MONASH University

MONASH BUSINESS SCHOOL

DEPARTMENT OF ECONOMETRCS AND BUSINESS STATISTICS

ISSN 1440-771X

WORKING PAPER SERIES

Improving Out-of-Sample Forecasts of Stock Price Indexes with Forecast Reconciliation and Clustering

Raffaele Mattera, George Athanasopoulos, Rob J Hyndman

Working Paper No. 17/23
July 2023

Improving out-of-sample forecasts of stock price indexes with forecast reconciliation and clustering

Raffaele Mattera[a], George Athanasopoulos[b], Rob J Hyndman[b]

[a]*Department of Social and Economic Sciences, Sapienza University of Rome, Italy*

[b]*Department of Econometrics and Business Statistics, Monash University, Australia*

Abstract

This paper discusses the use of forecast reconciliation with stock price time series and the corresponding stock index. The individual stock price series may be grouped using known meta-data or other clustering methods. We propose a novel forecasting framework that combines forecast reconciliation and clustering, to lead to better forecasts of both the index and the individual stock price series. The proposed approach is applied to the Dow Jones Industrial Average Index and its component stocks. The results demonstrate empirically that reconciliation improves forecasts of the stock market index and its constituents.

Keywords: financial time series, hierarchical forecasting, clustering, unsupervised learning, prediction, machine learning, finance

JEL classification: C53, C10

1. Introduction

In this paper we aim to improve forecasts of stock market indexes and their component prices, using forecast reconciliation and clustering. Accurate stock price forecasts are important in finance: they allow investors to make more informed decisions, they help traders construct more profitable investment strategies, and they assist policy-makers to better monitor the evolution of financial markets (e.g. see Pesaran and Timmermann, 2002, Marquering and Verbeek, 2004, Kong et al., 2011, Chen et al., 2015). Forecasts are particularly useful for trading activity. Forecast-based investment strategies can outperform static ones if they anticipate future market conditions (Merton, 1981, Pesaran and Timmermann, 2002). A simple forecast-based trading strategy is, for example, to go long if the market index price is predicted to increase in the next h time periods, and to go short otherwise (Anatolyev and Gerko, 2005, Blaskowitz and Herwartz, 2011). Forecasts can also be successfully used for the implementation of momentum-based strategies (e.g. see Daniel and Moskowitz, 2016) and those based on portfolio optimization (e.g. see Marquering and Verbeek, 2004, Cenesizoglu and Timmermann, 2012, Tsiakas et al., 2020).

Stock price forecasting is difficult, as financial markets are complex and turbulent systems. Indeed, stocks can be characterized by volatility clustering, non-linear relationships, long memory and hierarchical structures (e.g., see Barnett and Serletis, 2000, Lo, 1991, Mantegna, 1999, Tumminello et al., 2010). These difficulties, along with the efficient market hypothesis, have lead many authors to conclude that

Email addresses: `raffaele.mattera@uniroma1.it` (Raffaele Mattera), `George.Athanasopoulos@monash.edu` (George Athanasopoulos), `Rob.Hyndman@monash.edu` (Rob J Hyndman)

it is impossible to obtain better stock price forecasts than those from a random walk model (Fama, 1995, Timmermann and Granger, 2004). Those that have proposed more complex approaches have usually had limited empirical success (e.g. Yu and Yan, 2020, Wang et al., 2011).

The success of random walk forecasts suggests that future values of the time series cannot be predicted with the currently available information only. For this reason, researchers and practitioners often use additional variables, known as factors or predictive signals, to aid in predicting prices and return (e.g. see Green et al., 2013). In this paper, we provide evidence that, if the hierarchical structure of the stock market is taken into consideration via forecast reconciliation, we can obtain more accurate forecasts than the random walk. For this aim, approaches for forecasting hierarchical time series need to be adopted.

Forecast reconciliation is a statistical technique dealing with multivariate time series following a hierarchical structure (Hyndman et al., 2011), or that more generally adhere to linear constraints (Hyndman et al., 2016, Panagiotelis et al., 2021). Hierarchical time series are common in many application domains. For example, national level retail sales (Makridakis et al., 2022), tourism flows (Athanasopoulos et al., 2009), or electricity demand (Panagiotelis et al., 2023) can be disaggregated by state, regional or an even finer geographical grids. GDP is constructed using income or expense components (Bisaglia et al., 2020).

The motivation behind forecast reconciliation is that the forecasts for disaggregated and aggregated series are not necessarily coherent with the hierarchical structure. In other words, while time series data naturally aggregate according to a hierarchical structure, forecasts usually do not. This issue has been traditionally addressed by using forecasts at only one level of the hierarchy, from which the remaining forecasts are computed.

For example, in the bottom-up approach, forecasts are computed for the most disaggregated, bottom-level, series first. These are then aggregated to obtain forecasts for aggregate series (Dunn et al., 1976). However, the resulting forecasts of the aggregated series are often not as accurate as forecasting the aggregated series directly. The opposite approach is top-down forecasting, where forecasts of the most aggregated series are computed first, and these are then disaggregated down to obtain forecasts of series at lower levels of the hierarchy (e.g. see Gross and Sohl, 1990). However, this often results in poor forecast accuracy at the disaggregated levels, and will always produce biased forecasts even if the original forecasts are unbiased (Hyndman et al., 2011).

Hyndman et al. (2011) proposed a least squares reconciliation approach providing an ex post adjustment to a set of "base" forecasts. That is, base forecasts (an initial set of forecasts) are produced for all disaggregate and aggregate series. These are then adjusted to ensure they are coherent. Wickramasuriya et al. (2019) clarified and generalized the results leading to *Minimum Trace (MinT)* reconciliation, guaranteeing minimum variance unbiased forecasts. Panagiotelis et al. (2021) shows that the predictive accuracy of reconciled forecasts cannot be worse than unreconciled forecasts in the mean squared error sense.

In a recent study, Hollyman et al. (2021) discuss the connection between forecast reconciliation and forecast combination, which is a widely used technique in financial forecasting (Rapach et al., 2010, Alves et al., 2023). Therefore, forecast reconciliation offers two crucial advantages to the forecasting process. First, it ensures that forecasts are coherent with the hierarchical structure. Second, it allows getting more accurate predictions, according to a mechanism similar to the one behind the combination of alternative forecasts.

Successful applications of reconciliation techniques have been proposed in tourism (Athanasopoulos et al., 2009), macroeconomics (Athanasopoulos et al., 2020, Eckert et al., 2021, Lila et al., 2022), demography (Yang et al., 2022), and energy (Jeon et al., 2019, Di Fonzo and Girolimetto, 2023). Some early papers that discuss hierarchical forecasting for stock price indexes have been Lee and Swaminathan (1999) who used a bottom-up approach to forecast the Dow Jones Industrial Average Index, and Darrough and Russell (2002) who provided a comparison between bottom-up and top-down approaches for this aim. However, to the best of our knowledge, forecast reconciliation has not been used when forecasting stock prices. In the context of financial forecasting, Li and Tang (2019) recently adopted MinT reconciliation, but for predicting mortality bond indexes rather than stock market indexes, while Caporin et al. (2023) proposed a reconciliation procedure for realized volatility. (Hyndman and Athanasopoulos, 2021, Chapter 11) provide an introductory exposition to hierarchical forecasting and forecast reconciliation. Athanasopoulos et al. (2023) provide a comprehensive literature review.

It is natural to consider stocks as hierarchical time series, based on the linear aggregation of the common stocks included in a stock market index (Lee and Swaminathan, 1999, Darrough and Russell, 2002). Moreover, there is evidence that stocks are also characterized by hierarchical (or clustering) structures of unknown form (e.g. see Brown and Goetzmann, 1997, Tumminello et al., 2010, Zhang et al., 2020).

Therefore, it may be beneficial to first identify, estimate and forecast such hierarchies, and then apply forecast reconciliation to exploit the hierarchical structures of the series. Multiple hierarchical configurations are possible, based on different clustering tools and different dissimilarity measures applied to the stock price data. We discuss how multiple hierarchies can be used in reconciling stock price and index forecasts.

The contribution of this paper is twofold. First, we apply optimal forecast reconciliation to a new domain, that is financial markets. Second, we develop a novel forecasting framework that combines reconciliation and clustering, by implementing optimal ex post adjustment to the forecasts with the aim of making them coherent with the underlying hierarchical structure, based on the clustering of individual stock price time series.

As an empirical experiment, we consider the Dow Jones Industrial Average (DJIA) index and its constituents. The DJIA is the oldest stock market index in the United States and, for this reason, it is an established benchmark for tracking the overall market performance (Brown et al., 1998, Lee et al., 1999, Kim et al., 2011). Indeed, although it includes thirty stocks, the index has a broad market coverage with different sectors, so shocks affecting specific sectors are reflected by the index. Moreover, the DJIA represents a classical hierarchical time series because, as happens with any equally-weighted index, the top-level series is obtained as a simple linear function of the bottom series.

Clustering is introduced using both available meta-data (such as industry group or stock exchange), and empirical clustering algorithms based on several dissimilarity measures. Forecasts are obtained and evaluated for both the bottom-level series (common stocks) and the top-level (index) series. The results demonstrate empirically that forecast reconciliation, and furthermore forecast reconciliation based on hierarchical structures identified by clustering, is useful for predicting the stock market index and its constituents.

The rest of the paper is structured as follows. Section 2 discusses the methodology employed in the paper, considering the combination of optimal forecast reconciliation and clustering. The data adopted for the empirical experiment and the forecasting set-up are described in Section 3. Section 4 shows the results of cluster analysis used for the estimation of the alternative hierarchies, while Section 5 provides a discussion of the forecasting results. Section 6 discusses the usefulness of the proposed forecasting approach for investment purposes. We conclude with some final remarks in Section 7.

2. Forecast reconciliation with clustering structure

Let $\boldsymbol{p}_i = [p_{i,1}, \dots, p_{i,T}]'$ be the daily closing price time series of the ith stock. The DJIA stock price index at time t, denoted by y_t, is obtained by the sum of its N constituents:

$$y_t = \frac{\sum_{i=1}^{N} p_{i,t}}{d_t}, \tag{1}$$

discounted by a factor d_t — equal for all the stocks — which accounts for market operations such as changes in the index composition and stock splits. The adjusted price series are given by $b_{i,t} = p_{i,t}/d_t$, and so we have the linear constraint:

$$y_t = \sum_{i=1}^{N} b_{i,t}. \tag{2}$$

2.1. *Groups and clusters of stocks*

Let $\boldsymbol{b}_t$ be the vector of all N stocks of interest observed at time t, and let $\boldsymbol{a}_t$ be a corresponding vector of n_a aggregated time series:

$$\boldsymbol{a}_t = \boldsymbol{A}\boldsymbol{b}_t. \tag{3}$$

The first element of $\boldsymbol{a}_t$ is the stock index y_t. Other elements of $\boldsymbol{a}_t$ are aggregations based on subsets of stocks. For example, suppose we aggregate the prices for each of the n_1 exchanges on which they are traded. Let $c_{i,j} = 1$ if stock j is traded on exchange i, and 0 otherwise, and define $\boldsymbol{\mathcal{C}}_1$ to be the $n_1 \times N$ matrix with element $c_{i,j}$ in row i and column j. Then $\boldsymbol{\mathcal{C}}_1\boldsymbol{b}_t$ gives the aggregated prices for all exchanges at time t. We can similarly define $\boldsymbol{\mathcal{C}}_2$ to denote the grouping of stocks based on industries, where each row corresponds to a different industry group. Other groups or clusters of stocks can also be defined. This leads to the aggregation matrix

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{1}' \\ \boldsymbol{\mathcal{C}}_1 \\ \vdots \\ \boldsymbol{\mathcal{C}}_L \end{bmatrix}, \tag{4}$$

where each $\boldsymbol{\mathcal{C}}_\ell$ denotes a grouping or clustering of stocks. The "aggregation" matrix $\boldsymbol{A}$ has dimension $n_a \times N$ and specifies how the stock price series $\boldsymbol{b}_t$ aggregate to form $\boldsymbol{a}_t$. The first row of $\boldsymbol{A}$ defines the sum of all stocks making up the stock index, while other rows define sums of different groups of stocks.

Some components of $\boldsymbol{a}_t$, and therefore $\boldsymbol{A}$, can be formed using a data-driven hierarchical (or partitioning) clustering method. Time series clustering can be divided into three well-known classes (Maharaj et al., 2019): observation-based, feature-based and model-based. In this study, we will use observation-based and model-based clustering. Observation-based approaches group time series according to their observed values. Given a pair of bottom time series $\boldsymbol{b}_i = [b_{i1}, \ldots, b_{iT}]'$ and $\boldsymbol{b}_j = [b_{j1}, \ldots, b_{jT}]'$, a simple observation-based approach involves the use of the standard Euclidean distance:

$$d_{\text{EUCL}}(\boldsymbol{b}_i, \boldsymbol{b}_j) = \sqrt{\sum_{t=1}^{T} (b_{it} - b_{jt})^2}. \tag{5}$$

Observation-based approaches can be useful for clustering short time series, although they impose strong stationarity conditions on the original series. Notice that returns-based clustering approaches are widespread in the asset pricing literature (Brown and Goetzmann, 1997, Brown et al., 2012). Correlation-based approaches (Mantegna, 1999, Brida and Risso, 2010, Raffinot, 2017), instead, use pairwise correlations for computing dissimilarities across stocks.

Model-based approaches define distances based on parameter estimates from statistical models. A well-known example is the ARIMA-based distance. Given two time series $\boldsymbol{b}_i$ and $\boldsymbol{b}_j$, Piccolo (1990) defined the distance between two invertible ARIMA processes as the Euclidean distance between the AR(∞) representation of the two series, i.e.:

$$d_{\text{ARIMA}}(\boldsymbol{b}_i, \boldsymbol{b}_j) = \sqrt{\sum_{k=1}^{K} (\pi_{i,k} - \pi_{j,k})^2}, \tag{6}$$

where $\pi_{i,k}$ denotes the kth "π weight" (Box et al., 2016, p51) for the ith stock.

