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

Raffaele Mattera[a], George Athanastopulos[b] and Rob Hyndman[b]

[a]Sapienza University of Rome, Italy
[b]Monash University, Australia

Outline

1. Introduction
2. Stock market data
3. Methodology
4. Forecasting experiment: main results
5. Investing with reconciled forecasts
6. Final remarks

Introduction

Background

- Forecasting stock index prices is very important in finance:
  1. Monitoring the state of the stock market (e.g. Dow Jones, EUROSTOXX, etc.);
  2. Studying stock market efficiency;
  3. Forecasting with factor models;
  4. Timing-based investment strategies;
  5. etc.
- Stock price forecasting is difficult because financial markets are complex and turbulent systems: volatility clustering, non-linearities, long memory and **hierarchical structures** [Barnett and Serletis, 2000, Lo, 1991, Mantegna, 1999, Tumminello et al., 2010];
- Taking these characteristics into account, we can virtually improve stock price predictability.

Stocks as hierarchical time series

- To outperform random walk forecasts, practitioners use additional variables, known as factors or predictive signals [e.g. see Green et al., 2013];
- In this paper we take a different perspective and propose a novel framework for forecasting stock indices and their constituents;
- We consider the stock market as hierarchical time series, because common stocks linearly aggregate [Panagiotelis et al., 2021] to their market . Examples:
  1. Dow Jones born as an equally weighted index;
  2. S&P500 is volume-weighted;
  3. There are S&P500 index variants based on equal weighting also commonly used by investors;
  4. But "market" could also be the equally weighted portfolio...
  5. etc.

Main contribution

In this paper we introduce two main novelties:

1. *We apply forecast reconciliation to the financial domain*

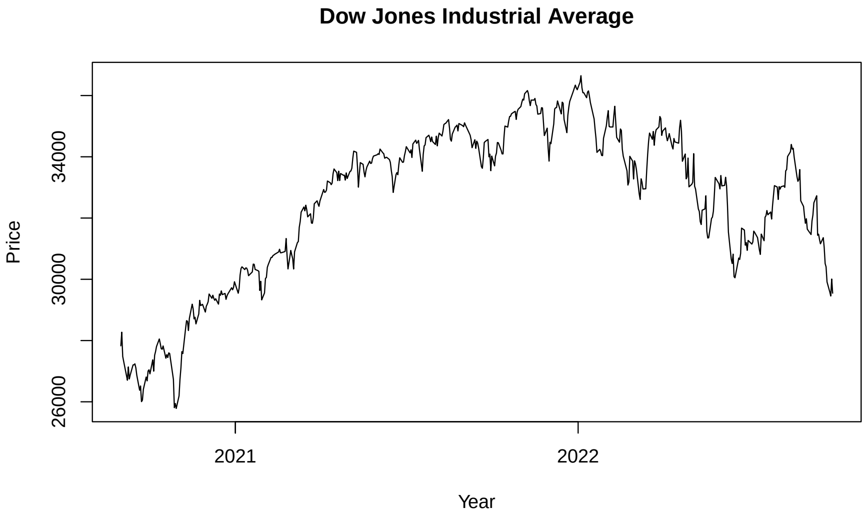
Some papers discussing hierarchical forecasting for stock price indexes are Lee and Swaminathan [1999], Darrough and Russell [2002], which use bottom-up and top-down approaches to forecast the Dow Jones. No studies use optimal reconciliation [except to Caporin et al., 2023, but for RV];

2. *We develop a novel forecasting framework that combines reconciliation and clustering*

