the data encoding. In light of these findings, the following sections outline our approaches for recovering news sentiment representations from neural networks.

3.2.1 Feedforward neural networks

The feedforward network can be seen as a main building block of many neural network architectures (see Goodfellow et al. (2016) for details). Let $\mathbf{a}^{(1)}, ...\mathbf{a}^{(M)}$ denote the M hidden layers of a neural network, where each layer is comprised of $\dim(\mathbf{a}^{(l)})$ scalar hidden units.

Our analysis of feedforward nets is centered around the last hidden layer, $\mathbf{a}^{(M)}$, which we take to be the representation of the sentiment topic data. Figure 1 gives details for our feedforward architecture where the sentiment data are encoded to a 5-dimensional hidden layer where each hidden unit of $\mathbf{a}^{(M)}$ is connected to just one output. In its most general form, the output layer is 'fully connected' to the last hidden layer. However, allowing for all possible connections between $\mathbf{a}^{(M)}$ and outputs would interfere with the desired interpretability of the sentiment representations. To support interpretability, we instead operate in a multi-output setting and impose a single connection from every hidden unit in $\mathbf{a}^{(M)}$ to just one particular univariate output variable.[1]

[1] To give a concise description of the architecture's layer structure in terms of dimensions, we write $(p, [\dim(\mathbf{a}^{(1)}), \dim(\mathbf{a}^{(2)}), ...], \dim(\hat{\mathbf{y}}))$, where p is the input dimension.

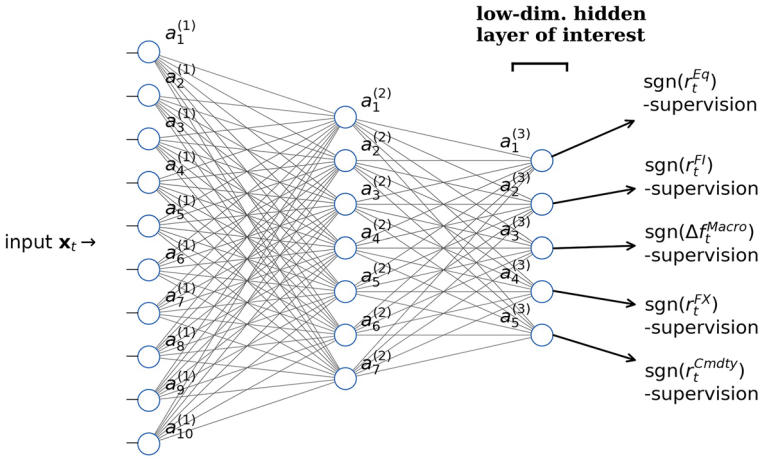

Fig. 1 The architecture of an example feedforward ANN with financial market supervision losses for the hidden representation of interest. Blue circles denote hidden units; black lines represent weighted connections of the data path through the network. Own illustration on top of blank ANN from http://alexlenail.me/NN-SVG/index.html

In classification tasks, training of the neural network weights, $\boldsymbol{\theta}$, is based on minimizing the cross-entropy between training data and model distribution. Let $y_i \in \{0, 1\}$ be the ith output label and let t denote the example time. The cross-entropy loss over all training examples $\mathbf{x}_t$ is given by

$$J_{ce}(\boldsymbol{\theta};\boldsymbol{\lambda}) = -\frac{1}{n} \sum_{t:\mathbf{x}_t \in \mathbf{X}} \sum_{i=1}^{\dim(a^{(M)})} \lambda_i \big(y_{it} \ln \widehat{p}_{it} + (1 - y_{it}) \ln(1 - \widehat{p}_{it}) \big), \tag{1}$$

where the ith component of $\boldsymbol{\lambda}$, and λ_i, is the weight of the ith supervision task.

Our specific architecture design (Fig. 1) has the following implications. First, only lower level layers share information on all losses. In consequence, hidden units of the layer of interest $\mathbf{a}^{(M)}$ are predominantly characterized by their single relation to the single output they are connected to. Second, we could in principle utilize the outputs $\widehat{\mathbf{y}}$ directly as representations. However, our empirical work shows that adding more losses and tasks directly to the output layer often results in worse training outcomes than attaching these losses to hidden and outcome layers separately. Our feedforward encoder can conveniently be expanded to include a reconstruction loss, for instance.

3.2.2 (Supervised) autoencoders

Autoencoder (AE) networks are a special case of (unsupervised) feedforward networks trained to copy the input data to the output. In the spirit of dimension reduction, so-called undercomplete autoencoders are designed to learn internal representations with a lower dimensionality than the original input.

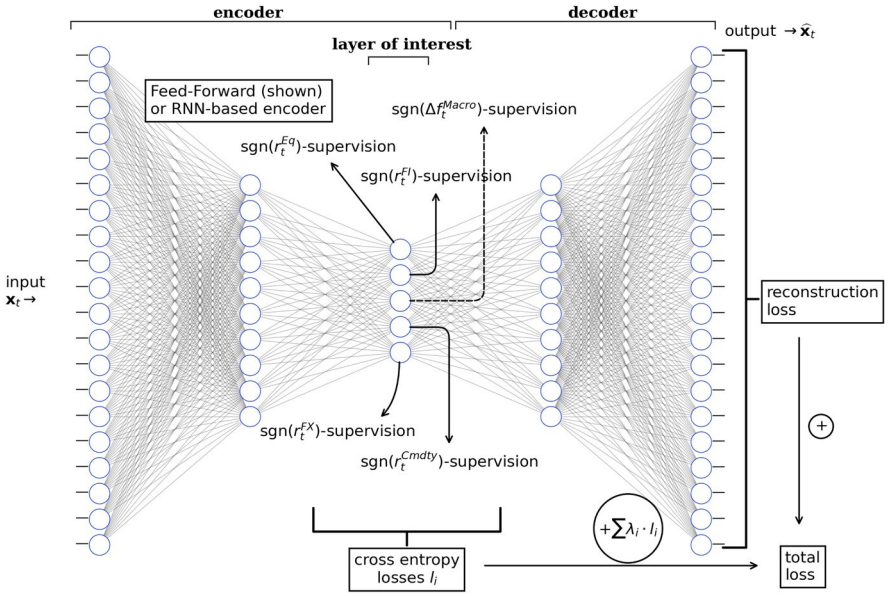

Fig. 2 The architecture of an autoencoder with supervision losses for the hidden representation of interest. Arrows and text fields highlight the position of the regularizing losses. Own illustration with blank background ANN constructed at http://alexlenail.me/NN-SVG/index.html

The architecture of an undercomplete autoencoder consists first of a so-called encoder network, mapping the input to a lower dimensional hidden layer of interest, the bottleneck layer, $\mathbf{a}^{(k)}$, $k < M$. The bottleneck layer is followed by a decoder network, mapping the low-dimensional representation back to the output (reconstruction). Figure 2 exemplifies a standard symmetric feedforward architecture underlying an undercomplete autoencoder. In the depicted design, the feedforward encoder part consists of a mapping of the input data $\mathbf{x}$ to an initial 20-dimensional hidden layer, followed by a 10-dimensional hidden layer. The output of the encoder is the 5-dimensional bottleneck $\mathbf{a}^{(k)}$ to be extracted as the lower dimensional representation of the input. Next, the 5-dimensional $\mathbf{a}^{(k)}$ is decoded by another feedforward network (decoder) to ultimately reconstruct the input as $\hat{\mathbf{x}}$. The decoder's architecture mirrors the encoder's architecture.

Training of undercomplete autoencoders minimizes the reconstruction cost based on the mean squared error between input and reconstruction as

$$J_r(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t:\mathbf{x}_t \in \mathbf{X}} \frac{1}{K} \|\hat{\mathbf{x}}_t - \mathbf{x}_t\|_2^2. \tag{2}$$

In their work on supervised autoencoders (SAE henceforth), Le et al. (2018) propose to add supervision losses to the bottleneck layer of an undercomplete

autoencoder. The bottleneck layer now produces both the input to the subsequent decoder and feeds an output layer where predictions $\hat{\mathbf{y}}$ are formed.

In analogy to the PLS, the two loss components of the supervised autoencoder promise benefits to our analysis of high-dimensional sentiment topic data. First, training the reconstruction loss will help to encode general underlying patterns of the data regardless of subsequent prediction studies. Second, training the supervision tasks helps to guide the encoding, such that the elements of the representation pertain to common financial market characteristics each.

Training of SAEs will optimize both the reconstruction cost and the cross-entropy losses (1). Given fixed weightings λ_r, λ for the reconstruction loss and the cross-entropy losses, the total cost is

$$J(\boldsymbol{\theta}) = J_{ce}(\boldsymbol{\theta};\lambda) + \lambda_r \cdot J_r(\boldsymbol{\theta}). \tag{3}$$

As per Le et al. (2018), there are theoretical guarantees that the SAE approach yields superior generalization compared to training either part of the cost alone.

We study two encoder architectures of the SAE.

1. SAE (FF): An SAE with feedforward encoder. This is depicted in Fig. 2.
2. SAE (GRU): Next to feedforward encoders, we consider recurrent neural networks (RNN) as encoder architecture in the SAE. RNNs like the long short-term memory network (LSTM) introduced in Hochreiter and Schmidhuber (1997) as well as the gated recurrent unit (GRU), see Chung et al. (2014), can pick up temporal dependence in data. Our study focuses on the GRU which shares many properties of the LSTM but exhibits a lower number of weights. To employ the GRU as encoder in the SAE, we configure its hidden state to have five dimensions. Visually speaking, the hidden state will serve as the bottleneck layer ('layer of interest') in Fig. 2.

3.3 Model training and specification details

3.3.1 Artificial neural networks

The neural network models are trained via stochastic gradient descent in its 'Adam' variant; see Kingma and Ba (2014). While the combination of reconstruction and supervision losses already introduces regularization, we make use of further techniques to mitigate overfitting risks.[2] Based on validation set insight, we employ:

- Dropout, i.e., zeroing out nodes during training, with probability 0.2.
- Early stopping: All networks are trained for a maximum of 30 epochs, i.e., complete passes over the randomly shuffled data.
- Weight initialization: uniform.

[2] A validation data set reveals widely detrimental effects of L2 losses (weight decay) as well as batch normalization techniques. We do not consider these henceforth.

- Learning rate: 0.005.
- Mini-batch sizes: 50.
- Activations: leaky ReLU.
- For the SAE(GRU) we use a GRU with one recurrent layer, a hidden layer size of 5, and an input sequence length of 21 days.

