

## Author response to reviewers

We would like to thank both the referees for their valuable comments and positive feedback on our paper. We have now thoroughly revised our paper to address your comments. We believe that this has resulted in a much-improved paper. We hope that you agree. Please find below our responses to your comments (in blue font).

### Author responses to comments by reviewer 1

1. Forecast combinations with non-negative weights, for example applied to disaggregate forecasts to compute an aggregate forecasts, share similar problems with forecast reconciliation. The major difference is probably that weights in forecast reconciliations do not sum to one as they do in forecast combinations. It will be worth to make this clear and **explain what this implies in terms of computational time**.

Forecast reconciliation has some similarities with forecast combinations, but with the following differences: (1) the combination weights are partially determined by the aggregation structure of the time series; (2) we only produce one forecast for each series; (3) the forecast of each series is a combination of the forecasts of all series; (4) the forecast weights are not restricted to add to 1 or to be non-negative. Typically, in forecast combination problems, the combination weights are free to be selected or estimated by the user, there are multiple forecasts for each series, the final forecast of each series is a combination only of forecasts for that series, and the forecast weights are normally restricted to be non-negative and to sum to one.

We have now added this paragraph at the end of Section 2.1.

While the problems share some similarities, the differences noted above make any direct comparisons of computational time largely meaningless.

2. The new method adds an important step to the MinT approach. However, forecast gains are almost negligible in simulation and empirical exercises. Even OLS produce accurate forecasts. Why is this the case? I will work out an exercise where optimal non-negative forecast reconciliation gives large gains otherwise the reader will be confused on the utility of the new method.

The purpose of non-negative forecast reconciliation is not necessarily to improve forecast accuracy but to align the forecasts with their use in practice. Many forecasts are of inherently non-negative variables such as sales volume or expenditure. In such cases, it is important to users that the forecasts are also non-negative. In cases where forecasts are provided at several levels of aggregation, it is also important that the forecasts are coherent. If forcing non-negative coherence on the forecasts improves accuracy, it is a side benefit, not the main purpose.

In any case, Table 7 does not reflect forecast gains through reconciliation. It only shows the percentage relative improvement of non-negative reconciled forecasts over the same reconciliation method without the non-negativity constraints. Hence, we cannot expect large gains in forecast accuracy to be seen in this table.

The forecasting performance results for each of these non-negative reconciliation approaches are given in Table 1. Each negative (positive) entry shows the percentage (decrease) increase in average RMSEs relative to the base forecasts. The top panel shows the results using the ARIMA base forecasts, while the bottom panel shows the results using the ETS base forecasts. It is clear immediately from both panels that minT(Shrink) is the best performing forecast reconciliation approach. The WLS<sub>r</sub> and OLS also perform well, generally showing improvements over the base forecasts.

This information has now been added to ???

Table 1: Out-of-sample forecast performances for Australian domestic tourism flows.