Given the distance measure, we also need to choose the clustering algorithm to be used. In this paper, we use the Partition Around Medoids algorithm (PAM, Kaufman and Rousseeuw, 1990), which provides an iterative solution to the following minimization problem:

$$\min: \quad \sum_{i=1}^{N} \sum_{c=1}^{C} d^2(\boldsymbol{b}_i, \boldsymbol{b}_c), \tag{7}$$

where $d^2(\boldsymbol{b}_i, \boldsymbol{b}_c)$ is the squared distance between the ith unit and the cth cluster centroid time series. (Any of the above distances measures may be used.) In what follows we focus on the PAM algorithm rather than C-means, because it is more interpretable (using a real time series as the centroid rather than an average), and is more robust to outliers.

The main drawback of the PAM clustering approach lies in the *a priori* selection of the number of clusters C. To address this issue, we follow Arbelaitz et al. (2013) and Batool and Hennig (2021) and use the Average Silhouette Width (ASW), a well-known cluster validity index for evaluating the quality of a partition, measuring the within-cluster cohesion and inter-cluster dispersion. The Silhouette for the ith object can be computed as

$$SW_i = \frac{g_i - f_i}{\max\{g_i, f_i\}} \tag{8}$$

where f_i is the average distance of the ith unit to the other units belonging to the same cluster, and g_i is the average distance of the same unit to others belonging to the closest different cluster. The number of clusters is commonly selected as the number C maximizing the Silhouette in (8).

Once the number of clusters, C, has been selected, the result of cluster analysis on the individual stock time series is the membership matrix $\boldsymbol{\mathcal{C}}$ of dimension $C \times N$, with element $\mathcal{C}_{c,j} = 1$ when $\boldsymbol{b}_j$ belongs to cluster c, and 0 otherwise.

2.2. *Forecast reconciliation*

The full vector of time series at time t is given by

$$\boldsymbol{y}_t = \begin{bmatrix} \boldsymbol{a}_t \\ \boldsymbol{b}_t \end{bmatrix} = \boldsymbol{S}\boldsymbol{b}_t, \tag{9}$$

where $\boldsymbol{S} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{I}_N \end{bmatrix}$ denotes the "summation" matrix of dimension $n \times N$, where $n = n_a + N$. Let $\hat{\boldsymbol{y}}_h$ be the vector of h-step-ahead forecasts obtained with a generic forecasting model. Base forecasts $\hat{\boldsymbol{y}}_h$ generally do not sum up to the top levels, so we say they are not "coherent". Forecast reconciliation methods aim at making forecasts coherent across the aggregation structure. We denote coherent forecasts as $\tilde{\boldsymbol{y}}_h$. Linear reconciliation can be written as follows:

$$\tilde{\boldsymbol{y}}_h = \boldsymbol{M}\hat{\boldsymbol{y}}_h, \tag{10}$$

where $\boldsymbol{M} = \boldsymbol{S}\boldsymbol{G}_h$ is a $n \times n$ mapping matrix, whose role is to project the base forecasts $\hat{\boldsymbol{y}}_h$ onto a coherent subspace (Panagiotelis et al., 2021). For example, in the bottom-up approach, we define $\boldsymbol{G}_h = [\boldsymbol{0} \quad \boldsymbol{I}_N]$, with $\boldsymbol{0}$ denoting a vector of zeros. The optimal least squares approach (known as MinT for Minimum Trace) is obtained (Wickramasuriya et al., 2019) with

$$\boldsymbol{G}_h = \left(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}, \tag{11}$$

where $\boldsymbol{W}$ is the $n \times n$ covariance matrix of the h-step base forecast errors.

We note that there were no negative reconciled forecasts in the empirical experiment. Hence, there was not need to implement non-negativity constraints in the MinT reconciliation procedure to ensure positive prices. The proposed procedure could easily be extended to account for non-negativity constraints, following Wickramasuriya et al. (2020).

3. Stock market data and experimental set up

We evaluate the usefulness of reconciliation in forecasting the prices of the Dow Jones Industrial Average (DJIA) and of its constituents. We use clustering to determine some possible grouping structures as explained in Section 2.

The dataset used for the empirical analysis consists of the daily time series associated with the DJIA index from 1-09-2020 to 30-09-2022. We made this choice because the DJIA composition (see Table 1) changed on 31 August 2020 and has not been revised since; the divisor takes the constant value $d_t = 0.15$ during this period. The DJIA price time series is shown in Figure 1, while Figure 2 shows the time series of its $N = 30$ constituents.

Since we are working with non-stationary integrated price time series, we forecast the series using ARIMA models. A rolling-window procedure is used to obtain the forecasts, where at each step of the recursion we choose the best ARIMA model by means of the automatic procedure described in Hyndman and Khandakar (2008). We use the $\boldsymbol{G}$ matrix resulting from the MinT approach of Wickramasuriya et al. (2019), shown in (11).

The experimental design can be outlined as follows. The time series in Figure 2 have length $T = 524$, and we leave the last $R = 124$ observations for out-of-sample testing. The clustering structures are estimated within the training set, i.e., considering only the first 400 observations. Forecasts at $h = \{1, 3, 6, 12\}$ steps ahead are produced, so the out-of-sample length is equal to $R - h$ ($r = 1, \ldots, R - h$). At each r recursion, an estimation window of length 400 is considered for the model

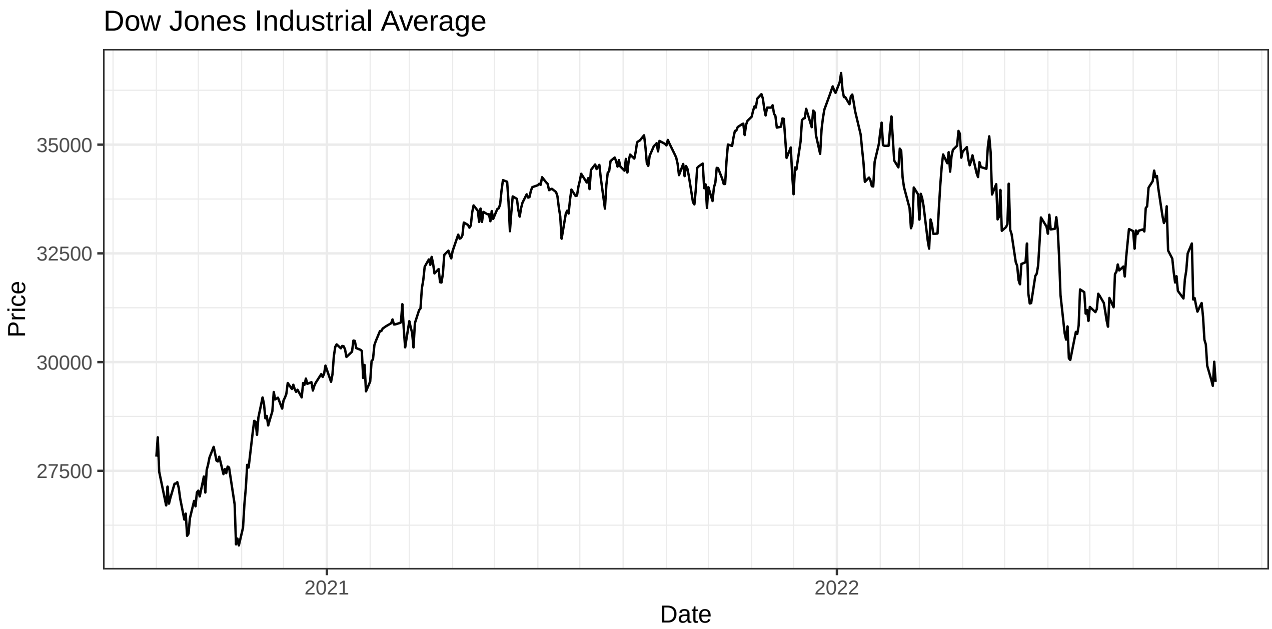

Figure 1: *Dow Jones Industrial Average (DJIA) index: price time series*

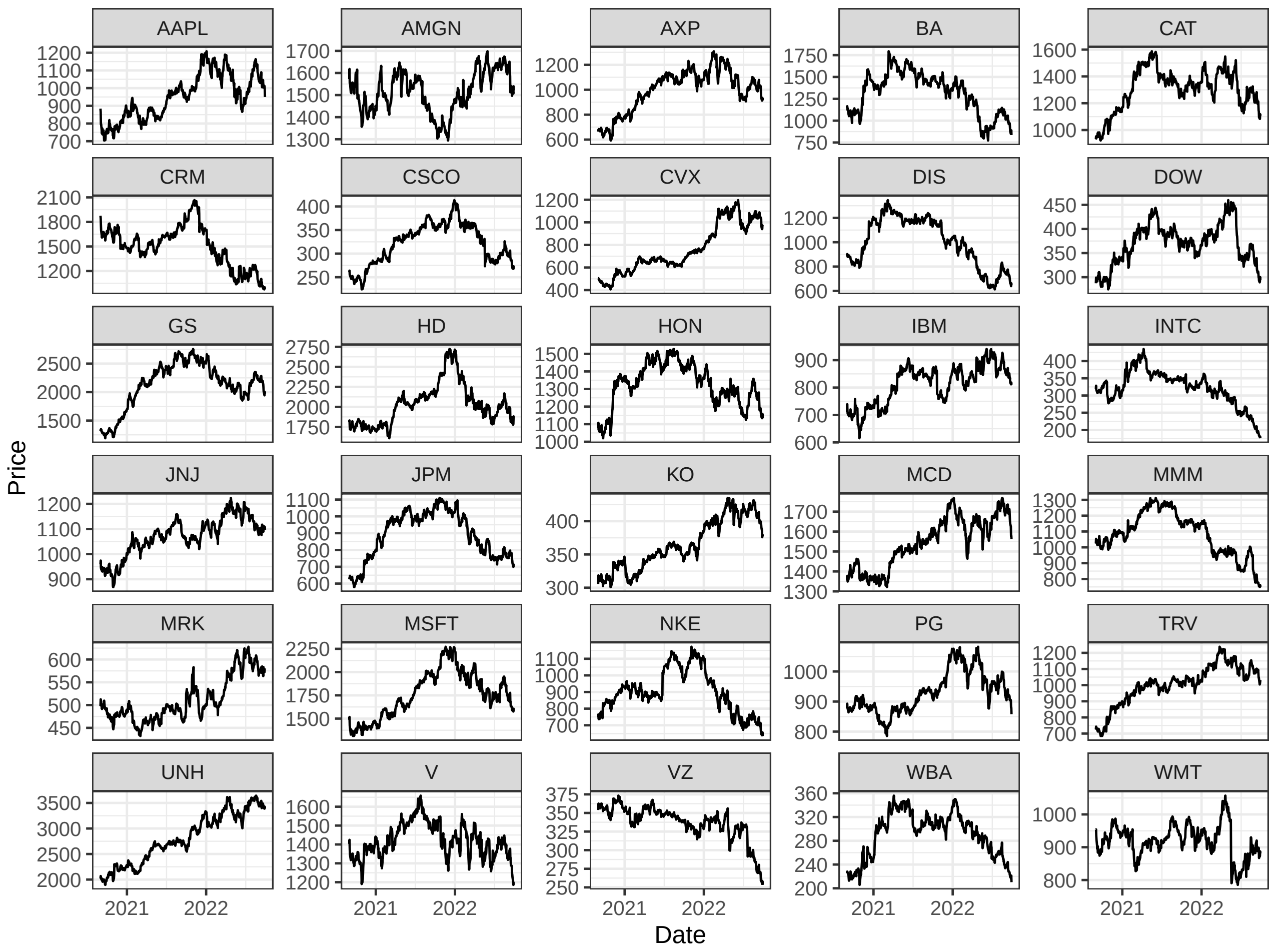

Figure 2: *DJIA constituents: price time series*

Table 1: *DJIA composition after last revision occurred the 31/08/2020*

Company	**Exchange**	**Symbol**	**Industry**
Procter & Gamble	NYSE	PG	Fast-moving consumer goods
3M	NYSE	MMM	Conglomerate
IBM	NYSE	IBM	Information technology
Merck	NYSE	MRK	Pharmaceutical industry
American Express	NYSE	AXP	Financial services
McDonald's	NYSE	MCD	Food industry
Boeing	NYSE	BA	Aerospace and defense
Coca-Cola	NYSE	KO	Drink industry
Caterpillar	NYSE	CAT	Construction and Mining
Disney	NYSE	DIS	Broadcasting and entertainment
JPMorgan Chase	NYSE	JPM	Financial services
Johnson & Johnson	NYSE	JNJ	Pharmaceutical industry
Walmart	NYSE	WMT	Retailing
Home Depot	NYSE	HD	Home Improvement
Intel	NASDAQ	INTC	Semiconductor industry
Microsoft	NASDAQ	MSFT	Information technology
Verizon	NYSE	VZ	Telecommunications industry
Chevron	NYSE	CVX	Petroleum industry
Cisco	NASDAQ	CSCO	Information technology
Travelers	NYSE	TRV	Insurance
UnitedHealth	NYSE	UNH	Managed health care
Goldman Sachs	NYSE	GS	Financial services
Nike	NYSE	NKE	Clothing industry
Visa	NYSE	V	Financial services
Apple	NASDAQ	AAPL	Information technology
Walgreens Boots Alliance	NASDAQ	WBA	Retailing
Dow	NYSE	DOW	Chemical industry
Amgen	NASDAQ	AMGN	Biopharmaceutical
Honeywell	NASDAQ	HON	Conglomerate
Salesforce	NYSE	CRM	Information technology

selection procedure and to make the h-step-ahead forecasts with the selected ARIMA model. We evaluate if the different MinT reconciliation approaches improve with respect to base forecasts, random walk and bottom-up reconciliation.