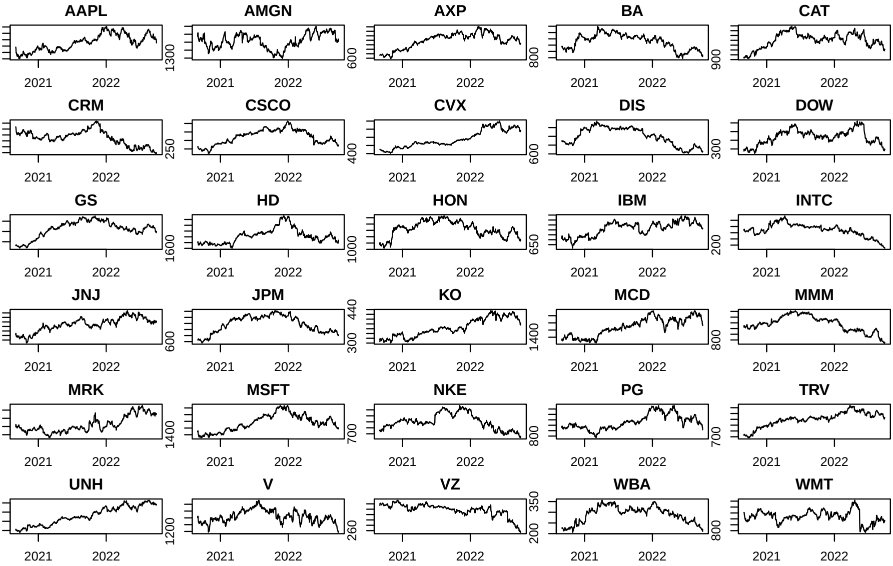
According to many authors [Mantegna, 1999, Brida and Risso, 2010, Wang et al., 2018], stocks are characterized by unknown hierarchical structures that need to be uncovered. We propose to estimate these with clustering and use them within a novel reconciliation framework.

Stock market data

Dow Jones Industrial Average Index (DJIA)



Common stocks included in the DJIA



Dow Jones as hierarchical time series

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

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

holds.

Data: some details more

- 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 changed on 31 August 2020 and has not been revised since;
- The divisor takes the constant value $d_t = 0.15$ during this period.

Methodology

Forecast reconciliation with clustering

- 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;
- We note that the aggregation matrix $\boldsymbol{A}$ is not known when we deal with stock prices;
- Let us denote with $\boldsymbol{\mathcal{C}}$ a grouping or clustering of stocks, that takes the form of a matrix of order $n_a \times G$ with G the number of clusters. Each element of $\boldsymbol{\mathcal{C}}$, called $c_{i,g}$ takes value of $c_{i,g} = 1$ is the ith stock belongs to the gth cluster and $c_{i,g} = 0$ otherwise.

Forecast reconciliation with clustering (cont'd)

- As such, we assume that stocks aggregate as follows

$$\boldsymbol{A} = \left[\begin{array}{c} \mathbf{1}' \\ \boldsymbol{\mathcal{C}}' \end{array} \right]. \tag{4}$$

- How do we determine $\boldsymbol{\mathcal{C}}$?
  1. Stock prices metadata (e.g. exchange, industry sector, etc.);
  2. Unsupervised learning.
- In the paper, we propose the use of Partition Around Medoids (PAM) clustering along with metadata-based clustering;
- The PAM [Kaufman and Rousseeuw, 1990], algorithm provides an iterative solution to the following minimization problem:

$$\text{min}: \quad \sum_{i=1}^{N}\sum_{g=1}^{G} d^2(\boldsymbol{b}_i, \boldsymbol{b}_g), \tag{5}$$

where $d^2(\boldsymbol{b}_i, \boldsymbol{b}_g)$ is the squared distance between the ith unit and the gth cluster centroid time series.

Time series clustering - observation-based

- Time series clustering can be divided into three well-known classes [Maharaj et al., 2019]: observation-based, feature-based and model-based;
- 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:

$$d_{\text{EUCL}}(\boldsymbol{b}_i, \boldsymbol{b}_j) = \sqrt{\sum_{t=1}^{T} (b_{it} - b_{jt})^2}. \tag{6}$$

  Observation-based approaches can be useful for clustering short time series, although they impose strong stationarity conditions on the original series.