	$h = 1$	2	3	6	12	1-6	1-12	$h = 1$	2	3	6	12	1-6	1-12
ARIMA														
Australia								Australia by purpose of travel						
OLS	-3.5	-2.2	-2.7	-2.4	-2.9	-2.9	-2.8	-0.4	-1.4	-0.7	-1.8	-1.8	-0.9	-0.9
WLS <sub>v</sub>	-0.6	1.5	-0.1	0.2	-4.2	-0.6	-0.9	4.8	3.7	3.5	2.4	1.0	3.4	3.0
MinT	<b>-8.0</b>	<b>-5.9</b>	<b>-8.0</b>	-6.2	<b>-8.5</b>	<b>-7.5</b>	<b>-7.2</b>	<b>-5.0</b>	<b>-5.5</b>	<b>-6.4</b>	<b>-6.4</b>	<b>-6.2</b>	<b>-5.9</b>	<b>-5.6</b>
States								States by purpose of travel						
OLS	0.8	0.2	0.5	-0.6	1.2	0.3	-0.1	1.6	0.9	0.9	1.5	1.9	1.3	1.6
WLS <sub>v</sub>	-0.1	-0.3	0.2	-0.9	-1.1	-0.4	-0.8	-0.3	-1.3	-1.2	-0.4	-1.1	-0.9	-0.8
MinT	<b>-5.2</b>	<b>-5.6</b>	<b>-5.4</b>	<b>-5.6</b>	<b>-4.5</b>	<b>-5.4</b>	<b>-5.5</b>	<b>-4.4</b>	<b>-5.1</b>	<b>-5.4</b>	<b>-4.2</b>	<b>-4.1</b>	<b>-4.8</b>	<b>-4.5</b>
Zones								Zones by purpose of travel						
OLS	-3.8	-3.9	-3.3	-3.5	-2.4	-3.5	-3.2	0.8	0.6	0.6	0.5	1.2	0.7	0.7
WLS <sub>v</sub>	-6.4	-6.2	-5.5	-5.6	-5.6	-5.7	-5.6	-2.6	-2.7	-2.6	-2.5	-2.5	-2.6	-2.6
MinT	<b>-8.5</b>	<b>-8.7</b>	<b>-8.0</b>	<b>-7.7</b>	<b>-7.0</b>	<b>-8.0</b>	<b>-7.7</b>	<b>-4.4</b>	<b>-4.4</b>	<b>-4.5</b>	<b>-4.1</b>	<b>-3.7</b>	<b>-4.3</b>	<b>-4.2</b>
Regions								Regions by purpose of travel						
OLS	-2.8	-2.4	-2.5	-2.5	-1.8	-2.4	-2.2	-0.1	0.2	0.3	0.0	0.5	0.2	0.2
WLS <sub>v</sub>	-5.4	-5.0	-4.7	-4.5	-4.6	-4.8	-4.6	-3.4	-3.2	-3.0	-3.0	-2.8	-3.1	-3.1
MinT	<b>-7.2</b>	<b>-6.6</b>	<b>-6.5</b>	<b>-5.9</b>	<b>-5.8</b>	<b>-6.4</b>	<b>-6.2</b>	<b>-4.6</b>	<b>-4.3</b>	<b>-4.1</b>	<b>-3.9</b>	<b>-3.7</b>	<b>-4.2</b>	<b>-4.1</b>
ETS														
Australia								Australia by purpose of travel						
OLS	-2.0	-2.0	-2.5	-2.3	-2.9	-2.3	-2.3	0.9	0.6	1.7	1.4	0.8	1.4	1.1
WLS <sub>v</sub>	-3.8	-2.8	-3.6	<b>-3.4</b>	<b>-5.0</b>	-3.9	-3.3	-0.5	-0.1	0.8	0.8	-1.5	0.5	0.4
MinT	<b>-4.2</b>	<b>-3.3</b>	<b>-4.1</b>	<b>-3.4</b>	-4.9	<b>-4.1</b>	<b>-3.6</b>	<b>-1.0</b>	<b>-0.5</b>	<b>0.4</b>	<b>0.6</b>	<b>-1.8</b>	<b>0.2</b>	<b>0.0</b>
States								States by purpose of travel						
OLS	1.3	1.5	1.6	2.0	2.3	1.6	1.6	0.3	0.6	0.7	0.4	1.3	0.5	0.6
WLS <sub>v</sub>	-2.1	-0.9	-1.0	<b>-0.4</b>	-1.8	-1.3	-1.1	-2.1	-1.5	-1.4	-1.8	-2.1	-1.8	-1.8
MinT	<b>-2.3</b>	<b>-1.1</b>	<b>-1.2</b>	<b>-0.4</b>	<b>-1.9</b>	<b>-1.4</b>	<b>-1.3</b>	<b>-2.3</b>	<b>-1.6</b>	<b>-1.5</b>	<b>-1.9</b>	<b>-2.3</b>	<b>-1.9</b>	<b>-1.9</b>
Zones								Zones by purpose of travel						
OLS	0.8	0.3	0.6	0.8	1.5	0.7	0.7	0.9	0.7	0.6	0.8	1.4	0.8	0.8
WLS <sub>v</sub>	<b>-0.8</b>	<b>-1.0</b>	<b>-0.8</b>	-0.3	-0.5	-0.6	-0.6	-0.5	-0.5	-0.7	-0.6	-0.3	-0.6	-0.5
MinT	<b>-0.8</b>	<b>-1.0</b>	<b>-0.8</b>	<b>-0.4</b>	<b>-0.8</b>	<b>-0.7</b>	<b>-0.8</b>	<b>-0.6</b>	<b>-0.7</b>	<b>-0.8</b>	<b>-0.7</b>	<b>-0.6</b>	<b>-0.7</b>	<b>-0.8</b>
Regions								Regions by purpose of travel						
OLS	0.1	-0.3	-0.3	-0.4	0.0	-0.2	-0.2	0.5	0.5	0.5	0.5	0.9	0.5	0.5
WLS <sub>v</sub>	-1.0	<b>-1.2</b>	-1.2	<b>-1.2</b>	-1.5	-1.1	-1.2	-0.6	-0.5	-0.5	-0.6	-0.4	-0.5	-0.5
MinT	<b>-1.1</b>	<b>-1.2</b>	<b>-1.3</b>	<b>-1.2</b>	<b>-1.7</b>	<b>-1.2</b>	<b>-1.3</b>	<b>-0.8</b>	<b>-0.7</b>	<b>-0.7</b>	<b>-0.8</b>	<b>-0.7</b>	<b>-0.8</b>	<b>-0.8</b>