Let us define $e_t = \hat{y}_{t+h} - y_{t+h}$ the h step ahead forecasting error. Forecast accuracy is evaluated in terms of both absolute and squared errors. For the top level and bottom level series MAE and RMSE are considered, that is:

$$\text{RMSE} = \sqrt{\frac{1}{R-h} \sum_{t=T-R+h}^{T} e_t^2} \tag{12}$$

and:

$$\text{MAE} = \frac{1}{R-h} \sum_{t=T-R+h}^{T} |e_t| \tag{13}$$

Then, predictive accuracy tests of Diebold and Mariano (2002) and van Dijk and Franses (2003) are considered to evaluate the statistical significance of the forecasting error differences. Let us define $d_t = g(e_{1,t}) - g(e_{2,t})$ the error differential between two forecasting approaches up to some transformation $g(\cdot)$; in this paper, we use squares $g(e_{1,t}) = e_{1,t}^2$ and absolute values $g(e_{1,t}) = |e_{1,t}|$. Assuming covariance stationarity of the loss differential series d_t, Diebold and Mariano (2002) show that the sample mean of the loss differential,

$$\bar{d} \equiv \frac{1}{R-h} \sum_{t=T-R+h}^{T} d_t, \tag{14}$$

follows an asymptotically standard Normal distribution. Therefore, testing the null hypothesis of equal forecast accuracy can be obtained by calculating the following statistic:

$$\mathrm{DM} = \bar{d}[V(\bar{d})]^{-1/2}, \tag{15}$$

where $V(\bar{d})$ is consistently estimated assuming a certain autocorrelation structure of the forecasting errors. From Figure 1, we might expect that forecasting models provide poor forecasts in the presence of trend reversions or changes of regimes. Therefore, we also consider an alternative testing approach, proposed in van Dijk and Franses (2003), that is based on the introduction of a weighting scheme providing lower weights to "less relevant" information. The resulting statistic is given by:

$$\bar{d}_w \equiv \frac{1}{R-h} \sum_{t=T-R+h}^{T} \frac{\hat{f}(y_t)}{\max \hat{f}(y_t)} d_t, \tag{16}$$

where $\hat{f}_t(\cdot)$ denotes the density function of y_t estimated by means of a standard Nadaraya-Watson kernel estimator with a Gaussian kernel. In doing so, we provide less weight to the forecasting errors occurring when the realized value y_{t+h} is in the tails of the distribution. The resulting weighted Diebold-Mariano statistic follows a standard Normal distribution asymptotically. In this case the covariance $V(\bar{d}_w)$ is computed considering the weighted loss differentials and their autocovariances.

4. Finding clustering structures: results

In this section, we explore various possible sets of clusters to be used in the forecast reconciliation.

4.1. No clustering structure

A natural starting point in our setting is the use of the MinT forecast reconciliation approach which employs no clustering structure at all. In other words, in this case we use (11) considering $\boldsymbol{A} = \mathbf{1}'$.

4.2. Industry-based clustering

There is a long tradition in financial economics of considering the presence of industry-based clusters of stocks. The rationale behind this clustering approach is that stocks belonging to the same industry sector are affected by common shocks (e.g. King, 1966, Livingston, 1977). In the case of the Dow-Jones stocks in our sample, we have $C = 20$ Industry-based clusters (see Table 1). Several groups are based on singletons, due to the relatively large number of industries relative to the number of stocks included in the DJIA. The implied hierarchy is shown in Figure 3. The Financial Services and IT industries include 4 and 5 stocks respectively, more than any other industry groups, which is why their aggregate values are larger than for the other industries.

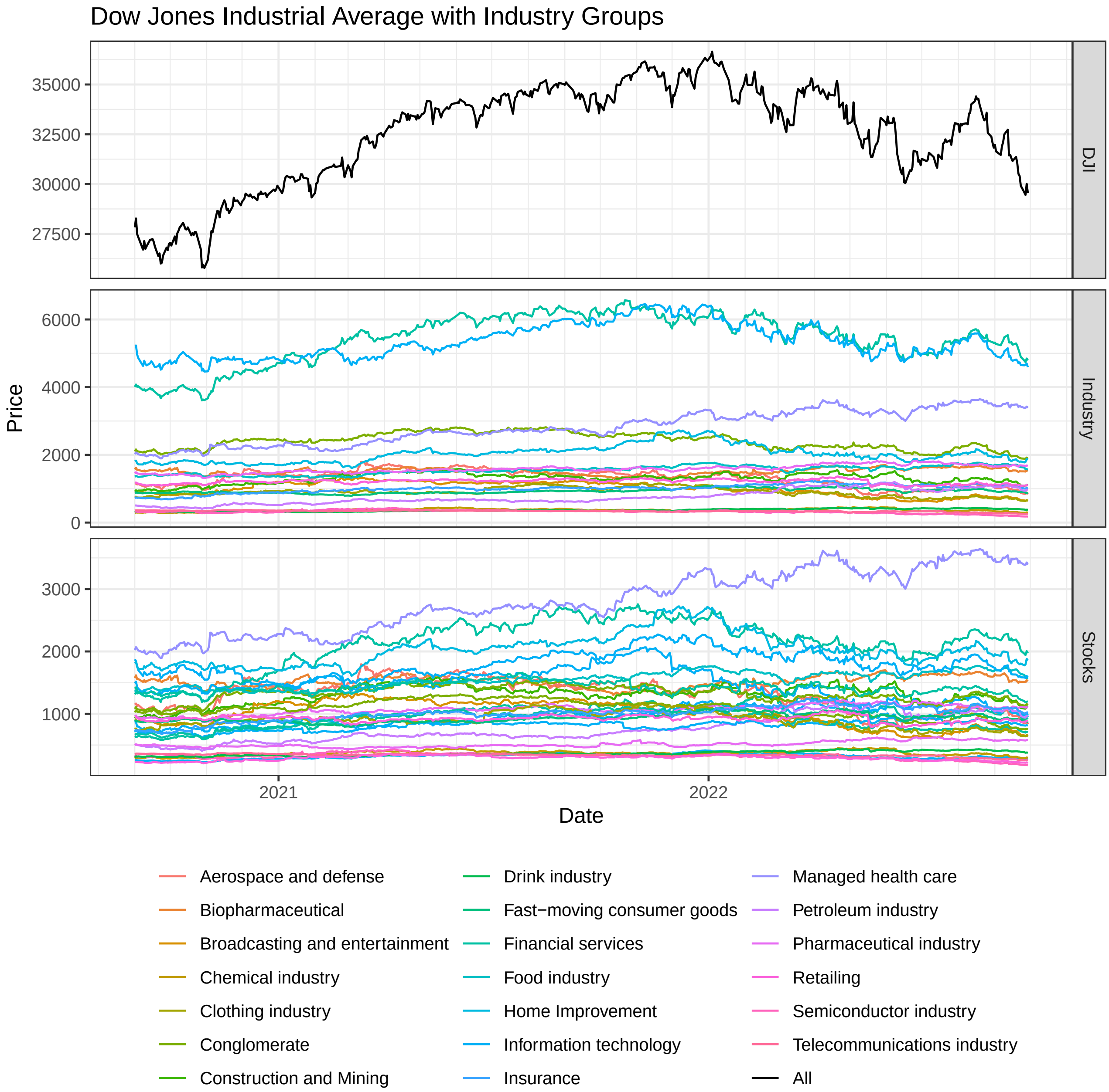

Figure 3: *Aggregated series from the hierarchical structure implied by the Industry-based clustering approach*

4.3. Exchange-based clustering

This clustering approach involves $C = 2$ groups, because DJIA stocks are traded at NYSE and NASDAQ only. The groups are quite unbalanced because most of the stocks included in the DJIA index are traded on the NYSE exchange. The NYSE is known to have a higher average market capitalization of listed companies compared to the NASDAQ so that, on average, the companies listed on the NYSE are larger and more established than those listed on the NASDAQ. Moreover, NASDAQ is known for its focus on IT companies, while NYSE lists a wider range of industries. The exchange-based clusters are shown in Table 1. The implied hierarchy is shown in Figure 4.

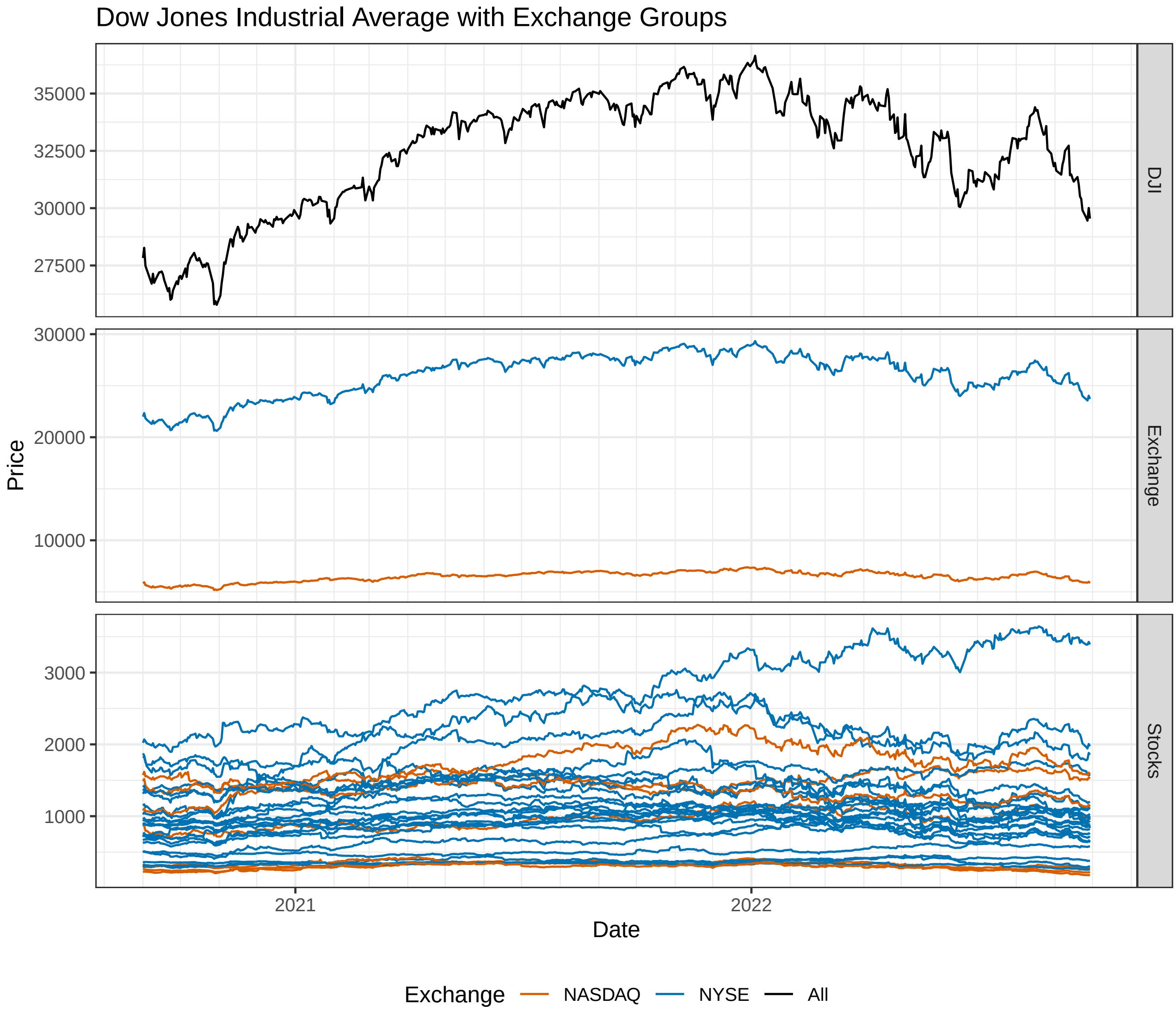

Figure 4: *Aggregated series from the hierarchical structure implied by the Exchange-based clustering approach*

4.4. Observation-based clustering

We construct clusters of stocks considering the standard Euclidean distance between log-returns, that is:

$$x_{i,t} = \log(p_{i,t}) - \log(p_{i,t-1}) \tag{17}$$

We use log returns rather than prices, so that we consider stationary series. As explained in Section 2, we choose the number of clusters C maximizing ASW. Figure 5 shows the ASW associated with $C \in \{1, \ldots, 10\}$ different number of clusters, with $C = 3$ giving the maximum ASW. The resulting clusters are shown in Table 2, and the aggregated series are shown in Figure 6.