Training of the supervised autoencoder poses a particular challenge in that two different types of losses are to be minimized. To address the problem of training both unsupervised and supervised losses, we follow Erhan et al. (2010) as well as Bengio et al. (2013) and split the training process into two phases: An unsupervised pre-training for some epochs (runs) will train the network based on the unsupervised loss alone. The subsequent supervised fine-tuning takes the pre-trained network and adds the supervised loss to the unsupervised loss during the last training epochs. We reserve the first ten epochs to train the reconstruction loss alone, while the latter 20 epochs additionally train the cross-entropy losses. The weights of the losses for the SAE, i.e., the parameters λ_i, $(i = 1, \dots, 5)$, and λ_r in equation 3 are calibrated with the validation set. Upon observing little differences for grids of parameters, we set the parameters to yield comparable magnitudes of individual losses. This means $\lambda_i = 0.05$ for all i and $\lambda_r = 10$.

An important decision concerns the choice between regression and classification. Joint training of both a reconstruction error and several regression tasks for the supervised autoencoder proved challenging on the validation data set. For this reason, we complement the reconstruction loss of the autoencoder with classification losses. This is also the approach taken in Le et al. (2018). To be consistent across neural network models, the feedforward network is based on classification tasks, too.

3.3.2 Linear approaches

In case of the PLS, we observed stronger validation data results for the regression setting compared to a PLS discriminant analysis for classification. As a direct consequence, the PLS is based on a regression loss.

3.4 Experimental design

3.4.1 Rolling window representation retrieval

We split the data into training and test sets to form time series of out-of-sample sentiment representations. The model training takes place in a rolling fashion every half year, with the first training data set given by the data between 01/01/2001 and 06/30/2003. In each run, the training (and testing) endpoints are moved forward in time by half a year. We expand the training data by half a year until a maximum of 7 year training data is covered. The hyperparameters and architectures of our models are calibrated on a global validation set on the window 07/01/2003 through 12/31/2003. Figure 3 depicts the training, validation, and testing windows used in our analysis. After training a model up until the end of a half year, the trained model

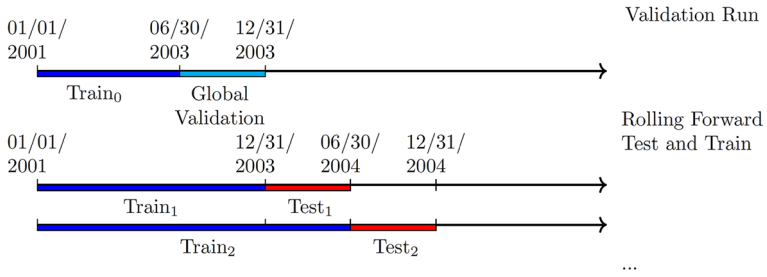

Fig. 3 The rolling window setup used in our study

with parameters $\widehat{\theta}(\text{Train}_k)$ is applied to the next half year as test data. Time series of sentiment representations are formed via the concatenation of the model output on the semi-annual test data sets post the validation set, ranging from 01/01/2004 to 12/31/2021. With training and test sets defined in Fig. 3, time series of the five-dimensional sentiment representations are given as

$$\mathbf{s}_t = f(\mathbf{x}_t; \widehat{\theta}(\text{Train}_k)), \qquad t \in \text{Test}_k, \quad k = 0, ..., 36. \tag{4}$$

The functional form f is determined by the model considered, i.e., will be a linear mapping for the PCA and PLS. In case of the neural networks, f outputs the hidden layer of interest in a chain of nonlinear transformations.

3.4.2 Comparing models with sentiment representations

In the following, let i be the index of the representation learning approach and let $\mathbf{s}_t^{(i)}$ denote the corresponding representation. We explore three applications for models employing the different representation variants. Each of the applications implies a distinct loss function in terms of the individual representation time series, $L_{i,t} := L(\mathbf{s}_t^{(i)}, \widetilde{\mathbf{y}}_t, \widetilde{\mathbf{x}}_t, ...)$, where $\widetilde{\mathbf{y}}_t$ and $\widetilde{\mathbf{x}}_t$ are application-specific dependent variables and controls. The applications and their losses are defined as follows.

1. We analyze the in-sample fits of asset return panel regressions with sentiment representations as regressors. Here, the squared residuals represent the loss.
2. We consider a performance evaluation of time series momentum investment strategies based on $\mathbf{s}_t$. In this case, the loss function is taken to be the negative difference of a strategy's return series to a passive benchmark return.[3]
3. Our third exercise reviews GDP growth nowcasting with sentiment representations. We specify the loss function as the squared forecast errors.

To tackle our first research question about the equivalence of models with variants of learned sentiment representations, we turn to the 'model confidence set'

[3] A related loss function evaluation can be found in Aparicio and López de Prado (2018).

Table 2 Overview of model identifiers

Identifier	Description	Loss types	Architecture
SAE(FF)	Sup. AE; FF encoder	reconstr.+class.	(1005, [10, 5, 10], 1005)
FF	Feedforward classifier	classification	(1005, [10, 5])
SAE(GRU)	Sup. AE; GRU encoder	reconstr.+class.	(1005, [GRU[5], 10], 1005)
PLS	Partial Least Squares	regression	-
PCA	Principal Components	reconstr.	-

(MCS) procedure of Hansen et al. (2011). Similar to the SPA test of Hansen (2005), the MCS approach addresses data snooping issues (see White (2000)) related to multiple model comparisons. However, two advantages make the MCS compelling for our study. For one, the MCS procedure does not require to formulate benchmark models. Second, in contrast to the SPA testing, the MCS approach can be used for model selection.

Let the model involving the ith representation variant be identified with i. The MCS framework iteratively evaluates two models from our model candidate set $\mathcal{M} = \{1, 2, 3, 4, 5\}$ in terms of their relative loss performances $d_{ij,t} := L_{i,t} - L_{j,t}$, $i, j \in \mathcal{M}$. Given a size of the test, the final MCS is defined as

$$\mathcal{M}^* := \{i \in \mathcal{M} : \mathbb{E}(d_{ij,t}) \leq 0 \text{ for all } j \in \mathcal{M}\}. \tag{5}$$

Hansen et al. (2011) devise a sequence of tests which eliminate the worst model at each step. As a result, $\mathcal{M}^*$ contains only models that have not been found inferior to another model and can be considered statistically equivalent.

4 Results

Our analysis encodes the daily sentiment data into time series of five-dimensional representations for each of the seven countries/regions (AU, CA, CN, EU, JP, UK, US). Table 2 gives an overview of the encoding approaches considered and the corresponding identifiers used henceforth. We provide additional details on the design of the neural networks in appendix B, which also contains a link to the implementation code repository.

The vectors $\mathbf{s}_t$ are created in rolling train-test splits to avoid including hindsight information (see Eq. (4)). Individual elements of representation vectors from approaches with supervision losses can be interpreted as encoding the five financial dimensions outlined in section 3. After characterizing the representations, we first explore their explanatory power for daily asset returns in an *in-sample* regression analysis. Second, a time series momentum strategy attempts to exploit the forecasting power of $\mathbf{s}_t$ for investment on the daily data sample from 1/2004-12/2021. Finally, we utilize the $\mathbf{s}_t$ to nowcast quarterly GDP

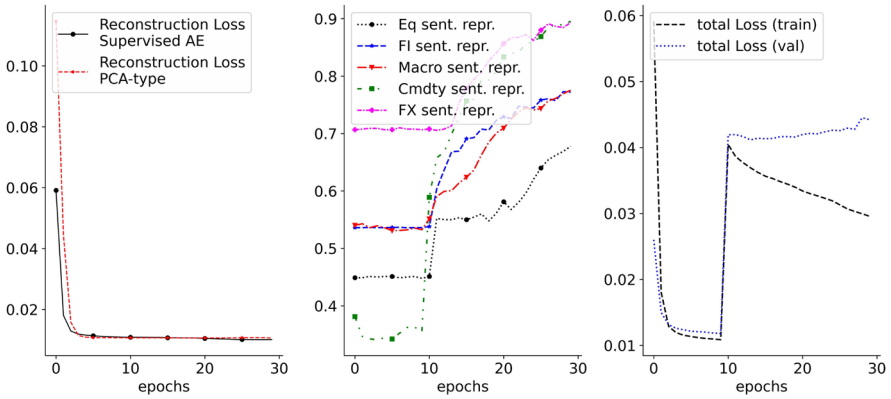

Fig. 4 Loss analysis of the first training run of the supervised autoencoder in the US case. Left plot: reconstruction loss evolution of the supervised AE vs. shallow PCA-type AE (train). Middle: accuracy evolution of hidden representations (train). Right: total loss evolution (train and val)

Table 3 Contemporaneous $t \leftrightarrow t$ accuracies and correlations (in brackets) between representation elements and their targets. Averages across regions shown for the supervised AE compared to the PLS

Model	Eq repr.	FI repr.	Macro repr.	Cmdty repr.	FX repr.
PLS	0.6 (0.24)	0.6 (0.25)	0.47 (0.19)	0.58 (0.23)	0.57 (0.37)
SAE(FF)	**0.67** (**0.34**)	**0.67** (**0.33**)	**0.58** (**0.22**)	**0.66** (**0.33**)	**0.76** (**0.49**)

Highest values are in bold

growth. This particular analysis is based on a static train-test split where time series models are trained and calibrated in the quarters Q1/2004–Q4/2010 and evaluated on Q1/2011–Q4/2021.

4.1 Characteristics of news sentiment representations

For ease of exposition, the following descriptive results concentrate on the the US region and the SAE(FF) as core architecture.