Time series clustering - feature-based

- The feature-based approaches, on the other hand, aim to cluster time series with similar characteristics. A commonly employed approach quantifies the dissimilarity among different series on the basis of their correlation:

$$d_{\mathrm{COR}}(\boldsymbol{b}_i, \boldsymbol{b}_j) = \sqrt{2(1 - \rho_{i,j})}, \tag{7}$$

  with $\rho_{i,j}$ denoting the correlation coefficient between the i and j time series;
- This approach is common for clustering financial return time series [e.g. see Mantegna, 1999, Bonanno et al., 2001] as the correlation coefficient represents one of the most important features from the financial perspective [e.g. portfolio diversification, Raffinot, 2017].

Time series clustering - model-based

- 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.:

$$d_{\text{ARIMA}}(\boldsymbol{b}_i, \boldsymbol{b}_j) = \sqrt{\sum_{k=1}^{K} (\pi_{i,k} - \pi_{j,k})^2}, \tag{8}$$

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

Choosing the number of clusters

- The main drawback of the PAM clustering approach lies in the *a priori* selection of the number of clusters G;
- 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.

Combining different clustering structures

- For example, suppose we aggregate the prices for each of the n_1 $(g = 1, \dots, n_1; n_1 = G)$ exchanges on which they are traded. Let $c_{i,g} = 1$ if stock i is traded on exchange g, and 0 otherwise, and define $\boldsymbol{\mathcal{C}}_1$ to be the $N \times n_1$ matrix with element $c_{i,g}$ in row i and column g. 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 column corresponds to a different industry group;
- The combination of different clustering structure leads to the aggregation matrix

$$\boldsymbol{A} = \begin{bmatrix} \mathbf{1}' \\ \boldsymbol{\mathcal{C}}_1' \\ \vdots \\ \boldsymbol{\mathcal{C}}_L' \end{bmatrix}, \tag{9}$$

where each $\boldsymbol{\mathcal{C}}_\ell$ denotes a grouping or clustering of stocks.

Optimal forecast reconciliation

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

$$\boldsymbol{y}_t = \begin{bmatrix} \boldsymbol{a}_t \\ \boldsymbol{b}_t \end{bmatrix} = \boldsymbol{S}\boldsymbol{b}_t, \tag{10}$$

  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.

Optimal forecast reconciliation (cont'd)

- We denote coherent forecasts as $\tilde{\boldsymbol{y}}_h$. Linear reconciliation can be written as follows:

  $$\tilde{\boldsymbol{y}}_h = \boldsymbol{M}\hat{\boldsymbol{y}}_h, \tag{11}$$

  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

  $$\boldsymbol{G}_h = \left(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}, \tag{12}$$

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

Forecasting experiment: main results

Forecasting experiment: set-up

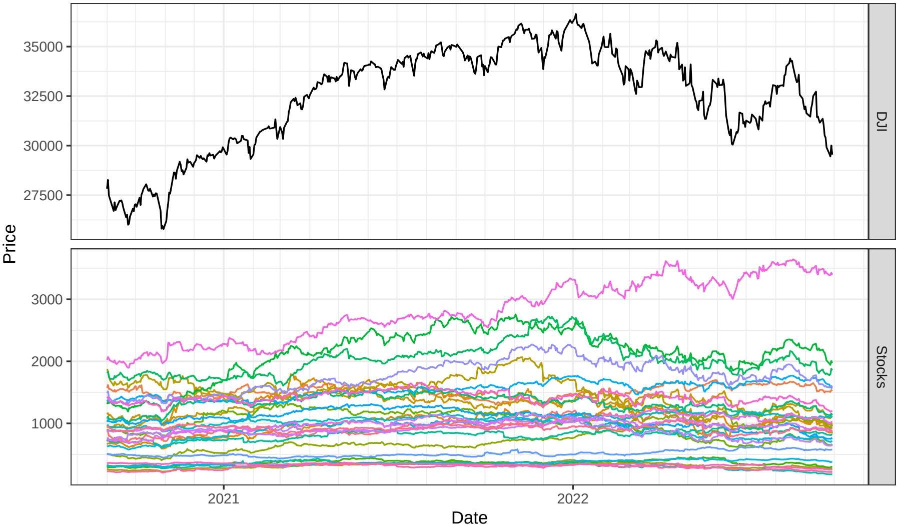
- The common stocks have length $T = 524$, and we leave the last $R = 124$ observations for out-of-sample testing;
- The clustering structures are estimated within the training set, i.e., considering only the first 400 observations;
- A rolling-window procedure is used to obtain the forecasts, where at each step of the recursion we choose the best ARIMA model by means of the automatic procedure described in Hyndman and Khandakar [2008];
- Forecasts at $h = \{1, 3, 6, 12\}$ steps ahead are produced, so the out-of-sample length is equal to $R - h$ $(r = 1, \dots, R - h)$;
- We evaluate if the different MinT reconciliation approaches improve with respect to base forecasts, random walk and bottom-up reconciliation.