Notes: A negative (positive) entry shows the percentage decrease (increase) in average RMSEs relative to the base reconciled forecasts. Bold entries identify the best performing approaches.

## Author responses to comments by reviewer 2

1. The abstract is too long and does not get to the point in a concise way, please reformulate.

The sum of forecasts of disaggregated time series are often required to equal the forecast of the aggregate, giving a set of coherent forecasts. The least squares solution for finding coherent forecasts uses a reconciliation approach known as MinT, proposed by Wickramasuriya, Athanasopoulos, and Hyndman (2019). The MinT approach and its variants do not guarantee that the coherent forecasts are non-negative, even when all of the original forecasts are non-negative in nature. This has become a serious issue in applications that are inherently non-negative such as with sales data or tourism numbers. While overcoming this difficulty, we reconsider the least squares minimization problem with non-negativity constraints to ensure that the coherent forecasts are strictly non-negative.

The constrained quadratic programming problem is solved using three algorithms. They are the block principal pivoting (BPV) algorithm, projected conjugate gradient (PCG) algorithm, and scaled gradient projection (SGP) algorithm. A Monte Carlo simulation is performed to evaluate the computational performances of these algorithms as the number of time series increases. The results demonstrate that the BPV algorithm clearly outperforms the rest, and PCG is the second best. The superior performance of the BPV algorithm can be partially attributed to the alternative representation of the weight matrix in the MinT approach.

An empirical investigation is carried out to assess the impact of imposing non-negativity constraints on forecast reconciliation. It is observed that slight gains in forecast accuracy have occurred at the most disaggregated level. At the aggregated level slight losses are also observed. Although the gains or losses are negligible, the procedure plays an important role in decision and policy implementation processes.

2. The matrix  $\mathbf{C}$  after eq (2) is not defined.

Consider the 2-level structure given in Figure 1. In this example,  $m = 7, n = 4$ ,  $\mathbf{y}_t = [y_t, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]$  and

$$\mathbf{S} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

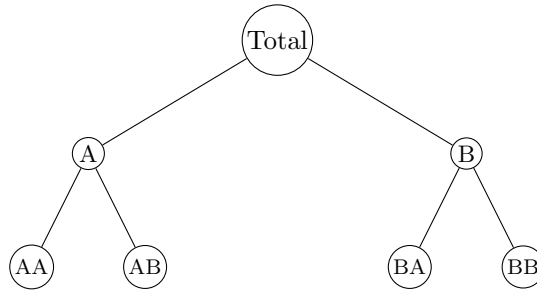


Figure 1: A two-level tree diagram.

We can include this on page 2, just after the first paragraph on Section 2.1.