Figure 4 shows the loss evolutions as well as the training accuracies of the individual (US) representation components when explaining their corresponding target. The equivalence of the PCA to a shallow AE (see Bourlard and Kamp (1988)) allows to compare the reconstruction loss of the supervised AE to a PCA-type reconstruction loss. In specific, the first panel of Fig. 4 gives the loss evolution of a (1005, [5], 1005) AE with identity activation in the bottleneck layer. Interestingly, we observe that the reconstruction loss of the supervised AE closely matches the PCA-type loss even when classification tasks are added after 10 epochs. We conjecture that the reconstruction task still leaves enough leeway

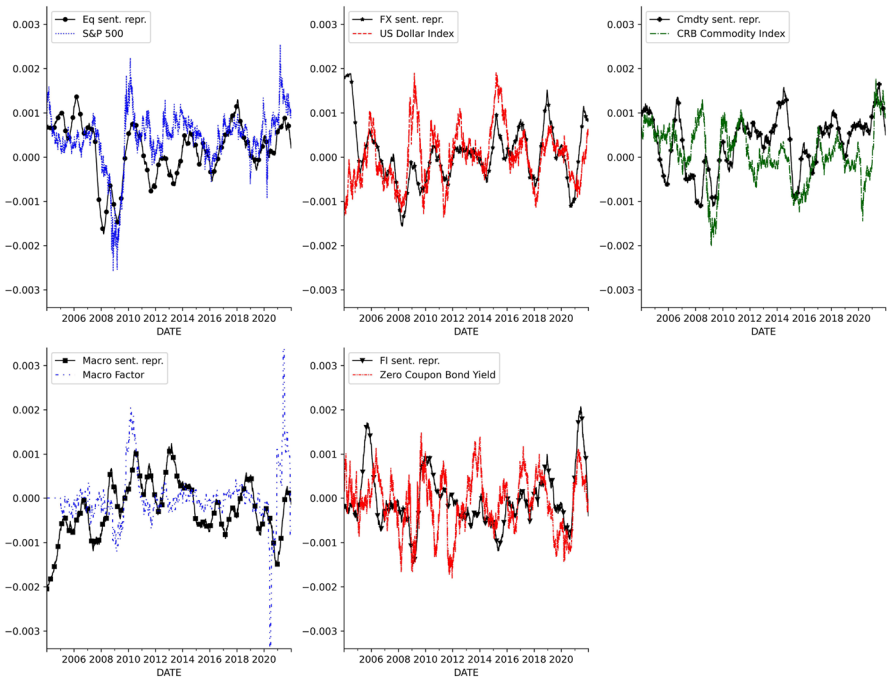

Fig. 5 Final representations of the supervised AE plotted against their supervision target in the US case. 252-day averaging applied

for training the classification tasks while sacrificing little of the reconstruction quality.

Table 3 gives full sample accuracies and correlations of the individual elements of the sentiment representations for the supervised AE and PLS representations. For both the correlations and the accuracies, the supervised AE representations surpass the PLS-based representations in all market dimensions covered. This observation is noteworthy, given that the supervision type is different in both approaches: regression for PLS and classification for the supervised AE.

To give a qualitative understanding of the representation, Fig. 5 traces out the time series of sentiment representations for the US against the supervision target. We observe that the representations closely track their hard economic data theme. This behavior is most pronounced during large market swings.

Next to the visual inspection, we rely on the SHAP (SHapley Additive exPlanations) approach put forward in Lundberg and Lee (2017) to explain the output of the supervised AE. SHAP has roots in cooperative game theory and gives explanations of model outputs based on Shapley values (Shapley (1953)). The approach casts the output generation of a model as a game where features act as players contributing to the output value. Computed in different coalitions of players, i.e., features, Shapley values per feature give the average contribution of a feature value to the total prediction over the average prediction. The attractiveness of using Shapley values for model output interpretation lies in the sound theory underlying it.

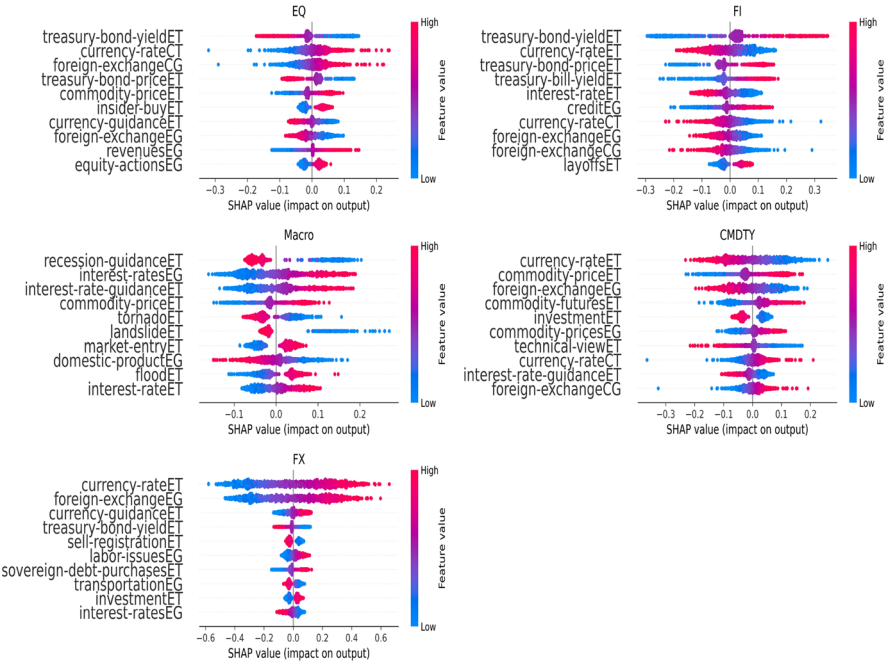

Fig. 6 Plots show SHAP contributions for ten most important (mean absolute SHAP value) features for individual data points. The results are given for 1000 random samples drawn from the sentiment data covering 01/2004-12/2021. US case shown

The panels in Fig. 6 show the Shapley values per daily data point for the five sentiment representation themes in the US case. In each panel, the ten most important features in terms of the mean absolute Shapley value are selected. The SHAP results for each of the sentiment representations Eq, FI, FX, Cmdty, and Macro reveal that the individual representations can be traced back to sentiments for corresponding economic topics. While some topics occur in each representation's top ten list, individual topics exclusive to a representation match its theme well: Eq. representation for instance draws on equity actions, revenues, and insider buying topics, while FI distinctively depends on treasury bill yield and credit topics. Naturally, the Cmdty representation exclusively loads on commodity futures and FX on currency guidances. Among the ten most important topics for the Macro representation, we find recession guidance, domestic product, as well as nationwide disaster topics that can be linked to the overarching theme.

SHAP also allows to assess the directional impact of features per data point on the output. However, in contrast to the absolute importance, the directionalities in general do not lend themselves to straightforward interpretations. Still, for the idiosyncratic factors exclusive to each representation theme, we find that higher sentiments correspond to a rising supervision variable. To give an example, higher sentiments for the commodity prices topic corresponds to a positive return in the commodity index. For the FI theme, our sample supports a negative correlation

Table 4 Effects of the reconstruction loss regularization in terms of the standard deviation of the SHAP values per representation element (columns 1-5) and the standard deviation of the complete ANN weights (last column)

sd()	Eq	FI	Macro	Cmdty	FX	weights
FF Class.	2.36	2.54	2.65	2.65	2.36	47
SAE(FF)	1.80	1.66	2.61	1.41	2.18	45
GRU	0.12	0.07	0.03	0.10	0.10	58
SAE(GRU)	0.08	0.08	0.03	0.05	0.03	44

between bonds and stocks, such that positive sentiments, i.e., a friendly stock market environment tends to coincide with rising yields. Interestingly, positive sentiments for topics like recession and domestic product have a negative impact on the macro-themed representation and are thus at odds with economic rationale. We attribute this observation to the frequency mismatch of the supervision variable (the monthly hard-data macro-factor) and the daily sentiment data.

The SHAP values and weights of the ANN present an opportunity to highlight the effect of adding the reconstruction loss. Table 4 gives standard deviations for architecture variants with and without reconstruction loss. In general, we observe a lower variance in weights and SHAP values when adding the reconstruction loss. This finding indicates that the autoencoder-based regularization indeed discourages extreme weights and hence adds stability.

4.2 Explaining short-term financial market returns

In the first application, we employ the news sentiment representations in a cross-sectional study of asset return time series. We report panel regression results for three dependent variables on a daily frequency covering 01/01/2004-12/31/2021: stock market index returns (Eq), daily 10-year government bond yield changes (FI), and foreign exchange forward (FXF) returns. We consider the panel of 7 regions/countries in case of stock returns as dependent variable (AU, CA, CN, EU, UK, JP, US). For FXF returns, the panel reduces to 5 regions as we measure currencies against the US dollar as base and the Chinese Renminbi cannot be considered free-floating in our sample. In the FI case, we drop China from the panel due to the impaired free tradability of its main government bonds.

The Hausman test points to a random-effects model. To also account for return autocorrelation, we estimate a dynamic panel regression for daily returns y_{it} according to

$$y_{it} = c + \alpha(L)y_{it} + \boldsymbol{\beta}(L)'\mathbf{x}_{it} + \boldsymbol{\gamma}(L)'\mathbf{s}_{it} + u_i + \varepsilon_{it}, \tag{6}$$

where the error component u_i captures the random heterogeneity of the i-th region and ε_{it} denotes the error that is random across the individual and time dimension. α, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are lag polynomials without an order-zero term.

The regressors in Eq. (6) are defined as follows: $\mathbf{s}_{it}$ stands for the sentiment representation vector in the ith country/region. We further include control variables

Table 5 Panel regressions for Equities, FI, and FX. In the Var.Name column, r_d and r_y denote daily and annual return variables, and vol_m stands for monthly realized volatility estimates. sgn() is the signum function. The s_d denote a representation element or a PC. The leftmost column 'Type' gives the context of the variables. ∗∗∗, ∗∗, ∗ are the 1%, 5% and 10% significance levels, respectively. *loglik* and $LB_1 0(p)$ give the log-likelihood value and Ljung-Box statistic (lag 10)