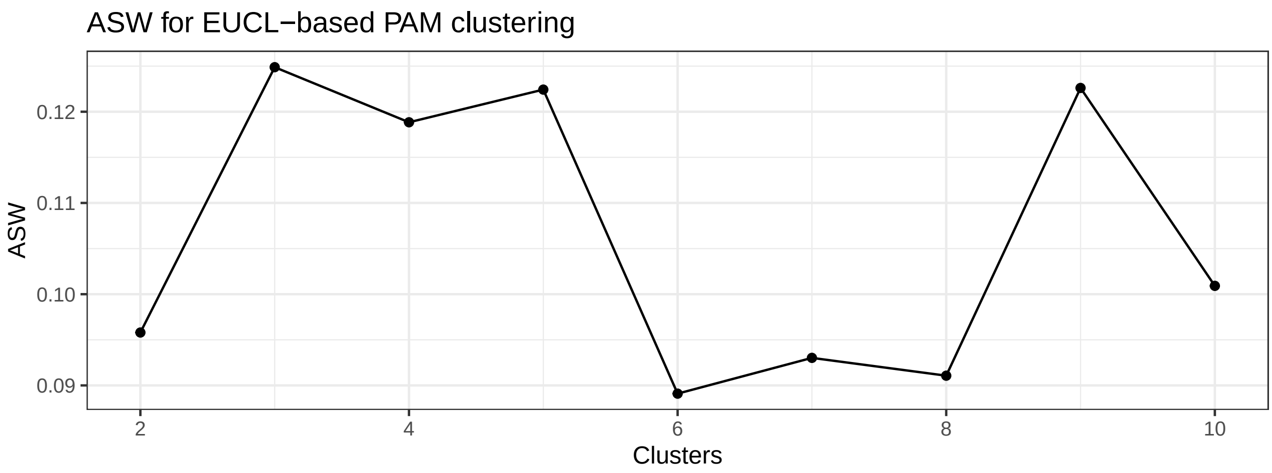

Figure 5: *Average Silhouette Width with EUCL-based PAM partition*

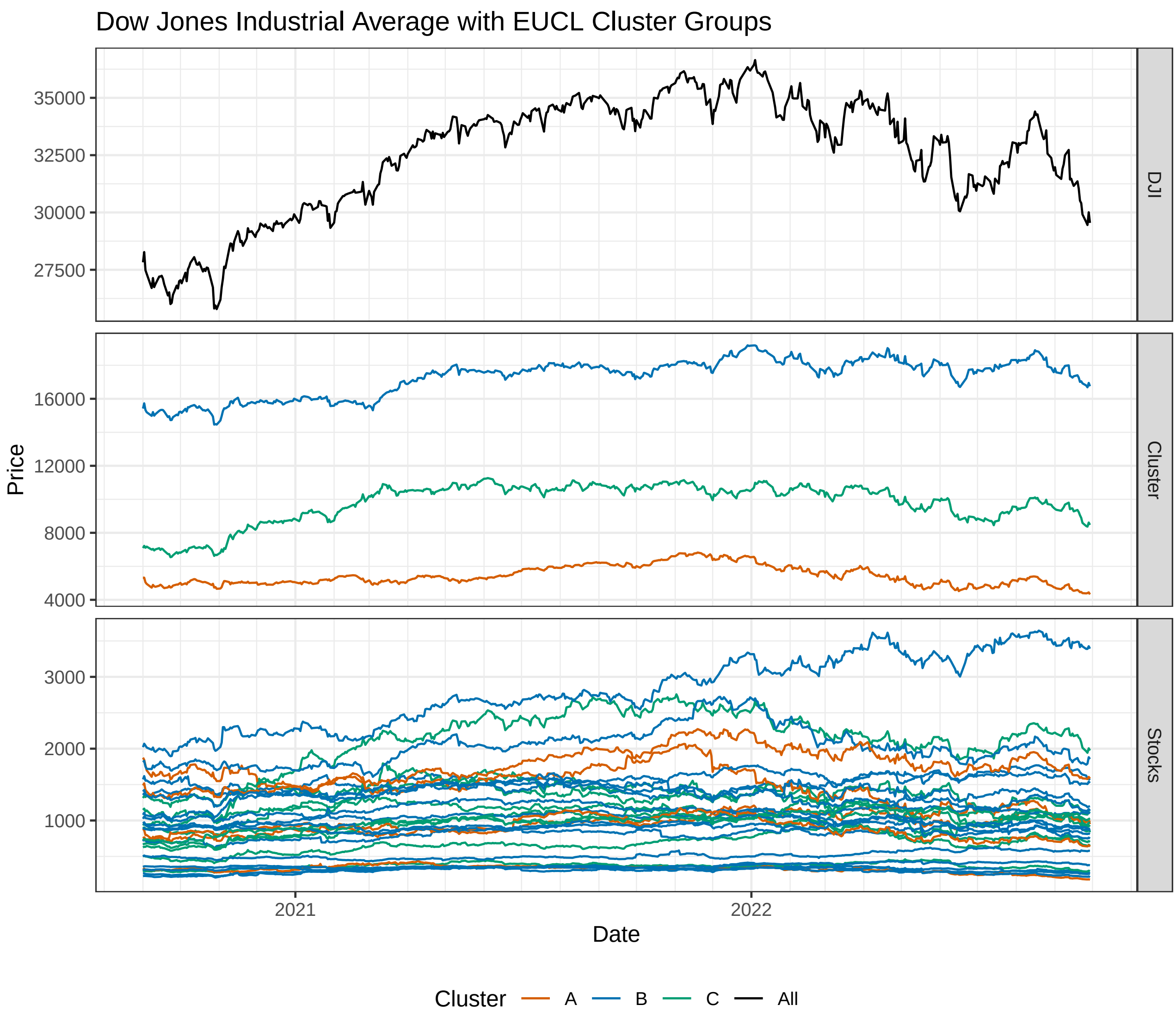

Figure 6: *Aggregated series from the hierarchical structure implied by the EUCL-based PAM clustering algorithm*

Stock	AAPL	AMGN	AXP	BA	CAT	CRM	CSCO	CVX	DIS	DOW
Cluster	A	B	C	C	C	A	A	C	C	C
Stock	GS	HD	HON	IBM	INTC	JNJ	JPM	KO	MCD	MMM
Cluster	C	A	C	C	A	B	C	B	A	B
Stock	MRK	MSFT	NKE	PG	TRV	UNH	V	VZ	WBA	WMT
Cluster	B	A	A	B	C	B	A	B	C	B

Table 2: *EUCL-based PAM clustering: partition*

4.5. Correlation-based clustering

We construct clusters of stocks based on the correlation between log-returns. The correlation-based approach is one of the most widely adopted for financial time series clustering. Figure 7 shows the ASW associated with different number of clusters, again showing $C = 3$ as the maximum value (although the ASW value is relatively lower compared to the previous observation based clustering). The resulting clusters are shown in Table 3 and the aggregated series are shown in Figure 8.

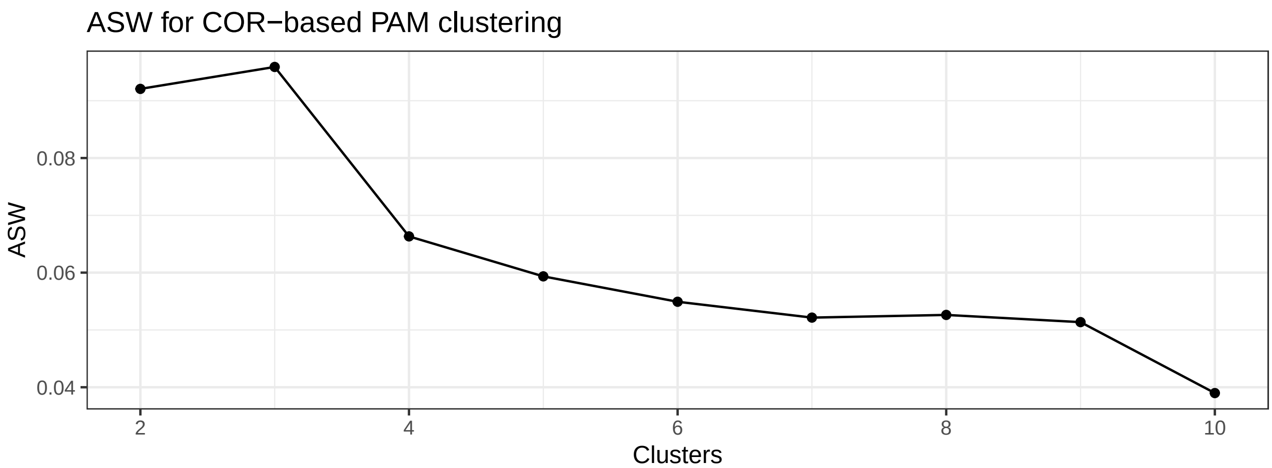

Figure 7: *Average Silhouette Width with COR-based PAM partition*

Stock	AAPL	AMGN	AXP	BA	CAT	CRM	CSCO	CVX	DIS	DOW
Cluster	A	B	C	C	C	A	B	C	C	C
Stock	GS	HD	HON	IBM	INTC	JNJ	JPM	KO	MCD	MMM
Cluster	C	B	B	B	A	B	C	B	B	B
Stock	MRK	MSFT	NKE	PG	TRV	UNH	V	VZ	WBA	WMT
Cluster	B	A	A	B	C	B	B	B	B	B

Table 3: *COR-based PAM clustering: partition*

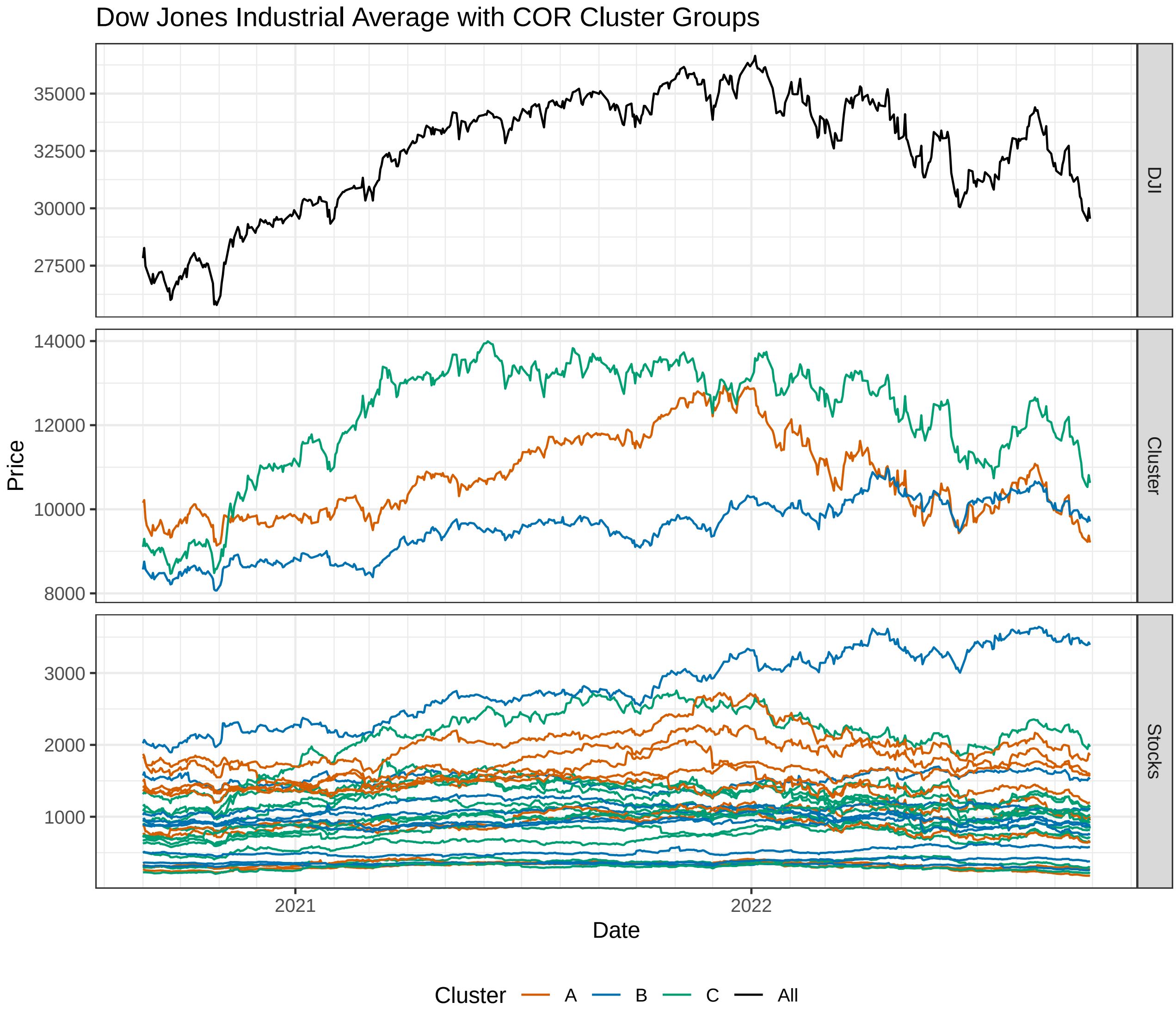

Figure 8: *Aggregated series from the hierarchical structure implied by the COR-based PAM clustering algorithm*

4.6. Model-based clustering

The first step of ARIMA-based clustering requires estimating the best fitting models within the training set. The resulting models identified using the automatic ARIMA algorithm of Hyndman and Khandakar (2008) are shown in Table 4.

The stocks with random walk models (shown as ARIMA(0,1,0)) are grouped in a separate cluster, and we apply cluster analysis on the remaining stocks using the Piccolo (1990) distance across their AR(∞) coefficients. The ASW values are shown in Figure 9, with the maximum given by $C = 4$. The resulting clusters are given in Table 5, along with the random walk cluster labeled E.

Figure 10 shows the AR(∞) weights for each stock, coloured by cluster. Cluster E includes stocks with zero coefficients (random walk processes), while Cluster D contains two stocks with persistent AR(∞) coefficients. Clusters A and C are characterized by similar patterns of the coefficients, but the parameters of stocks in Cluster A decay to zero faster than those included in Cluster C.

Figure 11 shows the aggregated series resulting from these clusters.