3. Non-negative least squares is not new can you provide references since the beginning?

Yes, we agree that non-negative least squares is not a new field. We have mentioned on Page 3 that this area started with the work of Lawson and Hanson (1974) and have included a good reference (Chen and Plemmons (2009)) for a detailed review of this area. Therefore, we believe that a further review of non-negative least squares seems not relevant for this paper.

4. (a) Assuming that  $\mathbf{\Lambda}_h$  is positive definite does not seem so straightforward, that means that forecast errors are linearly independent, but if variables are highly collinear this might not happen. I think it all depends on how you compute  $\hat{\mathbf{y}}_t(h)$ . If forecast capture all comovements than I can believe the assumption. Some more explanation should be given.

Yes, the assumption of positive definiteness depends on how we compute  $\hat{\mathbf{y}}_t(h)$ . Unless we use very simple forecasting methods such as (seasonal) naive or linear trend models, it is very unlikely that the covariance matrix of the base forecast errors is positive semi-definite.

We can write a sentence on page 3 around line 41.

- (b) I see you compute univariate forecasts (via ARMA) but why not using multivariate methods in high-dimensions, e.g. factor models (see Stock and Watson, 2002), this would control better  $\mathbf{\Lambda}_h$ . There are many multivariate models, such as factor models, VARMA, multivariate state space models, etc. Theoretically, we can expect these models to perform better than univariate models. However, the number of unknown parameters to be estimated is huge for a large structure. The estimation can also be prohibitive if there are insufficient number of observations. Even with a reasonable number of observations, the computing algorithms can be extremely slow.

One of the main reasons for reconciliation is to avoid multivariate forecasting. We do univariate forecasts for each series, and then the relationships between the series are captured by estimating the covariance matrix of base forecast errors. It would largely defeat the purpose and simplicity of our approach to add multivariate forecasting.

5. “large size of the structures that typically arise if forecast reconciliation” what are the orders of magnitude? Your application has 555 series which can be handled by multivariate forecasting methods.

Please refer to the previous comment about multivariate forecasting models. One of the authors of this paper has applied reconciliation to around 1 million series in an industry project based in UK.

6. When you define  $\boldsymbol{\lambda}^*$  that is equal the derivative of  $q$  wrt  $\mathbf{b}$  while you are writing the opposite fix the notation.

Thanks for pointing this, now we have corrected it in the revised document.

7. Why in (9) you use the norm to define a quadratic loss? Keep the same notation as before.

We have corrected this in the revised document.

8. (a) Given the computational cost of running a constrained minimization with respect to the unconstrained case, what would happen if once we compute (1) we just throw away the negative forecasts and compute  $\tilde{\mathbf{y}}$  with the remaining ones?

If we throw away the negative reconciled forecasts from unconstrained minimization and compute  $\tilde{\mathbf{y}}$  with the remaining, then it will not provide any guarantee that the elements of  $\tilde{\mathbf{y}}$  are non-negative. Your suggestion is similar to the first iteration of the block principal pivoting algorithm if  $\mathbf{\Lambda}_h$  is diagonal. However, the block principal pivoting algorithm deletes and add variables until it reaches the convergence - similar to stepwise regression.

- (b) Are we really after all  $m$  forecasts? Can you provide more motivation?

Yes, we are interested in all  $m$  forecasts. Generating accurate forecasts for each of the 555 time series within the Australian tourism grouping structure is imperative for the planning purposes of National, State and Local tourism authorities. Ensuring that these forecasts are reconciled across all geographical disaggregation levels and purposes of travel lead to coherent decision making across the various levels of management.

- (c) And anyway what happens if you just impose the negative forecast to be zero. Can you repeat Table 7 in this case? Still the numbers there do not seem dramatically large.

If we set the negative reconciled forecasts to zero without performing any NNLS algorithms will lead to a set of incoherent forecasts.

9. What is an ETS model?

ETS is a class of state space models, and stands for Error, Trend, and Seasonality. We have now defined the acronym on page 9.

10. You say you log-transform data to compute forecasts and then you back-transform but doesn't this amount to taking an exponential and therefore you must get positive values? This part should be clarified.

Yes, the base forecasts are positive because of the exponentials. However, this back-transformation does not make the reconciled forecasts to be non-negative. Therefore, NNLS algorithms are important.