Eq Panel:		PCA	PLS	SAE(FF)	FF Class.	SAE(GRU)
Type	Var.Name					
Cmdty:	$r_{d,t-1}$	0.98^{***}	0.98^{***}	0.97^{***}	0.97^{***}	0.97^{***}
	$r_{d,t-2}$	– 0.12	– 0.11	– 0.13	– 0.13	– 0.12
Eq:	$\text{vol}_{m,t-1}$	10.98^{**}	11.07^{**}	11.03^{**}	11.02^{**}	10.96^{**}
	$\text{vol}_{m,t-2}$	-15.72^{**}	-15.74^{**}	-15.72^{**}	-15.68^{**}	-15.66^{**}
	$r_{d,t-1}$	2.21^{***}	2.2^{***}	2.22^{***}	2.22^{***}	2.22^{***}
	$\text{sgn}(r_{d,t-1})$	0.01^{**}	0.01^{**}	0.01^{**}	0.01^{**}	0.01^{**}
	$r_{d,t-2}$	0.32	0.32	0.32	0.32	0.32
	$\text{sgn}(r_{d,t-2})$	– 0.0	– 0.0	– 0.0	– 0.0	– 0.0
	$r_{y,t-1}$	-0.67^{***}	-0.66^{***}	-0.67^{***}	-0.67^{***}	-0.67^{***}
	$r_{y,t-2}$	1.15^{***}	1.15^{***}	1.15^{***}	1.15^{***}	1.15^{***}
FI:	$r_{d,t-1}$	0.07^{***}	0.07^{***}	0.07^{***}	0.07^{***}	0.07^{***}
	$r_{d,t-2}$	– 0.01	– 0.01	– 0.01	– 0.01	– 0.01
FX:	$r_{d,t-1}$	0.76^{***}	0.77^{***}	0.74^{***}	0.73^{***}	0.73^{***}
	$r_{d,t-2}$	– 0.03	– 0.05	– 0.06	– 0.06	– 0.04
R-Eq/PC1:	$s_{d,t-1}$	– 0.01	0.14^{***}	– 0.0	– 0.0	0.01
	$s_{d,t-2}$	0.01	-0.01	0.05^{***}	0.03^{***}	0.03^{**}
R-FI/PC2:	$s_{d,t-1}$	– 0.0	0.01	– 0.0	– 0.01	-0.02^{*}
	$s_{d,t-2}$	0.01	– 0.01	0.01	0.01	0.04^{**}
R-Macro/PC3:	$s_{d,t-1}$	– 0.02	-0.24^{**}	– 0.0	– 0.0	– 0.01
	$s_{d,t-2}$	0.02^{*}	– 0.04	0.0	0.0	0.01
R-Cmdty/PC4:	$s_{d,t-1}$	– 0.01	– 0.04	0.01	0.02^{**}	0.03^{*}
	$s_{d,t-2}$	0.01	0.21^{**}	– 0.01	0.01	– 0.01
R-FX/PC5:	$s_{d,t-1}$	0.01	– 0.04	0.02	0.02^{**}	0.03^{**}
	$s_{d,t-2}$	– 0.02	0.1	– 0.0	0.0	– 0.01
Stats:	loglik, $LB_{10}(p)$	(**137733**, .98)	(**137761**, .97)	(**137744**, .98)	(**137750**, .98)	(**137742**, .98)
FI Panel:		PCA	PLS	SAE(FF)	FF Class.	SAE(GRU)
Type	Var.Name					
Cmdty:	$r_{d,t-1}$	0.03	0.03	0.03	0.03	0.0
	$r_{d,t-2}$	0.44	0.43	0.43	0.41	0.42
Eq:	$\text{vol}_{m,t-1}$	24.4^{*}	24.43^{*}	24.27^{*}	24.31^{*}	24.23^{*}
	$\text{vol}_{m,t-2}$	– 15.47	– 15.35	– 15.01	– 15.12	– 15.13
	$r_{d,t-1}$	– 0.25	– 0.26	– 0.23	– 0.24	– 0.25
	$r_{d,t-2}$	– 0.39	– 0.4	– 0.43	– 0.43	– 0.43
	$r_{y,t-1}$	-0.96^{***}	-0.96^{***}	-0.96^{***}	-0.96^{***}	-0.96^{***}
	$r_{y,t-2}$	1.05^{**}	1.06^{**}	1.05^{**}	1.05^{**}	1.06^{**}

Table 5 (continued)

FI Panel: Type	Var.Name	PCA	PLS	SAE(FF)	FF Class.	SAE(GRU)
FI:	$r_{d,t-1}$	2.59^{***}	2.6^{***}	2.6^{***}	2.6^{***}	2.6^{***}
	$\text{sgn}(r_{d,t-1})$	– 0.01	– 0.01	– 0.01	– 0.01	– 0.01
	$r_{d,t-2}$	0.59^{***}	0.59^{***}	0.58^{***}	0.58^{***}	0.58^{***}
	$\text{sgn}(r_{d,t-2})$	0.0	0.0	0.0	0.0	0.0
FX:	$r_{d,t-1}$	1.45^{**}	1.43^{*}	1.25^{*}	1.28^{*}	1.46^{*}
	$r_{d,t-2}$	0.25	0.26	0.36	0.31	0.36
R-Eq/PC1:	$s_{d,t-1}$	0.01	0.24	-0.08^{*}	– 0.02	– 0.0
	$s_{d,t-2}$	-0.03	0.07	0.03	0.04	0.07
R-FI/PC2:	$s_{d,t-1}$	-0.06	0.04	-0.05	-0.05^{*}	-0.02
	$s_{d,t-2}$	-0.02	0.01	0.12^{***}	0.07^{***}	0.14^{**}
R-Macro/PC3:	$s_{d,t-1}$	0.02	– 0.37	0.03	0.02	– 0.02
	$s_{d,t-2}$	0.04	0.2	– 0.01	– 0.02	– 0.03
R-Cmdty/PC4:	$s_{d,t-1}$	0.11^{**}	– 0.22	0.11^{**}	0.05	0.1^{*}
	$s_{d,t-2}$	– 0.02	0.28	– 0.03	– 0.0	– 0.08
R-FX/PC5:	$s_{d,t-1}$	0.06	– 0.12	0.09^{**}	0.06^{*}	0.01
	$s_{d,t-2}$	0.01	0.02	-0.07^{*}	– 0.04	– 0.05
Stats:	loglik, $LB_{10}(p)$	(**80192**, 1.0)	(**80188**, 1.0)	(**80199**, 1.0)	(**80196**, 1.0)	(**80195**, 1.0)
FX Panel: Type	Var.Name	PCA	PLS	SAE(FF)	FF Class.	SAE(GRU)
Cmdty:	$r_{d,t-1}$	-0.2^{***}	-0.2^{***}	-0.2^{***}	-0.2^{***}	-0.2^{***}
	$r_{d,t-2}$	0.06	0.06	0.07	0.07	0.06
Eq:	$\text{vol}_{m,t-1}$	0.47	0.49	0.49	0.48	0.48
	$\text{vol}_{m,t-2}$	– 0.46	– 0.48	– 0.46	– 0.45	– 0.46
	$r_{d,t-1}$	– 0.02	– 0.02	– 0.02	– 0.02	– 0.02
	$r_{d,t-2}$	– 0.03	– 0.03	– 0.03	– 0.03	– 0.03
	$r_{y,t-1}$	– 0.03	– 0.04	– 0.04	– 0.04	– 0.03
	$r_{y,t-2}$	0.11^{***}	0.11^{***}	0.11^{***}	0.11^{***}	0.11^{***}
FI:	$r_{d,t-1}$	-0.03^{**}	-0.03^{**}	-0.03^{**}	-0.03^{**}	-0.03^{**}
	$r_{d,t-2}$	0.01	0.01	0.01	0.01	0.01
FX:	$r_{d,t-1}$	2.85^{***}	2.85^{***}	2.83^{***}	2.83^{***}	2.83^{***}
	$\text{sgn}(r_{d,t-1})$	– 0.0	– 0.0	– 0.0	– 0.0	– 0.0
	$r_{d,t-2}$	0.97^{***}	0.97^{***}	0.94^{***}	0.95^{***}	0.96^{***}
	$\text{sgn}(r_{d,t-2})$	– 0.0	– 0.0	– 0.0	– 0.0	– 0.0
R-Eq/PC1:	$s_{d,t-1}$	0.01	0.04	– 0.0	– 0.01	-0.01^{**}
	$s_{d,t-2}$	– 0.01	0.03	0.0	0.0	– 0.0
R-FI/PC2:	$s_{d,t-1}$	0.0	-0.02^{**}	0.01	0.01	0.01
	$s_{d,t-2}$	– 0.0	– 0.01	0.0	0.0	– 0.0
R-Macro/PC3:	$s_{d,t-1}$	– 0.0	– 0.04	– 0.01	– 0.0	0.01
	$s_{d,t-2}$	– 0.01	– 0.02	0.0	0.0	0.0

Table 5 (continued)

FX Panel: Type	Var.Name	PCA	PLS	SAE(FF)	FF Class.	SAE(GRU)
R-Cmdty/PC4:	$s_{d,t-1}$	0.01	0.01	0.0	0.01	– 0.0
	$s_{d,t-2}$	0.01	– 0.0	– 0.01	– 0.0	0.0
R-FX/PC5:	$s_{d,t-1}$	– 0.01	– 0.01	0.01^{**}	0.01^{**}	0.02^{***}
	$s_{d,t-2}$	0.01^{*}	0.01	0.01^{*}	0.01^{**}	0.0
Stats:	loglik, $LB_{10}(p)$	(**171600**, .99)	(**171613**, .99)	(**171607**, .99)	(**171609**, .99)	(**171607**, .99)

A bold-faced *loglik* indicates inclusion in the MCS (5%)

$\mathbf{x}_{it}$ inspired by the monthly regional panel regression in Calomiris and Mamaysky (2019). The control variables $\mathbf{x}_{it}$ that enter our three model setups contain lags of the monthly realized stock market volatility as well as of the annual stock market returns. Further, we include lags of the dependent variable as well as signs thereof to control for the use of classifiers in the representation learning. However, we abstain from including low-frequency macro-variables like quarterly GDP growth as these exhibit little correlation to daily returns.

The three Tables 5 show the regression summaries for Eq, FX returns, and government bond yield changes (FI) as dependent variables. For each asset class, we report results for the inclusion of different representation variants—based on the PCA, the PLS, the supervised AE, the Feedforward Classifier, and the supervised AE with GRU encoder. All representation variables employed are constructed on the semi-annually held out test data sets (see subsection 3.4.1) and are hence out-of-sample. News sentiment representations $\mathbf{s}_{it}$ for the models with supervision losses each consist of the five interpretable dimensions, denoted 'R-Eq', 'R-FI', 'R-FX', 'R-Cmdty', and 'R-Macro'. Results for principal components 1–5 are subsumed under the same rows. We constrain the lag length to three for all variables. This number of lags is sufficient to eliminate error autocorrelation while not restricting the validity of later results. For sake of brevity, the tables show only estimates for the first two lags of $\mathbf{s}_{it}$ and $\mathbf{x}_{it}$.

Given the large time series dimension and limited number of cross-sectional units (large T, small N) in our study, endogeneity problems addressed by, e.g., Arellano and Bond (1991) tend to diminish. Hence, we abstain from a GMM estimation of the panel model. However, we do safeguard against potential inconsistent estimators arising due to omitted common effects in the cross-section. Importantly, conclusions of the regression are virtually unchanged when employing the common correlated effects estimator of Pesaran (2006). To compute standard errors we rely on a sandwich estimator for the covariance matrix of coefficients. In specific, we use the Driscoll and Kraay (1998) HAC-type estimator adapted to a panel context.