Table 4: *Best ARIMA models using the Hyndman-Khandakar algorithm with the AICc criterion*

Stock	Model	Stock	Model	Stock	Model
AAPL	ARIMA(3,1,1)	GS	ARIMA(0,1,1)	MRK	ARIMA(1,1,1)
AMGN	ARIMA(1,1,1)	HD	ARIMA(2,1,3)	MSFT	ARIMA(2,1,2)
AXP	ARIMA(0,1,0)	HON	ARIMA(2,1,0)	NKE	ARIMA(0,1,0)
BA	ARIMA(0,1,0)	IBM	ARIMA(2,1,0)	PG	ARIMA(1,1,0)
CAT	ARIMA(0,1,0)	INTC	ARIMA(0,1,0)	TRV	ARIMA(2,1,1)
CRM	ARIMA(2,1,1)	JNJ	ARIMA(2,1,0)	UNH	ARIMA(0,1,0)
CSCO	ARIMA(2,1,0)	JPM	ARIMA(0,1,0)	V	ARIMA(0,1,0)
CVX	ARIMA(0,1,0)	KO	ARIMA(1,1,1)	VZ	ARIMA(2,1,1)
DIS	ARIMA(0,1,0)	MCD	ARIMA(1,1,3)	WBA	ARIMA(0,1,0)
DOW	ARIMA(1,1,0)	MMM	ARIMA(1,1,2)	WMT	ARIMA(2,1,3)

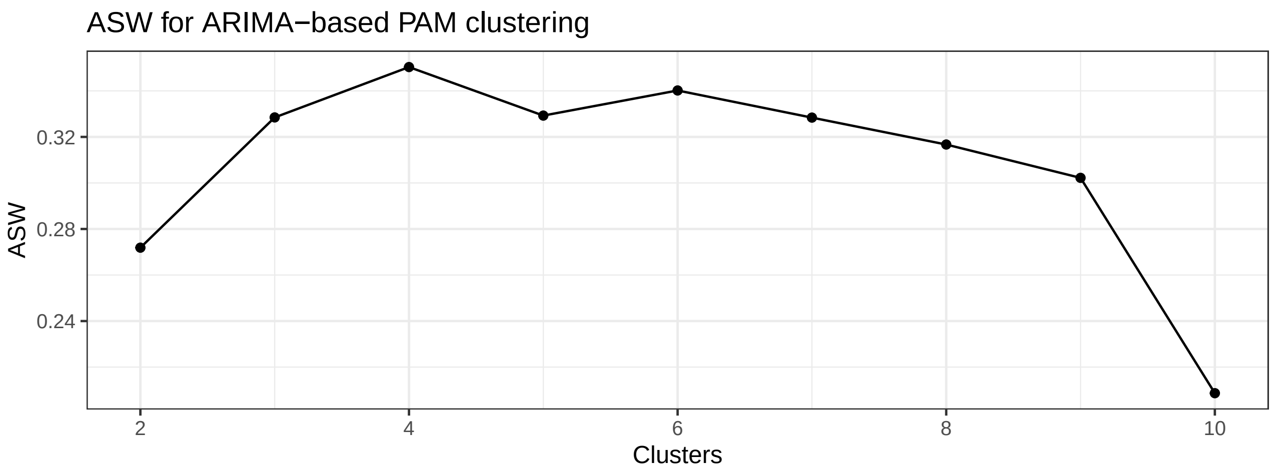

Figure 9: *Average Silhouette Width with ARIMA-based PAM partition*

Stock	AAPL	AMGN	AXP	BA	CAT	CRM	CSCO	CVX	DIS	DOW
Cluster	A	A	E	E	E	B	B	E	E	A
Stock	GS	HD	HON	IBM	INTC	JNJ	JPM	KO	MCD	MMM
Cluster	C	C	B	B	E	A	E	A	C	B
Stock	MRK	MSFT	NKE	PG	TRV	UNH	V	VZ	WBA	WMT
Cluster	A	A	E	A	D	E	E	D	E	C

Table 5: *ARMA-based PAM clustering: partition*

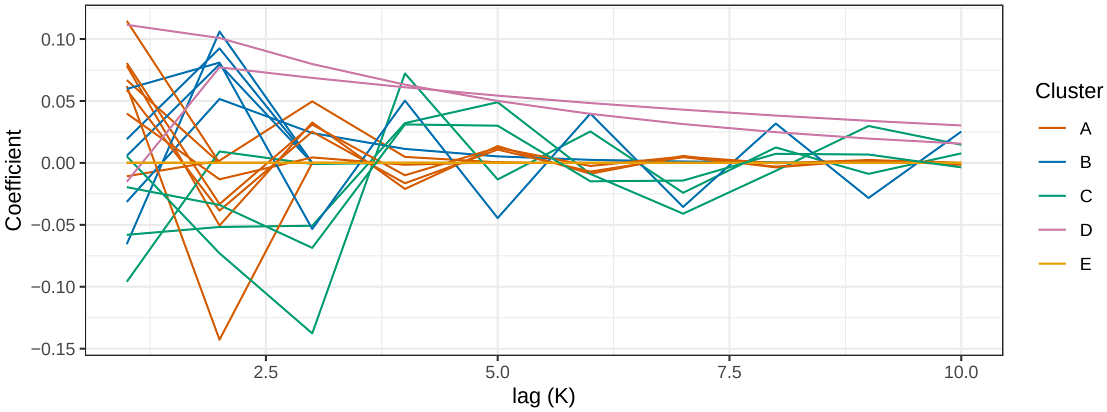

Figure 10: *$AR(\infty)$ weights of the clustered time series*

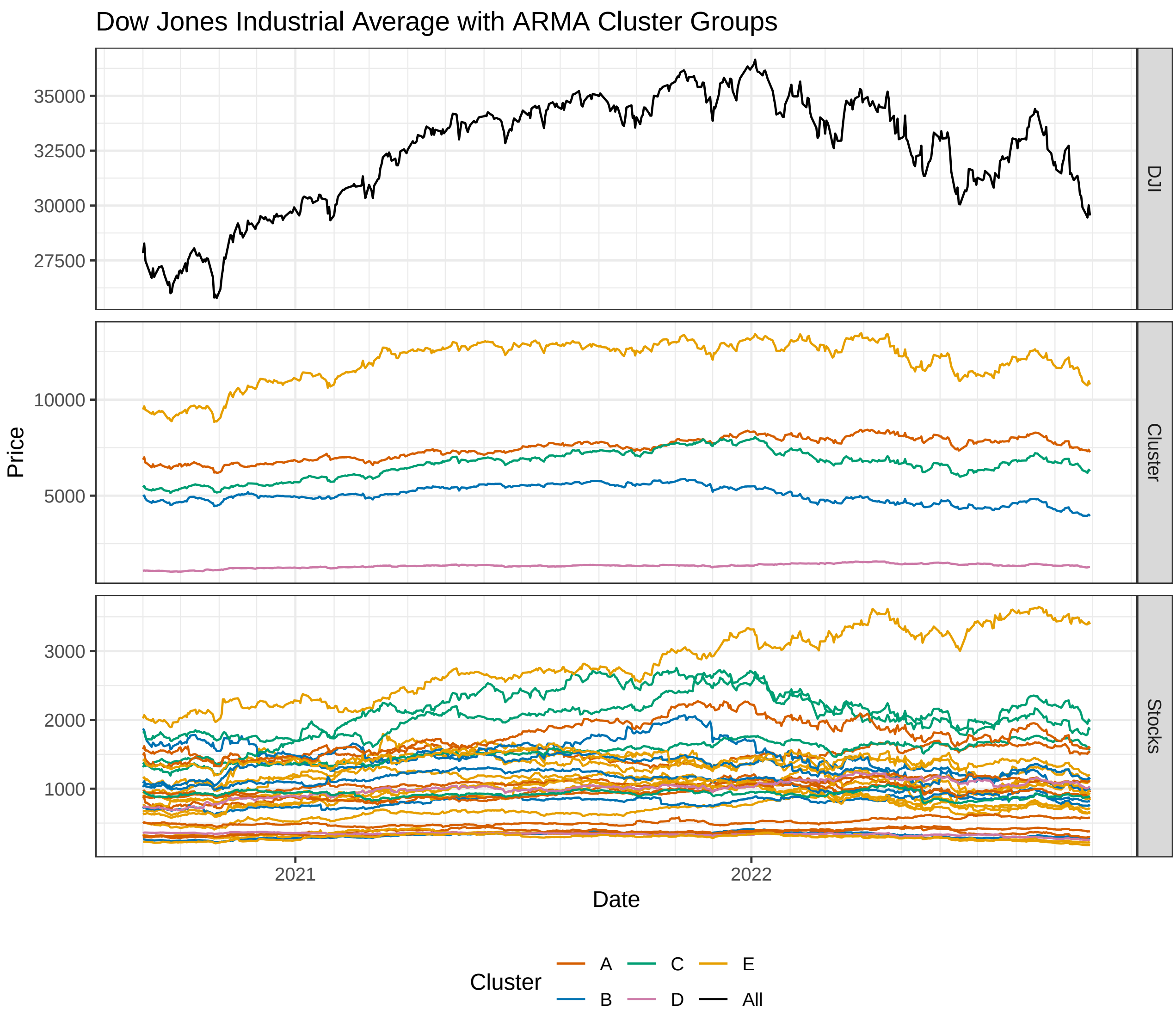

Figure 11: *Aggregated series from the hierarchical structure implied by the ARIMA-based PAM clustering algorithm*

5. Forecast reconciliation: results

In this section we provide details about the forecasting experiments. We compare the forecast accuracy of implementing forecast reconciliation using the various hierarchical structures based on clustering, versus various benchmarks. We consider as benchmarks: Base (unreconciled) forecasts; random walk (RW) forecasts, bottom-up (BU) forecasts; and MinT reconciled forecasts without using any clustering (MinT). Base and random walk forecasts provide natural benchmarks without reconciliation. BU shows the usefulness of MinT reconciliation versus using a traditional the single level approach, while MinT without clustering allows us to evaluate the usefulness of clustering within the MinT framework. In Section 5.1 we discuss the results in terms of the market index, while in Section 5.2 we analyze the results for the common individual stocks included in the index.

5.1. Forecasting top series: Dow Jones Index Average

Table 6 shows the out-of-sample average errors at different forecasting horizons. Panel A shows the accuracy metric in terms of absolute errors, while Panel B is in terms of squared errors. In each case we are always able to find a reconciliation approach that is more accurate than the considered benchmarks, although the rankings of the reconciliation approaches vary. This indicates that reconciliation can be successfully employed to improve forecasts of the stock market index. In terms of MAE loss, the reconciliation approaches combining all the different clustering structures — denoted by MinT: ALL — provides the most accurate forecasts for $h = 1, 6, 12$ horizons. Exchange-based clustering dominates the alternatives for $h = 3$ in terms of MAE loss and is the best approach for $h = 3, 6$ in terms of RMSE. ARIMA-based clustering provides the most accurate out-of-sample forecasts considering RMSE for

Panel A: MAE loss	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Base (unreconciled)	333.65	625.31	925.13	1436.55
RW	320.15	588.88	889.49	1344.02
BU	321.12	588.62	873.97	1309.89
MinT	320.62	588.39	873.20	1307.99
MinT: IND	320.59	589.20	871.64	1299.28
MinT: EXCH	320.02	**588.39**	872.82	1308.80
MinT: EUCL	321.00	588.50	874.69	1302.19
MinT: COR	320.88	589.20	872.10	1302.65
MinT: ARMA	319.40	590.07	873.83	1307.92
MinT: ALL	**319.34**	590.92	**868.85**	**1288.51**
Panel B: RMSE loss	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Base (unreconciled)	439.87	794.42	1212.65	1831.28
RW	424.80	749.47	1091.46	1570.48
BU	426.09	747.00	1076.13	1517.53
MinT	425.86	746.92	1074.50	1513.70
MinT: IND	426.24	750.88	1079.38	1512.85
MinT: EXCH	424.37	**745.96**	**1073.09**	1513.67
MinT: EUCL	426.32	748.78	1076.69	1510.22
MinT: COR	426.33	748.72	1075.55	**1508.20**
MinT: ARMA	**424.86**	747.45	1074.01	1511.86
MinT: ALL	424.93	752.71	1078.47	1508.26

Table 6: *Accuracy metric: MAE and RMSE results. Best model with bold font.*

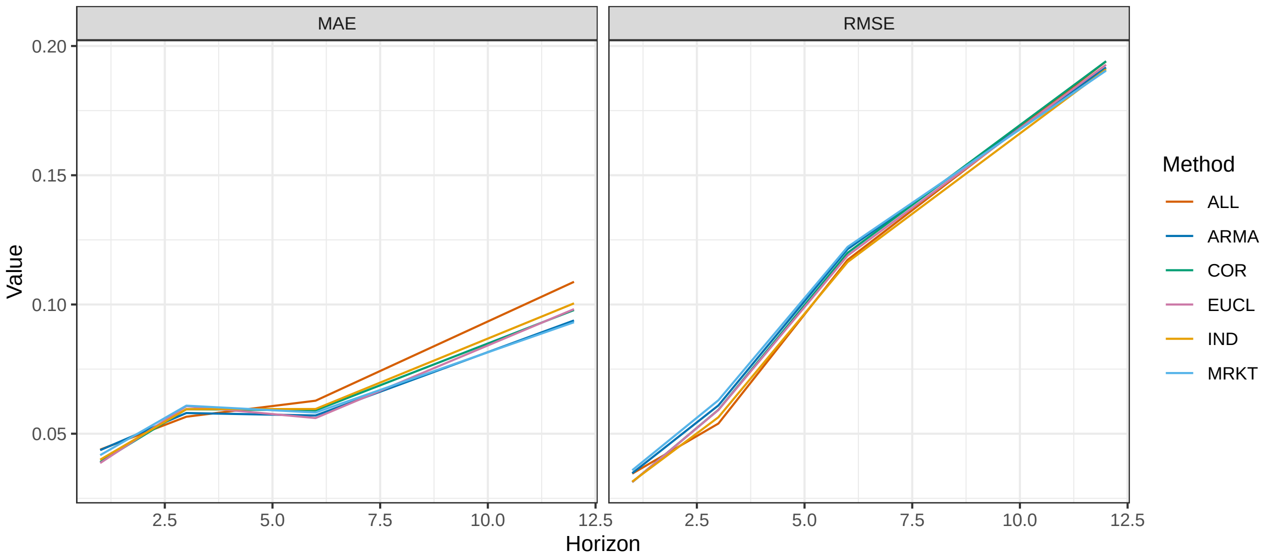

Figure 12: *Average relative accuracy of reconciliation methods compared to the base unreconciled forecasts. Values above zero indicate higher (an improvement in) forecast accuracy. Left: Forecast accuracy is given by the log of the Mean Absolute Errors (MAE) of the reconciled forecasts relative to the base forecasts. Right: Forecast accuracy is given by the log of the Root Mean Squared Errors (RMSE) of the reconciled forecasts relative to the base forecasts.*

$h = 1$ step ahead. Correlation-based clustering seems to be the best approach for $h = 12$ under RMSE loss. Hence, there is no clear clustering structure dominating the others across all forecasting horizons and error measures. However, MinT: ALL provides the most consistent results in terms of MAE loss. As this approach also represents a simple way of hedging against the wrong specification of the clustering structure, we consider this approach as the best among the considered alternatives.