Clustering structures in the DJIA

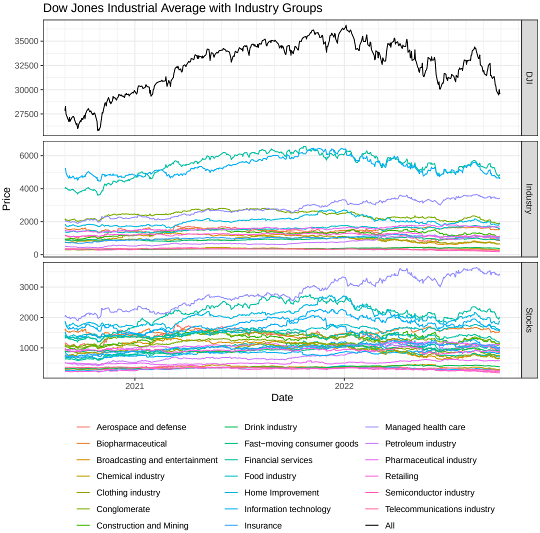
- A natural starting point in our setting is the use of the MinT forecast reconciliation approach which employs no clustering structure at all, i.e. $\boldsymbol{A} = \mathbf{1}'$;
- Then, we consider two clustering structures not involving PAM approach:
  1. Industry-based. The rationale behind this clustering approach is that stocks belonging to the same industry sector are affected by common shocks [e.g. King, 1966, Livingston, 1977];
  2. Exchange-based. This clustering approach involves $G = 2$ 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. Note that NASDAQ is known for its focus on IT companies, while NYSE lists a wider range of industries.

Single-level hierarchy structure without clustering structure

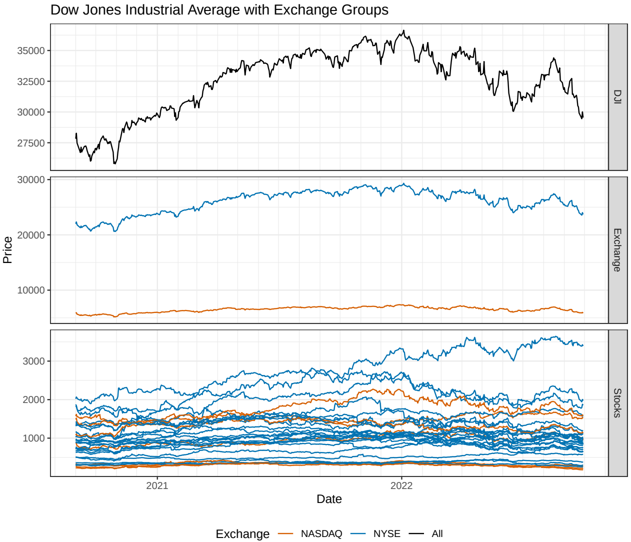
Dow Jones Industrial Average with Component Stocks