We can summarize the following findings: First, we can significantly ($\alpha = 0.01$) link major asset class returns to sentiment representations extracted from neural networks. In this context, the interdependence is interpretable in an economic sense. Focusing on the panel results for the neural network representations

(columns SAE(FF), FF Class., SAE(GRU)), we notice that stock index returns depend significantly positive on lagged equity-themed sentiments. Similar results, albeit weaker, are observed for government bond returns which can be traced back positively to lagged FI-themed sentiments. Regarding currency forward returns, we can establish a sensible significant ($\alpha = 0.05$) positive connection to FX-themed sentiment representations. Second, significant and interpretable interdependencies between every single one of the three asset classes and news sentiment representations can only be made for both supervised AE variants and the neural network classifier. This finding suggests that only expressive encodings allow to uncover the explanation power of news sentiment representations for future returns in all three asset classes. Third, for all asset classes, all models—as identified by the representation variant in the regression—are included in the MCS at 5% size based on the panel residuals. This establishes the statistical equivalence of the representation approaches when employed in the panels.

4.3 Trend following for news sentiment representations

A long-standing popular investment strategy for futures and forwards is predicated on the notion that past price trends often persist. These so-called trend-following or time series momentum strategies look at the direction (sign) of an asset's past return and take up a position of the same sign going forward. By design, such strategies are able to capture the momentum of markets. Moskowitz et al. (2012) and Hurst et al. (2013) give an in-depth analysis of time series momentum using past price return data for signal construction. Given the close link between financial returns and news events, more and more investment strategies aim to capitalize on measures of news sentiment as given in semi-structured datasets. Examples in this vein are the investment strategies outlined in Tetlock et al. (2008), Larsen and Thorsrud (2022) as well as Groß-Klußmann et al. (2019).

To uncover potential economic gains, we explore basic trading strategies that utilize news sentiment representations as stand-alone signal.

We consider a strategy that trades stock index futures for all 7 regions or countries under consideration, 10-year government bond futures in all regions but China and currency forwards (USD as quote currency) for all regions except China. Every single instrument can be uniquely associated with a country/region and hence with a sentiment representation for that country/region. Details on the financial data are given in Sect. 2.2. We devise strategies for the asset classes Eq, FI and FX as well as an average strategy across asset classes. To keep the strategy simple while maximizing the effect of the signals, we impose zero transaction costs and reshuffle the hypothetical portfolio on a daily basis. We further require that the sentiment representation used for the signal of an asset class instrument is of the same theme (as defined at the outset of section 3): This means that the sentiment representation element s_{it} for instrument i will stand for the Eq-themed representation element when trading stock index futures, FI-themed for government futures and FX-themed in case of FX forward investments. Due to the inverse relationship of yields (our

supervision) and bonds (the instrument), we multiply the FI-themed representation with -1 before signal construction.

Inspired by the trend-following signals for futures as in, e.g., Moskowitz et al. (2012), we compute exponential moving averages for sentiment representations. The exponential moving average (EMA) filter applied to a time series of representation elements $s_{i,t}$ is defined recursively as follows:

$$\bar{s}_{i,t} = \begin{cases} s_{i,t_0}, & t = t_0 \\ \alpha s_{i,t} + (1-\alpha)\bar{s}_{i,t-1}, & t > t_0, \end{cases} \tag{7}$$

where α is a smoothing factor, $0 < \alpha < 1$, controlling the weight distribution. Further, it allows to compute the half-life HL, the time at which the decay $(1-\alpha)$ reaches 0.5, $HL = -\frac{\ln 2}{\ln(1-\alpha)}$. An EMA filter is best understood in terms of the half-life measured in days, with a low value indicating a very steep weight decay, i.e., most of the weight mass attached to recent observations.

Weights for the *i*th financial instrument are based on the lagged sign of the EMA of the sentiment representation. To mitigate the effect of more volatile markets, we further scale the weight by the inverse lagged $2 \cdot 252$ day sample standard deviation of the return for the instrument traded. The weights per futures or forward contract for portfolio formation in $t-1$ are hence given by

$$w_{i,t-1} = \operatorname{sgn}(\bar{s}_{i,t-2}) \frac{1}{\hat{\sigma}_{i,t-2}}. \tag{8}$$

The additional signal lag ensures that all signals are available at the end of the $(t-1)$th trading day and autocorrelation effects are mitigated.

Finally, each strategy per asset class is an average of the individual sub-strategies in single instruments. The time-t investment strategy returns under zero transaction costs and daily end-of-day trading are given by

$$r^{trend}_{t,ac} = \frac{1}{N_{ac}} \sum_{i=1}^{N} w_{i,t-1,ac} r_{i,t,ac}, \tag{9}$$

where $r_{i,t,ac}$ is the futures or forward return and N_{ac} denotes the number of instruments. The *ac*-subscript identifies the asset class among Eq, FI, and FX. The cross-asset class strategy is an average of the three asset-class strategies.

Trend following strategies hold the promise of producing returns uncorrelated to global asset class returns. To analyze this claim, Moskowitz et al. (2012) regress hypothetical strategy returns on controls like representative stock market, government bond, and currency rate returns. Similarly, we inspect the sign and significance of the intercept in the regression

$$r^{trend}_{t,ac} = \alpha + \beta_1 \mathrm{EQMKT}_t + \beta_2 \mathrm{GOV}_t + \beta_3 \mathrm{FX}_t + \varepsilon_t, \tag{10}$$

where EQMKT_t is the average stock index futures return for our region sample, GOV_t is the average government bond return and FX_t stands for the return of the dollar index (see Sect. 2.2).

Table 6 Information ratios per half-life ('HL') for the return-based momentum (sub-panel 'return') as well as the supervised AE-based momentum (sub-panel 'SAE(FF)'). Table shows information ratios per asset class (Eq, FI, FX) and an equal-weighted strategy in the asset classes (eq.-wgt). ∗∗∗, ∗∗, ∗ (1%, 5% and 10%) are significance levels of the t test of significance of α in the regression (10). Maximum values bold-faced

	Return				SAE(FF)			
HL	eq.-wgt.	Eq	FI	FX	eq.-wgt.	Eq	FI	FX
1	– 0.18	– 0.18	– 0.1	– 0.05	0.51***	**0.25** *	**0.35** *	0.31
2	– 0.25	– 0.23	– 0.1	– 0.17	**0.54** ***	0.22**	0.32	**0.44** **
3	– 0.3	– 0.24	– 0.14	– 0.24	0.46***	0.21**	0.25	0.37
4	– 0.28	– 0.23	– 0.11	– 0.23	0.41**	0.15*	0.22	0.39
5	– 0.21	– 0.13	– 0.05	– 0.3	0.42**	0.12	0.27	0.39
10	– 0.12	– 0.08	0.07	– 0.27	0.28	0.02	0.32	0.19
20	– 0.05	– 0.12	0.21	– 0.2	0.24	0.07	0.19	0.17
30	0.08	0.02	0.21	– 0.07	0.25	0.13	0.2	0.09
40	0.23	0.14	0.31	0.03	0.16	0.08	0.18	0.01
50	0.28	**0.17**	**0.32**	0.1	0.18	0.07	0.14	0.09
60	0.23	0.13	0.29	0.07	0.14	0.05	0.26	– 0.1
70	0.27	0.12	0.31	0.15	0.13	0.04	0.27	– 0.11
80	0.28	0.16	0.27	**0.17**	0.14	0.04	0.28	– 0.1
90	**0.29**	0.15	0.28	0.16	0.15	0.03	0.31	– 0.09

Table 7 Information ratios shown for competitor models in case of an strategy based on the average returns of the equally weighted sub-strategies for all short-term EMAs (HL 1-10) and long-term EMAs (HL 20-90). Bold-faced values indicate exclusion from the MCS at the 5% level

	PCA	PLS	SAE(FF)	FF Class.	SAE(GRU)	Return
short term	**– 0.16**	0.26	0.48	0.63	0.47	– 0.24
long term	0.12	-0.03	0.15	0.21	0.05	0.29

Table 6 shows information ratios, defined as the annualized return divided by the annualized standard standard deviation, for short-term and longer term trend-following strategies with half-lifes ranging from 1 to 90 days. We contrast representative results for exponentially weighted sentiment representations from the SAE(FF) with results for exponentially weighted instrument returns. We can summarize the following findings: First, we observe that news sentiment representations favor short-term signal constructions as information ratios are highest for EMA half-lifes between 1 and 7 days. When pure asset returns are employed as competing signals in the EMA computations, only slower EMA signals with half-lifes from 30 to 90 days generate positive information ratios. We interpret this finding as indicative of quickly changing news environments which often obscure longer lasting narratives. Second, the regression results for Eq. (9) show that only the sentiment strategy produces returns significantly surpassing market variables (indicated by the asterisks). An inspection

of the individual asset class strategy components (columns Eq, FX, and FI) reveals that effects are strongest for the shorter term stock market futures strategy (HL 1-4). In FI and FX, information ratios can widely be explained away by general market returns with the exception of two strategies at half-lifes 1 and 2. Benefitting from diversification, a strategy comprised of the three equally weighted asset class sub-strategies (eq.-wgt.) attains highest information ratios and a positive α that stays significant up to HL 5.

Table 7 gives a compact overview of the equally weighted strategy performances for all competing models and a pure return-based strategy. We group the half-lifes into two main strategies—a short-term strategy utilizing half-lifes 1–10 and a longer term strategy for the HL ranging from 20 to 90. The table reports information ratios for average returns in the half-life groups. Summing up, we notice consistently superior information ratios for the short-term strategy involving the neural network-encoded sentiment representations when compared to both the linear counterparts (PLS, PCA) and the signals based on asset returns alone. Notably, the autoencoder-regularized strategies (SAE(FF), SAE(GRU)) yield slightly worse short-term information ratios than the FF classifier. We conjecture that—absent control variates—this particular analysis benefits the more greedy, only mildly regularized FF classifier. For the long-term EMAs, the return-based strategy dominates all models, highlighting the short-term nature of daily news sentiment aggregates and representations in our setup.

To construct the model confidence set, we compute losses as the negative strategy returns over a passive benchmark with the same volatility. Except for the PCA approach, all representation variants are included in the MCS (5%), thus establishing the statistical equivalence of neural and PLS representations in this application.

4.4 Nowcasting real GDP growth

The third application focuses on nowcasting real GDP growth in our region sample excluding China using daily news sentiment representations. Data for the real GDP growth rate in one quarter will gradually build up until preliminary GDP numbers get published at the end of the quarter. These data are often subject to revisions such that a final reading can take several more quarters to form. It is hence worthwhile for policymakers and other financial market practitioners to closely track macroeconomic data to get an early assessment of the real GDP growth. In this vein, Andreou et al. (2013) demonstrate that exploiting daily financial data to forecast quarterly GDP growth is superior to using monthly and quarterly data alone. In terms of alternative data sources, Ellingsen et al. (2022) show that topic time series mined from textual news data have similar predictive power for real GDP forecasts as hard economic data signals.