The most important result, overall, is that reconciliation provides more accurate forecasts than unreconciled approaches. Figure 12 shows the average relative accuracy of reconciliation methods in terms of MAE and RMSE, respectively. Values greater than zero indicate higher forecast accuracy relative to the base unreconciled forecasts. In particular, we observe how the benefit of reconciliation becomes larger with increasing forecasting horizon h. For $h = 1$ step forecasts, the benefit of reconciliation is around 5%, while for $h = 12$ it is larger than 10% for MAE loss and around 20% under RMSE loss. Moreover, we notice that with increasing forecasting horizons, all the reconciliation based forecasts provide similar improvements in accuracy compared with base forecasts.

Table 7 shows the results of predictive accuracy tests, with absolute errors and squared errors, respectively. The p-values of the modified Diebold and Mariano (2002) test of van Dijk and Franses (2003) are reported. Under the null hypothesis, the average error of the benchmark, shown across the columns, is equal to the average error from the reconciliation approach, shown down the rows. The results in terms of standard Diebold and Mariano (2002), which provide similar results, are shown in the Appendix.

Base. The first important result to highlight is that for most forecast horizons the reconciliation approaches provide statistically more accurate forecasts than the base benchmark, regardless of the loss employed for the tests. This suggests that, whatever the forecasting horizon is, forecast reconciliation should be used to improve forecasting of the market index.

Absolute errors	Base	RW	BU	MinT	**Squared errors**	Base	RW	BU	MinT
				$h = 1$ step ahead					
MinT: IND	**0.02**	0.96	0.09	0.27	MinT: IND	**0.02**	0.99	0.46	0.51
MinT: EXCH	**0.03**	0.60	**0.04**	**0.10**	MinT: EXCH	**0.02**	0.22	**0.08**	**0.09**
MinT: EUCL	**0.03**	0.95	**0.03**	0.11	MinT: EUCL	**0.02**	0.98	0.16	0.16
MinT: COR	**0.03**	0.99	0.39	0.80	MinT: COR	**0.02**	1.00	0.93	0.96
MinT: ARMA	**0.02**	0.13	**0.00**	**0.00**	MinT: ARMA	**0.02**	0.65	**0.00**	**0.00**
MinT: ALL	**0.02**	**0.06**	**0.01**	**0.03**	MinT: ALL	**0.02**	0.12	**0.04**	**0.05**
				$h = 3$ step ahead					
MinT: IND	**0.02**	0.19	0.12	0.27	MinT: IND	**0.02**	0.32	0.79	0.87
MinT: EXCH	**0.02**	0.33	**0.04**	0.29	MinT: EXCH	**0.02**	0.15	**0.07**	0.13
MinT: EUCL	**0.02**	0.28	**0.02**	0.12	MinT: EUCL	**0.02**	0.21	0.36	0.72
MinT: COR	**0.02**	0.31	**0.06**	0.40	MinT: COR	**0.02**	0.20	0.39	0.74
MinT: ARMA	**0.02**	0.43	0.56	0.73	MinT: ARMA	**0.02**	0.13	0.47	0.66
MinT: ALL	**0.02**	0.14	0.15	0.23	MinT: ALL	**0.02**	0.23	0.36	0.43
				$h = 6$ step ahead					
MinT: IND	**0.02**	**0.01**	**0.06**	**0.05**	MinT: IND	**0.02**	**0.02**	**0.09**	0.27
MinT: EXCH	**0.03**	**0.06**	0.11	**0.08**	MinT: EXCH	**0.02**	**0.04**	**0.03**	**0.08**
MinT: EUCL	**0.03**	**0.05**	0.53	0.73	MinT: EUCL	**0.02**	**0.04**	0.40	0.90
MinT: COR	**0.02**	**0.04**	**0.10**	**0.09**	MinT: COR	**0.02**	**0.03**	**0.06**	0.26
MinT: ARMA	**0.02**	**0.03**	0.46	0.57	MinT: ARMA	**0.02**	**0.02**	0.20	0.54
MinT: ALL	**0.02**	**0.01**	**0.06**	**0.06**	MinT: ALL	**0.02**	**0.02**	**0.09**	0.15
				$h = 12$ step ahead					
MinT: IND	**0.01**	**0.04**	**0.10**	**0.10**	MinT: IND	**0.01**	**0.01**	0.19	0.32
MinT: EXCH	**0.01**	0.14	0.57	0.96	MinT: EXCH	**0.01**	**0.03**	0.19	0.63
MinT: EUCL	**0.01**	0.11	**0.01**	**0.00**	MinT: EUCL	**0.01**	**0.03**	**0.01**	**0.00**
MinT: COR	**0.01**	**0.09**	**0.09**	**0.06**	MinT: COR	**0.01**	**0.02**	**0.02**	**0.01**
MinT: ARMA	**0.01**	0.11	0.52	0.61	MinT: ARMA	**0.01**	**0.02**	0.20	0.33
MinT: ALL	**0.01**	**0.01**	**0.04**	**0.03**	MinT: ALL	**0.01**	**0.00**	0.12	0.17

Table 7: *P-values from the modified Diebold and Mariano (2002) test of van Dijk and Franses (2003). Under the null, the difference in the forecasting errors is equal to zero. Entries in bold indicate a rejection of the null at a 10% level of significance.*

Random Walk. Not all the reconciliation approaches provide statistically significant improvements in forecasting the top series compared with the naïve approach. Nevertheless, under absolute error loss for $h = 1, 6, 12$, MinT: ALL reconciliation, results to statistically more accurate forecasts compared to the random walk. Some statistically significant improvements are also observed for other approaches, especially for the longer forecast horizons. The results show that random walk forecasts remain a valid (but less accurate) alternative, but only for shorter forecast horizons and under squared loss. In the other settings, reconciliation provides statistically more accurate forecasts.

Bottom-up. For all scenarios we are able to find a clustering-based MinT reconciliation approach which performs statistically better than bottom-up. The strongest results are shown for MinT: ALL, which provides statistically more accurate forecasts than bottom-up for $h = 1, 6, 12$ for absolute loss and for $h = 1, 6$ for squared loss. The overall evidence suggests that the clustering-based reconciliation approaches provide more accurate forecasts than bottom-up.

MinT without clustering. In general, we observe that clustering-based approaches are more accurate than when no clustering is considered, with many statistically significant entries in the MinT columns. In particular, under absolute error loss for $h = 1, 6, 12$ the MinT: ALL approach provides the more accurate forecasts (as shown in Table 6) which are also statistically significantly better than all the benchmarks, including MinT with no clustering. Using squared loss, MinT: ARMA and MinT: ALL provide statistically lower forecast errors than MinT with no clustering, while for $h = 3$, the best model is MinT: EXCH which gives statistically lower forecast errors. For $h = 12$ the MinT: COR is the most accurate (as shown in Table 6) and statistically better than all the benchmarks including MinT with no clustering.

These findings provide strong evidence that not only is MinT reconciliation useful for financial forecasting, but considering and exploring clustering structures in stocks, allows further improvements for forecasting the market index.

5.2. *Forecasting bottom series: common stocks*

In this section we evaluate the effect of forecast reconciliation in improving the accuracy of forecasting the individual stock prices. As these form the bottom-level series we remove from the benchmarks the bottom-up forecasts, as these are identical to the Base (unreconciled) forecasts.

Table 8 shows the results in terms of average loss across the 30 stocks included the index. Regardless of the loss and the forecast horizon, we are always able to find a reconciliation approach that performs better than the base forecasts. This indicates that forecast reconciliation can be successfully employed for more accurately forecasting both the index and its constituents. Most clustering approaches improve on MinT with no clustering. Considering squared loss, the random walk provides the most

Panel A: average MAE loss	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Base (unreconciled)	15.40	27.65	39.81	57.11
RW	15.35	**27.55**	39.77	57.38
MinT	15.40	27.64	39.77	57.01
MinT: IND	15.40	27.64	39.77	56.85
MinT: EXCH	15.38	27.61	**39.70**	56.89
MinT: EUCL	15.42	27.65	39.77	56.86
MinT: COR	15.41	27.66	39.75	56.86
MinT: ARMA	**15.37**	27.61	39.72	**56.80**
MinT: ALL	17.78	28.50	42.72	58.52
Panel B: average RMSE loss	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Base (unreconciled)	20.23	35.72	50.49	70.57
RW	**20.14**	**35.59**	50.37	70.73
MinT	20.23	35.71	50.42	70.38
MinT: IND	20.23	35.77	50.38	69.99
MinT: EXCH	20.18	35.65	50.25	70.09
MinT: EUCL	20.24	35.73	50.37	70.09
MinT: COR	20.24	35.75	50.34	70.06
MinT: ARMA	20.20	35.65	**50.21**	**69.90**
MinT: ALL	22.79	35.74	51.30	77.37

Table 8: *Accuracy metric: MAE and RMSE results. Best model with bold font.*

Absolute errors	Base	RW	MinT	**Squared errors**	Base	RW	MinT
			$h = 1$ step ahead				
MinT: IND	**0.08**	1.00	0.19	MinT: IND	0.42	1.00	0.64
MinT: EXCH	**0.05**	1.00	0.12	MinT: EXCH	**0.08**	1.00	0.13
MinT: EUCL	0.66	1.00	0.79	MinT: EUCL	0.64	1.00	0.84
MinT: COR	0.34	1.00	0.59	MinT: COR	0.70	1.00	0.96
MinT: ARMA	**0.00**	0.99	**0.00**	MinT: ARMA	**0.00**	1.00	**0.00**
MinT: ALL	0.15	0.88	0.19	MinT: ALL	0.34	0.99	0.42
			$h = 3$ step ahead				
MinT: IND	0.27	1.00	0.36	MinT: IND	0.81	1.00	0.85
MinT: EXCH	**0.06**	0.99	0.12	MinT: EXCH	**0.06**	0.95	**0.10**
MinT: EUCL	0.45	0.99	0.53	MinT: EUCL	0.47	0.98	0.57
MinT: COR	0.68	1.00	0.80	MinT: COR	0.77	1.00	0.87
MinT: ARMA	**0.10**	1.00	0.12	MinT: ARMA	**0.01**	0.97	**0.01**
MinT: ALL	0.47	0.95	0.50	MinT: ALL	0.81	0.99	0.84
			$h = 6$ step ahead				
MinT: IND	0.16	0.80	0.34	MinT: IND	0.31	0.71	0.51
MinT: EXCH	**0.02**	0.70	**0.06**	MinT: EXCH	**0.03**	0.56	**0.06**
MinT: EUCL	0.14	0.80	0.37	MinT: EUCL	**0.01**	0.62	**0.05**
MinT: COR	0.12	0.77	0.25	MinT: COR	**0.03**	0.61	**0.08**
MinT: ARMA	**0.01**	0.72	**0.02**	MinT: ARMA	**0.00**	0.48	**0.00**
MinT: ALL	**0.10**	0.61	0.14	MinT: ALL	0.16	0.59	0.26
			$h = 12$ step ahead				
MinT: IND	0.11	0.28	0.34	MinT: IND	**0.01**	0.32	**0.09**
MinT: EXCH	**0.00**	0.29	**0.02**	MinT: EXCH	**0.01**	0.35	**0.03**
MinT: EUCL	**0.00**	0.26	**0.00**	MinT: EUCL	**0.00**	0.31	**0.00**
MinT: COR	**0.02**	0.25	**0.06**	MinT: COR	**0.00**	0.29	**0.00**
MinT: ARMA	**0.00**	0.22	**0.00**	MinT: ARMA	**0.00**	0.26	**0.00**
MinT: ALL	**0.03**	0.17	**0.06**	MinT: ALL	**0.01**	0.26	**0.03**

Table 9: *P-values from the standard Diebold and Mariano (2002) test. Under the null, the difference in the forecasting errors is equal to zero. Entries in bold indicate a rejection of the null at a 10% level of significance.*

accurate forecasts for short horizons, $h = 1, 3$, while MinT: ARMA provides the most accurate forecasts for the longer horizons $h = 6, 12$. Considering absolute error loss, the MinT reconciliation approaches are generally more accurate than the random walk.

Table 9 shows the results of predictive accuracy tests, with absolute errors and squared errors, respectively. Under the null hypothesis, the average error of the benchmark, shown across the columns, is equal to the average error from the reconciliation approach, shown down the rows. In contrast to the tests applied to the market index, here we report the results of the Diebold and Mariano (2002) test, because average errors from different time series are considered. The results of the modified test van Dijk and Franses (2003) are reported in the Appendix for the sake of consistency.

Base. In general, the MinT reconciliation approaches including clustering, provide statistically significant improvements over the base forecasts for both absolute and squared error losses. The number of statistically significant improvements seems to increase as the forecast horizon increases. In particular, for $h = 12$ all the clustering-based MinT approaches, with the exception of MinT: IND for absolute loss, provide statistically more accurate forecasts than base forecasts, which do not involve reconciliation. Hence, the results send a strong signal that forecast reconciliation improves the forecast accuracy for individual stock price series.

Random Walk. Most reconciliation procedures allow reducing the forecast error compared with a random walk, especially for longer forecast horizons. However, in contrast to forecasting the market index, we do not find any statistically significant differences in forecast accuracy between the reconciliation approaches and the random walk for forecasting the individual stock price series. That is, we are not able to reject the null hypothesis of the test for any of the forecasting horizons h and error losses.