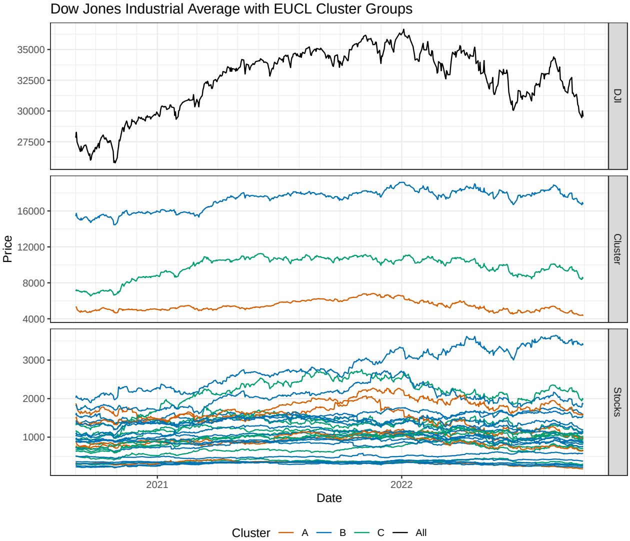
Two-level hierarchy structure with Industry-based clustering structure



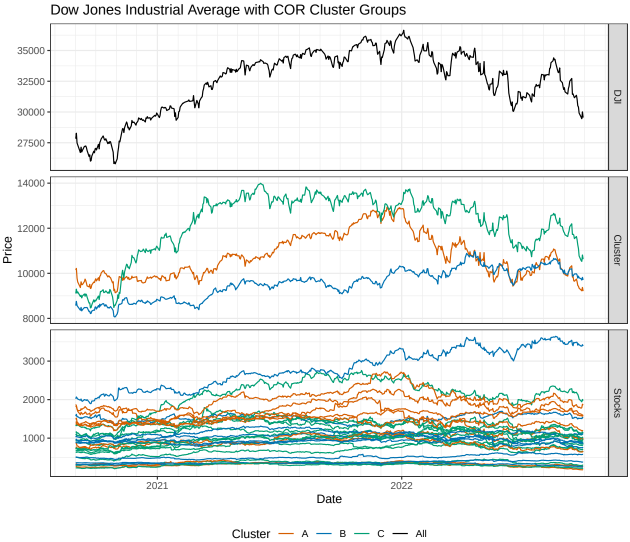
Two-level hierarchy structure with Exchange-based clustering structure



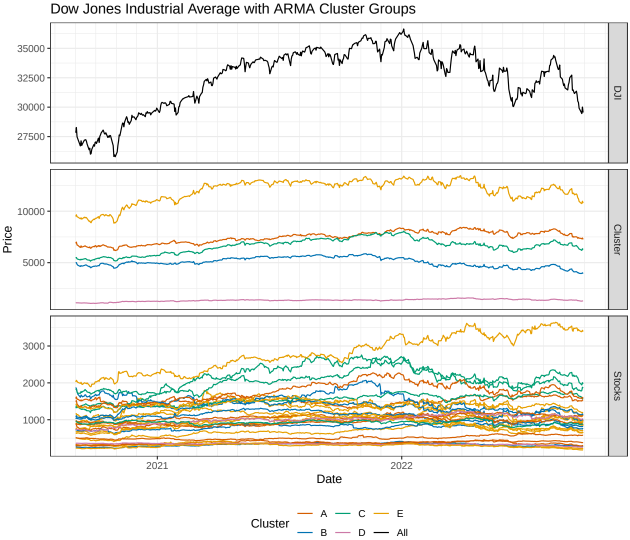
Clustering structures in the DJIA: hierarchical structure implied by the EUCL-based PAM clustering algorithm



Clustering structures in the DJIA: hierarchical structure implied by the COR-based PAM clustering algorithm



Clustering structures in the DJIA: hierarchical structure implied by the ARMA-based PAM clustering algorithm



Forecasting results: top series

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**

Table: Accuracy metric: MAE results. Best model with bold font.

Forecasting results: top series

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: Accuracy metric: RMSE results. Best model with bold font.