Our study draws on the mixed-data-sampling (MIDAS) regression put forth in Ghysels et al. (2007), Andreou et al. (2010) and Andreou et al. (2011) as well as the accompanying R-package (see Ghysels et al. (2016)). MIDAS regressions are econometric frameworks able to effectively connect data observed on higher, e.g., daily data frequencies to a dependent variable observed on a lower, say quarterly,

frequency. The conventional approach to regressing quarterly data onto data on a higher frequency would be to take quarterly averages of the high-frequency data and use these aggregates in a time series regression. The idea behind MIDAS time series models is instead to aggregate the higher frequency in a data-driven way, where aggregation weights for each data point are parameters to be estimated alongside other coefficients.

In the following, let t be the time index for the main data frequency of interest, i.e., quarterly in case of quarterly real GDP growth forecasting. Let individual data points $x^D_{i,t}$ stand for a scalar element (theme) of a daily sentiment representation. The subscript dimensions i, t give the quarter t of the data point as well as the day i in the quarter. The number of days in each quarter is m. With this notation, we denote the data item for the last day of a quarter as $x^D_{m,t}$ and the first day in quarter t as $x^D_{1,t}$.

To leverage within-quarter information in the explanatory variables, we employ the augmented distributed lag (ADL) MIDAS with leads model put forward in Andreou et al. (2013). The ADL-MIDAS with leads constructs quarterly forecasts with $J^D \in \{1, 2, 3\}$ monthly blocks of data from within the quarter to be predicted. As such, the model can capture the gradual information build-up and give an early GDP growth prediction at the start of the last month in the quarter. Using the above notation, the ADL-MIDAS(p^Q,q^D,J^D) for the quarter on quarter real GDP percentage growth, y^Q_t, with daily variables $x^D_{i,t}$ is based on

$$y^Q_{t+1} = \mu + \alpha(L)y^Q_{t+1} + \beta \Bigg[\sum_{i=(3-J^D)\cdot m/3}^{m-1} w^{\theta}_{i-m} x^D_{m-i,t+1} + \sum_{j=0}^{q^D-1} \underbrace{\sum_{i=0}^{m-1} w^{\theta}_{i+j\cdot m} x^D_{m-i,t-j}}_{\tilde{x}^Q_{t-j}} \Bigg] + \varepsilon_{t+1}, \tag{11}$$

where α is a lag polynomial of degree p^Q without order-zero terms. The first bracketed sum concerns the within quarter 'leads', given in J^D blocks of monthly explanatory data. The second bracketed term sums up q^D lags of the higher frequency variable taken in quarterly blocks ($\tilde{x}^Q_t$). Equation (11) shows the formula defining the ADL-MIDAS with leads for a daily predictor variable $x^D_{\cdot,t}$.

One of the main innovations in MIDAS regressions is to parsimoniously parameterize the aggregation of higher frequency data. In Eq. (11), the implicit aggregation to the quarterly $\tilde{x}^Q_t$ is based on weights w^{θ}_i with parameters $\boldsymbol{\theta} = \{\theta_0, \theta_1\}$. Our weight construction uses the exponential Almon lag polynomial where ith weights are given as

$$w_i(\theta_0, \theta_1) = \frac{\exp(\theta_0 i + \theta_1 i^2)}{\sum_{k=1}^{m} \exp(\theta_0 k + \theta_1 k^2)}. \tag{12}$$

Ghysels et al. (2005) assert that the Almon polynomial can mimic a wide range of weight function shapes. The parameters $\boldsymbol{\theta}$ are estimated jointly with other coefficients, making the implicit aggregation of the ADL-MIDAS data-driven.

Table 8 Values are percentage fractions of the RMSE of the AR(1)-benchmark. Values for the sample up to the COVID-19 pandemic given in the last column

Region		Q1/2011 - Q4/2021	Q1/2011 - Q4/2019
US	SAE(FF)	0.512	0.904
	FF Classifier	0.528	0.988
	SAE(GRU)	0.546	1.005
	PLS	0.554	0.946
	PCA (sentiment)	0.574	**1.095**
	PCA (Macro Var.)	0.939	**1.185**
EU	SAE(FF)	0.637	0.882
	FF Classifier	0.646	1.043
	SAE(GRU)	0.578	0.997
	PLS	0.89	**1.171**
	PCA (sentiment)	0.63	1.063
	PCA (Macro Var.)	0.914	**1.184**
All	SAE(FF)	0.634	0.951
(Av.)	FF Classifier	0.647	1
	SAE(GRU)	0.639	1.001
	PLS	0.691	**1.441**
	PCA (sentiment)	0.646	**1.153**
	PCA (Macro Var.)	0.947	**1.121**

Bold-faced values indicate model instances that are *excluded* from the MCS at the 5% level

In the application of the ADL-MIDAS model, we follow Andreou et al. (2013) who estimate single predictor models where the external higher frequency predictor variable $x_{\cdot,\cdot}$ is one-dimensional. The final forecast will be formed by a a combination of single predictor model forecasts. The reasons for this approach are twofold and rooted in the analysis of Timmermann (2006). First, in a survey of forecast combination methods, the author demonstrates that forecast combinations typically improve forecast accuracy. Second, and relatedly, forecast combinations tend to be robust toward regime changes and other model instabilities.

Our single predictor ADL-MIDAS models are each based on one of the five (economically themed) representations we recover from our model variants, i.e., $x^D_{\cdot,\cdot} \equiv s_{i,t}$. Let now denote $\widehat{Y}^Q_{i,t+1}$ the forecasts of the quarterly GDP growth in $t+1$ from a single predictor ADL-MIDAS. We compute the combined forecast for each model variant based on its five single predictor forecasts according to

$$\widehat{Y}^Q_{t+1} = \sum_{i=1}^{5} \omega_{i,t} \widehat{Y}^Q_{i,t+1}, \tag{13}$$

where the weights $\omega_{i,t}$ are computed with the discounted mean squared forecast error (MSFE) method. The discounted MSFE method assigns weights that are inversely proportional to the square of the MSFE, where a discount factor smoothes the

weights by giving greater weight to more recent inverse squared forecast errors. A detailed overview of the discounted MSFE method can be found in appendix C.

We estimate the single-predictor ADL-MIDAS (11) on the sample spanning Q1/2004 to Q4/2010 and subsequently compute one-quarter-ahead forecast combinations. To separate out the large effect of the COVID-19 pandemic on the macroeconomic data, we present forecast results for two out-of-sample periods, Q1/2011 to Q4/2021 as well as Q1/2011 to Q4/2019. Using the AIC criterion, we select an ADL-MIDAS(p^Q=1,q^D=0,J^D=1)-specification for the ADL-MIDAS with daily predictors. This means the first 2 months of daily data in the current quarter to be predicted enter the equation.

Each row of Table 8 describes forecast results from a model average of the single-predictor ADL-MIDAS based on one specific five-dimensional sentiment representation. Next to our sentiment representations, we additionally include the first five PCs derived from a rolling PCA applied to our macroeconomic variables (denoted Macro Var.) as classical nowcasting predictors. In spirit of the nowcasting literature (see, e.g., Andreou et al. (2013)), we use an AR(1) model as benchmark. All numbers are the RMSE fractions against the AR(1), $\frac{RMSE(model)}{RMSE(AR(1))}$. For sake of brevity, Table 8 gives results for the two regions US and EU as well as an average of all regions. The appendix table 9 contains results for the remaining regions.

We can summarize the following findings. First, forecast errors for ANN-based representations are widely lower than for linear benchmarks across the board. However, the model confidence set approach applied to the squared forecast errors shows that models are equivalent on the full sample. An exception is the sample excluding the COVID-19 period. In this case, only neural network representations enter the MCS in all regions shown, thus highlighting the usefulness of the approaches. Second, in line with findings of Andreou et al. (2013), we observe an improvement in the RMSE when daily predictors are employed over both the benchmark. Notably, the nowcasts from daily sentiment representations capture the COVID-19 pandemic well as the low RMSE fractions for the sample including the COVID-19 pandemic show. Third, remarkably, prior to COVID-19, on average, only MIDAS regressions employing sentiment representations from the SAE(FF) produce errors lower than the AR(1) benchmark. The success of the SAE(FF) supports the theoretical assertion in Le et al. (2018) that a combination of the reconstruction and supervision loss results in a more robust generalization to unseen tasks.

5 Conclusions

News analytics data like time series of sentiment measures cover a vast range of financial topics potentially relevant to the markets. Hence, compressing these high-dimensional data for subsequent use is of utmost importance. Neural networks have long been known to learn expressive and compact internal data representations in their hidden layers. In appreciation of this, our study leverages the flexibility offered by neural networks to obtain interpretable multi-purpose data representations that generalize well. At the core of our analysis is a supervised autoencoder architecture

which blends the standard reconstruction loss with financial market-inspired supervision losses in a multi-task setting.

We consider two data sets of daily news sentiment scores for more than 1,000 topics with mostly macroeconomic themes. After extracting low-dimensional hidden representations from neural networks and linear competitor models, we include the sentiment representations in three distinct macro-financial applications. In response to the two research questions raised, our analysis shows that models with neural network-derived sentiment representations are statistically at least equal to models employing the PLS- or PCA-based representations. Notably, the performances of the SAE(FF) variant in particular are stable across all applications and country/regions considered. In more detail, the following findings can be summarized. First, in three panel regressions explaining daily regional stock index returns, currency forward returns, as well as government bond yield changes, we observe that only representations based on neural networks significantly add explanatory power beyond controls. Remarkably, this finding is robust across all three asset classes considered. The analysis highlights the merits of interpretability as asset class returns can be traced back to elements of the neural sentiment representations that match the asset class theme. Second, when employed as signals in short-term time series momentum investment strategies, our neural network-based sentiment representations attain higher risk adjusted returns than both representations from linear competitor models and plain returns. Third, an MIDAS nowcasting exercise confirms the favorable properties of neural sentiment representations. Despite being trained mainly with short-term return supervisions, the representations generalize well to the problem of forecasting quarterly real GDP growth.