MinT without clustering. In all cases, there exist clustering-based reconciliation approaches that generate statistically significant forecast accuracy improvements compared to reconciliation with no clustering. In particular, MinT: ARMA and MinT: EXCH consistently provide statistically more accurate forecasts for all forecast horizons and error losses. Similar to the comparison with the base forecasts, the number of statistically significant improvements increases as the forecast horizon increases. Hence, coupling forecast reconciliation with clustering structures seems to also be beneficial for improving on average the forecasting accuracy for the bottom-level series.

6. Investing with forecast reconciliation

Forecasts are commonly used for constructing profitable trading strategies in finance. For this reason, an alternative approach of comparing forecast methods is by studying the profitability of investment strategies based on the obtained forecasts. While the previous section highlights the advantageous properties of forecast reconciliation and clustering in reducing out-of-sample forecast errors, in this section we evaluate the performance of the developed approach from a financial investment viewpoint.

We compare the performance of two investment strategies that involve buying the market index if the price is predicted to increase in the next h time periods, and selling it otherwise. The two strategies differ in the forecast method employed. In particular, we compare the performance of alternative reconciliation and clustering approaches against the base ARIMA model, bottom-up reconciliation and MinT without clustering.

Let us define as $\hat{y}_{t+h}$ the market index price forecast at time $t + h$ and as y_t the actual price at time t. The predicted return at time $t + h$ implied by the price forecast is computed as,

$$\hat{x}_{t+h} = \frac{\hat{y}_{t+h} - y_t}{y_t}, \tag{18}$$

while the actual return observed at time $t + h$ is given by,

$$x_{t+h} = \frac{y_{t+h} - y_t}{y_t}. \tag{19}$$

Following Anatolyev and Gerko (2005), the realized return of this investment strategy at time $t + h$ is given by,

$$r_{t+h} = \operatorname{sign}(\hat{x}_{t+h}) x_{t+h}, \tag{20}$$

where $\operatorname{sign}(\hat{x}_{t+h})$ is a function taking the value of 1 if $\hat{x}_{t+h} \geq 0$ and -1 otherwise. Therefore, assuming that short selling is allowed, investors realize positive returns if they buy the market index at time t and its price increases at time $t + h$, but also if they sell it at time t and the price is lower at time $t + h$.

Given two alternative forecasting methods, A and B, we construct two alternative investment strategies with different ex-post realized returns, $r_{A,t+h}$ and $r_{B,t+h}$. We compare the forecast methods in terms of their implied financial performance, measured by the Sharpe ratio (Sharpe, 1963), which is a commonly used metric for evaluating the performance of an investment strategy in terms of its return/risk trade-off. The Sharpe ratio of the strategy employing forecasts from method A is given by,

$$\mathrm{SR}_A = \frac{\hat{\mu}_A}{\hat{\sigma}_A}, \tag{21}$$

where $\hat{\mu}_A$ is the average return of $r_{A,t+h}$ and $\hat{\sigma}_A$ is its risk, computed with the standard deviation. The Sharpe ratio of the strategy using forecast from method B is similarly defined. We then test if the two strategies lead to statistically different Sharpe ratios using the procedure proposed by Jobson and Korkie (1981) and Memmel (2003).

Table 10 shows the difference in Sharpe ratio of investment strategies based on alternative forecasting methods. We report the results for the top-level series for the out-of-sample forecasting evaluation as defined in Section 3. The results are evaluated at forecast horizons $h = 1, 3, 6, 12$. We consider base ARIMA, bottom-up, and MinT reconciliation without clustering as the benchmarks. The alternative forecasting methods based on MinT reconciliation under different clustering structures are presented in the rows. Positive entries indicate that the investment strategy based on the forecasting approach in the row provides a higher Sharpe ratio than the benchmark considered in the column.

First, we highlight that for $h = 1$ and $h = 6$ step-ahead forecasts, the investment strategy based on MinT: ALL forecasts is the only one that provides a statistically higher Sharpe ratio compared with base ARIMA forecasts. The Sharpe ratio is 14% higher than the one obtained using ARIMA-based forecasts.For the other forecast horizons, we do not reject the null hypothesis of the test. The results obtained for $h = 1$ step-ahead forecasts are arguably the most interesting from a financial viewpoint, as trading strategies are usually constructed and evaluated considering short-term forecasts. Forecasts

$h = 1$	Base	BU	MinT	$h = 3$	Base	BU	MinT
MinT: IND	-4.16	**14.30**	**11.40**	MinT: IND	-2.40	-2.07	-2.07
MinT: EXCH	-8.87	**9.59**	**6.69**	MinT: EXCH	-2.43	-2.43	-2.43
MinT: EUCL	-11.47	**6.99**	**4.09**	MinT: EUCL	-0.32	0.00	0.00
MinT: COR	-15.43	**3.03**	0.13	MinT: COR	-0.32	0.00	0.00
MinT: ARMA	-0.12	**18.34**	**15.44**	MinT: ARMA	0.36	**0.68**	**0.68**
MinT: ALL	**13.97**	**32.44**	**29.54**	MinT: ALL	-2.40	-2.07	-2.07
$h = 6$	Base	BU	MinT	$h = 12$	Base	BU	MinT
MinT: IND	-2.37	-0.87	-0.87	MinT: IND	-3.57	0.00	0.00
MinT: EXCH	0.22	0.00	0.00	MinT: EXCH	0.52	0.00	0.00
MinT: EUCL	-1.70	-0.20	-0.20	MinT: EUCL	-1.39	**2.18**	**2.18**
MinT: COR	0.22	**1.72**	**1.72**	MinT: COR	-3.57	0.00	0.00
MinT: ARMA	-1.50	0.00	0.00	MinT: ARMA	-3.57	0.00	0.00
MinT: ALL	**1.53**	**3.03**	**3.03**	MinT: ALL	-1.39	**2.18**	**2.18**

Table 10: *Difference of Sharpe ratios (%) between two forecast-based investment strategies. Positive values indicate that the forecasting method in the row provides a higher Sharpe ratio than the benchmark method in the column. Entries in bold indicate a rejection of the null hypothesis at a 10% level of significance.*

about financial prices are indeed less accurate for longer horizons, as we also have highlighted in the previous section.

Another interesting result in Table 10 is that, for $h = 1$, all the reconciliation approaches provide statistically higher Sharpe ratios compared with both bottom-up and MinT reconciliation without clustering. Therefore, the combined use of clustering and MinT reconciliation is also useful in constructing more profitable trading strategies. Specifically, MinT: ALL has a Sharpe ratio 32.4% higher than the one obtained with bottom-up forecasts, and 29.5% higher than forecasts based on MinT reconciliation without clustering.

Considering longer horizons h, we find a reconciliation procedure outperforming the bottom-up and MinT benchmarks in all cases. In particular, MinT: ALL is the best model in most scenarios, namely $h = 1, 6$ and 12. For $h = 3$ forecasting horizon, MinT: ARMA is the only model providing a statistically higher Sharpe ratio than bottom-up and MinT reconciliation without clustering.

7. Conclusions

The objective of this research has been to explore the potential benefits of employing forecast reconciliation for forecasting stock market indexes and their underlying constituents. Both meta-data groups and empirical clustering techniques have been used to determine the underlying structure of the price time series. The study makes two contributions. First, to the best of our knowledge, it applies forecast reconciliation to the financial domain for the first time. Second, it combines cluster analysis with forecast reconciliation. This approach offers insights into the efficacy of reconciliation within the context of latent hierarchical structures.

To evaluate our proposed approach, we employ the Dow Jones Industrial Average index and its constituents as the basis of our empirical experiment. ARIMA models are used to generate forecasts for the time series using a rolling-window procedure. The reconciliation approach combines MinT reconciliation (Wickramasuriya et al., 2019) with cluster analysis. Three different dissimilarity measures are utilized for PAM-based clustering: raw returns, return correlation, and ARIMA distance, as proposed by Piccolo (1990). We show how all the clusters can be used simultaneously within a forecast reconciliation context. Furthermore, we investigate the usefulness of clustering by considering a reconciliation approach without a clustering structure, where stocks aggregate directly to the market index.

We evaluate the usefulness of reconciliation and clustering in terms of out-of-sample forecasting accuracy. Our results suggest that reconciliation is a useful tool for forecasting both the stock market index and its underlying constituents, even without clustering. But with the clustering of stocks included in the reconciliation procedure, even better forecasts are obtained. We also evaluate the usefulness of forecast reconciliation and clustering from a financial viewpoint, considering investment strategies build on alternative forecasts. By comparing the out-of-sample Sharpe ratios associated with alternative reconciliation procedures at different forecasting horizons, we find that MinT: ALL (i.e., MinT reconciliation combining different clustering structures) provides the best financial performance compared with other approaches. Thus we show that reconciliation and clustering can be successfully used for constructing profitable trading strategies based on forecasts.

References

Alves, R. P., de Brito, D. S., Medeiros, M. C. and Ribeiro, R. M. (2023), 'Forecasting large realized covariance matrices: The benefits of factor models and shrinkage', *Journal of Financial Econometrics* .

Anatolyev, S. and Gerko, A. (2005), 'A trading approach to testing for predictability', *Journal of Business & Economic Statistics* **23**(4), 455–461.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M. and Perona, I. (2013), 'An extensive comparative study of cluster validity indices', *Pattern Recognition* **46**(1), 243–256.

Athanasopoulos, G., Ahmed, R. A. and Hyndman, R. J. (2009), 'Hierarchical forecasts for Australian domestic tourism', *International Journal of Forecasting* **25**(1), 146–166.

Athanasopoulos, G., Gamakumara, P., Panagiotelis, A., Hyndman, R. J. and Affan, M. (2020), Hierarchical forecasting, *in* P. Fuleky, ed., 'Macroeconomic Forecasting in the Era of Big Data. Advanced Studies in Theoretical and Applied Econometrics.', vol 52 edn, Springer, Cham, chapter 21, pp. 689–719.

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N. and Panagiotelis, A. (2023), Forecast reconciliation: A review.
URL: *robjhyndman.com/publications/hfreview.html*

Barnett, W. A. and Serletis, A. (2000), 'Martingales, nonlinearity, and chaos', *Journal of Economic Dynamics and Control* **24**(5-7), 703–724.

Batool, F. and Hennig, C. (2021), 'Clustering with the average silhouette width', *Computational Statistics & Data Analysis* **158**, 107190.

Bisaglia, L., Di Fonzo, T. and Girolimetto, D. (2020), Fully reconciled GDP forecasts from income and expenditure sides, *in* A. Pollice, N. Salvati and F. Schirripa Spagnolo, eds, 'Book of Short Papers SIS 2020', Pearson, pp. 951–956.

Blaskowitz, O. and Herwartz, H. (2011), 'On economic evaluation of directional forecasts', *International Journal of Forecasting* **27**(4), 1058–1065.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M. (2016), *Time Series Analysis: Forecasting and Control*, 5th edn, John Wiley and Sons.

Brida, J. G. and Risso, W. A. (2010), 'Hierarchical structure of the German stock market', *Expert Systems with Applications* **37**(5), 3846–3852.

Brown, S. J. and Goetzmann, W. N. (1997), 'Mutual fund styles', *Journal of Financial Economics* **43**(3), 373–399.

Brown, S. J., Goetzmann, W. N. and Kumar, A. (1998), 'The Dow theory: William peter hamilton's track record reconsidered', *The Journal of Finance* **53**(4), 1311–1333.

Brown, S. J., Lajbcygier, P. and Wong, W. W. (2012), 'Estimating the cost of capital with basis assets', *Journal of Banking & Finance* **36**(11), 3071–3079.

Caporin, M., Di Fonzo, T. and Girolimetto, D. (2023), Exploiting intraday decompositions in realized volatility forecasting: A forecast reconciliation approach.
URL: *https://arxiv.org/abs/2306.02952*

Cenesizoglu, T. and Timmermann, A. (2012), 'Do return prediction models add economic value?', *Journal of Banking & Finance* **36**(11), 2974–2987.

Chen, J., Jiang, F. and Tu, J. (2015), 'Asset allocation in the Chinese stock market: The role of return predictability', *Journal of Portfolio Management* **41**(5), 71–83.

Daniel, K. and Moskowitz, T. J. (2016), 'Momentum crashes', *Journal of Financial Economics* **122**(2), 221–247.

Darrough, M. N. and Russell, T. (2002), 'A positive model of earnings forecasts: Top down versus bottom up', *Journal of Business* **75**(1), 127–152.

Di Fonzo, T. and Girolimetto, D. (2023), 'Spatio-temporal reconciliation of solar forecasts', *Solar Energy* **251**, 13–29.

Diebold, F. X. and Mariano, R. S. (2002), 'Comparing predictive accuracy', *Journal of Business & Economic Statistics* **20**(1), 134–144.

Dunn, D. M., Williams, W. H. and DeChaine, T. (1976), 'Aggregate versus subaggregate models in local area forecasting', *Journal of the American Statistical Association* **71**(353), 68–71.

Eckert, F., Hyndman, R. J. and Panagiotelis, A. (2021), 'Forecasting Swiss exports using Bayesian forecast reconciliation', *European Journal of Operational Research* **291**(2), 693–710.

Fama, E. F. (1995), 'Random walks in stock market prices', *Financial Analysts Journal* **51**(1), 75–80.

Green, J., Hand, J. R. and Zhang, X. F. (2013), 'The supraview of return predictive signals', *Review of Accounting Studies* **18**(3), 692–730.

Gross, C. W. and Sohl, J. E. (1990), 'Disaggregation methods to expedite product line forecasting', *Journal of Forecasting* **9**(3), 233–254.

Hollyman, R., Petropoulos, F. and Tipping, M. E. (2021), 'Understanding forecast reconciliation', *European Journal of Operational Research* **294**(1), 149–160.