Forecasting results: top series (cont'd)

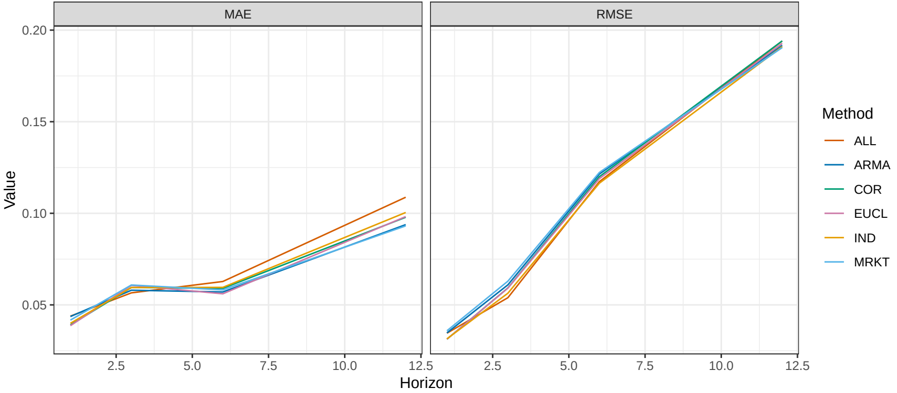


Figure: Average relative accuracy of reconciliation methods compared to the base unreconciled forecasts. Values above zero indicate higher (an improvement in) forecast accuracy. Left: Forecast accuracy is given by the log of the Mean Absolute Errors (MAE) of the reconciled forecasts relative to the base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

Forecasting results: bottom series

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

Table: Accuracy metric: MAE results. Best model with bold font.

Forecasting results: bottom series

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: Accuracy metric: RMSE results. Best model with bold font.

Investing with reconciled forecasts

Forecast-based investment strategy on market index

- 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,

$$\hat{x}_{t+h} = \frac{\hat{y}_{t+h} - y_t}{y_t}, \tag{13}$$

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

$$x_{t+h} = \frac{y_{t+h} - y_t}{y_t}. \tag{14}$$

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

$$r_{t+h} = \text{sign}(\hat{x}_{t+h}) x_{t+h}, \tag{15}$$

  where $\text{sign}(\hat{x}_{t+h})$ is a function taking the value of 1 if $\hat{x}_{t+h} \geq 0$ and -1 otherwise.

Evaluating the economic significance of forecasting

- 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]

$$\mathrm{SR}_A = \frac{\hat{\mu}_A}{\hat{\sigma}_A}, \tag{16}$$

  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;
- 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].

Reconciliation-based investing: results

$h = 1$	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: 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.

Final remarks

Final remarks

- In this paper we apply optimal forecast reconciliation to the financial domain;
- Considering that stocks aggregate into clusters, we propose a novel reconciliation framework that combines MinT and clustering;
- We show that more accurate forecasts can be achieved for both top (index) and bottom level (common stocks) series;
- We shed light on the economic significance of forecast reconciliation;
- ALL method seems to be recommendable due to good performance in terms of forecasting accuracy and larger performance in investment terms.

References I

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

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

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

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

G. Bonanno, F. Lillo, and R. Mantegna. High-frequency cross-correlation in a set of stocks. *Quantitative Finance*, 1(1):96–104, 2001.

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

References II

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

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

M. N. Darrough and T. Russell. A positive model of earnings forecasts: Top down versus bottom up. *Journal of Business*, 75(1):127–152, 2002.

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

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

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

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

References III

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

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

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

A. W. Lo. Long-term memory in stock market prices. *Econometrica*, pages 1279–1313, 1991.

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

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

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

References IV

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

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

T. Raffinot. Hierarchical clustering-based asset allocation. *Journal of Portfolio Management*, 44(2):89–99, 2017.

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

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

References V

G.-J. Wang, C. Xie, and H. E. Stanley. Correlation structure and evolution of world stock markets: Evidence from pearson and partial correlation-based networks. *Computational Economics*, 51:607–635, 2018.

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