Overall, the results show that the most balanced architecture choice is an autoencoder with additional supervision losses based on important financial market metrics like asset class returns. Several features make this model particularly attractive: One, the supervision losses act as regularizers and discipline the dimension reduction by leveraging external data. This is a direct consequence of the multi-task approach which prevents a too strong focus on just one optimization objective in training. The inclusion of multiple losses allows to aim at constructing 'general purpose' representations that generalize beyond single use cases. Two, in a rolling window setting, directionality in X can naturally be preserved by adding a supervision via the direction in y. Relatedly, the added finance-themed supervisions improve interpretability of their representation counterparts. Three, the architecture can be conveniently modified and scale up to complex data characteristics. While our applications favor a shallow feedforward network underlying the autoencoder, the FF encoder can optionally be replaced by an RNN to account for temporal dependence in data.

Appendix A: Topics

A.1 Ravenpack Global Macro Sentiment Dataset

The Ravenpack RPA 1.0 sentiment data provide two distinct sentiment measures. First, an event sentiment score (ESS) is designed to capture news that can be traced back to events. The events are identified by the system based on external data and internal algorithms. An example for an event is the publication of a GDP data surprise which typically triggers heavy news flow. Second, a composite sentiment score (CSS) is comprised of sentiment measures for more general news that cannot be attributed to specific events. The following topics hence each yield two time series of scores corresponding to the ESS and CSS.

GROUP-subclass

acquisitions-mergers, aid, assets, balance-of-payments, bankruptcy, business-activity, civil-unrest, commodity-prices, consumption, corporate-responsibility, credit, credit-ratings, crime, domestic-product, earnings, economic-union, elections, employment, equity-actions, exploration, foreign-exchange, foreign-relations, government, health, housing, industrial-accidents, insider-trading, interest-rates, inventory, labor-issues, legal, marketing, migration, natural-disasters, partnerships, pollution, production, products-services, public-finance, public-opinion, regulatory, revenues, security, social-relations, taxes, technical-analysis, transportation, war-conflict

TYPE-subclass

accelerated-approval-application, accelerated-approval-designation, acquisition, acquisition-bid, acquisition-regulation, air-pollution, aircraft-accident, airspace-closure, airspace-open, airspace-violation, animal-attack, animal-infestation, antitrust-investigation, antitrust-settlement, antitrust-suit, appeal, approval-rating, assassination, asylum, austerity-measures, automobile-accident, avalanche, award, balance-of-payments, balance-of-payments-deficit, balance-of-payments-guidance, bankruptcy, bankruptcy-unit, blackmail, blizzard, board-member-appointment, board-member-death, board-member-firing, board-member-health, board-member-resignation, board-member-retirement, board-member-salary, bombing, border-control, breakthrough-therapy-application, breakthrough-therapy-designation, business-confidence, business-confidence-guidance, business-contract, buybacks, cabinet, campaign-ad, candidacy, capital-punishment, central-bank-meeting, central-bank-meeting-minutes, civil-unrest, clinical-trials, clinical-trials-patient-enrollment, cold-wave, commodity-assets, commodity-futures, commodity-price, competition, composite-pmi, composite-pmi-guidance, conference, confidentiality-pact, congressional-testimony, construction-pmi, consumer-confidence, consumer-confidence-guidance, consumer-price-index, consumer-price-index-guidance, consumer-spending, consumer-spending-guidance, copyright-infringement, corporation-tax, corporation-tax-guidance, corruption, coup-d-etat, credit-rating-change, credit-rating-outlook, credit-rating-watch, currency-adoption, currency-adoption-guidance, currency-guidance, currency-rate, currency-valuation, current-account, current-account-deficit, current-account-guidance,

current-account-guidance-deficit, current-account-surplus, cyber-attacks, cyclone, dam-accident, defamation, defense-budget, defense-budget-guidance, deflation, deflation-guidance, demand, demand-guidance, diplomatic-recall, diplomatic-visit, discrimination, divorce, donation, drilling, drought, durable-goods, durable-goods-guidance, early-election, earnings, earnings-estimate, earnings-per-share-guidance, earthquake, economic-growth, economic-growth-guidance, economic-union-application, economic-union-membership, economic-union-withdrawal, elections, embargo, embargo-guidance, embezzlement, emergency-landing, employment, employment-guidance, endorsement, epidemic, evacuation, exchange-compliance, exchange-noncompliance, executive-appointment, executive-compensation, executive-death, executive-firing, executive-health, executive-incentives, executive-resignation, executive-retirement, executive-salary, executive-scandal, executive-search, expenses, explosion, export-tax, export-tax-guidance, exports, exports-guidance, facility, facility-accident, factory-accident, fast-track-application, fast-track-designation, flood, force-majeure, fraud, freight-transport-accident, going-private, government-administration, government-bailout, government-budget, government-budget-deficit, government-budget-guidance, government-budget-guidance-deficit, government-budget-guidance-surplus, government-budget-surplus, government-contract, government-official, government-power, government-treaty, grant, gross-domestic-product, gross-domestic-product-guidance, hail-storm, headquarters-change, heat-wave, hijacking, hirings, home-sales-existing, home-sales-existing-guidance, home-sales-new, home-sales-new-guidance, hostage-situation, house-prices, house-prices-guidance, human-stampede, hurricane, ice-storm, immigration, impeachment, import-tax, import-tax-guidance, imports, imports-guidance, industrial-production, industrial-production-guidance, inflation, inflation-guidance, initial-public-offering, initial-public-offering-issuance, initial-public-offering-lock-up, initial-public-offering-price, initial-public-offering-unit, insider-buy, insider-gift, insider-sell, insider-surrender, insider-trading-lawsuit, interest-rate, interest-rate-guidance, interest-rate-overnight, interest-rate-overnight-guidance, international-aid, inventories, inventories-guidance, investment, ipo-regulatory-approval, ipo-regulatory-scrutiny, jobless-claims, jobless-claims-guidance, joint-venture, judiciary, kidnapping, landslide, law-enforcement, layoffs, legal-issues, legislative, legislature, loan, manufacturing-index, manufacturing-index-guidance, manufacturing-pmi, manufacturing-pmi-guidance, market-entry, market-exit, market-guidance, market-share, marriage, merger, merger-regulation, military-action, mine-accident, minimum-wage, minimum-wage-guidance, monarchy, monsoon, murder, non-farm-payrolls, non-farm-payrolls-guidance, non-manufacturing-pmi, non-manufacturing-pmi-guidance, official-visit, orphan-drug-application, orphan-drug-designation, pandemic, partnership, party-nomination, patent, patent-infringement, peace-process, pipeline-accident, pipeline-bombing, piracy, platform-accident, political-campaign, political-endorsement, poll-survey, polls, power-outage, power-plant-accident, presidency, press-conference, primary-election, primary-election-polls, primary-elections, priority-review-application, priority-review-designation, private-credit, producer-price-index, producer-price-index-guidance, product-catastrophe, product-development, product-discontinued, product-enhancement,

product-fault, product-outage, product-pricing, product-promotion, product-recall, product-release, product-resumed, product-review, product-side-effects, product-support, project-abandoned, protest, public-offering, public-transport-accident, real-gross-domestic-product, real-gross-domestic-product-guidance, recession, recession-guidance, referendum, refinery-accident, regulatory-investigation, regulatory-product-application, regulatory-product-approval, regulatory-product-review, regulatory-product-warning, regulatory-stress-test, relative-strength-index, reorganization, reorganization-unit, resource-discovery, retail-sales, retail-sales-guidance, revenue, revenue-estimate, revenue-guidance, revenue-volume, robbery, same-store-sales, same-store-sales-guidance, sanctions, sanctions-guidance, sand-storm, sell-registration, services-pmi, services-pmi-guidance, settlement, shooting, short-selling-ban, sink-hole, snow-storm, solar-flare, sovereign-debt, sovereign-debt-guidance, sovereign-debt-purchases, sovereign-debt-purchases-guidance, spill, spin-off, sponsorship, stagflation, stagflation-guidance, stagnation, stagnation-guidance, state-of-emergency, state-visit, storm, suicide, suicide-bombing, supply, supply-guidance, tanker-accident, tax-break, tax-break-guidance, tax-evasion, technical-price-level, technical-view, terrorism, thunder-storm, tornado, trade-balance, trade-balance-deficit, trade-balance-guidance, trade-balance-guidance-deficit, trade-balance-surplus, trading, transportation-disruption, travel-warning, treasury-bill-auction, treasury-bill-yield, treasury-bond-auction, treasury-bond-price, treasury-bond-yield, treasury-note-auction, treasury-note-yield, tropical-storm, tsunami, typhoon, unemployment, unemployment-guidance, union-pact, unit-acquisition, unit-acquisition-regulation, vandalism, verdict, violence, volcanic-ash-cloud, volcanic-eruption, war-declaration, war-demonstration, water-contamination, water-shortage, weapons-testing, wild-fire, workers-strike, workforce-salary

A.2 GDELT economic topics

agriculture, alliance, armedconflict, austerity, capital markets, central bank, competition, conflict, consumption, corruption, cost, crisis agg, currencies, currencies agg, currency, currency exchange rate, debt, debt agg, deflation, democracy, determinants of econ growth, earningsreport, econ bankruptcy, econ budget deficit, econ cost of living, econ counterfeitmoney, econ currency reserves, econ cutoutlook, econ debt, econ deregulation, econ dieselprice, econ earningsreport, econ electricaldemand, econ electricalgrid, econ electricalprice, econ emergingecon, econ foreignbanks, econ foreigninvest, econ gasolineprice, econ goldprice, econ growth analytics, econ growth policy, econ heatingoil, econ heatingoilprice, econ housing prices, econ inflation, econ natgasprice, econ nationalize, econ oilprice, econ pay cuts, econ pricecontrol, econ propane, econ propaneprice, econ sovereign debt, econ stockmarket, econ suspicious activity report, econ transport cost, econ workingclass, economic crisis, economic growth, economic growth agg, election, emergingecon, euro, fi agg, finance and growth, fiscal policy, fuelprice, growth, human capital, inclusive growth, industrial accident, industry policy, inflation, infrastructure and growth, innovation and growth, insurgency, interest rates, jobless

growth, kill, labor intensive growth, labor markets, labor markets agg, leader, macroeconomic, military, monetary policy, monetary policy agg, negotiations, politics, poverty, price agg, prices, protest, public finance, public finance agg, rebellion, relevant currencies, resignation, sanctions, science, shocks, stockmarket agg, strike, terror, unemployment, us dollar, wb capital markets, wb industry policy and real sectors, wb labor markets, wb manufacturing, wb public finance, world currencies

Appendix B: Neural network setup: details and code repository

The code for the Pytorch (Paszke et al. (2019)) implementation of the neural networks can be found at

https://github.com/agk18/newsrepresentations.