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G. and Shang, H. L. (2011), 'Optimal combination forecasts for hierarchical time series', *Computational Statistics & Data Analysis* **55**(9), 2579–2589.

Hyndman, R. J. and Athanasopoulos, G. (2021), *Forecasting: Principles and Practice*, 3rd edn, OTexts, Melbourne, Australia.
URL: *OTexts.com/fpp3*

Hyndman, R. J. and Khandakar, Y. (2008), 'Automatic time series forecasting: the forecast package for R', *Journal of Statistical Software* **27**, 1–22.

Hyndman, R. J., Lee, A. J. and Wang, E. (2016), 'Fast computation of reconciled forecasts for hierarchical and grouped time series', *Computational Statistics & Data Analysis* **97**, 16–32.

Jeon, J., Panagiotelis, A. and Petropoulos, F. (2019), 'Probabilistic forecast reconciliation with applications to wind power and electric load', *European Journal of Operational Research* **279**(2), 364–379.

Jobson, J. D. and Korkie, B. M. (1981), 'Performance hypothesis testing with the sharpe and treynor measures', *Journal of Finance* pp. 889–908.

Kaufman, L. and Rousseeuw, P. J. (1990), *Finding groups in data: an introduction to cluster analysis*, John Wiley & Sons.

Kim, J. H., Shamsuddin, A. and Lim, K.-P. (2011), 'Stock return predictability and the adaptive markets hypothesis: Evidence from century-long US data', *Journal of Empirical Finance* **18**(5), 868–879.

King, B. F. (1966), 'Market and industry factors in stock price behavior', *Journal of Business* **39**(1), 139–190.

Kong, A., Rapach, D. E., Strauss, J. K. and Zhou, G. (2011), 'Predicting market components out of sample: asset allocation implications', *Journal of Portfolio Management* **37**(4), 29–41.

Lee, C. M. C. and Swaminathan, B. (1999), 'Valuing the Dow: A bottom-up approach', *Financial Analysts Journal* **55**(5), 4–23.

Lee, C. M., Myers, J. and Swaminathan, B. (1999), 'What is the intrinsic value of the Dow?', *The Journal of Finance* **54**(5), 1693–1741.

Li, H. and Tang, Q. (2019), 'Analyzing mortality bond indexes via hierarchical forecast reconciliation', *ASTIN Bulletin* **49**(3), 823–846.

Lila, M. F., Meira, E. and Oliveira, F. L. C. (2022), 'Forecasting unemployment in Brazil: A robust reconciliation approach using hierarchical data', *Socio-Economic Planning Sciences* **82**, 101298.

Livingston, M. (1977), 'Industry movements of common stocks', *Journal of Finance* **32**(3), 861–874.

Lo, A. W. (1991), 'Long-term memory in stock market prices', *Econometrica* pp. 1279–1313.

Maharaj, E. A., D'Urso, P. and Caiado, J. (2019), *Time series clustering and classification*, Chapman and Hall/CRC.

Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2022), 'M5 accuracy competition: Results, findings, and conclusions', *International Journal of Forecasting* **38**(4), 1346–1364.

Mantegna, R. N. (1999), 'Hierarchical structure in financial markets', *European Physical Journal B-Condensed Matter and Complex Systems* **11**(1), 193–197.

Marquering, W. and Verbeek, M. (2004), 'The economic value of predicting stock index returns and volatility', *Journal of Financial and Quantitative Analysis* **39**(2), 407–429.

Memmel, C. (2003), 'Performance hypothesis testing with the sharpe ratio', *Finance Letters* **1**, 21–23.

Merton, R. C. (1981), 'On market timing and investment performance. I. an equilibrium theory of value for market forecasts', *Journal of Business* pp. 363–406.

Panagiotelis, A., Athanasopoulos, G., Gamakumara, P. and Hyndman, R. J. (2021), 'Forecast reconciliation: A geometric view with new insights on bias correction', *International Journal of Forecasting* **37**(1), 343–359.

Panagiotelis, A., Gamakumara, P., Athanasopoulos, G. and Hyndman, R. J. (2023), 'Probabilistic forecast reconciliation: Properties, evaluation and score optimisation', *European Journal of Operational Research* **306**(2), 693–706.

Pesaran, M. H. and Timmermann, A. (2002), 'Market timing and return prediction under model instability', *Journal of Empirical Finance* **9**(5), 495–510.

Piccolo, D. (1990), 'A distance measure for classifying ARIMA models', *Journal of Time Series Analysis* **11**(2), 153–164.

Raffinot, T. (2017), 'Hierarchical clustering-based asset allocation', *Journal of Portfolio Management* **44**(2), 89–99.

Rapach, D. E., Strauss, J. K. and Zhou, G. (2010), 'Out-of-sample equity premium prediction: Combination forecasts and links to the real economy', *The Review of Financial Studies* **23**(2), 821–862.

Sharpe, W. F. (1963), 'A simplified model for portfolio analysis', *Management Science* **9**(2), 277–293.

Timmermann, A. and Granger, C. W. (2004), 'Efficient market hypothesis and forecasting', *International Journal of Forecasting* **20**(1), 15–27.

Tsiakas, I., Li, J. and Zhang, H. (2020), 'Equity premium prediction and the state of the economy', *Journal of Empirical Finance* **58**, 75–95.

Tumminello, M., Lillo, F. and Mantegna, R. N. (2010), 'Correlation, hierarchies, and networks in financial markets', *Journal of Economic Behavior & Organization* **75**(1), 40–58.

van Dijk, D. and Franses, P. H. (2003), 'Selecting a nonlinear time series model using weighted tests of equal forecast accuracy', *Oxford Bulletin of Economics and Statistics* **65**, 727–744.

Wang, J.-Z., Wang, J.-J., Zhang, Z.-G. and Guo, S.-P. (2011), 'Forecasting stock indices with back propagation neural network', *Expert Systems with Applications* **38**(11), 14346–14355.

Wickramasuriya, S. L., Athanasopoulos, G. and Hyndman, R. J. (2019), 'Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization', *Journal of the American Statistical Association* **114**(526), 804–819.

Wickramasuriya, S. L., Turlach, B. A. and Hyndman, R. J. (2020), 'Optimal non-negative forecast reconciliation', *Statistics and Computing* **30**(5), 1167–1182.

Yang, Y., Shang, H. L. and Raymer, J. (2022), 'Forecasting Australian fertility by age, region, and birthplace', *International Journal of Forecasting* .

Yu, P. and Yan, X. (2020), 'Stock price prediction based on deep neural networks', *Neural Computing and Applications* **32**, 1609–1628.

Zhang, J.-B., Gao, Y.-C. and Cai, S.-M. (2020), 'The hierarchical structure of stock market in times of global financial crisis', *Physica A: Statistical Mechanics and its Applications* **542**, 123452.

Appendix A. More predictive accuracy tests

Appendix A.1. Top level series: standard Diebold and Mariano (2002) test

Absolute Errors	Base	RW	BU	MinT	**Squared Errors**	Base	RW	BU	MinT
				$h = 1$ step ahead					
MinT: IND	**0.01**	0.89	0.04	0.47	MinT: IND	**0.01**	0.98	0.62	0.74
MinT: EXCH	**0.02**	0.38	**0.02**	0.12	MinT: EXCH	**0.01**	0.34	**0.07**	0.11
MinT: EUCL	**0.02**	0.97	0.41	0.82	MinT: EUCL	**0.02**	1.00	0.70	0.89
MinT: COR	**0.02**	0.99	0.21	0.89	MinT: COR	**0.01**	1.00	0.88	0.99
MinT: ARMA	**0.01**	**0.03**	**0.00**	**0.01**	MinT: ARMA	**0.01**	0.55	**0.01**	**0.04**
MinT: ALL	**0.02**	0.28	0.12	0.21	MinT: ALL	**0.02**	0.53	0.26	0.31
				$h = 3$ step ahead					
MinT: IND	**0.02**	0.57	0.66	0.72	MinT: IND	**0.02**	0.76	0.94	0.94
MinT: EXCH	**0.02**	0.43	0.29	0.49	MinT: EXCH	**0.02**	0.14	0.13	0.16
MinT: EUCL	**0.02**	0.44	0.42	0.59	MinT: EUCL	**0.03**	0.40	0.97	0.98
MinT: COR	**0.02**	0.55	0.75	0.85	MinT: COR	**0.02**	0.39	0.98	0.98
MinT: ARMA	**0.02**	0.86	0.74	0.79	MinT: ARMA	**0.02**	0.13	0.65	0.70
MinT: ALL	**0.02**	0.81	0.79	0.83	MinT: ALL	**0.04**	0.80	0.97	0.97
				$h = 6$ step ahead					
MinT: IND	**0.02**	**0.03**	0.16	0.24	MinT: IND	**0.02**	**0.05**	0.81	0.88
MinT: EXCH	**0.02**	**0.05**	0.16	0.17	MinT: EXCH	**0.02**	**0.03**	**0.04**	0.11
MinT: EUCL	**0.02**	**0.05**	0.64	0.83	MinT: EUCL	**0.02**	**0.04**	0.72	0.98
MinT: COR	**0.02**	**0.04**	0.18	0.20	MinT: COR	**0.02**	**0.04**	0.31	0.88
MinT: ARMA	**0.02**	**0.03**	0.48	0.59	MinT: ARMA	**0.02**	**0.02**	0.12	0.36
MinT: ALL	**0.02**	**0.02**	0.16	0.17	MinT: ALL	**0.02**	**0.07**	0.71	0.83
				$h = 12$ step ahead					
MinT: IND	**0.01**	**0.04**	**0.05**	**0.09**	MinT: IND	**0.01**	**0.01**	0.16	0.43
MinT: EXCH	**0.01**	0.12	0.31	0.87	MinT: EXCH	**0.01**	**0.03**	**0.07**	0.48
MinT: EUCL	**0.01**	**0.09**	**0.00**	**0.00**	MinT: EUCL	**0.01**	**0.03**	**0.01**	**0.00**
MinT: COR	**0.01**	**0.08**	**0.02**	**0.02**	MinT: COR	**0.01**	**0.02**	**0.01**	**0.00**
MinT: ARMA	**0.01**	**0.09**	0.36	0.49	MinT: ARMA	**0.01**	**0.02**	**0.10**	0.25
MinT: ALL	**0.01**	**0.02**	**0.02**	**0.02**	MinT: ALL	**0.01**	**0.01**	**0.10**	0.20

Table A.11: *P-values from the standard Diebold and Mariano (2002) test. Under the null, the difference in the forecasting errors is equal to zero. Entries in bold indicate a rejection of the null at a 10% level of significance.*

Appendix A.2. Bottom level series: modified van Dijk and Franses (2003) test

Absolute errors	Base	RW	MinT	Squared errors	Base	RW	MinT
			$h = 1$ step ahead				
MinT: IND	0.15	1.00	0.07	MinT: IND	0.47	1.00	0.53
MinT: EXCH	**0.08**	1.00	**0.09**	MinT: EXCH	**0.08**	0.99	**0.09**
MinT: EUCL	0.18	1.00	0.04	MinT: EUCL	**0.10**	1.00	**0.10**
MinT: COR	0.73	1.00	0.64	MinT: COR	0.86	1.00	0.91
MinT: ARMA	**0.00**	1.00	**0.00**	MinT: ARMA	**0.00**	1.00	**0.00**
MinT: ALL	0.02	0.88	0.02	MinT: ALL	**0.04**	1.00	**0.06**
			$h = 3$ step ahead				
MinT: IND	**0.08**	0.99	0.19	MinT: IND	0.35	1.00	0.57
MinT: EXCH	**0.04**	0.93	0.12	MinT: EXCH	**0.04**	0.89	**0.09**
MinT: EUCL	**0.04**	0.95	**0.10**	MinT: EUCL	**0.04**	0.97	0.12
MinT: COR	0.19	0.99	0.46	MinT: COR	0.21	1.00	0.49
MinT: ARMA	**0.03**	0.97	**0.07**	MinT: ARMA	**0.00**	0.93	**0.01**
MinT: ALL	**0.08**	0.72	0.11	MinT: ALL	0.17	0.92	0.25
			$h = 6$ step ahead				
MinT: IND	**0.01**	0.73	**0.01**	MinT: IND	**0.01**	0.67	**0.02**
MinT: EXCH	**0.02**	0.77	**0.05**	MinT: EXCH	**0.03**	0.69	**0.05**
MinT: EUCL	**0.07**	0.92	0.25	MinT: EUCL	**0.00**	0.82	**0.02**
MinT: COR	**0.05**	0.86	**0.10**	MinT: COR	**0.00**	0.73	**0.01**
MinT: ARMA	**0.01**	0.83	**0.01**	MinT: ARMA	**0.00**	0.63	**0.00**
MinT: ALL	**0.02**	0.44	**0.02**	MinT: ALL	**0.01**	0.46	**0.02**
			$h = 12$ step ahead				
MinT: IND	**0.07**	0.36	0.21	MinT: IND	**0.01**	0.35	**0.04**
MinT: EXCH	**0.01**	0.39	**0.04**	MinT: EXCH	**0.01**	0.41	**0.04**
MinT: EUCL	**0.00**	0.37	**0.01**	MinT: EUCL	**0.00**	0.38	**0.00**
MinT: COR	**0.04**	0.35	**0.08**	MinT: COR	**0.00**	0.34	**0.00**
MinT: ARMA	**0.00**	0.28	**0.00**	MinT: ARMA	**0.00**	0.29	**0.00**
MinT: ALL	**0.02**	0.17	**0.04**	MinT: ALL	**0.00**	0.24	**0.01**

Table A.12: *P-values from the modified Diebold and Mariano (2002) test of van Dijk and Franses (2003). Under the null, the difference in the forecasting errors is equal to zero. Entries in bold indicate a rejection of the null at a 10% level of significance.*