SAE(FF). In our model validation exercise, the supervised autoencoder favors a rather shallow network structure. We consider a grid of several specifications

$$[1005, [l_2, l_1, 5, l_1, l_2], 1005]$$

with l_1 from $\{10, 20, 30\}$ and an optional layer $l_2 \in \{30, 40, 50\}$.

The dev set results point to a [1005, [10, 5, 10], 1005]-architecture which is used in the results section.

FF. For comparison reasons, we report the feedforward network for an architecture that mirrors the encoder part of the SAE(FF), i.e., [1005, [10], 5].

SAE(GRU). recurrent layers: 1

hidden size: 5

input sequence length: 21.

The fixed input sequence length equals 1 month of trading days and is sufficient to pick up the sample autocorrelation.

Appendix C: The discounted MSFE forecast averaging

The discounted MSFE forecast averaging (DMSFE henceforth) computes weighted averages across M individual forecasts

$$\widehat{Y}^{Q}_{t+1} = \sum_{i=1}^{M} w_{i,t} \widehat{Y}^{Q}_{i,t+1}, \tag{C.1}$$

where the $w_{i,t}$ are the forecast combination weights formed at t. The weights for the ith forecast are computed according to

$$w_{i,t} = \frac{(\lambda_{i,t}^{-1})^{\kappa}}{\sum_{j=1}^{n} (\lambda_{j,t}^{-1})^{\kappa}}, \tag{C.2}$$

with $\lambda_{i,t}$ denoting the trailing sum of squared foreccast errors

$$\lambda_{i,t} = \sum_{\tau=T_0}^{t-1} \delta^{t-1-\tau} (Y^Q_{\tau+1} - \widehat{Y}^Q_{i,\tau+1})^2, \tag{C.3}$$

where $\kappa = 2$ and $\delta = 0.9$. T_0 is the point at which the first (pseudo) out-of-sample forecast is computed. Hence, the DMSFE attaches highest weights to the forecasts with lowest trailing sum of squared foreccast errors.

Appendix D: Nowcasting GDP growth: details for additional countries

Table 9 Values are percentage fractions of the RMSE of the AR(1)-benchmark. Lowest fractions (i.e., lowest RMSEs) bold-faced. Values for the regime up to the COVID-19 pandemic given in extra columns

region		Q1/2011 - Q4/2021	Q1/2011 - Q4/2019
AU	SAE(FF)	0.995	**1.203**
	FF Classifier	1.057	1.281
	SAE(GRU)	**0.968**	1.272
	PLS	0.997	1.368
	PCA (sentiment)	1.004	1.37
	PCA (Macro Var.)	1.008	1.307
CA	SAE(FF)	0.547	0.847
	FF Classifier	0.553	**0.814**
	SAE(GRU)	0.562	0.973
	PLS	0.552	0.869
	PCA (sentiment)	**0.543**	1.322
	PCA (Macro Var.)	0.941	0.958
GB	SAE(FF)	0.384	0.856
	FF Classifier	**0.377**	0.928
	SAE(GRU)	0.385	0.855
	PLS	0.388	**0.761**
	PCA (sentiment)	0.388	0.939
	PCA (Macro Var.)	0.928	1.069
JP	SAE(FF)	0.729	1.015
	FF Classifier	**0.722**	0.945
	SAE(GRU)	0.794	**0.907**
	PLS	0.768	3.531
	PCA (sentiment)	0.74	1.125
	PCA (Macro Var.)	0.955	1.021

Data availability The core alternative sentiment data that support the findings of this study are available from the GDELT project and the Ravenpack company. While the GDELT dataset is free to use, restrictions apply to the availability of the data from Ravenpack, which were used under license for the current study, and so are not publicly available. Macroeconomic time series from the OECD and the ECB are publicly available. As regards the financial market data, restrictions also apply to the data from Bloomberg and Refinitiv. Free supporting data alternatives are available from the author upon reasonable request. A curated example data set is already provided alongside the code repository.

Declarations

Disclaimer The views, methods, and opinions outlined in this study and accompanying web sources are those of the author and do not necessarily reflect those of Quoniam Asset Management GmbH or its employees.

References

Andreou, E., Ghysels, E., & Kourtellos, A. (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics, 158*(2), 246–261.

Andreou, E., Ghysels, E., & Kourtellos, A. (2011). *Forecasting with mixed-frequency data. The Oxford handbook of economic forecasting*. Oxford University Press.

Andreou, E., Ghysels, E., & Kourtellos, A. (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics, 31*(2), 240–251.

Aparicio, D., & López de Prado, M. (2018). How hard is it to pick the right model? MCS and backtest overfitting. *Algorithmic Finance, 7*(1–2), 53–61.

Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies, 58*(2), 277–297.

Bair, E., Hastie, T., Paul, D., & Tibshirani, R. (2006). Prediction by supervised principal components. *Journal of the American Statistical Association, 101*(473), 119–137.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence., 35*(8), 1798–1828.

Bourlard, H., & Kamp, Y. (1988). Auto-association by multilayer perceptrons and singular value decomposition. *Biological Cybernetics, 59*, 291–294.

Bybee, L., B.T. Kelly, A. Manela, and D. Xiu. 2021. Business news and business cycles. Available at SSRN: https://ssrn.com/abstract=3446225. Accessed 10 Dec 2023.

Bybee, L., B.T. Kelly, and Y. Su. 2022. Narrative asset pricing: Interpretable systematic risk factors from news text. Available at SSRN: https://ssrn.com/abstract=3895277. Accessed 10 Dec 2023.

Calomiris, C. W., & Mamaysky, H. (2019). How news and its context drive risk and returns around the world. *Journal of Financial Economics, 133*(2), 299–336.

Chung, J., C. Gulcehre, K. Cho, and Y. Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. Presented in NIPS 2014 Deep Learning and Representation Learning Workshop. Available at http://arxiv.org/abs/1412.3555.

Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics, 80*(4), 549–560.

Ellingsen, J., Larsen, V. H., & Thorsrud, L. A. (2022). News media versus fred-md for macroeconomic forecasting. *Journal of Applied Econometrics, 37*(1), 63–81.

Erhan, D., A. Courville, Y. Bengio, and P. Vincent 2010, 13–15 May. Why does unsupervised pre-training help deep learning? In Y. W. Teh and M. Titterington (Eds.), Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Volume 9 of Proceedings of Machine Learning Research, Chia Laguna Resort, Sardinia, Italy. PMLR. pp. 201–208.

Fan, J., Liao, Y., & Wang, W. (2016). Projected principal component analysis in factor models. *The Annals of Statistics, 44*(1), 219–254.

Ghysels, E., Kvedaras, V., & Zemlys, V. (2016). Mixed frequency data sampling regression models: The r package midasr. *Journal of Statistical Software, 72*(4), 1–35.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of Financial Economics, 76*(3), 509–548.

Ghysels, E., Sinko, A., & Valkanov, R. (2007). Midas regressions: Further results and new directions. *Econometric Reviews, 26*(1), 53–90.

Giglio, S., D. Xiu, and D. Zhang. 2021. Test assets and weak factors. Chicago Booth Research Paper 21-04, Available at SSRN: https://ssrn.com/abstract=3768081. Accessed 10 Dec 2023.

Goodfellow, I.J., Y. Bengio, and A. Courville. 2016. Deep Learning. Cambridge, MA, USA: MIT Press. http://www.deeplearningbook.org. Accessed 10 Dec 2023.

Groß-Klußmann, A., König, S., & Ebner, M. (2019). Buzzwords build momentum: Global financial twitter sentiment and the aggregate stock market. *Expert Systems with Applications, 136*, 171–186.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies, 33*(5), 2223–2273.

Gu, S., Kelly, B., & Xiu, D. (2021). Autoencoder asset pricing models. *Journal of Econometrics., 222*(1 Part B), 429–450.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics, 23*(4), 365–380.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica, 79*(2), 453–497.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

Hurst, B., Hua Ooi, Y., & Heje Pedersen, L. (2013). Demystifying managed futures. *Journal of Investment Management, 11*(3), 42–58.

Kelly, B.T., S. Malamud, and K. Zhou. (2021). The virtue of complexity in return prediction. Available at SSRN: https://ssrn.com/abstract=3984925. Accessed 10 Dec 2023.

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

Kingma, D. P., & Ba. J. (2014). Adam: A method for stochastic optimization. *CoRR*, abs.1412.6980.

Larsen, V. H., & Thorsrud, L. A. (2022). Asset returns, news topics, and media effects. *The Scandinavian Journal of Economics, 124*(3), 838–868.

Le, L., Patterson, A., & White, M. (2018). Supervised autoencoders: Improving generalization performance with unsupervised regularizers. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems.* (Vol. 31). Curran Associates Inc.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance, 66*(1), 35–65.

Lundberg, S.M. and S.I. Lee. 2017. A unified approach to interpreting model predictions, In Advances in Neural Information Processing Systems 30, eds. Guyon, I., U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Curran Associates, Inc. 4765–4774.

Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics, 104*, 228–250.

Paszke, A., S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. 2019. Pytorch: An imperative style, high-performance deep learning library, Advances in Neural Information Processing Systems. Curran Associates, Inc. 32: 8024–8035.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica, 74*(4), 967–1012.

Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the Theory of Games (AM-28)* (Vol. II, pp. 307–318). Princeton: Princeton University Press.
Spilak, B. and W.K. Härdle. (2023). Portfolio tail-risk protection with non-linear latent factors. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4483490. Accessed 10 Dec 2023.
Ter Ellen, S., V.H. Larsen, and L.A. Thorsrud. (2021). Narrative monetary policy surprises and the media. Journal of Money, Credit and Banking .
Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance, 63*(3), 1437–1467.
Thorsrud, L. A. (2020). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics, 38*(2), 393–409.
Timmermann, A. (2006). Forecast combinations, Volume 1 of Handbook of Economic Forecasting. Elsevier. pp 135–196. .
White, H. (2000). A reality check for data snooping. *Econometrica, 68*(5), 1097–1126.
Wold, H. (1966). Estimation of principal components and related models by iterative least squares. *Journal of Multivariate Analysis*.
Wold, H. (1975). Path models with latent variables: The nipals approach, In Quantitative Sociology, eds. Blalock, H., A. Aganbegian, F. Borodkin, R. Boudon, and V. Capecchi, International Perspectives on Mathematical and Statistical Modeling. Academic Press. pp. 307–357.
Zhuang, F., X. Cheng, P. Luo, S.J. Pan, and Q. He 2015. Supervised representation learning: Transfer learning with deep autoencoders. In Q. Yang and M. J. Wooldridge (Eds.), Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015. AAAI Press. pp. 4119–4125.